5-2022

Some Improved Markov Chain Convergence Rates

Fun Choi John Chan
fchan@clemson.edu

Some Improved Markov Chain Convergence Rates

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Mathematical Science

---

by
Fun Choi John Chan
May 2022

Accepted by:
Dr. Robert Lund and Dr. Peter Kiessler, Committee Chair
Dr. Andrew Brown
Dr. Brian Fralix

Abstract

Explicit convergence rates to equilibrium are established for non reversible Markov chains not having an atom via coupling methods. We consider two Markov chains having the same transition function but different initial conditions on the same probability space, that is, a coupling. A random time is constructed so that subsequent to the random time the two processes are identical. Exploiting a shadowing condition, we show that it is possible to bound the tail distribution of the random time using only one of the chains. This bound gives the convergence rate to equilibrium for the Markov chain. The method is then applied to two examples; a storage model and a Gaussian auto regressive model.

Acknowledgments

I would like to express my deepest gratitude to my supervisors, Dr. Robert Lund and Dr. Peter Kiessler, who made this work possible. Their advice and guidance allowed me go through my PhD. studies with grace while conducting this research.

I am very grateful to my dissertation committee members, Dr. Andrew Brown and Dr. Brain Fralix, for their suggestions, comments, and advice, allowing my defense to be a memorable occasion.

Moreover, I really want to thank my family and Fan Guo for their understanding and support, allowing me to focus on my research and complete this dissertation. I would also like to thank my friends Haodong Li and Benwen Zhang for their help and support. I also thank everyone who contributed to this work (apologies to anyone I forgot to mention).

Lastly, I would like to thank God. His everyday guidance through the difficulties I faced facilitated completion of my Ph.D degree. I will continue to follow you and believe in you.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Let $\{X_n\}_{n=0}^{\infty}$ be an ergodic time-homogeneous Markov chain on the state space $(\mathbb{R}, \mathcal{A})$, where $\mathcal{A}$ are the Borel measurable subsets of $\mathbb{R}$ and $\mathbb{R}$ denotes the real numbers. The chain's transition kernel will be denoted by P. Proposition 6.3 of [32] shows that $\{X_n\}_{n=0}^{\infty}$ is ergodic if and only if there exists an invariant probability measure π on $(\mathbb{R}, \mathcal{A})$ satisfying

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

for all initial points $X_0 = x$. Here, $|| \cdot ||$ is the total variation norm, defined in (2.8).

The rate at which convergence takes place in the above is of considerable importance. Convergence is said to be geometric at rate $r \in (1, \infty)$ if there exists a finite constant $M(x)$ for each initial starting point $X_0 = x$ such that

$$||P^n(x, \cdot) - \pi(\cdot)|| \leq M(x) r^{-n}, \qquad n \in \{0, 1, 2, \cdots\}. \tag{1.1}$$

1.1 Literature Review

Explicitly determining values of $M(x)$ and r in (1.1) has been an active area of research [32, 24, 30, 37, 8]. Convergence of Markov chain Monte Carlo (MCMC) generated chains, for example, has immense ramifications for the simulation community [14, 33, 40]. Indeed, some MCMC chains do not rapidly converge (mix) [14, 12, 11, 10, 13].

Three texts studying Markov chain convergence on general state spaces are [32], [30], and [8]. A tactic used there to analyze chain convergence is the so-called splitting technique of [32]. In splitting methods, one constructs a so-called split chain from the original chain's one-step-ahead dynamics. Split chains retain the marginal distributions of the original chain at each fixed time, but have an atom in their state space — a set of states to which the chain will repeatedly return. Convergence rates for the original chain are obtained from those for the split chain, which is easier to analyze due to its atom. If the chain does not have an atom, a split chain is introduced to effectively create one. One then shows that the first return (hitting) time of the atom has a geometric tail, from which a convergence rate can be extracted. This has been classically accomplished by establishing a drift condition [30] for the chain.

Finite state space aperiodic irreducible Markov chains are known to converge at a geometric rate out to the second largest eigenvalue of the transition matrix P in absolute magnitude [23, 2]. Unfortunately, the scenario becomes more difficult with countably infinite state space chains, where counterexamples to the above result arise [47]. The scenario becomes even more unwieldy for chains on uncountably infinite state spaces since the chain may not return to any fixed singleton state, precluding the existence of any single-state atom. These issues not withstanding, convergence rates for general chains were established in [31] using renewal recursions and the spectral theory of bounded linear operators. The rates obtained there, while quite general, are often so poor that they are of little practical use. Explicit values of r are identified in [34] by analyzing the generating functions of successive return times of the chain to some atom; while these rates are compared to those in [31] and [37]), values of $M(x)$ are eschewed. This reference constructs a coupling time T of the chain from a bivariate drift condition, which can be used to bound $\mathbb{P}(T > n)$ for any starting point. The references [34, 20] further discuss relationships between geometric convergence rates, minorization techniques, and drift conditions. Other prominent papers studying Markov chain convergence are [7, 37, 39].

Authors have also obtained convergence rates for chains having some special structures. A dependent sample path coupling is used by [28] to obtain convergence rates of stochastically ordered chains taking values in $[0, \infty)$ that have an atom at state 0. The obtained values of r are shown to be the best possible in many cases. These results are extended to continuous time Markov processes in [27]. Both of these papers require the chain to have an atom at state 0, implying that it will return to its minimal element of the state space. While such a property holds for many storage, queueing, and reflected random walk processes, other chains do not have any atom whatsoever. Rates for other stochastic orders, such as chains having new worse than used and decreasing hazard rate states, are obtained in [23] and [29]. Applications of [27] to reflected jump diffusions were considered in [41]; M/M/1 queues were analyzed in [9]. The paper [42] also extends [28] by permitting the chain to drift to an interval set, but still assumes that state $\{0\}$ is an atom.

Markov chains that are reversible are another heavily researched stochastic structure. Bounds on the second largest eigenvalues of reversible chains with finite or countably infinite states spaces are derived in [6] and [5]; [4] studies reversible random walks on finite groups. The reference [1] is a good treatise for convergence rates of general reversible chains; [17] improved the results of [1] for reversible chains with non-negative eigenvalues and extended [1] to more general cases $m \geq 1$ in minorization conditions. Moreover, [36] shows that reversible chains have the same convergence rate when distances from the limiting distribution are measured in total variation (as in (2.8)), or via another L^2-based norm. MCMC methods, which are perhaps the most important applications of these methods, are dominated by two types of chains: Gibbs chains and Metropolis–Hastings chains. Geometric convergence rates of Gibbs chains, which are reversible, are studied in [25], [37], [38], [20], [21] and [18]. Geometric convergence rates of Metropolis algorithms, which are also reversible, are studied in [19], [16], and [31].

Authors have also studied chain convergence rates that are not geometric in nature. Subgeometric convergence rates for general chains are considered in [46], [15], and [26] in various settings.

1.2 Main Problems in this dissertation

In this dissertation, we assume that $\{X_n\}_{n=0}^{\infty}$ is a aperiodic, $\phi-$irreducible (ϕ is Lebesgue measure) and positive recurrent (Harris recurrent) time-homogeneous Markov chain on $\mathbb{R}$. Much of the analysis for discrete state chains extends to chains on $\mathbb{R}$ having a proper atom, as defined in section 2.4. In general, a chain $\{X_n\}_{n=0}^{\infty}$ on $\mathbb{R}$ does not have an atom; however, one can construct a new chain $\{\tilde{X}_n\}_{n=0}^{\infty}$ having an atom in section 2.4 by using splitting techniques in [32, 30]. For example, a simple storage model on the state space $[0, \infty)$, is defined by $X_0 = x_0$ and

$$X_n = (X_{n-1} + I_{n-1})U_n, \quad n \geq 1,$$

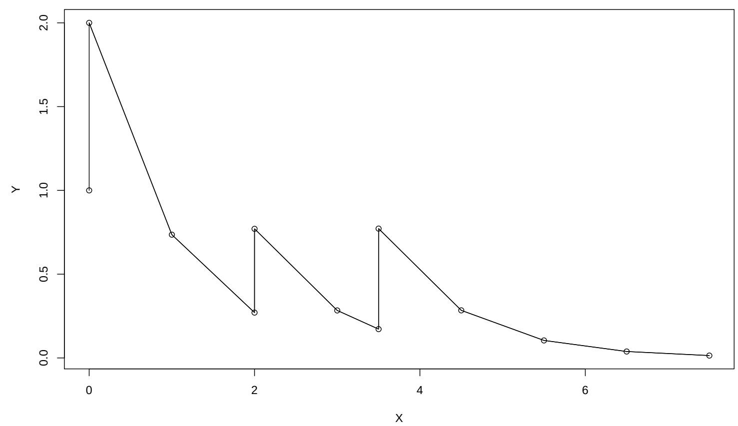

Figure 1.1: A Storage Model Chain

where U_ns are IID uniform $(0, 1)$ variables, and I_ns are IID uniform $(0, \beta)$ variables. This will never enter $\{0\}$ (see Figure 1.1); hence, the rate results in [28] do not apply. In this dissertation, we extend results in [28] to cases like the above. This work considers cases where the chain possesses some structure, such as contracting sample paths or is stochastically ordered. To illustrate the techniques, two detailed examples are presented: 1) a storage process chain, which is stochastically ordered in its initial state but cannot return to the minimal element $\{0\}$ in its state space, and 2) a first order Gaussian autoregressive model where sample paths of the chain contract in their initial level. Also, this dissertation will identify conditions under which the chain will converge quickly to a stationary (limiting) measure $\pi(\cdot)$.

The rest of this dissertation proceeds as follows. The next chapter details the background needed. Chapter 3 presents our main results. Chapter 4 and 5 consider applications of the methods to storage model chains and first order autoregressive chains.

Chapter 2

Background

In this chapter, we will setup the background for this dissertation. In this dissertation, it assumes that $\{X_n\}_{n=0}^{\infty}$ is an aperiodic, irreducible (ϕ-irreducible), and positive recurrent (Harris recurrent) time-homogeneous Markov chain on a general state space that is a subset of $\mathbb{R}$. Our overarching goal is to identify under what conditions the chains will have a nice convergence rate (say converge geometrically rapidly) or converge quickly to a stationary (limiting) measure $\pi(\cdot)$.

A time homogeneous Markov chain on $\mathbb{R}$ is a sequence of real valued random variables $\{X_n\}_{n=0}^{\infty}$ such that for each $n = 1, 2, \cdots$ and $x_0, x_1, \cdots, x_{n-1} = x \in \mathbb{R}$ and $A \subset \mathbb{R}$,

$$\begin{aligned}\mathbb{P}(X_n \in A | X_{n-1} = x_{n-1}, \cdots, X_0 = x_0) &= \mathbb{P}(X_n \in A | X_{n-1} = x) \\ &= \mathbb{P}(X_1 \in A | X_0 = x).\end{aligned}$$

We will use the notation

$$P(x, A) = \mathbb{P}(X_1 \in A | X_0 = x),$$

and

$$P^{n+1}(x, A) = \mathbb{P}_x(X_{n+1} \in A),$$

where $P(x, A)$ is transition probability.

Our definition resembles the one typically used for discrete state time-homogenous Markov chains. The main difference is that the set A replaces points in E. The main tool for studying discrete state Markov chains is the one-step-ahead transition matrix. This will be replaced by a transition probability (kernel), which we now define.

Unless otherwise specified, all sets encountered are assumed Borel measurable subsets of $\mathbb{R}$ (this class is called $\mathcal{A}$) and all functions are assumed to be Borel measurable on $\mathbb{R}$.

Definition 1 *A transition probability (kernel) on $\mathbb{R}$ is a map $P : \mathbb{R}\times\mathcal{A}\to \bar{\mathbb{R}}_+$ satisfying the following two conditions:*

1. *for any fixed set $A \in \mathcal{A}$, the function $P(x, A)$ is measurable in x;*

2. *for any fixed state $x \in \mathbb{R}$, the set function $P(x, \cdot)$ is a probability measure on $\mathbb{R}$.*

The first condition allows us to take iterates of P. That is, since $P(x, A)$ is measurable in x for each fixed A, we can integrate it with respect to the probability $P(x, dy)$; specifically,

$$P^2(x, A) = \int_{\mathbb{R}} P(x, dy)P(y, A).$$

P^2 is a two-step-ahead transition probability (kernel). We can thus inductively define the n-step-ahead transition probability (kernel) P^n, $n = 3, 4, \cdots$, by

$$P^n(x, A) = \int_{\mathbb{R}} P(x, dy)P^{n-1}(y, A) = \int_{\mathbb{R}} P^{n-1}(x, dy)P(y, A).$$

In the center expression, P^{n-1} is being viewed as a measurable function; in the rightmost expression, it is being used as a probability measure.

2.1 Irreducibility and Aperiodicity

We will need notions of irreducibility and aperiodicity for chains on $\mathbb{R}$. Toward this, we make the following definitions.

Recall that a discrete state Markov chain is called irreducible if for every pair of states i, j, there is a nonnegative integer n such that $p_{i,j}^{(n)} = P(X_n = j|X_0 = i) > 0$. That is, one can get to any state from any other state. This is not true for general Markov chains on $\mathbb{R}$. Indeed, for all $x, y \in \mathbb{R}$, it may be the case that $P^n(x, \{y\}) = 0$ for all n. To see this, suppose that $\{X_n\}_{n=0}^{\infty}$ is a sequence of i.i.d. uniform$(0, 1)$ random variables. In this case, for each x, $P^n(x, \cdot)$ is the uniform distribution on $(0, 1)$ and $P^n(x, \{y\}) = 0$ for all n. However, it may also be the case that for each $x \in \mathbb{R}$ and each open set U, there exists an n such that $P^n(x, U) > 0$. This says that for each pair $x, y \in \mathbb{R}$, that starting in x, the chain can get arbitrarily close to y. (We could call this irreducibility since in most examples we encounter, this will be the case.)

We now present a more rigorous definition. Let ϕ be a measure on $\mathbb{R}$; for example, ϕ may be the Lebesgue measure. We say that ϕ is a maximal irreducibility measure if

(i) For every $x \in \mathbb{R}$ and every set A with $\phi(A) > 0$, there is an n such that $P^n(x, A) > 0$. We denote this by $x \to A$.

(ii) If ψ is any other measure satisfying (i), then ψ is absolutely continuous with respect to ϕ.

In this case, we say that the chain is ϕ-irreducible. In the discrete state case, the ϕ-irreduciblity measure for an irreducible chain is the counting measure. We will call a ϕ-irreducible Markov chain irreducible; that is, we drop ϕ from further notations.

Let $P(\cdot, A)$ be the one-step-ahead transition probability of a Markov chain $\{X_n\}_{n=0}^{\infty}$ on $\mathbb{R}$. Then, $P(x, \mathbb{R}) = 1, \text{ for all } x \in \mathbb{R}$.

