

RESEARCH ARTICLE

Convergence Rates for Shadowing Markov Chains

Fun Choi John Chan¹, Peter Kiessler¹ and Robert Lund²¹School of Mathematical and Statistical Sciences, Clemson University, Clemson, 29634, SC, U.S..²Baskin School of Engineering, University of California, Santa Cruz, Santa Cruz, 95064, CA, U.S..**Corresponding author:** Fun Choi John Chan; Email: funcchan@gmail.com.**Keywords:** Coupling, Drift, Minorization, Shadowing, Splitting, Stochastic Monotonicity

Abstract

This paper derives explicit convergence rates for Markov chains possessing a so-called shadowing structure. Shadowing means that when one trajectory of the chain lies within a particular set of states, another trajectory of the chain can be constructed that is guaranteed to lie within another set of states. Shadowing techniques accommodate chains with contracting sample paths and stochastically ordered chains. Our methods are based on classic coupling, drift, and minorization techniques. To illustrate the methods, two detailed examples are presented: 1) a storage chain, which is stochastically ordered in its initial state but never returns to the minimal element in its state space, and 2) a first-order Gaussian autoregression, where sample paths of the chain contract from different initial levels.

1. Introduction

Let $\{X_n\}_{n=0}^\infty$ be an ergodic (aperiodic, irreducible, and positive recurrent) time-homogeneous Markov chain on the state space $(\mathbb{R}, \mathcal{A})$, where $\mathcal{A}$ is the Borel measurable subsets of the real numbers $\mathbb{R}$. While extensions to $\mathbb{R}^k$ are easily considered for $k \in \{2, 3, \dots\}$, we isolate on the univariate case for simplicity. The chain's n -step-ahead transition kernel will be denoted by $P^n(x_0, A) = \mathbb{P}[X_n \in A | X_0 = x_0]$; the superscript is omitted when $n = 1$.

Proposition 6.3 of [24] 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

$$\lim_{n \rightarrow \infty} \|P^n(x_0, \cdot) - \pi(\cdot)\| = 0$$

for all initial points $X_0 = x_0$. Here, $\|P^n(x_0, \cdot) - \pi(\cdot)\| = \sup_{A \in \mathcal{A}} |P^n(x_0, A) - \pi(A)|$ indicates a total variation norm: $\|X - X'\| = \sup_{A \in \mathcal{A}} |\mathbb{P}(X \in A) - \mathbb{P}(X' \in A)|$ is the total variational distance between the random variables X and X' . While other norms can be considered, total variational distances are classic in Markov chain analyses.

The rate at which $\|P^n(x_0, \cdot) - \pi(\cdot)\|$ converges to zero is of considerable importance. Convergence is called geometric at rate $r \in (1, \infty)$ if there exists a finite “first-constant” $M(x_0)$ for each deterministic starting point $X_0 = x_0$ such that

$$\|P^n(x_0, \cdot) - \pi(\cdot)\| \leq M(x_0)r^{-n}, \quad n \in \{0, 1, 2, \dots\}. \quad (1.1)$$

Explicitly identifying values of r and $M(x_0)$ in (1.1) has been an active area of probabilistic research [24, 17, 22, 28, 3]. Convergence of Markov chain Monte Carlo (MCMC) generated chains, for example, has important ramifications for the simulation community [8, 25, 31]. Indeed, some MCMC chains do not rapidly converge (mix) at all [8, 6, 5, 4, 7].

Three texts studying Markov chain convergence on general state spaces are [24], [22], and [3]. A tactic used there to analyze chain convergence is the splitting technique of [24]. Splitting constructs a

so-called split chain from the original chain's one-step-ahead dynamics. Split chains retain the marginal distributions of the chain at each fixed time, but have an atom in their state space, which is a set of states the chain repeatedly enters. Convergence rates for the original chain are obtained from 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 create one. One then shows that the first return time of the chain to the atom has a geometric tail, from which a geometric convergence rate can be extracted. This has been classically accomplished by establishing drift and minorization conditions for the chain [22].

Convergence rates for general chains were established in [23] using renewal recursions and the spectral theory of bounded linear operators. The rates obtained there, while quite general, are often poor and of little practical use. Improving these, explicit values of r in (1.1) are identified in [26] from the probability generating function of the return times of the chain to some atom; while these rates improve those in [23] and [28], values of $M(x_0)$ are eschewed. [28] constructs a so-called coupling (meeting) time T of two chain trajectories, one of which is stationary, from a bivariate drift condition. This can be used to bound $\mathbb{P}(T > n)$ and extract convergence rate information for the chain. [26] and [13] further discuss relationships between geometric convergence rates, minorization techniques, and drift conditions. Other prominent papers studying Markov chain convergence include [2], [28], and [30].

Authors have also obtained explicit convergence rates for chains having additional structure. A dependent sample path coupling is used in [20] to obtain convergence rates of stochastically ordered chains taking values in $[0, \infty)$. While these chains are assumed to have an atom at the minimal 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 [19]. Having a state $\{0\}$ atom implies that the chain will repeatedly return to state zero. While this holds for many storage, queueing, and reflected random walk processes, other chains do not naturally have any atom in their state spaces (Section 4 here gives an example). Rates for chains having other stochastic orderings, such as new worse than used and decreasing hazard rate states, are discussed in [16] and [21].

Reversible Markov chains are another heavily researched stochastic structure. The reference [1] is a good treatise for rates of general reversible chains; these are improved in [10]. MCMC simulated chains, a setting where convergence is sometime nebulous, 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 [18], [28], [29], [13], [14], and [11]. Geometric convergence rates of Metropolis algorithms, which are also reversible, are studied in [12], [9], and [23].

The rest of this paper proceeds as follows. The next section narrates technical preliminaries and introduces shadowing chains. Section 3 proves our main results. Sections 4 and 5 consider applications to storage model chains and first-order autoregressions, respectively. Section 6 concludes with some remarks.

2. Background

We assume that $\{X_n\}_{n=0}^\infty$ is an aperiodic, irreducible (ϕ -irreducible), and positive recurrent (Harris recurrent) time-homogeneous Markov chain on a state space that is a subset of $\mathbb{R}$. We refer the reader to [24], [22], and [3] for general background.

Our notation uses $\mathbb{P}_x(X_1 \in A) = \mathbb{P}[X_{n+1} \in A | X_n = x] := P(x, A)$ for the one-step-ahead transition probabilities and $P^n(x_0, A) := \mathbb{P}[X_n \in A | X_0 = x_0]$ for n -step-ahead transition probabilities for $n \geq 2$. Here, P is assumed to be a probability kernel. A kernel is a nonnegative mapping K from the product space of $\mathbb{R}$ and $\mathcal{A}$ such that for each $A \in \mathcal{A}$, $K(\cdot, A)$ is a measurable function on $\mathbb{R}$ and for each $x \in \mathbb{R}$, $K(x, \cdot)$ is a measure on $\mathcal{A}$. If f is a measurable mapping on $\mathbb{R}$, we use the notation

$$Kf(x) = \int_{\mathbb{R}} f(y)K(x, dy).$$

For initial condition notation, we take $X(0) = x_0$ when the chain starts at the deterministic initial level x_0 ; should the initial level be random, we write X_0 for $X(0)$ and $\mu_{X_0}(A) = P(X_0 \in A)$ for the initial measure.

