Discussion Paper Series

RIEB

Kobe University

DP2017-02

Some Unified Results for Classical and Monotone Markov Chain Theory

Takashi KAMIHIGASHI

John STACHURSKI

March 6, 2017

Research Institute for Economics and Business Administration

Kobe University

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

SOME UNIFIED RESULTS FOR CLASSICAL AND MONOTONE MARKOV CHAIN THEORY

TAKASHI KAMIHIGASHI AND JOHN STACHURSKI

ABSTRACT. This paper expedites integration of two strands of the literature on stability of Markov chains: conventional, total variation based results and more recent order-theoretic results. First we introduce a complete metric on the set of Borel probability measures based on "partial" stochastic dominance. We then show that many conventional results framed in the setting of total variation distance have natural generalizations to the partially ordered setting when this metric is adopted. The conventional results can be recovered as a special case.

Keywords: Total variation, Markov chains, stochastic domination, coupling

1. INTRODUCTION

Many classical results in Markov chain theory are based on fundamental connections between total variation distance, Markov chains and couplings, as illuminated by Wolfgang Doeblin [4, 5, 6] and many subsequent authors (for overviews see, e.g., [7, 32, 18, 23, 27, 21]). One foundation stone for this theory is the simple coupling inequality

$$\|\mu - \nu\| \leqslant 2\mathbb{P}\{X \neq Y\}, \tag{1}$$

where μ and ν are probability measures, $\|\cdot\|$ is total variation distance and X and Y are random elements with distributions μ and ν respectively. This inequality can be applied directly bound the distance between the time t distributions μ_t and ν_t of Markov chains $\{X_t\}$ and $\{Y_t\}$ with common laws of motion. Moreover, Doeblin showed that (1) can easily be improved to the more significant bound

$$\|\mu_t - \nu_t\| \leqslant 2\,\mathbb{P}\{X_j \neq Y_j \text{ for any } j \leqslant t\}. \tag{2}$$

Date: March 4, 2017.

This paper was written in part while the second author was visiting RIEB at Kobe University. We have benefited from financial support from the Japan Society for the Promotion of Science and Australian Research Council Discovery Grant DP120100321.

(See, e.g., [4, 33, 22, 18, 32].) Even when the state space is uncountable, the right hand side of (2) can often be shown to converge to zero by manipulating the joint distribution of (X_j, Y_j) to increase the chance of a meeting. The work of Deoblin was built upon by Doob [8], Harris [13], Orey [24], Pitman [25] and many other authors.

While these and related convergence results based around total variation distance (see, e.g., chapters 13–16 of [21]) have clear and enduring importance, there are fields and applications where their assumptions feel excessive. To give an example from the field of economics, consider a model of household wealth with state space $\mathbb{R}_+$ (see, e.g., [28]). On one hand, bounding the right hand side of (2) usually requires mixing assumptions framed in terms of irreducibility that have no natural economic interpretation. On the other hand, for this kind of application, it *is* natural to consider mixing conditions that are framed in terms of order. For example, do initially poor households become richer than initially rich households over a given time horizon with positive probability? Is the converse also true?

Recognizing the value of order theoretic treatments for applications such as the one described above, many authors have studied the stability of Markov chains and Markov processes using order theoretic methods. Examples include [9, 34, 1, 19, 19, 15, 30]. It should be noted that none of these results deliver total variation convergence, and indeed their assumptions are insufficient to do so. Rather, they deliver weaker modes of convergence.

In this paper we develop order theoretic results that recover several existing findings from both the conventional and monotone Markov chain literature as special cases, as well as generating several new implications. The starting point is introduction of what is shown to be a complete metric γ on the set of Borel probability measures. This metric can be thought of as a natural extension of total variation distance. We then show that many fundamental concepts from conventional Markov chain theory using total variation distance and coupling have direct generalizations to the partially ordered setting when this new metric is adopted.

To give one example, when the Markov kernel is monotone, we can construct versions of $\{X_t\}$ and $\{Y_t\}$ such that

$$\gamma(\mu_t, \nu_t) \leqslant \mathbb{P}\{X_j \npreceq Y_j \text{ for any } j \leqslant t\} + \mathbb{P}\{Y_j \npreceq X_j \text{ for any } j \leqslant t\}. \tag{3}$$

Here γ is the metric discussed above and $\preceq$ is a partial order on the state space. The right hand side of (3) is no larger than the right hand side of (2) and can be much smaller. To give some feeling for how (3) can be used in applications, consider again a model of household wealth, where $\{X_t\}$ and $\{Y_t\}$ track the wealth of two households taking values in $\mathbb{R}_+$. Let $\preceq$ be the usual order $\leqslant$. Intuitively, if dynamics are such that initially rich households will become poorer than initially poor households with a sufficiently long sequence of bad luck, then the right hand side of (3) will converge to zero in t.

One interesting facet of the results in this paper is that, in addition to encompassing some of the more familiar results in monotone Markov chain theory, they also encompass significant elements of the traditional theory. In every result we present, a standard result from the classical theory can be recovered by setting the partial order to equality. For example, if we take $\preceq$ to be equality ($x \preceq y$ iff $x = y$), then γ reduces to ordinary total variation (see below) and (3) reduces to (2).

After preliminaries, we begin with a discussion of "ordered" affinity, which generalizes the usual notion of affinity for measures. The concept of ordered affinity is then used to define the total ordered variation metric. Throughout the paper, longer proofs are deferred to the appendix. The conclusion contains many suggestions for future work.

2. Preliminaries

Let S be a Polish (i.e., separable and completely metrizable) space, let $\mathcal{O}$ be the open sets, let $\mathcal{C}$ be the closed sets and let $\mathcal{B}$ be the Borel sets. Let $\mathscr{M}_s$ denote the set of all finite signed measures on $(S, \mathcal{B})$. In other words, $\mathscr{M}_s$ is all countably additive set functions from $\mathcal{B}$ to $\mathbb{R}$. Let $\mathscr{M}$ and $\mathscr{P}$ be the finite measures and probability measures in $\mathscr{M}_s$ respectively. If κ and λ are in $\mathscr{M}_s$, then $\kappa \leqslant \lambda$ means that $\kappa(B) \leqslant \lambda(B)$ for all $B \in \mathcal{B}$.

Let bS be the set of all bounded $\mathcal{B}$-measurable functions from S into $\mathbb{R}$. If $h \in bS$ and $\lambda \in \mathscr{M}_s$, then $\lambda(h) := \int h \, d\lambda$. The total variation norm of $\lambda \in \mathscr{M}_s$ is

$$\|\lambda\| := \sup_{h \in H} |\lambda(h)|.$$

For f and g in bS, the statement $f \leqslant g$ means that $f(x) \leqslant g(x)$ for all $x \in S$. Let

$$H := \{h \in bS : -1 \leqslant h \leqslant 1\} \quad \text{and} \quad H_0 := \{h \in bS : 0 \leqslant h \leqslant 1\}.$$

Given μ and ν in $\mathscr{P}$, a random element (X, Y) taking values in $S \times S$ and defined on a common probability space $(\Omega, \mathscr{F}, \mathbb{P})$ is called a coupling of (μ, ν) if $\mu = \mathbb{P} \circ X^{-1}$ and $\nu = \mathbb{P} \circ Y^{-1}$ (i.e., if the distribution of (X, Y) has marginals μ and ν respectively—see, e.g., [18] or [32]). The set of all couplings of (μ, ν) is denoted below by $\mathscr{C}(\mu, \nu)$.

A sequence $\{\mu_n\} \subset \mathscr{P}$ converges to $\mu \in \mathscr{P}$ weakly if $\mu_n(h) \to \mu(h)$ as $n \to \infty$ for all continuous $h \in bS$. In this case we write $\mu_n \stackrel{w}{\to} \mu$.

Given μ and $\nu \in \mathscr{M}$, their measure theoretic infimum $\mu \wedge \nu$ is the largest element of $\mathscr{M}$ dominated by both μ and ν. It can be defined by taking f and g to be densities of μ and ν respectively under the dominating measure $\lambda := \mu + \nu$ and defining $\mu \wedge \nu$ by $(\mu \wedge \nu)(B) := \int_B \min\{f(x), g(x)\}\lambda(\mathrm{d}x)$ for all $B \in \mathcal{B}$. The total variation distance between μ and ν is related to $\mu \wedge \nu$ via $\|\mu - \nu\| = \|\mu\| + \|\nu\| - 2\|\mu \wedge \nu\|$. See, for example, [26]. For probability measures we also have

$$\sup_{B \in \mathcal{B}}\{\mu(B) - \nu(B)\} = \sup_{B \in \mathcal{B}} |\mu(B) - \nu(B)| = \|\mu - \nu\|/2. \tag{4}$$

The affinity between two measures μ, ν in $\mathscr{M}$ is the value

$$\alpha(\mu, \nu) := (\mu \wedge \nu)(S).$$

The following properties are elementary:

Lemma 2.1. *For all $(\mu, \nu) \in \mathscr{M} \times \mathscr{M}$ we have*

(a) $0 \leqslant \alpha(\mu, \nu) \leqslant \min\{\mu(S), \nu(S)\}$
(b) $\alpha(\mu, \nu) = \mu(S) = \nu(S)$ *if and only if* $\mu = \nu$.
(c) $\alpha(c\mu, c\nu) = c\alpha(\mu, \nu)$ *for all* $c \geqslant 0$.

There are several other common representations of affinity. For example, when μ and ν are both probability measures, we have

$$\alpha(\mu, \nu) = 1 - \sup_{B \in \mathcal{B}} |\mu(B) - \nu(B)| = \max_{(X,Y) \in \mathscr{C}(\mu, \nu)} \mathbb{P}\{X = Y\}. \tag{5}$$

(See, e.g., [26, 18].) The second equality in (5) states that, if $(X, Y) \in \mathscr{C}(\mu, \nu)$, then $\mathbb{P}\{X = Y\} \leqslant \alpha(\mu, \nu)$, and, moreover, there exists a $(X, Y) \in \mathscr{C}(\mu, \nu)$ such that equality is attained. Any such coupling is called a *maximal* or *gamma* coupling. See theorem 5.2 of [18]. From (4) and (5) we obtain

$$\|\mu - \nu\| = 2(1 - \alpha(\mu, \nu)). \tag{6}$$

3. Ordered Affinity

We next introduce a generalization of affinity when S has a partial order. We investigate its properties in detail, since both our metric and the stability theory presented below rely on this concept.

3.1. Preliminaries.

As before, let S be a Polish space. A closed partial order $\preceq$ on S is a partial order $\preceq$ such that its graph

$$\mathbb{G} := \{(x, y) \in S \times S : x \preceq y\}$$

is closed in the product topology. In the sequel, a *partially ordered Polish space* is any such pair $(S, \preceq)$, where S is nonempty and Polish, and $\preceq$ is a closed partial order on S. When no confusion arises, we denote it simply by S.

