



**HAL**  
open science

# Computable Convergence Rates for Subgeometrically Ergodic Markov Chains

Randal Douc, Éric Moulines, Philippe Soulier

► **To cite this version:**

Randal Douc, Éric Moulines, Philippe Soulier. Computable Convergence Rates for Subgeometrically Ergodic Markov Chains. *Bernoulli*, 2007, 13 (3), 10.3150/07-BEJ5162. hal-00013769

**HAL Id: hal-00013769**

**<https://hal.science/hal-00013769v1>**

Submitted on 10 Nov 2005

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# COMPUTABLE CONVERGENCE RATES FOR SUBGEOMETRICALLY ERGODIC MARKOV CHAINS

RANDAL DOUC\*, ERIC MOULINES, AND PHILIPPE SOULIER

ABSTRACT. In this paper, we give quantitative bounds on the  $f$ -total variation distance from convergence of an Harris recurrent Markov chain on an arbitrary under drift and minorisation conditions implying ergodicity at a sub-geometric 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 on two examples from queueing theory and Markov Chain Monte Carlo.

AMS 2000 MSC 60J10

Stochastic monotonicity; rates of convergence; Markov chains

## 1. INTRODUCTION

Let  $P$  be a Markov transition kernel on a state space  $X$  equipped with a countably generated  $\sigma$ -field  $\mathcal{X}$ . For a control function  $f : X \rightarrow [1, \infty)$ , the  $f$ -total variation or  $f$ -norm of a signed measure  $\mu$  on  $\mathcal{X}$  is defined as

$$\|\mu\|_f := \sup_{|g| \leq f} |\mu(g)| .$$

When  $f \equiv 1$ , the  $f$ -norm is the total variation norm, which is denoted  $\|\mu\|_{TV}$ . We assume that  $P$  is aperiodic positive Harris recurrent with stationary distribution  $\pi$ . Our goal is to obtain *quantitative* bounds on convergence rates, *i.e.* rate of the form

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

where  $f$  is a control function  $f : X \rightarrow [1, \infty)$ ,  $\{r(n)\}_{n \geq 0}$  is a non-decreasing sequence, and  $g : X \rightarrow [0, \infty]$  is a function which can be computed explicitly. As emphasized in (Roberts and Rosenthal, 2004, section 3.5), quantitative bounds have a substantial history in Markov chain theory. Applications are numerous

---

*Date:* November 10, 2005 (draft version).

including convergence analysis of Markov Chain Monte Carlo (MCMC) methods, transient analysis of queueing systems or storage models, etc. With few exception however, these quantitative bounds were derived under conditions which imply geometric convergence, *i.e.*  $r(n) = \beta^n$ ,  $n \geq 1$ . (see for instance Meyn and Tweedie (1994), Rosenthal (1995), Roberts and Tweedie (1999), Roberts and Rosenthal (2004), and Baxendale (2005)).

In this paper, we study conditions under which (1.1) hold for sequences in the set  $\Lambda$  of subgeometric rate functions from Nummelin and Tuominen (1983), defined as the family of sequences  $\{r(n)\}_{n \geq 0}$  such that  $r(n)$  is non decreasing and  $\log r(n)/n \downarrow 0$  as  $n \rightarrow \infty$ . Without loss of generality, we assume that  $r(0) = 1$  whenever  $r \in \Lambda$ . These rates of convergence have been only scarcely considered in the literature. Let us briefly summarize the results available for convergence at subgeometric rate for general state-space chain. To our best knowledge, the first result for subgeometric sequence has been obtained by Nummelin and Tuominen (1983), who derive sufficient conditions for  $\|\xi P^n - \pi\|_{\text{TV}}$  to be of order  $o(r^{-1}(n))$ . The basic condition involved in this work is the ergodicity of order  $r$  (or  $r$ -ergodicity), defined as

$$\sup_{x \in B} \mathbb{E}_x \left[ \sum_{k=0}^{\tau_B-1} r(k) \right] < \infty. \quad (1.2)$$

where  $\tau_B \stackrel{\text{def}}{=} \inf\{n \geq 1, X_n \in B\}$  (with the convention that  $\inf \emptyset = \infty$ ) is the return time to some accessible some small set  $B$  (*i.e.*  $\pi(B) > 0$ ). These results were later extended by Tuominen and Tweedie (1994) to  $f$ -norm for general control functions  $f : \mathsf{X} \rightarrow [1, \infty)$  under  $(f, r)$ -ergodicity, which states that

$$\sup_{x \in B} \mathbb{E}_x \left[ \sum_{k=0}^{\tau_B-1} r(k) f(X_k) \right] < \infty \quad (1.3)$$

for some accessible small set  $B$ . These contributions do not provide computable expressions for the bounds in (1.1).

A direct route to quantitative bounds for subgeometric sequences has been opened by Veretennikov (1997, 1999), based on coupling techniques (see Gulinsky and Veretennikov (1993) and Rosenthal (1995) for the coupling construction of Harris recurrent Markov chains). This method consists in relating the bounds (1.1) to a moment of the *coupling time* through Lindvall's inequality

Lindvall (1979, 1992). Veretennikov (1997, 1999) 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  $g : \mathbb{R}^d \rightarrow \mathbb{R}^d$  is a Borel function and  $(W_n)_{n \geq 0}$  is an i.i.d. sequence, and provides expressions of the bounds in (1.1) with the total variation distance ( $f \equiv 1$ ) and polynomial rate functions  $r(n) = n^\beta$ ,  $n \geq 1$ . These results have later been extended, using similar techniques, to *truly* subgeometric sequence, *i.e.*  $\{r(n)\}_{n \geq 0} \in \Lambda$  satisfying  $\lim_{n \rightarrow \infty} r(n)n^{-\kappa} = \infty$  for any  $\kappa$ , in Klovov and Veretennikov (2004), for a more general class of functional autoregressive process.

