

Discussion Paper Series

RIEB

Kobe University

DP2024-32

**Quantitative Convergence Rates for
Stochastically Monotone Markov Chains**

**Takashi KAMIHIGASHI
John STACHURSKI**

October 3, 2024



Research Institute for Economics and Business Administration

Kobe University

2-1 Rokkodai, Nada, Kobe 657-8501 JAPAN

QUANTITATIVE CONVERGENCE RATES FOR STOCHASTICALLY MONOTONE MARKOV CHAINS

TAKASHI KAMIHIGASHI AND JOHN STACHURSKI

ABSTRACT. For Markov chains and Markov processes exhibiting a form of stochastic monotonicity (larger states shift up transition probabilities in terms of stochastic dominance), stability and ergodicity results can be obtained using order-theoretic mixing conditions. We complement these results by providing quantitative bounds on deviations between distributions. We also show that well-known total variation bounds can be recovered as a special case.

1. INTRODUCTION

Quantitative bounds on the distance between distributions generated by Markov chains have many applications in statistics and the natural and social sciences (see, e.g., [22, 18]). One approach uses total variation distance and exploits minorization conditions (see, e.g., [21, 7, 12, 2]). Another branch of the literature bounds deviations using Wasserstein distance [6, 19, 20]. These bounds require some form of uniform continuity with respect to a metric on the state space.

In some applications, Markov chains lack both the minorization and continuity properties discussed above, making total variation and Wasserstein-type bounds difficult or impossible to apply. Fortunately, some of these models also possess valuable structure in the form of stochastic monotonicity. Such monotonicity can be exploited to obtain stability and ergodicity via order-theoretic versions of mixing conditions [5, 3, 9, 4, 13, 8, 14]. In this paper we complement these stability and ergodicity results by providing a theorem on quantitative bounds for stochastically monotone Markov chains.

There already exist several results that use stochastic monotonicity to bound the distributions generated by Markov chains [17, 10]. However, these bounds are typically stated in terms of total variation distance, which again requires traditional minorization conditions (as opposed to the order-theoretic mixing conditions discussed in the last paragraph). In this paper, we aim to fully exploit the monotonicity by instead bounding Kolmogorov distance between distributions. This works well because Kolmogorov distance respects order structure on the state space.

Date: October 2, 2024.

Our main theorem is closely related to the total variation bound in Theorem 1 of [21], which is representative of existing work on total variation bounds and supplies a simple and elegant proof. The main differences between that theorem and the one presented below is that we use Kolmogorov distance instead of total variation distance and an order-theoretic mixing condition instead of a standard minorization condition. At the same time, it is possible to recover Theorem 1 of [21] from the result we present below by a particular choice of partial order (see Sections 4.1 and 4.4). Thus our work can be viewed as a generalization of existing total variation results.

2. SET UP

In this section we recall basic definitions and state some preliminary results.

2.1. Environment. Throughout this paper, \mathbb{X} is a Polish space, \mathcal{B} is its Borel sets, and \preceq is a closed partial order on \mathbb{X} . The last statement means that the graph of \preceq , denoted by

$$\mathbb{G} := \{(x', x) \in \mathbb{X} \times \mathbb{X} : x' \preceq x\},$$

is closed under the product topology on $\mathbb{X} \times \mathbb{X}$. A map $h: \mathbb{X} \rightarrow \mathbb{R}$ is called *increasing* if $x \preceq x'$ implies $h(x) \leq h(x')$. We take $p\mathcal{B}$ to be the set of all probability measures on \mathcal{B} and let $b\mathcal{B}$ be the bounded Borel measurable functions sending \mathbb{X} into \mathbb{R} . The symbol $ib\mathcal{B}$ represents all increasing $h \in b\mathcal{B}$.

Given μ, ν in $p\mathcal{B}$, we say that μ is *stochastically dominated* by ν and write $\mu \preceq_s \nu$ if $\mu(h) \leq \nu(h)$ for all $h \in ib\mathcal{B}$. In addition, we set

$$\kappa(\mu, \nu) := \sup \left\{ \left| \int h d\mu - \int h d\nu \right| : h \in ib\mathcal{B} \text{ and } 0 \leq h \leq 1 \right\}, \quad (1)$$

which corresponds to the Kolmogorov metric on $p\mathcal{B}$ [14, 11].

A function $Q: (\mathbb{X}, \mathcal{B}) \rightarrow \mathbb{R}$ is called a *stochastic kernel* on $(\mathbb{X}, \mathcal{B})$ if Q is a map from $\mathbb{X} \times \mathcal{B}$ to $[0, 1]$ such that that $x \mapsto Q(x, A)$ is measurable for each $A \in \mathcal{B}$ and $A \mapsto Q(x, A)$ is a probability measure on \mathcal{B} for each $x \in \mathbb{X}$. At times we use the symbol Q_x to represent the distribution $Q(x, \cdot)$ at given x . A stochastic kernel Q on $(\mathbb{X}, \mathcal{B})$ is called *increasing* if $Qh \in ib\mathcal{B}$ whenever $h \in ib\mathcal{B}$.

