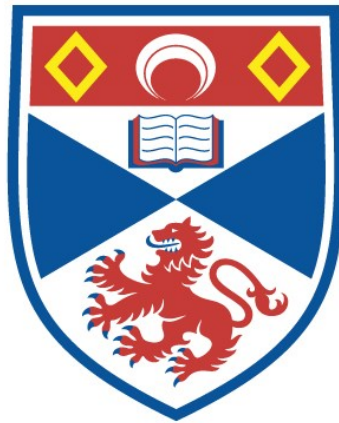# CONTRACTIVE MARKOV SYSTEMS

Ivan Werner

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews

2004

Full metadata for this item is available in
St Andrews Research Repository
at:
http://research-repository.st-andrews.ac.uk/

Please use this identifier to cite or link to this item:
http://hdl.handle.net/10023/15173

CONTRACTIVE MARKOV SYSTEMS

By

Ivan Werner

ProQuest Number: 10166158

ProQuest 10166158

TL
E729

UNIVERSITY OF ST ANDREWS

SCHOOL OF MATHEMATICS AND STATISTICS

(i) I, Ivan Werner, hereby certify that this thesis, which is approximately 25000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

date 23/11/04   signature of candidate ....

(ii) I was admitted as a research student in February 2003 and as a candidate for the degree of Doctor of Philosophy in February 2003; the higher study for which this is a record was carried out in the University of St Andrews in 2003-2004.

date 23/11/04   signature of candidate ....

(iii) I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Doctor of Philosophy in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

date 1/12/04   signature of supervisor ....

UNIVERSITY OF ST ANDREWS

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and abstract will be published, and that a copy of the work may be made and supplied to any *bona fide* library or research worker.

date 23/11/04   signature of candidate ....

iii

To my grandmother
Людмиле Петровне Мартыненко.

Contents

v

vi

I MUST SAY TO YOU YOUNG PEOPLE

"DON'T FOLLOW WHAT SENIOR MATHEMATICIANS SAY TO YOU
IF YOU, YOURSELF, ARE NOT FULLY CONVINCED,"...

Masamichi Takesaki (senior mathematician) [23]

Abstract

We introduce a theory of *contractive Markov systems (CMS)* which provides a unifying framework in so-called "fractal" geometry. It extends the known theory of *iterated function systems (IFS) with place dependent probabilities* [1][8] in a way that it also covers *graph directed constructions* of "fractal" sets [18]. Such systems naturally extend finite Markov chains and inherit some of their properties.

In Chapter 1, we consider iterations of a Markov system and show that they preserve the essential structure of it.

In Chapter 2, we show 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 restrictions of the probability functions on their vertex sets are Dini-continuous and bounded away from zero, and the system satisfies a condition of a contractiveness on average. This generalizes a result from [1]. Furthermore, we show that the rate of convergence to the stationary state is exponential in the aperiodic case with constant probabilities and a compact state space.

In Chapter 3, we construct a coding map for a contractive Markov system.

In Chapter 4, we calculate Kolmogorov-Sinai entropy of the generalized Markov shift.

In Chapter 5, we prove an ergodic theorem for Markov chains associated with the contractive Markov systems. It generalizes the ergodic theorem of Elton [8].

Acknowledgements

I would like to thank: EPSRC and School of Mathematics and Statistics of University of St Andrews for providing me with a scholarship and excellent working conditions in St Andrews, Professor K. J. Falconer and Professor A. Manning for a constructive criticism and various corrections to the thesis, 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, Barry Ridge for proofreading the manuscript.

St Andrews   Ivan Werner
November 22, 2004

Introduction

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 of Onicescu and Mihoc [19]. The reader is referred to Kaijser [14], Barnsley et al. [1] and Stenflo [21] for historical reviews.

Our work can be seen as a continuation of works of Barnsley et al. [1] and Elton [8], which were motivated by computer modelling of "fractal" measures. This addresses a 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 $(p_e)_{e \in E}$ (i.e. $p_e(x) \geq 0$ 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

1

the following form

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

Obviously, for any Borel subset B, $U1_B(x)$ defines a transition probability from the point x into the set B. Such systems have been employed for modelling different Markov processes long before (see the literature above) and were rediscovered by Hutchinson [12] (though 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 [8]. However, as we will see further (Remarks 2.1.1), their setting does *not* extend the case of a finite metric space, which is already very well understood. Related to the constructions of "fractal" sets, Mauldin and Williams [18] 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.

The theory does not claim to provide the most general model concerning its probabilistic phenomenon, since there is a general theory of "dependence with complete connections" [13] 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.

Notation

We use the following notation.

(K, d) is a metric space. All the following spaces of functions on K are real. $Lip(K)$ denotes the space of all Lipschitz functions, $C_C(K)$ denotes the space of all continuous functions with compact support, $C_B(K)$ denotes the space of all bounded continuous functions, $C(K)$ denotes the space of all continuous functions, $\mathcal{L}^0(K)$ 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 $f \in C_B(K)$, $\|f\|$ 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, $\xrightarrow{w^*}$ means "converges weakly* (weakly) to". We use the abbreviation "iff" for "if and only if".

Chapter 1

Markov systems

1.1 Main definitions

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

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

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

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

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

4
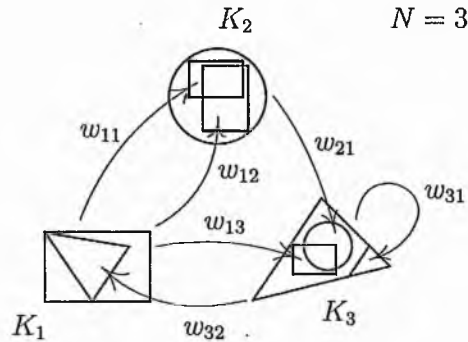
Fig. 1

Remark 1.1.1. (*i*) Case $N = 1$ covers the framework from [1] and [8].

(*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 1.1.4). This is another way to see how the framework from [1] and [8] can be embedded into ours.

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

Definition 1.1.1. We call the set $V := \{1, ..., N\}$ *the set of vertices* and the subsets $K_1, ..., K_N$ are called *the vertex sets*. Further, we call the set

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

the set of edges and we use the following notations:

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

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

$$i : E \longrightarrow V$$
$$(j, n) \longmapsto j.$$

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

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

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

Definition 1.1.2. We call the family $\mathcal{M} := \left(K_{i(e)}, w_e, p_e \right)_{e \in E}$ a *Markov system*, and we call the family without probabilities, $\left(K_{i(e)}, w_e \right)_{e \in E}$, a *topological Markov system*.

Definition 1.1.3. A Markov system is called *irreducible* iff its directed graph is irreducible, i.e. there is a path from any vertex to any other. An irreducible Markov system is said to have a *period d* iff its directed graph has a period d, i.e. the set of vertices can be partitioned into d non-empty subsets $\Omega_1, \Omega_2, ..., \Omega_d$ such that

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

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

Definition 1.1.4. We define the Markov operator on $\mathcal{L}^0(K)$ associated with the Markov system by

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

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

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

Definition 1.1.5. We say a probability measure μ is an *invariant probability measure* of the Markov system iff it is a stationary initial distribution of the associated Markov process, i.e.

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

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 1.1.6. Set

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

and

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

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

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

and

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

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

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

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

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

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 represent the Markov process, associated with the CMS, with the initial distribution δ_x. Moreover, obviously

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

1.2 Iterations of a Markov system

In contrast to the trivial case of finite Markov chains, here can be considered the following iterations of a Markov system.

Definition 1.2.1. Let $\mathcal{M} := \left(K_i, \left(w_{ij} \right)_{j \in J_i}, \left(p_{ij} \right)_{j \in J_i} \right)_{i \in I}$ 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 n-th 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 $w_{ij}^n \left(K_i^n \right)$, $i \in I^n$, $j \in J_i^n$.

That can be done by the following algorithm:

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

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

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

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

and

$$
\begin{aligned}
\Xi_m(s) \ &:= \ \Xi_{m-1}(s) \cup A_m(s), \ \text{where} \\
A_m(s) \ &:= \ \begin{cases} w_{e_m}\left(K_{i(e_m)}\right) &, \text{if } \Xi_{m-1}(s) \cap w_{e_m}\left(K_{i(e_m)}\right) \neq \emptyset \\ \emptyset &, \text{else} \end{cases}
\end{aligned}
$$

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

3. Set

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

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

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

So, $J_i^{n+1} := J_{\hat{i}}^n$. Through it, $\mathcal{M}^{n+1}$ is well-defined up to the indices.

Example 1.2.1. *The Fig. 2 shows the 1-st iteration of the Markov system from Fig.
1.*
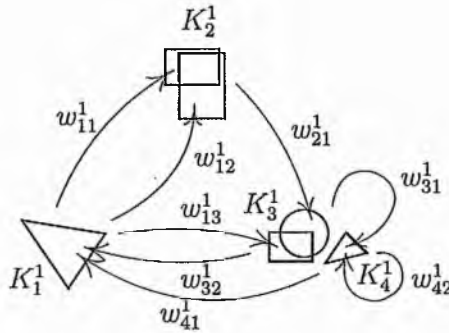
Fig. 2

Proposition 1.2.1. *A measure is invariant w.r.t. a Markov system iff it is invariant
w.r.t. one of its iterations.*

PROOF. Obvious by the definition of the iterations.   □

Remark 1.2.1. 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 [17]). It
is not difficult to see that the *shifts of finite type* defined by two directed graphes
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 sub-shift space as the original directed graph. And the difference between them is only in what we consider as separate vertex sets.

Lemma 1.2.2. *Suppose* $\left(K_{i(e)}, w_e, p_e\right)_{e \in E}$ *is an irreducible Markov system with an invariant probability measure* μ. *Then* $\mu\left(K_i\right) > 0$ *for all* $i = 1, ..., N$.

PROOF. Let $i_0 \in V$ such that $\mu\left(K_{i_0}\right) > 0$. 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

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

it follows that $\mu\left(K_j\right) > 0$. 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 $\mu\left(K_{j_0}\right) > 0$. $\qquad \square$

Chapter 2

Contractive Markov systems

In this chapter we assume that (K, d) is a metric space in which sets of finite diameter are relatively compact. It implies that (K, d) is a complete locally compact separable metric space.

2.1 Introduction

If we try to represent a Bernoulli process on a finite state space, say $\{1, ..., N\}$, 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 maps the whole space $\{1, ..., N\}$ 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 2.1.1 (CMS). We call a Markov system $\left(K_{i(e)}, w_e, p_e\right)_{e \in E}$ *contractive* iff it satisfies the following *condition of a contractiveness on average*: there exists

$0 < a < 1$ such that

$$\sum_{e \in E} p_e(x) d(w_e(x), w_e(y)) \le ad(x, y) \text{ for all } x, y \in K_i \text{ and } i \in \{1, ..., N\} \qquad (2.1.1)$$

(it is understood here that p_e's are extended on the whole space by zero and w_e's arbitrarily). We call a Markov chain with values in K *contractive* iff it is determined by a contractive Markov system. We call the constant a an *average contracting rate* of the Markov system.

Definition 2.1.2. We call a function $f : (X, d) \longrightarrow \mathbb{R}$ *Dini-continuous* iff 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, i.e.

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

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 2.1.1. *Let $0 < c < 1$ and $b > 0$. A real function f is Dini-continuous iff*

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

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 an increasing function,

$$\phi(bc^{n+1})(1-c) \leq \int\limits_{bc^{n+1}}^{bc^n} \frac{\phi(t)}{t}dt \leq \phi(bc^n)\left(\frac{1}{c}-1\right)$$

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

$$(1-c)\sum_{n=1}^{\infty} \phi(bc^n) \leq \int\limits_0^b \frac{\phi(t)}{t}dt \leq \sum_{n=0}^{\infty} \phi(bc^n)\left(\frac{1}{c}-1\right).$$

$\square$

Remark 2.1.1. Elton in [8] 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 a contractiveness on average: there exists $0 < r_1 < 1$ such that

$$\prod_{e \in E} d(w_e(x), w_e(y))^{p_e(x)} \leq r_1 d(x,y) \text{ for all } x,y \in K. \qquad (2.1.2)$$

There is a widely spread view in the literature that demanding condition (2.1.2) rather than (2.1.1) (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 (2.1.1) (with $N = 1$) in place of (2.1.2).

PROOF. First, observe that the condition (2.1.1) and the boundedness away from zero of the probability functions (i.e. there exists $\delta > 0$ such that $p_e \geq \delta$ for all $e \in E$) imply that the maps $(w_e)_{e \in E}$ are Lipschitz. Taking the logarithm of (2.1.2) and using its concavity reveals that (2.1.1) implies (2.1.2).

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

$$\sum_{e \in E} p_e(x) d(w_e(x), w_e(y))^q \leq r d(x, y)^q \text{ for all } x, y \in K. \qquad (2.1.3)$$

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 a loss of generality, to the condition (2.1.1).   □

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

$$\sum_{e \in E: \; d(w_e(x), w_e(y)) \leq r d(x, y)} p_e(x) p_e(y) \geq \delta^2 > 0 \text{ for all } x, y \in K. \qquad (2.1.4)$$

In fact, now conditions (2.1.4) and (2.1.3) 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, the condition (2.1.4) would not work for the Elton's proof of the corresponding ergodic theorem in [8]. So, 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 2.1.2. (i) A similar structure was discovered by Kaijser in setup of Random Systems with Complete Connections (RSCC) in [14]. However, what he calls *weakly distance diminishing RSCC* covers only aperiodic CMS's with compact state space.

(ii) Of course, the contractiveness of a Markov system can be weakened, just as it is done some times for maps, by demanding that a contraction on average happens

eventually not after one but after a number of iterations, i.e. there exist $r \in \mathbb{N}$ and $0 < a < 1$ such that

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

where P_x is a probability measure which represents the Markov process starting in x (see Def. 1.1.6). However, such systems, again just as in the case of maps, are not expected to exhibit a substantially new behavior, but a decrease of transparency of the proofs, for such systems, can be expected.

2.2 Contractive Markov systems with continuous probabilities

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

Definition 2.2.1. We call the partition $K_1, ..., K_N$ of K *open* iff every K_i, $i = 1, ..., N$, is an open subset of K. Of course, it means that K must be disconnected.

Theorem 2.2.1. *Suppose* $\left(K_{i(e)}, w_e, p_e\right)_{e \in E}$ *is a CMS with an average contracting rate* $0 < a < 1$ *such that the family* $K_1, ..., K_N$ *is an open partition of* K *and each* p_e *is continuous on* $K_{i(e)}$. *Then:*

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

(ii) The CMS has an invariant Borel probability measure μ.

(iii) The invariant probability measure μ is unique iff

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

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

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

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

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

and let $C > 0$ be such that

$$\max_{e \in E} d\left(w_e x_{i(e)}, x_{t(e)}\right) \le C.$$

We show inductively that

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

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

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

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

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

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

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

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

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

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

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

where $\bar{B}_{\rho/\epsilon}(y)$ denotes the closed ball of radius ρ/ϵ and center y. Then Q_ϵ is compact (note that, since Z_k^x are measurable, all sets considered here are measurable) and

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

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

(ii) Define an operator

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

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. So, 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,

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

However, since

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

We conclude that

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

i.e. μ is an invariant Borel probability measure on the CMS.

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

$$\frac{1}{n} \sum_{k=1}^{n} U^k g(x) \to \int g \, d\mu \text{ for all } x \in K \text{ and } g \in C_B(K). \qquad (2.2.1)$$

Conversely, if (2.2.1) holds true, by Lebesgue's Dominated Convergence Theorem, it implies that

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

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 $i = 1, ..., N$. Let ν be the Borel probability measure on K given by

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

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 proof of (i),

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

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

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

for all $n \in \mathbb{N}$ and $R > 0$. By (*iii*) and Lebesgue's Dominated Convergence Theorem, this implies that

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

By Levi's Theorem, we conclude that

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

as desired. $\qquad \square$

2.3 Contractive Markov systems with Dini-continuous probabilities

We intend to show here that Feller contractive Markov chains with probability functions which are Dini-continuous and bounded away from zero on their vertex sets exhibit a mixing behavior which is similar to the finite Markov chains.

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

Lemma 2.3.1. *Suppose* $\left(K_{i(e)}, w_e, p_e\right)_{e \in E}$ *is a CMS 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* $f \in C_G(K)$, *the functions* $(U^n f|_{K_i})_{n \in \mathbb{N} \cup \{0\}}$ *are uniformly equicontinuous for all* $i = 1, ..., N$.