Fort and Moulines (2003b) 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 (2000) later extended by Jarner and Roberts (2001). This method is based on a recursive computation of the polynomial moment of the coupling time (see (Fort and Moulines, 2003a, proposition 7)) which is related to the moments of the hitting time of a bivariate chain to a set where coupling might occur. This proof is tailored to the polynomial case and cannot be easily adapted to the general subgeometric case (see Fort (2001) for comments).

The objective of this paper is to generalize the results mentioned above in two directions. We consider Markov chains over general state space and we study general subgeometrical rates of convergence instead of polynomial rates Fort and Moulines (2003b). We establish a family of convergence bound (with a trade-off between the rate and the norm) extending to the subgeometrical case the computable bounds obtained in the geometrical case by Rosenthal (1995) and later refined by Roberts and Tweedie (1999) and Douc et al. (2004b) (see (Roberts and Rosenthal, 2004, Theorem 12) and the references therein). The method, based on coupling associated, provides a short and nearly self-contained proof of the results presented in Nummelin and Tuominen (1983) and Tuominen and Tweedie (1994): this allows for intuitive understanding of these results, while also avoiding various analytic technicalities of the previous proofs of these theorems.

The paper is organized as follows. In section 2, we present our assumptions and state our main results. In section 2.1, we specialize our result to stochastically

monotone Markov chains and derive bounds which extends results reported earlier by Scott and Tweedie (1996) and Roberts and Tweedie (2000). Examples from queueing theory and MCMC are discussed in section 3 to support our findings and illustrate the numerical computations of the bounds.

## 2. STATEMENTS 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 exist a set  $C \in \mathcal{X}$ , a constant  $\epsilon > 0$  and a probability measure  $\nu$  such that, for all  $x \in C$ ,  $P(x, \cdot) \geq \epsilon\nu(\cdot)$ .

For simplicity, only one-step minorisation is considered in this paper. Adaptations to  $m$ -step minorisation can be carried out as in Rosenthal (1995) (see also Fort (2001) and Fort and Moulines (2003b)).

Let  $\check{P}$  be a Markov transition kernel on  $\mathsf{X} \times \mathsf{X}$  such that, for all  $A \in \mathcal{X}$ ,

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

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

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

$$Q(x, A) = \begin{cases} (1 - \epsilon)^{-1} (P(x, A) - \epsilon\nu(A)) & 0 < \epsilon < 1 \\ \nu(A) & \epsilon = 1 \end{cases} \quad (2.3)$$

One may for example set

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

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 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 set:  $R(n) \stackrel{\text{def}}{=} \sum_{k=0}^{n-1} r(k)$ . Denote by  $\sigma_{C \times C} \stackrel{\text{def}}{=} \inf\{n \geq 0, (X_n, X'_n) \in C \times C\}$  the first hitting time of  $C \times C$  and let

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

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

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

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

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

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

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

We will establish that  $R$  is the maximal rate of convergence (that can be deduced from assumptions (A1)-(A3)) and that this rate is associated to convergence in 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  $f(x) + f(x') \leq V(x, x')$  for any  $(x, x') \in \mathbb{X} \times \mathbb{X}$ . Using an interpolation technique, we will derive rate of convergence  $1 \leq s \leq r$  associated to some  $g$ -norm,  $0 \leq g \leq f$ . To construct such interpolation, we consider pair of positive functions  $(\alpha, \beta)$  satisfying, for some  $0 \leq \rho \leq 1$ ,

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

Functions satisfying this condition can be obtained from Young's inequality. Let  $\psi$  be a real valued, continuous, strictly increasing function on  $\mathbb{R}^+$  such that  $\psi(0) = 0$ ; then for any  $(a, b) > 0$ ,

$$ab \leq \mathcal{P}(a) + \mathcal{F}(b), \quad \text{where } \mathcal{P}(a) \stackrel{\text{def}}{=} \int_0^a \psi(x) dx \quad \text{and} \quad \mathcal{F}(b) \stackrel{\text{def}}{=} \int_0^b \psi^{-1}(x) dx,$$

where  $\psi^{-1}$  is the inverse function of  $\psi$ . If we set  $\alpha(u) \stackrel{\text{def}}{=} \mathcal{P}^{-1}(\rho u)$  and  $\beta(v) = \mathcal{F}^{-1}((1 - \rho)v)$ , then the pair  $(\beta, \alpha)$  satisfies (2.9). Taking  $\psi(x) = x^{p-1}$  for some  $p \geq 1$  gives the special case  $\{(p\rho u)^{1/p}, (p(1 - \rho)v/(p - 1))^{(p-1)/p}\}$ .

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

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

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

$$\|P^n(x, \cdot) - P^n(x', \cdot)\|_{\text{TV}} \leq \frac{U(x, x') + M_U}{R(n) + M_U} \quad (2.11)$$

$$\|P^n(x, \cdot) - P^n(x', \cdot)\|_f \leq V(x, x') + M_V, \quad (2.12)$$

for any non-negative function  $f$  satisfying, for any  $(x, x') \in \mathbf{X} \times \mathbf{X}$ ,  $f(x) + f(x') \leq V(x, x') + M_V$ . Let  $(\alpha, \beta)$  be two positive functions satisfying (2.9) for some  $0 \leq \rho \leq 1$ . Then, for any  $(x, x') \in \mathbf{X} \times \mathbf{X}$  and  $n \geq 1$ , :

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

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

The proof is postponed to section 4.

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

*Remark 2.* When assumption (A2), then (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  $(x, x') \in \mathbf{X} \times \mathbf{X}$ ,

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

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

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

However, in particular when using drift functions, it is sometimes easier to apply theorem 2.1 with function a function  $v$  which does not coincide with  $v_r$ .

