On Mealy-Moore coding and images of Markov measures

Rostislav Grigorchuk*, Roman Kogan and Yaroslav Vorobets

Abstract

We study the images of the Markov measures under transformations generated by the Mealy automata. We find conditions under which the image measure is absolutely continuous or singular relative to the Markov measure. Also, we determine statistical properties of the image of a generic sequence.

11 references, 7 figures. UDC: 517.987. MSC: 37A50, 37B10.

Key words and phrases: Markov measure, Mealy automaton, Moore diagram, regular rooted tree endomorphism, activity growth, pushforward measure, asymptotic frequency.

1 Introduction

Finite *Mealy-type automata* (and closely related to them *Moore-type automata*) play an important role in computer science. Such automata also play remarkable role in algebra, dynamical systems, theory of random walks, spectral theory of graphs, operator algebras, holomorphic dynamics and other areas of mathematics (see for instance [6] and references therein).

The main feature of an initial deterministic automaton $\mathcal{A}_q$ with a finite input alphabet X and output alphabet Y is that it transforms finite words (strings) over X into words of the same length over Y, and this transformation can be extended to a map $\hat{\mathcal{A}}_q : X^{\mathbb{N}} \to Y^{\mathbb{N}}$ defined on the space $X^{\mathbb{N}}$ of infinite words. If the input and output alphabets coincide then one can iterate the map $\hat{\mathcal{A}}_q$ which leads to the dynamics on the space X^* of all finite words over X as well as on the space $X^{\mathbb{N}}$ of infinite words. Also, we can compose different maps of this kind, which leads to the

*The first author graciously acknowledges support from the Simons Foundation through Collaboration Grant #527814.

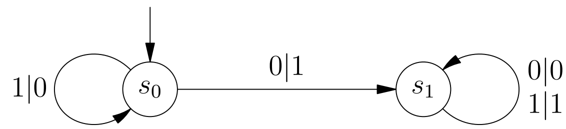


Figure 1: The odometer, also known as the adding machine

automaton semigroups or groups (in the invertible case). The space $X^{\mathbb{N}}$ is endowed with the natural product topology which makes it homeomorphic to a Cantor set. The map $\hat{\mathcal{A}}_q$ is continuous with respect to this topology and the pair $(\hat{\mathcal{A}}_q, X^{\mathbb{N}})$ is a topological dynamical system. A famous example of this sort is given by *the odometer* (see Figure 1).

The map $\hat{\mathcal{A}}_q : X^{\mathbb{N}} \to X^{\mathbb{N}}$ can be viewed as transducer that transforms the input sequence of symbols into output sequence and it can be considered as a coding map. One may be interested in what happens statistically with the sequence after such transformation. For instance, what happens with the frequency of occurrence of a fixed symbol from the alphabet? Or what happens with the probability distribution on the space of sequences if the input sequence is random?

Ryabinin [10] (see also [8], where more details are given) raised the above questions in the case of the binary alphabet when different input symbols have identical and independent distribution. In terms of ergodic theory, this means that the input sequence is generic with respect to a Bernoulli measure μ. It was observed that the output sequence is in general distributed according to a different law, just because of the change of the frequencies of symbols. Moreover, given the frequency p of symbol 1 in a μ-generic sequence, a formula was suggested for the frequency $f(p)$ of 1 in a sequence generic with respect to the image measure $(\hat{\mathcal{A}}_q)_*\mu$. The function f was suggested to be called a *stochastic function*. No justification of the formula for $f(p)$ was given. A formula with a heuristic argument was presented in the book [8].

The shift map σ acts on sequences by deleting the first symbol and hence statistical properties of a μ-generic sequence are impacted by ergodic properties of σ relative to μ. While the Bernoulli measures are shift-invariant and ergodic, their images under $\hat{\mathcal{A}}_q$ are usually not. The nature of the images was thoroughly studied by Kravchenko [7] in the case of the alphabet of arbitrary cardinality and a formula for the frequencies of symbols in the output sequences was found and justified. Moreover, he showed that in the case when automaton $\mathcal{A}_q$ has polynomially bounded activity (as defined by S. Sidki [11]), the image measure is absolutely continuous with respect to μ, while in the case of a *strongly connected* automaton it is typically singular with respect to μ. Singularity was proved by comparing the frequencies at which various symbols occur in a μ-generic input sequence and the corresponding output sequence.

This paper generalizes and extends the results of Kravchenko to the case when the measure μ belongs to a more general class of *Markov measures*. First we show that the image $(\hat{\mathcal{A}}_q)_*\mu$ is absolutely continuous with respect to μ if the automaton $\mathcal{A}_q$ is of low activity (Theorem 3.4). This result is extended to non-invertible automata as well as to transformations generated by automata with infinitely many states. We do have a restriction though. Indeed, the Markov measures can have forbidden words, when certain transitions in the corresponding Markov chain have zero probability. The automaton must not transform any allowed word into a forbidden one. Our next result is a formula that expresses the frequencies of symbols in a sequence generic with respect to the image measure (Theorem 4.7), which is then generalized to calculate frequencies of arbitrary words (Theorem 4.8). In these two theorems, the automaton is supposed to be strictly connected. The proof uses the idea of converting the Markov chain on the alphabet X of the automaton $\mathcal{A}_q$ (given by a stochastic matrix L that defines μ) into a Markov chain on the product $S \times X$ (where S is a set of states of $\mathcal{A}_q$). The latter is determined by a stochastic matrix T constructed as a kind of a skew product. This allows to represent $(\hat{\mathcal{A}}_q)_*\mu$ as composed of pieces of a measure Q that is the image under a 1-block factor map of a Markov measure on $(S \times X)^{\mathbb{N}}$. Special attention is devoted to the case when the stationary probability vector $\boldsymbol{t}$ of the matrix T decomposes as a tensor product of the stationary probability vector $\boldsymbol{l}$ of L with another probability vector $\boldsymbol{k}$. This condition holds, e.g., when μ is a Bernoulli measure or the automaton is reversible. For our last result, under assumptions that $\boldsymbol{t} = \boldsymbol{k} \otimes \boldsymbol{l}$ and the automaton is invertible and strongly connected, we prove that the Markov measure μ and its image $(\hat{\mathcal{A}}_q)_*\mu$ are either singular or the same (particular cases lead to Theorems 7.5 and 7.9).

The Bernoulli and Markov measures belong to the class of finite-state (or self-similar) measures, which was considered by the authors in [5]. This looks like a natural class for future generalizations.

Now let us discuss possible applications of the results obtained in this paper. First there is very interesting group theory related to the finite automata of Mealy type. Namely, given a non-initial invertible automaton $\mathcal{A}$, we can set any state q as initial. Hence the automaton generates several invertible transformations $\hat{\mathcal{A}}_q : X^{\mathbb{N}} \to X^{\mathbb{N}}$, which in turn generate a transformation group $\mathcal{G}(\mathcal{A})$. Groups of this kind are called automaton (or self-similar) groups [9] and they play an important role in group theory as they were used to solve a number of famous problems and find applications in many areas of mathematics [1]. The results of the present paper allow for a deeper study of such groups and their relation to the dynamics and information theory.

Secondly, if the automaton $\mathcal{A}$ is of polynomial activity, then the Markov measures are quasi-invariant with respect to the group $\mathcal{G}(\mathcal{A})$ and therefore can be used to build a Koopman type unitary representations in the Hilbert space $L^2(X^{\mathbb{N}}, \mu)$. Study of such representations was initiated by A. Dudko and the first author in [4], where it was shown that there are many pairwise disjoint representations of this type, they are irreducible and possess a number of interesting and useful properties.

2 Preliminaries

2.1 The shift and the Markov measures

Let X be a finite set consisting of more than one element. We refer to X as the *alphabet*. Elements of X are referred to as letters, symbols or characters. Let X^* denote the set of all finite strings $x_1x_2\dots x_n$ of letters from X (including the empty one $\varnothing$). We refer to them as *words* over X and write without any delimiters. Elements of X are identified with one-letter words in X^*. Let $X^{\mathbb{N}}$ denote the set of all infinite sequences (or infinite words) $\omega = \omega_1\omega_2\omega_3\dots$ over X. Given $u, w \in X^*$ and $\omega \in X^{\mathbb{N}}$, we can naturally define the concatenations $uw \in X^*$ and $u\omega \in X^{\mathbb{N}}$. Then u is called a *prefix* of the word uw and the sequence $u\omega$.

The set $X^{\mathbb{N}}$ is endowed with the product topology. The topology is generated by the *cylinders*, which are sets of the form $uX^{\mathbb{N}} = \{u\omega \mid \omega \in X^{\mathbb{N}}\}$, where $u \in X^*$. The **shift** over the alphabet X is a transformation $\sigma : X^{\mathbb{N}} \to X^{\mathbb{N}}$ given by $(\sigma(\omega))_n = \omega_{n+1}$ for all $\omega \in X^{\mathbb{N}}$ and $n \in \mathbb{N}$. The shift is continuous and non-invertible. To simplify notation, we use the same symbol σ even when dealing simultaneously with shifts over different alphabets.

Suppose μ is a Borel probability measure on $X^{\mathbb{N}}$. By Kolmogorov's theorem, μ is uniquely determined by its values on the cylinders. Conversely, for any function $f : X^* \to [0, \infty)$ satisfying $f(\varnothing) = 1$ and $\sum_{x\in X} f(wx) = f(w)$ for all $w \in X^*$, there is a (unique) Borel probability measure ν on $X^{\mathbb{N}}$ such that $\nu(wX^{\mathbb{N}}) = f(w)$ for all $w \in X^*$. The measure μ is *shift-invariant* if $\mu(\sigma^{-1}(E)) = E$ for any Borel set $E \subset X^{\mathbb{N}}$. A necessary and sufficient condition for this is that $\sum_{x\in X} \mu(xwX^{\mathbb{N}}) = \mu(wX^{\mathbb{N}})$ for all $w \in X^*$. The shift-invariant measure μ is *ergodic* if any Borel set $E \subset X^{\mathbb{N}}$ satisfying $\sigma^{-1}(E) = E$ has measure 0 or 1.

Let Y be another alphabet and $g : X^{\mathbb{N}} \to Y^{\mathbb{N}}$ be a Borel measurable map. Given a Borel probability measure μ on $X^{\mathbb{N}}$, the **pushforward** of μ by g, denoted $g_*\mu$, is a Borel probability measure on $Y^{\mathbb{N}}$ given by $g_*\mu(E) = \mu(g^{-1}(E))$ for all Borel sets $E \subset Y^{\mathbb{N}}$. If the map g intertwines the shifts on $X^{\mathbb{N}}$ and $Y^{\mathbb{N}}$, that is, $g\sigma = \sigma g$, then the pushforward measure $g_*\mu$ is shift-invariant whenever μ is shift-invariant. In the case g is continuous, it satisfies $g\sigma = \sigma g$ if and only if there exist an integer $k \geq 1$ and a function $\phi : X^k \to Y$ such that $(g(\omega))_n = \phi(\omega_n, \omega_{n+1}, \dots, \omega_{n+k-1})$ for all $\omega \in X^{\mathbb{N}}$ and $n \in \mathbb{N}$. Such a map is called a **k-block factor** map. Note that the shift itself is a 2-block factor map.

Any function $p : X \to \mathbb{R}$ can be interpreted as a vector $p = (p_x)_{x \in X}$ which coordinates are indexed by symbols in X. We use both $p(x)$ and p_x as notation for the coordinates. If the set X is naturally ordered, we can write p as a usual row vector. The vector p is a *probability vector* if $p_x \geq 0$ for all x and $\sum_x p_x = 1$. The probability vector defines a **Bernoulli measure** μ on $X^{\mathbb{N}}$ by $\mu(x_1 x_2 \dots x_n X^{\mathbb{N}}) = p_{x_1} p_{x_2} \dots p_{x_n}$ for any $x_1, x_2, \dots, x_n \in X$. Any Bernoulli measure is shift-invariant and ergodic.

Any function $L : X \times X \to \mathbb{R}$ can be interpreted as a matrix $L = (L_{xy})_{x,y \in X}$ which rows and columns are indexed by symbols in X. We use both $L(x,y)$ and L_{xy} as notation for the entries. If the set X is naturally ordered, we can write L as a usual matrix. The matrix L is *stochastic* if all entries are nonnegative and $\sum_y L_{xy} = 1$ for all y. The stochastic matrix defines a Markov chain on X such that L_{xy} is the probability of transition from x to y. The stochastic matrix L is called *irreducible* if the Markov chain is irreducible, which means that for any $x, y \in X$ we can find $x_1 = x, x_2, \dots, x_n = y$ such that $L_{x_i x_{i+1}} > 0$ for $1 \leq i \leq n-1$. Given a stochastic matrix $L = (L_{xy})_{x,y \in X}$ and a probability vector $\boldsymbol{l} = (\boldsymbol{l}_x)_{x \in X}$, we define a **Markov measure** μ on $X^{\mathbb{N}}$ by

$$\mu(x_1 x_2 x_3 \dots x_n X^{\mathbb{N}}) = \boldsymbol{l}_{x_1} L_{x_1 x_2} L_{x_2 x_3} \dots L_{x_{n-1} x_n}$$

for any $x_1, x_2, x_3, \dots, x_n \in X$. The Bernoulli measures are a particular case of the Markov measures, when each row of the matrix L coincides with $\boldsymbol{l}$. The Markov measure μ is shift-invariant if and only if $\boldsymbol{l}$ is a *stationary probability vector* of the matrix L, which means that $\boldsymbol{l}L = \boldsymbol{l}$, i.e., $\sum_x \boldsymbol{l}_x L_{xy} = \boldsymbol{l}_y$ for all y. If, additionally, L is irreducible then the Markov measure is ergodic. Moreover, the stationary probability vector of an irreducible stochastic matrix is unique and positive. For more details on Markov measures, see, e.g., [3]. In what follows we consider Markov measures that are shift-invariant but not necessarily ergodic.

Given a sequence $\omega \in X^{\mathbb{N}}$ and a letter $x \in X$, let $N(n)$ be the number of times x occurs among the first n terms of ω. The limit of $N(n)/n$ as $n \to \infty$, if it exists, yields the **asymptotic frequency** at which x occurs in the sequence ω. We denote this limit by $\text{freq}_\omega(x)$. If the limit does not exist then $\text{freq}_\omega(x)$ is not defined. Let $\chi_{xX^{\mathbb{N}}}$ be the characteristic function of the cylinder $xX^{\mathbb{N}}$. It is easy to see that

$$\text{freq}_\omega(x) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \chi_{xX^{\mathbb{N}}}(\sigma^i(\omega)).$$

Similarly, for any nonempty word $u \in X^*$ the limit

$$\mathrm{freq}_\omega(u) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \chi_{uX^{\mathbb{N}}}(\sigma^i(\omega)),$$

if it exists, yields the asymptotic frequency at which u occurs as a subword in ω (compared with other words of the same length in X^*).

Suppose μ is a Borel probability measure on $X^{\mathbb{N}}$. If μ is shift-invariant then it follows from the Birkhoff ergodic theorem that $\mathrm{freq}_\omega(u)$ is defined for μ-almost all $\omega \in X^{\mathbb{N}}$. If, additionally, μ is ergodic then $\mathrm{freq}_\omega(u) = \mu(uX^{\mathbb{N}})$ for μ-almost all ω.

