Check for updates

*J. Appl. Probab.* **63,** 798–810 (2026)
*doi:*10.1017/jpr.2025.10055

# QUANTITATIVE CONVERGENCE RATES FOR STOCHASTICALLY MONOTONE MARKOV CHAINS

TAKASHI KAMIHIGASHI,* *Kobe University*
JOHN STACHURSKI,** *Australian National University*

**Abstract**

For Markov chains and Markov processes exhibiting a form of stochastic monotonicity (higher states have higher transition probabilities in terms of stochastic dominance), stability and ergodicity results can be obtained with the use of 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.

*Keywords:* Markov chains; stochastic monotonicity; convergence rates; total ordered variation

2020 Mathematics Subject Classification: Primary 60J05
Secondary 37A25; 60J10

## 1. Introduction

Quantitative bounds on the distance between distributions generated by Markov models have many applications in statistics, computer science, and the natural and social sciences (see, e.g., [20, 25]). One approach to producing such bounds uses total variation distance and exploits minorization conditions (see, e.g., [2, 8, 13, 24). Another branch of the literature bounds deviations using Wasserstein distance (see, e.g., [4, 5, 7, 21, 22]). In general, total variation bounds require relatively strong mixing conditions on the law of motion in some 'attracting' region of the state space, while Wasserstein bounds rely on some degree of continuity of the laws of motion with respect to a specified metric. Total variation bounds can also be studied within the setting of Wasserstein distance by choosing the ground metric on the state space to be the discrete metric (see, e.g., [4]).

Although this research covers many important applications, there are also significant cases where Markov chains lack both the minorization and the 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 [3, 9, 10, 15, 16]. In this article, we complement these

Received 1 October 2024; accepted 23 September 2025.
* Postal address: Center for Computational Social Science and Research Institute for Economics and Business Administration, Kobe University, Kobe, 657-8501, Japan. Email: tkamihigashi@rieb.kobe-u.ac.jp
** Postal address: Research School of Economics, Australian National University, Canberra ACT 2601, Australia. Email: john.stachurski@anu.edu.au

© The Author(s), 2026. Published by Cambridge University Press on behalf of Applied Probability Trust. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.



https://doi.org/10.1017/jpr.2025.10055 Published online by Cambridge University Press

stability and ergodicity results by providing quantitative bounds for stochastically monotone Markov chains.

While there are some existing results that use stochastic monotonicity to bound the distributions generated by Markov chains [11, 19], these bounds are stated in terms of total variation distance, which again requires traditional minorization conditions (as opposed to the order-theoretic mixing conditions discussed in the previous paragraph). In this article, we aim to fully exploit monotonicity by instead bounding total *ordered* variation distance [16] between distributions. This works well because (a) our mixing conditions are stated in terms of order and (b) total ordered variation distance respects order structure on the state space.

Our main theorem is closely related to the total variation bound in Theorem 1 in [24], which is representative of existing work on total variation bounds and provides a simple and elegant proof. The main differences between that theorem and the one presented below is that we use total ordered variation distance instead of total variation distance and an order-theoretic mixing condition instead of a standard minorization condition. At the same time, as we show in Section 5.1, it is possible to recover Theorem 1 in [24] from our main theorem by adopting a particular choice of partial order.

Our work is also related to Wasserstein bounds on the deviation between distributions for Markov models, as found, for example, in [4, 22]. However, rather than bounding Wasserstein distance, our main theorem bounds deviations measured in terms of two directed Wasserstein semimetrics, each of which is connected to the same partial order on the state space. Further details are given in Section 5.2.

## 2. Setup

We first recall key definitions and state some preliminary results.

### 2.1. Environment

Throughout this article, $\mathbb{X}$ 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

$$\mathbb{G} = \{(x, x') \in \mathbb{X} \times \mathbb{X} : x \preceq x'\}, \tag{1}$$

is closed under the product topology on $\mathbb{X} \times \mathbb{X}$. A map $h\colon \mathbb{X} \to \mathbb{R}$ is called *increasing* if $x \preceq x'$ implies $h(x) \leqslant h(x')$. 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 $\mathbb{R}$. Given $h \in b\mathcal{B}$ and $\mu \in p\mathcal{B}$ we set $\mu(h) := \int h \, \mathrm{d}\mu$. The symbol $ib\mathcal{B}$ represents all increasing $h \in b\mathcal{B}$.

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

$$\rho(\mu, \nu) := \sup_{I \in i\mathcal{B}} (\mu(I) - \nu(I)) + \sup_{I \in i\mathcal{B}} (\nu(I) - \mu(I)). \tag{2}$$

This is the *total ordered variation* metric on $p\mathcal{B}$. A proof that $\rho$ is indeed a metric can be found in Lemma 4.1 in [16]. Positive definiteness follows from the fact that $\rho(\mu, \nu) = 0$ implies $\mu \preceq_s \nu$ and $\nu \preceq_s \mu$. Since $\preceq_s$ is antisymmetric on $p\mathcal{B}$ [14, Lemma 1], we then have $\mu = \nu$. Connections between $\rho$ and the total variation and Wasserstein metrics are discussed in Section 5.

A function $Q\colon (\mathbb{X}, \mathcal{B}) \to \mathbb{R}$ is called a *transition kernel* on $\mathbb{X}$ if $Q$ is a map from $\mathbb{X} \times \mathcal{B}$ to [0,1] such 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

https://doi.org/10.1017/jpr.2025.10055 Published online by Cambridge University Press

$Q(x, \cdot)$ at given $x$. A transition kernel $Q$ on $\mathbb{X}$ is called *increasing* if $Qh \in ib\mathcal{B}$ whenever $h \in ib\mathcal{B}$. Equivalently,

$$Q_x \preceq_s Q_{x'} \quad \text{whenever} \quad x \preceq x'.$$

For a transition kernel $Q$ on $\mathbb{X}$, we define the *left and right Markov operators* generated by $Q$ via

