

COMPRESSIBILITY AND STOCHASTIC STABILITY OF MONOTONE MARKOV CHAINS

BY SERGEY FOSS^{1,a}  AND MICHAEL SCHEUTZOW^{2,b} 

¹*School of MACS and Maxwell Institute, Heriot-Watt University, s.foss@hw.ac.uk*

²*Institut für Mathematik, Technische Universität Berlin, ms@math.tu-berlin.de*

For a stochastically monotone Markov chain taking values in a Polish space, we present a number of conditions for existence and for uniqueness of its stationary regime, as well as for closeness of its transient trajectories. In particular, we generalise a basic result by Bhattacharya and Majumdar (*Random Dynamical Systems: Theory and Applications* (2007) Cambridge Univ. Press) where a certain form of mixing, or *splitting* condition was assumed uniformly over the state space. We do not rely on continuity properties of transition probabilities.

1. Introduction. The paper deals with various problems related to existence and uniqueness of the stationary distribution for stochastically monotone Markov chains. Namely, we consider a Markov chain on a Polish (complete and separable) metric space $(\mathcal{E}, d)$ and assume that it is equipped with a partial ordering that is compatible with the metric. We assume further that the Markov transition probability kernel preserves monotonicity (see Section 2 for further details). We introduce a concept of *compressibility* (which may be viewed as an analogue of the contractiveness property) and provide several sufficient conditions for that to hold. Further, we provide sufficient conditions for the existence and for the uniqueness of the stationary distribution based on certain coupling constructions. We complement our results with a number of examples and counter-examples.

Our research is motivated by a stability result by Bhattacharya and Majumdar [1] for a stochastically monotone Markov chain taking values in an interval of the real line (see also Bhattacharya and Majumdar [2] and Bhattacharya and Waymire [3], Corollary 19.8 and Theorem 19.9, for an extension to a closed subset of $\mathbb{R}^k$). The authors assume that a certain mixing, or *splitting* condition holds uniformly over the whole state space. The result was intensively used in a variety of applications. Motivated by models in economics, Foss, Shneer, Thomas and Worrall [11] obtained analogous results for a class of monotone Markov-modulated Markov chains and also for certain non-Markovian models. Matias and Silva [19] extended results of [1] to a more general class of orderings in $\mathbb{R}^d$. Butkovsky and Scheutzow [5, 6] considered continuous-time monotone Markov processes and provided criteria for the existence, uniqueness and weak (exponential) convergence to a stationary measure under a Lyapunov-type condition plus a *splitting* condition. Their splitting condition was similar to ours. Roughly speaking it says that for initial conditions $x \preceq y$ the chains starting from x and y can be coupled in such a way that after some finite random time the two chains are ordered in the opposite direction, so the order is *swapped*.

There are a number of surveys on monotone Markov chains and monotone iterative maps that quote and discuss the results of [1, 2]; see, for example, [12] and references therein.

The rest of our paper is structured as follows. In Section 2 we introduce basic definitions and notation, recall known results related to monotonicity of Markov chains, and provide

Received March 2024; revised November 2025.

MSC2020 subject classifications. Primary 60J05; secondary 06A06, 06F30.

Key words and phrases. Markov chain, stochastic ordering, stochastic monotonicity, existence and uniqueness of stationary regime, stability, compressibility, coupling.

useful properties. Then, in Section 3, we discuss various conditions for compressibility of Markov chains. Section 4 deals with conditions for existence, and Section 5 with conditions for uniqueness of a stationary distribution. In Section 6 we summarise the obtained results and highlight their relation to those of Bhattacharya and Majumdar [1], and provide further comments and examples. In Section 7 we comment on the representation of a Markov chain as a stochastic recursion and on relations between splitting conditions and stability. Finally, in Section 8, we mention possible generalisations of our results.

2. Couplings and monotone Markov chains.

2.1. Couplings. In this subsection, we recall and discuss several coupling constructions related to probability measures, random variables and Markov chains. Let $(\mathcal{E}_1, \mathcal{E}_1)$ and $(\mathcal{E}_2, \mathcal{E}_2)$ be measurable spaces.

Coupling of probability measures. For probability measures μ on $(\mathcal{E}_1, \mathcal{E}_1)$ and ν on $(\mathcal{E}_2, \mathcal{E}_2)$, a coupling is a probability measure λ on the product space $(\mathcal{E}_1 \times \mathcal{E}_2, \mathcal{E}_1 \otimes \mathcal{E}_2)$ having μ and ν as marginals, that is, $\mu(A) = \lambda(A \times \mathcal{E}_2)$ and $\nu(B) = \lambda(\mathcal{E}_1 \times B)$, for any $A \in \mathcal{E}_1$ and $B \in \mathcal{E}_2$.

Coupling of random variables. For two random variables, ξ and η , that are defined, in general, on two different probability spaces and take values in $(\mathcal{E}_1, \mathcal{E}_1)$ and $(\mathcal{E}_2, \mathcal{E}_2)$, respectively, a coupling is a pair (ξ', η') of random variables on some probability space $(\Omega, \mathcal{F}, \mathbf{P})$ such that $\xi =_d \xi'$ and $\eta =_d \eta'$.

Consider a time-homogeneous Markov chain (MC) $(X_n)_{n \in \mathbb{N}_0}$ on a Polish space $(\mathcal{E}, d)$ equipped with its Borel σ -algebra $\mathcal{E}$ (generated by the open sets in $\mathcal{E}$), and with transition kernel $P(x, A) = \mathbf{P}(X_1 \in A | X_0 = x)$, $x \in \mathcal{E}$, $A \in \mathcal{E}$. We assume the usual measurability assumptions to be in force (for each x , $P(x, \cdot)$ is a probability measure and, for each A , $P(x, A)$ is measurable in x) but we do not assume the Feller property (i.e., continuity of the map $x \mapsto \int f(y)P(x, dy)$ for each bounded and continuous function f). For $n = 1, 2, \dots$, denote by $P^{(n)}(x, A)$ the n -step transition kernel.

From now on, we use the notation $(X_n^x)_{n \in \mathbb{N}_0}$ for the Markov chain that starts from initial state $X_0^x = x \in \mathcal{E}$.

Coupling of two Markov chains. One can couple two chains $(X_n^x)_{n \in \mathbb{N}_0}$ and $(X_n^y)_{n \in \mathbb{N}_0}$ with the same transition kernel and initial states $X_0^x = x$ and $X_0^y = y$ on a common probability space as a collection of pairs $(X_n^x, X_n^y)_{n \in \mathbb{N}_0}$ of random variables. Note that we require each of the sequences $(X_n^x)_{n \in \mathbb{N}_0}$ and $(X_n^y)_{n \in \mathbb{N}_0}$ to form Markov chains, but their joint distribution may be arbitrary. In particular, the sequence $(X_n^x, X_n^y)_{n \in \mathbb{N}_0}$ may or may not form a Markov chain on the product space $\mathcal{E} \times \mathcal{E}$. If it does, then the coupling is called a *bivariate Markovian coupling* (or *bi-Markovian coupling*, for short). In this case, the coupling is characterized by a Markov kernel $\bar{P}$ on $\mathcal{E} \times \mathcal{E}$ with both marginals equal to P . A standard example of a bi-Markovian coupling is the *independent coupling* where, for any $A, B \in \mathcal{E}$ and for any $n = 0, 1, \dots$,

$$\mathbf{P}((X_{n+1}^x, X_{n+1}^y) \in A \times B | (X_m^x, X_m^y)_{m \leq n}) = \mathbf{P}(X_{n+1}^x \in A | X_n^x) \cdot \mathbf{P}(X_{n+1}^y \in B | X_n^y) \quad \text{a.s.}$$

Another popular coupling, a *monotone bi-Markovian coupling*, will be considered in the next subsection.

Sometimes, we consider couplings for a particular pair $(x, y) \in \mathcal{E} \times \mathcal{E}$ of initial conditions rather than for all pairs. In this case we call the corresponding coupling of the Markov chains starting in x and y an (x, y) -coupling. In the following we will mostly work with the concept of a marginally Markovian coupling which is a weaker concept than a bi-Markovian coupling.

DEFINITION 2.1. Let $x, y \in \mathcal{E}$ and let $(X_n^x, X_n^y)_{n \in \mathbb{N}_0}$ be an (x, y) -coupling of two Markov chains with transition kernel P . The coupling is called a *marginally Markovian* (x, y) -coupling if it satisfies

$$\begin{aligned} \mathbf{P}(X_{n+1}^x \in \cdot | \sigma(X_k^x, X_k^y)_{k \leq n}) &= P(X_n^x, \cdot) \quad \text{and} \\ \mathbf{P}(X_{n+1}^y \in \cdot | \sigma(X_k^x, X_k^y)_{k \leq n}) &= P(X_n^y, \cdot) \end{aligned}$$

almost surely for any $n \in \mathbb{N}_0$.

Clearly, every bi-Markovian coupling is marginally Markovian.

We will now consider couplings for more than one pair of initial conditions. To do this, we use two upper indices for the chain starting in x , the first is x and the second one refers to the process starting in y and which we want to couple with the first process.

PROPOSITION 2.2. Fix $x, y \in \mathcal{E}$ and suppose that $(X_n^x, X_n^y)_{n \in \mathbb{N}_0}$ is a marginally Markovian (x, y) -coupling. Assume that, for each $\hat{x}, \hat{y} \in \mathcal{E}$, $(\hat{X}_n^{\hat{x}, \hat{y}}, \hat{X}_n^{\hat{y}, \hat{x}})_{n \in \mathbb{N}_0}$ is a marginally Markovian $(\hat{x}, \hat{y})$ -coupling, all defined on a common probability space with $(X_n^x, X_n^y)_{n \in \mathbb{N}_0}$ independent of all $\hat{X}$ -variables (whatever their upper and lower indices). Let τ be a stopping time with respect to $\sigma(X_k^x, X_k^y)_{k \leq n}$, $n \in \mathbb{N}_0$. Then the sequence

$$(\tilde{X}_n, \tilde{Y}_n) := \begin{cases} (X_n^x, X_n^y) & \text{if } n \leq \tau, \\ (\hat{X}_{n-\tau}^{u, v}, \hat{X}_{n-\tau}^{v, u}) & \text{if } n > \tau \text{ and } (X_\tau^x, X_\tau^y) = (u, v) \end{cases}$$

is a marginally Markovian (x, y) -coupling.

