J. Appl. Prob. **56,** 1–22 (2019)
*doi:*10.1017/jpr.2019.2

A UNIFIED STABILITY THEORY FOR CLASSICAL AND MONOTONE MARKOV CHAINS

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

Abstract

In this paper we integrate two strands of the literature on stability of general state Markov chains: conventional, total-variation-based results and more recent order-theoretic results. First we introduce a complete metric over 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.

Keywords: Total variation; Markov chain; stochastic domination; coupling

2010 Mathematics Subject Classification: Primary 60J05; 60J99
Secondary 54E50; 06A06

1. Introduction

Following the work of Wolfgang Doeblin ([6], [7], [8]), many classical results from Markov chain theory have built on fundamental connections between total variation distance, Markov chains and couplings. For some models, however, total variation convergence is too strong. In response, researchers have developed an alternative methodology based on monotonicity ([1], [10], [36]). In this line of research, transition probabilities are assumed to have a form of monotonicity not required in the classical theory. At the same time, mixing conditions are generally weaker, as is the notion of convergence to the stationary distribution. Further contributions to this approach can be found in [2], [16], [20], and [30]. For some recent extensions and applications in economics, see [21].

To give one example, consider a Markov chain $\{X_t\}$ defined by

$$X_{t+1} = \frac{X_t + W_{t+1}}{2}, \tag{1.1}$$

where $\{W_t\}_{t\geq 1}$ is an independent and identically distributed (i.i.d.) Bernoulli($\frac{1}{2}$) random sequence, taking values 0 and 1 with equal probability. For the state space take $S = [0, 1]$. Let $P^t(x, \cdot)$ be the distribution of X_t given $X_0 = x \in S$. Clearly, if X_t is a rational number in S then so is X_{t+1}. Similarly, if X_t is irrational then so is X_{t+1}. Thus, if x and y are rational and irrational, respectively, the distributions $P^t(x, \cdot)$ and $P^t(y, \cdot)$ are concentrated on disjoint sets,

Received 20 April 2017; revision received 12 November 2018.
* Postal address: Research Institute of Economics and Business, Kobe University, Japan.
Email address: tkamihig@rieb.kobe-u.ac.jp
** Postal address: Research School of Economics, Australian National University, Australia.
Email address: john.stachurski@anu.edu.au

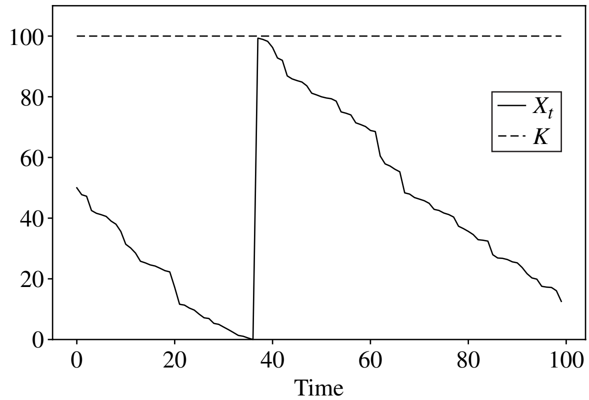

FIGURE 1: A time series from the inventory model

and, hence, when $\|\cdot\|$ is the total variation norm,

$$\|P^t(x, \cdot) - P^t(y, \cdot)\| = 2 \quad \text{for all } t \in \mathbb{N}.$$

Total variation convergence fails for this class of models.

At the same time, the right-hand side of (1.1) is increasing in the current state for each fixed value of the shock W_{t+1}. Moreover, trajectories mix in a monotone sense: a trajectory starting at $X_0 = 0$ can approach 1 with a suitable string of shocks and a trajectory starting at 1 can approach 0. Using these facts, one can show using the results in [1], say, that a unique stationary distribution exists and the distribution of X_t converges to it in a complete metric defined over the Borel probability measures that is weaker than total variation convergence.

This is one example where monotone methods can be used to establish some form of stability, despite the fact that the classical conditions based around total variation convergence fail. Conversely, there are many models that the monotone methods developed in [1] and related papers cannot accommodate, while the classical theory based around total variation convergence handles them easily. One example is the simple 'inventory' model

$$X_{t+1} = \begin{cases} (X_t - W_{t+1})_+ & \text{if } X_t > 0, \\ (K - W_{t+1})_+ & \text{if } X_t = 0, \end{cases} \tag{1.2}$$

where $x_+ := \max\{x, 0\}$. Again, $\{W_t\}$ is i.i.d. Assume that $\ln W_t$ is standard normal. The state space we take to be $S = [0, K]$. Figure 1 shows a typical trajectory when $K = 100$ and $X_0 = 50$.

On the one hand, the monotone methods in [1] cannot be applied here because of a failure of monotonicity with respect to the standard ordering of $\mathbb{R}$. On the other hand, the classical theory based around total variation convergence is straightforward to apply. For example, one can use Doeblin's condition (see, e.g. [25, Theorem 16.2.3]) to show the existence of a unique stationary distribution to which the distribution of X_t converges in total variation, regardless of the distribution of X_0. In the terminology of [25], the process is uniformly ergodic.

The purpose of this paper is to show that both of these stability results (i.e. the two sets of results concerning the two models (1.1) and (1.2)), which were based on two hitherto separate approaches, can be derived from the same theoretical framework. More generally, we construct stability results that encompasses all uniformly ergodic models in the sense of [25] and all monotone models shown to be stable in [1], as well as extending to other monotone or partially monotone models on state spaces other than $\mathbb{R}^n$.

We begin our analysis by introducing what is shown to be a complete metric γ on the set of Borel probability measures on a partially ordered Polish space that includes total variation distance, the Kolmogorov uniform distance, and the Bhattacharya distance ([1], [3]) as special cases. We 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. Then, by varying the choice of partial order, we recover key aspects of both classical total-variation-based stability theory and monotone methods as special cases.

Prior to commencing we note that, in terms of existing literature, there is an additional line of research that deals with Markov models for which the classical conditions of irreducibility and total variation convergence fail. Irreducibility is replaced by an assumption that the law of motion for the state is itself contracting 'on average', and this contractivity is then passed on to an underlying metric over distributions that conforms in some way to the topology of the state space. See, for example, [4], [17], or [35]. Such results can be used to show stability of our first example, which contracts on average with respect to the usual metric on $\mathbb{R}$. On the other hand, it cannot be directly applied to our second (i.e. inventory) example using the same metric, since the law of motion contains a jump. In [1] and [16] one can find other applications where monotone methods can be used—including the results developed here—while the 'contraction on average' conditions of [4] and [35] do not in general hold. See, for example, the growth model studied in Section 6.B of [16], where a high marginal product of capital near zero generates rapid growth when the capital stock is low. One consequence is that the law of motion does not typically exhibit contraction on average in the neighbourhood of zero.

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 Appendix A. The conclusion contains 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 $\mathfrak{C}$ be the closed sets, and let $\mathcal{B}$ be the Borel sets. Let $\mathcal{M}_s$ denote the set of all finite signed measures on $(S, \mathcal{B})$. In other words, $\mathcal{M}_s$ is all countably additive set functions from $\mathcal{B}$ to $\mathbb{R}$. Let $\mathcal{M}$ and $\mathcal{P}$ be the finite measures and probability measures in $\mathcal{M}_s$, respectively. If κ and λ are in $\mathcal{M}_s$, then $\kappa \leq \lambda$ means that $\kappa(B) \leq \lambda(B)$ for all $B \in \mathcal{B}$. The symbol δ_x denotes the probability measure concentrated on $x \in S$.

