Convergence Rates of Attractive-Repulsive MCMC Algorithms

by (in alphabetical order)

Yu Hang Jiang, Tong Liu, Zhiya Lou, Jeffrey S. Rosenthal,

Shanshan Shangguan, Fei Wang, and Zixuan Wu

Department of Statistical Sciences, University of Toronto

December, 2020

Abstract: We introduce a class of MCMC algorithms which include both attractive and repulsive forces, making their convergence analysis challenging. We prove that a version of these algorithms on a bounded state space is uniformly ergodic, while a version on an unbounded state space is still geometrically ergodic. We then use the method of shift-coupling to obtain explicit quantitative upper bounds on the convergence rate of the unbounded state space version.

1 Introduction

Markov Chain Monte Carlo (MCMC) algorithms are an indispensable tool for researchers and scientists across a wide spectrum of fields, ranging from machine learning and Bayesian inference to systems biology and mathematical finance, to sample from complicated distributions in high dimensions (see e.g. [8],[5],[3]). When running MCMC, one important question is the number of steps a Markov chain requires to converge. There are various approaches to analyzing this difficult problem. In this paper, we describe a challenging MCMC example, and show ways of deriving a quantitative mathematical bound using techniques related to coupling.

We shall focus on the following model. Suppose we have n particles randomly located in the $\mathbb{R}^2$ plane (so the state space $\mathcal{X} = \mathbb{R}^{2n}$), and the unnormalized density of each configuration is given by

$$\pi(x) = \exp\left(-\left[c_1 \sum_{i=1}^{n} ||x_i|| + c_2 \sum_{i<j} ||x_i - x_j||^{-1}\right]\right), \tag{1}$$

where c_1, c_2 are positive constants and $||\cdot||$ is the usual Euclidean (L^2) norm on $\mathbb{R}^2$. Since the density is fairly complicated, it is hard to compute expected values with respect to this distribution, such as the average distance of the particles to the origin. Therefore, a more feasible solution is to simulate this distribution using an MCMC algorithm. We shall use componentwise versions of the Metropolis-Hastings algorithm [8, 6], in which the multiple particles are updated one at a time in a sequential order, each with a proposal followed by an accept/reject step. (For a graphical illustration of this algorithm on these densities, see [15].) By running the algorithm for many iterations, we can approximately sample from π, and thus find good estimates of its expected values.

The density function (1) is designed so the first summation "pulls" the particles towards the origin, while the second summation "pushes" them away from each other. Hence, we call this an *attractive-repulsive* point process. The combination of attractive and repulsive forces mean that the MCMC algorithm does not satisfy simple monotonicity or other properties which would simplify its convergence analysis, so that more careful techniques are required. Nevertheless, for certain special cases of this density, we will derive both qualitative and quantitative convergence bounds herein.

This paper is organised as follows. In Section 2, we consider a version of our algorithm within a bounded domain, and show that it is uniformly ergodic by means of an explicit uniform minorization condition. In Section 3, we expand the state space to all of $\mathbb{R}^2$, and show that a version of our algorithm is still geometrically ergodic since it satisfies an explicit univariate drift condition. In Section 4, we discuss the challenges of computing a quantitative convergence bound for our algorithm, and use a *shift coupling* construction to overcome these problems and obtain an explicit quantitative bound. In Section 5, we provide proofs of all of the theorems in this paper.

1.1 Background: Minorization and Drift Conditions

We are interested in bounding the *total variation distance*

$$\|P^n(x,\cdot)-\pi(\cdot)\| := \sup_{S\subseteq\mathcal{X}} |P^n(x,S)-\pi(S)| = \sup_{S\subseteq\mathcal{X}} |P(X_n\in S\,|\,X_0=x)-\pi(S)|$$

between the n-step distribution $P^n(x,\cdot)$ and the stationary distribution $\pi(\cdot)$ of a Markov chain. One method involves coupling via minorization and drift conditions. A Markov chain with a state space $\mathcal{X}$ and transition probabilities $P(x,\cdot)$ satisfies a *minorization condition* if there is a measurable subset $C\subseteq\mathcal{X}$, a probability measure Q on $\mathcal{X}$, a constant $\epsilon>0$, and a positive integer n_0, such that

$$P^{n_0}(x, \cdot) \geq \epsilon\, Q(\cdot), \quad x \in C. \tag{2}$$

We call such C a *small set*, and refer to it (n_0, ϵ_{n_0}, Q)-small. In particular, if $C = \mathcal{X}$ (i.e., C is the entire state space), then we say the Markov chain satisfies a *uniform* minorization condition, also referred to as *Doeblin's condition* (see [4]). It then follows (see e.g. [9, 11]) that the chain is *uniformly ergodic*, i.e. there are fixed $\rho < 1$ and $M < \infty$ such that

$$\|P^n(x, \cdot) - \pi(\cdot)\| \;\leq\; M\, \rho^n, \quad n \in \mathbb{N}, \quad x \in \mathcal{X}.$$

Indeed, we have:

Proposition 1: If a Markov chain with stationary distribution $\pi(\cdot)$ has the property that the entire state space $\mathcal{X}$ is (n_0, ϵ_{n_0}, Q)-small, then the chain is uniformly ergodic, with

$$\|P^n(x, \cdot) - \pi(\cdot)\| \leq (1 - \epsilon_{n_0})^{\lfloor \frac{n}{n_0} \rfloor}, \quad n \in \mathbb{N}\,.$$

In Section 2, we prove uniform ergodicity for a bounded version of our algorithm. Unfortunately, many Markov chains are not uniformly ergodic. Instead, a Markov chain with stationary $\pi(\cdot)$ is *geometrically ergodic* if there are fixed $\rho < 1$ and π-a.e.-finite function $M : \mathcal{X} \to [0, \infty]$ such that

$$\|P^n(x, \cdot) - \pi(\cdot)\| \;\leq\; M(x)\, \rho^n, \quad n \in \mathbb{N}, \quad x \in \mathcal{X},$$

i.e. if the multiplier M can depend on the initial state x. Also, a Markov chain with a small set C satisfies a *univariate drift condition* if there are constants $0 < \lambda < 1$ and $b < \infty$, and a π-a.e. finite function $V : \mathcal{X} \to [1, \infty]$ such that

$$PV(x) \;:=\; E[V(X_1)\,|\,X_0 = x] \;\leq\; \lambda\, V(x) + b\, \mathbf{1}_C(x), \quad x \in \mathcal{X}. \tag{3}$$

The combination of a minorization condition (2) and a drift condition (3) guarantees (see e.g. [9, 11]) that the chain is geometrically ergodic:

Proposition 2: If a Markov chain with stationary distribution $\pi(\cdot)$ and small set $C \subset \mathcal{X}$ satisfies the univariate drift condition (3) for some $0 < \lambda < 1$, $b < \infty$, and π-a.e.-finite function $V : \mathcal{X} \to [1, \infty]$, then it is geometrically ergodic.