For a given stochastic kernel Q on $(\mathbb{X}, \mathcal{B})$, we define the *left and right Markov operators* generated by Q via

$$\mu Q(A) := \int Q(x, A) \mu(dx) \quad \text{and} \quad Qf(A) := \int f(y) Q(x, dy).$$

(The left Markov operator $\mu \mapsto \mu Q$ maps $p\mathcal{B}$ to itself, while the right Markov operator $f \mapsto Qf$ acts on bounded measurable functions.) A discrete-time \mathbb{X} -valued stochastic process $(X_t)_{t \geq 0}$ on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t \geq 0})$ is called *Markov-(Q, μ)* if $X_0 \stackrel{d}{=} \mu$ and

$$\mathbb{E}[h(X_{t+1}) \mid \mathcal{F}_t] = Qh(X_t) \text{ with probability one for all } t \geq 0 \text{ and } h \in b\mathcal{B}.$$

2.2. Couplings. A *coupling* of $(\mu, \nu) \in p\mathcal{B} \times p\mathcal{B}$ is a probability measure ρ on $\mathcal{B} \otimes \mathcal{B}$ satisfying $\rho(A \times \mathbb{X}) = \mu(A)$ and $\rho(\mathbb{X} \times A) = \nu(A)$ for all $A \in \mathcal{B}$. Let $\mathcal{C}(\mu, \nu)$ denote the set of all couplings of (μ, ν) and let

$$\alpha(\mu, \nu) := \sup_{\rho \in \mathcal{C}(\mu, \nu)} \rho(\mathbb{G}) \quad ((\mu, \nu) \in p\mathcal{B} \times p\mathcal{B}). \quad (2)$$

The value $\alpha(\mu, \nu)$ lies in $[0, 1]$ and can be understood as a measure of “partial stochastic dominance” of ν over μ [15]. By the Polish assumption and Strassen’s theorem [23, 16] we have

$$\alpha(\mu, \nu) = 1 \quad \text{whenever} \quad \mu \preceq_s \nu. \quad (3)$$

Let Q be a stochastic kernel on $(\mathbb{X}, \mathcal{B})$ and let \hat{Q} be a stochastic kernel on $(\mathbb{X} \times \mathbb{X}, \mathcal{B} \otimes \mathcal{B})$. We call \hat{Q} a *Markov coupling* of Q if $\hat{Q}_{(x, x')}$ is a coupling of Q_x and $Q_{x'}$ for all $x, x' \in \mathbb{X}$. We call \hat{Q} a *\preceq -maximal Markov coupling* of Q if \hat{Q} is a Markov coupling of Q and, in addition,

$$\hat{Q}((x, x'), \mathbb{G}) = \alpha(Q_x, Q_{x'}) \quad \text{for all } (x, x') \in \mathbb{X} \times \mathbb{X}. \quad (4)$$

Lemma 2.1. *For any stochastic kernel Q on $(\mathbb{X}, \mathcal{B})$, there exists a \preceq -maximal Markov coupling of Q .*

Proof. By Theorem 1.1 of [24], given lower semicontinuous $\varphi: \mathbb{X} \times \mathbb{X} \rightarrow \mathbb{R}$, there exists a stochastic kernel \hat{Q} on $(\mathbb{X} \times \mathbb{X}, \mathcal{B} \otimes \mathcal{B})$ such that \hat{Q} is a Markov coupling of Q and, in addition

$$(\hat{Q}\varphi)(x, x') = \inf \left\{ \int \varphi \, d\rho : \rho \in \mathcal{C}(Q_x, Q_{x'}) \right\}.$$

As \mathbb{G} is closed, this equality is attained when $\varphi = 1 - \mathbb{1}_{\mathbb{G}}$. Since $\hat{Q}_{(x, x')}$ and ρ are probability measures, we then have

$$\hat{Q}((x, x'), \mathbb{G}) = \sup \{ \rho(\mathbb{G}) : \rho \in \mathcal{C}(Q_x, Q_{x'}) \}.$$

