J. London Math. Soc. (2) 71 (2005) 236–258
DOI: 10.1112/S0024610704006088

CONTRACTIVE MARKOV SYSTEMS

IVAN WERNER

Abstract

Certain discrete-time Markov processes on locally compact metric spaces which arise from graph-directed constructions of fractal sets with place-dependent probabilities are studied. Such systems naturally extend finite Markov chains and inherit some of their properties. It is shown that the Markov operator defined by such a system has a unique invariant probability measure in the irreducible case and an attractive probability measure in the aperiodic case if the vertex sets form an open partition of the state space, the restrictions of the probability functions on their vertex sets are Dini-continuous and bounded away from zero, and the system satisfies a condition of contractiveness on average.

1. *Introduction and main definitions*

The study of Markov processes on metric spaces associated with a random iteration of maps has a long history which can be traced back to a paper by Onicescu and Mihoc [**15**]. The reader is referred to Kaijser [**10**, **11**], Barnsley *et al.* [**1**] and Stenflo [**16**] for historical reviews.

Our work can be seen as a continuation of work by Barnsley *et al.* [**1**] and Elton [**5**], the motivation for which was computer modelling of 'fractal' measures. This addresses the following heuristic question. What is the most general randomly driven finite mechanical structure on a metric space which determines a Markov operator with a unique invariant Borel probability measure?

If the metric space is finite, then one would immediately think about a directed graph with probability weights which determines a stochastic matrix, the only possible Markov operator in this case. A good candidate for such a mechanical structure handleable by a computer in a general case is a finite family of Lipschitz maps $(w_e)_{e\in E}$ on the metric space with some probability functions $(p_e)_{e\in E}$ (that is, $p_e(x) \geqslant 0$ for every $e \in E$ and $\sum_{e\in E} p_e(x) = 1$ for all x). The Markov operator which arises from it has the following form.

$$Uf := \sum_{e\in E} p_e f \circ w_e \quad \text{for all Borel measurable functions } f.$$

Obviously, for any Borel subset B, $U1_B(x)$ defines a transition probability from the point x into the set B. Such systems have already been used for modelling different stochastic processes for a long time (see the literature cited above), and they were rediscovered by Hutchinson [**8**] (although he considered only constant probability functions) for constructions of so-called self-similar or fractal sets and measures supported by them. Such systems in a general setting were studied by Barnsley *et al.* [**1**] and Elton [**5**]. However, as we will see below (Remark 3), their setting does *not* extend the case of a finite metric space, which is already very well understood. In connection with the constructions of fractal sets, Mauldin and Williams [**14**]

Received 9 October 2003; revised 12 April 2004.

2000 *Mathematics Subject Classification* 60J10, 28A80, 37A50.

introduced a finite mechanical structure which generalizes that used by Hutchinson and extends what is known on finite metric spaces. It is called a *graph-directed construction*.

We introduce a theory of systems which unifies those studied by Barnsley *et al.* and Elton with the graph-directed constructions. It has already been developed further in papers [**19**–**22**].

The theory does not claim to provide the most general model with respect to its probabilistic phenomenon, since there is a general theory of 'dependence with complete connections' [**9**] which aims at that. However, as far as the author is aware, none of the probabilistic results presented here are covered by the existing theory.

We use the following notation in the paper.

NOTATION 1. (K,d) is a metric space. All the following spaces of functions on K are real. $\mathrm{Lip}(K)$ denotes the space of all Lipschitz functions, $C_{\mathrm{C}}(K)$ denotes the space of all continuous functions with compact support, $C_{\mathrm{B}}(K)$ denotes the space of all bounded continuous functions, $C(K)$ denotes the space of all continuous functions, and $\mathcal{L}^0(K)$ denotes the space of all bounded Borel measurable functions. For a map u defined on K and $Q \subset K$, $u|_Q$ denotes the restriction of u on Q. For $f \in C_{\mathrm{B}}(K)$, $\|f\|$ is the supremum norm of f, and $\|f\|_Q$ denotes the supremum norm of $f|_Q$ for $Q \subset K$. $P(K)$ denotes the set of all Borel probability measures on K, δ_x is a Dirac probability measure concentrated on x, and $\stackrel{\mathrm{w}^*}{\to}$ means 'converges weakly* (weakly) to'.

Let $K_1, K_2, \dots, K_N$ be a partition of a metric space K into non-empty Borel subsets (we do not exclude the case $N = 1$). Furthermore, for each $i \in \{1,2,\dots,N\}$, let

$$w_{i1}, w_{i2}, \dots, w_{iL_i} : K_i \longrightarrow K$$

be a family of Borel measurable maps such that for each $j \in \{1,2,\dots,L_i\}$, there exists $n \in \{1,2,\dots,N\}$ such that $w_{ij}(K_i) \subset K_n$ (Figure 1). Finally, for each $i \in \{1,2,\dots,N\}$, let

$$p_{i1}, p_{i2}, \dots, p_{iL_i} : K_i \longrightarrow \mathbb{R}^+$$

be a family of positive Borel measurable probability functions (associated with the maps), that is, $p_{ij} > 0$ for all j and $\sum_{j=1}^{L_i} p_{ij}(x) = 1$ for all $x \in K_i$.

REMARK 1. (i) Case $N = 1$ covers the framework from [**1**] and [**5**].

(ii) In the following, all probability functions p_{ij} can be seen to be extended on the whole space by zero, and all maps w_{ij} can be seen to be extended on the whole space arbitrarily. These extensions are necessary for the definition of the Markov operator U rather than for the definition of its adjoint U^* (see Definition 4). This is another way to see how the framework from [**1**] and [**5**] can be embedded into ours.

In any arrangement of the maps, a structure of a directed (multi)graph is easily recognized.

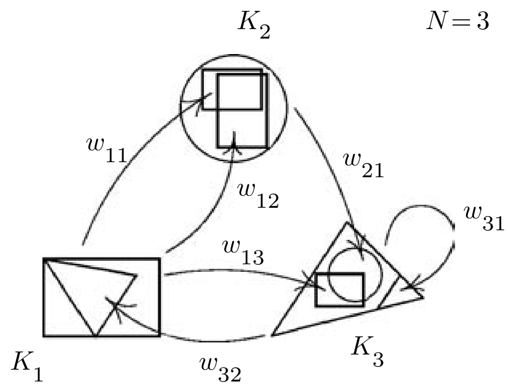

FIGURE 1.

DEFINITION 1. We call $V := \{1, \ldots, N\}$ the *set of vertices* and the subsets $K_1, \ldots, K_N$ are called the *vertex sets*. Further, we call

$$E := \{(i, n_i) : i \in \{1, \ldots, N\},\ n_i \in \{1, \ldots, L_i\}\}$$

the *set of edges* and we use the following notation.

$$p_e := p_{in} \quad \text{and} \quad w_e := w_{in} \quad \text{for } e := (i, n) \in E.$$

Each edge is provided with a direction (an arrow) by the marking of an *initial vertex* through the map

$$\begin{aligned} i : E &\longrightarrow V \\ (j, n) &\longmapsto j. \end{aligned}$$

The *terminal vertex* $t(j, n) \in V$ of an edge $(j, n) \in E$ is determined by the corresponding map through

$$t((j, n)) := k \quad :\Longleftrightarrow \quad w_{jn}(K_j) \subset K_k.$$

We call the quadruple $G := (V, E, i, t)$ a *directed* (*multi*) *graph* or *digraph*. A sequence (finite or infinite) $(\ldots, e_{-1}, e_0, e_1, \ldots)$ of edges which corresponds to a walk along the arrows of the digraph (that is, $t(e_k) = i(e_{k+1})$ for all k) is called a *path*.

DEFINITION 2. We call the family $\mathcal{M} := (K_{i(e)}, w_e, p_e)_{e \in E}$ a *Markov system*, and we call the family without probabilities, $(K_{i(e)}, w_e)_{e \in E}$, a *topological Markov system*.

DEFINITION 3. A Markov system is called *irreducible* if and only if its directed graph is irreducible, that is, there is a path from any vertex to any other. An irreducible Markov system is said to have a *period d* if and only if its directed graph has a period d, that is, the set of vertices can be partitioned into d nonempty subsets $\Omega_1, \Omega_2, \ldots, \Omega_d$ such that

$$i(e) \in \Omega_i \quad \Rightarrow \quad t(e) \in \Omega_{i+1} \mod d$$

for all $e \in E$, and d is the largest number with such a property. An irreducible Markov system with period 1 is called *aperiodic*.

DEFINITION 4. We define the Markov operator on $\mathcal{L}^0(K)$ associated with the Markov system by

$$Uf := \sum_{e\in E} p_e f \circ w_e \quad \text{for all } f \in \mathcal{L}^0(K),$$

and its adjoint operator on $P(K)$ by

$$U^*\nu(f) := \int U(f)\, d\nu \quad \text{for all } f \in \mathcal{L}^0(K) \text{ and } \nu \in P(K).$$

DEFINITION 5. We say that a probability measure μ is an *invariant probability measure* of the Markov system if and only if it is a stationary initial distribution of the associated Markov process, that is,

$$U^*\mu = \mu.$$

As in the case of a finite Markov chain, it is very useful to represent a Markov chain associated with a Markov system as a sequence of random variables defined on the product space of infinitely many copies of E.