In what follows, we call the marginally Markovian (x, y) -coupling $(\tilde{X}_n, \tilde{Y}_n)_{n \geq 0}$ a *concatenation* of the two couplings.

We will provide a straightforward proof of the proposition in the [Appendix](#).

DEFINITION 2.3. Let a set $H \in \mathcal{E} \otimes \mathcal{E}$ be given. Assume that, for any pair $(x, y) \in \mathcal{E} \times \mathcal{E}$, there is a marginally Markovian (x, y) -coupling $(X_n^x, X_n^y)_{n \geq 0}$ such that the hitting time of the set H ,

$$\tau_{x, y} \equiv \tau_{x, y}(H) = \min\{n \geq 0 : (X_n^x, X_n^y) \in H\}$$

is almost surely finite. Then the set H is *achievable* and the corresponding family of marginal Markovian couplings is called *H-successful*, or just *successful*.

REMARK 2.4. The independent and the monotone (see the next subsection) couplings are the most popular couplings in the literature. However, in some cases, there may exist a “better” coupling that makes a pair of Markov chains recurrent while the independent coupling is transient. We provide an example of a random walk with a discrete distribution of jumps. A similar example of a random walk in the plane for which the increments have a continuous distribution may be easily constructed, too.

EXAMPLE 2.5. Let $(X_n)_{n \in \mathbb{N}_0}$ be a simple random walk on the two-dimensional lattice $\mathbb{Z}^2$, $X_n^x = x + \sum_{i=1}^n \eta_i$ where $\eta_n = (\eta_{n,1}, \eta_{n,2})$, $n \in \mathbb{N}$ is an i.i.d. sequence of two-dimensional vectors with independent coordinates, with each of them taking three values, $-1, 0$ and 1 with equal probabilities $1/3$. This random walk forms an aperiodic null-recurrent Markov chain.

For $x \neq y$, an independent coupling of (X_n^x, X_n^y) is a zero-mean four-dimensional random walk which is known to be transient. On the other hand, we may use the following coupling that is a concatenation of independent and of “identical” couplings. For any

$x = x_0 \equiv (x_{0,1}, x_{0,2})$, $y = y_0 = (y_{0,1}, y_{0,2})$, any n and $x_k = (x_{k,1}, x_{k,2})$, $y_k = (y_{k,1}, y_{k,2})$, $k = n, n+1$, we let

$$\begin{aligned} & \mathbf{P}(X_{n+1}^x = x_{n+1}, X_{n+1}^y = y_{n+1} | X_n^x = x_n, X_n^y = y_n) \\ &= \mathbf{P}(X_{n+1}^x = x_{n+1} | X_n^x = x_n) \cdot \mathbf{P}(X_{n+1}^y = y_{n+1} | X_n^y = y_n) \end{aligned}$$

if $x_n \neq y_n$ (so the jumps are independent) and

$$\begin{aligned} & \mathbf{P}(X_{n+1}^x = x_{n+1}, X_{n+1}^y = y_{n+1} | X_n^x = x_n, X_n^y = y_n) \\ &= \mathbf{P}(X_{n+1}^x = x_{n+1} | X_n^x = x_n) \cdot \mathbf{1}_{x_{n+1}=y_{n+1}} \end{aligned}$$

if $x_n = y_n$ (so the jumps are identical).

This means that we proceed with independent coupling (X_n^x, X_n^y) (where $X_n^x = (X_{n,1}^x, X_{n,2}^x)$ and $X_n^y = (X_{n,1}^y, X_{n,2}^y)$) until time $\nu_1 = \nu_1(x, y)$ which is the first time when $X_{n,1}^x = X_{n,1}^y$ and $X_{n,2}^x = X_{n,2}^y$, or, equivalently, the first time when $X_n^x - X_n^y = (0, 0)$. Since a sequence $Y_n = X_n^x - X_n^y$ is a two-dimensional zero-mean random walk with independent coordinates having a symmetric distribution on the set $\{-2, -1, 0, 1, 2\}$, such a random walk is recurrent and, therefore, $\nu_1(x, y)$ is finite a.s., for any x and y from the two-dimensional lattice.

Then, starting from time ν_1 , we continue with the identical coupling, so $X_n^x = X_n^y$ for all $n \geq \nu_1$. Since X_n^x is a two-dimensional simple zero-mean random walk which is recurrent, there exists an a.s. finite time $\nu_2 \geq \nu_1$ when $X_n^x = X_n^y = (0, 0)$, for all x and y (or, equivalently, this coupling is H -successful, with $H = \{(0, 0)\} \times \{(0, 0)\}$). Note that this is a bi-Markovian coupling.

2.2. Stochastic monotonicity. Until the end of the paper, we deal with a Polish state space $(\mathcal{E}, d)$ equipped with the σ -algebra $\mathcal{E}$ generated by the open sets. We assume further that the state space is equipped with a *partial ordering* $\leq$ that is compatible with the topology generated by the metric d (see, e.g., [17] or [10]), namely, that

- (1) the set $M := \{(x, y) : x \leq y\}$ is closed in $\mathcal{E} \times \mathcal{E}$.

We call such a space an *ordered Polish space*.

We say that a set $I \in \mathcal{E}$ is *increasing* if, for any $x \in I$ and $y \in \mathcal{E}$, the inequality $x \leq y$ implies $y \in I$. We denote the set of all measurable and increasing sets by $\mathcal{I}$.

A real-valued measurable function $g : \mathcal{E} \rightarrow \mathbb{R}$ is called *monotone* or *increasing* if $x \leq y$ implies $g(x) \leq g(y)$. We denote by $\mathcal{G}$ the class of all bounded and monotone measurable functions taking values in $[0, 1]$ (restricting the range to the unit interval is convenient and does not lead to any loss of generality).

We denote the set of all probability measures on $(\mathcal{E}, d)$ by $\mathcal{M}_1(\mathcal{E})$. For $\mu_1, \mu_2 \in \mathcal{M}_1(\mathcal{E})$ we say that μ_1 is *stochastically smaller than* μ_2 and write $\mu_1 \leq \mu_2$ if $\mu_1(I) \leq \mu_2(I)$, for any set $I \in \mathcal{I}$. Note that $\mu_1 \leq \mu_2$ implies $\int g(x) d\mu_1(x) \leq \int g(x) d\mu_2(x)$ for every $g \in \mathcal{G}$ which is easy to see by considering nonnegative finite linear combinations of indicator functions and their limits.

We say that a Markov chain with transition kernel P is *stochastically monotone* (monotone, for short) if the mapping $\mathcal{G} \ni g \mapsto Pg$ preserves monotonicity, that is, $g \in \mathcal{G}$ implies that the function $Pg(x) := \int_{\mathcal{E}} g(y) P(x, dy)$ belongs to $\mathcal{G}$ as well. Note that it is enough to ensure monotonicity by considering indicator functions $g = \mathbf{1}_I$, $I \in \mathcal{I}$ only. Denoting $\mu P(\cdot) := \int P(x, \cdot) d\mu(x)$ for a probability measure μ , it is clear that P is monotone iff for every pair of probability measures $\mu_1 \leq \mu_2$, it follows that $\mu_1 P \leq \mu_2 P$.

PROPOSITION 2.6 (Strassen's theorem; see [15] or [17]). *Assume that $\mathcal{E}$ is an ordered Polish space. If $\mu_1, \mu_2 \in \mathcal{M}_1(\mathcal{E})$ satisfy $\mu_1 \leq \mu_2$ then there is a coupling λ of μ_1 and μ_2 for which $\lambda(M) = 1$.*

PROPOSITION 2.7 ([15], Proposition 4). *Assume that the $\mathcal{E}$ -valued Markov chain $(X_n)_{n \in \mathbb{N}_0}$ is monotone. For any finite or countable sequence $x_1 \leq x_2 \leq \dots \leq x_k \leq \dots$ of elements of $\mathcal{E}$, there exists a coupling of the Markov chains $(X_n^{x_1})_{n \in \mathbb{N}_0}, (X_n^{x_2})_{n \in \mathbb{N}_0}, \dots, (X_n^{x_k})_{n \in \mathbb{N}_0}, \dots$ such that $X_n^{x_1} \leq X_n^{x_2} \leq \dots \leq X_n^{x_k} \leq \dots$ a.s., for all n .*

We will need the following facts. The first one is a special case of [14], Lemma 1, while the second one is essentially [14], Theorem 2.

