On Mealy-Moore coding and images of Markov measures

Rostislav Grigorchuk, Roman Kogan, Yaroslav Vorobets

January 8, 2020

1 Introduction

Finite Mealy-type automata (and closely related to them Moore-type automata) play an important role in computer science. They frequently are called transducers, or sequential machines, and are used for various type of coding and related topics.

Such automata also play remarkable role in algebra (semigroup theory, group theory) [Gr1980, Gr1983, Gr1991, Kyoto Congress, Gupta-Sidki, GNS, Steinberg-Vorobets]. Since 2000 they started to play also a significant role in combinatorics (graph theory) [Barth.Gr 00, Bondarenko-..., Nagnibeda], fractals [BGS02, GNS2015], dynamical systems (in particular holomorphic dynamics) [GNS00, Gr2011, Nekr(book), Gr.Savchu2015?].

The main feature of initial deterministic automaton \( A_q \) with finite input alphabet \( A \) of the cardinality \( |A| \geq 2 \) and finite output alphabet \( B \) is that it transforms finite words (strings) over \( A \) into the words of the same length over \( B \), and the map (which we denoted \( \hat{A}_q \) can be extended to the map \( \hat{A}_q : A^N \to B^N \) defined on the space \( A^N \) of infinite words. If the input and output alphabets coincide then one can iterate the map \( \hat{A}_q \), which lead to the dynamics on the space \( A^* \) of all finite words over \( A \), or on the space \( A^N \) of infinite words. The space \( A^N \) has a natural product topology which makes it homeomorphic to a Cantor set. The map \( \hat{A}_q \) is continuous with respect to this topology and the pair \( (\hat{A}_q, A^N) \) is what is called in topological dynamics a topological dynamical system. A famous example of this sort is given the odometer, or adding machine, which is a map that can be defined by a 2-state automaton shown in Figure 1. The odometer gives a quite clear picture of the dynamics. For every \( n \in \mathbb{N} \), it permutes cyclically the words of of length \( n \) according to adding 1 \( \mod 2^n \) in the \( d \)-adic expansion of numbers \( i \), \( = \leq i \leq 2^n - 1 \). It also clearly acts on the space
of infinite sequences by the addition of 1 in the ring of $d$-adic integers (if we identify infinite words with them).

![Odometer Diagram]

Figure 1: The odometer, also known as the adding machine

The situation becomes much more complicated in the general case when the problem of obtaining the decomposition into ergodic components become hard, as demonstrated in [, GrigSavchuk2015]. The transformations we consider have a pure points spectrum (in the sense of Koopman), and hence are conjugate to the translations in an abelian group. But this observation does not help much.

In [, Ryabin ] (see [, KudryavcevAlesh]... where more details are given) , Ryabinin touched a subject of what happens with the frequencies of symbols of a Bernoulli measure on $\{0,1\}^\mathbb{N}$ under the transformation $\hat{A}_q$. If $\mu_p$ is Bernoulli measure given by probabilistic distribution $\{1 - p, p\}$ on the alphabet $\{0,1\}$, the a formula was suggested for frequency $f(p)$ of symbol 1 in the typical with respect to the image measure $(\hat{A}_q)_*\mu_p$ (such function is called in [?]he “stochastic function”. No justification to the formula for $f(p)$ was given in any of the cited sources (a note and a book). In the described situation the functions $f(p)$ is rational. Also in [, KudtAl] a theorem (attributed to Ryabinin [, Ryab]) is stated and supplied buy the arguments that lists a necessary and sufficient conditions for a rational function $P(x)/Q(x)$ to be a a stochastic function. A general formula in the case of arbitrary cardinality of the alphabet $A$ given by R.Kravchenko in [] shows that the formula from [?]s correct in the case when $A_q$ is invertible, and we observe that it is incorrect without the assumption of invertibility of $A_q$ (see example ??). This allows to conclude that the theorem ?? from [?]udAl also is not correct (at least its proof), as no assumption of invertibility is used in the construction of the automaton that realizes given $f(p) = P(x)/Q(x)$.

A natural question arises if there are $T$-quasi-invariant (i.e. their images are absolutely continuous with respect to preimages) Bernoulli measures or more generally $T$-quasi-invariant Markov measures on $\Omega$ different from $\mu_{un}$. R.Kravchenko, who put this type of consideration on rigorous rooting (based on the use of methods and results of ergodic theory), in addition to getting the correct formula for frequencies
in the case of arbitrary $d = |A|$, considered a problem of how the image measure is related to the original one $\mu_p$ given by a probabilistic distribution $p = (p_1, \ldots, p_d)$ on $A$. He showed that for the so-called polynomially bounded (in the sense of S.Sidki [] around 2000) the image measure is absolutely continuous with respect to $\mu_p$. On the contrary, in the case of strongly connected automaton (when any two states in the Moore diagram of the automaton can be connected by a path) and $p$ is not a uniform distribution on $A$, the image $(\hat{A}_q)\mu_p$ is singular with respect to the $\mu_p$ (in fact, this result requires a small correction as demonstrated in Thm. 6.11).

A singularity of the image with respect to the original measure is given by the comparison of frequencies of symbols from the alphabet in typical sequences.

This paper generalizes the results of Kravchenko on Markov measures. We show that their images are absolutely continuous with respect to the original measures in the case of polynomially bounded automata (Theorem 4.4). We get a formula that express frequencies of symbols with respect to the image measure via the data that determine Markov measure (a stochastic matrix denoted by us $L$ and a stationary vector $l$ for it). Under quite general conditions on a measure and automaton, we show the singularity of the image measure with respect to the origin. The proof uses the idea of converting the Markov chain on $A$ given by $L$ to a Markov chain on $S \times A$ (where $S$ is a set of states of the automaton $A_q$) given by a matrix, denoted $T$ in this paper, that is constructed in a special way that perhaps can be called a skew product. The main idea of the trick is to get an image measure as a “piece” of the measure (denoted in our paper by $Q$ that is an image under the 1 : 1 block map of a Markov measure over the alphabet $S \times A$). We pay some attention to the nature of the projection measure $Q$ and show that in some cases it is a Gibbs measure (Theorem 9.2). A special attention is paid to the case when a stationary vector $t$ for $T$ is a tensor product of vectors $l$ and a vector $k$, which is a stationary vector of a matrix $K$ indexed by elements of the set of $S$ which is constructed using vector $l$ and Moore diagram of the automaton $A_q$.

What is the use of the above results and their generalizations provided in this paper? First, there is a very interesting algebra related to finite automata of Mealy type. Namely, given a non-initial invertible automaton $A$ one can consider a group $G(A)$ generated by $A$ (in a sense to be explained later). The groups of this form are called automaton groups or self-similar groups and they play important role in group theory as they were used to solve a number of famous problems and find applications in many areas of mathematics [].

If the automaton $A$ is polynomially bounded then by the results of Kravchenko and of this paper the Bernoulli and more generally Markov measures are $G(A)$-quasi-invariant and therefore can be used to build a Koopman type unitary representations.
Study of such representations was initiate by A. Dudko and the first author in [], where it was shown that there are many pairwise disjoint representations of this type, they are irreducible and posses some other useful and interesting properties.

The paper is organized as follows. First, the necessary preliminaries are given, and new definitions are introduced. The case of automorphisms of polynomial growth is considered first. Then the strongly-connected case is considered. This is followed by examples, special cases, and considerations of the cases when the image is Gibbsian.

2 Preliminaries

Let $X$ be a finite alphabet, and $(X^N, \mathcal{B}, \mu)$ a probability measure space such that the shift $\sigma : X^N \to X^N$ is a probability-preserving transformation. If $\sigma$ is an ergodic probability-preserving transformation on $(X^N, \mathcal{B}, \mu)$, Birkhoff’s Pointwise Ergodic Theorem is a tool that can be used to calculate frequencies: for almost all $w \in X^N$ and $a \in X$,

$$\text{freq}(a) = \lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} \chi_{\sigma^k(w)} = \int \chi_a d\mu,$$

where $\chi_a$ is the characteristic function on the cylinder $aX^N$, and the left hand side of the equation is the frequency of $a$.

