YET ANOTHER QUANTITATIVE HARRIS THEOREM

CHRISTOPHER DUPRE ⓘ

ABSTRACT. In this paper we develop a quantitative Harris theorem with effective control over the constants. A benefit of our methodology is the decoupling of the small set and Lyapunov-Foster Drift conditions. Our methodology allows any small set and any set in the Lyapunov-Foster condition as long as the second satisfies a so-called "quantitative petiteness" condition. The theorem relies on a novel proof of a quantitative Kendall-type theorem which is inspired by the techniques of Markov Chains on general state spaces.

1. INTRODUCTION

In this paper, we will present a novel proof of a quantitative mixing theorem for Markov chains on general state spaces. Such limiting statements for the evolution of Markov chains on general state spaces are collectively referred to as Harris Theorems in honor of the pioneering work of T.E. Harris in the fifties [1]. We follow the work of Meyn and Tweedie in [2] by extending the ideas of Doeblin in the creation of a "small" set to a general state space. This is then combined with a technique known as Numellin splitting [3] to reduce the problem to a statement of renewal theory. Our particular approach relies on mixing with respect to a weigthed norm. This technique was introduced in [4].

In the mixing case, this approach relies on a key theorem pioneered by Kendall [5]. Most proofs in the literature seem to rely on techniques from complex analysis [2, 6]. While these proofs are beautiful in their own right, we present a new proof of a major implication in Kendall's theorem which we need for our estimates. The proof relies on techniques generated for general state space Markov chains and a coupling approach. As far as we can tell, this seems to be the first such coupling proof in the literature.

This technique has been successful in the creation of both qualitative [2, 7] and quantitative[6, 8, 9] estimates on mixing rates. There have also been several impressive applications of these ideas to rigorous or "honest" error bounds for Markov Chain Monte Carlo algorithms [10, 11, 12].

The motivation for this work came from considering uniform mixing rates a parametric family of random diffeomorphisms on a compact state space as an extension of [13]. In our case, the alignment of a small set and the sub-level set was unnatural and so we created these tools. We hope they may be of independent interest.

2. SETTING AND MAIN THEOREM

In this paper, we will be considering discrete-time Markov Chains on general state spaces. By this we mean that we have a measurable space $(X, \mathcal{B})$ and a kernel $P(x, \cdot)$ defined such that

(1) For every $x \in X$, $P(x, \cdot)$ is a probability measure

(2) For every $A \in \mathcal{B}$, $P(x, A)$ is a non-negative measurable function.

This kernel encodes the one-step transition probabilities of our Markov process. The n-step transition probabilities are then defined inductively via the Chapman-Kolmogorov equations.

$$P^n(x, \cdot) := \int_X P^{n-1}(x, dy) P(y, \cdot).$$

We can define an action of our kernel on measurable functions f and measures λ as follows.

$$P^n(\lambda, \cdot) := \lambda P^n(\cdot) := \int_X \lambda(dy) P^N(y, \cdot) \qquad P^n(x, f) := P^n f(x) := \int_X P^N(x, dy) f(y).$$

Given any initial measure λ, we then define the Markov chain $\{\Phi^n\}_{n \in \mathbb{N}}$ as the stochastic process such that for every increasing sequence of indices $i_1 < i_2 < \ldots i_n$ we have the following condition on the law.

$$\mathbb{P}\left(\Phi^{i_1} \in A_1, \ldots \Phi^{i_n} \in A_n\right) = \int_{A_1} \int_{A_2} \ldots \int_{A_{n-1}} P^{i_1}(\lambda, dy) P^{i_2 - i_1}(y, dz) \ldots P^{i_n - i_{n-1}}(w, A_n).$$

We will use the notation $\mathbb{P}_x$ to refer the law induced by the Markov chain with initial measure δ_x. We will use $\mathbb{E}_x$ to denote integrating with respect to this law. For any $C \in \mathcal{B}$, let τ_C denote the first hitting time.

Our goal will be to obtain quantitative estimates on the mixing rate for our Markov chain. If $f : X \to \mathbb{R}$ is a measurable function on our space, we will use the notation of $f(\Phi^n)$ to mean the composition of the n-th step in our Markov chain and the function f. An important construction of our work will be the so called V-weighted norm. For every measurable $V : X \to [1, \infty)$, we define a norm on measurable functions from X as follows.

$$\|\phi\|_V := \sup_{x \in X} \frac{|\phi(x)|}{V(x)}.$$

In particular, every bounded measurable function is also $\|\cdot\|_V$ bounded.