2.1.1 Periodicity

Consider the discrete state chain in [3] having state space $\{1, 2, 3, 4, 5, 6, 7\}$ and transition matrix

$$P = \begin{bmatrix} 0 & 0 & \frac{1}{2} & \frac{1}{4} & \frac{1}{4} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & \frac{2}{3} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{3} & \frac{2}{3} \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & \frac{3}{4} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{3}{4} & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

The state space can be partitioned into three disjoint sets $A_1 = \{1, 2\}$, $A_2 = \{3, 4, 5\}$, and

$A_3 = \{6, 7\}$, such that for $i \in A_1$,

$$\begin{cases} P(i, A_2) = 1, \\ P^2(i, A_3) = 1, \\ P^3(i, A_1) = 1. \end{cases}$$

A similar cyclic property holds for $i \in A_2$ and $i \in A_3$. It follows that for $k \in \{1, 2, 3\}$, $i \in A_k$, and $n \in \mathbb{N}$,

$$P^{3n}(i, A_j) = \begin{cases} 1, & \text{if } j = k, \\ 0, & \text{otherwise.} \end{cases}$$

Thus, the chain cycles through A_1 to A_2 to A_3 and then repeats. Hence, $\lim_{n\to\infty} P^{(n)}$ does not exist. One can verify that

$$P^3 = \begin{bmatrix} \frac{71}{192} & \frac{121}{192} & 0 & 0 & 0 & 0 & 0 \\ \frac{29}{72} & \frac{43}{72} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{14}{36} & \frac{3}{36} & \frac{19}{36} & 0 & 0 \\ 0 & 0 & \frac{19}{48} & \frac{3}{32} & \frac{49}{96} & 0 & 0 \\ 0 & 0 & \frac{13}{32} & \frac{7}{64} & \frac{31}{64} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{157}{288} & \frac{131}{288} \\ 0 & 0 & 0 & 0 & 0 & \frac{111}{192} & \frac{81}{192} \end{bmatrix}$$

is a block diagonal matrix and hence cannot be the transition matrix for an irreducible Markov chain. Note that $\lim_{n\to\infty} P^{(3n)}$ does exist but this limit depends on X_0.

A ϕ-irreducible Markov chain on $\mathbb{R}$ is said to have period d if there exists a partition $\{A_i\}_{i=0}^{d-1}$ of $\mathbb{R}$ with the smallest integer d such that for $i = 0, \ldots, d-2$, $x \in A_i$, we have $P(x, A_{i+1}) = 1$; and if $x \in A_{d-1}$, we have $P(x, A_0) = 1$. From the above, if $x \in A_i$, $i = 0, \ldots, d-1$,

$$P^n(x, A_i) = 1 \text{ if and only if } n \text{ is a multiple of } d.$$

Definition 2 *An irreducible Markov chain* $\{X_n\}_{n=0}^{\infty}$ *on* $\mathbb{R}$ *is called aperiodic when* $d \equiv 1$.

We will mostly be interested in irreducible Markov chains on $\mathbb{R}$ that are aperiodic.

For a discrete state chain that is irreducible, either all states are recurrent or all states are transient. Irreducible chains on $\mathbb{R}$ obey a similar dichotomy. A set B is called transient if starting from any $x \in B$, the probability the chain returns to B only finitely many times equals 1. The Markov chain is called recurrent if whenever $\phi(A) > 0$, for ϕ almost every $x \in A$,

$$\mathbb{P}_x(X_n \in A \text{ for infinitely many } n) = 1.$$

This dichotomy implies that either the chain is recurrent or $\mathbb{R}$ can be written as a countable union of transient sets. A slightly stronger notion of recurrence is Harris recurrence, which essentially drops the ϕ almost every x from the definition of recurrence.

Definition 3 *An irreducible chain* $\{X_n\}_{n=0}^{\infty}$ *on* $\mathbb{R}$ *is called Harris recurrent if for all* $B \subset \mathbb{R}$ *with* $\phi(B) > 0$*, and all* $x \in \mathbb{R}$, $h_B^{\infty}(x) \equiv 1$*, where*

$$h_B^{\infty}(x) = \mathbb{P}(X_n \in B \textit{ for infinitely many } n | X_0 = x).$$

We will be primarily interested in irreducible Harris recurrent chains.

Suppose an irreducible Markov on a countable state space is recurrent. Then, for all states x and y the Markov chain will visit y infinite often many times with probability one starting from x and hence is also Harris recurrent. If the state space is not countable, there may be a $\phi-$null set N such that there exists A with $\phi(A) > 0$ and an $x \in N$ for which A cannot be reach from x.

2.2 Renewal and Regenerative Techniques and Coupling Methods

Two main methods have been used to obtain convergence rate bounds: 1) renewal and regenerative techniques [32] and [31], and 2) coupling methods [37, 24, 28]. Other methods do exist; for example, [43] develop iterative random functions and the so-called one-shot coupling method, but even this is based heavily on 1) and 2). We review renewal theory methods first.

Let π denote the stationary distribution of the chain and let $\tau_C := \inf\{n > 0 : X_n \in C\}$ be the first time the chain enters a fixed set of states C. Then [32] establishes an inequality of form

$$||P^n(x,\cdot) - \pi(\cdot)|| \leq P_x(\tau_C \geq n) + \text{extra positive terms.} \tag{2.1}$$

While the particluar form of the extra terms is not important now, significant work is usually required to deal with them.

Coupling methods work a bit more cleanly when they work. The general idea is to take two chains $\{X_n\}_{n=0}^{\infty}$ and $\{X'_n\}_{n=0}^{\infty}$ having the same transition kernel P, but with different starting conditions. For these, we start X_0 at $X_0 = x$ and allow X'_0 to be random having the stationary distribution π. By stationarity, X'_n also has distribution π for each $n \geq 1$. Coupling works by analyzing the time at which the two chains meet: $T := \inf\{n \geq 0 : X_n = X'_n\}$. Then define a new chain $\{X''_n\}_{n=1}^{\infty}$ via

$$X''_n = \begin{cases} X_n, & n > T, \\ X'_n, & n \leq T. \end{cases}$$

One can show that $\{X''_n\}_{n=1}^{\infty}$ is stationary. Since $X_n = X''_n$ when $T < n$, we have

$$\mathbb{P}_{x,\pi}(X_n \in A, T \leq n) = \mathbb{P}_{x,\pi}(X''_n \in A, T \leq n).$$

Note that $\mathbb{P}_{x,\pi}(T < \infty)$ may not equal unity.

The famous coupling inequality for strong Markov chains (see e.g. [24]) is

$$||P^n(x,\cdot) - \pi(\cdot)|| \leq \mathbb{P}_{x,X'_0}(T > n), \tag{2.2}$$

which follows from the following argument.

$$
\begin{aligned}
&||P^n(x,\cdot)-\pi(\cdot)|| = \sup_A |P^n(x,A)-\pi(A)| \\
=&\sup_A |\mathbb{P}_x(X_n \in A) - \mathbb{P}_{X_0'}(X_n^{''} \in A)| \\
=&\sup_A |\mathbb{P}_{x,X_0'}(X_n \in A, T > n) + \mathbb{P}_{x,X_0'}(X_n \in A, T \leq n) \\
&- \mathbb{P}_{x,X_0'}(X_n^{''} \in A, T > n) - \mathbb{P}_{x,X_0'}(X_n^{''} \in A, T \leq n)| \\
=&\sup_A |\mathbb{P}_{x,X_0'}(X_n \in A, T > n) - \mathbb{P}_{x,X_0'}(X_n^{''} \in A, T > n)| \\
\leq&\mathbb{P}_{x,X_0'}(T > n).
\end{aligned} \tag{2.3}
$$

Here, the subscripts of x and X_0' demarcate the starting points of the two chains. Significant convergence rate information can be extracted from the coupling time T. For example, if $\boldsymbol{E}_{x,X_0'}[r^T]$ can be proven to be finite for some $r > 1$, then convergence in (2.10) holds at the geometric rate r^{-1}.

Complications in the above, however, arise. First and foremost, on continuous state spaces, T may not be finite (may never occur). It can also be difficult to calculate the distribution of T exactly. However, if $\boldsymbol{E}_{x_0,X_0'}[r^T]$ can be proven finite for an $r > 1$, then Markov's inequality gives

$$
||P^n(x,\cdot)-\pi(\cdot)|| \leq r^{-n}\boldsymbol{E}_{x_0,X_0'}[r^T]
$$

as $n \to \infty$. This identifies r as a geometric convergence rate and $M(x) = \boldsymbol{E}_{x_0,X_0'}[r^T]$ as a first constant.

2.3 Other Stochastic Structures

Some additional stochastic structure, such as chain reversibility or stochastic monotonicity, is often present in applications and can be invoked to obtain good convergence rates. A stochastically ordered chain is a chain that is higher at all times when it starts higher in a stochastic sense. Quantifying this, if $\{X_n\}_{n=0}^{\infty}$ and $\{X_n'\}_{n=0}^{\infty}$ are copies of the chain with X_0' stochastically larger than X_0, then X_n' is stochastically larger than X_n for all $n \geq 1$. Here, X is stochastically larger than Y means that $\mathbb{P}(X > x) \geq \mathbb{P}(Y > x)$ for all real x. The texts [24], [23], and [44] are good places to read about stochastic orderings.

Every Markov chain can be written in its simple representation form:

$$X_{n+1} = g(X_n, U_{n+1}), \quad n \geq 1, \tag{2.4}$$

for some function g and initial state X_0. In this representation, $\{U_i\}_{i=1}^{\infty}$ is IID. For a stochastically ordered chain, for each fixed u, $g(x,u)$ must be monotone non-decreasing in x.

Suppose that $g(x,u)$ is nondecreasing in x for each fixed u and consider two trajectories of a stochastically ordered chain — say $\{X_n\}$ and $\{X'_n\}$ — that are governed by (2.4) and use the same $\{U_n\}_{n=1}^{\infty}$. For starting conditions, we assume that $X_0 \geq X'_0$ pointwise. Then for all n, $X_n \geq X'_n$. This is essentially Strassen's Theorem [45] that every stochastically ordered chain can be made pathwise ordered.

2.4 Splitting Techniques

Much of the analysis of irreducible recurrent chains on a countable state space is performed by exploiting the fact that returns to a fixed state form a recurrent renewal process. Before showing how to extend this approach to chains on $\mathbb{R}$, we consider the following example. Let $U_0, U_1, \ldots$ be a sequence of iid uniform (0,1) random variables. This sequence forms an irreducible recurrent Markov chain. Let $x \in (0,1)$. Then the probability of returning to state x at any fixed time (and hence ever) is 0. Thus, the return times to state x form a transient renewal process: with probability one, the chain does not return to state x.

Suppose on the other hand that there exists $A \subset \mathbb{R}$ that satisfy:

(i) For any points $x, y \in A$,

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

(ii) For all $z \in \mathbb{R}$, $z \to A$ as defined on page 7 in section 2.1.

The subset A is called a proper atom of the chain. Return times to the set A form a renewal process and much of the analysis for discrete state chains extends to chains on $\mathbb{R}$ having a proper atom.

Typically, a chain $\{X_n\}_{n=0}^{\infty}$ on $\mathbb{R}$ does not have an atom; however, one can construct a new chain $\{\tilde{X}_n\}_{n=0}^{\infty}$ on $\mathbb{R}$ such that

(i) $\{\tilde{X}_n\}_{n=0}^{\infty}$ exhibits a regenerative structure, and

(ii) the distribution of $\{\tilde{X}_n\}_{n=0}^{\infty}$ is the same as that of $\{X_n\}_{n=0}^{\infty}$.

This can be accomplished using a so-called minorization condition in [32, 30]. Suppose that one can find an $\delta > 0$, some set $C \subset \mathbb{R}$, and some probability measure ν with $\nu(C) = 1$ that satisfies

$$P(x, A) \geq \delta 1_C(x)\nu(A) \equiv (\delta 1_C \otimes \nu)(x, A). \tag{2.5}$$

Then we call $(\delta 1_C, \nu)$ an atom. We need to define a sequence $\{Y_n\}_{n=0}^{\infty}$ of Bernoulli trials, constructed from $\{X_n\}_{n=0}^{\infty}$ to define the $\{\tilde{X}_n\}_{n=0}^{\infty}$ chain. First, to do this, initialize $\tilde{X}_0 = X_0$. Suppose $X_0 = x$. Then, If $x \notin C$, we do not flip the coin and $\tilde{X}_1$ follows the transition probability $P(x, \cdot)$. We flip a coin that has probability δ of being heads if $x \in C$. If the toss results in heads, we set $Y_0 = 1$ and choose $\tilde{X}_1$ according to the distribution ν . If the result of the toss is tails, we set $Y_0 = 0$ and generate $\tilde{X}_1$ according to the transition function

$$Q(x, \cdot) = \frac{P(x, \cdot) - \delta\nu(\cdot)}{1 - \delta}. \tag{2.6}$$

Now suppose that $\tilde{X}_0, \ldots, \tilde{X}_{n-1}$, and $\tilde{X}_n = x_n$ have been constructed. As in the initial step, we flip a coin independently of $(\tilde{X}_0, Y_0), \ldots, (\tilde{X}_{n-1}, Y_{n-1})$ having probability δ of being heads when $X_n = x_n \in C$. If the toss results in heads, we set $Y_n = 1$ and generate $\tilde{X}_{n+1}$ according to the distribution ν. If the result of the toss is tails, we set $Y_n = 0$ and generate $\tilde{X}_{n+1}$ from the transition function

$$Q(x_n, \cdot) = \frac{P(x_n, \cdot) - \delta\nu(\cdot)}{1 - \delta}.$$

.