Let bS be the set of all bounded $\mathcal{B}$-measurable functions from S into $\mathbb{R}$. If $h \in bS$ and $\lambda \in \mathcal{M}_s$, then $\lambda(h) := \int h \, \mathrm{d}\lambda$. For f and g in bS, the statement $f \leq g$ means that $f(x) \leq g(x)$ for all $x \in S$. Let

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

The *total variation norm* of $\lambda \in \mathcal{M}_s$ is $\|\lambda\| := \sup_{h \in H} |\lambda(h)|$. Given μ and ν in $\mathcal{P}$, a random element (X, Y) taking values in $S \times S$ and defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ 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. [22] or [34]). The set of all couplings of (μ, ν) is denoted below by $\mathcal{C}(\mu, \nu)$. A sequence $\{\mu_n\} \subset \mathcal{P}$ converges to $\mu \in \mathcal{P}$ *weakly* if $\mu_n(h) \to \mu(h)$ as $n \to \infty$ for all continuous $h \in bS$. In this case we write $\mu_n \xrightarrow{\text{w}} \mu$.

https://doi.org/10.1017/jpr.2019.2 Published online by Cambridge University Press

Given μ and $\nu \in \mathcal{M}$, their measure-theoretic *infimum* $\mu \wedge \nu$ is the largest element of $\mathcal{M}$ 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, [29]. For probability measures, we also have

$$\sup_{B\in\mathcal{B}}\{\mu(B) - \nu(B)\} = \sup_{B\in\mathcal{B}} |\mu(B) - \nu(B)| = \tfrac{1}{2}\|\mu - \nu\|. \tag{2.1}$$

The *affinity* between two measures μ, ν in $\mathcal{M}$ is the value $\alpha(\mu, \nu) := (\mu \wedge \nu)(S)$. The following properties are elementary.

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

(a) $0 \leq \alpha(\mu, \nu) \leq \min\{\mu(S), \nu(S)\}$,

(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* $c \geq 0$.

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

$$\alpha(\mu, \nu) = 1 - \sup_{B\in\mathcal{B}} |\mu(B) - \nu(B)| = \max_{(X,Y)\in\mathcal{C}(\mu,\nu)} \mathbb{P}\{X = Y\}. \tag{2.2}$$

(See, e.g. [22] and [29].) The second equality in (2.2) states that, if $(X, Y) \in \mathcal{C}(\mu, \nu)$, then $\mathbb{P}\{X = Y\} \leq \alpha(\mu, \nu)$, and, moreover, there exists a $(X, Y) \in \mathcal{C}(\mu, \nu)$ such that equality is attained. Any such coupling is called a *maximal* or *gamma* coupling. See Theorem 5.2 of [22]. From (2.1) and (2.2) we obtain

$$\|\mu - \nu\| = 2(1 - \alpha(\mu, \nu)). \tag{2.3}$$

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. Henceforth, 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. The partial order is assumed to be closed in the theory developed below because we build a metric over Borel probability measures that depends on this partial order and closedness of the order is required to show that the metric is complete.

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

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

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 $\mathcal{M}$, we say that μ is *stochastically dominated* by ν and write $\mu \preceq_{\rm sd} \nu$ if $\mu(S) = \nu(S)$ and $\mu(I) \leq \nu(I)$ for all $I \in i\mathcal{B}$. Equivalently, $\mu(S) = \nu(S)$ and $\mu(h) \leq \nu(h)$ for all h in iH or iH_0. See [18].

One important partial order on S is the *identity order*, where $x \preceq y$ if and only if $x = y$. Then $i\mathcal{B} = \mathcal{B}$, $ibS = bS$, $iH = H$, $iH_0 = H_0$, and $\mu \preceq_{\rm sd} \nu$ if and only if $\mu = \nu$.

Remark 3.1. Since S is a partially ordered Polish space, for any μ, ν in $\mathcal{P}$, we have $\mu = \nu$ whenever $\mu(C) = \nu(C)$ for all $C \in i\mathfrak{C}$, or, equivalently, $\mu(h) = \nu(h)$ for all continuous $h \in ibS$. See [18, Lemma 1]. Hence, $\mu \preceq_{\rm sd} \nu$ and $\nu \preceq_{\rm sd} \mu$ imply that $\mu = \nu$.

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

$$\sup_{h \in iH} |\lambda(h)| = \max\left\{ \sup_{h \in iH} \lambda(h),\ \sup_{h \in iH} (-\lambda)(h) \right\}. \tag{3.1}$$

The proof is given in Appendix A.1. We can easily check that

$$\lambda \in \mathcal{M}_s \quad \text{and} \quad \lambda(S) = 0 \quad \Longrightarrow \quad \sup_{h \in iH} \lambda(h) = 2 \sup_{h \in iH_0} \lambda(h). \tag{3.2}$$

3.2. Definition of ordered affinity

For each pair $(\mu, \nu) \in \mathcal{M} \times \mathcal{M}$, let

$$\Phi(\mu, \nu) := \{(\mu', \nu') \in \mathcal{M} \times \mathcal{M}\colon \mu' \leq \mu,\ \nu' \leq \nu,\ \mu' \preceq_{\rm sd} \nu'\}.$$

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, since $(\mu \wedge \nu, \mu \wedge \nu)$ is an element of $\Phi(\mu, \nu)$.

Example 3.1. If μ and ν are two measures satisfying $\mu \preceq_{\rm sd} \nu$, then $(\mu, \nu) \in \Phi(\mu, \nu)$.

Example 3.2. Given Bernoulli distributions $\mu = (\delta_1 + \delta_2)/2$ and $\nu = (\delta_0 + \delta_1)/2$, we have $(\mu', \nu') \in \Phi(\mu, \nu)$ when $\mu' = \nu' = \delta_1/2$.

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

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

(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

$$\alpha_O(\mu, \nu) := \sup\{\mu'(S)\colon (\mu', \nu') \in \Phi(\mu, \nu)\}, \tag{3.3}$$

https://doi.org/10.1017/jpr.2019.2 Published online by Cambridge University Press

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 $(\mu, \nu) \in \mathcal{M} \times \mathcal{M}$.*

Proof. Fix $(\mu, \nu) \in \mathcal{M} \times \mathcal{M}$, and let $s := \alpha_O(\mu, \nu)$. From the definition, we can take sequences $\{\mu'_n\}$ and $\{\nu'_n\}$ in $\mathcal{M}$ such that $(\mu'_n, \nu'_n) \in \Phi(\mu, \nu)$ for all $n \in \mathbb{N}$ and $\mu'_n(S) \uparrow s$. Since $\mu'_n \le \mu$ and $\nu'_n \le \nu$ for all $n \in \mathbb{N}$, Prokhorov's theorem [11, 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 $\mu', \nu' \in \mathcal{M}$. We claim that (μ', ν') is a maximal ordered component pair.