Geometric ergodicity is a helpful property, since it implies the chain converges geometrically quickly, and also implies certain other results such as central limit theorems (see e.g. [11]). We establish it for an unbounded version of our algorithm in Section 3. Unfortunately, qualitative bounds such as uniform or geometric ergodicity can still be quite weak in many cases, and do not necessarily imply that the Markov chain converges in a short time. For example, if $\mathcal{X} = \{0, 1\}$, with $X_0 = 1$ and

$$P = \begin{pmatrix} 1 & 0 \\ 1-z & z \end{pmatrix}$$

for some fixed $z \in (0, 1)$, then $\pi = (1, 0)$, and the chain satisfies a uniform minorization condition with $\epsilon = 1-z$ and $Q = (1, 0)$. So, it is both uniformly and geometrically ergodic, and in fact $\|P^n(x, \cdot) - \pi(\cdot)\| = z^n$. However, it converges arbitrarily slowly for z near 1, indicating that geometric ergodicity does not really imply fast convergence. Due to these limitations, it is best to find a *quantitative* bound, i.e. explicit bounds on $\|P^n(x, \cdot) - \pi(\cdot)\|$ which provide a value of n that guarantees that this distance will be sufficiently small. We consider this problem for an unbounded version of our attractive-repulsive processes in Section 4 below.

2 Particles in a Square: Uniform Ergodicity

In this section, we study the attractive-repulsive point process density (1) in a compact setting. Suppose we have three particles randomly located in the square $U = [0, 1]^2 \subset \mathbb{R}^2$, with each particle denoted by $(x_i)_{i=1,2,3} = (x_{i1}, x_{i2})_{i=1,2,3}$, so the state space $\mathcal{X} = [0, 1]^6$.

We use a componentwise Metropolis algorithm with systematic scan, in which we repeatedly update the three particles in order (see e.g. [8, 3, 15]). Specifically, given a configuration $X_n = x$, we first "propose" a new location for the first particle x_1 from the uniform (Lebesgue) measure on $\mathcal{X}$, to obtain a new particle location y_1, and hence a new configuration $y = (y_1, x_2, x_3)$. Then with probability $\alpha(x, y) = \min\left[1, \frac{\pi(y)}{\pi(x)}\right]$, we "accept" this proposal and update x_1 to y_1. Otherwise, we "reject" this proposal and leave the original x_1 unchanged. We then similarly update x_2 and then x_3. That entire procedure represents one iteration of our algorithm, which we then repeat n times to obtain a final configuration X_n.

For this algorithm, we show (all theorems are proved in Section 5):

Theorem 1. The above Markov Chain is uniformly ergodic, and satisfies a uniform minorization condition with $n_0 = 1$ and $\epsilon = (0.48)e^{-c_1(8.49)-c_2(19.76)}$.

For example, if $c_1 = c_2 = 1/10$, then we can take $\epsilon = 0.028$. By Proposition 1, we have $\|P(X_n, \cdot) - \pi(\cdot)\| \le 0.972^n$. This shows that after 163 steps, the total variation distance between the n-step distribution and the stationary distribution $\pi(\cdot)$ of this Markov chain will be within 0.01.

3 One particle in $\mathbb{R}^2$: Geometric Ergodicity

We now extend our state space to the whole $\mathbb{R}^2$ plane, but limit the number of particles to one. Specifically, suppose we have a particle randomly located at $\mathbb{R}^2$, denoted by $x = (x_1, x_2)$, with unnormalized density given by

$$\pi(x) \;=\; e^{-H(x)}, \quad \text{where} \quad H(x) \;=\; ||x|| + \frac{1}{||x||} \;:=\; r_x + \frac{1}{r_x},$$

where $r_x := ||x||$ is again the L^2 norm. We use the following Metropolis-Hastings algorithm on this distribution. For any $x = (x_1, x_2) \in \mathbb{R}^2$, let

$$B_x \;=\; \{z \in \mathbb{R}^2 : |r_x - 1| < \|z\| < r_x + 1\}\,.$$

Thus, B_x is an annulus of width $2\min(r_x, 1)$, which contains x unless $r_x < 0.5$; see Figure 1. And, $vol(B_x) = \pi(r_x+1)^2 - \pi|r_x-1|^2 = 4\pi r_x$. We then let the proposal density $q(x, \cdot)$ be the uniform distribution on B_x, i.e.

$$q(x, dy) \;=\; \mathbf{1}_{B_x}(y)\, \frac{dy}{4\pi r_x}\,, \quad x, y \in R^2\,.$$

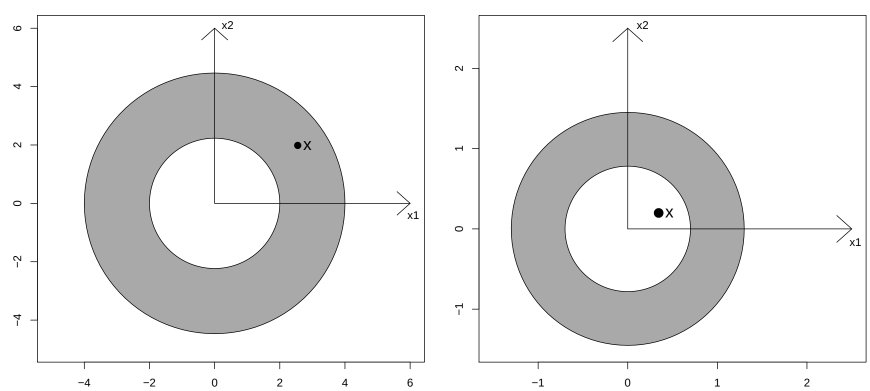

Figure 1: Illustration of the region B_x, when $r_x = 3$ (left) or 0.3 (right).

Note that $y \in B_x$ if and only if $x \in B_y$ (since for $r_x, r_y \le 1$ this is equivalent to $r_x + r_y < 1$; and for $r_x < 1 < r_y$ or $r_y < 1 < r_x$ this is equivalent to $\min[r_x, r_y] < \max[r_x, r_y]+1$; and for $r_x, r_y \geq 1$ this is equivalent to $|r_x - r_y| < 1$). Hence, these $q(x, dy)$ are valid proposal distributions for a Metropolis-Hastings algorithm. The corresponding acceptance rate is

$$\alpha(x, y) \; = \; \min\left\{1, \; \frac{\pi_u(y)\, q(y, x)}{\pi_u(x)\, q(x, y)}\right\} \; = \; \min\left\{1, \; \frac{e^{H(x)}\, r_x}{e^{H(y)}\, r_y}\right\}.$$

We shall prove the following:

Theorem 2. The Markov chain constructed above satisfies:
(a) the minorization condition

$$P^2(x, \cdot) \; \geq \; (3.5 \times 10^{-5})\, Q(\cdot), \quad x \in C,$$

where $C = \{x \in \mathbb{R}^2, \; \frac{1}{4} \leq ||x|| \leq 4\} \subseteq \mathcal{X}$.
(b) the univariate drift condition

$$PV(x) \; \leq \; 0.995\, V(x) + (e^{2.7} - 0.995)\, \mathbf{1}_C, \quad x \in \mathcal{X},$$

where $V(x) = e^{\frac{1}{2}H(x)}$. Furthermore, $\sup_{x \in C} PV(x) \leq e^{2.7}$.
Hence, by Proposition 2, this chain is geometrically ergodic.