**Example 2.1.** Let $X = \{0, 1\}$, and $\mu$ be a Bernoulli measure defined by the probability vector $q = (p, 1-p)$. On the cylinders, $\mu$ is given as follows: $\mu(0wX^N) = p\mu(wX^N)$ and $\mu(1wX^N) = (1-p)\mu(wX^N)$ for any finite word $w$, and $\mu(X^N) = 1$. $\triangle$

This measure can be understood as a (stochastic) process of flipping a coin, possibly biased, which gives heads with probability $p$. One would expect to get $\text{freq}(0) = p$; and indeed, by the ergodic theorem we have almost everywhere:

$$\text{freq}(0) = \int \chi_0 d\mu = \mu(0X^N) = p\mu(X^N) = p.$$

We then may ask the question: how do actions on the space affect frequencies? A natural object that acts on infinite sequences is a Mealy machine: an initial finite-state automaton with output, so we restrict our attention to the action $w \mapsto gw$ given by an automaton transformation, and examine

$$\lim_{n \to \infty} \sum_{k=0}^{n-1} \chi_{\sigma^k(gw)}.$$
The case for $\mu$ - a Bernoulli process (on a two-letter alphabet) was solved by Ryabinin in [2], and for $\mu$ - a Bernoulli measure on a finite alphabet was answered by Kravchenko in [1]. We generalize these result to the case of $\mu$ - a Markov measure.

3 Definitions

Let $X = \{a_1, \ldots, a_k\}$ be a finite alphabet, and $L$ a $k$-by-$k$ irreducible stochastic matrix (with entries indexed by $X$) with a stationary probability vector $l$. That is, $\sum_b L_{ab} = 1$ for all $a \in X$, $lL = l$, and the directed graph whose incidence matrix is given by nonzero entries of $L$ is path-connected. Let $\mu$ be the Markov measure on $X^\mathbb{N}$ induced by $L, l$; $\mu$ is given on the cylinders by

$$\mu(x_1x_2\ldots x_nX^\mathbb{N}) = l_{x_1}L_{x_1,x_2}L_{x_2,x_3}\ldots L_{x_{n-1},x_n}.$$

Hereafter, we let $\mathcal{B}$ be the sigma-algebra generated by cylinder sets $wX^\mathbb{N}$, for $w$ - finite word.

The shift operator $\sigma$ acts on $X^\mathbb{N}$ by “eating” the first character: $\sigma(aw) = w$, for $a \in X$, $w \in X^\mathbb{N}$.

A probability measure $\nu$ on $X^\mathbb{N}$ is invariant (with respect to the shift $\sigma$) if $\nu(\sigma^{-1}(E)) = \nu(E)$ for all $E \in \mathcal{B}$. We call an probability measure $\nu$ ergodic if for all invariant sets $E \in \mathcal{B}$ (sets such that $\sigma^{-1}(E) = E$), either $\nu(E) = 1$ or $\nu(E) = 0$ holds.

Equivalently, we may say that the shift $\sigma$ is an invariant and ergodic transformation on $(X, \mathcal{B}, \nu)$, or that $(X^\mathbb{N}, \mathcal{B}, \nu)$ is a probability-preserving transformation (ppt for short).

Example 3.1. The Markov measure $\mu$ defined above is invariant and ergodic [3]. Bernoulli measures are specific cases of Markov measures, and so are invariant and ergodic. △

Let $A = (X, S, \pi, \lambda)$ be a Mealy machine (an automaton with output) with the set of states $S$, a transition function $\pi : S \times X \to S$ and output function $\lambda : S \times X \to X$. These functions can be extended to $X^\mathbb{N}$ by the recursive rules: for $x \in X$, $w \in x^\mathbb{N}$, set

$$\pi(s, xw) = \pi(s, x)\pi(\pi(s, x), w);$$
$$\lambda(s, xw) = \lambda(s, x)\lambda(\pi(s, x), w).$$

A choice of an initial state $g \in S$ defines the action of the automaton on the space $X^\mathbb{N}$: we define $g(w) := \lambda(g, w)$.  

We write
\[ s_0 \xrightarrow{x_1} s_1 \xrightarrow{x_2} \cdots \xrightarrow{x_n} s_n \]
when the automaton takes an input string \( x_1 x_2 \ldots x_n \), outputs \( y_1 y_2 \ldots y_n \) while going through a sequence of states \( s_0 s_2 \ldots s_n \). That is, \( \pi(s_{i-1}, x_i = s_i) \) and \( \lambda(s_{i-1}, x_i = y_i) \) for \( i \in 1..n \). We call the above a path in the automaton (as it is indeed a path in the directed graph where the edges are given by the set of states \( S \), the transition function \( \pi \) providing the incidence matrix, and \( \lambda \) giving edge labels).

An automaton \( A \) is invertible if \( \lambda \) is injective. To an invertible automaton \( A \) with states \( s_1, \ldots, s_n \), we associate an automaton \( A^{-1} \) with states \( s_1^{-1}, \ldots, s_n^{-1} \) defined as follows:

\[
\lambda(s_i^{-1}, \lambda(s_i, x)) := x; \\
\pi(s_i^{-1}, \lambda(s_i, x)) := \pi(s_i, x).
\]

This is well-defined, and for \( w \in X^N \), \( s_i^{-1} \circ s_i(w) = s_i \circ s_i^{-1}(w) = w \). Since the actions preserve word length, they are isomorphisms of the regular tree where the children of every node are indexed by \( X \). We can now talk about the group of tree automorphisms generated by the states of \( A \).

Conversely, to a tree automorphism \( g \) we can associate the automaton of restrictions with an initial state labeled by \( g \) with the same action: we take \( S = \{ g|_w : w \in X^N \} \), and define \( \pi(h, x) := h|_x; \lambda(h, x) = h(x) \). A tree automorphism is called finite-state whenever its automaton is finite.

We call an automaton \( A \) strongly-connected if for every pair of states \( s, t \in S \) there exists a path that starts in \( s \) and ends in \( t \). A tree automorphism is strongly-connected whenever its automaton of restrictions is.

We define a stronger notion of connectedness (\( L \)-strongly-connectedness) in a later section, which we use as a sufficient condition for ergodicity of certain measures.

A tree automorphism \( g \) is said to have polynomial growth if the number of words \( w \) of length \( n \) such that \( g|_w X^N \) is nontrivial is at most polynomial in \( n \).

If \( g \) has polynomial growth, then \( g|_w \) is trivial for most \( w \). For such \( w \), \( \tilde{g} = \pi(g, w) \) acts trivially; we call such states trivial, and all others - nontrivial. Note that in the diagram of that automaton, all paths going from a trivial state can only go to trivial states; so if the automaton is reduced, it may only have at most one trivial state \( I \), with \( \pi(I, x) = I \) and \( \lambda(I, x) = x \) for all \( x \in X \).

From the above, it follows that if \( g \) is a tree automorphism generated by a reduced finite automaton \( A \) (with initial state \( g \)), then \( A \) has a unique trivial state.

Note that a tree automorphism can either be of polynomial growth, or be strongly-connected. Therefore, we consider these two cases separately.
From here onwards, we assume the automorphisms and the automata are all finite-state.

4 Automorphisms of polynomial activity

In the polynomial growth case, the pushforward of any non-atomic measure (measures $\nu$ such that $\nu(S) = 0$ for any countable set $S$) is easy to find. This comes from the following lemmas:

Lemma 4.1 Let $A = (X, S, \pi, \lambda)$ be a strongly connected automaton, and $g(w) := \lambda(g, w)$ as above. A nontrivial simple cycle through $s \in S$ is a path in $A$ that starts and ends in $s$ and consists of nontrivial states. Then $g$ has polynomial growth only if there do not exist two distinct nontrivial cycles through any state $s \in S$.

Proof: Suppose that there exists a state $s$ with two distinct cycles, defined by input words $a$ and $b$, through it:

$$
\begin{align*}
s \xrightarrow{a_1} \lambda(s, a_1) & \xrightarrow{a_2} \ldots \xrightarrow{a_n} \lambda(s, a_n) \xrightarrow{s_{n+1}} s; \\
s \xrightarrow{b_1} \lambda(s, b_1) & \xrightarrow{b_2} \ldots \xrightarrow{b_m} \lambda(s, b_m) \xrightarrow{t_{m+1}} t.
\end{align*}
$$

Furthermore, note that since $A$ is strongly connected, there exists a path defined by an input word $w_{gs}$ that goes from $g$ to $s$.

