SOFIC MEASURES

CHARACTERIZATIONS OF HIDDEN MARKOV CHAINS BY LINEAR ALGEBRA, FORMAL LANGUAGES, AND SYMBOLIC DYNAMICS

KARL PETERSEN

Mathematics 210, Spring 2006
University of North Carolina at Chapel Hill

NOTES BY
Rika Hagihara, Jessica Hubbs, Nathan Pennington, Yuki Yayama

1. Sofic Measures

In symbolic dynamics and the theory of stationary processes, Bernoulli and Markov measures are among the first examples of measures examined. They have been thoroughly studied and many of their properties are known. Each is specified by a finite set of data, and its entropy is given by an explicit finite formula. In contrast, images of these measures under continuous factor maps (which we here call *sofic measures* and are also called *hidden Markov measures* and *functions of Markov chains*) are not well understood. This is a bit surprising, since these measures result from finitely-specified ones by applying simple finite algorithms, such as clumping some symbols in an alphabet. For example, a Markov measure μ may be defined on a state space $\{a_1, a_2, b_1, b_2\}$, but maybe only the symbols a, b without subscripts are recorded, resulting in an output measure ν on the set of sequences on the alphabet $\{a, b\}$. Certain properties of ν follow automatically from those of μ, but the entropy of ν, for example, can be difficult to determine.

Sofic measures are appearing frequently in the mathematics of information transfer and also in the sciences, where what is measured is often an imperfect image of what one wants to study. It can be important to decide whether a given measure is sofic;

Keywords: Formal series, hidden Markov chain, sofic measure.

and if it is, finding the Markov measure of which it is a factor will identify the controlling parameters of the process.

In part of a graduate course taught in spring 2006 at the University of North Carolina at Chapel Hill, we sought to understand the connections among several equivalent characterizations of sofic measures, following mainly the book of Berstel and Reutenauer [1] and a paper of Hansel and Perrin [10]. Background for the following, including assumed terminology, can be found in [14].

Let (X, S) and (Y, T) be two topological dynamical systems. A map $\pi : X \to Y$ is called a *factor map* if it is a continuous surjective map and $\pi \circ S = T \circ \pi$.

Let $\mathcal{B}(X)$ be the σ-algebra of Borel subsets of X. Let $(X, \mathcal{B}(X), \mu)$ be a measure space. Then π induces a measure ν on $(Y, \mathcal{B}(Y))$ by $\nu(B) = \mu(\pi^{-1}B)$ for all $B \in \mathcal{B}(Y)$. We write $\nu = \pi\mu = \mu\pi^{-1}$.

Definition 1.1. *A* sofic measure *is a measure which is the image of a Markov measure on a shift of finite type (SFT) under a continuous shift-commuting map.*

Note: Because we can always pass to higher block representations, we may assume that the Markov measure is one-step and the factor map is a one-block map or clumping of the alphabet.

An outstanding problem is to characterize sofic measures and study their dynamical properties.

Example 1.2. Blackwell's Examples [4]

Case 1:

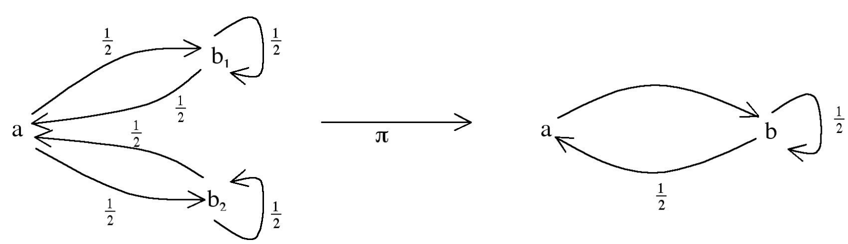

FIGURE 1. Blackwell's Example, Case 1

Consider the Markov measure μ on X given by the transition matrix

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

and the unique probability vector $p = (1/3, 1/3, 1/3)$ with $pA = p$. Then π induces the sofic measure $\nu = \mu\pi^{-1}$ on $Y = \pi(X)$. This measure is also Markov, and it is given by the transition matrix

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

and the fixed probability vector $(1/3, 2/3)$.

Case 2:

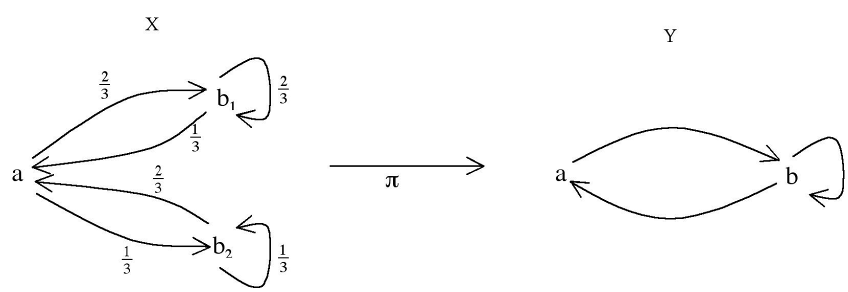

FIGURE 2. Blackwell's Example, Case 2

Let μ be the Markov measure on X given by the transition matrix

$$A = \begin{pmatrix} 0 & 2/3 & 1/3 \\ 1/3 & 2/3 & 0 \\ 2/3 & 0 & 1/3 \end{pmatrix}$$

and the probability vector $(2/7, 4/7, 1/7)$. The sofic measure ν is defined as $\nu(B) = \mu(\pi^{-1}(B))$ for all cylinder sets B.

Claim 1 : ν is not Markov (of any order).

Proof : Our proof proceeds in two stages. First we will show that ν is not one-step Markov and then show that it is not m-step Markov for any m. Assume for contradiction that ν is one-step Markov. Then we have

$$\nu(ab^n|a) = \frac{\nu[ab^n]_0}{\nu[a]_0} = \frac{\nu\{y \in \{a,b\}^{\mathbb{Z}} : y = \cdots \dot{a}\overbrace{bb\cdots b}^{n}\}}{\nu\{y \in \{a,b\}^{\mathbb{Z}} : y = \cdots \dot{a} \cdots\}}.$$

If ν is Markov, say given by the stochastic matrix

$$P = \begin{pmatrix} P_{aa} & P_{ab} \\ P_{ba} & P_{bb} \end{pmatrix}$$

and the fixed vector $p = (p_a, p_b)$, then

$$\frac{\nu[...\dot{a}\overbrace{bbb\cdots b}^{n}]}{\nu[\dot{a}]} = P_{ab}\overbrace{P_{bb}...P_{bb}}^{n-1} = P_{ab}P_{bb}^{n-1} = P_{bb}^{n-1}.$$

(Note that $P_{aa} = 0$ so $P_{ab} = 1$.)

Since $\nu = \mu\pi^{-1}$, we also have

$$\begin{aligned} \nu(ab^n|a) &= \frac{\mu\pi^{-1}[ab^n]_0}{\mu\pi^{-1}[a]_0} \\ &= \frac{\mu[ab_1^n]_0 + \mu[ab_2^n]_0}{\mu[a]_0} \\ &= \mu[ab_1^n|a] + \mu[ab_2^n|a] \\ &= \frac{2}{3}(\frac{2}{3})^{n-1} + \frac{1}{3}(\frac{1}{3})^{n-1}. \end{aligned}$$

So we conclude that

$$\frac{2}{3}(\frac{2}{3})^{n-1} + \frac{1}{3}(\frac{1}{3})^{n-1} = P_{bb}^{n-1} \text{ for all } n.$$

This in turn implies that

$$\frac{2}{3}\left(\frac{2/3}{P_{bb}}\right)^{n-1} + \frac{1}{3}\left(\frac{1/3}{P_{bb}}\right)^{n-1} = 1 \text{ for all } n.$$

If $P_{bb} < 2/3$, the left-hand side tends to infinity as n increases, while if $P_{bb} \geq 2/3$ the left-hand side is less than 1 for large n. Since there is no $P_{bb} > 0$ such that the above equation holds, we finally conclude by contradiction that ν is not one-step Markov.

For the more general case, if ν is Markov, its m-block representation is one-step Markov, and so is σ^m on its m-block representation. Then there are numbers t_0 and T_{b^m,b^m} such that

$$\nu(\dot{a}b^{mn}|b^m\dot{a}) = t_0 T_{b^m,b^m}^{mn-1} \text{ for all } n.$$

As in the one-step case,

$$\nu(\dot{a}b^{mn}|b^m\dot{a}) = \frac{2}{3}\left(\frac{2}{3}\right)^{mn-1} + \frac{1}{3}\left(\frac{1}{3}\right)^{mn-1} \text{ for all } n,$$

and this again leads to a contradiction.

2. Formal Series

2.1. **Basic definitions.** Let A be a finite alphabet, A^* the set of all finite words on A, and A^+ the set of all finite nonempty words on A. Let ϵ denote the empty word. A *language* on A is any subset $\mathcal{L} \subset A^*$.

Recall that a *monoid* is a set S with a binary operation $S \times S \to S$ which is associative and has a neutral element (identity). This means we can think of A^* as the multiplicative free monoid generated by A, where the operation is concatenation and the neutral element is ϵ.

A *formal series* (non-negative real-valued, based on A) is a function $s : A^* \to \mathbb{R}_+$. For all $w \in A^*$, $s(w) = (s, w) \in \mathbb{R}_+$, which can be thought of as the coefficient of w in the series s. We will think of this s as $\sum_{w \in A^*} s(w)w$, and this will be justified later.