Theorem 2.1 (Quantitative Harris Theorem, [2]). *Let P be a Markov transition kernel defined on a measurable space $(X, \mathcal{B})$ and assume the following:*

(a) (Small set) There exists a measurable set U, a $\delta > 0$, and a probability measure μ on X such that $\mu(U) = 1$ and

$$P(x, B) \geq \delta \mu(B) \text{ for all } x \in U, B \in \mathcal{B}.$$

(b) (Drift condition) There exists a measurable function $V : X \to [1, \infty)$, a constant $0 < \lambda < 1$, a constant $b < \infty$, and a set $C \in \mathcal{B}$ such that

$$PV \leq \lambda V + b \mathbb{1}_C.$$

(c) (Quantitative Petiteness) There exists an $N_0 \in \mathbb{N}$ and a $c > 0$ such that

$$\inf_{x \in C} \mathbb{P}_x \{\tau_U \leq N_0\} \geq c.$$

Let $M_U = \sup\limits_{x \in U} V(x)$ and let $M_C = \sup\limits_{x \in C} V(x)$. Then there exists an invariant measure π and a pair of effective constants $D < \infty, 0 < \gamma < 1$ which depend only (and explicitly) on $\delta, \lambda, b, N_0, c, M_U, M_C$ such that

$$\left| P^n \phi(x) - \int \phi d\pi \right| \leq DV(x) \|\phi\|_V \gamma^n.$$

Note that several qualifications for this theorem are unnecessary. U can be a small set for the n-step chain for any n. The measure μ need not give U a positive mass, but only needs to satisfy an aperiodicity requirement. In general, the proof can be expanded to the reader's exhaustion. We take the above requirements for the sake of clarity.

A key technique in the study of ergodicity in general state Markov chains is the reduction to a problem in renewal theory [2]. As such, we have developed a new proof of a key result related to Kendall's theorem. While Kendall's theorem is typically richer in structure than our result, this result will be all we need for our purposes.

Theorem 2.2. *Let $p(n)$ be an increment distribution such that $p(1) \geq \beta > 0$ and let $u(n)$ be the associated renewal sequence with 0 delay. If there exists a $r > 1$ such that*

$$\sum_{n=1}^{\infty} p(n) r^n \leq B$$

then there exists an invariant distribution for the forward recurrence time chain $e(n)$, a constant $\pi(1) = e * u(n)$, and an effective bound ρ depending only (and explicitly) on r, β, B such that for any $r_2 \leq \rho$ we can find an effective bound L depending only (and explicitly) on r_2, B, β, ρ such that

$$\sum_{n=1}^{\infty} \left| u(n) - \pi(1) \right| r_2^n \leq L.$$

This is again restricted by clarity instead of necessity. In fact the condition $p(n) > 0$ and $p(m) > 0$ for n, m co-prime is sufficient. This is clear from our proof and does not essentially affect finding the necessary quantitative bounds. Our proof appears to be novel in that it appears to be the first coupling proof of this implication of Kendall's theorem in the literature.

Appendix A. Main Tools

In this section we introduce the main tools which we will use in the statement of both theorems. The first tool gives explicit bounds on the moment generating function of r^{τ_C} where C is the set in our Lyapunov Foster drift condition. The result is effectively a consequence of Dynkin's Formula.

Theorem 2.3 (Theorem 15.2.5 in [2]). *If $V : X \to [1, \infty)$ satisfies the Lyapunov-Foster Drift condition*

$$PV \leq \lambda V + b \mathbb{1}_C$$

for $\lambda < 1, b < \infty, C \in \mathcal{B}$, then for any $r \in \left(1, \frac{1}{\lambda}\right)$

$$\frac{\mathbb{E}_x[r^{\tau_C}] - 1}{r - 1} \leq \mathbb{E}_x \left[\sum_{k=0}^{\tau_C - 1} V(\Phi^k) r^k \right] \leq \frac{1 + rb}{1 - \lambda r} V(x).$$

This allows us to use bounds on the function in our Lyapunov-Foster drift condition to obtain explicit bounds on the first hitting times to the set in our drift condition. However, we will need to transfer such bounds from set to set. For this, we will use the following result which follows from a close reading of Theorem 15.2.1 in [2].

Corollary 2.4. *Suppose there exists an $r_1 > 1$ such that*

$$\sup_{x \in C} \mathbb{E}_x \left[\sum_{k=0}^{\tau_C - 1} V(\Phi^k) r_1^k \right] \leq M_0 < \infty.$$

Further, assume that for $B \in \mathcal{B}$ there exists an $N_0 \geq 1$ and a $c > 0$ such that

$$\inf_{x \in C} \mathbb{P}_x \left\{ \tau_B \leq N_0 \right\} \geq c.$$

Then, there exists an effective ρ depending only (and explicitly) on r_1, M_0, c, N_0 such that for all $r_2 < \rho$ we can find an effective constant D depending only (and explicitly) on r_1, M_0, c, N_0, r_2 such that

$$\frac{\mathbb{E}_x \left[r_2^{\tau_B} \right] - 1}{r_2 - 1} \leq \mathbb{E}_x \left[\sum_{k=0}^{\tau_B - 1} V(\Phi^k) r_2^k \right] \leq D V(x)$$