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

(The left Markov operator $\mu \mapsto \mu Q$ maps $p\mathcal{B}$ to itself, while the right Markov operator $f \mapsto Qf$ acts on $f \in b\mathcal{B}$.) A discrete-time $\mathbb{X}$-valued stochastic process $(X_t)_{t\geqslant 0}$ on a filtered probability space $(\Omega, \mathscr{F}, \mathbb{P}, (\mathscr{F}_t)_{t\geqslant 0})$ is called Markov-$(Q, \mu)$ if $X_0 \overset{d}{=} \mu$ and $\mathbb{E}[h(X_{t+1}) \mid \mathscr{F}_t] = Qh(X_t)$ with probability 1 for all $t \geqslant 0$ 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 $\pi$ on $\mathcal{B} \otimes \mathcal{B}$ satisfying $\pi(A \times \mathbb{X}) = \mu(A)$ and $\pi(\mathbb{X} \times A) = \nu(A)$ for all $A \in \mathcal{B}$. Let $\mathscr{C}(\mu, \nu)$ denote the set of all couplings of $(\mu, \nu)$ and let

$$\alpha(\mu, \nu) = \sup_{\pi \in \mathscr{C}(\mu,\nu)} \pi(\mathbb{G}). \tag{3}$$

The value $\alpha(\mu, \nu)$ can be understood as a measure of 'partial stochastic dominance' of $\nu$ over $\mu$ [17]. In line with this interpretation, and applying Strassen's theorem [18, 26], we have

$$\alpha(\mu, \nu) = 1 \quad \text{whenever} \quad \mu \preceq_s \nu. \tag{4}$$

Let $Q$ be a transition kernel on $\mathbb{X}$. A *Markov coupling* of $Q$ is a real-valued function $\hat{Q}$ on $(\mathbb{X} \times \mathbb{X}) \times (\mathcal{B} \otimes \mathcal{B})$ such that

(i) $(x, x') \mapsto \hat{Q}((x, x'), E)$ is measurable for each $E \in \mathcal{B} \otimes \mathcal{B}$ and

(ii) $\hat{Q}_{(x,x')}$ is a coupling of $Q_x$ and $Q_{x'}$ for all $x, x' \in \mathbb{X}$.

In other words, $\hat{Q}$ is a transition kernel on $\mathbb{X} \times \mathbb{X}$ that couples the distributions $Q_x$ and $Q_{x'}$ at every pair of points in the state space.

We call $\hat{Q}$ a $\preceq$-maximal Markov coupling of $Q$ if $\hat{Q}$ is a Markov coupling of $Q$ and, in addition,

$$\hat{Q}((x, x'), \mathbb{G}) = \alpha(Q_x, Q_{x'}) \quad \text{for all } (x, x') \in \mathbb{X} \times \mathbb{X}. \tag{5}$$

Informally, $\hat{Q}$ serves as a transition kernel of a 'joint' chain $((X_t, X'_t))_{t\geqslant 0}$ that maximizes the probability of attaining $X_t \preceq X'_t$ at each step. Below we use the following existence result.

**Lemma 1.** *If $Q$ is a transition kernel on $\mathbb{X}$ then $Q$ has at least one $\preceq$-maximal Markov coupling.*

*Proof.* Let $Q$ be a transition kernel on $\mathbb{X}$. By Theorem 1.1 in [27], given lower semicontinuous $g\colon \mathbb{X} \times \mathbb{X} \to \mathbb{R}$, there exists a transition kernel $\hat{Q}$ on $\mathbb{X} \times \mathbb{X}$ such that $\hat{Q}$ is a Markov coupling of $Q$ and, in addition

$$(\hat{Q}g)(x, x') = \inf \left\{ \int g \, \mathrm{d}\pi : \pi \in \mathscr{C}(Q_x, Q_{x'}) \right\}.$$

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

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

This shows that $\hat{Q}$ is a $\preceq$-maximal Markov coupling of $Q$. □

https://doi.org/10.1017/jpr.2025.10055 Published online by Cambridge University Press

The following simple lemma will be important for our bounds.

**Lemma 2.** *If $Q$ is increasing and $\hat{Q}$ is a $\preceq$-maximal Markov coupling of $Q$, then $\mathbb{G}$ is absorbing for $\hat{Q}$.*

*Proof.* Let $Q$ and $\hat{Q}$ be as stated. If $(x, x')$ is in $\mathbb{G}$ then, since $Q$ is increasing, we have $Q_x \preceq_s Q_{x'}$. Combining this inequality with (4), we obtain $\alpha(Q_x, Q_{x'}) = 1$. Application of the property (5) produces $\hat{Q}((x, x'), \mathbb{G}) = 1$; hence, $\mathbb{G}$ is absorbing for $\hat{Q}$. □

## 3. Main Result

In this section we state our main result concerning rates of convergence.

### 3.1. A quantitative bound

Let $Q$ be an increasing transition kernel on $\mathbb{X}$ and let $\hat{Q}$ be a $\preceq$-maximal Markov coupling of $Q$. Let $W$ be a measurable function from $\mathbb{X} \times \mathbb{X}$ to $[1, \infty)$. Suppose that, for some measurable set $C$ in $\mathbb{X}$ and some strictly increasing convex function $\delta\colon \mathbb{R}_+ \to \mathbb{R}_+$ with $\delta(0) = 0$, the kernel $\hat{Q}$ obeys the drift condition

$$\hat{Q}W(x, x') \leqslant \delta(W(x, x')) \quad \text{for all } (x, x') \notin C \times C. \tag{6}$$

We set

$$B = \max\left\{1,\ \delta^{-1}(B_0)\right\}, \quad \text{where} \quad B_0 := \sup_{(x,x') \in C \times C} \hat{R}W(x, x'), \tag{7}$$

and

$$\hat{R}W(x, x') := \int \int W(y, y') \mathbb{1}\{y \not\preceq y'\} \hat{Q}((x, x'), \mathrm{d}(y, y')). \tag{8}$$

In addition, let

$$\varepsilon = \inf\left\{\alpha(Q_x, Q_{x'}) : (x, x') \in C \times C\right\}. \tag{9}$$

Given $\mu$, $\mu'$ in $p\mathcal{B}$, we set

$$(\mu \times \mu')(W) = \int \int W(x, x') \mu(\mathrm{d}\, x) \mu'(\mathrm{d}\, x').$$

In (7), $\delta^{-1}$ is the inverse of $\delta$. Below, $\delta^{-t}$ indicates $t$ compositions of $\delta^{-1}$ with itself. We are now ready to state the main result on quantitative bounds.