It is sometimes useful to consider formal series with values in any semiring K, which is just a ring without subtraction. That is, K is a set with operations $+$ and $\cdot$ such that $(K,+)$ is a commutative monoid with identity element 0, $(K,\cdot)$ is a monoid with identity element 1; the product distributes over the sum; and for $k \in K$, $0k = k0 = 0$.

We denote the set of all K-valued formal series based on A by $K\langle\langle A\rangle\rangle$ or $\mathcal{F}_K(A)$. We further abbreviate $\mathbb{R}_+\langle\langle A\rangle\rangle = \mathcal{F}(A)$.

Then $\mathcal{F}(A)$ is a semiring in a natural way: For $s_1, s_2 \in \mathcal{F}(A)$, define

(1) $(s_1 + s_2)(w) = s_1(w) + s_2(w)$
(2) $(s_1 s_2)(w) = \sum_{u,v\in A^*} s_1(u)s_2(v),$ where the sum is over all $uv = w,$ a finite sum.

The neutral element in $\mathcal{F}$(A) is $s_1(w) = \begin{cases} 1 & \text{if } w = \epsilon \\ 0 & \text{otherwise.} \end{cases}$

Note that:

- $\mathbb{R}_+$ acts on $\mathcal{F}$(A) on both sides:
  $(ts)(w) = ts(w),\ (st)(w) = s(w)t,$ for all $w \in A^*$, for all $t \in \mathbb{R}_+$.
- There is a natural injection $A^* \hookrightarrow \mathcal{F}$(A) as a multiplicative submonoid: For $w \in A^*$ and $v \in A^*$, define

$$w(v) = \delta_{wv} = \begin{cases} 1 & \text{if } w = v \\ 0 & \text{otherwise.} \end{cases}$$

  This is a 1-term series.

So the neutral element for multiplication in $\mathcal{F}$(A) is ϵ as a function on A^*:

$$\epsilon(v) = \begin{cases} 1 & \text{if } \epsilon = v \\ 0 & \text{otherwise.} \end{cases}$$

There is a natural injection $\mathbb{R}_+ \hookrightarrow \mathcal{F}$(A) by $t \mapsto t\epsilon$ for all $t \in \mathbb{R}_+$. So in $\mathcal{F}$(A), $s_1 = \epsilon = 1$ under these interpretations.

Definition 2.1. *The* support *of a formal series* $s \in \mathcal{F}(\mathrm{A})$ *is*

$$\sup(s) = \{w \in A^* : s(w) \neq 0\}.$$

Note that $\sup(s)$ is a language. Given any language $\mathcal{L}$, define

$$s(w) = \begin{cases} 1 & \text{if } w \in \mathcal{L} \\ 0 & \text{if } w \notin \mathcal{L}. \end{cases}$$

This is called the *characteristic series* of $\mathcal{L}$, and its support is $\mathcal{L}$. Thus a language corresponds to a series with coefficients 0 and 1.

Definition 2.2. *A* polynomial *is an element of* $\mathcal{F}(\mathrm{A})$ *whose support is a finite subset of* A^*.

Denote the K*-valued polynomials based on* A *by* $\wp_K(A) = K\langle A\rangle$*. The* degree *of a polynomial* p *is* $\deg(p) = \max\{|w| : p(w) \neq 0\}$ *and is* $-\infty$ *if* $p \equiv 0$.

Definition 2.3. *A family* $\{f_\lambda : \lambda \in \Lambda\} \subset \mathcal{F}(\mathrm{A})$ *of series is called* locally finite *if for all* $w \in A^*$ *there are only finitely many* $\lambda \in \Lambda$ *for which* $f_\lambda(w) \neq 0$*. A series* $f \in \mathcal{F}(\mathrm{A})$ *is called* proper *if* $f(\epsilon) = 0$.

Proposition 2.4. *If* $f \in \mathcal{F}(\mathrm{A})$ *is proper, then* $\{f^n : n = 0, 1, 2, \dots\}$ *is locally finite.*

Proof. If $n > |w|$, then $f^n(w) = 0$, because

$$f^n(w) = \sum_{\substack{u_1 \dots u_n = w \\ u_i \in A^*,\ i=1,\dots,n}} f(u_1) \dots f(u_n)$$

and at least one u_i is ϵ. □ □

Definition 2.5. *If* $f \in \mathcal{F}(\mathrm{A})$ *is proper, define*

$$f^* = \sum_{n=0}^{\infty} f^n \text{ and } f^+ = \sum_{n=1}^{\infty} f^n \text{ (pointwise finite sum)},$$

with $f^0 = 1 = 1 \cdot \epsilon = \epsilon$.

2.2. **Rational series and languages.**

Definition 2.6. *The* rational operations *in* $\mathcal{F}(\mathrm{A})$ *are sum* $(+)$*, product* $(\cdot)$*, multiplication by real numbers* (tw)*, and* $* : f \to f^*$*. The family of* rational series *consists of those* $f \in \mathcal{F}(\mathrm{A})$ *that can be obtained by starting with a finite set of polynomials in* $\mathcal{F}(\mathrm{A})$ *and applying a finite number of rational operations.*

Definition 2.7. *A language $\mathcal{L} \subset A^*$ is* rational *if and only if its characteristic series*

$$F(w) = \begin{cases} 1 & \text{if } w \in \mathcal{L} \\ 0 & \text{if } w \notin \mathcal{L} \end{cases} \text{ is rational.}$$

Proposition 2.8. *A language $\mathcal{L}$ is rational if and only if it is* regular. *Thus an insertive and extractive language is rational if and only if it is the language of a sofic subshift.*

Recall that regular languages correspond to regular expressions: Include A, ϵ, ϕ and closed under $+$, $\cdot$, $*$. (The language consists of words obtained by reading off sequences of edge labels on a finite labeled directed graph.)

2.3. **Distance and topology in $\mathcal{F}(\mathrm{A})$.** If $f_1, f_2 \in \mathcal{F}(\mathrm{A})$, define

$$D(f_1, f_2) = \inf\{n \geq 0 : \text{ there is } w \in A^n \text{ such that } f_1(w) \neq f_2(w)\}$$

and

$$d(f_1, f_2) = \frac{1}{2^{D(f_1,f_2)}}.$$

Note that $d(f_1, f_2)$ defines an *ultrametric* on $\mathcal{F}(\mathrm{A})$:

$$d(f, h) \leq \max\{d(f, g), d(g, h)\}$$

$$(\leq d(f, g) + d(g, h)).$$

Proposition 2.9. *With respect to the metric d, $f_k \to f$ if and only if $f_k(w) \to f(w)$ in the discrete topology on $\mathbb{R}$, i.e. pointwise.*

A possibly interesting question is to determine what would happen if we were to use the usual topology of $\mathbb{R}$ instead of this new topology.

Proposition 2.10. (1) *$\mathcal{F}(\mathrm{A})$ is* complete *with respect to the metric d.*

(2) *$\mathcal{F}(\mathrm{A})$ is a* topological semiring *with respect to the metric d (that is, $+$ and $\cdot$ are continuous as functions of two variables).*

Definition 2.11. *A family $\{F_\lambda : \lambda \in \Lambda\}$ of formal series is called* summable *if there is a series $F \in \mathcal{F}(\mathrm{A})$ such that for every $\varepsilon > 0$ there is a finite set $\Lambda_\varepsilon \subset \Lambda$ such that for each finite set $I \subset \Lambda$ with $\Lambda_\varepsilon \subset I$, $d(\sum_{i \in I} F_i, F) < \varepsilon$. Then F is called the* sum *of the series and we write* $F = \sum_{\lambda\in\Lambda} F_\lambda$.

Proposition 2.12. *If* $\{F_\lambda : \lambda \in \Lambda\}$ *is locally finite, then it is summable.*

Proof. Given such a locally finite family, define F pointwise on A^* by $F(w) = \sum_{\lambda\in\Lambda} F_\lambda(w)$. Note that this is a a finite sum. We claim that $F = \sum_{\lambda\in\Lambda} F_\lambda$.

Let $\varepsilon > 0$. Returning to the definition of d, there is a corresponding N such that if $F_1(w) = F_2(w)$ for all $w \in \mathrm{A}^*$ with $|w| \le N$, then $d(F_1, F_2) < \varepsilon$. Choose

$\Lambda_\varepsilon = \{\lambda \in \Lambda : F_\lambda(w) \ne 0$ for some $w \in \mathrm{A}^*$ with $|w| \le N\}$, a finite set.

If $\Lambda_\varepsilon \subset I$ and if $|w| \le N$, then

$$|\sum_{\lambda\in\Lambda} F_{\lambda\in\Lambda}(w) - \sum_{i\in I} F_i(w)| = 0,$$

so that $d(F, \sum_{i\in I} F_i) < \varepsilon$. □

□

So now we can write any $F \in \mathcal{F}(\mathrm{A})$ as $F = \sum_{w\in \mathrm{A}^*} F(w)w$, where the formal series is a convergent infinite series of polynomials in the metric of $\mathcal{F}(\mathrm{A})$. Recall that

$$(F(w)w)(v) = \begin{cases} F(w) & \text{if } w = v \\ 0 & \text{if } w \ne v, \end{cases}$$

where $F(w)w \in \mathcal{F}(\mathrm{A})$ and $w \in \mathrm{A}^*$, so that $\{F(w)(w) : w \in \mathrm{A}^*\}$ is a locally finite, and hence summable, subfamily of $\mathcal{F}(\mathrm{A})$.

We note here that the set of $\wp(A)$ of all polynomials is dense in $\mathcal{F}(\mathrm{A})$.