PROPOSITION 2.8. *Let π_1 and π_2 be two probability measures on $\mathcal{E}$ which agree on $\mathcal{I}$. Then $\pi_1 = \pi_2$.*

PROPOSITION 2.9. *The space $\mathcal{M}_1(\mathcal{E})$ of probability measures equipped with a complete metric generating weak convergence and the stochastic order $\leq$ is an ordered Polish space.*

PROOF. The fact that for each Polish space $\mathcal{E}$ the space $\mathcal{M}_1(\mathcal{E})$ is itself Polish with respect to some complete metric which generates the topology of weak convergence is well known. Antisymmetry of the partial order follows from Proposition 2.8 and the fact that this space satisfies property (1) is the content of [15], Proposition 3. $\square$

PROPOSITION 2.10. *If $\mu, \mu_1, \mu_2, \dots \in \mathcal{M}_1(\mathcal{E})$ satisfy $\mu_1 \leq \mu_2 \leq \dots$ and $\mu_i \leq \mu, i \in \mathbb{N}$, then there exist $\mathcal{E}$ -valued random variables $X, X_1, X_2, \dots$ on some probability space with respective laws $\mu, \mu_1, \mu_2, \dots$ such that $X_1 \leq X_2 \leq \dots \leq X$ almost surely.*

PROOF. Due to [15], Proposition 4, there exist random variables $X_1 \leq X_2 \leq \dots$ on some space with respective laws $\mu_1, \mu_2, \dots$. Define $F := \mathcal{E}^{\mathbb{N}}$ equipped with the product topology and componentwise ordering. Then F is again an ordered Polish space (see [15], p.901). Denote the joint law of $(X_1, X_2, \dots)$ on F by ν . Let $\bar{\nu}$ be the law of $Y, Y, \dots$ where Y has law μ . Note that, for each $i \in \mathbb{N}$, the law of $(X_1, \dots, X_i)$ is dominated by the law of $(Y, \dots, Y)$ on the space of probability measures $\mathcal{M}_1(\mathcal{E}^i)$. Hence, by [15], Proposition 2, we have $\nu \leq \bar{\nu}$. Applying Strassen's theorem yields the conclusion of the proposition. $\square$

We introduce the following condition:

(BMC, bounded-monotone-convergence) If an increasing sequence $x_1 \leq x_2 \leq \dots \leq x_n \dots$ of elements from $\mathcal{E}$ is bounded above by $y \in \mathcal{E}$ (i.e., $x_n \leq y$ for all n), then the sequence $(x_n)_{n \in \mathbb{N}}$ converges to some $x \in \mathcal{E}$.

Note that condition (BMC) and condition (1) imply that $x \leq y$ and $x_n \leq x$, for all n .

EXAMPLE 2.11. An example where condition (BMC) fails is the space $C([0, 1], \mathbb{R})$ of real-valued continuous functions on $[0, 1]$ equipped with the sup-norm and order $f \leq g$ if $f(x) \leq g(x)$ for all $x \in [0, 1]$. Indeed, a pointwise limit of an increasing sequence of continuous functions may have jumps. Another example is stated in Remark 4.2.

REMARK 2.12. A sufficient (but not necessary) condition for (BMC) to hold is that for each $a \leq b$, the interval $[a, b] := \{x : a \leq x \leq b\}$ is compact. Indeed, if $a_1 \leq a_2 \leq \dots \leq b$, then, by assumption, $[a_1, b]$ is compact and therefore each subsequence has a further subsequence which converges. If $\bar{a}$ and $\tilde{a}$ are two limit points of such subsequences, then $\bar{a} \leq \tilde{a}$ and $\tilde{a} \leq \bar{a}$, so $\bar{a} = \tilde{a}$. Therefore, the whole sequence converges and (BMC) holds.

The following is an immediate consequence of Proposition 2.10.

COROLLARY 2.13. *Assume that (BMC) holds for $\mathcal{E}$. Then a similar condition holds for $\mathcal{M}_1(\mathcal{E})$: if $\mu_1 \leq \mu_2 \leq \dots$ is an increasing sequence of probability measures which is bounded by the probability measure λ , then there exists a measure μ to which the sequence converges weakly.*

We will generally not assume condition (BMC) to hold. Whenever we want it to hold, we will explicitly say so.

The following auxiliary result will be used later.

LEMMA 2.14. *Let $\mu_n, n \in \mathbb{N}$ be a sequence of probability measures on $\mathcal{E}$ which converges weakly to $\tilde{\pi}$ and assume that π is a probability measure such that $\lim_{n \rightarrow \infty} \mu_n(I) = \pi(I)$ for every $I \in \mathcal{J}$. Then $\tilde{\pi} = \pi$.*

PROOF. If $I \in \mathcal{J}$ is closed, then

$$\tilde{\pi}(I) \geq \limsup_{n \rightarrow \infty} \mu_n(I) = \pi(I),$$

so [15], Theorem 1, implies $\tilde{\pi} \succeq \pi$. Similarly, if $I \in \mathcal{J}$ is open, then $\tilde{\pi}(I) \leq \pi(I)$ and, since complements of increasing open sets are decreasing closed sets, we can (by symmetry) again invoke [15], Theorem 1, to obtain $\tilde{\pi} \preceq \pi$, so $\tilde{\pi} = \pi$. $\square$

We conclude this section by introducing (bivariate) monotone couplings. Since monotone couplings which are not bivariate will not be relevant for us, we will make the bivariate property part of the definition.

DEFINITION AND PROPOSITION 2.15 (see [18], Theorem 2.3.). *Let P be the Markov kernel of a monotone Markov chain. There exists a Markov kernel $\overline{P}$ on $\mathcal{E} \times \mathcal{E}$ with both marginals equal to P with the property that $\overline{P}((x, y), M) = 1$ whenever $x \preceq y$. The corresponding bi-Markovian coupling is called monotone bi-Markovian coupling or simply monotone coupling.*

3. Compressibility and its sufficient conditions. As before, $(\mathcal{E}, \preceq)$ is an ordered Polish space with Borel σ -algebra $\mathcal{E}$ and P is a monotone Markov kernel on $\mathcal{E}$. Recall that $M = \{(x, y) : x \preceq y\}$ is closed. For subsets A and B of $\mathcal{E}$, we write $A \preceq B$ if $x \preceq y$ for every $x \in A$ and $y \in B$. For $x \in \mathcal{E}$ and $B \subseteq \mathcal{E}$, we write $x \preceq B$ if $x \preceq y$ for any $y \in B$.

In part (iii) of the following theorem we will need an additional assumption which we call *KS-normal* since it is similar (but not equivalent) to the concept of *normality* of an ordered Polish space (see, e.g., [10]) and it was introduced by Kamihigashi and Stachurski in [16]. For a subset K of E , we define $i(K) := \{x \in E : x \succeq y \text{ for some } y \in K\}$ and $d(K) := \{x \in E : x \preceq y \text{ for some } y \in K\}$. If K is compact then it is easy to see (and well known) that $i(K)$ and $d(K)$ are closed sets.

DEFINITION 3.1 ([16], Assumption 4.1). The ordered Polish space E is called *KS-normal* if for every compact set $C \subseteq E$, the set $i(C) \cap d(C)$ is also compact.

Recall that a family $\mathcal{M}$ of probability measures on a Polish space $\mathcal{E}$ is called *tight* if for every $\varepsilon > 0$ there exists a compact set K in $\mathcal{E}$ such that $\mu(K) \geq 1 - \varepsilon$ for every $\mu \in \mathcal{M}$.

3.1. Generic results.

THEOREM 3.2. *For a given transition kernel P , assume that the set M is achievable. Then:*

(i) *for any $x, y \in \mathcal{E}$ and as $n \rightarrow \infty$,*

$$(2) \quad \sup_{g \in \mathcal{G}} |\mathbf{E}g(X_n^x) - \mathbf{E}g(X_n^y)| \leq \max\{\mathbf{P}(\tau_{x,y}(M) > n), \mathbf{P}(\tau_{y,x}(M) > n)\} \rightarrow 0.$$

If we additionally assume that there exists a stationary distribution π for the Markov chain, then π is unique and the following hold:

(ii) *For each $x \in \mathcal{E}$,*

$$(3) \quad \lim_{n \rightarrow \infty} \sup_{g \in \mathcal{G}} \left| \mathbf{E}g(X_n^x) - \int g(y) d\pi(y) \right| \leq \lim_{n \rightarrow \infty} \sup_{g \in \mathcal{G}} \int |\mathbf{E}g(X_n^x) - \mathbf{E}g(X_n^y)| d\pi(y) = 0.$$

(iii) *If, in addition, $\mathcal{E}$ is KS-normal, then, for every $y \in \mathcal{E}$, the distribution of X_n^y converges to π weakly as $n \rightarrow \infty$.*

(iv) *If, for some $x \in \mathcal{E}$, the laws of $X_n^x, n \in \mathbb{N}_0$ are tight, then they converge to π weakly (even if $\mathcal{E}$ is not KS-normal).*

REMARK 3.3. In what follows, we call property (2) a *compressibility property*.

PROOF OF THEOREM 3.2. The proposed proof uses the concatenation of marginally Markovian couplings.

Proof of (i). Take any pair $x \neq y$. Consider an M -successful marginally Markovian (x, y) -coupling of (X_n^x) and (X_n^y) . Then $\tau \equiv \tau_{x,y}(M) < \infty$ a.s. We concatenate it with an independent monotone bi-Markovian coupling. By Proposition 2.2 the concatenation is a marginally Markovian (x, y) -coupling which, for ease of notation, we denote by $(X_n^x, X_n^y)_{n \in \mathbb{N}_0}$ (skipping the tilde on top of X).

For any function $g \in \mathcal{G}$,

$$\begin{aligned} \mathbf{E}(g(X_n^x)) &= \mathbf{E}(g(X_n^x)\mathbf{1}_{\{n < \tau\}}) + \mathbf{E}(g(X_n^x)\mathbf{1}_{\{n \geq \tau\}}) \\ &\leq \mathbf{P}(n < \tau) + \mathbf{E}(g(X_n^y)\mathbf{1}_{\{n \geq \tau\}}) \\ &\leq \mathbf{P}(n < \tau) + \mathbf{E}(g(X_n^y)). \end{aligned}$$

Thus,

$$\mathbf{E}(g(X_n^x)) - \mathbf{E}(g(X_n^y)) \leq \mathbf{P}(\tau_{x,y}(M) > n).$$

Now we may swap x and y in the previous construction, to obtain

$$\mathbf{E}(g(X_n^y)) - \mathbf{E}(g(X_n^x)) \leq \mathbf{P}(\tau_{y,x}(M) > n).$$

Therefore,

$$|\mathbf{E}(g(X_n^y)) - \mathbf{E}(g(X_n^x))| \leq \max\{\mathbf{P}(\tau_{x,y}(M) > n), \mathbf{P}(\tau_{y,x}(M) > n)\}.$$

Since the right-hand side of the above inequality does not depend on the function g and since both random variables $\tau_{x,y}$ and $\tau_{y,x}$ are proper, we conclude that

$$(4) \quad \sup_{g \in \mathcal{G}} |\mathbf{E}(g(X_n^y)) - \mathbf{E}(g(X_n^x))| \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

This proves (2).