Two main methods have been used to obtain convergence rates: 1) renewal and regenerative techniques [24, 23], and 2) coupling methods [28, 17, 20]. Other methods exist; for one recent example, [32] develops iterative random functions and the so-called one-shot coupling method, but even this is based heavily on 1) and 2).

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 the set C . Then [24] establishes an inequality of form

$$\|P^n(x_0, \cdot) - \pi(\cdot)\| \leq \mathbb{P}_{x_0}(\tau_C \geq n) + \text{extra positive terms.} \quad (2.1)$$

While the form of the extra positive terms is unimportant here, significant work is usually required to bound them. Coupling methods tend to work a bit more cleanly (caveat: when they work at all). The general idea is to take two chain trajectories $\{X_n\}_{n=0}^\infty$ and $\{X'_n\}_{n=0}^\infty$ having the same transition kernel P , but with different starting conditions. Many times, X'_0 will be random and taken to have the stationary distribution π . Then X'_n also has distribution π for each $n \geq 1$ and $\{X'_n\}_{n=0}^\infty$ is a stationary chain.

Coupling works by analyzing the time where the two chains first meet: $T := \inf\{n \geq 0 : X_n = X'_n\}$. The famous coupling inequality is [17]

$$\|P^n(x_0, \cdot) - \pi(\cdot)\| \leq \mathbb{P}_{x_0, X'_0}(T > n). \quad (2.2)$$

Our notation here places the subscripts x_0 and X'_0 on quantities to bookkeep initial conditions of the two chains. Here, X'_0 is taken to have the stationary distribution π so that $\{X'_n\}_{n=0}^\infty$ is stationary.

Significant convergence rate information can be extracted from T . For example, if $\mathbf{E}_{x_0, X'_0}[r^T]$ can be proven finite for some $r > 1$, then (1.1) will hold at the geometric rate r^{-1} (see (2.3) below). Complications in the above, however, arise. Foremost, on continuous state spaces, T may not be finite (may never occur). It can also be difficult to calculate the distribution of T . However, if $\mathbf{E}_{x_0, X'_0}[r^T]$ can be proven finite for an $r > 1$, then Markov's inequality applied to r^T and (2.2) give

$$\|P^n(x_0, \cdot) - \pi(\cdot)\| \leq r^{-n} \mathbf{E}_{x_0, X'_0}[r^T]. \quad (2.3)$$

This identifies r as a geometric convergence rate and $M(x_0) = \mathbf{E}_{x_0, X'_0}[r^T]$ as a first constant in (1.1).

Some additional stochastic structure, such as chain reversibility or stochastic monotonicity, is often present in applications and may be used to obtain improved convergence rates. A stochastically ordered chain is a chain that is higher at all times when it starts higher initially: if $\{X_n\}_{n=0}^\infty$ and $\{X'_n\}_{n=0}^\infty$ are copies of the chain starting 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 X' means that $\mathbb{P}[X > x] \geq \mathbb{P}[X' > x]$ for all real x . The texts [17], [16], and [33] are good references for stochastic orderings.

Every Markov chain $\{X_n\}_{n=0}^\infty$ has the simple representation

$$X_{n+1} = g(X_n, U_{n+1}), \quad n \geq 0, \quad (2.4)$$

for some measurable function g and initial state X_0 . In this representation, $\{U_n\}_{n=1}^\infty$ is independent and identically distributed (IID) and independent of X_0 . For a stochastically ordered chain, for each fixed u , $g(x, u)$ must be monotone non-decreasing in x (almost everywhere in x relative to π).

Suppose that $g(x, u)$ is non-decreasing 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 Straussen's Theorem [34] that every stochastically ordered chain can be made pathwise ordered on a different probability space.

2.1. Shadowing Chains

The new aspect of this paper is the property of shadowing.

Definition: A Markov chain $\{X_n\}_{n=0}^\infty$ is said to shadow if there exist two sets of states, say C and C' , and two trajectories of the chain, say $\{X_n\}_{n=0}^\infty$ and $\{X'_n\}_{n=0}^\infty$, such that $X_n \in C$ implies that $X'_n \in C'$. A chain is said to weakly shadow if it shadows after a random initial time; that is, the shadowing property holds for all times $n > \tau$, where τ is some stopping time.

Any versions of the chains $\{X_n\}$ and $\{X'_n\}$ satisfying the shadowing condition can be used; in fact, X_n and X'_n can be highly dependent. The sets C and C' are called shadowing sets.

Every stochastically ordered chain shadows. To see this, if $\{X_n\}$ and $\{X'_n\}$ are two chain copies of a stochastically ordered chain with $X_0 \geq X'_0$, then $X_n \geq X'_n$ for all $n \geq 1$; hence, when $X_n \in (-\infty, c]$ for any c , then $X'_n \in (-\infty, c]$ as well and $C = C' = (-\infty, c]$ can be used as shadowing sets.

For an example of a chain that weakly shadows but is not stochastically ordered, consider the Markov recursion

$$X_{n+1} = \frac{1}{2} \sin(X_n) + Z_{n+1}, \quad n \geq 0.$$

Here, $\{Z_n\}_{n=1}^\infty$ is IID and standard normal, with Z_n taken independent of $X_n, X_{n-1}, \dots$. This chain has a limit distribution (we will not investigate this limit here), but is not stochastically ordered due to the sin component. However, this chain satisfies a weak shadowing condition with the shadowing sets $C = [-1, 1]$ and $C' = [-2, 2]$. To see this, note that if the same $\{Z_n\}_{n=1}^\infty$ sequence is used to drive two chains — say $\{X_n\}_{n=0}^\infty$ and $\{X'_n\}_{n=0}^\infty$, with the initial deterministic starting levels x_0 and x'_0 respectively — then

$$|X_{n+1} - X'_{n+1}| = \frac{1}{2} |\sin(X_n) - \sin(X'_n)| \leq \frac{1}{2} |X_n - X'_n|.$$

Hence, the two chains contract by a factor of $1/2$ at each iteration. Thus, $|X_n - X'_n| \leq 2^{-n} |X_0 - X'_0|$ and for large n , $|X_n - X'_n| < 1$. For such n , whenever $X_n \in [-1, 1] = C$, $X'_n \in [-2, 2] = C'$.

Expanding the above example, in some cases, one can construct chain trajectories $\{X_n\}_{n=0}^\infty$ and $\{X'_n\}_{n=0}^\infty$ that contract in the sense that $X_n - X'_n$ converges to zero. Section 5 below gives an autoregressive example of this “contraction”. Suppose that n is large enough so that $|X_n - X'_n| \leq b$. Then for such n , if $X_n \in [-c, c]$ for some $c > 0$, then $X'_n \in [-(c+b), (c+b)]$ and weakly shadowing sets are $C = [-c, c]$ and $C' = [-(c+b), (c+b)]$.