3. Recognizable (linearly representable) series

3.1. Definition and examples.

Definition 3.1. $F \in \mathcal{F}(\mathrm{A})$ *is* linearly representable *if there exists an* $n \geq 1$ *(the* dimension *of the representation) such that we have a* $1 \times n$ *row vector* $x \in \mathbb{R}_+^n$*, an* $n \times 1$ *column vector* $y \in \mathbb{R}_+^n$*, and a morphism of multiplicative monoids* φ *:* $\mathrm{A}^* \to \mathbb{R}_+^{n \times n}$ *(the multiplicative monoid of* $n \times n$ *matrices) such that for all* $w \in \mathrm{A}^*$*,* $F(w) = x\varphi(w)y$ *(matrix multiplication).*

Example 3.2. Consider a Bernoulli measure $\mathcal{B}(p_0, p_1, \ldots, p_{d-1})$ on $\Omega_+(A) = A^{\mathbb{Z}_+}$ where $A = \{0, 1, \ldots, d-1\} = \{a_0, a_1, \ldots, a_{d-1}\}$, and $p = (p_0, p_1, \ldots, p_{d-1})$ is a probability vector. Let $f = \sum_{i=0}^{d-1} p_i a_i \in \mathcal{F}(\mathrm{A})$. Then

$$f(w) = \begin{cases} p_i & \text{if } w = a_i \\ 0 & \text{if } w \neq a_i. \end{cases}$$

Define $F_p = f^* = \sum_{n \geq 0} f^n$. Note that f is proper since we have $f(\epsilon) = 0$. For $w = a_2 a_0$, $f^*(w) = f^0(w) + f(w) + f^2(w) + f^3(w) + \ldots$, $f^0(w) = 0$ and $f(w) = 0$. For $n \geq 3$, we have $f^n(w) = 0$ because any factorization $w = u_1 u_2 u_3$ includes ϵ and $f(\epsilon) = 0$. We also have $f^2(w) = \sum_{uv=w} f(u)f(v) = f(a_2)f(a_0) = p_2 p_0$. Continuing in this way, we see that $F_p(w_1 w_2 \ldots w_n) = p_{w_1} p_{w_2} \ldots p_{w_n}$, where $w_i \in A$.

Example 3.3. Consider a Markov measure μ on $\Omega_+(A)$ defined by a $d \times d$ stochastic matrix P and a d-dimensional probability row vector $p = (p_0, p_1, \cdots, p_{d-1})$.

This means that for $a_j \in A$, $\mu[a_j]_0 = \sum_{i=0}^{d-1} p_i P_{ij} = p_j$ (if $pP = p$, $\mu[a_i]_0 = p_i$, and μ is σ-invariant). We also have that for all $w \in \mathrm{A}^*$ and $a_i, a_j \in A$, $\mu[w a_i a_j]_0 = \mu[w a_i]_0 P_{ij}$.

Put $y = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}_+^d$, $x = p \in \mathbb{R}_+^d$, and let φ be generated by $\varphi(a_j)$, $j = 0, 1, ..., d-1$,

where

$$\varphi(a_j) = \begin{pmatrix} 0 & \cdots & P_{0j} & 0 & \cdots & 0 \\ 0 & \cdots & P_{1j} & 0 & \cdots & 0 \\ \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 0 & \cdots & P_{d-1,j} & 0 & \cdots & 0 \end{pmatrix} \text{ for each } a_j \in A.$$

Then $F_{p,P} \in \mathcal{F}(\mathrm{A})$ is defined for all $w = w_1 \cdots w_n \in \mathrm{A}^*$ by

$$\begin{aligned} F_{p,P}(w) &= x\varphi(w)y \\ &= p\varphi(w_1)\varphi(w_2)\ldots\varphi(w_n)y \\ &= p\varphi(w_1)\ldots\varphi(w_{n-1}) \begin{pmatrix} 0 & \cdots & P_{0j_n} & 0 & \cdots & 0 \\ 0 & \cdots & P_{1j_n} & 0 & \cdots & 0 \\ \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 0 & \cdots & P_{d-1j_n} & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} a_{j_n}, \end{aligned}$$

where $w_i = a_{j_i}$ for $i = 1, \cdots, n$. So

$$p\varphi(w_1)\ldots\varphi(w_{n-1}) \begin{pmatrix} P_{0j_n} \\ P_{1j_n} \\ \vdots \\ P_{d-1,j_n} \end{pmatrix}$$

$$= p\varphi(w_1)\cdots\varphi(w_{n-2}) \begin{pmatrix} 0 & \cdots & P_{0j_{n-1}} & 0 & \cdots & 0 \\ 0 & \cdots & P_{1j_{n-1}} & 0 & \cdots & 0 \\ \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 0 & \cdots & P_{d-1,j_{n-1}} & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} P_{0j_n} \\ P_{1j_n} \\ \vdots \\ P_{d-1,j_n} \end{pmatrix}$$

$$= p\varphi(w_1)\cdots\varphi(w_{n-2}) \begin{pmatrix} P_{0j_{n-1}}P_{j_{n-1},j_n} \\ P_{1j_{n-1}}P_{j_{n-1},j_n} \\ \vdots P_{d-1,j_{n-1}}P_{j_{n-1},j_n} \end{pmatrix}$$

$$= p_0P_{0j_1}P_{j_1j_2}\cdots P_{j_{n-1}j_n} + p_1P_{1j_1}P_{j_1j_2}\cdots P_{j_{n-1}j_n} + \ldots + p_{d-1}P_{d-1,j_1}P_{j_1j_2}\cdots P_{j_{n-1}j_n}$$

$$= (\sum_{i=0}^{d-1} p_iP_{ij_1})P_{j_1j_2}P_{j_2j_3}\cdots P_{j_{n-1}j_n} = \mu[w_1\cdots w_n].$$

The last equality holds because we have $\mu[w_1\ldots w_n] = \mu[w_1]_0P_{j_1j_2}P_{j_2j_3}\cdots P_{j_{n-1}j_n}$. This shows that the triple x, φ, y represents the given Markov measure μ.

Example 3.4. We now consider how the preceding analysis works on a Blackwell type example, where the Markov measure is given by the graph in Example 1 (see Figure 3).

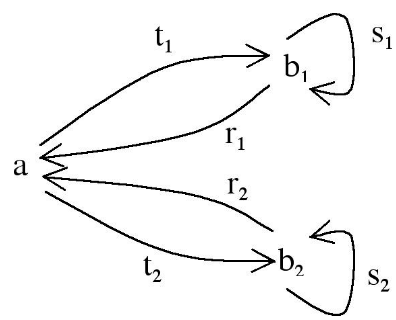

FIGURE 3. Blackwell-type example

We require $t_1 + t_2 = 1$, $r_1 + s_1 = 1$, and $r_2 + s_2 = 1$. Then the linear representation is given by $(p,\ P)$, where $p = (p_0, p_1, p_2)$ is the a priori probability vector and

$P = \begin{pmatrix} 0 & t_1 & t_2 \\ r_1 & s_1 & 0 \\ r_2 & 0 & s_2 \end{pmatrix}$ is the transition matrix.

The measure is σ-invariant if and only if $pP = p$. We have $x = p$, $y = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$,

$$\varphi(a) = \begin{pmatrix} 0 & 0 & 0 \\ r_1 & 0 & 0 \\ r_2 & 0 & 0 \end{pmatrix},\ \varphi(b_1) = \begin{pmatrix} 0 & t_1 & 0 \\ 0 & s_1 & 0 \\ 0 & 0 & 0 \end{pmatrix},\ \varphi(b_2) = \begin{pmatrix} 0 & 0 & t_2 \\ 0 & 0 & 0 \\ 0 & 0 & s_2 \end{pmatrix}.$$

Suppose $pP = p$. Then

$$F_{p,P}(ab_1{}^k ab_2{}^l a) = x\varphi(a)\begin{pmatrix}0 & t_1 & 0\\ 0 & s_1 & 0\\ 0 & 0 & 0\end{pmatrix}^k \varphi(a)\begin{pmatrix}0 & 0 & t_2\\ 0 & 0 & 0\\ 0 & 0 & s_2\end{pmatrix}^l \varphi(0)\begin{pmatrix}1\\1\\1\end{pmatrix}$$

$$= x\varphi(a)\begin{pmatrix}0 & t_1 s_1^{k-1} & 0\\ 0 & s_1^k & 0\\ 0 & 0 & 0\end{pmatrix}\varphi(a)\begin{pmatrix}0 & 0 & t_2 s_2^{l-1}\\ 0 & 0 & 0\\ 0 & 0 & s_2^l\end{pmatrix}\varphi(a)\begin{pmatrix}1\\1\\1\end{pmatrix}$$

$$= r_1 s_1^{k-1} t_2 s_2^{l-1} r_2 (p_1 r_1 + p_2 r_2).$$

We can easily see that the last equality holds if we plug in

$$x\varphi(a) = p_1 r_1 t_1 s_1^{k-1} r_1 t_2 s_2^{l-1} r_2 + p_2 r_2 t_1 s_1^{k-1} r_1 t_2 s_2^{l-1} r_2$$

and also note that $p_1 r_1 + p_2 r_2 = p_0$, because $pP = p$ implies that

$$pP = (p_0, p_1, p_2)\begin{pmatrix}0 & t_1 & t_2\\ r_1 & s_1 & 0\\ r_2 & 0 & s_2\end{pmatrix} = (p_1 r_1 + p_2 r_2, \ldots, \ldots) = p.$$

Example 3.5. *A sofic measure (hidden Markov chain) as in Example 1.2.* We examine a one-block factor of a Blackwell-type example. Define $\pi : \{a_1, b_1, b_2\} \to \{a, b\}$ by $\pi(a) = a$, $\pi(b_1) = \pi(b_2) = b$, and extend π as a monoid morphism to $\pi : \{a_1, b_1, b_2\}^* \to \{a, b\}^*$ and also as a factor map between shifts of finite type:

$$\begin{array}{ccc} & X \subset \Omega_+\{a, b_1, b_2\} & \\ \pi & \downarrow & \text{one-block map} \\ & Y \subset \Omega_+\{a, b\} & \end{array}$$

Y is actually the golden mean shift shown below.