Thus, \hat{Q} is a \preceq -maximal Markov coupling of Q . □

2.3. Drift. Consider the geometric drift condition

$$QV(x) \leq \lambda V(x) + \beta \quad \text{for all } x \in \mathbb{X}, \quad (5)$$

where Q is a stochastic kernel on $(\mathbb{X}, \mathcal{B})$, V is a measurable function from \mathbb{X} to $[1, \infty)$, and λ and β are nonnegative constants. We fix $d \geq 1$ and set

$$\gamma := \lambda + \frac{2\beta}{d} \quad \text{and} \quad C := \{x \in \mathbb{X} : V(x) \leq d\}. \quad (6)$$

Fix μ, μ' in $p\mathcal{B}$ and set

$$H(\mu, \mu') := \frac{1}{2} \left[\int V d\mu + \int V d\mu' \right]. \quad (7)$$

Let \hat{Q} be a Markov coupling of Q and let $((X_t, X'_t))_{t \geq 0}$ be Markov- $(\hat{Q}, \mu \times \mu')$ on $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t \geq 0})$. We are interested in studying the number of visits to $C \times C$, as given by

$$N_t := \sum_{j=0}^t \mathbb{1}\{(X_j, X'_j) \in C \times C\}.$$

Lemma 2.2. *If Q satisfies the geometric drift condition (5), then, for all $t \in \mathbb{N}$ and all $j \in \mathbb{N}$ with $j \leq t$, we have*

$$\mathbb{P}\{N_t < j\} \leq \gamma^t d^{j-1} H(\mu, \mu').$$

The result in Lemma 2.2 has already been used in other sources. To make the paper more self-contained, we provide a proof in the appendix. Our proof is based on arguments in [21].

3. CONVERGENCE RATES

Let V be a measurable function from \mathbb{X} to $[1, \infty)$ and let Q be a stochastic kernel on $(\mathbb{X}, \mathcal{B})$ satisfying the geometric drift condition (5). Fix $d \in \mathbb{R}_+$ and let C and γ be as defined in (6). Let $H(\mu, \mu')$ be as given in (7). Let

$$\varepsilon := \inf \{\alpha(Q_x, Q_{x'}) : (x, x') \in C \times C\}. \quad (8)$$

We now state the main result.

Theorem 3.1. *If Q is increasing, then, for any $j, t \in \mathbb{N}$ with $j \leq t$, we have*

$$\kappa(\mu Q^t, \mu' Q^t) \leq (1 - \varepsilon)^j + \gamma^t d^{j-1} H(\mu, \mu').$$

Proof. Given Q in Theorem 3.1, we let \hat{Q} be a \preceq -maximal Markov coupling of Q (existence of which follows from Lemma 2.1). Let $((X_t, X'_t))_{t \geq 0}$ be Markov- $(\hat{Q}, \mu \times \mu')$ on $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t \geq 0})$. We observe that the graph \mathbb{G} of \preceq is absorbing

for \hat{Q} . Indeed, if $(x, x') \in \mathbb{G}$, then, since Q is increasing, $Q(x, \cdot) \preceq_s Q(x', \cdot)$. Hence, by (3), we have $\alpha(Q(x, \cdot), Q(x', \cdot)) = 1$. Applying (4) yields $\hat{Q}((x, x'), \mathbb{G}) = 1$.

Let τ be the stopping time $\tau := \inf\{t \geq 0 : X'_t \preceq X_t\}$ with $\inf \emptyset = \infty$. Since \mathbb{G} is absorbing for \hat{Q} , we have $\mathbb{P}\{X'_t \preceq X_t\} = 1$ whenever $t \geq \tau$. Let h be any element of $ib\mathcal{B}$ with $0 \leq h \leq 1$. Since $((X_t, X'_t))_{t \geq 0}$ is Markov- $(\hat{Q}, \mu \times \mu')$ and $\hat{Q}((x, x'), \cdot)$ is a coupling of $Q(x, \cdot)$ and $Q(x', \cdot)$, we have

$$\begin{aligned} (\mu'Q^t)(h) - (\mu Q^t)(h) &= \mathbb{E}h(X'_t) - \mathbb{E}h(X_t) \\ &= \mathbb{E}[h(X'_t) - h(X_t)]\mathbb{1}\{X'_t \preceq X_t\} + \mathbb{E}[h(X'_t) - h(X_t)]\mathbb{1}\{X'_t \not\preceq X_t\}^c. \end{aligned}$$

Since h is increasing, this leads to

$$(\mu'Q^t)(h) - (\mu Q^t)(h) \leq \mathbb{E}[h(X'_t) - h(X_t)]\mathbb{1}\{X'_t \preceq X_t\}^c \leq \mathbb{P}\{X'_t \preceq X_t\}^c.$$