Consider words $w$ of the form

$$
w = w_{gs} A_1 A_2 \ldots A_{2n},$$

where $A_i \in \{a, b\}$, and where $a$ occurs $n$ times (i.e. $|\{i : A_i = a\}| = n$). All these words have the same length:

$$
|w_{gs} A_1 A_2 \ldots A_{2n}| = |w_{gs}| + n(|a| + |b|),
$$

and this length is linear in $n$.

However, there are at least $2^n$ distinct such words. Indeed, there are $2^n$ distinct words in the alphabet $\{a, b\}$; it suffices to show that they give distinct words in $X$. 

7
Since input words \(a\) and \(b\) define distinct paths in \(A\), there exist \(i, j\) such that \(s_i \neq t_j\). Given an input word \(w\), define a function \(F : X^N \to \{0, 1\}^N\) as follows:

\[
F(wx) = \begin{cases} 
F(w)0, & \text{if } \pi(s, wx) = s_i; \\
F(w)1, & \text{if } \pi(s, wx) = t_j; \\
F(w) & \text{otherwise.}
\end{cases}
\]

That is, \(F\) simply writes down 0 when \(s_i\) is encountered, and 1 when \(t_j\) is encountered while going along the path defined by \(w\).

Let \(R\) be the rewriting of a word in alphabet \(\{a, b\}\) in \(X\). Then \(F \circ R\) is injective by the assumption that \(s_i\) and \(t_j\) are distinct states such that \(s_i\) is not on the path defined by \(b\), and \(t_j\) is not on the path defined by \(a\). Therefore, \(R\) must be injective as well.

Finally, we note that restriction of \(g\) to \(w\) is nontrivial, since \(\pi(g, w) = s\), a nontrivial state.

We thus produced, for arbitrary \(n\), at least \(2^n\) words of length \(|w_{gs} + n(|a| + |b|)|\) such that \(g|_w\) is nontrivial; by definition, this implies exponential growth. The lemma holds by contradiction. □.

**Lemma 4.2** \(A^{-1}\) has polynomial growth whenever \(A\) has polynomial growth.

**Proof:** This is by definition from the following observation:

\[g|_w \neq 1 \iff g^{-1}|_{g(w)} \neq 1.\]

(In other words, if \(g\) leaves suffixes of \(w\) unchanged, so does \(g^{-1}\)).

**Lemma 4.3** (Kravchenko) Let \(V, V_{max}\) be given by

\[
V = \{w \in X^N : g^{-1}|_w = 1\}
\]

\[
V_{max} = \{w \in V : w = vw', |w'| > 0 \Rightarrow v \notin V\}.
\]

Then \(\mu\) is supported on \(\sqcup_{v \in V_{max}} vX^N\).

That is, \(V\) is the set of words giving trivial sections, and words in \(V_{max}\) are words in \(V\) whose proper prefixes are not in \(V\) (i.e. yield non trivial sections). In a tree diagram of the automorphism \(g^{-1}\), \(v \in V_{max}\) if next-to-last node on the path given by \(v\) is a switch node, and there are no switches in its subtree.

**Proof:** First, note that the cylinders \(vX^N\) are disjoint for \(v \in V_{max}\): if \(v_1X^N\) and \(v_2X^N\) intersect, than either \(v_1\) starts with \(v_2\), or \(v_2\) starts with \(v_1\) - neither is possible by construction. Therefore, the union of sets \(vX^N\) for \(v \in V_{max}\) is disjoint.
Now, to prove the lemma, we show that

\[ W = X^N - \bigcup_{v \in V_{\text{max}}} vX^N \]

is at most countable. Indeed, let \( w \in W \). Then the path in the statues of that automaton of \( g^{-1} \) defined by \( w \) must consist of nontrivial states: if \( \pi(g, w_1 w_2 w_3 \ldots w_n) \) is a trivial state, then so are all the subsequent states on the path defined by \( w \), and so \( g^{-1}|_{w_1 w_2 \ldots w_n} \) is trivial.

By assumption, \( A \) (and thus \( A^{-1} \)) is finite, and so the path in \( A^{-1} \) defined by \( w \) must pass through some non-trivial state \( s \) infinitely often. By Lemma 4.2, \( A^{-1} \) has polynomial growth. There, Lemma 4.1 applies, and so there is at most one nontrivial cycle passing through \( s \). Therefore \( w \) is eventually periodic. The set of such words is countable. Since for Markov measures, \( \mu(S) = 0 \) when \( S \) is countable, the lemma follows. □

To proceed further, we introduce a technical definition. Given a measure \( \mu \) and an automorphism \( g \), we say \( g \) is \( \mu \)-\( V_{\text{max}} \)-compatible if \( \mu(vX^\infty) \neq 0 \) for all \( v \in V_{\text{max}} \).

If \( \mu \) is a Markov measure induced by a matrix \( L \), this means that all paths in the automaton of restrictions \( g \) that lead to a trivial state must be induced by words \( w \) that are not forbidden in \( L \) (i.e. \( L_{w/w+1} > 0 \) for \( i = 0, \ldots, |w| - 1 \)).

We can now show the following:

**Theorem 4.4** Let \( \mu \) be a Markov measure induced by \( L, l \), and \( g \) - an automorphism of polynomial growth. If \( g \) is \( \mu \)-\( V_{\text{max}} \)-compatible, then \( g^* \mu \) is absolutely continuous w.r.t. \( \mu \), with the derivative given by

\[
\frac{dg^* \mu}{d\mu} = \sum_{v \in V_{\text{max}}, a \in X} \frac{\mu(g^{-1}(vX^\infty)) L(g^{-1}(v), a)}{\mu(vX^\infty) L(v, a)} \chi_{vaX^N}.
\]

**Proof:** Note that the above expression is well-defined iff \( g \) is \( \mu \)-\( V_{\text{max}} \)-compatible.

Now, extending the approach of Kravchenko to Markov measures, let

\[
g' = \sum_{v \in V_{\text{max}}, a \in X} \frac{\mu(g^{-1}(vX^\infty)) L(g^{-1}(v), a)}{\mu(vX^\infty) L(v, a)} \chi_{vaX^N}.
\]

By construction, the measure \( g'd\mu \) is supported on cylinder sets \( vX^N \), for \( v \in V_{\text{max}} \). From Lemma 4.3, it suffices to show that \( dg^* \mu \) and \( g'd\mu \) agree on these cylinder sets for the theorem to hold.

Since both measures are continuous, the above will follow if

\[
\int_{wX^N} g'd\mu = \int_{wX^N} dg^* \mu
\]
for all $w \in V$.

Now, if $w \in V$, then either $w \in V_{\max}$, or $w = vaw'$ where $v \in V_{\max}$ and $a \in X$. In the latter case:

\[
\int_{wX^N} g'd\mu = \frac{\mu(g^{-1}(vX^N))}{\mu(vX^N)} \cdot \frac{L(g^{-1}(v|v|), a)}{L(v|v|, a)} \cdot \mu(vaw'X^N)
\]

\[
= \frac{\mu(g^{-1}(vX^N))}{\mu(vX^N)} \cdot \frac{L(g^{-1}(v|v|), a)}{L(v|v|, a)} \cdot \mu(vX^N) \cdot \mu(vaw|v|, a) \cdot L(a, w') \cdots L(w'_{|w'|-1}, w'_{|w'|})
\]

\[
= \frac{\mu(g^{-1}(vX^N))}{\mu(vX^N)} \cdot \frac{L(g^{-1}(v|v|), a)}{L(v|v|, a)} \cdot \mu(vX^N)
\]

\[
= \mu(g^{-1}(vaw'X^N)) = \mu(g^{-1}(vaw'X^N))
\]

\[
= \mu(g^{-1}(wX^N))
\]

\[
= \int_{wX^N} g_*\mu.
\]

In the above, equality 1 follows from the assumption that $v \in V_{\max}$, and therefore

\[
g^{-1}(vaw') = g^{-1}(v)aw'
\]

for all $w'$.