For such a space S, we call $I \subset S$ *increasing* if $x \in I$ and $x \preceq y$ implies $y \in I$. We call $h \colon S \to \mathbb{R}$ *increasing* if $x \preceq y$ implies $h(x) \leqslant h(y)$. We let $i\mathcal{B}$, $i\mathcal{O}$ and $i\mathcal{C}$ denote the increasing Borel, open and closed sets respectively, while ibS is the increasing functions in bS. In addition,

- $iH := H \cap ibS = \{h \in ibS : -1 \leqslant h \leqslant 1\}$ and
- $iH_0 := H_0 \cap ibS = \{h \in ibS : 0 \leqslant h \leqslant 1\}$.

If $B \in \mathcal{B}$, then $i(B)$ is all $y \in S$ such that $x \preceq y$ for some $x \in B$, while $d(B)$ is all $y \in S$ such that $y \preceq x$ for some $x \in B$. Given μ and ν in $\mathscr{M}$, we say that μ is *stochastically dominated* by ν and write $\mu \preceq_{sd} \nu$ if $\mu(S) = \nu(S)$ and $\mu(I) \leqslant \nu(I)$ for all $I \in i\mathcal{B}$. We can equivalently say that $\mu(S) = \nu(S)$ and $\mu(h) \leqslant \nu(h)$ for all h in either iH or iH_0.

Remark 3.1. Since S is a partially ordered Polish space, for any μ, ν in $\mathscr{P}$ we have $\mu = \nu$ whenever $\mu(C) = \nu(C)$ for all $C \in i\mathcal{C}$, or, equivalently, $\mu(h) = \nu(h)$ for all continuous $h \in ibS$. See [16, lemma 1]. One implication is that $\mu \preceq_{sd} \nu$ and $\nu \preceq_{sd} \mu$ together imply $\mu = \nu$.

Remark 3.2. We consider several alternative specifications for the partial order $\preceq$ on S. One important special case is when $\preceq$ is equality, so that $x \preceq y$ if and only if $x = y$. Then $i\mathcal{B} = \mathcal{B}$, $ibS = bS$, $iH = H$, $iH_0 = H_0$ and $\mu \preceq_{sd} \nu$ if and only if $\mu = \nu$.

Lemma 3.1. *If $\lambda \in \mathscr{M}_s$, then $\sup_{I \in i\mathcal{B}} \lambda(I) = \sup_{h \in iH_0} \lambda(h)$ and*

$$\sup_{h \in iH} |\lambda(h)| = \max \left\{ \sup_{h \in iH} \lambda(h), \; \sup_{h \in iH} (-\lambda)(h) \right\}. \tag{7}$$

One can easily check that

$$\lambda \in \mathscr{M}_s \text{ and } \lambda(S) = 0 \quad \Longrightarrow \quad \sup_{h \in iH} \lambda(h) = 2 \sup_{h \in iH_0} \lambda(h). \tag{8}$$

3.2. **Definition of Ordered Affinity.** For each pair $(\mu, \nu) \in \mathscr{M} \times \mathscr{M}$, let

$$\Phi(\mu, \nu) := \{(\mu', \nu') \in \mathscr{M} \times \mathscr{M} : \mu' \leqslant \mu, \; \nu' \leqslant \nu, \; \mu' \preceq_{sd} \nu'\}.$$

We call $\Phi(\mu, \nu)$ the set of *ordered component pairs* for (μ, ν). Here "ordered" means ordered by stochastic dominance. The set of ordered component pairs is always nonempty. For example, $(\mu \wedge \nu, \mu \wedge \nu)$ is an element of $\Phi(\mu, \nu)$. In the case where $\mu \preceq_{sd} \nu$ we have $(\mu, \nu) \in \Phi(\mu, \nu)$.

We call an ordered component pair $(\mu', \nu') \in \Phi(\mu, \nu)$ a *maximal ordered component pair* if it has greater mass than all others; that is, if

$$\mu''(S) \leqslant \mu'(S) \quad \text{for all } \; (\mu'', \nu'') \in \Phi(\mu, \nu).$$

(We can restate this by replacing $\mu'(S)$ and $\mu''(S)$ with $\nu'(S)$ and $\nu''(S)$ respectively, since the mass of ordered component pairs is equal by the definition of stochastic dominance.) We let $\Phi^*(\mu, \nu)$ denote the set of maximal ordered component pairs for (μ, ν). Thus, if

$$\alpha_O(\mu, \nu) := \sup\{\mu'(S) : (\mu', \nu') \in \Phi(\mu, \nu)\}. \tag{9}$$

then

$$\Phi^*(\mu, \nu) = \{(\mu', \nu') \in \Phi(\mu, \nu) : \mu'(S) = \alpha_O(\mu, \nu)\}.$$

Using the Polish space assumption, one can show that maximal ordered component pairs always exist:

Proposition 3.1. *The set $\Phi^*(\mu, \nu)$ is nonempty for all $(\mu, \nu) \in \mathscr{M} \times \mathscr{M}$.*

Proof. Fix $(\mu, \nu) \in \mathscr{M} \times \mathscr{M}$ and let $s := \alpha_O(\mu, \nu)$. From the definition, we can take sequences $\{\mu'_n\}$ and $\{\nu'_n\}$ in $\mathscr{M}$ such that $(\mu'_n, \nu'_n) \in \Phi(\mu, \nu)$ for all $n \in \mathbb{N}$ and $\mu'_n(S) \uparrow s$. Since $\mu'_n \leqslant \mu$ and $\nu'_n \leqslant \nu$ for all $n \in \mathbb{N}$, Prohorov's theorem [10, theorem 11.5.4] implies that these sequences have convergent subsequences with $\mu'_{n_k} \xrightarrow{w} \mu'$ and $\nu'_{n_k} \xrightarrow{w} \nu'$ for some $\mu', \nu' \in \mathscr{M}$. We claim that (μ', ν') is a maximal ordered component pair.

Using $\mu'_n \leqslant \mu$ and $\nu'_n \leqslant \nu$ for all $n \in \mathbb{N}$ and theorem 1.5.5 of [14], we can strengthen weak convergence setwise convergence over $\mathcal{B}$. It follows that, for any $B \in \mathcal{B}$, we have $\mu'(B) \leqslant \mu(B)$ and $\nu'(B) \leqslant \nu(B)$. Moreover, the definition of $\Phi(\mu, \nu)$ and stochastic dominance imply that $\mu'_n(S) = \nu'_n(S)$ for all $n \in \mathbb{N}$, and therefore $\mu'(S) = \nu'(S)$. Also, for any $I \in i\mathcal{B}$, the fact that $\mu'_n(I) \leqslant \nu'_n(I)$ for all $n \in \mathbb{N}$ gives us $\mu'(I) \leqslant \nu'(I)$. Thus, $\mu' \preceq_{sd} \nu'$. Finally, $\mu'(S) = s$, since $\mu'_n(S) \uparrow s$. Hence (μ', ν') lies in $\Phi^*(\mu, \nu)$. □

The value $\alpha_O(\mu, \nu)$ defined in (9) gives the mass of the maximal ordered component pair. We call it the *ordered affinity* from μ to ν. On an intuitive level, we can think of $\alpha_O(\mu, \nu)$ as the "degree" to which μ is dominated by ν in the sense of stochastic dominance. Since the pair $(\mu \wedge \nu, \mu \wedge \nu)$ is an ordered component pair for (μ, ν), we always have

$$0 \leqslant \alpha(\mu, \nu) \leqslant \alpha_O(\mu, \nu), \tag{10}$$

where $\alpha(\mu, \nu)$ is the standard affinity defined in section 2. In fact $\alpha_O(\mu, \nu)$ generalizes the standard the notion of affinity by extending it to arbitrary partial orders, as shown in the next lemma.

Lemma 3.2. *If $\preceq$ is equality, then $\alpha_O = \alpha$ on $\mathscr{M} \times \mathscr{M}$.*

Proof. Fix $(\mu, \nu) \in \mathscr{M} \times \mathscr{M}$ and let $\preceq$ be equality ($x \preceq y$ iff $x = y$). Then $\preceq_{sd}$ is also equality, from which it follows that the supremum in (9) is attained by $\mu \wedge \nu$. Hence $\alpha_O(\mu, \nu) = \alpha(\mu, \nu)$. □

3.3. **Properties of Ordered Affinity.** Let's list some elementary properties of α_O. The following list should be compared with lemma 2.1. It shows that analogous results hold for α_O as hold for α. (Lemma 2.1 is in fact a special case of lemma 3.3 with the partial order taken to be equality.)

Lemma 3.3. *For all $(\mu, \nu) \in \mathscr{M} \times \mathscr{M}$, we have*

(a) $0 \leqslant \alpha_O(\mu, \nu) \leqslant \min\{\mu(S), \nu(S)\}$,
(b) $\alpha_O(\mu, \nu) = \mu(S) = \nu(S)$ *if and only if* $\mu \preceq_{sd} \nu$, *and*

(c) $c\alpha_O(\mu, \nu) = \alpha_O(c\mu, c\nu)$ *whenever* $c \geqslant 0$.

Proof. Fix $(\mu, \nu) \in \mathscr{M} \times \mathscr{M}$. Claim (a) follows directly from the definitions. Regarding claim (b), suppose first that $\mu \preceq_{sd} \nu$. Then $(\mu, \nu) \in \Phi(\mu, \nu)$ and hence $\alpha_O(\mu, \nu) = \mu(S)$. Conversely, if $\alpha_O(\mu, \nu) = \mu(S)$, then, since the only component $\mu' \leqslant \mu$ with $\mu'(S) = \mu(S)$ is μ itself, we must have $(\mu, \nu') \in \Phi(\mu, \nu')$ for some $\nu' \leqslant \nu$ with $\mu \preceq_{sd} \nu'$. But then $\mu(I) \leqslant \nu'(I) \leqslant \nu(I)$ for any $I \in i\mathcal{B}$. Hence $\mu \preceq_{sd} \nu$.