Furthering this logic, return to the simple representation in (2.4) and suppose that $g(x, u)$ satisfies

$$\sup_x |g(x+h, u) - g(x, u)| \leq Mh,$$

for $h > 0$ and $M < 1$. Then by driving two chain trajectories $\{X_n\}_{n=0}^\infty$ and $\{X'_n\}_{n=0}^\infty$ with the same $\{U_n\}$ sequence (different initial starting conditions are still allowed), $|X_n - X'_n|$ converges monotonically to zero as $n \rightarrow \infty$ and the two sample paths contract to zero.

2.2. Splitting, Drift, and Minorization

Moving to splitting techniques, suppose that a state set A exists such that for any two starting points $x, y \in A$, $P(x, B) = P(y, B)$ for all Borel measurable sets B . Then A is called a proper atom of the chain. Return times to A form a renewal process and much of the analysis for discrete state chains applies to chains on $\mathbb{R}$ with an atom.

Unfortunately, many chains $\{X_n\}_{n=0}^\infty$ on continuous state spaces do not have atoms. This said, one can construct a new chain $\{X'_n\}_{n=0}^\infty$ such that (i) $\{X'_n\}_{n=0}^\infty$ exhibits a regenerative structure, and (ii) the marginal distribution of X'_n is the same as that of X_n for each $n \geq 1$.

Atom construction can be accomplished using a so-called minorization condition (see [24], [22], and many others). A minorization condition supposes that one can find a $\delta > 0$, a set of states C , and a

probability measure ν with

$$P(x, A) \geq \delta 1_C(x) \nu(A) \quad (2.5)$$

for all sets A . In [24], the chain's atom is referred to as $(\delta 1_C, \nu)$ and $P - \delta 1_C \nu$ is a legitimate non-negative kernel.

We will often construct a sequence $\{Y_n\}_{n=0}^\infty$ of Bernoulli trials from $\{X_n\}_{n=0}^\infty$. These "coin flips" are usually made at times when the chain lies in C . More specifically, suppose that $X_0 = x_0$. If $x_0 \notin C$, the coin is not flipped (or the coin flip has probability zero of heads) and X_1 follows the transition probability $P(x_0, \cdot)$. If $x_0 \in C$, we flip a coin with heads probability δ . If the toss results in heads, set $Y_0 = 1$ and choose X'_1 according to ν . If the coin toss is tails, we set $Y_0 = 0$ and generate X'_1 from the probability kernel $[P(x, A) - \delta \nu(A)] / (1 - \delta)$. An analogous coin flip is made at all times $n \geq 1$, depending whether or not $X_n \in C$. This yields a $Y_n \in \{0, 1\}$ for all $n \geq 1$ (if $X_n \notin C$, set $Y_n = 0$). It is a fundamental fact that this construction leaves X'_n with the same marginal distribution as X_n for each $n \geq 1$ [24, 22].

Many convergence analyses exploit drift to some state set C . Quantifying this, suppose that $V(\cdot)$ is a measurable function. Define

$$PV(x) = \int_{\mathbb{R}} V(y) P(x, dy) = \mathbf{E}_x[V(X_1)]. \quad (2.6)$$

The chain $\{X_n\}_{n=0}^\infty$ is said to drift to the set C if there exists a V with $V(x) \geq 1$ for all x and constants $r_0 > 1$ and $0 \leq b < \infty$ satisfying

$$PV(x) \leq r_0^{-1} V(x) + b 1_C(x) \quad (2.7)$$

for all x . The following fundamental result was proven in [25].

Theorem 2.1. *Consider a ϕ -irreducible, aperiodic Markov chain $\{X_n\}_{n=0}^\infty$ on $\mathbb{R}$ having the stationary distribution $\pi(\cdot)$. Suppose that (2.5) is satisfied for some $C \subset \mathbb{R}$, $\delta > 0$, and probability measure $\nu(\cdot)$. Suppose further that the drift condition (2.7) holds for some constant $r_0 > 1$, $0 \leq b < \infty$, and a drift function $V(x) \geq 1$ with $V(x) < \infty$ for at least one $x \in \mathbb{R}$. Then the chain is geometrically ergodic in the sense of (1.1) for some $r > 1$ and finite $M(x_0)$.*

Although Theorem 2.1 tells us that the chain is geometrically ergodic, it does not tell us how to find specific geometric convergence rates and first constants. Indeed, this is where the bulk of our work lies. In general, is not true that the chain will converge at the geometric rate r_0^{-1} .

3. Convergence Rates for Shadowing Chains

For chains $\{X_n\}_{n=0}^\infty$ and $\{X'_n\}_{n=0}^\infty$ having the same transition kernel P , our construction produces a coupling time T such that $X_n = X'_n$ for all $n > T$. By the last section, geometric convergence will be obtained out to the radius of convergence of $\mathbf{E}_{x_0, X'_0}[r^T]$ (in r).

Suppose that there exist sets $C, C' \in \mathcal{A}$, a $\delta > 0$, and a probability measure ν on $(\mathbb{R}, \mathcal{A})$ for which the minorizations

$$P(x, A) \geq \delta 1_C(x) \nu(A), \quad P(x, A) \geq \delta 1_{C'}(x) \nu(A) \quad (3.1)$$

hold for all $x \in \mathbb{R}$ and $A \in \mathcal{A}$. In most applications, ν is concentrated on C or C' , but this need not be the case. Our construction below makes use of the transition kernels Q and Q' defined by

$$Q(x, A) = \frac{P(x, A) - \delta 1_C(x) \nu(A)}{1 - \delta 1_C(x)}, \quad Q'(x, A) = \frac{P(x, A) - \delta 1_{C'}(x) \nu(A)}{1 - \delta 1_{C'}(x)}. \quad (3.2)$$

3.1. The coupling

Our coupling scheme is similar to that in [24]. In particular, at each n where $X_n \in C$ and $X'_n \in C'$, we toss a coin (independently of all else) with success probability δ . Should this coin toss result in a success ($Y_n = 1$), the $(n + 1)$ st values of the primed and unprimed chains are generated as the same draw from ν . This forces $X_{n+1} = X'_{n+1}$, but does not change the marginal distribution of X_{n+1} or X'_{n+1} . After this meeting time T , all future transitions for both chains are generated from the same g and $\{U_n\}$ in (2.4). This keeps $X_k = X'_k$ for all $k > T$, where T is the time of the first successful coin toss ($Y_n = 1$).

If the coin toss is a failure ($Y_n = 0$), then X_{n+1} is generated from Q and X'_{n+1} is generated from Q' . In this process, g and the same U_{n+1} are used to generate both Q and Q' in (3.2). This does not change the marginal distributions of X_{n+1} or X'_{n+1} . Future transitions are subsequently generated from P , using the same g and $\{U_n\}$ again, until the two chain trajectories reenter C and C' , respectively. By the shadowing property, this takes place when the unprimed chain enters C .