Since $\tau \leq t$ implies $X'_t \preceq X_t$ we have $\{X'_t \preceq X_t\}^c \subset \{\tau > t\}$, and hence

$$(\mu'Q^t)(h) - (\mu Q^t)(h) \leq \mathbb{P}\{\tau > t\}. \quad (9)$$

Now define $N_t := \sum_{j=0}^t \mathbb{1}\{(X'_j, X_j) \in C \times C\}$. Fixing $j \in \mathbb{N}$ with $j \leq t$, we have

$$\mathbb{P}\{\tau > t\} = \mathbb{P}\{\tau > t, N_t < j\} + \mathbb{P}\{\tau > t, N_t \geq j\}. \quad (10)$$

To bound the first term in (10), we set $W(x, x') := [V(x) + V(x')]/2$. Since $\hat{Q}((x, x'), \cdot)$ is a coupling of $Q(x, \cdot)$ and $Q(x', \cdot)$, we have

$$\hat{Q}W(x, x') = \frac{QV(x) + QV(x')}{2} \leq \lambda W(x, x') + \beta. \quad (11)$$

Hence, applying Lemma 2.2 to \hat{Q} yields

$$\mathbb{P}\{\tau > t, N_t < j\} \leq \mathbb{P}\{N_t < j\} \leq \gamma^t d^{j-1} H(\mu, \mu') \quad (12)$$

Regarding the second term in (10), we claim that

$$\mathbb{P}\{\tau > t, N_t \geq j\} \leq (1 - \varepsilon)^j. \quad (13)$$

To see this, suppose $(J_i)_{i \geq 1}$ is the times of the successive visits of (X_t, X'_t) to $C \times C$. That is, J_1 is the time of the first visit and

$$J_{i+1} := \inf\{m \geq J_i + 1 : (X_m, X'_m) \in C \times C\} \quad \text{for } i \geq 1.$$

It is not difficult to see that $\{N_t > j\} \subset \{J_j \leq t - 1\}$. As a result,

$$\mathbb{P}\{\tau > t, N_t > j\} \leq \mathbb{P}\{\tau > t, J_j + 1 \leq t\}. \quad (14)$$

Consider the set $\{\tau > t, J_j + 1 \leq t\}$. If a path is in this set, then as $\tau > t$, for any index j with $j \leq t$ we have $X'_j \not\preceq X_j$. In addition, $J_i + 1 \leq J_j + 1 \leq t$ for any $i \leq j$, so $X'_{J_i+1} \not\preceq X_{J_i+1}$ for every $i \leq j$.

$$\therefore \mathbb{P}\{\tau > t, J_j + 1 \leq t\} \leq \mathbb{P} \bigcap_{i=1}^j \{X'_{J_i+1} \not\preceq X_{J_i+1}\}. \quad (15)$$

Observe that

$$\mathbb{P} \cap_{i=1}^j \{X'_{J_i+1} \not\leq X_{J_i+1}\} = \mathbb{P} \left[\cap_{i=1}^{j-1} \{X'_{J_i+1} \not\leq X_{J_i+1}\} \mathbb{P}[X'_{J_j+1} \not\leq X_{J_j+1} \mid \mathcal{F}_{J_j}] \right].$$

By the definition of J_j we have $(X_{J_j}, X'_{J_j}) \in C \times C$. Using this fact, the strong Markov property and the definition of \hat{Q} (see (4)) yields

$$\mathbb{P}[X'_{J_j+1} \preceq X_{J_j+1} \mid \mathcal{F}_{J_j}] = \hat{Q}((X'_{J_j}, X_{J_j}), \mathbb{G}) = \alpha(Q(X'_{J_j}, \cdot), Q(X_{J_j}, \cdot)).$$

Applying the definition of ε in (8), we obtain $\mathbb{P}[X_{J_j+1} \not\leq X'_{J_j+1} \mid \mathcal{F}_{J_j}] \leq 1 - \varepsilon$, so

$$\mathbb{P} \cap_{i=1}^j \{X_{J_i+1} \not\leq X'_{J_i+1}\} \leq (1 - \varepsilon) \mathbb{P} \cap_{i=1}^{j-1} \{X_{J_i+1} \not\leq X'_{J_i+1}\}.$$

Continuing to iterate backwards in this way yields $\mathbb{P} \cap_{i=1}^j \{X_{J_i+1} \not\leq X'_{J_i+1}\} \leq (1 - \varepsilon)^j$. Combining this inequality with (14) and (15) verifies (13).

Combining (9), (10), (12), and (13) yields

$$(\mu' Q^t)(h) - (\mu Q^t)(h) \leq (1 - \varepsilon)^j + \gamma^t d^{j-1} H(\mu, \mu').$$

Reversing the roles of μ and μ' does not change the value on the right-hand side of this bound, and hence

$$|(\mu' Q^t)(h) - (\mu Q^t)(h)| \leq (1 - \varepsilon)^j + \gamma^t d^{j-1} H(\mu, \mu')$$

also holds. Taking the supremum over all $h \in \text{ib}\mathcal{B}$ with $0 \leq h \leq 1$ completes the proof of Theorem 3.1. \square

4. EXAMPLES AND APPLICATIONS

In this section we consider some special cases, with a focus on (a) connections to the existing literature and (b) how to obtain an estimate of the value ε in (8).