Now that we have verified that $g'd\mu$ and $dg_*\mu$ agree on cylinders $wX^N$ for $w \in V$, the theorem holds. □

### 5 Subexponential case

The result of the Theorem 4.4 also holds in the case of $g$ having subexponential growth, i.e. $g$ satisfying the following condition:

\[
|\{w : |w| = n, g|w \neq 1\}| < C^n \text{ for all } C > 1.
\]

Indeed, a more general version of Lemma 4.3 holds:

**Lemma 5.1** The conclusion of Lemma 4.3 holds in the subexponential case.
Proof: All we need to demonstrate is that $\mu(X^n - \sqcup_{v \in V_{max}} v X^n) = 0$. Let $V_n = \{ w : |w| = n, g|_w \neq 1 \}$. Following [Grigorchuk-Dudko], let

$$M_n = \mu(\sqcup_{w \in V_n} w X^n),$$

and let $L_{max} = \max L_{ij} < 1$. Then

$$M_n \leq |V_n| \max\{ \mu(w X^n) : w \in V_n \}$$

$$< |V_n| L_{max}^n.$$

For $w \in X^n$, $w \neq vw'$ for some $v \in V_{max}$ if and only if $w_1 w_2 \ldots w_n \in V_n$ for $n = 1..|w|$, and so

$$\mu(X^n - \sqcup_{v \in V_{max}} v X^n) = \lim_{n \to \infty} M_n = 0$$

by the definition of subexponential growth of an automaton. □

Aside from Lemma 4.3, the proof of Theorem 4.4 does not depend on whether $g$ has polynomial growth or not; all that matters is that the conclusion of Lemma 4.3 holds. Therefore, we have:

**Corollary 5.2** The conclusion of Theorem 4.4 holds in the subexponential case.

6 Strongly-connected tree automorphisms

The measure given by $\mu$ is invariant and ergodic. Naively, one could hope that the automaton action preserves this, and Birkhoff’s Pointwise Ergodic Theorem could be applied directly by a change of variable in the integral:

$$\text{freq}(x) = \lim_{n \to \infty} \sum_{k=0}^{n-1} \chi_x \circ \sigma^k(gw) \equiv \int \chi_x dg \ast \mu.$$

However, the pushforward measure $g_*\mu$ is not a-priori invariant and ergodic, and the measure is not easy to deal with directly. On the cylinders, it is given by

$$g_*\mu(y_1 y_2 \ldots y_n X^n) = \sum_{g \xrightarrow{x_1} y_1 \xrightarrow{x_2} \ldots \xrightarrow{x_n} y_n \xrightarrow{s_n}} l(g) L(g, x_1) L(x_1, x_2) \ldots L(x_{n-1}, x_n),$$

where the summation is over all paths $g \xrightarrow{x_1} y_1 \xrightarrow{x_2} \ldots \xrightarrow{x_n} y_n \xrightarrow{s_n}$ in the automaton.
To address this difficulty, we keep track of the states $A$ goes through along with the output. We will define maps to obtain a commutative diagram:

$$
\begin{align*}
X^N & \xrightarrow{g} g(X^N) \subset X^N \\
\pi_g & \xrightarrow{\lambda} (S \times X)^N \\
\end{align*}
$$

with $g_*\mu = \tilde{\lambda} \circ \tilde{\pi} _g \mu$. We then define an invariant, ergodic measure $Q$ on $X^N$ which is a pushforward of a Markov measure under the 1-block factor map $\tilde{\lambda}$ such that $g_*\mu \ll Q$. We are then able to apply the pointwise ergodic theorem with respect to measure $Q$, and the result will hold for $\mu$-almost-all input words $w$.

Let $\tilde{\pi} : S \times X^N \to (S \times X)^N$ be given by

$$
\tilde{\pi}(g,xw) = (g,x)\tilde{\pi}(\pi(g,x),w),
$$

and write $\tilde{\pi}_g(w) := \tilde{\pi}(g,w)$.

Given a sequence $\tilde{\pi}(g,w) \in (S \times X)^\infty$, we can extract the output $g(w) = \lambda(g,w)$ simply by looking at the states; so we define $\tilde{\lambda} : (S \times X)^\infty \to X^\infty$ recursively by

$$
\tilde{\lambda}((s,x)w) = \lambda(s,x)\tilde{\lambda}(w)
$$

for $w \in (S \times X)^\infty$.

That is, for each path

$$
g \xrightarrow{x_1} s_1 \xrightarrow{x_2} \ldots \xrightarrow{x_n} s_n
$$

we have

$$
\tilde{\pi}_g(x_1,x_2,\ldots,x_n) = (g,x_1)(s_1,x_2)\ldots(s_{n-1},x_n);
\tilde{\lambda}(s_0,x_1)(s_1,x_2)\ldots(s_{n-1},x_n) = y_1y_2\ldots y_n.
$$

By construction, we have $\tilde{\lambda} \circ \tilde{\pi}_g(w) = g(w)$, and thus have the commutative diagram from the previous section.

The measure $\tilde{\pi}_g \mu$ on $(S \times X)^N$ is easier to work with than the pushforward $g_* \mu$ on $X^N$. On the cylinders, it is given as follows:

$$
\tilde{\pi}_g \mu((g,x_1)(s_1,x_2)\ldots(s_{n-1},x_n)(S \times X)^\infty) = \begin{cases} 
    l(x_1)Lx_1x_2\ldots Lx_{n-1}x_n, & \text{if } g \xrightarrow{x_1} s_1 \xrightarrow{x_2} \ldots \xrightarrow{x_n} s_n \text{ is a path} \\
    0, & \text{otherwise.}
\end{cases}
$$
This is *almost* a Markov measure! The product matches a Markov measure except for the first term, $l(x_i)$. It is piecewise-Markov, scaled by constants on cylinders $(s,x)(S \times X)^\infty$.

To make this statement more precise, let $T = T_{L,A}$ be a transition matrix with entries indexed by elements of $S \times X$, and values given by

$$T_{(s_0,x_0),(s_1,x_1)} = \begin{cases} L(x_0, x_1), & \text{if } \pi(s_0, x_0) = s_1; \\ 0, & \text{otherwise.} \end{cases}$$

$T$ then is a stochastic matrix:

$$\sum_{(t,y)} T_{(s,x)(t,y)} = \sum_y L(x, y) = 1,$$

since for $T_{(s,x)(t,y)} \neq 0$, $t$ must be uniquely given by $t = \pi(s, x)$.

**Example 6.1.** Let $X = \{0, 1, 2\}$, $\mu$ be induced by the matrix

$$L = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \end{pmatrix},$$

and let $A$ have the transition function given by the diagram in Figure 2.

$$\text{Figure 2: Any path from } A \text{ to } C \text{ ends in } 02$$

Note that $02$ is a forbidden word in the subshift defined by $L$ (that is, $\mu(02X^\mathbb{N}) = 0$). However, the only arrow going into $B$ is labeled by 0. Therefore, $T$ has a zero column at $(C, 2)$, and so the Markov measure induced by it is not irreducible. $\triangle$

The example above calls for a restriction on $A$ that would make $T$ irreducible.

**Definition:** We call an automaton $A$ $L$-strongly-connected, if for every pair of states $s, t \in S$ and every pair of symbols $x, y \in X$, there exists word $w \in X^*$ such that $\pi(s, xw) = t$, and $xwy$ is not a forbidden word in the subshift defined by the nonzero entries of $L$ (that is, if $w = w_1 \ldots w_n$, then $L_{x,w_1}$, $L_{w_n,y}$, and $L_{w_i,w_{i+1}}$ for $i = 1..n - 1$ are all nonzero).
Lemma 6.2 The Markov chain defined by $T$ is irreducible if and only if the automaton $A$ is $L$-strongly connected.

Proof: This follows directly from the definitions given above.

Let $t$ be the stationary probability vector of $T$, so $tT = t$, which exists since $T$ is irreducible; and let $P$ be the associated Markov measure. Then on the cylinders, $P$ is given by

$$P((s_0, x_0) \ldots (s_n, x_n)(S \times X)^\infty) = \begin{cases} 
  t((s_0, x_0))T((s_0, x_0)(s_0, x_1)) \ldots T((s_{n-1}, x_{n-1})(s_n, x_n)) 
  = t((s_0, x_0))Lx_0x_1 \ldots Lx_{n-1}x_n, & \text{if } S_0 \xrightarrow{x_0} \ldots \xrightarrow{x_n} s_{n+1} \text{ is a path} \\
  0, & \text{otherwise}.
\end{cases}$$

The measure $P$ is ergodic, since $T$ is irreducible. Furthermore, its definition starts to look like the value of $\tilde{\pi}_{g*}\mu$ on the cylinders. Now we can make precise the statement about $\tilde{\pi}_{g*}\mu$ being piecewise-Markov. Note that the vectors $l$ and $t$ are positive, as they are stationary probability vectors of ergodic Markov chains [3]. Therefore, on the cylinders $\Omega_{g,x} = (g, x)(S \times X)^N$, we have:

$$P((g, x)w(S \times X)^N) = \frac{t(g, x)}{l(x)} \tilde{\pi}_{g*}\mu(w(S \times X)^N),$$

that is,

$$\tilde{\pi}_{g*}\mu|_{\Omega_{g,x}} = \frac{l(x)}{l(g, x)} P|_{\Omega_{g,x}}$$

and since $\tilde{\pi}_{g*}\mu$ is supported on $\Omega_{g,x}$,

$$\tilde{\pi}_{g*}\mu = \sum_x \frac{l(x)}{l(g, x)} P|_{\Omega_{g,x}}.$$

From this observation we have the following:

Lemma 6.3 $\tilde{\pi}_{g*}\mu \ll P$.