To check assumptions (A2) and (A3) it is often useful to use a drift conditions. Drift conditions implying convergence at polynomial rates have been recently proposed in Jarner and Roberts (2001). These conditions have later been extended to general subgeometrical rates by Douc et al. (2004a). Define by  $\mathcal{C}$  the set of functions

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

For  $\phi \in \mathcal{C}$ , define the function  $H_\phi : [1, \infty) \rightarrow [0, \infty)$  as  $H_\phi(v) \stackrel{\text{def}}{=} \int_1^v \frac{dx}{\phi(x)}$ . Since  $\phi$  is non decreasing,  $H_\phi$  is a non decreasing concave differentiable function on  $[1, \infty)$  and  $\lim_{v \rightarrow \infty} H_\phi(v) = \infty$ . The inverse  $H_\phi^{-1} : [0, \infty) \rightarrow [1, \infty)$  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  $H_\phi$  is increasing and  $\phi'$  is decreasing,  $\phi \circ H_\phi^{-1}$  is log-concave, which implies that the sequence

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

belongs to the set of subgeometric sequences  $\Lambda$ . Consider the following assumption

- (A4) There exists a function  $W : \mathsf{X} \times \mathsf{X} \rightarrow [1, \infty)$ , a function  $\phi \in \mathcal{C}$  and a constant  $b$  such that  $\check{P}W(x, x') \leq W(x, x') - \phi \circ W(x, x')$  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. (2004a) that under (A4), (A2) and (A3) are satisfied with the rate sequence  $r_\phi$  and the control function  $v = \phi \circ W$ . In addition, it is possible to deduce explicit bounds for the constants  $B_U, b_U, B_V$  and  $b_V$  from the constants appearing in the drift condition.



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

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

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

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

$$b_V \leq \sup_{C \times C} \phi \circ W + \sup_{C \times C} \check{P}W. \quad (2.19)$$

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 establish directly 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, 2004, Proposition 11) for a similar construction for geometrically ergodic Markov chain). Consider the following assumption:

(A5) There exists a function  $W_0 : \mathsf{X} \times \mathsf{X} \rightarrow [1, \infty)$ , 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*

$$W(x, x') = W_0(x) + W_0(x') - 1 \quad (2.20)$$

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

$$\sup_{C \times C} \check{P}W \leq 2(1 - \epsilon)^{-1} \left\{ \sup_C PW_0 - \epsilon \nu(W_0) \right\} - 1. \quad (2.22)$$

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

The proof is postponed to the appendix.

*Remark 3.* Since the function  $\phi_0$  is non-decreasing and  $\lim_{v \rightarrow \infty} \phi_0(v) = \infty$ , one may always find  $d$  such that the condition  $\phi_0(1) + \phi_0(d) \geq b_0(1 - \alpha)^{-1} + 2$  is fulfilled. The assumptions of the theorem above are satisfied provided that the associated level set  $\{V_0 \leq d\}$  is small. This will happen of course if all the level sets are 1-small, which may appear to be a rather strong requirement. More

realistic conditions may be obtained by using small sets associated to the iterate  $P^m$  of the kernel (see *e.g.* Rosenthal (1995), Fort (2001) and Fort and Moulines (2003b)).

**2.1. Stochastically ordered chains.** In this section, we show how to define the kernel  $\check{P}$  and obtain a drift condition for stochastically ordered Markov chain. Let  $\mathsf{X}$  be a totally ordered set, and denote  $\preceq$  the order relation. For  $a \in \mathsf{X}$ , denote  $(-\infty, a] = \{x \in \mathsf{X} : x \preceq a\}$  and  $[a, +\infty) = \{x \in \mathsf{X} : a \preceq x\}$ . A transition kernel  $P$  on  $\mathsf{X}$  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 network, Markov Monte-Carlo methods, storage models, etc. Stochastically ordered Markov chains have been considered in Lund and Tweedie (1996), Lund et al. (1996), Scott and Tweedie (1996) and Roberts and Tweedie (2000). In the first two papers, it is assumed that there exists an atom at the bottom of the state space. Lund et al. (1996) cover only geometric convergence; subgeometric rate of convergence are considered in Scott and Tweedie (1996). Roberts and Tweedie (2000) covers the case where the bottom of the space is a small set but restrict their attentions to conditions implying 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*, *i.e.* 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  $\mathsf{X}$  denote by  $G_K^-(x, u)$  the quantile function associated to 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

$$\begin{aligned} \mathbb{1}_{(C \times C)^c(x, x')} \check{P}(x, x'; A) &= \int_0^1 \mathbb{1}_A(G_P^-(x, u), G_P^-(x', u)) du \\ &\quad + \mathbb{1}_{C \times C}(x, x') \int_0^1 \mathbb{1}_A(G_Q^-(x, u), G_Q^-(x', u)) du , \end{aligned}$$

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  $v_0 : \mathsf{X} \rightarrow [1, \infty)$  be a measurable function and define:

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

Consider the following assumptions:

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

(B3) 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)-(B3) holds 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 obvious and omitted for brevity. As mentioned above, drift conditions often provide an easy path to prove conditions such as (B2) and (B3). Consider the following assumption:

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

Using, as above Douc et al. (2004a), it may be shown that this assumption implies (B2) and (B3) and allows to compute explicitly the constants.

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

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

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

$$b_{U_0} \leq 1 + \frac{r_\phi(1)}{\phi(1)} \left( (1 - \epsilon)^{-1} \left\{ \sup_C PW_0 - \epsilon \nu(W_0) \right\} - 1 \right) \quad (2.27)$$

$$b_{V_0} \leq \sup_C \phi \circ W_0 + (1 - \epsilon)^{-1} \left\{ \sup_C PW_0 - \epsilon \nu(W_0) \right\}. \quad (2.28)$$

The proof is entirely similar to Proposition 2.2 and is omitted.