It may be helpful to think of a discrete state space chain on $\{0, 1, \dots\}$ with two transition probability rows, say i and j , having the same one-step-ahead distribution (these two states are termed lumpable). Should $X_n = i$ and $X'_n = j$, we simply move both the primed and unprimed chains to the same state generated by P at time $n + 1$, forcing them to agree. Again, this does not change marginal distributions. The chief nuance beyond lumpability is that two minorization conditions in (3.1) are invoked now since the row distributions of states i and j may not be identical.

We now need some facts about our coupling time $T := \inf\{n \geq 0 : Y_n = 1\}$. First, note that shadowing gives

$$T = \inf\{n \geq 0 : X_n \in C \cap X'_n \in C' \cap Y_n = 1\} = \inf\{n \geq 0 : X_n \in C \cap Y_n = 1\}.$$

To bound the generating function of T , note that $P - \delta 1_C \nu$ satisfies

$$[P - \delta 1_C \nu]^n(x, A) = \mathbb{P}_x(X_n \in A \cap T > n - 1). \quad (3.3)$$

It follows from (4.21) of [24] that

$$\mathbb{P}_{x_0, x'_0}(T = n) = \delta ([P - \delta 1_C \nu]^n 1_C)(x_0) \quad (3.4)$$

for $n \geq 0$, the “extra δ ” arising from the probability of a successful coin flip. For chains that shadow, we hence have

$$E_{x_0, x'_0}[r^T] \leq E_{x_0}[r^T] = \sum_{r=0}^{\infty} r^n \delta ([P - \delta 1_C \nu]^n 1_C)(x_0). \quad (3.5)$$

Note that for shadowing chains, we only have to “track” the $\{X_n\}$ chain to couple: any time $X_n \in C$, the shadowing property guarantees that $X'_n \in C'$. Phrased another way, for shadowing chains, the bound in (3.5) for $E_{x_0, x'_0}[r^T]$ is uniform in x'_0 .

3.2. Drift Relationships

If the coin toss heads probability is $\delta = 1$, then $\{X_n\}$ and $\{X'_n\}$ will couple at the first coin toss; however, when $\delta < 1$, this may not be the case. By requiring that $\{X_n\}$ drifts to C a bit quicker than in (2.7), some nice rates can be achieved. Furthering these ideas, consider a kernel $r[P - \delta 1_C \nu]$. In addition to (2.7), suppose that V also satisfies

$$V(x) \geq \delta 1_C(x) + r[P - \delta 1_C \nu]V(x) \quad (3.6)$$

for some $r \in (1, r_0]$. Then Theorem 3.1 (iii) of [24] states that $E_{x_0}[r^T] \leq V(x_0)$ for all x_0 . Translating into the terminology of [24], $h = V$, $g(x) = \delta 1_C(x)$, and $K = r(P - \delta 1_C \nu)$. By the definitions of V and g , g and h belong to Nummelin’s class $\mathcal{E}_+$, and h is a superharmonic function.

In our future examples, a useful drift function will have the form

$$V(x) = a|x|1_{C^c}(x) + \alpha, \quad (3.7)$$

where $\alpha \geq 1$ and $a > 0$. The following result provides conditions on α and a so that such a V will satisfy (3.6).

Lemma 3.1. *Suppose that (2.7) holds for a V of form (3.7), with ν being concentrated on C . Then $V(x) = \alpha$ for $x \in C$ and $\nu(V) = \alpha$. Suppose that $\alpha > (b + \delta)/(1 + \delta - r_0^{-1})$. Then for $r \leq \min(\kappa, r_0)$, where $\kappa = (\alpha - \delta)/(r_0^{-1}\alpha + b - \delta\alpha)$, $\mathbf{E}_{x_0}[r^T] \leq V(x_0)$.*

Proof. For $x \notin C$, $(\delta 1_{C^c} \nu)V(x) = 0$ and (2.7) implies that one can choose $r \in (1, r_0]$ to get (3.6). For $x \in C$, using $r \leq (\alpha - \delta)/(r_0^{-1}\alpha + b - \delta\alpha)$ gives

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

Thus, (3.6) is again satisfied. In either case, Theorem 3.1 (iii) of [24] gives the result. $\square$

We now give our primary convergence rate result for shadowing chains.

Theorem 3.2. *Let μ_{X_0} and $\mu_{X'_0}$ be the general initial measures of a shadowing Markov chain having the trajectories $\{X_n\}_{n=0}^\infty$ and $\{X'_n\}_{n=0}^\infty$. Assume that the drift and minorization relations in (2.6) and (3.1) hold and that r and V satisfy (3.6). Then if V is both μ_{X_0} and $\mu_{X'_0}$ integrable,*

$$\|X_n - X'_n\| \leq \left[\mu_{X_0}(V) + \mu_{X'_0}(V) \right] r^{-n}. \quad (3.8)$$

Proof. Since $X_n = X'_n$ for all $n > T$, the coupling inequality gives

$$|\mathbb{P}_{X_0}(X_n \in A) - \mathbb{P}_{X'_0}(X'_n \in A)| \leq \mathbb{P}_{X_0, X'_0}(T > n). \quad (3.9)$$

Then for an r that satisfies Lemma 3.1, Markov's inequality gives

$$\mathbb{P}_{X_0, X'_0}(T > n) = \mathbb{P}_{\mu_{X_0}}(r^T > r^n) \leq \mathbf{E}_{\mu_{X_0}}[r^T] r^{-n} \leq V(X_0) r^{-n}.$$

Conditioning on X_0 , applying shadowing and Lemma 1, and integrating over the distribution of X_0 gives

$$\|P^n(X_0, A) - \pi(A)\| \leq \mathbf{E}[V(X_0)] r^{-n}.$$

Now use the triangle inequality $\|X_n - X'_n\| \leq \|X_n - \pi\| + \|X'_n - \pi\|$ and argue similarly to bound $\|X'_n - \pi\|$ and complete the proof. $\square$

Now suppose that $\{X_n\}$ has the initial measure μ_{X_0} , $\{X'_n\}$ has the initial measure $\mu_{X'_0}$, and that the two chains shadow. Under the conditions of Lemma 1, Theorem 2 gives

$$\|X_n - X'_n\| \leq [\mu_{X_0}(V) + \mu_{X'_0}(V)] r^{-n}, \quad (3.10)$$

for r satisfying the Lemma. When μ_0 is the Dirac measure at the point x and μ'_0 is the stationary measure π , $\mu_{X_0}(V) = V(x)$ and $\mu_{X'_0}(V) = \pi(V)$. Integrating both sides of (2.7) with respect to π gives

$$\pi(V) = \int_{-\infty}^{\infty} V(z) \pi(dz) \leq \frac{b}{1 - r_0^{-1}} \quad (3.11)$$

(both sides of the equation can be shown to be finite — see [22]). Thus,

$$\mu_{X_0}(V) + \mu_{X'_0}(V) = V(x) + \pi(V) \leq V(x) + \frac{b}{1 - r_0^{-1}}, \quad (3.12)$$

giving a “clean” first-constant to the geometric rate. Inserting (3.12) into (3.10) gives the following result.