Proof of (ii). Fix $x \in \mathcal{E}$ and a stationary distribution π . Then,

$$\begin{aligned} \sup_{g \in \mathcal{G}} \left| \mathbf{E}g(X_n^x) - \int g(y) d\pi(y) \right| &= \sup_{g \in \mathcal{G}} \left| \mathbf{E}g(X_n^x) - \int \mathbf{E}g(X_0^y) d\pi(y) \right| \\ &= \sup_{g \in \mathcal{G}} \left| \mathbf{E}g(X_n^x) - \int \mathbf{E}g(X_n^y) d\pi(y) \right| \\ &= \sup_{g \in \mathcal{G}} \left| \int (\mathbf{E}g(X_n^x) - \mathbf{E}g(X_n^y)) d\pi(y) \right| \\ &\leq \sup_{g \in \mathcal{G}} \int |\mathbf{E}g(X_n^x) - \mathbf{E}g(X_n^y)| d\pi(y) \rightarrow 0, \end{aligned}$$

as $n \rightarrow \infty$, by (2) and dominated convergence. Proposition 2.8 implies that π is unique.

(iii) Let $\varepsilon > 0$ and choose a compact set C such that $\pi(C) > 1 - \varepsilon$. By assumption, the set $\overline{C} := i(C) \cap d(C)$ is compact. Since $i(C)$ and $i(C) \setminus d(C)$ are in $\mathcal{I}$ and $d(C)^c$ and $(d(C) \setminus i(C))^c$ are in $\mathcal{J}$ it follows from part (ii) that for each $y \in E$ the sequence $\mathbf{P}(X_n^y \in \overline{C})$ converges to $\pi(\overline{C})$ which is larger than $1 - \varepsilon$. Since $\varepsilon > 0$ is arbitrary it follows that the laws of X_n^y , $n \in \mathbb{N}_0$ are tight and the claim follows from part (iv).

(iv) Tightness implies that every subsequence of $\mathcal{L}(X_n^x)$, $n \in \mathbb{N}_0$, has a further subsequence which converges weakly to some probability measure $\tilde{\pi}$ (this follows from Prokhorov's theorem). By (ii) and Lemma 2.14, we have $\tilde{\pi} = \pi$, so $\tilde{\pi}$ is independent of the choice of the subsequence and the claim follows. $\square$

REMARK 3.4. Tightness of the laws of X_n^x , $n \in \mathbb{N}_0$ does not imply the existence of a stationary distribution π (not even if $\mathcal{E}$ is KS-normal):

Consider the space $\mathcal{E} = \{0, 1, 1/2, 1/3, 1/4, \dots\}$ equipped with the Euclidean metric and the trivial order (i.e., no pair of distinct states is comparable). This defines a KS-normal ordered Polish space. Consider the chain on $\mathcal{E}$ which remains in the current state with probability $1/2$ and moves to the next state (in the order given above) with probability $1/2$. The chain is trivially monotone, $M = \{(x, x), x \in \mathcal{E}\}$ is achievable and all transition probabilities converge weakly to a Dirac measure on 0. In particular, all transition probabilities are tight but, obviously, there is no stationary measure.

REMARK 3.5. The conditions of Theorem 3.2 plus existence of a stationary distribution π do *not* imply weak convergence of all transition probabilities to π as the following example shows.

Equip $\mathcal{E} = \mathbb{N}_0 \times [-1, 1]$ with the metric induced by the Euclidean metric in $\mathbb{R}^2$ and consider the chain with transitions

$$\begin{aligned} (0, x) &\mapsto \left(0, \frac{x}{2} + \mathcal{U}\right), \\ (j, x) &\mapsto \left(j+1, \frac{x}{2} + \mathcal{U}\right), \quad j \geq 1, \end{aligned}$$

for $x \in [-1, 1]$ and $\mathcal{U}$ uniformly distributed on $\{-\frac{1}{2}, \frac{1}{2}\}$.

Obviously, the chain has a unique stationary distribution π (concentrated on the states with first component 0; the stationary distribution is in fact the uniform distribution on those states) and no sequence of transition probabilities with first component larger than 0 converges to π weakly.

Our aim is to equip $\mathcal{E}$ with a partial order which makes $(\mathcal{E}, \preceq)$ an ordered Polish space for which the chain is monotone and M is achievable.

Define

$$\begin{aligned} (i, x) < (i, y) &\quad \text{iff } x < y, \\ (i, x) < (j, y) &\quad \text{iff } j > i \text{ and } y \geq x + h_{i,j}, \\ (j, y) < (i, x) &\quad \text{iff } j > i \text{ and } y \leq x - h_{i,j}, \end{aligned}$$

where $h_{i,j}$ are strictly positive numbers which we specify now. We choose $h_{0,j} = 2^{-j}$ and $h_{i,j} = \sum_{k=i+1}^j 2^{-k}$ in case $i \geq 1$. This defines a partial order and $(\mathcal{E}, \preceq)$ is an ordered Polish space. To see that the chain is monotone, we couple all one-step transitions by choosing the same realization of the random variable $\mathcal{U}$ for all initial conditions. Then, the states after one step are almost surely ordered for any given pair of initial conditions which are ordered. Finally, let A be the set of all pairs in $\mathcal{E}$ for which the second coordinate is at most $-\frac{3}{4}$ and B the set for which the second coordinate is at least $\frac{3}{4}$. Clearly, we have $A \prec B$ and the set $A \times B$ and hence M are achievable.

3.2. *Sufficient conditions for achievability of the set M .* In this subsection, we present two sets of sufficient conditions for achievability of the set M . We start with conditions formulated in terms of coupled Markov chains.

LEMMA 3.6. *Assume that there exists a set $C \in \mathcal{C}$ such that*

(5) *the set $C \times C$ is achievable;*

and that there exist two measurable sets $A \preceq B$ and numbers $\varepsilon > 0$ and $N \geq 1$ such that

(6) *$P^{(N)}(x, A) \geq \varepsilon$ and $P^{(N)}(x, B) \geq \varepsilon$ for any $x \in C$.*

Then the set M is achievable.

Here, condition (5) is formulated in terms of a coupling of two Markov chains, and it may not be easy to verify it. Below we provide sufficient conditions for the conclusion of Lemma 3.6 to hold, that are formulated in terms of a “marginal” Markov chain.

LEMMA 3.7. *Assume there is an integer $N \geq 1$ such that condition (6) holds with the following additional constraint:*

(7) *there exists $z_0 \in C$ such that $A \preceq z_0 \preceq B$;*

and that

(8) *$\tau_x(C) := \inf\{n \geq 1 : X_{Nn}^x \in C\} < \infty$ a.s., for all $x \in \mathcal{E}$*

and, moreover, there exists a positive and increasing function $g : \mathbb{N} \rightarrow \mathbb{R}$ such that

(9)
$$\sum_{n=1}^\infty 1/g(n) < \infty$$

and

(10)
$$\sup_{x \in C} \mathbf{E} g(\tau_x(C)) < \infty.$$

Then the set $A \times B$ is achievable, and therefore the set M is achievable, too.

REMARK 3.8. Examples of functions g satisfying (9) include $g(n) = n^\alpha$ and $g(n) = n(\log(n + 1))^\alpha$, for some $\alpha > 1$.

REMARK 3.9. One can get upper bounds on the expected time to achieve the set $A \times B$ starting from states x and y . For example, if $\mathbf{E}(\tau_x(C))^\alpha$, $\mathbf{E}(\tau_y(C))^\alpha$ and $\sup_{z \in C} \mathbf{E}(\tau_z(C))^\alpha$ are all finite for some $\alpha > 1$, then there exists a coupling such that the hitting time of the set $A \times B$ has a finite α 's moment, too. See, for example, the proof of Theorem 7.3 in Chapter 10 of [20] for a similar consideration. Similarly, finiteness of exponential moments of the former three terms imply finiteness of an exponential moment of the hitting time.

PROOF OF LEMMA 3.6. The proposed proof is based on a particular coupling construction that combines iteratively two types of couplings.

Take any $x \neq y$. We construct a coupling of (X_n^x) and (X_n^y) and a random time $U < \infty$ a.s. such that $X_U^x \leq X_U^y$ a.s. Then it will follow that $\tau_{x,y}(M) \leq U$ is a.s. finite too.

Step 1. We start with a $C \times C$ -successful coupling of $(X_n^x)_{n \in \mathbb{N}_0}$ and $(X_n^y)_{n \in \mathbb{N}_0}$ until time $T_1 = \tau_{x,y}(C \times C)$. Then we concatenate it with a coupling $(\hat{X}_n^x, \hat{X}_n^y)_{n \in \mathbb{N}_0}$ with independent components as in Proposition 2.2. We continue to use the letter X instead of $\tilde{X}$ for the concatenated coupling.

If the event

$$(11) \quad E_1 = \{X_{T_1+N}^x \in A, X_{T_1+N}^y \in B\}$$

occurs, then, clearly, $X_{T_1+N}^x \leq X_{T_1+N}^y$, and we stop the procedure, letting $U = T_1 + N$. Otherwise, we proceed with

Step 2: starting at time $T_1 + N$ from values $\hat{x} = X_{T_1+N}^x$ and $\hat{y} = X_{T_1+N}^y$, we run another $C \times C$ -successful coupling until time $T_1 + N + T_2$,

$$T_2 = \min\{n > T_1 + N : (X_n^x, X_n^y) \in C \times C\}$$

and then another independent coupling at times $n = T_1 + N + T_2 + 1, \dots, T_1 + N + T_2 + N$.