Appendix B. Proof of Theorem 2.2

We begin by introducing the language of renewal theory. Let $p(n) : \mathbb{N} \to \mathbb{R}_{\geq 0}$ and $d(n) : \mathbb{N} \cup \{0\} \to \mathbb{R}_{\geq 0}$ be sequences of non-negative numbers such that

$$\sum_{n=1}^{\infty} p(n) = \sum_{n=0}^{\infty} d(n) = 1.$$

We will refer to $p(n)$ as the increment distribution and $d(n)$ as the delay distribution.

We let $\{Y_i\}_{i=1}^{\infty}$ be independent copies of discrete random variables with the law $\mathbb{P}\{Y_i = n\} = p(n)$ and let Y_0 denote a random variable with values in $\mathbb{N} \cup \{0\}$ with law $d(n)$ which is independent of all other Y_i. We then define $S_n^d = \sum_{i=0}^{n} Y_i$ and

$$V_n^d = \begin{cases} 1 & S_m^d = n \text{ for some m} \\ 0 & \text{else} \end{cases}.$$

The associated renewal process is

$$v_d(n) := \mathbb{P}\left\{ V_n^d = 1 \right\} = \sum_{m=0}^{\infty} d * p^{*m}(n).$$

Of special importance are the renewal sequences with $d(0) = 1$. We denote such a sequence with the symbol $u(n) := v_0(n)$. We can the write a renewal sequence with any delay in terms of the zero delay sequence as $v_d(n) = d * u(n)$. For more details into renewal sequences we refer the reader to [14].

Notice that $\sum_{n=1}^{\infty} p(n)r^n < \infty$ implies that $\sum_{n=1}^{\infty} p(n)n < \infty$ as in particular we have that $\frac{r}{r-1}r^n \geq n$ for all $r > 1$. Let $e(n)$ denote a delay distribution with the law

$$e(n) := \frac{\sum_{j=n+1}^{\infty} p(j)}{\sum_{n=1}^{\infty} p(n)n} = \frac{1 - \sum_{j=1}^{n} p(j)}{\sum_{n=1}^{\infty} p(n)n}, n \in \mathbb{N} \cup \{0\}.$$

This can be easily seen to be the invariant probability measure of the forward recurrence time chain (see [2, 14] for more details). In particular, we have that

$$\pi(1) := \mathbb{P}\{V_n^e = 1\} = e * u(n) = \frac{1}{\sum_{n=1}^{\infty} np(n)}$$

By definition of the renewal sequences, we have the reduction that

$$\left|u(n) - \pi(1)\right| = \left|\mathbb{P}\left\{V^0(n) = 1\right\} - \mathbb{P}\left\{V^e(n) = 1\right\}\right|.$$

To control this difference, we introduce a coupling.

Definition 2.5. The **coupling time** of $V^0(n), V^e(n)$ is defined as the stopping time

$$T_{0,e} = \min\left\{n \in \mathbb{N} \cup \{0\} \,\middle|\, V^0(n) = V^e(n) = 1\right\}.$$

The coupling time represents the first simultaneous renewal of the two sequences. Note that after a renewal, the distribution of the next renewal time is uniquely determined by the increment sequence. Thus we will see that the coupling time allows us to control the difference between the two sequences after any given time.

To formalize this, let us define the random variables

$$V^{0,e}(n) = \begin{cases} V^0(n) & n < T_{0,e} \\ V^e(n) & n \geq T_{0,e}. \end{cases}$$

Notice that for any n (due to the distributions being identical after the first simultaneous renewal) we have that

$$\mathbb{P}\left(V^{0,e}(n) = 1\right) = \mathbb{P}\left(V^0(n) = 1\right).$$

We then have that

$$\left|u(n) - \pi(1)\right| = \left|\mathbb{P}\left\{V^0(n) = 1\right\} - \mathbb{P}\left\{V^e(n) = 1\right\}\right| = \left|\mathbb{P}\left\{V^{0,e}(n) = 1\right\} - \mathbb{P}\left\{V^e(n) = 1\right\}\right|$$

$$= \left|\mathbb{P}\left\{V^0(n) = 1, T_{0,e} > n\right\} + \mathbb{P}\left\{V^e(n) = 1, T_{0,e} \leq n\right\}\right.$$

$$\left. - \mathbb{P}\left\{V^e(n) = 1, T_{0,e} > n\right\} - \mathbb{P}\left\{V^e(n) = 1, T_{0,e} \leq n\right\}\right|$$

$$\leq \mathbb{P}\{T_{0,e} > n\}.$$

Thus, multiplying by the appropriate power of r to be chosen later and summing, we have that

$$(2.1) \qquad \sum_{n=1}^{\infty} \left|u(n) - \pi(1)\right| r^n \leq \sum_{n=1}^{\infty} r^n \mathbb{P}\{T_{0,e} > n\} \leq \frac{r\mathbb{E}[r^{T_{0,e}}]}{r - 1} = \frac{r}{r - 1} \sum_{n=1}^{\infty} e(n)\mathbb{E}[r^{T_{0,n}}].$$

Thus it suffices for us to obtain geometrically bounded tails for the first coupling time of a zero delay distribution and a constant delay distribution.