The measure we are interested in, $g*\mu$, is given as a pushforward: $g*\mu = \tilde{\lambda}_s \tilde{\pi}_{g*}\mu$.

We now set $Q := \tilde{\lambda}_s P$. After observing that $Q$ is invariant, ergodic, and $g*\mu \ll Q$, we are able to apply the Birkhoff Pointwise Ergodic theorem with the measure $Q$ to compute the frequencies, and state that the result holds almost everywhere w.r.t measure $\mu$.

Lemma 6.4 $Q$ is an invariant, ergodic measure whenever $P$ is.
Proof: These properties are carried over by projections which commute with the shift. Note that the following diagram commutes:

\[(S \times X)^N \xrightarrow{\sigma} (S \times X)^N\]

\[\downarrow \lambda \downarrow \sigma \]

\[X^N \xrightarrow{\sigma} X^N\]

Therefore,

\[Q(\sigma^{-1}S) = P(\tilde{\lambda}^{-1}(\sigma^{-1}S))\]

\[= P(\sigma^{-1}(\tilde{\lambda}^{-1}(S)))\]

\[= P(\tilde{\lambda}^{-1}(S)) \text{ since } P \text{ is invariant by assumption}\]

\[= Q(S),\]

so \(Q\) is invariant.

For ergodicity, we need to show that whenever \(\sigma^{-1}(E) = E\), \(Q(E) = 0\) or \(1\). If \(E\) is shift-invariant (\(\sigma^{-1}(E) = E\)), then so is \(\tilde{\lambda}^{-1}(E)\):

\[\tilde{\lambda}^{-1}(E) = \tilde{\lambda}^{-1}(\sigma^{-1}(E)) = \sigma^{-1}(\tilde{\lambda}^{-1}(E)),\]

where the last equality follows from the commutative diagram above. Now

\[Q(E) = P(\tilde{\lambda}^{-1}(E)) = 0 \text{ or } 1,\]

since \(P\) is ergodic, and \(\tilde{\lambda}^{-1}(E)\) is shift-invariant. □

Lemma 6.5 \(g_*\mu \ll Q\)

Proof: since \(g_*\mu = \tilde{\lambda}_*\tilde{\pi}_g\mu\), and \(Q = \tilde{\lambda}_*P\), it suffices to verify that \(\tilde{\pi}_g\mu \ll P\). But this was shown in Lemma 6.3. □

Finally we have the following:

Theorem 6.6 Let \(L\) be a stochastic matrix defining an irreducible Markov measure \(\mu\), \(A\) - an \(L\)-strongly-connected automaton with initial state \(g\), \(T_{L,A}\) - the stochastic matrix defined above. Then for all \(x \in X\) and almost all \(w \in X^N:\)

\[\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \chi_x \sigma^n(gw) = \sum_{s_0 \to x, s_1} t(s_0, y_0),\]

where \(t\) is the stationary distribution vector of \(T_{L,A}\).
Proof: We have shown the measure $Q$ to be invariant and ergodic. Following Kravchenko: by Pointwise Birkhoff Ergodic theorem, for $Q$-almost all $v \in X^N$,

$$
\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \chi_{x \sigma^n(v)} = \int_{X^N} \chi_x \, dQ
= Q(xX^N)
= (\tilde{\lambda} \ast P)(xX^N)
= \sum_{s_0 \to x} t(s_0, y_0).
$$

Let $V \subset X^N$ be the set of sequences for which the above does not hold. Since $g_*\mu \ll Q$, $g_*\mu(V) = 0$ as well. Therefore, for $g_*\mu$-almost all $v \in X^N$:

$$
\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \chi_{x \sigma^n(v)} = \sum_{s_0 \to x} t(s_0, y_0).
$$

Now we let $v = g(w)$. The above equation becomes

$$
\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \chi_{g(w) \sigma^n(w)} = \sum_{s_0 \to x} t(s_0, y_0),
$$

and it holds for $\mu$-almost all $w$ (if $W$ is the set of $w \in X^N$ for which 2 does not hold, then $W = g^{-1}(V)$, and $\mu(W) = \mu(g^{-1}(V)) = g_*\mu(V) = 0$).

The above theorem immediately generalizes to frequencies of words $w$. Indeed, replacing $x \in X$ with $u = u_0u_1 \ldots u_{k-1} \in X^k$, we obtain:

**Corollary 6.7** Let $\mu$, $g$, $T$, etc. be as in Theorem 6.6. Then

$$
\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \chi_{u \sigma^n(gw)} = \sum_{s_0 \to x} t(s_0, x_0) \prod_{j=1}^{k-1} T_{(s_{j-1}, x_{j-1})}(s_j, x_j)
$$

for $\mu$-almost-all $w \in X^N$ (where $\chi_u = \chi(uX^N)$).
Proof: As in the proof of Theorem 6.6,
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \chi_{x^k} = \int_{X^\mathbb{N}} \chi_x \, dQ
\]
\[
= Q(x^\mathbb{N})
\]
\[
= (\tilde{\lambda}_s P)(u^\mathbb{N})
\]
\[
= \sum_{s_0 \xrightarrow{x_0} s_1 \xrightarrow{x_1} \ldots \xrightarrow{x_{k-1}} s_k} t(s_0, x_0) T(s_0, x_0)(s_1, x_1) \ldots T(s_{k-2}, x_{k-2})(s_{k-1}, x_{k-1}).
\]
The rest of the proof applies without change. □

Example 6.8. Let \( X = \{0, 1\} \), \( A \) be the Aleshin automaton with initial state \( A \), and let \( L \) define a Bernoulli measure with probabilities \( (1/3, 2/3) \):
\[
L = \begin{pmatrix} 1/3 & 2/3 \\ 1/3 & 2/3 \end{pmatrix}
\]
\[
A =
\]
\[
\begin{array}{c}
C \\
A \\
B \\
\end{array}
\]
Then \( T, t \) and \( l \) are as follows:
\[
T = \begin{pmatrix}
0 & 0 & 1/3 & 2/3 & 0 & 0 \\
0 & 0 & 0 & 0 & 1/3 & 2/3 \\
0 & 0 & 0 & 0 & 1/3 & 2/3 \\
0 & 0 & 1/3 & 2/3 & 0 & 0 \\
1/3 & 2/3 & 0 & 0 & 0 & 0 \\
\end{pmatrix},
\]
\[
t = \begin{pmatrix}
1/9 & 2/9 & 1/9 & 2/9 & 1/9 & 2/9
\end{pmatrix},
\]
\[
l = \begin{pmatrix}
1/3 & 2/3
\end{pmatrix}.
\]
and the frequencies of 0 and 1 in the output are 5/9 and 4/9, respectively. △

Kravchenko observed in [1] that in the case of Bernoulli measures, the vector \( t \) can be written as \( t = r \otimes l \), where \( l \) is the 1-dimensional distribution of the measure, and \( r \) is the stationary probability vector of a chain defined by a matrix \( S \times S \to \mathbb{R} \) which depends on \( A \) and \( l \). In the example above,
\[
t = \begin{pmatrix}
1/3 & 1/3 & 1/3
\end{pmatrix} \otimes \begin{pmatrix}
1/3 & 2/3
\end{pmatrix}
\]
17
With this automaton, this will happen for all matrices $L$ generating a Markov measure.