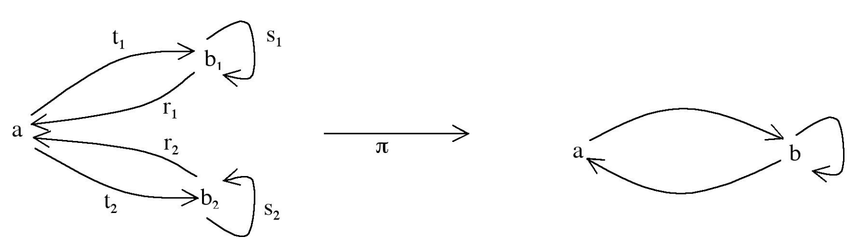

FIGURE 4

Take any shift-invariant measure μ on X. It projects to a (unique) shift-invariant measure $\nu = \pi\mu = \mu\pi^{-1}$ on Y: $\nu(B) = \mu(\pi^{-1}B)$ for all Borel sets $B \subset Y$. In particular, take $\mu = \mu_{p,P}$ to be a σ-invariant one-step Markov measure on X, as in Example 4 above. Let (x, φ, y) be the linear representation of the Markov measure μ.

We will now show that ν is linearly representable.

We begin by defining $x = (p_a, p_{b_1}, p_{b_2})$, $y = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$, and

$$\psi(b) = \varphi(b_1) + \varphi(b_2) = \begin{pmatrix} 0 & t_1 & t_2 \\ 0 & s_1 & 0 \\ 0 & 0 & s_2 \end{pmatrix}.$$

We now wish to show that (x, ψ, y) corresponds to a linearly representable measure ν.

The first step is to rewrite $\psi(b)$ as $\psi(b) = \psi(ab) + \psi(bb)$, where

$$\psi(ab) = \begin{pmatrix} 0 & t_1 & t_2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

represents the probability of moving from a to b and

$$\psi(bb) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & s_1 & 0 \\ 0 & 0 & s_2 \end{pmatrix}$$

represents the probability of moving from b to b.

We observe the following facts about $\psi(ab)$ and $\psi(bb)$:

(1) $\psi(ab)^2 = \mathbf{0}$
(2) $\psi(bb) \cdot \psi(ab) = \mathbf{0}$

(3) $\psi(ab)\cdot\psi(bb) = \begin{pmatrix} 0 & t_1s_1 & t_2s_2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$

(4) $\psi(bb)^k = \begin{pmatrix} 0 & 0 & 0 \\ 0 & s_1^k & 0 \\ 0 & 0 & s_2^k \end{pmatrix}$

We use these four facts to compute, for arbitrary k, $\psi(b)^k = [\psi(ab) + \psi(bb)]^k$. Since

$$[\psi(ab) + \psi(bb)][\psi(ab) + \psi(bb)] = \psi(ab)^2 + \psi(ab)\psi(bb) + \psi(bb)\psi(ab) + \psi(bb)^2$$

$$= \psi(ab)\psi(bb) + \psi(bb)^2,$$

we may use induction to conclude that

$$\psi(b)^k = [\psi(ab) + \psi(bb)]^k$$

$$= \psi(ab)\psi(bb)^{k-1} + \psi(bb)^k$$

$$= \begin{pmatrix} 0 & t_1s_1^{k-1} & t_2s_2^{k-1} \\ 0 & s_1^k & 0 \\ 0 & 0 & s_2^k \end{pmatrix}.$$

With a matrix representation of $\psi(b)^k$ in hand, we can check that (x, ψ, y) is a linear representation of ν. To do this, we will first show that (x, ψ, y) gives the right answer for $\nu[b^k]_0$. In fact,

$$x\psi(b^k)y = (p_a, p_{b_1}, p_{b_2})\psi(b^k)\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

$$= p_at_1s_1^{k-1} + p_{b_1}s_l^k + p_at_2s_2^{k-1} + p_{b_2}s_2^k$$

$$= s_1^{k-1}(p_at_1 + p_{b_1}s_1) + s_2^{k-1}(p_at_2 + p_{b_2}s_2).$$

Since we know $p_a t_1 + p_{b_1} s_1 = p_{b_1}$ and $p_a t_2 + p_{b_2} s_2 = p_{b_2}$, we have

$$s_1^{k-1}(p_a t_1 + p_{b_1} s_1) + s_2^{k-1}(p_a t_2 + p_{b_2} s_2) = p_{b_1} s_1^{k-1} + p_{b_2} s_2^{k-1}.$$

From the construction of μ we have that $\mu[b_1^k]_0 = p_{b_1} s_1^{k-1}$ and $\mu[b_2^k]_0 = p_{b_2} s_2^{k-1}$, so we finally conclude that

$$x\psi(b^k)y = \mu[b_1^k]_0 + \mu[b_2^k]_0 = \mu\pi^{-1}[b^k]_0.$$

By definition, $\nu[b^k]_0 = \mu\pi^{-1}[b^k]_0$, so we have now shown that $\nu[b^k]_0 = x\psi(b^k)y$. Now we obtain a matrix representation of $\psi(b^k a)$. Since ψ is a morphism, we have

$$\begin{aligned}
\psi(b^k a) &= \psi(b^k)\psi(a) \\
&= \begin{pmatrix} 0 & t_1 s_1^{k-1} & t_2 s_2^{k-1} \\ 0 & s_1^k & 0 \\ 0 & 0 & s_2^k \end{pmatrix} \cdot \begin{pmatrix} 0 & 0 & 0 \\ r_1 & 0 & 0 \\ r_2 & 0 & 0 \end{pmatrix} \\
&= \begin{pmatrix} r_1 t_1 s_1^{k-1} + r_2 t_2 s_2^{k-1} & 0 & 0 \\ r_1 s_1^k & 0 & 0 \\ r_2 s_2^k & 0 & 0 \end{pmatrix}.
\end{aligned}$$

Then

$$\begin{aligned}
x\psi(ab^k ab^l a)y &= x\psi(a)\psi(b^k a)\psi(b^l a)y \\
&= (r_1 p_{b_1} + r_2 p_{b_2}, 0, 0) \begin{pmatrix} r_1 t_1 s_1^{k-1} + r_2 t_2 s_2^{k-1} & 0 & 0 \\ r_1 s_1^k & 0 & 0 \\ r_2 s_2^k & 0 & 0 \end{pmatrix} \cdot \\
&\quad \begin{pmatrix} r_1 t_1 s_1^{l-1} + r_2 t_2 s_2^{l-1} & 0 & 0 \\ r_1 s_1^l & 0 & 0 \\ r_2 s_2^l & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \\
&= ((r_1 p_{b_1} + r_2 p_{b_2})(r_1 t_1 s_1^{k-1} + r_2 t_2 s_2^{k-1}), 0, 0) \begin{pmatrix} r_1 t_1 s_1^{l-1} + r_2 t_2 s_2^{l-1} \\ r_1 s_1^l \\ r_2 s_2^l \end{pmatrix} \\
&= (r_1 p_{b_1} + r_2 p_{b_2})(r_1 t_1 s_1^{k-1} + r_2 t_2 s_2^{k-1})(r_1 t_1 s_1^{l-1} + r_2 t_2 s_2^{l-1}).
\end{aligned}$$

Using induction we can then conclude that

$$x\psi(ab^{k_1}ab^{k_2}...ab^{k_n}a)y = x\psi(a)\psi(b^{k_1}a)\cdot\psi(b^{k_2}a)\cdot...\cdot\psi(b^{k_n}a)y$$

$$= (r_1p_{b_1}+r_2p_{b_2})(r_1t_1s_1^{k_1-1}+r_2t_2s_2^{k_1-1})(r_1t_1s_1^{k_2-1}+r_2t_2s_2^{k_2-1})\cdots$$

$$(r_1t_1s_1^{k_n-1}+r_2t_2s_2^{k_n-1}).$$

Since $p_a = r_1p_{b_1} + r_2p_{b_2}$, the above sum can be rewritten as

$$\sum_{(i_1,...,i_n)} p_a(r_{i_1}t_{i_1}s_{i_1}^{k_1-1})...(r_{i_n}t_{i_n}s_{i_n}^{k_1-1}),$$

where the sum is taken over all n-tuples $(i_1,...,i_n)$ such that i_j is either 1 or 2. But this is nothing more than

$$\sum_{(i_1,...,i_n)} \mu[ab_{i_1}^{k_1}a...b_{i_n}^{k_n}a]_0.$$

Viewed as a set, $\pi^{-1}[ab^{k_1}ab^{k_2}...ab^{k_n}a]_0$ is exactly the set of all $[ab_{i_1}^{k_1}a...b_{i_n}^{k_n}a]_0$ such that i_j is either 1 or 2. So we can finally conclude that

$$x\psi(ab^{k_1}ab^{k_2}...ab^{k_n}a)y = \sum_{(i_1,...,i_n)} \mu[ab_{i_1}^{k_1}a...b_{i_n}^{k_n}a]_0$$

$$= \mu\pi^{-1}[ab^{k_1}ab^{k_2}...ab^{k_n}a]_0,$$

and since the two measures agree on all the cylinder sets of the form $[ab_{i_1}^{k_1}a...b_{i_n}^{k_n}a]_0$, and these generate the full σ-algebra, we have that $\nu = \mu\pi^{-1}$. Therefore, (x,ψ,y) represents $\nu = \mu\pi^{-1}$.