Inspired by methods in the theory of Markov chains on general state spaces (such as Theorem 2.3 and Corollary 2.4), we note that the boundedness of the moment generating function of a return time is related to the existence of a Lypaunov function. Thus we would like to cast our problem as a Markov chain on some state space, and then find an appropriate Lyapunov function.

We first consider the so-called forward recurrence time chain. This is a Markov chain defined on the state space $\mathbb{N}$ with the discrete σ-algebra with the transition kernel

$$P(n, n-1) = 1 \text{ if } n > 1, P(1, m) = p(m).$$

Considering the evolution of our two renewal sequences independently, we can then imagine the state of our system a a point in $\mathbb{N} \times \mathbb{N}$. Initial delays of $0, n$ corresponds to an initial point of $(1, n+1)$. The system then induces a Markov chain on $\mathbb{N} \times \mathbb{N}$ with kernel defined as follows

$$P\big((a,b),(a-1,b-1)\big) = 1 \text{ if } a, b \neq 1$$
$$P\big((1,b),(n,b-1)\big) = p(n) \text{ if } a = 1, b \neq 1$$
$$P\big((a,1),(a-1,n)\big) = p(n) \text{ if } a \neq 1, b = 1$$
$$P\big((1,1),(n,m)\big) = p(n)p(m) \text{ if } a, b = 1.$$

A simultaneous renewal is equivalent to this Markov chain reaching $(1,1)$. Thus we have that $\mathbb{E}[r^{T_{0,n}}] = \mathbb{E}_{1,n+1}[r^{\tau_{1,1}}]$.

Now consider the function $V(a,b) = \frac{r^{a-1}+r^{b-1}}{2}$. By direct computation, we have the following relations.

$$P\big((a,b),V\big) = \frac{1}{r}V(a,b) \qquad\qquad\qquad \text{if } a, b \neq 1$$

$$P\big((1,b),V\big) = \frac{1}{2r}\left(\sum_{n=1}^{\infty} p(n)r^n + r^{b-1}\right) \qquad a = 1, b \neq 1$$

$$P\big((a,1),V\big) = \frac{1}{2r}\left(r^{a-1} + \sum_{n=1}^{\infty} p(n)r^n\right) \qquad a = 1, b \neq 1$$

$$P\big((1,1),V\big) = \frac{1}{r}\sum_{n=1}^{\infty} p(n)r^n \qquad\qquad\quad a = 1, b \neq 1.$$

We have that $\sum_{n=1}^{\infty} p(n)r^n < \infty$ so at the very least each of these is well-defined. We wish the set C that we will construct in the Lyapunov function to be finite for reasons we shall get to about controlling access from our set C to $(1,1)$. Thus, the point $(1,1)$ is of no concern but each of the rays $(1,n), (n,1)$ are seemingly problematic. However, we have that

$$\frac{P\big((a,1),V\big)}{V(a,1)} = \frac{r^{a-1} + \sum_{n=1}^{\infty} p(n)r^n}{r^a + r}.$$

Using $\sum_{n=1}^{\infty} p(n)r^n \geq r$ for all $r > 1$, we can rearrange the condition that this fraction is less than $\frac{1}{r} < \eta < 1$ as follows.

$$\frac{P\left((a,1),V\right)}{V(a,1)} \leq \eta \iff \frac{r^{a-1} + \sum_{n=1}^{\infty} p(n)r^n}{r^a + r} \leq \eta$$

$$\iff \sum_{n=1}^{\infty} p(n)r^n - \eta r \leq r^{a-1}\left(\eta r - 1\right)$$

$$\iff a \geq \frac{1}{\ln(r)} \ln\left(\frac{\sum_{n=1}^{\infty} p(n)r^n - \eta r}{\eta r - 1}\right) + 1.$$

The same calculation applies in the situation when $a = 1, b \neq 0$. Thus we have the Lyapunov-Foster Drift condition as follows.