3. APPLICATIONS

3.1. **the embedded M/G/1 queue.** In a M/G/1 queue, customers arrive into a service operation according to a Poisson process with parameter  $\lambda$ . Customers bring jobs requiring a service times which are independent of each others 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

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

where for each  $j \geq 0$ ,  $a_j \stackrel{\text{def}}{=} \int_0^\infty \{e^{-\lambda t} (\lambda t)^j / j!\} dB(t)$  (see (Meyn and Tweedie, 1993, 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  $u > 0$ ,  $m_u \stackrel{\text{def}}{=} \int t^u dB(t)$ . Applying the results derived above, we will compute explicit bounds (depending on  $\lambda$ ,  $x$  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  $x = 1, 2, \dots$ ,  $\tau_{x-1} \leq \tau_0$ ,  $\mathbb{P}_x$ -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  $\tau_{x-1}$  is the first return time of the state  $x - 1$ . For all  $x = 1, 2, \dots$ , 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

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

By differentiating the previous relation with respect to  $s$  and taking the limit as  $s \rightarrow 1$ , the previous relation implies that:  $\mathbb{E}_1[\tau_0] = (1 - \rho)^{-1}$ . 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  $\alpha$  and  $\beta$  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  $\pi$  (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  $\epsilon \stackrel{\text{def}}{=} \sum_{y=x_0-1}^{\infty} a_y$  and  $\nu(y) = \epsilon^{-1} a_y \mathbb{1}_{\{y \geq x_0-1\}}$ . Taking again  $r(k) \equiv 1$  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 Theorems 2.4, we finally compute a bound for  $b_{U_0} = \sup_C Q U_0 = (1 - \epsilon)^{-1} [\sup_C P U_0 - \epsilon \nu(U_0)]$ , which can be obtained by combining a bound for  $\sup_C P U_0$  and the expression of  $\nu(U_0)$ . An expression  $\nu(U_0)$  is computed by a direct application of the definitions. The bound for  $\sup_C P U_0$  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

$$\begin{aligned} P U_0(x) &= \mathbb{E}_x[\tau_C] = 1 + \mathbb{E}_x [\mathbb{E}_{X_1}[\tau_C] \mathbb{1}_{\{\tau_C > 1\}}] = 1 + \mathbb{E}_x [\mathbb{E}_{X_1}[\tau_{x_0}] \mathbb{1}_{\{X_1 \notin C\}}] \\ &= 1 + (1 - \rho)^{-1} \sum_{y=x_0+1}^{\infty} (y - x_0) P(x, y) \leq 1 + (1 - \rho)^{-1} \sum_{y=x_0+1}^{\infty} (y - x_0) a_{y-x_0+1}. \end{aligned}$$

We provide some numerical illustrations of the bounds described above. We use the distribution of service time suggested by in Roughan et al. (1998) given by

$$b(x) = \begin{cases} \alpha B^{-1} e^{-\frac{\alpha}{B} x} & x \leq B \\ \alpha B^\alpha e^{-\alpha x} x^{-\alpha+1} & x > B \end{cases} \quad (3.2)$$

where  $B$  marks where the tail begins. The mean of the service distribution is  $m_1 = B \{1 + e^{-\alpha} / (\alpha - 1)\} / \alpha$  and its Laplace transform,  $G(s) = \int_0^\infty e^{-st} dB(t)$ ,

$s \in \mathbb{C}$ ,  $\text{Re}(s) \geq 0$ , is given by

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

where  $\Gamma(x, z)$  is the incomplete  $\Gamma$  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 figures 1 and 2, we display the convergence bound  $\|P^n(x, \cdot) - \pi\|_{\text{TV}}$  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 better uniformly in the iteration index  $n$ . There is a trade off between the number of visits to the small set where coupling might 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 the convergence bound from a larger coupling set improves the bound (this effect is even more noticeable for a critically loaded system).

Insert figures 1 and 2 approximately here

**3.2. The Independence Sampler.** This second example is borrowed from Jarner and Roberts (2001). It is an example of a Markov chain which is stochastically monotone w.r.t a non-standard ordering of the state and 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  $\pi$  (with respect to some  $\sigma$ -finite measure  $\mu$  on  $\mathbf{X}$ ), 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  $\mu$ . The move is accepted with probability  $a(x, y) \stackrel{\text{def}}{=} \frac{q(x) \pi(y)}{\pi(x) q(y)} \wedge 1$ . The transition kernel of the algorithm is thus given by

$$P(x, A) = \int_A a(x, y)q(y) \mu(dy) + \mathbb{1}_A(x) \int_{\mathbf{X}} (1 - a(x, y))q(y) \mu(dy), \quad x \in \mathbf{X}, A \in \mathcal{X}.$$

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  $x \in \mathbf{X}$  and that  $q > 0$   $\pi$ -a.s.. For all  $\eta > 0$ , define the set

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

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

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

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

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

- (1) *the function  $\phi(v) = vK^{-1}(v)$  is differentiable, increasing and concave on  $[1, \infty)$ ,  $\lim_{v \rightarrow \infty} \phi(v) = \infty$ , and  $\lim_{v \rightarrow \infty} \phi'(v) = 0$ .*
- (2)  *$\int_0^{+\infty} uK(u)d\psi(u) < \infty$ , where for  $\eta > 0$ ,  $\psi(\eta) \stackrel{\text{def}}{=} 1 - \pi(C_\eta)$ .*

*Then, for any  $\eta^*$  satisfying*

$$\{1 - \psi(\eta^*)\} \phi(1) > \int_0^\infty (u \wedge \eta^*) K(u) d\psi(u)$$

*assumption (B4) is satisfied with  $W_0 = K \circ (q/\pi)$ ,  $C = C_{\eta^*}$  and*

$$\phi_0(v) = \{1 - \psi(\eta^*)\} \phi(v) - \int_0^\infty (u \wedge \eta^*) K(u) d\psi(u).$$

*In addition,*

$$\sup_{x \in C_{\eta^*}} PW_0 \leq \int_0^{+\infty} uK(u)d\psi(u) + K(\eta^*).$$

To illustrate our results, we evaluate the convergence bounds in the case where the target density  $\pi$  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 mean to derive a drift condition of the form  $PW_0 \leq W_0 - \phi \circ W_0$  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  $\psi$  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

$\int uK(u)d\psi(u) = \frac{(r+1)^{-\alpha}}{r(-\alpha+1/r+1)}$  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 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 figures 3 and 4, 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  $\eta^* = 0.25$  for  $r = 2$  and  $\alpha = 1.5$  and  $\eta^* = 0.5$  for  $r = 1/2$ . When ( $r = 2, \alpha = 1.1$ ) the convergence to stationarity is quite slow, which is not surprising since the instrumental density does not match well the target density at zero  $x = 0$ : according to our computable bounds, 500 iterations are required to get 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