Since $\mu'_{n_k} \xrightarrow{w} \mu'$ and $\mu'_n \le \mu$ for all $n \in \mathbb{N}$, Theorem 1.5.5 of [15] implies that, for any Borel set B, we have $\mu'_{n_k}(B) \to \mu'(B)$ in $\mathbb{R}$. Hence, $\mu'(B) \le \mu(B)$ and, in particular, $\mu' \le \mu$. An analogous argument gives $\nu' \le \nu$. 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 $\mu'_n(I) \le \nu'_n(I)$ for all $n \in \mathbb{N}$ gives $\mu'(I) \le \nu'(I)$. Thus, $\mu' \preceq_{\text{sd}} \nu'$. 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 (3.3) 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 $(\mu \wedge \nu, \mu \wedge \nu)$ is an ordered component pair for (μ, ν), we have

$$0 \le \alpha(\mu, \nu) \le \alpha_O(\mu, \nu), \tag{3.4}$$

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 the identity order then $\alpha_O = \alpha$ on $\mathcal{M} \times \mathcal{M}$.*

Proof. Fix $(\mu, \nu) \in \mathcal{M} \times \mathcal{M}$, and let '$\preceq$' be the identity order ($x \preceq y$ if and only if $x = y$). Then '$\preceq_{\text{sd}}$' also corresponds to equality, from which it follows that the supremum in (3.3) is attained by $\mu \wedge \nu$. Hence, $\alpha_O(\mu, \nu) = \alpha(\mu, \nu)$. □

3.3. Properties of ordered affinity

Let us 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 set to the identity order.)

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

(a) $0 \le \alpha_O(\mu, \nu) \le \min\{\mu(S), \nu(S)\}$,

(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* $c \ge 0$.

Proof. Fix $(\mu, \nu) \in \mathcal{M} \times \mathcal{M}$. Claim (a) follows directly from the definitions. Regarding claim (b), suppose first that $\mu \preceq_{\text{sd}} \nu$. Then $(\mu, \nu) \in \Phi(\mu, \nu)$ and, hence, $\alpha_O(\mu, \nu) = \mu(S)$.

https://doi.org/10.1017/jpr.2019.2 Published online by Cambridge University Press

Conversely, if $\alpha_O(\mu, \nu) = \mu(S)$ then, since the only component $\mu' \leq \mu$ with $\mu'(S) = \mu(S)$ is μ itself, we must have $(\mu, \nu') \in \Phi(\mu, \nu')$ for some $\nu' \leq \nu$ with $\mu \preceq_{\text{sd}} \nu'$. But then $\mu(I) \leq \nu'(I) \leq \nu(I)$ for any $I \in i\mathcal{B}$. Hence, $\mu \preceq_{\text{sd}} \nu$.

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

$$c\alpha_O(\mu, \nu) = c\mu'(S) \leq \alpha_O(c\mu, c\nu). \tag{3.5}$$

For reverse inequality, we can apply (3.5) again to get

$$\alpha_O(c\mu, c\nu) = c\left(\frac{1}{c}\right)\alpha_O(c\mu, c\nu) \leq c\alpha_O(\mu, \nu).$$ □

3.4. Equivalent representations

In (2.2) 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* $(\mu, \nu) \in \mathcal{P} \times \mathcal{P}$, *we have*

$$\alpha_O(\mu, \nu) = 1 - \sup_{I \in i\mathcal{B}} \{\mu(I) - \nu(I)\} = \max_{(X,Y) \in \mathcal{C}(\mu,\nu)} \mathbb{P}\{X \preceq Y\}. \tag{3.6}$$

Evidently (2.2) is a special case of (3.6) because (3.6) reduces to (2.2) 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 (2.1). 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 $d\mathfrak{C}$. In particular,

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

One of the assertions of Theorem 3.1 is the existence of a coupling $(X, Y) \in \mathcal{C}(\mu, \nu)$ 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 3.3. For $(x, y) \in S \times S$, we have

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

as can easily be verified from the definition or either of the alternative representations in (3.6). The map $(x, y) \mapsto \mathbf{1}_{\mathbb{G}}(x, y)$ is measurable due to the Polish assumption. As a result, for any $(X, Y) \in \mathcal{C}(\mu, \nu)$, we have

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

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

3.5. Comments on Theorem 3.1

The existence of an order maximal coupling shown in Theorem 3.1 implies two well-known results that are usually treated separately. One is the Nachbin–Strassen theorem (see, e.g. Theorem 1 of [19] or Chapter IV of [22]), which states the existence of a coupling $(X, Y) \in \mathcal{C}(\mu, \nu)$ attaining $\mathbb{P}\{X \preceq Y\} = 1$ whenever $\mu \preceq_{\text{sd}} \nu$. The existence of an order maximal coupling for each (μ, ν) in $\mathcal{P} \times \mathcal{P}$ implies this statement, since, under the hypothesis that $\mu \preceq_{\text{sd}} \nu$, 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 the existence of an order maximal coupling is the existence of a maximal coupling in the standard sense (see the discussion of maximal couplings after (2.2) and the result on page 19 of [22]). Indeed, if we take '$\preceq$' to be the identity order then (3.6) reduces to (2.2), as already discussed.

4. Total ordered variation

Let S be a partially ordered Polish space. Consider the function on $\mathcal{P} \times \mathcal{P}$ given by

$$\gamma(\mu, \nu) := 2 - \alpha_O(\mu, \nu) - \alpha_O(\nu, \mu).$$

We call $\gamma(\mu, \nu)$ the *total ordered variation distance* between μ and ν. The natural comparison is with (2.3), 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 (3.4)), we have $\gamma(\mu, \nu) \leq \|\mu - \nu\|$ for all $(\mu, \nu) \in \mathcal{P} \times \mathcal{P}$.

Other, equivalent, representations of γ are available. For example, in view of (3.6), for any $(\mu, \nu) \in \mathcal{P} \times \mathcal{P}$, we have

$$\gamma(\mu, \nu) = \sup_{I \in i\mathcal{B}} (\mu - \nu)(I) + \sup_{I \in i\mathcal{B}} (\nu - \mu)(I). \tag{4.1}$$

By combining Lemma 3.1 and (3.2), we also have

$$2\gamma(\mu, \nu) = \sup_{h \in iH} (\mu - \nu)(h) + \sup_{h \in iH} (\nu - \mu)(h). \tag{4.2}$$

It is straightforward to show that

$$\sup_{I \in i\mathcal{B}} |\mu(I) - \nu(I)| \leq \gamma(\mu, \nu) \quad \text{and} \quad \sup_{D \in d\mathcal{B}} |\mu(D) - \nu(D)| \leq \gamma(\mu, \nu). \tag{4.3}$$

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

Proof. The claim that γ is a metric follows in a straightforward way from the definition or the alternative representation (4.1). For example, the triangle inequality is easy to verify using (4.1). Also, $\gamma(\mu, \nu) = 0$ implies that $\mu = \nu$ by (4.1) and Remark 3.1. □

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

$$\beta(\mu, \nu) := \sup_{h \in iH} |\mu(h) - \nu(h)|. \tag{4.4}$$

See [1] and [3]. (In [3] 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 '$\preceq$' is the identity order (since iH equals H under this order). From (3.1) we have

$$\frac{1}{2}\Big[\sup_{h\in iH}\lambda(h)+\sup_{h\in iH}(-\lambda)(h)\Big]\le\sup_{h\in iH}|\lambda(h)|\le\sup_{h\in iH}\lambda(h)+\sup_{h\in iH}(-\lambda)(h),$$

and from this and (4.2) we have

$$\gamma(\mu,\nu)\le\beta(\mu,\nu)\le 2\gamma(\mu,\nu). \tag{4.5}$$

Hence, β and γ are equivalent metrics.