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

$$\phi_0(t) := \left\{ \begin{array}{ll} t & , \ 0 \leq t \leq 1 \\ 1 & , \ t > 1 \end{array} \right.$$

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

Let $f \in Lip(K)$ and $\|f\| \leq 1$. Then there is $C \geq 2$ such that

$$|f(x) - f(y)| \leq Cd(x,y) \ \forall x,y \in K.$$

Set $L := \max\{L_1, ..., L_N\}$ 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 the Sublemma from [2], increasing ϕ if necessary, we can assume that β is concave. Further,

$$\beta(t) - \beta(at) = \frac{L \vee C}{1 - a} \int\limits_{t}^{ta^{-1}} \frac{\phi(u)}{u} du \geq \frac{L}{1 - a} \frac{\phi(t)}{ta^{-1}} t(a^{-1} - 1) = L\phi(t).$$

Hence

$$\beta(at) + L\phi(t) \leq \beta(t) \text{ for all } t \geq 0.$$

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

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

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

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

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

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

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

Since $Lip(K) \cap C_C(K)$ is a dense subset of $(C_C(K), \|.\|)$, the claim follows by an $\epsilon/3$-argument. $\square$

We will need to know more about properties of irreducible directed graphs. The following Lemma is a generalization of Lattice Theorem (see Theorem 4.3 in [6]).

Lemma 2.3.2 (Lattice Theorem). *Let an irreducible directed graph with period d be given. Then for every finite path $(e_1, ..., e_n)$ of the digraph, there exists $m_0 \geq 0$ such that for all integers $m \geq m_0$ there exists a closed path of the length md which has $(e_1, ..., e_n)$ as a part 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 the length k which has $(e_1, ..., e_n)$ as a part 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 of positive multiples of d (see Theorem 1.1 of the Appendix in [6]). In other words, there exists $m_0 \in \mathbb{N}$ such that for all $m \geq m_0$ there exists a path of the length md which has $(e_1, ..., e_n)$ as a part and starts with e_1. $\square$

Lemma 2.3.3. *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 ordered 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 pair $(\alpha, \beta) \in \mathcal{V}_i$ i is accessible from α and β by paths of the same length less than or equal to dr.*

PROOF. Let $(\alpha, \beta) \in \mathcal{V}_i$. Then there exist paths s_α and s_β respectively from i to α and from i to β of the same length, say $n_{\alpha\beta}$. By the Lattice Theorem, there exists $m_\alpha \in \mathbb{N}$ such that for all integers $m \geq m_\alpha$ there exists a closed path of the length md

which starts in i and has s_α as a part. Analogously, there exists $m_\beta \in \mathbb{N}$ such that for all integers $m \geq m_\beta$ there exists a closed path of the length md which starts in i and has s_β as a part. 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 the length dr which start in i and one of them has s_α as a part and the other has s_β as a part. Hence, there exist two paths of the same length $dr - n_{\alpha\beta} \leq dr$ where one of them is from α to i and the other is from β to i. $\square$

The next lemma is a generalization of Lemma 2.7 from [1]. 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 [14] and [15] for more on that.

Lemma 2.3.4. *Suppose* $\left(K_{i(e)}, w_e, p_e\right)_{e\in E}$ *is an irreducible CMS 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* $p_e|_{K_{i(e)}} \geq \delta$ *for all* $e \in E$. *Then:*

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

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

and the convergence is uniform on bounded subsets.

(ii) If in addition the CMS is aperiodic, then for every $f \in C_C(K)$

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

and the convergence is again uniform on bounded subsets.

PROOF. Let $S \subset K$ be bounded. We can assume $S \cap K_i \neq \emptyset$ for all $i = 1, ..., N$. 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

$$\max_{e \in E} d \left(w_e x_{i(e)}, x_{t(e)} \right) \leq C$$

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

$$\Sigma^* := \Sigma^+ \times \Sigma^+ = \left\{ \bar{e} := (e_1, \tilde{e}_1, e_2, \tilde{e}_2, ...) \mid (e_1, e_2, ...) \in \Sigma^+, (\tilde{e}_1, \tilde{e}_2, ...) \in \Sigma^+ \right\}$$

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

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

then $Z_n^{x_i}$ and $\tilde{Z}_n^{y_j}$ are independent Markov processes with initial distributions respectively δ_{x_i} and δ_{y_j} and $U^n f(x_i) = E(f \circ Z_n^{x_i})$ for all $f \in C_B(K)$, 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

$$\exists \tilde{i} \text{ s.t. } Z_m^{x_i}(\bar{e}), \tilde{Z}_m^{y_j}(\bar{e}) \in K_{\tilde{i}}, \ d \left(Z_m^{x_i}(\bar{e}), \tilde{Z}_m^{y_j}(\bar{e}) \right) \leq \alpha \text{ and } d \left(Z_l^{x_i}(\bar{e}), \tilde{Z}_l^{y_j}(\bar{e}) \right) > \alpha$$

$$\text{for all } l < m.$$

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

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

Denote by $\mathcal{B}_m$ the σ-Algebra in Σ^* generated by $Z_1^{x_i}, ..., Z_m^{x_i}, \tilde{Z}_1^{y_j}, ..., \tilde{Z}_m^{y_j}$. Then $G_{\alpha,m} \in$

$\mathcal{B}_m$. Now, for $f \in C_C(K)$,

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

Further, note that for $n > m$

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

Therefore,

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

Let $\epsilon > 0$ and choose, by Lemma 2.3.1, $\alpha > 0$ such that for all $u, v \in K_l$, $l = 1, \dots, N$,

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

Then

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

Thus, the proof of (ii) will be complete when we prove the following

Sublemma 2.3.5. *Suppose the CMS is irreducible and*

(i) $i = j$ or

(ii) the CMS is aperiodic.

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

PROOF. First, observe that

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

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

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

Repeating that we are led to

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

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

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

Repeating it we get

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

Hence

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

Set

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

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

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

So, by Markov inequality,

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

Analogously,

$$P^*\left(d(\tilde{Z}_{n_2}^{y_j}, y_{t(e_{n_2})}) > \lambda\middle| Z_{n_1}^{y_j}\right) \leq \frac{1}{2} \ P^*\text{-a.e..}$$

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

$$P^*\left(d(Z_{n_2}^{x_i}, x_{t(e_{n_2})}) \leq \lambda \text{ and } d(\tilde{Z}_{n_2}^{y_j}, y_{t(e_{n_2})}) \leq \lambda\middle| Z_{n_1}^{x_i}, Z_{n_1}^{y_j}\right) \geq \frac{1}{4} \ P^*\text{-a.e..}$$

Note that the average contractiveness condition,

$$\sum_{e\in E} p_e(u)d(w_e u, w_e v) \leq ad(u,v) \text{ for all } u,v \in K_i, \ i = 1, ..., N,$$

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

Now, in case (i), by Lemma 2.3.3, there exists $r \geq 0$ 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 less than or equal to dr, where d is the period of the CMS. In case (ii),

i.e. $d = 1$, 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 implies that

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

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

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

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

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

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 so large that $a^k \left(2s^{dr} \lambda + \rho_{dr} \right) < \alpha$. Let $n_2 \geq \gamma n_1 + dr + k$. Then

$$P^* \left(e_{n_2-l} = \tilde{e}_{n_2-l} \text{ and } d(Z_{n_2-l}^{x_i}, \tilde{Z}_{n_2-l}^{y_j}) \leq a d(Z_{n_2-l-1}^{x_i}, \tilde{Z}_{n_2-l-1}^{y_j}) \right.$$
$$\left. \text{for all } l = 0, \dots, k-1 \,\Big|\, Z_{n_1}^{x_i}, \tilde{Z}_{n_1}^{y_j} \right) \geq \delta^{2(k+dr)} \ P^*\text{-a.e. (see Fig. 3).}$$
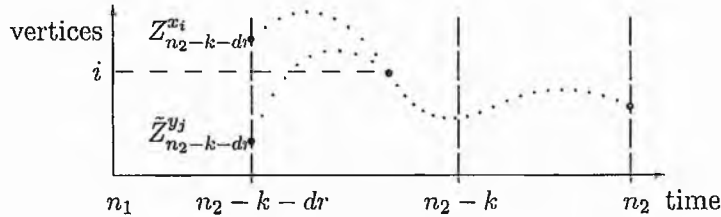
Fig. 3

Then, by the above and the Markov property,

$$P^* \left(e_{n_2-l} = \tilde{e}_{n_2-l}, \ d(Z_{n_2-l}^{x_i}, \tilde{Z}_{n_2-l}^{y_j}) \leq a d(Z_{n_2-l-1}^{x_i}, \tilde{Z}_{n_2-l-1}^{y_j}) \ \text{for all} \right.$$
$$l = 0, \dots, k-1, \ d(Z_{n_2-k-dr}^{x_i}, x_{t(e_{n_2-k-dr})}) \leq \lambda \text{ and}$$
$$\left. d(\tilde{Z}_{n_2-k-dr}^{y_j}, y_{t(e_{n_2-k-dr})}) \leq \lambda \,\Big|\, Z_{n_1}^{x_i}, \tilde{Z}_{n_1}^{y_j} \right) \geq \frac{1}{4} \delta^{2(k+dr)} \ P^*\text{-a.e..}$$

Observe that

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

P^*-a.e.. Hence,

$$
P^*\left(d(Z^{x_i}_{n_2}, \tilde{Z}^{y_j}_{n_2}) \leq a^k(2s^{dr}\lambda + \rho_{dr}) \,\Big|\, Z^{x_i}_{n_1}, \tilde{Z}^{y_j}_{n_1}\right) \geq \frac{1}{4}\delta^{2(k+dr)} \ P^*\text{-a.e..}
$$

Thus

$$
P^*\left(d(Z^{x_i}_{n_2}, \tilde{Z}^{y_j}_{n_2}) > \alpha \,\Big|\, Z^{x_i}_{n_1}, \tilde{Z}^{y_j}_{n_1}\right) \leq 1 - \frac{1}{4}\delta^{2(k+dr)} \ P^*\text{-a.e..}
$$

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

$$
P^*\left(d(Z^{x_i}_{n_t}, \tilde{Z}^{y_j}_{n_t}) > \alpha, \ t = 1, \ldots, m\right) \leq \left(1 - \frac{1}{4}\delta^{2(k+dr)}\right)^{m-1} \text{ for all } m \in \mathbb{N}.
$$

Hence

$$
P^*\left(B_{\alpha,n}\right) \leq \left(1 - \frac{1}{4}\delta^{2(k+dr)}\right)^{m-1} \text{ if } n \geq n_m.
$$

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

Definition 2.3.1. A measure $\mu \in P(X)$ is called *the attractive measure of the CMS* iff

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

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

Theorem 2.3.6. *Suppose* $\left(K_{i(e)}, w_e, p_e\right)_{e \in E}$ *is an irreducible CMS such that* $K_1, ..., K_N$ *is an open partition of* K, $p_e|_{K_{i(e)}}$ *is Dini-continuous and there exists* $\delta > 0$ *such that* $p_e|_{K_{i(e)}} \geq \delta$ *for all* $e \in E$. *Then:*

(i) The CMS has a unique invariant Borel probability measure μ.

(ii) If in addition the CMS is aperiodic, then

$$U^n f(x) \to \mu(f) \text{ for all } x \in K \text{ and } f \in C_B(K)$$

and the convergence is uniform on bounded subsets, i.e. μ *is an attractive probability measure.*

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

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

Since, by Lemma 2.3.4 (i), for every $f \in C_C(K)$,

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

we conclude that for every $f \in C_C(K)$

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

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

$$\lim_{k \to \infty} \frac{1}{n_k} \sum_{l=1}^{n_k} U^l f(x) = \sum_{i=1}^{N} \mu_i(f) 1_{K_i}(x) \text{ for all } x \in K \text{ and all } f \in C_B(K). \quad (2.3.1)$$

Define a linear operator $Q : C_B(K) \longrightarrow C_B(K)$ by

$$Q(f) := \sum_{i=1}^{N} \mu_i(f) 1_{K_i} \text{ for all } f \in C_B(K). \qquad (2.3.2)$$

Then, by (2.3.1),

$$QU = Q$$

and therefore

$$Q^2 = Q. \qquad (2.3.3)$$

Now, by the definition of μ_i, $U^* \mu_i = \mu_i$ for all $i = 1, ..., N$. Since the CMS is irreducible, this implies, by Lemma 1.2.2, that $\mu_i(K_j) > 0$ for all $i, j = 1, ..., N$. Now, let $f \in C_B(K)$ with $f \geq 0$. Then, by (2.3.3),

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

i.e.

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

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

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

which obviously can not be true. Hence

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

Let $\mu := \mu_1$. Since $f \in C_B(K)$ with $f \geq 0$ was arbitrary, we conclude that all μ_i, $i = 1, ..., N$, are equal to μ. Hence,

$$\lim_{k \to \infty} \frac{1}{n_k} \sum_{l=1}^{n_k} U^l f(x) = \mu(f) \text{ for all } x \in K \text{ and } f \in C_B(K). \qquad (2.3.4)$$

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

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

but applying Lebesgue's Dominated Convergence Theorem to (2.3.4) implies that

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

Thus, $\lambda = \mu$, i.e. μ is a unique invariant Borel probability measure of the CMS.

(ii) Let $x \in K$. By Theorem 2.2.1 (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, i.e. $U^{n_k}f(x) \to \mu(f)$ $(k \to \infty)$ for all $f \in C_B(K)$. Since, by Lemma 2.3.4 (ii), $|U^{n_k}f(x) - U^{n_k}f(y)| \to 0$ for all $y \in K$ and for all $f \in C_C(K)$, it follows that $U^{n_k}f(y) \to \mu(f)$ for all $y \in K$ and $f \in C_C(K)$.

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

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

for all $g \in C_B(K)$ and all $n \in \mathbb{N}$. Let $f \in C_C(K)$. Since, by Lemma 2.3.1, the functions $(U^{n_k}f|_{K_i})_{k \in \mathbb{N}}$ are equicontinuous for each $i = 1, ..., N$, by Arzela-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

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

Thus, by the above,

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

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

$$
U^n f(x) \rightarrow \int f d\mu \text{ for all } x \in K \text{ and } f \in C_C(K).
$$

This also implies that

$$
U^n f(x) \rightarrow \int f d\mu \text{ for all } x \in K \text{ and } f \in C_B(K),
$$

the convergence is uniform on bounded subsets by Lemma 2.3.4 (ii). By Lebesgue's Dominated Convergence Theorem, we conclude that

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

$\square$

Example 2.3.1. *Every irreducible finite Markov chain is a contractive Markov chain satisfying the hypothesis of Theorem 2.3.6.*

Example 2.3.2. *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 $\mathbb{R}^2$*

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

$$
= \begin{pmatrix} y \\ -\frac{1}{3}x + \frac{7}{6} \end{pmatrix}, \ w_4 \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ \frac{2}{3}y - \frac{1}{3} \end{pmatrix}, \ w_5 \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{1}{2}y \\ -\frac{2}{3}x + \frac{4}{3} \end{pmatrix}
$$

with probability functions

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

An easy calculation shows that they define a CMS with an average contracting rate 8/9 on $K_1 \cup K_2 \cup K_3$, as it is shown on Fig. 4, which satisfies the hypothesis of Theorem 2.3.6 (ii) and does not satisfy the hypothesis of Theorem 2.1 in [1].
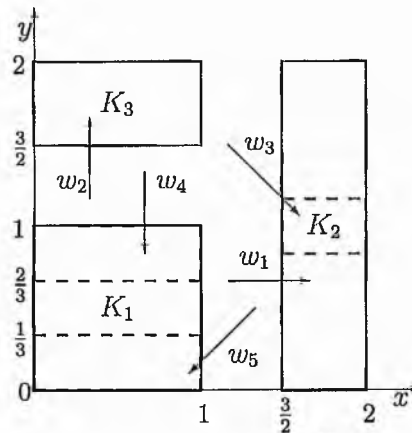
Fig. 4

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 2.3.3. *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 $\sigma := (..., \sigma_{-1}, \sigma_0)$ 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 $k < i \leq 0$. Let g be a positive, Dini-continuous function on Σ_G such that

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

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 $e \in E$,

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

Obviously, maps $(w_e)_{e \in E}$ are contractions. Therefore, $\left(K_{i(e)}, w_e, p_e\right)_{e \in E}$ defines a CMS which satisfies the hypothesis of Theorem 2.3.6 and does not satisfy the hypothesis of Theorem 2.1 in [1]. Hence, Theorem 2.3.6 (ii) covers Theorem 3.1 in [24] (there, it was assumed that $\sum_{n=0}^{\infty} \phi(1/(1+n)) < \infty$ where ϕ is the modulus of uniform continuity of $\log g$ w.r.t. 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 w.r.t. metric d, since $\log x \leq x - 1$). The invariant measure of such a CMS is called a g-measure. This notion was introduced by M. Keane [16]. See [4], [10], [11], [22] for more on that.

2.4 Contractive Markov systems with constant probabilities

For many applications, it is sufficient to consider the subsets $K_1, ..., K_N$ to be compact and each probability function p_e to be constant and positive on $K_{i(e)}$ (as in Example

2.3.2). For such systems, an easy proof can be given to show that their long term behaviour is analogous to the finite Markov chains by using some well known facts about stochastic matrices. We are going to present such a proof here. In addition to the results in the previous section, we show here that the rate of convergence to the stationary state is exponential in the aperiodic case with the above assumptions.

So, let the subsets $K_1, ..., K_N$ be compact and each probability function p_e be constant and positive on $K_{i(e)}$ and zero on the complement to $K_{i(e)}$.