2.4.1 $\{\tilde{X}_n\}_{n=0}^{\infty}$ is a Markov chain having the same distribution as $\{X_n\}_{n=0}^{\infty}$

This essentially follows from the observation that Y_n is conditionally independent of $\tilde{X}_0, \ldots, \tilde{X}_n$, $Y_0, \ldots, Y_{n-1}$ given $\tilde{X}_n$, and $\tilde{X}_{n+1}$ is conditionally independent of $\tilde{X}_0, \ldots, \tilde{X}_{n-1}$, $Y_0, \ldots, Y_{n-1}$ given $(\tilde{X}_n, Y_n)$. We first note that $\tilde{X}_0 = X_0$ by construction. All that remains to do is to show that

$\{\tilde{X}_n\}_{n=0}^{\infty}$ is Markov chain having the same transtion kernel as $\{X_n\}_{n=0}^{\infty}$. Using the tower property of conditional expectation to obtain the first equality and the law of total probability to obtain the second

$$
\begin{aligned}
&\mathbb{P}(\tilde{X}_{n+1} \in B | \tilde{X}_0, \dots, \tilde{X}_{n-1}, \tilde{X}_n) \\
&= \boldsymbol{E}[\mathbb{P}(\tilde{X}_{n+1} \in B | (\tilde{X}_0, Y_0), \dots, (\tilde{X}_{n-1}, Y_{n-1}), \tilde{X}_n) | \tilde{X}_0, \dots, \tilde{X}_{n-1}, \tilde{X}_n] \\
&= \boldsymbol{E}[\mathbb{P}(\tilde{X}_{n+1} \in B, Y_n = 1 | (\tilde{X}_0, Y_0), \dots, (\tilde{X}_{n-1}, Y_{n-1}), \tilde{X}_n) | \tilde{X}_0, \dots, \tilde{X}_{n-1}, \tilde{X}_n] \\
&+ \boldsymbol{E}[\mathbb{P}(\tilde{X}_{n+1} \in B, Y_n = 0 | (\tilde{X}_0, Y_0), \dots, (\tilde{X}_{n-1}, Y_{n-1}), \tilde{X}_n) | \tilde{X}_0, \dots, \tilde{X}_{n-1}, \tilde{X}_n] \\
&= \boldsymbol{E}[\delta 1_C(\tilde{X}_n) \nu(B) | \tilde{X}_0, \dots, \tilde{X}_{n-1}, \tilde{X}_n] \\
&+ \boldsymbol{E}[P(\tilde{X}_n, B) - \delta 1_C(\tilde{X}_n) \nu(B) | \tilde{X}_0, \dots, \tilde{X}_{n-1}, \tilde{X}_n] \\
&= P(\tilde{X}_n, B),
\end{aligned}
$$

where in the last equality we note $P(\tilde{X}_n, B) - \delta 1_C(\tilde{X}_n)\nu(B)$ is $\tilde{X}_0, \dots, \tilde{X}_{n-1}, \tilde{X}_n$ measurable. Thus, $\{\tilde{X}_n\}_{n=0}^{\infty}$ is Markov chain having transition kernel P in [32, 30]. and the proof is complete. It is also the case that $\{(\tilde{X}_n, Y_n)\}_{n=0}^{\infty}$ is Markov chain having transition kernel

$$R((x,i); B \times \{1\}) = \delta 1_C(x) P(x; B)$$

and

$$R((x,i); B \times \{0\}) = (1 - \delta 1_C(x)) P(x; B).$$

This follows from the proof above,

$$\mathbb{P}(Y_{n+1} = 1 | (\tilde{X}_0, Y_0), \dots, (\tilde{X}_n, Y_n), \tilde{X}_{n+1}) = \delta 1_C(\tilde{X}_{n+1})$$

and

$$\mathbb{P}(Y_{n+1} = 0 | (\tilde{X}_0, Y_0), \dots, (\tilde{X}_n, Y_n), \tilde{X}_{n+1}) = 1 - \delta 1_C(\tilde{X}_{n+1}).$$

Since, for all $x \in C$,

$$R((x,1); B \times \{1\}) = \nu(B)\delta$$

and

$$R((x,1); B \times \{0\}) = \nu(B)(1 - \delta),$$

it follows that $C \times \{1\}$ is a proper atom for the Markov chain $\{(\tilde{X}_n, Y_n)\}_{n=0}^{\infty}$.

2.5 Convergence Rate Definitions

Consider a Markov chain $\{X_n\}_{n=0}^{\infty}$ on $\mathbb{R}$ with transition probability P and suppose that $V(\cdot)$ is a function on $\mathbb{R}$. Define

$$PV(x) = \int_{\mathbb{R}} V(y)P(x, dy) = \boldsymbol{E}_x[V(X_1)].$$

Definition 4 *A chain $\{X_n\}_{n=0}^{\infty}$ on $\mathbb{R}$ is said to satisfy a drift condition if there exists a measurable function $V(x) \geq 1$, a set C, and constants $1 < r_0$ and $0 < b < \infty$, such that*

$$PV(x) \leq r_0^{-1}V(x) + b1_C(x), \qquad \textit{for all } x. \tag{2.7}$$

Our convergence rates will be phrased in terms of total variational norm, which is defined as follows:

Definition 5 *The total variation norm between the two probability measures on $\mathbb{R}$, denoted by $\nu_1(\cdot)$ and $\nu_2(\cdot)$, is*

$$||\nu_1(\cdot) - \nu_2(\cdot)|| = \sup_A |\nu_1(A) - \nu_2(A)|, \tag{2.8}$$

where the supremum is taken over all Borel sets A.

From this definition, it follows that if f is any Borel measurable function on $\mathbb{R}$ bounded by [0,1], then

$$||\nu_1(\cdot) - \nu_2(\cdot)|| = \sup_f \left| \int f d\nu_1 - \int f d\nu_2 \right|.$$

Significant issues arise in the treatment of chains on continuous state spaces. For but one simple but troublesome example, consider the first order autoregressive chain

$$X_n = \frac{1}{2}X_{n-1} + Z_n, \quad n \geq 1,$$

with $X_0 = 0$ and $\{Z_n\}_{n=1}^{\infty}$ IID with $P(Z_n = 0) = P(Z_n = 1/2) = 1/2$. Induction verifies that this recursion produces an X_n that is uniformly distributed over the discrete set $\{0, 1/2^n, \ldots, (2^{n-1} -$

$1)/2^n\}$. It follows that X_n converges to a uniformly distributed limiting distribution π over [0,1] as $n \to \infty$. However, with I being the set of all irrational numbers in [0,1], $\mathbb{P}(X_n \in I) = 0$ and $\pi(A) = 1$ for every finite n. Hence,

$$\sup_A |P(X_n \in A) - \pi(A)| = 1$$

for every $n \geq 1$. This chain does not converge in total variation at all, although it is a nice and stable stochastic recursion. The problem here is that irrational points in the state space cannot be reached when X_0 is rational-valued; that is, this chain is not ϕ-irreducible. This illustrates some of the complications that arise when working on continuous spaces.

The following theorem was proven in [33]:

Theorem 6 *If a chain $\{X_n\}_{n=0}^{\infty}$ on $\mathbb{R}$ is ϕ-irreducible and aperiodic and has a stationary distribution $\pi(\cdot)$, then for π-a.e. $x \in \mathbb{R}$,*

$$\lim_{n\to\infty} ||P^n(x,\cdot) - \pi(\cdot)|| = 0. \tag{2.9}$$

Unfortunately, Theorem 6 does not say anything about convergence rates. To quantify geometric convergence, we make the following definitions.

Definition 7 *The chain $\{X_n\}_{n=0}^{\infty}$ with stationary distribution $\pi(\cdot)$ is uniformly geometrically ergodic if*

$$\sup_x ||P^n(x,\cdot) - \pi(\cdot)|| \leq M\rho^n, \qquad n = 1, 2, 3, \cdots,$$

for some $\rho < 1$ and $M < \infty$.

Our next definition, a weaker condition than uniform ergodicity, is called geometric ergodicity and allows for a different M for each initial chain state x.

Definition 8 *The chain $\{X_n\}_{n=0}^{\infty}$ on $\mathbb{R}$ with stationary distribution $\pi(\cdot)$ is geometrically ergodic if*

$$||P^n(x,\cdot) - \pi(\cdot)|| \leq M(x)\rho^n, \qquad n = 0, 1, 2, 3, \cdots, \tag{2.10}$$

for some $\rho < 1$ and $0 \leq M(x) < \infty$ for π a.e $x \in \mathbb{R}$.

The following theorem was proven in [33]:

Theorem 9 *Suppose that* $\{X_n\}_{n=0}^{\infty}$ *is a* ϕ*-irreducible, aperiodic Markov chain on* $\mathbb{R}$ *with stationary distribution* $\pi(\cdot)$*; satisfying the minorization condition (2.5) for some* $C \subset \mathbb{R}$*,* $\delta > 0$*, and probability measure* $\nu(\cdot)$*, and satisfying the drift condition (2.7) for some constant* $r_0 > 1$*,* $0 < b < \infty$*, and a drift function* $V(x) \geq 1$ *with* $V(x) < \infty$ *at least one* $x \in \mathbb{R}$*. Then the chain is geometrically ergodic.*

Although Theorem 9 tells us that the chain is geometrically ergodic, it does not tell us how to find $M(x)$. Later, we will improve Theorem 9 via the coupling method defined in Section 3.1 and split techniques and identify $M(x)$. Split chains will be used to bound $\boldsymbol{E}_{x,\pi}[r^T]$, thereby identifying good convergence rates. In the next chapter, we apply coupling and splitting techniques with minorization and drift conditions to contracting sample paths or stochastic orderings paths to identify $M(x)$ and good convergence rates $r > 1$.

Chapter 3

New Results

This chapter constructs a coupling of two chains $\{X_n\}_{n=0}^{\infty}$ and $\{X'_n\}_{n=0}^{\infty}$ on $\mathbb{R}$ having the same transition kernel P. Our construction yields a split chain $\{(X_n, X'_n, Y_n)\}_{n=0}^{\infty}$, where $Y_n \in \{0,1\}$ and $\{X_n\}_{n=0}^{\infty}$ and $\{X'_n\}_{n=0}^{\infty}$ have transition kernel P. We will construct a coupling (stopping) time T such that $X_n = X'_n$ for all $n > T$ (It is slightly different from the T defined in section 2.2). By the last section, geometric convergence will be obtained at rates out to at least the radius of convergence of $\boldsymbol{E}_{x,X'_0}[r^T]$. We will show that if there exists two small sets, say C_1 and C_2, such that $X_n \in C_1$ implies that $X'_n \in C_2$, then only one chain needs to be tracked. This leads to a drift condition that can be used to bound the generating function of the stopping time.

3.1 Our coupling

Let $P : \mathbb{R} \times \mathcal{A} \to [0,1]$ be a chain transition kernel satisfying:

(i) P is $\phi-$irreducible, where ϕ is the Lebesgue measure;

(ii) The exists sets $C_1, C_2 \in \mathcal{A}$, a $\delta > 0$, and a probability measure ν on $(\mathbb{R}, \mathcal{A})$ for which

$$P(x, A) \geq \delta 1_{C_i}(x)\nu(A),$$

for all $x \in \mathbb{R}, A \in \mathcal{A}$ and $i = 1, 2$;

(iii) The chain satisfies the drift condition in (2.7).

Consider the transition kernels $Q_i : (\mathbb{R} \times \mathbb{R}, \mathcal{A}) \to [0,1], i = 1,2$ given by

$$Q_i((x_1,x_2),A) = \frac{P(x_i,A) - \delta 1_{C_1 \times C_2}(x_1,x_2)\nu(A)}{1 - \delta 1_{C_1 \times C_2}(x_1,x_2)}.$$

Lemma 2.22 in [22] shows that there exists measurable functions $g_i : \mathbb{R}^2 \times \mathbb{R}^d \to \mathbb{R}, i = 1,2,3$, and a d-variate random variable U such that $g_i((x_1,x_2),U)$ has distribution $Q_i((x_1,x_2),\cdot)$ for $i = 1,2$ and $g_3((x_1,x_2),U)$ has distribution ν, which does not depend on x_1 or x_2. Because of this, we write $g_3((x_1,x_2),U) = g_3(U)$. The value of the dimension d is not particularly relevant here; our future examples will elaborate on this aspect.

Let $(\Omega, \mathcal{F}, P)$ be a probability space containing the following independent random elements:

1. A bivariate random variable (X_0, X_0') representing the initial conditions of the two chains;
2. A sequence $V_0, V_1, \ldots$ of independent and uniform$(0,1)$ random variables;
3. A sequence of independent and identically distributed d-variate random variables $U_1, U_2, \ldots$.

Set $\mathcal{F}_0 = \sigma((X_0, X_0'))$ and $\mathcal{G}_0 = \sigma((X_0, X_0'), V_0)$. When $n \geq 1$, define the histories $\mathcal{F}_n = \sigma((X_0, X_0'), V_0, (U_i, V_i), i = 1, \ldots, n-1, U_n)$ and $\mathcal{G}_n = \sigma(\mathcal{F}_n, V_n)$. Given $\mathcal{F}_n$, define the coin tosses

$$Y_n = \begin{cases} 0, & V_n \leq 1 - \delta 1_{C_1 \times C_2}(X_n, X_n') \\ 1, & V_n > 1 - \delta 1_{C_1 \times C_2}(X_n, X_n') \end{cases}.$$

Suppose that we are given $\mathcal{G}_n$. On the set $\{Y_n = 0\}$, define $X_{n+1} = g_1(X_n, X_n', U_{n+1})$ and $X_{n+1}' = g_2(X_n, X_n', U_{n+1})$. On the set $\{Y_n = 1\}$, define $X_{n+1} = X_{n+1}' = g_3(U_{n+1})$. The first time where $Y_n = 1$ is special as $X_\ell = X_\ell'$ for all $\ell > n$ since both chains are driven by the same $\{U_n\}$ sequence thereafter.

Let $\zeta : \mathbb{R}^2 \times \{0,1\} \to \mathbb{R}$ be any measurable function. Define four functions as follows:

$$\begin{aligned}
G_{0,0}(x,x',u) &= 1_{C_1 \times C_2}(x,x')\zeta(g_1(x,x',u), g_2(x,x',u), 0), \\
G_{0,1}(x,x',u) &= 1_{C_1 \times C_2}(x,x')\zeta(g_1(x,x',u), g_2(x,x',u), 1), \\
G_{1,0}(x,x',u) &= 1_{C_1 \times C_2}(x,x')\zeta(g_3(u), g_3(u), 0), \\
G_{1,1}(x,x',u) &= 1_{C_1 \times C_2}(x,x')\zeta(g_3(u), g_3(u), 1).
\end{aligned}$$

Also define

$$H_{0,0}(x,x',u) = 1_{(C_1\times C_2)^c}(x,x')\zeta(g_1(x,x',u),g_2(x,x',u),0),$$
$$H_{1,0}(x,x',u) = 1_{(C_1\times C_2)^c}(x,x')\zeta(g_3(u),g_3(u),0).$$

These six functions are explained as follows. Suppose the current state is (x,x',i). Let (y,y',j) be the next state. If $i=0$, the reseult of the coin toss at time n is tails. Thus (y,y') are generated using the functions g_1 and g_2. If $(y,y')\in C_1\times C_2$, then j is 0 or 1 depending on whether or not of the next coin toss is tails or heads. This gives the functions $G_{0,0}$ and $G_{0,1}$. If, on the other hand $(y,y')\notin C_1\times C_2$ then the next coin toss results in tails. This gives the function $H_{0,0}$. If $i=1$, the logic above applies except that (y,y') is generated using g_3. This gives the functions $G_{1,0}$, $G_{1,1}$ and $H_{1,0}$.

Lemma 1 *The process $\{(X_n,X'_n,Y_n)\}_{n=0}^{\infty}$ is a Markov chain with respect to the filtration $\mathcal{G}_n$ whose kernel is*

$$\begin{aligned}\boldsymbol{E}_{x,x',i}[\zeta(X_1,X'_1,Y_1)] = &\int_{\mathbb{R}^d}\left[(1-\delta)G_{i,0}(x,x',u)+H_{i,0}(x,x',u)\right]\mu_U(du)\\ &+\int_{\mathbb{R}^d}\delta G_{i,1}(x,x',u)\mu_U(du).\end{aligned} \tag{3.1}$$

where $x,x'\in\mathbb{R}$, $i\in\{0,1\}$, and μ_U is the distribution of U_1.

