Lyapunov Analysis for Rates of Convergence in Markov Chains and Random-Time State-Dependent Drift

by

Ramiro A. Zurkowski

A thesis submitted to the
Department of Mathematics and Statistics
in conformity with the requirements for
the degree of Master of Science

Queen's University
Kingston, Ontario, Canada
October 2013

Abstract

In this paper we survey approaches to studying the ergodicity of aperiodic and irreducible Markov chains [3], [18], [5], [12], [19]. Various results exist for subgeometric and geometric ergodicity with different techniques. Roberts and Rosenthal [19] show using the Coupling Inequality and Nummelin's Splitting Technique how geometric ergodicity follows from a simple drift condition. Subgeometric ergodicity is characterized with a theorem introduced by Tuominen and Tweedie [22], which show the equivalence of a variety of criteria that imply subgeometric ergodicity. Concave functions and the class of pairs of ultimately nondecreasing functions are used by Douc, Fort, Moulines, and Soulier [18] and Hairer [5] to extend and construct practical criteria that imply subgeometric ergodicity. In all these results petite sets and drift conditions play a crucial role, which allows us to unify these results in a common context and notation. We end by using the known results to show ergodicity when random time drift conditions are satisfied on a set of stopping times. We explore how the rate of ergodicity and the expectation between stopping times relate, motivated by the possible applications in network control and event triggered control systems.

Acknowledgments

I would like to thank my advisors, Tamas Linder and Serdar Yüksel, for their encouragement, support and patience in the long process of revision. Their guidance was instrumental in helping give structure and coherence to this thesis.

Contents

Chapter 1

Introduction

The theory of Markov chains offers an intuitive model for noisy systems with short term feedback, describing the randomness of the future conditioned only on the immediate past. The theory lends itself well to many real life applications as it naturally describes a system with probabilistic updating dynamics based on a limited memory of the past. In this thesis we consider the long term effects of probabilistic updating dynamics by focusing on the asymptotic behavior of the transition probabilities for a discrete time indexed Markov chain.

We collect and present three different approaches to ergodic theory of Markov chains: a discussion of geometric ergodicity by Rosenthal [19] using Coupling Inequalities and Nummelin's Splitting technique [15]; a characterization of subgeometric ergodicity by Meyn and Tweedie [13], and Tuominen and Tweedie [22] in the context of regularity and the decomposition of transition probabilities; and an extension of the applicability of subgeometric ergodicity by Douc, Fort, Moulines and Soulier [18], and Hairer [5] exploiting concavity and a class of paired ultimately increasing functions. Throughout, we use the concepts of drift conditions and petite sets to connect the different styles of the three methods and provide some cohesion of the ideas.

1.1 Definitions and Background

Definition (Convolutions)**.** *The convolution of two functions $f, g : \mathbb{N} \to \mathbb{R}$ is denoted as $f * g : \mathbb{N} \to \mathbb{R}$ and defined as*

$$f * g(n) = \sum_{k=0}^{n} f(k)g(n-k)$$

for all $n \in \mathbb{N}$

Note that if $f, g \geq 0$, $\lim_{n\to\infty} g(n) = g(\infty)$ exists and $\sum_n f(n) < \infty$ then $\lim_{n\to\infty} f * g(n) = \sum_{k=0}^{\infty} f(k)g(\infty)$ follows from the dominated convergence theorem.

1.1.1 Markov chains

A discrete time Markov chain is described by

(a) a sequence of random variables $\{x_t\}_{t\in\mathbb{N}}$ that take values in some complete separable metric space $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$, where $\mathcal{B}(\mathcal{X})$ is the Borel σ-field of $\mathcal{X}$ assumed countably generated (p110 of [12])

(b) a family of transition probabilities $\{P_t\}$ defined on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$, where $P_t(x_t, \cdot)$ is the probability distribution of x_{t+1} given x_t.

(c) The transition probabilities satisfy the Markov Property

$$P(x_{t+1} \in B : x_t, x_{t-1}, ..x_0) = P(x_{t+1} \in B : x_t) = P_t(x_t, B) \quad \text{for all } B \in \mathcal{B}(\mathcal{X}),$$

conditioning on the immediate past is the same as conditioning on the whole past.

The one step transition probabilities act on measures μ on the measure space

$(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ and functions $\phi : \mathcal{X} \to \mathbb{R}$ by

$$(\mu P)(A) = \int_{\mathcal{X}} P(x, A)\mu(dx), \qquad (P\phi)(x) = \int_{\mathcal{X}} \phi(y)P(x, dy)$$

respectively for any $A \in \mathcal{B}(\mathcal{X})$ and any $x \in \mathcal{X}$. A Markov chain is time homogeneous if the transition probabilities are independent of the time index. For a time homogeneous Markov chain the notation normally used for the transition probabilities follows

$$P(x_{t+m} : x_t) = P^m(x_t, x_{t+m})$$

where P^m is the mth step transition probability of the Markov chain and is defined by chaining the one step transition probability measures together. An m-skeleton of $\{x_t\}$ is a Markov chain constructed by sampling $\{x_t\}$ at every mth step; so $\{x_t\}$ has $m-1$ m-skeletons $\{x_{tm+k}\}$, $k = 0, 1, \dots m-1$.

Definition (Invariant Distribution)**.** *An invariant distribution (or measure) is a measure μ such that $\mu P(B) = \mu(B)$ for all $B \in \mathcal{B}(\mathcal{X})$.*

Definition (Irreducibility)**.** *A Markov chain $\{x_t\}$ with transition probabilities P, defined on a measurable space $(X, \mathcal{B}(\mathcal{X}))$, is said to be irreducible if there exists a finite measure ψ on $(X, \mathcal{B}(\mathcal{X}))$ such that*

$$\psi(A) > 0 \Rightarrow \sum_{n=0}^{\infty} P^n(x, A) > 0 \quad \textit{for all } x \in \mathcal{X}$$

More specifically, the Markov chain is called ψ-irreducible.

A maximal irreducibility measure ψ is an irreducibility measure such that for all other irreducibility measures ϕ, we have $\psi(B) = 0 \Rightarrow \phi(B) = 0$ for any $B \in \mathcal{B}(\mathcal{X})$.

We can see from the definition of irreducibility that any measure of the form

$$\int \sum_{n=0}^{\infty} \phi(dx) P^n(x, B) 2^{-n-1} \quad \text{for } B \in \mathcal{B}(\mathcal{X}) \tag{1.1}$$

is a maximal irreducibility measure, where ϕ is any finite irreducibility measure. For an irreducible Markov chain we denote by $\mathcal{B}^+(\mathcal{X})$ the family of sets $A \in \mathcal{B}(\mathcal{X})$ such that $\sum_{n=0}^{\infty} P^n(x, A) > 0$ for all $x \in \mathcal{X}$. Equivalently $\mathcal{B}^+(\mathcal{X}) = \{A \in \mathcal{B}(\mathcal{X}) : \psi(A) > 0\}$ where ψ is a maximal irreducibility measure.

Definition (Aperiodicity)**.** *An irreducible Markov chain $\{x_t\}$ with transition probabilities P is said to be aperiodic if for any fixed $x \in \mathcal{X}$ and $B \in \mathcal{B}^+(\mathcal{X})$ there exists $N(x, B) > 0$ such that*

$$P^n(x, B) > 0 \quad \textit{for all } n \geq N(x, B)$$

A set B where $\{n > 0 : P^n(x, B) > 0\}$ is nonempty is said to be reachable from x, in that there is a positive probability of the Markov chain hitting that set in finite time. We note that $\mathcal{B}^+(\mathcal{X})$ is the collection of sets reachable from any point.

For the rest of this thesis we only deal with time homogeneous Markov chains, and use the notation $\{x_t\}$ for Markov chains and P as the one step transition probability. In the study of Markov chains there are several important types of sets that are connected to the transition probabilities.

Definition (Full and Absorbing Sets)**.** *A set $A \in \mathcal{B}(\mathcal{X})$ is full if $\psi(A^C) = 0$ for a maximal irreducibility measure ψ. A set $A \in \mathcal{B}(\mathcal{X})$ is absorbing if $P(x, A) = 1$ for all $x \in A$.*

In an irreducible Markov chain every absorbing set is full. We use the concept of a full absorbing set to restrict our attention and not have to worry about sets that will never be reached.

Definition (Atom). *A set $\alpha \in \mathcal{B}^+(\mathcal{X})$ is an atom if*

$$P(x, \cdot) = P(y, \cdot)$$

for all $x, y \in \alpha$.

The concept of an atom is extremely important as it gives us some sort of fundamental unit, where all the points of a reachable set act together.

Definition (Small sets). *A set $C \in \mathcal{B}^+(\mathcal{X})$ is (m, ϵ, ν)-small if*

$$P^m(x, B) \geq \epsilon\nu(B) \quad \forall B \in \mathcal{B}(\mathcal{X}), x \in C$$

where $m \in \mathbb{N}$, $\epsilon \in (0, 1)$ and $\nu(\cdot)$ is a positive measure on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$.

Definition (Petite Sets). *A set $C \in \mathcal{B}^+(\mathcal{X})$ is called κ-petite if there is a measure κ on $\mathcal{B}(\mathcal{X})$ and a probability distribution $a(\cdot)$ on $\mathbb{Z}_{\geq 0}$ such that*

$$\sum_{n=0}^{\infty} a(n)P^n(x, B) \geq \kappa(B) \quad \forall B \in \mathcal{B}(\mathcal{X}), x \in C \tag{1.2}$$

The next lemma follows a similar proof as Lemma 5.5.2 of [12] and allows us to assume without loss of generality, that for an irreducible Markov chain, if a set is κ-petite then κ can be replaced by maximal irreducibility measure (or equivalently κ can be assumed maximal).

Lemma 1.1.1. *If an irreducible Markov chain has some set $C \in \mathcal{B}^+(\mathcal{X})$ that is κ-petite for some distribution $a(\cdot)$, then C is ψ-petite for the distribution $a * f(n)$ where $f(n) = 2^{-n-1}$ and ψ is a maximal irreducibility measure.*

Proof. Since $a(\cdot)$ and $f(\cdot)$ are proper probability distributions on $\mathbb{N}$, $a * f$ is a proper distribution on $\mathbb{N}$. We then have for all $B \in \mathcal{B}^+(\mathcal{X})$

$$\begin{aligned}
\sum_{n=0}^{\infty} P^n(x,B) a * f(n) &= \sum_{n=0}^{\infty} P^n(x,B) \sum_{j=0}^{n} 2^{-j-1} a(n-j) \\
&= \sum_{n=0}^{\infty} \sum_{j=0}^{n} \int P^j(x,dy) a(j) P^{n-j}(y,B) 2^{j-n-1} \\
&= \int \sum_{j=0}^{\infty} P^j(x,dy) a(j) \sum_{n=0}^{\infty} P^n(x,B) 2^{-n-1} \\
&\geq \sum_{n=0}^{\infty} \int \kappa(dy) P^n(x,B) 2^{-n-1}.
\end{aligned}$$

Comparing (1.2) with the definition of irreducible measures we see that κ is an irreducibility measure. Therefore letting $\psi(\cdot) = \sum_{n=0}^{\infty} \int \kappa(dy) P^n(x,\cdot) 2^{-n-1}$ implies that ψ is a maximal irreducibility measure of the form (1.1) and that C is ψ-petite, which completes the proof. $\square$

A central fact in the study of Markov chains is the existence of small sets. Nummelin's Splitting Technique [15] and some important ergodic theorem (see Theorem 2 of [22]) rely on that fact.

Fact 1.1.2. *For an irreducible Markov chain, every set $A \in \mathcal{B}^+(\mathcal{X})$ contains a small set in $\mathcal{B}^+(\mathcal{X})$.*

The proof of the fact requires measure-theoretic techniques that are outside the scope of this thesis. A detailed proof can be found in Theorem 5.2.1 of Meyn and Tweedie [12].

An interesting result is the equivalence of small sets and petite sets.

Theorem 1.1.3 (Theorem 5.5.3 of [12], lemma 17 [19])**.** *For an aperiodic and irreducible Markov chain $\{x_t\}$ every petite set is small.*

Proof. Let A be a petite set for some distribution $a(\cdot)$. We can assume by Lemma 1.1.1 that A is κ-petite where κ is a maximal irreducibility measure. By Fact 1.1.2 we know that there exists a $(m, \epsilon, \nu(\cdot))$-small set $C \in \mathcal{B}^+(\mathcal{X})$.

By aperiodicity there exists $N(x, C)$ for each $x \in \mathcal{X}$ such that $P^n(x, C) > 0$ for all $n \geq N(x, C)$, which leads to the inequalities

$$P^{N(x,C)+m}(x, B) \geq \int_C P^{N(x,C)}(x, dy)P^m(y, B) \geq P^{N(x,C)}(x, C)\epsilon\nu(B)$$

for all $B \in \mathcal{B}(\mathcal{X})$. Thus we have for all $x \in \mathcal{X}$

$$P^{n(x)}(x, B) \geq \delta(x)\nu(B), \quad B \in \mathcal{B}^+\mathcal{X} \tag{1.3}$$

for some functions $n : \mathcal{X} \to \mathbb{N}$, $\delta : \mathcal{X} \to (0, \infty)$. We now define the sets $A_m = \{x : n(x) = m\}$ and the constants $\delta_m = \int_{A_m} \nu(dx)\delta(x)$ for all $m \in \mathbb{N}$. We note that by (1.3) that $\int_{A_n} \nu(dx)P^n(x, \cdot) \geq \delta_m\nu(\cdot)$.

Letting T be the set $\left\{n \geq 1 : \exists c_n > 0 \text{ such that } \int \nu P^n(x, \cdot) \geq c_n\nu(\cdot)\right\}$, we show that T is non empty and additive. By the irreducibility assumption $\cup_{n=0}^{\infty} A_n$ is a full absorbing set, and the A_n are disjoint so that $\sum_{n=0}^{\infty} \kappa(A_n) = 1$ implying T is non empty. For all $n, k \in T$ then

$$\begin{aligned}\int \nu(dx)P^{n+k}(x,\cdot) &\geq \int \nu(dx)\int P^n(x,dy)P^k(y,\cdot)\\ &\geq \int \nu(dx)\int \delta(x)\nu(dy)P^k(y,\cdot)\\ &= \int \nu(dx)\delta(x)\int \nu(dy)P^k(y,\cdot)\\ &\geq c_n c_k \nu(\cdot)\end{aligned}$$

implying that $n+k$ is in T and so T is additive.

If $\gcd(T) = d > 1$ then for some N^* the sets $\{\cup_n A_{N^*+nd+i}\}$, $i = 1, ...d$ are positive measure sets that are periodic in the Markov chain $\{x_t\}$. However, this contradicts the assumption that the Markov chain is aperiodic and so we must conclude that $\gcd(T) = 1$.

With $\gcd(T) = 1$ we have that $n \in T$ for all $n \geq N^*$ for some N^*. With the (m, ϵ, ν)-small set C we have for all $n \in T$

$$\begin{aligned}P^{m+n}(x,\cdot) &= \int P^m(x,dy)P^n(y,\cdot)\\ &\geq \int 1_C(x)\epsilon\nu(dy)P^n(y,\cdot)\\ &\geq 1_C(x)\epsilon c_n \nu(\cdot)\end{aligned} \tag{1.4}$$

which implies that the (m, ϵ, ν)-small set C is actually $(m+n, \epsilon c_n, \nu)$-small for all $n \geq N^*$.

With A being κ-petite for a distribution $a(\cdot)$ we can choose a positive integer $N(C)$ such that $\sum_{n=N(C)+1}^{\infty} a(n) \leq \kappa(C)/2$ so that $\sum_{n=1}^{N(C)} a(n)P^n(x,C) \geq \kappa(C) - \sum_{n=N(C)+1}^{\infty} a(n) \geq \kappa(C)/2$. Without loss of generality $N^* \geq m$, which implies for all

$k \le N(C)$ that $N(C) + 2N^* - k \ge m + N^*$. Using (1.4) and choosing $\gamma = \min\{c_n : m + N^* \le n \le N(C) + 2N^*\} > 0$ yields for all x in the petite set A

$$\begin{aligned} P^{N(C)+2N^*}(x,\cdot) &\ge \sum_{k=1}^{N(C)} P^{N(C)+2N^*}(x,\cdot)a(k) \\ &\ge \sum_{k=1}^{N(C)} \int_C a(k)P^k(x,dy)P^{N(C)+2N^*-k}(y,\cdot) \\ &\ge \sum_{k=1}^{N(C)} a(k)P^k(x,C)\gamma\nu(\cdot) \\ &\ge \frac{1}{2}\kappa(C)\gamma\nu(\cdot). \end{aligned}$$

The proof is completed by noting that the petite set A is $\big(N(C) + 2N^*, \frac{1}{2}\kappa(C)\gamma, \nu(\cdot)\big)$-small. $\square$

The equivalence of small sets and petite sets can be used cleverly to show that all petite sets are petite for some distribution that has finite mean. The next theorem follows from Proposition 5.5.5 and 5.5.6 of [12].

Theorem 1.1.4. *For an aperiodic and irreducible Markov chain every petite set is petite with a maximal irreducibility measure for a distribution with finite mean.*

Proof. Suppose $C \in \mathcal{B}^+(\mathcal{X})$ is κ-petite. By the previous theorem we have that C is (m, ϵ, ν)-small, so that C is $\epsilon\nu$-petite for the impulse distribution $\delta_m(\cdot) = 1_m(\cdot)$. We can then apply Lemma 1.1.1 with $f(n) = 2^{-n-1}1_{\{n\ge 0\}}$ to show that C is ψ_{δ_m}-petite for the distribution $\delta_m * f(\cdot) = 2^{m-1-(\cdot)}1_{\{\cdot\ge m\}}$ so that $\psi_{\delta_m} = \sum_{k=m}^{\infty} \int \nu(dx)P^k(x,\cdot)2^{m-1-k}$ is a maximal irreducibility measure. The only part left to prove is that the distribution $\delta_m * f$ has a finite mean, but that follows by noting that $\sum_{k=1}^{\infty} k(\delta_m * f)(k) = \sum_{k=m}^{\infty} k2^{m-1-k} < \infty$. $\square$

We require one last definition relating to Markov chains, that of stopping times which allow us to sample the chain at times when events occur, most notably the times of first hitting or returning to a set.

Definition (Stopping Times)**.** *A stopping time with respect to a Markov chain* $\{x_t\}$ *is a random variable* τ *such that for each* n*, the events* $\{\tau \leq n\}$ *are measurable with respect to the* σ*-field generated by* $\{x_0, x_1, ...x_n\}$*.*

The most common stopping time we use is the hitting time and we use the standard notation that $\tau_B = \min\{t > 0 : x_t \in B\}$ for $B \in \mathcal{B}(\mathcal{X})$.

1.1.2 Regularity and Ergodicity

Regularity and ergodicity are concepts closely related through the work of Meyn and Tweedie [12], [13] and Tuominen and Tweedie [22], which will be explored in Section 2, and specifically in Theorem 3.1.1. Although we do not use the concepts in the first sections, we still try to provide some interpretation in the context of regularity and ergodicity. The definitions below are in terms of functions $f : \mathcal{X} \to [1, \infty)$ and $r : \mathbb{N} \to (0, \infty)$.