Remark 2.4.1. If $N = 1$ and all maps w_e are contractive, then we get the case considered by Hutchinson [12].

Since with each edge e there is an associated probability weight p_e, the directed graph describes in particular a finite Markov chain with the state space V and transition probabilities

$$a_{ij} := \sum_{e \in E, \ i(e)=i, \ t(e)=j} p_e \text{ for all } i, j \in V,$$

provided that an initial probability distribution $r := (r_1, ..., r_N)$ is given on V. Then the probability distribution on V at each following time is calculated by multiplying the distribution at the previous time as a row vector from the left with the transition matrix

$$A := (a_{ij})_{1 \le i, j \le N} . \qquad (2.4.1)$$

At this point, it is appropriate to remind ourselves of some definitions and facts about finite Markov chains. Good references for that are e.g. [5],[6].

Definition 2.4.1. (i) A finite Markov chain and its transition matrix A are called *irreducible* iff for all $i, j \in \{1, ..., N\}$ there is $n \in \mathbb{N}$ such that $a_{ij}(n) > 0$ where

$(a_{ij}(n))_{1 \leq i,j \leq N} := A^n$. Since $a_{ij}(n)$ is the probability for a transition with n steps from i to j, the irreducibility means that for all $i, j \in V$ there is always a finite path from i to j in the directed graph.

(ii) An irreducible finite Markov chain and its transition Matrix are said to have a *period* d iff their directed graph has a period d, i.e. the set of vertices can be partitioned into d non-empty subsets $\Omega_1, \Omega_2, ..., \Omega_d$ such that

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

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

Theorem 2.4.1. *Let A be an irreducible, stochastic $N \times N$-matrix. Then there exists a unique probability vector r_0 such that $r_0 A = r_0$. Furthermore, $r_{0i} > 0$ for all $i = 1, ..., N$. If the matrix is in addition aperiodic, then there exists $\lambda \in [0, 1)$ such that*

$$\|rA^n - r_0\|_1 \leq \lambda^n \chi(r, r_0) \text{ for any probability vector } r \text{ and } n \in \mathbb{N}, \qquad (2.4.2)$$

where

$$\chi(r, r_0) := \left(\sum_{i=1}^{N} \frac{(r_i - r_{0i})^2}{r_{0i}} \right)^{\frac{1}{2}}$$

and λ is the positive square root of the second-largest eigenvalue of the matrix $A\tilde{A}$ where $\tilde{A} := D^{-1}AD$ and $D := diag\{r_{01}, ..., r_{0N}\}$ if $N \geq 2$, or $\lambda = 0$ if $N = 1$.

Definition 2.4.2. A finite Markov chain with the property (2.4.2) is called historically *geometrically ergodic*.

For a proof of Theorem 2.4.1 we refer to [6] (Theorem 3.4, p. 211).

Now, consider an equivalence relation on $P(K)$ given by

$$\mu \sim \nu :\Longleftrightarrow \mu(K_i) = \nu(K_i) \ \forall i \in \{1, ..., N\}.$$

Let $\bigcup_{r \in R} M_r = P(K)$ be the partition of $P(K)$ which is imposed by the equivalence relation with the set of equivalence classes

$$P(K)/_\sim \ \cong \ R := \left\{ (r_1, ..., r_N) \in \mathbb{R}^N : \sum_{i=1}^N r_i = 1, \ r_i \geq 0 \ \forall i \right\}.$$

For convenience we consider R to be normed by $\|.\|_1$,

$$\|r\|_1 := \sum_{i=1}^N |r_i|, \ r \in R.$$

Further, we define a metric $\tilde{L}$ on each equivalence class M_r which generates the weak*-topology on it. Set

$$S(K) := \{f \in C(K) : \forall 1 \leq i \leq N \ \forall x, y \in K_i \ |f(x) - f(y)| \leq d(x, y)\}$$

and

$$\tilde{L}(\mu, \nu) := \sup_{f \in S(K)} |\mu(f) - \nu(f)| \text{ for } \mu, \nu \in M_r, \ r \in R.$$

Remark 2.4.2. Obviously $\tilde{L}(\mu, \nu) \geq L(\mu, \nu)$ for $\mu, \nu \in P(K)$ where L is the metric used by J. Hutchinson (see [12])

$$L(\mu, \nu) = \sup_{Lip(f) \leq 1} |\mu(f) - \nu(f)|,$$

where $Lip(f)$ is the Lipschitz constant of f. It is well known that L generates the weak* topology on $P(K)$.

If μ and ν are from different equivalence classes then $\tilde{L}(\mu, \nu)$ is infinite.

Proposition 2.4.2. $\tilde{L}$ *is a metric on M_r which generates the weak*-topology on it for every $r \in R$, and*

$$\tilde{L}(\mu, \nu) \leq 2 \max_i diam(K_i) \text{ for all } \mu, \nu \in M_r, \ r \in R.$$

Proof. Let $r \in R$. We show first that $\tilde{L}$ is finite on M_r.

Let $\mu, \nu \in M_r$, $f \in S(K)$ and $x_i \in K_i$ $\forall 1 \leq i \leq N$. Then

$$
\begin{aligned}
|\mu(f) - \nu(f)| \quad &\leq \quad \sum_{i=1}^{N} |\mu(1_{K_i}f) - \nu(1_{K_i}f)| \\
&\overset{(\mu,\nu\in M_r)}{\leq} \quad \sum_{i=1}^{N} [|\mu(1_{K_i}f - 1_{K_i}f(x_i))| + |\nu(1_{K_i}f - 1_{K_i}f(x_i))|] \\
&\leq \quad \sum_{i=1}^{N} [diam(K_i)\,\mu(K_i) + diam(K_i)\,\nu(K_i)] \\
&\leq \quad 2 \max_i diam(K_i).
\end{aligned}
$$

This shows that $\tilde{L}(\mu, \nu) \leq 2 \max_i diam(K_i)$.

By Remark 2.4.2,

$$\tilde{L}(\mu, \nu) = 0 \ \Rightarrow \mu = \nu.$$

The remaining metric properties are obvious.

Now, we verify the equivalence

$$\mu_k \overset{w^*}{\to} \mu \iff \tilde{L}(\mu_k, \mu) \to 0, \text{ for } \mu_k, \mu \in M_r.$$

The direction "$\Leftarrow$" holds true by Remark 2.4.2. For "$\Rightarrow$", let $\mu_k, \mu \in M_r$ with $\mu_k \overset{w^*}{\to} \mu$. Suppose $\lim_{k\to\infty} \tilde{L}(\mu_k, \mu) \neq 0$. Then there are $\epsilon > 0$ and a subsequence, without a loss of generality, $(\mu_k)_{k\in\mathbb{N}}$ such that $\tilde{L}(\mu_k, \mu) > \epsilon$ $\forall k \in \mathbb{N}$. Hence, there exists a sequence $(f_k)_{k\in\mathbb{N}} \subset S(K)$ such that

$$|\mu_k(f_k) - \mu(f_k)| > \epsilon \ \forall k \in \mathbb{N}. \tag{2.4.3}$$

Fix $x_i \in K_i$ for each $i = 1, ..., N$. Then the sequence $(f_k - f_k(x_i))$ is equicontinuous and bounded on K_i for each $1 \leq i \leq N$. By Arzelà-Ascoli Theorem, it follows that there exist $g_i \in C(K_i)$ for every $1 \leq i \leq N$ and a subsequence, without loss of generality, $(f_k)_{k \in \mathbb{N}}$ such that $\|(f_k - f_k(x_i)) - g_i\|_{K_i} \to 0$ for all i. Define $g := \sum_{i=1}^{N} g_i 1_{K_i}$ (with an arbitrary extension of g_i on K). Then $g \in C(K)$ and

$$\left\| f_k - \sum_{i=1}^{N} f_k(x_i) 1_{K_i} - g \right\| \leq \sum_{i=1}^{N} \|(f_k - f_k(x_i) - g_i) 1_{K_i}\| \to 0.$$

Thus

$$|\mu_k(f_k) - \mu(f_k)| \overset{\mu_k, \mu \in M_r}{=} \left| \mu_k \left(f_k - \sum_{i=1}^{N} f_k(x_i) 1_{K_i} \right) - \mu \left(f_k - \sum_{i=1}^{N} f_k(x_i) 1_{K_i} \right) \right|$$

$$\leq \quad 2 \left\| f_k - \sum_{i=1}^{N} f_k(x_i) 1_{K_i} - g \right\| + |\mu_k(g) - \mu(g)| \to 0$$

which is a contradiction to (2.4.3).   □

Lemma 2.4.3. (i) *For any* $r \in R$ *there exists* $s \in R$ *such that* $U^* \mu \in M_s$ *for all* $\mu \in M_r$. *Thus, the operator* U^* *defines a map* T *through*

$$T : \quad R \longrightarrow R$$

$$r \longmapsto (U^* \mu(K_1), ..., U^* \mu(K_N)), \text{ where } \mu \in M_r.$$

(ii) *For all* $r \in R$

$$T(r) = rA,$$

i.e. $U^* \nu \in M_{rA}$ *for all* $\nu \in M_r$, *where* A *is the transition matrix (2.4.1).*

Proof. Let $\mu(K_i) = \nu(K_i) =: r_i$ for all $i = 1, ..., N$. We show $U^* \mu(K_i) = U^* \nu(K_i)$ for all $i = 1, ..., N$.

For each $i \in \{1, ..., N\}$,

$$U^* \mu(K_i) = \sum_{e \in E} p_e \int_{K_{i(e)}} 1_{K_i} \circ w_e d\mu = \sum_{e \in E, \ t(e)=i} p_e \mu \left(K_{i(e)} \right)$$

which implies (i) with

$$(Tr)_i = \sum_{e \in E, \ t(e)=i} p_e r_{i(e)} \text{ for all } i = 1, ..., N.$$

Thus, the map T is linear. Applying T to the natural basis $(e_j)_{j=1}^N \subset R$,

$$e_j := (0, ..., 0, \overset{\underset{j-\text{th position}}{\downarrow}}{1}, 0, ..., 0), \text{ gives}$$

$$(Te_j)_i = \sum_{e \in E, \ i(e)=j, \ t(e)=i} p_e = a_{ji} \text{ for all } i, j \in \{1, ..., N\}.$$

This shows that the representation matrix of T is exactly the transposed matrix of A. Hence

$$T(r) = rA \text{ for all } r \in R.$$

$\square$

Definition 2.4.3. We call a general Markov chain on a compact metric space K and its transfer operator U^* *geometrically ergodic* iff there exist a fixed point μ_0 of U^* and numbers $H > 0$, $\lambda \in [0, 1)$ such that

$$L \left(U^{*n} \nu, \mu_0 \right) \leq \lambda^n C(\nu, \mu_0) \text{ for all } \nu \in P(K), \ n \geq H \text{ and some } C(\nu, \mu_0) > 0.$$

Note that the Definition 2.4.3 is consistent with the Definition 2.4.2.

Theorem 2.4.4. (i)

$$\bar{L}(U^{*n} \mu, U^{*n} \nu) \leq a^n \bar{L}(\mu, \nu) \text{ for all } \mu, \nu \in M_r, \ r \in R, \ n \in \mathbb{N}.$$

(ii) For all $r \in R$,

$$rA = r \iff \quad \text{there exists a unique } \mu \in M_r \text{ such that } U^*\mu = \mu.$$

(iii) If the matrix A is aperiodic, then there exists a probability measure $\mu_0 \in P(K)$ such that for m large enough

$$L\left(U^{*m}\nu, \mu_0\right) \leq C(r, r_0)\lambda_1{}^m$$

where

$$C(r, r_0) := diam(K)\left(\chi(r, r_0) + 2a^{-1/2}\right),$$

$$r_0 = (\mu_0(K_1), ..., \mu_0(K_N))$$

and

$$\lambda_1 := \max\{\lambda, a\}^{1/2}$$

where λ is the positive square root of the second-largest eigenvalue of the matrix $A\tilde{A}$ (as in Theorem 2.4.1) if $N \geq 2$, or $\lambda = 0$ if $N = 1$. In other words, the operator U^* is geometrically ergodic.

Proof. Let $\mu, \nu \in M_r$ for some $r \in R$ and $f \in S(K)$. Then, for all $i \in \{1, ..., N\}$ and $x, y \in K_i$,

$$\left| \frac{1}{a} \sum_{e \in E,\ i(e)=i} p_e f \circ w_e(x) - \frac{1}{a} \sum_{e \in E,\ i(e)=i} p_e f \circ w_e(y) \right| \leq \frac{1}{a} \sum_{e \in E,\ i(e)=i} p_e d(w_e(x), w_e(y))$$
$$\leq d(x, y),$$

i.e.

$$\frac{1}{a} \sum_{e \in E} p_e f \circ w_e \in S(K).$$

Therefore

$$
\begin{aligned}
|U^*\mu(f) - U^*\nu(f)| &= a\left|\mu\left(\frac{1}{a}\sum_{e\in E}p_e f\circ w_e\right) - \nu\left(\frac{1}{a}\sum_{e\in E}p_e f\circ w_e\right)\right| \\
&\leq a\tilde{L}(\mu,\nu).
\end{aligned}
$$

Hence

$$
\tilde{L}(U^*\mu, U^*\nu) \leq a\tilde{L}(\mu,\nu).
$$

Thus the claim in (i) follows inductively by Lemma 2.4.3.

The claim in (ii) follows from (i) by Banach's Fixed Point Theorem and Lemma 2.4.3 because $\sim$-equivalence classes are weakly*-closed.

Now, let the matrix A be irreducible and aperiodic. Then by Theorem 2.4.1, there exists a unique fixed point $r_0 \in R$ of A and it has the following properties: $r_{0i} > 0$ for all $i = 1, ..., N$ and there exists $\lambda \in [0, 1)$ such that

$$
\|rA^n - r_0\|_1 \leq \chi(r, r_0)\lambda^n \text{ for all } r \in R, \ n \in \mathbb{N}.
$$

Therefore, by (ii), the operator U^* has a unique fixed point $\mu_0 \in M_{r_0}$. Now, let $\nu \in P(K)$ be arbitrary and $r \in R$ such that $\nu \in M_r$. Since $r_{0i} > 0$ for all $i = 1, ..., N$, there exists $H \in \mathbb{N}$ such that

$$
U^{*n}\nu(K_i) > 0 \text{ for all } i = 1, ..., N \text{ and } n \geq H.
$$

Define a sequence $(\nu_n)_{n\geq H}$ in M_{r_0} by

$$
\nu_n(B \cap K_i) := \frac{U^{*n}\nu(B \cap K_i)}{U^{*n}\nu(K_i)}r_{0i} \text{ for all Borel } B \subset K \text{ and } i = 1, ..., N.
$$

Denote by $(e_1, ..., e_k)^*$ a path of the directed graph of length k. Then, for $f \in C(K)$

with $Lip(f) \leq 1$ and $k, n \in \mathbb{N}$ with $n \geq H$,

$$
\left| U^{*k+n}\nu(f) - U^{*k}\nu_n(f) \right| = \left| \sum_{(e_1,\dots,e_k)^*} p_{e_1}\dots p_{e_k} U^{*n}\nu \left(f \circ w_{e_k} \circ \dots \circ w_{e_1} \right) \right.
$$
$$
\left. - \sum_{(e_1,\dots,e_k)^*} p_{e_1}\dots p_{e_k} \nu_n \left(f \circ w_{e_k} \circ \dots \circ w_{e_1} \right) \right| \quad (*).
$$

Set, for an abbreviation, $g_{e_1\dots e_k} := f \circ w_{e_k} \circ \dots \circ w_{e_1} - f \circ w_{e_k} \circ \dots \circ w_{e_1}(x_0)$ for some $x_0 \in K$. Then $|g_{e_1\dots e_k}| \leq diam(K)$ and

$$
(*) = \left| \sum_{(e_1,\dots,e_k)^*} p_{e_1}\dots p_{e_k} U^{*n}\nu \left(g_{e_1\dots e_k} \right) - \sum_{(e_1,\dots,e_k)^*} p_{e_1}\dots p_{e_k} \nu_n \left(g_{e_1\dots e_k} \right) \right|
$$
$$
\leq \sum_{(e_1,\dots,e_k)^*} p_{e_1}\dots p_{e_k} \sum_{i=1}^{N} \left| \int_{K_i} g_{e_1\dots e_k} dU^{*n}\nu - \frac{r_{0i}}{U^{*n}\nu(K_i)} \int_{K_i} g_{e_1\dots e_k} dU^{*n}\nu \right|
$$
$$
\leq \sum_{(e_1,\dots,e_k)^*} p_{e_1}\dots p_{e_k} \sum_{i=1}^{N} \frac{|U^{*n}\nu(K_i) - r_{0i}|}{U^{*n}\nu(K_i)} \int_{K_i} |g_{e_1\dots e_k}| dU^{*n}\nu
$$
$$
\leq \sum_{(e_1,\dots,e_k)^*} p_{e_1}\dots p_{e_k} \| r A^n - r_0 \|_1 \, diam(K)
$$
$$
\leq diam(K)\chi(r, r_0)\lambda^n.
$$

Hence

$$
L\left(U^{*k+n}\nu, U^{*k}\nu_n \right) \leq diam(K)\chi(r, r_0)\lambda^n.
$$

Thus, by Remark 2.4.2,

$$
L\left(U^{*k+n}\nu, \mu_0 \right) \leq L\left(U^{*k+n}\nu, U^{*k}\nu_n \right) + L\left(U^{*k}\nu_n, \mu_0 \right)
$$
$$
\leq diam(K)\chi(r, r_0)\lambda^n + \tilde{L}\left(U^{*k}\nu_n, U^{*k}\mu_0 \right).
$$

By (i) and Proposition 2.4.2,

$$
L\left(U^{*k+n}\nu, \mu_0 \right) \leq diam(K)\left(\chi(r, r_0)\lambda^n + 2a^k \right).
$$

Now, let $m \in \mathbb{N}$ such that $[m/2] \geq H$ where $[m/2]$ denotes the greatest natural number less than or equal to $m/2$. Then $m - [m/2] \geq m/2$ and $[m/2] \geq m/2 - 1/2$. Therefore, by the above,

$$\begin{aligned} L\left(U^{*m}\nu, \mu_0\right) &\leq diam(K)\left(\chi(r, r_0)\lambda^{m-[m/2]} + 2a^{[m/2]}\right) \\ &\leq diam(K)\left(\chi(r, r_0)\lambda^{\frac{m}{2}} + 2a^{\frac{m}{2}-\frac{1}{2}}\right). \end{aligned}$$

Define

$$C(r, r_0) := diam(K)\left(\chi(r, r_0) + 2a^{-1/2}\right)$$

and

$$\lambda_1 := \max\{\lambda, a\}^{1/2}.$$

Then

$$L\left(U^{*m}\nu, \mu_0\right) \leq C(r, r_0)\lambda_1{}^m.$$

This proves the claim in *(iii)*.   □

Now, we are able to calculate a relative rate of convergence to the stationary state in Example 2.3.2.