**Theorem 1.** *For all $j, t \in \mathbb{N}$ with $j \leqslant t$, we have*

$$\rho(\mu Q^t, \mu' Q^t) \leqslant 2(1 - \varepsilon)^j + \frac{B^{j-1}}{\delta^{-t}(1)} [(\mu \times \mu')(W) + (\mu' \times \mu)(W)].$$

Theorem 1 provides a total ordered variation bound on the deviation between the time-$t$ distributions $\mu Q^t$ and $\mu' Q^t$ generated by iteration with the Markov operator $Q$, taking as given an arbitrary pair of initial distributions $\mu$, $\mu'$ in $p\mathcal{B}$. Convergence of the bound in Theorem 1 to zero as $t \to \infty$ requires that $\varepsilon > 0$ and $\delta^{-1}(1) \to \infty$.

When $\delta$ is linear we obtain the geometric case. Since this case is important we state it as a corollary.

**Corollary 1.** *If the conditions above hold with (6) replaced by*

$$\hat{Q}W(x, x') \leqslant \gamma W(x, x') \quad \text{for all } (x, x') \notin C \times C \tag{10}$$

https://doi.org/10.1017/jpr.2025.10055 Published online by Cambridge University Press

*for some positive constant $\gamma$, then, for all $j, t \in \mathbb{N}$ with $j \leqslant t$,*

$$\rho(\mu Q^t, \mu' Q^t) \leqslant 2(1-\varepsilon)^j + \gamma^t B^{j-1}[(\mu \times \mu')(W) + (\mu' \times \mu)(W)].$$

Convergence of the bound in Corollary 1 to zero as $t \to \infty$ requires that the constants $\varepsilon$ and $\gamma$ obey $\varepsilon > 0$ and $\gamma \in [0, 1)$. The bound can be viewed as an order-theoretic version of the geometric total variation bound in Theorem 1 in [24]. Further comparisons are given in Section 5.1.

### 3.2. Sketch of the proof

The bound in Theorem 1 is obtained by tracking a joint chain $((X_t, X'_t))_{t \geqslant 0}$ generated by a $\preceq$-maximal Markov coupling $\hat{Q}$ and started from initial condition $\mu \times \mu'$. Because $\hat{Q}$ is a Markov coupling of $Q$, the individual chains $(X_t)_{t \geqslant 0}$ and $(X'_t)_{t \geqslant 0}$ are Markov-$(Q, \mu)$ and Markov-$(Q, \mu')$, respectively. Taking $\tau$ to be the first time that $X_t \preceq X'_t$ occurs, we can use the fact that $\mathbb{G}$ is absorbing for $\hat{Q}$ (Lemma 2) to obtain

$$\tau \leqslant t \quad \text{if and only if} \quad X_t \preceq X'_t. \tag{11}$$

Next, an order-theoretic version of a standard total variation coupling argument is used to generate the bound $(\mu Q^t)(I) - (\mu' Q^t)(I) \leqslant \mathbb{P}\{X_t \not\preceq X'_t\}$ for all $I \in i\mathcal{B}$. In view of (11), the left-hand side is also bounded above by $\mathbb{P}\{\tau > t\}$. This in turn is bounded with the use of the drift to $C \times C$ implied by (6), and the $\varepsilon$-probability of the joint chain $((X_t, X'_t))_{t \geqslant 0}$ entering $\mathbb{G}$ after $C \times C$ implied by (9). Reversing the roles of $\mu$ and $\mu'$ and then adding the two inequalities leads to the bound in Theorem 1. Details are given in Section 4.

### 3.3. Univariate drift

In applications, drift conditions on the underlying kernel $Q$ are usually easier to test and interpret than drift conditions on a joint kernel such as (6). Fortunately, there are relatively straightforward ways to map the former (let us call them 'univariate' drift conditions) to the latter ('joint' drift conditions). For example, suppose that $Q$ is a transition kernel on $\mathbb{X}$ and $V$ is a measurable function from $\mathbb{X}$ to $\mathbb{R}_+$. Suppose there exist $\lambda, \beta \in \mathbb{R}_+$ such that $\lambda < 1$ and

$$QV(x) \leqslant \lambda V(x) + \beta \qquad \text{for all } x \in \mathbb{X}. \tag{12}$$

In this setting, we can attain (10) by setting $C = \{x \in \mathbb{X} : V(x) \geqslant d\}$ for some fixed $d \geqslant 1$, and then

$$W(x, x') = 1 + V(x) + V(x') \quad \text{and} \quad \gamma = \frac{1 + \lambda d + 2\beta}{1 + d}.$$

A proof that (10) holds with these definitions can be found in Theorem 12 in [23].

Alternatively, if $V \geqslant 1$ then we can choose $C$ in the same way and then set

$$W(x, x') = \frac{V(x) + V(x')}{2} \quad \text{and} \quad \gamma = \lambda + \frac{2\beta}{d}.$$

Indeed, since $\hat{Q}$ is a Markov coupling of $Q$, an application of (12) yields

$$\hat{Q}W(x, x') = \frac{QV(x) + QV(x')}{2} \leqslant \lambda W(x, x') + \beta = \left(\lambda + \frac{\beta}{W(x, x')}\right) W(x, x').$$

The drift condition (10) now follows from $d/2 \leqslant W$ on the complement of $C \times C$.

https://doi.org/10.1017/jpr.2025.10055 Published online by Cambridge University Press

## 4. Proof of Theorem 1

In this section we prove Theorem 1. Throughout, $Q$ is an increasing transition kernel on $\mathbb{X}$, $\hat{Q}$ is a $\preceq$-maximal Markov coupling of $Q$, and the conditions in Section 3.1 are in force. We fix $\pi \in p(\mathcal{B} \otimes \mathcal{B})$ and take $((X_t, X'_t))_{t \geqslant 0}$ to be Markov-$(\hat{Q}, \pi)$ on a filtered probability space $(\Omega, \mathscr{F}, \mathbb{P}, (\mathscr{F}_t)_{t \geqslant 0})$. Let $\tau$ be the stopping time $\tau = \inf\{t \geqslant 0 : X_t \preceq X'_t\}$ with $\inf \varnothing = \infty$. Let

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

count the number of visits of this joint chain to $C \times C$. In addition, we set $N_{-1} := 0$.