If the event

$$(12) \quad E_2 = \{X_{T_1+N+T_2+N}^x \in A, X_{T_2+N-1}^y \in B\}$$

occurs, we stop and let $U = T_1 + N + T_2 + N$, otherwise continue with

Step 3: another $C \times C$ -successful coupling, followed by an independent coupling of length N . An induction argument then completes the construction.

Note that the probability of the event E_1 is not smaller than ε^2 . Further, given that E_1 does not occur, the probability of E_2 is again not smaller than ε^2 , and so on. Therefore, we stop after a random number ν of steps where ν is bounded above by a random variable having a geometric distribution with parameter ε^2 and, therefore, is finite a.s. Thus,

$$U = \sum_{i=1}^{\nu} T_i + \nu N$$

is a.s. finite too. $\square$

PROOF OF LEMMA 3.7. We start with two observations.

Observation 1: Without loss of generality, we may and will assume that the sets A and B are of the form

$$(13) \quad A = \{x \in \mathcal{E} : x \leq z_0\} \quad \text{and} \quad B = \{y \in \mathcal{E} : z_0 \leq y\}.$$

Then, in particular, A and B are closed and A is decreasing and B is increasing.

Observation 2: By stochastic monotonicity of the transition kernel, $\mathbf{P}(X_N^x \in A) \geq \varepsilon$, for any $x \in A$ and $\mathbf{P}(X_N^y \in B) \geq \varepsilon$, for any $y \in B$. In particular, these properties hold for $x = z_0$ and for $y = z_0$ and monotonicity and the fact that A is decreasing imply that for any $x \in A$ we

have $\mathbf{P}(X_N^x \in A) \geq \mathbf{P}(X_N^{z_0} \in A) \geq \varepsilon$ and, since B is increasing, $\mathbf{P}(X_N^y \in B) \geq \mathbf{P}(X_N^{z_0} \in B) \geq \varepsilon$ for any $y \in B$.

Now we turn to the main part of the proof. We first assume that $N = 1$.

For $x \in \mathcal{E}$, write for short $\tau_x = \tau_x(C)$. For $k \geq 1$, let $\chi_k^x = \min\{n \geq k : X_n^x \in C\} - k$. Further, let $R := \sup_{z \in C} \mathbf{E}g(\tau_z)$, which is finite by the assumptions in the lemma.

Clearly, for any $m \geq 0$,

$$\begin{aligned} \mathbf{P}(\chi_k^x > m) &= \mathbf{P}(\tau_x > k + m) + \sum_{i=0}^{k-1} \mathbf{P}(X_{i+1}^x \in C, \chi_i^x > k - i + m - 1) \\ &= \mathbf{P}(\tau_x > k + m) + \sum_{i=0}^{k-1} \mathbf{P}(\chi_{i+1}^x > k - i + m - 1 | X_i^x \in C) \mathbf{P}(X_i^x \in C) \\ &\leq \mathbf{P}(\tau_x > k + m) + \sum_{i=0}^{k-1} \mathbf{P}(\chi_{i+1}^x \geq k - i + m | X_i^x \in C) \\ &\leq \mathbf{P}(\tau_x > k + m) + \sum_{i=0}^{k-1} \sup_{z \in C} \mathbf{P}(\tau_z \geq k + m - i) \\ &\leq \mathbf{P}(\tau_x > k + m) + \sum_{i=0}^{k-1} \frac{R}{g(k + m - i)}. \\ &\leq \mathbf{P}(\tau_x > k + m) + R \sum_{j=0}^{\infty} 1/g(j + m + 1), \end{aligned}$$

where we applied Markov's inequality in the penultimate step. By condition (9), we may choose $k_0 = k_0(x)$ and m_0 (not depending on x !) such that, for $k \geq k_0$, we have

$$(14) \quad \mathbf{P}(\chi_k^x > m_0) \leq 1/2.$$

Hence, for $m \geq 0$ and $k \geq 1$, we obtain

$$\begin{aligned} \mathbf{P}(X_{k+m+1}^x \in A) &\geq \sum_{i=0}^m \mathbf{P}(\chi_k^x = i, X_{k+m+1}^x \in A) \\ &= \sum_{i=0}^m \mathbf{P}(X_{k+m+1}^x \in A | \chi_k^x = i) \mathbf{P}(\chi_k^x = i) \\ &\geq \sum_{i=0}^m \varepsilon^{m-i} \varepsilon \mathbf{P}(\chi_k^x = i) \\ &\geq \varepsilon^{m+1} \mathbf{P}(\chi_k^x \leq m) \end{aligned}$$

which, by (14), is bounded from below by $\frac{1}{2}\varepsilon^{m_0+1}$ in case $m = m_0$ and $k \geq k_0$. The same lower bound holds for $\mathbf{P}(X_{k+m+1}^y \in B)$.

To complete the proof of the lemma in case $N = 1$, we proceed similar to Step 2 in the proof of the previous lemma. Consider $x, y \in \mathcal{E}$ and take an *independent* coupling of the chains X^x and X^y . We showed that

$$\mathbf{P}(X_n^x \in A, X_n^y \in B) \geq \frac{1}{4} \varepsilon^{2m_0+2}$$

for all sufficiently large n (depending on x and y). Choosing $T(x, y)$ such that $\mathbf{P}(X_n^x \in A, X_n^y \in B) \geq \frac{1}{8}\varepsilon^{2m_0+2}$ for $n \geq T(x, y)$ and then, recursively, $T_1 = T(x, y)$, $T_i = T_{i-1} + T(X_{T_{i-1}}^x, X_{T_{i-1}}^y)$, $i = 2, 3, \dots$, we see that $A \times B$ is achievable.

Finally, we treat the case $N \geq 2$. We define $Y_n^x := X_{nN}^x$, $n \in \mathbb{N}_0$. The assumptions in the lemma (for X) imply that the assumptions also hold for Y with $N = 1$ and therefore, $A \times B$ is achievable for Y . Since, as we showed above, $A \times B$ is achievable via independent coupling, the same holds for X and the proof is complete. $\square$

4. Conditions for existence of a stationary regime.

PROPOSITION 4.1. *Suppose that condition (BMC) holds. Suppose in addition that the following condition holds:*

(Condition (PR)) *There are two points $x_0 \leq y_0$ such that the hitting times*

$$v^- \equiv v_{x_0}^- = \min\{n \geq 1 : X_n^{x_0} \geq x_0\} \quad \text{and} \quad v^+ \equiv v_{y_0}^+ = \min\{n \geq 1 : X_n^{y_0} \leq y_0\}$$

have finite expectations.

Then the Markov chain admits at least one stationary distribution.

PROOF OF PROPOSITION 4.1. We introduce two auxiliary Markov chains, the “lower” and the “upper”. Start with $L_0 := X_0^{x_0} = x_0$ and let v^- be the first time $n \geq 1$ when $X_n^{x_0} \geq x_0$ and define $L_n = X_n^{x_0}$ for $n < v^-$. Let $L_{v^-} = x_0$ and assume that the process follows the transition probabilities of the original chain after time v^- until it hits the set $\{x \geq x_0\}$ next time after time v^- , say, at time v_2^- , and let $L_{v_2^-} = x_0$. Continue in this way by induction. Then we obtain a regenerative Markov chain $(\tilde{L}_n)_{n \geq 0}$ with an atom at x_0 and with a finite mean return time $\mathbf{E}v^-$ to x_0 . (There will be no need to specify the transition probabilities of this chain starting from states which cannot be reached from x_0).

Define the probability distribution

$$\pi^-(\cdot) = \frac{\mathbf{E}(\sum_{n=0}^{v^- - 1} \mathbf{1}_{\{X_n^{x_0} \in \cdot\}})}{\mathbf{E}v^-}$$

and let $\{\tilde{L}_n\}$, $n \in \mathbb{N}_0$ be the associated stationary chain, that is, we start at time 0 with a π^- -distributed random variable $\tilde{L}_0$ and follow the dynamics to the chain $(\tilde{L}_n)$.

From the classical theory of regenerative Markov chains, it is known that $\{\tilde{L}_n\}$ is a stationary Markov chain and that, for any measurable set $B \in \mathcal{E}$,

$$(15) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \sum_1^n \mathbf{1}_{\{L_n \in B\}} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_1^n \mathbf{P}(L_n \in B) = \pi^-(B) \quad \text{a.s.}$$

Analogously, we introduce an “upper” chain $(U_n)_{n \in \mathbb{N}_0}$ that starts from state $U_0 = y_0$ and which moves as $X_n^{y_0}$ until it hits the set $\{y \leq y_0\}$ at time v_+ . Then let $U_{v_+} = y_0$ and restart from y_0 . Then use induction to continue the construction, with restarting from y_0 each time when hitting the set $\{y \leq y_0\}$. Then define its stationary version $\hat{U}_n$, $n \geq 0$ with stationary measure

$$\pi^+(\cdot) = \frac{\mathbf{E}(\sum_{n=0}^{v^+ - 1} \mathbf{1}_{\{X_n^{y_0} \in \cdot\}})}{\mathbf{E}v^+}.$$

As above, we conclude that

$$(16) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \sum_1^n \mathbf{1}_{\{U_n \in B\}} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_1^n \mathbf{P}(U_n \in B) = \pi^+(B) \quad \text{a.s.}$$

Recall that $x_0 \leq y_0$ and, due to monotonicity of the kernel of the original Markov chain, we get that $\mathbf{P}(L_n \in D) \leq \mathbf{P}(U_n \in D)$, for any $n \geq 0$ and any set $D \in \mathcal{J}$. Based on the second equalities in (15) and (16), we may conclude that $\pi^- \leq \pi^+$.