Claim (c) is trivial if $c = 0$, so suppose instead that $c > 0$. Fix $(\mu', \nu') \in \Phi(\mu, \nu)$ such that $\alpha_O(\mu, \nu) = \mu'(S)$. It is clear that $(c\mu', c\nu') \in \Phi(c\mu, c\nu)$, implying that

$$c\alpha_O(\mu, \nu) = c\mu'(S) \leqslant \alpha_O(c\mu, c\nu). \tag{11}$$

For reverse inequality, we can apply (11) again to get

$$\alpha_O(c\mu, c\nu) = c(1/c)\alpha_O(c\mu, c\nu) \leqslant c\alpha_O(\mu, \nu). \qquad \square$$

3.4. **Equivalent Representations.** In (5) we noted that the affinity between two measures has several alternative representations. In our setting these results generalize as follows:

Theorem 3.1. *For all* $(\mu, \nu) \in \mathscr{P} \times \mathscr{P}$, *we have*

$$\alpha_O(\mu, \nu) = 1 - \sup_{I \in i\mathcal{B}} \{\mu(I) - \nu(I)\} = \max_{(X,Y) \in \mathscr{C}(\mu,\nu)} \mathbb{P}\{X \preceq Y\}. \tag{12}$$

Evidently (5) is a special case of (12) because (12) reduces to (5) when $\preceq$ is set to equality. For example, when $\preceq$ is equality,

$$\sup_{I \in i\mathcal{B}} \{\mu(I) - \nu(I)\} = \sup_{B \in \mathcal{B}} \{\mu(B) - \nu(B)\} = \sup_{B \in \mathcal{B}} |\mu(B) - \nu(B)|.$$

where the last step is from (4). Note also that, as shown in the proof of theorem 3.1, the supremum can also be written in terms of the open increasing sets $i\mathcal{O}$ or the closed decreasing sets $d\mathcal{C}$. In particular,

$$\sup_{I \in i\mathcal{B}} \{\mu(I) - \nu(I)\} = \sup_{I \in i\mathcal{O}} \{\mu(I) - \nu(I)\} = \sup_{D \in d\mathcal{C}} \{\nu(D) - \mu(D)\}.$$

One of the assertions of theorem 3.1 is the existence of a coupling $(X, Y) \in \mathscr{C}(\mu, \nu)$ attaining $\mathbb{P}\{X \preceq Y\} = \alpha_O(\mu, \nu)$. Let us refer to any such coupling as an *order maximal* coupling for (μ, ν).

Example 3.1. Let δ_z be the dirac probability concentrated on z. For $(x, y) \in S \times S$, We have

$$\alpha_O(\delta_x, \delta_y) = \mathbb{1}\{x \preceq y\} = \mathbb{1}_{\mathbb{G}}(x, y),$$

as can easily be verified from the definition or either of the alternative representations in (12). The map $(x, y) \mapsto \mathbb{1}_{\mathbb{G}}(x, y)$ is measurable due to the Polish assumption. As a result, for any $(X, Y) \in \mathscr{C}(\mu, \nu)$ we have

$$\mathbb{E}\, \alpha_O(\delta_X, \delta_Y) = \mathbb{P}\{X \preceq Y\} \leqslant \alpha_O(\mu, \nu),$$

with equality when (X, Y) is an order maximal coupling.

4. Total Ordered Variation

Consider the function on $\mathscr{P} \times \mathscr{P}$ given by

$$\gamma(\mu, \nu) := 2 - \alpha_O(\mu, \nu) - \alpha_O(\nu, \mu). \tag{13}$$

We call $\gamma(\mu, \nu)$ the *total ordered variation distance* between μ and ν. The natural comparison is with (6), which renders the same value if α_O is replaced by α. In particular, when $\preceq$ is equality, ordered affinity reduces to affinity, and total ordered variation distance reduces to total variation distance.

Since ordered affinities dominate affinities (see (10)), we have

$$\gamma(\mu, \nu) \leqslant \|\mu - \nu\| \quad \text{for all} \quad (\mu, \nu) \in \mathscr{P} \times \mathscr{P}.$$

Other, equivalent, representations are available. For example, in view of (12), for any $(\mu, \nu) \in \mathscr{P} \times \mathscr{P}$ we have

$$\gamma(\mu, \nu) = \sup_{I \in i\mathcal{B}} (\mu - \nu)(I) + \sup_{I \in i\mathcal{B}} (\nu - \mu)(I), \tag{14}$$

By combining lemma 3.1 and (8), we also have

$$2\gamma(\mu, \nu) = \sup_{h \in iH} (\mu - \nu)(h) + \sup_{h \in iH} (\nu - \mu)(h). \tag{15}$$

It is straightforward to show that

$$\sup_{I \in i\mathcal{B}} |\mu(I) - \nu(I)| \leqslant \gamma(\mu, \nu) \quad \text{and} \quad \sup_{D \in d\mathcal{B}} |\mu(D) - \nu(D)| \leqslant \gamma(\mu, \nu). \tag{16}$$

Lemma 4.1. *The function γ is a metric on $\mathscr{P}$.*

Proof. The claim that γ is a metric follows in a straightforward way from the definition or the alternative representation (14). For example, the triangle inequality is easy to verify using (14). Also, $\gamma(\mu, \nu) = 0$ implies $\mu = \nu$ by (14) and remark 3.1. $\square$

4.1. **Connection to Other Modes of Convergence.** As well as total variation, the metric γ is closely related to the so-called *Bhattacharya metric*, which is given by

$$\beta(\mu, \nu) := \sup_{h \in iH} |\mu(h) - \nu(h)|. \tag{17}$$

See [1, 2]. (In [2] the metric is defined by taking the supremum over iH_0 rather than iH, but the two definitions differ only by a positive scalar.) The Bhattacharya metric can be thought of as an alternative way to generalize total variation distance, in the sense that, like γ, the metric β reduces to total variation distance when the partial order $\preceq$ is equality (since iH equals H in this setting). From (7) we have

$$\frac{1}{2}\left[\sup_{h \in iH} \lambda(h) + \sup_{h \in iH} (-\lambda)(h)\right] \leqslant \sup_{h \in iH} |\lambda(h)| \leqslant \sup_{h \in iH} \lambda(h) + \sup_{h \in iH} (-\lambda)(h), \tag{18}$$

and from this and (15) we have

$$\gamma(\mu, \nu) \leqslant \beta(\mu, \nu) \leqslant 2\gamma(\mu, \nu). \tag{19}$$

Hence β and γ are equivalent metrics.

The metric γ is also connected to the Wasserstein metric [11, 12]. If ρ metrizes the topology on S, then the Wasserstein distance between probability measures μ and ν is

$$w(\mu, \nu) := \inf_{(X,Y) \in \mathscr{C}(\mu,\nu)} \mathbb{E}\, \rho(X, Y).$$

The total ordered variation metric can be compared as follows. Consider the "directed semimetric" $\hat\rho(x, y) := \mathbb{1}\{x \npreceq y\}$. In view of (12) we have

$$\gamma(\mu, \nu) = \inf_{(X,Y) \in \mathscr{C}(\mu,\nu)} \mathbb{E}\, \hat\rho(X, Y) + \inf_{(X,Y) \in \mathscr{C}(\mu,\nu)} \mathbb{E}\, \hat\rho(Y, X).$$

Thus, $\gamma(\mu, \nu)$ is found by summing two partial, "directed Wasserstein deviations." Summing the two directed differences from opposite directions yields a metric.

Finally, here is one useful connection between γ and weak convergence:

Proposition 4.1. *If $\{\mu_n\} \subset \mathscr{P}$ is tight and $\gamma(\mu_n, \mu) \to 0$ for some $\mu \in \mathscr{P}$, then $\mu_n \xrightarrow{w} \mu$.*

Proof. Let $\{\mu_n\}$ and μ satisfy the conditions of the proposition. Take any subsequence of $\{\mu_n\}$ and observe that by Prohorov's theorem, this subsequence has a subsubsequence converging weakly to some $\nu \in \mathscr{P}$. Along this subsubsequence, for any continuous $h \in ibS$ we have both $\mu_n(h) \to \mu(h)$ and $\mu_n(h) \to \nu(h)$. This is sufficient for $\nu = \mu$ by remark 3.1. Thus, every subsequence of $\{\mu_n\}$ has a subsubsequence converging weakly to μ, and hence so does the entire sequence. $\square$

4.2. **Completeness.** To obtain completeness of $(\mathscr{P}, \gamma)$, we adopt the following additional assumption.

Assumption 4.1. If $K \subset S$ is compact, then $i(K) \cap d(K)$ is also compact.

Assumption 4.1 is satisfied if, say, compact sets are order bounded (i.e., lie in order intervals) and order intervals are compact. For example, $\mathbb{R}^n$ with the usual pointwise partial order has this property.

Theorem 4.1. *If assumption 4.1 holds, then $(\mathscr{P}, \gamma)$ is complete.*

Remark 4.1. In [2] it was shown that β is a complete metric when $S = \mathbb{R}^n$. Due to equivalence of the metrics, theorem 4.1 extends this result to partially ordered Polish spaces where assumption 4.1 is satisfied.

5. Applications

In this section we show that certain results in classical and monotone Markov chain theory, hitherto treated separately, can be derived from the same set of results based around total ordered variation and ordered affinity. (As in [21], we use the term "Markov chain" to refer to a stochastic process with discrete time parameter and general state space.)

5.1. **Applications of Theorem 3.1.** First we make some comments on theorem 3.1, which states the existence of an order maximal coupling for any pair of probabilities (μ, ν); that is, a $(X, Y) \in \mathscr{C}(\mu, \nu)$ such that $\mathbb{P}\{X \preceq Y\} = \alpha_O(\mu, \nu)$. The existence of an order maximal coupling implies two well-known results that are usually treated separately.

The first is the Nachbin–Strassen theorem (see, e.g., thm. 1 of [17] or ch. IV of [18]), which states the existence of a coupling $(X, Y) \in \mathscr{C}(\mu, \nu)$ attaining $\mathbb{P}\{X \preceq Y\} = 1$

whenever $\mu \preceq_{sd} \nu$. The existence of an order maximal coupling for each (μ, ν) in $\mathscr{P} \times \mathscr{P}$ implies this statement, since, under the hypothesis that $\mu \preceq_{sd} \nu$, we also have $\alpha_O(\mu, \nu) = 1$. Hence any order maximal coupling satisfies $\mathbb{P}\{X \preceq Y\} = 1$.

The other familiar result implied by existence of an order maximal coupling is is existence of a maximal coupling in the standard sense (see the discussion of maximal couplings after (5) and the result on p. 19 of [18]). Indeed, if we take $\preceq$ to be equality, then (12) reduces to (5), as already discussed.

5.2. **Markov Kernels.** Let $\{S_i\}$ be partially ordered Polish spaces over $i = 0, 1, 2, \ldots$, with Borel sets $\mathcal{B}_i$, bounded Borel measurable functions bS_i, probability measures $\mathscr{P}_i$ and so on. To simplify notation we use $\preceq$ for the partial order $\preceq_i$ on any of these spaces. On finite and infinite products of these spaces we use the product topology and partial order. For example, if (x_0, x_1) and (y_0, y_1) are points in $S_0 \times S_1$, then $(x_0, x_1) \preceq (y_0, y_1)$ means that $x_0 \preceq y_0$ and $x_1 \preceq y_1$. Note that, once again, the same symbol $\preceq$ is used for the partial order. No confusion should arise.