Definition (Regular Sets)**.** *A set* $A \in \mathcal{B}(\mathcal{X})$ *is called* (f, r)*-regular if*

$$\sup_{x \in A} E_x[\sum_{k=0}^{\tau_B - 1} r(k) f(x_k)] < \infty$$

for all $B \in \mathcal{B}^+(\mathcal{X})$

Definition (Regular Measures)**.** *A finite measure* ν *on* $\mathcal{B}(\mathcal{X})$ *is called* (f, r)*-regular if*

$$E_\nu[\sum_{k=0}^{\tau_B - 1} r(k) f(x_k)] < \infty$$

for all $B \in \mathcal{B}^+(\mathcal{X})$, and a point x is called (f,r)-regular if the measure δ_x is (f,r)-regular.

This leads to a lemma relating regular distributions to regular atoms.

Lemma 1.1.5. *If a Markov chain $\{x_t\}$ has an atom $\alpha \in \mathcal{B}^+(\mathcal{X})$ and an (f,r)-regular distribution λ, then α is an (f,r)-regular set.*

Proof.

$$\begin{aligned}\lambda(\alpha)E_\alpha[\sum_{n=0}^{\tau_B-1} r(n)f(x_n)] &= \int_\alpha \lambda(dx)E_x[\sum_{n=0}^{\tau_B-1} r(n)f(x_n)] \\ &\le \int_{\mathcal{X}} \lambda(dx)E_x[\sum_{n=0}^{\tau_B-1} r(n)f(x_n)] \\ &= E_\lambda[\sum_{n=0}^{\tau_B-1} r(n)f(x_n)]\end{aligned}$$

The expectation is finite for all $B \in \mathcal{B}^+(\mathcal{X})$ since λ is assumed (f,r)-regular, and so α is an (f,r)-regular set. □

To make sense of ergodicity we first need to define the f-norm, denoted $\|.\|_f$.

Definition (f-norm)**.** *For a function $f : \mathcal{X} \to [1,\infty)$ the f-norm of a measure μ defined on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ is given by*

$$\|\mu\|_f = \sup_{g \le f} |\int \mu(dx)g(x)|.$$

The commonly used Total Variation norm, or TV-norm, is the f-norm when $f = 1$, denoted $\|.\|_{TV}$.

Definition (Ergodicity)**.** *A Markov chain $\{x_t\}$ with invariant distribution π is (f,r)-ergodic if it satisfies the conditions (i)-(iv) of Theorem 3.1.1. A more useful classification of ergodicity comes from the fact (see Theorem 3.2.3) that (f,r)-ergodicity implies*

$$r(n)\|P^n(x,\cdot)-\pi(\cdot)\|_f \to 0 \quad \textit{as } n\to\infty \textit{ for all } x\in\mathcal{X}. \tag{1.5}$$

If (1.5) is satisfied for a geometric r and $f = 1$ then the Markov chain $\{x_t\}$ is geometrically ergodic.

Before getting to Theorem 3.1.1 we reference ergodicity for convenience, however for clarity's sake the concept that we allude to is the convergence of the transition probabilities to an invariant distribution (i.e. $r(n)\|P^n(x,\cdot)-\pi(\cdot)\|_f \to 0$)

1.1.3 Foster-Lyapunov Drift Condition

The main tool used in studying Markov chains and ergodicity is drift conditions as they provide bounds on return times to reachable sets, which allows some sort of control on the dynamics of a Markov chain by focusing on when the chain hits a specific set.

Theorem 1.1.6 (See Chap. 11 of [12])**.** *Let $\{x_t\}$ be an irreducible Markov chain on state space $\mathcal{X}$. If there exists a function $V:\mathcal{X}\to(0,\infty)$, a small set C and constants $\epsilon>0,\ b<\infty$ such that*

$$E[V(x_{t+1}):x_t=x]\le V(x)-\epsilon+b\mathbf{1}_C(x) \quad \forall x\in\mathcal{X}$$

then for any fixed $B\in\mathcal{B}^+(\mathcal{X})$, $\epsilon E_{x_0}[\tau_B]\le V(x_0)+c(B)$. Further, there exists a petite set A such that $\sup_{x\in A}E_x[\tau_A]<\infty$, and $\{x_t\}$ is Positive Harris Recurrent (i.e. the *chain is* $(1, 1)$*-ergodic and converges to an invariant distribution).*

Proof. We construct a martingale sequence using the drift condition and then use it to bound return times to the small set. The Recurrence result follows after by constructing the invariant distribution using Nummelin's Splitting Technique and considering the small set as an atom. We first fix $B \in \mathcal{B}^+(\mathcal{X})$ and define stopping times

$$\tau_B = \min(t > 0 : x_t \in B)$$
$$\tau^N = \min(\tau_B, N)$$

and define a sequence $\{M_t\}_{t\in\mathbb{N}}$ by

$$M_0 = V(x_0)$$
$$M_t = V(x_t) + \sum_{k=0}^{t-1}(\epsilon - b1_C(x_k)).$$

Then $\{M_t\}$ is a supermartingale sequence since

$$\begin{aligned} E[M_{t+1} : x_t, \ldots, x_0] &= E[V(x_{t+1}) + \sum_{k=0}^{t}(\epsilon - b1_C(x_k)) : x_t, \ldots, x_0] \\ &\leq V(x_t) + \sum_{k=0}^{t-1}(\epsilon - b1_C(x_k)) \text{ (by hypothesis)} \\ &= M_t \quad \text{(by definition)} \end{aligned}$$

By Doob's Optional Sampling theorem

$$
\begin{aligned}
E_{x_0}[M_{\tau^N}] &\leq M_0 \\
E_{x_0}[V(x_{\tau^N}) + \sum_{k=0}^{\tau^N-1}(\epsilon - b1_C(x_k))] &\leq V(x_0) \\
E_{x_0}[\sum_{k=0}^{\tau^N-1}\epsilon] &\leq V(x_0) - E_{x_0}[V(x_{\tau^N})] + bE_{x_0}[\sum_{k=0}^{\tau^N-1}1_C(x_k)]
\end{aligned}
$$

To continue we note by Theorem 1.1.4 that C is κ-petite for some finite mean distribution $a(\cdot)$, which gives the inequality

$$
1_C(x) \leq \frac{\sum_{n=1}^{\infty} P^n(x, B)}{\kappa(B)}
$$

for any $B \in \mathcal{B}^+(\mathcal{X})$, and so with

$$
\begin{aligned}
E_{x_0}[P^n(x_k, B)] &= \int P^k(x_0, dy)P^n(y, B) \\
&= \int P^{n+k}(x_0, dx)1_B(x) = E_{x_0}[1_B(x_{k+n})]
\end{aligned}
$$

we have

$$
\begin{aligned}
\epsilon E_{x_0}[\tau^N] &\leq V(x_0) + b\frac{\sum\limits_{n=1}^{\infty} a(n)E_{x_0}[\sum\limits_{k=0}^{\tau^N-1} 1_B(x_{k+n})]}{\kappa(B)} \\
&\leq V(x_0) + b\frac{\sum_{n=1}^{\infty} a(n)n}{\kappa(B)}
\end{aligned}
$$

Since $\tau^N \uparrow \tau_B$ we have by the monotone convergence theorem

$$E_{x_0}[\tau^N] \uparrow E_{x_0}[\tau_B] = \sup_N E_{x_0}[\tau^N] \leq V(x_0) + b\frac{\sum_{n=1}^{\infty} a(n)n}{\kappa(B)}. \tag{1.6}$$

With $a(\cdot)$ assumed to have finite mean by the hypothesis we obtain the first claim that $\epsilon E_{x_0}[\tau_B] \leq V(x_0) + c(B)$ for any $B \in \mathcal{B}^+(\mathcal{X})$.

We now turn to the second claim of a petite set with finite return time. Noting $\mathcal{X} = \{V < \infty\}$ we have that for some $M > 0$, $\{V \leq M\} \cap C \in \mathcal{B}^+(\mathcal{X})$ and with Fact 1.1.2 there exists a petite set A contained in $\{V \leq M\} \cap C$. Since V is bounded on petite $A \in \mathcal{B}^+(\mathcal{X})$, our second claim

$$\sup_{x \in A} E_x[\tau_A] < \infty$$

follows directly from (1.6).

The claims of ergodicity and recurrence follows from Theorem 4.1 of [11], and more explicitly from Nummelin's splitting technique and its implications described below, turning A into an atom. $\square$

1.2 Nummelin's Splitting Technique

Nummelin's splitting technique [15] is a widely used method in the study of Markov chains; see 4.2 of [19], Chapter 5 of [12], Proposition 3.7 and Theorem 4.1 of [22], [24], [13], and in the appendix of [3]. With an irreducible, aperiodic Markov chain $\{x_t\}$ on state space $\mathcal{X}$ with transition probability P and a (m, δ, ν)-small set C with finite return time, we construct an atom for the Markov chain in order to construct an invariant distribution for the chain.

We first show the splitting technique for the case $m = 1$ (i.e. C is a $(1, \delta, \nu$-small set). Construct a new Markov chain $\{z_t\}$ on $\mathcal{X} \times \{0, 1\}$ by $z_t = (x_t, a_t)$ where $\{a_t\}$ are a sequence of random variables on $\{0, 1\}$, independent of $\{x_t\}$, except for when $x_t \in C$.

1. If $x_t \notin C$ then $x_{t+1} \sim P(x_t, \cdot)$
2. If $x_t \in C$ then

   with probability $\delta : a_t = 1$ and $x_{t+1} \sim \nu(\cdot)$

   with probability $(1-\delta) : a_t = 0$ and $x_{t+1} \sim \frac{P(x_t,\cdot)-\delta\nu(\cdot)}{1-\delta}$

So that the distribution of x_t given z_t is

$$P(x_{t+1} \in B : z_t = (x_t, a_t) \in C \times \{1\}) = \nu(B)$$

$$P(x_{t+1} \in B : z_t = (x_t, a_t) \in C \times \{0\}) = \frac{P(x_t, B) - \delta\nu(B)}{1-\delta}$$

Note that $\frac{P(x_t,\cdot)-\delta\nu(\cdot)}{1-\delta} \geq 0$ is a valid probability measure since C is $(1, \delta, \nu)$-small. If $x_t \in C$ then

$$x_{t+1} \sim \delta\nu(\cdot) + (1-\delta)\frac{P(x_t, \cdot) - \delta\nu(\cdot)}{1-\delta} = P(x_t, \cdot)$$

so the one-step transition probabilities are unchanged for $\{x_t\}$.

This allows us to define $S = C \times \{1\}$ as an accessible atom for $\{z_t\}$, and to construct an invariant distribution for $\{x_t\}$ using $\{z_t\}$.

We specified the technique for the one step transition probability, but the same construction equally applies for (m, ϵ, ν)-small sets where $m > 1$ with the only change being that the $m-1$ steps after hitting C at x_t are distributed conditionally on

x_t and x_{t+m} (see Section 4.2 of [19]). In the case with $m > 1$ the Markov chain $\{z_t\}$ does not have an atom, instead it has an "m-step atom" in the sense that $P^m((x,1),\cdot) = P^m((y,1),\cdot)$ for all $x, y \in \mathcal{X}$

1.3 Invariant Distribution

The invariant distribution of a Markov chain is appears as the limit of transition probabilities when Markov chains are ergodic. We use, without proving, a Theorem from Meyn and Tweedie [12] to provide conditions that give insight into the existence and properties of the invariant distribution.

Theorem 1.3.1 (Theorem 10.0.1 of [12]). *An aperiodic and irreducible Markov chain* $\{x_t\}$ *admits a unique invariant probability measure* $\pi(\cdot)$ *if there exists a petite set* C *such that* $\sup_{x\in C} E_x[\tau_C] < \infty$. *The invariant measure* π *has the representation, for any* $A \in \mathcal{B}^+(\mathcal{X})$

$$\pi(B) = \int_A \pi(dy) E_y[\sum_{n=1}^{\tau_A} 1_B(x_n)], \quad B \in \mathcal{B}(\mathcal{X}). \tag{1.7}$$

The case when $\{x_t\}$ has an atom α such that $\sup_{x\in\alpha} E_x[\tau_\alpha] < \infty$ is interesting because it allows the invariant distribution π to have a more explicit representation. In the case with an atom α,

$$\pi(\mathcal{X}) = 1 = \pi(\alpha) E_\alpha[\tau_\alpha]$$

and so

$$\pi(B) = \pi(\alpha) E_\alpha[\sum_{n=1}^{\tau_\alpha} 1_B(x_n)] = \frac{E_\alpha[\sum_{n=1}^{\tau_\alpha} 1_B(x_n)]}{E_\alpha[\tau_\alpha]}$$

which is Kac's Theorem (see Theorem 10.2.2 of [12]). We note that for any $A \in \mathcal{B}^+(\mathcal{X})$,

$$\pi(V) = \int_A \pi(dy) E_y[\sum_{n=1}^{\tau_A} V(x_n)] \tag{1.8}$$

and if $\{x_t\}$ satisfies The Foster-Lyapunov Drift condition $PV \leq V - \epsilon + b1_C$ then $\{V < \infty\}$ is a full absorbing set and we can choose $A = \{V < Q\} \in \mathcal{B}^+(\mathcal{X})$ for Q large enough.

Therefore with the supermartingale sequence $M_t = V(x_t) + \sum_{k=0}^{t-1}(\epsilon - b1_C(x_k))$ we have that for $\tau^N = \min\{N, \tau_A\}$

$$\begin{aligned}
E_y[\sum_{n=1}^{\tau_A} V(x_n)] &= E_y[\sum_{n=1}^{\tau_A} E_y[V(x_{\tau^n}) : x_n]] \\
&\leq E_y[\sum_{n=1}^{\tau_A} V(x_{\tau^{n-1}}) + b1_C(x_{\tau^{n-1}})]
\end{aligned}$$

so that by iteterating conditional expectations we have

$$\begin{aligned}
&\leq E_y[\sum_{n=1}^{\tau_A} V(y) + bE_y[\sum_{k=1}^{\tau^n - 1} 1_C(x_k) : \tau_A]] \\
&\leq E_y[\sum_{n=1}^{\tau_A} V(y) + \sum_{s=0}^{\infty} a(s) \frac{bE_y[\sum_{k=0}^{\tau^n-1} 1_A(x_{k+s}) : \tau_A]}{\kappa(A)}] \quad \text{as in Theorem 1.1.6} \\
&\leq E_y[\sum_{n=1}^{\tau_A} V(y) + \sum_{s=0}^{\infty} a(s) \frac{bs}{\kappa(A)}] \\
&\leq E_y[\tau_A(V(y) + c(A))] \\
&\leq (V(y) + c(A))^2/\epsilon
\end{aligned}$$

and because V is bounded on $A = \{V < Q\}$ we have that $\pi(V) < \infty$.

It is interesting to note how the drift condition can give us a more explicit representation of π by starting with Kac's Theorem. If $\{x_t\}$ satisfies a Foster-Lyapunov drift condition with a (m, δ, ν)-small set C then using Nummelin's Splitting Technique to create the Markov chain $\{z_t\} = \{(x_t, a_t)\}$ where the m-skeleton $\{z_{tm}\}$ has an atom, we can use (1.4) to get $P^{km}(x, \cdot) \geq m^{-1} \sum_{i=0}^{m-1} P^i 1_C(x)\nu(\cdot)$. Therefore the m-skeleton $\{z_{tm}\}$ satisfies a drift condition with a (k, a, ν)-small set $\{\sum_{i=0}^{m-1} P^i 1_C \geq a\}$ (i.e. Lemma 14.2.8 of [12])and so the m-skeleton by Theorem 1.1.6 has a bounded return time to an atom. We can then apply Kac's Theorem to get an invariant measure π_m for $\{x_{tm}\}$ and an invariant measure $\pi = m^{-1} \sum_{i=0}^{m-1} P^i \pi_m$ for $\{x_t\}$. A more detailed explanation, and a proof, can be found in Chapter 10 of [12], by Meyn and Tweedie.

1.3.1 Bounds on Stopping Time Moments

If a Markov chain has an invariant distribution π of the form of (1.7) it can be used to bound the expected hitting time of any set $B \in \mathcal{B}(\mathcal{X})$ such that $\pi(B) > 0$. The result is similar to Theorem 14.2.4 of [12] which is more general and uses regularity.

Proposition 1.3.2. *Suppose $\{x_t\}$ is an aperiodic and irreducible Markov chain. If $\sup_{x \in C} E_x[\tau_C] < \infty$ for some small set C then*

$$\sup_{x \in C} E_x[\tau_B] < \infty$$

for all $B \in \mathcal{B}^+(\mathcal{X})$. (i.e. C is a (1,1)-regular set).

Proof. Similarly to the proof of Theorem 3.1.1, for any bounded function $f : \mathcal{X} \to$

$[1, M]$ we can define

$$V(x) = E_x[\sum_{k=1}^{\tau_C} f(x_k)]1_{C^C}(x) + f(x)$$

which satisfies a drift condition

$$\begin{aligned}
PV(x) &= E_x[\sum_{k=2}^{\tau_C} f(x_k)1_{C^C}(x_1)] + Pf(x) \\
&= E_x[\sum_{k=2}^{\tau_C} f(x_k)1_{C^C}(x_1)] + E_x[f(x_1)1_C(x_1)] + E_x[f(x_1)1_{C^C}(x_1)] \\
&= E_x[\sum_{k=1}^{\tau_C} f(x_k)1_{C^C}(x_1)] + E_x[f(x_1)1_C(x_1)] \\
&= E_x[\sum_{k=1}^{\tau_C} f(x_k)] \\
&= V(x) - f(x) + 1_C(x)E_x[\sum_{k=1}^{\tau_C} f(x_k)] \\
&\leq V(x) - 1 + 1_C(x)\sup_{x\in C} ME_x[\tau_C].
\end{aligned}$$

With V bounded on C by M, the proof is completed by an application of Theorem 1.1.6, giving $\sup_{x\in C} E_x[\tau_B] \leq M + c(B)$.

□

1.4 Coupling Inequality and Parallel Chains

The main idea behind the coupling inequality is to bound the total variation distance between the distributions of two random variables by the probability they are different; the more coupled the two random variables are the more their distributions match up. Let X, Y be two jointly distributed random variables on a space $\mathcal{X}$ with distributions $\mathcal{L}(X)$ and $\mathcal{L}(Y)$ respectively. Then we can bound the total variation between the distributions by the probability the two variables are not equal.

$$\begin{aligned}
\|\mathcal{L}(X)-\mathcal{L}(Y)\|_{TV} &= \sup_A |P(X\in A)-P(Y\in A)| \\
&= \sup_A |P(X\in A, X=Y)+P(X\in A, X\neq Y) \\
&\quad - P(Y\in A, X=Y)-P(Y\in A, X\neq Y)| \\
&\leq \sup_A |P(X\in A, X\neq Y)-P(Y\in A, X\neq Y)| \\
&\leq P(X\neq Y)
\end{aligned}$$

The coupling inequality is useful in discussions of ergodicity when used in conjunction with parallel Markov chains, as in 4.1 of [19], and 4.2 of [5]. We try to create two Markov chains having the same one-step transition probability and as high as possible probability of being equal.

Let $\{x_n\}$ and $\{x'_n\}$ be two Markov chains that have probability transition kernel $P(x,\cdot)$, and let C be an (m,δ,ν)-small set. We use the coupling construction provided by Roberts and Rosenthal.

Let $x_0 = x$ and $x'_0 \sim \pi(\cdot)$ where $\pi(\cdot)$ is the invariant distribution of both Markov chains.