$$M := \left\lceil \frac{1}{\ln(r)} \ln\left(\frac{\sum_{n=1}^{\infty} p(n)r^n - \eta r}{\eta r - 1}\right) + 1 \right\rceil$$

$$C := \{(1,b) \in \mathbb{N}^2 \text{ such that } b \leq M\} \cup \{(a,1) \in \mathbb{N}^2 \text{ such that } a \leq M\}$$

$$PV \leq \eta V + \frac{\max\left\{2\sum_{n=1}^{\infty} p(n)r^n, \sum_{n=1}^{\infty} p(n)r^n + r^{M-1}\right\}}{2r} \mathbb{1}_C.$$

In particular, using Theorem (2.3) we have the following bound.

$$\sup_{x \in C} \mathbb{E}_x\left[\sum_{k=0}^{\tau_C - 1} V(\Phi^k)r^k\right] \leq \frac{2 + \max\left\{2\sum_{n=1}^{\infty} p(n)r^n, \sum_{n=1}^{\infty} p(n)r^n + r^{M-1}\right\}}{4(1 - \eta r)}(r^{M-1} + 1).$$

We now wish to transfer this bound from a bound on τ_C to a bound on $\tau_{1,1}$. From the construction of our Markov chain, it is clear that $\inf_{x \in C} \mathbb{P}_x\{\tau_{1,1} \leq M\} \geq p(1)^M$. This bound will act as a "quantitative petiteness" condition as in the statement of our Harris theorem which allows us to transfer bounds on τ_C to $\tau_{1,1}$ due to Corollary 2.4. We thus now have an effective constant D and an effective new r with a bound

$$\mathbb{E}[r^{T_{0,n}}] = \mathbb{E}_{1,n+1}[r^{\tau_{1,1}}] \leq DV(1, n+1) = \frac{D(1 + r^n)}{2} \leq \frac{D}{2}\left(\frac{1}{r-1} + 1\right)r^n = \frac{Dr}{2(r-1)}r^n.$$

We have used the fact that $\frac{r^n}{r-1} > 1$ for all $n \geq 1$ for all $r > 1$. Summing these up then gives the following.

$$\mathbb{E}[r^{T_{0,e}}] = \sum_{n=0}^{\infty} e(n)\mathbb{E}[r^{T_{0,n}}] \leq \frac{Dr}{2(r-1)}\sum_{n=0}^{\infty} e(n)r^n = \frac{Dr}{2(r-1)} \frac{1}{\sum_{n=1}^{\infty} np(n)}\left[\sum_{n=0}^{\infty}\sum_{j=n+1}^{\infty} p(j)r^n\right]$$

$$\leq \frac{Dr}{2(r-1)}\left[\sum_{j=1}^{\infty}\sum_{n=0}^{j-1} p(j)r^n\right] \leq \frac{Dr}{2(r-1)^2}\left[\sum_{j=1}^{\infty} p(j)r^j\right].$$

Plugging this back into (2.1) gives the result.

Appendix C. Proof of Theorem 2.1

First notice that by scaling it is sufficient to consider ϕ such that $\|\phi\|_V \leq 1$. We adopt a coupling based approach. We wish to control two separate quantities

(1) Geometric tails on τ_U
(2) The construction of a so called "atom" which can translate our problem into a renewal theory problem.

Our main tools in Theorem 2.3 and Corollary 2.4 give an effective way to achieve the first result. Combining M_U, M_C, c, N_0 in the theorem, we get geometric tails for the first hitting time for U from an arbitrary initial point.

We now return to the second of our goals. We recall the definition of an atom.

Definition 2.6. Let $P(x, dy)$ be a transition kernel on a measurable space $(X, \mathcal{B})$. A set $\alpha \in \mathcal{B}$ is called an **atom** if there exists a probability measure ν such that that

$$P(x, B) = \nu(B) \text{ for all } x \in \alpha.$$

We will reserve the symbol α for an atom. When α is an atom, we will write $P(\alpha, B)$ to mean $P(x, B)$ for any $x \in \alpha$ or equivalently $\nu(B)$. If a stopping time refers to an atom, we are referring to a Markov chain with an atom. We will construct such a chain in the next paragraph using the technique of Nummelin Splitting [3]. We will then transfer the geometric tails to the newly created atom.

We then split the chain at U with a $\delta/2$ splitting. What we mean by this is that we consider the extended space $X \times \{0, 1\}$ equipped with the product σ-algebra. We will use the notation that $A_0 = A \times \{0\}, A_1 = A \times \{1\}$ $A \in \mathcal{B}$, $x_i = x \times \{i\}$ where $x \in X$. Given any measure λ on $(X, \mathcal{B})$, we will construct a measure on this new space as

$$\lambda^*(A_0) = \left(1 - \frac{\delta}{2}\right) \lambda(A \cap U) + \lambda(A \cap U^c)$$

$$\lambda^*(A_1) = \frac{\delta}{2}\lambda(A \cap U).$$

Likewise, given any function $V : X \to \mathbb{R}$, we can define the function $\hat{V} : X \times \{0, 1\} \to \mathbb{R}$ on the split chain as $\hat{V}(x_i) := V(x)$. We now are able to extend measures into this new space, so now we turn to extending transition probabilities. In this way, we will be able to construct a Markov Chain on this new space from our original chain.