**Example 6.9.** When $A$ is the Aleshin automaton, as above, and $L$ is a stochastic matrix, so

$$L = \begin{pmatrix} p & 1-p \\ q & 1-q \end{pmatrix},$$

the vector $t' = (q, 1-p, q, 1-p, q, 1-p)$ is an eigenvector of $T$ with eigenvalue 1, so

$$t = \frac{1}{3(1-p+q)}(1, 1, 1) \otimes (1, 1-p).$$

In general, this will not be the case.

**Example 6.10.** Let $X = \{1, 2, 3\}$, and $L, A$ (with initial state $a$) be given as follows:

$$L = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix} \quad A = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} \end{pmatrix}$$

Then $A$ is $L$-strongly-connected, and we have

$$T = \begin{pmatrix} 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}$$

$$t = \begin{pmatrix} \frac{2}{15} & \frac{2}{15} & \frac{1}{5} & \frac{1}{15} & \frac{1}{15} & \frac{2}{15} & \frac{2}{15} & \frac{1}{15} & \frac{1}{15} \end{pmatrix}$$

$$l = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}.$$
The frequencies of 0, 1 and 2 in $aw$ for $w \in X$ are $(4/15, 1/3, 2/5)$, respectively, for $\mu$-almost all $w$.

In this case, $t \neq v \otimes l$ for any vector $v$.

Note that if we modified $A$, for example, by making $\pi(a, 2) = a$, then $A$ would no longer be $L$-strongly-connected: since $L_{1,3} = \mu(13X) = 0$, this modification disconnects $(a, 1)$ and $(b, 3)$ in the graph defined by nonzero entries of $T$. In this case, $t$ is not positive:

$$t = \left( \frac{2}{9}, \frac{2}{9}, \frac{1}{3}, \frac{1}{9}, \frac{1}{9}, 0, 0, 0 \right).$$

This example shows that in the case of Markov measures, $t$ cannot depend only on $A$ and $l$, like in the case of Bernoulli measures. For a Bernoulli measure with the same 1-dimensional distribution, $t$ would be be positive.

**Singularity**

In the case of subexponential growth, $g_*\mu$ is absolutely continuous with respect to $\mu$. For tree automorphisms generated by strongly-connected automata, this is not necessarily the case. In [1], Kravchenko considers a sufficient condition for $g_*\mu$ and $\mu$ to be singular when $\mu$ is a Markov measure. We reproduce his argument, fixing a technical error.

First, let us define $K$ to be a matrix $S \times S \rightarrow \mathbb{R}$ with entries given by

$$K_{s,s'} := \sum_{\pi(s,x) = s'} l(x),$$

and let $k$ be its stationary probability vector ($kK = k; \sum k_i = 1$).

**Theorem 6.11** Suppose $\mu$ is a Bernoulli measure with probability vector $p$, and $g$ is a strongly-connected tree automorphism. Suppose that there exist $x, y \in X$ such that $p(y) = \max p$, $p(x) < p(y)$, and for some state $s$, $\lambda(s, x) = y$. Then $\mu$ and $g_*\mu$ are singular.

**Proof:** In the case of Bernoulli measures, the vector $t$ falls apart as a tensor product:

$$t = k \otimes p : t(s, x) = k(s)p(x)$$

where $p$ is the probability vector of the Bernoulli measure, and $k$ is the stationary probability vector of the Markov chain generated by $K$ defined above (see [1], Lemmas 4 and 5).
Let $x, y, s$ be as above. Then the frequency of $y$ in under $g \ast \mu$ is

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \chi_{y} \sigma^{n}(gw) = \sum_{s \in S} t(s, x') = \sum_{s \in S} q(s)p(x')$$

$$< p(y) \sum_{s \in S} q(s)$$

$$= p(y) \sum_{s \in S} q(s)$$

$$= p(y),$$

which is the frequency of $y$ under $\mu$. As a consequence of the ergodic theorem, $g \ast \mu$ and $\mu$ are singular.

\[ \square \]

If $g$ is trivial, or $p = (1/|X|, \ldots, 1/|X|)$, then $g \ast \mu = \mu$. However, it is not sufficient for $g$ to be nontrivial and $p \neq (1/|X|, \ldots, 1/|X|)$ for $g \ast \mu$ and $\mu$ to be singular.

**Example 6.12.** Let $X = \{1, 2, 3\}$, $\mu$ - a Bernoulli measure with probability vector $p = \{1/2, 1/4, 1/4\}$, and let $g$ be generated by an automaton with a single state $s$ with $\pi(s, x) = s$ for $x \in X$, $\lambda(g, 1) = 1$, $\lambda(g, 2) = 3$, $\lambda(g, 3) = 2$. Then $g \ast \mu = \mu$. \[ \triangle \]

This is a counter-example to Theorem 10 of [1], where the listed conditions were insufficient to guarantee singularity, and shows that additional conditions are needed.

In spite of Example 6, there are many cases when the vector $t$ falls apart as a tensor product. In particular, we show the following:

**Lemma 6.13** Suppose that $A$ is such that for any $s \in S$ and $x \in X$, there is $s' \in S$ such that $\pi(s', x) = s$. Then $k = \frac{1}{|S|}(1, 1, 1, \ldots, 1)$, and $t = k \otimes l$.

**Proof:** The condition forces $s'$ to be unique, by pigeonhole principle.

Therefore, if the condition holds, the columns of $K$ sum to 1, and so $(1, \ldots, 1)K = (1, 1, \ldots, 1)$. Normalizing by the sum, we obtain

$$k = \frac{1}{|S|}(1, 1, \ldots, 1).$$

To show the second part, note that the columns of $T$ then have exactly $|X|$ nonzero entries: for any $x \in X, T(s', x)(s, y) = L(x, y)$ only for the unique $s'$ such that
\[ \pi(s', x) = y. \] Then, for all \( s \in S \) and \( y \in X \),

\[
|S| k \otimes l \cdot T_{(s, y)} = (l_1, l_2, \ldots, l_{|X|}, \ldots, l_1, l_2, \ldots, l_{|X|}) \cdot T_{(s, y)} = \sum_{s' \in S} \sum_{x \in X} l_x \cdot \begin{cases} L(x, y), & \text{if } \pi(s', x) = y; \\ 0, & \text{otherwise} \end{cases} = \sum_{x \in X} l_x L(x, y).
\]

Therefore, \( k \otimes l \cdot T_{(s, y)} = \frac{1}{|S|} l_y \), and

\[
k \otimes l \cdot T = \frac{1}{|S|} (l_1, l_2, \ldots, l_{|X|}, \ldots, l_1, l_2, \ldots, l_{|S|}) = k \otimes l.
\]

Since

\[
\sum_{s \in S} \sum_{x \in X} (k \otimes l)_{(s, x)} = \sum_{s \in S} \left( \sum_{x \in X} \frac{l_x}{S} \right) = \sum_{s \in S} \frac{1}{S} = 1,
\]

\( k \otimes l \) is the stationary vector of \( T \), and the result follows. \( \Box \)

The automata for which the condition of Lemma 6.13 holds are called **reversive** automata. We thus have the following

**Corollary 6.14** Suppose \( A \) is reversive, or \( \mu \) is Bernoulli. Then \( t = k \otimes l \).

Since the proof of Theorem 6.11 follows from \( t = k \otimes l \), we obtain a more general version for Markov measures:

**Theorem 6.15** Suppose \( A \) is reversive, or \( \mu \) is Bernoulli. If \( l \neq \frac{1}{|X|} (1, 1, \ldots, 1) \), then \( \mu \) and \( g_\ast \mu \) are singular.
Examples on a 2-letter alphabet

In these examples, \( X = \{0, 1\} \), and \( L \) is in general form:

\[
L = \begin{bmatrix} 1 - p & p \\ q & 1 - q \end{bmatrix},
\]

\( l \) is the stationary vector of \( L \) (\( ll = L \) and \( \sum l_i = 1 \)), so

\[
l = \left( \frac{q}{p+q}, \frac{p}{p+q} \right),
\]

and \( \mu \) is the Markov measure induced by \( L \). \( K \) is a matrix \( S \times S \to \mathbb{R} \) defined by

\[
K_{s,s'} := \sum_{\pi(s,x)=s'} l(x)
\]

and \( k \) is its stationary probability vector (\( kk = k, \sum k_i = 1 \)).