DEFINITION 6. Set

$$\Sigma := E^{\mathbb{Z}} := \{(\ldots, \sigma_{-1}, \sigma_0, \sigma_1, \ldots) : \sigma_i \in E,\ i \in \mathbb{Z}\}$$

and

$$\Sigma^+ := E^{\mathbb{N}} := \{(\sigma_1, \sigma_2, \ldots) : \sigma_i \in E,\ i \in \mathbb{N}\}.$$

We call Σ^+ the *future* of Σ. Consider Σ and Σ^+ as provided with the product topology. Further, set

$${}_m[e_m, \ldots, e_n] := \{\sigma \in \Sigma : \sigma_m = e_m, \sigma_{m+1} = e_{m+1}, \ldots, \sigma_n = e_n\} \quad \text{for all integers } m \leqslant n,$$

and

$${}_1[e_1, \ldots, e_n]^+ := \{\sigma \in \Sigma^+ : \sigma_1 = e_1, \sigma_2 = e_2, \ldots, \sigma_n = e_n\} \quad \text{for all } n \in \mathbb{N}.$$

We call ${}_m[e_m, \ldots, e_n]$ and ${}_1[e_1, \ldots, e_n]^+$ *thin cylinder sets*. Now, for any $x \in K$ and ${}_1[e_1, \ldots, e_n]^+ \subset \Sigma^+$, define

$$P_x({}_1[e_1, \ldots, e_n]^+) := p_{e_1}(x)\, p_{e_2}\big(w_{e_1}x\big) \ldots\, p_{e_n}\big(w_{e_{n-1}} \circ \ldots \circ w_{e_1}x\big).$$

Then P_x extends uniquely to a Borel probability measure on Σ^+. Finally, for any $x \in K$ and $k \in \mathbb{N}$, set

$$Z_k^x(\sigma) := w_{\sigma_k} \circ w_{\sigma_{k-1}} \circ \ldots \circ w_{\sigma_1}(x) \quad \text{for all } \sigma \in \Sigma^+.$$

It is easy to check that the sequence of random variables $(Z_k^x)_{k\in\mathbb{N}}$ with respect to the measure P_x represents the Markov process associated with the Markov system with the initial distribution δ_x. Moreover, obviously

$$U^k f(x) = \int f \circ Z_k^x\, dP_x \quad \text{for all } x \in K,\ f \in C_{\mathrm{B}}(K) \text{ and } k \in \mathbb{N}.$$

2. *Iterations of a Markov system*

In contrast to the trivial case of finite Markov chains, we consider here the following iterations of a Markov system.

Definition 7. Let

$$\mathcal{M} := (K_i, (w_{ij})_{j\in J_i}, (p_{ij})_{j\in J_i})_{i\in I}$$

be a Markov system. Set $K_i^0 := K_i$, $w_{ij}^0 := w_{ij}$, $p_{ij}^0 := p_{ij}$ for all $i \in I^0 := I$, $j \in J^0 := J$ and

$$\mathcal{M}^0 := \mathcal{M}.$$

Let the nth iteration of $\mathcal{M}$ be defined by a Markov system

$$\mathcal{M}^n := \left(K_i^n, \left(w_{ij}^n\right)_{j\in J_i^n}, \left(p_{ij}^n\right)_{j\in J_i^n}\right)_{i\in I^n}$$

for some $n \in \mathbb{N}\cup\{0\}$. First we define the vertex sets of the $(n+1)$-iteration $\mathcal{M}^{n+1}$ by forming lumps of intersecting subsets $w_{ij}^n(K_i^n),\ i\in I^n,\ j\in J_i^n$.

This can be done using the following algorithm.

(1) Order the set of edges $E^n := \{(i,j) : i \in I^n,\ j\in J_i^n\}$ arbitrarily, say as

$$E^n = \{e_1, \dots, e_k\}, \qquad k\in\mathbb{N}.$$

(2) For each $s = 1,\dots,k$, construct recursively a set $\Xi_k(s) \subset K$ by setting

$$\Xi_0(s) := w_{e_s}\left(K_{i(e_s)}\right)$$

and

$$\Xi_m(s) := \Xi_{m-1}(s)\bigcup A_m(s),$$

where

$$A_m(s) := \begin{cases} w_{e_m}\left(K_{i(e_m)}\right) & \Xi_{m-1}(s)\cap w_{e_m}\left(K_{i(e_m)}\right)\neq\varnothing\\ \varnothing & \text{otherwise,}\end{cases}$$

for all $m = 1,\dots,k$.

(3) Set

$$\left\{K_i^{n+1}\,|\, i\in I^{n+1}\right\} := \{\Xi_k(1),\dots,\Xi_k(k)\}$$

by an arbitrary counting (without distinguishing the same elements in the right set).

Finally, we define on each vertex set $K_i^{n+1},\ i\in I^{n+1}$, the family of maps and probability functions. For each $i\in I^{n+1}$, there exists a unique index $\hat{i}\in I^n$ such that $K_i^{n+1}\subset K_{\hat{i}}^n$. Define

$$w_{ij}^{n+1} := w_{\hat{i}j}^n\Big|_{K_i^{n+1}}$$

and

$$p_{ij}^{n+1} := p_{\hat{i}j}^n\Big|_{K_i^{n+1}} \quad \text{for all } j\in J_{\hat{i}}^n.$$

Therefore $J_i^{n+1} := J_{\hat{i}}^n$. Thus $\mathcal{M}^{n+1}$ is well defined up to the indices.

Example 1. Figure 2 shows the first iteration of the Markov system from Figure 1.

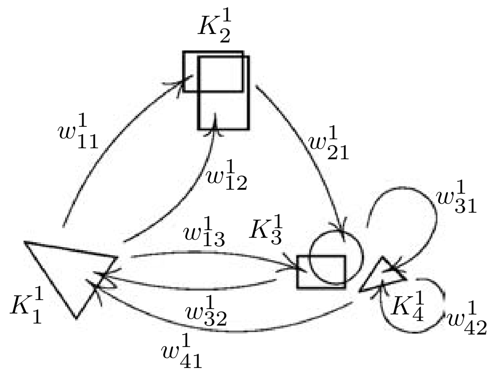

FIGURE 2.

PROPOSITION 1. *A measure is invariant wit respect to a Markov system if and only if it is invariant with respect to one of its iterations.*

Proof. This is obvious by the definition of the iterations. □

REMARK 2. Trivially in the case of finite Markov chains, such iterations do not change anything in the structure. It is known that the essential structure is preserved by such iterations in a general case as well. The directed graph associated with an iteration of a Markov system is exactly that obtained from the original directed graph by a procedure which is known in *symbolic dynamics* as *state-splitting* (see [**13**]). It is not difficult to see that the *shifts of finite type* defined by two directed graphs where one is obtained from the other by state-splitting are conjugate. It means, in particular, that such iterations of an irreducible Markov system produce irreducible Markov systems with the same period. If we decide to label the edges of the directed graph of an iteration of a Markov system simply by giving them the names of the maps of the original Markov system to which they correspond, then each iteration produces a *sofic* system, but not a proper one, because it defines the same subshift space as the original directed graph. The difference between them is only in what we consider to be separate vertex sets.

LEMMA 1. *Suppose that $(K_{i(e)}, w_e, p_e)_{e\in E}$ is an irreducible Markov system with an invariant probability measure μ. Then $\mu(K_i) > 0$ for all $i = 1, \dots, N$.*

Proof. Let $i_0 \in V$ such that $\mu(K_{i_0}) > 0$. Let $j \in V$ such that there is an edge e_0 from i_0 to j. Since all probability functions are positive on their vertex sets, $\int_{K_{i(e_0)}} p_{e_0}\, d\mu > 0$. Then, by

$$\begin{aligned}\mu(K_j) &= U^*\mu\,(K_j)\\ &= \sum_{e\in E}\int_{K_{i(e)}} p_e 1_{K_j}\circ w_e\, d\mu\\ &\geqslant \int_{K_{i(e_0)}} p_{e_0}\, d\mu,\end{aligned}$$

it follows that $\mu(K_j) > 0$. Now let $j_0 \in V$ be arbitrary. Then, by the irreducibility, there is a path from i_0 to j_0 in G. Therefore, we see, through a finite repetition of the above argument, that $\mu(K_{j_0}) > 0$. $\square$

3. *Contractive Markov systems*

If we try to represent a Bernoulli process on a finite state space, say $\{1,\ldots,N\}$, as a Markov process arising from a Markov system, then we find that the underlying Markov system consists of N contractive maps, each of them mapping the whole space $\{1,\ldots,N\}$ on a single point, and some constant probability functions corresponding to them. Any other Markov chain on this state space can be obtained by changing only the probability functions. It turns out that the contractiveness of the maps has deep roots.

Definition 8 (contractive Markov system). We call a Markov system $(K_{i(e)}, w_e, p_e)_{e\in E}$ *contractive* if and only if it satisfies the following *condition of contractiveness on average*. There exists $0 < a < 1$ such that

$$\sum_{e\in E} p_e(x)d(w_e(x), w_e(y)) \leqslant ad(x,y) \quad \text{for all } x,y \in K_i \text{ and } i \in \{1,\ldots,N\} \tag{3.1}$$

(it is understood here that the p_e are extended on the whole space by zero and the w_e arbitrarily). We call the constant a an *average contracting rate* of the Markov system.

We intend to show here that contractive Markov systems, under some conditions on the probability functions and the state space (which are a little bit stronger than is necessary just to make the Markov chain have the Feller property), exhibit a mixing behavior which is similar to that of the finite Markov chains.

From here we assume that (K,d) is a metric space in which sets of finite diameter are relatively compact. This implies that (K,d) is a locally compact separable metric space.

Definition 9. We call a function $f : (X,d) \longrightarrow \mathbb{R}$ *Dini-continuous* if and only if there is $c > 0$ such that

$$\int_0^c \frac{\phi(t)}{t}\,dt < \infty,$$

where ϕ is the *modulus of uniform continuity* of f, that is,

$$\phi(t) := \sup\{|f(x) - f(y)| : d(x,y) \leqslant t,\ x,y \in X\}.$$

It is easily seen that the Dini-continuity is weaker than the Hölder and stronger than the uniform continuity. There is a well known characterization of the Dini-continuity, which will be useful later.

Lemma 2. *Let $0 < c < 1$ and $b > 0$. A real function f is Dini-continuous if and only if*

$$\sum_{n=0}^{\infty} \phi(bc^n) < \infty,$$

where ϕ is the modulus of uniform continuity of f.