4.1. Connection to Total Variation Results. One special case is when \preceq is the identity order, so that $x \preceq y$ if and only if $x = y$. For this order we have $\text{ib}\mathcal{B} = \text{b}\mathcal{B}$, so every stochastic kernel is increasing, and the Kolmogorov metric (see (1)) becomes the total variation distance. In this setting total variation setting, Theorem 3.1 is similar to standard geometric bounds for total variation distance, such as Theorem 1 in [21].

It is worth noting that, in the total variation setting, ε in (8) is at least as large as the analogous term ε in Theorem 1 in [21]. Indeed, in [21], the value ε , which we now write as $\hat{\varepsilon}$ to avoid confusion, comes from an assumed minorization condition: there exists a $\nu \in \text{p}\mathcal{B}$ such that

$$\hat{\varepsilon}\nu(B) \leq Q(x, B) \quad \text{for all } B \in \mathcal{B} \text{ and } x \in C. \quad (16)$$

To compare $\hat{\varepsilon}$ with ε defined in (8), suppose that this minorization condition holds and define the residual kernel $R(x, B) := (Q(x, B) - \hat{\varepsilon}\nu(B))/(1 - \hat{\varepsilon})$. Fixing $(x, x') \in C \times C$, we draw (X, X') as follows: With probability ε , draw $X \sim \nu$ and set $X' = X$. With probability $1 - \varepsilon$, independently draw $X \sim R(x, \cdot)$ and $X' \sim R(x', \cdot)$. Simple arguments confirm that X is a draw from $Q(x, \cdot)$ and X' is a draw from $Q(x', \cdot)$. Recalling that \preceq is the identity order, this leads to $\hat{\varepsilon} = \mathbb{P}\{I \leq \hat{\varepsilon}\} \leq \mathbb{P}\{X = X'\} = \mathbb{P}\{X \preceq X'\} \leq \alpha(Q(x, \cdot), Q(x', \cdot))$. (The last bound is by the definition of α in (2) and the fact that the joint distribution of (X, X') is a coupling of $Q(x, \cdot)$ and $Q(x', \cdot)$.) Since, in this discussion, the point (x, x') was arbitrarily chosen from $C \times C$, we conclude that $\hat{\varepsilon} \leq \varepsilon$, where ε is as defined in (8).

4.2. Stochastic Recursive Sequences. The preceding section showed that Theorem 3.1 reduces to existing results for bounds on total variation distance when the partial order \preceq is the identity order. Now we show how Theorem 3.1 leads to new results other settings, such as when \preceq is a pointwise partial order. To this end, consider the process

$$X_{t+1} = F(X_t, W_{t+1}) \quad (17)$$

where $(W_t)_{t \geq 1}$ is an IID shock process taking values in some space \mathbb{W} , and F is a measurable function from $\mathbb{X} \times \mathbb{W}$ to \mathbb{X} . The common distribution of each W_t is denoted by φ . We suppose that F is increasing, in the sense that $x \preceq x'$ implies $F(x, w) \preceq F(x', w)$ for any fixed $w \in \mathbb{W}$. We let Q represent the stochastic kernel corresponding to (17), so that $Q(x, B) = \varphi\{w \in \mathbb{W} : F(x, w) \in B\}$ for all $x \in \mathbb{X}$ and $B \in \mathcal{B}$. Since F is increasing, the kernel Q is increasing. Hence Theorem 3.1 applies. We can obtain a lower bound on ε in (8) by calculating

$$e := \inf \left\{ \int \int \mathbb{1}\{F(x', w') \leq F(x, w)\} \varphi(dw) \varphi(dw') : (x, x') \in C \times C \right\}. \quad (18)$$

Indeed, if W and W' are drawn independently from φ , then $X = F(x, W)$ is a draw from $Q(x, \cdot)$ and $X' = F(x', W)$ is a draw from $Q(x', \cdot)$. Hence

$$e = \mathbb{P}\{X' \preceq X\} \leq \alpha(Q(x, \cdot), Q(x', \cdot)) \leq \varepsilon. \quad (19)$$