2.2 The Mealy-Moore coding

In this article we consider the Mealy automata, which are the simplest type of transducers with input and output (for a detailed exposition, see [6]). By definition, a **Mealy automaton** (or simply an **automaton**) is a quadruple $\mathcal{A} = (X, S, \pi, \lambda)$ consisting of two nonempty finite sets X (the input/output alphabet) and S (the set of states), and two functions, the transition function $\pi : S \times X \to S$ and the output function $\lambda : S \times X \to X$. (One can consider a more general construction where the automaton has separate input and output alphabets X and Y; then λ takes values in Y.) These functions are naturally extended to functions on $S \times X^*$ by $\pi(s, \varnothing) = s$, $\lambda(s, \varnothing) = \varnothing$, and recursive rules

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

where $x \in X$ and $w \in X^*$. The same recursive rules allow to extend π and λ to functions on $S \times X^{\mathbb{N}}$, but this time $w \in X^{\mathbb{N}}$.

Selecting a state $g \in S$ as initial makes $\mathcal{A}$ into an *initial automaton*. The initial automaton generates transformations of X^* and of $X^{\mathbb{N}}$, both given by $w \mapsto \lambda(g, w)$ and referred to as the action of the state g or, more generally, as an **automaton transformation**. The state g is **nontrivial** if the action is nontrivial. Note that the action on X^* uniquely determines the action on $X^{\mathbb{N}}$, and vice versa. By overloading notation, we use g to denote either transformation.

All automaton transformations of $X^{\mathbb{N}}$ are continuous. An automaton with one state generates a 1-block factor map. No block factor map that is not a 1-block factor map can be generated by an automaton.

Any automaton $\mathcal{A} = (X, S, \pi, \lambda)$ can be pictured using its **Moore diagram**, which is a directed graph with labeled edges. The vertices are the states and the edges correspond to transition routes (loops and multiple edges are possible). Every edge carries a label consisting of two fields. The top (or left) field is the input letter that invokes that particular transition. The bottom (or right) field is the output letter produced during that. Hence every edge is of the form $s \underset{y}{\overset{x}{\to}} s'$ or $s \overset{x|y}{\to} s'$, where $\pi(s,x) = s'$ and $\lambda(s,x) = y$. Multiple edges can be pictured as a single edge with multiple labels. The action of a state g on X^* can be described using paths in the Moore diagram. Namely, given an input word $x_1x_2\dots x_n \in X^*$, we need to find a path of the form

$$g \underset{y_1}{\overset{x_1}{\longrightarrow}} s_1 \underset{y_2}{\overset{x_2}{\longrightarrow}} \dots \underset{y_n}{\overset{x_n}{\longrightarrow}} s_n.$$

Such a path exists and is unique. Then $g(x_1x_2\dots x_n) = \lambda(g, x_1x_2\dots x_n) = y_1y_2\dots y_n$ and $\pi(g, x_1x_2\dots x_n) = s_n$. Likewise, the action of g on $X^{\mathbb{N}}$ can be described using infinite paths.

The automaton $\mathcal{A}$ is called **strongly connected** if its Moore diagram is a strongly connected graph, which means that there is a path from any state to any other state.

The automaton $\mathcal{A} = (X, S, \pi, \lambda)$ is called **invertible** if each state acts on X by a permutation, that is, the function $\lambda(s,\cdot) : X \to X$ is invertible for any $s \in S$. Assuming this, let $\lambda'(s,x)$ be a unique letter such that $\lambda(s, \lambda'(s,x)) = x$. Also, let $\pi'(s,x) = \pi(s, \lambda'(s,x))$. Then $\mathcal{A}' = (X, S, \pi', \lambda')$ is called the **inverse automaton** of $\mathcal{A}$. In terms of the Moore diagrams, the automaton $\mathcal{A}'$ is obtained from $\mathcal{A}$ by interchanging the two fields of each label. The action of any state $s \in S$ on X^* (or on $X^{\mathbb{N}}$) generated by $\mathcal{A}'$ is the inverse of the action of the same state generated by $\mathcal{A}$. It follows that the automaton is invertible if and only if the action of each state on X^* (or on $X^{\mathbb{N}}$) is invertible. In the case the automaton is strongly connected, it is enough to know that the action of one state is invertible.

2.3 Endomorphisms of a regular rooted tree

Given an alphabet X, let $\mathcal{T}(X)$ be a graph with the vertex set X^* in which two vertices are connected by an edge if and only if one of them is obtained by adding one letter at the end of the other. Then $\mathcal{T}(X)$ is an *m-regular rooted tree*, where $m = |X|$, the number of letters in X. The root is the empty word. All words of a fixed length k form the *k-th level* of the tree as they are at distance k from the root. An invertible map $g : X^* \to X^*$ is an **automorphism** of the tree $\mathcal{T}(X)$ if it maps adjacent vertices to adjacent vertices. Any automorphism fixes the root and hence preserves each level of $\mathcal{T}(X)$. An arbitrary map $g : X^* \to X^*$ is called an **endomorphism** of the tree $\mathcal{T}(X)$ if it maps adjacent vertices to adjacent vertices and also preserves each level. An equivalent condition is that g preserves the length of any word and does not decrease the length of the longest common prefix of any two words. In particular, any automaton transformation of X^* is a tree endomorphism.

The set $X^{\mathbb{N}}$ of infinite sequences is naturally identified with the *boundary* of the rooted tree $\mathcal{T}(X)$, which consists of infinite paths without backtracking that start at the root. Consequently, any tree endomorphism $h : X^* \to X^*$ induces a unique transformation $\tilde{h} : X^{\mathbb{N}} \to X^{\mathbb{N}}$ such that $h(u)$ is a prefix of $\tilde{h}(\omega)$ whenever $u \in X^*$ is a prefix of $\omega \in X^{\mathbb{N}}$. If h is an automaton transformation, then $\tilde{h}$ is generated by the same initial automaton. Note that $\tilde{h}$ does not decrease the length of the longest common prefix of any two sequences. Moreover, any transformation of $X^{\mathbb{N}}$ with the latter property is induced by a unique tree endomorphism. In view of this, we refer to $\tilde{h}$ itself as a tree endomorphism and also as the action of h on $X^{\mathbb{N}}$.

Given a tree endomorphism $g : X^* \to X^*$, for any word $u \in X^*$ there exists a unique map $g|_u : X^* \to X^*$ such that $g(uw) = g(u)\, g|_u(w)$ for all $w \in X^*$. The map $g|_u$, which is also a tree endomorphism, is called the **restriction** (or **section**) of g by the word u. The restriction $g|_u$ describes how g acts inside a subtree of $\mathcal{T}(X)$ with the vertex set $uX^* = \{uw \mid w \in X^*\}$, which is canonically isomorphic to the entire tree. Likewise, we can define restrictions for a tree endomorphism $g : X^{\mathbb{N}} \to X^{\mathbb{N}}$ (but this time $w \in X^{\mathbb{N}}$). If a tree endomorphism g is generated by an automaton $\mathcal{A} = (X, S, \pi, \lambda)$ with initial state g, then any restriction $g|_u$ is the action of another state of the same automaton, namely, $\pi(g, u)$. In the case the automaton is strongly connected, all states are restrictions of one another.

A tree endomorphism is called **finite-state** if it has only finitely many distinct restrictions. Given a finite-state tree endomorphism $g : X^* \to X^*$, we associate to it the *automaton of restrictions* $\mathcal{A} = (X, S, \pi, \lambda)$, where $S = \{g|_w : w \in X^*\}$, $\pi(s, x) = s|_x$ and $\lambda(s, x) = s(x)$ for all $s \in S$ and $x \in X$. Then g is generated by $\mathcal{A}$ with initial state g. An arbitrary endomorphism of $\mathcal{T}(X)$ could be similarly generated by the automaton of restrictions if we allowed automata with infinitely many states (see [6]), which we do not.

3 Tree endomorphisms of low activity

Suppose a transformation $g : X^* \to X^*$ is an endomorphism of the regular rooted tree $\mathcal{T}(X)$. For any integer $n \geq 0$ let $R_g(n)$ denote the number of words $w \in X^*$ of length n such that the restriction $g|_w$ is nontrivial (i.e., not the identity map). The function R_g describes the **activity growth** of g as the length of input increases. It is not uncommon that only few (if any) restrictions of g are trivial, in which case $R_g(n)$ grows exponentially in n, i.e., $R_g(n) \geq c^n$ for some $c > 1$ and all sufficiently large n. For example, if g is generated by an automaton with no trivial state, then all restrictions are nontrivial so that $R_g(n) = |X|^n$ for any n. However in this section we are looking for transformations with much slower activity growth.

We say that the endomorphism g is of **polynomial activity growth** (or simply of **polynomial activity**) if the function $R_g(n)$ grows at most polynomially in n, that is, $R_g(n) \leq cn^\alpha$ for some $c, \alpha > 0$ and all n. Similarly, we can consider endomorphisms g of **bounded activity**, when the function R_g is bounded (they form a smaller class), and of **subexponential activity growth**, when $R_g(n) \leq c^n$ for any fixed $c > 1$ and all sufficiently large n (those form a larger class). Note that $(g|_u)|_w = g_{uw}$ for all $u, w \in X^*$. Therefore $R_{g|_u}(n) \leq R_g(n+k)$, where k is the length of u. It follows that all three classes are closed under taking restrictions.

If a restriction $g|_u$ is nontrivial, then so is the restriction of g by any prefix of u. As a consequence, $R_g(n+1) \leq |X|R_g(n)$ for all n. Conversely, if a function $f : \mathbb{N} \cup \{0\} \to \mathbb{N} \cup \{0\}$ satisfies $f(0) \leq 1$ and $f(n+1) \leq |X|f(n)$ for all $n \geq 0$, then f is the activity growth function of some tree endomorphism. Hence various tree endomorphisms exhibit a huge variety of activity growths including intermediate between polynomial and exponential. As there are only countably many finite-state tree endomorphisms, their activity growth cannot be so diverse. In fact, any finite-state endomorphism has either polynomial or exponential activity growth. There is an elegant criterion, due to Sidki [11] who introduced the notion of activity growth, that allows to distinguish between these two possibilities.

Proposition 3.1. *All tree endomorphisms generated by an automaton $\mathcal{A}$ have polynomial activity growth if and only if the Moore diagram of $\mathcal{A}$ does not admit two distinct simple cycles through any nontrivial state.*

Given a tree endomorphism $g : X^* \to X^*$, let us associate to it two sets of finite words. The set $V(g)$ consists of all $w \in X^*$ such that the restriction $g|_u$ is trivial whenever $g(u) = w$. This includes a possibility that no such words u exist. If g is invertible, then $w \in V(g)$ if and only if $g^{-1}|_w$ is trivial. The set $V_{\max}(g)$ is a subset of $V(g)$. A word $w \in V(g)$ belongs to $V_{\max}(g)$ if no word in $V(g)$ is a proper prefix of w.

In the case g is invertible, it is an automorphism of the regular rooted tree $\mathcal{T}(X)$, and so is the inverse g^{-1}. In this case, any word $w \in V(g)$ corresponds to a subtree wX^* such that the action of g^{-1} inside wX^* is trivial. Words in $V_{\max}(g)$ correspond to maximal subtrees of that kind.

Now we turn to the action of g on $X^{\mathbb{N}}$. By definition of the set $V_{\max}(g)$, the cylinders $wX^{\mathbb{N}}$, $w \in V_{\max}(g)$ are disjoint subsets of $X^{\mathbb{N}}$. Let us consider the complement

$$\Omega_g = X^{\mathbb{N}} \setminus \bigcup_{w \in V_{\max}(g)} wX^{\mathbb{N}}.$$

The size of the set Ω_g depends on the activity growth of g.

Lemma 3.2. *Suppose $g : X^{\mathbb{N}} \to X^{\mathbb{N}}$ is a finite-state tree endomorphism of polynomial activity. Then the sets Ω_g and $g^{-1}(\Omega_g)$ are at most countable.*

Proof. Let $\mathcal{A} = (X, S, \pi, \lambda)$ be the automaton of restrictions of g. Then every state of $\mathcal{A}$ is of polynomial activity. By Proposition 3.1, the Moore diagram of $\mathcal{A}$ does not admit two distinct simple cycles through any nontrivial state.

Let Ω'_g denote the set of all sequences $\omega \in X^{\mathbb{N}}$ such that the restriction of g by any prefix of ω is nontrivial. Given $\omega = x_1x_2x_3\ldots \in \Omega'_g$, consider a sequence of states $s_0, s_1, s_2, \ldots$ that the automaton $\mathcal{A}$ with initial state g goes through while processing the input ω. We have $s_0 = g$ and $s_n = \pi(s_{n-1}, x_n)$ for $n \geq 1$. Each s_n is nontrivial since $\omega \in \Omega'_g$. As there are only finitely many states, some $s \in S$ is visited infinitely often. If $s_k = s_n = s$ for some k and n, $k < n$, then $s|_{x_{k+1}x_{k+2}\ldots x_n} = s$. Let u be the shortest nonempty word in X^* such that $s|_u = s$. Since the Moore diagram of $\mathcal{A}$ does not admit two distinct simple cycles through any nontrivial state, it follows that $s|_w = s$ if and only if the word w is obtained by repeating u several times. We conclude that some tail of the sequence ω coincides with the periodic sequence $uuu\ldots$ so that ω is eventually periodic. As there are only countably many eventually periodic sequences in $X^{\mathbb{N}}$, the set Ω'_g is at most countable.

Next we show that $\Omega_g \subset g(\Omega'_g)$, which will imply that Ω_g is also at most countable. Indeed, let $\omega = x_1x_2x_3\ldots$ be in Ω_g. Then no prefix $x_1x_2\ldots x_n$ of ω belongs to $V(g)$. Hence there is a word $u^{(n)}$ of length n such that $g(u^{(n)}) = x_1x_2\ldots x_n$ and the restriction $g|_{u^{(n)}}$ is nontrivial. Since X is a finite set, we can build inductively a sequence $\omega' \in X^{\mathbb{N}}$ such that any prefix of ω' is also a prefix for infinitely many words $u^{(n)}$. If a word w occurs as a prefix for another word u, then $g(w)$ is a prefix for $g(u)$ and $g|_w$ is nontrivial whenever $g|_u$ is nontrivial. It follows that $g(\omega') = \omega$ and $\omega' \in \Omega'_g$ so that $\omega \in g(\Omega'_g)$.

Let $\omega \in \Omega_g$ and suppose ω' is a pre-image of ω under the transformation g. If ω' is not in Ω'_g then there is a prefix u of ω' such that the restriction $g|_u$ is trivial. This implies that g does not change the tail of ω' following the prefix u. Hence ω can be obtained from ω' by changing some letters in the prefix u. We conclude that any element of $g^{-1}(\Omega_g) \setminus \Omega'_g$ coincides with some element of Ω_g up to finitely many terms. Note that for any $\omega \in X^{\mathbb{N}}$ there are only countably many sequences in $X^{\mathbb{N}}$

that coincide with ω up to finitely many terms. Since the sets Ω_g and Ω'_g are at most countable, it follows that $g^{-1}(\Omega_g)$ is at most countable as well. $\square$

A Borel measure on $X^{\mathbb{N}}$ is called **non-atomic** if every one-element set has measure zero. Under the assumptions of Lemma 3.2, we have $\mu(\Omega_g) = \mu(g^{-1}(\Omega_g)) = 0$ for any non-atomic measure μ.