Proof. Note that in any case it holds true that

$$\int_0^b \frac{\phi(t)}{t}\,dt = \sum_{n=0}^{\infty} \int_{bc^{n+1}}^{bc^n} \frac{\phi(t)}{t}\,dt.$$

As ϕ is a non-decreasing function,

$$\phi(bc^{n+1})(1-c) \leqslant \int_{bc^{n+1}}^{bc^n} \frac{\phi(t)}{t}\,dt \leqslant \phi(bc^n)\left(\frac{1}{c}-1\right)$$

for all $n \in \mathbb{N} \cup \{0\}$. Hence

$$(1-c)\sum_{n=1}^{\infty}\phi(bc^n) \leqslant \int_0^b \frac{\phi(t)}{t}\,dt \leqslant \sum_{n=0}^{\infty}\phi(bc^n)\left(\frac{1}{c}-1\right).$$

□

Remark 3. Elton in [**5**] and Barnsley *et al.* in [**1**] considered the case $N = 1$ with Dini-continuous probability functions $(p_e)_{e\in E}$ which are bounded away from zero, and Lipschitz-continuous maps $(w_e)_{e\in E}$ such that the system satisfies the following condition of contractiveness on average. There exists $0 < r_1 < 1$ such that

$$\prod_{e\in E} d(w_e(x), w_e(y))^{p_e(x)} \leqslant r_1 d(x,y) \quad \text{for all } x, y \in K. \tag{3.2}$$

There is a widely spread view in the literature that demanding condition (3.2) rather than (3.1) (with $N = 1$) would give a weaker assumption. However, this is not quite true.

The above Elton–Barnsley setup is equivalent to that with condition (3.1) (*with* $N = 1$) *in place of* (3.2).

Proof of Lemma 2. First, observe that condition (3.1) and the boundedness away from zero of the probability functions (that is, there exists $\delta > 0$ such that $p_e \geqslant \delta$ for all $e \in E$) imply that the maps $(w_e)_{e\in E}$ are Lipschitz. Taking the logarithm of (3.2) and using its concavity reveals that (3.1) implies (3.2).

On the other hand, by [**1** Lemma 2.6], the Elton–Barnsley setup implies that there exist $r_1 < r < 1$ and $0 < q \leqslant 1$ such that

$$\sum_{e\in E} p_e(x) d(w_e(x), w_e(y))^q \leqslant r d(x,y)^q \quad \text{for all } x, y \in K. \tag{3.3}$$

By performing a remetrization $\tilde{d}(x,y) := d(x,y)^q$, which preserves the Dini-continuity of the probability functions, we can reduce it, without loss of generality, to condition (3.1). □

In [**1**], Barnsley *et al.* realized that for the proof of the attractiveness of the invariant probability measure, the condition of a uniform boundedness away from zero for the probability functions can be weakened. They came up with the following condition. There exists $\delta > 0$ such that

$$\sum_{e\in E: d(w_e(x), w_e(y)) \leqslant r d(x,y)} p_e(x) p_e(y) \geqslant \delta^2 > 0 \quad \text{for all } x, y \in K. \tag{3.4}$$

In fact, conditions (3.4) and (3.3) now also cover some finite Markov chains where some transition probabilities between the states can be zero, but still very few of those which are known to possess an attractive probability measure. Moreover, condition (3.4) would not work for Elton's proof of the corresponding ergodic theorem in [**7**]. Thus an incompleteness of their setup is obvious and there is a need for an extension of it. Contractive Markov systems provide it in a satisfactory way.

Remark 4. (i) A similar structure was discovered by Kaijser in the setup of random systems with complete connections (RSCC) in [**10**]. However, what he calls *weakly distance diminishing random systems with complete connections* cover only aperiodic contractive markov systems with compact state space.

(ii) Of course, the contractiveness of a Markov system can be weakened, just as is done sometimes for maps, by the demand that a contraction on average happens eventually not after one but after a number of iterations, that is, there exist $r \in \mathbb{N}$ and $0 < a < 1$ such that

$$\int d\big(w_{\sigma_r}\dots w_{\sigma_1}x, w_{\sigma_r}\dots w_{\sigma_1}y\big)\, dP_x(\sigma) \leqslant ad(x,y) \qquad \text{for all } x, y \in K_i \text{ and } i \in \{1,\dots,N\},$$

where P_x is a probability measure which represents the Markov process starting in x (Definition 6). However, such systems, again just as in the case of maps, are not expected to exhibit a substantially new behavior, but a decrease in transparency of the proofs for such systems can be expected.

Now we are able to prove the first theorem which shows that contractive Markov system, under reasonable topological assumptions, which allow the associated Markov operator to map continuous functions on continuous functions, has some nice properties.

Definition 10. We call the partition $K_1,\dots,K_N$ of K *open* if and only if every K_i is an open subset of K. Of course this means that K must be disconnected.

Theorem 1. *Suppose that $(K_{i(e)}, w_e, p_e)_{e\in E}$ is a contractive Markov system with an average contracting rate $0 < a < 1$ such that the family $K_1,\dots,K_N$ is an open partition of K and each p_e is continuous on $K_{i(e)}$. Then the following hold.*

(i) *The sequence $(U^{*k}\delta_x)_{k\in\mathbb{N}}$ is tight for all $x \in K$, that is, for all $\epsilon > 0$, there exists a compact subset $Q \subset K$ such that $U^{*k}\delta_x(Q) \geqslant 1 - \epsilon$ for all $k \in \mathbb{N}$.*

(ii) *The contractive Markov system has an invariant Borel probability measure μ.*

(iii) *The invariant probability measure μ is unique if and only if*

$$\frac{1}{n}\sum_{k=1}^{n} U^k g(x) \longrightarrow \int g\, d\mu \quad \text{for all } x \in K \text{ and } g \in C_B(K).$$

(iv) *If the invariant probability measure is unique, then*

$$\sum_{i=1}^{N} \int_{K_i} d(x,x_i)d\mu(x) < \infty \quad \text{for all } x_i \in K_i,\ i = 1,\dots,N.$$

Proof. (i) Fix $x_i \in K_i$ for each $i = 1, \dots, N$. Define

$$f(x) := \sum_{i=1}^{N} 1_{K_i}(x) d(x, x_i) \quad \text{for all } x \in K,$$

and let $C > 0$ be such that

$$\max_{e \in E} d\big(w_e x_{i(e)}, x_{t(e)}\big) \leqslant C.$$

We show inductively that

$$U^k f(x_i) \leqslant C \frac{1 - a^k}{1 - a}$$

for all $k \in \mathbb{N}$ and all $i = 1, \dots, N$. First, observe that, for any $i \in \{1, \dots, N\}$,

$$\begin{aligned} Uf(x_i) &= \sum_{e \in E} p_e(x_i) f \circ w_e(x_i) = \sum_{j=1}^{N} \sum_{e \in E} p_e(x_i) 1_{K_j}(w_e x_i) d(w_e x_i, x_j) \\ &= \sum_{j=1}^{N} \sum_{e \in E, t(e) = j} p_e(x_i) d\big(w_e x_i, x_{t(e)}\big) \leqslant \sum_{e \in E} p_e(x_i) C = C. \end{aligned}$$

Suppose that $U^{k-1} f(x_i) \leqslant C(1 - a^{k-1})/(1 - a)$ for some k. Denote by $(e_1, \dots, e_k)^*$ a path starting in i. Then

$$\begin{aligned} U^k f(x_i) &= \sum_{(e_1, \dots, e_k)^*} p_{e_1}(x_i) \dots p_{e_k}\big(w_{e_{k-1}} \dots w_{e_1} x_i\big) \\ &\quad \times \sum_{j=1}^{N} 1_{K_j}\big(w_{e_k} \dots w_{e_1} x_i\big) d\big(w_{e_k} \dots w_{e_1} x_i, x_j\big) \\ &\leqslant \sum_{(e_1, \dots, e_k)^*} p_{e_1}(x_i) \dots p_{e_k}\big(w_{e_{k-1}} \dots w_{e_1} x_i\big) d\big(w_{e_k} \dots w_{e_1} x_i, w_{e_k} x_{i(e_k)}\big) \\ &\quad + \sum_{j=1}^{N} \sum_{(e_1, \dots, e_k)^*, t(e_k) = j} p_{e_1}(x_i) \dots p_{e_k}\big(w_{e_{k-1}} \dots w_{e_1} x_i\big) d\big(w_{e_k} x_{i(e_k)}, x_j\big) \\ &\leqslant \sum_{(e_1, \dots, e_{k-1})^*} \sum_{j=1}^{N} 1_{K_j}\big(w_{e_{k-1}} \dots w_{e_1} x_i\big) \sum_{e_k, i(e_k) = j} p_{e_1}(x_i) \dots \\ &\quad \times p_{e_k}\big(w_{e_{k-1}} \dots w_{e_1} x_i\big) d\big(w_{e_k} \dots w_{e_1} x_i, w_{e_k} x_j\big) + C \\ &\leqslant a \sum_{(e_1, \dots, e_{k-1})^*} \sum_{j=1}^{N} 1_{K_j}\big(w_{e_{k-1}} \dots w_{e_1} x_i\big) p_{e_1}(x_i) \dots p_{e_{k-1}}\big(w_{e_{k-2}} \dots w_{e_1} x_i\big) \\ &\quad \times d\big(w_{e_{k-1}} \dots w_{e_1} x_i, x_j\big) + C \\ &= a U^{k-1} f(x_i) + C \leqslant aC \frac{1 - a^{k-1}}{1 - a} + C = C \frac{1 - a^k}{1 - a}. \end{aligned}$$