Next, we consider the sequence of probability measures $\mu_0 := \pi^-$, $\mu_{n+1} := \mu_n P = \pi^- P^{(n)}$, $n \in \mathbb{N}_0$. From the definition of (L_n) we see that $\pi^- P \geq \pi^-$ and since P is monotone, it follows that the sequence μ_n , $n \geq 0$ is increasing. We claim that the sequence is bounded by π^+ . To see this, recall that $\pi^- \leq \pi^+$ and, by monotonicity of P , $\pi^+ \geq \pi^+ P^{(n)} \geq \pi^- P^{(n)} = \mu_n$. By the (BMC) property and Corollary 2.13, we see that the sequence μ_n , $n \geq 0$ has a weak limit $\mu = \mu_\infty$. Since we did not assume P to enjoy the Feller property we cannot conclude that the limit is invariant under P but we can at least conclude that $\mu P \geq \mu$ and, moreover, $\mu P^{(n+1)} \geq \mu P^{(n)}$ for all $n \geq 0$. We will now apply Zorn's lemma to show the existence of a stationary measure.

Recall that the space $\mathcal{M}_1(\mathcal{E})$ equipped with the topology of weak convergence and the stochastic order is an ordered Polish space (Proposition 2.9). Let $\mathcal{M}$ be the subset of probability measures μ on $\mathcal{E}$ for which all $\mu P^{(n)}$, $n \in \mathbb{N}_0$ are bounded above by π^+ and which are superinvariant, that is, $\mu P^{(n+1)} \geq \mu P^{(n)}$, for all $n \geq 0$. Note that $\mathcal{M}$ contains π^- and that $\mu \in \mathcal{M}$ implies $\mu P \in \mathcal{M}$. In order to apply Zorn's lemma, we need to verify that for each totally ordered subset $\mathcal{T}$ of $\mathcal{M}$ there exists some upper bound in $\mathcal{M}$.

Assume this is not the case and that there exists a totally ordered subset $\mathcal{T}$ that does not have an upper bound. As a subspace of the separable metric space $\mathcal{M}_1(\mathcal{E})$, the space $\mathcal{T}$ is itself separable, so we can find a countable dense subset $\mathcal{A}$ of $\mathcal{T}$. If $\mathcal{T}$ has a maximal element, then this element is an upper bound and we are done, so assume that $\mathcal{T}$ does not possess a maximal element. Let $t \in \mathcal{T}$. We claim that there exists some $a \in \mathcal{A}$ which satisfies $a \geq t$. Since t is not maximal there exists some $u \in \mathcal{T}$ which satisfies $u \succ t$. Let a_n be a sequence in $\mathcal{A}$ which converges to u . If any of the a_n is greater or equal to t then it is an upper bound of t . Otherwise all of the a_n are smaller than t . Since $\mathcal{M}_1(\mathcal{E})$ is an ordered Polish space, the limit u of the sequence satisfies $u \leq t$ which is a contradiction. Next, we take any strictly increasing sequence $\bar{a}_n$ of elements in $\mathcal{A}$ with the property that for each $b \in \mathcal{A}$ there is some a in the sequence so that $a \geq b$. This sequence has the property that for each $t \in \mathcal{T}$ there is an element in the sequence which dominates t . Since $\bar{a}_n \leq \pi^+$, (BMC) implies that the sequence converges weakly to some $\bar{a}$. Then, $\bar{a} \in \mathcal{M}$.

Now, Zorn's lemma shows that there is at least one maximal element μ_∞ in $\mathcal{M}$. In particular, μ_∞ is superinvariant. Assume that μ_∞ is not invariant. Then $\mu_\infty P^{(2)} \geq \mu_\infty P \succ \mu_\infty$, so $\mu_\infty P \in \mathcal{M}$ is strictly larger than μ_∞ contradicting our assumption that μ_∞ is not invariant, that is, μ_∞ is invariant and the proposition is proved. $\square$

REMARK 4.2. Proposition 4.1 does not hold without the (BMC) condition. As an example, let $\mathcal{E} = \{\pm \frac{1}{n}, n \in \mathbb{N}\}$ equipped with the discrete metric and the order induced by that of the real line. Consider the order preserving Markov kernel $P(\frac{1}{n}, \{\frac{1}{n+1}\}) = P(-\frac{1}{n}, \{-\frac{1}{n+1}\}) = 1$, $n \in \mathbb{N}$. All conditions of Lemma 4.1 hold except for the (BMC) condition. Obviously, there is no stationary distribution.

REMARK 4.3. The conditions of Proposition 4.1 do not guarantee uniqueness of a stationary distribution. As an example, let $\mathcal{E} = [0, 1]$ with the natural order and metric. Let P be the identity, that is, $P(x, \{x\}) = 1$ for every $x \in [0, 1]$. All conditions of Proposition 4.1 hold and there are infinitely many stationary distributions.

5. Conditions for uniqueness of the stationary regime. Consider again a monotone Markov chain $(X_n)_{n \in \mathbb{N}_0}$ on an ordered Polish state space $\mathcal{E}$. The following theorem establishes uniqueness of a stationary distribution under different conditions than Theorem 3.2. Here, we do not assume that M is achievable.

THEOREM 5.1. *Suppose that the following condition holds:*

For any $x, y \in \mathcal{E}$, there exist two measurable sets $A_{x,y}$ and $B_{x,y}$ such that $A_{x,y} \preceq B_{x,y}$ and the following two random variables are finite a.s.:

$$(17) \quad \tau_{x,y}^x = \inf\{n \geq 0 : X_n^x \in A_{x,y}\} \quad \text{and} \quad \tau_{x,y}^y = \inf\{n \geq 0 : X_n^y \in B_{x,y}\}.$$

Then the Markov chain possesses at most one stationary distribution.

REMARK 5.2. For any $x, y \in \mathcal{E}$, if

$$(18) \quad \hat{\tau} := \inf\{n \geq 0 : X_n^y \succeq x\} < \infty \quad \text{a.s.,}$$

then condition (17) follows. Indeed, it is enough to take $A_{x,y} = \{z \in \mathcal{E} : z \preceq x\}$ and $B_{x,y} = \{z \in \mathcal{E} : z \succeq x\}$. Then $\tau_{x,y}^x \equiv 0$ and $\tau_{x,y}^y = \hat{\tau}$,

REMARK 5.3. The finiteness of random variables in (17) is equivalent to the following condition: there exists a coupling of two Markov chains X_n^x and X_n^y such that

$$(19) \quad \mathbf{P}(\tau_{x,y}^x < \infty, \tau_{x,y}^y < \infty) = 1.$$

Indeed, condition (19) implies a.s. finiteness of each of the two random variables. On the other hand, given (17), one can take the independent coupling of the two Markov chains to arrive at (19).

REMARK 5.4. The assumptions of Theorem 5.1 do not imply compressibility and if a stationary measure π exists, then weak convergence of all transition probabilities does *not* necessarily hold. As an example, take $\mathcal{E} = \{0, 1\}$ with the trivial order and the chain which jumps from 0 to 1 and from 1 to 0 with probability one. The chain is monotone and $A = B = \{0\}$ is recurrent. Clearly, (2) fails, for example, for $g = \mathbf{1}_{\{0\}}$ and $x \neq y$. The uniform distribution π on $\mathcal{E}$ is the unique stationary measure but there is no weak convergence of transition probabilities. Note that in this example the set M is not achievable.

PROOF OF THEOREM 5.1. Assume that π_1 and π_2 are two distinct stationary measures. Then there exist two distinct *ergodic* stationary measures (see, e.g., [13]). Hence we can and will assume that π_1 and π_2 are distinct and ergodic. Fix an arbitrary set $I \in \mathcal{J}$. By Birkhoff–Khinchin’s ergodic theorem, the set S_I^j of initial conditions $x \in \mathcal{E}$ for which $\frac{1}{n} \sum_{i=0}^{n-1} \mathbf{1}_{X_i^x \in I}$ converges to $\pi_j(I)$ satisfies $\pi_j(S_I^j) = 1$, $j \in \{1, 2\}$. Fix some $x \in S_I^1$ and $y \in S_I^2$. Let τ_1 be the first time when the chain $X := X^x$ hits the set $A_{x,y}$ and let τ_2 be the first time when the chain $Y := X^y$ hits the set $B_{x,y}$. Couple the chains such that chain X until time τ_1 and the chain Y until time τ_2 are independent. Afterwards, couple the chains such that $X_{\tau_1+n} \leq Y_{\tau_2+n}$ almost surely for all $n \in \mathbb{N}$. This, together with the argument above, implies $\pi_1(I) \leq \pi_2(I)$. Interchanging x and y in the arguments above leads to the conclusion that $\pi_1(I) = \pi_2(I)$ for every $I \in \mathcal{J}$. Hence $\pi_1 = \pi_2$ by Lemma 2.8 contradicting our assumption. $\square$

6. Stochastic stability. In this section, we formulate criteria for the stability of a monotone Markov chain by combining results from previous sections. By *stability*, we mean the ability of the Markov chain to *stabilise in time*, namely, the existence of a unique stationary distribution π and convergence of the distribution of X_n^x to π (in some sense) as $n \rightarrow \infty$, for any initial value $x \in \mathcal{E}$. As a corollary (formulated in Remark 6.3), we obtain a version of a basic result by Bhattacharya and Majumdar [1], Theorem 5.1 in Section 3.5; see also [3].