3.2. **Linearly representable probability measures.** Our goal is to prove a theorem giving several characterizations of sofic measures. The first step in this process is to prove the following Proposition:

Proposition 3.6. *A formal series $\mathcal{F}(A)$ corresponds to a linearly representable shift-invariant probability measure μ on $\Omega_+(A)$ if and only if F has a linear representation (x,ϕ,y) with x a probability vector, $P=\sum_{a\in A}\phi(a)$ a stochastic matrix, y a column vector of all 1's, and $xP=x$.*

Proof. We begin by proving the "if" portion of the proposition.

Given the hypothesis of the proposition, define the functions $p_n : A^n \to \mathbb{R}_+$ by $p_n(w) = x\varphi(w)y$ for all $w \in A^n$. Then we have, for all $w \in A^n$, the following four relations:

(1) $p_n(w) = x\phi(w)y \geq 0$.
(2)

$$\begin{aligned}\sum_{a\in A} p_{n+1}(wa) &= \sum_{a\in A} x\phi(wa)y \\ &= \sum_{a\in A} x\phi(w)\phi(a)y = x\phi(w)(\sum_{a\in A}\phi(a)y).\end{aligned}$$

Since P is stochastic and y is a column vector of all 1's, $Py = y$, so

$$\sum_{a\in A} p_{n+1}(wa) = x\phi(w)Py = x\phi(w)y = p_n(w).$$

(3) $\sum_{a\in A} p_1(a) = \sum_{a\in A} x\phi(a)y = xPy = xy = 1$, since x is a probability vector.

(4) $\sum_{a\in A} p_{n+1}(aw) = p_n(w)$. The details of this are very similar to the details of (2).

Since the x, ϕ, y, and p of our hypothesis satisfy these four relations, the Kolmogorov Consistency Theorem (see [2], Chapter 7) gives a probability measure μ. Since $xP = x$, we also get that μ is shift-invariant.

We now turn our attention to proving the "only if" direction of the Proposition. We begin by proving several necessary claims and then use them to prove the Proposition.

Claim 1: If $\mathcal{F}(A)$ corresponds to a linearly representable shift-invariant probability measure μ on $\Omega_+(A)$, then for all $w \in A^*$, $\mu[w]_0 = x\phi(w)P^k y = xP^k\phi(w)y$ for all natural numbers k, where $P = \sum_{a\in A}\phi(a)$.

Proof: We prove this by induction. Since F corresponds to a linearly representable shift-invariant probability measure μ on $\Omega_+(A)$, we have that, for $k = 0$, $x\phi(w)P^0 y =$

$x\phi(w)y = \mu[w]_0$. Now assume the statement holds for all $k \leq M$. Then

$$\begin{aligned} x\phi(w)P^{M+1}y &= x\phi(w)\sum_{a\in A}\phi(a)P^M y \\ &= \sum_{a\in A} x\phi(w)\phi(a)P^M y \\ &= \sum_{a\in A} x\phi(wa)P^M y. \end{aligned}$$

By the induction hypothesis, $\mu[wa]_0 = x\phi(wa)P^M y$, so we have that

$$x\phi(w)P^{M+1}y = \sum_{a\in A}\mu[wa]_0.$$

Since μ is a probability measure, it satisfies the Kolmogorov consistency theorem, so $\sum_{a\in A}\mu[wa]_0 = \mu[w]_0$. We conclude that $x\phi(w)P^{M+1}y = \mu[w]_0$, so by induction we have proved $\mu[w]_0 = x\phi(w)P^k y$. Since μ is shift-invariant, a similar argument shows that $\mu[w]_0 = xP^k\phi(w)y$, so we have proved Claim 1.

Claim 2: There exists a linear representation (x, ϕ, y) of μ such that each entry of x and y is nonzero, and with P defined as $P = \sum_{a\epsilon A}\phi(a)$, $xP = x$ and $Py = y$.

Proof: Since (x, ϕ, y) is a linear representation of the measure μ, we have that for all $w \in A^*$, $\mu[w]_0 = x\phi(w)y = \sum_{i,j} x_i\phi(w)_{ij}\, y_j$. Thus, if $x_i = 0$; we may assume that the i'th row of $\phi(w)$ is identically 0, and similarly if $y_j = 0$, then we may assume that the j'th column of $\phi(w)$ is identically 0.

Note that

$$\mu[\epsilon] = 1 = x\phi(\epsilon)(P)^k y = xP^k y = \sum_{i,j} x_i (P^k)_{i,j}\, y_j.$$

Thus if $x_i > 0$ and $y_j > 0$, the set $\{(P^k)_{ij} : k = 0, 1, 2, \cdots\}$ is bounded. So we can find a sequence k_m such that $\{\frac{1}{k_m}\sum_{k=1}^{k_m} P^k\}$ converges to some Q_{ij} for all such i, j. If $x_i y_j = 0$, put $Q_{ij} = 0$. Now define $Q = (Q_{ij})$.

From Claim 1, we have that $\mu[w]_0 = xP^k\phi(w)P^k y$ for any k, so $\mu[w]_0 = xQ\phi(w)Qy$. This implies that by setting $\tilde{x} = xQ$ and $\tilde{y} = Qy$, we have that $(\tilde{x}, \phi, \tilde{y})$ is a linear representation of μ. We note that by the construction of Q, $\tilde{x}P = \tilde{x}$ and $P\tilde{y} = \tilde{y}$.

Moreover, now $\tilde{x}_j = 0$ if and only if $\tilde{y}_j = 0$. Eliminating all coordinates j such that $\tilde{y}_j = 0$ from $\tilde{x}, \tilde{y}$, and Q gives us a linear representation of μ given by $(\tilde{x}, \phi, \tilde{y})$, where $\tilde{x}P = \tilde{x}$, $P\tilde{y} = \tilde{y}$, and both $\tilde{x}$ and $\tilde{y}$ have no entries equal to 0. This proves Claim 2.

To finish the proof of Proposition 3.6, let (x, ϕ, y) be the linear representation of μ given by Claim 2. Define the diagonal matrix D by $D_{ii} = y_i$ for all i. Now define $x' = xD$, $\phi' = D^{-1}\phi D$, and $y' = D^{-1}y$. Then for any $w \in A^*$,

$$x'\phi'(w)y' = (xD)(D^{-1}\phi(w)D)(D^{-1}y) = x\phi(w)y,$$

so (x', ϕ', y') is a linear representation of μ. We have that $(D^{-1}y)_i = 1$ for all i, so y' is a column of 1's.

To show that x' is a probability vector, we note that the sum of the entries of x' is equal to

$$x'y' = (xD)(D^{-1}y) = xy = x\phi(\epsilon)y = \mu[\epsilon]_0 = 1.$$

To show that $P' = \sum_{a\in A}\phi'(a)$ is a stochastic matrix, we note that P' is a stochastic matrix if and only if $P'y'$ is a column vector of 1's, and that

$$P'y' = \sum_{a\in A}\phi(a)y = \sum_{a\in A} D^{-1}\phi(a)DD^{-1}y$$

$$= \sum_{a\in A} D^{-1}\phi(a)y = D^{-1}\sum_{a\in A}\phi(a)y = D^{-1}Py.$$

By Claim 2, $Py = y$, so we have that

$$P'y' = D^{-1}Py = D^{-1}y = y',$$

showing that P' is a stochastic matrix.

Lastly, we see that since

$$x'P' = xD\sum_{a\in A}\phi'(a) = xD\sum_{a\in A}D^{-1}\phi(a)D = xPD,$$

and we know from Claim 2 that $xP = x$, we can conclude that

$$x'P' = x'PD = xD = x'.$$

$\square$

In [10], Proposition 2.6 is stated without the requirement that the probability measure be shift-invariant. In Remark 4.9, we sketch how this could be proved. Part of the difficulty is due to the lack at this stage of any apparent relationship between coordinates in the linear representation and "states" (vertices) for a Markov measure.

4. Characterization of sofic measures

Kleene [12] characterized rational languages as the linearly representable ones, and this was generalized to formal series by Schützenberger [17]. In the study of stochastic processes, functions of Markov chains were analyzed by Gilbert [9], Furstenberg [8], Dharmadhikari [6], Heller [11], and others. Hansel and Perrin [10] explained the connection between rational series and continuous images of Markov chains. These characterizations are collected in the following theorem, which involves some terminology (stable, submodule) that is yet to be defined.