**Lemma 3.** *The process $(M_t)_{t \geqslant 0}$ defined by*

$$M_t = B^{-N_{t-1}} \delta^{-t} [W(X_t, X'_t)] \mathbb{1}\{\tau > t\}$$

*is a supermartingale.*

*Proof.* In the argument below, we use the implication

$$(x, x') \in C \times C \quad \Longrightarrow \quad B^{-1} \delta^{-(t+1)} [\hat{R} W(x, x')] \leqslant \delta^{-t}(W(x, x')), \tag{13}$$

which holds for all $t \geqslant 0$. To establish (13), we fix $t$ and use $(x, x') \in C \times C$ and the definition of $B$ to obtain $\delta^{-1}(\hat{R}W(x, x')) \leqslant B$. Use of $W \geqslant 1$ now produces $\delta^{-1}(\hat{R}W(x, x')) \leqslant BW(x, x')$. Since $B \geqslant 1$ and $\delta^{-1}$ is increasing and concave with $\delta^{-1}(0) = 0$, application of $\delta^{-1}$ to both sides of the previous bound and use of the scaling inequality gives

$$\delta^{-2}(\hat{R}W(x, x')) \leqslant \delta^{-1}(BW(x, x')) \leqslant B\delta^{-1}(W(x, x')).$$

Continued iteration in the same way yields (13).

Now we show that $(M_t)_{t \geqslant 0}$ is an $(\mathscr{F}_t)$-supermartingale. Clearly $(M_t)_{t \geqslant 0}$ is adapted. In proving $\mathbb{E}[M_{t+1} \mid \mathscr{F}_t] \leqslant M_t$, we can and do assume that $\tau > t$, since $\tau \leqslant t$ implies $\tau \leqslant t + 1$, in which case the inequality is trivial. Let us first consider the case $(X_t, X'_t) \in C \times C$. When this holds, we have $N_t = N_{t-1} + 1$, so

$$\begin{aligned}
\mathbb{E}\big[M_{t+1} \mid \mathscr{F}_t\big] &= B^{-N_{t-1}-1} \mathbb{E}\big[\delta^{-(t+1)}\big(W\big(X_{t+1}, X'_{t+1}\big)\big) \mathbb{1}\{\tau > t+1\} \mid \mathscr{F}_t\big] \\
&= B^{-N_{t-1}-1} \mathbb{E}\big[\delta^{-(t+1)}\big(W\big(X_{t+1}, X'_{t+1}\big)\big) \mathbb{1}\big\{X_{t+1} \not\preceq X'_{t+1}\big\} \mid \mathscr{F}_t\big] \\
&= B^{-N_{t-1}-1} \mathbb{E}\big[\delta^{-(t+1)}\big(W\big(X_{t+1}, X'_{t+1}\big) \mathbb{1}\big\{X_{t+1} \not\preceq X'_{t+1}\big\}\big) \mid \mathscr{F}_t\big].
\end{aligned}$$

The second equality follows from the identity in (11), while the third follows from $\delta(0) = \delta^{-1}(0) = 0$. Since $\delta^{-1}$ is concave, using the previous chain of equalities and Jensen's inequality for conditional expectations, along with the definition of $\hat{R}$ in (8), we have

$$\begin{aligned}
\mathbb{E}\left[M_{t+1} \mid \mathscr{F}_t\right] &\leqslant B^{-N_{t-1}} B^{-1} \delta^{-(t+1)} \big[\mathbb{E}\big[W\big(X_{t+1}, X'_{t+1}\big) \mathbb{1}\big\{X_{t+1} \not\preceq X'_{t+1}\big\} \mid \mathscr{F}_t\big]\big] \\
&= B^{-N_{t-1}} B^{-1} \delta^{-(t+1)} \big[\hat{R} W\big(X_t, X'_t\big)\big] \\
&\leqslant B^{-N_{t-1}} \delta^{-t} \big[W\big(X_t, X'_t\big)\big] \\
&= M_t,
\end{aligned}$$

where the second inequality is by (13), as well as the restriction $(X_t, X'_t) \in C \times C$. The last equality holds because we are specializing to $\tau > t$.

https://doi.org/10.1017/jpr.2025.10055 Published online by Cambridge University Press

Now we turn to the case $(X_t, X'_t) \notin C \times C$. In this case we have $N_t = N_{t-1}$, so

$$\begin{aligned}
\mathbb{E}[M_{t+1} \mid \mathscr{F}_t] &= B^{-N_{t-1}} \mathbb{E}\big[\delta^{-(t+1)}\big[W\big(X_{t+1}, X'_{t+1}\big)\big]\mathbb{1}\{\tau > t+1\} \mid \mathscr{F}_t\big] \\
&\leqslant B^{-N_{t-1}} \mathbb{E}\big[\delta^{-(t+1)}\big[W\big(X_{t+1}, X'_{t+1}\big)\big] \mid \mathscr{F}_t\big] \\
&\leqslant B^{-N_{t-1}} \delta^{-(t+1)}\big[\mathbb{E}\big[W\big(X_{t+1}, X'_{t+1}\big) \mid \mathscr{F}_t\big]\big] \\
&\leqslant B^{-N_{t-1}} \delta^{-(t+1)}\big[\delta\big[W\big(X_t, X'_t\big)\big]\big] \\
&= B^{-N_{t-1}} \delta^{-t}\big[W\big(X_t, X'_t\big)\big] \\
&= M_t,
\end{aligned}$$

where the second inequality uses Jensen's inequality again and the third inequality uses the drift condition (6). As before, the last equality holds because we are specializing to $\tau > t$. We have now shown that $\mathbb{E}[M_{t+1} \mid \mathscr{F}_t] \leqslant M_t$ holds almost surely, so $(M_t)$ is a supermartingale, as claimed. □

**Lemma 4.** *If $j, t \in \mathbb{N}$ with $j \leqslant t$ then*

$$\mathbb{P}\{\tau > t, N_{t-1} < j\} \leqslant \frac{B^{j-1}}{\delta^{-t}(1)} \int W \, \mathrm{d}\pi.$$

*Proof.* Fix $t \in \mathbb{N}$ and $j \leqslant t$. Since $B \geqslant 1$, we have

$$\mathbb{P}\{\tau > t,\ N_{t-1} < j\} = \mathbb{P}\{\tau > t,\ N_{t-1} \leqslant j-1\} = \mathbb{P}\big\{\tau > t,\ B^{-N_{t-1}} \geqslant B^{-(j-1)}\big\}.$$