THEOREM 6.1. *Assume that conditions (BMC) and (PR) hold and that the set M is achievable. Then there exists a unique stationary measure π and the statements (2) and (ii) to (iv) of Theorem 3.2 hold.*

If, in addition, for $\tau_{x,y}(M)$, $x, y \in \mathcal{E}$ as in (2), we have

$$(20) \quad \sup_{x,y \in \mathcal{E}} \mathbf{P}(\tau_{x,y}(M) > n) \leq \beta_n,$$

then

$$(21) \quad \sup_{g \in \mathcal{G}} \left| \mathbf{E}g(X_n^x) - \int g(y) d\pi(y) \right| \leq \beta_n$$

and, in particular,

$$\sup_{I \in \mathcal{I}} |\mathbf{P}(X_n^x \in I) - \pi(I)| \leq \beta_n.$$

PROOF. The first claim is an immediate consequence of Proposition 4.1 and Theorem 3.2. Further, (21) follows from (2), (3), and (20). $\square$

COROLLARY 6.2. *Assume that condition (BMC) holds. Assume further that the following splitting condition holds: there exist $c \in \mathcal{E}$, $\varepsilon > 0$ and $N \in \mathbb{N}$ such that, for $A_c = \{x \in \mathcal{E} : x \preceq c\}$ and $B_c = \{x \in \mathcal{E} : x \succeq c\}$, and for any $z \in \mathcal{E}$,*

$$(22) \quad P^{(N)}(z, A_c) \equiv \mathbf{P}(X_N^z \in A_c) \geq \varepsilon \quad \text{and} \quad P^{(N)}(z, B_c) \geq \varepsilon.$$

Then the Markov chain admits a unique stationary distribution π , and, for any $x \in \mathcal{E}$, $n \in \mathbb{N}$,

$$(23) \quad \sup_{I \in \mathcal{I}} |\mathbf{P}(X_n^x \in I) - \pi(I)| \leq (1 - \varepsilon)^{\lfloor \frac{n}{N} \rfloor}.$$

In particular, there is geometric convergence of the transition probabilities to the stationary distribution in the uniform (Kolmogorov) metric.

PROOF OF COROLLARY 6.2. Note that M is achievable using an independent coupling and that (PR) holds with $x_0 = y_0 = c$. Using independent coupling, (20) holds with $\beta_n = (1 - \varepsilon^2)^{\lfloor \frac{n}{N} \rfloor}$, so (23) holds with $1 - \varepsilon^2$ instead of $1 - \varepsilon$.

To see that the convergence rate can be improved to $1 - \varepsilon$, we first consider the case $N = 1$. We can find a coupling on $\mathcal{E} \times \mathcal{E}$ such that $P((x, y), A_c \times B_c) \geq \varepsilon$ for all $x, y \in \mathcal{E}$: for given $x \neq y$, toss a coin which comes up heads with probability ε and in this case let $X_1^x \in A_c$ and $X_1^y \in B_c$ in such a way that both have the correct law. Then M is achievable and so (23) holds in case $N = 1$.

For $N \geq 2$ we argue as follows. Considering the chain X_{kN}^x for $k = 0, 1, \dots$ we just saw that

$$|\mathbf{P}(X_{kN}^x \in D) - \pi(D)| \leq (1 - \varepsilon)^k, \quad k \in \mathbb{N}_0, x \in \mathcal{E}, D \in \mathcal{I}.$$

For $m \in \{0, \dots, N - 1\}$, $x \in \mathcal{E}$, denote the law of X_m^x by $\nu_{x,m}$. Using the Markov property we obtain, for $D \in \mathcal{I}$,

$$\begin{aligned} |\mathbf{P}(X_{kN+m}^x \in D) - \pi(D)| &= |\mathbf{E}[\mathbf{P}(X_{kN}^{X_m^x} \in D) | X_1^x, \dots, X_m^x] - \pi(D)| \\ &= |\mathbf{E}[\mathbf{P}(X_{kN}^{X_m^x} \in D) | X_m^x] - \pi(D)| \\ &= \left| \int_{\mathcal{E}} \mathbf{P}(X_{kN}^y \in D) - \pi(D) d\nu_{x,m}(y) \right| \end{aligned}$$

$$\begin{aligned} &\leq \int_{\mathcal{E}} |\mathbf{P}(X_{kN}^y \in D) - \pi(D)| \, dv_{x,m}(y) \\ &\leq (1 - \varepsilon)^k, \end{aligned}$$

so (23) holds for general $N \in \mathbb{N}$. $\square$

REMARK 6.3. The set-up of Corollary 6.2 contains the case in which $\mathcal{E}$ is a G_δ subset of $\mathbb{R}^k$ for some $k \in \mathbb{N}$ (i.e., $\mathcal{E}$ is a countable intersection of open sets in $\mathbb{R}^k$). Such a set is a Polish space with respect to the trace topology (see [7], Proposition 8.1.5). If, in addition, $\preceq$ is a partial order on $\mathcal{E}$ (e.g., the componentwise order) and (BMC) holds, then Corollary 6.2 can be applied. Particular cases of G_δ sets in $\mathbb{R}^k$ equipped with the componentwise order which satisfy (BMC) are closed subsets of $\mathbb{R}^k$.

Results of this type were initially obtained in Dubins and Freedman [8] under the additional assumption of continuity of the transition kernel. The original proofs in [1, 2] are different from ours.

7. Discussion.

7.1. A Markov chain as a stochastic recursion. *Coupling representation of a Markov chain.* It is known (and follows, say, from [4], Section 2) that any Markov chain $(X_n)_{n \in \mathbb{N}_0}$ on a Polish space $(\mathcal{E}, \mathcal{E})$ may be represented as

$$(24) \quad X_{n+1} = f(X_n, \eta_n) \quad \text{for all } n,$$

where the *driving sequence* $\{\eta_n\}$ is i.i.d. and may be taken real-valued with the uniform $U(0, 1)$ distribution, and $f : \mathcal{E} \times (0, 1) \rightarrow \mathcal{E}$ a measurable function on the product space. We will say that (24) is a *stochastic recursion*, or an *SR (coupling) representation* of a Markov chain.

SR representation of two Markov chains. One can couple the chains (X_n^x) and (X_n^y) with initial states $x \neq y$ on a common probability space using recursions $X_{n+1}^x = f(X_n^x, \eta_n^x)$ and $X_{n+1}^y = f(X_n^y, \eta_n^y)$, where each of the driving sequences $\{\eta_n^x\}$ and $\{\eta_n^y\}$ is i.i.d., but their joint distribution at any time n may be arbitrary and, in particular, may depend on n and on the history and on the future of the sequence.

The following result holds.

COROLLARY 7.1. *Under the conditions of Proposition 2.7, there exists a coupling representation of two Markov chains, (X_n^x) and (X_n^y) as stochastic recursions (24) such that $X_n^x \preceq X_n^y$ a.s., for all n . Moreover, thanks to Proposition 4 from [15], a similar coupling construction is valid for any linearly ordered sequence of initial states $x_1 \preceq x_2 \preceq \dots \preceq x_n \preceq \dots$.*

This allows us to provide alternative formulations and proofs of our results using a parallel “language” of stochastic recursions. There are pros and cons for doing that. We decided to follow a more standard terminology, with stochastic kernels.

REMARK 7.2. One might expect that if a Markov chain is monotone on a partially ordered state space, then there always exists a coupling representation (24) with function f monotone in the first argument. However, this is not true, in general, as was shown in [9], Example 1.1, with a finite state space $\mathcal{E}$ containing only four elements, $|\mathcal{E}| = 4$. On the other hand, this fact is correct in the case of linear (complete) ordering (see again [9]).

7.2. Splitting conditions and stability. In this subsection, we provide several examples that may help to clarify relations between the splitting conditions and stability.

It was pointed out in [1] that the splitting conditions are close to necessary for stability in the case of finite-dimensional Euclidean spaces. Here we provide two examples that show that this is not always the case. Further, we provide an example that shows that, in general, under the assumptions of Theorem 6.1, we cannot expect convergence in a stronger sense than in the uniform metric.

EXAMPLE 7.3. Let $\mathcal{E} = [0, 1]$ and $(U_n)_{n \geq 0}$ an i.i.d. sequence of uniform on $[0, 1]$ random variables, and let

$$X_{n+1}^x = X_n^x + \frac{U_n}{2} \left(\frac{1}{2} - X_n^x \right) \quad \text{for } X_0^x = x \in [0, 1] \text{ and } n = 0, 1, \dots$$

Here (X_n^x) is stochastically monotone and weakly converges to its unique stationary distribution that is degenerate at point $1/2$. However, if $x \neq 1/2$, then $X_n^x \neq 1/2$ for all n and there is no convergence in the uniform metric and, therefore, there is no convergence in total variation. The splitting conditions of Theorem 6.1 and Corollary 6.2 do not hold for this Markov chain.

Assume now that $\mathcal{E}$ is a closed subset of $\mathbb{R}^k$ with $k > 1$. Then a Markov chain may converge to the unique limiting distribution in a strong sense (say, in total variation), while the splitting conditions from Theorem 6.1 do not hold. We provide a very simple example.

EXAMPLE 7.4. Consider the interval

$$\mathcal{E} = \{(x_1, x_2) : x_1 \geq 0, x_2 \geq 0, x_1 + x_2 = 1/\sqrt{2}\}.$$

Assume that we use the standard component-wise partial ordering: $x = (x_1, x_2) \leq y = (y_1, y_2)$ iff $x_1 \leq y_1$ and $x_2 \leq y_2$. Then the partial ordering in $\mathcal{E}$ is trivial: any point of $\mathcal{E}$ is comparable with itself only. Assume that $P(x, A) = l(A)$ for any measurable set $A \in \mathcal{E}$ where $l(A)$ is its Lebesgue measure. Then the Markov chain converges to its stationary (uniform) distribution, but there is no splitting.

We end this subsection with one more example. In Theorem 6.1, we show the convergence to the stationary distribution in the uniform metric $\sup_{I \in \mathcal{I}} |\mathbf{P}(X_n^x \in I) - \pi(I)|$. However, the stronger convergence in the total variation may not follow, in general.