Let $\rho := C/(1 - a)$ and $\epsilon > 0$. Then, by the above,

$$\begin{aligned} \rho \geqslant U^k f(x_i) &= \int f \circ Z_k^{x_i} \, dP_{x_i} = \int \sum_{j=1}^{N} 1_{K_j}\big(Z_k^{x_i}\big) d\big(Z_k^{x_i}, x_j\big) \, dP_{x_i} \\ &\geqslant \frac{\rho}{\epsilon} P_{x_i}\left(d\big(Z_k^{x_i}, x_j\big) > \frac{\rho}{\epsilon} \text{ for all } j = 1, \dots, N \right) \end{aligned}$$

for all $k \in \mathbb{N}$ and for all $i = 1, \dots, N$. Thus

$$P_{x_i}\left(d\big(Z_k^{x_i}, x_j\big) > \frac{\rho}{\epsilon} \text{ for all } j = 1, \dots, N\right) \leqslant \epsilon$$

for all $k \in \mathbb{N}$ and for all $i = 1, \dots, N$. Set

$$Q_\epsilon := \bigcup_{j=1}^{N} \bar{B}_{\rho/\epsilon}(x_j),$$

where $\bar{B}_{\rho/\epsilon}(y)$ denotes the closed ball of radius ρ/ϵ and center y. Then Q_ϵ is compact, by the assumption on the metric space, and

$$\begin{aligned} U^{*k}\delta_{x_i}(Q_\epsilon) = U^k 1_{Q_\epsilon}(x_i) &= \int 1_{Q_\epsilon} \circ Z_k^{x_i}\, dP_{x_i} = P_{x_i}\big(Z_k^{x_i} \in Q_\epsilon\big) \\ &= 1 - P_{x_i}\big(Z_k^{x_i} \in K \backslash Q_\epsilon\big) = 1 - P_{x_i}\left(d\big(Z_k^{x_i}, x_j\big) > \frac{\rho}{\epsilon} \text{ for all } j = 1, \dots, N\right) \\ &\geqslant 1 - \epsilon \end{aligned}$$

for all $k \in \mathbb{N}$ and $i = 1, \dots, N$, as desired.

(ii) Define an operator

$$U_n := \frac{1}{n}\sum_{k=1}^{n} U^k \quad \text{for every } n \in \mathbb{N},$$

and let ${U_n}^*$ be its adjoint operator on $P(K)$. Fix $x \in K$. By (i), the sequence $({U_n}^*\delta_x)_{n\in\mathbb{N}}$ is tight also. Therefore it has a subsequence ${U_{n_m}}^*\delta_x$ which converges weakly* to a Borel probability measure, say μ. By the hypothesis of the theorem, the Markov operator U maps continuous functions to continuous functions. Therefore, its adjoint operator U^* is weakly* continuous. Hence

$$U^*\big({U_{n_m}}^*\delta_x\big) \xrightarrow{\text{w}^*} U^*\mu \quad \text{as } m \to \infty.$$

However, since

$$\left|\frac{1}{n_m}\sum_{k=2}^{n_m+1} U^k g(x) - U_{n_m} g(x)\right| \leqslant \frac{1}{n_m} 2\|g\| \quad \text{for all } g \in C_B(K),$$

we conclude that

$$U^*\mu = \mu,$$

that is, μ is an invariant Borel probability measure on the contractive Markov system.

(iii) Suppose that μ is the unique invariant probability measure. Then, by the above,

$$\frac{1}{n}\sum_{k=1}^{n} U^k g(x) \to \int g\, d\mu \quad \text{for all } x \in K \text{ and } g \in C_B(K). \tag{3.5}$$

Conversely, if (3.5) holds true, by Lebesgue's dominated convergence theorem, it implies that

$$\frac{1}{n}\sum_{k=1}^{n} U^{*k}\lambda \xrightarrow{\text{w}^*} \mu \quad \text{for all } \lambda \in P(K).$$

Again, by the weak*-continuity of U^*, this implies that μ is the unique invariant Borel probability measure.

(iv) Fix $x_i \in K_i$ for each $i = 1, \dots, N$. Let ν be the Borel probability measure on K given by

$$\nu(A) := \sum_{i=1}^{N} \delta_{x_i}(A) \quad \text{for all } A \in \mathcal{B}(K).$$

Define $f_R := \min\{f, R\}$ for $R > 0$, where f is the function from (i). Then every f_R is a bounded continuous function on K by the assumption of the theorem and, as in the proof of (i),

$$\int U^k f_R \, d\nu \leqslant \sum_{i=1}^{N} U^k f(x_i) \leqslant N\rho$$

for all $k \in \mathbb{N}$ and $R > 0$. Therefore,

$$\int \frac{1}{n} \sum_{k=1}^{n} U^k f_R \, d\nu \leqslant N\rho$$

for all $n \in \mathbb{N}$ and $R > 0$. By (iii) and Lebesgue's dominated convergence theorem, this implies that

$$\int f_R \, d\mu \leqslant N\rho \quad \text{for all } R > 0.$$

By Levi's theorem, we conclude that

$$\int f \, d\mu \leqslant N\rho,$$

as desired. □

The next lemma is a generalization of [**1**, Lemma 2.5].

LEMMA 3. *Suppose that $(K_{i(e)}, w_e, p_e)_{e \in E}$ is a contractive Markov system with an average contracting rate $0 < a < 1$ such that $p_e|_{K_{i(e)}}$ is Dini-continuous for all $e \in E$. Then, for every $f \in C_{\mathrm{C}}(K)$, the functions $(U^n f|_{K_i})_{n \in \mathbb{N} \cup \{0\}}$ are uniformly equicontinuous for all $i = 1, \dots, N$.*

Proof. Let ϕ_e be the modulus of uniform continuity of $p_e|_{K_{i(e)}}$ for all $e \in E$. Note that each ϕ_e is non-decreasing and that $\phi_e(t) \leqslant 1$ for all t. Set

$$\phi_0(t) := \begin{cases} t & 0 \leqslant t \leqslant 1 \\ 1 & t > 1 \end{cases}$$

and $\phi := \max_{e \in E \cup \{0\}}\{\phi_e\}$. It is clear that ϕ is also non-decreasing and that it satisfies Dini's condition.

Let $f \in \mathrm{Lip}(K)$ and $\|f\| \leqslant 1$. Then there is $C \geqslant 2$ such that

$$|f(x) - f(y)| \leqslant C d(x, y) \quad \text{for all } x, y \in K.$$

Set $L := \max\{L_1, \dots, L_N\}$ and

$$\beta(t) := \frac{L \vee C}{1 - a} \int_0^{ta^{-1}} \frac{\phi(u)}{u} \, du,$$

where $L \vee C := \max\{L, C\}$. Then $\beta(0) = 0$, and β is continuous and increasing. By [**2**, Sublemma], increasing ϕ if necessary, we can assume that β is concave. Further,

$$\beta(t) - \beta(at) = \frac{L \vee C}{1-a} \int_t^{ta^{-1}} \frac{\phi(u)}{u}\, du \geqslant \frac{L}{1-a} \frac{\phi(t)}{ta^{-1}} t(a^{-1} - 1) = L\phi(t).$$

Hence

$$\beta(at) + L\phi(t) \leqslant \beta(t) \quad \text{for all } t \geqslant 0.$$

Note that, for $0 \leqslant t \leqslant 1$,

$$\beta(t) \geqslant C \int_0^t \frac{\phi(u)}{u}\, du \geqslant C \int_0^t du = Ct,$$

and, for $t > 1$, $\beta(t) \geqslant \beta(1) \geqslant C \geqslant 2$. Therefore,

$$|f(x) - f(y)| \leqslant \beta(d(x, y)) \quad \text{for all } x, y \in K.$$

As an induction hypothesis for some $n \in \mathbb{N}$, assume that $\left|U^{n-1}f(x) - U^{n-1}f(y)\right| \leqslant \beta(d(x, y))$ for all $x, y \in K_i$, $i = 1, \ldots, N$. Let $x, y \in K_i$ for some $i \in \{1, \ldots, N\}$. Then, since β is increasing and concave,

$$\begin{aligned}
|U(U^{n-1}f)(x) - U(U^{n-1}f)(y)| &\leqslant \sum_{j=1}^{L_i} p_{ij}(x)|U^{n-1}f(w_{ij}(x)) - U^{n-1}f(w_{ij}(y))| \\
&\quad + \sum_{j=1}^{L_i} |p_{ij}(x) - p_{ij}(y)||U^{n-1}f(w_{ij}(y))| \\
&\leqslant \sum_{j=1}^{L_i} p_{ij}(x)\beta(d(w_{ij}(x), w_{ij}(y)) + L_i\phi(d(x, y))) \\
&\leqslant \beta(ad(x, y)) + L\phi(d(x, y)) \\
&\leqslant \beta(d(x, y)).
\end{aligned}$$

Hence the $(U^n f|_{K_i})_{n \in \mathbb{N} \cup \{0\}}$ are uniformly equicontinuous for each $i = 1, \ldots, N$.

Since $\mathrm{Lip}(K) \cap C_{\mathrm{C}}(K)$ is a dense subset of $(C_{\mathrm{C}}(K), \|.\|)$, the claim follows by an $(\epsilon/3)$-argument. □

We need to know more about properties of irreducible directed graphs. The following lemma is a generalization of the lattice theorem (see [**3**, Theorem 4.3]).