Proof. Observe that $\mathcal{G}_n\subset\mathcal{F}_{n+1}$ and V_{n+1} is independent of $\mathcal{F}_{n+1}$. Since (X_{n+1},X'_{n+1}) is $\mathcal{F}_{n+1}$ measurable and $\mathbb{P}(V_{n+1}>1-\delta)=\delta$,

$$\mathbb{P}(Y_{n+1}=1|\mathcal{F}_{n+1})=\delta 1_{C_1\times C_2}(X_{n+1},X'_{n+1}),\quad \mathbb{P}(Y_{n+1}=0|\mathcal{F}_{n+1})=1-\delta 1_{C_1\times C_2}(X_{n+1},X'_{n+1}).$$

Using the tower property of conditional expectations, namely $\boldsymbol{E}[\cdot|\mathcal{G}_n]=\boldsymbol{E}[\boldsymbol{E}[\cdot|\mathcal{F}_{n+1}|\mathcal{G}_n]]$, we obtain

$$\boldsymbol{E}_{x,x',i}[\zeta(X_{n+1}, X'_{n+1}, Y_{n+1})|\mathcal{G}_n] = \boldsymbol{E}_{x,x',i}[1_{C_1 \times C_2}(X_{n+1}, X'_{n+1}) \times$$
$$(\delta\zeta(X_{n+1}, X'_{n+1}, 1) + (1-\delta)\zeta(X_{n+1}, X'_{n+1}, 0))$$
$$+1_{(C_1 \times C_2)^c}(X_{n+1}, X'_{n+1})\zeta(X_{n+1}, X'_{n+1}, 0)|\mathcal{G}_n].$$

By definition, when $Y_n = 0, X_{n+1} = g_1(X_n, X'_n, U_{n+1}), X'_{n+1} = g_2(X_n, X'_n, U_{n+1})$, and $Y_n = 1, X_{n+1} = X'_{n+1} = g_3(U_{n+1})$. The result now follows by noting that X_n, X'_n, and Y_n are $\mathcal{G}_n$ measurable, that U_{n+1} is independent of $\mathcal{G}_n$, and the definitions of $G_{i,j}$ and $H_{i,j}$. □

Corollary 1 *The processes $\{X_n\}_{n=0}^{\infty}$ and $\{X'_n\}_{n=0}^{\infty}$ are Markov chains with respect to the filtration $\{\mathcal{F}_n\}$ having the same transition kernel P.*

Proof. We only prove the result for $\{X_n\}_{n=0}^{\infty}$, the proof for $\{X'_n\}_{n=0}^{\infty}$ being identical. Set $\kappa(x) = \zeta(x, x', i)$ where ζ is as above. In this case, $G_{i,0} = G_{i,1} = H_{i,0}$ for $i = 0, 1$. From Lemma 1,

$$\boldsymbol{E}_{x,x'}[\kappa(X_{n+1})|\mathcal{G}_n] = 1_{[Y_n=0]}\boldsymbol{E}_{x,x'}[\kappa(g_1(X_n, X'_n, U_{n+1}))] + 1_{[Y_n=1]}\boldsymbol{E}_{x,x'}[\kappa(g_3(U_{n+1}))].$$

Since $g_1(X_n, X'_n, U_{n+1})$ is distributed as the kernel $Q_1((X_n, X'_n), \cdot)$ and $g_3(U_{n+1})$ is distributed as the kernel $\nu(\cdot)$, the above equation becomes

$$1_{[Y_n=0]} \int_{\mathbb{R}} Q_1((x, x'), dy)\kappa(y) + 1_{[Y_n=1]} \int_{\mathbb{R}} \nu(dy)\kappa(y).$$

Noting that (X_n, X'_n) is $\mathcal{F}_n$ measurable,

$$\mathbb{P}(Y_n = 0|\mathcal{F}_n) = 1 - \delta 1_{C_1 \times C_2}(X_n, X'_n) \text{ and } \mathbb{P}(Y_n = 1|\mathcal{F}_n) = \delta 1_{C_1 \times C_2}(X_n, X'_n),$$

and the tower property of conditional expectations gives

$$\begin{aligned}
&\boldsymbol{E}_{x,x'}[\kappa(X_{n+1})|\mathcal{F}_n] = \boldsymbol{E}_{x,x'}[\boldsymbol{E}_{x,x'}[\kappa(X_{n+1})|\mathcal{G}_n]|\mathcal{F}_n] \\
&\quad =(1-\delta 1_{C_1\times C_2}(X_n, X_n')) \int_{\mathbb{R}} Q_1((x,x'),dy)\kappa(y) \\
&\quad +\delta 1_{C_1\times C_2}(X_n, X_n') \int_{\mathbb{R}} \nu(dy)\kappa(y) \\
&\quad = \int_{\mathbb{R}} P(X_n, dy)\kappa(y),
\end{aligned}$$

which completes the proof. □

3.2 Stochastically Ordered Chains

Suppose that our chains satisfy the shadowing condition that when $X_n \in C_1$, $X_n' \in C_2$. We show that to find the coupling time, we only need the process $\{(X_n, Y_n)\}$ and that this process is precisely the split chain developed in [32]. In this case, our coupling time will be

$$\begin{aligned}
T =& \inf\{n \geq 0 : (X_n, X_n', Y_n) \in C_1 \times C_2 \times \{1\}\} \\
=& \inf\{n \geq 0 : (X_n, Y_n) \in C_1 \times \{1\}\} \\
=& \inf\{n \geq 0 : Y_n = 1\} := \eta_{C_1}.
\end{aligned}$$

Note that we do not have to observe $\{X_n'\}$ to determine when the process couples. The process $\{(X_n, X_n', Y_n)\}_{n=0}^{\infty}$ is a Markov chain by Lemma 1. To determine its transition function, note that the dynamics of the chain do not depend on the value of X_1'. Let $\xi : \mathbb{R}\times\{0,1\} \to \mathbb{R}$ be a measurable function. Hence, $g_1(x,x',u) \in C_1$ implies that $g_2(x,x',u) \in C_2$, yielding

$$G_{0,j}(x,x',u) = 1_{C_1}(g_1(x,x',u))\xi(g_1(x,x',u),j),$$

for $j = 0, 1$, and

$$H_{0,0}(x,x',u) = 1_{C_1^c}(g_1(x,x',u))\xi(g_1(x,x',u),0).$$

Using the functions $G_{0,j}$ when $X_1 \in C_1$, $j = 0, 1$, and $H_{0,0}$ otherwise, we obtain

$$\boldsymbol{E}_{x,x',0}[\xi(X_1, Y_1)] = \delta \int_{C_1} Q_1((x, x'), dz)\xi(z, 1)$$
$$+(1-\delta) \int_{C_1} Q_1((x, x'), dz)\xi(z, 0) + \int_{C_1^c} Q_1((x, x'), dz)\xi(z, 0),$$

where

$$Q_1((x, x'), A) = \frac{P(x, A) - \delta 1_{C_1}(x)\nu(A)}{1 - \delta 1_{C_1}(x)}.$$

Since Q_1 does not depend on x', the Markov chain $\{(X_n, Y_n)\}_{n=0}^{\infty}$ is precisely the split chain in Chapter 4 of [32] with the atom $\tilde{\rho} = (\delta 1_{C_1}, \nu)$. Equation (4.21) of [32] gives

$$\mathbb{P}_x(\eta_{C_1} = n) = (P - \delta 1_{C_1} \otimes \nu)^n \delta 1_{C_1}(x_1),$$

for $n \geq 1$, where $\otimes$ defined in (2.5) and

$$(P - \delta 1_{C_1} \otimes \nu))^n \delta 1_{C_1}(x) := \delta \int_{\mathbb{R}} 1_{C_1}(z)(P - \delta 1_{C_1} \otimes \nu)^n(x, dz) \text{ for each } n \geq 1.$$

Since $X_n \in C_1$ implies $X'_n \in C_2$ and Y_n depends on X_n, it now follows that

$$\boldsymbol{E}_{x,x'}[r^T] = \boldsymbol{E}_{x,x'}[r^{\eta_{c_1}}] = \boldsymbol{E}_x[r^{\eta_{c_1}}] = \sum_{r=0}^{\infty} r^n (P - \delta 1_{C_1} \otimes \nu)^n \delta 1_{C_1}(x).$$

Note that the first equality above is following by the fact η_{C_1} is a coupling time when the chain is shadowing, and the second equality above is followed by the fact "η_{C_1} is only depended on x, not x'" since when $X_n \in C_1$, $X'_n \in C_2$.

Before continuing, we look at an example. Consider the case when

$$X_{n+1} = G(X_n, U_{n+1}),$$

where $U_1, U_2, \ldots$ are independent and identically distributed random elements taking values in a set $\mathbb{U}$ and $G : \mathbb{R} \times \mathbb{U} \to \mathbb{R}$, such that for each $u \in \mathbb{U}$, $G(\cdot, u)$ is increasing in x. In this case, if $x_0 \geq x'_0$, then for all n, $X_n \geq X'_n$ and the Markov chain $\{X_n\}_{n=0}^{\infty}$ is stochastically larger than the $\{X'_n\}_{n=0}^{\infty}$. For convenience take the state space is $[0, \infty)$. We can choose a small set $C_1 = [0, c]$, for some $c > 0$.

In this scenario, we see that whenever $X_n \in C_1$, then $X'_n \in C_1$.

Consider the kernel $r(P - \delta 1_{C_1} \otimes \nu)$. Using (4.21) in [32], we obtain

$$\boldsymbol{E}_x[r^{\eta_{C_1}}] = \sum_{n=0}^{\infty}(r(P - \delta 1_{C_1} \otimes \nu))^n \delta 1_{C_1}(x) \equiv G_x(r).$$

Lemma 2 *Let V be a nonnegative measurable function such that*

$$V \geq \delta 1_{C_1} + r(P - \delta \otimes \nu)V, \tag{3.2}$$

then, for all x

$$V(x) \geq G_x(r).$$

Proof. The proof is a direct application of Theorem 3.1 (iii) of [32]. □

In our examples to follow, we use $V(x) = a|x|1_{C_1^c}(x) + \alpha$, where $\alpha \geq 1$ and $a > 0$. Lemma 3 below provides conditions on α and a that satisfy 3.2.

Lemma 3 *Suppose $\boldsymbol{E}_x[V(X_1)] \leq r_0^{-1}V(x) + b1_{C_1}(x)$ with $r_0 > 1$, with $\nu(V) = \alpha$. If $1 < r < \min(\delta/(\delta\alpha - b), r_0)$, then*

$$\delta 1_{C_1}(x) + r\left(P - (\delta 1_{C_1} \otimes \nu)\right)V(x) \leq V(x).$$

Proof. For $x > c$, it follows from drift condition in (2.7) if $1 < r \leq r_0$. For $x \leq c$,

$$\begin{aligned} &\delta 1_{C_1}(x) + r\left(P - (\delta 1_{C_1} \otimes \nu)\right)V(x) \\ \leq &\delta + r(r_0^{-1}V(x) + b) - r\delta\alpha \\ \leq &V(x) + r(b - \delta\alpha) + \delta \\ \leq &V(x). \end{aligned}$$

The last inequality is given by $r < \delta/(\delta\alpha - b)$. □

Definition 10 *A chain on $\mathbb{R}$ is said to satisfy a shadowing condition if two trajectories, $\{X_n\}_{n=0}^{\infty}$ and $\{X'_n\}_{n=0}^{\infty}$, have the property that $X_n \in C_1$ implies that $X'_n \in C_2$.*

Theorem 11 *Let μ be the initial distribution of $\{X_n\}_{n=0}^{\infty}$ and let μ' be that of $\{X'_n\}_{n=0}^{\infty}$. Suppose that V is both μ and μ' integrable, that is,*

$$\mu(V) \equiv \int_{\mathbb{R}} V(x)\mu(dx) < \infty \qquad \text{and} \qquad \mu'(V) \equiv \int_{\mathbb{R}} V(x)\mu'(dx) < \infty.$$

Suppose that shadowing condition holds and r satisfies (3.2), then

$$||X_n - X'_n|| \leq (\mu(V) + \mu'(V))r^{-n}. \tag{3.3}$$

Proof. Let $B \subset \Omega$ be the set such that when $X_n \in C_1$, then $X'_n \in C_2$ and let $B' \subset \Omega$ be the set such that when $X'_n \in C_1$, then $X_n \in C_2$. By assumption, $B \cup B' = \Omega$, however, the two sets are not necessarily disjoint. Since, $X_n = X'_n$ for all $n > T$, for $A \in \mathcal{B}(\mathbb{R})$,

$$\begin{aligned}
&|\mathbb{P}(X_n \in A) - \mathbb{P}(X'_n \in A)| \\
&= |\mathbb{P}(X_n \in A, T \geq n) - \mathbb{P}(T \geq n, X'_n \in A)| \\
&\leq \mathbb{P}(T \geq n).
\end{aligned}$$

By Markov's inequality, Lemma 2, and the definition of the set B,

$$\mathbb{P}(\{T \geq n\} \cap B) \leq \mathbb{P}_{\mu}(\eta_{C_1} \geq n) \leq \mu(V)r^{-n}.$$

Similarly, by the definition of B'

$$\mathbb{P}(\{T \geq n\} \cap B') \leq \mathbb{P}_{\mu'}(\eta_{C_1} \geq n) \leq \mu'(V)r^{-n}.$$

Inserting these two inequalities into (3.4) gives (3.3) and completes the proof. □

Let μ be a Dirac measure at a single point x with probability one, and let μ' be a stationary measure π. Then, $\mu(V) = V(x)$ and $\mu'(V) = \pi(V)$. By Theorem 14.3.7 from Meyn and Tweedie (1993) with $f(x) = (1 - r_0^{-1})V(x)$ and $s(x) = b1_{C_1}(x)$,

$$\pi(V) = \int_{-\infty}^{\infty} V(z)\pi(dz) \leq \frac{b}{1 - r_0^{-1}} < \infty. \tag{3.4}$$

Thus,

$$\mu(V) + \mu'(V) = V(x) + \pi(V) \leq V(x) + \frac{b}{1 - r_0^{-1}}. \tag{3.5}$$

Theorem 12 *Suppose that $\{X_n\}_{n=0}^{\infty}$ is a ϕ-irreducible, aperiodic Markov chain with the stationary distribution $\pi(\cdot)$ satisfying the minorization condition (2.5) for some set C_1 and $\delta \in (0,1)$. Suppose that shadowing condition holds. Then for any $1 < r < \min(\delta/(\delta\alpha - b), r_0)$,*

$$||P^n(x, \cdot) - \pi(\cdot)|| \leq M(x) r^{-n},$$

where $M(x) = V(x) + b/(1 - r_0^{-1})$.

Proof. Applying Theorem 11, with μ is a Dirac measure at a point x and $\mu' = \pi$, Combining (3.3), and (3.5), the result is followed. □

Definition 13 *A measurable map $\Theta : \Omega \to \Omega$ is said to be a shift operator for the Markov chain $\{X_n\}$ if*

$$X_n(\Theta\omega) = X_{n+1}(\omega) \quad \text{for all } \omega \in \Omega, n \geq 0.$$

The iterates $\Theta_n, n \geq 0$, of Θ are defined by $\Theta_0 = I_\Omega$, the identity operator on Ω and iteratively,

$$\Theta_n = \Theta \circ \Theta_{n-1} \text{ for } n \geq 1.$$

Define $\Theta_T(\omega) : \Omega \to \Omega$ by

$$\Theta_T(\omega) := \Theta_{T(\omega)}(\omega) \text{ for all } \omega \in \Omega.$$

We now relax the shadowing condition slightly. Let τ be a "stopping" time depending on X_0 and X_0' so that for all $n > \tau$ when one of the processes $\{X_n\}_{n=0}^{\infty}$ or $\{X_n'\}_{n=0}^{\infty}$ is in C_1 the other must be in C_2. Now the coupling works by running the two processes independently until τ occurs and then couple as done in beginning of this section. The coupling now occurs at the stopping time $\tau + T \circ \Theta_\tau$ by the definition of Θ.