On $\tau > t$ we have $B^{-N_{t-1}} \delta^{-t}[W(X_t, X'_t)] = M_t$, so the final term in the previous expression is dominated by

$$\begin{aligned}
\mathbb{P}\big\{\tau > t,\ M_t \geqslant B^{-(j-1)} \delta^{-t}\big[W\big(X_t, X'_t\big)\big]\big\} &\leqslant \mathbb{P}\big\{\tau > t,\ M_t \geqslant B^{-(j-1)} \delta^{-t}(1)\big\} \\
&\leqslant \mathbb{P}\big\{M_t \geqslant B^{-(j-1)} \delta^{-t}(1)\big\} \\
&\leqslant \frac{B^{j-1}}{\delta^{-t}(1)} \mathbb{E}[M_t]
\end{aligned}$$

Here the first inequality is by $W \geqslant 1$ and the last is by Markov's inequality. Collecting terms and using the supermartingale property, we obtain

$$\mathbb{P}\{\tau > t,\ N_{t-1} < j\} \leqslant \frac{B^{j-1}}{\delta^{-t}(1)} \mathbb{E}[M_0].$$

Since $\mathbb{E}[M_0] = \int W \, \mathrm{d}\pi$, the claim in Lemma 4 is proved. □

**Lemma 5.** *If $j, t \in \mathbb{N}$ with $j \leqslant t$ then*

$$\mathbb{P}\{\tau > t,\ N_{t-1} \geqslant j\} \leqslant (1-\varepsilon)^j. \tag{14}$$

*Proof.* Fix $j, t \in \mathbb{N}$ with $j \leqslant t$. Let $(J_i)_{i \geqslant 1}$ be 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

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

Note that $N_{t-1} > j$ implies $J_j < t-1$. As a result,

$$\mathbb{P}\{\tau > t,\ N_{t-1} > j\} \leqslant \mathbb{P}\{\tau > t,\ J_j + 1 < t\}. \tag{15}$$

https://doi.org/10.1017/jpr.2025.10055 Published online by Cambridge University Press

Fix $i \leqslant j$ and consider all paths in the set $\{\tau > t,\ J_j + 1 < t\}$. Since $i \leqslant j \leqslant t < \tau$, we have $J_i + 1 \leqslant J_j + 1 < t < \tau$ and hence $X_{J_i+1} \not\preceq X'_{J_i+1}$. Therefore,

$$\mathbb{P}\{\tau > t,\ J_j + 1 < t\} \leqslant \mathbb{P} \cap_{i=1}^{j} \{X_{J_i+1} \not\preceq X'_{J_i+1}\}. \tag{16}$$

Observe that, with $L_i := \mathbb{1}\{X_{J_i+1} \not\preceq X'_{J_i+1}\}$, we have

$$\mathbb{P} \cap_{i=1}^{j} \{X_{J_i+1} \not\preceq X'_{J_i+1}\} = \mathbb{E} \prod_{i=1}^{j} L_i = \mathbb{E}\left[\prod_{i=1}^{j-1} L_i \cdot \mathbb{E}\left[L_j \mid \mathscr{F}_{J_j}\right]\right].$$

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 (5)), we find that

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

Applying the definition of $\varepsilon$ in (9), we obtain $\mathbb{P}[X_{J_j+1} \not\preceq X'_{J_j+1} \mid \mathscr{F}_{J_j}] \leqslant 1 - \varepsilon$, so

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

Iteration backwards in this way yields $\mathbb{P} \cap_{i=1}^{j} \{X_{J_i+1} \not\preceq X'_{J_i+1}\} \leqslant (1-\varepsilon)^j$. Combining this inequality with (15) and (16), we verify (14). □

We now complete the proof of Theorem 1. The proof uses an order-theoretic version of a standard total variation coupling argument [18, 24].

*Proof of Theorem* 1. We continue to take $((X_t, X'_t))_{t \geqslant 0}$, the stopping time $\tau$, and the process $(N_t)_{t \geqslant 0}$ as defined at the start of Section 4. In addition, we specialize to the case where the initial distribution $\pi$ has the form $\mu \times \mu'$ for fixed $\mu, \mu' \in p\mathcal{B}$. Fix $h$ in $ib\mathcal{B}$ with $0 \leqslant h \leqslant 1$. Since $((X_t, X'_t))_{t \geqslant 0}$ is Markov-$(\hat{Q}, \mu \times \mu')$ and $\hat{Q}_{(x,x')}$ is a coupling of $Q_x$ and $Q_{x'}$, the random element $X_t$ has distribution $\mu Q^t$ and $X'_t$ has distribution $\mu' Q^t$. As a result,

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

Since $h$ is increasing and $0 \leqslant h \leqslant 1$, the previous expression leads to

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

Application of (11) produces

$$(\mu Q^t)(h) - (\mu' Q^t)(h) \leqslant \mathbb{P}\{\tau > t\} \quad \text{for all } t \geqslant 0. \tag{17}$$

Fixing $j \in \mathbb{N}$ with $j \leqslant t$, we decompose the right-hand side of (17) to get

$$\mathbb{P}\{\tau > t\} = \mathbb{P}\{\tau > t, N_{t-1} < j\} + \mathbb{P}\{\tau > t, N_{t-1} \geqslant j\}.$$

Use of Lemmas 4 and 5 allows us to obtain

$$\mathbb{P}\{\tau > t\} \leqslant (1-\varepsilon)^j + \frac{B^{j-1}}{\delta^{-t}(1)}(\mu \times \mu')(W). \tag{18}$$

https://doi.org/10.1017/jpr.2025.10055 Published online by Cambridge University Press

Combining (17) and (18), we obtain

$$\sup_{I \in i\mathcal{B}} \left\{ (\mu Q^t)(I) - (\mu' Q^t)(I) \right\} \leqslant (1-\varepsilon)^j + \frac{B^{j-1}}{\delta^{-t}(1)} (\mu \times \mu')(W). \tag{19}$$

Reversing the roles of $\mu$ and $\mu'$, we obtain

$$\sup_{I \in i\mathcal{B}} \left\{ (\mu' Q^t)(I) - (\mu Q^t)(I) \right\} \leqslant (1-\varepsilon)^j + \frac{B^{j-1}}{\delta^{-t}(1)} (\mu' \times \mu)(W). \tag{20}$$

Adding the last two inequalities and using the definition of $\rho$ in (2), we generate the bound in Theorem 1. □

## 5. Related Convergence Results

In this section we discuss connections between Theorem 1 and convergence results in other metrics.