1. If $x_n = x'_n$ then $x_{n+1} = x'_{n+1} \sim P(x_n,\cdot)$
2. Else, if $(x_n, x'_n) \in C\times C$ then

   with probability δ, $x_{n+m} = x'_{n+m} \sim \nu(\cdot)$

   with probability $1-\delta$ then independently

   $x_{n+m} \sim \frac{1}{1-\delta}(P^m(x_n,\cdot)-\delta\nu(\cdot))$

$x'_{n+m} \sim \frac{1}{1-\delta}(P^m(x'_n, \cdot) - \delta\nu(\cdot))$

3. Else, independently $x_{n+m} \sim P^m(x_n, \cdot)$ and $x'_{n+m} \sim P^m(x'_n, \cdot)$.

The in-between states $x_{n+1}, ...x_{n+m-1}, x'_{n+1}, ...x'_{n+m-1}$ are distributed conditionally given $x_n, x_{n+m}, x'_n, x'_{n+m}$, as mentioned in Section 1.2.

We would like to use this coupling between the $\{x_n\}$ and the stationary $\{x'_n\}$ to bound the normed difference between distribution of x_n and the invariant distribution $\pi(\cdot)$. By the Coupling Inequality and the previous discussion with Nummelin's Splitting technique we have $\|P^n(x, \cdot) - \pi(\cdot)\|_{TV} \leq P(x_n \neq x'_n)$.

1.5 Commentary

We introduced the important concepts that will be used throughout this thesis: aperiodic and irreducible time homogeneous Markov chains $\{x_t\}$, petite sets, drift conditions, and Nummelin's Splitting technique. We showed the equivalence of petite sets and small sets, how a petite set and drift condition bound stopping times, and how using a small set a Markov chain can be split to create a sort of pseudo-atom. The ergodicity results in the following sections use a similar concept as the coupling inequality to show how bounded return times of a petite set resulting from a drift condition implies ergodicity.

Chapter 2

Geometric Ergodicity

In this section we follow the results given by Roberts and Rosenthal in [19] to show how a strong type of ergodicity, geometric ergodicity, follows from a simple drift condition. The proofs are simple and self-contained and rely heavily on the Coupling Inequality and Nummelin's Splitting technique, showcasing how powerful the concepts are.

2.1 Univariate Drift Condition

A variant of the Foster-Lyapunov condition, the univariate drift condition is stronger and in fact implies the Foster-Lyapunov drift condition. An irreducible Markov chain satisfies the univariate drift condition if there are constants $\lambda \in (0,1)$ and $b < \infty$, along with a function $W : \mathcal{X} \to [1,\infty)$, and a small set C such that

$$PW \leq \lambda W + b\mathbf{1}_C.$$

We can check in the same way as we did for the Foster-Lyapunov drift condition that the functions $M_n = \lambda^{-n}V(x_n) - \sum_{k=0}^{n-1} b1_C(x_k)$ are supermartingale if the univariate drift condition is satisfied, and so we get the inequality

$$E_{x_0}[\lambda^{-n}V(x_n)] \leq V(x_0) + E_{x_0}[\sum_{k=0}^{n-1} b1_C(x_k)] \tag{2.1}$$

for all n and $x_0 \in \mathcal{X}$, and with Theorem 1.1.4 we also have

$$E_x[\lambda^{\tau_B}] \leq V(x) + c(B)$$

for all $B \in \mathcal{B}^+(\mathcal{X})$. The last inequality is proven in the same way as Theorem 1.1.6.

The univariate drift condition interestingly allows us to assume that V is bounded on C, which simplifies working with this particular drift condition.

Lemma 2.1.1 (Lemma 14 of [19])**.** *If a function $V : \mathcal{X} \to [1, \infty)$, constants $\lambda \in (0, 1)$, $b > 0$ and a small set C satisfy the univariate drift condition*

$$PV \leq \lambda V + b1_C$$

then we can assume that V is bounded on C.

Proof. We aim to replace λ and C so the statement is true. Choosing $\lambda_0 \in (\lambda, 1)$ and setting $K = b/(\lambda - \lambda_0)$ we define the set $C_0 = C \cap \{V \leq K\}$. We wish now to show the univariate drift condition

$$PV \leq \lambda_0 V + b1_{C_0} \tag{2.2}$$

holds for all x and with C_0 a small set. We note that by definition of a small set $C_0 \subseteq C$ is small just as C is small, and that the univariate drift condition (2.2) holds for any $x \in C_0$ and $x \notin C$, so we need only check the drift condition holds for $x \in C \setminus C_0$. Indeed for any $x \in C \setminus C_0$ we have that $V(x) \geq K$ so that

$$\begin{aligned}
PV(x) \leq & \lambda V + b \\
= & \lambda_0 V - (\lambda_0 - \lambda)V + b \\
\leq & \lambda_0 V - (\lambda_0 - \lambda)K + b \\
= & \lambda_0 V
\end{aligned}$$

and the univariate drift condition (2.2) holds for all $x \in \mathcal{X}$. Therefore we can assume either V is bounded on C or that V satisfies a different univariate condition with C_0 where V is bounded on C_0. □

Roberts and Rosenthal [19] prove using the coupling inequality that geometric ergodicity follows from the univariate drift condition.

Theorem 2.1.2 (Theorem 9 of [19])**.** *Suppose $\{x_t\}$ is an aperiodic, irreducible Markov chain with invariant distribution $\pi(\cdot)$. Suppose C is a $(1, \epsilon, \nu)$-small set and $V :\to [1, \infty)$ satisfies the univariate drift condition with constants $\lambda \in (0, 1)$ and $b < \infty$ with $V(x) < \infty$ for some $x \in \mathcal{X}$. Then $\{x_t\}$ is geometrically ergodic.*

The proof of this theorem is the focus of this chapter and will be explicitly done in steps, and the reasoning will be clear.

2.2 Bivariate Drift Condition

The bivariate drift condition is satisfied for two independent copies of a Markov chain with a small set C if there exists a function $h : \mathcal{X} \times \mathcal{X} \to [1, \infty)$ and $\alpha > 1$ such that

$$\begin{aligned}\bar{P}h(x, y) \leq & h(x, y)/\alpha \quad (x, y) \notin C \times C \\ \bar{P}h(x, y) < & \infty \quad (x, y) \in C \times C\end{aligned}$$

where

$$\bar{P}h(x, y) = \int_{\mathcal{X}} \int_{\mathcal{X}} h(z, w) P(x, dz) P(y, dw)$$

Now we explore a connection between the univariate and bivariate drift conditions.

Proposition 2.2.1 (Proposition 11 of [19])**.** *Suppose the univariate drift condition is satisfied for* $V : \mathcal{X} \to [1, \infty)$ *and constants* $\lambda \in (0, 1)$ $b < \infty$ *and small set* C*. Letting* $d = \inf_{x \in C^C} V(x)$*, if* $d > \frac{b}{1-\lambda} - 1$*, then the bivariate drift condition is satisfied for* $h(x, y) = \frac{1}{2}(V(x) + V(y))$ *and* $\alpha^{-1} = \lambda + b/(d + 1) > 1$*.*

Proof. Having $(x, y) \notin C \times C$ implies by the univariate drift condition that $h(x, y) = \frac{1}{2}(V(x) + V(y)) \geq (d + 1)/2$ so

$$\begin{aligned}\bar{P}h(x, y) =& \frac{1}{2}(PV(x) + PV(y)) \\ \leq& \frac{1}{2}(\lambda V(x) + \lambda V(y) + b) \\ =& \lambda h(x, y) + \frac{b}{2} \\ \leq& \lambda h(x, y) + \frac{b}{2}(h(x, y)/((1 + d)/2)\end{aligned}$$

$$=(\lambda + b/(d+1))h(x,y).$$

Therefore if $d > (b/(1-\lambda)) - 1$ the bivariate drift condition is satisfied with $\alpha^{-1} = \lambda + b/(d+1) > 1$. $\square$

Assuming two copies of a Markov chain with a $(m.\epsilon, \nu)$-small set C are run in parallel as described in Section 1.4 we have for all $(x_k, x'_k) \in C \times C$.

$$\bar{P^m} h(x_k, x'_k) 1_{(x_{k+1} \neq x_{k+1})} = (1-\epsilon) \int_{\mathcal{X}} \int_{\mathcal{X}} (1-\epsilon)^{-2} h(z,w)(P^m(x,dz) - \epsilon\nu(dz))(P^m(y,dw) - \epsilon\nu(dw))$$

and we can define

$$B_m = \max[1, \alpha(1-\epsilon) \sup_{C \times C} \bar{R}h]$$

where $(1-\epsilon)\bar{R}h(x_k, x'_k) = \bar{P}h(x_k, x'_k)1_{(x_{k+1} \neq x_{k+1})}$

2.3 Geometric Ergodicity

That geometric ergodicity follows from the univariate drift condition with a small set C is proven by Roberts and Rosenthal by using the coupling inequality to bound the TV-norm, but an alternate proof is given by Meyn and Tweedie [12] with Theorem 15.4.1, Theorem 15.4.3 using a regenerative decomposition to bound the TV-norm. The results are extremely similar in that they both make use of bounded return times to a small set, however Meyn and Tweedie have a stronger result proving geometric ergodicity under the V-norm where V satisfies the univariate drift condition. We note that Hairer [5] gives a concise proof of geometric ergodicity using the coupling inequality and a bivariate drift in the same spirit as Roberts and Rosnethal in Section 4.2 of [5], although for the continuous time case.

Theorem 2.3.1 (Theorem 12 of [19])**.** *If V satisfies the univariate drift condition and h satisfies the bivariate drift condition then for the two Markov chains $\{x_k\}$ and $\{x'_k\}$ run in parallel*

$$\|P^k(x_0, x_k) - P^k(x'_0, x'_k)\|_{TV} \le (1-\epsilon)^j + \alpha^{-k} B^{j-1} E[h(x_0, x'_0)]$$

for all $1 \le j \le k$ where B and α are defined as the bivariate drift condition, ϵ is defined from the drift condition for V where C is a (m, ϵ, ν)-small set.

Proof. We first prove the result for the case when $m = 1$ and for ease denote B_1 as B. Let $N_k = \#\{m : 0 \le m \le k, (x_m, x'_m) \in C \times C\}$ then

$$\begin{aligned} P(x_k \neq x'_k) =& P(x_k \neq x'_k, N_{k-1} \ge j) + P(x_k \neq x'_k, N_{k-1} < j) \\ \le& (1-\delta)^j + P(x_k \neq x'_k, N_{k-1} < j) \end{aligned} \tag{2.3}$$

Then we bound the second term with martingales using $M_k = \alpha^k B^{-N_{k-1}-1} h(x_k, x'_k) 1_{(x_k \neq x'_k)}$ with $N_{-1} = 0$. The proof that $\{M_k\}$ is supermartingale is given in the next lemma. We assume for the moment that $\{M_k\}$ is supermartingale and proceed.

$$\begin{aligned} P(x_k \neq x'_k, N_{k-1} < j) \le& P(x_k \neq x'_k, B^{-N_{k-1}} \ge B^{-(j-1)}) \text{ (since } B \ge 1) \\ =& P(1_{(x_k \neq x'_k)} B^{-N_{k-1}} \ge B^{-(j-1)}) \\ \le& B^{j-1} E[1_{(x_k \neq x'_k)} B^{-N_{k-1}}] \quad \text{(By Markov's Iequality)} \\ \le& B^{j-1} E[1_{(x_k \neq x'_k)} B^{-N_{k-1}} h(x_k, x'_k)] \quad (\text{as } h \ge 1) \\ =& \alpha^{-k} B^{j-1} E[M_k] \quad (\text{by definition of } M_k) \\ \le& \alpha^{-k} B^{j-1} E[M_0] \quad (\text{since } \{M_k\} \text{ is supermartingale}) \end{aligned}$$

$$=\alpha^{-k}B^{j-1}E[h(x_0,x_0')]$$

Thus the result holds, and choosing $j=\lfloor rk\rfloor$ for sufficiently small r gives geometric convergence of $\|P^k(x_0,x_k)-P^k(x_0',x_k')\|_{TV}$.

The case for $m>1$ can be handled in the same way with only a few changes to the definitions of the terms. The main difference between the case when $m=1$ and $m>1$ is that for the latter case we wish to ignore the Markov chains for the $m-1$ steps after hitting $C\times C$ since those times are "filled in" by conditioning on the Markov chain at the hitting time of $C\times C$ and m steps afterwards (see Sections 1.4, 1.2).

We define the sequence hitting times of $C\times C$ avoiding the "filled in" times as

$$\tau_i=\min\{t\geq\tau_{i-1}+m:(x_t,x_t')\in C\times C\}$$

with $\tau_0=0$. We change the term N_k to count the number of hitting terms we care about,

$$N_k=\max\{n:\tau_n\leq k\}.$$

and we replace N_k by N_{k-m} in (2.3) and in the definition of M_k. Finally we have that M_k is not supermartingale but $M_{t(k)}$ is; $t(k)$ represent the latest current time we care about and is defined as

$$t(k)=\begin{cases}k, & \text{if } (x_n,x_n')\notin C\times C \text{ for } k-m\leq m\leq k-1\\ t(k-1), & \text{else}\end{cases}$$

With these changes to the terms the proof works exactly as before. $\square$

Lemma 2.3.2. *$\{M_k\}$ defined in the proof of the previous theorem is a supermartingale.*

Proof. If $(x_k, x'_k) \notin C \times C$ then $N_k = N_{k-1}$ and

$$
\begin{aligned}
&E[M_{k+1} : x_0, \ldots, x_k, x'_0, \ldots, x'_k] \\
&\quad =\alpha^{k+1} B^{-N_{k-1}} E[h(x_{k+1}, x'_{k+1}) 1_{(x_{k+1} \neq x'_{k+1})} : x_k, x'_k] \\
&\quad \leq\alpha^{k+1} B^{-N_{k-1}} E[h(x_{k+1}, x'_{k+1}) : x_k, x'_k] 1_{(x_k \neq x'_k)} \\
&\quad =\alpha^{k+1} B^{-N_{k-1}} \bar{P}h(x_k, x'_k) 1_{(x_k \neq x'_k)} \quad \text{(by definition of } h) \\
&\quad =M_k \alpha \bar{P}h(x_k, x'_k)/h(x_k, x'_k) \quad \text{(by definition of } M_k) \\
&\quad \leq M_k
\end{aligned}
$$

where the last inequality holds since h satisfies the bivariate drift condition $\bar{P}h(x, y) \leq h(x, y)/\alpha$ for all $(x, y) \notin C \times C$.

If $(x_k, x'_k) \in C \times C$ then $N_k = N_{k-1} + 1$ and

$$
\begin{aligned}
&E[M_{k+1} : x_0, \ldots, x_k, x'_0, \ldots, x'_k] \\
&\quad =\alpha^{k+1} B^{-N_{k-1}-1} E[h(x_{k+1}, x'_{k+1}) 1_{(x_{k+1} \neq x'_{k+1})} : x_k, x'_k] \\
&\quad \leq\alpha^{k+1} B^{-N_{k-1}-1} (1-\epsilon) \bar{R}h(x_k, x'_k) \\
&\quad =\alpha^{k+1} B^{-N_{k-1}} \bar{P}h(x_k, x'_k) 1_{(x_k \neq x'_k)} \\
&\quad =M_k \alpha B^{-1} (1-\epsilon) \bar{R}h(x_k, x'_k)/h(x_k, x'_k) \\
&\quad \leq M_k/h(x_k, x'_k) \quad \text{(by definition of B)} \\
&\quad \leq M_k \quad (\text{since } h \geq 1)
\end{aligned}
$$

The second last inequality holds since $B = \max\{1, \alpha(1-\epsilon)\sup_{C\times C}\bar{R}h\}$, thus $\{M_k\}$ is supermartingale. $\square$

Corollary 2.3.3. *Suppose V satisfies the univariate drift condition with corresponding h satisfying the bivariate drift condition on two copies of a Markov chain $\{x_k\}$ and $\{x'_k\}$ run in parallel such that $x'_0 \sim \pi(\cdot)$ where $\pi(\cdot)$ is the invariant distribution of $\{x_k\}$. Then the distribution of the Markov chain converges geometrically to the invariant distribution under the TV-norm.*

Although the results of this section require that $\inf_{x\in C^C} V(x) > \frac{b}{1-\lambda} - 1$ for V and petite set C satisfying the univariate drift condition, Lemma 18 of [19] by Rosenthal shows how this assumption is satisfied. What is interesting about the results in this section is that a Markov chain satisfying the simple univariate drift condition 2.1 is geometrically ergodic, which is a fast convergence rate. If we wish explore more varied ergodic rates we would try to alter or weaken the drift condition, and in fact the next sections are dedicated to finding conditions for subgeometric ergodicity.

2.4 V-uniform Ergodicity

We attempt to find a simple proof where we extend Roberts and Rosenthal's results into uniform ergodicity without any of the analysis done by Meyn and Tweedie. We look towards Lemma 4.6 by Hairer [5] for inspiration.

Theorem 2.4.1. *Suppose an irreducible and aperiodic Markov chain $\{x_t\}$ satisfies the univariate drift condition 2.1 with $\lambda < 1$, a constant b, a small set C and a function $V : \mathcal{X} \to [1, \infty)$. If $\pi(V) < \infty$, where π is the invariant distribution, then* *the Markov chain is V-uniform ergodic, which means that*

$$\|P^n(x_0,\cdot)-\pi(\cdot)\|_V \leq BV(x_0)\gamma^n$$

for all $x_0 \in \mathcal{X}$ where B is a constant and $\gamma < 1$.

Proof. From the discussion of Rosenthal's techniques, in particular 2.3.1 we have that the Markov chain is geometrically ergodic, and that

$$\|P^n(x_0,\cdot)-\pi(\cdot)\|_{TV} \leq MV(x_0)\rho_1^n$$

for some constants M and $\rho_1 < 1$. From the definition of $\|.\|_V$ and the triangle inequality we have for all $R > 0$ and all n

$$\|P^n(x_0,\cdot)-\pi(\cdot)\|_V \leq R\|P^n(x_0,\cdot)-\pi(\cdot)\|_{TV} + \int_{V\geq R} V(y)(P^n(x_0,dy)+\pi(dy)).$$

We can treat this inequality as separate inequalities for each n which allows us to replace the constant R by a function on $\mathbb{N}$. We choose a constant ρ_2 so that we have the strict chain of inequalities $1 < \rho_2 < \min\{\rho_1^{-1}, \lambda^{-1}\}$, and define a function $r : \mathbb{N} \to (0,\infty)$ by $r(n) = \rho_2^n$.

From the univariate drift condition we have the inequality $E_{x_0}[V(x_n)] \leq V(x_0) + b(n+1)$ but with $1 < \rho_2$ we know that for some N_b that $\rho_2^n \geq (n+1)b$ for all $n \geq N_b$, which yields

$$E_{x_0}[\lambda^{-n}V(x_n)] \leq V(x_0) + E_{x_0}[\rho_2^n V(x_n)]$$

for all $n \geq N_b$. By choice we have that $\lambda\rho_2 < 1$ and so there exists N^* such that

$1-(\lambda\rho_2)^n \geq 2^{-1}$ for all $n \geq N^*$, so that

$$E_{x_0}[V(x_n)2^{-1}] \leq E_{x_0}[(1-\lambda^n\rho_2^n)V(x_n)] \leq \lambda^n V(x_0)$$

for all $n \geq N^*$. Applying all these inequalities we get

$$\|P^n(x_0,\cdot)-\pi(\cdot)\|_V \leq \rho_2^n\|P^n(x_0,\cdot)-\pi(\cdot)\|_{TV} + \int_{V\geq\rho_2^n} V(y)(P^n(x_0,dy)+\pi(dy)) \tag{2.4}$$

$$\leq MV(x_0)(\rho_1\rho_2)^n + 2\lambda^n V(x_0) + \pi(V \geq \rho_2^n) \tag{2.5}$$

for all $n \geq \max N_b, N^*$. By Chebyshev's inequality $\pi(V \geq \rho_2^n) \leq \pi(V)\rho_2^{-n}$ and with the assumption $\pi(V) < \infty$ we can take $\gamma = \max\{\rho_1\rho_2, \lambda, \rho_2^{-1}\} < 1$ to achieve the bound

$$\|P^n(x_0,\cdot)-\pi(\cdot)\|_V \leq (M+2+\pi(V))V(x_0)\gamma^{-n}$$

which shows V-uniform ergodicity and completes the proof since x_0 was arbitrary.