Lemma 3.3. *If $g : X^{\mathbb{N}} \to X^{\mathbb{N}}$ is a tree endomorphism of subexponential activity growth, then $\mu(\Omega_g) = \mu(g^{-1}(\Omega_g)) = 0$ for any non-atomic Markov measure μ on $X^{\mathbb{N}}$.*

Proof. Let μ be an arbitrary non-atomic Markov measure on $X^{\mathbb{N}}$. It is defined by a stochastic matrix L with stationary probability vector $\boldsymbol{l}$. First we need to estimate measures of cylinders. Let α_0 be the largest entry of L different from 1. Let m be the number of letters in X. We claim that $\mu(wX^{\mathbb{N}}) \leq c\alpha^n$ for any word $w \in X^*$ of length $n \geq 1$, where $c = \alpha_0^{-1-1/m}$ and $\alpha = \alpha_0^{1/m}$ (note that $\alpha < 1$). Assume the contrary: $\mu(wX^{\mathbb{N}}) > c\alpha^n$ for some $w = x_1x_2\dots x_n$, where each $x_i \in X$. The measure is given by $\mu(wX^{\mathbb{N}}) = \boldsymbol{l}_{x_1}L_{x_1x_2}\dots L_{x_{n-1}x_n}$, where no factor in the product exceeds 1. Since $c\alpha^n = \alpha_0^{(n-1)/m-1}$, the sequence $L_{x_1x_2}, L_{x_2x_3}, \dots, L_{x_{n-1}x_n}$ contains no more than $(n-1)/m - 1$ numbers different from 1. Also, $n - 1 > m$ since $\alpha_0^{(n-1)/m-1} < \mu(wX^{\mathbb{N}}) \leq 1$. It follows that the sequence admits m consecutive 1s. That is, $L_{x_ix_{i+1}} = 1$ for $k \leq i \leq k+m-1$, where $1 \leq k \leq n-m$. Then some letter $x \in X$ occurs more than once in the word $x_kx_{k+1}\dots x_{k+m}$, that is, $x_j = x_{j'} = x$ for some j and j', $k \leq j < j' \leq k+m$. Let us take the first j letters of w and append to them the word $x_{j+1}x_{j+2}\dots x_{j'}$ repeated infinitely many times. We obtain an infinite sequence $\omega = y_1y_2y_3\dots$ in $X^{\mathbb{N}}$. By construction, $y_i = x_i$ for $1 \leq i \leq j$ and $L_{y_iy_{i+1}} = 1$ for $i \geq j$. As a consequence, a cylinder $y_1y_2\dots y_iX^{\mathbb{N}}$ has the same measure $M = \boldsymbol{l}_{x_1}L_{x_1x_2}\dots L_{x_{j-1}x_j}$ for all $i \geq j$. This measure is not zero since $\mu(wX^{\mathbb{N}})$ is not zero. The cylinders $y_1y_2\dots y_iX^{\mathbb{N}}$ are nested and their intersection is $\{\omega\}$. It follows that $\mu(\{\omega\}) = M \neq 0$, which contradicts with μ being a non-atomic measure.

For any $n \geq 1$ let W'_n denote the set of all words $w \in X^*$ of length n such that the restriction $g|_w$ is nontrivial. Further, let $W_n = g(W'_n)$. For any k, $0 \leq k \leq n$, let $W_{n,k}$ be the set of all words of length n that coincide with a word in W_n up to changing some of the first k letters. The cardinality of the set W'_n is $|W'_n| = R_g(n)$. Then $|W_n| \leq |W'_n| = R_g(n)$ and $|W_{n,k}| \leq |X|^k|W_n| \leq m^kR_g(n)$.

Let Ξ'_n be the union of cylinders $wX^{\mathbb{N}}$ over all $w \in W'_n$, let Ξ_n be the union of $wX^{\mathbb{N}}$ over all $w \in W_n$, and let $\Xi_{n,k}$ be the union of $wX^{\mathbb{N}}$ over all $w \in W_{n,k}$. Since $\mu(wX^{\mathbb{N}}) \leq c\alpha^n$ for any word w of length n, it follows that $\mu(\Xi'_n) \leq c\alpha^nR_g(n)$, $\mu(\Xi_n) \leq c\alpha^nR_g(n)$ and $\mu(\Xi_{n,k}) \leq c\alpha^nm^kR_g(n)$. By assumption, $R_g(n)$ grows subexponentially in n. Since $\alpha < 1$, we conclude that $\mu(\Xi'_n)$, $\mu(\Xi_n)$ and $\mu(\Xi_{n,k})$ all tend to 0 as $n \to \infty$.

Just like in the proof of Lemma 3.2, consider the set Ω'_g of all sequences $\omega \in X^{\mathbb{N}}$ such that the restriction of g by any prefix of ω is nontrivial. Clearly, $\Omega'_g \subset \Xi'_n$ for all n. Since $\mu(\Xi'_n) \to 0$ as $n \to \infty$, the set Ω'_g has measure zero. Just like in the proof of Lemma 3.2, we can show that $\Omega_g \subset g(\Omega'_g)$. Then $\Omega_g \subset \Xi_n$ for all n. Since $\mu(\Xi_n) \to 0$ as $n \to \infty$, the set Ω_g has measure zero. Further, we can show that any pre-image under g of any $\omega \in \Omega_g$ either belongs to Ω'_g or coincides with ω up to finitely many terms. Hence any element of $g^{-1}(\Omega_g) \setminus \Omega'_g$ belongs to $\Xi_{n,k}$ for some k (depending on the element) and all n. Since $\mu(\Xi_{n,k}) \to 0$ as $n \to \infty$ for any fixed k, it follows that $\mu(g^{-1}(\Omega_g) \setminus \Omega'_g) = 0$. We already know that $\mu(\Omega'_g) = 0$. Thus $\mu(g^{-1}(\Omega_g)) = 0$. □

Lemmas 3.2 and 3.3 suggest that a tree endomorphism of slow activity growth changes only finitely many terms in a generic infinite sequence $\omega \in X^{\mathbb{N}}$. This observation leads to the following result.

Theorem 3.4. *Let μ be a non-atomic Markov measure on $X^{\mathbb{N}}$ and $g : X^{\mathbb{N}} \to X^{\mathbb{N}}$ be a tree endomorphism of subexponential activity growth. Then the measure $g_*\mu$ is absolutely continuous with respect to μ if and only if $\mu(wxX^{\mathbb{N}}) = 0$ implies $\mu(g^{-1}(wxX^{\mathbb{N}})) = 0$ for all $w \in V_{\max}(g)$ and $x \in X$. If this is the case, then the Radon-Nikodym derivative is given by*

$$\frac{dg_*\mu}{d\mu} = \sum_{\substack{w \in V_{\max}(g),\, x \in X: \\ \mu(wxX^{\mathbb{N}}) \neq 0}} \frac{\mu(g^{-1}(wxX^{\mathbb{N}}))}{\mu(wxX^{\mathbb{N}})}\, \chi_{wxX^{\mathbb{N}}}. \tag{3.1}$$

Proof. If the measure $g_*\mu$ is absolutely continuous with respect to μ, then $\mu(E) = 0$ implies $\mu(g^{-1}(E)) = g_*\mu(E) = 0$ for any measurable set $E \subset X^{\mathbb{N}}$. Hence the conditions of the theorem are clearly necessary. Now assume they hold. We need to show that $g_*\mu = D\mu$, where the function $D : X^{\mathbb{N}} \to \mathbb{R}$ is given by (3.1).

Consider any $w \in V_{\max}(g)$ and $x \in X$ such that $\mu(wxX^{\mathbb{N}}) \neq 0$. Let $\delta_{w,x} = \mu(g^{-1}(wxX^{\mathbb{N}}))/\mu(wxX^{\mathbb{N}})$. First we are going to show that $g_*\mu(C) = \delta_{w,x}\mu(C)$ for any cylinder $C \subset wxX^{\mathbb{N}}$. The cylinder C is of the form $wxw'X^{\mathbb{N}}$, where $w' \in X^*$. Let U_w be the set of all words $u \in X^*$ such that $g(u) = w$. The pre-image $g^{-1}(wX^{\mathbb{N}})$ is the disjoint union of cylinders $uX^{\mathbb{N}}$, $u \in U_w$. Since the restriction $g|_u$ is trivial for each $u \in U_w$, it follows that $g^{-1}(wxX^{\mathbb{N}})$ is the union of cylinders $uxX^{\mathbb{N}}$, $u \in U_w$, while $g^{-1}(C)$ is the union of cylinders $uxw'X^{\mathbb{N}}$, $u \in U_w$. Let $w = x_1x_2\dots x_n$ and $w' = x'_1x'_2\dots x'_k$ ($x_i, x'_j \in X$). The Markov measure μ is defined by a stochastic matrix L with stationary probability vector $\boldsymbol{l}$. We have

$$\begin{aligned} \mu(wxX^{\mathbb{N}}) &= \boldsymbol{l}_{x_1}L_{x_1x_2}\dots L_{x_{n-1}x_n}L_{x_nx}, \\ \mu(C) &= \boldsymbol{l}_{x_1}L_{x_1x_2}\dots L_{x_{n-1}x_n}L_{x_nx}L_{xx'_1}L_{x'_1x'_2}\dots L_{x'_{k-1}x'_k}, \end{aligned}$$

which implies that $\mu(C) = \mu(wxX^{\mathbb{N}})L_{xx'_1}L_{x'_1x'_2}\dots L_{x'_{k-1}x'_k}$. Similarly,

$$\mu(uxw'X^{\mathbb{N}}) = \mu(uxX^{\mathbb{N}})L_{xx'_1}L_{x'_1x'_2}\dots L_{x'_{k-1}x'_k}$$

for all words $u \in X^*$. Summing up the latter equality over $u \in U_w$, we obtain

$$\mu(g^{-1}(C)) = \mu(g^{-1}(wxX^{\mathbb{N}}))L_{xx'_1}L_{x'_1x'_2}\dots L_{x'_{k-1}x'_k}.$$

It follows that

$$g_*\mu(C) = \mu(g^{-1}(C)) = \delta_{w,x}\mu(wxX^{\mathbb{N}})L_{xx'_1}L_{x'_1x'_2}\dots L_{x'_{k-1}x'_k} = \delta_{w,x}\mu(C).$$

To prove that $g_*\mu = D\mu$, it is enough to show that the two measures agree on all cylinders. Take any cylinder $C \subset X^{\mathbb{N}}$. The set $X^{\mathbb{N}}$ is the disjoint union of Ω_g and all cylinders of the form $wxX^{\mathbb{N}}$, where $w \in V_{\max}(g)$ and $x \in X$. By definition, the function D takes a constant value on each $wxX^{\mathbb{N}}$, which is $\delta_{w,x}$ if $\mu(wxX^{\mathbb{N}}) \neq 0$ and 0 otherwise. Also, D is zero on Ω_g. It follows that

$$\int_C D(\omega)\,d\mu(\omega) = \sum_{\substack{w\in V_{\max}(g),\, x\in X:\\ \mu(wxX^{\mathbb{N}})\neq 0}} \delta_{w,x}\mu(C\cap wxX^{\mathbb{N}}).$$

Since C is a cylinder, the intersection $C \cap wxX^{\mathbb{N}}$ is either a cylinder or the empty set. By the above, $\delta_{w,x}\mu(C\cap wxX^{\mathbb{N}}) = g_*\mu(C\cap wxX^{\mathbb{N}})$. Further, if $\mu(wxX^{\mathbb{N}}) = 0$ for some $w \in V_{\max}(g)$ and $x \in X$, then $g_*\mu(wxX^{\mathbb{N}}) = 0$ by assumption. As a consequence, $g_*\mu(C\cap wxX^{\mathbb{N}}) = 0$. Finally, $g_*\mu(\Omega_g) = 0$ due to Lemma 3.3. Hence $g_*\mu(C\cap\Omega_g) = 0$. We conclude that

$$\int_C D(\omega)\,d\mu(\omega) = g_*\mu(C\cap\Omega_g) + \sum_{w\in V_{\max}(g),\, x\in X} g_*\mu(C\cap wxX^{\mathbb{N}}) = g_*\mu(C),$$

which completes the proof. □

The set $V_{\max}(g)$ is rarely finite. Therefore the conditions of Theorem 3.4 might not be easy to verify, especially if g is not invertible. We can replace them with simpler but somewhat stronger conditions.

Corollary 3.5. *Let μ be a Markov measure on $X^{\mathbb{N}}$ defined by a stochastic matrix L with stationary vector $\boldsymbol{l}$, and $g : X^{\mathbb{N}} \to X^{\mathbb{N}}$ be a tree endomorphism of subexponential activity growth. If all coordinates of $\boldsymbol{l}$ and all entries of L are positive, then the* *measure $g_*\mu$ is absolutely continuous with respect to μ, with the Radon-Nikodym derivative given by*

$$\frac{dg_*\mu}{d\mu} = \sum_{w\in V_{\max}(g),\, x\in X} \frac{\mu(g^{-1}(wxX^{\mathbb{N}}))}{\mu(wxX^{\mathbb{N}})}\,\chi_{wxX^{\mathbb{N}}}.$$

Proof. Since all coordinates of $\boldsymbol{l}$ and all entries of L are positive, every cylinder has nonzero measure. Besides, all entries of L are less than 1, which implies that the measure μ is non-atomic. It remains to apply Theorem 3.4. □

Corollary 3.6. *Let μ be a non-atomic Markov measure on $X^{\mathbb{N}}$ defined by a stochastic matrix L with stationary vector $\boldsymbol{l}$, and $g : X^{\mathbb{N}} \to X^{\mathbb{N}}$ be a tree endomorphism of polynomial activity generated by an automaton $\mathcal{A} = (X, S, \pi, \lambda)$. Suppose that $\boldsymbol{l}_x = 0$ whenever $\boldsymbol{l}_{\lambda(g,x)} = 0$ and $L_{x,y} = 0$ whenever $L_{\lambda(s,x),\,\lambda(\pi(s,x),y)} = 0$ (for all $s \in S$ and $x, y \in X$). Then the measure $g_*\mu$ is absolutely continuous with respect to μ, with the Radon-Nikodym derivative given by* (3.1).

Proof. In view of Theorem 3.4, we only need to show that $\mu(wxX^{\mathbb{N}}) = 0$ implies $\mu(g^{-1}(wxX^{\mathbb{N}})) = 0$ for all $w \in V_{\max}(g)$ and $x \in X$. We are going to show more, namely, $\mu(wX^{\mathbb{N}}) = 0$ implies $\mu(g^{-1}(wX^{\mathbb{N}})) = 0$ for all $w \in X^*$. Since the pre-image $g^{-1}(wX^{\mathbb{N}})$ is the union of cylinders $uX^{\mathbb{N}}$ over all words u such that $g(u) = w$, it is enough to show that $\mu(wX^{\mathbb{N}}) = 0$ and $g(u) = w$ implies $\mu(uX^{\mathbb{N}}) = 0$.