Theorem 3.3. *Suppose that $\{X_n\}_{n=0}^\infty$ is a shadowing ϕ -irreducible, aperiodic Markov chain with the stationary distribution π satisfying the conditions in Theorem 2. Then*

$$\|P^n(x_0, \cdot) - \pi(\cdot)\| \leq \left[V(x_0) + \frac{b}{1 - r_0^{-1}} \right] r^{-n}.$$

Moving to weak shadowing, a coupling works by running the two processes independently until the stopping time τ ; thereafter, we couple as in beginning of this section. This coupling occurs at the time $T^* = \inf\{n \geq \tau : Y_n = 1\}$.

Theorem 3.4. *Consider the setup of Theorem 3. Suppose that the initial distributions of the chains are μ_{X_0} and $\mu_{X'_0}$ and that both $r^\tau V(X_\tau)$ and $r^\tau V(X'_\tau)$ are integrable. Then*

$$\|X_n - X'_n\| \leq r^{-n} \mathbf{E}_{X_0, X'_0} \left[r^\tau (V(X_\tau) + V(X'_\tau)) \right].$$

Proof. The coupling inequality is $\|X_n - X'_n\| \leq \mathbb{P}_{X_0, X'_0}(T^* > n)$ and Markov’s inequality applied to r^{T^*} provides $\mathbb{P}_{X_0, X'_0}(T^* > n) \leq r^{-n} \mathbf{E}_{X_0, X'_0}[r^{T^*}]$. Conditioning on (X_τ, X'_τ) , using the strong Markov property, and applying Lemma 1 gives the result. $\square$

Lemma 3.5. *Suppose that $\{X_n\}_{n=0}^\infty$ is a chain satisfying the drift condition in (2.7). Then*

$$P^m V(x) \leq r_0^{-m} V(x) + b(1 - r_0^{-m})/(1 - r_0^{-1}) \quad (3.13)$$

for all $m \geq 1$.

Proof. Use (2.7) for $m = 1$ and inductively apply the inequality. $\square$

Our next task is to bound $\mathbf{E}_{x_0, \pi}[r^\tau (V(X_\tau) + V(X'_\tau))]$. For this, let $A_{x_0, k} = \{z : \tau = k \text{ when } X_0 = x_0 \text{ and } X'_0 = z\}$. Then

$$\begin{aligned} \mathbf{E}_{x_0, \pi}[r^\tau (V(X_\tau) + V(X'_\tau))] &= \int_{-\infty}^{\infty} \mathbf{E}_{x_0, z}[r^\tau (V(X_\tau) + V(X'_\tau))] \pi(dz) \\ &= \sum_{k=0}^{\infty} \int_{A_{x_0, k}} \mathbf{E}_{x_0, z}[r^\tau V(X_\tau)] \pi(dz) \\ &\quad + \sum_{k=0}^{\infty} \int_{A_{x_0, k}} \mathbf{E}_{x_0, z}[r^\tau V(X'_\tau)] \pi(dz). \end{aligned}$$

Simplifying further and applying (3.13) gives

$$\begin{aligned}
 E_{x_0, \pi}[r^\tau (V(X_\tau) + V(X'_\tau))] &\leq \sum_{k=0}^{\infty} r^k P^k V(x_0) \pi(A_{x_0, k}) + \sum_{k=0}^{\infty} \int_{A_{x_0, k}} r^k P^k V(z) \pi(dz) \\
 &\leq \sum_{k=0}^{\infty} \left(V(x_0) + b \frac{r_0^k - 1}{1 - r_0^{-1}} \right) \pi(A_{x_0, k}) \\
 &\quad + \sum_{k=0}^{\infty} \int_{A_{x_0, k}} \left(V(z) + b \frac{r_0^{-k} - 1}{1 - r_0^{-1}} \right) \pi(dz) \\
 &\leq V(x_0) + \int_{-\infty}^{\infty} V(z) \pi(dz) + 2b \frac{E_{x_0, \pi}[r_0^\tau] - 1}{1 - r_0^{-1}} \\
 &\leq V(x_0) + 2 \frac{b}{1 - r_0^{-1}} E_{x_0, \pi}[r_0^\tau]
 \end{aligned}$$

after the strong Markov property and our bound for $\pi(V)$ in (3.11) are applied.

Let $M(x) = V(x) + 2b/(1 - r_0^{-1})E_{x, \pi}[r_0^\tau]$. If $E_{x, \pi}[r_0^\tau] < \infty$, then $M(x) < \infty$ and we arrive at the following result.

Corollary 3.1. *Suppose that $\{X_n\}_{n=0}^\infty$ is a ϕ -irreducible, aperiodic Markov chain with the stationary distribution $\pi(\cdot)$ satisfying (2.5) for some set C and $\delta \in (0, 1)$. If the weak shadowing condition holds, $r \in (1, r_0]$, and (2.7) and (3.6) apply, then the rate in (1.1) holds with rate r and the above $M(x_0)$.*

4. A Storage Chain Example

Our first example considers a discrete-time storage chain 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 n th input, I_n is added to the store's content. The I_n s are IID positive random variables, independent of the Poisson process. The release rate of the store is x when the store's content is x . Let X_n be the storage level just prior to the n th input, with x_0 being the initial level. 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_n s are uniform $[0, 1]$. 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_n s are uniform $[0, \beta]$ for some parameter $\beta > 0$. The cumulative transition distribution is then

$$\mathbb{P}_x(X_1 \leq y) = \mathbb{P}\left((x + I_0)U_1 \leq y\right) = \begin{cases} \frac{(y-x)^+}{\beta} + \frac{y}{\beta} \ln\left(\frac{x+\beta}{\max\{x, y\}}\right), & x + \beta \leq y \\ 1, & x + \beta > y \end{cases}, \quad (4.1)$$

where $x^+ = \max(x, 0)$.

Differentiating (4.1) yields the transition density

$$P(x, dy) = \begin{cases} \frac{1}{\beta} \ln\left(\frac{x+\beta}{\max\{x, y\}}\right) dy, & x + \beta \leq y \\ 0, & x + \beta > y \end{cases}. \quad (4.2)$$

We leave it to the reader to show that the chain is ϕ -irreducible, where ϕ is the Lebesgue measure. Below, it is shown that this chain satisfies a drift condition. From this, it follows that $\{X_n\}_{n=0}^\infty$ is Harris positive recurrent [24].