The metric γ is also connected to the Wasserstein metric ([12], [13]). If ρ metrizes the topology on S then the Wasserstein distance between probability measures μ and ν is

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

The total ordered variation metric can be compared as follows. Consider the 'directed semimetric' $\hat{\rho}(x,y):=\mathbf{1}\{x\npreceq y\}$. In view of (3.6) we have

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

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

Proposition 4.1. *If $\{\mu_n\}_{n\ge 0}\subset\mathcal{P}$ is tight and $\gamma(\mu_n,\mu_0)\to 0$, then $\mu_n\xrightarrow{\text{w}}\mu_0$.*

Proof. Let $\{\mu_n\}$ and $\mu:=\mu_0$ satisfy the conditions of the proposition. Take any subsequence of $\{\mu_n\}$ and observe that, by Prokhorov's theorem, this subsequence has a subsubsequence converging weakly to some $\nu\in\mathcal{P}$. 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. □

4.2. Completeness

To obtain completeness of $(\mathcal{P},\gamma)$, we adopt the following additional assumption, which 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.)

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

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

Remark 4.1. In [3] 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 many 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.

Regarding notation, if $\{S_i\}$ are partially ordered Polish spaces over $i = 0, 1, 2, \ldots$, we often use a common symbol '$\preceq$' for the partial order on any of these spaces. On products of these spaces we use the product topology and pointwise partial order. Once again, the symbol '$\preceq$' is used for the 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 $x_1 \preceq y_1$.

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 $\mathcal{P}_1$ for all $x \in S_0$. If $S_0 = S_1 = S$, we will call P a *Markov kernel on* S, or just a Markov kernel. Following standard conventions (see, e.g. [25]), for any Markov kernel P from S_0 to S_1, any $h \in bS_1$, and $\mu \in \mathcal{P}_0$, we define $\mu P \in \mathcal{P}_1$ and $Ph \in bS_0$ via

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

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

$$(\mu \otimes P)(A \times B) = \int_A P(x, B)\mu(\mathrm{d}x).$$

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

5.1. Order affinity and monotone Markov kernels

Let S be a Polish space partially ordered by '$\preceq$'. A Markov kernel P is called *monotone* if $Ph \in ibS_0$ whenever $h \in ibS_1$. An equivalent condition is that $\mu P \preceq_{\mathrm{sd}} \nu P$ whenever $\mu \preceq_{\mathrm{sd}} \nu$; or just $P(x, \cdot) \preceq_{\mathrm{sd}} P(y, \cdot)$ whenever $x \preceq y$. It is well known (see, e.g. Proposition 1 of [19]) that if $\mu \preceq_{\mathrm{sd}} \nu$ and P is monotone, then $\mu \otimes P \preceq_{\mathrm{sd}} \nu \otimes P$. Note that, when '$\preceq$' is the identity order, 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 $\mathcal{P}_0$, then*

$$\mu' \preceq_{sd} \mu \quad \textit{and} \quad \nu \preceq_{sd} \nu' \quad \Longrightarrow \quad \alpha_O(\mu P, \nu P) \le \alpha_O(\mu' P, \nu' P).$$

Proof. Let P, μ, μ', ν, and ν' have the stated properties. In view of the equivalently representation in (3.6), the claim will be established if

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

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*

$$\alpha_O(\mu P, \nu P) \ge \alpha_O(\mu, \nu) \quad \textit{for any } \mu, \nu \textit{ in } \mathcal{P}_0.$$

https://doi.org/10.1017/jpr.2019.2 Published online by Cambridge University Press

Proof. Fix μ, ν in $\mathcal{P}_0$, 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,

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

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*

$$\alpha_O(\mu \otimes P, \nu \otimes P) = \alpha_O(\mu, \nu) \quad \textit{for any } \mu, \nu \textit{ in } \mathcal{P}_0.$$

Proof. Fix μ, ν in $\mathcal{P}_0$, 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

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

Taking the supremum over all couplings in $\mathcal{C}(\mu \otimes P, \nu \otimes P)$ 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 $\hat{\mu} \otimes P \preceq_{\text{sd}} \hat{\nu} \otimes P$. 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,

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

5.2. Monotone Markov chains

Given $\mu \in \mathcal{P}$ and Markov kernel P on S, a stochastic process $\{X_t\}_{t \geq 0}$ 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

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

(The meaning of the right-hand side is clarified in [22, Section III.8], [19, p. 903], and [25, Section 3.4] for example.) If P is a monotone Markov kernel then $(x, B) \mapsto \boldsymbol{Q}_x(B) := \boldsymbol{Q}_{\delta_x}(B)$ is a monotone Markov kernel from S to S^∞. See Propositions 1 and 2 of [19].

There are various useful results about representations of Markov chains that are ordered almost surely. One is that, if the initial conditions satisfy $\mu \preceq_{\text{sd}} \nu$ 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 [19].) 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 $\boldsymbol{Q}_x$ is likewise monotone. By Lemma 5.3 we have

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

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

$$\mathbb{P}\{X_t \preceq Y_t \text{ for all } t \geq 0\} = \alpha_O(\mu, \nu).$$

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

5.3. Nonexpansiveness

It is well known that every Markov kernel is nonexpansive with respect to the total variation norm, so that

$$\|\mu P - \nu P\| \leq \|\mu - \nu\| \quad \text{for all } (\mu, \nu) \in \mathcal{P} \times \mathcal{P}. \tag{5.1}$$

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

$$\gamma(\mu P, \nu P) \leq \gamma(\mu, \nu) \quad \text{for all } (\mu, \nu) \in \mathcal{P} \times \mathcal{P}. \tag{5.2}$$

The bound (5.2) follows directly from Lemma 5.2. Evidently (5.1) be recovered from (5.2) 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.

5.4. An order coupling bound for Markov chains

Doeblin [7] established and exploited the coupling inequality

$$\|\mu_t - \nu_t\| \leq 2\,\mathbb{P}\{X_j \neq Y_j \text{ for any } j \leq t\}, \tag{5.3}$$

where μ_t and ν_t are the time t distributions of Markov chains $\{X_t\}$ and $\{Y_t\}$ generated by common Markov kernel P. Even when the state space is uncountable, the right-hand side of (5.3) 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 ([9], [14], [22], [25], [26], [27], [28], [31], [34]).

Consider the following generalization: given a monotone Markov kernel P on S and arbitrary $\mu, \nu \in \mathcal{P}$, we can construct Markov chains $\{X_t\}$ and $\{Y_t\}$ with common kernel P and respective initial conditions μ and ν such that

$$\gamma(\mu_t, \nu_t) \leq \mathbb{P}\{X_j \npreceq Y_j \text{ for any } j \leq t\} + \mathbb{P}\{Y_j \npreceq X_j \text{ for any } j \leq t\}. \tag{5.4}$$

This generalizes (5.3) because both the right- and left-hand sides of (5.4) reduce to (5.3) if we take '$\preceq$' to be equality.