□

The problem with this proof is that the rate of convergence (i.e. γ) was not explicitly determined since it is dependent on the rate of convergence in the TV-norm. However, Rosenthal [19] provides a somewhat concrete formulation of the rate of convergence in the TV-norm, and is included as Theorem 2.3.1.

2.5 Commentary

In this section we followed the paper [19], by Roberts and Rosenthal to show, using the coupling inequality, how a simple drift criteria $PV \leq \lambda V + b1_C$ implies geometric ergodicity for an aperiodic and irreducible Markov chain. The rate of ergodicity is made somewhat explicit with Theorem 2.3.1, and the results are extended to V-uniform ergodicity with a undetermined rate. To explore different rates of convergence we alter the drift condition in the following sections.

Chapter 3

Subgeometric Ergodicity

We wish to find conditions under which an aperiodic and irreducible Markov chain is ergodic with a subgeometric rate. To that effect (f, r)-regularity, ergodicity defined in the beginning in Section 1.1.2 are needed, as well as an understanding of the class of subgeometric rate functions.

In this section we focus on current research on the class of subgeometric rate functions(see section 4 of [5], section 5 of [3], [13], [11], [12], [18], [22]). Let Λ_0 be the family of functions $r : \mathbb{N} \to \mathbb{R}_{>0}$ such that

$$r \text{ is non-decreasing}, \quad r(1) \geq 2$$

and

$$\frac{\log r(n)}{n} \downarrow 0 \quad \text{as } n \to \infty$$

The second condition implies that for all $r \in \Lambda_0$ if $n > m > 0$ then

$$n \log r(n+m) \leq n \log r(n) + m \log r(n) \leq n \log r(n) + n \log r(m)$$

so that

$$r(m+n) \le r(m)r(n) \quad \text{for all } m, n \in \mathbb{N}. \tag{3.1}$$

For two functions $r, u : \mathbb{N} \to \mathbb{R}_+$ we define an equivalence relation by $r \asymp u$ if there exist constants $c_1, c_2 > 0$ such that for large $n, c_1 u(n) \le r(n) \le c_2 u(n)$. The class of subgeometric rate functions, denoted by Λ, is the class of functions $r : \mathbb{N} \to \mathbb{R}_{>0}$ such that $r \asymp u$ for some $u \in \Lambda_0$.

If $r \asymp u$ then (f, r)-regularity $\Rightarrow$ (f, u)-regularity,. As a consequence, without loss of generality we can restrict our attention to rate functions r such that $r \in \Lambda_0$.

We also note for any $r \in \Lambda_0$, defining the partial sums $R(n) = \sum_{k=0}^{n-1} r(k)$ we have that

$$\frac{r(k-1)}{R(k)} + \frac{R(k-1)}{R(k)} = 1$$

so that $\limsup_k \frac{r(k-1)}{R(k)} = c \le 1$. Further, since

$$\frac{r(k-1)}{R(k)} = \frac{r(k-1)}{r(k-1) + R(k-1)}$$

we have $c = \frac{c}{c+1}$ which implies $c^2 = 0$. Therefore

$$\lim_k \frac{r(k)}{R(k)} \le r(1) \lim_k \frac{r(k-1)}{R(k)} = 0 \tag{3.2}$$

for all $r \in \Lambda_0$ and so $r(k)/R(k) \to 0$ for any $r \in \Lambda$.

3.1 Characterization of Subgeometric Ergodicity

The main theorem we use to try to construct conditions on subgeometric rates of convergence is due to Tuominen and Tweedie [22].

Theorem 3.1.1 (Theorem 2.1 of [22]). *Suppose that $\{x_t\}_{t\in\mathbb{N}}$ is an irreducible and aperiodic Markov chain on state space $\mathcal{X}$ with stationary transition probabilities given by P. Let $f : \mathcal{X} \to [1,\infty)$ and $r \in \Lambda$ be given. The following are equivalent:*

(i) there exists a petite set $C \in \mathcal{B}(\mathcal{X})$ such that

$$\sup_{x\in C} E_x[\sum_{k=0}^{\tau_C-1} r(k)f(x_k)] < \infty$$

(ii) there exists a sequence (V_n) of functions $V_n : \mathcal{X} \to [0,\infty]$, a petite set $C \in \mathcal{B}(\mathcal{X})$ and $b \in \mathbb{R}_+$ such that V_0 is bounded on C,

$$V_0(x) = \infty \Rightarrow V_1(x) = \infty,$$

and

$$PV_{n+1} \leq V_n - r(n)f + br(n)1_C, \quad n \in \mathbb{N}$$

(iii) there exists an (f,r)-regular set $A \in \mathcal{B}^+(\mathcal{X})$.

(iv) there exists a full absorbing set S which can be covered by a countable number of (f,r)-regular sets.

To prove Theorem 3.1.1 we first need a fact of (f,r)-regularity. We first fix $f : \mathcal{X} \to [1,\infty)$ and $r \in \Lambda_0$, and denote the partial sums of r as

$$r^0(n) = \sum_{k=0}^{n} r(k).$$

Proposition 3.1.2 (Proposition 3.1 of [22]). *Suppose the conditions of (ii) in Theorem 3.1.1 are satisfied with the sequence of functions (V_n), $b \in \mathbb{R}_{>0}$ and small set C. Then,* *(i) For any $B \in \mathcal{B}^+(\mathcal{X})$ there exists a constant $c(B) < \infty$ such that for all $x \in \mathcal{X}$*

$$E_x[\sum_{k=0}^{\tau_B-1} r(k) f(x_k)] \leq c(B)(V_0(x)+1)]$$

(ii) If V_0 is bounded on $A \in \mathcal{B}^+(\mathcal{X})$, then A is (f,r)-regular.

(iii) If ν is a probability measure such that $\nu(V_0) < \infty$ then ν is (f,r)-regular.

Proof. We may assume $r \in \Lambda_0$. Theorem 3.1.1(ii) is a Lyapunov drift criteria, so by using the same reasoning as in Theorem 1.1.6 we have for any $B \in \mathcal{B}(\mathcal{X})$

$$E_x[\sum_{k=0}^{\tau_B-1} r(k) f(x_k)] \leq V_0(x) + bE_x[\sum_{k=0}^{\tau_B-1} r(k) 1_C(x_k)]. \tag{3.3}$$

Using (3.1), (3.3) implies

$$E_x[r^0(\tau_C)] \leq c_r E_x[r^0(\tau_C - 1)] \leq \ c_r(V_0(x) + b) < \infty \tag{3.4}$$

with $c_r = r(1) + 1$. Fixing $B \in \mathcal{B}^+(\mathcal{X})$ and using Proposition 1.3.2 gives

$$c_0(B) = \sup_{x \in C} E_x[r^0(\tau_B)] < \infty.$$

Using again (3.1) and $\tau_B \leq \tau_C + \theta^{\tau_C}\tau_B$, where θ is the shift operator, yields

$$\begin{aligned} E_x[r^0(\tau_B)] \leq & E_x[r^0(\tau_C)] + E_x[\sum_{k=\tau_C}^{\tau_C+\tau_B} r(k)] \\ \leq & E_x[r^0(\tau_C)] + \sup_{y \in C} E_y[r^0(\tau_B)] E_x[r(\tau_C)]. \end{aligned}$$

Now using (3.3) and (3.4) gives

$$
\begin{aligned}
E_x[\sum_{k=0}^{\tau_B-1} r(k)f(x_k)] \leq & V_0(x) + bE_x[r^0(\tau_B)] \\
\leq & V_0(x) + b(E_x[r^0(\tau_C)] + \sup_{y\in C} E_y[r^0(\tau_B)]E_x[r(\tau_C)]) \\
\leq & V_0(x) + b(V_0(x) + b)(1 + c_0(B))
\end{aligned}
$$

which completes the proof of (i). The other two results (ii) and (iii) follow immediately. $\square$

We return our attention to Theorem 3.1.1, now that we are able to prove it.

3.1.1 Proof of Theorem 3.1.1

We may assume that $r \in \Lambda_0$. We first prove the hardest implication, (i)$\Rightarrow$(ii). We define two sequences $\{V_n\}$, $\{W_n\}$ of functions $\mathcal{X} \to [0, \infty)$ by

$$
\begin{aligned}
V_n(x) = E_x[\sum_{k=1}^{\tau_C} r(n+k)f(x_k)]1_{C^C}(x) + r(n)f(x) \\
W_n = E_x[\sum_{k=1}^{\tau_C} r(n+k)f(x_k)], \quad n = 0, 1, 2...
\end{aligned}
$$

From (3.1) the functions satisfy the bounds

$$
\begin{aligned}
V_0 \leq & V_n \leq r(n)V_0 \\
W_0 \leq & W_n \leq r(n)W_0
\end{aligned}
$$

for all $n \in \mathbb{N}$, and by (i) V_0 is bounded on C since $V_0(x) = r(1)f(x)$ $\forall x \in C$. In addition, (i) directly implies that

$$\sup_{x\in C} W_0(x) = b < \infty.$$

Now noting that $V_n = r(n)f + 1_{C^c}W_n$ we get

$$\begin{aligned}
PV_{n+1}(x) = &\int_{\mathcal{X}} P(x,dy)V_{n+1}(y) \\
= &\int_{C^C} P(x,dy)W_{n+1}(y) + r(n+1)Pf(x) \\
= &\int_{C^C} P(x,dy)E_y[\sum_{k=1}^{\tau_C} r(n+1+k)f(x_k)] + r(n+1)Pf(x) \\
= &\int_{C^C} P(x,dy)\left(E_y[\sum_{k=1}^{\tau_C} r(n+1+k)f(x_k)] + r(n+1)f(y)\right) + r(n+1)P(f1_C)(x) \\
= &E_x[\sum_{k=1}^{\tau_C} r(n+k)f(x_k)] \\
= &W_n(x)
\end{aligned}$$

which implies, together with $W_n \leq r(n)W_0$ and $\sup_{x\in C} W_0(x) = b < \infty$, that $\{V_n\}$ satisfies the set of drift conditions in (ii)

$$\begin{aligned}
PV_{n+1} = &W_n \\
= &V_n - r(n)f + 1_C W_n \\
\leq &V_n - r(n)f + 1_C r(n) W_0 \\
\leq &V_n - r(n)f + r(n)b
\end{aligned}$$

thereby completing the proof that (i)⇒(ii).To show that (ii)⇒(iv) we let

$$S = \{x \in \mathcal{X} : V_0(x) < \infty\}$$

and then using the drift condition of (ii) in conjunction with $V_0(x) = \infty \Rightarrow V_1(x) = \infty$ it follows that S is absorbing and so, by irreducibility, full. Applying Proposition 3.1.2(ii) the sets

$$S_n = \{x : V_0(x) \leq n\} \quad \text{n=0,1,2,...}$$

are a countable collection of (f, r)-regular sets that cover S, which completes the proof that (ii)⇒(iv). To finish the proof of Theorem 3.1.1 we note that (iv) ⇒ (iii) is trivial and (iii)⇒(i) follows by choosing a small set $C \in \mathcal{B}(\mathcal{X})$ such that $C \subset A$. □

3.2 Ergodicity as a Rate of Convergence

If a Markov chain $\{x_t\}$ satisfies Theorem 3.1.1 for (f, r) then $r(n)\|P^n(x_0, \cdot) - \pi(\cdot)\|_f \rightarrow 0$ as n increases; see [22], [18]. The approach taken does not rely on the coupling inequality, but instead on a first-entrance last-exit decomposition [12] of the transition probabilities.

Lemma 3.2.1 (Section 13.2.3 of [12])**.** *If a Markov chain has an atom $\alpha \in \mathcal{B}^+(\mathcal{X})$ the transition probability over n steps from state x to set B is given by*

$$\begin{aligned} P^n(x, B) = &P_x(x_n \in B, \tau_\alpha \geq n) \\ &+ \sum_{j=1}^{n-1} \sum_{k=1}^{j} P_x(\tau_\alpha = k, x_j \in \alpha) P_\alpha(x_{n-j} \in B, \tau_\alpha \geq n - j) \end{aligned}$$

Proof.

$$\begin{aligned} P^n(x,B) = & P_x(x_n \in B, \tau_\alpha \geq n) \\ & + \sum_{j=1}^{n-1} P_x(\tau_\alpha \leq j, x_j \in \alpha, x_n \in B, x_k \notin \alpha, k = j+1, ..., n-1) \end{aligned}$$

The equality holds since the events $\{\tau_\alpha, x_n \in B, x_k \notin \alpha, k - j + 1, ..., n-1\}$ are disjoint for each $j \leq n$ and their union over all $j \leq n$ is $\{\tau_\alpha \leq n, x_n = y\}$. Since the event $\{\tau_\alpha \leq j, x_j \in \alpha\}$ is measurable with respect to $\{x_0, ..., x_j\}$ and the event $\{x_n = y, x_k \notin \alpha k = j+1, ..., n-1\}$ is measurable with respect to $\{x_{j+1}, .., x_n\}$, we can use the Markov property to condition the latter events on $x_j \in \alpha$ to get

$$\begin{aligned} P^n(x,B) = & P_x(x_n \in B, \tau_\alpha \geq n) \\ & + \sum_{j=1}^{n-1} P_x(\tau_\alpha \leq j, x_j \in \alpha) P_\alpha(x_{n-j} \in B, \tau_\alpha \geq n-j). \end{aligned}$$

Now noting that the events $\{\tau_\alpha = k\}$ are disjoint completes the proof

$$\begin{aligned} P^n(x,B) = & P_x(x_n \in B, \tau_\alpha \geq n) \\ & + \sum_{j=1}^{n-1} \sum_{k=1}^{j} P_x(\tau_\alpha = k, x_j \in \alpha) P_\alpha(x_{n-j} \in B, \tau_\alpha \geq n-j). \end{aligned}$$

□

Keeping in mind our goal of finding criteria for a Markov chain to be (f, r)-ergodic, we represent the results of this lemma in a more useful way using convolutions, yielding similar results to Theorem 13.2.5 of [12].

Corollary 3.2.2. *Suppose a Markov chain with atom $\alpha \in \mathcal{B}^+(\mathcal{X})$ and invariant distribution π satisfies Lemma 3.2.1. Define for any $f : \mathcal{X} \to \mathbb{R}_{>0}$ and $n \in \mathbb{N}$*

$$\begin{cases} a_x(n) = P_x(\tau_\alpha = n) \\ u(n) = P^n(\alpha, \alpha) \\ t_f(n) = E_\alpha[f(x_n)1_{\tau_\alpha \geq n}.] \end{cases} \tag{3.5}$$

Then we have for all $B \in \mathcal{B}(\mathcal{X})$

$$P^n(x, B) = P_x(x_n \in B, \tau_\alpha \geq n) + a_x * u * t_{1_B}(n) \tag{3.6}$$

and more specifically,

$$\begin{aligned} \|P^n(x, \cdot) - \pi(\cdot)\|_f \leq & E_x[f(x_n)1_{\tau_\alpha \geq n}] \\ & + |a_x * u - \pi(\alpha)| * t_f(n) + \pi(\alpha) \sum_{j=n+1}^{\infty} t_f(j) \end{aligned} \tag{3.7}$$

Proof. Noting that $a_x * u(j) = P_x(\tau_\alpha \leq j, x_j \in \alpha)$, (3.6) follows immediately from the definitions (3.5).

By the characterization of the invariant distribution (1.7) for a Markov chain with an atom $\alpha \in \mathcal{B}^+(\mathcal{X})$ we have $\pi(\alpha) = (E_\alpha[\tau_\alpha])^{-1}$ so that using the definitions (3.5) we have

$$\begin{aligned} \pi(B) = & \pi(\alpha) \sum_{k=0}^{\infty} E_\alpha[1_B(x_k)1_{\tau_\alpha \geq k}] \\ = & \pi(\alpha) \sum_{k=0}^{\infty} t_{1_B}(k) \\ = & \lim_{k\to\infty} (\pi(\alpha) * t_{1_B})(k). \end{aligned}$$

Now applying (3.6) we get the inequality

$$\begin{aligned}
|P^n(x,B)-\pi(B)| =&|P^n(x,B)-\pi(\alpha)\sum_{k=0}^{\infty} t_{1_B}(k)| \\
\leq&P_x(x_n\in B,\tau_\alpha\geq n) \\
&+|a_x*u*t_{1_B}(n)-\pi(\alpha)\sum_{k=0}^{n} t_{1_B}(k)|+\pi(\alpha)\sum_{k=n+1}^{\infty} t_{1_B}(k) \\
=&P_x(x_n\in B,\tau_\alpha\geq n) \\
&+|(a_x*u-\pi(\alpha))*t_{1_B}(n)|+\pi(\alpha)\sum_{k=n+1}^{\infty} t_{1_B}(k) \\
\leq&P_x(x_n\in B,\tau_\alpha\geq n) \\
&+|a_x*u-\pi(\alpha)|*t_{1_B}(n)+\pi(\alpha)\sum_{k=n+1}^{\infty} t_{1_B}(k)
\end{aligned}$$

for all $x\in\mathcal{X}$ and $B\in\mathcal{B}(\mathcal{X})$.

Since the Lebesgue integral is defined as the limit of sums of simple functions, the result (3.7) follows from noting that the above inequality holds for all measurable set $B\in\mathcal{B}(\mathcal{X})$ and that

$$\|P^n(x,\cdot)-\pi\|_f=\sup_{g\leq f}|\int(P^n(x,dy)-\pi(dy))g(y)|.$$

$\square$

We are now set to connect Theorem 3.1.1 to (f,r)-ergodicity by showing through the inequality (3.7) that $r(n)\|P^n(x,\cdot)-\pi\|_f\to 0$. Although Lemma 3.2.1, and therefore (3.7), only apply to Markov chains with an atom we extend the results to Markov chains without an atom by applying Nummelin's Splitting Technique as in

Section 1.2.

Theorem 3.2.3 (Theorem 4.1 of [22]). *Suppose an aperiodic and irreducible Markov chain $\{x_t\}$ satisfies (ii)-(iv) of Theorem 3.1.1 with $f : \mathcal{X} \to [1, \infty)$ and $r \in \Lambda$. Then the Markov chain is (f, r)-ergodic so as in (1.5),*

$$r(n)\|P^n(x, \cdot) - \pi\|_f \to 0$$

Proof. We may assume without loss of generality that $r \in \Lambda_0$. Let $S(f, r) = \{x \in \mathcal{X} : x \text{ is } (f, r)\text{-regular}\}$. By Proposition 3.1.2, $S(f, r) = \{x : V_0(x) < \infty\}$ is a full and absorbing set where V_0 is a function as described in Theorem 3.1.1(ii).