We define the split kernel $\hat{P}(x_i, A)$ for $x_i = (x, i) \in X \times \{0, 1\}$ and $A \in \mathcal{B} \times \{0, 1\}$ as follows.

$$\hat{P}(x_i, A) = \begin{cases} P^*(x, A) & x_i \in X_0 \backslash U_0 \\ \frac{2P^*(x,A) - \delta\mu^*(A)}{2 - \delta} & x_i \in U_0 \\ \mu^*(A) & x_i \in X_1. \end{cases}$$

An absolutely vital aspect of this construction is that our original chain is exactly the marginal of this constructed chain and thus $\tau_{A_1 \cup A_0}$ for the split chain is exactly equal to τ_A for the original chain. Another vital aspect is that U_1 now forms an atom for this new Markov chain. For more

information on the splitting method in the context of Markov Chains, we refer the reader to Chapter 5 of [2].

Another important aspect of the construction is the following estimate.

$$x_i \in U_1 \implies \hat{P}(x_i, U_1) = \mu^*(U_1) = \frac{\delta}{2}$$

$$x_i \in U_0 \implies \hat{P}(x_i, U_1) = \frac{2P^*(x, U_1) - \delta\mu^*(U_1)}{2 - \delta}$$

$$= \frac{\delta\left(P(x, U) - \frac{\delta}{2}\mu(U)\right)}{2 - \delta} \geq \frac{\delta}{2 - \delta}\frac{\delta}{2}.$$

In either case, we have $\mathbb{P}_{x_i}\left(\tau_{U_1} = 1\right) \geq \frac{\delta}{2-\delta}\frac{\delta}{2}$ (note this is always less than or equal to $\frac{\delta}{2}$ as necessarily $0 < \delta \leq 1$).

Now we have quantitative control from $U_0 \cup U_1$ to U_1 and so we can again apply Corollary 2.4 to show that U_1 is an atom with $\mathbb{E}_x\left[r^{\tau_{U_1}}\right] < \infty$ for all $x \in X_0 \cup X_1$ and bounded on U_1 for some effective constant r with effective bounds. The goal is to relate this control of geometric tails on the first occupation time for an atom to a control over the V-weighted distance. The key tool for this part of the proof will be the so-called "regenerative" or "first entrance-last exit" decomposition.

Definition 2.7. Given a transition kernel $P(x, dy)$ generating a Markov chain $\{\Phi^n\}_{n=1}^\infty$ on a measurable space $(X, \mathcal{B})$ and a $B \in \mathcal{B}$, the **taboo probability** for the set is the probability of reaching another set while avoiding B. More formally, it is defined as

$$_BP^N(x, A) = \mathbb{P}_x\left\{\Phi^n \in A, \tau_B \geq n\right\}.$$

This can also be expressed inductively according the the following relations.

$$_BP(x, A) = P(x, A)$$

$$_BP^{N+1}(x, A) = \int_{B^c} {_BP^{N+1}}(x, dy)P(y, A).$$

Definition 2.8. Given a transition kernel $P(x, dy)$ on a measurable space $(X, \mathcal{B})$ and a $B \in \mathcal{B}$, we define the **regenerative decomposition** of the kernel as follows

$$P^n(x, A) = {_BP^n}(x, C) + \sum_{j=1}^{n-1}\int_B\left[\sum_{k=1}^j\int_B {_BP^k}(x, dv)P^{j-k}(v, dw)\right] {_BP^{n-j}}(w, A).$$

Intuitively, the regenerative decomposition tracks the first entrance and last exit from a given set. Every trajectory can be divided into never meeting the set in n steps or meeting in less than n steps. These trajectories can again be divided by how long until they have their last visit to B before leaving and never returning within the first n iterates. Notice that the last exit time in general is not a stopping time. The fact that we only consider the first n steps at a time is therefore crucial.

If the set $B = \alpha$ is an atom and we are considering the Markov chain that we have defined on the split chain, this simplifies from a series of integrals into a sum as follows.

$$\hat{P}^n(x, C) = {}_\alpha\hat{P}^n(x, C) + \sum_{j=1}^{n-1} \sum_{k=1}^{j} {}_\alpha\hat{P}^k(x, \alpha)\hat{P}^{j-k}(\alpha, \alpha) {}_\alpha\hat{P}^{n-j}(\alpha, C)$$

We can then consider the difference with a fixed measure (say the invariant measure). We must first verify that the invariant measure exists. As we already know that $\mathbb{E}_\alpha[r_\alpha^\tau]$ is bounded, we have that $\mathbb{E}_\alpha[\tau_\alpha]$ is also finite. Thus by Theorem 10.2.1 in [2], there exists an invariant measure for the split chain. A key additional fact is that $\hat{\pi}(\alpha) = \mathbb{E}_\alpha[\tau_\alpha]^{-1}$.