A function $P \colon (S_0, \mathcal{B}_1) \to [0,1]$ is called a *Markov kernel* from S_0 to S_1 if $x \mapsto P(x, B)$ is $\mathcal{B}_0$-measurable for each $B \in \mathcal{B}_1$ and $B \mapsto P(x, B)$ is in $\mathscr{P}_1$ for all $x \in S_0$. If $S_0 = S_1 = S$, we will simply call P are Markov kernel on S, or just a Markov kernel. Following standard conventions, for any Markov kernel P from S_0 to S_1, any $h \in bS_1$ and $\mu \in \mathscr{P}_0$, we define $\mu P \in \mathscr{P}_1$ and $Ph \in bS_0$ via

$$(\mu P)(B) = \int P(x, B) \mu(dx) \quad \text{and} \quad (Ph)(x) = \int h(y) P(x, dy).$$

Also, $\mu \otimes P$ denotes the joint distribution on $S_0 \times S_1$ defined by

$$(\mu \otimes P)(A \times B) = \int_A P(x, B) \mu(dx).$$

To simplify notation, we frequently use P_x to represent the measure $\delta_x P = P(x, \cdot)$. Also, P^m is the m-th composition of P with itself.

5.3. **Order Affinity and Monotone Markov Kernels.** Let S be a partially ordered Polish space. A Markov kernel P is called *monotone* if $Ph \in ibS_0$ whenever $h \in ibS_1$. An equivalent condition is that $\mu P \preceq_{sd} \nu P$ whenever $\mu \preceq_{sd} \nu$; or just $P(x, \cdot) \preceq_{sd} P(y, \cdot)$ whenever $x \preceq y$. It is well-known (see, e.g., proposition 1 of [17]) that if $\mu \preceq_{sd} \nu$ and P is monotone, then $\mu \otimes P \preceq_{sd} \nu \otimes P$. Note that, when the partial order is equality, every Markov kernel is monotone.

Lemma 5.1. *If P is a monotone Markov kernel from S_0 to S_1 and μ, μ', ν and ν' are probabilities in $\mathscr{P}_0$, then*

$$\mu' \preceq_{sd} \mu \text{ and } \nu \preceq_{sd} \nu' \quad \implies \quad \alpha_O(\mu P, \nu P) \leqslant \alpha_O(\mu' P, \nu' P).$$

Proof. Let P, μ, μ', ν and ν' have the stated properties. In view of the equivalently representation in (12), the claim will be established if

$$\sup_{I \in i\mathcal{B}} \{(\mu P)(I) - (\nu P)(I)\} \geqslant \sup_{I \in i\mathcal{B}} \{(\mu' P)(I) - (\nu' P)(I)\}.$$

This holds by the monotonicity of P and the order of μ, μ', ν and ν'. □

Lemma 5.2. *If P is a monotone Markov kernel from S_0 to S_1, then, for any μ, ν in $\mathscr{P}_0$,*

$$\alpha_O(\mu P, \nu P) \geqslant \alpha_O(\mu, \nu).$$

Proof. Fix μ, ν in $\mathscr{P}_0$ and let $(\hat{\mu}, \hat{\nu})$ be a maximal ordered component pair for (μ, ν). From monotonicity of P and the fact the Markov kernels preserve the mass of measures, it is clear that $(\hat{\mu} P, \hat{\nu} P)$ is an ordered component pair for $(\mu P, \nu P)$. Hence

$$\alpha_O(\mu P, \nu P) \geqslant (\hat{\mu} P)(S) = \hat{\mu}(S) = \alpha_O(\mu, \nu). \qquad \square$$

On the other hand, for the joint distribution, the ordered affinity of the initial pair is preserved.

Lemma 5.3. *If P is a monotone Markov kernel from S_0 to S_1, then, for any μ, ν in $\mathscr{P}_0$,*

$$\alpha_O(\mu \otimes P, \nu \otimes P) = \alpha_O(\mu, \nu).$$

Proof. Fix μ, ν in $\mathscr{P}_0$ and let (X_0, X_1) and (Y_0, Y_1) be random pairs with distributions $\mu \otimes P$ and $\nu \otimes P$ respectively. We have

$$\mathbb{P}\{(X_0, X_1) \preceq (Y_0, Y_1)\} \leqslant \mathbb{P}\{X_0 \preceq Y_0\} \leqslant \alpha_O(\mu, \nu).$$

Taking the supremum over all couplings in $\mathscr{C}(\mu \otimes P, \nu \otimes P)$ shows that $\alpha_O(\mu \otimes P, \nu \otimes P)$ is dominated by $\alpha_O(\mu, \nu)$.

To see the reverse inequality, let $(\hat{\mu}, \hat{\nu})$ be a maximal ordered component pair for (μ, ν). Monotonicity of P now gives $\hat{\mu} \otimes P \preceq_{sd} \hat{\nu} \otimes P$. Using this and the fact the Markov kernels preserve the mass of measures, we see that $(\hat{\mu} \otimes P, \hat{\nu} \otimes P)$ is an ordered component pair for $(\mu \otimes P, \nu \otimes P)$. Hence

$$\alpha_O(\mu P, \nu P) \geqslant (\hat{\mu} \otimes P)(S_0 \times S_1) = \hat{\mu}(S_0) = \alpha_O(\mu, \nu). \qquad \square$$

5.4. **Monotone Markov Chains.** Given $\mu \in \mathscr{P}$ and Markov kernel P on S, a stochastic process $\{X_t\}_{t \geqslant 0}$ taking values in $S^\infty := \times_{t=0}^{\infty} S$ will be called a Markov chain with initial distribution μ and kernel P if the distribution of $\{X_t\}$ on S^∞ is

$$\mathbf{Q}_\mu := \mu \otimes P \otimes P \otimes P \otimes \cdots$$

(The meaning of the right hand side is clarified in, e.g., §III.8 of [18], p. 903 of [17], §3.4 of [21].) If P is a monotone Markov kernel, then $(x, B) \mapsto \mathbf{Q}_x(B) := \mathbf{Q}_{\delta_x}(B)$ is a monotone Markov kernel from S to S^∞. See propositions 1 and 2 of [17].

There are various useful results about representations of Markov chains that are ordered almost surely. One is that, if the initial conditions satisfy $\mu \preceq_{sd} \nu$ and P is a monotone Markov kernel, then we can find Markov chains $\{X_t\}$ and $\{Y_t\}$ with initial distributions μ and ν and kernel P such that $X_t \preceq Y_t$ for all t almost surely. (See, e.g., theorem 2 of [17].) This result can be generalized beyond the case where μ and ν are stochastically ordered, using the results presented above. For example, let μ and ν be arbitrary initial distributions and let P be monotone, so that $\mathbf{Q}_x$ is likewise monotone. By lemma 5.3 we have

$$\alpha_O(\mathbf{Q}_\mu, \mathbf{Q}_\nu) = \alpha_O(\mu \otimes \mathbf{Q}_x, \nu \otimes \mathbf{Q}_x) = \alpha_O(\mu, \nu).$$

In other words, the ordered affinity of the entire processes is given by the ordered affinity of the initial distributions. It now follows from theorem 3.1 that there exist Markov chains $\{X_t\}$ and $\{Y_t\}$ with initial distributions μ and ν and Markov kernel P such that

$$\mathbb{P}\{X_t \preceq Y_t, \ \forall t \geqslant 0\} = \alpha_O(\mu, \nu).$$

The standard result is a special case, since $\mu \preceq_{sd} \nu$ implies $\alpha_O(\mu, \nu) = 1$, and hence the sequences are ordered almost surely.

5.5. **Nonexpansiveness.** It is well-known that every Markov kernel is nonexpansive with respect to total variation norm, so that

$$\|\mu P - \nu P\| \leqslant \|\mu - \nu\| \quad \text{for all } (\mu, \nu) \in \mathscr{P} \times \mathscr{P}. \tag{20}$$

An analogous result is true for γ when P is monotone. That is,

$$\gamma(\mu P, \nu P) \leqslant \gamma(\mu, \nu) \quad \text{for all } (\mu, \nu) \in \mathscr{P} \times \mathscr{P}. \tag{21}$$

The bound (21) follows directly from lemma 5.2. Evidently (20) be recovered from (21) by setting $\preceq$ to equality.

Nonexpansiveness is interesting partly in its own right (we apply it in proofs below) and partly because it suggests that, with some additional assumptions, we can strengthen it to contractiveness. We expand on this idea below.

5.6. **An Order Coupling Bound for Markov Chains.** We now turn to the claim that, given a monotone Markov kernel P on S and arbitrary $\mu, \nu \in \mathscr{P}$, we can construct Markov chains $\{X_t\}$ and $\{Y_t\}$ with common kernel P and respective initial conditions μ and ν such that (3) holds. In proving (3), we need only show that

$$1 - \alpha_O(\mu P^t, \nu P^t) \leqslant \mathbb{P}\{X_j \npreceq Y_j \text{ for any } j \leqslant t\}, \tag{22}$$

since, with (22) is established, we can reverse the roles of $\{X_t\}$ and $\{Y_t\}$ in (22) to obtain $1 - \alpha_O(\nu P^t, \mu P^t) \leqslant \mathbb{P}\{Y_j \npreceq X_j \text{ for any } j \leqslant t\}$ and then add this inequality to (22) to produce (3).

If $\{X_t\}$ and $\{Y_t\}$ are Markov chains with kernel P and initial conditions μ and ν, then (12) yields $\alpha_O(\mu P^t, \nu P^t) \geqslant \mathbb{P}\{X_t \preceq Y_t\}$. Therefore, we need only construct such chains with the additional property that

$$\mathbb{P}\{X_j \npreceq Y_j \text{ for any } j \leqslant t\} = \mathbb{P}\{X_t \npreceq Y_t\}. \tag{23}$$

Intuitively we can do so by using a "conditional" version of the Nachbin–Strassen theorem, producing chains that, once ordered, remain ordered almost surely. This can be formalized as follows: By [19, theorem 2.3], there exists a Markov kernel M on $S \times S$ such that $\mathbb{G}$ is absorbing for M (i.e., $M((x,y), \mathbb{G}) = 1$ for all (x,y) in $\mathbb{G}$),

$$P(x, A) = M((x,y),\, A \times S) \quad \text{and} \quad P(y, B) = M((x,y),\, S \times B)$$

for all $(x,y) \in S \times S$ and all $A, B \in \mathcal{B}$. Given M, let η be a distribution on $S \times S$ with marginals μ and ν, let $\mathbf{Q}_\eta := \eta \otimes M \otimes M \otimes \cdots$ be the induced joint distribution, and let $\{(X_t, Y_t)\}$ have distribution $\mathbf{Q}_\eta$ on $(S \times S)^\infty$. By construction, X_t has distribution μP^t and Y_t has distribution νP^t. Moreover, (23) is valid because $\mathbb{G}$ is absorbing for M, and hence $\mathbb{P}\{(X_j, Y_j) \notin \mathbb{G} \text{ for any } j \leqslant t\} = \mathbb{P}\{(X_t, Y_t) \notin \mathbb{G}\}$.