Example 2.4.1. *By Theorem 2.4.4 (iii), a simple calculation shows that for the CMS from Example 2.3.2*

$$L\left(U^{*m}\nu, \mu_0\right) \leq C(r, r_0)\left(\frac{\sqrt{10 + \sqrt{19}}}{3\sqrt{2}}\right)^{\frac{m}{2}} \text{ for all } \nu \in P(K),$$

where $r = (\nu(K_1), \nu(K_2), \nu(K_3))$ and $r_0 = (4/10, 3/10, 3/10)$.

Chapter 3

Coding map for a contractive Markov system

In this chapter we continue development of the theory of contractive Markov systems. We show here that the coding map is well defined for a contractive Markov system. First, we construct it with respect to an outer measure on the code space. Then we prove that it is also defined almost everywhere with respect to a generalized Markov measure if the probability functions are Dini-continuous and bounded away from zero on their vertex sets.

3.1   Introduction

In Chapter 2, we introduced a theory of *contractive Markov systems (CMS)* which provides a unifying framework in so-called "fractal" geometry.

The coding map is an important tool in "fractal" geometry which allows one to represent a constructed set as an image of a code space under this map, that is, to code

47

elements in this set by infinite sequences of elements of E (e.g. n-adic coding of the set $[0, 1]$ or binary coding of Cantor set). Such a coding map can be easily constructed for any contractive Markov system if all its maps are contractive. However, some fundamental difficulties arise in case of a general contractive Markov system. We overcome these difficulties by constructing an outer measure on the code space with respect to which we then define the coding map.

This coding map will play an important role in next two chapters.

3.2 Construction with respect to an outer measure

Let $\left(K_{i(e)}, w_e, p_e\right)_{e \in E}$ be a contractive Markov system with the average contracting rate $0 < a < 1$ and an invariant Borel probability measure μ. We assume $K_1, ..., K_N$ to be any disjoint non-empty Borel subsets of a complete metric space such that $K := \bigcup_{e \in E} K_{i(e)}$ is closed, the set of edges E to be finite. We do not impose any conditions on the directed graph.

Let $\Sigma := \{(..., e_{-1}, e_0, e_1, ...) : e_i \in E \; \forall i \in \mathbb{Z}\}$ and S be the left shift on Σ. Denote by $\mathcal{A}$ the finite σ-algebra generated by the partition $\{_0[e] : e \in E\}$ of Σ, and define, for each integer $m \le 1$,

$$\mathcal{A}_m := \bigvee_{i=m}^{+\infty} S^{-i}\mathcal{A},$$

which is the smallest σ-algebra containing all finite σ-algebras $\bigvee_{i=m}^{n} S^{-i}\mathcal{A}$, $n \ge m$. Let $x \in K$. For each integer $m \le 1$ let P_x^m be the probability measure on the σ-algebra $\mathcal{A}_m$ given by

$$P_x^m(_m[e_m, ..., e_n]) = p_{e_m}(x)p_{e_{m+1}}(w_{e_m}(x))...p_{e_n}(w_{e_{n-1}} \circ ... \circ w_{e_m}(x))$$

for all thin cylinders $_m[e_m, ..., e_n]$, $n \geq m$.

Lemma 3.2.1. *Let $m \leq 1$ and $A \in \mathcal{A}_m$. Then $x \longmapsto P_x^m(A)$ is a Borel measurable function on K.*

Proof. Set

$$\mathcal{D} := \{A \in \mathcal{A}_m : K \ni x \longmapsto P_x^m(A) \text{ Borel measurable }\}.$$

Then, by definition of P_x^m, $\mathcal{D}$ contains all thin cylinders of the form $_m[e_m, ..., e_n]$, $n \geq m$, which generate $\mathcal{A}_m$. Furthermore, obviously it holds true that

$$\Sigma \in \mathcal{D},$$

$$A \in \mathcal{D} \Rightarrow \Sigma \setminus A \in \mathcal{D}$$

and, for any pairwise disjoint family $(A_n)_{n \in \mathbb{N}} \subset \mathcal{D}$,

$$\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{D},$$

i.e. $\mathcal{D}$ is a Dynkin-system. Hence, $\mathcal{D}$ contains the Dynkin-system which is generated by the thin cylinders. Since the set of the thin cylinders is $\cap$-stable, it follows that $\mathcal{D} = \mathcal{A}_m$. $\qquad \qquad \square$

Definition 3.2.1. Let $\nu \in P(K)$. We call a probability measure $\Phi_m(\nu)$ on $(\Sigma, \mathcal{A}_m)$ given by

$$\Phi_m(\nu)(A) := \int P_x^m(A) d\nu(x), \ A \in \mathcal{A}_m,$$

the m-th lift of ν.

Definition 3.2.2. Set

$$\mathcal{C}(B) := \left\{ (A_m)_{m=0}^{-\infty} : A_m \in \mathcal{A}_m \ \forall m \text{ and } B \subset \bigcup_{m=0}^{-\infty} A_m \right\}$$

for $B \subset \Sigma$. Let $\nu \in P(K)$. We call a set function given by

$$\Phi(\nu)(B) := \inf \left\{ \sum_{m=0}^{-\infty} \Phi_m(\nu)(A_m) : (A_m)_{m \leq 0} \in \mathcal{C}(B) \right\}, B \subset \Sigma,$$

the lift of ν.

Lemma 3.2.2. *Let $\nu, \lambda \in P(K)$. Then*

(i) $\Phi(\nu)$ is an outer measure on Σ.

(ii) If $\Phi_m(\nu) \ll \Phi_m(\lambda)$ for all $m \leq 0$, then $\Phi(\nu) \ll \Phi(\lambda)$.

Proof. It is obvious that $\Phi(\nu)(\emptyset) = 0$.

Let $B_1 \subset B_2 \subset \Sigma$. Then $\mathcal{C}(B_1) \supset \mathcal{C}(B_2)$ and therefore

$$\Phi(\nu)(B_1) \leq \Phi(\nu)(B_2).$$

Now, we show

$$\Phi(\nu) \left(\bigcup_{i=1}^{\infty} B_i \right) \leq \sum_{i=1}^{\infty} \Phi(\nu)(B_i)$$

for all $B_i \subset \Sigma$, $i \in \mathbb{N}$. We can assume that the right hand side is finite. Let $\epsilon > 0$. Then for every $i \in \mathbb{N}$ there exists $(A_{im})_{m \leq 0} \in \mathcal{C}(B_i)$ such that

$$\Phi(\nu)(B_i) > \sum_{m=0}^{-\infty} \Phi_m(\nu)(A_{im}) - \epsilon 2^{-i}.$$

Since $\left(\bigcup_{i=1}^{\infty} A_{im}\right)_{m \le 0} \in \mathcal{C}\left(\bigcup_{i=1}^{\infty} B_i\right)$, it follows that

$$
\begin{aligned}
\Phi(\nu)\left(\bigcup_{i=1}^{\infty} B_i\right) &\le \sum_{m=0}^{-\infty} \Phi_m(\nu)\left(\bigcup_{i=1}^{\infty} A_{im}\right) \\
&\le \sum_{m=0}^{-\infty} \sum_{i=1}^{\infty} \Phi_m(\nu)\left(A_{im}\right) \\
&\le \sum_{i=1}^{\infty} \Phi(\nu)(B_i) + \epsilon.
\end{aligned}
$$

For (ii), let $\Phi(\lambda)(B) = 0$. Then for every $n \in \mathbb{N}$ there exists $(A_{nm})_{m \le 0} \in \mathcal{C}(B)$ such that

$$
\sum_{m=0}^{-\infty} \Phi_m(\lambda)(A_{nm}) < 2^{-n}.
$$

Set

$$
D_m := \bigcap_{k=1}^{\infty} \bigcup_{n \ge k} A_{nm} \text{ for each } m \le 0.
$$

Then, for each $m \le 0$,

$$
\Phi_m(\lambda)(D_m) \le \Phi_m(\lambda)\left(\bigcup_{n \ge k} A_{nm}\right) \le \sum_{n \ge k} 2^{-n}
$$

for all $k \ge 0$. Hence, $\Phi_m(\lambda)(D_m) = 0$ for all $m \le 0$. By the hypothesis, this implies that

$$
0 = \Phi_m(\nu)(D_m) = \lim_{k \to \infty} \Phi_m(\nu)\left(\bigcup_{n \ge k} A_{nm}\right) \ge \limsup_{k \to \infty} \Phi_m(\nu)\left(A_{km}\right)
$$

for all $m \le 0$. Hence

$$
\limsup_{k \to \infty} \sum_{m=0}^{-\infty} \Phi_m(\nu)\left(A_{km}\right) = 0.
$$

Thus, $\Phi(\nu)(B) = 0$. $\qquad\square$

We use further the following notation.

Notation 3.2.1. *Fix $x_i \in K_i$ for each $i \in \{1, ..., N\}$ and set*

$$P^m_{x_1...x_N} := \Phi_m \left(\frac{1}{N} \sum_{i=1}^{N} \delta_{x_i} \right) \text{ and } P_{x_1...x_N} := \Phi \left(\frac{1}{N} \sum_{i=1}^{N} \delta_{x_i} \right)$$

for every $m \in \mathbb{Z} \setminus \mathbb{N}$.

Now, for every $m \leq 0$ and $n \geq m$ define a random variable

$$Y^{x_1...x_N}_{mn} : \Sigma \longrightarrow K$$

$$\sigma \longmapsto w_{\sigma_n} \circ w_{\sigma_{n-1}} \circ ... \circ w_{\sigma_m}(x_{i(\sigma_m)})$$

with respect to the measure $P^m_{x_1...x_N}$.

Now, we are going to prove the main lemma which enables us to define the coding map. The proof of it involves a kind of Borel-Cantelli argument, which was also used by Barnsley and Elton in [3] (they considered the case $N = 1$ with constant probabilities). However, their key point, the reversion of the order of finite sequences of the maps, does not work here. That is why we first needed to construct the outer measure $P_{x_1...x_N}$.

Lemma 3.2.3. *Let $x_i, y_i \in K_i$ for each $1 \leq i \leq N$. Then*

(i)

$$\lim_{m \to -\infty} d\left(Y^{x_1...x_N}_{m0}, Y^{y_1...y_N}_{m0}\right) = 0 \ P_{x_1...x_N}\text{-}a.e.,$$

(ii)

$$\lim_{m \to -\infty} Y^{x_1...x_N}_{m0} \text{ exists } P_{x_1...x_N}\text{-}a.e.,$$

and by (i) $\lim_{m \to -\infty} Y^{x_1...x_N}_{m0} = \lim_{m \to -\infty} Y^{y_1...y_N}_{m0} \ P_{x_1...x_N}\text{-}a.e..$

Proof. Applying the average contractiveness condition $-m+1$ times gives

$$\sum_{e_m,...,e_0} \frac{1}{N} p_{e_m}(x_{i(e_m)})...p_{e_0}(w_{e_{-1}} \circ ... \circ w_{e_m}(x_{i(e_m)}))$$

$$\times \quad d(w_{e_0} \circ ... \circ w_{e_m}(x_{i(e_m)}), w_{e_0} \circ ... \circ w_{e_m}(y_{i(e_m)})) \leq a^{-m+1} \frac{1}{N} \sum_{i=1}^{N} d(x_i, y_i),$$

i.e.

$$\int d\left(Y_{m0}^{x_1...x_N}, Y_{m0}^{y_1...y_N}\right) dP_{x_1...x_N}^m \leq a^{-m+1} \frac{1}{N} \sum_{i=1}^{N} d(x_i, y_i).$$

So, by Markov inequality,

$$P_{x_1...x_N}^m \left(d\left(Y_{mn}^{x_1...x_N}, Y_{mn}^{y_1...y_N}\right) > a^{\frac{-m+1}{2}} \frac{1}{N} \sum_{i=1}^{N} d(x_i, y_i) \right) \leq a^{\frac{-m+1}{2}}.$$

Set $A_m := \left\{ \sigma \in \Sigma : \ d\left(Y_{m0}^{x_1...x_N}(\sigma), Y_{m0}^{y_1...y_N}(\sigma)\right) > a^{\frac{-m+1}{2}} \frac{1}{N} \sum_{i=1}^{N} d(x_i, y_i) \right\}$ and

$$A := \bigcap_{l=0}^{-\infty} \bigcup_{m=l}^{-\infty} A_m.$$

Then

$$P_{x_1...x_N}(A) \leq P_{x_1...x_N} \left(\bigcup_{m=l}^{-\infty} A_m \right) \leq \sum_{m=l}^{-\infty} P_{x_1...x_N}^m(A_m) \leq \sum_{m=l}^{-\infty} a^{\frac{-m+1}{2}},$$

since $(\emptyset, ..., \emptyset, A_l, A_{l-1}, ...) \in \mathcal{C}\left(\bigcup_{m=l}^{-\infty} A_m\right)$ for all $l \leq 0$. Hence $P_{x_1...x_N}(A) = 0$ and for every $\sigma \in \Sigma \setminus A$

$$d\left(Y_{m0}^{x_1...x_N}(\sigma), Y_{m0}^{y_1...y_N}(\sigma)\right) \leq a^{\frac{-m+1}{2}} \frac{1}{N} \sum_{i=1}^{N} d(x_i, y_i)$$

for all m except finitely many. This implies (i).

Now, for part (ii) set $C := \max_{e \in E} d(x_{t(e)}, w_e(x_{i(e)}))$. Then applying the average contractiveness condition $-m+1$ times reveals that

$$\int d\left(Y_{m0}^{x_1...x_N}, Y_{(m-1)0}^{x_1...x_N}\right) dP_{x_1...x_N}^m \ \leq \ a^{-m+1} \frac{1}{N} \sum_{e \in E} p_e(x_{i(e)}) d(x_{t(e)}, w_e(x_{i(e)}))$$

$$\leq \ a^{-m+1} C.$$

So, by Markov inequality,

$$P_{x_1...x_N}^m \left(d \left(Y_{m0}^{x_1...x_N}, Y_{(m-1)0}^{x_1...x_N} \right) > a^{\frac{-m+1}{2}} C \right) \le a^{\frac{-m+1}{2}}.$$

Set $B_m := \left\{ \sigma \in \Sigma : \ d \left(Y_{m0}^{x_1...x_N}(\sigma), Y_{(m-1)0}^{x_1...x_N}(\sigma) \right) > a^{\frac{-m+1}{2}} C \right\}$ and

$$B := \bigcap_{l=0}^{-\infty} \bigcup_{m=l}^{-\infty} B_m.$$

Then

$$P_{x_1...x_N}(B) \le P_{x_1...x_N} \left(\bigcup_{m=l}^{-\infty} B_m \right) \le \sum_{m=l}^{-\infty} P_{x_1...x_N}^m(B_m) \le \sum_{m=l}^{-\infty} a^{\frac{-m+1}{2}},$$

since $(\emptyset, ..., \emptyset, B_l, B_{l-1}, ...) \in \mathcal{C} \left(\bigcup_{m=l}^{-\infty} B_m \right)$ for all $l \le 0$. Hence $P_{x_1...x_N}(B) = 0$ and for every $\sigma \in \Sigma \setminus B$

$$\sum_{m=0}^{-\infty} d \left(Y_{m0}^{x_1...x_N}(\sigma), Y_{(m-1)0}^{x_1...x_N}(\sigma) \right) < \infty.$$

This implies that $\left(Y_{m0}^{x_1...x_N}(\sigma) \right)_{m \in \mathbb{Z} \setminus \mathbb{N}}$ is a Cauchy sequence for $P_{x_1...x_N}$-a.e. $\sigma \in \Sigma$, and so $\lim_{m \to -\infty} Y_{m0}^{x_1...x_N}$ exists $P_{x_1...x_N}$-a.e.. By part (i) $\lim_{m \to -\infty} Y_{m0}^{x_1...x_N} = \lim_{m \to -\infty} Y_{m0}^{y_1...y_N}$ $P_{x_1...x_N}$-a.e.. $\qquad \square$

Definition 3.2.3. We call

$$F_{x_1...x_N} : \Sigma \ \longrightarrow \ K$$

$$\sigma \ \longmapsto \ \lim_{m \to -\infty} Y_{m0}^{x_1...x_N}(\sigma)$$

the *coding map*. By Lemma 3.2.3, $F_{x_1...x_N}$ is defined $P_{x_1...x_N}$-a.e. and $F_{x_1...x_N} = F_{y_1...y_N}$ $P_{x_1...x_N}$-a.e..