Lemma 4 (lattice theorem). *Let an irreducible directed graph with period d be given. Then, for every finite path $(e_1, \ldots, e_n)$ of the digraph, there exists $m_0 \geqslant 0$ such that for all integers $m \geqslant m_0$ there exists a closed path of length md which has $(e_1, \ldots, e_n)$ as a part of it and starts with e_1.*

Proof. Let A be the set of all $k \in \mathbb{N}$ such that there exists a closed path of length k which has $(e_1, \ldots, e_n)$ as a part of it and starts with e_1. Then A is closed under addition. Since the digraph has period d, the greatest common divisor of numbers from A is d. Therefore, the set A contains all but a finite number of positive multiples of d (see [**3**, Appendix, Theorem 1.1]). In other words, there exists $m_0 \in \mathbb{N}$ such that for all $m \geqslant m_0$, there exists a path of length md which has $(e_1, \ldots, e_n)$ as a part of it and starts with e_1. □

LEMMA 5. *Let an irreducible directed graph with the set of vertices V and period d be given. Fix $i \in V$ and let $\mathcal{V}_i$ be the set of all pairs of vertices $(\alpha, \beta) \in V \times V$ which are accessible from i by paths of the same length. Then there exists $r \in \mathbb{N}$ such that, for each $(\alpha, \beta) \in \mathcal{V}_i$, i is accessible from α and β by paths of the same length which are less than or equal to dr.*

Proof. Let $(\alpha, \beta) \in \mathcal{V}_i$. Then there exist paths s_α and s_β from i to α and from i to β, respectively, of the same length, say $n_{\alpha\beta}$. By the lattice theorem, there exists $m_\alpha \in \mathbb{N}$ such that, for all integers $m \geqslant m_\alpha$, there exists a closed path of length md which starts in i and has s_α as a part of it. Analogously, there exists $m_\beta \in \mathbb{N}$ such that, for all integers $m \geqslant m_\beta$, there exists a closed path of length md which starts in i and has s_β as a part of it. Set $r_{\alpha\beta} := \max\{m_\alpha, m_\beta\}$ and $r := \max_{(\alpha,\beta)\in\mathcal{V}_i} r_{\alpha\beta}$. Then there exist two closed paths of length dr which start in i, and one of them has s_α as a part of it and the other has s_β as a part of it. Hence there exist two paths of the same length $dr - n_{\alpha\beta} \leqslant dr$, where one of them is from α to i and the other is from β to i. □

The next lemma is a generalization of [**1**, Lemma 2.7]. It uses a well known technique of coupling, the main idea of which is to put as much mass as possible close to the diagonal of two processes (see [**10**, **11**] for more on this).

LEMMA 6. *Suppose that $(K_{i(e)}, w_e, p_e)_{e\in E}$ is an irreducible contractive Markov system with an average contracting rate $0 < a < 1$ such that $p_e|_{K_{i(e)}}$ is Dini-continuous and there exists $\delta > 0$ such that $p_e|_{K_{i(e)}} \geqslant \delta$ for all $e \in E$. Then the following hold.*

(i) *For every $f \in C_C(K)$,*

$$\lim_{n\to\infty} |U^n f(x) - U^n f(y)| = 0 \quad \text{for all } x, y \in K_i \text{ and } i \in \{1, \dots, N\}$$

and the convergence is uniform on bounded subsets.

(ii) *If in addition the contractive Markov system is aperiodic, then, for every $f \in C_{\mathrm{C}}(K)$,*

$$\lim_{n\to\infty} |U^n f(x) - U^n f(y)| = 0 \quad \text{for all } x, y \in K,$$

and the convergence is again uniform on bounded subsets.

Proof. Let $S \subset K$ be bounded. We can assume that $S \cap K_i \neq \varnothing$ for all $i = 1, \dots, N$. Since each probability function p_e is bounded away from zero on $K_{i(e)}$, the average contractiveness condition implies that each map $w_e|_{K_{i(e)}}$ is Lipschitz. Hence there exists $C > 0$ such that

$$\max_{e\in E} d\big(w_e x_{i(e)}, x_{t(e)}\big) \leqslant C$$

for all $x_i \in S \cap K_i$, $i = 1, \dots, N$. Let $x_i, y_i \in S \cap K_i$ for each $i = 1, \dots, N$. Fix $i, j \in \{1, \dots, N\}$. Set

$$\Sigma^* := \Sigma^+ \times \Sigma^+ = \{\bar{e} := (e_1, \tilde{e}_1, e_2, \tilde{e}_2, \dots) |\ (e_1, e_2, \dots) \in \Sigma^+,\ (\tilde{e}_1, \tilde{e}_2, \dots) \in \Sigma^+\},$$

and let $P^* := P_{x_i} \otimes P_{y_j}$ be the product measure on Σ^*. Thus, if we define

$$Z_n^{x_i}(\bar{e}) := w_{e_n} \circ \dots \circ w_{e_1}(x_i) \quad \text{and} \quad \tilde{Z}_n^{y_j}(\bar{e}) := w_{\tilde{e}_n} \circ \dots \circ w_{\tilde{e}_1}(y_j) \quad \text{on } \Sigma^*,$$

then $Z_n^{x_i}$ and $\tilde{Z}_n^{y_j}$ are independent Markov processes with initial distributions δ_{x_i} and δ_{y_j} respectively and $U^n f(x_i) = E(f \circ Z_n^{x_i})$ for all $f \in C_{\mathrm{B}}(K)$, where the expectation means with respect to the measure P^*. Let $\alpha > 0$ and for each $m \in \mathbb{N}$, let $G_{\alpha,m}$ be the set of all $\bar{e} \in \Sigma^*$ such that there exists $\tilde{i}$ such that

$$Z_m^{x_i}(\bar{e}), \tilde{Z}_m^{y_j}(\bar{e}) \in K_{\tilde{i}}, \quad d\big(Z_m^{x_i}(\bar{e}), \tilde{Z}_m^{y_j}(\bar{e})\big) \leqslant \alpha, \quad d\big(Z_l^{x_i}(\bar{e}), \tilde{Z}_l^{y_j}(\bar{e})\big) > \alpha \quad \text{for all } l < m.$$

Then the $(G_{\alpha,m})_{m\in\mathbb{N}}$ are disjoint. Further, for each $n \in \mathbb{N}$, set

$$B_{\alpha,n} := \Sigma^* \Big\backslash \bigcup_{m=1}^{n} G_{\alpha,m}.$$

Denote by $\mathcal{B}_m$ the σ-algebra in Σ^* generated by $Z_1^{x_i}, \ldots, Z_m^{x_i}, \tilde{Z}_1^{y_j}, \ldots, \tilde{Z}_m^{y_j}$. Then $G_{\alpha,m} \in \mathcal{B}_m$. Now, for $f \in C_{\mathrm{C}}(K)$,

$$\begin{aligned} U^n f(x_i) - U^n f(y_j) &= Ef\big(Z_n^{x_i}\big) - Ef\big(\tilde{Z}_n^{y_j}\big) \\ &= \sum_{m=1}^{n} E\big[1_{G_{\alpha,m}}\big(f\big(Z_n^{x_i}\big) - f\big(\tilde{Z}_n^{y_j}\big)\big)\big] \\ &\quad + E\big[1_{B_{\alpha,n}}\big(f\big(Z_n^{x_i}\big) - f\big(\tilde{Z}_n^{y_j}\big)\big)\big] \\ &= \sum_{m=1}^{n} E\big[1_{G_{\alpha,m}}\big(E\big(f\big(Z_n^{x_i}\big)\big|\mathcal{B}_m\big) - E\big(f\big(\tilde{Z}_n^{y_j}\big)\big|\mathcal{B}_m\big)\big)\big] \\ &\quad + E\big[1_{B_{\alpha,n}}\big(f\big(Z_n^{x_i}\big) - f\big(\tilde{Z}_n^{y_j}\big)\big)\big]. \end{aligned}$$

Further, note that for $n > m$,

$$\begin{aligned} E\big(f\big(Z_n^{x_i}\big)\big|\mathcal{B}_m\big) &= \sum_{(e_{m+1},\ldots,e_n)} p_{e_{m+1}}\big(Z_m^{x_i}\big)\ldots p_{e_n}\big(w_{e_{n-1}} \circ \ldots \circ w_{e_{m+1}} Z_m^{x_i}\big) \\ &\quad \times f\big(w_{e_n} \circ \ldots \circ w_{e_{m+1}} Z_m^{x_i}\big) \\ &= U^{n-m} f\big(Z_m^{x_i}\big). \end{aligned}$$

Therefore,

$$\begin{aligned} U^n f(x_i) - U^n f(y_j) = &\sum_{m=1}^{n} E\big[1_{G_{\alpha,m}}\big(U^{n-m} f\big(Z_m^{x_i}\big) - U^{n-m} f\big(\tilde{Z}_m^{y_j}\big)\big)\big] \\ &+ E\big[1_{B_{\alpha,n}}\big(f\big(Z_n^{x_i}\big) - f\big(\tilde{Z}_n^{y_j}\big)\big)\big]. \end{aligned}$$

Let $\epsilon > 0$ and choose, by Lemma 3, $\alpha > 0$ such that, for all $u, v \in K_l$, $l = 1, \ldots, N$,

$$d(u,v) \leqslant \alpha \;\Rightarrow\; |U^n f(u) - U^n f(v)| < \epsilon \quad \text{for all } n \in \mathbb{N}.$$

Then

$$|U^n f(x_i) - U^n f(y_j)| \leqslant \sum_{m=1}^{n} E\big[1_{G_{\alpha,m}}\epsilon\big] + E\big[1_{B_{\alpha,n}} 2\|f\|\big] \leqslant \epsilon + 2\|f\| P^*(B_{\alpha,n}).$$

Thus the proof of (i) and (ii) will be complete when we prove the following. □

SUBLEMMA 1. *Suppose that the contractive Markov system is irreducible and one of the following hold.*

(i) $i = j$.

(ii) *The contractive Markov system is aperiodic.*

Then $P^*(B_{\alpha,n}) \to 0$ *as* $n \to \infty$ *for all* $\alpha > 0$, *and the convergence is uniform on* S.