Definition 14 *A chain on $\mathbb{R}$ are said to satisfy a strong shadowing condition if for any two trajectories $\{X_n\}_{n=0}^{\infty}$ and $\{X_n'\}_{n=0}$, there exist a τ (depends on X_0 and X_0') such that for all $n > \tau$ when $X_n \in C_1$, $X_n' \in C_2$.*

Note that if X_0 and X_0' are given, τ is fixed.

Definition 15 *Define $L^1(\mu \times \mu^{'})$ to be the collections of all measurable functions $f : \mathbb{R}^2 \to \mathbb{R}$ satisfying*

$$||f||_{L^1(\mu\times\mu')} = \int_{\mathbb{R}} \int_{\mathbb{R}} f(x,z)\mu(dx)\mu'(dz) < \infty.$$

Theorem 16 *Suppose the initial distribution of $\{X_n\}_{n=0}^{\infty}$ is μ and that of $\{X_n'\}_{n=0}^{\infty}$ is μ'. Suppose that strong shadowing condition holds, and both $r^\tau V(X_\tau)$ and $r^\tau V(X_\tau')$ are in $L^1(\mu\times\mu')$. Then*

$$||X_n - X_n'|| \le r^{-n} \boldsymbol{E}_{\mu,\mu'}[r^\tau \left(V(X_\tau) + V(X_\tau')\right)].$$

Proof. As in the proof of Theorem 11, we have

$$||X_n - X_n'|| \le \mathbb{P}(\tau + T\circ\Theta_\tau \ge n).$$

By Markov's inequality and the strong Markov property applied at τ,

$$\begin{aligned}\mathbb{P}(\tau{+}T\circ\Theta_\tau \ge n) &\le r^{-n}\boldsymbol{E}_{\mu,\mu'}[r^{\tau+T\circ\Theta_\tau}] \\ &= r^{-n}\boldsymbol{E}_{\mu,\mu'}[r^\tau \boldsymbol{E}_{X_\tau,X_\tau'}[r^T]].\end{aligned}$$

With B and B' as in the proof of Theorem 2, the above is no greater than

$$\begin{aligned}& r^{-n}\boldsymbol{E}_{\mu,\mu'}[r^\tau\left(\boldsymbol{E}_{X_\tau,X_\tau'}[r^T 1_B] + \boldsymbol{E}_{X_\tau,X_\tau'}[r^T 1_{B'}]\right)] \\ &\qquad \le r^{-n}\boldsymbol{E}_{\mu,\mu'}[r^\tau\left(V(X_\tau)+V(X_\tau')\right)],\end{aligned}$$

where the last inequality follows as in Theorem 11. □

Lemma 4 *Suppose that $\{X_n\}_{n=0}^{\infty}$ is a Markov chain satisfies the drift condition (2.7) with $V(x)\ge 1$. Then $P^mV(x) \le V(x) + b(r_0^m - 1)/(1-r_0^{-1})$ for all $x\in\mathbb{R}$, and for all integers $m \ge 1$.*

Proof.

$$P^mV(x) \le r_0 P^{m-1}V(x) + b \le r_0^{-2}P^{m-2}V(x) + r_0^{-1}b + b = \cdots \le r_0^{-m}V(x) + b\frac{1-r_0^{-m}}{1-r_0^{-1}}. \quad (3.6)$$

□

Suppose the chain is strongly shadowing. Let μ be a Dirac measure at a single point x with probability one, and let μ' be a stationary measure π.

And then, we have

$$
\begin{aligned}
&\boldsymbol{E}_{x,\pi}[r^{\tau}\left(V(X_\tau)+V(X'_\tau)\right)] = \int_{-\infty}^{\infty} \boldsymbol{E}_{x,z}[r^{\tau}\left(V(X_\tau)+V(X'_\tau)\right)]\pi(dz) \\
&= \int_{-\infty}^{\infty} \boldsymbol{E}_{x,z}[r^{\tau}V(X_\tau)]\pi(dz) + \int_{-\infty}^{\infty} \boldsymbol{E}_{x,z}[r^{\tau}V(X'_\tau)]\pi(dz) \\
&= \int_{-\infty}^{\infty} r^{\tau}P^{\tau}V(x)\pi(dz) + \int_{-\infty}^{\infty} r^{\tau}P^{\tau}V(z)\pi(dz) \\
&\leq \int_{-\infty}^{\infty} \left(V(x) + b\frac{r_0^{\tau}-1}{1-r_0^{-1}}\right)\pi(dz) + \int_{-\infty}^{\infty} \left(V(z) + b\frac{r_0^{\tau}-1}{1-r_0^{-1}}\right)\pi(dz) \\
&\leq V(x) + \int_{-\infty}^{\infty} V(z)\pi(dz) + 2\int_{-\infty}^{\infty} \left(b\frac{r_0^{\tau}-1}{1-r_0^{-1}}\right)\pi(dz).
\end{aligned}
\tag{3.7}
$$

The first inequality above is given by Lemma 4 with $m = \tau$. The first integral of last inequality is applying the result in (3.4). Thus,

$$\boldsymbol{E}_{x,\pi}[r^{\tau}\left(V(X_\tau)+V(X'_\tau)\right)] \leq V(x) + 2\int_{-\infty}^{\infty} \frac{b}{1-r_0^{-1}} r_0^{\tau}\mu(dz).$$

If the moment generating function of π exists for $s < 1$, then

$$\int_{-\infty}^{\infty} r_0^{\tau}\pi(dz) < \infty. \tag{3.8}$$

So, we have

$$\boldsymbol{E}_{x,\pi}[r^{\tau}\left(V(X_\tau)+V(X'_\tau)\right)] \leq V(x) + 2\int_{-\infty}^{\infty} \frac{b}{1-r_0^{-1}} r_0^{\tau}\pi(dz) < \infty. \tag{3.9}$$

Corollary 2 *Suppose that $\{X_n\}_{n=0}^{\infty}$ is a ϕ-irreducible, aperiodic Markov chain with the stationary distribution $\pi(\cdot)$ satisfying the minorization condition (2.5) for some set C_1 and $\delta \in (0,1)$. Suppose the strong shadowing condition holds. Then for any $1 < r$, satisfies $r \leq r_0$ and (3.2),*

$$||P^n(x,\cdot) - \pi(\cdot)|| \leq M(x)r^{-n}, \tag{3.10}$$

where $M(x) = V(x) + 2b/\left(1-r_0^{-1}\right)\int_{-\infty}^{\infty} r_0^{\tau}\pi(dz)$.

Proof. Combining (2.2), (3.7), (3.8) and Lemma 4, gives (3.10). □

Chapter 4

A Storage Chain Example

Our first example considers a discrete time storage process constructed as follows. The inputs to the store arrive via a Poisson process $\{T_n\}_{n=1}^{\infty}$ with arrival rate $\lambda > 0$. At the time of the nth input, an amount I_n is added to the store's content. The I_ns are IID positive random variables, independent of the Poisson process. The release rate of the store is e^{-x} when the store's content is x. Let x_0 be the initial storage level (if the initial state is random, it is assumed independent of all else). Let X_n be the storage level just prior to the nth input. Then

$$X_{n+1} = (X_n + I_n)e^{-\tau_{n+1}},$$

where $\tau_{n+1} = T_{n+1} - T_n$ (take $T_0 = 0$). Then $U_n := e^{-\tau_n}$ are IID beta variates with parameters λ and 1. When $\lambda = 1$, the U_ns are uniform $(0, 1)$ variables. The above recursion becomes $X_{n+1} = (X_n + I_n)U_{n+1}$, which constitutes a Markov chain on $[0, \infty)$.

To proceed further, we suppose that $\lambda = 1$ and that the I_ns are uniform $(0, \beta)$ for some parameter $\beta > 0$. The cumulative transition distribution is

$$\mathbb{P}_x(X_1 \le y) = \mathbb{P}\Big((x + I_0)U_1 \le y\Big) = \frac{(y-x)^+}{\beta} + \frac{y}{\beta} \ln\Big(\frac{x+\beta}{\max\{x, y\}}\Big). \tag{4.1}$$

Differentiating (4.1) yields the transition density

$$P(x, dy) = \frac{1}{\beta} \ln\Big(\frac{x+\beta}{\max\{x, y\}}\Big) dy. \tag{4.2}$$

We leave it to the reader to show that the chain is $\phi-$irreducible, where ϕ is the Lebesgue measure. Below, it is shown that this chain satisfies a drift condition. It now follows that $\{X_n\}$ is Harris positive recurrent [32].

The moment generating function of the limit distribution of the chain exists and can be shown to have form

$$M_{X_\infty}(t) = \boldsymbol{E}[e^{tX_\infty}] = \exp\left\{\int_0^t \frac{e^{\zeta\beta} - 1 - \zeta\beta}{\zeta^2\beta} d\zeta\right\}.$$

When $x > y$, it is easy to check that $\mathbb{P}_x(X_1 > z) \geq \mathbb{P}_y(X_1 > z)$. Thus, $\{X_n\}$ is stochastically ordered in its initial state. However, this chain may not be reversible: let $y = x + 2$. Then if $dy \in [x+2-0.01, x+2+0.01]$, $P(x, dy) = 0$ and $P(y, dx) > 0$. Thus, $\pi(dx)P(x, dy) \neq \pi(dy)P(y, dx)$. While this chain is stochastically ordered on $[0, \infty)$, it will never enter the minimal state space element of $\{0\}$; hence, the rate results in [28] do not apply.

4.1 A Minorization Condition

To identify a minorization and drift function for the chain, our first step is to find a probability measure $\nu(\cdot)$ and $\delta > 0$ such that $P(x, A) \geq \delta\nu(A)$ for all $x \in C_1 := [0, c]$ and Borel measurable sets A for some fixed c. Our δ and ν will depend on c. Later, we employ a numerical method to select the best c for each fixed β.

Lemma 5 *Define $y_0 = \beta c/(\beta + c)$, $\delta = \beta/(\beta + c)$, and*

$$\nu(dy) = \frac{1}{\delta}\left[1_{(0,y_0]}(y)\frac{1}{\beta}\ln\left(\frac{c+\beta}{c}\right)dy + 1_{(y_0,\beta]}(y)\frac{1}{\beta}\ln\left(\frac{\beta}{y}\right)dy\right].$$

Then, for $0 \leq x \leq c$, δ and ν satisfy the minorization condition $P(x, dy) \geq \delta\nu(dy)$.

Proof. To prove the lemma, it is sufficient to establish that

$$P(x, dy) \geq 1_{(0,y_0]}(y)\frac{1}{\beta}\ln\left(\frac{c+\beta}{c}\right)dy + 1_{(y_0,\beta]}(y)\frac{1}{\beta}\ln\left(\frac{\beta}{y}\right)dy \tag{4.3}$$

for all $y > 0$ and $x \in C$. We do this by cases. First, suppose that $y_0 < x$. We will consider four subcases. First, when $y < y_0$, y is also smaller than x and (4.2) gives

$$P(x, dy) = \frac{1}{\beta}\ln\left(\frac{x+\beta}{x}\right)dy \geq \frac{1}{\beta}\ln\left(\frac{c+\beta}{c}\right)dy.$$

Since the indicator $1_{(y_0,\beta]}(y) = 0$, (4.3) holds.

For the second subcase, when $y_0 < y \le x$, the indicator $1_{(0,y_0]}(y)$ is still zero and the choice of y_0 gives

$$\begin{aligned} P(x, dy) &\ge \frac{1}{\beta} \ln\left(\frac{c+\beta}{c}\right) dy \\ &= \frac{1}{\beta} \ln\left(\frac{\beta}{y_0}\right) dy \\ &\ge \frac{1}{\beta} \ln\left(\frac{\beta}{y}\right) dy, \end{aligned}$$

implying (4.3) again.

Our third subcase moves to $y \in (x, \beta]$. Then (4.2) gives

$$P(x, dy) = \frac{1}{\beta} \ln\left(\frac{x+\beta}{y}\right) dy \ge \frac{1}{\beta} \ln\left(\frac{\beta}{y}\right) dy,$$

implying (4.3) again.

Finally, when $y > \beta$, (4.3) trivially holds since the right hand side of (4.3) is zero.

Our second case considers $y_0 \ge x$. Again, we partition work into four subcases. When $y \le x$, (4.2) gives

$$P(x, dy) = \frac{1}{\beta} \ln\left(\frac{x+\beta}{x}\right) dy \ge \frac{1}{\beta} \ln\left(\frac{c+\beta}{c}\right) dy,$$

implying (4.3). When, $y \in (x, y_0]$, we have

$$P(x, dy) = \frac{1}{\beta} \ln\left(\frac{x+\beta}{y}\right) dy \ge \frac{1}{\beta} \ln\left(\frac{x+\beta}{y_0}\right) dy \ge \frac{1}{\beta} \ln\left(\frac{\beta}{y_0}\right) dy = \frac{1}{\beta} \ln\left(\frac{c+\beta}{c}\right) dy.$$

Since the indicator $1_{(y_0,\beta]}(y) = 0$, (4.3) holds. When $y \in (y_0, \beta]$, our bound is

$$P(x, dy) = \frac{1}{\beta} \ln\left(\frac{x+\beta}{y}\right) dy \ge \frac{1}{\beta} \ln\left(\frac{\beta}{y}\right) dy,$$

implying (4.3) again. Finally, when $y > \beta$, (4.3) holds since its right hand side is zero. □

Corollary 3 *Under the assumptions in Lemma (5), $y_0 \to \beta$ and $\delta \to 0$ as $c \to \infty$.*

Proof. The limit for y_0 follows for its definition in the previous Lemma. For the result on δ, to make ν defined in the last lemma a probability measure,

$$\delta = \int_0^{y_0} \frac{1}{\beta} \ln\Big(\frac{c+\beta}{c}\Big) dy + \int_{y_0}^{\beta} \frac{1}{\beta} \ln\Big(\frac{\beta}{y}\Big) dy = \frac{\beta}{\beta + c}, \tag{4.4}$$

from which the limit claim about δ follows. □

4.2 A Drift Condition

We next establish a drift condition for the storage model. Define $V_c(x) = \alpha + 3x1_{(c,\infty)}(x)$, where $c > 0$ and $\alpha > 1$.

Lemma 6 *If $x > c$, then*

$$\boldsymbol{E}_x[V_c(X_1)] = \alpha + 3\left[\frac{-c^2/2}{\beta} \ln\left(1 + \frac{\beta}{x}\right) + \frac{1}{2}\left(x + \frac{\beta}{2}\right)\right].$$

If $0 \leq x \leq c$, then

$$\boldsymbol{E}_x[V_c(X_1)] = \alpha + 3\left[\frac{-c^2}{2\beta} \ln\left(\frac{x+\beta}{c}\right) - \frac{c^2 - x^2}{4\beta} + \frac{1}{2}\left(x + \frac{\beta}{2}\right)\right] 1_{(x+\beta>c)}(x).$$

Proof. When $x > c$, (4.2) gives

$$\begin{aligned} \boldsymbol{E}_x[V_c(X_1)] &= \int_0^\infty \alpha P(x,dy) + \int_0^\infty 3yP(x,dy) - \int_0^c 3yP(x,dy) \\ &= \alpha + 3\boldsymbol{E}_x[X_1] - 3\int_0^c y\frac{1}{\beta} \ln\left(\frac{x+\beta}{x}\right) dy \\ &= \alpha + 3\left[\frac{1}{2}\left(x + \frac{\beta}{2}\right) + \frac{-c^2/2}{\beta} \ln\left(\frac{x+\beta}{x}\right)\right], \end{aligned}$$

establishing the first claim.