4 Quantitative Bounds and Shift Coupling

We next consider quantitative bounds for the algorithm in the previous section. There are many potential ways to obtain quantitative bounds for MCMC algorithms. However, not all methods are feasible for our attractive-repulsive process.

For example, one approach uses minorization conditions and bivariate drift conditions (e.g. [13]). Theorem 3 already provides a minorization condition and a univariate drift condition, and there are ways to derive a bivariate drift condition from a univariate one if certain conditions are satisfied (see e.g. Proposition 11 of [11]). However, to obtain a bivariate drift condition for our processes, we would have to prove a multi-step minorization condition on a much larger subset, which would be very challenging and lead to extremely weak bounds. Alternatively, minorization and univariate drift conditions give good quantitative bounds for Markov chains which are *stochastically monotone* (see [7, 12]), but the attractive-repulsive nature of our processes seems to preclude any stochastic monotonicity condition.

Instead, we shall use a particular coupling method called *shift coupling* [1, 10] to derive a quantitative bound for the point process. This construction only requires a univariate drift condition (not a bivariate one), and does not require aperiodicity. In the shift coupling construction, just like ordinary coupling, we will jointly define two Markov chains to obtain a bound on the rate of convergence. The key point in which shift coupling differs from the ordinary method is that we allow the chains to couple at different times.

Let $P(\cdot, \cdot)$ be the transition probabilities for a Markov chain on a state space $\mathcal{X}$. Assume the chain is ϕ-irreducible, with stationary distribution $\pi(\cdot)$. Let $\{X_k\}_{k=0}^{\infty}$ and $\{X'_k\}_{k=0}^{\infty}$ be two different copies of the chain, defined jointly. Suppose T and T' are two random variables taking values in $\mathbb{Z}_{\geq 0} \cup \{\infty\}$, such that for any non-negative integer n, $X_{T+n} = X_{T'+n}$. Ordinary coupling requires $T = T'$, but shift coupling allows the two Markov chains to become equal at different times, thus making it easier for the chains to couple. We can then combine this shift-coupling bound with minorization and univariate drift conditions, leading to the following (which generalizes Theorem 4 of [10] to the case $n_0 > 1$):

Theorem 3: Suppose a Markov chain on a state space $\mathcal{X}$, with initial distribution $\nu(\cdot)$, transition probabilities $P(\cdot, \cdot)$, and stationary distribution $\pi(\cdot)$, satisfies the minorization condition (2), and the drift condition (3), such that $C = \{x \in \mathcal{X} : V(x) \leq d\}$ for some fixed $d \geq 0$. Then setting $A := \sup_{x \in C} E(V(X_1)|X_0 = x)$ (so $A \leq \lambda d + b$), for any $0 < r < 1$ such that $\lambda^{(1-n_0 r)} A^r < 1$, we have

$$\begin{aligned} &\left\| \frac{1}{n} \sum_{k=1}^{n} P(X_k \in \cdot) - \pi(\cdot) \right\| \\ &\leq \frac{1}{n} \left[\frac{2(1-\epsilon)^r}{1-(1-\epsilon)^r} + \frac{\lambda^{-n_0+1-n_0 r} A^r}{1-\lambda^{1-n_0 r} A^r} \left(E_\nu(V) + \frac{b}{1-\lambda} \right) \right]. \end{aligned}$$

We now apply this shift-coupling bound to the attractive-repulsive point processes of Section 3. By Theorem 2, we can take $\epsilon = 3.5 \times 10^{-5}$, $n_0 = 2$, $\lambda = 0.995$, $b = e^{2.7} - 0.995$, $d = e^{17/8}$, and $A = e^{2.7}$. Assume the chain starts from the point $(1, 0)$, so $E_\nu(V) = V((1,0)) = e^{\frac{1}{2}(1+\frac{1}{1})} = e$. Choosing $r = 0.0016$, we have $\lambda^{(1-n_0 r)} A^r \doteq 0.9993 < 1$, and we compute from Theorem 3 that

$$\left\| \frac{1}{n} \sum_{k=1}^{n} P(X_k \in .) - \pi(.) \right\| \;\leq\; \frac{39,900,000}{n}.$$

This bound is certainly far from tight. However, it does show that shift-coupling can provide explicit quantitative bounds on the distance to stationarity, even for the attractive-repulsive processes that we consider herein.

Finally, we note that the left-hand side of the bound in Theorem 3 differs somewhat from the conventional total variation distance between the n-step distribution and the stationary distribution. This raises the question of the meaning of the quantity we are bounding. An interpretation is given by:

Theorem 4: For $n \in \mathbb{N}$, let $F_n(S)$ be the expected fraction of time from 1 to n that the chain is inside S. Then

$$\sup_S |F_n(S) - \pi(S)| = \left\| \frac{1}{n} \sum_{k=1}^{n} P(X_k \in \cdot) - \pi(\cdot) \right\| \leq \frac{1}{n} \sum_{1}^{n} \left\| P(X_k \in \cdot) - \pi(\cdot) \right\|.$$

Theorem 4 provides context for Theorem 3. It shows that the bound of Theorem 3 in turn provides an upper bound on the difference between the expected occupation fraction of S and the target probability $\pi(S)$, uniformly over choice of subset S. So, if the bound is small, then the chain spends approximately the target fraction of time in every subset, on average.

Theorem 4 also gives us a way to relate the shift coupling result to more conventional results. In particular, note that $||P(X_k \in \cdot) - \pi(\cdot)||$ is the usual total variation distance discussed in previous sections. Hence, $\frac{1}{n}\sum_1^n ||P(X_k \in \cdot) - \pi(\cdot)||$ is the average of the total variation distances between the k-step distribution and the stationary distribution, averaged over $k = 1, 2, \ldots, n$.