5.7. **Uniform Ergodicity.** Let S be a partially ordered Polish space satisfying assmption 4.1, and let P be a monotone Markov kernel on S. A distribution π is called *stationary* for P if $\pi P = \pi$. Consider the value

$$\sigma(P) := \inf_{(x,y) \in S \times S} \alpha_O(P_x, P_y),$$

which can be understood as an order-theoretic extension of the Markov–Dobrushin coefficient of ergodicity [3, 29]. It reduces to the usual notion when $\preceq$ is equality.

Theorem 5.1. *If P is monotone, then*

$$\gamma(\mu P, \nu P) \leqslant (1 - \sigma(P))\, \gamma(\mu, \nu) \quad \text{for all } (\mu, \nu) \in \mathscr{P} \times \mathscr{P}. \tag{24}$$

Thus, strict positivity of $\sigma(P)$ implies that $\mu \mapsto \mu P$ is a contraction map on $(\mathscr{P}, \gamma)$. Moreover, in many settings, the bound in (24) cannot be improved upon. For example,

Lemma 5.4. *If P is monotone, S is not a singleton and any x, y in S have a lower bound in S, then*

$$\forall\, \xi > \sigma(P), \;\; \exists\, \mu, \nu \in \mathscr{P} \;\text{ s.t. }\; \gamma(\mu P, \nu P) > (1 - \xi)\, \gamma(\mu, \nu). \tag{25}$$

The significance of theorem 5.1 is summarized in the next corollary.

Corollary 5.1. *Let P be monotone and let S satisfy assumption 4.1. If there exists an $m \in \mathbb{N}$ such that $\sigma(P^m) > 0$, then P has a unique stationary distribution π in $\mathscr{P}$, and*

$$\gamma(\mu P^t, \pi) \leqslant (1 - \sigma(P^m))^{\lfloor t/m \rfloor}\, \gamma(\mu, \pi) \quad \text{for all } \mu \in \mathscr{P}, \; t \geqslant 0. \tag{26}$$

Here $\lfloor x \rfloor$ is the largest $n \in \mathbb{N}$ with $n \leqslant x$.

Proof. Let P and μ be as in the statement of the theorem. The existence of a fixed point of $\mu \mapsto \mu P$, and hence a stationary distribution $\pi \in \mathscr{P}$, follows from theorem 5.1 applied to P^m, Banach's contraction mapping theorem, and the completeness of $(\mathscr{P}, \gamma)$ shown in theorem 4.1. The bound in (26) follows from (24) applied to P^m and the nonexpansiveness of P in the metric γ (see (21)). $\square$

As a first application of these results, consider the standard notion of uniform ergodicity. A Markov kernel P on S is called *uniformly ergodic* if it has a stationary distribution π and $\sup_{x \in S} \|P_x^t - \pi\| \to 0$ as $t \to \infty$. Uniform ergodicity was studied by Markov [20] in a countable state space and by Doeblin [5], Yoshida and Kakutani [35], Doob [8] and many subsequent authors in a general state space. It is defined and reviewed in chapter 16 of [21]. One of the most familiar equivalent conditions for uniform ergodicity [21, thm. 16.0.2] is the existence of an $m \in \mathbb{N}$ and a nontrivial $\phi \in \mathscr{M}$ such that $P_x^m \geqslant \phi$ for all x in S.

One can recover this result using corollary 5.1. Take $\preceq$ to be equality, in which case every Markov operator is monotone, γ is total variation distance and assumption 4.1 is always satisfied. Moreover, $\sigma(P^m)$ reduces to the ordinary ergodicity coefficient of P^m, evaluated using the standard notion of affinity, and hence

$$\sigma(P^m) = \inf_{(x,y)\in S\times S} \alpha(P_x^m, P_y^m) = \inf_{(x,y)\in S\times S} (P_x^m \wedge P_y^m)(S) \geqslant \phi(S) > 0.$$

Thus, all the conditions of corollary 5.1 are satisfied, and

$$\sup_{x\in S} \|P_x^t - \pi\| = \sup_{x\in S} \gamma(P_x^t, \pi) \leqslant 2(1-\sigma(P^m))^{\lfloor t/m \rfloor} \to 0 \qquad (t\to\infty).$$

Now consider the setting of Bhattacharya and Lee [1], where $S = \mathbb{R}^n$, $\preceq$ is the usual pointwise partial order $\leqslant$ for vectors, and $\{g_t\}$ is a sequence of IID random maps from S to itself, generating $\{X_t\}$ via $X_t = g_t(X_{t-1}) = g_t \circ \cdots \circ g_1(X_0)$. The corresponding Markov kernel is $P(x,B) = \mathbb{P}\{g_1(x) \in B\}$. The random maps are assumed to be order preserving on S, so that P is monotone. Bhattacharya and Lee use a "splitting condition," which assumes existence of a $\bar{x} \in S$ and $m \in \mathbb{N}$ such that

(a) $s_1 := \mathbb{P}\{g_m \circ \cdots \circ g_1(y) \leqslant \bar{x},\ \forall y \in S\} > 0$ and
(b) $s_2 := \mathbb{P}\{g_m \circ \cdots \circ g_1(y) \geqslant \bar{x},\ \forall y \in S\} > 0$.

Under these assumptions, they show that $\sup_{x\in S} \beta(P_x^t, \pi)$ converges to zero exponentially fast in t, where β is the Bhattacharya metric introduced in (17). This finding extends earlier results by Dubins and Freedman [9] and Yahav [34] to multiple dimensions.

This result can be obtained as a special case of corollary 5.1. Certainly S is a partially ordered Polish space and assumption 4.1 is satisfied. Moreover, the ordered ergodicity coefficient $\sigma(P^m)$ is strictly positive. To see this, suppose that the splitting condition is satisfied at $m \in \mathbb{N}$. Pick any $x, y \in S$ and let $\{X_t\}$ and $\{Y_t\}$ be independent copies of the Markov chain, starting at x and y respectively. We have

$$\sigma(P^m) \geqslant \mathbb{P}\{X_m \leqslant Y_m\} \geqslant \mathbb{P}\{X_m \leqslant \bar{x} \leqslant Y_m\} = \mathbb{P}\{X_m \leqslant \bar{x}\}\mathbb{P}\{\bar{x} \leqslant Y_m\} \geqslant s_1 s_2.$$

The last term is strictly positive by assumption. Hence all the conditions of corollary 5.1 are satisfied, a unique stationary distribution π exists, and $\sup_{x\in S} \gamma(P_x^t, \pi)$ converges to zero exponentially fast in t. We showed in (19) that $\beta \leqslant 2\gamma$, so the same convergence holds for the Bhattacharya metric.

We can also recover a related convergence result due to Hopenhayn and Prescott [15, theorem 2] that is routinely applied to stochastic stability problems in economics. They assume that S is a compact metric space with a closed partial order and a least element a and greatest element b. They suppose that P is monotone, and that there exists an $\bar{x}$ in S and an $m \in \mathbb{N}$ such that

$$P^m(a, [\bar{x}, b]) > 0 \quad \text{and} \quad P^m(b, [a, \bar{x}]) > 0. \tag{27}$$

In this setting, they show that P has a unique stationary distribution π and $\mu P^t \xrightarrow{w} \pi$ for any $\mu \in \mathscr{P}$ as $t \to \infty$. This result can be obtained from corollary 5.1. Under the stated assumptions, S is Polish and assumption 4.1 is satisfied. The coefficient $\sigma(P^m)$ is strictly positive because, if we let $\{X_t\}$ and $\{Y_t\}$ be independent copies of the Markov chain starting at b and a respectively, then, since $(X_m, Y_m) \in \mathscr{C}(P^m_b, P^m_a)$, we have

$$\alpha_O(P^m_b, P^m_a) \geqslant \mathbb{P}\{X_m \leqslant Y_m\} \geqslant \mathbb{P}\{X_m \leqslant \bar{x} \leqslant Y_m\} = \mathbb{P}\{X_m \leqslant \bar{x}\}\mathbb{P}\{\bar{x} \leqslant Y_m\}.$$

The last term is strictly positive by (27). Positivity of $\sigma(P^m)$ now follows from lemma 5.1, since $a \preceq x, y \preceq b$ for all $x, y \in S$. Hence, by corollary 5.1, there exists a unique stationary distribution π and $\gamma(\mu P^t, \pi) \to 0$ as $t \to \infty$ for any $\mu \in \mathscr{P}$. This convergence implies weak convergence by proposition 4.1 and compactness of S.

6. Final Comments

The conditions for uniform ergodicity (see section 5.7) are often excessively strict when S is unbounded. This led to the study, initiated by Harris [13], of Markov chains that, regardless of their starting point, return to some set $C \subset S$ infinitely often, and where, for some $\delta > 0$ and $m \in \mathbb{N}$, we have $\alpha(P^m_x, P^m_y) \geqslant \delta$ whenever $(x, y) \in C \times C$. The idea is that, provided aperiodicity also holds, independent copies $\{X_t\}$ and $\{Y_t\}$ of the Markov chain will return to C at the same time infinitely often, providing infinitely many opportunities to obtain $X_t = Y_t$ with a probability at least as large as δ. Stability results can then be obtained via (2).

It seems likely that these ideas can also be extended to the order theoretic setting of the present paper if α is replaced by α_0 in the paragraph above, and $X_t = Y_t$ is replaced with $X_t \preceq Y_t$. In particular, both $X_t \preceq Y_t$ and $Y_t \preceq X_t$ will be true with high probability when t is large, and a stability result using the metric γ can be obtained via (3).

Apart from Harris chains, there are various other notions of and results on ergodicity that use the total variation metric, such as f-ergodicity and geometry ergodicity (see, e.g., chapters 13–16 of [21]). It seems likely that, just as for uniform ergodicity, many of these notions and results have natural extensions to the order theoretic context, replacing total variation with total order variation. All of these conjectures are left for future research.

7. Appendix

The appendix collects remaining proofs. Throughout, in addition to notation defined above, cbS_0 denotes all continuous functions $h\colon S \to [0,1]$, while

$$g(\mu,\nu) := \|\mu\| - \alpha_O(\mu,\nu)$$

for each $\mu,\nu \in \mathscr{M}$.