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

For any probability measure  $(x, x') \in \mathsf{X} \times \mathsf{X}$ , denote  $\bar{\mathbb{P}}_{x, x'}$  and  $\bar{\mathbb{E}}_{x, x'}$  the probability measure and the expectation on associated to the Markov chain  $\{(X_n, X'_n, d_n)\}_{n \geq 0}$  with transition kernel  $\bar{P}$  starting from  $(X_0, X'_0, 0) = (x, x', 0)$ . In words, the coupling construction proceeds as follows. If  $d_n = 0$  and  $(X_n, X'_n) \notin C \times C$ , we draw  $(X_{n+1}, X'_{n+1})$  according to  $\check{P}(x, x', \cdot)$  and set  $d_{n+1} = 0$ . If  $d_n = 0$  and  $(X_n, X'_n) \in C \times C$ , we draw a coin with probability of heads  $\epsilon$ . If the coin comes up head, then we draw  $X_{n+1}$  from  $\nu$  and set  $X'_{n+1} = X_{n+1}$  and  $d_{n+1} = 1$  (the coupling is said to be successful); if the coin comes up tails, then we draw



$(X_{n+1}, X'_{n+1})$  from  $\check{P}(X_n, X'_n, \cdot)$  and we set  $d_{n+1} = 1$ . Finally, if  $d_n = 1$ , we draw  $X_{n+1}$  from  $P(X_n, \cdot)$  and set  $X_{n+1} = X'_{n+1}$ .

By construction, for any  $n$ ,  $(x, x') \in \mathsf{X} \times \mathsf{X}$  and  $(A, A') \in \mathcal{X} \times \mathcal{X}$ ,

$$\begin{aligned} \bar{\mathbb{P}}_{x,x',0}(Z_n \in A \times \mathsf{X} \times \{0, 1\}) &= \bar{\mathbb{P}}_{x,x',0}(X_n \in A) = P^n(x, A) \quad \text{and} \\ \bar{\mathbb{P}}_{x,x',0}(Z_n \in \mathsf{X} \times A' \times \{0, 1\}) &= \bar{\mathbb{P}}_{x,x',0}(X'_n \in A') = P^n(x', A'). \end{aligned}$$

By (Douc et al., 2004b, 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 process  $(\chi_k)_{k \geq 0}$  and  $(x, x') \in \mathsf{X} \times \mathsf{X}$ ,

$$\bar{\mathbb{E}}_{x,x',0}[\chi_n \mathbb{1}_{\{T > n\}}] = \check{\mathbb{E}}_{x,x'}[\chi_n (1 - \epsilon)^{N_n-1}] \quad , \quad (4.1)$$

where  $N_n$  is the number of visit to the set  $C \times C$  before time  $n$ ,

$$N_n = \sum_{j=0}^{\infty} \mathbb{1}_{\{\sigma_j \leq n\}} = \sum_{i=0}^n \mathbb{1}_{C \times C}(X_i, X'_i) \quad . \quad (4.2)$$

Let  $f : \mathsf{X} \rightarrow [0, \infty)$  and let  $g : \mathsf{X} \rightarrow \mathbb{R}$  be any Borel function such that  $\sup_{x \in \mathsf{X}} |g(x)|/f(x) < \infty$ . The classical coupling inequality (see e.g. (Thorisson, 2000, Chapter 2, section 3)) implies that

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

and (4.1) shows the following key coupling inequality:

$$\|P^n(x, \cdot) - P^n(x', \cdot)\|_f \leq \check{\mathbb{E}}_{x,x'}\{(f(X_n) + f(X'_n))(1 - \epsilon)^{N_n-1}\} \quad . \quad (4.3)$$

Because by definition  $\alpha(u)\beta(v) \leq \rho u + (1 - \rho)v$  for all  $(u, v) \in \mathbb{R}^+ \times \mathbb{R}^+$  and 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 (4.3) shows that

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

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

$$\check{P}U_{n+1} \leq U_n - r(n) + b_U r(n) \mathbb{1}_{C \times C}, \quad (4.4)$$

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

$$\begin{aligned} W_n^{(0)} &\stackrel{\text{def}}{=} U_n(X_n, X'_n) + \sum_{k=0}^{n-1} r(k) + M_U, \\ W_n^{(1)} &\stackrel{\text{def}}{=} V(X_n, X'_n) + \sum_{k=0}^{n-1} v(X_k, X'_k) + M_V. \end{aligned}$$

with the convention  $\sum_u^v = 0$  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,

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

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

$$T_n^{(0)} \stackrel{\text{def}}{=} \prod_{i=0}^{n-1} \frac{W_i^{(0)} + b_U r(i) \mathbb{1}_{C \times C}(X_i, X'_i)}{W_i^{(0)}} \quad \text{and} \quad T_n^{(1)} \stackrel{\text{def}}{=} \prod_{i=0}^{n-1} \frac{W_i^{(1)} + b_V \mathbb{1}_{C \times C}(X_i, X'_i)}{W_i^{(1)}}. \quad (4.5)$$

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

$$T_n^{(0)} = \prod_{i=0}^{N_{n-1}-1} \frac{W_{\sigma_i}^{(0)} + b_U r(\sigma_i)}{W_{\sigma_i}^{(0)}} \quad \text{and} \quad T_n^{(1)} = \prod_{i=0}^{N_{n-1}-1} \frac{W_{\sigma_i}^{(1)} + b_V}{W_{\sigma_i}^{(1)}} \quad (4.6)$$

where  $\sigma_i$  are the successive hitting time 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  $1 + b_U r(n) / \{R(n+1) + M_U\} \leq 1/(1 - \epsilon)$ , for  $N_{n-1} > 0$ , we have

$$T_n^{(0)} (1 - \epsilon)^{N_{n-1}} \leq \prod_{i=0}^{N_{n-1}-1} \left( \left\{ 1 + \frac{b_U r(\sigma_i)}{R(\sigma_i + 1) + M_U} \right\} (1 - \epsilon) \right) \leq 1. \quad (4.7)$$