As before, \( T \) is given by

\[
T_{(s_0,x_0)(s_1,x_1)} = \begin{cases} L(x_0, x_1), & \text{if } \pi(s_0, x_0) = s_1; \\ 0, & \text{otherwise}, \end{cases}
\]

and \( t \) is its stationary probability vector (\( tt = t, \sum t_i = 1 \)).

Finally, \( f \) is the vector of frequencies under the action of \( A \) (with any starting state):

\[
f_x = \lim_{n \to \infty} \frac{1}{n} \sum_{m=0}^{n-1} \chi_{x_{x_0}^n \sigma^m} w
\]

for \( \mu \)-almost all \( w \) (\( \sigma \) is the shift).
7.1 Aleshin

This is a generalization of Example 6 to an arbitrary Markov chain on a 2-letter alphabet. This is a reversive automaton (in fact, bireversive: its dual is reversive), and so $t = k \otimes l$.

\[
K = \begin{pmatrix}
0 & \frac{p}{p+q} & \frac{q}{p+q} \\
0 & \frac{q}{p+q} & \frac{p}{p+q} \\
1 & 0 & 0
\end{pmatrix}
\]

\[
k = \begin{pmatrix}
\frac{1}{3} & \frac{1}{3} & \frac{1}{3}
\end{pmatrix}
\]

\[
T = \begin{pmatrix}
0 & 0 & 0 & 0 & 1-p & p \\
0 & 0 & q & 1-q & 0 & 0 \\
0 & 0 & 1-p & p & 0 & 0 \\
0 & 0 & 0 & 0 & q & 1-q \\
1-p & p & 0 & 0 & 0 & 0 \\
q & 1-q & 0 & 0 & 0 & 0
\end{pmatrix}
\]

\[
t = \begin{pmatrix}
\frac{q}{3(p+q)} & \frac{p}{3(p+q)} & \frac{q}{3(p+q)} & \frac{p}{3(p+q)} & \frac{q}{3(p+q)} & \frac{p}{3(p+q)}
\end{pmatrix}
\]

\[
k \otimes l = \begin{pmatrix}
\frac{q}{3(p+q)} & \frac{p}{3(p+q)} & \frac{q}{3(p+q)} & \frac{p}{3(p+q)} & \frac{q}{3(p+q)} & \frac{p}{3(p+q)}
\end{pmatrix}
\]

\[
f = \begin{pmatrix}
\frac{2p+q}{3(p+q)} & \frac{p+2q}{3(p+q)}
\end{pmatrix}
\]

Incidence graph of $T$:
### 7.2 Bellaterra

Since Bellaterra is the product of Aleshin with an automaton that switches 0 and 1, the entries of $f$ are switched, and the rest stays the same:

$$K = \begin{pmatrix} 0 & \frac{p}{p+q} & \frac{q}{p+q} \\ \frac{p}{p+q} & \frac{p}{p+q} & 1 \\ \frac{p}{p+q} & \frac{p}{p+q} & 0 \end{pmatrix}$$

$$k = \left( \frac{1}{3} \frac{1}{3} \frac{1}{3} \right)$$

$$T = \begin{pmatrix} 0 & 0 & 0 & 1-p & p \\ 0 & 0 & q & 1-q & 0 & 0 \\ 0 & 0 & 1-p & p & 0 & 0 \\ 0 & 0 & 0 & 0 & q & 1-q \\ 1-p & p & 0 & 0 & 0 & 0 \\ q & 1-q & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$t = \left( \frac{q}{3(p+q)} \frac{p}{3(p+q)} \frac{q}{3(p+q)} \frac{p}{3(p+q)} \frac{q}{3(p+q)} \frac{p}{3(p+q)} \right)$$

$$k \otimes l = \begin{pmatrix} \frac{q}{3(p+q)} & \frac{p}{3(p+q)} & \frac{q}{3(p+q)} & \frac{p}{3(p+q)} & \frac{q}{3(p+q)} & \frac{p}{3(p+q)} \\ \frac{p+2q}{3(p+q)} & \frac{2p+q}{3(p+q)} & \frac{p+2q}{3(p+q)} & \frac{2p+q}{3(p+q)} & \frac{p+2q}{3(p+q)} & \frac{2p+q}{3(p+q)} \end{pmatrix}$$

Incidence graph of $T$: [Diagram of the incidence graph]

---

24
7.3 Lamplighter

The Lamplighter automaton is interesting in that the output frequencies don’t depend on the input frequencies. This is again a reversible automaton.

\[
K = \begin{pmatrix}
\frac{q}{p+q} & \frac{p}{p+q} \\
\frac{p}{p+q} & \frac{q}{p+q}
\end{pmatrix}
\]

\[
k = \begin{pmatrix}
\frac{1}{2} & \frac{1}{2}
\end{pmatrix}
\]

\[
T = \begin{pmatrix}
1 - p & p & 0 & 0 \\
0 & 0 & q & 1 - q \\
0 & 0 & 1 - p & p \\
q & 1 - q & 0 & 0
\end{pmatrix}
\]

\[
t = \begin{pmatrix}
\frac{q}{2(p+q)} & \frac{p}{2(p+q)} & \frac{q}{2(p+q)} & \frac{p}{2(p+q)}
\end{pmatrix}
\]

\[
k \otimes l = \begin{pmatrix}
\frac{q}{2(p+q)} & \frac{p}{2(p+q)} & \frac{q}{2(p+q)} & \frac{p}{2(p+q)}
\end{pmatrix}
\]

\[
f = \begin{pmatrix}
\frac{1}{2} & \frac{1}{2}
\end{pmatrix}
\]

Incidence graph of \(T\):

Note: even though the frequencies of individual characters don’t depend on \(p\) and \(q\), this is not the case for words of length 2. The input and output frequencies are as follows:

Input frequency: \[
\begin{pmatrix}
\frac{q-pq}{q} & \frac{pq}{p+q} & \frac{pq}{p+q} & \frac{p-pq}{q} \\
\frac{q}{2(p+q)} & \frac{p}{2(p+q)} & \frac{p}{2(p+q)} & \frac{q}{2(p+q)}
\end{pmatrix}
\]

Output frequency: \[
\begin{pmatrix}
\frac{q-pq}{q} & \frac{pq}{p+q} & \frac{pq}{p+q} & \frac{p-pq}{q} \\
\frac{q}{2(p+q)} & \frac{p}{2(p+q)} & \frac{p}{2(p+q)} & \frac{q}{2(p+q)}
\end{pmatrix}
\]
7.4 Case when $t \neq k \otimes l$

This can happen even with a 2-character alphabet. This automaton differs from Aleshin only by one arrow, but that change made the automaton non-reversive.

\[ K = \begin{pmatrix} 0 & 1 & 0 \\ 0 & \frac{q}{p+q} & \frac{p}{p+q} \\ 1 & 0 & 0 \end{pmatrix}, \quad k = \begin{pmatrix} p & \frac{p+q}{3p+q} & \frac{p}{3p+q} \\ \frac{p}{3p+q} & \frac{p+q}{3p+q} & \frac{p}{3p+q} \\ 0 & 0 & 1 \end{pmatrix}, \quad T = \begin{pmatrix} 0 & 0 & 1-p & p & 0 & 0 \\ 0 & 0 & q & 1-q & 0 & 0 \\ 0 & 0 & 1-p & p & 0 & 0 \\ 0 & 0 & 0 & 0 & q & 1-q \\ 1-p & p & 0 & 0 & 0 & 0 \\ q & 1-q & 0 & 0 & 0 & 0 \end{pmatrix} \]

\[ t = \begin{pmatrix} \frac{pq(p+q-2)}{q(p^2+2q^2-3q+3)p+(q^2-3q+3)q} & \frac{p(q^2+(p-2)q+1)}{q(p^2+2q^2-3q+3)p+(q^2-3q+3)q} & \frac{q(p^2+2q^2-3q+3)p+(q^2-3q+3)q}{q(p^2+2q^2-3q+3)p+(q^2-3q+3)q} \\ \frac{p}{q(p^2+2q^2-3q+3)p+(q^2-3q+3)q} & \frac{pq}{q(p^2+2q^2-3q+3)p+(q^2-3q+3)q} & \frac{p-pq}{q(p^2+2q^2-3q+3)p+(q^2-3q+3)q} \end{pmatrix} \]

\[ k \otimes l = \begin{pmatrix} \frac{pq}{(p+q)(3p+q)} & \frac{p^2}{(p+q)(3p+q)} & \frac{q}{3p+q} & \frac{p}{3p+q} & \frac{pq}{(p+q)(3p+q)} & \frac{p^2}{(p+q)(3p+q)} \\ \frac{p}{q(p^2+2q^2-3q+3)p+(q^2-3q+3)q} & \frac{q}{3p+q} & \frac{p}{3p+q} & \frac{pq}{(p+q)(3p+q)} & \frac{p^2}{(p+q)(3p+q)} & \frac{p^2}{(p+q)(3p+q)} \end{pmatrix} \]

Incidence graph of $T$:  

Incidence graph of $T$:  

26
8 Reversive automata

An automaton is **reversive** if it satisfies the condition of Lemma 6.13: if one can arrive to any state by any letter of the alphabet. Equivalently, that means that $X$ acts on the states of the automaton as a group (by $x \cdot g = \pi(g, x)$).