7.1. **Proofs of Section 3 Results.**

Proof of lemma 3.1. For the first equality, fix $\lambda \in \mathscr{M}_s$ and let

$$s(\lambda) := \sup_{I \in i\mathcal{B}} \lambda(I) \quad \text{and} \quad b(\lambda) := \sup_{h \in iH_0} \lambda(h).$$

Since $\mathbb{1}_I \in ibS$ for all $I \in i\mathcal{B}$, we have $b(\lambda) \geqslant s(\lambda)$. To see the reverse inequality, let $h \in iH_0$. Fix $n \in \mathbb{N}$. Let $r_j := j/n$ for $j = 0,\dots,n$. Define $h_n \in iH_0$ by

$$h_n(x) = \max\{r \in \{r_0,\dots,r_n\} : r \leqslant h(x)\}.$$

Since $h \leqslant h_n + 1/n$, we have

$$\lambda(h) \leqslant \lambda(h_n) + \frac{\|\lambda\|}{n}. \tag{28}$$

For $j = 0,\dots,n$, let $I_j := \{x \in S : h_n(x) \geqslant r_j\} \in i\mathcal{B}$. Note that

$$I_n = \{x \in S : h_n(x) = 1\} \subset I_{n-1} \subset \cdots \subset I_0 = S. \tag{29}$$

We have

$$\lambda(h_n) = \lambda(I_n) + \sum_{j=1}^{n} r_{n-j}\lambda(I_{n-j} \setminus I_{n-j+1}).$$

We define $f_0,\dots,f_{n-1} \in iH_0$ and $A_0,\dots,A_{n-1} \in i\mathcal{B}$ as follows. Define $f_0 = h_n$ and $A_0 = I_n$. We trivially have

$$\lambda(f_0) \geqslant \lambda(h_n), \quad \forall x \in A_0,\ f_0(x) = 1, \quad \forall x \in I_{n-1} \setminus A_0,\ f_0(x) = r_{n-1}.$$

Now suppose that for some $j \in \{0, 1, \dots, n-2\}$, we have

$$\lambda(f_j) \geqslant \lambda(h_n), \quad \forall x \in A_j,\ f_j(x) = 1, \quad \forall x \in I_{n-j-1} \setminus A_j,\ f_j(x) = r_{n-j-1}. \tag{30}$$

If $\lambda(I_{n-j-1} \setminus A_j) > 0$, then define

$$f_{j+1}(x) = \begin{cases} 1 & \text{if } x \in I_{n-j-1} \setminus A_j, \\ f_j(x) & \text{otherwise,} \end{cases} \qquad \text{and} \qquad A_{j+1} = I_{n-j-1}.$$

Note that in this case

$$\lambda(f_{j+1}) - \lambda(f_j) = (1 - r_{n-j-1})\lambda(I_{n-j-1} \setminus A_j) > 0,$$
$$\forall x \in I_{n-j-2} \setminus A_{j+1}, \quad f_{j+1}(x) = r_{n-j-2}.$$

If $\lambda(I_{n-j-1} \setminus A_j) \leqslant 0$, then define

$$f_{j+1}(x) = \begin{cases} r_{n-j-2} & \text{if } x \in I_{n-j-1} \setminus A_j, \\ f_j(x) & \text{otherwise,} \end{cases} \qquad \text{and} \qquad A_{j+1} = A_j.$$

In this case we have

$$\lambda(f_{j+1}) - \lambda(f_j) = (r_{n-j-2} - r_{n-j-1})\lambda(I_{n-j-1} \setminus A_j) \geqslant 0.$$

We also have (7.1) by construction. Thus in both cases, we have (30) with j replaced by $j+1$. Continuing this way, we see that (30) holds for all $j = 0, \dots, n-1$.

Let $j = n-1$ in (30). From the definition of r_j and (29) we have $r_0 = 0$ and $I_0 = S$. Thus

$$\lambda(f_{n-1}) \geqslant \lambda(h_n), \quad \forall x \in A_{n-1},\ f_{n-1}(x) = 1, \quad \forall x \in S \setminus A_{n-1},\ f_{n-1}(x) = 0.$$

Since $f_{n-1} = \mathbb{1}_{A_{n-1}}$ and $A_{n-1} = I_j$ for some $j \in \{0, \dots, n-1\}$, recalling (28) we have

$$\lambda(h) - \frac{\|\lambda\|}{n} \leqslant \lambda(h_n) \leqslant \lambda(A_{n-1}) \leqslant s(\lambda).$$

Applying $\sup_{h \in iH_0}$ to the leftmost side, we see that $b(\lambda) - 1/n \leqslant s(\lambda)$. Since this is true for any $n \in \mathbb{N}$, we obtain $b(\lambda) \leqslant s(\lambda)$.

The claim (7) follows from $|a| = \max\{a, -a\}$ and interchange of max and sup. $\square$

Proof of theorem 3.1. Let (X, Y) be a coupling of (μ, ν), and define

$$\mu'(B) := \mathbb{P}\{X \in B,\ X \preceq Y\} \quad \text{and} \quad \nu'(B) := \mathbb{P}\{Y \in B,\ X \preceq Y\}.$$

Clearly $\mu' \leqslant \mu$, $\nu' \leqslant \nu$ and $\mu'(S) = \mathbb{P}\{X \preceq Y\} = \nu'(S)$. Moreover, for any increasing set $I \in \mathcal{B}$ we clearly have $\mu'(I) = \nu'(I)$. Hence $(\mu', \nu') \in \Phi(\mu, \nu)$ and $\mathbb{P}\{X \preceq Y\} = \mu'(S) \leqslant \alpha_O(\mu, \nu)$. We now exhibit a coupling such that equality is attained. In doing so, we can assume that $a := \alpha_O(\mu, \nu) > 0$.[1]

To begin, observe that, by proposition 3.1, there exists a pair $(\mu', \nu') \in \Phi(\mu, \nu)$ with $\mu'(S) = \nu'(S) = a$. Let $\mu^r := \frac{\mu - \mu'}{1-a}$ and $\nu^r := \frac{\nu - \nu'}{1-a}$. By construction, $\mu^r, \nu^r, \mu'/a$ and ν'/a are probability measures satisfying

$$\mu = (1-a)\mu^r + a(\mu'/a) \quad \text{and} \quad \nu = (1-a)\nu^r + a(\nu'/a).$$

We construct a coupling (X, Y) as follows. Let U, X', Y', X^r and Y^r be random variables on a common probability space such that

(a) $X' \stackrel{\mathscr{D}}{=} \mu'/a$, $Y' \stackrel{\mathscr{D}}{=} \nu'/a$, $X^r \stackrel{\mathscr{D}}{=} \mu^r$ and $Y^r \stackrel{\mathscr{D}}{=} \nu^r$
(b) U is uniform on $[0, 1]$ and independent of (X', Y', X^r, Y^r) and
(c) $\mathbb{P}\{X' \preceq Y'\} = 1$.

The pair in (c) can be constructed via the Nachbin–Strassen theorem [17, thm. 1], since $\mu'/a \preceq_{sd} \nu'/a$. Now let

$$X := \mathbb{1}\{U \leqslant a\}X' + \mathbb{1}\{U > a\}X^r \quad \text{and} \quad Y := \mathbb{1}\{U \leqslant a\}Y' + \mathbb{1}\{U > a\}Y^r.$$

Evidently $(X, Y) \in \mathscr{C}(\mu, \nu)$. Moreover, for this pair, we have

$$\mathbb{P}\{X \preceq Y\} \geqslant \mathbb{P}\{X \preceq Y,\ U \leqslant a\} = \mathbb{P}\{X' \preceq Y',\ U \leqslant a\}.$$

By independence the right hand side is equal to $\mathbb{P}\{X' \preceq Y'\}\mathbb{P}\{U \leqslant a\} = a$, so $\mathbb{P}\{X \preceq Y\} \geqslant a := \alpha_O(\mu, \nu)$. We conclude that

$$\alpha_O(\mu, \nu) = \max_{(X,Y) \in \mathscr{C}(\mu,\nu)} \mathbb{P}\{X \preceq Y\}. \tag{31}$$

Next, observe that, for any $(X, Y) \in \mathscr{C}(\mu, \nu)$ and $h \in ibS$, we have

$$\begin{aligned}
\mu(h) - \nu(h) &= \mathbb{E}\, h(X) - \mathbb{E}\, h(Y) \\
&= \mathbb{E}\,[h(X) - h(Y)]\mathbb{1}\{X \preceq Y\} + \mathbb{E}\,[h(X) - h(Y)]\mathbb{1}\{X \npreceq Y\} \\
&\leqslant \mathbb{E}\,[h(X) - h(Y)]\mathbb{1}\{X \npreceq Y\}.
\end{aligned}$$

[1] If not, then for any $(X, Y) \in \mathscr{C}(\mu, \nu)$ we have $0 \leqslant \mathbb{P}\{X \preceq Y\} \leqslant \alpha_O(\mu, \nu) = 0$.

Specializing to $h = \mathbb{1}_I$ for some $I \in i\mathcal{B}$, we have $\mu(I) - \nu(I) \leqslant \mathbb{P}\{X \npreceq Y\} = 1 - \mathbb{P}\{X \preceq Y\}$. From this bound and (31), the proof of (12) will be complete if we can show that

$$\sup_{(X,Y) \in \mathscr{C}(\mu,\nu)} \mathbb{P}\{X \preceq Y\} \geqslant 1 - \sup_{I \in i\mathcal{B}} \{\mu(I) - \nu(I)\}. \tag{32}$$

To prove (32), let $\mathcal{B} \otimes \mathcal{B}$ be the product σ-algebra on $S \times S$ and let π_i be the i-th coordinate projection, so that $\pi_1(x, y) = x$ and $\pi_2(x, y) = y$ for any $(x, y) \in S \times S$. As usual, given $Q \subset S \times S$, we let $\pi_1(Q)$ be all $x \in S$ such that $(x, y) \in Q$, and similarly for π_2. Recall that $\mathcal{C}$ is the closed sets in S and $d\mathcal{C}$ is the decreasing sets in $\mathcal{C}$. Strassen's theorem [31] implies that, for any $\epsilon \geqslant 0$ and any closed set $K \subset S \times S$, there exists a probability measure ξ on $(S \times S, \mathcal{B} \otimes \mathcal{B})$ with marginals μ and ν such that $\xi(K) \geqslant 1 - \epsilon$ whenever

$$\nu(F) \leqslant \mu(\pi_1(K \cap (S \times F))) + \epsilon, \qquad \forall\, F \in \mathcal{C}.$$

Note that if $F \in \mathcal{C}$, then, since $\preceq$ is a closed partial order, so is the smallest decreasing set $d(F)$ that containts F. Let $\epsilon := \sup_{D \in d\mathcal{C}}\{\nu(D) - \mu(D)\}$, so that

$$\epsilon \geqslant \sup_{F \in \mathcal{C}} \{\nu(d(F)) - \mu(d(F))\} \geqslant \sup_{F \in \mathcal{C}} \{\nu(F) - \mu(d(F))\}.$$

Noting that $d(F)$ can be expressed as $\pi_1(\mathbb{G} \cap (S \times F))$, it follows that, for any $F \in \mathcal{C}$,

$$\nu(F) \leqslant \mu(\pi_1(\mathbb{G} \cap (S \times F))) + \epsilon.$$

Since $\preceq$ is closed, $\mathbb{G}$ is closed, and Strassen's theorem applies. From this theorem we obtain a probability measure ξ on the product space $S \times S$ such that $\xi(\mathbb{G}) \geqslant 1 - \epsilon$ and ξ has marginals μ and ν.