(i) First we deal with the case where the Markov chain $\{x_t\}$ has an atom $\alpha \in \mathcal{B}(\mathcal{X})$. Therefore by Lemma 3.2.1, the inequality (3.7) holds and, with the definitions (3.5), we have

$$\begin{aligned} r(n)\|P^n(x, \cdot) - \pi(\cdot)\|_f \leq & r(n)E_x[f(x_n)1_{\tau_\alpha \geq n}] \\ & + r(n)|a_x * u - \pi(\alpha)| * t_f(n) + r(n)\pi(\alpha) \sum_{j=n+1}^{\infty} t_f(j). \end{aligned}$$

To show the convergence of the first term we note that by the definition of $S(f, r)$

$$\sum_{n=0}^{\infty} r(n)E_x[f(x_n)1_{\tau_\alpha \geq n}] = E_x[\sum_{n=0}^{\tau_\alpha} r(n)f(x_n)] < \infty$$

for all $x \in S(f, r)$, therefore

$$r(n)E_x[f(x_n)1_{\tau_\alpha \geq n}] \to 0 \text{ as } n \to \infty.$$

To show the convergence of the third term we have

$$\sum_{n=0}^{\infty} r(n)t_f(n) = E_\alpha[\sum_{n=0} \tau_\alpha r(n)f(x_n)] = E_y[\sum_{n=0}^{\tau_\alpha} r(n)f(x_n)]$$

for any $y \in \alpha$. The sum is finite since $S(f,r)$ is full and absorbing and $\alpha \in \mathcal{B}^+(\mathcal{X})$, which implies that $\sum_{j=n+1}^{\infty} r(j)t_f(j) \to 0$ as $n \to \infty$.

Since $\sum_{n=0}^{\infty} r(n)t_f(n)$ is finite, the second term $r(n)|a_x * u - \pi(\alpha)| * t_f(n)$ converges to 0 if $r(n)|a_x * u(n) - \pi(\alpha)| \to 0$. Noting that for all $x \in S(f,r)$

$$\sum_{n=0}^{\infty} a_x(n) = \sum_{n=0}^{\infty} P_x(\tau_\alpha = n) = P_x(\tau_\alpha < \infty) = 1 \tag{3.8}$$

since $x \in S(f,r) \Rightarrow E_x[\tau_\alpha] < \infty$. It then follows from (3.8) that

$$\begin{aligned} |a_x * u(n) - \pi(\alpha)| =& |\sum_{k=0}^{n} a_x(k)(u(n-k) - \pi(\alpha))| + \sum_{k=n+1}^{\infty} a_x(k)\pi(\alpha) \\ =& |a_x * (u - \pi(\alpha))(n)| + \sum_{k=n+1}^{\infty} a_x(k)\pi(\alpha) \\ \leq& a_x * |u - \pi(\alpha)|(n) + \sum_{k=n+1}^{\infty} a_x(k)\pi(\alpha). \end{aligned}$$

The summability of $\sum_{n=0}^{\infty} a_x(n)r(n) = E_\alpha(r(\tau_\alpha))$ implies the RHS of the inequality converges to 0 as $n \to \infty$ if $r(n)|u(n) - \pi(\alpha)| \to 0$. Theorem 4.1 of [22] and Proposition 2.1 of [16] both give $r(n)|u(n) - \pi(\alpha)| \to 0$ as a result; the derivation is omitted here but the proof of Theorem 3.2.5 offers a similar result. This completes the proof for the case when Markov chain $\{x_t\}$ has an atom.

(ii) We now deal with the case when the Markov chain $\{x_t\} \subset \mathcal{X}$ does not have an atom. A common method in [13], [22], [19], [3] is to use Nummelin's Splitting

Technique as in Section 1.2 to create an atom for the Markov chain $\{z_t\} = \{x_t, a_t\} \subset \mathcal{X} \times \{0, 1\}$. The new Markov chain $\{z_t\}$ is (f, r)-ergodic as in (i), and it follows that $\{x_t\}$ is (f, r)-ergodic as well.

We show explicitly how the Nummelin's Technique is to be used to show $\{x_t\}$ is (f, r)-ergodic. By Theorem 3.1.1(i) $\{x_t\}$ has (m, δ, ν)-small set C, so as in Section 1.2 we define a new Markov chain $\{z_t\} = \{(x_t, a_t)\}$ on $\mathcal{X} \times \{0, 1\}$. The created Markov chain $\{z_t\}$ satisfies:

$\{a_t\}$ are a sequence of random variables on $\{0, 1\}$ independent of $\{x_t\}$, except when $x_t \in C$.

1. If $x_t \notin C$ then $x_{t+1} \sim P(x_t, \cdot)$
2. If $x_t \in C$ then

   with probability $\delta : a_t = 1$ and $x_{t+m} \sim \nu(\cdot)$

   with probability $(1 - \delta) : a_t = 0$ and $x_{t+m} \sim \frac{P^m(x_t, \cdot) - \delta\nu(\cdot)}{1 - \delta}$

The in-between steps x_{t+k}, $k = 1, 2, ..., m - 1$ are conditions on x_t, x_{t+m}. What is important to note is that the one step transition probabilities of the $\{x_t\}$ in the new chain $\{z_t\}$ are left unchanged.

Knowing the distribution of a_t when $x_t \in C$ we can take that distribution to be that of a_t regardless of x_t, so that for all $t \in \mathbb{N}$ $a_t \sim h$ where

$$\begin{aligned} P(a_t = 1) =& h(1) = \delta \\ P(a_t = 0) =& h(0) = 1 - \delta. \end{aligned} \tag{3.9}$$

This forces a_t to be independent from x_t including when x_t lands in C. Therefore, we have for the transition probabilities of $\{z_t\}$

$$P(z_{t+1} : x_t, a_t \sim h) = P(x_{t+1} : x_t, a_t \sim h) P(a_{t+1} : x_{t+1}, x_t, a_t \sim h) \tag{3.10}$$

$$= P(x_t, x_{t+1}) h(a_{t+1}) \tag{3.11}$$

since $\{a_t\}$ are i.i.d. by assumption, and $P(x_{t+1} : x_t, a_t \sim h) = P(x_t, x_{t+1})$ as in Section 1.2.

We now prove by induction that for all $n \in \mathbb{N}$

$$P^n(z_n : x_0, a_0 \sim h) = P^n(x_0, x_n) h(a_n). \tag{3.12}$$

Assume that (3.12) holds for all $k \leq n$, then

$$\begin{aligned} P^{n+1}(z_n : x_0, a_0 \sim h) = &\int_{\mathcal{X}} h(1) P(dy, 1 : x_0) P^n(z_n : y, 1) \\ &+ h(0) P(dy, 0 : x_0) P^n(z_n : y, 0) \end{aligned}$$

by the induction hypothesis and independence of a_t from x_t

$$\begin{aligned} = &\int_{\mathcal{X}} h(1) P(x_0, dy) P^n(y, x_n) h(a_n) \\ &+ h(0) P(x_0, dy) P^n(y, x_n) h(a_n) \\ = &P^{n+1}(x_0, x_n) h(a_n) \end{aligned}$$

so by induction, (3.12) holds for all n.

All we really need to finish is for the transition probabilities of $\{z_t\}$ to converge to $\pi_z(., \cdot)$. Unfortunately we can not just apply (i) since $\{z_t\}$ does not have an atom. With being C a (m, δ, ν)-small set, instead the m-skeleton $\{z_{tm}\}$ has an atom. Defining

$$f^{(m)} = \sum_{i=0}^{m-1} P^i f$$
$$r_m(n) = r(nm)$$

we can get our desired result by applying (i) to $\{z_{tm}\}$ if we can show $\{z_{tm}\}$ satisfies Theorem 3.1.1 with $(f^{(m)}, r_m)$. Fortunately Theorem 3.9 of [22] and Theorem 14.2.10, Theorem 14.3.3 of [12] state that an aperiodic and irreducible Markov chain $\{x_t\}$ is (f, r)-ergodic if and only if every m-skeleton $\{x_{tm+k}\}$ $k = 0, 1, 2, ..., m-1$ is $(f^{(m)}, r_m)$-ergodic.

So by (i) we know that $\{z_t\}$ has some invariant distribution $\pi_z(., \cdot)$ and is (f, r)-ergodic, so that

$$\begin{aligned}\lim_{n\to\infty} P^n((x,a) : x_0, a_0 \sim h) = &\lim_{n\to\infty} h(0)P^n((x,a) : x_0, 0) + h(1)P^n((x,a) : x_0, 1) \\ =&(h(0) + h(1))\pi_z(x,a) = \pi_z(x,a).\end{aligned}$$

However, by (3.12) we have

$$\lim_{n\to\infty} P^n((x,a) : x_0, a_0 \sim h) = \lim_{n\to\infty} P^n(x_0, x)h(0) = \lim_{n\to\infty} P^n(x_0, x)h(1)$$

which implies, with the assumed (f, r)-ergodicity of $\{z_t\}$, that $\{x_t\}$ is (f, r)-ergodic and that $P^n(x_0, \cdot)$ has a limit which is an invariant distribution $\pi(\cdot) = \frac{\pi_z(.,a)}{h(a)}$. $\square$

The last step in the proof suggests a simple relation between the case where a

Markov chain $\{x_t\}$ satisfying Theorem 3.1.1 has an atom or not. Indeed we have the following the result which allows us to focus only on Markov chains that have atoms.

Suppose $\{x_t\}$ satisfies the conditions of Theorem 3.1.1 with some $f : \mathcal{X} \to [1.\infty)$ and $r \in \Lambda$. Therefore by Theorem 1.1.3 it has a small set C and so Nummelin's Splitting Technique can be applied. Let $\{z_t = (x_t, a_t)\} \subset \mathcal{X} \times \{0,1\}$ be the Markov chain as in Section 1.2 where $\{a_t\}$ are i.i.d. with distribution h.

Theorem 3.2.4. *With the notation as above, $\{z_t\}$ and $\{x_t\}$ are (f,r)-ergodic, $\{z_t\}$ has invariant distribution π_z, $\{x_t\}$ has invariant distribution π, and there exists a constant $c(h) \in (0,\infty)$ such that*

$$r(n)\|P^n(x_0,\cdot) - \pi(\cdot)\|_f \le c(h)r(n)\|P^n((.,a) : x_0, a_0 \sim h) - \pi_z((.,a))\|_f \tag{3.13}$$

for any $a \in \{0,1\}$.

Proof. The (f,r)-ergodicity of both $\{x_t\}$ and $\{z_t\}$ follow from Theorem 3.2.3. In the proof of Theorem 3.2.3 we have that $\pi_z(.,a)/h(a) = \pi(\cdot)$ which implies, together with (3.12) that

$$r(n)\|P^n(x_0,\cdot) - \pi(\cdot)\|_f = \frac{1}{h(a)} r(n)\|P^n((.,a) : x_0, a_0 \sim h) - \pi_z((.,a))\|_f$$

for any $a \in \{0,1\}$. The proof is finished by selecting $c(h) = \max\{\frac{1}{h(0)}, \frac{1}{h(1)}\}$. $\square$

The above theorem showcases the power of Nummelin's Splitting Technique as it allows us to restrict our focus on ergodicity to Markov chains with atoms.

We have used Theorem 3.1.1 to characterize (f,r)-ergodicity when a Markov chain has an (f,r)-regular set (i.e. it is condition (iii) of Theorem 3.1.1). We wish now to use the same methods as before to show (f,r)-ergodicity when a Markov chain has an (f,r)-regular initial distribution.

Theorem 3.2.5 (Theorem 4.2 of [13], Proposition 2.5 of [18])**.** *Suppose that $\{x_t\}$ is (f,r)-ergodic. If λ and μ are (f,r)-regular, then*

$$\sum_{n=0}^{\infty} r(n) \int\int \lambda(dx)\mu(dy)\|P^n(x,\cdot) - P^n(y,\cdot)\|_f < \infty. \tag{3.14}$$

Proof. We may assume $r \in \Lambda_0$.

(i) We first deal with the case where $\{x_t\}$ has an atom $\alpha \in \mathcal{B}^+(\mathcal{X})$. By the triangle inequality it is sufficient to show that

$$\sum_{n=0}^{\infty} r(n) \int \lambda(dx)\|P^n(x,\cdot) - P^n(\alpha,\cdot)\|_f < \infty.$$

Using the definitions (3.5) and the equality (3.6) we have

$$\|P^n(x,\cdot) - P^n(\alpha,\cdot)\|_f \leq E_x(f(x_n)1_{\tau_\alpha \geq n}) + |a_x * u - u| * t_f(n)$$

where $u * t_f(n) = \sum_{j=0}^{n} P^j(\alpha,\alpha)E_\alpha[f(x_{n-j})1_{\tau_\alpha \geq n-j}] = \|P^n(\alpha,\cdot)\|_f$.

Multiplying by $r(n)$ and summing over n for the first term yields

$$\sum_{n=0}^{\infty} r(n)E_x[f(x_n)1_{\tau_\alpha \geq n}] = E_x[\sum_{n=0}^{\tau_\alpha} r(n)f(x_n)]$$

and for the second term, with $r(n) \leq r(n-j)r(j)$ by (3.1) yields

$$\sum_{n=0}^{\infty} r(n)|a_x * u - u| * t_f(n) = \sum_{n=0}^{\infty} r(n) \sum_{j=0}^{n} |a_x * u - u|(j)t_f(n-j)$$

$$
\begin{aligned}
&\leq \sum_{n=0}^{\infty}\sum_{j=0}^{n}|a_x * u - u|(j)r(j)t_f(n-j)r(n-j) \\
&= \sum_{n=0}^{\infty}|a_x * u - u|r * rt_f(n) \\
&\leq \left(\sum_{n=0}^{\infty} r(n)|a_x * u - u|(n)\right)\left(\sum_{n=0}^{\infty} r(n)t_f(n)\right).
\end{aligned}
$$

To ensure the finiteness of the second term we note that

$$
\begin{aligned}
\sum_{n=0}^{\infty} r(n)|a_x * u - u|(n) =& \sum_{n=0}^{\infty} r(n)|P_\alpha(x_n \in \alpha, \tau_\alpha \leq n) - P_\alpha(x_n \in \alpha)| \\
=& \sum_{n=0}^{\infty} r(n)|P_\alpha(x_n \in \alpha)P_\alpha(\tau_\alpha < n) \\
&+ P_\alpha(\tau_\alpha = n) - P_\alpha(x_n \in \alpha)| \\
=& \sum_{n=0}^{\infty} r(n)\left(u(n)\sum_{k=n}^{\infty} a_x(k) + a_x(n)\right) \\
\leq& \sum_{n=0}^{\infty} r(n)\sum_{k=n}^{\infty} 2a_x(k) = 2\sum_{k=0}^{\infty} a_x(k)\sum_{n=0}^{k} r(n) \\
=& 2E_x[\sum_{n=0}^{\tau_\alpha} r(n)]
\end{aligned}
$$

We now apply the bounds we attained in conjunction with the (f, r)-regular distribution λ to get

$$
\begin{aligned}
\int \lambda(dx)\sum_{n=0}^{\infty} r(n)\|P^n(x,\cdot) - P^n(\alpha,\cdot)\|_f \leq& \int \lambda(dx)E_x[\sum_{n=0}^{\tau_\alpha} r(n)f(x_n)] \\
&+ 2E_x[\sum_{n=0}^{\tau_\alpha} r(n)]\sum_{n=0}^{\infty} r(n)t_f(n) \\
=& E_\lambda[\sum_{n=0}^{\tau_\alpha} r(n)f(x_n)]
\end{aligned}
$$

$$+ 2E_\lambda[\sum_{n=0}^{\tau_\alpha} r(n)]E_\alpha[\sum_{n=0}^{\tau_\alpha} r(n)f(x_n)].$$

which is finite since λ is (f, r)-regular and the atom α is an (f, r)-regular set by lemma 1.1.5. Therefore we obtain the result we want in case $\{x_t\}$ has an atom.

(ii) The case when $\{x_t\}$ does not have an atom follows immediately from (i) and Theorem 3.2.4. $\square$

3.3 Commentary

This section characterized subgeometric ergodicity with the particulalry important Theorem 3.1.1 by Tuominen and Tweedie [22], which equates (f, r)-ergodicity with both a dirft function $PV_{k+1} \leq V_k - r(k)f + r(k)b1_C$ and bounded return time of $r(k)f(x_k)$ to a petite set. The proofs follow a similar concept to Roberts and Rosnethal's use of coupling inequality in Chapter 1; a first entrance last exit decomposition of transition probabilities was used to prove ergodicity for Markov chains with atoms, and then for atom-less Markov chains by making use of Nummelin's Splitting technique. Although the results of this chapter provide a useful and robust model, the conditions are hard to check. The next section provides a construction that uses Theorem 3.1.1 to provide more practical conditions to imply ergodicity for a large class of functions.

Chapter 4

Practical Drift Conditions for Subgeometric Rates

In this section we discuss the methods of Douc et al. [18] and Hairer [5] that extend the subgeometric ergodicity results of the last section. The results rely less on renewal process or the coupling inequality and more on concavity of rate functions. It will we be useful to assume one condition [18] hereafter.

CONDITION $\mathbf{D}(\phi, V, C)$ There exists a function $V : \mathcal{X} \to [1, \infty]$, a concave monotone nondecreasing differentiable function $\phi : [1, \infty] \to (0, \infty]$, a set $C \in \mathcal{B}(\mathcal{X})$ and a constant $b \in \mathbb{R}$ such that

$$PV + \phi \circ V \leq V + b 1_C.$$

If an aperiodic and irreducible Markov chain $\{x_t\}$ satisfies $\mathbf{D}(\phi, V, C)$ with a petite set C, and if $V(x_0) < \infty$, then $\{x_t\}$ satisfies Theorem 3.1.1(ii). Therefore $\{x_t\}$ has invariant distribution π and is $(\phi \circ V, 1)$-ergodic so that

$$\lim_{n \to \infty} \|P^n(x, \cdot) - \pi(\cdot)\|_{\phi \circ V} = 0$$

for all x in the set of π-measure 1, $\{x : V(x) < \infty\}$. By using the results of the previous section we can show a summability condition.

Proposition 4.0.1. *If an aperiodic and irreducible Markov chain $\{x_t\}$ satisfies $\mathbf{D}(\phi, V, C)$ with a small set C and $\pi(V) < \infty$, then there exists some constant $B > 0$ such that*

$$\sum_{n=0}^{\infty} \|P^n(x, \cdot) - \pi(\cdot)\|_{\phi \circ V} \leq B(1 + V(x))$$

for all x in $\{V < \infty\}$, where π is the invariant distribution of $\{x_t\}$.