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

$$\begin{aligned} \mathbb{E}_{x,x'}^* [W_n^{(0)}(1 - \epsilon)^{N_{n-1}}] &\leq \mathbb{E}_{x,x'}^* [W_n^{(0)}\{T_n^{(0)}\}^{-1}] , \\ \mathbb{E}_{x,x'}^* [W_n^{(1)}(1 - \epsilon)^{N_{n-1}}] &\leq \mathbb{E}_{x,x'}^* [W_n^{(1)}\{T_n^{(1)}\}^{-1}] . \end{aligned}$$

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

$$\begin{aligned} &\check{\mathbb{E}}_{x,x'} [W_n^{(0)}\{T_n^{(0)}\}^{-1} \mid \mathcal{F}_{n-1}] \\ &= \check{\mathbb{E}}_{x,x'} [W_n^{(0)} \mid \mathcal{F}_{n-1}] \frac{W_{n-1}^{(0)}}{W_{n-1}^{(0)} + b_U r(n-1) \mathbb{1}_{C \times C}(X_{n-1}, X'_{n-1})} \{T_{n-1}^{(0)}\}^{-1}, \quad (4.8) \end{aligned}$$

where  $\mathcal{F}_n = \sigma\{(X_0, X'_0), \dots, (X_n, X'_n)\}$ . Now, (4.4) yield:

$$\check{\mathbb{E}}_{x,x'} [W_n^{(0)} \mid \mathcal{F}_{n-1}] \leq W_{n-1}^{(0)} + b_U r(n-1) \mathbb{1}_{C \times C}(X_{n-1}, X'_{n-1}). \quad (4.9)$$

Combining (4.8) and (4.9) shows that  $\left\{W_n^{(0)}\{T_n^{(0)}\}^{-1}\right\}_{n \geq 0}$  is a  $\mathcal{F}$ -supermartingale.

Thus,

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

Similarly,  $\mathbb{E}_{x,x'}^* [W_n^{(1)}(1 - \epsilon)^{N_{n-1}}] \leq V(x, x') + M_V$ , 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, 1993) and (Douc et al., 2004a, Proposition 2.2), we obtain the following inequalities. Then, for all  $(x, x') \in \mathbf{X} \times \mathbf{X}$ ,

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

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

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

obtain:

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

showing (2.16). Similarly,

$$\begin{aligned} \check{\mathbb{E}}_{x,x'} \left[ \sum_{k=0}^{\sigma_{C \times C}} \phi \circ W(X_k, X'_k) \right] &= \phi \circ W(x, x') \mathbb{1}_{C \times C}(x, x') \\ + \check{\mathbb{E}}_{x,x'} \left[ \sum_{k=0}^{\tau_{C \times C} - 1} \phi \circ W(X_k, X'_k) \right] &\mathbb{1}_{(C \times C)^c}(x, x') + \check{\mathbb{E}}_{x,x'} [\phi \circ W(X_\tau, X'_\tau)] \mathbb{1}_{(C \times C)^c}(x, x') \end{aligned}$$

showing (2.17).  $\square$

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

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

where we have used the inequality: for any  $u \geq 1$  and  $v \geq 1$ ,  $\phi_0(u+v-1) - \phi_0(u) \leq \phi_0(v) - \phi_0(1)$ . For  $(x, x') \notin C$ ,  $b_0 \leq (1 - \lambda)\phi_0(d) \leq (1 - \lambda)\phi_0 \circ W_0(x, x')$  and the previous inequality implies  $\check{P}W(x, x') \leq W(x, x') - \phi_0 \circ W(x, x')$ .  $\square$

## APPENDIX A. PROOF OF PROPOSITION 3.1

Let  $W$  be any measurable non negative function on  $\mathsf{X}$ . Then, for  $\eta > 0$  and  $x \notin C_\eta$ ,

$$\begin{aligned} PW(x) - W(x) &= \int_{\mathsf{X}} a(x, y) \{W(y) - W(x)\} q(y) \mu(dy) \\ &\leq \int_{\mathsf{X}} \left( \eta \wedge \frac{q(y)}{\pi(y)} \right) W(y) \pi(y) \mu(dy) - W(x) \int_{\mathsf{X}} a(x, y) q(y) \mu(dy). \end{aligned}$$

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, we have:

$$\int_{\mathbf{X}} a(x, y)q(y)\mu(dy) \geq \int_{C_\eta} a(x, y)q(y)\mu(dy) = \frac{q(x)}{\pi(x)}\pi(C_\eta) = \frac{q(x)}{\pi(x)}(1 - \psi(\eta)).$$

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

$$PW(x) - W(x) \leq \int_{\mathbf{X}} \left( \eta \wedge \frac{q(y)}{\pi(y)} \right) W(y)\pi(y)\mu(dy) - \{1 - \psi(\eta)\} \frac{q(x)}{\pi(x)} W(x). \quad (\text{A.1})$$

Applying the definition of  $W_0$ , we now have:

$$\begin{aligned} & \int_{\mathbf{X}} \left( \eta \wedge \frac{q(y)}{\pi(y)} \right) W_0(y)\pi(y)\mu(dy) \\ &= \int_{\mathbf{X}} \left( \eta \wedge \frac{q(y)}{\pi(y)} \right) K \left( \frac{q(y)}{\pi(y)} \right) \pi(y)\mu(dy) = \int_0^\infty (\eta \wedge u)K(u)d\psi(u) < \infty. \end{aligned} \quad (\text{A.2})$$