Suppose $w = x_1x_2 \dots x_n$ and $u = y_1y_2 \dots y_n$ are words of length $n \geq 1$ such that $g(u) = w$. We have $\mu(wX^{\mathbb{N}}) = \boldsymbol{l}_{x_1}L_{x_1x_2}\dots L_{x_{n-1}x_n}$ and $\mu(uX^{\mathbb{N}}) = \boldsymbol{l}_{y_1}L_{y_1y_2}\dots L_{y_{n-1}y_n}$. Let $s_1 = g$ and $s_i = \pi(g, y_1y_2 \dots y_{i-1})$ for $2 \leq i \leq n$. Then $\lambda(s_i, y_i) = x_i$ for $1 \leq i \leq n$ and $\pi(s_i, y_i) = s_{i+1}$ for $1 \leq i \leq n-1$. It follows that $\boldsymbol{l}_{x_1} = 0$ implies $\boldsymbol{l}_{y_1} = 0$ and $L_{x_ix_{i+1}} = 0$ implies $L_{y_iy_{i+1}} = 0$ for any i, $1 \leq i \leq n-1$. Thus $\mu(wX^{\mathbb{N}}) = 0$ implies $\mu(uX^{\mathbb{N}}) = 0$. □

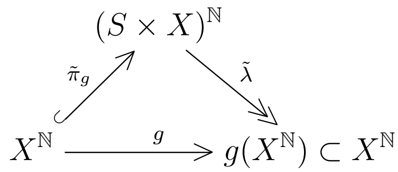
4 Strongly connected automata

Suppose μ is a shift-invariant, ergodic Markov measure on $X^{\mathbb{N}}$ defined by an irreducible stochastic matrix L with stationary vector $\boldsymbol{l}$. Let $g : X^{\mathbb{N}} \to X^{\mathbb{N}}$ be an automaton transformation and ω be a μ-generic sequence in $X^{\mathbb{N}}$. To learn about statistical properties of the sequence $g(\omega)$, we should study the pushforward measure $g_*\mu$. Unfortunately, the measure $g_*\mu$ need not be shift-invariant, let alone ergodic. On the cylinders, it is given by

$$g_*\mu(y_1y_2\dots y_nX^{\mathbb{N}}) = \sum_{g\xrightarrow[y_1]{x_1}s_1\xrightarrow[y_2]{x_2}\dots\xrightarrow[y_n]{x_n}s_n} \boldsymbol{l}(x_1)L(x_1,x_2)\dots L(x_{n-1},x_n),$$

where the sum is over all paths of the form $g \xrightarrow[y_1]{x_1} s_1 \xrightarrow[y_2]{x_2} \dots \xrightarrow[y_n]{x_n} s_n$ in the Moore diagram of the automaton $\mathcal{A} = (X, S, \pi, \lambda)$ generating g.

One way to address this difficulty is to keep track of the states $\mathcal{A}$ goes through along with the output. We are going to define maps to obtain a commutative diagram



so that $g_*\mu = \tilde{\lambda}_* \tilde{\pi}_{g*}\mu$. Then we introduce a shift-invariant measure Q on $X^{\mathbb{N}}$ that is the pushforward of a Markov measure under the 1-block factor map $\tilde{\lambda}$. Under some assumptions on the automaton $\mathcal{A}$ and the matrix L, the measure Q is ergodic while the measure $g_*\mu$ is absolutely continuous with respect to Q.

Let us begin with defining a map $\tilde{\pi} : S \times X^{\mathbb{N}} \to (S \times X)^{\mathbb{N}}$ recursively by

$$\tilde{\pi}(s, x\omega) = (s, x)\, \tilde{\pi}(\pi(s, x), \omega)$$

for all $s \in S$, $x \in X$ and $\omega \in X^{\mathbb{N}}$. Then for any state $s \in S$ we define a map $\tilde{\pi}_s : X^{\mathbb{N}} \to (S \times X)^{\mathbb{N}}$ by $\tilde{\pi}_s(\omega) = \tilde{\pi}(s, \omega)$, $\omega \in X^{\mathbb{N}}$.

Recall that g is one of the states of the automaton $\mathcal{A}$. Given a sequence $\tilde{\pi}_g(\omega) = \tilde{\pi}(g, \omega) \in (S \times X)^{\mathbb{N}}$, we can extract the output $g(\omega) = \lambda(g, \omega)$ simply by looking at the states. Hence we define a map $\tilde{\lambda} : (S \times X)^{\mathbb{N}} \to X^{\mathbb{N}}$ recursively by

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

for all $s \in S$, $x \in X$ and $\tilde{\omega} \in (S \times X)^{\mathbb{N}}$. Note that $\tilde{\lambda}$ is a 1-block factor map. Now for every infinite path

$$g \xrightarrow[y_1]{x_1} s_1 \xrightarrow[y_2]{x_2} \dots \xrightarrow[y_n]{x_n} s_n \xrightarrow[y_{n+1}]{x_{n+1}} \dots$$

in the Moore diagram of the automaton $\mathcal{A}$ we have

$$\tilde{\pi}_g(x_1 x_2 \dots x_n \dots) = (g, x_1)(s_1, x_2) \dots (s_{n-1}, x_n) \dots,$$
$$\tilde{\lambda}\big((s_0, x_1)(s_1, x_2) \dots (s_{n-1}, x_n) \dots\big) = y_1 y_2 \dots y_n \dots$$

In particular, $\tilde{\lambda}(\tilde{\pi}_g(\omega)) = g(\omega)$ for all $\omega \in X^{\mathbb{N}}$. Hence we do have the commutative diagram.

The measure $\tilde{\pi}_{g*}\mu$ on $(S \times X)^{\mathbb{N}}$ is easier to treat than the measure $g_*\mu$ on $X^{\mathbb{N}}$. On the cylinders, the former is given by

$$\tilde{\pi}_{g*}\mu\big((g, x_1)(s_1, x_2) \dots (s_{n-1}, x_n)(S \times X)^{\mathbb{N}}\big) = \boldsymbol{l}_{x_1} L_{x_1 x_2} \dots L_{x_{n-1} x_n}$$

if $g \xrightarrow[y_1]{x_1} s_1 \xrightarrow[y_2]{x_2} \dots \xrightarrow[y_n]{x_n} s_n$ is a valid path for some $y_1, y_2, \dots, y_n \in X$ and $s_n \in S$. Otherwise the cylinder has measure 0.

In a sense, the measure $\tilde{\pi}_{g*}\mu$ is "piecewise" Markov, scaled by constants on cylinders $(s,x)(S \times X)^{\mathbb{N}}$. To make this statement more precise, let us introduce a matrix $T = T_{L,\mathcal{A}}$ with rows and columns indexed by elements of $S \times X$, and entries given by

$$T_{(s_0,x_0)(s_1,x_1)} = \begin{cases} L(x_0,x_1) & \text{if } \pi(s_0,x_0) = s_1, \\ 0 & \text{otherwise.} \end{cases} \tag{4.1}$$

This matrix is stochastic. Indeed,

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

since $T_{(s,x)(r,y)} \neq 0$ for at most one choice of r, $r = \pi(s,x)$.

Let $\boldsymbol{t}$ be a stationary probability vector of the stochastic matrix T. Recall that $\boldsymbol{t}$ is a row vector which coordinates are indexed by elements of $S \times X$. All coordinates are nonnegative and add up to 1. The vector $\boldsymbol{t}$ satisfies the matrix identity $\boldsymbol{t}T = \boldsymbol{t}$. Let P denote the Markov measure on $(S \times X)^{\mathbb{N}}$ with transition matrix T and initial probability distribution $\boldsymbol{t}$. On the cylinders, the measure P is given by

$$\begin{aligned} P\big((s_0,x_0)\dots(s_n,x_n)(S \times X)^{\mathbb{N}}\big) &= \boldsymbol{t}_{(s_0,x_0)} T_{(s_0,x_0)(s_0,x_1)} \dots T_{(s_{n-1},x_{n-1})(s_n,x_n)} \\ &= \boldsymbol{t}_{(s_0,x_0)} L_{x_0x_1} \dots L_{x_{n-1}x_n} \end{aligned}$$

if $s_0 \xrightarrow[y_0]{x_0} s_1 \xrightarrow[y_1]{x_1} \dots \xrightarrow[y_n]{x_n} s_{n+1}$ is a valid path for some $y_0, y_1, \dots, y_n \in X$ and $s_{n+1} \in S$. Otherwise the cylinder has measure 0.

The measure P is shift-invariant since $\boldsymbol{t}$ is a stationary vector of T. If the Markov chain defined by the matrix T is irreducible, then the vector $\boldsymbol{t}$ is unique and positive, and the measure P is ergodic, but we do not make this assumption yet. We do know that the vector $\boldsymbol{l}$ is positive. Hence for every finite word $\tilde{w} \in (S \times X)^*$,

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

Therefore for each cylinder $\Omega_{g,x} = (g,x)(S \times X)^{\mathbb{N}}$ we obtain

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

Since the image of the map $\tilde{\pi}_g$ is contained in the union of the cylinders $\Omega_{g,x}$, $x \in X$, the measure $\tilde{\pi}_{g*}\mu$ is supported on that union. It follows that

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

provided that $\boldsymbol{t}(g,x) > 0$ for all $x \in X$. From this observation we derive the following lemma.

Lemma 4.1. *If $\boldsymbol{t}(g,x) > 0$ for all $x \in X$, then the measure $\tilde{\pi}_{g*}\mu$ is absolutely continuous with respect to P.*

Next we introduce the measure $Q = \tilde{\lambda}_* P$. Note that $g_*\mu = \tilde{\lambda}_* \tilde{\pi}_{g*}\mu$. Since $\tilde{\lambda}$ is a 1-block factor map, properties of the measures P and $\tilde{\pi}_{g*}\mu$ translate into analogous properties of Q and $g_*\mu$.

Lemma 4.2. *The measure Q is shift-invariant. It is ergodic whenever P is ergodic.*

Proof. Since $\tilde{\lambda}$ is a block factor map, it intertwines the shifts on $(S \times X)^{\mathbb{N}}$ and $X^{\mathbb{N}}$ so that we have the following commutative diagram:

$$\begin{array}{ccc} (S \times X)^{\mathbb{N}} & \xrightarrow{\sigma} & (S \times X)^{\mathbb{N}} \\ {\scriptstyle\tilde{\lambda}}\big\downarrow & & {\scriptstyle\tilde{\lambda}}\big\downarrow \\ X^{\mathbb{N}} & \xrightarrow{\quad\sigma\quad} & X^{\mathbb{N}} \end{array}$$

By construction, the measure P is shift-invariant, that is, $P(\sigma^{-1}(\widetilde{E})) = P(\widetilde{E})$ for any measurable set $\widetilde{E} \subset (S \times X)^{\mathbb{N}}$. Then for any measurable set $E \subset X^{\mathbb{N}}$,

$$Q(\sigma^{-1}(E)) = P(\tilde{\lambda}^{-1}(\sigma^{-1}(E))) = P(\sigma^{-1}(\tilde{\lambda}^{-1}(E))) = P(\tilde{\lambda}^{-1}(E)) = Q(E).$$

Hence Q is shift-invariant as well.

Now assume that P is ergodic, that is, for any measurable set $\widetilde{E} \subset (S \times X)^{\mathbb{N}}$ invariant under the shift, $\sigma^{-1}(\widetilde{E}) = \widetilde{E}$, we have $P(\widetilde{E}) = 0$ or 1. Let E be a measurable subset of $X^{\mathbb{N}}$ invariant under the shift. Then $\widetilde{E} = \tilde{\lambda}^{-1}(E)$ is also invariant under the shift and $Q(E) = P(\widetilde{E})$. Hence $Q(E) = 0$ or 1. Thus the measure Q is ergodic as well. □

Lemma 4.3. *If $\boldsymbol{t}(g,x) > 0$ for all $x \in X$, then the measure $g_*\mu$ is absolutely continuous with respect to Q.*

Proof. We need to show that $Q(E) = 0$ implies $g_*\mu(E) = 0$ for any measurable set $E \subset X^{\mathbb{N}}$. Let $\widetilde{E} = \tilde{\lambda}^{-1}(E)$. Then $P(\widetilde{E}) = Q(E) = 0$. By Lemma 4.1, the measure $\tilde{\pi}_{g*}\mu$ is absolutely continuous with respect to P. Hence $\tilde{\pi}_{g*}\mu(\widetilde{E}) = 0$. Since $g_*\mu = \tilde{\lambda}_*\tilde{\pi}_{g*}\mu$, it follows that $g_*\mu(E) = \tilde{\pi}_{g*}\mu(\widetilde{E}) = 0$. □

Now let us discuss when the Markov chain defined by the matrix T is irreducible. An obvious necessary condition is that the automaton $\mathcal{A}$ be strongly connected. If all entries of the matrix L are positive (for example, if the measure μ is Bernoulli), this condition is also sufficient. However it need not be so for a general Markov measure.

Example 4.4. Let $X = \{0, 1, 2\}$, the measure μ be defined by the matrix

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

and the automaton $\mathcal{A}$ have the transition function given by the diagram in Figure 2.

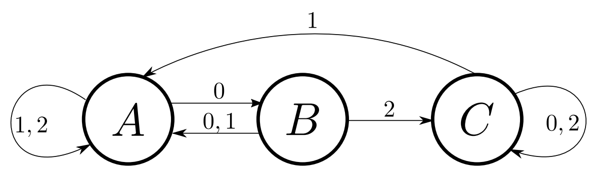


Figure 2: Any path from A to C ends in 02

Note that 02 is a forbidden word in the Markov chain defined by L, that is, $\mu(wX^{\mathbb{N}}) = 0$ whenever 02 is a subword of w. On the other hand, any input word that takes the automaton from state A to state C must end in 02. Therefore in the Markov chain defined by T there is zero chance to get from (A, x) to (C, y) in any number of steps. Thus the Markov chain is not irreducible. △

The above example motivates the following definition.

Definition 4.5. We say that the automaton $\mathcal{A}$ is **L-strongly connected** if for any pair of states $s, r \in S$ and any pair of symbols $x, y \in X$, there exists a word $w \in X^*$ such that $\pi(s, xw) = r$ and xwy is not a forbidden word in the Markov chain defined by the matrix L (that is, if $w = w_1 \dots w_n$ then L_{xw_1}, $L_{w_i w_{i+1}}$ for $1 \le i \le n-1$, and $L_{w_n y}$ are all nonzero).

Lemma 4.6. *The Markov chain defined by the matrix $T = T_{L,\mathcal{A}}$ is irreducible if and only if the automaton $\mathcal{A}$ is L-strongly connected.*

Proof. This follows directly from the definitions. □

Finally we can formulate the main results of this section.