From a direct application of Theorem 10.4.9 in [2], we have that

$$\pi(g) = \pi(\alpha)\mathbb{E}_\alpha\left[\sum_{k=0}^{\tau_\alpha-1} g\left(\hat{\Phi}^k\right)\right] = \pi(\alpha)\sum_{k=1}^{\infty}\mathbb{E}_\alpha\left[g\left(\hat{\Phi}^k\right)\mathbb{1}_{\tau_\alpha \geq k}\right] = \pi(\alpha)\sum_{k=1}^{\infty} {}_\alpha\hat{P}^k(\alpha, g).$$

Based on these two decompositions, we introduce three sequences of value.

$$a_x(n) := \mathbb{P}_x\{\tau_\alpha = n\} = {}_\alpha\hat{P}^n(x, \alpha)$$

$$u(n) := \mathbb{P}_\alpha\left\{\hat{\Phi}^n \in \alpha\right\} = \hat{P}^n(\alpha, \alpha)$$

$$t_g(n) := \mathbb{E}_\alpha\left[g\left(\hat{\Phi}^n\right)\mathbb{1}_{\tau_\alpha \geq n}\right] = {}_\alpha\hat{P}^n(\alpha, g)$$

Combining these terms, we have the following bound.

$$\hat{P}^n(x, g) = {}_\alpha\hat{P}^n(x, g) + a_x * u * t_g(n)$$

$$\hat{\pi}(g) = \hat{\pi}(\alpha)\sum_{k=1}^{\infty} {}_\alpha\hat{P}^k(\alpha, g) = \hat{\pi}(\alpha) * t_g(n) + \hat{\pi}(\alpha)\sum_{k=n+1}^{\infty} {}_\alpha\hat{P}^k(\alpha, g)$$

We have used a small abuse of notation by saying that $\hat{\pi}(\alpha) * t_g(n)$ is the convolution of $t_g(n)$ with the constant sequence of value $\hat{\pi}(\alpha)$. We then obtain the following useful bound.

$$\left|\hat{P}^n(x, g) - \hat{\pi}(g)\right| \leq {}_\alpha P^n(x, g) + \left|a_x * u - \hat{\pi}(\alpha)\right| * t_g(n) + \hat{\pi}(\alpha)\sum_{k=n+1}^{\infty} {}_\alpha P^k(\alpha, g).$$

Multiplying by r^n and summing, this gives the following.

$$\sum_{n=1}^{\infty}\left|\hat{P}^n(x, g) - \hat{\pi}(g)\right| r^n \leq \sum_{n=1}^{\infty} {}_\alpha\hat{P}^n(x, g)r^n + \sum_{n=1}^{\infty}\left|a_x * u - \hat{\pi}(\alpha)\right| * t_g(n)r^n$$

$$+ \hat{\pi}(\alpha)\sum_{n=1}^{\infty}\sum_{k=n+1}^{\infty} {}_\alpha\hat{P}^k(\alpha, g)r^n.$$

Our goal is to now bound these terms. Following the bounds in pages 360-361 of [2], we achieve the following for all $\|g\|_{\hat{V}} \leq 1$.

$$\sum_{n=1}^{\infty} \left| \hat{P}^n(x,g) - \hat{\pi}(g) \right| r^n \leq \mathbb{E}_x \left[\sum_{n=1}^{\tau_\alpha} \hat{V}(\Phi^n) r^n \right] + \frac{1}{r-1} \mathbb{E}_\alpha \left[\sum_{n=1}^{\tau_\alpha} \hat{V}(\Phi^n) r^n \right]$$

$$+ \mathbb{E}_\alpha[r^{\tau_\alpha}] \left[\frac{r}{r-1} + \sum_{n=1}^{\infty} \left| u(n) - \hat{\pi}(\alpha) \right| r^n \right]$$

Our goal will now be to bound these terms.

First notice that

$$\mathbb{E}_x \left[\sum_{n=1}^{\tau_\alpha} \hat{V}(\Phi^n) r^n \right] \leq \sup_{x \in \alpha} \hat{V}(x) \mathbb{E}_x[r^{\tau_\alpha}] + \mathbb{E}_x \left[\sum_{n=0}^{\tau_\alpha - 1} \hat{V}(\Phi^n) r^n \right]$$

$$\mathbb{E}_\alpha \left[\sum_{n=1}^{\tau_\alpha} \hat{V}(\Phi^n) r^n \right] \leq \sup_{x \in \alpha} \hat{V}(x) \mathbb{E}_\alpha[r^{\tau_\alpha}] + \mathbb{E}_\alpha \left[\sum_{n=0}^{\tau_\alpha - 1} \hat{V}(\Phi^n) r^n \right]$$

By definition,
$$\sup_{x \in \alpha} \hat{V}(x) = \sup_{x \in U_1} \hat{V}(x) = \sup_{x \in U} V(x) = M_U.$$

The other terms are then bounded by effective constants multiplied by $V(x)$.