5 Theorem Proofs

5.1 Proof of Theorem 1

Let

$$\mathcal{X}' = \{(x_1, x_2, x_3) \in \mathcal{X} : \forall 1 \leq i < j \leq 3, |x_i - x_j| \geq 1/4\}.$$

Since $\mathcal{X}'$ is compact and $\pi(\cdot)$ is continuous and positive on $\mathcal{X}'$, it must achieve its minimum ratio $m := \min_{x,y \in \mathcal{X}'} \frac{\pi(y)}{\pi(x)} > 0$. Then for any $x = (x_1, x_2, x_3) \in [0,1]^6$ and measurable $A \subseteq \mathcal{X}$,

$$\begin{aligned}
P(x, A) &= \int_A P(x, dy) \\
&\geq \int_A P_1((x_1, x_2, x_3), dy_1)\, P_2((y_1, x_2, x_3), dy_2)\, P_3((y_1, y_2, x_3), dy_3) \\
&\geq \int_{A \cap \mathcal{X}'} \min\left[1, \frac{\pi(y_1, x_2, x_3)}{\pi(x_1, x_2, x_3)}\right] \min\left[1, \frac{\pi(y_1, y_2, x_3)}{\pi(y_1, x_2, x_3)}\right] \min\left[1, \frac{\pi(y_1, y_2, y_3)}{\pi(y_1, y_2, x_3)}\right] dy,
\end{aligned}$$

where $P_1((x_1, x_2, x_3), B) = P((x_1, x_2, x_3), B \times \{x_2\} \times \{x_3\})$ for any measurable $B \subset [0, 1]^2$ (and similarly for $i = 2, 3$). Denote the three acceptance rates by α_1, α_2, α_3 respectively.

If $\alpha_i = 1$ for some i (say $\alpha_1 = 1$), then $\alpha_1\alpha_2\alpha_3 = \alpha_2\alpha_3 \geq m^2$. On the other hand, if $\alpha_i < 1$ for $i = 1, 2, 3$, then $\alpha_1\alpha_2\alpha_3 = \frac{\pi(y_1,y_2,y_3)}{\pi(x_1,x_2,x_3)} \geq m \geq m^2$ (since $m \leq 1$). So

$$P(x, A) \;\geq\; \int_{A\cap\mathcal{X}'} m^2\, dx \;=\; m^2\,\mathrm{Leb}(A \cap \mathcal{X}'),$$

where Leb is Lebesgue measure on $\mathbb{R}^2$. It follows that our algorithm satisfies a uniform minorization condition, with $\epsilon = m^2\,\mathrm{Leb}(\mathcal{X}')$ and $Q(A) = \frac{\mathrm{Leb}(A\cap\mathcal{X}')}{\mathrm{Leb}(\mathcal{X}')}$. Hence, by Proposition 1, this chain is uniformly ergodic.

To obtain a quantitative bound, we need to compute m^2 and $\mathrm{Leb}(\mathcal{X}')$. For any $x \in \mathcal{X}'$, we must have $0 \leq |x_i| \leq \sqrt{2}$ and $1/4 \leq |x_i - x_j| \leq \sqrt{2}$, thus

$$0 \leq \sum_i |x_i| \leq 3\sqrt{2}, \text{ and } \frac{3}{\sqrt{2}} \leq \sum_{i<j} |x_i - x_j|^{-1} \leq 12.$$

Then

$$m = \frac{\min_{\mathcal{X}'} \pi(\cdot)}{\max_{\mathcal{X}'} \pi(\cdot)} = \frac{e^{-c_1(3\sqrt{2})-c_2(12)}}{e^{-c_1(0)-c_2(3/\sqrt{2})}} = e^{-c_1(3\sqrt{2})-c_2(12-3/\sqrt{2})}.$$

Thus

$$m^2 \geq \left(e^{-c_1(3\sqrt{2})-c_2(12-3/\sqrt{2})}\right)^2 \geq e^{-c_1(8.49)-c_2(19.76)}.$$

Lastly we need to compute $\mathrm{Leb}(\mathcal{X}')$. To make $(x_1, x_2, x_3) \in \mathcal{X}$, we can choose any $x_1 \in [0, 1]^2$ (with area 1), then any $x_2 \in [0, 1]^2 \setminus B(x_2, 1/4)$ (with area $\geq 1 - 3.14(1/4)^2$), then any x_3 $[0, 1]^2 \setminus (B(x_1, 1/4) \cup B(x_2, 1/4))$(with area $\geq 1 - 3.14(1/4)^2 - 3.14(1/4)^2$). Hence

$$\mathrm{Leb}(\mathcal{X}') \geq (1)\left(1 - \frac{\pi}{16}\right)\left(1 - \frac{\pi}{8}\right) \geq 0.48.$$

Therefore

$$\epsilon \;=\; m^2\mathrm{Leb}(\mathcal{X}') \;\geq\; (0.48)e^{-c_1(8.49)-c_2(19.76)}.$$

□

5.2 Proof of Theorem 2 (a)

Recall that

$$\alpha(x, y) \;=\; \min\left\{1, \frac{e^{H(x)}r_x}{e^{H(y)}r_y}\right\} \;=\; \min\left\{1, \frac{f(r_x)}{f(r_y)}\right\},$$

where $f:\mathbb{R}\to\mathbb{R}$ by $f(x)=xe^{x+\frac{1}{x}}$. We then have

$$f'(x)=e^{x+\frac{1}{x}}+x(1-\frac{1}{x^2})e^{x+\frac{1}{x}}=(x-\frac{1}{x}+1)e^{x+\frac{1}{x}},$$

so that

$$f'(x)=0 \iff x=\frac{-1\pm\sqrt{5}}{2},$$

and hence $f(x)$ is decreasing on $(0,\frac{\sqrt{5}-1}{2})$ and increasing on $(\frac{\sqrt{5}-1}{2},\infty)$.

Next, let

$$C_1=\{x\in C:1/4\le r_x\le 2\},$$

$$C_2=\{x\in C:2\le r_x\le 4\},$$

$$D=\{x\in\mathbb{R}^2, 2\le r_x\le 9/4\},$$

$$E_1=\{x\in\mathbb{R}^2, 1\le r_x\le 5/4\},$$

and

$$E_2=\{x\in\mathbb{R}^2, 3\le r_x\le 13/4\}.$$

We shall show that $P^2(x,\cdot)$ has an overlap on D for all $x\in C$. In particular, we will consider the case when the x first jumps into E_1 and then enters D for $x\in C_1$ (similarly for E_2).

We know

$$\alpha(x,y)=\min\left\{1,\frac{f(r_x)}{f(r_y)}\right\},$$

and we have shown f takes its minimum at $\frac{\sqrt{5}-1}{2}$ and is increasing on $(\frac{\sqrt{5}-1}{2},\infty)$. Therefore

$$m_1:=\min_{C_1\times E_1}\alpha(x,y)=\frac{f(\frac{\sqrt{5}-1}{2})}{f(\frac{5}{4})}\ge 0.59,\ m_2:=\min_{C_2\times E_2}\alpha(x,y)=\frac{f(2)}{f(\frac{13}{4})}\ge 0.21,$$

$$m_1':=\min_{E_1\times D}\alpha(x,y)=\frac{f(1)}{f(\frac{9}{4})}\ge 0.22,\ m_2':=\min_{E_2\times D}\alpha(x,y)=\min\{\frac{f(3)}{f(\frac{9}{4})},1\}=1.$$

For any $x\in C_1$, $y\in D$, take $M_y=\{z\in\mathbb{R}^2, r_y-1\le r_z\le 5/4\}\subset E_1$. Then for any $z\in M_y$, $r_z\le 5/4\le r_x+1$ and $r_z\ge 2-1=1\ge|r_x-1|$. Thus