Theorem 4.7. *Let μ be a Markov measure on $X^{\mathbb{N}}$ defined by an irreducible stochastic matrix L. Suppose a transformation $g : X^{\mathbb{N}} \to X^{\mathbb{N}}$ is generated by an automaton $\mathcal{A}$. If the automaton $\mathcal{A}$ is L-strongly connected then for any $x \in X$ and μ-almost all $\omega \in X^{\mathbb{N}}$,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \chi_{xX^{\mathbb{N}}}(\sigma^i g(\omega)) = \sum_{s_0 \xrightarrow[x]{x_0} s_1} \boldsymbol{t}(s_0, x_0),$$

where $\boldsymbol{t}$ is the stationary probability vector of the stochastic matrix $T = T_{L,\mathcal{A}}$ defined in (4.1) *and the sum is over edges in the Moore diagram of $\mathcal{A}$.*

Proof. We are going to use the measures P and Q defined above. Since the automaton $\mathcal{A}$ is L-strongly connected, the Markov chain defined by the matrix T is irreducible due to Lemma 4.6. It follows that the stationary vector $\boldsymbol{t}$ is unique and positive. Besides, the Markov measure P defined by T and $\boldsymbol{t}$ is ergodic. Then Lemma 4.2 implies that the measure Q is also shift-invariant and ergodic. By the Birkhoff ergodic theorem, for Q-almost all $v \in X^{\mathbb{N}}$ we have

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \chi_{xX^{\mathbb{N}}}(\sigma^i(v)) &= \int_{X^{\mathbb{N}}} \chi_{xX^{\mathbb{N}}}\, dQ = Q(xX^{\mathbb{N}}) = (\tilde{\lambda}_* P)(xX^{\mathbb{N}}) \\ &= \sum_{s_0 \xrightarrow[x]{x_0} s_1} P\big((s_0, x_0)(S \times X)^{\mathbb{N}}\big) = \sum_{s_0 \xrightarrow[x]{x_0} s_1} \boldsymbol{t}(s_0, x_0),\end{aligned}$$

where the last two sums are over valid edges in the Moore diagram of the automaton $\mathcal{A}$. Since all coordinates of the vector $\boldsymbol{t}$ are positive, Lemma 4.3 implies that the measure $g_*\mu$ is absolutely continuous with respect to Q. Therefore the above equality also holds for $g_*\mu$-almost all $v \in X^{\mathbb{N}}$. In other words, if $v = g(\omega)$ then the equality holds for μ-almost all $\omega \in X^{\mathbb{N}}$. □

Theorem 4.7 allows to calculate frequencies with which various symbols $x \in X$ appear in a sequence $g(\omega)$, where ω is μ-generic. A generalization to frequencies of arbitrary words over the alphabet X is straightforward.

Theorem 4.8. *Let μ be a Markov measure on $X^{\mathbb{N}}$ defined by an irreducible stochastic matrix L. Suppose a transformation $g : X^{\mathbb{N}} \to X^{\mathbb{N}}$ is generated by an automaton*

$\mathcal{A}$. If the automaton $\mathcal{A}$ is L-strongly connected then for any nonempty word $u = u_1u_2\dots u_k \in X^*$ and μ-almost all $\omega \in X^{\mathbb{N}}$,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \chi_{uX^{\mathbb{N}}}(\sigma^i g(\omega)) = \sum_{s_0 \xrightarrow[u_1]{x_0} s_1 \dots \xrightarrow[u_k]{x_{k-1}} s_k} \boldsymbol{t}_{(s_0,x_0)} L_{x_0x_1}\dots L_{x_{k-2}x_{k-1}},$$

where $\boldsymbol{t}$ is the stationary probability vector of the stochastic matrix $T = T_{L,\mathcal{A}}$ defined in (4.1) *and the sum is over paths in the Moore diagram of $\mathcal{A}$.*

The proof is completely analogous to that of Theorem 4.7 and we omit it. For examples of calculations using Theorems 4.7 and 4.8, see Section 6 below.

Suppose $K = (K_{ss'})_{s,s'\in S}$ is a stochastic matrix that defines a Markov chain on the set S. The tensor product $K \otimes L$ is an array of numbers indexed by two states $s, s' \in S$ and two symbols $x, x' \in X$, and given by $(K\otimes L)_{(s,x)(s',x')} = K_{ss'}L_{xx'}$. We regard $K\otimes L$ as a matrix which rows and columns are indexed by elements of $S\times X$, not as a 4-dimensional array. Then $K\otimes L$ is stochastic and defines a Markov chain on $S\times X$. Suppose $\boldsymbol{k}$ is a stationary probability vector of the matrix K. The tensor product $\boldsymbol{k}\otimes\boldsymbol{l}$ is an array of numbers indexed by pairs $(s,x)\in S\times X$ and given by $(\boldsymbol{k}\otimes\boldsymbol{l})_{(s,x)} = \boldsymbol{k}_s\boldsymbol{l}_x$. We regard it as a vector, not as a matrix. Then $\boldsymbol{k}\otimes\boldsymbol{l}$ is a stationary probability vector of the matrix $K\otimes L$.

Recall that the matrix $T = T_{L,\mathcal{A}}$ defines a Markov chain on $S\times X$. Unfortunately, T cannot be represented as the tensor product of L with another stochastic matrix. Nevertheless, in some cases the stationary probability vector $\boldsymbol{t}$ does decompose as the tensor product of $\boldsymbol{l}$ with another probability vector, which allows to simplify the formulas in Theorems 4.7 and 4.8.

Let us define a matrix $K = K_{\boldsymbol{l},\mathcal{A}}$ by

$$K_{s_0s_1} = \sum_{x\colon \pi(s_0,x)=s_1} \boldsymbol{l}(x) = \sum_{s_0 \xrightarrow[y]{x} s_1} \boldsymbol{l}(x) \tag{4.2}$$

for all $s_0, s_1 \in S$ (in the second formula, the sum is over valid edges in the Moore diagram of the automaton $\mathcal{A}$). For any $s\in S$ we have $\sum_r K_{sr} = \sum_x \boldsymbol{l}(x) = 1$ so that K is indeed a stochastic matrix. Since the vector $\boldsymbol{l}$ is positive, it follows that K is irreducible if and only if the automaton $\mathcal{A}$ is strongly connected.

Lemma 4.9. *Suppose $\boldsymbol{k}$ is a stationary probability vector of K. If the Markov measure defined by L is Bernoulli, then $\boldsymbol{k}\otimes\boldsymbol{l}$ is a stationary probability vector of T.*

Proof. The vector $\boldsymbol{k} = (\boldsymbol{k}_s)_{s\in S}$ satisfies $\sum_s \boldsymbol{k}_s K_{ss'} = \boldsymbol{k}_{s'}$ for all $s' \in S$. We need to show that $\sum_{s,x} \boldsymbol{k}_s \boldsymbol{l}_x T_{(s,x)(s',x')} = \boldsymbol{k}_{s'} \boldsymbol{l}_{x'}$ for all $(s', x') \in S \times X$. The Markov measure defined by L is Bernoulli if $L_{xx'} = \boldsymbol{l}_{x'}$ for all $x, x' \in X$. Then $T_{(s,x)(s',x')} = \boldsymbol{l}_{x'}$ if $\pi(s, x) = s'$ and 0 otherwise. It follows that $\sum_x \boldsymbol{l}_x T_{(s,x)(s',x')} = K_{ss'} \boldsymbol{l}_{x'}$. Consequently, $\sum_{s,x} \boldsymbol{k}_s \boldsymbol{l}_x T_{(s,x)(s',x')} = \sum_s \boldsymbol{k}_s K_{ss'} \boldsymbol{l}_{x'} = \boldsymbol{k}_{s'} \boldsymbol{l}_{x'}$. □

Combining Lemma 4.9 with Theorem 4.8, we obtain the following result (in the case of one-letter words, it was proved by Kravchenko [7]).

Theorem 4.10. *Let μ be a Bernoulli measure on $X^{\mathbb{N}}$ defined by a positive probability vector $\boldsymbol{l}$. Suppose a transformation $g : X^{\mathbb{N}} \to X^{\mathbb{N}}$ is generated by an automaton $\mathcal{A}$. If the automaton $\mathcal{A}$ is strongly connected then for any nonempty word $u = u_1 u_2 \dots u_k \in X^*$ and μ-almost all $\omega \in X^{\mathbb{N}}$,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \chi_{uX^{\mathbb{N}}}(\sigma^i g(\omega)) = \sum_{s_0 \xrightarrow[u_1]{x_0} s_1 \dots \xrightarrow[u_k]{x_{k-1}} s_k} \boldsymbol{k}_{s_0} \boldsymbol{l}_{x_0} \boldsymbol{l}_{x_1} \dots \boldsymbol{l}_{x_{k-1}},$$

where $\boldsymbol{k}$ is the stationary probability vector of the stochastic matrix $K = K_{\boldsymbol{l},\mathcal{A}}$ defined in (4.2) *and the sum is over paths in the Moore diagram of $\mathcal{A}$.*

If the Markov measure defined by L is not Bernoulli, the vector $\boldsymbol{t}$ need not decompose as $\boldsymbol{k} \otimes \boldsymbol{l}$ (see Example 6.4 below). However there is a large, well known class of automata for which Lemma 4.9 does hold for a general stochastic matrix L. We consider that class in the next section.

5 Reversible automata

Definition 5.1. An automaton $\mathcal{A} = (X, S, \pi, \lambda)$ is called **reversible** if for any $s \in S$ and any $x \in X$ there exists a unique state $s_0 \in S$ such that $\pi(s_0, x) = s$.

Suppose $\mathcal{A} = (X, S, \pi, \lambda)$ is a reversible automaton. For any $s \in S$ and any $x \in X$ let $\overleftarrow{\pi}(s, x)$ be a unique state such that $\pi\big(\overleftarrow{\pi}(s, x), x\big) = s$. Also, let $\overleftarrow{\lambda}(s, x) = \lambda\big(\overleftarrow{\pi}(s, x), x\big)$. Then $\overleftarrow{\mathcal{A}} = (S, X, \overleftarrow{\pi}, \overleftarrow{\lambda})$ is called the **reverse automaton** of $\mathcal{A}$. In terms of the Moore diagrams, the automaton $\overleftarrow{\mathcal{A}}$ is obtained from $\mathcal{A}$ by reversing all edges. That is, every edge of the form $s_0 \xrightarrow[y]{x} s_1$ is replaced by $s_1 \xrightarrow[y]{x} s_0$. The automaton $\overleftarrow{\mathcal{A}}$ is also reversible and its reverse automaton is $\mathcal{A}$.

Suppose L is a stochastic matrix that defines a Markov chain on X and $\boldsymbol{l}$ is a stationary probability vector of L.

Lemma 5.2. *If an automaton $\mathcal{A} = (X, S, \pi, \lambda)$ is reversible then the constant vector $\boldsymbol{k} = \frac{1}{|S|}(1, 1, \ldots, 1)$ is a stationary probability vector of the stochastic matrix $K = K_{\boldsymbol{l},\mathcal{A}}$ defined in* (4.2) *while $\boldsymbol{k} \otimes \boldsymbol{l}$ is a stationary probability vector of the stochastic matrix $T = T_{L,\mathcal{A}}$ defined in* (4.1).

Proof. For any $s, s' \in S$,

$$K_{\boldsymbol{l},\mathcal{A}}(s, s') = \sum_{s \xrightarrow[y]{x} s'} \boldsymbol{l}(x).$$

It follows that the transpose of the matrix $K_{\boldsymbol{l},\mathcal{A}}$ is $K_{\boldsymbol{l},\overleftarrow{\mathcal{A}}}$. As a consequence, the transpose is stochastic as well. Then $\sum_s K_{\boldsymbol{l},\mathcal{A}}(s, s') = 1$ for all $s' \in S$, which implies that $\boldsymbol{k} K_{\boldsymbol{l},\mathcal{A}} = \boldsymbol{k}$.

To prove the second statement of the lemma, it is enough to show that

$$\sum_{s,x} \boldsymbol{l}_x T_{(s,x)(s',x')} = \boldsymbol{l}_{x'}$$

for all $(s', x') \in S \times X$. Note that $T_{(s,x)(s',x')} = L_{xx'}$ if $s = \overleftarrow{\pi}(s', x)$ and 0 otherwise. It follows that $\sum_s T_{(s,x)(s',x')} = L_{xx'}$. Then $\sum_{s,x} \boldsymbol{l}_x T_{(s,x)(s',x')} = \sum_x \boldsymbol{l}_x L_{xx'} = \boldsymbol{l}_{x'}$. □

Combining Lemma 5.2 with Theorem 4.8, we obtain the following result.

Theorem 5.3. *Let μ be a Markov measure on $X^{\mathbb{N}}$ defined by an irreducible stochastic matrix L with stationary probability vector $\boldsymbol{l}$. Suppose a transformation $g : X^{\mathbb{N}} \to X^{\mathbb{N}}$ is generated by an automaton $\mathcal{A}$. If the automaton $\mathcal{A}$ is L-strongly connected and reversible, then for any nonempty word $u = u_1 u_2 \ldots u_k \in X^*$ and μ-almost all $\omega \in X^{\mathbb{N}}$,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \chi_{uX^{\mathbb{N}}}(\sigma^i g(\omega)) = \frac{1}{N} \sum_{s_0 \xrightarrow[u_1]{x_0} s_1 \ldots \xrightarrow[u_k]{x_{k-1}} s_k} \boldsymbol{l}_{x_0} L_{x_0 x_1} \ldots L_{x_{k-2} x_{k-1}},$$

where N is the number of states in $\mathcal{A}$ and the sum is over paths in the Moore diagram of $\mathcal{A}$.

A remarkable feature of the reversible automata is that their states act naturally on bi-infinite sequences over the alphabet. Suppose $\mathcal{A} = (X, S, \pi, \lambda)$ is a reversible automaton and let $w = \ldots x_{-2} x_{-1} x_0 . x_1 x_2 x_3 \ldots$ be a bi-infinite sequence in $X^{\mathbb{Z}}$ (the dot between x_0 and x_1 serves as a reference point). Given a state $g \in S$, we need to find a bi-infinite path in the Moore diagram of $\mathcal{A}$ of the form

$$\ldots \xrightarrow[y_{-2}]{x_{-2}} s_{-2} \xrightarrow[y_{-1}]{x_{-1}} s_{-1} \xrightarrow[y_0]{x_0} g \xrightarrow[y_1]{x_1} s_1 \xrightarrow[y_2]{x_2} s_2 \xrightarrow[y_3]{x_3} \ldots$$

Since the automaton $\mathcal{A}$ is reversible, such a path exists and is unique. Then, by definition, $g(w) = \dots y_{-2}y_{-1}y_0.y_1y_2y_3\dots$. Let g^+ denote the action of the same state g on $X^{\mathbb{N}}$ and g^- denote the action of g on $X^{\mathbb{N}}$ when g is regarded as a state of the reverse automaton $\overleftarrow{\mathcal{A}}$. Then $y_1y_2y_3\dots = g^+(x_1x_2x_3\dots)$ and $y_0y_{-1}y_{-2}\dots = g^-(x_0x_{-1}x_{-2}\dots)$.

The **two-sided shift** on $X^{\mathbb{Z}}$ (still denoted by σ) is defined by

$$\sigma(\dots x_{-2}x_{-1}x_0.x_1x_2x_3\dots) = \dots x_{-1}x_0x_1.x_2x_3x_4\dots$$

Unlike the shift on $X^{\mathbb{N}}$, it is invertible. Given a stochastic matrix $L = (L_{xx'})_{x,x'\in X}$ with a stationary probability vector $\boldsymbol{l} = (\boldsymbol{l}_x)_{x\in X}$, a Markov measure μ on $X^{\mathbb{Z}}$ is defined on the cylinders by

$$\mu\big(\{\dots w_{-2}w_{-1}w_0.w_1w_2\dots \mid w_i = x_i,\ m\le i\le n\}\big) = \boldsymbol{l}_{x_m}L_{x_mx_{m+1}}\dots L_{x_{n-1}x_n}$$

for any $m,n\in\mathbb{Z}$, $m\le n$ and any $x_m,x_{m+1},\dots,x_n\in X$. The measure μ is shift-invariant. It is ergodic if L is irreducible.