To prove (5.4), we need only show that

$$1 - \alpha_O(\mu P^t, \nu P^t) \leq \mathbb{P}\{X_j \npreceq Y_j \text{ for any } j \leq t\}, \tag{5.5}$$

since, with (5.5) is established, we can reverse the roles of $\{X_t\}$ and $\{Y_t\}$ in (5.5) to obtain $1 - \alpha_O(\nu P^t, \mu P^t) \leq \mathbb{P}\{Y_j \npreceq X_j \text{ for any } j \leq t\}$ and then add this inequality to (5.5) to produce (5.4).

If $\{X_t\}$ and $\{Y_t\}$ are Markov chains with kernel P and initial conditions μ and ν, then (3.6) yields $\alpha_O(\mu P^t, \nu P^t) \geq \mathbb{P}\{X_t \preceq Y_t\}$. Therefore, we need only construct such chains with the additional property that

$$\mathbb{P}\{X_j \npreceq Y_j \text{ for any } j \leq t\} = \mathbb{P}\{X_t \npreceq Y_t\}. \tag{5.6}$$

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 [23, 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 $\boldsymbol{Q}_\eta := \eta \otimes M \otimes M \otimes \cdots$ be the induced joint distribution, and let $\{(X_t, Y_t)\}$ have distribution $\boldsymbol{Q}_\eta$ on $(S \times S)^\infty$. By construction, X_t has distribution μP^t and Y_t has distribution νP^t. Moreover, (5.6) is valid because $\mathbb{G}$ is absorbing for M, and, hence, $\mathbb{P}\{(X_j, Y_j) \notin \mathbb{G}$ for any $j \leq t\} = \mathbb{P}\{(X_t, Y_t) \notin \mathbb{G}\}$.

5.5. Uniform ergodicity

Let S be a partially ordered Polish space satisfying Assumption 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

$$\sigma(P) := \inf_{(x,y) \in S \times S} \alpha_O(P_x, P_y), \tag{5.7}$$

which can be understood as an order-theoretic extension of the Markov–Dobrushin coefficient of ergodicity ([5], [32]). It reduces to the usual notion when '$\preceq$' is equality.

Theorem 5.1. *If P is monotone then*

$$\gamma(\mu P, \nu P) \leq (1 - \sigma(P))\, \gamma(\mu, \nu) \quad \textit{for all } (\mu, \nu) \in \mathcal{P} \times \mathcal{P}. \tag{5.8}$$

Thus, strict positivity of $\sigma(P)$ implies that $\mu \mapsto \mu P$ is a contraction map on $(\mathcal{P}, \gamma)$. Moreover, in many settings, the bound in (5.8) cannot be improved upon, as we now see.

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,*

$$\textit{for all } \xi > \sigma(P), \ \textit{there exists } \mu, \nu \in \mathcal{P} \textit{ such that } \gamma(\mu P, \nu P) > (1 - \xi)\gamma(\mu, \nu). \tag{5.9}$$

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 $\mathcal{P}$, and*

$$\gamma(\mu P^t, \pi) \leq (1 - \sigma(P^m))^{\lfloor t/m \rfloor}\, \gamma(\mu, \pi) \quad \textit{for all } \mu \in \mathcal{P}, t \geq 0. \tag{5.10}$$

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

Proof of Corollary 5.1. 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 $\pi \in \mathcal{P}$, follows from Theorem 5.1 applied to P^m, Banach's contraction mapping theorem, and the completeness of $(\mathcal{P}, \gamma)$ shown in Theorem 4.1. The bound in (5.10) follows from (5.8) applied to P^m and the nonexpansiveness of P in the metric γ (see (5.2)). □

5.6. Applications from the introduction

Let us see how Corollary 5.1 can be used to show stability of the two models discussed in the Introduction, beginning with the monotone model in (1.1). The state space is $S = [0, 1]$ with its standard order and all assumptions are as per the discussion immediately following (1.1). We let P be the corresponding Markov kernel and consider the following coupling of (P_1, P_0). Let W be a draw from the Bernoulli$(\frac{1}{2})$ distribution, and let $V = 1 - W$. Then P_1 and P_0 are respectively the distributions of

$$X := \frac{1 + W}{2} \quad \text{and} \quad Y := \frac{0 + V}{2} = \frac{1 - W}{2}.$$

https://doi.org/10.1017/jpr.2019.2 Published online by Cambridge University Press

Since $X \leq Y$ if and only if $W = 0$, we have, by Lemma 5.1 and (3.6),

$$\alpha_O(P_x, P_y) \geq \alpha_O(P_1, P_0) \geq \mathbb{P}\{X \preceq Y\} = \tfrac{1}{2}$$

for all $x, y \in S$. From the definition in (5.7) we then have $\sigma(P) \geq \frac{1}{2}$, and globally stability in the γ metric follows from Corollary 5.1.

Next we turn to the inventory model in (1.2), with state space $S = [0, K]$ and $\{W_t\}$ an i.i.d. process satisfying $\mathbb{P}\{W_t \geq w\} > 0$ for any real w. Let $\{X_t\}$ and $\{Y_t\}$ be generated by (1.2), with common shock sequence $\{W_t\}$, and respective initial conditions x, y in S. In view of (2.2), we have

$$\alpha(P_x, P_y) \geq \mathbb{P}\{X_1 = Y_1\} \geq \mathbb{P}\{W_1 \geq K\} =: \kappa > 0.$$

Letting '$\preceq$' be equality on S, we have $\alpha_O(P_x, P_y) = \alpha(P_x, P_y) \geq \kappa$, and hence $\sigma(P) \geq \kappa$. Globally stability in the γ metric follows from Corollary 5.1.

5.7. General applications

Next we show that Theorem 5.1 and in particular Corollary 5.1 cover as special cases two well-known results on Markov chain stability from the classical literature on one hand and the more recent monotone Markov chain literature on the other.

Consider first 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 [24] in a countable state space, and by Doeblin [6], Yosida and Kakutani [37], Doob [9], and many subsequent authors in a general state space. It is defined and reviewed in Chapter 16 of [25]. One of the most familiar equivalent conditions for uniform ergodicity [25, Theorem 16.0.2] is the existence of an $m \in \mathbb{N}$ and a nontrivial $\phi \in \mathcal{M}$ such that $P_x^m \geq \phi$ 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,

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

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

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

Now consider the setting of Bhattacharya and Lee [1], where $S = \mathbb{R}^n$, '$\preceq$' is the usual pointwise partial order '$\leq$' for vectors, and $\{g_t\}$ is a sequence of i.i.d. 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 used a 'splitting condition', which assumes existence of a $\bar{x} \in S$ and $m \in \mathbb{N}$ such that

(a) $s_1 := \mathbb{P}\{g_m \circ \cdots \circ g_1(y) \leq \bar{x} \text{ for all } y \in S\} > 0$, and

(b) $s_2 := \mathbb{P}\{g_m \circ \cdots \circ g_1(y) \geq \bar{x} \text{ for all } y \in S\} > 0$.

Under these assumptions, they show that $\sup_{x \in S} \beta(P^t_x, \pi)$ converges to zero exponentially fast in t, where β is the Bhattacharya metric introduced in (4.4). This finding extends earlier results by Dubins and Freedman [10] and Yahav [36] 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

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

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 $\sup_{x \in S} \gamma(P^t_x, \pi)$ converges to zero exponentially fast in t. We showed in (4.5) that $\beta \leq 2\gamma$, so the same convergence holds for the Bhattacharya metric.