Proof. The invariant distribution π of $\{x_t\}$ exists by $\mathbf{D}(\phi, V, C)$ and Theorem 1.1.6. Assuming $\pi(V) < \infty$, then by Jensen's Inequality $\pi(\phi \circ V) \leq \phi(\pi(V)) < \infty$ implying π is a $(\phi \circ V, 1)$-regular distribution. $\mathbf{D}(\phi, V, C)$ is a drift condition so by Theorem 1.1.6 there exists a petite A set with finite return time and by Proposition 1.3.2 the petite set A is $(\phi \circ V, 1)$-regular, thus Theorem 3.1.1(i) is satisfied for $(\phi \circ V, 1)$ and $\{x_t\}$ is $(\phi \circ V, 1)$-ergodic. Using the methods of Theorem 1.1.6 we also have that the distribution $\delta_x = 1_x$ is $(\phi \circ V, 1)$-regular since $E_x[\sum_{n=0}^{\tau_B} \phi \circ V(x_n)] \leq V(x) + c(B)$. The proof is completed by applying Theorem 3.2.5 with the $(\phi \circ V, 1)$-regular distributions π and δ_x for any x in the full absorbing set $\{V < \infty\}$. $\square$

4.1 Concave Rate Functions

To extend the results of Tuominen and Tweedie in the style of Douc et al. [18] and Harier [5] we consider concave functions. Let $\phi : [1, \infty) \to (0, \infty)$ be a concave nondecreasing differentiable function. Define

$$H_\phi(s) = \int_1^s \frac{dx}{\phi(x)}. \tag{4.1}$$

Then H_ϕ is a nondecreasing concave differentiable function on $[1,\infty)$. Since ϕ is concave we have that $\phi(s) \leq \phi(1) + \phi'(1)(s-1)$ for all $s \geq 1$, which implies that H_ϕ increases to infinity. Thus we can define the inverse $H_\phi^{-1} : [0,\infty) \to [1,\infty)$ which is also an increasing and differentiable function with $(H_\phi^{-1})'(x) = \phi \circ H_\phi^{-1}(x)$. We also define for $k \in \mathbb{N}, z \geq 0$ and $s \geq 1$

$$\begin{aligned} r_\phi(z) =& (H_\phi^{-1})'(z) = \phi \circ H_\phi^{-1}(z), \\ H_k(s) =& \int_0^{H_\phi(s)} r_\phi(w+k)dw = H_\phi^{-1}(H_\phi(s)+k) - H_\phi^{-1}(k), \\ V_k =& H_k \circ V. \end{aligned} \tag{4.2}$$

We will assume these definitions throughout this section. Starting from $\mathbf{D}(\phi, V, C)$ we will prove a string of statements on ergodicity and regularity using the definitions in (4.2) and Theorem 3.1.1.

Proposition 4.1.1 (Proposition 2.1 of [18]). *Assume* $\mathbf{D}(\phi, V, C)$ *. Then* r_ϕ *is log concave and for all* $k \geq 0, H_k$ *is concave and*

$$PV_{k+1} \leq V_k - r_\phi(k) + \frac{br_\phi(k+1)}{r_\phi(0)} 1_C. \tag{4.3}$$

Proof. Noting that $r'_\phi(z)/r_\phi(z) = \phi' \circ H_\phi^{-1}(z)$ is non-increasing and positive since H_ϕ^{-1} is increasing and $\phi' > 0$ is non-increasing, it follows that r_ϕ is log concave. This implies that for any $k \in \mathbb{N}$, the function $z \mapsto r_\phi(z+k)/r_\phi(z)$ is a decreasing function and so with

$$\begin{aligned}H_k'(s)\phi(s) =& H_k'(s)(H_\phi^{-1})' \circ H_\phi(s) = \left.\frac{dH_k(H_\phi^{-1}(x))}{dx}\right|_{x=H_\phi(s)} \\ =& r_\phi(H_\phi(s)+k)\end{aligned} \tag{4.4}$$

we have that $H_k'(s) = r_\phi(H_\phi(s)+k)/r_\phi(H_\phi(s))$ is decreasing and non negative which means that H_k is concave for each k.

We now turn our focus to two inequalities. Applying (4.4) and that r_ϕ is increasing we have

$$\begin{aligned}H_{k+1}(s) - H_k(s) =& \int_0^1 r_\phi(H_\phi(s)+k+u) - r_\phi(k+u)du \\ \leq& r_\phi(H_\phi(s)+k+1) - r_\phi(k) \\ =& \phi(s)H_{k+1}'(s) - r_\phi(k)\end{aligned}$$

so that we have the following inequality:

$$H_{k+1}(s) - \phi(s)H_{k+1}'(s) \leq H_k(s) - r_\phi(k). \tag{4.5}$$

Noting that for a concave differentiable function g on $[1,\infty)$, if $s+x \geq 1$ with $s \geq 1$ and $x \in \mathbb{R}$ then

$$g(s+x) - g(s) \leq g'(s)x \tag{4.6}$$

since g' is decreasing. Applying (4.6) to the concave function H_{k+1}, and Jensen's

Inequality we get for all $k \geq 0$ and $x \in \{V < \infty\}$

$$\begin{aligned} PV_{k+1}(x) =& PH_{k+1} \circ V(x) \leq H_{k+1} \circ PV(x) \\ \leq& H_{k+1}(V(x) - \phi \circ V(x) + b1_C(x)) \\ \leq& H_{k+1}(V(x)) - \phi \circ V(x) H'_{k+1}(V(x)) + b1_C(x) H'_{k+1}(V(x)) \\ \leq& H_{k+1}(V(x)) - \phi \circ V(x) H'_{k+1}(V(x)) + b1_C(x) H'_{k+1}(1). \end{aligned}$$

To the complete the proof we use (4.4) and (4.5) to show that $H'_{k+1}(1) = r_\phi(1 + k)/r_\phi(0)$ and

$$PV_{k+1}(x) \leq V_k(x) - r_\phi(k) + \frac{br_\phi(k+1)}{r_\phi(0)} 1_C(x)$$

holds for all $x \in \{V < \infty\}$, which is a full absorbing set by assumption. □

To be able use these new functions we need a way to check if r_ϕ is indeed a valid subgeometric rate function.

Lemma 4.1.2 (Lemma 2.3 of [18]). *If* $\lim_{t\to\infty} \phi'(t) = 0$ *then* r_ϕ *defined as in (4.2) belongs to the class of subgeometric functions* Λ.

Proof. We have by (4.1) and (4.2) that

$$\begin{aligned} r'_\phi(x)/r_\phi(x) =& \phi' \circ H_\phi^{-1}(x)(H_\phi^{-1})'(x)/\phi \circ H_\phi^{-1}(x) \\ =& \phi' \circ H_\phi^{-1}(x). \end{aligned}$$

With H_ϕ increasing and ϕ' non-increasing we have

$$\frac{\log r_\phi(n) - \log r_\phi(0)}{n} = \frac{1}{n} \int_0^n \frac{r'_\phi(s)}{r_\phi(s)} ds$$

$$= \frac{1}{n}\int_0^n \phi' \circ H_\phi^{-1}(s)ds,$$

and by fixing a $k < n$

$$\begin{aligned}
&\leq \frac{k}{n}\phi' \circ H_\phi^{-1}(0) + \frac{n-k}{n}\phi' \circ H_\phi^{-1}(k) \\
&\to \phi' \circ H_\phi^{-1}(k) \quad \text{as } n \to \infty \\
&\to 0 \quad \text{as } k \to \infty
\end{aligned}$$

sSince by the assumption in the hypothesis $\lim_{t\to\infty} \phi'(t) = 0$ and that by definition H_ϕ^{-1} increases to infinity. □

We now link Theorem 3.1.1 with the Condition $\mathbf{D}(\phi, V, C)$.

Proposition 4.1.3 (Proposition 2.5 of [18], Theorem 4.1(3) of [5])**.** *Let $\{x_t\}$ be an aperiodic and irreducible Markov chain that satisfies $\mathbf{D}(\phi, V, C)$ with a petite set C, a function ϕ with $\lim_{t\to\infty} \phi'(t) = 0$, and a function V such that $\{V < \infty\} \neq \emptyset$. Then there exists an invariant probability measure π for the Markov chain and for all x in the full absorbing set $\{V < \infty\}$*

$$\lim_{n\to\infty} r_\phi(n)\|P^n(x,\cdot) - \pi(\cdot)\|_{TV} = 0$$

(i.e. the Markov chain is $(1, r_\phi)$-ergodic). In addition, $\{x_t\}$ satisfies the conditions of Theorem 3.1.1 and for any probability measure λ such that $\lambda(V) < \infty$ is $(1, r_\phi)$-regular.

Proof. Since $\mathbf{D}(\phi, V, C)$ is a drift condition, by Theorem 1.1.6 the invariant distribution π of $\{x_t\}$ exists. Further, by Proposition 4.1.1 and Lemma 4.1.2 $\{x_t\}$ satisfies the drift condition

$$PV_{k+1} \leq V_k - r_\phi(k) + \frac{br_\phi(k+1)}{r_\phi(0)} 1_C$$

with $r_\phi \in \Lambda$. As in Chapter 3 we can assume $r_\phi \in \lambda_0$. By Corollary 1.1.4 we know that the petite set C is κ-petite for a distribution $a(\cdot)$ with finite mean, where κ is a maximal irreducibility measure. Therefore we have the inequality

$$1_C(x) \leq \frac{\sum_{k=1}^{\infty} P^k(x,B)}{\kappa(B)}$$

for all $B \in \mathcal{B}^+(\mathcal{X})$ and with the drift condition above by Theorem 1.1.6 we have

$$\begin{aligned} E_x[\sum_{n=0}^{\tau_B - 1} r_\phi(n)] \leq & V_0(x) + \frac{bE_x[\sum_{n=0}^{\tau_B-1} r_\phi(n+1) 1_C(x_n)]}{r_\phi(0)} \\ \leq & H_0 \circ V(x) + \frac{b\sum_{k=0}^{\infty} a(k) E_x[\sum_{n=0}^{\tau_B-1} r_\phi(n+1) 1_B(x_{n+k})]}{\kappa(B) r_\phi(0)} \\ \leq & H_0 \circ V(x) + \frac{b\sum_{k=0}^{\infty} a(k) k E_x[r_\phi(\tau_B)]}{\kappa(B) r_\phi(0)} \end{aligned}$$

where $\kappa(B) > 0$ for all sets reachable with positive probability, since $\kappa(\cdot)$ is a maximal irreducibility measure , and $\sum_{k=1}^{\infty} a(k)k = m_a < \infty$, since $a(\cdot)$ is a distribution with finite mean. To continue, we wish to compare the ratio between the partial sums of r_ϕ and r_ϕ itself, and so for ease we define the partial sum

$$R_\phi(k) = \sum_{n=0}^{k-1} r_\phi(n). \tag{4.7}$$

Since r_ϕ is subgeometric we have that by (3.2)

$$\lim_{k\to\infty}\frac{r_\phi(k)}{R_\phi(k)}=0.$$

Therefore for all $\delta>0$ there exists a constant $c(\delta)$ such that $r_\phi(k)\le \delta R_\phi(k)+c(\delta)$ for all $k\ge 1$, and so

$$E_x[R_\phi(\tau_B)]\le H_0\circ V(x)+bm_a\frac{(\delta E_x[R_\phi(\tau_B)]+c(\delta))}{\kappa(B)r_\phi(0)}$$

which means that for a small enough choice of δ we have that

$$E_x[R_\phi(\tau_B)]\le \frac{H_0\circ V(x)+c(\delta)bm_a\kappa(B)^{-1}r_\phi(0)^{-1}}{1-\delta bm_a\kappa(B)^{-1}r_\phi(0)^{-1}}. \tag{4.8}$$

The above implies that any set on which V is bounded is $(1,r_\phi)$-regular. Therefore $\{V\le n\}$, $n\in\mathbb{N}$ are a countable collection of $(1,r_\phi)$-regular sets whose union is a the full absorbing set $\{V<\infty\}$. Thus Theorem 3.1.1(iv) is satisfied for $(1,r_\phi)$, so the Markov chain is $(1,r_\phi)$-ergodic.

The only other part of the proof is that if $\lambda(V)<\infty$ then λ is $(1,r_\phi)$-regular, which follows directly from (4.8).

$\square$

4.2 Extension to Ultimately Non-decreasing Functions

Let $\mathscr{Y}$ be the class of pairs of ultimately nondecreasing functions Ψ_1 and Ψ_2 defined on $[1, \infty)$ such that $\lim_{x\to\infty} \Psi_i(x) = \infty$ for at least one $i \in \{1, 2\}$, and for all $x, y \in [1, \infty)$

$$\Psi_1(x)\Psi_2(y) \leq x + y \tag{4.9}$$

We need to check now that these functions are valid subgeometric rate functions.

Lemma 4.2.1 (Lemma 2.7 of [18]). *Using the notation as in* $\mathbf{D}(\phi, V, C)$ *, assume* $\phi : [1, \infty) \to (0, \infty)$ *is a non decreasing, differentiable, concave function with* $\lim_{t\to\infty} \phi'(t) = 0$*. For any nondecreasing function* Ψ *such that* $\Psi(x) \leq ax$ *for some constant* a *we have* $\Psi \circ r_\phi \in \Lambda$*, where* r_ϕ *is as in (4.2).*

Proof. By lemma 4.1.2 we know $r_\phi \in \Lambda$ so then $\Psi \circ r_\phi(x) \leq ar_\phi(x)$, implying $\Psi \circ r_\phi \in \Lambda$. $\square$

The final theorem in the paper by Douc et al. [18] summarizes their results, and provides a flexible theorem with seemingly many applications.

Theorem 4.2.2 (Theorem 2.8 of [18]). *Let* $\{x_t\}$ *be an aperiodic and irreducible Markov chain. Assume that* $\mathbf{D}(\phi, V, C)$ *holds with* $\phi'(t) \to 0$ *as* $t \to \infty$ *and a petite set* C *such that* V *is bounded on* C*. If* $(\Psi_1, \Psi_2) \in \mathscr{Y}$ *then there exists an invariant measure* $\pi(\cdot)$ *and for all* x *in the full absorbing set* $\{V < \infty\}$

$$\lim_n \Psi_1(r_\phi(n))\|P^n(x, \cdot) - \pi(\cdot)\|_{\Psi_2(\phi\circ V)} = 0.$$

In addition if for any probability measure λ*,* $\lambda(V) < \infty$ *then* λ *is* $(\Psi_2(\phi \circ V), \Psi_1(r_\phi))$*-regular.*

Proof. With $(\Psi_1, \Psi_2) \in \mathscr{Y}$ we have $\Psi_2(x)\Psi_1(x) \leq 2x$ so that for large enough x, $\Psi_i(x) \leq 2x$ for $i = 1, 2$. Then by the previous lemma we have that $\Psi_1 \circ r_\phi \in \Lambda$.

$$\begin{aligned} E_x[\sum_{n=0}^{\tau_C-1} \Psi_1(r_\phi(n))\Psi_2(\phi \circ V)] \leq & E_x[\sum_{n=0}^{\tau_C-1} r_\phi(n)\phi(V(x_n))] \\ \leq & E_x[\sum_{n=0}^{\tau_C-1} r_\phi(n)] E_x[\sum_{n=0}^{\tau_C-1} \phi(V(x_n))] \\ \leq & E_x[\sum_{n=0}^{\tau_C-1} r_\phi(n)] \left(V(x) + b1_C(x)\right) \end{aligned} \tag{4.10}$$

where the last line is due to $\mathbf{D}(\phi, V, C)$. By Proposition 4.1.1 we have that the $(V_k = H_k \circ V)$ satisfies a drift condition (4.3) and by the hypothesis C is bounded on V, which means that the drift condition (4.3) satisfies Theorem 3.1.1(ii) so the Markov chain is $(1, r_\phi)$-ergodic. Further, by satisfying Theorem 3.1.1(ii) with $(1, r_\phi)$ the hypothesis of Proposition 3.1.2 is satisfied and so implies that any set that is bounded on V is $(1, r_\phi)$-regular.

Therefore, V bounded on the petite set C implies that C is $(1, r_\phi)$-regular, and we have by (4.10) that C satisfies Theorem 3.1.1(i) with $(\Psi_2(\phi \circ V), \Psi_1(r_\phi))$, so the Markov chain is $(\Psi_2(\phi \circ V), \Psi_1(r_\phi))$-ergodic.

The other part of the proof with λ being $(\Psi_2(\phi \circ V), \Psi_1(r_\phi))$-regular if $\lambda(V) < \infty$ follows from (4.10) and noting that petite set C is $(\Psi_2(\phi \circ V), \Psi_1(r_\phi))$-regular.

□

The set of ultimately increasing functions includes is large class that include polynomials, logarithms, and sub exponential functions which are mentioned in [18].

4.3 Commentary

This section follows the results from [18] by Douc et al. and cleverly uses the results by Touminen and Tweedie [22] to show ergodicity for a large class of rates and norms by using ultimately nondecreasing functions in Theorem 4.2.2. Starting with $PV \leq V - \phi \circ V + b1_C$ subgeometric ergodicity is proved for $(r_\phi, \phi \circ V)$ with some conditions on ϕ in order to construct r_ϕ, then $(\Psi_1(r_\phi), \Psi_2(\phi \circ V))$ ergodicity is established for Ψ_1, Ψ_2 with some conditions like (4.9). This allows ergodicity to be checked for a wide variety of rates and norms while only needing to satisfy one drift condition $\mathbf{D}(\phi, V, C)$
.

Chapter 5

Ergodicity under Random-Time State-Dependant Drift Conditions

This section is devoted to results that follow from the discussed methods. We draw inspiration from the work of Meyn and Tweedie [12], Tuominen and Tweedie [22], and Connor and Fort [3] who all discuss ergodicity in the context of a state dependent drift condition on a deterministic time index. We wish to extend their work to obtain results on ergodicity for random time state dependent drift conditions, we draw heavily from their previous work and structure our proofs and results in a similar manner. Connor and Fort have especially interesting results studying ergodicity with a drift condition of the form $P^{n(x)}V(x) \leq \lambda V(x) + b1_C(x)$ for some function deterministic function $n : \mathcal{X} \to [1, \infty)$, which is similar to the drift conditions that we will study except for considering n to be a random stopping time.

Our motivation for studying the random time case is that for many applications information or control of a system is limited to random times. There has been significant research on stochastic stabilization [21],[10], with stochastic stability of adaptive quantizers studied in information theory [4],[6],[7] and control theory [9],[2],[14],[23].

A specific example of quantizer control over an erasure channel is given in [24] where a control of an adaptive quantizer is applied at event driven times and stochastic stability is shown using drift conditions and martingales. Another notable example comes from [17] which focuses on a control system where the transmission time from the controller to the plant has a random delay, the closed loop and open loop stability are related by studying a transmission time out $\tau_{\max}$. Event triggered feedback control systems [1],[8],[10],[20] where stopping times are event instances offers more applications for a random time drift model.

5.1 Subgeometric Ergodicity

The second condition of Theorem 3.1.1 supposes a deterministic sequence of functions (V_n) exists and satisfies a Foster-Lyapunov drift condition

$$PV_{n+1} \le V_n - r(n)f + br(n)1_C, \quad n \in \mathbb{N}.$$

We wish to apply Theorem 3.1.1 to the case where the Foster-Lyapunov drift condition holds not for every n but for a sequence of stopping times $\{\tau_n\}$. Our intention is to reveal a relation between the stopping times $\{\tau_n\}$ where a drift condition holds and the rate function r so as to imply (f,r)-ergodicity. To this end we borrow a hypothesis proposed by Yüksel and Meyn [24].

Unfortunately the techniques we used that rely on petite sets become unavailable in the random time drift setting as a petite set A for $\{x_n\}$ is not necessarily petite for $\{x_{\tau_n}\}$. We therefore rely on working with the small sets that are present in the drift conditions, and must insist that our drift function is bounded on that small set. To be able to relax the condition that V is bounded on C we can place a condition of independence on the stopping times.