For any nonempty word $u = u_1u_2\dots u_k\in X^*$ consider a cylinder $\Omega_u\subset X^{\mathbb{Z}}$ defined by $\Omega_u = \{\dots w_{-2}w_{-1}w_0.w_1w_2\dots \mid w_i = u_i,\ 1\le i\le k\}$. Given $w\in X^{\mathbb{Z}}$, the limit

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}\chi_{\Omega_u}(\sigma^i(w)),$$

if it exists, yields the (asymptotic) frequency at which the word u occurs in the right-hand half of the bi-infinite sequence w. Likewise, the limit

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}\chi_{\Omega_u}(\sigma^{-i}(w)),$$

if it exists, yields the frequency at which u occurs in the left-hand half of w.

Now we can formulate an analogue of Theorem 5.3 for bi-infinite sequences.

Theorem 5.4. *Let μ be a Markov measure on $X^{\mathbb{Z}}$ defined by an irreducible stochastic matrix L with stationary probability vector $\boldsymbol{l}$. Suppose a transformation $g : X^{\mathbb{Z}}\to X^{\mathbb{Z}}$ is generated by a reversible automaton $\mathcal{A}$. If the automaton $\mathcal{A}$ is L-strongly connected then for any nonempty word $u = u_1u_2\dots u_k\in X^*$ and μ-almost all $w\in X^{\mathbb{Z}}$,*

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}\chi_{\Omega_u}(\sigma^i g(w)) &= \lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}\chi_{\Omega_u}(\sigma^{-i} g(w))\\ &= \frac{1}{N}\sum_{s_0\xrightarrow[u_1]{x_0}s_1\dots\xrightarrow[u_k]{x_{k-1}}s_k}\boldsymbol{l}_{x_0}L_{x_0x_1}\dots L_{x_{k-2}x_{k-1}},\end{aligned}$$

where N is the number of states in $\mathcal{A}$ and the sum is over paths in the Moore diagram of $\mathcal{A}$.

Proof. We are going to use the transformations g^+ and g^- defined above. Let us also define maps $F^+, F^- : X^{\mathbb{Z}} \to X^{\mathbb{N}}$ by

$$F^+(\dots w_{-2}w_{-1}w_0.w_1w_2w_3\dots) = w_1w_2w_3\dots,$$
$$F^-(\dots w_{-2}w_{-1}w_0.w_1w_2w_3\dots) = w_0w_{-1}w_{-2}\dots$$

The maps F^+ and F^- are continuous. Consider the pushforward measures $\mu^+ = F^+_*\mu$ and $\mu^- = F^-_*\mu$ on $X^{\mathbb{N}}$. The measure μ^+ is clearly the Markov measure on $X^{\mathbb{N}}$ defined by the same matrix L and vector $\boldsymbol{l}$. By Theorem 5.3,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \chi_{uX^{\mathbb{N}}}(\sigma^i g^+(\omega)) = \frac{1}{N} \sum_{s_0 \xrightarrow[u_1]{x_0} s_1 \dots \xrightarrow[u_k]{x_{k-1}} s_k} \boldsymbol{l}_{x_0} L_{x_0x_1}\dots L_{x_{k-2}x_{k-1}}$$

for μ^+-almost all $\omega \in X^{\mathbb{N}}$. In other words, if $\omega = F^+(w)$ then the latter equality holds for μ-almost all $w \in X^{\mathbb{Z}}$. Since

$$\chi_{uX^{\mathbb{N}}}(\sigma^i g^+(F^+(w))) = \chi_{uX^{\mathbb{N}}}(\sigma^i F^+(g(w))) = \chi_{uX^{\mathbb{N}}}(F^+(\sigma^i g(w))) = \chi_{\Omega_u}(\sigma^i g(w))$$

for all $w \in X^{\mathbb{Z}}$ and $i \geq 0$, this establishes the first limit in the formulation of the theorem.

The second limit requires more work. The measure μ^- is given on the cylinders by

$$\mu^-(y_1y_2\dots y_mX^{\mathbb{N}}) = \boldsymbol{l}_{y_m} L_{y_my_{m-1}}\dots L_{y_2y_1}.$$

Consider a matrix $\overleftarrow{L} = (\overleftarrow{L}_{xx'})_{x,x'\in X}$ defined by $\overleftarrow{L}_{xx'} = \boldsymbol{l}_{x'}L_{x'x}/\boldsymbol{l}_x$ for all $x, x' \in X$ (note that the vector $\boldsymbol{l}$ is positive). The matrix $\overleftarrow{L}$ is stochastic. Indeed, $\sum_{x'} \overleftarrow{L}_{xx'} = \sum_{x'} \boldsymbol{l}_{x'}L_{x'x}/\boldsymbol{l}_x = \boldsymbol{l}_x/\boldsymbol{l}_x = 1$ for all $x \in X$. Also, $\boldsymbol{l}$ is a stationary probability vector of $\overleftarrow{L}$ since $\sum_x \boldsymbol{l}_x \overleftarrow{L}_{xx'} = \sum_x \boldsymbol{l}_{x'}L_{x'x} = \boldsymbol{l}_{x'}$ for all $x' \in X$. Now for any $y_1, y_2, \dots, y_m \in X$ we obtain

$$\begin{aligned}\boldsymbol{l}_{y_1}\overleftarrow{L}_{y_1y_2}\overleftarrow{L}_{y_2y_3}\dots\overleftarrow{L}_{y_{m-1}y_m} &= \boldsymbol{l}_{y_1}(\boldsymbol{l}_{y_2}L_{y_2y_1}/\boldsymbol{l}_{y_1})(\boldsymbol{l}_{y_3}L_{y_3y_2}/\boldsymbol{l}_{y_2})\dots(\boldsymbol{l}_{y_m}L_{y_my_{m-1}}/\boldsymbol{l}_{y_{m-1}})\\ &= \boldsymbol{l}_{y_m}L_{y_my_{m-1}}\dots L_{y_3y_2}L_{y_2y_1},\end{aligned}$$

which implies that μ^- is the Markov measure defined by the matrix $\overleftarrow{L}$ with stationary probability vector $\boldsymbol{l}$.

By construction, $\overleftarrow{L}_{xx'} > 0$ if and only if $L_{x'x} > 0$. Since the stochastic matrix L is irreducible, it follows that $\overleftarrow{L}$ is irreducible as well. Next let us show that the reverse automaton $\overleftarrow{\mathcal{A}}$ is $\overleftarrow{L}$-strongly connected. Given states $s, s' \in S$ and symbols $x, x' \in X$, we need to find symbols $x_0 = x, x_1, \dots, x_m = x'$ ($m \geq 1$) such that $\overleftarrow{\pi}(s, x_0x_1 \dots x_{m-1}) = s'$ and $\overleftarrow{L}_{x_ix_{i+1}} > 0$ for $0 \leq i \leq m-1$. Let $r = \overleftarrow{\pi}(s, x)$ and $r' = \overleftarrow{\pi}(s', x')$. Since the automaton $\mathcal{A}$ is L-strongly connected, there exist symbols $y_0 = x', y_1, \dots, y_j = x$ ($j \geq 1$) such that $\pi(r', y_0y_1 \dots y_{j-1}) = r$ and $L_{y_iy_{i+1}} > 0$ for $0 \leq i \leq j-1$. Then $s = \pi(s', y_1y_2 \dots y_j)$ so that $s' = \overleftarrow{\pi}(s, y_j \dots y_2y_1)$. Moreover, $\overleftarrow{L}_{y_iy_{i-1}} > 0$ for $1 \leq i \leq j$.

Applying Theorem 5.3 to the measure μ^-, the matrix $\overleftarrow{L}$, the transformation g^-, the automaton $\overleftarrow{\mathcal{A}}$ and the word $\overleftarrow{u} = u_ku_{k-1} \dots u_1$ (which is u written backwards), we obtain that for μ^--almost all $\omega \in X^{\mathbb{N}}$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \chi_{\overleftarrow{u}X^{\mathbb{N}}}(\sigma^i g^-(\omega)) = \frac{1}{N} \sum_{\overleftarrow{\mathcal{A}}:\, s_0 \xrightarrow[u_k]{x_0} s_1 \dots \xrightarrow[u_1]{x_{k-1}} s_k} \boldsymbol{l}_{x_0} \overleftarrow{L}_{x_0x_1} \dots \overleftarrow{L}_{x_{k-2}x_{k-1}},$$

where the sum is over paths in the Moore diagram of $\overleftarrow{\mathcal{A}}$. In other words, if $\omega = F^-(w)$ then the latter equality holds for μ-almost all $w \in X^{\mathbb{Z}}$. By construction of the automaton $\overleftarrow{\mathcal{A}}$, its Moore diagram admits a path $s_0 \xrightarrow[u_k]{x_0} s_1 \dots \xrightarrow[u_1]{x_{k-1}} s_k$ if and only if the Moore diagram of $\mathcal{A}$ admits the path $s_k \xrightarrow[u_1]{x_{k-1}} \dots s_1 \xrightarrow[u_k]{x_0} s_0$. By the above, $\boldsymbol{l}_{x_0} \overleftarrow{L}_{x_0x_1} \dots \overleftarrow{L}_{x_{k-2}x_{k-1}} = \boldsymbol{l}_{x_{k-1}} L_{x_{k-1}x_{k-2}} \dots L_{x_1x_0}$ for all $x_0, x_1, \dots, x_{k-1} \in X$. It follows that the right-hand side in the last formula (that is, the value of the limit) is the same as in the formulation of the theorem. As for the left-hand side, we have $\sigma^i g^-(F^-(w)) = \sigma^i F^-(g(w)) = F^-(\sigma^{-i} g(w))$ for all $w \in X^{\mathbb{Z}}$ and $i \geq 0$. Besides, $\chi_{\Omega_u}(\sigma^{-i}(\tilde{w})) = \chi_{\overleftarrow{u}X^{\mathbb{N}}}(F^-(\sigma^{-i+k}(\tilde{w})))$ for all $\tilde{w} \in X^{\mathbb{Z}}$ and $i \in \mathbb{Z}$ (here k is the length of the word u). It follows that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \chi_{\Omega_u}(\sigma^{-i} g(w)) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \chi_{\overleftarrow{u}X^{\mathbb{N}}}(\sigma^i g^-(F^-(w)))$$

whenever the latter limit exists. This completes the proof. □

6 Examples with strongly connected automata

In this section we consider several examples of automaton transformations generated by strongly connected automata and perform for them calculations related to results of Sections 4 and 5.

In each example, an automaton $\mathcal{A} = (X, S, \pi, \lambda)$ is given by its Moore diagram. A shift-invariant, ergodic Markov measure μ on $X^{\mathbb{N}}$ is defined by an irreducible stochastic matrix L with stationary probability vector $\boldsymbol{l}$ ($\boldsymbol{l}L = \boldsymbol{l}$ and $\sum_x \boldsymbol{l}_x = 1$). In most examples, the alphabet is $X = \{0, 1\}$ and the matrix L is in general form

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

that is, $p > 0$ is the probability of transition from 0 to 1 and $q > 0$ is the probability of transition from 1 to 0. Then

$$\boldsymbol{l} = \left(\frac{q}{p+q}, \frac{p}{p+q} \right).$$

In all examples, we compute the matrix $T = T_{L,\mathcal{A}}$ defined in (4.1) and find its stationary probability vector $\boldsymbol{t}$. Rows and columns of T as well as coordinates of $\boldsymbol{t}$ are indexed by elements of $S \times X$. The sets S and X are canonically ordered as their elements are either letters or digits. We impose the lexicographic order on the set $S \times X$, which allows us to write T as a usual matrix and $\boldsymbol{t}$ as a usual row vector.

In addition, we compute the matrix K defined in (4.2) and find its stationary probability vector $\boldsymbol{k}$ to check if the vector $\boldsymbol{t}$ decomposes as the tensor product $\boldsymbol{k} \otimes \boldsymbol{l}$.

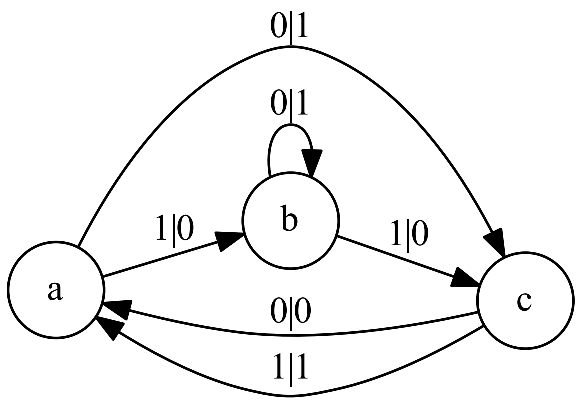
Finally, we calculate the vector $\boldsymbol{f} = (\boldsymbol{f}_x)_{x \in X}$ of frequencies of each character $x \in X$ after the action of $\mathcal{A}$ with an initial state g on a μ-generic sequence:

$$\boldsymbol{f}_x = \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \chi_{xX^{\mathbb{N}}}(\sigma^i g(\omega))$$

for μ-almost all $\omega \in X^{\mathbb{N}}$, where σ denotes the shift. If the automaton is L-strongly connected, then the vector $\boldsymbol{f}$ of output frequencies does not depend on the initial state g (as easily follows from Theorem 4.7).

6.1 Automaton generating a free group

The three states of this automaton generate a free nonabelian group. Moreover, this is essentially the only 3-state automaton over the alphabet $X = \{0, 1\}$ with that property (see [2]). What is more important for us is that the automaton is reversible (in fact, bireversible: its inverse is reversible as well), and hence $\boldsymbol{t} = \boldsymbol{k} \otimes \boldsymbol{l}$.



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

$$\boldsymbol{k} = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$$

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

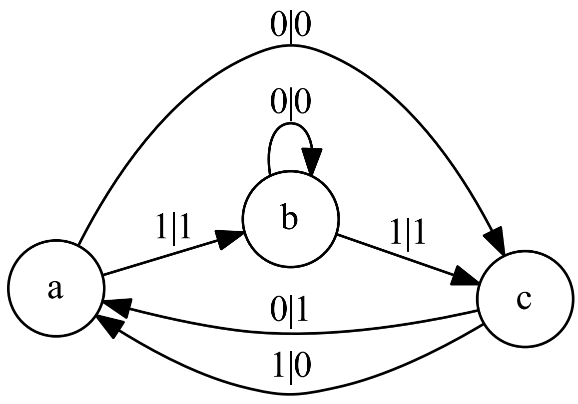
$$\boldsymbol{t} = \begin{pmatrix} \frac{q}{3(p+q)} & \frac{p}{3(p+q)} & \frac{q}{3(p+q)} & \frac{p}{3(p+q)} & \frac{q}{3(p+q)} & \frac{p}{3(p+q)} \end{pmatrix}$$

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

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

6.2 The Bellaterra automaton

The Bellaterra automaton is obtained by composing the automaton from the previous example with a one-state automaton that switches 0 and 1. In other words, all values of the transition function are retained while all values of the output function are switched. This significantly changes the character of transformations generated by the automaton: they are all involutions now (see [2]). However the action on Markov measures is not that much different: the coordinates of the vector $\boldsymbol{f}$ are interchanged while the other data remain the same.



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

$$\boldsymbol{k} = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$$

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

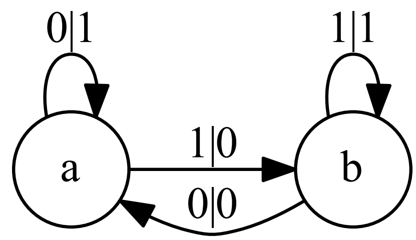
$$\boldsymbol{t} = \begin{pmatrix} \frac{q}{3(p+q)} & \frac{p}{3(p+q)} & \frac{q}{3(p+q)} & \frac{p}{3(p+q)} & \frac{q}{3(p+q)} & \frac{p}{3(p+q)} \end{pmatrix}$$

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

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

6.3 The lamplighter automaton

The two states of this automaton generate a group isomorphic to the lamplighter group $(\mathbb{Z}/2\mathbb{Z}) \wr \mathbb{Z}$ (see [6]). This is again a reversible automaton. An interesting feature of the lamplighter automaton is that the output frequencies of individual characters do not depend on the input frequencies.