EXAMPLE 7.5. Let $\mathcal{E} = [0, 1]$ and consider a Markov chain

$$X_{n+1} = \frac{1}{2}(X_n + \xi_{n+1}), \quad n = 0, 1, \dots,$$

where $\{\xi_n\}_{n \geq 1}$ is an i.i.d. sequence of Bernoulli random variables. Here the splitting condition from Theorem 6.1 clearly holds with $c = 1/2$, that is, $\mathbf{P}(X_1^x \geq 1/2) \geq 1/2$ and $\mathbf{P}(X_1^x \leq 1/2) \geq 1/2$, for all $x \in [0, 1]$. The limiting distribution is uniform on $[0, 1]$, it is (absolutely) continuous w.r.t. Lebesgue measure. However, for any n and any fixed x , the distribution of X_n^x is discrete. Therefore, there is no convergence in the total variation.

Similar conclusions may be made for many other discrete distributions of $\{\xi_n\}$ on $[0, 1]$. However, if the distribution of $\{\xi_n\}$ has an absolutely continuous component, then the convergence in the total variation follows.

8. Generalisations. One can expect the results of this paper to be generalised to non-Markovian settings and, in particular, to Markov-modulated (Markov-adapted, in another terminology) Markov chains and to stochastic recursions with stationary ergodic driving sequences; see, for example, [11] for terminology. For that, one may try to combine arguments from the current paper and from [11] where the case of a bounded state space was considered. However, such generalisations are not straightforward and require further work.

APPENDIX: PROOF OF PROPOSITION 2.2

PROOF OF PROPOSITION 2.2. We will prove that $(\tilde{X}_n, \tilde{Y}_n)_{n \geq 0}$ is a marginally Markovian (x, y) -coupling with transition kernel P . Note that τ is also a stopping time with respect to the filtration $\mathcal{G}_n := \sigma(\tilde{X}_k, \tilde{Y}_k)_{k \leq n}$, $n \in \mathbb{N}_0$. For fixed $n \in \mathbb{N}_0$ and $A \in \mathcal{E}$, we will show that

$$\mathbf{P}(\tilde{X}_{n+1} \in A | \mathcal{G}_n) = P(\tilde{X}_n, A) \quad \text{a.s.}$$

The right-hand side is $\mathcal{G}_n$ -measurable, so we have to show that, for each event $G \in \mathcal{G}_n$,

$$(25) \quad \mathbf{E}(\mathbf{1}_{\tilde{X}_{n+1} \in A} \mathbf{1}_G) = \mathbf{E}(P(\tilde{X}_n, A) \mathbf{1}_G)$$

(the corresponding equality for $\tilde{Y}$ will follow in the same way). It is enough to show (25) for G replaced by $G \cap \{\tau \geq n+1\}$ and for G replaced by $G \cap \{\tau = k\}$ for every $k \leq n$. Note that all of these events are in $\mathcal{G}_n$ since τ is a stopping time with respect to $\mathcal{G}_n$, $n \in \mathbb{N}_0$. We have

$$\begin{aligned} \mathbf{E}(\mathbf{1}_{\tilde{X}_{n+1} \in A} \mathbf{1}_{G \cap \{\tau \geq n+1\}}) &= \mathbf{E}(\mathbf{1}_{X_{n+1}^x \in A} \mathbf{1}_{G \cap \{\tau \geq n+1\}}) = \mathbf{E}(P(X_n^x, A) \mathbf{1}_{G \cap \{\tau \geq n+1\}}) \\ &= \mathbf{E}(P(\tilde{X}_n, A) \mathbf{1}_{G \cap \{\tau \geq n+1\}}), \end{aligned}$$

where the second equality holds since $(X_n^x, X_n^y)_{n \geq 0}$ is a marginally Markovian (x, y) -coupling. For $k \in \{0, \dots, n\}$ we have, by definition,

$$\mathbf{P}(\hat{X}_{n+1-k}^{X_k^x, X_k^y} \in A | \mathcal{G}_n) = P(\hat{X}_{n-k}^{X_k^x, X_k^y}, A) \quad \text{on the event } \{\tau = k\},$$

thus

$$\begin{aligned} \mathbf{E}(\mathbf{1}_{\tilde{X}_{n+1} \in A} \mathbf{1}_{G \cap \{\tau = k\}}) &= \mathbf{E}(\mathbf{1}_{\hat{X}_{n+1-k}^{X_k^x, X_k^y} \in A} \mathbf{1}_{G \cap \{\tau = k\}}) = \mathbf{E}(P(\hat{X}_{n-k}^{X_k^x, X_k^y}, A) \mathbf{1}_{G \cap \{\tau = k\}}) \\ &= \mathbf{E}(P(\tilde{X}_n, A) \mathbf{1}_{G \cap \{\tau = k\}}), \end{aligned}$$

so the claim follows. $\square$

Note added in proof. It is easy to check that KS-normality implies the BMC condition. We thank Martin Hairer for pointing this out.

Acknowledgments. The authors are very thankful to Jim Fill and Motoya Machida for discussions and valuable comments related to various coupling constructions. The authors also thank the anonymous referees and the editor for many helpful comments and suggestions.

Funding. The second author was supported in part by the London Mathematical Society, grant No. 22203.

REFERENCES

- [1] BHATTACHARYA, R. and MAJUMDAR, M. (2007). *Random Dynamical Systems: Theory and Applications*. Cambridge Univ. Press, Cambridge. [MR2290421](#) <https://doi.org/10.1017/CBO9780511618628>
- [2] BHATTACHARYA, R. and MAJUMDAR, M. (2010). Random iterates of monotone maps. *Rev. Econ. Des.* **14** 185–192. [MR2591810](#) <https://doi.org/10.1007/s10058-008-0064-5>
- [3] BHATTACHARYA, R. and WAYMIRE, E. C. (2022). *Stationary Processes and Discrete Parameter Markov Processes. Graduate Texts in Mathematics* **293**. Springer, Cham. [MR4744443](#) <https://doi.org/10.1007/978-3-031-00943-3>
- [4] BOROVKOV, A. A. and FOSS, S. G. (1992). Stochastically recursive sequences and their generalizations. *Siberian Adv. Math.* **2** 16–81. [MR1157423](#)
- [5] BUTKOVSKY, O. and SCHEUTZOW, M. (2020). Couplings via comparison principle and exponential ergodicity of SPDEs in the hypoelliptic setting. *Comm. Math. Phys.* **379** 1001–1034. [MR4163359](#) <https://doi.org/10.1007/s00220-020-03834-w>
- [6] BUTKOVSKY, O. and SCHEUTZOW, M. (2025). Correction to: Couplings via comparison principle and exponential ergodicity of SPDEs in the hypoelliptic setting. *Comm. Math. Phys.* **406** Paper No. 242, 5. [MR4951480](#) <https://doi.org/10.1007/s00220-025-05424-0>
- [7] COHN, D. L. (2013). *Measure Theory*, 2nd ed. *Birkhäuser Advanced Texts: Basler Lehrbücher*. [Birkhäuser Advanced Texts: Basel Textbooks]. Birkhäuser, New York. [MR3098996](#) <https://doi.org/10.1007/978-1-4614-6956-8>
- [8] DUBINS, L. E. and FREEDMAN, D. A. (1966). Invariant probabilities for certain Markov processes. *Ann. Math. Statist.* **37** 837–848. [MR0193668](#) <https://doi.org/10.1214/aoms/1177699364>
- [9] FILL, J. A. and MACHIDA, M. (2001). Stochastic monotonicity and realizable monotonicity. *Ann. Probab.* **29** 938–978. [MR1849183](#) <https://doi.org/10.1214/aop/1008956698>
- [10] FLANDOLI, F., GESS, B. and SCHEUTZOW, M. (2017). Synchronization by noise for order-preserving random dynamical systems. *Ann. Probab.* **45** 1325–1350. [MR3630300](#) <https://doi.org/10.1214/16-AOP1088>
- [11] FOSS, S., SHNEER, V., THOMAS, J. P. and WORRALL, T. (2018). Stochastic stability of monotone economies in regenerative environments. *J. Econom. Theory* **173** 334–360. [MR3737263](#) <https://doi.org/10.1016/j.jet.2017.11.004>
- [12] GHOSH, R. and MARECEK, J. Iterative function systems: a comprehensive survey. Available at [arXiv:2211.14661](#).
- [13] HAIRER, M. (2008). Ergodic Theory for Stochastic PDEs. Available at <http://www.hairer.org/notes/Imperial.pdf>.
- [14] KAMAE, T. and KRENGEL, U. (1978). Stochastic partial ordering. *Ann. Probab.* **6** 1044–1049. [MR0512419](#)
- [15] KAMAE, T., KRENGEL, U. and O'BRIEN, G. L. (1977). Stochastic inequalities on partially ordered spaces. *Ann. Probab.* **5** 899–912. [MR0494447](#) <https://doi.org/10.1214/aop/1176995659>
- [16] KAMIHIGASHI, T. and STACHURSKI, J. (2019). A unified stability theory for classical and monotone Markov chains. *J. Appl. Probab.* **56** 1–22. [MR3981142](#) <https://doi.org/10.1017/jpr.2019.2>
- [17] LINDVALL, T. (1999). On Strassen's theorem on stochastic domination. *Electron. Commun. Probab.* **4** 51–59. [MR1711599](#) <https://doi.org/10.1214/ECP.v4-1005>
- [18] MACHIDA, M. and SHIBAKOV, A. (2010). Monotone bivariate Markov kernels with specified marginals. *Proc. Amer. Math. Soc.* **138** 2187–2194. [MR2596058](#) <https://doi.org/10.1090/S0002-9939-10-10241-X>
- [19] MATIAS, E. and SILVA, E. (2021). Random iterations of maps on $\mathbb{R}^k$: Asymptotic stability, synchronisation and functional central limit theorem. *Nonlinearity* **34** 1577–1597. [MR4228028](#) <https://doi.org/10.1088/1361-6544/abe17b>
- [20] THORISSON, H. (2000). *Coupling, Stationarity, and Regeneration. Probability and Its Applications (New York)*. Springer, New York. [MR1741181](#) <https://doi.org/10.1007/978-1-4612-1236-2>