By Lebesgue's bounded convergence theorem,  $\lim_{\eta \rightarrow 0} \int_0^\infty (\eta \wedge u)K(u)d\psi(u) = 0$ . Since moreover  $\lim_{\eta \rightarrow 0} \psi(\eta) = 0$ , hence, for  $\eta$  small enough,  $\{1 - \psi(\eta)\}\phi(M) > \int_0^\infty (\eta \wedge u)K(u)d\psi(u)$ , hence  $\eta^*$  is well defined. Now, (A.1) and (A.2) yield, for all  $x \notin C_{\eta^*}$ ,

$$\begin{aligned} PW_0(x) - W_0(x) &\leq \int_0^\infty (\eta^* \wedge u)K(u)d\psi(u) - (1 - \psi(\eta^*))W_0(x)K^{-1} \circ W_0(x) \\ &= -\phi_0(W_0(x)). \end{aligned}$$

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

$$\begin{aligned} PW_0(x) &\leq \int_{\mathbf{X}} q(y)W_0(y)\mu(dy) + W_0(x) \\ &= \int_{\mathbf{X}} \frac{q(y)}{\pi(y)} K \left( \frac{q(y)}{\pi(y)} \right) \pi(y)\mu(dy) + W_0(x) \leq \int_0^\infty uK(u)d\psi(u) + K(\eta^*). \end{aligned}$$

## REFERENCES

- BAXENDALE, P. H. (2005). Renewal theory and computable convergence rates for geometrically ergodic Markov chains. *Ann. Appl. Probab.* **15** 700–738.
- DOUC, R., FORT, G., MOULINES, E. and SOULIER, P. (2004a). Practical drift conditions for subgeometric rates of convergence. *Ann. Appl. Probab.* **14** 1353–1377.

- DOUC, R., MOULINES, E. and ROSENTHAL, J. (2004b). Quantitative bounds for geometric convergence rates of Markov chains. *Annals of Applied Probability* **14** 1643–1665.
- 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.
- FORT, G. and MOULINES, E. (2000).  $V$ -subgeometric ergodicity for a Hastings-Metropolis algorithm. *Statist. Probab. Lett.* **49** 401–410.
- FORT, G. and MOULINES, E. (2003a). Convergence of the Monte Carlo expectation maximization for curved exponential families. *Ann. Statist.* **31** 1220–1259.
- FORT, G. and MOULINES, E. (2003b). Polynomial ergodicity of Markov transition kernels. *Stochastic Processes and Their Applications* **103** 57–99.
- GULINSKY, O. V. and VERETENNIKOV, A. Y. (1993). *Large deviations for discrete-time processes with averaging*. VSP, Utrecht.
- JARNER, S. and ROBERTS, G. O. (2001). Polynomial convergence rates of Markov chains. *Annals of Applied Probability* **12** 224–247.
- KLOKOV, S. A. and VERETENNIKOV, A. Y. (2004). Sub-exponential mixing rate for a class of Markov chains. *Math. Commun.* **9** 9–26.
- LINDVALL, T. (1979). On coupling of discrete renewal sequences. *Z. Wahrsch. Verw. Gebiete* **48** 57–70.
- LINDVALL, T. (1992). *Lectures on the Coupling Method*. Wiley, New-York.
- LUND, R. B., MEYN, S. P. and TWEEDIE, R. (1996). Computable exponential convergence rates for stochastically ordered Markov processes. *Annals of Applied Probability* **6** 218–237.
- LUND, R. B. and TWEEDIE, R. L. (1996). Geometric convergence rates for stochastically ordered Markov chains. *Mathematics of Operation Research* **21** 182–194.
- MEYN, S. P. and TWEEDIE, R. L. (1993). *Markov Chains and Stochastic Stability*. Springer, London.
- MEYN, S. P. and TWEEDIE, R. L. (1994). Computable bounds for convergence rates of Markov chains. *Annals of Applied Probability* **4** 981–1011.
- NUMMELIN, E. and TUOMINEN, P. (1983). The rate of convergence in Orey's theorem for Harris recurrent Markov chains with applications to renewal theory.

- Stochastic Processes and Their Applications* **15** 295–311.
- ROBERTS, G. O. and ROSENTHAL, J. S. (2004). General state space Markov chains and MCMC algorithms. *Probab. Surv.* **1** 20–71.
- ROBERTS, G. O. and TWEEDIE, R. L. (1999). Bounds on regeneration times and convergence rates for Markov chains. *Stochastic Processes and Their Applications* **80** 211–229.
- ROBERTS, G. O. and TWEEDIE, R. L. (2000). Rates of convergence of stochastically monotone and continuous time Markov models. *Journal of Applied Probability* **37** 359–373.
- ROSENTHAL, J. S. (1995). Minorization conditions and convergence rates for Markov chain Monte Carlo. *J. Am. Statist. Assoc.* **90** 558–566.
- ROUGHAN, M., VEITCH, D. and RUMSEWICZ, M. (1998). Computing queue-length distributions for power-law queues. In *Proceedings. IEEE INFOCOM '98*, vol. 1. IEEE.
- SCOTT, D. J. and TWEEDIE, R. L. (1996). Explicit rates of convergence of stochastically ordered Markov chains. In *Athens Conference on Applied Probability and Time Series: Applied Probability in Honor of J. M. Gani*, vol. 114 of *Lecture Notes in Statistics*. Springer.
- THORISSON, H. (2000). *Coupling, Stationarity and Regeneration*. Probability and its Applications, Springer-Verlag, New-York.
- TUOMINEN, P. and TWEEDIE, R. (1994). Subgeometric rates of convergence of  $f$ -ergodic Markov Chains. *Advances in Applied Probability* **26** 775–798.
- VERETENNIKOV, A. (1997). On polynomial mixing bounds for stochastic differential equations. *Stochastic Process. Appl.* **70** 115–127.
- VERETENNIKOV, A. (1999). On polynomial mixing and the rate of convergence for stochastic differential and difference equations. *Theory of probability and its applications* 361–374.