$M_y \subset B_x$, and then

$$
\begin{aligned}
P^2(x,dy) &= \int_{B_x} P(x,dz)P(z,dy) \geq \int_{M_y} P(x,dz)P(z,dy) \\
&= \frac{1}{4\pi|x|}\left(\int_{M_y} \alpha(x,z)\alpha(z,y)q(y,z)dz\right)dy \\
&\geq \frac{1}{8\pi}\left(\int_{M_y} m_1 m_1' \cdot \frac{1}{4\pi|z|}dz\right)dy \\
&= \frac{m_1 m_1'}{8\pi}\left(2\pi \int_{r_y-1}^{\frac{5}{4}} \frac{1}{4\pi}dr\right)dy \\
&= \frac{m_2 m_2'}{16\pi}(\frac{9}{4} - r_y)dy \geq \frac{0.13}{16\pi}(\frac{9}{4} - r_y)dy.
\end{aligned}
$$

For any $x \in C_2$, $y \in D$, take $N_y = \{z \in \mathbb{R}^2, 3 \leq r_z \leq r_y + 1\} \subset E_2$. Similarly we have

$$
\begin{aligned}
P^2(x,dy) &= \int_{B_x} P(x,dz)P(z,dy) \geq \int_{M_y} P(x,dz)P(z,dy) \\
&= \frac{1}{4\pi|x|}\left(\int_{M_y} \alpha(x,z)\alpha(z,y)q(y,z)dz\right)dy \\
&\geq \frac{1}{16\pi}\left(\int_{M_y} m_2 m_2' \cdot \frac{1}{4\pi|z|}dz\right)dy \\
&= \frac{m_1 m_1'}{32\pi}(r_y - 2)dy \geq \frac{0.1}{16\pi}(r_y - 2)dy.
\end{aligned}
$$

Then

$$
P^2(x,dy) \geq 1_D \frac{1}{16\pi} \min\left\{0.13(\frac{9}{4} - ||y||), 0.1(||y|| - 2)\right\} dy,
$$

where the size $\epsilon \geq 3.5 * 10^{-5}$. $\square$

5.3 Proof of Theorem 2 (b)

Since H and V only depend on r_x, we will regard them as functions of $r_x \in \mathbb{R}$ in the rest of the proof. We consider three different cases.

Case 1: $r_x > 4$.

Then $r_y > 4 - 1 > \frac{\sqrt{5}-1}{2}$ for any $y \in B_x$. So f is increasing on $(r_x - 1, r_x + 1)$. For any $y \in B_x$, we have $\alpha(x,y) = 1$ if and only if $r_y \leq r_x$. Let

$A_x = B(0, r_x) \setminus B(0, r_x - 1)$ (the inner part of the annulus). Then

$$\begin{aligned}
PV(x) &= \int_{R^2} V(y)P(x, dy) \\
&= \frac{1}{4\pi r_x}(\int_{A_x} V(y)dy + \int_{B_x\setminus A_x} V(y)\frac{f(r_x)}{f(r_y)} + \int_{B_x\setminus A_x} V(x)(1 - \frac{f(r_x)}{f(r_y)})dy \\
&= \frac{1}{4\pi r_x}(\int_{A_x} V(y)dy + \int_{B_x\setminus A_x} (V(x) + (V(y) - V(x))\frac{f(r_x)}{f(r_y)})dy).
\end{aligned}$$

Let

$$\begin{aligned}
I(x, y) &= V(x) + (V(y) - V(x))\frac{f(r_x)}{f(r_y)} = V(x)(1 + (\frac{V(y)}{V(x)} - 1)\frac{f(r_x)}{f(r_y)}) \\
&= V(x)(1 + (e^{\frac{1}{2}(H(y)-H(x))} - 1)\frac{e^{H(x)}r_x}{e^{H(y)}r_y}).
\end{aligned}$$

Let $u = H(y) - H(x)$, and set

$$I(x, y) = V(x)(1 + (e^{\frac{1}{2}u} - 1)e^{-u}\frac{r_x}{r_y}).$$

Then

$$\begin{aligned}
\int_{B_x\setminus A_x} I(x, y)dy &= V(x)\left(\int_{B_x\setminus A_x} dy + \int_{B_x\setminus A_x} (e^{\frac{1}{2}u} - 1)e^{-u}\frac{r_x}{r_y}dy\right) \\
&= V(x)(vol(B_x \setminus A_x) + r_x \int_{B_x\setminus A_x} (e^{-\frac{1}{2}u} - e^{-u})\frac{1}{r_y}dy).
\end{aligned}$$

Since u is a function of r_y (i.e. u only depends on the magnitude of y),

$$\int_{B_x\setminus A_x} (e^{-\frac{1}{2}u} - e^{-u})\frac{1}{r_y}dy = \int_0^{2\pi}\int_{r_x}^{r_x+1} (e^{-\frac{1}{2}u} - e^{-u})\frac{1}{r}rdrd\theta = 2\pi\int_{r_x}^{r_x+1} (e^{-\frac{1}{2}u} - e^{-u})dr.$$

Since $r_x \leq r_y \leq r_x + 1$, $u = H(y) - H(x) = r_y - r_x + \frac{1}{r_y} - \frac{1}{r_x} \leq r_y - r_x \leq 1$.
Note that $(e^{-\frac{1}{2}u} - e^{-u})$ is increasing for $u \in (0, 1)$. So

$$\begin{aligned}
\int_{r_x}^{r_x+1} (e^{-\frac{1}{2}u} - e^{-u})dr &\leq \int_{r_x}^{r_x+1} (e^{-\frac{1}{2}(r-r_x)} - e^{-(r-r_x)})dr \\
&= \int_0^1 (e^{-\frac{1}{2}t} - e^{-t})dt = 1 + e^{-1} - 2e^{-\frac{1}{2}}.
\end{aligned}$$

Denote $(1+e^{-1}-2e^{-\frac{1}{2}})$ by m_1. Then

$$\int_{B_x\setminus A_x} I(x,y)dy \le V(x)(vol(B_x\setminus A_x)+2\pi m_1 r_x) = 2\pi V(x)(r_x+\frac{1}{2}+m_1 r_x).$$

(since $vol(B_x\setminus A_x)=\pi(r_x+1)^2-\pi r_x^2=\pi(2r_x+1)$). Now consider the other part.

$$\int_{A_x} V(y)dy = 2\pi\int_{r_x-1}^{r_x} e^{\frac{1}{2}(r+\frac{1}{r})}rdr = 2\pi V(x)\int_{r_x-1}^{r_x} e^{\frac{1}{2}(r-r_x+\frac{1}{r}-\frac{1}{r_x})}rdr.$$

Note

$$r-r_x+\frac{1}{r}-\frac{1}{r_x} = r-r_x+\frac{r_x-r}{rr_x} \le r-r_x+\frac{r_x-r}{12} = \frac{11}{12}(r-r_x).$$