### 5.1. Connection to total variation results

One interesting special case of Theorem 1 is obtained by setting $\preceq$ to the identity order, so that $x \preceq y$ if and only if $x = y$. For this order we have $ib\mathcal{B} = b\mathcal{B}$, so every transition kernel is increasing, and, moreover, the total ordered variation distance becomes the total variation distance. In this setting, Theorem 1 becomes a version of well-known geometric bounds for total variation distance, such as Theorem 1 in [24].

In the total variation setting, $\varepsilon$ in (9) is at least as large as the analogous term $\varepsilon$ in Theorem 1 in [24]. Indeed, in [24], the value $\varepsilon$, which we now write as $\hat{\varepsilon}$ to avoid confusion, comes from an assumed minorization condition: there exists a $\nu \in p\mathcal{B}$ such that

$$\hat{\varepsilon}\, \nu(B) \leqslant Q(x, B) \quad \text{for all } B \in \mathcal{B} \text{ and } x \in C. \tag{21}$$

To compare $\hat{\varepsilon}$ with $\varepsilon$ defined in (9), suppose that this minorization condition holds and set $R(x, B) = (Q(x, B) - \hat{\varepsilon}\nu(B))/(1-\hat{\varepsilon})$. Fixing $(x, x') \in C \times C$, we draw $(X, X')$ as follows: With probability $\hat{\varepsilon}$, we draw $X \sim \nu$ and set $X' = X$. With probability $1 - \hat{\varepsilon}$, we 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, we obtain $\hat{\varepsilon} \leqslant \mathbb{P}\{X = X'\} = \mathbb{P}\{X \preceq X'\} \leqslant \alpha(Q(x, \cdot), Q(x', \cdot))$. Since, in this discussion, the point (x,x') was arbitrarily chosen from $C \times C$, we conclude that $\hat{\varepsilon} \leqslant \varepsilon$, where $\varepsilon$ is as defined in (9).

### 5.2. Connection to Wasserstein bounds

Theorem 1 is also connected to research on convergence rates for distributions in Wasserstein distance. To see this, recall that if $d$ is a metric on $\mathbb{X}$ then the induced Wasserstein distance between probability measures $\mu$ and $\nu$ is expressed as

$$W_d(\mu, \nu) := \inf_{\pi \in \mathscr{C}(\mu, \nu)} \int d(x, x') \pi(\mathrm{d}x, \mathrm{d}x'). \tag{22}$$

To connect this distance to the total ordered variation metric, we use Theorem 3.1 in [16] to write

$$\sup_{I \in i\mathcal{B}} (\mu(I) - \nu(I)) = \inf_{\pi \in \mathscr{C}(\mu, \nu)} \int \mathbb{1}\{x \not\preceq x'\} \pi(\mathrm{d}x, \mathrm{d}x').$$

https://doi.org/10.1017/jpr.2025.10055 Published online by Cambridge University Press

Thus, if

$$s(x, x') = \mathbb{1}\{x \not\preceq x'\} \quad \text{and} \quad W_s(\lambda, \kappa) = \inf_{\pi \in \mathscr{C}(\lambda,\kappa)} \int s(x, x')\pi(\mathrm{d}x, \mathrm{d}x'), \tag{23}$$

then, by the definition of $\rho$ in (2), we have

$$\rho(\mu, \nu) = W_s(\mu, \nu) + W_s(\nu, \mu). \tag{24}$$

We can understand $s$ as a 'directed semimetric' that fails symmetry and positive definiteness but obeys $s(x, x) = 0$ and the triangle inequality. The 'directed Wasserstein semimetric' $W_s$ inherits these properties. The sum of this directed Wasserstein semimetric and its reversed deviation creates a metric, as in (24). The inequalities (19) and (20), which we combined to prove Theorem 1, are just bounds on $W_s(\mu, \nu)$ and $W_s(\nu, \mu)$. For example, (19) tells us that

$$W_s(\mu Q^t, \mu' Q^t) \leqslant (1 - \varepsilon)^j + \frac{B^{j-1}}{\delta^{-t}(1)}(\mu \times \mu')(W).$$

The discussion above helps us understand the relationship between the order-theoretic mixing condition used in this article and the Wasserstein distance mixing condition in [4]. In the latter, the notion of $d$-small sets is introduced in order to study Wasserstein distance convergence rates for distributions: for transition kernel $Q$, a Borel set $C$ is called *d-small* if there exists an $\varepsilon > 0$ such that $W_d(Q_x, Q_{x'}) \leqslant (1 - \varepsilon)d(x, x')$ for all $(x, x') \in C \times C$. Here $d$ is an arbitrary ground metric on $\mathbb{X}$ and $W_d$ is defined as in (22). By analogy, we replace $d$ with $s$ from (23) and call $C$ *s-small* if there exists an $\varepsilon > 0$ such that

$$W_s(Q_x, Q_{x'}) \leqslant (1 - \varepsilon)s(x, x') \quad \text{for all } (x, x') \in C \times C.$$

Fixing $x, x' \in C$ and using the definition of $s$, we can equivalently write this as

$$\inf_{\pi} \pi(\mathbb{G}^c) \leqslant (1 - \varepsilon)\mathbb{1}\{x \not\preceq x'\}, \tag{25}$$

where $\mathbb{G}$ is as defined in (1) and the infimum is over all $\pi \in \mathscr{C}(Q_x, Q_{x'})$. Rearranging and using the definition of $\alpha$ in (3), we can also write (25) as

$$\alpha(Q_x, Q_{x'}) \geqslant \mathbb{1}\{x \preceq x'\} + \varepsilon\mathbb{1}\{x \not\preceq x'\} \quad \text{for all } (x, x') \in C \times C. \tag{26}$$

When $Q$ is increasing, as required in Theorem 1, we can use (4) to obtain $\alpha(Q_x, Q_{x'}) = 1$ whenever $x \preceq x'$. In this case, (26) is equivalent to $\alpha(Q_x, Q_{x'}) \geqslant \varepsilon$ whenever $(x, x') \in C \times C$. Thus, the requirement that $C$ is $s$-small is equivalent to the condition that we can extract a positive $\varepsilon$ in (9).

## 6. Examples and Applications

In this section we discuss several examples, focusing in particular on how to obtain an estimate of the value $\varepsilon$ in (9).