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

$$E[e^{tX_\infty}] := \lim_{n \rightarrow \infty} E[e^{tX_n}] = \exp \left\{ \int_0^t \frac{e^{\zeta\beta} - 1 - \zeta\beta}{\zeta^2\beta} d\zeta \right\}.$$

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. To see this, let $y = x + 2\beta$. Then if $dy \in [x + 2\beta - 0.01, x + 2\beta + 0.01]$, $P(x, dy) = 0$ and $P(y, dx) > 0$, implying that $\pi(dx)P(x, dy) \neq \pi(dy)P(y, dx)$. 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 [20] do not apply.

To identify minorization and drift conditions for the chain, our first step is to find a probability measure ν and $\delta > 0$ such that $P(x, A) \geq \delta 1_C(x)\nu(A)$ for $x \in \mathbb{R}$ and all sets A for some set C of form $C = [0, c]$, where $c > 0$. Our δ and ν will depend on c .

Lemma 4.1. *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

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

for all $y > 0$ and $x \in C$. We do this by cases. First, suppose that $y_0 < x$. Our work is partitioned into four further 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 $1_{(y_0, \beta]}(y) = 0$, (4.3) holds. In our second subcase, which is when $y_0 < y \leq x$, the indicator $1_{(0, y_0]}(y)$ is zero and the choice of y_0 gives

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

implying (4.3) again. Our third subcase moves to $y \in (x, \beta]$. Then (4.2) gives

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

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 $y_0 \geq x$. Again, we partition our work into four subcases. When $y \leq x$, (4.2) gives

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

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

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

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

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

implying (4.3) again. Finally, when $y > \beta$, (4.3) holds with a right hand side of zero. $\square$

Corollary 4.1. *Under the assumptions in Lemma (4.1), $y_0 \rightarrow \beta$ and $\delta \rightarrow 0$ as $c \rightarrow \infty$.*

Proof. The limit for y_0 follows from its definition in the previous Lemma. For the result on δ , to make ν a probability measure, we must have

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

from which the limit claim about δ follows. $\square$

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

Lemma 4.2. *If $x > c$, then*

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

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

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

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

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

establishing the first claim. For the case where $x < c$, use (4.2) to get

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

Integration by parts on the rightmost integral and algebraic simplifications now establish the second claim. $\square$

We will need an analysis of the function h_c defined by

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

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

Lemma 4.3. *Define $K(c) = \sup_{x>c} h_c(x)$. Then*

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

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

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

The above give

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

□

Corollary 4.2. *Under the assumptions of Lemma 4.3, $K(c) < 1$ for all $c > \beta/6$.*

Proof. When $\beta < 6c$, Lemma 4.3 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 4.4. *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 4.2 and the definition of $h_c(x)$ give

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

To identify a drift condition, use Lemma 4.3 to get

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

For $x \leq c$, $V_c(x) = \alpha$ and Lemma 4.2 give $PV_c(x) = h(x) = h(x) - \alpha r_0^{-1} + \alpha r_0^{-1}$. 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. □

By the definition of r_0 , if α is fixed, $\lim_{c \rightarrow \infty} r_0 = 2$. Unfortunately, if c is fixed, $\lim_{\alpha \rightarrow \infty} r_0 = 1$. We will need to balance these quantities to get good rates. We now numerically illustrate our convergence rates when $x_0 = 10$. Here, $C = [0, c]$, $V_c(x) = \alpha + 3x1_{(c, \infty)}(x)$, and $M(x) = V_c(x) + b/(1 - r_0^{-1})$. For a fixed β , we want to select a c that gives good convergence rates. While this choice is not optimized, we consider $c = \beta\sqrt{3}$ and $\alpha \in (1, 200]$ for the β values in $\{1, 3, 6, 16\}$. Our choices of δ, c , and α satisfy $\delta\alpha > b$ and $v(V) = \alpha$. Hence, Lemma 3.1 and Theorem 3.3 give geometric convergence at the rate $r = \max\{1, \min\{(\alpha - \delta)/(r_0^{-1}\alpha + b - \delta\alpha), r_0\}\}$.

Table displays our geometric convergence rates and first-constants for different β . These are better than the convergence rates in [27], which are labeled as “RT convergence rates” and are provided for comparison’s sake. The first constants are always reasonable. Our rates increase as β increases.

	$\beta = 1$	$\beta = 3$	$\beta = 6$	$\beta = 16$
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
Our Convergence Rate	1.298798	1.315168	1.319311	1.321902
RT Convergence Rate	1.169257	1.127226	1.107116	1.086219
$M(x)$	39.01684	92.17664	172.1353	438.6049
$M(x)r^{-1000}$	2.1581e-112	9.8407e-118	5.6552e-119	1.4677e-119

Table 1. A convergence rate comparison for our storage chain.

5. A First-order Autoregression

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

$$X_{n+1} = \varphi X_n + Z_{n+1}, \quad n \geq 0, \quad (5.1)$$

where $\varphi \in (-1, 1) \setminus \{0\}$ 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, two simplifying assumptions are made up front to inject tractability into the calculations. First, we work with a normally distributed process, which is achieved by positing that $\{Z_n\}$ is IID normal noise. This chain is reversible; hence, the results in [1] apply. Second, to scale the process, we take $\sigma^2 = 1$.

When $\varphi \in (0, 1)$ the chain is stochastically ordered in its initial state. Such a φ results in a positively correlated $\{X_n\}_{n=0}^\infty$. For $\varphi \in (-1, 0)$, the chain is no longer stochastically ordered. To handle this general case, we use a weak shadowing with shadowing sets $C = [-c, c]$ and $C' = [-2c, 2c]$ for some $c > 0$.

When $X_0 = x_0$, X_1 is normally distributed with mean φx_0 and unit variance:

$$P(x_0, dy) = \frac{1}{\sqrt{2\pi}} e^{-(y - \varphi x_0)^2/2} dy. \quad (5.2)$$

This chain is easily shown to be ϕ -irreducible, where ϕ is the Lebesgue measure. The stationary distribution of $\{X_n\}$ is normal with mean zero and variance $1/(1 - \varphi^2)$, which has a finite moment generating function of all orders: $E[e^{sX_\infty}] = e^{s^2/[2(1 - \varphi^2)]}$. Below, a drift condition is established for the chain. From this, it follows that $\{X_n\}_{n=0}^\infty$ is Harris positive recurrent [24].

Recurring (5.1) provides

$$X_n = \varphi^n X_0 + \sum_{j=0}^{n-1} \varphi^j Z_{n-j}, \quad (5.3)$$

showing that the chain is pathwise (and hence stochastically) ordered in its initial state when $\varphi \in [0, 1)$. 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 $|X_n - X'_n| \leq |\varphi|^n |x_0 - x'_0|$. We exploit this sample path contraction to invoke weak shadowing.

To find a probability measure ν and $\delta > 0$ satisfying (2.5), let $c > 0$ and $C = [-c, c]$.