(the inequality follows from the fact that $rr_x\ge(r_x-1)r_x\ge(4-1)4=12$).
So

$$\int_{A_x} V(y)dy \le 2\pi V(x)\int_{r_x-1}^{r_x} e^{\frac{11}{24}(r-r_x)}rdr = 2\pi V(x)\int_{-1}^{0} e^{\frac{11}{24}t}(t+r_x)dt$$

$$= 2\pi V(x)(\int_{-1}^{0} te^{\frac{11}{24}t}dt + r_x\int_{-1}^{0} e^{\frac{11}{24}t}dt) = 2\pi V(x)(\frac{840e^{-\frac{11}{24}}-576}{121}+\frac{24(1-e^{-\frac{11}{24}})}{11}r_x).$$

Denote this by $2\pi V(x)(m_2r_x+m_3)$. Then

$$\begin{aligned} PV(x) &\le \frac{1}{4\pi r_x}(2\pi V(x)(r_x+\frac{1}{2}+m_1r_x)+2\pi V(x)(m_2r_x+m_3)) \\ &= \frac{V(x)}{2}(1+\frac{1}{2r_x}+m_1+m_2+\frac{m_3}{r_x}) \\ &\le \frac{1}{2}(1+\frac{1}{8}+m_1+m_2+\frac{m_3}{4})V(x) \text{ (as } r_x>4) \\ &< 0.995V(x). \end{aligned}$$

Case 2: $r_x<1/4$.

In this case $|r_x-1|=1-r_x>1-1/4=3/4$, and $(r_x+1)<1/4+1=5/4$. So $B_x\subset(B(0,\frac{5}{4})\setminus B(0,\frac{3}{4}))$. Note

$$\max_{y\in B_x} H(y) \le \max\{H(\frac{3}{4}),H(\frac{5}{4})\} = \max\{\frac{3}{4}+\frac{4}{3},\frac{4}{5}+\frac{5}{4}\} = \frac{25}{12}.$$

And

$$H(x) \ge \frac{1}{4}+4 = \frac{17}{4}.$$

So for any $y \in B_x$,

$$V(y)/V(x) = e^{\frac{1}{2}(H(y)-H(x))} \leq e^{\frac{1}{2}(\frac{25}{12}-\frac{17}{4})} = e^{-\frac{13}{12}}.$$

Then we will show the acceptance rate is always 1. Recall

$$\alpha(x,y) = \min\{1, \frac{\pi_u(y)q(y,x)}{\pi_u(x)q(x,y)}\} = \min\{1, \frac{f(r_x)}{f(r_y)}\}.$$

We showed $f(x)$ is decreasing on $(0, \frac{\sqrt{5}-1}{2})$ and is increasing on $(\frac{\sqrt{5}-1}{2}, \infty)$. Since $\frac{1}{4} < \frac{\sqrt{5}-1}{2} < \frac{3}{4}$, we have

$$f(r_x) \geq f(\frac{1}{4}) = \frac{1}{4}e^{\frac{17}{4}}, \quad f(r_y) \leq f(\frac{5}{4}) = \frac{5}{4}e^{\frac{41}{20}}.$$

So

$$\frac{f(r_x)}{f(r_y)} \geq \frac{\frac{1}{4}e^{\frac{17}{4}}}{\frac{5}{4}e^{\frac{41}{20}}} > \frac{e^2}{5} > 1, \ y \in B_x.$$

Therefore

$$PV(x) = \int_{B_x} q(x,y)V(y)dy \leq \int_{B_x} q(x,y)e^{-\frac{13}{12}}V(x)dy = e^{-\frac{13}{12}}V(x) < 0.995V(x).$$

Case 3: $r_x \in [1/4, 4]$ (i.e., $x \in C$).

Let $E = B(0, \frac{1}{4})$. Note $r_y \leq 5$ for all $y \in B_x$. If $y \notin E$ is proposed, since $1/4 \leq r_y \leq 5$ and $V(1/4) = V(4) \leq V(5)$,

$$V(X_{n+1}) \leq \max\{V(x), V(5)\} \leq \max\{V(4), V(5)\} = e^{\frac{13}{5}}.$$

If $y \in E$ is proposed, first note this requires $|r_x - 1| < \frac{1}{4}$. So $r_x \in [\frac{3}{4}, \frac{5}{4}]$. Then

$$f(r_x) \leq f(\frac{5}{4}) = \frac{5}{4}e^{\frac{41}{20}}, \ f(r_y) \geq f(\frac{1}{4}) = \frac{1}{4}e^{\frac{17}{4}}.$$

So

$$\frac{f(r_x)}{f(r_y)} \leq \frac{\frac{5}{4}e^{\frac{41}{20}}}{\frac{1}{4}e^{\frac{17}{4}}} < 1, \ y \in E.$$

This implies

$$\alpha(x,y) = \frac{f(r_x)}{f(r_y)} < 1, \ y \in E \cap B_x.$$

Note

$$PV(x) = \int_{B_x \cap E} V(y)P(x,dy) + \int_{B_x \setminus E} V(y)P(x,dy).$$

Clearly if $r_x \notin [\frac{3}{4}, \frac{5}{4}]$ (i.e. $B_x \cap E = \emptyset$), then $PV(x) \leq V(5) = e^{\frac{13}{5}} = e^{2.6} < e^{2.7}$. Otherwise

$$\begin{aligned}
PV(x) &\leq \int_{B_x \cap E} q(x,y) \frac{f(r_x)}{f(r_y)} V(y) dy + \int_{B_x \setminus E} V(5) P(x, dy) \\
&= \frac{f(r_x)}{4\pi r_x} \int_{B_x \cap E} \frac{e^{\frac{1}{2}H(y)}}{e^{H(y)} r_y} dy + V(5) \\
&\leq \frac{f(r_x)}{4\pi r_x} 2\pi \int_0^{\frac{1}{4}} e^{-\frac{1}{2}(r + \frac{1}{r})} dr + V(5) \\
&< \frac{e^{(r_x + \frac{1}{r_x})} 0.001}{2} + V(5) \\
&\leq \frac{e^{4+\frac{1}{4}}}{2000} + V(5) < e^{2.7}.
\end{aligned}$$

Therefore

$$PV(x) \leq e^{2.7}, \quad x \in C.$$

On the other hand, we always have $V(x) = e^{\frac{1}{2}H(x)} \geq e^{\frac{1}{2}(1+\frac{1}{1})} = e$. So, for $x \in C$,

$$PV(x) \leq e^{2.7} \leq 0.995 V(x) + (e^{2.7} - 0.995) \mathbf{1}_C. \qquad \square$$

5.4 Proof of Theorem 3

It was shown in [2, 1] that if two copies $\{X_k\}_{k=0}^{\infty}$ and $\{X'_k\}_{k=0}^{\infty}$ of a time-inhomogeneous Markov chain have shift-coupling times T and T', then the total variation distance between the ergodic averages of their distributions can be bounded as:

$$\left\| \frac{1}{n} \sum_{k=1}^{n} P(X_k, \cdot) - \frac{1}{n} \sum_{k=1}^{n} P(X'_k, \cdot) \right\| \;\leq\; \frac{1}{n} E\Big[\min(\max(T, T'),\, n) \Big]. \qquad (4)$$

Thus, Theorem 3 will follow by constructing copies $\{X_k\}$ and $\{X'_k\}$, with the latter in stationarity, in such a way that we can bound these shift-coupling tail probabilities. To do this, we generalize the construction of $\{X_k\}$ and $\{X'_k\}$ from Section 3 of [10] to the case $n_0 > 1$.