Proof. First, observe that

$$\begin{aligned} E\big(d\big(Z_{n+1}^{x_i}, x_{t(e_{n+1})}\big)\big|Z_n^{x_i}\big) &= \sum_{e\in E} p_e\big(Z_n^{x_i}\big) d\big(w_e Z_n^{x_i}, x_{t(e)}\big) \\ &\leqslant \sum_{e\in E} p_e\big(Z_n^{x_i}\big)\big[d\big(w_e Z_n^{x_i}, w_e x_{i(e)}\big) + d\big(w_e x_{i(e)}, x_{t(e)}\big)\big] \\ &\leqslant a d\big(Z_n^{x_i}, x_{t(e_n)}\big) + C \end{aligned}$$

for all $n \in \mathbb{N}$. Therefore, for any natural numbers $n_2 > n_1$,

$$\begin{aligned} E\big(d\big(Z_{n_2}^{x_i}, x_{t(e_{n_2})}\big)\big|Z_{n_1}^{x_i}\big) &= E\big[E\big(d\big(Z_{n_2}^{x_i}, x_{t(e_{n_2})}\big)\big|Z_{n_2-1}^{x_i}\big)\big|Z_{n_1}^{x_i}\big] \\ &\leqslant aE\big[d\big(Z_{n_2-1}^{x_i}, x_{t(e_{n_2-1})}\big)\big|Z_{n_1}^{x_i}\big] + C. \end{aligned}$$

Repeating this, we are led to

$$E\big(d\big(Z_{n_2}^{x_i}, x_{t(e_{n_2})}\big)\big|Z_{n_1}^{x_i}\big) \leqslant \frac{C}{1-a} + a^{n_2-n_1} d\big(Z_{n_1}^{x_i}, x_{t(e_{n_1})}\big).$$

Now, let $s \geqslant 2$ be the largest Lipschitz constant of the maps $w_e|_{K_{i(e)}}$, $e \in E$. Then, for all $n \in \mathbb{N}$,

$$\begin{aligned} d\big(Z_n^{x_i}, x_{t(e_n)}\big) &\leqslant d\big(w_{e_n} Z_{n-1}^{x_i}, w_{e_n} x_{t(e_{n-1})}\big) + d\big(w_{e_n} x_{t(e_{n-1})}, x_{t(e_n)}\big) \\ &\leqslant s d\big(Z_{n-1}^{x_i}, x_{t(e_{n-1})}\big) + C \quad P^*\text{-a.e.} \end{aligned}$$

Repeating this, we get

$$d\big(Z_n^{x_i}, x_{t(e_n)}\big) \leqslant \frac{s^n - 1}{s-1} C \leqslant s^n C \quad P^*\text{-a.e.}$$

Hence

$$E\big(d\big(Z_{n_2}^{x_i}, x_{t(e_{n_2})}\big)\big|Z_{n_1}^{x_i}\big) \leqslant \frac{C}{1-a} + a^{n_2-n_1} s^{n_1} C \quad P^*\text{-a.e.}$$

Set

$$\gamma := \frac{\log \frac{s}{a}}{\log \frac{1}{a}},$$

and let $n_2 \geqslant \gamma n_1$. Then

$$E\big(d\big(Z_{n_2}^{x_i}, x_{t(e_{n_2})}\big)\big|Z_{n_1}^{x_i}\big) \leqslant \frac{2C}{1-a} =: \frac{\lambda}{2} \quad P^*\text{-a.e.}$$

Therefore, by Markov inequality,

$$P^*\big(d\big(Z_{n_2}^{x_i}, x_{t(e_{n_2})}\big) > \lambda\big|Z_{n_1}^{x_i}\big) \leqslant \frac{1}{2} \quad P^*\text{-a.e.}$$

Analogously,

$$P^*\big(d\big(\tilde{Z}_{n_2}^{y_j}, y_{t(e_{n_2})}\big) > \lambda\big|Z_{n_1}^{y_j}\big) \leqslant \tfrac{1}{2} \quad P^*\text{-a.e.}$$

Since $(Z_n^{x_i})_{n\in\mathbb{N}}$ and $(\tilde{Z}_n^{y_j})_{n\in\mathbb{N}}$ are independent processes,

$$P^*\big(d\big(Z_{n_2}^{x_i}, x_{t(e_{n_2})}\big) \leqslant \lambda \text{ and } d\big(\tilde{Z}_{n_2}^{y_j}, y_{t(e_{n_2})}\big) \leqslant \lambda\big|Z_{n_1}^{x_i}, Z_{n_1}^{y_j}\big) \geqslant \tfrac{1}{4} \quad P^*\text{-a.e.}$$

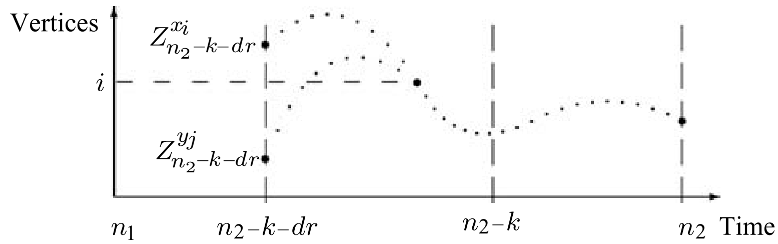

FIGURE 3.

Note that the average contractiveness condition

$$\sum_{e\in E} p_e(u)d(w_e u, w_e v) \leqslant a d(u,v) \quad \text{for all } u,v\in K_i,\ i=1,\dots,N,$$

implies that, for every $u,v\in K_i$, $i=1,\dots,N$, there exists $e_0\in E$ such that $d(w_{e_0}u, w_{e_0}v)\leqslant a d(u,v)$.

Now, in case (i), by Lemma 5, there exists $r\geqslant 0$ such that for any pair of vertices accessible from i by paths of the same length, there exist paths from them to i of an equal length that is less than or equal to dr, where d is the period of the contractive Markov system. In case (ii), that is, $d=1$, there also exists $r\in\mathbb{N}$ such that there are paths of length equal to r between any two vertices. In both these cases, it is implied that

$$P^*\big(\exists\,\tilde{i} \text{ s.t. } Z^{x_i}_n, \tilde{Z}^{y_j}_n \in K_{\tilde{i}} \,\big|\, Z^{x_i}_{n-dr}, \tilde{Z}^{y_j}_{n-dr}\big) > \delta^{2dr} \quad P^*\text{-a.e.}$$

for all $n\geqslant dr$. Therefore, by the Markov property,

$$P^*\big(\exists\,\tilde{i} \text{ s.t. } Z^{x_i}_{n_2}, \tilde{Z}^{y_j}_{n_2} \in K_{\tilde{i}} \,\big|\, Z^{x_i}_{n_1}, \tilde{Z}^{y_j}_{n_1}\big) > \delta^{2dr} \quad P^*\text{-a.e.}$$

for all $n_2\geqslant dr+n_1$. Since each $w_e|_{K_{i(e)}}$ is Lipschitz, there exists $\rho_{dr}>0$ such that

$$\max_{(e_1,\tilde{e}_1,\dots,e_{dr},\tilde{e}_{dr})^*} d\big(w_{e_{dr}}\circ\dots\circ w_{e_1}x_{i(e_1)}, w_{\tilde{e}_{dr}}\circ\dots\circ w_{\tilde{e}_1}x_{i(\tilde{e}_1)}\big) \leqslant \rho_{dr}$$

for all $x_i, y_i\in K_i\cap S$, $i=1,\dots,N$, where the maximum is taken over all paths $(e_1,\dots,e_{dr})^*$ and $(\tilde{e}_1,\dots,\tilde{e}_{dr})^*$ of the directed graph.

Now, choose k large enough so that $a^k(2s^{dr}\lambda+\rho_{dr})<\alpha$. Let $n_2\geqslant \gamma n_1+dr+k$. Then

$$\begin{aligned}P^*\big(&e_{n_2-l}=\tilde{e}_{n_2-l} \text{ and } d\big(Z^{x_i}_{n_2-l}, \tilde{Z}^{y_j}_{n_2-l}\big)\leqslant a d\big(Z^{x_i}_{n_2-l-1}, \tilde{Z}^{y_j}_{n_2-l-1}\big)\\ &\text{for all } l=0,\dots,k-1\big|Z^{x_i}_{n_1}, \tilde{Z}^{y_j}_{n_1}\big)\geqslant \delta^{2(k+dr)} \quad P^*\text{-a.e.}\end{aligned}$$

(see Figure 3).

Then, by the above and the Markov property,

$$\begin{aligned}P^*\big(&e_{n_2-l}=\tilde{e}_{n_2-l}, d\big(Z^{x_i}_{n_2-l}, \tilde{Z}^{y_j}_{n_2-l}\big)\leqslant a d\big(Z^{x_i}_{n_2-l-1}, \tilde{Z}^{y_j}_{n_2-l-1}\big) \text{ for all}\\ &l=0,\dots,k-1, d\big(Z^{x_i}_{n_2-k-dr}, x_{t(e_{n_2-k-dr})}\big)\leqslant\lambda \text{ and}\\ &d\big(\tilde{Z}^{y_j}_{n_2-k-dr}, y_{t(e_{n_2-k-dr})}\big)\leqslant\lambda\big|Z^{x_i}_{n_1}, \tilde{Z}^{y_j}_{n_1}\big)\geqslant \tfrac{1}{4}\delta^{2(k+dr)} \quad P^*\text{-a.e.}\end{aligned}$$

Observe that

$$\begin{aligned} d\big(Z^{x_i}_{n_2-k}, \tilde Z^{y_j}_{n_2-k}\big) &\leqslant d\big(Z^{x_i}_{n_2-k}, w_{e_{n_2-k}} \circ \ldots \circ w_{e_{n_2-k-dr+1}} x_{t(e_{n_2-k-dr})}\big) + \rho_{dr} \\ &\quad + d\big(w_{\tilde e_{n_2-k}} \circ \ldots \circ w_{\tilde e_{n_2-k-dr+1}} y_{t(\tilde e_{n_2-k-dr})}, \tilde Z^{y_i}_{n_2-k}\big) \\ &\leqslant s^{dr} d\big(Z^{x_i}_{n_2-k-dr}, x_{t(e_{n_2-k-dr})}\big) \\ &\quad + \rho_{dr} + s^{dr} d\big(y_{t(\tilde e_{n_2-k-dr})}, \tilde Z^{y_i}_{n_2-k-dr}\big) \end{aligned}$$