Lemma 5.1.1. *Suppose $\{x_t\}$ is an aperiodic and irreducible Markov chain. if there exists sequence of stopping times $\{\tau_n\}$ independent from $\{x_t\}$ then any C that is small for $\{x_t\}$ is petite for $\{x_{\tau_n}\}$*

Proof. Let C be (m, δ, ν)-small for $\{x_t\}$. Assuming $N \geq m$, we have that

$$P^{\tau_1}(x, \cdot) = \sum_{k=1}^{\infty} P(\tau_1 = k) P^k(x, \cdot) \tag{5.1}$$
$$\geq \sum_{k=m}^{\infty} P(\tau_1 = k) \int P^m(x, dy) P^{k-m}(y, \cdot) \tag{5.2}$$
$$\geq \sum_{k=m}^{\infty} P(\tau_1 = k) \int 1_C(x) \delta \nu(dy) P^{k-m}(y, \cdot) \tag{5.3}$$

which is a well defined measure. Therefore defining

$$\kappa(\cdot) = \int \nu(dy) \sum_{k=m}^{\infty} P(\tau_1 = k) P^{k-m}(y, \cdot)$$

we have that C is $(1, \delta, \kappa(\cdot))$-small for $\{x_{\tau_n}\}$. □

Independence of stopping times $\{\tau_n\}$ from $\{x_t\}$ is a restrictive condition that event triggered systems can not fulfill. One useful example where independence of stopping times can be enforced is given in [17] where a control system over an unreliable network is affected by variable transmission delays between controller and plant. We suspect there are less restrictive methods to preserve petiteness in random sampling.

Proposition 5.1.2. *Let $\{x_t\}$ be an aperiodic and irreducible Markov chain with a small set C. Suppose there are functions $V : \mathcal{X} \to (0, \infty)$ with V bounded on C,*

$f, g : \mathcal{X} \to [1, \infty)$, a constant $b \in \mathbb{R}$ and $r \in \Lambda$ such that for an increasing sequence of stopping times $\{\tau_n\}$

$$E[V(x_{\tau_{n+1}}) : x_{\tau_n}] \leq V(x_{\tau_n}) - \delta(x_{\tau_n}) + b1_C(x_{\tau_n}) \tag{5.4}$$
$$E[\sum_{k=\tau_n}^{\tau_{n+1}-1} f(x_k) r(k) : x_{\tau_n}, \tau_n] \leq \delta(x_{\tau_n}), \tag{5.5}$$

then $\{x_t\}$ satisfies Theorem 3.1.1 with (f, r) and is (f, r)-ergodic.

Proof. We may assume $r \in \Lambda_0$. We first note that a sampled Markov chain $\{x_{\tau_n}\}$ is still a Markov chain and so we can define sampled hitting times $\gamma_B = \min\{n > 0 : \tau_n \in B\}$ for all $B \in \mathcal{B}^+(\mathcal{X})$. Since $\{x_{\tau_n}\}$ satisfies a drift condition by Theorem 1.1.6 we have that

$$\begin{aligned} E_x[\sum_{n=0}^{\gamma_C-1} \delta(x_{\tau_n})] \leq & V(x) + bE_x[\sum_{n=0}^{\gamma_B-1} 1_C(x_{\tau_n})] \\ \leq & V(x) + b \end{aligned}$$

which is finite since V is bounded on C by assumption.

With $\tau_B \leq \tau_{\gamma_B}$ for all $B \in \mathcal{B}^+(\mathcal{X})$ by definition, we have

$$E_x[\sum_{n=0}^{\tau_C-1} f(x_n) r(n)] \leq E_x[\sum_{n=0}^{\gamma_C-1} \delta(x_{\tau_n})] < V(x) + b$$

so the set C is a petite set that satisfies

$$\sup_{x \in C} E_x[\sum_{n=0}^{\tau_C-1} r(n) f(x_n)] \leq \sup_{x \in C} V(x) + b < \infty$$

which means that the Markov chain $\{x_n\}$ satisfies Theorem 3.1.1(i) and is (f,r)-ergodic.

We note that if the stopping times satisfy Lemma 5.1.1 then C is petite for the sampled Markov chain $\{x_{\tau_n}\}$ as well. This gives an alternate proof where we can drop the condition that V is bounded on C, we just apply Theorem 1.1.6 to $\{x_{\tau_n}\}$ and note by (5.5) that $\{x_t\}$ satisfies Theorem 3.1.1(i) with (f,r). $\square$

The second inequality (5.5) may be hard to check as it does not provide a criteria to check the relation between the stopping times $\{\tau_n\}$ and the rate function r.

Theorem 5.1.3. *Let $\{x_t\}$ be an aperiodic and irreducible Markov chain with a small set C. Suppose there exists a function $V : \mathcal{X} \to (0,\infty)$ with $\inf_{x\in C} V(x) = \epsilon > 0$, constants $b \in \mathbb{R}$ and $\lambda \in (0,1)$ such that for an increasing sequence of stopping times $\{\tau_n\}$*

$$E[V(x_{\tau_{n+1}}) : x_{\tau_n}] \leq \lambda V(x_{\tau_n}) + b1_C(x_{\tau_n}). \tag{5.6}$$

Then for any $r \in \Lambda$ such that $\sup_k E[\sum_{n=\tau_k}^{\tau_{k+1}} r(n-\tau_k) : x_{\tau_k}] < \infty$ and $\sup_k E[r(\tau_{k+1} - \tau_k) : x_{\tau_k}] \leq \lambda^{-1}$ we have that $\{x_t\}$ satisfies Theorem 3.1.1 with $(1,r)$ and is $(1,r)$-ergodic.

Proof. We may assume $r \in \lambda_0$ or at least that r satisfies $r(m+n) \leq r(m)r(n)$. We note that $\{x_{\tau_n}\}$ is an aperiodic and irreducible Markov chain satisfying a univariate condition as in Section 2.1 so that by Lemma 2.1.1 V is bounded on C. Defining $\gamma_B = \min\{n > 0 : x_{\tau_n} \in B\}$ for $B \in \mathcal{B}(\mathcal{X})$ gives by (2.1) that

$$E_x[\lambda^{-\gamma_B} V(x_{\tau_{\gamma_B}})] \leq V(x) + E_x[\sum_{n=0}^{\gamma_B - 1} b1_C(x_{\tau_n})]$$

for any $B \in \mathcal{B}^+(\mathcal{X})$.

Since V is bounded above and below on C we have that $C \subseteq \{\epsilon \le V \le L\}$ for some L and so

$$\sup_{x\in A} E_x[\lambda^{-\gamma_C}\epsilon] \le L + b.$$

Therefore for any $r \in \Lambda$ such that $\sup_k E[\sum_{n=\tau_k}^{\tau_{k+1}} r(n-\tau_k) : x_{\tau_k}] = M < \infty$ we have

$$\begin{aligned}
\sup_{x\in C} E_x[\sum_{n=0}^{\tau_C-1} r(n)] \le& \sup_{x\in A} E_x[\sum_{n=0}^{\tau_{\gamma_C}-1} r(n)] \\
=& \sup_{x\in A} E_x[\sum_{k=0}^{\gamma_C-1} E[\sum_{n=\tau_k}^{\tau_{k+1}-1} r(n-\tau_k) r(\tau_k) : x_{\tau_k}, x_{\tau_{k-1}}, \ldots, x_0]] \\
\le& \sup_{x\in C} E_x[\sum_{k=0}^{\gamma_C-1} M r(\tau_k)]
\end{aligned}$$

and if $\sup_k E[r(\tau_{k+1}-\tau_k) : x_{\tau_k}] \le \lambda^{-1}$ then since $r(m+n) \le r(m)r(n)$ by (3.1) we have

$$\begin{aligned}
\sup_{x\in C} E_x[\sum_{n=0}^{\tau_B-1} r(n)] \le& \sup_{x\in A} E_x[\sum_{k=0}^{\gamma_C-1} M E[r(\tau_k-\tau_{k-1}) r(\tau_{k-1}) : x_{\tau_{k-1}}, x_{\tau_{k-2}}, \ldots, x_0]] \\
\le& \sup_{x\in C} E_x[\sum_{k=1}^{\gamma_C-1} M\lambda^{-1} r(\tau_{k-1}) + M\lambda^{-1}]
\end{aligned}$$

so by iterating the above two steps we get

$$\begin{aligned}
\le& \sup_{x\in C} E_x[\sum_{k=1}^{\gamma_C-1} M\lambda^{-k}] \\
\le& \sup_{x\in C} M E_x[\lambda^{-\gamma_C} - \lambda^{-1}]/(\lambda^{-1}-1) \\
\le& \frac{M}{\epsilon}\frac{L+b-\lambda^{-1}}{\lambda^{-1}-1}.
\end{aligned}$$

Therefore $C \in \mathcal{B}^+(\mathcal{X})$ is a petite set such that $\sup_{x\in C} E_x[\sum_{n=0}^{\tau_C} r(n)]$ and so $\{x_t\}$ satisfies Theorem 3.1.1(i) with $(1, r)$ and is $(1, r)$-ergodic.

We note just like in the previous theorem if the stopping times satisfy Lemma 5.1.1 then C is a petite set for $\{x_{\tau_n}\}$ and by Theorem 1.1.4 we can focus on bounding return times for a petite set $A \subseteq \{\epsilon \leq V \leq L\}$ instead of C. This allows us to get rid of the condition that V is below on C. $\square$

One difficulty is that $\sup_k E_x[r(\tau_{k+1} - \tau_k) : x_{\tau_k}] < \lambda^{-1}$ which may not be easy to satisfy. We consider some options of modifying r to relax this condition. Suppose $\sup_k E_x[r(\tau_{k+11} - \tau_k) : x_{\tau_k}] < M$ then by Jensen's inequality we have

$$E[r(\tau_{k+1} - \tau_k)^{1/s} : \tau_k]^s \leq E[r(\tau_{k+1} - \tau_k)^{s/s} : \tau_k] < M$$

for any $s > 1$. By taking s large enough such that $M^{1/s} \leq \lambda^{-1}$ we have that $r^{1/s}$ satisfies the bound. Further, if r is subgeometric then so is $r^{1/s}$ for any fixed s, and so a suitable rate is obtained. Note that just scaling r by $\lambda^{-1}M^{-1}$ will not work as the relation $r(m + n) \leq r(m)r(n)$ will not hold for the scaled rate.

We note that Theorem 5.1.3 above is useful for proving $(1, r)$-ergodicity and Proposition 5.1.2 is really only useful for proving $(f, 1)$-ergodicity, where r, f satisfy the respective hypothesis. To extend our ability to prove more rates we use results by Douc et al. [18] on the class of pairs of ultimately non decreasing functions defined in Chapter 4.

Proposition 5.1.4. *Suppose $\{x_t\}$ is an aperiodic and irreducible Markov chain that is both $(1, r)$-ergodic and $(f, 1)$-ergodic for some $r \in \Lambda$ and $f : \mathcal{X} \to [1, \infty)$. Suppose $\Psi_1, \Psi_2 : \mathcal{X} \to [1, \infty)$ are a pair of ultimately non decreasing functions, in other words* *they satisfy* $\Psi_1(x)\Psi_2(y) \leq x + y$ *and* $\Psi_i(x) \to \infty$ *for one of* $i = 1, 2$. *Then* $\{x_t\}$ *is* $(\Psi_1 \circ f, \Psi_2 \circ r)$*-ergodic.*

Proof. Since $\Psi_2(x) \leq \Psi_1(x)\Psi_2(x) \leq 2x$ we have that $\Psi_2 \circ r \leq 2r$ and so $\Psi_2 \circ r \in \Lambda$. Since $\{x_t\}$ is irreducible and aperiodic there is only one full absorbing set S, and with $\{x_t\}$ both $(1, r)$ and $(f, 1)$ ergodic we have by Theorem 3.1.1(iv) that the full absorbing set S has two countable partitions $\{R_n\}$ and $\{F_n\}$ where $\{R_n\}$ is a collection of $(1, r)$-regular sets and $\{F_n\}$ is a collection of $(1, f)$-regular sets. We can then intersect the two partitions by

$$G_{(n,m)} = R_n \cap F_m$$

to get a third countable partition $\{G_{(n,m)}\}$ of the full absorbing set S. We then note that since $\Psi_1(x)\Psi_2(y) \leq x + y$ we have

$$\sup_{x \in G_{(n,m)}} E_x[\sum_{n=0}^{\tau_B - 1} \Psi_2 \circ r(n)\Psi_1 \circ f(x_n)] \leq \sup_{x \in R_n} E_x[\sum_{n=0}^{\tau_B - 1} r(n)] + \sup_{x \in F_m} E_x[\sum_{n=0}^{\tau_B - 1} f(x_n)]$$

which is finite for any $B \in \mathcal{B}^+(\mathcal{X})$ since R_n is $(1, r)$-regular and F_m is $(f, 1)$-regular. Therefore $\{G_{(n,m)}\}$ is a countable collection of $(\Psi_1 \circ f, \Psi_2 \circ r)$-regular sets whose unions is the full absorbing set S, so $\{x_t\}$ satisfies Theorem 3.1.1(iv) with $(\Psi_1 \circ f, \Psi_2 \circ r)$ and is $(\Psi_1 \circ f, \Psi_2 \circ r)$-ergodic. □

5.2 Geometric Ergodicity

We now give attention to a Theorems that will be extremely helpful in proving geometric ergodicity. We note that the proof of Theorem 3.2.4 does not use properties that follow from the rates being subgeometric.

The theorem we re-introduce is just picking apart the useful parts of an explicit construction using Nummelin's Splitting Technique.

Theorem 5.2.1 (Adaptation of 3.2.4)**.** *Suppose $\{x_t\}$ is an aperiodic and irreducible Markov chain with a (m, δ, ν)-small set C such that $\sup_{x \in C} E_x[\tau_C] < \infty$. Then we have:*

(i) There exists an aperiodic and irreducible Markov chain $\{z_t\} = \{(x_t, a_t)\}$ on $\mathcal{X} \times \{0,1\}$ where $\{a_t\}$ is i.i.d., independent from $\{x_t\}$, and $a_t \sim h$. In particular with C (m, δ, ν)-small, $P(a_t = 1) = h(1) = \delta$.

(ii) There exists a reachable set $S = C \times 1$ for $\{z_t\}$ such that $P^m((x,1), \cdot) = P^m((y,1), \cdot)$ for all $x, y \in \mathcal{C}$, and $\sup_{x \in C} E_x[\tau_C] \leq \sup_{x \in C} E_{(x,a)}[\tau_S] < \infty$, and by Theorem 1.3.1, $\{z_t\}$ has an invariant distribution π_z, $\{x_t\}$ has an invariant distribution π.

(iii) $P(z_n : x_0, a_t \sim h) = P(x_n : x_0)h(a_n)$ for all n.

If in addition $\{z_t\}$(or $\{x_t\}$) is ergodic for some f-norm then $\{x_t\}$(or $\{z_t\}$) is ergodic with the same rate and for the same f-norm. In particular there exists a constant $c(h) > 0$ such that if $\|P((.,\cdot) : x_0, a_0 \sim h) - \pi_z((.,\cdot))\|_f \to 0$ then

$$\|P^n(x_0, \cdot) - \pi(\cdot)\|_f \leq c(h)\|P((.,\cdot) : x_0, a_0 \sim h) - \pi_z((.,\cdot))\|_f$$

Proof. (i) By Theorem 3.1.1(i) $\{x_t\}$ has (m, δ, ν)-small set C, so as in Section 1.2 we define a new Markov chain $\{z_t\} = \{(x_t, a_t)\}$ on $\mathcal{X} \times \{0,1\}$. The created Markov chain $\{z_t\}$ satisfies:

$\{a_t\} \subseteq \{0,1\}$ are a sequence random variables independent of $\{x_t\}$, except when

$x_t \in C$.

1. If $x_t \notin C$ then $x_{t+1} \sim P(x_t, \cdot)$
2. If $x_t \in C$ then

   with probability $\delta : a_t = 1$ and $x_{t+m} \sim \nu(\cdot)$

   with probability $(1-\delta) : a_t = 0$ and $x_{t+m} \sim \frac{P(x_t,\cdot)-\delta\nu(\cdot)}{1-\delta}$

The inbetween steps x_{t+k}, $k = 1, 2, ..., m-1$ are conditions on x_t, x_{t+m}. What is important to note is that the one step transition probabilities of the $\{x_t\}$ in the new chain $\{z_t\}$ are left unchanged.

Knowing the distribution of a_t when $x_t \in C$ we can take that distribution to be that of a_t regardless of x_t, so that for all $t \in \mathbb{N}$ $a_t \sim h$ where

$$\begin{aligned} P(a_t = 1) =& h(1) = \delta \\ P(a_t = 0) =& h(0) = 1 - \delta. \end{aligned} \tag{5.7}$$

Therefore, we have for the transition probabilities of $\{z_t\}$

$$P(z_{t+1} : x_t, a_t \sim h) = P(x_{t+1} : x_t, a_t \sim h) P(a_{t+1} : x_{t+1}, x_t, a_t \sim h) \tag{5.8}$$

$$= P(x_t, x_{t+1}) h(a_{t+1}) \tag{5.9}$$

since $\{a_t\}$ are i.i.d. by assumption, and $P(x_{t+1} : x_t, a_t \sim h) = P(x_t, x_{t+1})$ as in Section 1.2, which gives the result (i).

(iii) We now prove by induction that for all $n \in \mathbb{N}$

$$P^n(z_n : x_0, a_0 \sim h) = P^n(x_0, x_n) h(a_n). \tag{5.10}$$

Assume that (5.10) holds for all $k \leq n$, then

$$P^{n+1}(z_n : x_0, a_0 \sim h) = \int_{\mathcal{X}} h(1)P(dy, 1 : x_0)P^n(z_n : y, 1) + h(0)P(dy, 0 : x_0)P^n(z_n : y, 0)$$

by the induction hypothesis and independence of a_t from x_t

$$\begin{aligned} &= \int_{\mathcal{X}} h(1)P(x_0, dy)P^n(y, x_n)h(a_n) \\ &\quad + h(0)P(x_0, dy)P^n(y, x_n)h(a_n) \\ &= P^{n+1}(x_0, x_n)h(a_n) \end{aligned}$$

so by induction, (5.10) holds for all n.

(ii) To prove (ii) we note that $\sup_{x \in C} E_{(x,a)}[\tau_S] \leq \sum_{n=0}^{\infty}(1-\delta)^n \delta n(m + \sup_{x \in C} E_x[\tau_C])$, by Nummelin's splitting technique used in (i).

Now we deal with the case where $\{z_t\}$ is ergodic with some f-norm.

$$\begin{aligned} \lim_{n\to\infty} P^n((x,a) : x_0, a_0 \sim h) &= \lim_{n\to\infty} h(0)P^n((x,a) : x_0, 0) + h(1)P^n((x,a) : x_0, 1) \\ &= (h(0) + h(1))\pi_z(x,a) = \pi_z(x,a). \end{aligned}$$

However, by (5.10) we have

$$\lim_{n\to\infty} P^n((x,a) : x_0, a_0 \sim h) = \lim_{n\to\infty} P^n(x_0, x)h(0) = \lim_{n\to\infty} P^n(x_0, x)h(1)$$

which implies, with the assumed ergodicity of $\{z_t\}$(or $\{x_t\}$), that $\{x_t\}$(or $\{z_t\}$) is ergodic and that $P^n(x_0, \cdot)$ has a limit which is an invariant distribution $\pi(\cdot) = \frac{\pi_z(.,a)}{h(a)}$. Now $\pi_z(.,a)/h(a) = \pi(\cdot)$ implies, together with (3.12) that

$$\|P^n(x_0, \cdot) - \pi(\cdot)\|_f = \frac{1}{h(a)} \|P^n((.,a) : x_0, a_0 \sim h) - \pi_z((.,a))\|_f \tag{5.11}$$

for any $a \in \{0, 1\}$. The proof is finished by selecting $c(h) = \max\{\frac{1}{h(0)}, \frac{1}{h(1)}\}$. □

This theorem and its proof, especially the last step, give a relation between the ergodicity in the case where a Markov chain has an atom and the case where it does not. Indeed, the last statement of the theorem allows us to focus only on Markov chains that have atoms when dealing with ergodicity.