We can also recover a related convergence result due to Hopenhayn and Prescott [16, Theorem 2] that is routinely applied to stochastic stability problems in economics. They assumed that S is a compact metric space with a closed partial order, and a least element a and greatest element b. They supposed that P is monotone, and that there exists an $\bar{x}$ in S and an $m \in \mathbb{N}$ such that

$$P^m(a, [\bar{x}, b]) > 0 \quad \text{and} \quad P^m(b, [a, \bar{x}]) > 0. \tag{5.11}$$

In this setting, they showed that P has a unique stationary distribution π and $\mu P^t \xrightarrow{\text{w}} \pi$ for any $\mu \in \mathcal{P}$ 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 $(X_m, Y_m) \in \mathcal{C}(P^m_b, P^m_a)$, we have

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

The last term is strictly positive by (5.11). 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 $\mu \in \mathcal{P}$. This convergence implies weak convergence by Proposition 4.1 and compactness of S.

Appendix A

In this appendix we collect the remaining proofs. Throughout, in addition to notation defined above, cbS_0 denotes all continuous functions $h \colon S \to [0, 1]$, while

$$g(\mu, \nu) := \|\mu\| - \alpha_O(\mu, \nu) \quad \text{for each } \mu, \nu \in \mathcal{M}.$$

A.1. Proofs of Section 3 results

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

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

https://doi.org/10.1017/jpr.2019.2 Published online by Cambridge University Press

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

$$h_n(x) = \max\{r \in \{r_0, \ldots, r_n\} \colon r \leq h(x)\}.$$

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

$$\lambda(h) \leq \lambda(h_n) + \frac{\|\lambda\|}{n}. \tag{A.1}$$

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

$$I_n = \{x \in S \colon h_n(x) = 1\} \subset I_{n-1} \subset \cdots \subset I_0 = S. \tag{A.2}$$

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 $f_0, \ldots, f_{n-1} \in iH_0$ and $A_0, \ldots, A_{n-1} \in i\mathcal{B}$ as follows. Let $f_0 = h_n$ and $A_0 = I_n$. Evidently,

$$\lambda(f_0) \geq \lambda(h_n), \qquad f_0(x) = 1 \quad \text{for all } x \in A_0, \qquad f_0(x) = r_{n-1} \quad \text{for all } x \in I_{n-1} \setminus A_0.$$

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

$$\begin{gathered} \lambda(f_j) \geq \lambda(h_n), \qquad f_j(x) = 1 \quad \text{for all } x \in A_j, \\ f_j(x) = r_{n-j-1} \quad \text{for all } x \in I_{n-j-1} \setminus A_j. \end{gathered} \tag{A.3}$$

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

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

If $\lambda(I_{n-j-1} \setminus A_j) \leq 0$ 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 $\lambda(f_{j+1}) - \lambda(f_j) = (r_{n-j-2} - r_{n-j-1}) \lambda(I_{n-j-1} \setminus A_j) \geq 0$. We also have (A.1) by construction. Thus, in both cases, we have (A.3) with j replaced by $j+1$. Continuing this way, we see that (A.3) holds for all $j = 0, \ldots, n-1$.

Let $j = n-1$ in (A.3). From the definition of r_j and (A.2) we have $r_0 = 0$ and $I_0 = S$. Thus,

$$\begin{gathered} \lambda(f_{n-1}) \geq \lambda(h_n), \qquad f_{n-1}(x) = 1 \quad \text{for all } x \in A_{n-1}, \\ f_{n-1}(x) = 0 \quad \text{for all } x \in S \setminus A_{n-1}. \end{gathered}$$

https://doi.org/10.1017/jpr.2019.2 Published online by Cambridge University Press

Since $f_{n-1} = \mathbf{1}_{A_{n-1}}$ and $A_{n-1} = I_j$ for some $j \in \{0, \ldots, n-1\}$, recalling (A.1) we have

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

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

The claim (3.1) follows from $|a| = \max\{a, -a\}$ and interchange of max and sup. □

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, $\mu' \leq \mu$, $\nu' \leq \nu$ 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 $\mathbb{P}\{X \preceq Y\} = \mu'(S) \leq \alpha_O(\mu, \nu)$. We now exhibit a coupling such that equality is attained. In doing so, we can assume that $a := \alpha_O(\mu, \nu) > 0$. If not then, for any $(X, Y) \in \mathcal{C}(\mu, \nu)$, we have $0 \leq \mathbb{P}\{X \preceq Y\} \leq \alpha_O(\mu, \nu) = 0$.

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

$$\mu^r := \frac{\mu - \mu'}{1 - a} \quad \text{and} \quad \nu^r := \frac{\nu - \nu'}{1 - a}.$$

By construction, μ^r, ν^r, μ'/a, and ν'/a are probability measures satisfying

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

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) $X' \stackrel{\mathrm{D}}{=} \mu'/a$, $Y' \stackrel{\mathrm{D}}{=} \nu'/a$, $X^r \stackrel{\mathrm{D}}{=} \mu^r$, and $Y^r \stackrel{\mathrm{D}}{=} \nu^r$,

(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 [19, Theorem 1], since $\mu'/a \preceq_{\mathrm{sd}} \nu'/a$. Now let

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

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

$$\mathbb{P}\{X \preceq Y\} \geq \mathbb{P}\{X \preceq Y,\ U \leq a\} = \mathbb{P}\{X' \preceq Y',\ U \leq a\}.$$

By independence, the right-hand side is equal to $\mathbb{P}\{X' \preceq Y'\}\mathbb{P}\{U \leq a\} = a$, so

$$\mathbb{P}\{X \preceq Y\} \geq a := \alpha_O(\mu, \nu).$$

We conclude that

$$\alpha_O(\mu, \nu) = \max_{(X,Y) \in \mathcal{C}(\mu,\nu)} \mathbb{P}\{X \preceq Y\}. \tag{A.4}$$

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

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

Specializing to $h = \mathbf{1}_I$ for some $I \in i\mathcal{B}$, we have $\mu(I) - \nu(I) \leq \mathbb{P}\{X \npreceq Y\} = 1 - \mathbb{P}\{X \preceq Y\}$. From this bound and (A.4), the proof of (3.6) will be complete if we can show that

$$\sup_{(X,Y)\in\mathcal{C}(\mu,\nu)} \mathbb{P}\{X \preceq Y\} \geq 1 - \sup_{I\in i\mathcal{B}}\{\mu(I) - \nu(I)\}. \tag{A.5}$$

To prove (A.5), let $\mathcal{B} \otimes \mathcal{B}$ be the product σ-algebra on $S \times S$ and let π_i be the ith 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 $\mathfrak{C}$ is the closed sets in S and $d\mathfrak{C}$ is the decreasing sets in $\mathfrak{C}$. Strassen's theorem [33] implies that, for any $\epsilon \geq 0$ 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 $\xi(K) \geq 1 - \epsilon$ whenever

$$\nu(F) \leq \mu(\pi_1(K \cap (S \times F))) + \epsilon \quad \text{for all } F \in \mathfrak{C}.$$

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

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

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

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

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 $\xi(\mathbb{G}) \geq 1 - \epsilon$ and ξ has marginals μ and ν.