Because complements of increasing sets are decreasing and vice versa, we have

$$\sup_{I \in i\mathcal{B}} \{\mu(I) - \nu(I)\} \geqslant \sup_{D \in d\mathcal{C}} \{\nu(D) - \mu(D)\} = \epsilon \geqslant 1 - \xi(\mathbb{G}). \tag{33}$$

Now consider the probability space $(\Omega, \mathscr{F}, \mathbb{P}) = (S \times S, \mathcal{B} \otimes \mathcal{B}, \xi)$, and let $X = \pi_1$ and $Y = \pi_2$. We then have $\xi(\mathbb{G}) = \xi\{(x, y) \in S \times S : x \preceq y\} = \mathbb{P}\{X \preceq Y\}$. Combining this equality with (33) implies (32). $\square$

7.2. **Proofs of Section 4 Results.** We begin with an elementary lemma:

Lemma 7.1. *For any $\mu, \nu \in \mathscr{M}$, we have $\mu \leqslant \nu$ whenever $\mu(h) \leqslant \nu(h)$ for all $h \in cbS_0$.*

Proof. Suppose that $\mu(h) \leqslant \nu(h)$ for all $h \in cbS_0$. We claim that

$$\mu(F) \leqslant \nu(F) \text{ for any closed set } F \subset S. \tag{34}$$

To see this, let ρ be a metric compatible with the topology of S and let F be any closed subset of S. Let $f_\epsilon(x) := \max\{1 - \rho(x, F)/\epsilon, 0\}$ for $\epsilon > 0$, $x \in S$, where $\rho(x, F) = \inf_{y \in F} \rho(x, y)$. Since $\rho(\cdot, F)$ is continuous and $0 \leqslant f_\epsilon \leqslant 1$, we have $f_\epsilon \in cbS_0$. Let $F_\epsilon = \{x \in S : \rho(x, F) < \epsilon\}$ for $\epsilon > 0$. Note that $f_\epsilon(x) = 1$ for all $x \in F$, and that $f_\epsilon(x) = 0$ for all $x \notin F_\epsilon$. Thus,

$$\mu(F) \leqslant \mu(f_\epsilon) \leqslant \nu(f_\epsilon) \leqslant \nu(F_\epsilon). \tag{35}$$

Since $F = \cap_{\epsilon>0} F_\epsilon$, we have $\lim_{\epsilon \downarrow 0} \nu(F_\epsilon) = \nu(F)$, so letting $\epsilon \downarrow 0$ in (35) yields $\mu(F) \leqslant \nu(F)$. Hence (34) holds.

Let $B \in \mathcal{B}$ and fix $\epsilon > 0$. Since all probability measures on a Polish space are regular, there exists a closed set $F \subset B$ such that $\mu(B) \leqslant \mu(F) + \epsilon$. Thus by (34), we have $\mu(B) \leqslant \mu(F) + \epsilon \leqslant \nu(F) + \epsilon \leqslant \nu(B) + \epsilon$. Since $\epsilon > 0$ is arbitrary, this yields $\mu(B) \leqslant \nu(B)$. Hence $\mu \leqslant \nu$. □

Proof of theorem 4.1. Let $\{\mu_n\}$ be a Cauchy sequence in $(\mathscr{P}, \gamma)$. Our first claim is that $\{\mu_n\}$ is tight. To show this, fix $\epsilon > 0$. Let $\mu := \mu_N$ be such that

$$n \geqslant N \implies \gamma(\mu, \mu_n) < \epsilon. \tag{36}$$

Let K be a compact subset of S such that $\mu(K) \geqslant 1 - \epsilon$ and let $\bar{K} := i(K) \cap d(K)$. We have

$$\mu_n(\bar{K}^c) = \mu_n(i(K)^c \cup d(K)^c) \leqslant \mu_n(i(K)^c) + \mu_n(d(K)^c).$$

For $n \geqslant N$, this bound, (16), (36) and the definition of K yield

$$\mu_n(\bar{K}^c) < \mu(i(K)^c) + \mu(d(K)^c) + 2\epsilon \leqslant \mu(K^c) + \mu(K^c) + 2\epsilon \leqslant 4\epsilon.$$

Hence $\{\mu_n\}_{n \geqslant N}$ is tight. It follows that $\{\mu_n\}_{n \geqslant 1}$ is likewise tight. As a result, by Prohorov's theorem, it has a subsequence that converges weakly to some $\mu^* \in \mathscr{P}$. We aim to show that $\gamma(\mu_n, \mu^*) \to 0$.

To this end, fix $\epsilon > 0$ and let n_ϵ be such that $\gamma(\mu_m, \mu_{n_\epsilon}) < \epsilon$ whenever $m \geqslant n_\epsilon$. Fix $m \geqslant n_\epsilon$ and let $\nu := \mu_m$. For all $n \geqslant n_\epsilon$, we have $\gamma(\nu, \mu_n) < \epsilon$. Fixing any such

$n \geqslant n_\epsilon$, we observe that since $g(\mu_n, \nu) < \epsilon$, there exists $(\tilde{\mu}_n, \tilde{\nu}_n) \in \Phi(\mu_n, \nu)$ with $\|\tilde{\mu}_n\| = \|\tilde{\nu}_n\| > 1 - \epsilon$. Multiplying $\tilde{\mu}_n$ and $\tilde{\nu}_n$ by $(1-\epsilon)/\|\tilde{\mu}_n\| < 1$, denoting the resulting measures by $\tilde{\mu}_n$ and $\tilde{\nu}_n$ again, we have

$$\tilde{\mu}_n \leqslant \mu_n, \quad \tilde{\nu}_n \leqslant \nu, \quad \|\tilde{\mu}_n\| = \|\tilde{\nu}_n\| = 1 - \epsilon, \quad \tilde{\mu}_n \preceq_{sd} \tilde{\nu}_n. \tag{37}$$

Note that $\{\tilde{\nu}_n\}$ is tight. Since $\{\mu_n\}$ is tight, so is $\{\tilde{\mu}_n\}$. Thus there exist subsequences $\{\mu_{n_i}\}_{i\in\mathbb{N}}, \{\tilde{\mu}_{n_i}\}_{i\in\mathbb{N}}$, and $\{\tilde{\nu}_{n_i}\}_{i\in\mathbb{N}}$ of $\{\mu_n\}, \{\tilde{\mu}_n\}$, and $\{\tilde{\nu}_n\}$ respectively such that, for some $\tilde{\mu}^*, \tilde{\nu}^* \in \mathscr{M}$ with $\|\tilde{\mu}^*\| = \|\tilde{\nu}^*\| = 1 - \epsilon$, we have

$$\mu_{n_i} \xrightarrow{w} \mu^*, \quad \tilde{\mu}_{n_i} \xrightarrow{w} \tilde{\mu}^*, \quad \tilde{\nu}_{n_i} \xrightarrow{w} \tilde{\nu}^*, \quad \forall i \in \mathbb{N}, \; \tilde{\mu}_{n_i} \preceq_{sd} \tilde{\nu}_{n_i}.$$

Given $h \in cbS_0$, since $\tilde{\mu}_{n_i}(h) \leqslant \mu_{n_i}(h)$ and $\tilde{\nu}_{n_i}(h) \leqslant \nu(h)$ for all $i \in \mathbb{N}$ by (37), we have $\tilde{\mu}^*(h) \leqslant \mu^*(h)$ and $\tilde{\nu}^*(h) \leqslant \nu(h)$ by weak convergence. Thus $\tilde{\mu}^* \leqslant \mu^*$ and $\tilde{\nu}^* \leqslant \nu$ by lemma 7.1. We have $\tilde{\mu}^* \preceq_{sd} \tilde{\nu}^*$ by [17, proposition 3]. It follows that $(\tilde{\mu}^*, \tilde{\nu}^*) \in \Phi(\mu^*, \nu)$. We have $g(\mu^*, \nu) \leqslant 1 - \|\tilde{\mu}^*\| = \epsilon$.

By a symmetric argument, we also have $g(\nu, \mu^*) \leqslant \epsilon$. Hence $\gamma(\nu, \mu^*) \leqslant 2\epsilon$. Recalling the definition of ν, we have now shown that, $\forall m \geqslant n_\epsilon$, $\gamma(\mu_m, \mu^*) \leqslant 2\epsilon$. Since ϵ was arbitrary this concludes the proof. □

7.3. **Proofs of Section 5 Results.** We begin with some lemmata.

Lemma 7.2. *If P is monotone, then $\sigma(P) = \inf_{(\mu,\nu)\in\mathscr{P}\times\mathscr{P}} \alpha_O(\mu P, \nu P)$.*

Proof. Let P be a monotone Markov kernel. It suffices to show that the inequality $\sigma(P) \leqslant \inf_{(\mu,\nu)\in\mathscr{P}\times\mathscr{P}} \alpha_O(\mu P, \nu P)$ holds, since the reverse inequality is obvious. By the definition of $\sigma(P)$ and the identities in (12), the claim will be established if we can show that

$$\sup_{x,y} \sup_{I\in i\mathcal{B}} \{P(x,I) - P(y,I)\} \geqslant \sup_{\mu,\nu} \sup_{I\in i\mathcal{B}} \{\mu P(I) - \nu P(I)\}. \tag{38}$$

where x and y are chosen from S and μ and ν are chosen from $\mathscr{P}$. Let s be the value of the right hand side of (38) and let $\epsilon > 0$. Fix $\mu, \nu \in \mathscr{P}$ and $I \in i\mathcal{B}$ such that $\mu P(I) - \nu P(I) > s - \epsilon$, or, equivalently,

$$\int \{P(x,I) - P(y,I)\}(\mu \times \nu)(dx, dy) > s - \epsilon.$$

From this expression we see that there are $\bar{x}, \bar{y} \in S$ such that $P(\bar{x}, I) - P(\bar{y}, I) > s - \epsilon$. Hence $\sup_{x,y} \sup_{I\in i\mathcal{B}} \{P(x,I) - P(y,I)\} \geqslant s$, as was to be shown. □

Lemma 7.3. *If $\mu, \nu \in \mathscr{M}$ and $(\tilde{\mu}, \tilde{\nu})$ is an ordered component pair of (μ, ν), then*

$$g(\mu P, \nu P) \leqslant g((\mu - \tilde{\mu})P, (\nu - \tilde{\nu})P).$$

Proof. Fix μ, ν in $\mathscr{M}$ and $(\tilde{\mu}, \tilde{\nu}) \in \Phi(\mu, \nu)$. Consider the residual measures $\hat{\mu} := \mu - \tilde{\mu}$ and $\hat{\nu} := \nu - \tilde{\nu}$. Let (μ', ν') be a maximal ordered component pair of $(\hat{\mu}P, \hat{\nu}P)$, and define

$$\mu^* := \mu' + \tilde{\mu}P \quad \text{and} \quad \nu^* := \nu' + \tilde{\nu}P.$$