Lemma 5.1. *Define*

$$\delta = \int_{-c}^0 \frac{1}{\sqrt{2\pi}} e^{-(y-|\varphi|c)^2/2} dy + \int_0^c \frac{1}{\sqrt{2\pi}} e^{-(y+|\varphi|c)^2/2} dy \quad (5.4)$$

and

$$\nu(dy) = \frac{1}{\delta} \left[1_{(-c,0)}(y) \frac{1}{\sqrt{2\pi}} e^{-(y-|\varphi|c)^2/2} dy + 1_{[0,c)}(y) \frac{1}{\sqrt{2\pi}} e^{-(y+|\varphi|c)^2/2} dy \right]. \quad (5.5)$$

Then, for $x \in C$, 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 $x \in C$,

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

which follows from $|\varphi x| \leq |\varphi|c$. Similar arguments handle the case where $y \geq 0$. $\square$

To establish a drift condition, define $V_c(x) = 2|x|1_{[c,\infty)}(|x|) + \alpha$, 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 $h_c(x) = \mathbf{E}_x[V_c(X_1)]/(2|x| + \alpha)$.

Lemma 5.2. *Define $K(c) = \sup_{|x|>c} h_c(x)$. Then*

$$K(c) \leq \frac{(1-|\varphi|)\alpha + 2\sqrt{2/\pi}}{2c + \alpha} + |\varphi|. \quad (5.6)$$

Proof. When $X_0 = x_0$ and $|X_1| \geq c$, $V_c(X_1) \leq 2(|\varphi||x_0| + |Z_1|) + \alpha$; when $|X_1| < c$, $V_c(X_1) \leq \alpha$. Hence, $V_c(X_1) \leq 2(|\varphi||x_0| + |Z_1|) + \alpha$. Standard normality gives $E[|Z_1|] = \sqrt{2/\pi}$, implying

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

from which (5.6) follows. $\square$

It is easy to see that $K(c) < 1$ for all $c > \sqrt{2/\pi}/(1-|\varphi|)$. Hence, there exists a $c > 1$ such that $K(c) < 1$.

Lemma 5.3. *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 the contraction parameter r_0^{-1} and constant b .*

Proof. The definition of $h_c(x)$ gives $\mathbf{E}_x[V_c(X_1)] = (2|x| + \alpha)h_c(x)$. To identify a drift condition, when $|x| > c$, use Lemma 5.2 to get

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

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

To obtain convergence rates, set $F_x(y) = F(x, y) = Q(x, (-\infty, y])$, where Q is as in (3.2). 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 $H_x(\cdot) = H(x, \cdot)$. Lemma 3.22 of [15] shows that if θ is a 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 and set $g(x) = H_x(u)$. Then g is differentiable and $F(x, g(x)) = u$. Taking a derivative with respect to x with 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 i th component of F . To generate X_1 and X'_1 , generate θ and if $\theta = u$, set $X_1 = H_{x_0}(u) = g(x_0)$ and $X'_1 = H_{x'_0}(u) = g(x'_0)$. Thus, when $|x_0 - x'_0| \leq c$,

$$|X_1 - X'_1| = |g(x_0) - g(x'_0)| = \int_{x'_0}^{x_0} |g'(t)| dt \leq |x_0 - x'_0| \leq c$$

if we can show that $|g'(x)| \leq 1$. The following Lemma with $g'(x) = \varphi P(x, dy) / (P(x, dy) - \delta' v(dy))$ establishes this.

Lemma 5.4. For δ' with $0 < \delta' \leq \delta(1 - |\varphi|)$,

$$\frac{|\varphi| P(x, dy)}{P(x, dy) - \delta' v(dy)} \leq 1, \quad (5.7)$$

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

Proof. Since $\delta' \leq \delta(1 - |\varphi|)$, Lemma 5.1 provides

$$P(x, dy) \geq \delta v(dy) \geq \frac{\delta'}{1 - |\varphi|} v(dy),$$

proving (5.7). $\square$

Next, we bound $M(x)$ to get our first-constant. Define

$$\tau = \inf\{n \geq 0 : |X_n - X'_n| \leq c\}. \quad (5.8)$$

Given $X_0 = x$ and $X'_0 = z$ with $x > z$, there exist a τ depending on x and z such that $|X_{1+\tau} - X'_{1+\tau}| \leq |\varphi|^\tau |x - z|$; thus, if $|\varphi|^\tau |x - z| \leq c$, Lemma 5.4 implies that for all $n > \tau$, if $X_n \in C$, then $X'_n \in C'$. This is our weak shadowing. For AR(1) chains, $T \leq \tau + T_C$. Corollary 3.1 implies that we need to bound $\int_{-\infty}^{\infty} r_0^\tau \pi(dz)$ to get good convergence rates.

To bound $\int_{-\infty}^{\infty} r_0^\tau \pi(dz)$, (5.8) and $0 < |\varphi| < 1$ provide,

$$\tau \leq \frac{\ln(c) - \ln(|x - z|)}{\ln(|\varphi|)} = \frac{\ln(c)}{\ln(|\varphi|)} + \frac{-1}{\ln(|\varphi|)} (\ln(|x - z|)) \leq \frac{\ln(c)}{\ln(|\varphi|)} + \frac{-1}{\ln|\varphi|} |x - z|. \quad (5.9)$$

The last inequality follows from $\ln(|x - z|) \leq |x - z|$ and $|\varphi| < 1$. Equation (5.9) gives

$$r_0^\tau \leq r_0^{\frac{\ln(c)}{\ln(|\varphi|)} + \frac{-1}{\ln(|\varphi|)} |x - z|} = r_0^{\frac{\ln(c)}{\ln(|\varphi|)}} r_0^{\frac{-1}{\ln(|\varphi|)} |x - z|}. \quad (5.10)$$

	$\varphi = 0.1$	$\varphi = 0.25$	$\varphi = 0.5$	$\varphi = 0.75$
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	2.224894	2.809701	2.891592	2.69765
Our Convergence Rate	2.564565	1.551759	1.144701	1.030529
Bax Convergence 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 2. A convergence rate comparison for AR(1) chains..

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

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

Applying this and splitting the absolute value into cases gives

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

where the last inequality follows from $\Psi_\pi(s) := E[e^{sX_\infty}] = e^{s^2/(2(1-|\varphi|^2))}$, and the facts that $\Psi_\pi(\ln(r_0)/\ln(|\varphi|))$ and $\Psi_\pi(-\ln(r_0)/\ln(|\varphi|))$ are finite. Hence, $E_{x,\pi}[r_0^\tau] < \infty$ and $M(x_0) < \infty$ for all x_0 . Corollary 3.1 now gives our convergence rates.

For each $\varphi \in \{0.1, 0.25, 0.5, 0.75\}$, we take $x_0 = 10$ and select a c to yield the good convergence rate $r_0 = 1/K(c)$. For numerical purposes, $c \in (1.5, 4]$ and $\alpha \in (2, 20]$ values are chosen for each φ . Our choices of δ , c , and α imply that $\delta\alpha > b$ and $\nu(V) = \alpha$. Lemma 3.1 and Corollary 3.1 give the geometric rate $r = \max\{1, \min\{(\alpha - \delta)/(r_0^{-1}\alpha + b - \delta\alpha), r_0\}\}$. Table 2 numerically displays our convergence results. Here, we compare to convergence rates in [1], which are labeled as ‘‘Bax Rates’’. Our convergence rates are often better than those in [1], although not strikingly so. For AR(1) convergence rates obtained by completely different methods, see [32].