Theorem 4.1. *Let A be a finite alphabet. For a formal series $F \in F_{\mathbb{R}_+}(A)$ that corresponds to a shift-invariant probability measure μ in $\Omega_+(A)$, the following are equivalent:*

1. *F is linearly representable.*
2. *F is a member of a stable finitely-generated submodule of $F_{\mathbb{R}_+}(A)$.*
3. *F is rational.*
4. *The measure μ is the image under a one-block map of a one-step Markov shift-invariant probability measure.*

4.1. **Linearly representable=finite dimensional=rational.**

In this section we prove the equivalence of the first three items in the foregoing theorem. Note that $\mathcal{F}$(A) is an $\mathbb{R}_+$-*module* in a natural way. On this module we have an *action of A^** , defined as follows:

For $F \in \mathcal{F}(A)$ and $w \in A^*$, define $(w, F) \to w^{-1}F$ by

$$(w^{-1}F)(v) = F(wv) \text{ for all } v \in A^*.$$

Thus

$$w^{-1}F = \sum_{v \in A^*} F(wv)v.$$

If $F = u \in A^*$, then

$$(w^{-1}F)(v) = u(wv) = \begin{cases} 1 & \text{if } wv = u \\ 0 & \text{if } wv \neq u. \end{cases}$$

So $w^{-1}u \neq 0$ if and only if $u = wv$ for some $v \in A^*$. Then $w^{-1}u = v$ (in the sense that they are the same function on A^*). So $w^{-1}v$ erases w from v if v has w as a prefix, otherwise $w^{-1}v$ gives 0. Note also that this is a *monoid action* :

$$(vw)^{-1}F = w^{-1}(v^{-1}F).$$

Definition 4.2. *A submodule M of $\mathcal{F}(A)$ is called* stable *if $w^{-1}F \in M$ for all $F \in M$, i.e. $w^{-1}M \subset M$, for all $w \in A^*$.*

Proposition 4.3. *A formal series $F \in \mathcal{F}(A)$ is linearly representable if and only if it is a member of a stable finitely-generated submodule of $\mathcal{F}(A)$*

Proof. Suppose that F is linearly representable by (x, φ, y). For each $i = 1, 2, \cdots n$, (where n is the dimension of the representation) and each $w \in A^*$, define

$$F_i(w) = [\varphi(w)y]_i.$$

Let $M = \langle F_1, \cdots, F_n \rangle$ be the span of the F_i with coefficients in $\mathbb{R}_+$, which is a submodule of $\mathcal{F}(A)$. Since

$$F(w) = x\varphi(w)y = \sum_{i=1}^{n} x_i[\varphi(w)y]_i = \sum_{i=1}^{n} x_i F_i(w),$$

we have that $F = \sum_{i=1}^{n} x_i F_i$, which means $F \in M$.

We next show that M is stable. Let $w \in A^*$. Then for $u \in A^*$,

$$(w^{-1}F_i)(u) = F_i(wu) = [\varphi(wu)y]_i = [\varphi(w)\varphi(u)y]_i$$

$$= \sum_{j=1}^{n} \varphi(w)_{ij}[\varphi(u)y]_j = \sum_{j=1}^{n} \varphi(w)_{ij} F_j(u).$$

Since $\varphi(w)_{ij} \in \mathbb{R}_+$, we have $\sum_{j=1}^{n} \varphi(w)_{ij} F_j(u) \in M$, so

$$w^{-1}F_i = \sum_{j=1}^{n} x_i \varphi(w)_{ij} F_j \in \langle F_1, ...F_n \rangle = M.$$

Conversely, let M be a stable finitely-generated left submodule, and assume that $F \in < F_1, \cdots, F_n >= M$. Then there are $x_1, \cdots, x_n \in \mathbb{R}_+$ such that $F = \sum_{i=1}^{n} x_i F_i$.

Since M is stable, for each $a \in A$ and each $i = 1, 2, \cdots, n,$, we have that $a^{-1}F_i \in < F_1, ...F_n >$. So there exist $c_{ij} \in \mathbb{R}_+, j = 1, 2, \cdots, n$, such that $a^{-1}F_i = \sum_{j=1}^{n} c_{ij} F_j$.

Define $\varphi(a)_{ij} = c_{ij}$, for $i, j = 1, 2, \cdots n$. Extend φ to a monoid morphism from $A^* \to \mathbb{R}_+^{n \times n}$. Because the action of A^* on $\mathcal{F}(A)$ is monoidal (which means $(a_1 a_2)^{-1} = a_2^{-1} a_1^{-1}$), we have that for all $w \in A^*$,

$$w^{-1}F_i = \sum_{j=1}^{n} \varphi(w)_{ij} F_j.$$

Define $y_j = F_j(1)$ for $j = 1, 2, \cdots, n$. Then

$$F_i(w) = w^{-1}F_i(1) = \sum_{j=1}^{n} \varphi(w)_{ij} F_j(1) = \sum_{j=1}^{n} \varphi(w)_{ij} y_j = [\varphi(w)y]_i \text{ for all } i, w.$$

So

$$F(w) = \sum_{i=1}^{n} x_i F_i(w) = \sum_{i=1}^{n} x_i [\varphi(w)y]_i = x\varphi(w)y \text{ for all } w \in A^*,$$

showing that (x, φ, y) is a linear representation for F. □

An immediate consequence of this Proposition is the following Corollary.

Corollary 4.4. *Any finite linear combination of linearly representable series is also a linearly representable series.*

We are now ready to begin the proof of the major theorems of this subject. Our first theorem is the following:

Theorem 4.5 ([12, 17]). *A formal series is linearly representable if and only if it is rational i.e., in the closure of the polynomials under the rational operations + (union), · (concatenation), ∗, and multiplication by elements of* $\mathbb{R}_+$*.*

Proof. We break the proof of this theorem up into several parts. We begin by assuming that F is rational and prove the theorem in that direction. This requires the use of several claims.

Claim 1: Every polynomial is linearly representable.

Proof : If $w \in \mathrm{A}^*$ and $|w|$ is greater than the degree of the polynomial F, then $w^{-1} \equiv 0$. Let $S = \{w^{-1}F : w \in A^*\}$. Then S is finite and stable, hence S spans a finitely-generated stable submodule M to which F belongs. (Take $\epsilon^{-1}F = F$). So by the preceding Proposition, F is linearly representable, and the claim is proved.

Claim 2: If F_1 and F_2 are in stable finitely-generated submodules of $\mathcal{F}(A)$ and $t \in \mathbb{R}_+$, then $(F_1 + F_2)$ and (tF_1) are in finitely-generated submodules.

Proof : This follows immediately from the definition of stability.

Claim 3: Suppose that for $i = 1, 2$, $F_i \in M_i$, where each M_i is a stable, finitely-generated submodule. Let $M = M_1F_2+M_2$. Then M is finitely-generated and stable.

Proof : The fact that M is finitely-generated is immediate. The proof that M is stable is a consequence of the following lemma.

Lemma 4.6. *For* $F, G \in \mathcal{F}(\mathrm{A})$ *and* $a \in A$*,* $a^{-1}(FG) = (a^{-1}F)G + F(\epsilon)a^{-1}G$*.*

Proof. For any $w \in A^*$,

$$(a^{-1}(FG))(w) = (FG)(aw)$$

$$= \sum_{uv=aw} F(u)G(v)$$

$$= F(\epsilon)G(aw) + \sum_{u'v'=w} F(au')G(v')$$

$$= F(\epsilon)G(aw) + \sum_{u'v'=w} (a^{-1}F)(u')G(v')$$

$$= F(\epsilon)(a^{-1}G)(w) + ((a^{-1}F)(G))(w).$$

□

We can now show that M is stable. First, let $f_1F_2 + f_2$ be an arbitrary element of M and let $a \in A$ also be arbitrary. Then

$$a^{-1}(f_1F_2 + f_2) = (a^{-1}f_1)F_2 + f_1(\epsilon)(a^{-1}F_2) + a^{-1}f_2.$$

Note that $a^{-1}f_1 \in M_1$ and $a^{-1}f_2$, $a^{-1}F_2 \in M_2$. Thus $f_1(\epsilon)(a^{-1}F_2) + f_2 \in M_2$, so we conclude that M is stable.

Claim 4: Suppose M_1 is finitely-generated and stable, and that $F_1 \in M_1$ is proper, that is $F_1(\epsilon) = 0$. Then F_1^* is in a finitely-generated stable submodule.

Proof : Define $M = \mathbb{R}_+ + M_1F_1^*$. Since $F_1^* = \sum_{n\geq 0} F_1^n$, we have

$$F_1^* = 1 + \sum_{n\geq 1} F_1^n = (1 + F_1F_1^*) \in M.$$

Also M is finitely generated (generators 1 and $f_iF_1^*$ if the f_i generate M_1). To show that M is stable, we need a lemma.

Lemma 4.7. *If F is proper and $a \in A$, then $a^{-1}(F^*) = (a^{-1}F)F^*$.*

Proof. $a^{-1}(F^*) = a^{-1}(1 + FF^*) = a^{-1}(\epsilon + FF^*) = a^{-1}\epsilon + (a^{-1}F)F^* + F(\epsilon)a^{-1}(F^*)$.

Because $(a^{-1}\epsilon)(w) = \epsilon(aw) = 0$ for all $w \in A^*$ and $F(\epsilon) = 0$, we get that $a^{-1}F^* = (a^{-1}F)F^*$. $\square$

Now the stability of M is an immediate consequence. For if $t \in \mathbb{R}_+$ and $a \in A$, then for any $u \in A^*$ we have $(a^{-1}t)(u) = t(au) = 0$, so $a^{-1}t = 0 \in \mathbb{R}_+$. And for any $f_1 \in F_1$ and $a \in A$, $a^{-1}(f_1^*) = (a^{-1}f_1)F_1^* + f_1(\epsilon)a^{-1}(F_1^*)$. Since M_1 is stable, $a^{-1}f_1 \in F_1^*$ and the first term is in $M_1F_1^*$. By the Lemma, the second term is $f_1(\epsilon)(a^{-1}F_1)F_1^*$, which is again in $M_1F_1^*$.

These four claims give that if F is rational, then F lies in a finitely-generated stable submodule, so the previous Proposition gives that F is linearly representable, and this proves one direction of the Theorem.