We claim that (μ^*, ν^*) is an ordered component pair for $(\mu P, \nu P)$. To see this, note that

$$\mu^* = \mu' + \tilde{\mu}P \leqslant \hat{\mu}P + \tilde{\mu}P = (\hat{\mu} + \tilde{\mu})P = \mu P,$$

and, similarly, $\nu^* \leqslant \nu P$. The measures μ^* and ν^* also have the same mass, since

$$\|\mu^*\| = \|\mu'\| + \|\tilde{\mu}P\| = \|\mu'\| + \|\tilde{\mu}\| = \|\nu'\| + \|\tilde{\nu}\| = \|\nu'\| + \|\tilde{\nu}P\| = \|\nu^*\|.$$

Moreover, since $\tilde{\mu} \preceq_{sd} \tilde{\nu}$ and P is monotone, we have $\tilde{\mu}P \preceq_{sd} \tilde{\nu}P$. Hence $\mu^* \preceq_{sd} \nu^*$, completing the claim that (μ^*, ν^*) is an ordered component pair for $(\mu P, \nu P)$. As a result,

$$\begin{aligned} g(\mu P, \nu P) &\leqslant \|\mu P\| - \|\mu^*\| = \|\mu P\| - \|\mu'\| - \|\tilde{\mu}P\| \\ &= \|\mu\| - \|\mu'\| - \|\tilde{\mu}\| = \|\hat{\mu}\| - \|\mu'\| = \|\hat{\mu}P\| - \|\mu'\| = g(\hat{\mu}P, \hat{\nu}P). \qquad \square \end{aligned}$$

Proof of theorem 5.1. Let $\mu, \nu \in \mathscr{P}$. Let $(\tilde{\mu}, \tilde{\nu})$ be a maximal ordered component pair for (μ, ν). Let $\hat{\mu} = \mu - \tilde{\mu}$ and $\hat{\nu} = \nu - \tilde{\nu}$ be the residuals. Since $\|\tilde{\mu}\| = \|\tilde{\nu}\|$, we have $\|\hat{\mu}\| = \|\hat{\nu}\|$. Suppose first that $\|\hat{\mu}\| > 0$. Then $\hat{\mu}P/\|\hat{\mu}\|$ and $\hat{\nu}P/\|\hat{\mu}\|$ are both in $\mathscr{P}$. Thus, by lemma 7.2,

$$1 - \alpha_O(\hat{\mu}P/\|\hat{\mu}\|, \hat{\nu}P/\|\hat{\mu}\|) \leqslant 1 - \sigma(P).$$

Applying the positive homogeneity property in lemma 3.3 yields

$$\|\hat{\mu}\| - \alpha_O(\hat{\mu}P, \hat{\nu}P) \leqslant (1 - \sigma(P))\,\|\hat{\mu}\|,$$

Note that this inequality trivially holds if $\|\hat{\mu}\| = 0$. From the definition of g, we can write the same inequality as $g(\hat{\mu}P, \hat{\nu}P) \leqslant (1 - \sigma(P))g(\mu, \nu)$. If we apply lemma 7.3 to the latter we obtain

$$g(\mu P, \nu P) \leqslant (1 - \sigma(P))g(\mu, \nu).$$

Reversing the roles of μ and ν, we also have $g(\nu P, \mu P) \leqslant (1 - \sigma(P))g(\nu, \mu)$. Thus

$$\gamma(\mu P, \nu P) = g(\mu P, \nu P) + g(\nu P, \mu P) \\ \leqslant (1 - \sigma(P))[g(\mu, \nu) + g(\nu, \mu)] = (1 - \sigma(P))\gamma(\mu, \nu),$$

verifying the claim in (24). □

Proof of lemma 5.4. To see that (25) holds, fix $\xi > \sigma(P)$ and suppose first that $\sigma(P) = 1$. Then (25) holds because the right hand side of (25) can be made strictly negative by choosing $\mu, \nu \in \mathscr{P}$ to be distinct. Now suppose that $\sigma(P) < 1$ holds. It suffices to show that

$$\forall \epsilon > 0, \ \exists x, y \in S \text{ such that } y \preceq x, \ x \npreceq y \text{ and } \alpha_0(P_x, P_y) < \sigma(P) + \epsilon. \tag{39}$$

Indeed, if we take (39) as valid, set $\epsilon := \xi - \sigma(P)$ and choose x and y to satisfy the conditions in (39), then we have $\gamma(P_x, P_y) = 2 - \alpha_0(P_x, P_y) - \alpha_0(P_y, P_x) > 2 - \xi - 1 = 1 - \xi = (1 - \xi)\gamma(\delta_x, \delta_y)$. Therefore (25) holds.

To show that (39) holds, fix $\epsilon > 0$. We can use $\sigma(P) < 1$ and the definition of $\sigma(P)$ as an infimum to choose an $\delta \in (0, \epsilon)$ and points $\bar{x}, \bar{y} \in S$ such that $\alpha_0(P_{\bar{x}}, P_{\bar{y}}) < \sigma(P) + \delta < 1$. Note that $\bar{x} \preceq \bar{y}$ cannot hold here, because then $\alpha_0(P_{\bar{x}}, P_{\bar{y}}) = 1$, a contradiction. So suppose instead that $\bar{x} \npreceq \bar{y}$. Let y be a lower bound of $\bar{x}$ and $\bar{y}$ and let $x := \bar{x}$. We claim that (39) holds for the pair (x, y).

To see this, observe that, by the monotonicity result in lemma 5.1 and $y \preceq \bar{y}$ we have $\alpha_0(P_x, P_y) = \alpha_0(P_{\bar{x}}, P_y) \leqslant \alpha_0(P_{\bar{x}}, P_{\bar{y}}) < \sigma(P) + \delta < \sigma(P) + \epsilon$. Moreover, $y \preceq x$ because $x = \bar{x}$ and y is by definition a lower bound of $\bar{x}$. Finally, $x \npreceq y$ because if not then $\bar{x} = x \preceq y \preceq \bar{y}$, contradicting our assumption that $\bar{x} \npreceq \bar{y}$. □

References

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

[2] Santanu Chakraborty and BV Rao. Completeness of the Bhattacharya metric on the space of probabilities. *Statistics & Probability Letters*, 36(4):321–326, 1998.

[3] Roland L Dobrushin. Central limit theorem for nonstationary Markov chains. i. *Theory of Probability & Its Applications*, 1(1):65–80, 1956.

[4] Wolfang Doeblin. Exposé de la théorie des chaınes simples constantes de markova un nombre fini détats. *Mathématique de lUnion Interbalkanique*, 2(77-105):78–80, 1938.

[5] Wolfgang Doeblin. Sur les propriétés asymptotiques de mouvements régis par certains types de chaînes simples. *Bull. Math. Soc. Roum. Sci.*, 39:(1) 57–115; (2), 3–61, 1937.

[6] Wolfgang Doeblin. Éléments d'une théorie générale des chaînes simples constantes de Markoff. In *Annales Scientifiques de l'Ecole Normale Supérieure*, volume 57, pages 61–111, 1940.

[7] Wolfgang Doeblin and Harry Cohn. *Doeblin and modern probability*, volume 149. American Mathematical Soc., 1993.

[8] Joseph L Doob. *Stochastic processes*. Wiley New York, 1953.

[9] Lester E Dubins and David A Freedman. Invariant probabilities for certain Markov processes. *The Annals of Mathematical Statistics*, 37(4):837–848, 1966.

[10] R.M. Dudley. *Real Analysis and Probability*, volume 74. Cambridge University Press, 2002.

[11] Alison L Gibbs and Francis Edward Su. On choosing and bounding probability metrics. *International statistical review*, 70(3):419–435, 2002.

[12] Clark R Givens, Rae Michael Shortt, et al. A class of wasserstein metrics for probability distributions. *The Michigan Mathematical Journal*, 31(2):231–240, 1984.

[13] Theodore E Harris. The existence of stationary measures for certain Markov processes. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, volume 2, pages 113–124, 1956.

[14] O. Hernández-Lerma and J.B. Lasserre. *Markov Chains and Invariant Probabilities*. Springer, 2003.

[15] Hugo A Hopenhayn and Edward C Prescott. Stochastic monotonicity and stationary distributions for dynamic economies. *Econometrica*, pages 1387–1406, 1992.

[16] T. Kamae and U. Krengel. Stochastic partial ordering. *The Annals of Probability*, 6(6):1044–1049, 1978.

[17] T. Kamae, U. Krengel, and G.L. O'Brien. Stochastic inequalities on partially ordered spaces. *The Annals of Probability*, pages 899–912, 1977.

[18] T. Lindvall. *Lectures on the coupling method*. Dover Publications, 2002.

[19] Motoya Machida and Alexander Shibakov. Monotone bivariate Markov kernels with specified marginals. *Proceedings of the American Mathematical Society*, 138(6):2187–2194, 2010.

[20] Andrey Andreyevich Markov. Extension of the law of large numbers to dependent quantities. *Izv. Fiz.-Matem. Obsch. Kazan Univ.(2nd Ser)*, 15:135–156, 1906.

[21] Sean P Meyn and Richard L Tweedie. *Markov chains and stochastic stability*. Springer Science & Business Media, 2012.

[22] James R Norris. *Markov chains*. Number 2. Cambridge university press, 1998.

[23] Esa Nummelin. *General irreducible Markov chains and non-negative operators*. Cambridge University Press, 2004.

[24] Steven Orey et al. Recurrent Markov chains. *Pacific J. Math*, 9(3):805–827, 1959.

[25] James W Pitman. Uniform rates of convergence for Markov chain transition probabilities. *Probability Theory and Related Fields*, 29(3):193–227, 1974.

[26] David Pollard. *A user's guide to measure theoretic probability*, volume 8. Cambridge University Press, 2002.

[27] Daniel Revuz. *Markov chains*, volume 11. Elsevier, 2008.

[28] Jack Schechtman. An income fluctuation problem. *Journal of Economic Theory*, 12(2):218–241, 1976.

[29] Eugene Seneta. Coefficients of ergodicity: structure and applications. *Advances in applied probability*, 11(03):576–590, 1979.

[30] N. Stokey and Jr. Lucas, R.E. *Recursive Methods in Economic Dynamics (with EC Prescott)*. Harvard University Press, 1989.

[31] V. Strassen. The existence of probability measures with given marginals. *The Annals of Mathematical Statistics*, pages 423–439, 1965.

[32] H. Thorisson. *Coupling, stationarity, and regeneration*. Springer New York, 2000.

[33] Hermann Thorisson. Coupling and shift-coupling random sequences. *Contemporary Mathematics*, 149:85–85, 1993.

[34] Joseph A Yahav. On a fixed point theorem and its stochastic equivalent. *Journal of Applied Probability*, pages 605–611, 1975.

[35] Kakutani S Yoshida . Operator theoretical treatment of Markov's process and mean ergodic theorem. *Ann. Math*, 42, 1941.

RIEB, KOBE UNIVERSITY

E-mail address: `tkamihig@rieb.kobe-u.ac.jp`

RESEARCH SCHOOL OF ECONOMICS, AUSTRALIAN NATIONAL UNIVERSITY

E-mail address: `john.stachurski@anu.edu.au`