3.3 Definition with respect to a generalized Markov measure

Our next aim is to show that the coding map is defined almost everywhere with respect to any outer measure $\Phi(\nu)$ if the restrictions of the probability functions on their vertex sets are Dini-continuous and bounded away from zero. For this, we only need to establish that $\Phi(\nu)$ is absolutely continuous with respect to $P_{x_1...x_N}$ in this case.

Lemma 3.3.1. *Let $x_i \in K_i$ for every $1 \le i \le N$ and $x \in K$. Let $i_0 \in \{1,...,N\}$ such that $x \in K_{i_0}$. Then for all integers $m \le 0$ and for all $\epsilon > 0$ there exist $k \ge m$ and $B \in \mathcal{A}_m$ such that $P_x^m(B) < \epsilon$ and*

$$n \ge k \;\Rightarrow\; d\left(Z_{mn}^x(\sigma), Y_{mn}^{x_1...x_N}(\sigma)\right) \le a^{\frac{n-m+1}{2}} d(x, x_{i_0})$$

for all $\sigma \in \Sigma \setminus B$.

Proof. Fix $m \le 0$. Applying the average contractiveness condition $n - m + 1$ times gives

$$\sum_{e_m,...,e_n} p_{e_m}(x)...p_{e_n}(w_{e_{n-1}} \circ ... \circ w_{e_m}(x)) d(w_{e_n} \circ ... \circ w_{e_m}(x), w_{e_n} \circ ... \circ w_{e_m}(x_{i(e_m)}))$$
$$\le\; a^{n-m+1} d\left(x, x_{i_0}\right),$$

i.e.

$$\int d\left(Z_{mn}^x, Y_{mn}^{x_1...x_N}\right) dP_x^m \le a^{n-m+1} d(x, x_{i_0}).$$

So, by Markov inequality,

$$P_x^m \left(d\left(Z_{mn}^x, Y_{mn}^{x_1...x_N}\right) > a^{\frac{n-m+1}{2}} d(x, x_{i_0}) \right) \le a^{\frac{n-m+1}{2}}.$$

Set $A_{mn} := \left\{ \sigma \in \Sigma : \ d\left(Z_{mn}^x(\sigma), Y_{mn}^{x_1 \ldots x_N}(\sigma)\right) > a^{\frac{n-m+1}{2}} d(x, x_{i_0}) \right\}$ for $m \leq n$ and $B_k :=$ $\bigcup_{n \geq k}^{\infty} A_{mn}$ for $k \geq m$. Then

$$\sum_{n=m}^{\infty} P_x^m(A_{mn}) \leq \sum_{l=1}^{\infty} a^{\frac{l}{2}} < \infty.$$

Therefore

$$P_x^m \left(\bigcap_{k=m}^{\infty} B_k \right) = 0.$$

Hence, for all $\epsilon > 0$ there exists $k \geq m$ such that $P_x^m(B_k) < \epsilon$ and

$$n \geq k \ \Rightarrow \ d\left(Z_{mn}^x(\sigma), Y_{mn}^{x_1 \ldots x_N}(\sigma)\right) \leq a^{\frac{n-m+1}{2}} d(x, x_{i_0})$$

for all $\sigma \in \Sigma \setminus B_k$. $\hspace{2cm}$ □

The next lemma is a generalization of Lemma 3 in [8].

Lemma 3.3.2. *Suppose that each probability function* $p_e|_{K_{i(e)}}$ *is Dini-continuous and there exists* $\delta > 0$ *such that* $p_e|_{K_{i(e)}} \geq \delta$ *for all* $e \in E$. *Let* $x_i \in K_i$ *for all* $1 \leq i \leq N$ *and* $x \in K$. *Then* P_x^m *is absolutely continuous with respect to* $P_{x_1 \ldots x_N}^m$ *for all* $m \leq 0$.

Proof. Fix $m \leq 0$. Let $A \in \mathcal{A}_m$ such that $P_{x_1 \ldots x_N}^m(A) = 0$ and $\epsilon > 0$. We show $P_x^m(A) < \epsilon$.

Let $i_0 \in \{1, \ldots, N\}$ such that $x \in K_{i_0}$. By Lemma 3.3.1, there exists $n_\epsilon \geq m$ and $B \in \mathcal{A}_m$ such that $P_x^m(B) < \epsilon/2$ and

$$n \geq n_\epsilon \ \Rightarrow \ d\left(Z_{mn}^x(\sigma), Y_{mn}^{x_1 \ldots x_N}(\sigma)\right) \leq a^{\frac{n-m+1}{2}} d(x, x_{i_0})$$

for all $\sigma \in \Sigma \setminus B$. Let ϕ_e be the modulus of uniform continuity of p_e on $K_{i(e)}$ for each $e \in E$ and $\phi := \max_{e \in E} \phi_e$. Since each p_e is Dini-continuous on $K_{i(e)}$, by Lemma

2.1.1, we can choose $l \geq n_\epsilon$ such that $\sum_{k=l+1}^{\infty} \phi\left(a^{(k-m)/2}d(x, x_{i_0})\right) < \delta/2$. Set

$$
Q_n := \begin{cases} \left\{\sigma \in \Sigma : d(Z_{mk}^x(\sigma), Y_{mk}^{x_1...x_N}(\sigma)) \leq a^{\frac{k-m+1}{2}} d(x, x_{i_0}) \; \forall l \leq k \leq n\right\}, \\ \text{if } l \leq n \\ \Sigma, \text{ else} \end{cases}
$$

and $Q := \bigcap_{n \geq m} Q_n$. Then $\Sigma \setminus B \subset Q$ and therefore $P_x(\Sigma \setminus Q) < \epsilon/2$. Now, for $\sigma \in Q_n$, if $l \leq n$ and $(\sigma_m, ..., \sigma_n)$ is a path of the digraph starting in i_0, then

$$
p_{\sigma_m}(x)...p_{\sigma_n}(w_{\sigma_{n-1}} \circ ... \circ w_{\sigma_m} x)
$$
$$
\leq \; p_{\sigma_m}(x_{i(\sigma_m)})...p_{\sigma_n}(w_{\sigma_{n-1}} \circ ... \circ w_{\sigma_m} x_{i(\sigma_m)}) \left(\frac{1-\delta}{\delta}\right)^{l-m+1}
$$
$$
\times \; \prod_{k=l+1}^{n} \left[1 + \frac{p_{\sigma_k}(w_{\sigma_{k-1}} \circ ... \circ w_{\sigma_m} x) - p_{\sigma_k}(w_{\sigma_{k-1}} \circ ... \circ w_{\sigma_m} x_{i(\sigma_m)})}{p_{\sigma_k}(w_{\sigma_{k-1}} \circ ... \circ w_{\sigma_m} x_{i(\sigma_m)})}\right]
$$
$$
\leq \; p_{\sigma_m}(x_{i(\sigma_m)})...p_{\sigma_n}(w_{\sigma_{n-1}} \circ ... \circ w_{\sigma_m} x_{i(\sigma_m)}) \left(\frac{1-\delta}{\delta}\right)^{l-m+1}
$$
$$
\times \; \prod_{k=l+1}^{n} \left[1 + \frac{\phi\left(a^{\frac{k-m}{2}} d(x, x_{i_0})\right)}{\delta}\right].
$$

Since $\prod_{k=l+1}^{n}\left[1 + \phi\left(a^{(k-m)/2}d(x, x_{i_0})\right)/\delta\right] \leq 1 + 2\sum_{k=l+1}^{\infty} \phi\left(a^{(k-m)/2}d(x, x_{i_0})\right)/\delta \leq 2$, it follows that

$$
p_{\sigma_m}(x)...p_{\sigma_n}(w_{\sigma_{n-1}} \circ ... \circ w_{\sigma_m} x) \leq 2N \left(\frac{1-\delta}{\delta}\right)^{l-m+1} \frac{1}{N} p_{\sigma_m}(x_{i(\sigma_m)})...p_{\sigma_n}(w_{\sigma_{n-1}} \circ ... \circ w_{\sigma_m} x_{i(\sigma_m)}).
$$

If $l > n$ or $(\sigma_m, ..., \sigma_n)$ is not a path of the digraph starting in i_0, then it holds trivially for any $\sigma \in \Sigma$.

Let $\triangle_m$ be the algebra every element of which is a finite union of thin cylinders of the form $_m[e_m, ..., e_n]$, $n \geq m$. By Caratheodory construction, there exists a sequence $(A_k)_{k \in \mathbb{N}} \subset \triangle_m$ such that $A \subset \bigcup_{k=1}^{\infty} A_k$ and

$$
\sum_{k=1}^{\infty} P_{x_1...x_N}^m(A_k) < \frac{\epsilon}{4N}\left(\frac{\delta}{1-\delta}\right)^{l-m+1}.
$$

We can write every finite union $\bigcup_{k=1}^{n} A_k$ as a disjoint union $\biguplus_{k=1}^{m_n} C_k$ of thin cylinders which generate $\triangle_m$. Let $C_k = {}_m[e_m, ..., e_n]$ with $m \leq n$. Then

$$P_x^m (Q \cap C_k)$$

$$\leq P_x^m (Q_n \cap C_k)$$

$$= \sum_{(\sigma_m,...,\sigma_n):\sigma \in Q_n \cap C_k} p_{\sigma_m}(x)...p_{\sigma_n}(w_{\sigma_{n-1}} \circ ... \circ w_{\sigma_m} x)$$

$$\leq 2N \left(\frac{1-\delta}{\delta}\right)^{l-m+1} \sum_{(\sigma_m,...,\sigma_n):\sigma \in Q_n \cap C_k} \frac{1}{N} p_{\sigma_m}(x_{i(\sigma_m)})...p_{\sigma_n}(w_{\sigma_{n-1}} \circ ... \circ w_{\sigma_m} x_{i(\sigma_m)})$$

$$\leq 2N \left(\frac{1-\delta}{\delta}\right)^{l-m+1} P_{x_1...x_n}^m (C_k).$$

Hence

$$\begin{aligned}
P_x^m(A) &= P_x^m(Q \cap A) + P_x^m(A \setminus Q) \\
&\leq \lim_{n\to\infty} P_x^m \left(\biguplus_{k=1}^{m_n} C_k \cap Q\right) + \frac{\epsilon}{2} \\
&= \lim_{n\to\infty} \sum_{k=1}^{m_n} P_x^m (C_k \cap Q) + \frac{\epsilon}{2} \\
&\leq 2N \lim_{n\to\infty} \left(\frac{1-\delta}{\delta}\right)^{l-m+1} \sum_{k=1}^{m_n} P_{x_1...x_n}^m (C_k) + \frac{\epsilon}{2} \\
&\leq 2N \left(\frac{1-\delta}{\delta}\right)^{l-m+1} \sum_{k=1}^{\infty} P_{x_1...x_n}^m (A_k) + \frac{\epsilon}{2} \\
&< \epsilon.
\end{aligned}$$

$\square$

Theorem 3.3.3. *Suppose that each probability function $p_e|_{K_{i(e)}}$ is Dini-continuous and there exists $\delta > 0$ such that $p_e|_{K_{i(e)}} \geq \delta$ for all $e \in E$. Let $x_i, y_i \in K_i$ for all $1 \leq i \leq N$ and $\nu \in P(K)$. Then the coding map $F_{x_1...x_N}$ is defined $\Phi(\nu)$-a.e. and $F_{x_1...x_N} = F_{y_1...y_N}$ $\Phi(\nu)$-a.e..*

Proof. By Lemma 3.3.2, $\Phi_m(\nu)$ is absolutely continuous with respect to $P^m_{x_1...x_N}$ for all $m \leq 0$. By Lemma 3.2.2 *(ii)*, this implies that $\Phi(\nu)$ is absolutely continuous with respect to $P_{x_1...x_N}$. The claim follows by Lemma 3.2.3. $\square$

Definition 3.3.1. We call

$$M := \Phi(\mu)$$

the *generalized Markov measure*, where μ an invariant Borel probability measure of the CMS. Denote the Borel σ-algebra on Σ by $\mathcal{B}(\Sigma)$.

Proposition 3.3.4. *M is a shift invariant Borel probability measure on Σ.*

Proof. First define a set function $\bar{M}$ on all thin cylinders of Σ by

$$\bar{M}(_m[e_m, ..., e_n]) := \Phi_m(\mu)(_m[e_m, ..., e_n]).$$

We show that $\bar{M}$ extends uniquely to a shift invariant Borel probability measure on Σ and

$$\bar{M}|_{\mathcal{A}_m} = \Phi_m(\mu) \text{ for all } m \leq 0.$$

We only need to check that

$$\bar{M}(_m[e_m, ..., e_n]) = \sum_{e_{n+1}} \bar{M}(_m[e_m, ..., e_n, e_{n+1}])$$

and that

$$\bar{M}(_m[e_m, ..., e_n]) = \sum_{e_{m-1}} \bar{M}(_{m-1}[e_{m-1}, e_m, ..., e_n]),$$

the rest follows by the standard extension argument. The first equation is obvious by

the definition of P_x. For the second we need the invariance of μ.

$$
\begin{aligned}
&\sum_{e_{m-1}} \bar{M}\left(_{m-1}[e_{m-1}, e_m, ..., e_n]\right) \\
={}& \sum_{e_{m-1}} \int p_{e_{m-1}}(x) p_{e_m}(w_{e_{m-1}}(x))...p_{e_n}(w_{e_{n-1}} \circ ... \circ w_{e_{m-1}}(x)) d\mu(x) \\
={}& U^* \mu(p_{e_m}...p_{e_n} \circ w_{e_{n-1}} \circ ... \circ w_{e_m}) \\
={}& \mu(p_{e_m}...p_{e_n} \circ w_{e_{n-1}} \circ ... \circ w_{e_m}) \\
={}& \bar{M}\left(_m[e_m, ..., e_n]\right).
\end{aligned}
$$

Now, we show that

$$
\bar{M} = M|_{\mathcal{B}(\Sigma)}.
$$

Let $B \in \mathcal{B}(\Sigma)$ and $\epsilon > 0$. Since $\mathcal{B}(\Sigma)$ is the smallest σ-algebra containing all $\mathcal{A}_m$, $m \leq 0$, it follows that for every $(A_m)_{m \leq 0} \in \mathcal{C}(B)$

$$
\bar{M}(B) \leq \bar{M}\left(\bigcup_{m \leq 0} A_m\right) \leq \sum_{m \leq 0} \bar{M}(A_m) = \sum_{m \leq 0} \Phi_m(\mu)(A_m).
$$

Hence $\bar{M}(B) \leq M(B)$. On the other hand, let $\triangle$ be the algebra generated by all thin cylinders in Σ. Then every $A \in \triangle$ is also an element of some $\mathcal{A}_m$, $m \leq 0$, and therefore $\bar{M}(A) \geq M(A)$. By Caratheodory's construction, there exists a sequence $(A_k)_{k \in \mathbb{N}} \subset \triangle$ such that $B \subset \bigcup_{k \in \mathbb{N}} A_k$ and

$$
\bar{M}(B) > \sum_{k \in \mathbb{N}} \bar{M}(A_k) - \epsilon.
$$

Therefore

$$
\bar{M}(B) \geq \sum_{k \in \mathbb{N}} M(A_k) - \epsilon \geq M\left(\bigcup_{k \in \mathbb{N}} A_k\right) - \epsilon \geq M(B) - \epsilon.
$$

Hence, $\bar{M}(B) = M(B)$. $\qquad\square$

Definition 3.3.2. We call the measure preserving transformation S of the probability space $(\Sigma, \mathcal{B}(\Sigma), M)$ a *generalized Markov shift.*

Now, we state explicitly the most important special case of Theorem 3.3.3.