We now turn to the last term. By construction, we have that $P(\alpha, \alpha) = p(1) \geq \frac{\delta}{2}$ and that $\mathbb{E}_\alpha[r^{\tau_\alpha}] < \infty$. Comparing forms directly, we also see that $\pi(1) = \hat{\pi}(\alpha)$. Thus by Theorem 2.2, we have that there exists an effective constant $D < \infty$ such that $\sum_{n=1}^{\infty} |u(n) - \hat{\pi}(\alpha)| r^n \leq D$. Again, we can bound $\mathbb{E}[r^{\tau_\alpha}]$ by an effective constant. Since $V(x) \geq 1$, we can multiply all relevant terms by it without loss to obtain an effective constant D with the following bound for all g such that $\|g\|_{\hat{V}} \leq 1$.

$$\sum_{n=1}^{\infty} \left| \hat{P}^n(x,g) - \hat{\pi}(g) \right| r^n \leq DV(x)$$

$$\sup_{g:\|g\|_{\hat{V}} \leq 1} \left| \hat{P}^n(x,g) - \hat{\pi}(g) \right| \leq DV(x) r^{-n}.$$

Using the fact that our original Markov chain is the marginal of this chain, we have the final bound exactly as in page 327 of [2]. π here indicates the marginal of $\hat{\pi}$. This is an invariant measure of the original Markov process by Theorem 10.4.1 in [2]. This then gives the theorem.

$$\sup_{g:\|g\|_V \leq 1} \left| P^n(x,g) - \pi(g) \right| \leq \sup_{g:\|g\|_{\hat{V}} \leq 1} \left| \hat{P}^n(x,g) - \hat{\pi}(g) \right| \leq DV(x) r^{-n}.$$

References

[1] Theodore E Harris. "The existence of stationary measures for certain Markov processes". In: *Matematika* 4.1 (1960), pp. 131–143.

[2]   Sean P Meyn and Richard L Tweedie. *Markov chains and stochastic stability*. Springer Science & Business Media, 2012.

[3]   Esa Nummelin. "A splitting technique for Harris recurrent Markov chains". In: *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 43.4 (1978), pp. 309–318.

[4]   Arie Hordijk and Flora Spieksma. "On ergodicity and recurrence properties of a Markov chain by an application to an open Jackson network". In: *Advances in applied probability* 24.2 (1992), pp. 343–376.

[5]   David G Kendall. "Unitary dilations of Markov transition operators, and the corresponding integral representations for transition-probability matrices". In: *Probability and statistics* (1959), pp. 139–161.

[6]   Peter H Baxendale. "Renewal theory and computable convergence rates for geometrically ergodic Markov chains". In: *The Annals of Applied Probability* 15.1B (2005), pp. 700–738.

[7]   Jeffrey S Rosenthal. "Minorization conditions and convergence rates for Markov chain Monte Carlo". In: *Journal of the American Statistical Association* 90.430 (1995), pp. 558–566.

[8]   Witold Bednorz. "The Kendall theorem and its application to the geometric ergodicity of Markov chains". eng. In: *Applicationes Mathematicae* 40.2 (2013), pp. 129–165. URL: http://eudml.org/doc/279924.

[9]   Yu Hang Jiang et al. "The Coupling/Minorization/Drift Approach to Markov Chain Convergence Rates". In: *Notices of the American Mathematical Society* 68.4 (2020).

[10]   Krzysztof Łatuszyński and Wojciech Niemiro. "Rigorous confidence bounds for MCMC under a geometric drift condition". In: *Journal of Complexity* 27.1 (2011), pp. 23–38.

[11]   Galin L Jones and James P Hobert. "Honest exploration of intractable probability distributions via Markov chain Monte Carlo". In: *Statistical Science* (2001), pp. 312–334.

[12]   Luke Tierney. "Markov chains for exploring posterior distributions". In: *the Annals of Statistics* (1994), pp. 1701–1728.

[13]   Alex Blumenthal, Michele Coti Zelati, and Rishabh S Gvalani. "Exponential mixing for random dynamical systems and an example of Pierrehumbert". In: *arXiv preprint arXiv:2204.13651* (2022).

[14]   Torgny Lindvall. *Lectures on the coupling method*. Courier Corporation, 2002.

Department of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332
Email address: cdupre3@gatech.edu