For the case where $x < c$, return to (4.2) to get

$$\begin{aligned}
\boldsymbol{E}_x[V_c(X_1)] &= \int_0^\infty \alpha P(x, dy) + \int_c^\infty 3y P(x, dy) \\
&= \alpha + 1_{x+\beta > c}(x) \int_c^{x+\beta} 3y \frac{1}{\beta} \ln\left(\frac{x+\beta}{y}\right) dy \\
&= \alpha + \frac{3}{\beta} 1_{x+\beta > c}(x) \left[\int_c^{x+\beta} y \ln(x+\beta)\, dy - \int_c^{x+\beta} y \ln(y)\, dy \right].
\end{aligned}$$

Integration by parts on the rightmost integral and algebraic simplifications now establish the second claim. □

We will need an analysis of the function h_c defined by

$$h_c(x) = \frac{-3c^2/2}{\beta} \frac{\ln(1+\beta/x)}{\alpha + 3x} + \frac{1}{2} + \frac{\alpha/2 + 3\beta/4}{\alpha + 3x}, \quad x > c.$$

In fact, convergence rates will be linked to extremums of h_c.

Lemma 7 *Define $K(c) = \sup_{x>c} h_c(x)$. Then*

$$K(c) < \frac{1}{2} + \frac{\alpha/2 + \beta/4}{\alpha + 3c}.$$

Proof. $\ln(1 + x/\beta)/x$ is decreasing on $x > c$ and approaches zero as $x \to \infty$. Hence,

$$\sup_{x>c} \left\{ \frac{-3c^2/2}{\beta} \frac{\ln(1+\beta/x)}{\alpha + 3x} \right\} = 0.$$

The above give

$$\begin{aligned}
K(c) &\le \sup_{x>c} \left\{ \frac{-3c^2/2}{\beta} \frac{\ln(1+\beta/x)}{\alpha + 3x} \right\} + \sup_{x>c} \left\{ \frac{1}{2} + \frac{\alpha/2 + \beta/4}{\alpha + 3x} \right\} \\
&= 0 + \sup_{x>c} \left\{ \frac{1}{2} + \frac{\alpha/2 + \beta/4}{\alpha + 3x} \right\} \\
&= \frac{1}{2} + \frac{\alpha/2 + \beta/4}{\alpha + 3c}.
\end{aligned}$$

□

Corollary 4 *Under the assumptions of Lemma 7. Then* $K(c) < 1$ *for all* $c > \beta/6$.

Proof. When $\beta < 6c$, and Lemma 7 gives

$$K(c) < \frac{1}{2} + \frac{\alpha/2 + \beta/4}{\alpha + 3c} < \frac{1}{2} + \frac{\alpha/2 + 3c/2}{\alpha + 3c} = 1.$$

When $\beta \geq 4$ and $c > \beta/2$, $c > 2$ and similar reasoning provides the result. □

Lemma 8 *Choose* c *to satisfy* $K(c) < 1$ *and* $b = \sup_{x \leq c} h(x) - r_0^{-1}\alpha$, *where*

$$h(x) = \alpha + 3\left(\frac{-c^2}{2\beta}\ln\left(\frac{x+\beta}{c}\right) - \frac{c^2 - x^2}{4\beta} + \frac{1}{2}\left(x + \frac{\beta}{2}\right)\right)1_{(x+\beta>c)}(x)$$

and $r_0 = 1/K(c)$. *Then* V_c *satisfies the drift condition (2.7) with contraction parameter* r_0^{-1} *and constant* b.

Proof. When $x > c$, Lemma 6 and the definition of $h_c(x)$ give

$$\boldsymbol{E}_x[V_c(X_1)] = \alpha + 3\left(\frac{-c^2/2}{\beta}\ln\left(1 + \frac{\beta}{x}\right) + \frac{1}{2}\left(x + \frac{\beta}{2}\right)\right) = (\alpha + 3x)h_c(x).$$

To identify a drift condition, use Lemma 7 to get

$$\boldsymbol{E}_x[V_c(X_1)] = (\alpha + 3x)h_c(x) \leq (\alpha + 3x)K(c) = r_0^{-1}V_c(x).$$

For $x \leq c$, $V_c(x) = \alpha$ and Lemma 6 gives

$$PV_c(x) = h(x) = h(x) - \alpha r_0^{-1} + \alpha r_0^{-1}.$$

Taking $b = sup_{x \leq c}h(x) - \alpha r_0^{-1}$ gives $PV_c(x) \leq r_0^{-1}V_c(x) + b1_C(x)$ and finishes our work. □

Lemma 9 *Under the assumptions of Lemma 8, if* $c > \beta$, $1 < r \leq \min(\delta/(\delta\alpha - b), r_0)$, *then*

$$\delta 1_{C_1}(x) + r\left(P - (\delta 1_{C_1} \otimes \nu)\right)V_c(x) \leq V_c(x). \tag{4.5}$$

Proof. For $|x| > c$, it immediately follows the result from Lemma 8 if $r < r_0$. For $x \in C_1$,

$V_c(x) = \alpha$ and $\nu(C_1) = \nu([0, \beta]) = 1$ since $c > \beta$. So, $\nu(V_c) = \alpha$. And, this gives

$$\begin{aligned} \delta + r(P - \delta\nu)V_(x) \leq& \delta + r\left(r_0^{-1}V_c(x) + b\right) - r\delta\alpha \\ \leq& V_c(x) + \delta + r(b - \delta\alpha). \end{aligned}$$

If $c > \beta$, $1 < r \leq \min(\delta/(\delta\alpha - b), r_0)$, then (4.5) follows. □

By the definition of r_0, if α fixed, $\lim_{c\to\infty} r_0 = 2$. Unfortunately, if c fixed, $\lim_{\alpha\to\infty} r_0 = 1$. We will have to balance these two quantities to get good convergence rates.

We now obtain convergence rates for the storage model by applying $X_0 = 10$, $C_1 = [0, c]$ with $V_c(x) = \alpha + 3x1_{(c,\infty)}(x)$ and $M(x) = V_c(x) + b/(1 - r_0^{-1})$. For a fixed β, we want to select the set $C_1 = [0, c]$ that gives a good convergence rate. For a numerical illustration, we consider $c = \beta\sqrt{3}$ and $\alpha \in (1, 200]$ for β values in $\{1, 3, 6, 16\}$. The following table displays some convergence results, with the located values of c. In this presentation, we compare to the convergence rates in [35], which are labeled as "RT rates".

	$\beta = 1$	$\beta = 3$	$\beta = 6$	$\beta = 16$
c	1.732051	5.196152	10.3923	27.71281
δ	0.3660254	0.3660254	0.3660254	0.3660254
α	11.7	33	65	171.7
r_0	1.298798	1.315168	1.319311	1.321902
b	3.983865	11.78474	23.5095	62.55994
CKL rates	1.298798	1.315168	1.319311	1.321902
RT rates	1.169257	1.127226	1.107116	1.086219
$M(x)$	39.01684	92.17664	172.1353	438.6049
$M(x)r^{-1000}$	1.121185e-112	9.622537e-118	7.740599658e-119	2.772939962e-119

Table 4.1: A Convergence Rate Comparison for our Storage Model Chain

The results show that our convergence rates are always better than the RT convergence rates. The first constants $M(x)$ are always reasonable. The obtained rates get faster as β increases.

Chapter 5

A First Order Autoregressive Chain

This chapter considers a first order causal autoregressive (AR(1)) process on the state space $\mathbb{R}$. Such a process obeys the stochastic difference equation

$$X_{n+1} = \varphi X_n + Z_{n+1}, \tag{5.1}$$

where $\varphi \in (-1, 1)$ and $\{Z_n\}_{n=1}^{\infty}$ are IID random variables with zero mean and variance $\sigma^2 > 0$. It is easy to check that $\{X_n\}_{n=0}^{\infty}$ is a Markov chain.

As with the last example, some simplifying assumptions are made up front to inject tractability into the ensuing calculations. First, we take $\varphi \in [0, 1)$ so that the chain will be stochastically ordered in its initial state. This results in a positively correlated $\{X_n\}_{n=0}^{\infty}$, which is the primary case encountered in time series practice. Should one encounter $\varphi \in (-1, 0)$, then we suggest examining $\{X_{2n}\}_{n=0}^{\infty}$, which is stochastically ordered. Second, we work with a normally distributed process, which is achieved by positing that $\{Z_n\}$ is independent and identically distributed normal noise. Finally, to scale the process, we take $\sigma^2 = 1$.

When $X_0 = x_0$, X_1 is normally distributed with mean φx_0 and unit variance. It follows that

$$P(x, dy) = \frac{1}{\sqrt{2\pi}} e^{-(y-\varphi x)^2/2} dy. \tag{5.2}$$

This chain is easily shown to be $\phi-$irreducible, where ϕ is the Lebesgue measure. The stationary distribution of $\{X_n\}$ is normal with mean 0 and variance $1/(1-\varphi^2)$. This limit law hence has a finite moment generating function of all orders: $E[\exp^{sX_\infty}] = e^{s^2/[2(1-\varphi^2]}$. Below, it is shown that the chain admits a drift condition. From this, it follows that $\{X_n\}$ is Harris positive recurrent [32].

Recursing (5.1) provides

$$X_n = \varphi^n X_0 + \sum_{j=0}^{n-1} \varphi^j Z_{n-j}. \tag{5.3}$$

When $\varphi \in [0,1]$, this shows that the chain is pathwise (and hence stochastically) ordered in its initial state. However, more can be extracted from (5.3): if $\{X_n\}$ and $\{X_n'\}$ are two chains driven by the same $\{Z_n\}$ but starting at x_0 and x_0', respectively, then

$$|X_n - X_n'| \leq \varphi^n |x_0 - x_0'|.$$

This sample path contraction will prove useful later.

5.1 A Minorization Condition

To find a probability measure ν and $\delta > 0$ satisfying (2.5), let $c > 0$ and define $C_1 = [-c, c]$. Below, we will choose c to optimize our convergence rate.

Lemma 10 *Define*

$$\delta = \int_{-c}^{0} \frac{1}{\sqrt{2\pi}} e^{-(y-\varphi c)^2/2} dy + \int_{0}^{c} \frac{1}{\sqrt{2\pi}} e^{-(y+\varphi c)^2/2} dy \tag{5.4}$$

and

$$\nu(dy) = \frac{1}{\delta}\Big(1_{(-c,0)}(y) \frac{1}{\sqrt{2\pi}} e^{-(y-\varphi c)^2/2} dy + 1_{[0,c)}(y) \frac{1}{\sqrt{2\pi}} e^{-(y+\varphi c)^2/2} dy\Big), \tag{5.5}$$

Then, for $x \in C_1$, the chain satisfies the minorization condition in (2.5).

Proof. We proceed in cases. First, when $y < 0$, by (5.2), (5.4), and (5.5), it is enough to show that for all $x \in C_1$,

$$\frac{1}{\sqrt{2\pi}} e^{-(y-\varphi x)^2/2} dy \geq \frac{1}{\sqrt{2\pi}} e^{-(y-\varphi c)^2/2} dy,$$

which follows from $|x| \leq c$. Similar arguments deal with the case where $y \geq 0$. □

5.2 A Drift Condition

We next establish a drift condition for the AR(1) model. Define the candidate drift function

$$V_c(x) = 2|x|1_{[c,\infty)}(|x|) + \alpha, \tag{5.6}$$

where $\alpha > 1$. As in the preceding section, convergence rates will be linked to the extrema of a function h_c, which in this case is $h_c(x) = \boldsymbol{E}_x[V_c(X_1)]/(2|x| + \alpha)$.

Lemma 11 *Define* $K(c) = \sup_{|x|>c} h_c(x)$. *Then*

$$K(c) \leq \frac{(1-\varphi)\alpha + 2\sqrt{2/\pi}}{2c+\alpha} + \varphi. \tag{5.7}$$

Proof. When $|X_1| \geq c$, $V_c(X_1) \leq 2(\varphi|x| + |Z_1|) + \alpha$ and when $|X_1| < c$, $V_c(X_1) \leq \alpha$. Hence, $V_c(X_1) \leq 2(\varphi|x| + |Z_1|) + \alpha$. Standard normality gives $E[|Z_1|] = \sqrt{2/\pi}$, implying that

$$h_c(x) \leq \frac{(1-\varphi)\alpha + 2\sqrt{2/\pi}}{2|x|+\alpha} + \varphi,$$

from which (5.7) follows. □

From the above, it is easy to see that $K(c) < 1$ for all $c > \sqrt{2/\pi}/(1-\varphi)$. Any such c will be larger than unity since $\varphi \in [0, 1)$.

Lemma 12 *Choose* $c > 1$ *to satisfy* $K(c) < 1$ *and* $b = \sup_{|x|\leq c}[h_c(x)(2|x| + \alpha)] - \alpha r_0^{-1}$, *where* $r_0 = 1/K(c)$. *Then* $V_c(x)$ *satisfies (2.7) with contraction parameter* r_0^{-1} *and constant* b.

Proof. By the definition of $h_c(x)$ gives $\boldsymbol{E}_x[V_c(X_1)] = (2|x|+\alpha)h_c(x)$. To identify a drift condition, when $|x| > c$, use Lemma 11 to get

$$\boldsymbol{E}_x[V_c(X_1)] = (2|x| + \alpha)h_c(x) \leq (2|x| + \alpha)K(c) = r_0^{-1}V_c(x).$$

When $|x| \leq c$, $V_c(x) = \alpha$ and $\boldsymbol{E}_x[V_c(X_1)] = PV_c(x) = h_c(x)(2|x| + \alpha) = h_c(x)(2|x| + \alpha) - \alpha r_0^{-1} + \alpha r_0^{-1}$. Taking $b = \sup_{|x|\leq c}[h_c(x)(2|x| + \alpha)] - r_0^{-1}$ gives the drift $PV_c(x) \leq r_0^{-1}V_c(x) + b1_{C_1}(x)$.