Now we turn our attention to proving the theorem in the other direction. So assume that $F \in \mathcal{F}(\mathrm{A})$ is linearly representable. Then $F(w) = x\varphi(w)y$ for all $w \in \mathrm{A}^*$ for some x, φ, y. Consider the semiring of formal series $\mathcal{F}_K(A) = K^{\mathrm{A}^*}$, where K is the semiring $\mathbb{R}_+^{n\times n}$ of $n \times n$ nonnegative real matrices and n is the dimension of the representation. Let $D = \sum_{a\in A}\phi(a)a \in \mathcal{F}_{\mathbb{R}_+^{n\times n}}(A)$. D is proper, so we can form

$$D^* = \sum_{h\geq 0} D^h = \sum_{h\geq 0}(\sum_{a\in A}\varphi(a)a)^h = \sum_{h\geq 0}(\sum_{w\in A^h}\varphi(w)w) = \sum_{w\in \mathrm{A}^*}\phi(w)w.$$

D^* is a rational element of $\mathcal{F}_{\mathbb{R}_+^{n\times n}}(A)$, since we started with a polynomial and formed its *. It can be checked that each entry $(D^*)_{ij}$ is rational in $\mathcal{F}_{\mathbb{R}_+}(A)$. (Start with the $\sum_{a\in A}\phi(a)_{ij}a$ and form sums, products, and *'s.)

So with D and D^* now defined, we have that

$$F(w) = x\phi(w)y = \sum_{i,j} x_i\phi(w)_{ij}y_j = \sum_{i,j} x_iD^*(w)_{ij}y_j,$$

and each $D^*(w)_{ij}$ is a rational series applied to w. So $F(w)$ is a finite linear combination of rational series D^*_{ij} applied to w and hence is rational, which proves this direction of the theorem. $\square$

4.2. **Rational = sofic.**

Theorem 4.8 ([8, 10, 11])**.** *A shift-invariant probability measure ν on $\Omega_+(A)$ corresponds to a rational formal series $F = F_\nu \in \mathcal{F}_{\mathbb{R}_+}(A)$ if and only if it is sofic - the image of a one-step shift-invariant Markov probability measure μ under a one-block map π.*

Proof. Suppose that ν is the image under a one-block map (determined by a map $\pi : A \to B$ between the alphabets) of a one-step Markov measure μ. Then the formal series $F_\mu \in \mathcal{F}(\mathrm{A})$ can be obtained by starting with a polynomial in $\mathcal{F}(\mathrm{A})$ and applying a finite number of rational operations. Replacing in the initial polynomials each $a \in A$ with $\pi(a)$ and applying the same finitely many rational operations will yield a rational series $F_\nu \in \mathcal{F}(B)$ that represents ν.

Alternatively, if F_μ is represented by x, φ, y then for each $w \in A^*$ we have

$$F_\mu(w) = \sum_{i,j} x_i \varphi(w)_{ij} y_j = \sum_{i,j} x_i [\sum_{a \in A} \varphi(a) a]^*(w) y_j.$$

For $u \in B^*$ define

$$F_\nu(u) = \sum_{i,j} x_i \sum_{b \in B} [(\sum_{a \in A, \pi(a) = b} \varphi(a)) b]^*(u) y_j$$

to see that F_ν is a linear combination of rational series and to see its linear representation. The previous example shows how this clumping of the linear representation works in case of Blackwell's examples.

We now prove the other direction. Suppose that ν corresponds to a rational (and hence linearly representable) formal series $F \in \mathcal{F}_{\mathbb{R}_+}(B)$ with dimension n. Let (x, ϕ, y) represent F. Then the theorem will be proved if we can construct a Markov measure μ on some $\Omega_+(A)$ and a one-block map $\pi : \Omega_+(A) \to \Omega_+(B)$ determined by some $\pi : A \to B$ satisfying $\nu = \mu\pi^{-1} = \pi\mu$.

We begin this construction by recalling that our alphabet B has a finite number of letters, say $|B| = d$ and $B = \{b_1, \dots, b_d\}$. From Proposition 3.6, there exists a representation (x, ϕ, y) such that $P = \sum_{b \in B} \varphi(b)$ is a stochastic $n \times n$ matrix, x is a $1 \times n$ probability vector, and $y = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$ is an $n \times 1$ column vector.

We make a $(n \times d \times n) \times (n \times d \times n)$ matrix M in the following manner:

For $n_1, n_2, n_1', n_2' = 1, \ldots, n$ and $r_1, r_2 = 1, \ldots, d$, put

$$M_{n_1 r_1 n_1', n_2 r_2 n_2'} = \begin{cases} \phi(b_{r_2})_{n_2 n_2'} & \text{if } n_2 = n_1' \\ 0 & \text{otherwise.} \end{cases}$$

We "split up" the probability vector x into an $n \times d \times n$-dimensional probability vector X so that for each $n_1' = 1, \ldots, n$, we have $\sum_{\substack{n_1=1,\ldots,n \\ r_1=1,\ldots,d}} x_{n_1 r_1 n_1'} = x_{n_1'}$.

The sum of the $n_1 r_1 n_1$'th row of M is

$$\sum_{n_2 r_2 n_2'} M_{n_1 r_1 n_1', n_2 r_2 n_2'} = \sum_{b, n_2'} M_{n_1 r_1 n_1', n_1' b n_2'}.$$

Since $n_1' = n_2$, we have

$$\sum_{n_2 r_2 n_2'} M_{n_1 r_1 n_1', n_2 r_2 n_2'} = \sum_{b \in B} \sum_{n_2'=1}^{n} \phi(b)_{n_1' n_2'} = \sum_{n_2'=1}^{n} (\sum_{b \in B} \phi(b))_{n_1' n_2'} = \sum_{n_2'=1}^{n} P_{n_1' n_2'}.$$

This is equal to the sum of the n_1''th row of P, which is equal to 1, since P is stochastic. It can also be checked that $XM = X$. (Use $xP = x$, which follows from ν being shift-invariant).

Now define the new alphabet $A = \{n_1 r_1 n_1' : n_1 n_1' = 1, \ldots, n; r_1 = 1, \ldots, d\}$. Let μ be the shift-invariant Markov measure on Ω_+ determined by the initial probability vector X and stochastic transition matrix M.

Claim 1 : $\pi\mu = \nu$, where $\pi : A \to B$ is defined by $\pi(n_1 r_1 n_1') = b_{r_1} \in B$.

Proof : We begin by showing that the claim holds on cylinders of length 1. We have

$$\begin{aligned}
(\mu\pi^{-1})[b]_0 &= \sum_{n_1,n_1'} \sum_{n_2 r_2 n_2'} X_{n_2 r_2 n_2'} M_{n_1 r_2 n_2', n_1 b n_1'} \\
&= \sum_{n_1,n_1'} \sum_{n_2,r_2} X_{n_2 r_2 n_1} M_{n_2 r_2 n_1, n_1 b n_1'} \\
&= \sum_{n_1,n_1'} \varphi(b)_{n_1 n_1'} \sum_{n_2,r_2} X_{n_2 r_2 n_1} \\
&= \sum_{n_1,n_1'} x_{n_1} \phi(b)_{n_1 n_1'} \\
&= x\varphi(b)y \\
&= \nu[b]_0.
\end{aligned}$$

A similar argument will show that for any $w \in B^*$, $(\mu\pi)[w]_0 = x\phi(w)y = \nu[w]_0$. This concludes the proof of the Theorem. □

Remark 4.9. If ν is a possibly not shift-invariant linearly representable probability measure on $\Omega_+(B)$, the above scheme for constructing a (possibly not shift-invariant) Markov measure μ on $\Omega_+(A)$ that projects to ν under a one-block map might not work as stated. The problem is that if ν is represented by (x, φ, y) and some $y_j = 0$, although we may still assume that all $\varphi(w)$ for $w \in B^*$ have j'th column 0, and we may still define Q as above, it is not clear that we may assume that the j'th row of Q is 0 and the j'th entry of x is 0, since they can play a role in the calculation of $x\varphi(w)y = x\varphi(w)Qy$ for some w. This difficulty is due to the lack of relation alluded to at the end of Section 3 between the coordinates $1, 2, \cdots, n$ of the representation and the states of any possible Markov measure above ν. However, as in Proposition 2.6 we may still define $Q_{ij} = 0$ if $x_i y_j = 0$ and

$$Q_{ij} = \lim \left(\frac{1}{k_m} \sum_{k=1}^{k_m} B_{ij}^k\right) \quad \text{if } x_i y_j \neq 0,$$

for a sequence (k_m) such that the limits exist. Then (x, φ, Qy) represents ν, so replace y by Qy. Now if some $y_j = 0$, we may assume that all $\varphi(w)$ have j'th column identically 0. Let $D_{ii} = y_i$ for all i, $D_{ij} = 0$ for $i \neq j$, and $E_{ij} = 1/D_{ij}$ if

$D_{ij} \neq 0$, $E_{ij} = 0$ otherwise. Replacing (x, φ, y) by $\tilde{x} = x$, $\tilde{\varphi} = \varphi D$, $\tilde{y} = Ey$ still leaves us a representation of ν. Proceed with the definition of the alphabet A, the matrix M, and the vector X as above. Define Y by splitting up $\tilde{y}$ similarly, and let

$$\Phi(a)_{n_1 r_1 n_1', n_2 r_2 n_2'} = \begin{cases} M_{n_1 r_1 n_1', n_2 r_2 n_2'} = \varphi(b_{r_2})_{n_2 n_2'} \text{ if } n_2 = n_1' \text{ and } a = n_2 r_2 n_2' \\ 0 \text{ otherwise.} \end{cases}$$

One can verify that

(1) (X, Φ, Y) represents a probability measure μ on $\Omega_+(A)$;
(2) μ maps to ν under $\pi(irj) = b_r$;
(3) μ is Markov with initial probability vector X and transition matrix M.