4.3. Example: TCP Window Size Process. To help illustrate how to compute e in (18), we consider the TCP window size process (see, e.g., [2]), which has embedded jump chain $X_{t+1} = a(X_t^2 + 2E_{t+1})^{1/2}$. Here $a \in (0, 1)$ and (E_t) is IID exponential with unit rate. If $C = [0, c]$, then drawing E, E' as independent standard exponentials and using (19),

$$\varepsilon \geq \inf_{0 \leq x, y \leq c} \mathbb{P}\{a\sqrt{y^2 + 2E'} \leq a\sqrt{x^2 + 2E}\} = \mathbb{P}\{\sqrt{c^2 + 2E'} \leq \sqrt{2E}\}.$$

Since $E' - E$ has the Laplace- $(0, 1)$ distribution, we get

$$1 - \varepsilon \leq \mathbb{P}\{c^2 + 2E' > 2E\} = \mathbb{P}\{E' - E > c^2/2\} = \frac{1}{2} \exp(-c^2/2).$$

4.4. Example: When Minorization Fails. We provide an elementary scenario where Theorem 3.1 provides a usable bound while the minorization based methods described in Section 4.1 do not. Let \mathbb{Q} be the rational numbers, let $\mathbb{X} = \mathbb{R}$, and assume that

$$X_{t+1} = \frac{X_t}{2} + W_{t+1} \quad \text{where } W_t \text{ is IID on } \{0, 1\} \text{ and } \mathbb{P}\{W_t = 0\} = 1/2.$$

Let C contain at least one rational and one irrational number. Let μ be a measure on the Borel sets of \mathbb{R} obeying $\mu(B) \leq Q(x, B) = \mathbb{P}\{x/2 + W \in B\}$ for all $x \in C$ and Borel sets B . If x is rational, then $x/2 + W \in \mathbb{Q}$ with probability one, so $\mu(\mathbb{Q}^c) \leq Q(x, \mathbb{Q}^c) = 0$. Similarly, if x is irrational, then $x/2 + W \in \mathbb{Q}^c$ with probability one, so $\mu(\mathbb{Q}) \leq Q(x, \mathbb{Q}) = 0$. Hence μ is the zero measure on \mathbb{R} . Thus, we cannot take a $\hat{\varepsilon} > 0$ and probability measure ν obeying the minorization condition (16). On the other hand, letting $V(x) = x + 1$ and $d = 1$, so that $C = \{V \leq 2\} = [0, 1]$, the value e from (18) obeys $e = \mathbb{P}\{1/2 + W \leq W'\} = \mathbb{P}\{W' - W \geq 1/2\} = \frac{1}{4}$. Hence, by (19), the constant ε in Theorem 3.1 is positive.

4.5. Example: Wealth Dynamics. Many economic models examine wealth dynamics in the presence of credit market imperfections (see, e.g., [1]). These often result in dynamics of the form

$$X_{t+1} = \eta_{t+1} G(X_t) + \xi_{t+1}, \quad (\eta_t) \stackrel{\text{iid}}{\sim} \varphi, \quad (\xi_t) \stackrel{\text{iid}}{\sim} \psi. \quad (20)$$

Here (X_t) is some measure of household wealth, G is a function from \mathbb{R}_+ to itself and (η_t) and (ξ_t) are independent \mathbb{R}_+ -valued sequences. The function G is increasing, since greater current wealth relaxes borrowing constraints and increases financial income. We assume that there exists a $\lambda < 1$ such that $\mathbb{E} \eta_t G(x) \leq \lambda x$ for all $x \in \mathbb{R}_+$, and, in addition, that $\bar{\xi} := \mathbb{E} \xi_t < \infty$.

Let Q be the stochastic kernel corresponding to (20). With $V(x) = x + 1$, we have

$$QV(x) = \mathbb{E}[\eta_{t+1} G(x) + \xi_{t+1} + 1] \leq \lambda x + \bar{\xi} + 1 \leq \lambda V(x) + \bar{\xi} + 1. \quad (21)$$

Fixing $d \in \mathbb{R}_+$ and setting $C = \{V \leq d\} = [0, d]$, we can obtain e in (18) via

$$e = \mathbb{P}\{\eta' G(d) + \xi' \leq \eta G(0) + \xi\} \quad \text{when} \quad (\eta', \xi', \eta, \xi) \sim \varphi \times \psi \times \varphi \times \psi.$$

This term will be strictly positive under suitable conditions, such as when ψ has a sufficiently large support. Combining (19) and (21) with the bound in

Theorem 3.1, we have, for any μ and μ' in $p\mathcal{B}$ and $j, t \in \mathbb{N}$ with $j \leq t$,

$$\kappa(\mu Q^t, \mu' Q^t) \leq (1 - e)^j + \frac{1}{2} \left(\lambda + \frac{2(\bar{\xi} + 1)}{d} \right)^t d^{j-1} H(\mu, \mu').$$

where $H(\mu, \mu') := (\int x\mu(dx) + \int x\mu'(dx)) / 2$.