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

$$\sup_{I\in i\mathcal{B}}\{\mu(I) - \nu(I)\} \geq \sup_{D\in d\mathfrak{C}}\{\nu(D) - \mu(D)\} = \epsilon \geq 1 - \xi(\mathbb{G}). \tag{A.6}$$

Now consider the probability space $(\Omega, \mathcal{F}, \mathbb{P}) = (S \times S, \mathcal{B} \otimes \mathcal{B}, \xi)$, 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 (A.6) implies (A.5). □

A.2. Proofs of Section 4 results

We begin with an elementary lemma.

Lemma A.1. *For any $\mu, \nu \in \mathcal{M}$, we have $\mu \leq \nu$ whenever $\mu(h) \leq \nu(h)$ for all $h \in cbS_0$.*

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

$$\mu(F) \leq \nu(F) \text{ for any closed set } F \subset S. \tag{A.7}$$

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 $0 \le f_\epsilon \le 1$, we have $f_\epsilon \in cbS_0$. Let $F_\epsilon = \{x \in S\colon \rho(x, F) < \epsilon\}$ 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,

$$\mu(F) \le \mu(f_\epsilon) \le \nu(f_\epsilon) \le \nu(F_\epsilon). \tag{A.8}$$

Since $F = \bigcap_{\epsilon > 0} F_\epsilon$, we have $\lim_{\epsilon \downarrow 0} \nu(F_\epsilon) = \nu(F)$, so letting $\epsilon \downarrow 0$ in (A.8) yields $\mu(F) \le \nu(F)$. Hence, (A.7) 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 $\mu(B) \le \mu(F) + \epsilon$. Thus, by (A.7), we have $\mu(B) \le \mu(F) + \epsilon \le \nu(F) + \epsilon \le \nu(B) + \epsilon$. Since $\epsilon > 0$ is arbitrary, this yields $\mu(B) \le \nu(B)$. Hence, $\mu \le \nu$. □

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

$$n \ge N \quad \Longrightarrow \quad \gamma(\mu, \mu_n) < \epsilon. \tag{A.9}$$

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

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

For $n \ge N$, this bound, (4.3), (A.9), and the definition of K yield

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

Hence, $\{\mu_n\}_{n \ge N}$ is tight. It follows that $\{\mu_n\}_{n \ge 1}$ is likewise tight. As a result, by Prokhorov's theorem, it has a subsequence that converges weakly to some $\mu^* \in \mathcal{P}$. 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 $m \ge n_\epsilon$. Fix $m \ge n_\epsilon$ and let $\nu := \mu_m$. For all $n \ge n_\epsilon$, we have $\gamma(\nu, \mu_n) < \epsilon$. Fixing any such $n \ge n_\epsilon$, 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

$$\tilde{\mu}_n \le \mu_n, \qquad \tilde{\nu}_n \le \nu, \qquad \|\tilde{\mu}_n\| = \|\tilde{\nu}_n\| = 1 - \epsilon, \qquad \tilde{\mu}_n \preceq_{\mathrm{sd}} \tilde{\nu}_n. \tag{A.10}$$

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 $\tilde{\mu}^*, \tilde{\nu}^* \in \mathcal{M}$ with $\|\tilde{\mu}^*\| = \|\tilde{\nu}^*\| = 1 - \epsilon$, we have

$$\mu_{n_i} \xrightarrow{\mathrm{w}} \mu^*, \qquad \tilde{\mu}_{n_i} \xrightarrow{\mathrm{w}} \tilde{\mu}^*, \qquad \tilde{\nu}_{n_i} \xrightarrow{\mathrm{w}} \tilde{\nu}^*, \qquad \tilde{\mu}_{n_i} \preceq_{\mathrm{sd}} \tilde{\nu}_{n_i} \quad \text{for all } i \in \mathbb{N}.$$

Given $h \in cbS_0$, since $\tilde{\mu}_{n_i}(h) \le \mu_{n_i}(h)$ and $\tilde{\nu}_{n_i}(h) \le \nu(h)$ for all $i \in \mathbb{N}$ by (A.10), we have $\tilde{\mu}^*(h) \le \mu^*(h)$ and $\tilde{\nu}^*(h) \le \nu(h)$ by weak convergence. Thus, $\tilde{\mu}^* \le \mu^*$ and $\tilde{\nu}^* \le \nu$ by Lemma A.1. We have $\tilde{\mu}^* \preceq_{\mathrm{sd}} \tilde{\nu}^*$ by [19, Proposition 3]. It follows that $(\tilde{\mu}^*, \tilde{\nu}^*) \in \Phi(\mu^*, \nu)$. We have $g(\mu^*, \nu) \le 1 - \|\tilde{\mu}^*\| = \epsilon$.

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

https://doi.org/10.1017/jpr.2019.2 Published online by Cambridge University Press

A.3. Proofs of Section 5 results

We begin with some lemmas.

Lemma A.2. *If P is monotone then* $\sigma(P) = \inf_{(\mu,\nu)\in\mathcal{P}\times\mathcal{P}} \alpha_O(\mu P, \nu P)$.

Proof. Let P be a monotone Markov kernel. It suffices to show that the inequality $\sigma(P) \leq \inf_{(\mu,\nu)\in\mathcal{P}\times\mathcal{P}} \alpha_O(\mu P, \nu P)$ holds, since the reverse inequality is obvious. By the definition of $\sigma(P)$ and the identities in (3.6), the claim will be established if we can show that

$$\sup_{x,y} \sup_{I\in i\mathcal{B}} \{P(x, I) - P(y, I)\} \geq \sup_{\mu,\nu} \sup_{I\in i\mathcal{B}} \{\mu P(I) - \nu P(I)\}, \tag{A.11}$$

where x and y are chosen from S and μ and ν are chosen from $\mathcal{P}$. Let s be the value of the right-hand side of (A.11) and let $\epsilon > 0$. Fix $\mu, \nu \in \mathcal{P}$ and $I \in i\mathcal{B}$ such that $\mu P(I) - \nu P(I) > s - \epsilon$, or, equivalently,

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

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, $\sup_{x,y} \sup_{I\in i\mathcal{B}}\{P(x, I) - P(y, I)\} \geq s$, as was to be shown. □

Lemma A.3. *If $\mu, \nu \in \mathcal{M}$ and $(\tilde{\mu}, \tilde{\nu})$ is an ordered component pair of (μ, ν), then*

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

Proof. Fix μ, ν in $\mathcal{M}$ 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

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

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

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

□

https://doi.org/10.1017/jpr.2019.2 Published online by Cambridge University Press

Proof of Theorem 5.1. Let $\mu, \nu \in \mathcal{P}$. 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 $\mathcal{P}$. Thus, by Lemma A.2,

$$1 - \alpha_O\left(\frac{\hat{\mu}P}{\|\hat{\mu}\|}, \frac{\hat{\nu}P}{\|\hat{\mu}\|}\right) \leq 1 - \sigma(P).$$

Applying the positive homogeneity property in Lemma 3.3 yields

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

Note that this inequality trivially holds if $\|\hat{\mu}\| = 0$. From the definition of g, we can write the same inequality as $g(\hat{\mu}P, \hat{\nu}P) \leq (1 - \sigma(P))g(\mu, \nu)$. If we apply Lemma A.3 to the latter we obtain