6. Discussion

This paper derived some ‘‘clean’’ total variational geometric convergence rate bounds for Markov chains satisfying a so-called shadowing property. The results were applied to storage chains and a first order autoregressive chain. While we do not know of ways to get better rates than those obtained here for our storage chain, alternative techniques exist to get first-order autoregression rates; see [32] for example. Another simple way to obtain first-order autoregressive chain rates proceeds from first principles, using $X_n = X'_n + \varphi^n(X_0 - X'_0)$ and a total variational distance between two univariate Gaussian variables when the same Gaussian errors are used to drive both chains.

Additional research is also needed. While we did not do so here, a broader definition of shadowing would allow for the possibility of $X_n \in C'$ implying that $X'_n \in C$. The same minorizing issues could then be invoked. Presumably, one would arrive at improved convergence rates, especially in cases where C and C' are not proper subsets of one and other. Extension to chains on $\mathbb{R}^k$ also merit consideration.

Acknowledgements

Fun Choi John Chan and Robert Lund acknowledge support from National Science Foundation Grant DMS 2113592.

References

- [1] Peter H. Baxendale, *Renewal theory and computable convergence rates for geometrically ergodic Markov chains*, The Annals of Applied Probability **15** (2005), no. 1B, 700–738.
- [2] R Douc, E Moulines, and J. Rosenthal, *Quantitative convergence rates for inhomogeneous Markov chains*, Electronic Communications in Probability **7** (2002), 123–128.
- [3] Randal Douc, Eric Moulines, Pierre Priouret, and Philippe Soulier, *Markov chains*, Springer, 2018.
- [4] Andrew Gelman, Frederic Bois, and Jiming Jiang, *Physiological pharmacokinetic analysis using population modeling and informative prior distributions*, Journal of the American Statistical Association **91** (1996), no. 436, 1400–1412.
- [5] Andrew Gelman and Donald B Rubin, *Inference from iterative simulation using multiple sequences*, Statistical Science **7** (1992), no. 4, 457–472.
- [6] ———, *A single sequence from the Gibbs sampler gives a false sense of security*, Bayesian Statistics **4** (1992), 625–631.
- [7] Charles J Geyer, *Practical Markov chain Monte Carlo*, Statistical Science **7** (1992), 473–483.
- [8] Walter R Gilks, Sylvia Richardson, and David Spiegelhalter, *Markov chain Monte Carlo in practice*, CRC Press, 1995.
- [9] Søren Fiig Jarner and Ernst Hansen, *Geometric ergodicity of Metropolis algorithms*, Stochastic Processes and their Applications **85** (2000), no. 2, 341–361.
- [10] Daniel C Jerison, *Quantitative convergence rates for reversible Markov chains via strong random times*, ArXiv Preprint ArXiv:1908.06459 (2019).
- [11] Zhumengmeng Jin and James P Hobert, *Dimension free convergence rates for Gibbs samplers for Bayesian linear mixed models*, Stochastic Processes and their Applications **148** (2022), 25–67.
- [12] Leif T Johnson and Charles J Geyer, *Variable transformation to obtain geometric ergodicity in the random-walk Metropolis algorithm*, The Annals of Statistics **40** (2012), 3050–3076.
- [13] Galin L. Jones and James P. Hobert, *Honest exploration of intractable probability distributions via Markov chain Monte Carlo*, Statistical Science **16** (2001), 312–334.
- [14] Galin L Jones and James P Hobert, *Sufficient burn-in for Gibbs samplers for a hierarchical random effects model*, The Annals of Statistics **32** (2004), no. 2, 784–817.
- [15] Olav Kallenberg, *Foundations of modern probability*, Springer, 2002.
- [16] Masaaki Kijima, *Markov processes for stochastic modeling*, CRC Press, 1997.

- [17] Torngny Lindvall, *Lectures on the coupling method*, Courier Corporation, 2002.
- [18] 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** (1995), no. 1, 157–169.
- [19] 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** (1996), no. 1, 218–237.
- [20] R. B. Lund and Richard L. Tweedie, *Geometric convergence rates for stochastically ordered Markov chains*, Mathematics of Operations Research **21** (1996), no. 1, 182–194.
- [21] Robert Lund, Ying Zhao, and Peter C Kiessler, *A monotonicity in reversible Markov chains*, Journal of Applied Probability **43** (2006), no. 2, 486–499.
- [22] S. P. Meyn and R. L. Tweedie, *Markov chains and stochastic stability*, Springer, 1993.
- [23] ———, *Computable bounds for geometric convergence rates of Markov chains*, The Annals of Applied Probability **4** (1994), no. 4, 981–1011.
- [24] Esa Nummelin, *General irreducible Markov chains and non-negative operators*, Cambridge University Press, 1984.
- [25] Gareth O. Roberts and Jeffrey S. Rosenthal, *General state space Markov chains and MCMC algorithms*, Probability Surveys **1** (2004), 20–71.
- [26] Gareth O Roberts and Richard L Tweedie, *Bounds on regeneration times and convergence rates for Markov chains*, Stochastic Processes and their Applications **80** (1999), no. 2, 211–229.
- [27] Gareth O. Roberts and Richard L. Tweedie, *Rates of convergence of stochastically monotone and continuous time Markov models*, Journal of Applied Probability **37** (2000), no. 2, 359–373.
- [28] Jeffrey S. Rosenthal, *Minorization conditions and convergence rates for Markov chain Monte Carlo*, Journal of the American Statistical Association **90** (1995), no. 430, 558–566.
- [29] Jeffrey S Rosenthal, *Rates of convergence for Gibbs sampling for variance component models*, The Annals of Statistics **23** (1995), no. 3, 740–761.
- [30] Jeffrey S. Rosenthal, *Quantitative convergence rates of Markov chains: A simple account*, Electronic Communications in Probability **7** (2002), 123–128.
- [31] Jeffrey S Rosenthal, *Asymptotic variance and convergence rates of nearly-periodic Markov chain Monte Carlo algorithms*, Journal of the American Statistical Association **98** (2003), no. 461, 169–177.
- [32] Sabrina Sixta and Jeffrey S Rosenthal, *Convergence rate bounds for iterative random functions using one-shot coupling*, Statistics and Computing **32** (2021), 71–.
- [33] Dietrich Stoyan, Wilfrid S Kendall, Sung Nok Chiu, and Joseph Mecke, *Stochastic geometry and its applications*, John Wiley & Sons, 2013.
- [34] Volker Strassen, *The existence of probability measures with given marginals*, The Annals of Mathematical Statistics **36** (1965), no. 2, 423–439.