An immediate corollary of Theorem 6.6 and Lemma 6.13 is the following

**Corollary 8.1** Let $g, A, k, l, \mu$, etc. be as before. If $A$ is **reversive**, then for all $x \in X$ and almost all $w \in X^\mathbb{Z}$:

$$
\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \chi(x) \sigma^n(gw) = \sum_{s_0 \to x \to s_1} k(s_0)l(y_0).
$$

These automata act on bi-infinite sequences $w : \mathbb{Z} \to X$. To define the action, for a reversive automaton $A$, let $\hat{A}$ be the automaton obtained by reversing the arrows in $A$, i.e.

$$
\pi_A(s, x) = t \iff \pi_{\hat{A}}(t, x) = s
$$

$$
\lambda_A(s, x) = y \iff \lambda_{\hat{A}}(\pi(s, x), x) = y.
$$

This is well-defined when $A$ is reversive; we call $\hat{A}$ the reverse of $A$.

As before, we can extend $\pi_A, \lambda_A$ to $X^{-\mathbb{N}}$: for $w \in X^{-\mathbb{N}}$ and $x \in X$, define

$$
\hat{\pi}(s, wx) = \hat{\pi}(\pi_A(s, x), w);
$$

$$
\hat{\lambda}(s, wx) = \hat{\lambda}(\pi_A(s, x), w).
$$

Now we can extend $\pi = \pi_A$ to $w \in X^\mathbb{Z}$: for $w = \ldots w_{-1}w_0w_1 \ldots$, set

$$
\pi(w) := \hat{\pi}(\ldots w_{-2}w_{-1}w_0)\pi(w_0w_1w_2 \ldots).
$$

Theorem 6.6 holds in this setting as well, after being reformulated for bi-infinite sequences:

**Theorem 8.2** If $A$ is reversive, then

$$
\lim_{n \to \infty} \frac{1}{2n + 1} \left( \sum_{k=1}^{n} \chi(x) \sigma^n(gw) + \sum_{k=1}^{n} \chi(x) \sigma^{-n}(gw) \right) = \sum_{s_0 \to x \to s_1} k(s_0)l(y_0).
$$

27
9 When the pushforward is Gibbsian

A sofic measure is an image of a Markov measure under a projection. Thus, by definition, the measure $Q$ is sofic.

It is interesting to consider when $Q$ belongs to a class that lies between sofic and Markovian measures: the so-called Bowen-Gibbs measures. These are measures $\nu$ such that for some $C, P \in \mathbb{R} (C > 0)$ and all $w \in X^n$,

$$\frac{1}{C} < \frac{\nu(w_0 \ldots w_{n-1} X^n)}{\exp(-nP + \sum_{i=0}^{n-1} f(\sigma^i w))} < C.$$ 

Chazottes and Ugalde have provided in [5] a sufficient condition for an image of a Markov measure under a factor map to be Gibbsian (we follow a rephrasing in [4]):

**Theorem 9.1 (Chazotte, Ugalde)** Let $A$ and $B$ be mixing 1-step one-sided shifts of finite type. Let $\phi : A \to B$ be a one-block factor map with the following properties:

- for a 2-block $b_1b_2$ in $B \times B$, a letter $a_1 \in A$ such that $\phi(a_1) = b_1$ can be extended to a two-block $a_1a_2$ in $X \times X$ such that $\phi(a_1a_2) = b_1b_2$;

- given a periodic point $b \in B$ with period $m$ less than or equal to the number of letters appearing in $B$, any pair of letters $a_0, a_{m-1}$ mapping to $b_0, b_{m-1}$, respectively, can be extended to a word $a_0, \ldots, a_{m-1}$ of length $m$ that maps to $b_0, \ldots, b_{m-1}$.

Then $\phi_*\mu$ is a Gibbs measure.

In our setup, $A$ is the subshift of $(S \times X)^\mathbb{N}$ defined by nonzero entries of $T$, $B$ is the full shift $X^\mathbb{N}$, and $\phi = \tilde{\lambda}$.

Since $B$ is the full shift, $B$ is 1-step and mixing.

$A$ is mixing whenever the automaton $A$ is $\mu$-strongly-connected.

The map $\lambda : (S \times X)^\mathbb{N} \to X^\mathbb{N}$ is induced by the map $\lambda$, which is as a map from the alphabet $S \times X$ to alphabet $X$. Therefore, $\tilde{\lambda}$ is, in fact, a 1-block factor map.

The first condition translates to the following: for every sequence $y_1y_2$, and for every $(s_1, x_1)$ such that $\lambda(s_1, x_1) = y_1$, there exists $(s_2, x_2)$ such that:

- $\pi(s_1, x_1) = s_2$;
- $\lambda(s_2, x_2) = y_2$;
- $L(x_1, x_2) > 0$. 

28
Note that the first two conditions are satisfied whenever the automaton \( A \) is invertible by setting \( s_2 = \pi(s_1, x_1) \) and \( x_2 = \lambda_{s_2}^{-1}(y_2) \). The addition of the third condition is satisfied by making \( A \) \( \mu \)-strongly-connected.

The second condition of is satisfied trivially for binary alphabets, i.e. when \(|X| = 2\).

The sufficient conditions of [5] translate to the following corollary:

**Corollary 9.2** Let \( X \) be a binary alphabet, and let \( A \) be invertible, \( \mu \)-strongly connected automaton.

Then \( Q \) is Gibbsian.

This condition is not satisfied for non-invertible automata.

Let \( X = 0, 1 \). Consider an automaton \( A \) with four states, \( A, B, C, D \), which is a Moore machine (i.e. its output only depends on the current state), where \( A, B \) output 0, and \( C, D \) output 1, and the transition is given by \( A \rightarrow_1 B \rightarrow_0 C \rightarrow_0 D \rightarrow_0 A \).

Then the matrix in the condition of Chazottes and Ugalde has zero rows.

In [4], Yoo has shown that if \( h : X^\infty \rightarrow Y^\infty \) is a 1-block factor map (i.e. induced by a map \( h : X \rightarrow Y \)), a sufficient condition for \( h_*\mu \) to be Gibbsian is the following:

**Theorem 9.3 (Yoo)** If there is a \( k \) such that for every pair \( u, v \in X^k \) satisfying \( h(u) = h(v) \), there is \( w \in X^k \) such that \( h(u) = h(v) = h(w) \), with \( w_1 = u_1 \) and \( w_k = v_k \), then \( h_*\mu \) is Gibbsian.

In our case, we have the map \( \tilde{\lambda} : (S \times X)^N \rightarrow X^\infty \). The the condition becomes the following: for every two paths with the same output:

\[
\begin{align*}
 s_1 \rightarrow^{x_1}_{y_1} s_2 \cdots \rightarrow^{x_n}_{y_n} \\
 \hat{s}_1 \rightarrow^{\hat{x}_1}_{y_1} s_2 \cdots \rightarrow^{\hat{x}_n}_{y_n}
\end{align*}
\]

there must exist a path

\[
\begin{align*}
 s'_1 \rightarrow^{x'_1}_{y_1} s'_2 \cdots \rightarrow^{x'_n}_{y_n}
\end{align*}
\]

with \((s'_1, x'_1) = (s_1, x_1)\), and \((s'_n, x'_n) = (\hat{s}_n, \hat{x}_n)\).

If \( A \) is invertible, two paths from the same initial state differ in the first output character already. The only way the condition can be satisfied is if all paths of length \( n \) that give the same output end up in the same state (for some \( n \)).
References


30