Theorem 5.2.2. *Let $\{x_t\}$ be an aperiodic and irreducible Markov chain with a small set C. If there exists a function $V : \mathcal{X} \to [1, \infty)$, constants $b \in \mathbb{R}$ and $0 < \lambda < 1$ such that for an increasing sequence of stopping times $\{\tau_n\}$*

$$E[V(x_{\tau_{n+1}})|x_{\tau_n}] \leq \lambda V(x_{\tau_n}) + b1_C(x_{\tau_n})$$

and for all n

$$P(\tau_{n+1} - \tau_n = k : x_{\tau_n}) \leq B\beta^k, \quad \beta < 1$$

then $\|P^n(x, \cdot) - P^n(y, \cdot)\|_{TV} \to 0$ with a geometric rate for all $x, y \in \mathcal{X}$. If in addition $\pi(V) < \infty$ for the invariant distribution $\pi(\cdot)$ of $\{x_t\}$, then x_t is geometrically ergodic.

Proof. We first note that a sampled Markov chain $\{x_{\tau_n}\}$ is still a Markov chain and so we can define sampled hitting times $\gamma_B = \min\{n > 0 : \tau_n \in B\}$ for all $B \in \mathcal{B}^+(\mathcal{X})$.

Since the sampled chain satisfies a univariate drift condition as in Section 2.1, $\{M_n\}$ is supermartingale where

$$M_n = \lambda^{-n} V(x_{\tau_n}) - \sum_{k=0}^{n-1} b1_C(x_{\tau_k}),$$

and so we have for all $B \in \mathcal{B}^+(\mathcal{X})$,

$$E_x[\lambda^{-\gamma_B} \leq] E_x[\lambda^{-\gamma_B} V(x_{\tau_{\gamma_B}})] \leq V(x) + \sum_{n=0}^{\gamma_B - 1} 1_C(x_{\tau_n}).$$

Since $\{x_{\tau_n}\}$ satisfies a drift condition, by Lemma 2.1.1 V is bounded on C by some $M > 0$ and therefore

$$\sup_{x \in C} E_x[\lambda^{-\gamma_A}] \leq M + b \tag{5.12}$$

If we fix a $\theta > 0$ small enough such that $(1-\theta) > \beta$, we have that for all $\rho \in (1, \frac{(1-\theta)}{\beta})$

$$\begin{aligned} E_x[\rho^{\tau_{n+1}-\tau_n}] &\leq \rho^N + \sum_{k=N}^{\infty} B(\beta\rho)^k \\ &\leq \rho^N + \frac{B(\beta\rho)^N}{(1-\beta\rho)} \\ &= \rho^N(1 + \frac{B\beta^N}{\theta}) \end{aligned}$$

for all $N > 0$. Now for some N^* large enough we have that

$$(1 + B\beta^{N^*}/\theta)\lambda < 1,$$

in particular we can take

$$N^* = \lceil \log_\beta \big((\lambda^{-1} - 1)\theta / B \big) \rceil$$

. Therefore for all ρ such that

$$\rho < \big((1 + B\beta^{N^*}/\theta)\lambda \big)^{\frac{-1}{N^*}} \tag{5.13}$$

and $\rho\beta < (1 - \theta)$ still holds, in particular

$$\rho < \min\{ \frac{1-\theta}{\beta}, \big((1 + B\beta^{N^*}/\theta)\lambda \big)^{\frac{-1}{N^*}} \}$$

we have that $E_x[\rho^{\tau_{n+1}-\tau_n}] < \lambda^{-1}$, and so

$$E_x[\rho^{\tau_{\gamma_C}}] = E_x[E[\rho^{\tau_N} : \gamma_C = N]] \leq E_x[\lambda^{-\gamma_C}]. \tag{5.14}$$

Combining (5.12) and (5.14) we have that $C \in \mathcal{B}^+(\mathcal{X})$ is a (m, δ, ν)-small set with bounded return time. Therefore we can apply Theorem 5.2.1 to get an aperiodic and irreducible Markov chain $\{z_t\} = \{(x_t, a_t)\}$ with a set $S = C \times \{1\}$ such that $P^m((x, 1), \cdot) = P^m((y, 1), \cdot)$ for all $x, y \in \mathcal{X}$, $P(a_t = 1) = \delta$ and satisfying $(i) - (iii)$ of Theorem 5.2.1.

We would like to use the first entrance last exit decomposition to bound our converge rates but that requires having a Markov chain with an atom. To this end we define a new time index $J : \mathbb{N} \to \mathbb{N}$ for $\{z_t\}$

$$J(k+1) = \begin{cases} J(k) + m, & \text{if } (x_k, a_k) \in S \\ J(k) + 1, & \text{else} \end{cases}$$

so that $\{z_{J(t)}\}$ is an aperiodic and irreducible Markov chain with S as an atom.

We note that for $\{z_{J(t)}\} = \{(x_{J(t)}, a_{J(t)})\}$ we have $J(\tau_S) = \tau_S$, so by defining a sequence of stopping times $\tau_C^{k+1} = \min\{n > \tau_C^k : x_n \in C\}$, $\tau_C^0 = \tau_C$ we have

$$\begin{aligned} E_{z_0}[\rho^{J(\tau_S)}] =& E_{(x_0,a_0)}[\rho^{\tau_S}] \\ =& \sum_{k=1}^{\infty} (1-\delta)^k \delta E_{(x,0)}[\rho^{\tau_C^k}] + \delta E_{x_0}[\rho^{\tau_C}] \\ \leq& \sum_{k=1}^{\infty} (1-\delta)^k \delta E_{(x,0)}[E_{(x_{\tau_C^{k-1}},0)}[\rho^{\tau_C^k - \tau_C^{k-1}}]\rho^{\tau_C^{k-1}}] + \delta E_{x_0}[\rho^{\tau_C}] \\ \leq& \sum_{k=1}^{\infty} (1-\delta)^k \delta E_{(x,0)}[\left(\sup_{x\in C} E_{(x,0)}[\rho^{\tau_C}]\right) \rho^{\tau_C^{k-1}}] + \delta E_{x_0}[\rho^{\tau_C}] \\ \leq& \sum_{k=1}^{\infty} k(1-\delta)^k \delta \left(\rho^m + \sup_{x\in C} E_x[\rho^{\tau_C} 1_{\{\tau_C \geq m\}}]\right)^k + \delta E_{x_0}[\rho^{\tau_C}] \end{aligned} \tag{5.15}$$

where m comes from C being (m, δ, ν)-small, so by Nummelin's Splitting technique we do not know explicitly how the Markov Chain acts for $m-1$ steps after hitting C. In fact we have that for all $n \geq m$

$$\begin{aligned} E_{z_0}[\rho^{J(\tau_S)}] \leq& \sum_{k=1}^{\infty} k(1-\delta)^k \delta \left(\rho^n + \rho^n \sup_{x\in C} E_x[\lambda^{-\tau_C + n} 1_{\{\tau_C \geq n\}}]\right)^k + \delta E_{x_0}[\rho^{\tau_C}] \\ \leq& \sum_{k=1}^{\infty} k(1-\delta)^k \delta \left[\rho^n + \rho^n \lambda^n (M+b)\right]^k + \delta E_{x_0}[\rho^{\tau_C}] \end{aligned} \tag{5.16}$$

Increasing n and shrinking ρ sufficiently will make the sum finite. We now prove that

$\{z_t\}$, and therefore $\{x_t\}$, is geometrically ergodic with a rate ρ defined as to make $E_{z_0}[\rho^{J(\tau_S)}]$ finite.

Using the definitions (3.5) repeated below

$$\begin{cases} a_z(n) = P_z(\tau_S = n) \\ u(n) = P^n(S,S) \\ t_f(n) = E_S[f(x_n)1_{\tau_S \geq n}.] \end{cases} \tag{5.17}$$

and the first entrance last exit equality (3.6) we have

$$\|P^{J(n)}(z,\cdot) - P^{J(n)}(S,\cdot)\|_{TV} \leq E_z(1_{\tau_S \geq n}) + |a_z * u - u| * t_1(J(n))$$

where $u * t_1(n) = \sum_{j=0}^{n} P^j(S,S)E_S[1_{\tau_S \geq n-j}] = \|P^n(S,\cdot)\|_{TV}$.

The first term will converge to 0 since $E_z[\rho^{\tau_S}]$ is bounded, but to handle the second term we use

$$\begin{aligned} |a_z * u - u|(J(n)) =& |P_S(z_{J(n)} \in S, \tau_S \leq J(n) - P_S(z_{J(n)} \in S)| \\ =& |P_S(z_{J(n)} \in S)P_S(\tau_S < J(n)) + P_S(\tau_S = J(n)) - P_S(z_{J(n)} \in S)| \\ =& \left(u(J(n)) \sum_{k=n}^{\infty} a_z(J(k)) + a_z(J(n)) \right) \\ \leq& \sum_{k=J(n)}^{\infty} 2a_z(k) \\ =& 2E_z[1_{(\tau_S \geq J(n))}] \end{aligned}$$

We note that by (5.16) we have $E_{z_0}[\rho^{J(\tau_S)}] < \infty$. Therefore with the bound above and the property of convolutions that $f * g(n) \to (\lim_n f(n))(\sum_n g(n))$ we have that

$$\begin{aligned}
\rho^{J(n)}\|P^{J(n)}(z,\cdot) - P^{J(n)}(S,\cdot)\|_{TV} \leq & E_{(x_0,a)}[\rho^{J(n)}1_{\tau_S \geq J(n)}] \\
& + 2E_{(x_0,a)}[\rho^{J(\cdot)}1_{\tau_S \geq J(\cdot)}] * E_S[\rho^{J(\cdot)}1_{\tau_S \geq J(\cdot)}](n) \\
& \to 0 + 2\left(\lim_n E_{(x_0,a)}[\rho^{J(n)}1_{\tau_S \geq J(n)}]\right) E_S[\sum_{n=0}^{\tau_S} \rho^{J(n)}] \\
& = 0
\end{aligned}$$

where $E_S[\sum\limits_{n=0}^{\tau_S} \rho^{J(n)}] = E_S[\sum\limits_{n=0}^{\tau_S} \rho^n] \leq \frac{E_S[\rho^{\tau_S}]-\rho}{\rho-1} < \infty$. This proves that by the triangle inequality and Theorem 5.2.1 that

$$\rho^{J(n)}\|P^{J(n)}(x,\cdot) - P^{J(n)}(y,\cdot)\|_{TV} \to 0$$

for all $x, y \in \mathcal{X}$.

To translate ergodicity of $\{x_{J(t)}\}$ to ergodicity of $\{x_t\}$ we use the m-skeleton $\{x_{tm}\}$. We note that $\lfloor J(t+1)/m \rfloor \leq \lfloor J(t)/m \rfloor + 1$ so that if $T(t) = \min\{k > 0 : J(k) > tm\}$

$$\rho^{tm}\|P^{tm}(x,\cdot) - P^{tm}(y,\cdot)\|_{TV} \leq \rho^m \rho^{\lfloor J(T(t))/m \rfloor m}\|P^{\lfloor J(T(t))/m \rfloor m}(x,\cdot) - P^{\lfloor J(T(t)) \rfloor m}(y,\cdot)\|_{TV}.$$

We then have for the original chain $\{x_t\}$ that

$$\rho^t\|P^t(x,\cdot) - P^t(y,\cdot)\|_{TV} \leq \rho^m \rho^{\lfloor t/m \rfloor m}\|P^{\lfloor t/m \rfloor m}(x,\cdot) - P^{\lfloor t/m \rfloor m}(y,\cdot)\|_{TV}.$$

Using that by (5.16), $E_{(x,a)}[\rho^{\tau_S}] \leq M_0 + E_x[\rho^{\tau_C}]$ for some $M_0 > 0$ we can achieve geometric ergodicity towards the invariant distribution $\pi(\cdot)$ if $E_\pi[\rho^{\tau_C}] < \infty$. From

(5.12) and (5.14) we get that

$$E_\pi[\rho^{\tau_C}] \leq \int \pi(dx)(V(x) + b))$$

so that if $\pi(V) < \infty$ or equivalently $\sup_{x \in B} E_x[\sum_{n=0}^{\tau_B} V(x_n)] < \infty$ for some small set B, we achieve

$$\rho^n \|P^n(x, \cdot) - \pi(\cdot)\|_{TV} \to 0.$$

$\square$

Unfortunately the rate of ergodicity relies on the constants m, δ for some (m, δ, ν)-small set A, so the rate ρ can not be made explicit using only the information in the drift condition. Naively by looking at (5.16), we could achieve faster ergodic rates if we could make δ arbitrarily close to 1 which corresponds to a small set C_δ such that $\|P^m(x, .) - P^m(y, .)\| < (1 - \delta)$ for all $x, y \in C_\delta$.

5.3 Commentary

The previous chapters provide a robust and meaningful characterization of geometric and subgeometric ergodicity, however to apply the results a drift condition must be satisfied at deterministic times. This section aims to use similar methods as previous results due to Roberts and Rosenthal [19], Tuominen and Tweedie [22], Meyn and Tweedie [13], and Douc et al. [18], to show how ergodicity can follow from random time drift conditions. We aim to improve the results of this chapter by trying to find a parallel of the results in [3] by Connor and Fort, which use the drift condition $P^{n(x)}V(x) \leq \lambda V(x) + b1_C(x)$, for random time drift conditions.

Chapter 6

Conclusion

We conclude by summarizing the three approaches taken to studying ergodicity of aperiodic and irreducible Markov chains.

Section 2 describes how geometric ergodicity under the TV-norm follows from a simple drift condition as in [19] by appealing to the Coupling Inequality and Nummelin's Splitting Technique. Geometric ergodicity is an extremely strong result; however, in practice, drift conditions are not usually straightforward to check. While the results of Section 2 are important to theory they do not provide a robust enough framework to explore varied ergodic rates in practical conditions.

Section 3 characterizes subgeometric ergodicity and relates it to the concept of regularity as in [13], [12], [22]; it also provides the equivalence of regularity to a range of conditions in Theorem 3.1.1. This allows a way to study a large class of ergodic rates and provides several ways of checking whether a Markov chain is ergodic with a certain rate or norm. The results from this section provide a robust theoretical framework and show how flexible the criteria for subgeometric ergodicity can be, but a discussion of the practicality of these conditions is not included.

Section 4 follows extensions of Section 3 to a large class of rate functions by exploiting concavity and offers constructive ways of fulfilling the criteria of the previous section. These results are more explicit and varied than the rest, and focus on practical criteria to check for subgeometric ergodicity.

The central concepts of drift conditions and petite sets connect the three methods and provide cohesion between the style and ideas used. We also provide a couple of results in Section 5, including a criteria to check for subgeometric ergodicity that relies on observations of a Markov chain only on a sequence of stopping times.

Bibliography

[1] K. J. Aström and B. Bernhardsson. Comparison of Riemann and Lebesgue sampling for first order stochastic systems. In *Proceedings of the IEEE Conference on Decision and Control*, pages 2011–2016, December 2002.

[2] R. Brockett and D. Liberzon. Quantized feedback stabilization of linear systems. *IEEE Transactions on Automatic Control*, 45:1279–1289, July 2000.

[3] S.B. Connor and G. Fort. State-dependent Foster-Lyapunov criteria for subgeometric convergence of Markov chains. *stochastic processes and their applications*, 119:4176–4193, October 2009.

[4] G. Fort, S. P. Meyn, E. Moulines, and P. Priouret. ODE methods for skip-free Markov chain stability with applications to MCMC. *Annals of Applied Probability*, pages 664–707, 2008.

[5] M. Hairer. Convergence of Markov processes. Lecture Notes, 2010.

[6] J. C. Kieffer. Stochastic stability for feedback quantization schemes. *IEEE Transactions on Information Theory*, 28:248–254, March 1982.

[7] J. C. Kieffer and J. G. Dunham. On a type of stochastic stability for a class of encoding schemes. *IEEE Transactions on Information Theory*, 29:793–797, November 1983.

[8] M. Lemmon. Event-triggered feedback in control, estimation, and optimization. In *Networked Control Systems, ser. Lecture Notes in Control and Information Sciences, A. Bemporad, M. Heemels, and M. Johansson, Eds.*, pages 293–358. Springer Verlag, 2010.

[9] D. Liberzon and D. Nešić. Input-to-state stabilization of linear systems with quantized state measurements. *IEEE Transactions on Automatic Control*, 52(5):767–781, May 2007.

[10] K.H. Johansson M. Rabi and M. Johansson. Optimal stopping for event-triggered sensing and actuation. In *Proceedings of the IEEE Conference on Decision and Control*, pages 3607–3612, December 2008.

[11] S.P. Meyn and R.L. Tweedie. Stability of Markovian processes i: Criteria for discrete-time chains. *Adv. Appl. Prob.*, 24:542–574, 1992.

[12] S.P. Meyn and R.L. Tweedie. *Markov Chains and Stochastic Stability*. Springer, 1993.

[13] S.P. Meyn and R.L. Tweedie. State-dependent criteria for convergence of Markov chains. *Annals Appl. Prob.*, 4(1):149–168, 1994.

[14] G. N. Nair and R. J. Evans. Stabilizability of stochastic linear systems with finite feedback data rates. *SIAM J. Control and Optimization*, 43:413–436, July 2004.

[15] E. Nummelin. A splitting technique for Harris recurrent Markov chains. *Z. Wahrscheinlichkeitstheoric verw. Gebiete*, 43:309–318, 1978.

[16] E. Nummelin and P. Tuominen. 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, 1983.

[17] D. Quevedo, E. I. Silva, and G. C. Goodwin. Control over unreliable networks affected by packet erasures and variable transmission delays. *IEEE J. on Selected Areas in Communications*, 26:672–685, 2008.

[18] E. Moulines R. Douc, G. Fort and P. Soulier. Practical drift conditions for subgeometric rates of convergence. *Annals of Appl. Prob.*, 14(3):1353–1377, 2004.

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

[20] P. Tabuada. Event-triggered real-time scheduling of stabilizing control tasks. *IEEE Transactions on Automatic Control*, 52:1680–1685, September 2007.

[21] S. Tatikonda and S. Mitter. Control under communication constraints. *IEEE Transactions on Automatic Control*, 49(7):1056–1068, 2004.

[22] P. Tuominen and R.L. Tweedie. Subgeometric rates of convergence of f-ergodic Markov chains. *Adv. Annals Appl. Prob.*, 26(3):775–798, September 1994.

[23] S. Yüksel. Stochastic stabilization of noisy linear systems with fixed-rate limited feedback. *IEEE Transactions on Automatic Control*, 55(5):1014–1029, December 2010.

[24] S. Yüksel and S.P. Meyn. Random-time, state-dependent stochastic drift for Markov chains and applications to stochastic stabilization over erasure channels. *IEEE Trans. Autom. Control*, 58(1):47–55, January 2013.