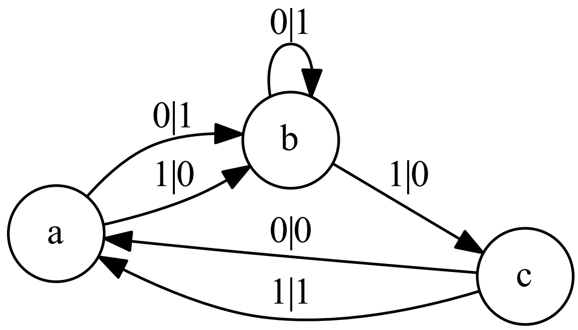
$$K = \begin{pmatrix} \frac{q}{p+q} & \frac{p}{p+q} \\ \frac{p}{p+q} & \frac{q}{p+q} \end{pmatrix}$$
$$\boldsymbol{k} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$
$$T = \begin{pmatrix} 1-p & p & 0 & 0 \\ 0 & 0 & q & 1-q \\ 0 & 0 & 1-p & p \\ q & 1-q & 0 & 0 \end{pmatrix}$$
$$\boldsymbol{t} = \begin{pmatrix} \frac{q}{2(p+q)} & \frac{p}{2(p+q)} & \frac{q}{2(p+q)} & \frac{p}{2(p+q)} \end{pmatrix}$$
$$\boldsymbol{k} \otimes \boldsymbol{l} = \begin{pmatrix} \frac{q}{2(p+q)} & \frac{p}{2(p+q)} & \frac{q}{2(p+q)} & \frac{p}{2(p+q)} \end{pmatrix}$$
$$\boldsymbol{f} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

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

Words:	00	01	10	11
Input frequency:	$\frac{q-pq}{p+q}$	$\frac{pq}{p+q}$	$\frac{pq}{p+q}$	$\frac{p-pq}{p+q}$
Output frequency:	$\frac{q}{2(p+q)}$	$\frac{p}{2(p+q)}$	$\frac{p}{2(p+q)}$	$\frac{q}{2(p+q)}$

6.4 Case when $\boldsymbol{t} \neq \boldsymbol{k} \otimes \boldsymbol{l}$

This can already happen with a two-character alphabet. The automaton in this example differs from the automaton in Example 6.1 only by one arrow (in the Moore diagram), but that change makes the automaton non-reversible.



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

$$\boldsymbol{k} = \begin{pmatrix} \frac{p}{3p+q} & \frac{p+q}{3p+q} & \frac{p}{3p+q} \end{pmatrix}$$

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

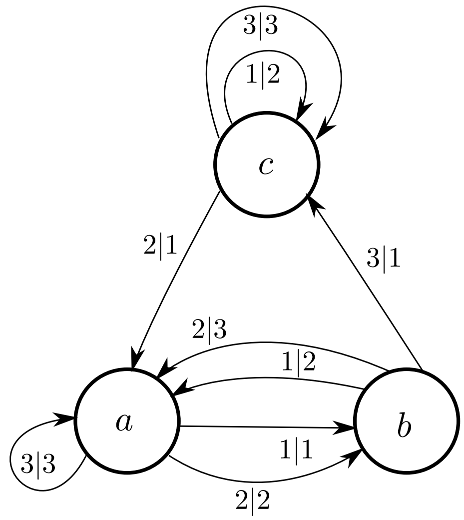
$$\boldsymbol{t} = \Big(-\frac{pq(p+q-2)}{qp^2+(2q^2-3q+3)p+(q^2-3q+3)q} \quad \frac{p\left(q^2+(p-2)q+1\right)}{qp^2+(2q^2-3q+3)p+(q^2-3q+3)q} \quad \frac{q\left(p^2+(2q-3)p+q^2-3q+3\right)}{qp^2+(2q^2-3q+3)p+(q^2-3q+3)q}$$
$$\frac{p}{qp^2+(2q^2-3q+3)p+(q^2-3q+3)q} \quad \frac{pq}{qp^2+(2q^2-3q+3)p+(q^2-3q+3)q} \quad \frac{p-pq}{qp^2+(2q^2-3q+3)p+(q^2-3q+3)q} \Big)$$

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

$$\boldsymbol{f} = \begin{pmatrix} \frac{p\left(q^2+(p-1)q+2\right)}{qp^2+(2q^2-3q+3)p+(q^2-3q+3)q} & \frac{p(q-1)^2+\left(q^2-3q+3\right)q}{qp^2+(2q^2-3q+3)p+(q^2-3q+3)q} \end{pmatrix}$$

6.5 Automaton over a three-character alphabet

In our final example, we consider an automaton $\mathcal{A}$ over a three-character alphabet $X = \{1, 2, 3\}$.



Let μ be the Markov measure on $X^{\mathbb{N}}$ defined by the matrix

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

with stationary probability vector $\boldsymbol{l} = (1/3, 1/3, 1/3)$. Then $\mathcal{A}$ is L-strongly connected and we have

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

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

$$\boldsymbol{f} = \begin{pmatrix} \frac{4}{15} & \frac{1}{3} & \frac{2}{5} \end{pmatrix}$$

We did not include the matrix K and the vector $\boldsymbol{k}$ in this example, but it is easy to see that $\boldsymbol{t} \neq \boldsymbol{v} \otimes \boldsymbol{l}$ for any vector $\boldsymbol{v}$.

Note that if we modified the automaton $\mathcal{A}$ by changing one arrow in the Moore diagram so that $\pi(a,2) = a$ (instead of b), then $\mathcal{A}$ would no longer be L-strongly-connected. Indeed, since $L_{1,3} = \mu(13X^{\mathbb{N}}) = 0$, there would be zero chance to get from $(a,1)$ and $(b,3)$ in the Markov chain defined by T. As a result, the vector $\boldsymbol{t}$ is no longer positive:

$$\boldsymbol{t} = \begin{pmatrix} \frac{2}{9} & \frac{2}{9} & \frac{1}{3} & \frac{1}{9} & \frac{1}{9} & 0 & 0 & 0 & 0 \end{pmatrix}.$$

This modification shows that in the case of Markov measures, $\boldsymbol{t}$ may not be uniquely determined by the automaton and the vector $\boldsymbol{l}$, like in the case of Bernoulli measures. Indeed, for a Bernoulli measure with the same probability distribution $\boldsymbol{l}$, the vector $\boldsymbol{t}$ would be positive.

7 Singularity

Suppose μ is a Markov measure on $X^{\mathbb{N}}$ and g is an automaton transformation of $X^{\mathbb{N}}$. If g has polynomial activity growth, the results of Section 3 suggest that we should expect the pushforward measure $g_*\mu$ to be absolutely continuous with respect to μ. In this section we study the relation between the measures μ and $g_*\mu$ in the case when g is generated by a strongly connected automaton. The relation turns out to be quite different, namely, we should expect μ and $g_*\mu$ to be singular (that is, concentrated on disjoint sets).

Kravchenko observed in [7] that if μ is a Bernoulli measure and the transformation g generated by a strongly connected automaton is invertible, then μ and $g_*\mu$ are singular except for a few cases, in which $g_*\mu = \mu$. We are going to correct his result fixing a minor error in the argument, and then further extend it.

One obvious exception is when g acts trivially. The second exception is when μ is the uniform Bernoulli measure (defined by a constant probability vector). Such a measure is preserved by any invertible automaton transformation. Unfortunately, another exceptional case (that kind of combines the said two) was overlooked in [7].

Example 7.1. Let $X = \{1,2,3\}$ and μ be a Bernoulli measure on $X^{\mathbb{N}}$ defined by a probability vector $\boldsymbol{l} = (1/2, 1/4, 1/4)$. Let $\mathcal{A} = (X, \{g\}, \pi, \lambda)$, where $\pi(g,x) = g$ for all $x \in X$, $\lambda(g,1) = 1$, $\lambda(g,2) = 3$ and $\lambda(g,3) = 2$. The only state g of the automaton $\mathcal{A}$ acts on $X^{\mathbb{N}}$ as a 1-block factor map that applies the transposition $(2\,3)$ to every term of a sequence. Since $\boldsymbol{l}_2 = \boldsymbol{l}_3$, we have $g_*\mu = \mu$ even though g does not act trivially and the measure μ is not uniform. $\triangle$

Lemma 7.2. *Let μ be a Bernoulli measure on $X^{\mathbb{N}}$ defined by a positive probability vector $\boldsymbol{l}$ and $g : X^{\mathbb{N}} \to X^{\mathbb{N}}$ be an invertible transformation generated by a strongly connected automaton $\mathcal{A} = (X, S, \pi, \lambda)$. Then the following conditions are equivalent: (i) $g_*\mu = \mu$; (ii) $\boldsymbol{l}_{x'} = \boldsymbol{l}_x$ whenever $\lambda(s, x) = x'$ for some $s \in S$.*

Proof. The action of g on X^* is invertible as well. We use g^{-1} to denote the inverse of both the action of g on $X^{\mathbb{N}}$ and on X^*. Consider an arbitrary word $u \in X^*$ of length $k \geq 1$. We have $u = x_1x_2 \dots x_k$ and $g^{-1}(u) = y_1y_2 \dots y_k$ for some $x_i, y_i \in X$, $1 \leq i \leq k$. Then $\mu(uX^{\mathbb{N}}) = \boldsymbol{l}_{x_1}\boldsymbol{l}_{x_2} \dots \boldsymbol{l}_{x_k}$ and $g_*\mu(uX^{\mathbb{N}}) = \mu(g^{-1}(uX^{\mathbb{N}})) = \mu\big(g^{-1}(u)X^{\mathbb{N}}\big) = \boldsymbol{l}_{y_1}\boldsymbol{l}_{y_2} \dots \boldsymbol{l}_{y_k}$. Note that $x_i = \lambda(s_i, y_i)$, where $s_1 = g$ and $s_i = \pi(g, y_1y_2 \dots y_{i-1})$ for $2 \leq i \leq k$. Assuming the condition (ii) holds, we obtain that $\boldsymbol{l}_{x_i} = \boldsymbol{l}_{y_i}$ for $1 \leq i \leq k$. Then $\mu(uX^{\mathbb{N}}) = g_*\mu(uX^{\mathbb{N}})$. Thus the measures μ and $g_*\mu$ coincide on the cylinders, which implies that $g_*\mu = \mu$.

Conversely, assume that $g_*\mu = \mu$ and suppose $\lambda(s, x) = x'$ for some $s \in S$. Since the automaton $\mathcal{A}$ is strongly connected, there exists a word $u \in X^*$ such that $\pi(g, u) = s$. Let $u' = g(u)$. Then $g(ux) = u'x'$. As a consequence, $\mu(uX^{\mathbb{N}}) = g_*\mu(u'X^{\mathbb{N}}) = \mu(u'X^{\mathbb{N}})$ and $\mu(uxX^{\mathbb{N}}) = g_*\mu(u'x'X^{\mathbb{N}}) = \mu(u'x'X^{\mathbb{N}})$. By definition of the measure μ, we have $\mu(uxX^{\mathbb{N}}) = \mu(uX^{\mathbb{N}})\boldsymbol{l}_x$ and $\mu(u'x'X^{\mathbb{N}}) = \mu(u'X^{\mathbb{N}})\boldsymbol{l}_{x'}$. Since $\boldsymbol{l}$ is a positive vector, the measure $\mu(uX^{\mathbb{N}}) = \mu(u'X^{\mathbb{N}})$ is not zero. It follows that $\boldsymbol{l}_{x'} = \boldsymbol{l}_x$. □

The main idea behind the proof of singularity is rather simple. Recall that the asymptotic frequency $\mathrm{freq}_\omega(u)$ with which a finite word $u \in X^*$ occurs in an infinite sequence $\omega \in X^{\mathbb{N}}$ is defined as a limit

$$\mathrm{freq}_\omega(u) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \chi_{uX^{\mathbb{N}}}(\sigma^i(\omega))$$

(it is not defined if the limit does not exist).

Lemma 7.3. *Let μ be a Borel probability measure on $X^{\mathbb{N}}$ that is invariant and ergodic with respect to the shift. Let $g : X^{\mathbb{N}} \to X^{\mathbb{N}}$ be a Borel measurable map. Suppose that $\mathrm{freq}_{g(\omega)}(u) \neq \mathrm{freq}_\omega(u)$ for some $u \in X^*$ and μ-almost all $\omega \in X^{\mathbb{N}}$. Then the measures μ and $g_*\mu$ are singular.*

Proof. Let E_1 be the set of all sequences $\omega \in X^{\mathbb{N}}$ such that $\mathrm{freq}_\omega(u) = \mu(uX^{\mathbb{N}})$. Let E_2 be the set of all $\omega \in X^{\mathbb{N}}$ such that $\mathrm{freq}_{g(\omega)}(u) \neq \mathrm{freq}_\omega(u)$. Both E_1 and E_2 are Borel measurable sets. We have $\mu(E_2) = 1$ by assumption and $\mu(E_1) = 1$ due to the Birkhoff ergodic theorem. As a consequence, $\mu(E_1 \cap E_2) = 1$. The image $g(E_1 \cap E_2)$ is clearly disjoint from E_1. It follows that $g_*\mu(X^{\mathbb{N}} \setminus E_1) \geq \mu(E_1 \cap E_2) = 1$. Hence

E_1 is a set of full measure for μ while $X^{\mathbb{N}} \setminus E_1$ is a set of full measure for $g_*\mu$. Thus μ and $g_*\mu$ are singular measures. $\square$

The next lemma is crucial for this section.

Lemma 7.4. *Let μ be a Markov measure on $X^{\mathbb{N}}$ defined by an irreducible stochastic matrix L with stationary probability vector $\boldsymbol{l}$ and $g : X^{\mathbb{N}} \to X^{\mathbb{N}}$ be an invertible transformation generated by an L-strongly connected automaton $\mathcal{A}$. Suppose that $\boldsymbol{t} = \boldsymbol{k} \otimes \boldsymbol{l}$, where $\boldsymbol{t}$ is the stationary probability vector of the matrix $T_{L,\mathcal{A}}$ defined in* (4.1) *and $\boldsymbol{k}$ is the stationary probability vector of the matrix $K_{\boldsymbol{l},\mathcal{A}}$ defined in* (4.2). *Then the measures μ and $g_*\mu$ are either singular or the same.*

Proof. Since g is invertible, all restriction of g are invertible as well. Since the automaton $\mathcal{A}$ is strongly connected, every state $s \in S$ is a restriction of g. Both the action of s on $X^{\mathbb{N}}$ and on X^* are invertible. We denote by s^{-1} the inverses of both actions.