P^*-a.e. Hence

$$P^*\big(d\big(Z^{x_i}_{n_2}, \tilde Z^{y_j}_{n_2}\big) \leqslant a^k(2s^{dr}\lambda + \rho_{dr})\big|\, Z^{x_i}_{n_1}, \tilde Z^{y_j}_{n_1}\big) \geqslant \tfrac14 \delta^{2(k+dr)} \quad P^*\text{-a.e.}$$

Thus

$$P^*\big(d\big(Z^{x_i}_{n_2}, \tilde Z^{y_j}_{n_2}\big) > \alpha \big| Z^{x_i}_{n_1}, \tilde Z^{y_j}_{n_1}\big) \leqslant 1 - \tfrac14 \delta^{2(k+dr)} \quad P^*\text{-a.e.}$$

Now choose a sequence of natural numbers $n_1, n_2, \ldots$ such that $n_{t+1} \geqslant \gamma n_t + dr + k$ for all $t \in \mathbb{N}$. Then, by the above and the Markov property,

$$P^*\big(d\big(Z^{x_i}_{n_t}, \tilde Z^{y_j}_{n_t}\big) > \alpha,\ t = 1, \ldots, m\big) \leqslant \big(1 - \tfrac14 \delta^{2(k+dr)}\big)^{m-1} \quad \text{for all } m \in \mathbb{N}.$$

Hence

$$P^*(B_{\alpha,n}) \leqslant \big(1 - \tfrac14 \delta^{2(k+dr)}\big)^{m-1} \quad \text{if } n \geqslant n_m.$$

Thus $P^*(B_{\alpha,n}) \to 0$ as $n \to \infty$ and convergence is uniform on S, since γ, r, k do not depend on the choice of $x_i, y_i \in S$, $i = 1, \ldots, N$. □

DEFINITION 11. A measure $\mu \in P(X)$ is called the *attractive measure of the contractive Markov system* if and only if

$$U^{*n}\nu \xrightarrow{\text{w}^*} \mu \quad \text{for all } \nu \in P(X).$$

Note that the attractive measure is the only invariant probability measure of the contractive Markov system if U^* is weakly* continuous, which is true if U maps continuous functions on continuous functions.

THEOREM 2. *Suppose that $(K_{i(e)}, w_e, p_e)_{e\in E}$ is an irreducible contractive Markov system such that $K_1, \ldots, K_N$ is an open partition of K, $p_e|_{K_{i(e)}}$ is Dini-continuous and there exists $\delta > 0$ such that $p_e|_{K_{i(e)}} \geqslant \delta$ for all $e \in E$. Then the following hold.*

(i) *The contractive Markov system has a unique invariant Borel probability measure μ.*

(ii) *If, in addition, the contractive Markov system is aperiodic, then*

$$U^n f(x) \to \mu(f) \quad \textit{for all } x \in K \textit{ and } f \in C_{\mathrm{B}}(K),$$

and the convergence is uniform on bounded subsets, that is, μ is an attractive probability measure of the contractive Markov system.

Proof. (i) Fix $x_i \in K_i$ for all $i = 1, \ldots, N$. Since the sequence $(U^{*l}\delta_{x_i})_{l\in\mathbb{N}}$ is tight, $(1/n \sum_{l=1}^n U^{*l}\delta_{x_i})_{n\in\mathbb{N}}$ is also tight for all $i = 1, \ldots, N$. Hence there exists an increasing sequence of natural numbers $(n_k)_{k\in\mathbb{N}}$ such that, for each $i = 1, \ldots, N$, $(1/n_k \sum_{l=1}^{n_k} U^{*l}\delta_{x_i})_{k\in\mathbb{N}}$ converges weakly* to a Borel probability measure, say μ_i, that is,

$$\lim_{k\to\infty}\frac{1}{n_k}\sum_{l=1}^{n_k}U^l f(x_i)=\mu_i(f)\quad\text{for all } f\in C_{\mathrm{B}}(K)\text{ and } i\in\{1,\dots,N\}.$$

Since, by Lemma 6(i), for every $f\in C_{\mathrm{C}}(K)$,

$$\lim_{n\to\infty}|U^n f(x_i)-U^n f(y_i)|=0\quad\text{for all } y_i\in K_i\text{ and } i\in\{1,\dots,N\},$$

we conclude that for every $f\in C_{\mathrm{C}}(K)$,

$$\lim_{k\to\infty}\frac{1}{n_k}\sum_{l=1}^{n_k}U^l f(x)=\sum_{i=1}^{N}\mu_i(f)1_{K_i}(x)\quad\text{for all } x\in K.$$

Since, for every x, we deal here with convergence of Radon probability measures on a locally compact metric space, it is implied that

$$\lim_{k\to\infty}\frac{1}{n_k}\sum_{l=1}^{n_k}U^l f(x)=\sum_{i=1}^{N}\mu_i(f)1_{K_i}(x)\quad\text{for all } x\in K\text{ and all } f\in C_{\mathrm{B}}(K).\tag{3.6}$$

Define a linear operator $Q:C_{\mathrm{B}}(K)\longrightarrow C_{\mathrm{B}}(K)$ by

$$Q(f):=\sum_{i=1}^{N}\mu_i(f)1_{K_i}\quad\text{for all } f\in C_{\mathrm{B}}(K).\tag{3.7}$$

Then, by (3.6),

$$QU=Q,$$

and therefore

$$Q^2=Q.\tag{3.8}$$

Now, by the definition of μ_i, $U^*\mu_i=\mu_i$ for all $i=1,\dots,N$. Since the contractive Markov system is irreducible, this implies, by Lemma 1, that $\mu_i(K_j)>0$ for all $i,j=1,\dots,N$. Now let $f\in C_{\mathrm{B}}(K)$ with $f\geqslant 0$. Then, by (3.8),

$$\sum_{i=1}^{N}\mu_i(f)1_{K_i}=\sum_{i=1}^{N}\mu_i\left(\sum_{j=1}^{N}\mu_j(f)1_{K_j}\right)1_{K_i}=\sum_{i,j=1}^{N}\mu_j(f)\mu_i(K_j)1_{K_i},$$

that is,

$$\mu_i(f)=\sum_{j=1}^{N}\mu_j(f)\mu_i(K_j)\quad\text{for all } i=1,\dots,N.$$

Suppose that there exists i_0 such that $\mu_{i_0}(f)<\max_{1\leqslant j\leqslant N}\mu_j(f)$. Then, by the above,

$$\mu_i(f)<\max_{1\leqslant j\leqslant N}\mu_j(f)\quad\text{for all } i=1,\dots,N,$$

which obviously cannot be true. Hence

$$\mu_i(f)=\mu_j(f)\quad\text{for all } i,j=1,\dots,N.$$

Let $\mu:=\mu_1$. Since $f\in C_{\mathrm{B}}(K)$ with $f\geqslant 0$ was arbitrary, we conclude that all μ_i, $i=1,\dots,N$, are equal to μ. Hence

$$\lim_{k\to\infty}\frac{1}{n_k}\sum_{l=1}^{n_k}U^l f(x)=\mu(f)\quad\text{for all } x\in K\text{ and } f\in C_{\mathrm{B}}(K).\tag{3.9}$$

Suppose that there exists $\lambda \in P(K)$ such that $U^*\lambda = \lambda$. Then

$$\frac{1}{n_k}\sum_{l=1}^{n_k} {U^*}^l\lambda = \lambda \quad \text{for all } k \in \mathbb{N}$$

also, but if we apply Lebesgue's dominated convergence theorem to (3.9), it is implied that

$$\frac{1}{n_k}\sum_{l=1}^{n_k} {U^*}^l\lambda \stackrel{\mathrm{w}^*}{\to} \mu.$$

Thus $\lambda = \mu$, that is, μ is a unique invariant Borel probability measure of the contractive Markov system.

(ii) Let $x \in K$. By Theorem 1(i), the sequence $({U^*}^n\delta_x)_{n\in\mathbb{N}}$ is tight. Therefore, there is a subsequence $({U^*}^{n_k}\delta_x)_{k\in\mathbb{N}}$ which converges weakly* to a Borel probability measure, say μ, that is, $U^{n_k}f(x) \to \mu(f)$ $(k \to \infty)$ for all $f \in C_{\mathrm{B}}(K)$. Since, by Lemma 6(ii), $|U^{n_k}f(x) - U^{n_k}f(y)| \to 0$ for all $y \in K$ and for all $f \in C_{\mathrm{C}}(K)$, it follows that $U^{n_k}f(y) \to \mu(f)$ for all $y \in K$ and $f \in C_{\mathrm{C}}(K)$.

Let $\epsilon > 0$. By the tightness of $({U^*}^n\delta_x)_{n\in\mathbb{N}}$, there exists a compact $Q \subset K$ such that ${U^*}^n\delta_x(K \setminus Q) < \epsilon$ for all $n \in \mathbb{N}$. Hence

$$\begin{aligned} |U^n g(x)| &= \left|\int g \circ Z_n^x \, dP_x\right| \leqslant \int\limits_{\{Z_n^x \in K\setminus Q\}} \left|g \circ Z_n^x\right| dP_x + \int\limits_{\{Z_n^x \in Q\}} \left|g \circ Z_n^x\right| dP_x \\ &\leqslant \|g\| \int 1_{K\setminus Q} \circ Z_n^x \, dP_x + \|g\|_Q = \|g\| {U^*}^n\delta_x(K \setminus Q) + \|g\|_Q \\ &\leqslant \|g\|\epsilon + \|g\|_Q \end{aligned}$$

for all $g \in C_{\mathrm{B}}(K)$ and all $n \in \mathbb{N}$. Let $f \in C_{\mathrm{C}}(K)$. Since, by Lemma 3, the functions $(U^{n_k}f|_{K_i})_{k\in\mathbb{N}}$ are equicontinuous for each $i = 1, \ldots, N$, by the Arzelà–Ascoli theorem, there exists a subsequence, without loss of generality $(U^{n_k}f)_{k\in\mathbb{N}}$, which converges uniformly on Q. Hence there exists $n_\epsilon > 0$ such that

$$\|U^{n_k}f - \mu(f)\|_Q < \epsilon \quad \text{for all } k \geqslant n_\epsilon.$$