□

Lemma 13 *Under assumption of Lemma 12, Then, if* $1 < r \leq \min(\delta/(\delta\alpha - b), r_0)$,

$$\delta 1_{C_1}(x) + r\left(P - (\delta 1_{C_1} \otimes \nu)\right) V_c(x) \leq V_c(x). \tag{5.8}$$

Proof. For $|x| > c$, the result immediately follows from Lemma 12 if $r < r_0$. For $x \in C_1$, $\nu(C_1) = 1$. So, $\nu(V_c) = \alpha$, and

$$\begin{aligned} \delta + r(P - \delta\nu)V_c(x) \leq & \delta + r\left(r_0^{-1}V_c(x) + b\right) - r\delta\alpha \\ \leq & V_c(x) + \delta + r(b - \delta\alpha). \end{aligned}$$

If $1 < r \leq \min(\delta/(\delta\alpha - b), r_0)$, then (5.8) follows. □

To obtain convergence rates for the AR(1) model, set

$$F_x(y) = F(x, y) = Q(x, (-\infty, y]).$$

Since F is continuously differentiable on $\mathbb{R}^2$ and $F(x, \cdot)$ is strictly increasing for each $x \in \mathbb{R}$, the implicit function theorem implies that for each x, $F(x, \cdot)$ has an inverse function $H(x, \cdot)$. From Lemma 3.22 of [22], it follows that if θ is an uniformly distributed random variable, $H_x(\theta)$has distribution F_x. Since F_x and H_x are inverses, $F_x(H_x(\theta)) = \theta$. Fix $\theta = u$, as in the implicit function theorem, set $g(x) = H_x(u)$. Then, g is differentiable and $F(x, g(x)) = u$. Taking a derivative with respect to x and applying the chain rule gives

$$g'(x) = \frac{-F_1(x, g(x))}{F_2(x, g(x))}$$

where F_i denotes the partial derivative with respect to the ith component of F. To generate X_1 and X_1', generate θ and if $\theta = u$. Then

$$X_1 = H(x, u) = g(x) \text{ and } X_1' = H(x', u) = g(x'),$$

Thus,

$$|X_1 - X_1'| = |g(x) - g(x')| = \int_{x'}^{x} |g'(t)| dt \le |x - x'|.$$

To show that $|X_1 - X_1'| \le |x - x'| < c$, it is sufficient to show that $g'(x) \le 1$. The following Lemma with $g'(x) = \varphi P(x, dy)/(P(x, dy) - \delta' \nu(dy))$ establishes this.

Lemma 14 *For δ' with $0 < \delta' \le \delta(1 - \varphi)$, then*

$$\frac{\varphi P(x, dy)}{P(x, dy) - \delta' \nu(dy)} \le 1, \tag{5.9}$$

where δ is given in (5.4) and $\nu(dy)$ is given in (5.5).

Proof. Since $\delta' \le \delta(1 - \varphi)$, $\delta'/(1 - \varphi) \le \delta$. Moreover, from Lemma 10,

$$P(x, dy) \ge \delta \nu(dy) \ge \frac{\delta'}{1 - \varphi} \nu(dy),$$

proving (5.9). □

Next, we bound for $M(x)$ and get the convergence rates for AR(1) model. Define τ by

$$\tau = \inf\{n \ge 0 : |X_n - X_n'| \le c\}. \tag{5.10}$$

Given $X_0 = x$ and $X_0' = z$ with $x > z$, there exist a τ depending on x and z such that

$$|X_{1+\tau} - X_{1+\tau}'| \le \varphi^{\tau} |x - z|; \tag{5.11}$$

thus, if $\varphi^{\tau}|x - z| \le c$, then Lemma 14 implies that for all $n > \tau$ if $X_n \in C_1$ then $X_n' \in C_2 := [-2c, c]$. For this case of AR(1) models, $T \le \tau + T \circ \Theta_\tau = \tau + \eta_{C_1} \circ \Theta_\tau$. A similar argument applies in the case $x \le z$. Therefore, the chain is strongly shadowing. And, according Corollary 2, it is sufficient to bound for $\int_{-\infty}^{\infty} r_0^{\tau} \pi(dz)$ to yields the better convergence rates for AR(1). Next, we bound for $\int_{-\infty}^{\infty} r_0^{\tau} \pi(dz)$. By the definition of τ in (5.10) and $0 < \varphi < 1$,

$$\tau \le \frac{\ln(c) - \ln(|x - z|)}{\ln(\varphi)} = \frac{\ln(c)}{\ln(\varphi)} + \frac{-1}{\ln(\varphi)} \big(\ln |x - z| \big) \le \frac{\ln(c)}{\ln(\varphi)} + \frac{-1}{\ln \varphi} |x - z|.$$

The last inequality follows since $\ln |x - z| \le |x - z|$ and $\varphi < 1$. Then, from the bound for τ, we have

$$r_0^\tau \le r_0^{\frac{\ln(c)}{\ln(\varphi)}+\frac{-1}{\ln\varphi}|x-z|} = r_0^{\frac{\ln(c)}{\ln(\varphi)}} r_0^{\frac{-1}{\ln(\varphi)}|x-z|}. \tag{5.12}$$

Now set $r_0^{-1/\ln(\varphi)} = e^t$. Then, $t = -\ln(r_0)/\ln(\varphi)$ and the above gives

$$r_0^\tau \le r_0^{\frac{\ln(c)}{\ln(\varphi)}} \exp\left(\frac{-\ln(r_0)}{\ln(\varphi)}|x-z|\right).$$

Hence,

$$\begin{aligned}
\int_{-\infty}^{\infty} r_0^\tau \pi(dz) \le & r_0^{\frac{\ln(c)}{\ln(\varphi)}} \int_{-\infty}^{\infty} \exp\Big(\frac{-\ln(r_0)}{\ln(\varphi)}|x-z|\Big)\pi(dz) \\
=& r_0^{\frac{\ln(c)}{\ln(\varphi)}} \Big(\exp\Big(\frac{-\ln(r_0)}{\ln(\varphi)}x\Big)\int_{-\infty}^{x} \exp\Big(\frac{\ln(r_0)}{\ln(\varphi)}z\Big)\pi(dz) \\
&+ \exp\Big(\frac{\ln(r_0)}{\ln(\varphi)}x\Big)\int_{x}^{\infty} \exp\Big(\frac{-\ln(r_0)}{\ln(\varphi)}z\Big)\pi(dz)\Big) \\
\le & r_0^{\frac{\ln(c)}{\ln(\varphi)}} \left[\exp\Big(\frac{-\ln(r_0)}{\ln(\varphi)}x\Big)\Psi_\pi\Big(\frac{\ln(r_0)}{\ln(\varphi)}\Big) + \exp\Big(\frac{\ln(r_0)}{\ln(\varphi)}x\Big)\Psi_\pi\Big(\frac{-\ln(r_0)}{\ln(\varphi)}\Big)\right].
\end{aligned}$$

The second equality above is established by considering two cases $(i)x < z$ and $(ii)x \ge z$. The last inequality above follows from the definition of $\Psi_\pi(s)$. Since $\Psi_\pi(s) = e^{s^2/(2(1-\varphi))}$, $\Psi_\pi(\ln r_0/\ln\varphi)$ and $\Psi_\pi(-\ln r_0/\ln\varphi)$ are finite. This establishes (3.8), which implies (3.9). We now apply Corollary 2 to get our convergence rate. For a fixed φ, we set $X_0 = 10$ and select a set $C_1 = [-c, c]$ that yields a good convergence rate $r_0 = 1/K(c)$. For a numerical illustration, we consider $c \in (1.3, 4]$ and $\alpha \in (2, 40]$ for φ values in $\{0.1, 0.25, 0.5, 0.75\}$. The following table displays some convergence results, with the identified values of c. In this presentation, we compare to convergence rates in [1], which are labeled as 'Bax rates'.

	$\varphi = 0.1$	$\varphi = 0.25$	$\varphi = 0.5$	$\varphi = 0.75$
c	2.853	2.18	1.655	1.39
α	3.585	7.111	15.518	39.21
δ	0.773715	0.5793234	0.3949073	0.2821849
$\delta' = \delta(1-\varphi)$	0.6963435	0.4344925	0.1974537	0.07054623
r_0	2.564616	1.551767	1.144701	1.030529
b	3.622764	7.392219	16.44797	40.74607
CKL rate	2.564565	1.551759	1.144701	1.030529
Bax rate	2.4885	1.442	1.1214	1.0354
$M(x)$	520.9311	845.0045	1782.097	8730.82
$M(x)r^{-1000}$	0	1.266852e-188	3.621976e-56	7.602448e-10

Table 5.1: A Convergence Rate Comparison for AR(1) Chains.

Appendices

Appendix A

δ in the Minorization Condition for the Storage Model

Corollary 5 *If*

$$\delta = \int_0^{y_0} \frac{1}{\beta} \ln\Big(\frac{c+\beta}{c}\Big) dy + \int_{y_0}^{\beta} \frac{1}{\beta} \ln\Big(\frac{\beta}{y}\Big) dy.$$

Then,

$$\delta = \frac{\beta}{\beta + c}$$

Proof.

$$\begin{aligned}
\delta &= \int_0^{y_0} \frac{1}{\beta} \ln\Big(\frac{c+\beta}{c}\Big) dy + \int_{y_0}^{\beta} \frac{1}{\beta} \ln\Big(\frac{\beta}{y}\Big) dy \\
&= \frac{y_0}{\beta} \ln \frac{c+\beta}{c} + \frac{1}{\beta} \int_{y_0}^{\beta} \ln \beta - \ln y dy \\
&= \frac{y_0}{\beta} \ln \frac{c+\beta}{c} + \frac{\beta - y_0}{\beta} \ln \beta - \frac{1}{\beta}\Big(y \ln y - y\Big)\Big|_{y_0}^{\beta} \\
&= \frac{y_0}{\beta} \ln \frac{c+\beta}{c} + \frac{\beta - y_0}{\beta} \ln \beta - \frac{1}{\beta}\Big(\beta \ln \beta - \beta - y_0 \ln y_0 + y_0\Big) \\
&= \frac{\frac{\beta c}{\beta + c}}{\beta} \ln \frac{c+\beta}{c} + \frac{\beta - \frac{\beta c}{\beta + c}}{\beta} \ln \beta - \frac{1}{\beta}\Big(\beta \ln \beta - \beta - \frac{\beta c}{\beta + c} \ln \frac{\beta c}{\beta + c} + \frac{\beta c}{\beta + c}\Big) \\
&= \frac{c}{\beta + c} \ln \frac{c+\beta}{c} + \frac{\beta}{\beta + c} \ln \beta - \ln \beta + 1 + \frac{c}{\beta + c} \ln \frac{\beta c}{\beta + c} - \frac{c}{\beta + c}.
\end{aligned} \tag{A.1}$$

Combine the term 1 and term 5, term 2 and term 3, term 4 and term 6, we have

$$\begin{aligned} \delta &= \frac{c}{\beta + c} \ln \beta - \frac{c}{\beta + c} \ln \beta + \frac{\beta}{\beta + c} \\ &= \frac{\beta}{\beta + c}. \end{aligned} \tag{A.2}$$

□

Appendix B

R Code for a Storage Model

```
Storage_CKL.rates <- function(c = 1.6, alpha = 7.5, coefs = 3, beta = 1){
  ds = seq(c,c+beta,0.1)
  h_c = seq(c,c+beta,0.1)
  m = length(ds)
  for (j in 1:m){
    d = ds[j]
    ###########################
    #####For x > c and V(x) = alpha + coefs * x1_{C^C}(x)
    h_c[j] = (coefs*(-log((d+beta)/d) *c^2/(2*beta)
    +d/2 +beta/4)+ alpha)/(alpha + coefs*d)
  }
  # print(h_c)
  # print("rho")
  rho = max(h_c)
  r_0_w = 1/rho
  r_0_w
  #r_0s[i] = min(r_0_w,1.5)
  dCs = seq(0, c,0.01)
  dri_c = seq(0,c,0.01)
```

```
bCs = seq(0,c,0.01)
mC = length(dCs)
#alpha = 2
delta = beta/(beta+c)
for (j in 1:mC){
  d = dCs[j]
  if (d + beta <= c){
    bCs[j]= alpha
  }else{
    ########################################
    #####For x < c and V(x) = alpha  + coefs x 1_(C^C)(x)
    bCs[j] = alpha  + (- c^2/2/beta * log((d+beta)/c)
    - (c^2 -d^2)/(4*beta) + 1/2*(d + beta/2))*coefs
  }
  dri_c[j] = delta + r_0_w *(bCs[j] - delta * alpha) - alpha
}
covs = max(dri_c)
b = max(bCs) - alpha /r_0_w
RTds = max(bCs) ####need to check it
#print(bCs)
Js = RTds + r_0_w*(b - delta)
eta = log(Js/(1-delta))/log(r_0_w)
RT_rs = max(1, 1/ (1-delta)^(1/eta))
CKL_rs = 1
if (covs < 0){
  CKL_rs = r_0_w
}
else{
  r_0_w = 1
}
```

```
  x = 10
  M_x = alpha + x + b/(1- r_0_w^{-1})
  M_x
  M_x/CKL_rs^1000
  pars = matrix(-99, nrow = 1, ncol = 11)
  pars[1] = c
  pars[2] = alpha
  pars[3] = delta
  pars[4] = r_0_w
  pars[5] = b
  pars[6] = CKL_rs
  pars[7] = M_x
  pars[8] = M_x*CKL_rs^(-1000)
  pars[9] = covs
  pars[10] = coefs
  pars[11] = RT_rs
  return (list("c" = pars[1], "alpha" = pars[2],"delta" = pars[3],
  "r_0" = pars[4],"b" = pars[5],"CKL_rs" = pars[6],
  "M_x" = pars[7],"bound" = pars[8],
  "covs" = pars[9],"coefs" = pars[10],"RT_rs" = pars[11]))
}

Storage_RT.rates <- function(c = 1.6, alpha = 7.5, coefs = 3, beta = 1){
  ds = seq(c,c+beta,0.1)
  h_c = seq(c,c+beta,0.1)
  m = length(ds)
  for (j in 1:m){
    d = ds[j]
    ############################
    #####For x > c and V(x) = alpha + coefs * x1_{C^C}(x)
    h_c[j] = (coefs*(-log((d+beta)/d) *c^2/(2*beta)
```

```
  +d/2 +beta/4)+ alpha)/(alpha + coefs*d)
}
# print(h_c)
# print("rho")
rho = max(h_c)
#rho = (alpha*(1 - beta) + sqrt(2/pi) )/(alpha + c) + beta
#rho = (beta/2*log(c) - beta/2*log(c+beta) +(c+beta)^2/(4*beta)
#-c^2/beta/4 + alpha)/(alpha + c)
r_0_w = 1/rho
r_0_w
#r_0s[i] = min(r_0_w,1.5)
dCs = seq(0, c,0.01)
dri_c = seq(0,c,0.01)
bCs = seq(0,c,0.01)
mC = length(dCs)
#alpha = 2
delta = beta/(beta+c)
for (j in 1:mC){
  d = dCs[j]
  if (d + beta <= c){
    bCs[j]= alpha
  }else{
    ##########################################
    #####For x < c and V(x) = alpha  + coefs x 1_(C^C)(x)
    bCs[j] = alpha  + (- c^2/2/beta * log((d+beta)/c)
    - (c^2 -d^2)/(4*beta) + 1/2*(d + beta/2))*coefs
  }
  dri_c[j] = delta + r_0_w *(bCs[j] - delta * alpha) - alpha
}
covs = max(dri_c)
```

```
  b = max(bCs) - alpha /r_0_w
  RTds = max(bCs) ####need to check it
  #print(bCs)
  Js = RTds + r_0_w*(b - delta)
  eta = log(Js/(1-delta))/log(r_0_w)
  RT_rs = max(1,min(r_0_w, 1/ (1-delta)^(1/eta)))
  pars = matrix(-99, nrow = 1, ncol = 6)
  pars[1] = c
  pars[2] = alpha
  pars[3] = delta
  pars[4] = r_0_w
  pars[5] = b
  pars[6] = RT_rs
  return (list("c" = pars[1], "alpha" = pars[2],"delta" = pars[3],
  "r_0" = pars[4],"b" = pars[5],"RT_rs" = pars[6]))
}
```