Assume $g_*\mu \neq \mu$. Then $g_*\mu(wX^{\mathbb{N}}) \neq \mu(wX^{\mathbb{N}})$ for some nonempty word $w \in X^*$. Note that $g_*\mu(wX^{\mathbb{N}}) = \mu(g^{-1}(wX^{\mathbb{N}})) = \mu\big(g^{-1}(w)X^{\mathbb{N}}\big)$. Let k denote the length of w. Consider all words $u \in X^*$ of length k such that $\mu\big(s^{-1}(u)X^{\mathbb{N}}\big) \neq \mu(uX^{\mathbb{N}})$ for some $s \in S$ (one such word is w) and choose among them one with the largest value of $\mu(uX^{\mathbb{N}})$. We claim that $\mu\big(s^{-1}(u)X^{\mathbb{N}}\big) \leq \mu(uX^{\mathbb{N}})$ for all $s \in S$ (by the choice of u, at least one of these inequalities is going to be strict). Indeed, take any $s \in S$ and let $u^{(0)} = u$, $u^{(1)}, u^{(2)}, \dots$ be a sequence of words such that $u^{(n+1)} = s^{-1}(u^{(n)})$ for all $n \geq 0$. Since the state s acts as a permutation on the finite set of all words of length k, it follows that the sequence is periodic. If $\mu(u^{(n)}X^{\mathbb{N}}) > \mu(uX^{\mathbb{N}})$ for some n, then $\mu(u^{(n+1)}X^{\mathbb{N}}) = \mu(u^{(n)}X^{\mathbb{N}})$ due to the choice of u. Therefore $\mu(u^{(1)}X^{\mathbb{N}}) > \mu(uX^{\mathbb{N}})$ would imply $\mu(u^{(n)}X^{\mathbb{N}}) = \mu(u^{(1)}X^{\mathbb{N}}) > \mu(uX^{\mathbb{N}})$ for all $n \geq 1$, which is not the case as u occurs infinitely often in the sequence.

Let $u = u_1u_2\dots u_k$, where each $u_i \in X$. By Theorem 4.8, for μ-almost all $\omega \in X^{\mathbb{N}}$ we have

$$\operatorname{freq}_{g(\omega)}(u) = \sum_{s_0 \xrightarrow[u_1]{x_0} s_1 \dots \xrightarrow[u_k]{x_{k-1}} s_k} \boldsymbol{t}_{(s_0,x_0)} L_{x_0x_1} \dots L_{x_{k-2}x_{k-1}},$$

where the sum is over paths in the Moore diagram of $\mathcal{A}$. Let Σ denote the value of the sum. Since $\boldsymbol{t} = \boldsymbol{k} \otimes \boldsymbol{l}$, we have

$$\boldsymbol{t}_{(s_0,x_0)} L_{x_0x_1} \dots L_{x_{k-2}x_{k-1}} = \boldsymbol{k}_{s_0} \boldsymbol{l}_{x_0} L_{x_0x_1} \dots L_{x_{k-2}x_{k-1}} = \boldsymbol{k}_{s_0} \mu(x_0x_1 \dots x_{k-1}X^{\mathbb{N}}).$$

For any choice of s_0 the Moore diagram of $\mathcal{A}$ admits a unique path of the form $s_0 \xrightarrow[u_1]{x_0} s_1 \dots \xrightarrow[u_k]{x_{k-1}} s_k$, with $x_0x_1\dots x_{k-1} = s_0^{-1}(u)$. It follows that

$$\Sigma = \sum_{s\in S} \boldsymbol{k}_s \mu\big(s^{-1}(u)X^{\mathbb{N}}\big).$$

Since the automaton $\mathcal{A}$ is strongly connected, the stochastic matrix $K_{\boldsymbol{l},\mathcal{A}}$ is irreducible. Therefore the vector $\boldsymbol{k}$ is positive. By the above, $\mu\big(s^{-1}(u)X^{\mathbb{N}}\big) \le \mu(uX^{\mathbb{N}})$ for all $s \in S$. Moreover, at least one of these inequalities is strict. It follows that $\Sigma < \sum_s \boldsymbol{k}_s \mu(uX^{\mathbb{N}}) = \mu(uX^{\mathbb{N}})$. In particular, $\mathrm{freq}_{g(\omega)}(u) < \mu(uX^{\mathbb{N}})$ for μ-almost all $\omega \in X^{\mathbb{N}}$. On the other hand, $\mathrm{freq}_{\omega}(u) = \mu(uX^{\mathbb{N}})$ for μ-almost all $\omega \in X^{\mathbb{N}}$ due to the Birkhoff ergodic theorem. Now Lemma 7.3 implies that the measures μ and $g_*\mu$ are singular. □

Theorem 7.5. *Let μ be a Bernoulli measure on $X^{\mathbb{N}}$ defined by a positive probability vector $\boldsymbol{l}$. Suppose $g : X^{\mathbb{N}} \to X^{\mathbb{N}}$ is an invertible transformation generated by a strongly connected automaton $\mathcal{A} = (X, S, \pi, \lambda)$. Then the measures μ and $g_*\mu$ are singular unless $\boldsymbol{l}_{\lambda(s,x)} = \boldsymbol{l}_x$ for all $s \in S$ and $x \in X$, in which case $g_*\mu = \mu$.*

Proof. The measure μ can be regarded as a Markov measure defined by a stochastic matrix L each row of which coincides with $\boldsymbol{l}$. Since all entries of L are positive, the automaton $\mathcal{A}$ is L-strongly connected. By Lemma 4.6, the stochastic matrix $T_{L,\mathcal{A}}$ defined in (4.1) is irreducible. Therefore its stationary probability vector $\boldsymbol{t}$ is unique. Lemma 4.9 implies that $\boldsymbol{t} = \boldsymbol{k} \otimes \boldsymbol{l}$, where $\boldsymbol{k}$ is the stationary probability vector of the stochastic matrix $K_{\boldsymbol{l},\mathcal{A}}$ defined in (4.2). By Lemma 7.4, the measures μ and $g_*\mu$ are either singular or the same. It follows from Lemma 7.2 that $g_*\mu = \mu$ if and only if $\boldsymbol{l}_{\lambda(s,x)} = \boldsymbol{l}_x$ for all $s \in S$ and $x \in X$. □

Example 7.6. Let $X = \{1, 2, 3\}$ and μ be a Bernoulli measure on $X^{\mathbb{N}}$ defined by a probability vector $\boldsymbol{l} = (1/2, 1/4, 1/4)$. Let $\mathcal{A} = (X, \{s_0, s_1\}, \pi, \lambda)$, where $\pi(s_i, x) = s_{1-i}$ for all $x \in X$ and $i \in \{0, 1\}$, $\lambda(s_0, 1) = 2$, $\lambda(s_1, 1) = 3$, and $\lambda(s_i, x) = 1$ for $x \in \{2, 3\}$ and $i \in \{0, 1\}$. Let g be either of the two states of the automaton $\mathcal{A}$. Then for μ-almost all $\omega \in X^{\mathbb{N}}$ any symbol $x \in X$ occurs with the same frequency $\boldsymbol{l}_x$ in ω and $g(\omega)$. If g were invertible, this would imply $g_*\mu = \mu$. In fact, the measures μ and $g_*\mu$ are singular, but we need to look at words of length 2 to be able to apply Lemma 7.3. Indeed, 22 and 33 occur with the same frequency 1/16 in a μ-generic sequence ω while not occurring at all in $g(\omega)$. △

In view of the previous example, we should expect the measures μ and $g_*\mu$ to be singular even if g is not invertible. There are exceptions, of course.

Example 7.7. Let X be any alphabet of more than one character. For any $x \in X$ and $\omega \in X^{\mathbb{N}}$ let $g_x(\omega) = x\omega$. All transformations g_x, $x \in X$ can be generated by a single automaton $\mathcal{A} = (X, S, \pi, \lambda)$, where $S = \{g_x \mid x \in X\}$, $\pi(g_x, y) = g_y$ and $\lambda(g_x, y) = x$ for all $x, y \in X$. If μ is a Bernoulli measure on $X^{\mathbb{N}}$ defined by a positive probability vector $\boldsymbol{l}$, then $\mu = \sum_x \boldsymbol{l}_x\,(g_x)_*\mu$. As a consequence, each measure $(g_x)_*\mu$ is absolutely continuous with respect to μ while not the same as μ. $\triangle$

To prove an analogue of Theorem 7.5 for general Markov measures, we need first to establish an analogue of Lemma 7.2.

Lemma 7.8. *Let μ be a Markov measure on $X^{\mathbb{N}}$ defined by an irreducible stochastic matrix L with stationary probability vector $\boldsymbol{l}$ and $g : X^{\mathbb{N}} \to X^{\mathbb{N}}$ be an invertible transformation generated by an L-strongly connected automaton $\mathcal{A} = (X, S, \pi, \lambda)$. Then the following conditions are equivalent: (i) $g_*\mu = \mu$; (ii) $\boldsymbol{l}_{x'} = \boldsymbol{l}_x$ whenever $\lambda(g, x) = x'$, and $L_{x'y'} = L_{xy}$ whenever $\lambda(s, xy) = x'y'$ for some $s \in S$.*

Proof. Consider an arbitrary word $u = x_1x_2\dots x_k \in X^*$ and let $g^{-1}(u) = y_1y_2\dots y_k$. Then $\mu(uX^{\mathbb{N}}) = \boldsymbol{l}_{x_1}L_{x_1x_2}\dots L_{x_{k-1}x_k}$ and $g_*\mu(uX^{\mathbb{N}}) = \mu(g^{-1}(uX^{\mathbb{N}})) = \mu\big(g^{-1}(u)X^{\mathbb{N}}\big) = \boldsymbol{l}_{y_1}L_{y_1y_2}\dots L_{y_{k-1}y_k}$. Clearly, $x_1 = \lambda(g, y_1)$ and $x_1x_2 = \lambda(g, y_1y_2)$. Besides, $x_ix_{i+1} = \lambda(s_i, y_iy_{i+1})$ for $2 \le i \le k-1$, where $s_i = \pi(g, y_1y_2\dots y_{i-1})$. Assuming the condition (ii) holds, we obtain that $\boldsymbol{l}_{x_1} = \boldsymbol{l}_{y_1}$ and $L_{x_ix_{i+1}} = L_{y_iy_{i+1}}$ for $1 \le i \le k-1$. Then $\mu(uX^{\mathbb{N}}) = g_*\mu(uX^{\mathbb{N}})$. Thus the measures μ and $g_*\mu$ coincide on the cylinders, which implies that $g_*\mu = \mu$.

Conversely, assume that $g_*\mu = \mu$. If $\lambda(g, x) = x'$ for some $x, x' \in X$, then $\boldsymbol{l}_{x'} = \mu(x'X^{\mathbb{N}}) = g_*\mu(x'X^{\mathbb{N}}) = \mu(xX^{\mathbb{N}}) = \boldsymbol{l}_x$. Now suppose $\lambda(s, xy) = x'y'$ for some $s \in S$. Since the automaton $\mathcal{A}$ is L-strongly connected, there exist symbols $x_0, x_1, \dots, x_k = x$ ($k \ge 1$) such that $\pi(g, x_0x_1\dots x_{k-1}) = s$ and $L_{x_ix_{i+1}} > 0$ for $0 \le i \le k-1$. Note that the vector $\boldsymbol{l}$ is positive since the stochastic matrix L is irreducible. Therefore $\mu(x_0x_1\dots x_kX^{\mathbb{N}}) = \boldsymbol{l}_{x_0}L_{x_0x_1}\dots L_{x_{k-1}x_k} > 0$. Let $u = x_0x_1\dots x_{k-1}$ and $u' = g(u)$. Since $\pi(g, u) = s$, we have $g(ux) = u'x'$ and $g(uxy) = u'x'y'$. As a consequence, $\mu(uxX^{\mathbb{N}}) = g_*\mu(u'x'X^{\mathbb{N}}) = \mu(u'x'X^{\mathbb{N}})$ and $\mu(uxyX^{\mathbb{N}}) = g_*\mu(u'x'y'X^{\mathbb{N}}) = \mu(u'x'y'X^{\mathbb{N}})$. By definition of the measure μ, we have $\mu(uxyX^{\mathbb{N}}) = \mu(uxX^{\mathbb{N}})L_{xy}$ and $\mu(u'x'y'X^{\mathbb{N}}) = \mu(u'x'X^{\mathbb{N}})L_{x'y'}$. By the above the measure $\mu(uxX^{\mathbb{N}}) = \mu(u'x'X^{\mathbb{N}})$ is not zero. It follows that $L_{x'y'} = L_{xy}$. $\square$

Theorem 7.9. *Let μ be a Markov measure on $X^{\mathbb{N}}$ defined by an irreducible stochastic matrix L with stationary probability vector $\boldsymbol{l}$. Suppose $g : X^{\mathbb{N}} \to X^{\mathbb{N}}$ is an invertible transformation generated by a reversible, L-strongly connected automaton $\mathcal{A} = (X, S, \pi, \lambda)$. Then the measures μ and $g_*\mu$ are singular unless $\boldsymbol{l}_{\lambda(g,x)} = \boldsymbol{l}_x$ and $L_{\lambda(s,x),\,\lambda(\pi(s,x),y)} = L_{x,y}$ for all $s \in S$ and $x, y \in X$, in which case $g_*\mu = \mu$.*

Proof. The theorem is proved in the same way as Theorem 7.5 but instead of Lemmas 4.9 and 7.2, one has to use respectively Lemmas 5.2 and 7.8. □

References

[1] L. Bartholdi, R. Grigorchuk, V. Nekrashevych, From fractal groups to fractal sets. *Fractals in Graz 2001*, 25–118. Trends Math., Birkhäuser, Basel, 2003. [arXiv:math/0202001]

[2] I. Bondarenko, R. Grigorchuk, R. Kravchenko, Y. Muntyan, V. Nekrashevych, D. Savchuk, Z. Šunić, On classification of groups generated by 3-state automata over a 2-letter alphabet. *Algebra Discrete Math.* (2008), no. 1, 1–163. [arXiv:0803.3555]

[3] P. Billingsley, Ergodic Theory and Information. John Wiley & Sons, 1965.

[4] A. Dudko, R. Grigorchuk, On spectra of Koopman, groupoid and quasi-regular representations. *J. of Modern Dynamics* **11** (2017), 99–123. [arXiv:1510.00897]

[5] R. Grigorchuk, R. Kogan, Y. Vorobets, Automatic logarithm and associated measures. Preprint, 2018. [arXiv:1812.00069]

[6] R. I. Grigorchuk, V. V. Nekrashevych, V. I. Sushchanskii, Automata, dynamical systems, and groups. *Proc. Steklov Inst. Math.* 2000, no. 4(231), 128–203.

[7] R. Kravchenko, The action of finite-state tree automorphisms on Bernoulli measures. *J. of Modern Dynamics* **4** (2010), no. 3, 443–451.

[8] V. B. Kudryavtsev, S. V. Aleshin, A. S. Podkolzin, Introduction to automata theory (Russian). Nauka, Moscow, 1985.

[9] V. Nekrashevych, Self-similar groups. *Mathematical Surveys and Monographs, 117.* Amer. Math. Soc., Providence, RI, 2005.

[10] A. V. Ryabinin, Stochastic functions of finite automata. *Algebra, Logic and Number Theory (Russian)*, 77–80, Moskov. Gos. Univ., Moscow, 1986.

[11] S. Sidki, Automorphisms of one-rooted trees: Growth, circuit structure, and acyclicity. *J. Math. Sci. (New York)* **100** (2000), 1925–1943.

Department of Mathematics
Mailstop 3368
Texas A&M University
College Station, TX 77843-3368

Email: grigorch@math.tamu.edu, romwell@math.tamu.edu,
yvorobet@math.tamu.edu