Specifically, we proceed as follows. We begin by choosing $X_0 \sim \nu$ and $X'_0 \sim \pi$ independently, and also generate an independent random variable $W \sim Q(\cdot)$. Then, whenever $V(X_n) \leq d$, we flip an independent coin with probability of heads equal to ϵ. If the coin comes up heads, we set

$X_{n+n_0} = W$ and $T = n + n_0$. If the coin comes up tails, we instead generate $X_{n+n_0} \sim \frac{1}{1-\epsilon}(P(X_n, \cdot) - \epsilon\, Q(\cdot))$, i.e. from the *residual* distribution. For completeness we then also "fill in" the values $X_{n+1}, \ldots, X_{n+n_0-1}$ by conditional probability, according to the Markov chain transition probabilities conditional on the already-constructed values of X_n and X_{n+n_0}. If instead $V(X_n) > d$, then we simply choose $X_{n+1} \sim P(X_n, \cdot)$ as usual. We continue this way until time T, i.e. until we get heads and set $X_T = W$.

We construct $\{X'_n\}$ and T' similarly, by flipping an independent ϵ-coin whenever $V(X'_n) \le d$, and setting either $X'_{n+n_0} = W$ or $X'_{n+n_0} \sim \frac{1}{1-\epsilon}(P(X'_n, \cdot) - \epsilon\, Q(\cdot))$ (and again we "fill in" $X'_{n+1}, \ldots, X'_{n+n_0-1}$ by conditional probability), up until the first head upon which we set $X'_{n+n_0} = W$ and $T' = n + n_0$.

This construction guarantees that $X_T = X'_{T'} = W \sim Q(\cdot)$. We then continue the two chains identically from W onwards, by choosing $X_{T+n} = X'_{T'+n} \sim P(X_{T+n-1}, \cdot)$ for $n = 1, 2, 3, \ldots$. Our construction ensures that each of $\{X_n\}$ and $\{X'_n\}$ each marginally follow the transition probabilities $P(\cdot, \cdot)$, and also that $X_{T+n} = X'_{T'+n}$ for $n = 0, 1, 2, \ldots$.

Now, combining the inequality (4) with the assumption that $P(X'_k \in \cdot) = \pi(\cdot)$ and the standard fact (see e.g. Proposition A.2.1 of [14]) that $E(Z) = \sum_{k=1}^{\infty} P(Z \ge k)$ for any non-negative integer-valued random variable Z, and noting that $P\left(\min[\max(T, T'),\, n] \ge k\right) \;\le\; P\left(\max(T, T') \ge k\right)$, yields the bound:

$$\left\| \frac{1}{n} \sum_{k=1}^{n} P(X_k \in \cdot) - \pi(\cdot) \right\| \;\le\; \frac{1}{n} \sum_{k=1}^{\infty} P\left(\max(T, T') \ge k\right). \tag{5}$$

We now bound $P\left(\max(T, T') \ge k\right)$ for any non-negative integer k. Let $t_1, t_2, \ldots$ be the times at which we flipped a coin for $\{X_n\}$, i.e. the times when $V(X_n) \le d$ excluding the "fill in" times. Then, let

$$N_k = \max\{i : t_i \le k\}$$

be the number of such coin-flip times up to and including time k. Since each coin-flip yields probability ϵ of reaching T after n_0 additional steps, we have for any integer $j \ge 1$ that $P(T \ge k, N_{k-n_0} \ge j) \le (1-\epsilon)^j$. Hence,

$$\begin{aligned} P(T \ge k) \;&=\; P(T \ge k,\, N_{k-n_0} \ge j) + P(T \ge k,\, N_{k-n_0} < j) \\ &\le\; (1-\epsilon)^j + P(N_{k-n_0} < j). \end{aligned} \tag{6}$$

Then since $\lambda < 1$, we have by Markov's inequality that

$$P(N_{k-n_0} < j) \;=\; P\left(t_j > k - n_0\right) \;=\; P\left(\lambda^{-t_j} > \lambda^{-k-n_0}\right) \;\le\; \lambda^{k-n_0}\, E\left[\lambda^{-t_j}\right].$$

To continue, let $\tau_1 = t_1$ and $\tau_i = t_i - t_{i-1}$ for $i \geq 2$. Then by Lemma 1 below,

$$\lambda^{k-n_0} E\left[\lambda^{-t_j}\right] = \lambda^{k-n_0} E\left[\lambda^{-(\tau_1+\ldots\tau_j)}\right] \leq \lambda^k E[V(X_0)] \, (\lambda^{-n_0} A)^{j-1}.$$

Hence, from (6),

$$P(T \geq k) \leq (1-\epsilon)^{[j]} + \lambda^{k-n_0(j-1)} A^{j-1} E_\nu(V).$$

Similarly,

$$P(T' \geq k) \leq (1-\epsilon)^{[j]} + \lambda^{k-n_0(j-1)} A^{j-1} E_\pi(V).$$

By Lemma 2 below, we have $E_\pi(V) \leq \frac{b}{1-\lambda}$. Hence,

$$\begin{aligned} P(\max(T,T') \geq k) &\leq P(T \geq k) + P(T' \geq k) \\ &\leq 2(1-\epsilon)^{[j]} + \lambda^{k-n_0(j-1)} A^{j-1} \left(E_\nu(V) + \frac{b}{1-\lambda} \right). \end{aligned}$$

Finally, choosing $j = \lfloor rk + 1 \rfloor \geq rk$ and using (5),

$$\begin{aligned} &\left\| \frac{1}{n} \sum_{k=1}^{n} P(X_k \in \cdot) - \pi(\cdot) \right\| \\ &\leq \frac{1}{n} \sum_{k=1}^{\infty} \left[2(1-\epsilon)^{rk} + \lambda^{-n_0} (\lambda^{1-n_0 r} A^r)^k \left(E_\nu(V) + \frac{b}{1-\lambda} \right) \right]. \end{aligned}$$

Since $(1-\epsilon)^r < 1$ and $\lambda^{(1-n_0 r)} A^r < 1$, the right hand side is a geometric sum which is equal to the claimed bound. $\square$

The above proof requires two lemmas. The first is a bound on expected values using a non-increasing expectation property, i.e. a partial supermartingale argument (similar to Lemma 4 of [13]):

Lemma 1. In the above proof of Theorem 3,
(a) $E\left[\lambda^{-\tau_1}\right] \leq E\left[V(X_0)\right]$, and
(b) for $i \geq 2$, $E\left[\lambda^{-\tau_i} | \tau_1, \ldots, \tau_{i-1}\right] \leq \lambda^{-n_0} A$.