Now if some $\tilde{y}_j = 0$, whenever $n_2' = j$ we have $\varphi(b_{r_2})_{n_2' r_2} = 0$ for all $n_2 r_2$. Thus $\Phi(a)$ is identically 0 whenever $a = n_2 r_2 j$, so any word that contains a symbol $a = n_2 r_2 j$ will have probability 0, and the symbol a can be discarded from the alphabet A. The difference is that now the states of the Markov measure μ coincide exactly with the coordinates $n_2 r_2 n_2'$ of the representation.

Example 4.10. We return to Blackwell's example to see how to build μ from ν in that case.

$$B = \{a, b\}, n = 3, x = (p_1, p_2, p_3), y = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix},$$

$$\psi(a) = \begin{pmatrix} 0 & 0 & 0 \\ r_1 & 0 & 0 \\ r_2 & 0 & 0 \end{pmatrix}, \psi(b) = \begin{pmatrix} 0 & t_1 & t_2 \\ 0 & s_1 & 0 \\ 0 & 0 & s_2 \end{pmatrix}.$$

Let $r \in \{a, b\},\ i, i', j, j' \in \{1, 2, 3\}$.

$$M_{irj, i'aj'} = \begin{cases} \psi(a)_{21} = r_1 & \text{if } i'j' = 21 \text{ and } j = i' \\ \psi(a)_{31} = r_2 & \text{if } i'j' = 31 \text{ and } j = i' \\ 0 & \text{otherwise.} \end{cases}$$

Note that this matrix works by discarding all states involving a except $2a_1$ and $3a_1$.

We also have $M_{irj,i'bj'} = \begin{cases} \psi(b)_{12} = t_1 & \text{if } i'j' = 12 \text{ and } j = i' \\ \psi(b)_{13} = t_2 & \text{if } i'j' = 13 \text{ and } j = i' \\ \psi(b)_{22} = s_1 & \text{if } i'j' = 22 \text{ and } j = i' \\ \psi(b)_{33} = s_2 & \text{if } i'j' = 33 \text{ and } j = i' \\ 0 & \text{otherwise.} \end{cases}$

This matrix works by discarding all states involving b except $1b_2, 1b_3, 2b_2, 3b_3$. Then we have a 2-block representation of Blackwell.

5. Other approaches and open questions

5.1. **Furstenberg's approach** [8]**.** We start with $F \in \mathcal{F}_{\mathbb{R}}(A)$ representing a shift-invariant probability measure and consider the set $\wp_{\mathbb{R}}(A)$, of *real-valued* polynomials (finitely-supported formal series with *real* coefficients). For $G \in \wp_{\mathbb{R}}(A)$, let $G^+, G^- \in \wp_{\mathbb{R}}(A)$ represent the positive and negative parts of G. Then we have

$$G = \sum_{w \in \mathrm{A}^*} G(w)w = \sum_{w \in \mathrm{A}^*} (G^+(w)w - G^-(w)w).$$

Define

$$|G|_F = \sum_{w \in \mathrm{A}^*} (G^+(w) + G^-(w))F(w),$$

a finite sum. This gives $\wp_{\mathbb{R}}(A)$ the structure of a *normed algebra* over $\mathbb{R}$.

Define $F(G) = \sum_{w \in \mathrm{A}^*} G(w)F(w)$, a continuous linear functional. Then define

$$K = \{G \in \wp_{\mathbb{R}}(A) : \parallel G \parallel_{\mathcal{F}} = 0\},$$

a two-sided ideal of $\mathcal{F}$-negligible words. For if $F(w) = 0$, then since $F(wa) \geq 0$ for all a and

$$F(w) = \sum_{a \in A} F(wa) = \sum_{a \in A} F(aw),$$

then $F(aw) = F(wa) = 0$ for all a in A. Define $\mathcal{A}(F) = \wp_{\mathbb{R}}(A)/K$, a finitely-generated algebra, with generators A and norm inherited from $\parallel \cdot \parallel_F$.

Theorem 5.1. *F corresponds to a (shift-invariant) sofic measure if and only if the quotient module $\mathcal{A}(F)/(\mathcal{A}(F)([\epsilon] - \sum_{a \in A}[a])$, is finite dimensional. (The denominator is a left ideal of $\mathcal{A}(F)$ and [ ] denotes equivalence class in $\mathcal{A}(F)$).*

5.2. **Heller's approach.** E. J. Gilbert [9] and A. Heller [11] developed an approach similar to Furstenberg's to the problem of identifying functions of Markov chains. Again A is a finite alphabet, and $\mathrm{Alg}_{\mathbb{R}}(A)$, the free associative alphabet over $\mathbb{R}$ generated by A, coincides with the set $\wp_{\mathbb{R}}(A)$ of polynomials on A with real coefficients. For a fixed shift-invariant probability measure μ on $\Omega_+(A)$, $F = F_\mu \in \mathcal{F}_{\mathbb{R}_+}(A)$ determines, as in Furstenberg, a linear functional

$$\xi_F(G) = \sum_{w \in A^*} G(w)F(w) \text{ for all } G \in \mathrm{Alg}_{\mathbb{R}}(A).$$

Denote by N the largest left ideal contained in $\ker(\xi_F)$ (so that $N \supset K$, the ideal in Furstenberg's approach).

One can check that

$$N = \{G \in \mathrm{Alg}_{\mathbb{R}}(A) : F(\mathrm{Alg}_R(G)) = 0\}$$

$$= \{G \in \mathrm{Alg}_{\mathbb{R}}(A) : F(A^*G) = 0\}.$$

Definition 5.2. *F (or μ) is* finitary *if $Alg_{\mathbb{R}}(A)/N$ is finite-dimensional over $\mathbb{R}$, as a vector space.*

Theorem 5.3 ([11]). *Every sofic measure is finitary, but there are finitary measures which are not sofic.*

We end with the statement of some open questions.

Question 5.4. *What about the entropy of a sofic measure?*

For a shift-invariant measure μ on $\Omega_+(A)$,

$$h(\mu) = \lim_{n\to\infty}[-\frac{1}{n}\sum_{w \in A^n} \mu[w]_0 \log(\mu[w]_0)].$$

If μ is Markov, determined by p and P, $h(\mu) = -\sum_{i,j} p_i P_{ij} \log P_{ij}$.

What about the entropy of $\nu = \pi\mu$? Blackwell [4] gives an integral formula, and these are procedures to approximate $h(\nu)$ and study its dependence on the parameters defining μ, but much remains unknown.

Question 5.5. *How could one characterize continuous images of countable-state Markov measures, or Gibbs measures* [5], [7]*? Or even Bernoulli measures ?*

References

1. J. Berstel and C. Reutenauer, *Rational Series and Their Languages*, Springer-Verlag, 1988.
2. P. Billingsley, *Probability and Measure*, John Wiley and Sons, Inc., 1995.
3. M. Binkowska and B. Kaminski, *Classification of ergodic finitary shifts*, Ann. Sci. Univ. Clermont-Ferrand II Probab. Appl. (1984), no. 2, 25–37.
4. D. Blackwell, *The entropy of functions of finite state Markov chains*, Trans. First Prague Conf. Inf. Theory, Stat. Dec. Functions, Random Proc., 1957, pp. 13–20.
5. J.-R. Chazottes and E. Ugalde, *Projection of Markov measures may be Gibbsian*, Journal of Stat Physics **111** (2003), 1245–1272.
6. S. W. Dharmadhikari, *Functions of finite Markov chains*, Ann. Math. Stat. **34** (1963), 1022–1032.
7. T. Downarowicz and R.D. Mauldin, *Some remarks on output measures*, Topology Applications **152** (2005), 11–25.
8. H. Furstenberg, *Stationary Processes and Prediction Theory*, Annals of Mathematics Studies, vol. 44, Princeton University Press, 1960.
9. E. J. Gilbert, *On the identifiability problem for functions of finite Markov chains*, Ann. Math. Stat. **30** (1959), 688–697.
10. G. Hansel and D. Perrin, *Rational probability measures*, Theor. Comp. Sci. **65** (1989), 171–188.
11. A. Heller, *On stochastic processes derived from Markov chains*, Ann. Math. Stat. **36** (1965), 1286–1291.
12. S. C. Kleene, *Representation of events in nerve nets and finite automata*, Automata Studies (C. E. Shannon and J. McCarthy, eds.), Princeton University Press, 1956, pp. 3–42.
13. B. Marcus, K. Petersen, and S. Williams, *Transmission rates and factors of Markov chains*, Contemp. Math. **26** (1984), 279–293.
14. K. Petersen, *Symbolic dynamics*, http://www.math.unc.edu/Faculty/petersen/m261s98.pdf, 1998.
15. J. B. Robertson, *The mixing propoerties of certain processes related to Markov chains*, Math. Sys. Th **7** (1973), 39–43.
16. ——, *A spectral representation of the states of a measure preserving transformation*, Z. Wahr. verw. Geb. **27** (1973), 185–194.
17. M. P. Schützenberger, *On the definition of a family of automata*, Inf. and Control **4** (1961), 245–270.

Department of Mathematics, University of North Carolina at Chapel Hill

E-mail address: `petersen@email.unc.edu`