Appendix C

R Code for Our AR(1) Gaussian Autoregressive Model

```
CKL.rates <- function(c = 1.6, alpha = 7.5, coefs = 1, varphi = 0.5){
 ds = seq(c,c+40,0.1)
 h_c = seq(c, c + 40,0.1)
 m = length(ds)
 ##########################################################################
 ##########################################################################
 #####For x > c, we calculate the E_x[ V(X_1)]
 ##########################################################################
 for (j in 1:m){
   d = ds[j]
   #integrand_density_b <- function(x) {1/(sqrt(2*pi))
   * exp(-(x-varphi*d)^2/2)}
   integrand_mean <- function(x) {1/(sqrt(2*pi))
   * exp(-(x-varphi*d)^2/2)*abs(x)}
   mean_up = integrate(integrand_mean, lower = c, upper = Inf);
   meanV = mean_up$value;
   meanV
```

```
  mean_low = integrate(integrand_mean, lower = -Inf, upper = -c)
  meanV = meanV + mean_low$value;
  meanV
  h_c[j] = (alpha + coefs*meanV)/(alpha + coefs*d)
}
rho = max(h_c)
r_0_w = 1/rho
CKL = r_0_w
##Calculate the Cumulative Distribution from -c to 0
delta = (pnorm(0,varphi * c, 1)- pnorm(-c,varphi * c, 1))
+ (pnorm(c, -varphi * c, 1) - pnorm(0,- varphi * c,1))
#print("pnorm",pnorm(0,varphi * c, 1)
#+ (1 - pnorm(0,- varphi * c,1)))
#delta = pnorm(0,varphi * c, 1)
#+ (1 - pnorm(0,- varphi * c,1))
dCs = seq(-c, c,0.01)
hC_c = seq(-c,c,0.01)
mC = length(dCs)
for (j in 1:mC){
  d = dCs[j]
  #integrand_density_b <- function(x) {1/(sqrt(2*pi))
  #* exp(-(x-varphi*d)^2/2)}
  integrand_mean <- function(x) {1/(sqrt(2*pi))
  * exp(-(x-varphi*d)^2/2)*abs(x)}
  mean_up = integrate(integrand_mean, lower = c, upper = Inf);
  meanV = mean_up$value;
  #mean
  mean_low = integrate(integrand_mean, lower = -Inf, upper = -c)
  meanV = meanV + mean_low$value;
  hC_c[j] = (alpha + coefs*meanV)
```

```
    #print(hC_c[j])
  }
  b = max(hC_c)
  delta_prime =  delta*(1-varphi)
  CKL = max(min(r_0_w, (alpha - delta_prime)/(b - delta_prime * alpha)), 1)
  CKL
  covs = delta_prime+ CKL *(b - delta_prime * alpha) - alpha
  x = 10
  val = r_0_w^(log(c)/log(varphi))*(exp(-log(r_0_w)/log(varphi) *x)
*exp((log(r_0_w)/log(varphi))^2/(2*(1-varphi)))
+ exp(log(r_0_w)/log(varphi) *x)
*exp((-log(r_0_w)/log(varphi))^2/(2*(1-varphi))) )
  M_x = alpha + x + 2*b/(1- r_0_w^{-1})*val
  pars = matrix(-99, nrow = 1, ncol = 10)
  pars[1] = c
  pars[2] = alpha
  pars[3] = delta
  pars[4] = delta*(1-varphi)
  pars[5] = r_0_w
  pars[6] = b
  pars[7] = CKL
  pars[8] = M_x
  pars[9] = M_x*CKL^(-1000)
  pars[10] = covs
  return (list("c" = pars[1], "alpha" = pars[2],"delta" = pars[3],
  "delta*(1-varphi)" = pars[4],
  "r_0" = pars[5],"b" = pars[6],
  "CKL_rs" = pars[7],"M_x" = pars[8],
  "bound" = pars[9],"covs" = pars[10]))
}
```

```
Bax.rates <- function(c = 1.6, alpha = 7.5, coefs = 2, varphi = 0.5){
  ds = seq(c,c+40,0.1)
  h_c = seq(c, c + 40,0.1)
  m = length(ds)
  ##############################################################################
  ##############################################################################
  #####For x > c, we calculate the E_x[ V(X_1)]
  ##############################################################################
  for (j in 1:m){
    d = ds[j]
    #integrand_density_b <- function(x) {1/(sqrt(2*pi))
    * exp(-(x-varphi*d)^2/2)}
    integrand_mean <- function(x) {1/(sqrt(2*pi))
    * exp(-(x-varphi*d)^2/2)*abs(x)}
    mean_up = integrate(integrand_mean, lower = c, upper = Inf);
    mean = mean_up$value;
    mean
    mean_low = integrate(integrand_mean, lower = -Inf, upper = -c)
    mean = mean + mean_low$value;
    mean
    h_c[j] = (alpha + coefs*mean)/(alpha + coefs*d)
  }
  rho = max(h_c)
  r_0_w = 1/rho
  CKL = r_0_w
  #print(CKL)
  ##Calculate the Cumulative Distribution from -c to 0
  delta = (pnorm(0,varphi * c, 1)- pnorm(-c,varphi * c, 1))
  + (pnorm(c, -varphi * c, 1) - pnorm(0,- varphi * c,1))
```

```
#print("pnorm",pnorm(0,varphi * c, 1) + (1 - pnorm(0,- varphi * c,1)))
#delta = pnorm(0,varphi * c, 1) + (1 - pnorm(0,- varphi * c,1))
dCs = seq(-c, c,0.01)
hC_c = seq(-c,c,0.01)
mC = length(dCs)
for (j in 1:mC){
  d = dCs[j]
  #integrand_density_b <- function(x) {1/(sqrt(2*pi))
  #* exp(-(x-varphi*d)^2/2)}
  integrand_mean <- function(x) {1/(sqrt(2*pi))
  * exp(-(x-varphi*d)^2/2)*abs(x)}
  mean_up = integrate(integrand_mean, lower = c, upper = Inf);
  mean = mean_up$value;
  #mean
  mean_low = integrate(integrand_mean, lower = -Inf, upper = -c)
  mean = mean + mean_low$value;
  hC_c[j] = (alpha + coefs*mean)
}
b = max(hC_c)
#####################################################################
#####################################################################
######################Calculate the Baxendale convergent rate with
######################the same drift condiction and minorization
#######################################################################
K = max(hC_c)
b_w = b - alpha /r_0_w
Js = K - delta
Js/(1-delta)
eta =1+ log(Js/(1-delta))/log(r_0_w)
Rts = min(r_0_w,1/ ((1-delta)^(1/eta)))
```

```
#print(Rts)
pars = matrix(-99, nrow = 1, ncol = 6)
rate = 1
if (Rts > 1){
  #print(Rts)
  RRS = seq(1,Rts,0.1)
  #rate_cri = 1 + 2* deltas[i] *Rts - delta*Rts
  #/(1- (1-deltas[i])*(Rts^etas[i]))
  rate_cri2 = (1 + 2* delta *Rts)*(1- (1-delta)*(Rts^eta) )
  - delta*Rts
  if (rate_cri2 < 0){
    mm = length(RRS)
    for (k in 1:mm){
      cri = 1 + 2* delta *RRS[k]
      -delta*RRS[k]/(1- (1-delta)*(RRS[k]^eta))
      if (cri > 0){
        rate = RRS[k]
    #print("TestingTestingTestingTesting")
      }
    }
  }
  else{
    rate = Rts
  }
Bax = rate
#pars = matrix(-99, nrow = 1, ncol = 6)
pars[1] = c
pars[2] = alpha
pars[3] = delta
pars[4] = r_0_w
```

```
    pars[5] = K
    pars[6] = Bax
    }
    return (list("c" = pars[1], "alpha" = pars[2],"delta" = pars[3],
    "r_0" = pars[4],"K" = pars[5],"Bax_rs" = pars[6]))
}
```

Bibliography

[1] Peter H. Baxendale. Renewal theory and computable convergence rates for geometrically ergodic Markov chains. *The Annals of Applied Probability*, 15(1B):700–738, 2005.

[2] Pierre Brémaud. Eigenvalues and nonhomogeneous Markov chains. In *Markov Chains*, pages 195–251. Springer, 1999.

[3] Erhan Cinlar. *Introduction to Stochastic Processes*. Courier Corporation, 1975.

[4] Persi Diaconis and Laurent Saloff-Coste. Comparison techniques for random walk on finite groups. *The Annals of Probability*, 3:2131–2156, 1993.

[5] Persi Diaconis and Laurent Saloff-Coste. Comparison theorems for reversible Markov chains. *The Annals of Applied Probability*, 3(3):696–730, 1993.

[6] Persi Diaconis and Daniel Stroock. Geometric bounds for eigenvalues of Markov chains. *The Annals of Applied Probability*, 1:36–61, 1991.

[7] R Douc, E Moulines, and J. Rosenthal. Quantitative convergence rates for inhomogeneous Markov chains. *Electronic Communications in Probability*, 7:123–128, 2002.

[8] Randal Douc, Eric Moulines, Pierre Priouret, and Philippe Soulier. *Markov Chains*. Springer, 2018.

[9] Julia Gaudio, Saurabh Amin, and Patrick Jaillet. Exponential convergence rates for stochastically ordered Markov processes under perturbation. *Systems & Control Letters*, 133:104515, 2019.

[10] Andrew Gelman, Frederic Bois, and Jiming Jiang. Physiological pharmacokinetic analysis using population modeling and informative prior distributions. *Journal of the American Statistical Association*, 91(436):1400–1412, 1996.

[11] Andrew Gelman and Donald B Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):457–472, 1992.

[12] Andrew Gelman and Donald B Rubin. A single sequence from the Gibbs sampler gives a false sense of security. *Bayesian Statistics*, 4:625–631, 1992.

[13] Charles J Geyer. Practical Markov chain Monte Carlo. *Statistical Science*, pages 473–483, 1992.

[14] Walter R Gilks, Sylvia Richardson, and David Spiegelhalter. *Markov chain Monte Carlo in Practice*. CRC press, 1995.

[15] Zhenting Hou, Yuanyuan Liu, and Hanjun Zhang. Subgeometric rates of convergence for a class of continuous-time Markov process. *Journal of Applied Probability*, 42(3):698–712, 2005.

[16] Søren Fiig Jarner and Ernst Hansen. Geometric ergodicity of Metropolis algorithms. *Stochastic Processes and their Applications*, 85(2):341–361, 2000.

[17] Daniel C Jerison. Quantitative convergence rates for reversible Markov chains via strong random times. *ArXiv Preprint ArXiv:1908.06459*, 2019.

[18] Zhumengmeng Jin and James P Hobert. Dimension free convergence rates for Gibbs samplers for Bayesian linear mixed models. *ArXiv Preprint ArXiv:2103.06324*, 2021.

[19] Leif T Johnson and Charles J Geyer. Variable transformation to obtain geometric ergodicity in the random-walk Metropolis algorithm. *The Annals of Statistics*, pages 3050–3076, 2012.

[20] Galin L. Jones and James P. Hobert. Honest exploration of intractable probability distributions via Markov chain Monte Carlo. *Statistical Science*, pages 312–334, 2001.

[21] Galin L Jones and James P Hobert. Sufficient burn-in for Gibbs samplers for a hierarchical random effects model. *The Annals of Statistics*, 32(2):784–817, 2004.

[22] Olav Kallenberg. *Foundations of Modern Probability*. Springer, 2002.

[23] Masaaki Kijima. *Markov Processes for Stochastic Modeling*, volume 6. CRC Press, 1997.

[24] Torgny Lindvall. *Lectures on the Coupling Method*. Courier Corporation, 2002.

[25] Jun S Liu, Wing H Wong, and Augustine Kong. Covariance structure and convergence rate of the Gibbs sampler with various scans. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):157–169, 1995.

[26] Yuanyuan Liu, Hanjun Zhang, and Yiqiang Zhao. Subgeometric ergodicity for continuous-time Markov chains. *Journal of Mathematical Analysis and Applications*, 368(1):178–189, 2010.

[27] R. B. Lund, S. P. Meyn, and R. L. Tweedie. Computable exponential convergence rates for stochastically ordered Markov processes. *The Annals of Applied Probability*, 6(1):218–237, 1996.

[28] R. B. Lund and Richard L. Tweedie. Geometric convergence rates for stochastically ordered Markov chains. *Mathematics of Operations Research*, 21(1):182–194, 1996.

[29] Robert Lund, Ying Zhao, and Peter C Kiessler. A monotonicity in reversible Markov chains. *Journal of Applied Probability*, 43(2):486–499, 2006.

[30] S. P. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability*. Springer Science & Business Media, 1993.

[31] S. P. Meyn and R. L. Tweedie. Computable bounds for geometric convergence rates of Markov chains. *The Annals of Applied Probability*, 4(4):981–1011, 1994.

[32] Esa Nummelin. *General Irreducible Markov Chains and Non-negative Operators*, volume 83. Cambridge University Press, 1984.

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

[34] Gareth O Roberts and Richard L Tweedie. Bounds on regeneration times and convergence rates for Markov chains. *Stochastic Processes and their Applications*, 80(2):211–229, 1999.

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

[36] Gareth O Roberts and Richard L Tweedie. Geometric L^2 and L^1 convergence are equivalent for reversible Markov chains. *Journal of Applied Probability*, 38(A):37–41, 2001.

[37] Jeffrey S. Rosenthal. Minorization conditions and convergence rates for Markov chain Monte Carlo. *Journal of the American Statistical Association*, 90(430):558–566, 1995.

[38] Jeffrey S Rosenthal. Rates of convergence for Gibbs sampling for variance component models. *The Annals of Statistics*, 23(3):740–761, 1995.

[39] Jeffrey S. Rosenthal. Quantitative convergence rates of Markov chains: A simple account. *Electronic Communications in Probability*, 7:123–128, 2002.

[40] Jeffrey S Rosenthal. Asymptotic variance and convergence rates of nearly-periodic Markov chain Monte Carlo algorithms. *Journal of the American Statistical Association*, 98(461):169–177, 2003.

[41] Andrey Sarantsev. Explicit rates of exponential convergence for reflected jump diffusions on the half-line. *Latin American Journal of Probability and Mathematical Statistics*, 16(2):1069–1093, 2016.

[42] D. Scott and R. Tweedie. Explicit rates of convergence of stochastically ordered Markov chains. In *Athens Conference on Applied Probability and Time Series Analysis*, pages 176–191. Springer, 1996.

[43] Sabrina Sixta and Jeffrey S Rosenthal. Convergence rate bounds for iterative random functions using one-shot coupling. *ArXiv Preprint ArXiv:2112.03982*, 2021.

[44] Dietrich Stoyan, Wilfrid S Kendall, Sung Nok Chiu, and Joseph Mecke. *Stochastic Geometry and its Applications*. John Wiley & Sons, 2013.

[45] Volker Strassen. The existence of probability measures with given marginals. *The Annals of Mathematical Statistics*, 36(2):423–439, 1965.

[46] Pekka Tuominen and Richard L Tweedie. Subgeometric rates of convergence of f-ergodic Markov chains. *Advances in Applied Probability*, 26(3):775–798, 1994.

[47] D Vere-Jones. On the spectra of some linear operators associated with queueing systems. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 2(1):12–21, 1963.