Proof. Let

$$g_i(k) = \begin{cases} \lambda^{-k} V(X_k), & k \leq t_i \\ 0, & k > t_i \end{cases}$$

For (a), we know that $X_k \notin C$ for any $k < t_1$, so the drift condition implies that $g_1(k)$ has non-increasing expectation as a function of k, and hence

$$E\left[\lambda^{-\tau_1}\right] = E\left[\lambda^{-t_1}\right] \leq E\left[\lambda^{-t_1} V(X_{t_1})\right] = E\left[g_1(t_1)\right] \leq E\left[g_1(0)\right] = E\left[V(X_0)\right].$$

For (b), for any $i \geq 2$ we know that $X_k \notin C$ if $t_{i-1} + n_0 \leq k < t_i$, so the drift condition implies that $g_i(k)$ has non-increasing expectation as a function of k for $k \geq t_{i-1} + n_0$. Hence,

$$\begin{aligned}
E\left[\lambda^{-\tau_i} | X_{t_{i-1}}\right] &= E\left[\lambda^{-(t_i - t_{i-1})} | X_{t_{i-1}}\right] \\
&\leq E\left[\lambda^{t_{i-1}} \lambda^{-t_i} V(X_{t_i}) | X_{t_{i-1}}\right] \\
&= E\left[\lambda^{t_{i-1}} g_i(t_i) | X_{t_{i-1}}\right] \\
&\leq E\left[\lambda^{t_{i-1}} g_i(t_{i-1} + n_0) | X_{t_{i-1}}\right] \\
&\leq \lambda^{-n_0} E\left[V(t_{i-1} + n_0) | X_{t_{i-1}}\right] \\
&\leq \lambda^{-n_0} \sup_{x \in C} E\left[V(X_1) | X_0 = x\right]. \qquad \square
\end{aligned}$$

We also require a lemma which bounds $\pi(V)$.

Lemma 2: Suppose a ϕ-irreducible Markov chain on a state space $\mathcal{X}$, with transition probabilities $P(\cdot, \cdot)$ and stationary distribution $\pi(\cdot)$, satisfies the minorization condition (2) and the drift condition (3). Then the expected value $E_\pi(V) \leq b/(1-\lambda)$.

Proof. By parts (i) and (iii) of Theorem 14.0.1 of [9] with the choice $f(x) = (1-\lambda)\, V(x)$, it follows that $E_\pi(V) < \infty$. (That result is stated assuming aperiodicity, but it still holds in the periodic case by passing to the lazy chain $\bar{P} = \frac{1}{2}(I + P)$, which has the same π, and minorisation and drift with $\bar{\epsilon} = \epsilon/2$ and $\bar{b} = b/2$ and $\bar{\lambda} = (1+\lambda)/2$.) Then, taking expected values with respect to π of both sides of the drift condition $PV \leq \lambda V + b$ yields that $E_\pi(V) \leq \lambda\, E_\pi(V) + b$, which implies the result. $\square$

Remark: Lemma 2 can also be derived from Theorem 14.3.7 of [9], with the choices $f(x) = (1-\lambda)\, V(x)$, and $s(x) = b$, after verifying that the chain is positive recurrent using their Theorem 14.0.1.

5.5 Proof of Theorem 4

For any measurable subset S,

$$|F_n(S) - \pi(S)| = \Big|\mathbb{E}[\text{fraction of time from 1 to n that the chain is in S}] - \pi(S)\Big|$$
$$= \left|\mathbb{E}\left[\frac{1}{n}\sum_{1}^{n} \mathbf{1}_{X_k \in S}\right] - \pi(S)\right| = \left|\frac{1}{n}\sum_{k=1}^{n} P(X_k \in S) - \pi(S)\right|.$$

Thus

$$\sup_S |F_n(S) - \pi(S)| = \sup_S \left|\frac{1}{n}\sum_{k=1}^{n} P(X_k \in S) - \pi(S)\right| = \left\|\frac{1}{n}\sum_{k=1}^{n} P(X_k \in \cdot) - \pi(\cdot)\right\|,$$

by definition of total variation distance. Also, by the triangle inequality,

$$\left\|\frac{1}{n}\sum_{k=1}^{n} P(X_k \in \cdot) - \pi(\cdot)\right\| \leq \frac{1}{n}\sum_{1}^{n} \|P(X_k, \cdot) - \pi(\cdot)\|.$$

This completes the proof. □

References

[1] D. J. Aldous and H. Thorisson. Shift-coupling. *Stochastic Processes and their Applications*, 44(1):1–14, 1993.

[2] S. Asmussen, P. W. Glynn, and H. Thorisson. Stationarity detection in the initial transient problem. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 2(2):130–157, 1992.

[3] S. Brooks, A. Gelman, G. Jones, and X. L. Meng, editors. *Handbook of Markov chain Monte Carlo*. CRC press, 2011.

[4] W. Doeblin. Exposé de la théorie des chaınes simples constantes de markov á un nombre fini d'états. *Mathématique de l'Union Interbalkanique*, 2(77–105):78–80, 1938.

[5] C.J. Geyer. Practical Markov chain Monte Carlo. *Statistical science*, 7:473–483, 1992.

[6] W.K. Hastings. Monte Carlo sampling methods using Markov chain Monte Carlo. *Biometrika*, 57:97–109, 1970.

[7] R.B. Lund, S.P. Meyn, and R.L. Tweedie. Computable exponential convergence rates for stochastically ordered Markov processes. *Ann. Appl. Probab.*, 6(1):218–237, 1996.

[8] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953.

[9] S.P. Meyn and R.L. Tweedie. *Markov Chains and Stochastic Stability*. Springer Science & Business Media, 2012.

[10] G.O. Roberts and J.S. Rosenthal. Shift-coupling and convergence rates of ergodic averages. *Stochastic Models*, 13(1):147–165, 1997.

[11] G.O. Roberts and J.S. Rosenthal. General state space Markov chains and MCMC algorithms. *Probability Surveys*, 1:20–71, 2004.

[12] G.O. Roberts and R.L. Tweedie. Rates of convergence of stochastically monotone and continuous time Markov models. *Journal of Applied Probability*, 37(2):359–373, 2000.

[13] J.S. Rosenthal. Minorization conditions and convergence rates for Markov chain Monte Carlo. *Journal of the American Statistical Association*, 90:558—-566, 1995.

[14] J.S. Rosenthal. *A First Look at Stochastic Processes*. World Scientific Publishing Co., 2019.

[15] J.S. Rosenthal. Point process MCMC JavaScript simulation. 2020. Available at: probability.ca/pointproc.