Corollary 3.3.5. *Suppose that each probability function $p_e|_{K_{i(e)}}$ is Dini-continuous and there exists $\delta > 0$ such that $p_e|_{K_{i(e)}} \geq \delta$ for all $e \in E$. Let $x_i, y_i \in K_i$ for all $1 \leq i \leq N$. Then the coding map $F_{x_1 \ldots x_N}$ is defined M-a.e. and $F_{x_1 \ldots x_N} = F_{y_1 \ldots y_N}$ M-a.e..*

Notation 3.3.1. *By Corollary 3.3.5, it would not be too heavy an abuse of notation if we denote the coding map on the measure space $(\Sigma, \mathcal{B}(\Sigma), M)$ simply by F.*

Chapter 4

Applications of the coding map

Let $\left(K_{i(e)}, w_e, p_e\right)_{e \in E}$ be a contractive Markov system with the average contracting rate $0 < a < 1$ and an invariant Borel probability measure μ. We assume that: (K, d) is a metric space in which sets of finite diameter are relatively compact and the family $K_1, ..., K_N$ partitions K into non-empty open subsets; each probability function $p_e|_{K_{i(e)}}$ is Dini-continuous and bounded away from zero by $\delta > 0$. Note that the assumption on the metric space implies that it is locally compact, separable and complete.

4.1 Main Lemma for the generalized Markov shift

We prove here what seems to be the main lemma for the generalized Markov shift associated with a contractive Markov system using the coding map constructed in the previous chapter. This lemma establishes a relation between the Markovian picture of a CMS studied in Chapter 2 and the dynamical picture given by the generalized Markov shift (Chapter 3).

62

For the proof, we need to define some measures on the product space $K \times \Sigma$.

Denote by $\mathcal{A}$ the finite σ-algebra generated by the partition $\{_0[e] : e \in E\}$ of Σ and define, for each integer $m \leq 1$,

$$\mathcal{A}_m := \bigvee_{i=m}^{+\infty} S^{-i}\mathcal{A},$$

which is the smallest σ-algebra containing all finite σ-algebras $\bigvee_{i=m}^{n} S^{-i}\mathcal{A}$, $n \geq m$. Let $x \in K$. For every integer $m \leq 1$, let P_x^m be a probability measure on σ-algebra $\mathcal{A}_m$ given by

$$P_x^m(_m[e_m, ..., e_n]) = p_{e_m}(x)p_{e_{m+1}}(w_{e_m}(x))...p_{e_n}(w_{e_{n-1}} \circ ... \circ w_{e_m}(x))$$

for all thin cylinders $_m[e_m, ..., e_n]$, $n \geq m$. By Lemma 3.2.1, $x \longmapsto P_x^m(Q)$ is a Borel measurable function on K. Therefore, we can define, for every integer $m \leq 0$,

$$\tilde{M}_m(A \times Q) := \int_A P_x^m(Q)\, d\mu(x)$$

for all $A \in \mathcal{B}(K)$ and all $Q \in \mathcal{A}_m$. Then $\tilde{M}_m$ extends uniquely to a probability measure on the product σ-algebra $\mathcal{B}(K) \otimes \mathcal{A}_m$ with

$$\tilde{M}_m(\Omega) = \int P_x^m(\{\sigma \in \Sigma : (x, \sigma) \in \Omega\})\, d\mu(x)$$

for all $\Omega \in \mathcal{B}(K) \otimes \mathcal{A}_m$. Note that the set of all $\Omega \in \mathcal{B}(K) \otimes \mathcal{A}_m$ for which the integrand in the above is measurable forms a Dynkin system which contains the set all rectangles $A \times Q$, $A \in \mathcal{B}(K)$, $Q \in \mathcal{A}_m$. As the latter is $\cap$-stable and generates $\mathcal{B}(K) \otimes \mathcal{A}_m$, the integrand is measurable for all $\Omega \in \mathcal{B}(K) \otimes \mathcal{A}_m$. Further, note that $P_x^m(\{\sigma \in \Sigma : (x, \sigma) \in \Omega\}) = \int 1_\Omega(x, \sigma) dP_x^m(\sigma)$ for all $\Omega \in \mathcal{B}(K) \otimes \mathcal{A}_m$. Therefore

$$\int s\, d\tilde{M}_m = \int \int s(x, \sigma) dP_x^m(\sigma) d\mu(x)$$

for all $\mathcal{B}(K) \otimes \mathcal{A}_m$-simple functions s. Now, let ψ be a $\mathcal{B}(K) \otimes \mathcal{A}_m$-measurable and $\tilde{M}_m$-integrable function on $K \times \Sigma$. Then the usual monotone approximation of positive and negative parts of ψ by simple functions and the B. Levi Theorem imply that

$$\int \psi d\tilde{M}_m = \int \int \psi(x, \sigma) dP_x^m(\sigma) d\mu(x).$$

Lemma 4.1.1. *Suppose* $C := \sum_{i=1}^{N} \int_{K_i} d(x, x_i) d\mu(x) < \infty$ *for some* $x_i \in K_i$, $i = 1, ..., N$. *Let* $_1[e_1, ..., e_n] \subset \Sigma$ *be a thin cylinder set. Let* $\mathcal{F} := \bigvee_{i=0}^{\infty} S^i \mathcal{A}$. *Then*

$$E_M \left(1_{1[e_1,...,e_n]} | \mathcal{F} \right) (\sigma) = P_{F(\sigma)}^1 \left(_1[e_1, ..., e_n] \right) \text{ for } M\text{-a.e. } \sigma \in \Sigma,$$

where $E_M(.|.)$ *denotes conditional expectation with respect to measure* M.

Proof. We can obviously assume that $(e_1, ..., e_n)$ is a path of the directed graph. Set $\mathcal{F}_m := \bigvee_{i=0}^{m} S^i \mathcal{A}$ for all $m \in \mathbb{Z} \backslash \mathbb{N}$. We denote further a $(-m+1)$-tuple by $(\sigma_m, ..., \sigma_0)^*$ if $M \left(_m[\sigma_m, ..., \sigma_0] \right) > 0$ (i.e. $(\sigma_m, ..., \sigma_0)$ is a path). Then obviously

$$E_M \left(1_{1[e_1,...,e_n]} | \mathcal{F}_m \right) (\tilde{\sigma}) = \sum_{(\sigma_m,...,\sigma_0)^*} \frac{\int_{m[\sigma_m,...,\sigma_0]} 1_{1[e_1,...,e_n]} dM}{M \left(_m[\sigma_m, ..., \sigma_0] \right)} 1_{m[\sigma_m,...,\sigma_0]}(\tilde{\sigma})$$

for M-a.e. $\tilde{\sigma} \in \Sigma$. Since $(\mathcal{F}_m)_{m \leq 0}$ is an increasing sequence of σ-algebras and $\mathcal{F}$ is the smallest σ-algebra containing all $\mathcal{F}_m$, it follows by Doob's Martingale Theorem (e.g. see [7] p. 199) that

$$E_M \left(1_{1[e_1,...,e_n]} | \mathcal{F}_m \right) \to E_M \left(1_{1[e_1,...,e_n]} | \mathcal{F} \right) \tag{4.1.1}$$

M-a.e..

Now, set

$$Z_m^x(\sigma) := w_{\sigma_0} \circ ... \circ w_{\sigma_m} x \text{ and } Y_m(\sigma) := w_{\sigma_0} \circ ... \circ w_{\sigma_m} x_{i(\sigma_m)}$$

for all $x \in K$, $\sigma \in \Sigma$ and $m \leq 0$. Further, define

$$X_m(\tilde{\sigma}) := \sum_{(\sigma_m, \ldots, \sigma_0)^*} \frac{\displaystyle\int_{K \times_m [\sigma_m, \ldots, \sigma_0]} d\left(Z_m^x(\bar{\sigma}), Y_m(\bar{\sigma})\right) d\tilde{M}_m(x, \bar{\sigma})}{\tilde{M}_m\left(K \times_m [\sigma_m, \ldots, \sigma_0]\right)} 1_{m[\sigma_m, \ldots, \sigma_0]}(\tilde{\sigma})$$

for all $\tilde{\sigma} \in \Sigma$. Then

$$\int X_m dM = \int d\left(Z_m^x(\sigma), Y_m(\sigma)\right) d\tilde{M}_m(x, \sigma)$$

$$= \int \sum_{\sigma_m, \ldots, \sigma_0} p_{\sigma_m}(x) \ldots p_{\sigma_0}(w_{\sigma_{-1}} \ldots w_{\sigma_m} x) d(w_{\sigma_0} \ldots w_{\sigma_m} x, w_{\sigma_0} \ldots w_{\sigma_m} x_{i(\sigma_m)}) d\mu(x)$$

$$\leq a^{-m+1} \sum_{i=1}^{N} \int_{K_i} d(x, x_i) d\mu(x) = a^{-m+1} C.$$

Set

$$\Omega_m := \left\{ \sigma \in \Sigma : X_m > a^{\frac{-m+1}{2}} C \right\}$$

for all $m \leq 0$ and $\Omega := \bigcap_{n \leq 0} \bigcup_{m \leq n} \Omega_m$. Then, by Markov inequality,

$$M(\Omega_m) \leq a^{\frac{-m+1}{2}}.$$

Hence, $M(\Omega) = 0$ and

$$X_m(\sigma) \to 0 \text{ for all } \sigma \in \Sigma \setminus \Omega.$$

Now, for $\sigma \in \Sigma$ with $M\left(_m[\sigma_m, \ldots, \sigma_0]\right) > 0$,

$$\left| \frac{\displaystyle\int_{m[\sigma_m, \ldots, \sigma_0]} 1_{1[e_1, \ldots, e_n]} dM}{M\left(_m[\sigma_m, \ldots, \sigma_0]\right)} - P_{F(\sigma)}^1 \left(_1[e_1, \ldots, e_n]\right) \right|$$

$$= \left| \frac{\int p_{\sigma_m}(x) \ldots p_{\sigma_0}(w_{\sigma_{-1}} \ldots w_{\sigma_m} x) p_{e_1}(Z_m^x(\sigma)) \ldots p_{e_n}(w_{e_{n-1}} \ldots w_{e_1} Z_m^x(\sigma)) d\mu(x)}{\int p_{\sigma_m}(x) \ldots p_{\sigma_0}(w_{\sigma_{-1}} \ldots w_{\sigma_m} x) d\mu(x)} \right.$$

$$\left. - p_{e_1}(F(\sigma)) \ldots p_{e_n}(w_{e_{n-1}} \ldots w_{e_1} F(\sigma)) \right|. \tag{4.1.2}$$

Set $p(x) := p_{e_1}(x) \ldots p_{e_n}(w_{e_{n-1}} \ldots w_{e_1} x)$, $x \in K$. Note that the average contractiveness condition and the boundedness away from zero of the probability functions (on their

vertex sets) imply that each map $w_e|_{K_{i(e)}}$ is Lipschitz. Since each $p_e|_{K_{i(e)}}$ is Dini-continuous, it follows that each function $p_{e_k} \circ w_{e_{k-1}} \circ \ldots \circ w_{e_1}|_{K_{i(e_1)}}$, $1 \le k \le n$, is Dini-continuous. As bounded Dini-continuous functions form an algebra, $p|_{K_{i(e_1)}}$ is also Dini-continuous. Let η be the modulus of uniform continuity of $p|_{K_{i(e_1)}}$. By the Sublemma from [2], there exists $\psi : [0,\infty) \longrightarrow [0,\infty)$ such that $\psi(t) \ge \eta(t)$ for all t, $\psi(t)/t$ is non-increasing, and $\int_0^1 \psi(t)/t\, dt < \infty$. Set

$$\beta(u) := \frac{1}{1-a} \int_0^{ua^{-1}} \frac{\psi(t)}{t} dt \text{ for } u \ge 0.$$

Then β is continuous, concave and $\beta(0) = 0$. Moreover,

$$\beta(u) \ge \frac{1}{1-a} \int_u^{ua^{-1}} \frac{\psi(t)}{t} dt \ge \frac{1}{1-a} \frac{\psi(u)}{ua^{-1}} u(a^{-1} - 1) = \psi(u) \ge \eta(u)$$

for all $u > 0$. Hence, $\eta(u) \le \beta(u)$ for all $u \ge 0$. Therefore,

$$
\begin{aligned}
(4.1.2) \quad &\le \left| \frac{\int_{K \times\, _m[\sigma_m,\ldots,\sigma_0]} p \circ Z_m^x(\bar\sigma) d\tilde{M}_m(x,\bar\sigma)}{\tilde{M}_m\left(K \times\, _m[\sigma_m,\ldots,\sigma_0]\right)} - p \circ Y_m(\sigma) \right| + |p \circ Y_m(\sigma) - p \circ F(\sigma)| \\
&\le \frac{\int_{K \times\, _m[\sigma_m,\ldots,\sigma_0]} |p \circ Z_m^x(\bar\sigma) - p \circ Y_m(\bar\sigma)|\, d\tilde{M}_m(x,\bar\sigma)}{\tilde{M}_m\left(K \times\, _m[\sigma_m,\ldots,\sigma_0]\right)} + |p \circ Y_m(\sigma) - p \circ F(\sigma)| \\
&\le \frac{\int_{K \times\, _m[\sigma_m,\ldots,\sigma_0]} \beta\left(d\left(Z_m^x(\bar\sigma), Y_m(\bar\sigma)\right)\right) d\tilde{M}_m(x,\bar\sigma)}{\tilde{M}_m\left(K \times\, _m[\sigma_m,\ldots,\sigma_0]\right)} + |p \circ Y_m(\sigma) - p \circ F(\sigma)| \\
&\le \beta\left(\frac{\int_{K \times\, _m[\sigma_m,\ldots,\sigma_0]} d\left(Z_m^x(\bar\sigma), Y_m(\bar\sigma)\right) d\tilde{M}_m(x,\bar\sigma)}{\tilde{M}_m\left(K \times\, _m[\sigma_m,\ldots,\sigma_0]\right)} \right) + |p \circ Y_m(\sigma) - p \circ F(\sigma)| \\
&= \beta \circ X_m(\sigma) + |p \circ Y_m(\sigma) - p \circ F(\sigma)|.
\end{aligned}
$$

Hence

$$\left| E_M\left(1_{1[e_1,\ldots,e_n]}|\mathcal{F}_m\right)(\sigma) - P_{F(\sigma)}^1\left(1[e_1,\ldots,e_n]\right) \right| \le \beta \circ X_m(\sigma) + |p \circ Y_m(\sigma) - p \circ F(\sigma)|$$

for M-a.e. $\sigma \in \Sigma$. By Corollary 3.3.5 and the continuity of p on K, the second term also converges to zero M-a.e.. Thus

$$\left| E_M \left(1_{1[e_1,...,e_n]} | \mathcal{F}_m \right)(\sigma) - P^1_{F(\sigma)} \left(1[e_1,...,e_n] \right) \right| \to 0 \text{ as } m \to -\infty$$

for M-a.e. $\sigma \in \Sigma$. With (4.1.1), this implies that

$$E_M \left(1_{1[e_1,...,e_n]} | \mathcal{F} \right)(\sigma) = P^1_{F(\sigma)} \left(1[e_1,...,e_n] \right)$$

for M-a.e. $\sigma \in \Sigma$. □

4.2 What is the image of the generalized Markov measure under the coding map?

The next proposition is the most important application of the main Lemma for the generalized Markov shift.

Proposition 4.2.1. *If invariant probability measure μ is unique, then*

$$F(M) = \mu.$$

Proof. Let U^* be the adjoint of the Markov operator associated with the CMS. It is sufficient to show that $U^*F(M) = F(M)$, since μ is the unique invariant probability measure. Let $f \in C_B(K)$. Then

$$U^*F(M)(f) = \sum_{e \in E} \int p_e f \circ w_e dF(M) = \sum_{e \in E} \int p_e \circ F f \circ w_e \circ F dM.$$

Let $e \in E$. By Theorem 2.2.1 (iv), $\sum_{i=1}^{N} \int_{K_i} d(x, x_i) d\mu(x) < \infty$. Therefore, by Lemma 4.1.1,

$$E_M \left(1_{1[e]} \big| \mathcal{F} \right) (\sigma) = P^1_{F(\sigma)} \left(1[e] \right) = p_e \circ F(\sigma)$$

for M-a.e. $\sigma \in \Sigma$. Since $f \circ w_e \circ F$ is bounded and $\mathcal{F}$-measurable, it follows by a well known property of the conditional expectation that

$$E_M \left(1_{1[e]} f \circ w_e \circ F \big| \mathcal{F} \right) (\sigma) = p_e \circ F(\sigma) f \circ w_e \circ F(\sigma)$$

for M-a.e. $\sigma \in \Sigma$. Hence, by the shift invariance of M,

$$
\begin{aligned}
U^* F(M)(f) &= \sum_{e \in E} \int 1_{1[e]}(\sigma) f \circ w_e \circ F(\sigma) dM(\sigma) \\
&= \sum_{e \in E} \int 1_{1[e]} \circ S^{-1}(\sigma) f \circ w_e \circ F \circ S^{-1}(\sigma) dM(\sigma) \\
&= \sum_{e \in E} \int 1_{0[e]}(\sigma) f \circ F(\sigma) dM(\sigma) \\
&= F(M)(f).
\end{aligned}
$$

$\square$

4.3 Kolmogorov-Sinai entropy of the generalized Markov shift

In this section we give a further application of the main Lemma for the generalized Markov shift. It turns out that the coding map is the key tool for the calculation of the Kolmogorov-Sinai entropy of the generalized Markov shift.