### 6.1. Stochastic recursive sequences

The preceding section showed that Theorem 1 reduces to existing results for bounds on total variation distance when the partial order $\preceq$ is the identity order. Next we illustrate how Theorem 1 can lead to new results in other settings. To this end, consider the process

$$X_{t+1} = F(X_t, \xi_{t+1}) \tag{27}$$

https://doi.org/10.1017/jpr.2025.10055 Published online by Cambridge University Press

where $(\xi_t)_{t\geqslant 1}$ is an independent and identically distributed (i.i.d.) shock process taking values in some space $\mathbb{Y}$, and $F$ is a measurable function from $\mathbb{X}\times\mathbb{Y}$ to $\mathbb{X}$. The common distribution of each $\xi_t$ is denoted by $\varphi$. We suppose that $F$ is increasing, in the sense that $x\preceq x'$ implies $F(x, y)\preceq F(x', y)$ for any fixed $y\in\mathbb{Y}$. We let $Q$ represent the transition kernel corresponding to (27), so that $Q(x, B)=\varphi\{y\in\mathbb{Y} : F(x, y)\in B\}$ for all $x\in\mathbb{X}$ and $B\in\mathcal{B}$. Since $F$ is increasing, the kernel $Q$ is increasing. Hence, Theorem 1 applies. We can obtain a lower bound on $\varepsilon$ in (9) by calculating

$$e := \inf\left\{\int\int \mathbb{1}\{F(x', y')\preceq F(x, y)\}\varphi(\,\mathrm{d}y)\varphi(\,\mathrm{d}y') : (x, x')\in C\times C\right\}. \tag{28}$$

To see this, fix $(x, x')\in C\times C$ and let $\xi$ and $\xi'$ be drawn independently from $\varphi$. Since $X=F(x, \xi)$ is a draw from $Q(x, \cdot)$ and $X'=F(x', \xi)$ is a draw from $Q(x', \cdot)$, we have $e\leqslant\mathbb{P}\{X'\preceq X\}\leqslant\alpha(Q(x, \cdot), Q(x', \cdot))$. As this inequality holds for all $(x, x')\in C\times C$, we obtain $e\leqslant\varepsilon$.

To illustrate how these calculations can be used, consider the TCP window size process (see, e.g., [2]) with embedded jump chain $X_{t+1}=a(X_t^2+2E_{t+1})^{1/2}$. Here $a\in(0, 1)$ and $(E_t)$ is i.i.d. exponential with unit rate. If $C=[0, c]$, then drawing $E$ and $E$' as independent standard exponentials and using (28), we obtain

$$e=\inf_{0\leqslant x,y\leqslant c}\mathbb{P}\left\{a(y^2+2E')^{1/2}\leqslant a(x^2+2E)^{1/2}\right\}=\mathbb{P}\{c^2+2E'\leqslant 2E\}.$$

Since $E'-E$ has the Laplace-(0,1) distribution, we can use $e\leqslant\varepsilon$ to get

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

### 6.2. Example: when minorization fails

We provide an elementary scenario where Theorem 1 provides a usable bound, while the minorization based methods described in Section 5.1 do not. Let $\mathbb{Q}$ be the rational numbers, let $\mathbb{X}=\mathbb{R}$, and assume that

$$X_{t+1}=\frac{X_t}{2}+\xi_{t+1},\quad \text{where } \xi_t \text{ is i.i.d. on } \{0, 1\} \text{ and } \mathbb{P}\{\xi_t=0\}=1/2.$$

Let $C$ contain at least one rational number and one irrational number. Let $\mu$ be a measure on the Borel sets of $\mathbb{R}$ obeying $\mu(B)\leqslant Q(x, B)=\mathbb{P}\{x/2+\xi\in B\}$ for all $x\in C$ and Borel sets $B$. If $x$ is rational then $x/2+\xi\in\mathbb{Q}$ with probability 1, so $\mu(\mathbb{Q}^c)\leqslant Q(x, \mathbb{Q}^c)=0$. Similarly, if $x$ is irrational then $x/2+\xi\in\mathbb{Q}^c$ with probability 1, so $\mu(\mathbb{Q})\leqslant Q(x, \mathbb{Q})=0$. Hence, $\mu$ is the zero measure on $\mathbb{R}$. Thus, we cannot take a $\hat{\varepsilon}>0$ and probability measure $\nu$ obeying the minorization condition (21). On the other hand, if we let $C=[0, 1]$, the value $e$ from (28) obeys $e=\mathbb{P}\{1/2+\xi\leqslant\xi'\}=\mathbb{P}\{\xi'-\xi\geqslant 1/2\}=\frac{1}{4}$. Since $e\leqslant\varepsilon$ (see the discussion after (28)), the constant $\varepsilon$ in (9) is positive.

### 6.3. 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

$$X_{t+1}=\eta_{t+1}\,G(X_t)+\xi_{t+1},\quad (\eta_t)\overset{\text{i.i.d.}}{\sim}\varphi,\quad (\xi_t)\overset{\text{i.i.d.}}{\sim}\psi. \tag{29}$$

Here $(X_t)$ is some measure of household wealth, $G$ is a function from $\mathbb{R}_+$ to itself, and $(\eta_t)$ and $(\xi_t)$ are independent $\mathbb{R}_+$-valued sequences. The function $G$ is increasing, since greater current

https://doi.org/10.1017/jpr.2025.10055 Published online by Cambridge University Press

wealth relaxes borrowing constraints and increases financial income. We assume that there exists a $\kappa < 1$ such that $\mathbb{E}\,\eta_t G(x) \leqslant \kappa x$ for all $x \in \mathbb{R}_+$, and, in addition, that $\bar{\xi} := \mathbb{E}\xi_t < \infty$.

Let $Q$ be the transition kernel corresponding to (29). With $V(x) = x$, we have

$$QV(x) = \mathbb{E}[\eta_{t+1}\, G(x) + \xi_{t+1}] \leqslant \kappa x + \bar{\xi} = \kappa V(x) + \bar{\xi}.$$

Fixing $d \in \mathbb{R}_+$ and setting $C = \{V \leqslant d\} = [0, d]$, we can obtain $e$ in (28) via

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

This term, which provides a lower bound for $\varepsilon$, will be strictly positive under suitable conditions, such as when $\psi$ has a sufficiently large support. By the discussion in Section 3.3, the drift condition (10) holds with $W(x, x') = 1 + V(x) + V(x')$ and $\gamma$ set to $(1 + \kappa d + 2\bar{\xi})/(1 + d)$. The function $W$ is bounded above by $2d + 1$ on $C \times C$, so we can set $B = 2d + 1$. With $\gamma$ and $B$ so defined, the bound in Corollary 1 is valid.