Thus, by the above,

$$\begin{aligned} |U^n f(x) - \mu(f)| &= |U^{n-n_k}(U^{n_k}f - \mu(f))(x)| \\ &\leqslant \epsilon(\|f\| + \mu(f)) + \|U^{n_k}f - \mu(f)\|_Q \\ &\leqslant \epsilon(\|f\| + \mu(f) + 1) \end{aligned}$$

for all $n \geqslant n_{n_\epsilon}$. Hence

$$U^n f(x) \to \int f \, d\mu \quad \text{for all } x \in K \text{ and } f \in C_{\mathrm{C}}(K).$$

This also implies that

$$U^n f(x) \to \int f \, d\mu \quad \text{for all } x \in K \text{ and } f \in C_{\mathrm{B}}(K);$$

the convergence is uniform on bounded subsets by Lemma 6(ii). By Lebesgue's dominated convergence theorem, we conclude that

$${U^*}^n\nu \stackrel{\mathrm{w}^*}{\to} \mu \quad \text{for all } \nu \in P(K). \qquad \square$$

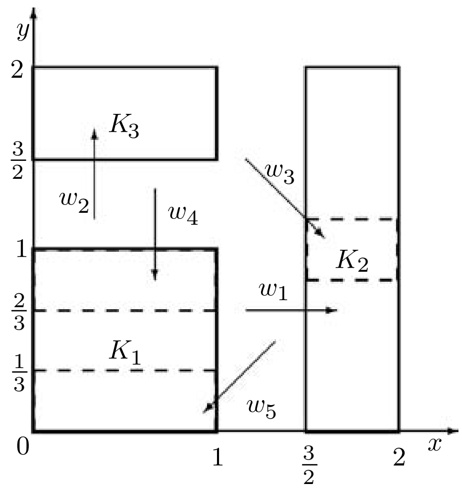

FIGURE 4.

EXAMPLE 2. Every irreducible finite Markov chain is a contractive Markov system satisfying the hypothesis of Theorem 2.

EXAMPLE 3. Consider, for simplicity, $\mathbb{R}^2$ to be normed by $\|.\|_1$. Let $K_1 := [0,1]\times[0,1]$, $K_2 := [0,1]\times[3/2,2]$ and $K_3 := [3/2,2]\times[0,2]$. Consider the following maps on $\mathbb{R}^2$:

$$w_1\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}\frac{1}{2}x+\frac{3}{2}\\2y\end{pmatrix},\quad w_2\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}x\\\frac{1}{2}y+\frac{3}{2}\end{pmatrix},\quad w_3\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}y\\-\frac{1}{3}x+\frac{7}{6}\end{pmatrix},$$

$$w_4\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}x\\\frac{2}{3}y-\frac{1}{3}\end{pmatrix},\quad w_5\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}\frac{1}{2}y\\-\frac{2}{3}x+\frac{4}{3}\end{pmatrix},$$

with probability functions

$$p_1 := \tfrac{1}{4}1_{K_1},\quad p_2 := \tfrac{3}{4}1_{K_1},\quad p_3 := \tfrac{2}{3}1_{K_3},\quad p_4 := \tfrac{1}{3}1_{K_3},\quad p_5 := 1_{K_2}.$$

An easy calculation shows that they define a contractive Markov system with an average contracting rate 8/9 on $K_1\cup K_2\cup K_3$, as is shown on Figure 4, which satisfies the hypothesis of Theorem 2(ii) and does *not* satisfy the hypothesis of [**1**, Theorem 2.1].

w_1 contracts K_1 in the x-direction, expands it in the y-direction, and maps it on K_2. w_2 contracts K_1 in the y-direction and maps it on K_3. w_3 contracts K_3 in the x-direction, rotates it 90° clockwise, and maps it on the middle dashed rectangle in K_2. w_4 contracts K_3 in the y-direction and maps it on the upper dashed rectangle in K_1. w_5 rotates K_2 90° clockwise, contracts it and maps it on the bottom dashed rectangle in K_1. Note that w_5 is the only contractive map here.

EXAMPLE 4. Let $G := (V,E,i,t)$ be a finite irreducible directed (multi)graph. Let Σ_G be the set of all one-sided infinite paths $\sigma := (\ldots,\sigma_{-1},\sigma_0)$ of G (*one-sided subshift of finite type* associated with G) provided with the metric $d(\sigma,\sigma') := 2^k$, where k is the smallest integer with $\sigma_i = \sigma'_i$ for all $k < i \leqslant 0$. Let g be a positive, Dini-continuous function on Σ_G such that

$$\sum_{y\in T^{-1}(\{x\})} g(y) = 1 \quad \text{for all } x \in \Sigma_G,$$

where T is the right shift map on Σ_G. Define, for every $i \in V$,

$$K_i := \{\sigma \in \Sigma_G : t(\sigma_0) = i\},$$

and, for every $e \in E$,

$$w_e(\sigma) := (\ldots, \sigma_{-1}, \sigma_0, e), \quad p_e(\sigma) := g(\ldots, \sigma_{-1}, \sigma_0, e) \quad \text{for all } \sigma \in K_{i(e)}.$$

Obviously, maps $(w_e)_{e\in E}$ are contractions. Therefore, $(K_{i(e)}, w_e, p_e)_{e\in E}$ defines a contractive Markov system which satisfies the hypothesis of Theorem 2 and does *not* satisfy the hypothesis of [**1**, Theorem 2.1]. Hence Theorem 2(ii) covers [**18**, Theorem 3.1] (there, it was assumed that $\sum_{n=0}^{\infty} \phi(1/(1+n)) < \infty$, where ϕ is the modulus of uniform continuity of $\log g$ with respect to metric $d'(\sigma, \sigma') = 1/(|k|+1)$ (k is the same as in the definition of d), which is equivalent to the Dini-continuity of g with respect to metric d, since $\log x \leqslant x - 1$). The invariant measure of such a contractive Markov system is called a *g-measure*. This notion was introduced by Keane [**12**]. See [**3**, **6**, **7**, **17**] for more on this.

Acknowledgements. I would like to thank the EPSRC and the School of Mathematics and Statistics, University of St Andrews, for providing me with a scholarship and excellent working conditions in St Andrews, Professor K. J. Falconer for valuable comments on the first draft of the paper, my supervisor Lars Olsen for his interest in my work, his support and many fruitful discussions, Örjan Stenflo for informing me about the theory of dependence with complete connections, and Barry Ridge for proofreading the paper.

References

1. M. F. Barnsley, S. G. Demko, J. H. Elton and J. S. Geronimo, 'Invariant measures for Markov processes arising from iterated function systems with place-dependent probabilities', *Ann. Inst. H. Poincaré* 24 (1988) 367–394.

2. M. F. Barnsley, S. G. Demko, J. H. Elton and J. S. Geronimo, 'Erratum: Invariant measures for Markov processes arising from iterated function systems with place-dependent probabilities', *Ann. Inst. H. Poincaré* 25 (1989) 589–590.

3. H. Berbee, 'Chains with infinite connections: uniqueness and Markov representation', *Probab. Theory Related Fields* 76 (1987) 243–253.

4. P. Brémaud, *Markov chains: Gibbs fields, Monte Carlo simulation, and queues* (Springer, 1999).

5. J. H. Elton, 'An ergodic theorem for iterated maps', *Ergod. Theory Dynam. Systems* 7 (1987) 481–488.

6. P. Hulse, 'Uniqueness and ergodic properties of attractive *g*-measures', *Ergod. Theory Dynam. Systems* 11 (1991) 65–77.

7. P. Hulse, 'A class of unique *g*-measures', *Ergod. Theory Dynam. Systems* 17 (1997) 1383–1392.

8. J. E. Hutchinson, 'Fractals and self-similarity', *Indiana Univ. Math. J.* 30 (1981) 713–747.

9. M. Iosifescu and S. Grigorescu, *Dependence with complete connections and its applications* (Cambridge University Press, 1990).

10. T. Kaijser, 'On a new contraction condition for random systems with complete connections', *Rev. Roumaine Math. Pures Appl.* XXVI (1981) 1075–1117.

11. T. Kaijser, 'On a theorem of Karlin', *Acta Appl. Math.* 34 (1994) 51–69.

12. M. Keane, 'Strongly mixing *g*-measures', *Invent. Math.* 16 (1972) 309–324.

13. D. Lind and B. Marcus, *An introduction to symbolic dynamics and coding* (Cambridge University Press, 1995).

14. R. D. Mauldin and S. C. Williams, 'Hausdorff dimension in graph directed constructions', *Trans. Amer. Math. Soc.* 309 (1988) 811–829.

15. O. Onicescu and G. Mihoc, 'Sur les chaînes de variables statistiques', *Bull. Sci. Math.* 59 (1935) 174–192.
16. Ö. Stenflo, 'Uniqueness of invariant measures for place-dependent random iterations of functions', *Fractals in Multimedia* (*Minneapolis, MN*, 2001), IMA Volumes in Mathematics and its Applications 132 (Springer, New York, 2002) 13–32.
17. Ö. Stenflo, 'Uniqueness in g-measures', *Nonlinearity* 16 (2003) 403–410.
18. P. Walters, 'Ruelle's operator theorem and g-measures', *Trans. Amer. Math. Soc.* 214 (1975) 375–387.
19. I. Werner, 'Contractive Markov systems with constant probabilities', Preprint, 2003.
20. I. Werner, 'Coding map for a contractive Markov system', Preprint, 2003.
21. I. Werner, 'Kolmogorov–Sinai entropy of a generalized Markov shift', Preprint, 2003.
22. I. Werner, 'Ergodic theorem for contractive Markov systems', *Nonlinearity* 17 (2004) 2303–2313.

Ivan Werner
School of Mathematics and Statistics
University of St Andrews
St Andrews
Fife KY16 9SS
United Kingdom

iw11@st-andrews.ac.uk