Theorem 4.3.1. *Let $h_M(S)$ be Kolmogorov-Sinai entropy of the generalized Markov shift associated with the contractive Markov system.*

(i) If $\sum_{i=1}^{N} \int_{K_i} d(x, x_i) d\mu(x) < \infty$ for some $x_i \in K_i$, $i = 1, ..., N$, then

$$h_M(S) = -\sum_{e \in E} \int_{K_{i(e)}} p_e \log p_e \, dF(M).$$

(ii) If the invariant probability measure μ is unique, then

$$h_M(S) = -\sum_{e \in E} \int_{K_{i(e)}} p_e \log p_e \, d\mu.$$

Proof. It is well known that $h_M(S) = h_M(S^{-1})$ (e.g. Theorem 4.13 in [25]) and, by the Kolmogorov-Sinai Theorem (e.g. Theorem 4.17 in [25]), $h_M(S^{-1}) = h_M(S^{-1}, \mathcal{A})$. Further, using the notion of conditional entropy (e.g. Theorem 4.3 (ix) and Theorem 4.14 in [25]),

$$h_M(S^{-1}, \mathcal{A}) = \mathcal{H}\left(S^{-1}\mathcal{A} \Big/ \bigvee_{i=1}^{\infty} S^{i-1}\mathcal{A} \right).$$

Set $\mathcal{F} := \bigvee_{i=0}^{\infty} S^i \mathcal{A}$. Hence

$$h_M(S) = -\sum_{e \in E} \int E\left(1_{1[e]} \big| \mathcal{F} \right) \log E\left(1_{1[e]} \big| \mathcal{F} \right) dM.$$

By the assumption in (i), Lemma 4.1.1 implies that

$$E\left(1_{1[e]} \big| \mathcal{F} \right) = p_e \circ F \ M\text{-a.e.}$$

for each $e \in E$. Hence, with $0 \log 0 = 0$, we have

$$h_M(S) = -\sum_{e \in E} \int p_e \circ F \log \left(p_e \circ F \right) dM = -\sum_{e \in E} \int_{K_{i(e)}} p_e \log p_e \, dF(M).$$

By the assumption in (ii), Theorem 2.2.1 (iv) and Proposition 4.2.1 imply that

$$h_M(S) = -\sum_{e \in E} \int_{K_{i(e)}} p_e \log p_e \, d\mu.$$

$\square$

Chapter 5

Empiricalness of the invariant measure of a CMS

In this chapter we prove an ergodic theorem for Markov chains associated with contractive Markov systems using the coding map. It is a generalization of the ergodic theorem from [8].

5.1 Introduction

It is very important for various applications (such as image compression) to know whether the sequence x, $w_{\sigma_1}x$, $w_{\sigma_2} \circ w_{\sigma_1}x, \dots$ "draws a picture" of the invariant measure of the CMS for any $x \in K$ and P_x-a.e. $\sigma \in \Sigma^+$, i.e. whether

$$\mu(B) = \lim_{n \to \infty} \frac{\sharp\{i : w_{\sigma_i} \circ \dots \circ w_{\sigma_1}x \in B,\ 1 \leq i \leq n\}}{n} \text{ for all open } B \subset K,$$

where $\sharp$ counts the number of element in the set. In the affirmative case, such a statement is called a *strong law of large numbers* or an *ergodic theorem*.

71

Remark 5.1.1. If the metric space K is compact, then such a strong law of large numbers holds true for any Markov operator with the Feller property which possesses a unique invariant probability measure. This was proved by Breiman in 1960 (see [6]). In 1987, Elton gave a proof of an ergodic theorem for IFS with place-dependent probabilities on a metric space in which sets of finite diameter are relatively compact. His proof is probabilistic in nature; it uses an argument from the general theory of Markov processes (see Lemma 1 in [8]). It is possible to generalize his proof in order to cover CMS as well.

Also, two other papers with similar titles should probably be mentioned here. However, the systems studied there overlap with ours only in the trivial case where $N = 1$ and the probability functions are constant. A strong law of large number proved by S. Grigorescu [9] allows an arbitrary index set E with a fixed probability measure on it such that the system is contractive on average. Both Elton's and Grigorescu's assumptions imply that the Markov operator has an attractive probability measure. In this case, the generalized Markov shift is strongly mixing (see Proposition 5.2.4) which is an unnecessarily strong assumption for the ergodic theorem. Ö. Stenflo [20] proved an ergodic theorem for Markov processes generated by a finite family of Lipschitz maps on a complete metric space which is driven randomly by an i.i.d. process and satisfies the average contraction condition. This generalizes the trivial setup where the driving process is Bernoulli. He suggests a use of such a setup (with possibly uncountably many maps) for a representation of an IFS with place dependent probabilities. (The reader must be aware that the driving process in the latter case is one with infinite memory.)

The proof of the ergodic theorem for CMS which we present here is completely analytic and it also uses a representation of the Markov process, but as a factor of the generalized Markov shift associated with the CMS via the coding map. The generalized Markov measure on the shift space Σ makes the sequence

$$(Y_0, Y_1, ...) : \quad \Sigma \quad \longrightarrow K^{\mathbb{N}}$$

$$\sigma \quad \longmapsto \left(F(\sigma), F(S\sigma), F(S^2\sigma), ... \right)$$

to a stationary stochastic process which is equivalent to the Markov process generated by the CMS with the stationary initial distribution μ. We exploit it in Theorem 5.2.5. Case $N = 1$ of it is exactly Elton's Ergodic Theorem (see Remark 2.1.1).

5.2   Ergodic theorem for contractive Markov chains

Let $\left(K_{i(e)}, w_e, p_e \right)_{e \in E}$ be a contractive Markov system with the average contracting rate $0 < a < 1$ and an invariant Borel probability measure μ. We assume that: (K, d) is a metric space in which sets of finite diameter are relatively compact and the family $K_1, ..., K_N$ partitions K into non-empty open subsets; each probability function $p_e|_{K_{i(e)}}$ is Dini-continuous and bounded away from zero by $\delta > 0$; the set of edges E is finite and the map $i : E \longrightarrow V$ is surjective. Note that the assumption on the metric space implies that it is locally compact separable and complete.

Before we move to the ergodic theorem, we need to clear up some technical details. We will use the abbreviation $w_{\sigma^k} := w_{\sigma_k} \circ ... \circ w_{\sigma_1}$ for all $\sigma \in \Sigma$ or $\sigma \in \Sigma^+$. Let P_x be the Borel probability measure on Σ^+ as in Definition 1.1.6 for $x \in K$. Let $\nu \in P(K)$.

Since $x \longmapsto P_x(Q)$ is Borel measurable for all $Q \in \mathcal{B}(\Sigma^+)$ (Lemma 3.2.1), we can define

$$\tilde{\phi}(\nu)(A \times Q) := \int_A P_x(Q) d\nu(x)$$

for all $A \in \mathcal{B}(K)$ and $Q \in \mathcal{B}(\Sigma^+)$. Then $\tilde{\phi}(\nu)$ extends uniquely to a Borel probability measure on $K \times \Sigma^+$ with

$$\tilde{\phi}(\nu)(\Omega) = \int P_x \left(\{ \sigma \in \Sigma^+ : (x, \sigma) \in \Omega \} \right) d\nu(x)$$

for all $\Omega \in \mathcal{B}(K \times \Sigma^+)$. Note that the set of all $\Omega \subset K \times \Sigma^+$ for which the integrand in the above is measurable forms a Dynkin system which contains all rectangles. Therefore, it is measurable for all $\Omega \in \mathcal{B}(K \times \Sigma^+)$.

Now, consider the following maps:

$$\gamma : \Sigma \longrightarrow \Sigma^+$$
$$\sigma \longmapsto (\sigma_1, \sigma_2, ...)$$

and

$$\xi : \Sigma \longrightarrow K \times \Sigma^+$$
$$\sigma \longmapsto (F(\sigma), \gamma(\sigma)).$$

Lemma 5.2.1. *Suppose the invariant probability measure μ is unique. Then*

$$\xi(M) = \tilde{\phi}(\mu).$$

Proof. We only need to check that

$$\xi(M) \left(A \times_1 [e_1, ..., e_n]^+ \right) = \tilde{\phi}(\mu) \left(A \times_1 [e_1, ..., e_n]^+ \right)$$

for all thin cylinders $_1[e_1, ..., e_n]^+$ and $A \in \mathcal{B}(K)$. For such sets,

$$\xi(M)\left(A \times {_1}[e_1, ..., e_n]^+\right) = M\left(F^{-1}(A) \cap \gamma^{-1}\left({_1}[e_1, ..., e_n]^+\right)\right)$$
$$= \int_{F^{-1}(A)} 1_{{_1}[e_1,...,e_n]} dM.$$

Let $\mathcal{F}$ be the sub-σ-algebra of $\mathcal{B}(\Sigma)$ generated by thin cylinders of the form $_m[\sigma_m, ..., \sigma_0]$, $m \in \mathbb{Z} \setminus \mathbb{N}$. Then $F^{-1}(A) \in \mathcal{F}$. By Theorem 2.2.1 (iv), the uniqueness of the invariant probability measure implies that $\sum_{i=1}^{N} \int_{K_i} d(x, x_i) d\mu(x) < \infty$. Hence, by Lemma 4.1.1,

$$E_M\left(1_{{_1}[e_1,...,e_n]} \big| \mathcal{F}\right) = P_{F(\sigma)}\left({_1}[e_1, ..., e_n]^+\right).$$

Therefore,

$$\int_{F^{-1}(A)} 1_{{_1}[e_1,...,e_n]} dM = \int_{F^{-1}(A)} P_{F(\sigma)}\left({_1}[e_1, ..., e_n]^+\right) dM(\sigma)$$
$$= \int_A P_x\left({_1}[e_1, ..., e_n]^+\right) dF(M)(x).$$

By Proposition 4.2.1, uniqueness of the invariant probability measure implies that $F(M) = \mu$. Thus,

$$\xi(M)\left(A \times {_1}[e_1, ..., e_n]^+\right) = \int_A P_x\left({_1}[e_1, ..., e_n]^+\right) d\mu(x)$$

as desired. $\qquad\square$

Lemma 5.2.2. *Let $x, y \in K_i$ for some $1 \leq i \leq N$.*

(i) For $\epsilon > 0$ there exist $n \in \mathbb{N}$ and $B_{xy} \in \mathcal{B}(\Sigma^+)$ such that $P_x(B_{xy}) < \epsilon$ and

$$k \geq n \Rightarrow d\left(w_{\sigma_k} \circ ... \circ w_{\sigma_1}(x), w_{\sigma_k} \circ ... \circ w_{\sigma_1}(y)\right) \leq a^{\frac{k}{2}} d(x, y)$$

for all $\sigma \in \Sigma^+ \setminus B_{xy}$.

(ii) There exists $H_{xy} \in \mathcal{B}(\Sigma^+)$ such that $P_x(H_{xy}) = 1$ and

$$\lim_{k \to \infty} d\left(w_{\sigma_k} \circ \ldots \circ w_{\sigma_1}(x), w_{\sigma_k} \circ \ldots \circ w_{\sigma_1}(y)\right) = 0$$

for all $\sigma \in H_{xy}$.

Proof. Applying the average contractiveness condition k times gives

$$\sum_{e_1,\ldots,e_k} p_{e_1}(x)\ldots p_{e_k}(w_{e_{k-1}} \circ \ldots \circ w_{e_1}(x)) d(w_{e_k} \circ \ldots \circ w_{e_1}(x), w_{e_k} \circ \ldots \circ w_{e_1}(y))$$
$$\leq \quad a^k d(x,y),$$

i.e.

$$\int d(w_{\sigma_k} \circ \ldots \circ w_{\sigma_1}(x), w_{\sigma_k} \circ \ldots \circ w_{\sigma_1}(y)) dP_x(\sigma) \leq a^k d(x,y).$$

So, by Markov inequality,

$$P_x\left(d(w_{\sigma_k} \circ \ldots \circ w_{\sigma_1}(x), w_{\sigma_k} \circ \ldots \circ w_{\sigma_1}(y)) > a^{\frac{k}{2}} d(x,y)\right) \leq a^{\frac{k}{2}}.$$

Set $A_k^{xy} := \left\{\sigma \in \Sigma^+ : d(w_{\sigma_k} \circ \ldots \circ w_{\sigma_1}(x), w_{\sigma_k} \circ \ldots \circ w_{\sigma_1}(y)) > a^{\frac{k}{2}} d(x,y)\right\}$ for all $k \in \mathbb{N}$, $B_n^{xy} := \bigcup_{k \geq n}^{\infty} A_k^{xy}$ for all $n \in \mathbb{N}$ and $B^{xy} := \bigcap_{n \geq 1} B_n^{xy}$. Then

$$P_x(B^{xy}) \leq P_x(B_n^{xy}) \leq \sum_{k=n}^{\infty} P_x(A_k^{xy}) \leq \sum_{k=n}^{\infty} a^{\frac{k}{2}}$$

for all $n \in \mathbb{N}$. Therefore, for all $\epsilon > 0$ there exists $n \in \mathbb{N}$ such that $P_x(B_n^{xy}) < \epsilon$ and

$$k \geq n \Rightarrow d\left(w_{\sigma_k} \circ \ldots \circ w_{\sigma_1}(x), w_{\sigma_k} \circ \ldots \circ w_{\sigma_1}(y)\right) \leq a^{\frac{k}{2}} d(x,y)$$

for all $\sigma \in \Sigma^+ \setminus B_n^{xy}$. Also, it follows that

$$P_x(B^{xy}) = 0$$

and

$$\lim_{k \to \infty} d\left(w_{\sigma_k} \circ \dots \circ w_{\sigma_1}(x), w_{\sigma_k} \circ \dots \circ w_{\sigma_1}(y)\right) = 0$$

for all $\sigma \in \Sigma^+ \setminus B^{xy}$. $\square$

The next lemma is a generalization of Lemma 3 in [8].

Lemma 5.2.3. *Let $\delta > 0$ be such that $p_e(x) \geq \delta$ for all $x \in K_{i(e)}$ and $e \in E$. Let $x, y \in K_i$ for some $1 \leq i \leq N$. Then P_x is absolutely continuous with respect to P_y.*

Proof. Let $A \in \mathcal{B}\left(\Sigma^+\right)$ be such that $P_y(A) = 0$ and $\epsilon > 0$. We show $P_x(A) < \epsilon$.

By Lemma 5.2.2 (i), there exists $n_\epsilon > 0$ and $B \in \mathcal{B}\left(\Sigma^+\right)$ such that $P_x(B) < \epsilon/2$ and

$$k \geq n_\epsilon \ \Rightarrow \ d\left(w_{\sigma_k} \circ \dots \circ w_{\sigma_1}(x), w_{\sigma_k} \circ \dots \circ w_{\sigma_1}(y)\right) \leq a^{\frac{k}{2}} d(x, y)$$

for all $\sigma \in \Sigma^+ \setminus B$. Since each $p_e|_{K_{i(e)}}$ is Dini-continuous, we can choose $l \geq n_\epsilon$ such that $\sum_{k=l}^{\infty} \phi(a^{k/2} d(x, y)) < \delta/2$, where ϕ is the maximum of all moduli of continuity of $p_e|_{K_{i(e)}}$, $e \in E$. For every $n \in \mathbb{N}$, set

$$Q_n := \begin{cases} \left\{ \sigma \in \Sigma^+ : d\left(w_{\sigma_k} \circ \dots \circ w_{\sigma_1}(x), w_{\sigma_k} \circ \dots \circ w_{\sigma_1}(y)\right) \leq a^{\frac{k}{2}} d(x, y) \ \forall l \leq k \leq n \right\}, \\ \text{if } l \leq n \\ \Sigma^+, \text{ otherwise} \end{cases}$$

and $Q := \bigcap_{n \geq 1} Q_n$. Then $\Sigma^+ \setminus B \subset Q$ and therefore $P_x(\Sigma^+ \setminus Q) < \epsilon/2$. Now, for

$\sigma \in Q_n$ if $l \leq n$ and $(\sigma_1, ..., \sigma_n)$ is a path of the digraph starting in i, then

$$p_{\sigma_1}(x)...p_{\sigma_n}(w_{\sigma_{n-1}} \circ ... \circ w_{\sigma_1} x)$$

$$\leq \quad p_{\sigma_1}(y)...p_{\sigma_n}(w_{\sigma_{n-1}} \circ ... \circ w_{\sigma_1} y) \left(\frac{1-\delta}{\delta}\right)^l$$

$$\times \quad \prod_{k=l+1}^{n} \left[1 + \frac{p_{\sigma_k}(w_{\sigma_{k-1}} \circ ... \circ w_{\sigma_1} x) - p_{\sigma_k}(w_{\sigma_{k-1}} \circ ... \circ w_{\sigma_1} y)}{p_{\sigma_k}(w_{\sigma_{k-1}} \circ ... \circ w_{\sigma_1} y)}\right]$$

$$\leq \quad p_{\sigma_1}(y)...p_{\sigma_n}(w_{\sigma_{n-1}} \circ ... \circ w_{\sigma_1} y) \left(\frac{1-\delta}{\delta}\right)^l \prod_{k=l+1}^{n} \left[1 + \frac{\phi\left(a^{\frac{k-1}{2}}d(x,y)\right)}{\delta}\right].$$

Since $\prod_{k=l+1}^{n} \left[1 + \phi\left(a^{(k-1)/2}d(x,y)\right)/\delta\right] \leq 1 + 2\sum_{k=l+1}^{\infty} \phi\left(a^{(k-1)/2}d(x,y)\right)/\delta \leq 2$, it follows that

$$p_{\sigma_1}(x)...p_{\sigma_n}(w_{\sigma_{n-1}} \circ ... \circ w_{\sigma_1} x) \leq 2\left(\frac{1-\delta}{\delta}\right)^l p_{\sigma_1}(y)...p_{\sigma_n}(w_{\sigma_{n-1}} \circ ... \circ w_{\sigma_1} y).$$

If $l > n$ or $(\sigma_1, ..., \sigma_n)$ is not a path of the digraph starting in i, then it holds trivially for any $\sigma \in \Sigma^+$.