Notice that, for this model, we cannot compute useful total variation bounds or Wasserstein distance bounds without additional assumptions.

REFERENCES

- [1] Antonio Antunes and Tiago Cavalcanti. Start up costs, limited enforcement, and the hidden economy,. *European Economic Review*, 51:203–224, 2007.
- [2] Jean-Baptiste Bardet, Alejandra Christen, Arnaud Guillin, Florent Malrieu, and Pierre-André Zitt. Total variation estimates for the TCP process. *Electron. J. Probab*, 18(10):1–21, 2013.
- [3] Rabi Bhattacharya and Mukul Majumdar. On a theorem of Dubins and Freedman. *Journal of Theoretical Probability*, 12:1067–1087, 1999.
- [4] Rabi Bhattacharya, Mukul Majumdar, and Nigar Hashimzade. Limit theorems for monotone Markov processes. *Sankhya A*, 72:170–190, 2010.
- [5] Rabi N Bhattacharya and Oesook Lee. Asymptotics of a class of Markov processes which are not in general irreducible. *The Annals of Probability*, pages 1333–1347, 1988.
- [6] Djalil Chafaï, Florent Malrieu, and Katy Paroux. On the long time behavior of the TCP window size process. *Stochastic Processes and their Applications*, 120(8):1518–1534, 2010.
- [7] Stephen B Connor and Gersende Fort. State-dependent Foster–Lyapunov criteria for subgeometric convergence of Markov chains. *Stochastic Processes and their Applications*, 119(12):4176–4193, 2009.
- [8] Sergey Foss, Vsevolod Shneer, Jonathan P Thomas, and Tim Worrall. Stochastic stability of monotone economies in regenerative environments. *Journal of Economic Theory*, 173:334–360, 2018.
- [9] Serguei Foss and Takis Konstantopoulos. An overview of some stochastic stability methods. *Journal of the Operations Research Society of Japan*, 47(4):275–303, 2004.
- [10] Julia Gaudio, Saurabh Amin, and Patrick Jaillet. Exponential convergence rates for stochastically ordered Markov processes with random initial conditions. *arXiv preprint arXiv:1810.07732v1*, 202118.
- [11] Robert E Gaunt and Siqi Li. Bounding Kolmogorov distances through Wasserstein and related integral probability metrics. *Journal of Mathematical Analysis and Applications*, 522(1):126985, 2023.

- [12] Yu Hang Jiang, Tong Liu, Zhiya Lou, Jeffrey S Rosenthal, Shanshan Shang-guan, Fei Wang, and Zixuan Wu. The coupling/minorization/drift approach to Markov chain convergence rates. *Notices of the American Mathematical Society*, 68(4), 2021.
- [13] Takashi Kamihigashi and John Stachurski. Stochastic stability in monotone economies. *Theoretical Economics*, 9(2):383–407, 2014.
- [14] Takashi Kamihigashi and John Stachurski. A unified stability theory for classical and monotone Markov chains. *Journal of Applied Probability*, 56(1):1–22, 2019.
- [15] Takashi Kamihigashi and John Stachurski. Partial stochastic dominance via optimal transport. *Operations Research Letters*, 48(5):584–586, 2020.
- [16] Torgny Lindvall. *Lectures on the coupling method*. Dover, 2002.
- [17] Robert B Lund, Sean P Meyn, and Richard L Tweedie. Computable exponential convergence rates for stochastically ordered Markov processes. *The Annals of Applied Probability*, 6(1):218–237, 1996.
- [18] Ravi Montenegro and Prasad Tetali. Mathematical aspects of mixing times in Markov chains. *Foundations and Trends in Theoretical Computer Science*, 1(3):237–354, 2006.
- [19] Qian Qin and James P Hobert. Geometric convergence bounds for Markov chains in Wasserstein distance based on generalized drift and contraction conditions. 58(2):872–889, 2022.
- [20] Yanlin Qu, Jose Blanchet, and Peter Glynn. Computable bounds on convergence of Markov chains in Wasserstein distance. *arXiv preprint arXiv:2308.10341*, 2023.
- [21] Jeffrey S Rosenthal. Quantitative convergence rates of Markov chains: A simple account. *Electronic Communications in Probability*, 7:123–128, 2002.
- [22] Jeffrey S Rosenthal. How Markov’s little idea transformed statistics. *Handbook of the History and Philosophy of Mathematical Practice*, pages 1–11, 2023.
- [23] Volker Strassen. The existence of probability measures with given marginals. *The Annals of Mathematical Statistics*, 36(2):423–439, 1965.
- [24] Shaoyi Zhang. Existence and application of optimal Markovian coupling with respect to non-negative lower semi-continuous functions. *Acta Mathematica Sinica*, 16(2):261–270, 2000.