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

## 7. Conclusion

We exploited monotonicity properties of certain discrete-time Markov models to provide quantitative bounds on total ordered variation distance between distributions over time. There are several avenues for future research on these topics. One would be to extend the results to continuous-time Markov processes. Another would be to investigate the connection between the conditions listed here and sample path results, such as the central limit theorem. A third would be to attempt to reframe, prove, and generalize our results using a variation on the analytical arguments in, say, [4, 6, 12]. This third avenue seems promising because, at least in the Polish space setting, the total ordered variation metric used in this article is complete [16].

## Acknowledgements

We are very grateful to the two referees for many detailed and insightful comments and suggestions.

## Funding information

There are no funding bodies to thank relating to the creation of this article.

## Competing interests

There are no competing interests to declare that arose during the preparation of or the publication process for this article.

## References

[1] Antunes, A. and Cavalcanti, T. (2007). Start up costs, limited enforcement, and the hidden economy. *European Economic Review* **51**, 203–224.

[2] Bardet, J.-B., Christen, A., Guillin, A., Malrieu, F. and Zitt, P.-A. (2013). Total variation estimates for the TCP process. *Electron. J. Probab* **18**, 1–21.

[3] Bhattacharya, R. N. and Lee, O. (1988). Asymptotics of a class of Markov processes which are not in general irreducible. *The Annals of Probability* 1333–1347.

[4] Butkovsky, O. (2014). Subgeometric rates of convergence of Markov processes in the Wasserstein metric. *Annals of Applied Probability* **24**, 526–552.

https://doi.org/10.1017/jpr.2025.10055 Published online by Cambridge University Press

[5] BUTKOVSKY, O. AND SCHEUTZOW, M. (2020). Couplings via comparison principle and exponential ergodicity of SPDEs in the hypoelliptic setting. *Communications in Mathematical Physics* **379**, 1001–1034.
[6] CAÑIZO, J. A. AND MISCHLER, S. (2023). Harris-type results on geometric and subgeometric convergence to equilibrium for stochastic semigroups. *Journal of Functional Analysis* **284**, 109830.
[7] CHAFAÏ, D., MALRIEU, F. AND PAROUX, K. (2010). On the long time behavior of the TCP window size process. *Stochastic Processes and their Applications* **120**, 1518–1534.
[8] CONNOR, S. B. AND FORT, G. (2009). State-dependent Foster–Lyapunov criteria for subgeometric convergence of Markov chains. *Stochastic Processes and their Applications* **119**, 4176–4193.
[9] FOSS, S. AND SCHEUTZOW, M. (2024). *Compressibility and stochastic stability of monotone Markov chains*. *arXiv preprint* arXiv:2403.15259v1.
[10] FOSS, S., SHNEER, V., THOMAS, J. P. AND WORRALL, T. (2018). Stochastic stability of monotone economies in regenerative environments. *Journal of Economic Theory* **173**, 334–360.
[11] GAUDIO, J., AMIN, S. AND JAILLET, P. (2018). Exponential convergence rates for stochastically ordered Markov processes with random initial conditions. *arXiv preprint* arXiv:1810.07732v1.
[12] HAIRER, M. AND MATTINGLY, J. C. (2011). Yet another look at Harris' ergodic theorem for Markov chains. In *Seminar on Stochastic Analysis, Random Fields and Applications VI: Centro Stefano Franscini, Ascona, May 2008*. Springer. pp. 109–117.
[13] JIANG, Y. H., LIU, T., LOU, Z., ROSENTHAL, J. S., SHANGGUAN, S., WANG, F. AND WU, Z. (2021). The coupling/minorization/drift approach to Markov chain convergence rates. *Notices of the American Mathematical Society* **68**.
[14] KAMAE, T. AND KRENGEL, U. (1978). Stochastic partial ordering. *The Annals of Probability* **6**, 1044–1049.
[15] KAMIHIGASHI, T. AND STACHURSKI, J. (2014). Stochastic stability in monotone economies. *Theoretical Economics* **9**, 383–407.
[16] KAMIHIGASHI, T. AND STACHURSKI, J. (2019). A unified stability theory for classical and monotone Markov chains. *Journal of Applied Probability* **56**, 1–22.
[17] KAMIHIGASHI, T. AND STACHURSKI, J. (2020). Partial stochastic dominance via optimal transport. *Operations Research Letters* **48**, 584–586.
[18] LINDVALL, T. (2002). *Lectures on the Coupling Method*. Dover.
[19] LUND, R. B., MEYN, S. P. AND TWEEDIE, R. L. (1996). Computable exponential convergence rates for stochastically ordered Markov processes. *The Annals of Applied Probability* **6**, 218–237.
[20] MONTENEGRO, R. AND TETALI, P. (2006). Mathematical aspects of mixing times in Markov chains. *Foundations and Trends in Theoretical Computer Science* **1**, 237–354.
[21] QIN, Q. AND HOBERT, J. P. (2022). Geometric convergence bounds for Markov chains in Wasserstein distance based on generalized drift and contraction conditions. **58**, 872–889.
[22] QU, Y., BLANCHET, J. AND GLYNN, P. (2025). Computable bounds on convergence of Markov chains in Wasserstein distance via contractive drift.
[23] ROSENTHAL, J. S. (1995). Minorization conditions and convergence rates for Markov chain Monte Carlo. *Journal of the American Statistical Association* **90**, 558–566.
[24] ROSENTHAL, J. S. (2002). Quantitative convergence rates of Markov chains: a simple account. *Electronic Communications in Probability* **7**, 123–128.
[25] ROSENTHAL, J. S. (2023). How Markov's little idea transformed statistics. *Handbook of the History and Philosophy of Mathematical Practice* 1–11.
[26] STRASSEN, V. (1965). The existence of probability measures with given marginals. *The Annals of Mathematical Statistics* **36**, 423–439.
[27] ZHANG, S. (2000). Existence and application of optimal Markovian coupling with respect to non-negative lower semi-continuous functions. *Acta Mathematica Sinica* **16**, 261–270.

https://doi.org/10.1017/jpr.2025.10055 Published online by Cambridge University Press