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

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

$$\begin{aligned}\gamma(\mu P, \nu P) &= g(\mu P, \nu P) + g(\nu P, \mu P)\\ &\leq (1 - \sigma(P))[g(\mu, \nu) + g(\nu, \mu)]\\ &= (1 - \sigma(P))\gamma(\mu, \nu),\end{aligned}$$

verifying the claim in (5.8). □

Proof of Lemma 5.4. To see that (5.9) holds, fix $\xi > \sigma(P)$ and suppose first that $\sigma(P) = 1$. Then (5.9) holds because the right-hand side of (5.9) can be made strictly negative by choosing $\mu, \nu \in \mathcal{P}$ to be distinct. Now suppose that $\sigma(P) < 1$ holds. It suffices to show that,

$$\text{for all } \epsilon > 0, \text{ there exists } x, y \in S \text{ such that } y \preceq x,\ x \npreceq y \text{ and } \alpha_0(P_x, P_y) < \sigma(P) + \epsilon. \tag{A.12}$$

Indeed, if we take (A.12) as valid, set $\epsilon := \xi - \sigma(P)$ and choose x and y to satisfy the conditions in (A.12), 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, (5.9) holds.

To show that (A.12) 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 (A.12) 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

$$\alpha_0(P_x, P_y) = \alpha_0(P_{\bar{x}}, P_y) \leq \alpha_0(P_{\bar{x}}, P_{\bar{y}}) < \sigma(P) + \delta < \sigma(P) + \epsilon.$$

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] BHATTACHARYA, R. AND LEE, O. (1988). Asymptotics of a class of Markov processes which are not in general irreducible. *Ann. Prob.* **16**, 1333–1347.
[2] BHATTACHARYA, R., MAJUMDAR, M. AND HASHIMZADE, N. (2010). Limit theorems for monotone Markov processes. *Sankhyā* A **72**, 170–190.
[3] CHAKRABORTY, S. AND RAO, B. (1998). Completeness of the Bhattacharya metric on the space of probabilities. *Statist. Probab. Lett.* **36**, 321–326.
[4] DIACONIS, P. AND FREEDMAN, D. (1999). Iterated random functions. *SIAM Review* **41**, 45–76.
[5] DOBRUSHIN, R. L. (1956). Central limit theorem for nonstationary Markov chains. *Theory Prob. Appl.* **1**, 65–80.
[6] DOEBLIN, W. (1937). 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.
[7] DOEBLIN, W. (1938). Exposé de la théorie des chaînes simples constantes de Markov à un nombre fini d'états. *Mathématique de l'Union Interbalkanique* **2**, 78–80.
[8] DOEBLIN, W. (1940). Éléments d'une théorie générale des chaînes simples constantes de Markoff. In *Ann. Sci. Ec. Norm. Supér.* **57**, 61–111.
[9] DOOB, J. L. (1953). *Stochastic Processes*. Wiley, New York.
[10] DUBINS, L. E. AND FREEDMAN, D. A. (1966). Invariant probabilities for certain Markov processes. *Ann. Math. Statist.* **37**, 837–848.
[11] DUDLEY, R. (2002). *Real Analysis and Probability*. Cambridge University Press.
[12] GIBBS, A. L. AND SU, F. E. (2002). On choosing and bounding probability metrics. *Int. Stat. Rev.* **70**, 419–435.
[13] GIVENS, C. R., SHORTT, R. M. (1984). A class of Wasserstein metrics for probability distributions. *Michigan Math. J.* **31**, 231–240.
[14] HARRIS, T. E. (1956). The existence of stationary measures for certain Markov processes. In *Proc. Third Berkeley Symp. on Mathematical Statistics and Probability*, Vol. 2, University of California Press, pp. 113–124.
[15] HERNÁNDEZ-LERMA, O. AND LASSERRE, J. (2003). *Markov Chains and Invariant Probabilities*. Springer.
[16] HOPENHAYN, H. A. AND PRESCOTT, E. C. (1992). Stochastic monotonicity and stationary distributions for dynamic economies. *Econometrica* **60**, 1387–1406.
[17] JARNER, S. AND TWEEDIE, R. (2001). Locally contracting iterated functions and stability of Markov chains. *J. Appl. Prob.* **38**, 494–507.
[18] KAMAE, T. AND KRENGEL, U. (1978). Stochastic partial ordering. *Ann. Prob.* **6**, 1044–1049.
[19] KAMAE, T., KRENGEL, U. AND O'BRIEN, G. (1977). Stochastic inequalities on partially ordered spaces. *Ann. Prob.* **5**, 899–912.
[20] KAMIHIGASHI, T. AND STACHURSKI, J. (2012). An order-theoretic mixing condition for monotone Markov chains. *Statist. Probab. Lett.* **82**, 262–267.
[21] KAMIHIGASHI, T. AND STACHURSKI, J. (2014). Stochastic stability in monotone economies. *Theoret. Econom.* **9**, 383–407.
[22] LINDVALL, T. (2002). *Lectures on the Coupling Method*. Dover.
[23] MACHIDA, M. AND SHIBAKOV, A. (2010). Monotone bivariate Markov kernels with specified marginals. *Proc. Amer. Math. Soc.* **138**, 2187–2194.
[24] MARKOV, A. A. (1906). Extension of the law of large numbers to dependent quantities. *Izv. Fiz.-Matem. Obsch. Kazan Univ.* (2nd ser.) **15**, 135–156.
[25] MEYN, S. P. AND TWEEDIE, R. L. (2012). *Markov Chains and Stochastic Stability*. Springer Science & Business Media.
[26] NUMMELIN, E. (2004). *General Irreducible Markov Chains and Non-Negative Operators*. Cambridge University Press.
[27] OREY, S. (1959). Recurrent Markov chains. *Pacific J. Math.* **9**, 805–827.
[28] PITMAN, J. W. (1974). Uniform rates of convergence for Markov chain transition probabilities. *Prob. Theory Relat. Fields* **29**, 193–227.
[29] POLLARD, D. (2002). *A User's Guide to Measure Theoretic Probability*. Cambridge University Press.
[30] RAZIN, A. AND YAHAV, J. A. (1979). On stochastic models of economic growth. *Internat. Econom. Review* **20**, 599–604.
[31] REVUZ, D. (2008). *Markov Chains*. Elsevier.
[32] SENETA, E. (1979). Coefficients of ergodicity: structure and applications. *Adv. Appl. Prob.* **11**, 576–590.
[33] STRASSEN, V. (1965). The existence of probability measures with given marginals. *Ann. Math. Statist.* **36**, 423–439.
[34] THORISSON, H. (2000). *Coupling, Stationarity, and Regeneration*. Springer, New York.
[35] WU, W. B. AND SHAO, X. (2004). Limit theorems for iterated random functions. *J. Appl. Prob.* **41**, 425–436.
[36] YAHAV, J. A. (1975). On a fixed point theorem and its stochastic equivalent. *J. Appl. Prob.* **12**, 605–611.
[37] YOSIDA, K. AND KAKUTANI, S. (1941). Operator-theoretical treatment of Markoff's process and mean ergodic theorem. *Ann. Math.* **42**, 188–228.

https://doi.org/10.1017/jpr.2019.2 Published online by Cambridge University Press