Let $\triangle$ be the algebra each element of which is a finite union of thin cylinders of the form $_1[e_1, ..., e_n]$. By the Caratheodory construction, there exists a sequence $(A_k)_{k \in \mathbb{N}} \subset \triangle$ such that $A \subset \bigcup_{k=1}^{\infty} A_k$ and

$$\sum_{k=1}^{\infty} P_y(A_k) < \frac{\epsilon}{4}\left(\frac{\delta}{1-\delta}\right)^l.$$

Each finite union $\bigcup_{k=1}^{n} A_k$ we can write as a disjoint union $\biguplus_{k=1}^{m_n} C_k$ of thin cylinders

which generate $\triangle$. Let $C_k = {}_1[e_1, ..., e_n]$. Then

$$
\begin{aligned}
P_x(Q \cap C_k) \quad &\leq \quad P_x(Q_n \cap C_k) \\
&= \sum_{(\sigma_1,...,\sigma_n):\sigma \in Q_n \cap C_k} p_{\sigma_1}(x)...p_{\sigma_n}(w_{\sigma_{n-1}} \circ ... \circ w_{\sigma_1} x) \\
&\leq \quad 2\left(\frac{1-\delta}{\delta}\right)^l \sum_{(\sigma_1,...,\sigma_n):\sigma \in Q_n \cap C_k} p_{\sigma_1}(y)...p_{\sigma_n}(w_{\sigma_{n-1}} \circ ... \circ w_{\sigma_1} y) \\
&\leq \quad 2\left(\frac{1-\delta}{\delta}\right)^l P_y(C_k).
\end{aligned}
$$

Hence

$$
\begin{aligned}
P_x(A) \quad &= \quad P_x(Q \cap A) + P_x(A \setminus Q) \leq \lim_{n \to \infty} P_x\left(\biguplus_{k=1}^{m_n} C_k \cap Q\right) + \frac{\epsilon}{2} \\
&= \quad \lim_{n \to \infty} \sum_{k=1}^{m_n} P_x(C_k \cap Q) + \frac{\epsilon}{2} \leq 2\lim_{n \to \infty} \left(\frac{1-\delta}{\delta}\right)^l \sum_{k=1}^{m_n} P_y(C_k) + \frac{\epsilon}{2} \\
&\leq \quad 2\left(\frac{1-\delta}{\delta}\right)^l \sum_{k=1}^{\infty} P_y(A_k) + \frac{\epsilon}{2} \\
&< \quad \epsilon.
\end{aligned}
$$

$\square$

Proposition 5.2.4. *(i) If μ is a unique invariant Borel probability measure of the CMS, then the generalized Markov shift associated with it is ergodic.*
(ii) If μ is an attractive Borel probability measure of the CMS, then the generalized Markov shift associated with it is strongly mixing.

Proof. Let ${}_1[a_1, ..., a_k]$ and ${}_1[c_1, ..., c_m]$ be two thin cylinder subsets of Σ. We can assume that $(a_1, ..., a_k)$ and $(c_1, ..., c_m)$ are paths of the directed graph. Then, for

sufficiently large n,

$$M\left({}_1[a_1, ..., a_k] \cap S^{-n}{}_1[c_1, ..., c_m]\right)$$

$$= \sum_{(b_1, ..., b_{n-k})} M\left({}_1[a_1, ..., a_k, b_1, ..., b_{n-k}, c_1, ..., c_m]\right)$$

$$= \int p_{a_1}(x)...p_{a_k}\left(w_{a_{k-1}}...w_{a_1}x\right) \sum_{(b_1, ..., b_{n-k})} p_{b_1}\left(w_{a^k}x\right)...p_{b_{n-k}}\left(w_{b_{n-k-1}}...w_{b_1}w_{a^k}x\right)$$

$$p_{c_1}\left(w_{b_{n-k}}...w_{b_1}w_{a^k}x\right)...p_{c_m}\left(w_{c_{m-1}}...w_{c_1}w_{b_{n-k}}...w_{b_1}w_{a^k}x\right)d\mu(x)$$

$$= \int p_{a_1}(x)...p_{a_k}\left(w_{a_{k-1}}...w_{a_1}x\right)$$

$$U^{n-k}\left(p_{c_1}p_{c_2} \circ w_{c_1}...p_{c_m} \circ w_{c_{m-1}} \circ ... \circ w_{c_1}\right)\left(w_{a^k}x\right)d\mu(x). \qquad (5.2.1)$$

Suppose, μ is a unique invariant Borel probability measure of the CMS. Then, by Theorem 2.2.1 (iii),

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

Therefore, $1/l \sum_{n=k}^{l} U^{n-k}\left(p_{c_1}p_{c_2} \circ w_{c_1}...p_{c_m} \circ w_{c_{m-1}} \circ ... \circ w_{c_1}\right)$ converges pointwise to $\int \left(p_{c_1}p_{c_2} \circ w_{c_1}...p_{c_m} \circ w_{c_{m-1}} \circ ... \circ w_{c_1}\right)d\mu$ as $l \to \infty$. Hence, by Lebesgue's dominated convergence theorem, (5.2.1) implies that

$$\lim_{l\to\infty}\frac{1}{l}\sum_{n=1}^{l} M\left({}_1[a_1, ..., a_k] \cap S^{-n}{}_1[c_1, ..., c_m]\right) = M\left({}_1[a_1, ..., a_k]\right) M\left({}_1[c_1, ..., c_m]\right).$$

By the standard extension argument (Theorem 1.17 in [25]), it follows that the generalized Markov shift associated with the CMS is ergodic.

Now, suppose that μ is an attractive Borel probability measure of the CMS, then

$$U^n g(x) = U^{*n}\delta_x(g) \to \int g d\mu \text{ for all } x \in K \text{ and } g \in C_B(K).$$

Hence, $U^{n-k}\left(p_{c_1}p_{c_2} \circ w_{c_1}...p_{c_m} \circ w_{c_{m-1}} \circ ... \circ w_{c_1}\right)$ converges pointwise to $\int \left(p_{c_1}p_{c_2} \circ w_{c_1}...p_{c_m} \circ w_{c_{m-1}} \circ ... \circ w_{c_1}\right) d\mu$ (as $n \to \infty$). Hence, by Lebesgue's dominated convergence theorem, (5.2.1) implies that

$$\lim_{n\to\infty} M\left({}_1[a_1,...,a_k] \cap S^{-n}{}_1[c_1,...,c_m]\right) = M\left({}_1[a_1,...,a_k]\right) M\left({}_1[c_1,...,c_m]\right).$$

This implies (Theorem 1.17 in [25]) that the generalized Markov shift associated with the CMS is strongly mixing.   □

Now, we prove the ergodic theorem. By Remark 2.1.1, case $N = 1$ of it is exactly Elton's Ergodic Theorem.

Theorem 5.2.5 (Ergodic Theorem). *Suppose that μ is a unique invariant Borel probability measure of the CMS and $\mu\left(K_i\right) > 0$ for all $i = 1,...,N$. Then for every $x \in K$*

$$\frac{1}{n}\sum_{k=1}^{n} \delta_{w_{\sigma^k}(x)} \xrightarrow{w^*} \mu \text{ for } P_x\text{-a.e. } \sigma \in \Sigma^+.$$

Proof. By Proposition 5.2.4 (i), the generalized Markov shift associated with the CMS is ergodic. Let $\Sigma_G := \{\sigma \in \Sigma : t(\sigma_i) = i(\sigma_{i+1}) \,\forall i \in \mathbb{Z}\}$. Since $M([e_1,...,e_n]) > 0$ for every path $(e_1,...,e_n)$, $M(\Sigma_G) = 1$ by the ergodicity. Therefore, we can restrict ourselves on Σ_G.

Let $f \in C_C(K)$. Further, let A be the set of all $\sigma \in \Sigma_G$ such that the limit defining $F(\sigma)$ exists and

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} f \circ w_{\sigma_k} \circ ... \circ w_{\sigma_1}\left(F\left(\sigma\right)\right) = \int f dF(M).$$

Since $F\left(S^k\sigma\right) = w_{\sigma_k} \circ ... \circ w_{\sigma_1}\left(F\left(\sigma\right)\right)$ for all such σ, it follows by Corollary 3.3.5 and Birkhoff's Ergodic Theorem that $M(A) = 1$. Applying the map ξ and Proposition 4.2.1, we deduce that

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} f \circ w_{\sigma_k} \circ ... \circ w_{\sigma_1}\left(y\right) = \int f d\mu$$

for all $(y, \sigma) \in \xi(A)$. By Lemma 5.2.1,   .

$$\tilde{\phi}(\mu)(\xi(A)) = \xi(M)(\xi(A)) = M\left(\xi^{-1}(\xi(A))\right) \geq M(A) = 1.$$

Since $\tilde{\phi}(\mu)$ is a probability measure,

$$
\begin{aligned}
1 &= \tilde{\phi}(\mu)(\xi(A)) = \int P_x\left(\{\sigma \in \Sigma^+ : (x,\sigma) \in \xi(A)\}\right) d\mu(x) \\
&= \sum_{i=1}^{N} \int_{K_i} P_x\left(\{\sigma \in \Sigma^+ : (x,\sigma) \in \xi(A)\}\right) d\mu(x).
\end{aligned}
$$

Furthermore, for each $i = 1, ..., N$, there exists $x_i \in K_i$ such that

$$
\begin{aligned}
&\int_{K_i} P_x\left(\{\sigma \in \Sigma^+ : (x,\sigma) \in \xi(A)\}\right) d\mu(x) \\
&= P_{x_i}\left(\{\sigma \in \Sigma^+ : (x_i,\sigma) \in \xi(A)\}\right)\mu(K_i).
\end{aligned}
$$

Set

$$Q_i := \left\{\sigma \in \Sigma^+ : (x_i, \sigma) \in \xi(A)\right\}$$

for each $i = 1, ..., N$. Then, for every i,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} f \circ w_{\sigma_k} \circ ... \circ w_{\sigma_1}\left(x_i\right) = \int f d\mu \text{ for all } \sigma \in Q_i$$

and, by the above,

$$\sum_{i=1}^{N} P_{x_i}\left(Q_i\right)\mu\left(K_i\right) = 1.$$

As $\mu(K_i) > 0$ for all $i = 1, ..., N$, this implies that $P_{x_i}(Q_i) = 1$ for all $i = 1, ..., N$.

Now, fix $x \in K_i$ for some $i \in \{1, ..., N\}$. Then, by Lemma 5.2.2 (ii), there exists $H_i \in \mathcal{B}(\Sigma^+)$ with $P_{x_i}(H_i) = 1$ such that

$$\lim_{k \to \infty} d\left(w_{\sigma_k} \circ ... \circ w_{\sigma_1}(x_i), w_{\sigma_k} \circ ... \circ w_{\sigma_1}(x)\right) = 0$$

for all $\sigma \in H_i$. Hence

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} f \circ w_{\sigma_k} \circ ... \circ w_{\sigma_1}(x) = \int f d\mu$$

for all $\sigma \in Q_i \cap H_i$. As $P_{x_i}(Q_i \cap H_i) = 1$, we deduce by Lemma 5.2.3 that $P_x(Q_i \cap H_i) = 1$. As $(C_C(K), \|.\|_\infty)$ is separable, it follows by an $\epsilon/3$−argument that

$$\frac{1}{n} \sum_{k=1}^{n} \delta_{w_{\sigma^k}(x)} \xrightarrow{w^*} \mu \text{ for } P_x\text{-a.e. } \sigma \in \Sigma^+.$$

$\square$

Finally, we give an important application of the Ergodic Theorem which allows an empirical calculation of Kolmogorov-Sinai entropy $h_M(S)$ of the generalized Markov shift associated with the CMS without explicitly knowing anything about its invariant measure.

We know by Theorem 4.3.1 that

$$h_M(S) = - \int \sum_{e \in E} p_e \log p_e d\mu$$

if μ is a unique invariant Borel probability measure of the CMS. Then, by Theorem 5.2.5, we deduce the following fact.

Corollary 5.2.6. *Suppose that μ is a unique invariant Borel probability measure of the CMS and $\mu\left(K_i\right) > 0$ for all $i = 1, ..., N$. Then for every $x \in K$*

$$h_M(S) = -\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \left(\sum_{e \in E} p_e \log p_e \right) \circ w_{\sigma^k}(x) \ for \ P_x\text{-}a.e. \ \sigma \in \Sigma^+.$$

Bibliography

[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. Henri Poincaré **24** (1988), no. 3, 367–394.

[2] _____, *Eratum; Invariant measures for Markov processes arising from iterated function systems with place-dependent probabilities*, Ann. Inst. Henri Poincaré **25** (1989), no. 4, 589–590.

[3] M. F. Barnsley and J. H. Elton, *A new class of Markov processes for image encoding*, Adv. Appl. Prob. **20** (1988), 14–32.

[4] H. Berbee, *Chains with infinite connections: Uniqueness and Markov representation*, Probab. Th. Rel. Fields **76** (1987), 243–253.

[5] Marc A. Berger, *An Introduction to Probability and Stochastic processes*, Springer-Verlag, New York, 1998.

[6] P. Brémaud, *Markov chains: Gibbs fields, Monte Carlo simulation, and queues*, Springer-Verlag, New York, 1998.

[7] J. L. Doob, *Measure Theory*, Springer-Verlag, New York, 1993.

[8] J. H. Elton, *An ergodic theorem for iterated maps*, Ergod. Th. & Dynam. Sys. **7** (1987), 481–488.

[9] S. Grigorescu, *Limit theorems for markov chains arising from iterated function systems*, Rev. Roumaine Math. Pures Appl. **37** (1992), no. 10, 887–899.

[10] P. Hulse, *Uniqueness and ergodic properties of attractive g-measures*, Ergod. Th. & Dyman. Sys. **11** (1991), 65–77.

[11] ———, *A class of unique g-measures*, Ergod. Th. & Dyman. Sys. **17** (1997), 1383–1392.

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

[13] M. Iosifescu and S. Grigorescu, *Dependence with complete connections and its applications*, Cambridge University Press, 1990.

[14] T. Kaijser, *On a new contraction condition for random systems with complete connections*, Rev. Roum. Math. Pures et Appl. **XXVI** (1981), no. 8, 1075–1117.

[15] ———, *On a Theorem of Karlin*, Acta Appl. Math. **34** (1994), 51–69.

[16] M. Keane, *Strongly mixing g-measures*, Inventiones math. **16** (1972), 309–324.

[17] D. Lind and B. Marcus, *An introduction to symbolic dynamics and coding*, Cambridge University Press, 1995.

[18] R. D. Mauldin and S. C. Williams, *Hausdorff dimension in graph directed constructions*, Tran. AMS **309** (1988), 811–829.

[19] O. Onicescu and G. Mihoc, *Sur les chaînes de variables statistiques*, Bull. Sci. Math. de France **59** (1935), 174–192.

[20] Ö. Stenflo, *Ergodic theorems for Markov chains represented by iterated function systems*, Bull. Polish Acad. Sci. Math. **49** (2001), 27–43.

[21] _____, *Uniqueness of invariant measures for place-dependent random iterations of functions*, IMA Vol. Math. Appl. **132** (2002), 13–32, Fractals in multimedia (Minneapolis, MN, 2001).

[22] _____, *Uniqueness in g-measures*, Nonlinearity **16** (2003), 403–410.

[23] M. Takesaki, *Twenty-Five Years in the Theory of Type III von Neumann Algebras*, AMS Contemporary Mathematics **167** (1994), 239.

[24] P. Walters, *Ruelle's operator theorem and g-measures*, Trans. Am. Math. Soc. **214** (1975), 375–387.

[25] _____, *An introduction to Ergodic Theory*, Springer-Verlag, New York, 1982.