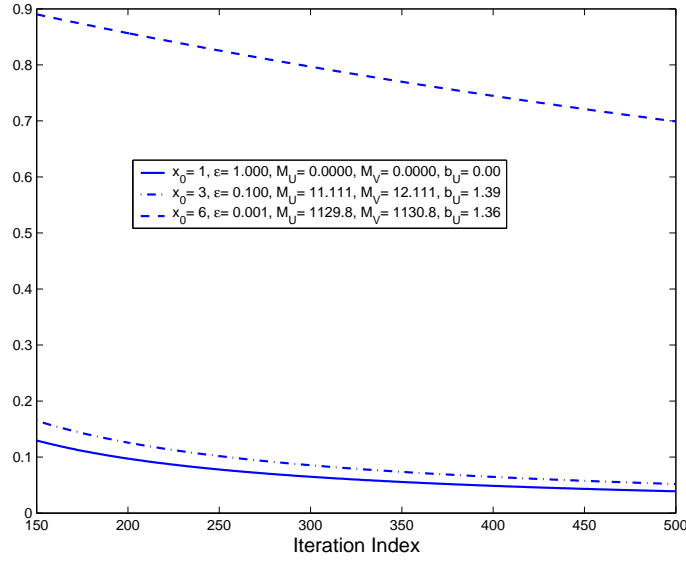


FIGURE 1. convergence bound for the total variation distance in the light-traffic case:  $\rho = 0.5, \alpha = 2.5$

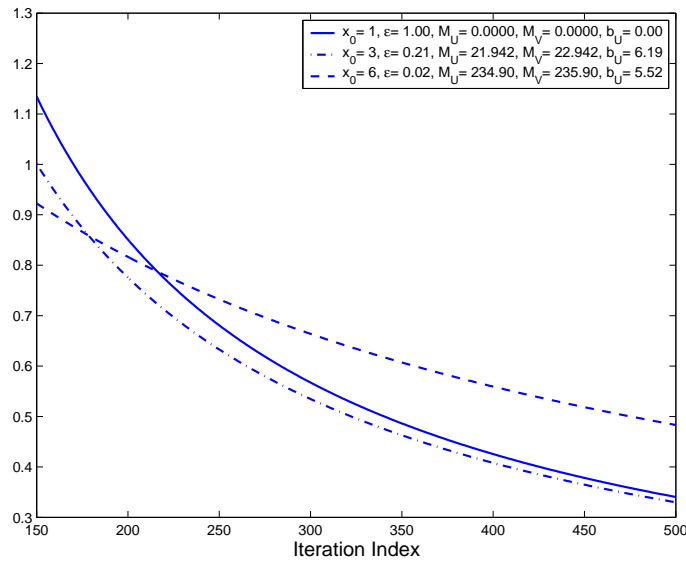


FIGURE 2. convergence bound for the total variation distance in the heavy traffic case:  $\rho = 0.9, \alpha = 2.5$



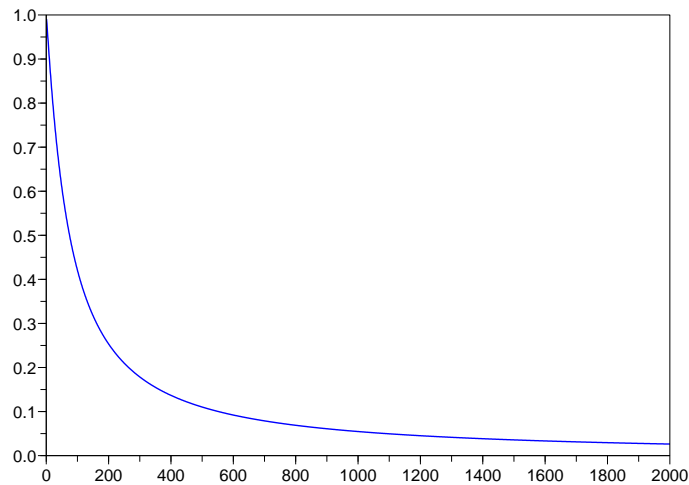


FIGURE 3. convergence bound for the total variation distance for the independence sampler with  $q(x) = 3x^2$

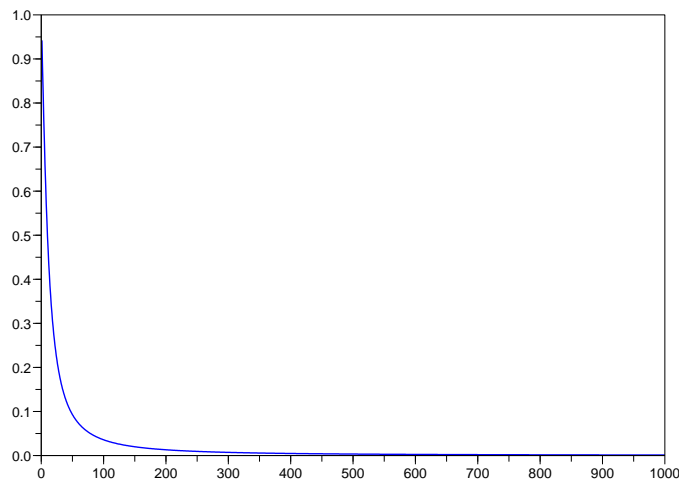


FIGURE 4. convergence bound for the total variation distance when  $q(x) = 1.5\sqrt{x}$

RANDAL DOUC (CORRESPONDING AUTHOR), CMAP, ECOLE POLYTECHNIQUE, 91128 PALAISEAU CEDEX, FRANCE

*E-mail address:* douc@cmapx.polytechnique.fr

ERIC MOULINES, ECOLE NATIONALE SUPÉRIEURE DES TÉLÉCOMMUNICATIONS, DÉPARTEMENT DE TRAITEMENT DU SIGNAL ET DES IMAGES, 46, RUE BARRAULT, 75634 PARIS CÉDEX 13, FRANCE

*E-mail address:* moulines@tsi.enst.fr

PHILIPPE SOULIER, DEPARTMENT OF MATHEMATICS OF UNIVERSITY PARIS X, BÂTIMENT G, BUREAU E15, 200 AVENUE DE LA RÉPUBLIQUE, 92000 NANTERRE CEDEX, FRANCE