APPENDIX A. PROOF OF LEMMA 2.2

Let the conditions of Lemma 2.2 hold and let Q , \hat{Q} and $((X_t, X'_t))_{t \geq 0}$ be as described above. We assume in what follows that $H(\mu, \mu')$ is finite, since otherwise

Lemma 2.2 is trivial. Setting $W(x, x') := (V(x) + V(x'))/2$, we have

$$\hat{Q}W(x, x') = \frac{QV(X_t) + QV(X'_t)}{2} \leq \lambda W(x, x') + \beta. \quad (22)$$

Now define

$$M_t := \gamma^{-t} d^{-N_{t-1}} W(X_t, X'_t) \quad \text{for } t \geq 0 \text{ with } N_{-1} := 1.$$

We claim that $(M_t)_{t \geq 0}$ is an (\mathcal{F}_t) -supermartingale. Clearly (M_t) is adapted. To see that $\mathbb{E}[M_{t+1} | \mathcal{F}_t] \leq M_t$ holds \mathbb{P} -almost surely,¹ let $F_t := \mathbb{1}\{(X_t, X'_t) \in C \times C\}$, and let $F_t^c := 1 - F_t$, so that

$$\mathbb{E}[M_{t+1} | \mathcal{F}_t] = \mathbb{E}[M_{t+1}F_t | \mathcal{F}_t] + \mathbb{E}[M_{t+1}F_t^c | \mathcal{F}_t]. \quad (23)$$

On $C \times C$ we have $W \leq d$, so, by (22), $\mathbb{E}[W(X_{t+1}, X'_{t+1}) | \mathcal{F}_t] \leq \lambda d + \beta = d\gamma$. Therefore,

$$\mathbb{E}[M_{t+1}F_t | \mathcal{F}_t] = \gamma^{-(t+1)} d^{-N_t} \mathbb{E}[W(X_{t+1}, X'_{t+1}) | \mathcal{F}_t] F_t \leq \gamma^{-t} d^{-N_{t-1}} F_t.$$

Also, on F_t we have $N_t = N_{t-1} + 1$. Using this fact and $1 \leq W$ yields

$$\mathbb{E}[M_{t+1}F_t | \mathcal{F}_t] \leq \gamma^{-t} d^{-N_{t-1}} F_t \leq M_t F_t. \quad (24)$$

Turning to the term $\mathbb{E}[M_{t+1}F_t^c | \mathcal{F}_t]$, observe that on F_t^c we have $W \geq d/2$, so, using (11) again,

$$\frac{\hat{Q}W}{W} \leq \lambda + \frac{\beta}{W} \leq \lambda + \frac{2\beta}{d} = \gamma.$$

Therefore, $\mathbb{E}[W(X_{t+1}, X'_{t+1})F_t^c | \mathcal{F}_t] \leq \gamma W(X_t, X'_t)F_t^c$. Combining this bound with the fact that $N_t = N_{t-1}$ on F_t^c yields

$$\mathbb{E}[M_{t+1}F_t^c | \mathcal{F}_t] = \gamma^{-(t+1)} d^{-N_t} \mathbb{E}[W(X_{t+1}, X'_{t+1}) | \mathcal{F}_t] F_t^c \leq \gamma^{-t} d^{-N_{t-1}} F_t^c \leq M_t F_t^c,$$

where the last inequality used $1 \leq W$. Together with (24) and (23), this inequality gives $\mathbb{E}[M_{t+1} | \mathcal{F}_t] \leq M_t$, so (M_t) is a supermartingale as claimed.

Now fix $t \in \mathbb{N}$ and $j \leq t$. Since $d \geq 1$ we have

$$\mathbb{P}\{N_t < j\} \leq \mathbb{P}\{N_{t-1} \leq j-1\} = \mathbb{P}\{d^{-N_{t-1}} \geq d^{-j+1}\}.$$

From Chebychev's inequality, $1 \leq W$ and the supermartingale property, the last term is dominated by

$$d^{j-1} \mathbb{E} d^{-N_{t-1}} \leq \gamma^t d^{j-1} \mathbb{E}[\gamma^{-t} d^{-N_{t-1}} W(X_t, X'_t)] = \gamma^t d^{j-1} \mathbb{E}[M_t] \leq \gamma^t d^{j-1} \mathbb{E}[M_0].$$

The last term is just $\gamma^t d^{j-1} H(\mu, \mu')$, so the claim in Lemma 2.2 is now proved.

¹This inequality implies integrability of M_t because then $\mathbb{E}M_t \leq \mathbb{E}M_0$ and $\mathbb{E}M_0 = H(\mu, \mu')$ is finite by assumption.