On the classification of Markov chains by finite equivalence

WILLIAM PARRY AND SELIM TUNCEL

Mathematics Institute, University of Warwick, Coventry CV4 7AL, England

In memory of Rufus Bowen (1947–78)

Abstract. We consider a certain analytic function $\beta(t)$ which is an invariant of finite equivalence between two finite state Markov chains. If two such chains P, Q have the same β-function we wish to prove that they are finitely equivalent. To this end we show that $U(t)P^t = Q^t U(t)$ has a nontrivial matrix solution U over the ring $\mathbb{Z}(\exp)$ of integral combinations of exponential functions. In fact we can force $U(t)$ to be strictly positive at any specified t_0. If $U(t)$ has entries from $\mathbb{Z}^+(\exp)$, the sub-semi-ring of positive integral combinations of exponential functions, then P, Q are finitely equivalent. Many examples reinforce the conjecture that $U(t)$ may always be chosen over $\mathbb{Z}^+(\exp)$ when P, Q have the same β-function. We relate the β-function to topological entropy, measure entropy and information variance.

0. *Introduction*

In recent years a number of ergodic theorists have directed their attention to classification problems which differ from the classical one by being restricted to shift spaces (usually with an invariant measure) and by the imposition of state-structure restraints on homomorphisms. We cite [1], [4], [7], [10], [14] and [15] as examples of papers which fall into this category. The first named author, in particular, has investigated a number of problems with the aid of what he has called the information cocycle. In [14] the second named author derived a new invariant by a limiting operation on the information cocycle and used it to investigate very strict forms of isomorphism for Markov chains. This invariant is an analytic function $\beta : \mathbb{R} \to \mathbb{R}^+$. In particular, if two Markov chains are finitely equivalent then they have the same β-function. Our aim is to study the β-function and to provide results which go some way towards establishing the converse of the last statement. We also illustrate the converse with examples which show that the invariant is well fitted (we conjecture perfectly fitted) to finite equivalence.

We start by giving some basic definitions. In § 2 we define the β-function and show that it is an invariant of finite equivalence. Examination of its derivatives reveals that it contains all invariants of finite equivalence known to us. Many of
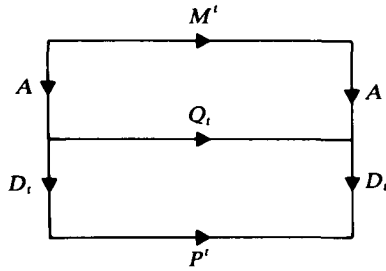
FIGURE 1

the properties established could be derived from the work of Ruelle in [12], but we give direct proofs for the sake of completeness.

Section 3 is purely topological and introduces the ideas and techniques that will be extended later to cover the measure-theoretic aspect of finite equivalence of Markov chains. Apart from the new proof of Furstenberg's lemma and some observations, this section is taken from [7].

In § 4 we confine our attention to so-called dichotomies. These are finite equivalences involving the measure-theoretic analogues of right and left resolving maps of Adler & Marcus [1] (amalgamations and divisions in [7] and [15]). Given two stochastic matrices M, P we ask for a commutative diagram which splits as in figure 1. Here, the entries of M, P are raised to the power of t, Q_t and D_t are variable matrices depending on t, A is an amalgamation matrix and D_t has division shape. This situation obtains precisely when M, P are finitely equivalent through right and left resolving maps. Analytically, it turns out that the commutative diagram (figure 2) can be split precisely when U_t has for its entries non-negative integral combinations of exponential functions. In § 5 we investigate the possibility of obtaining such a diagram when M, P have the same β-function. It turns out that we can always obtain a diagram in which the entries of U_t are integral combinations of exponential functions which reduce to positive integers at $t = 0$. This result is not quite strong enough to establish a completely general classification theory. We have not been able to obtain *non-negative* integral combinations of exponentials. However, in the final section we describe a number of examples which illustrate our theory and which tend to indicate the possibility of a future complete theory.
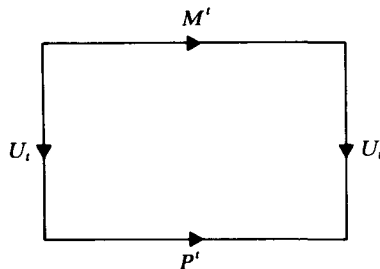

FIGURE 2

Before turning to definitions, a few words about the special nature of dichotomies are in order. It seems that much work needs to be done to understand the nature and structure of bounded-to-one measure preserving continuous maps. Right and left resolving maps are by no means the most general. Adler & Marcus have conjectured in [1], however, that all bounded-to-one continuous semi-conjugacies are compositions of finite sequences of such maps. (A recent unpublished result of B. Kitchens on the zeta function seems to be in line with this conjecture.) Right resolving (and left resolving) maps lie somewhere between the generality of bounded-to-one continuous maps and the speciality of bounded-to-one open maps. Indeed, it is not hard to prove that right resolving maps are (when interpreted on the level of one-sided shifts) precisely the bounded-to-one open maps semi-conjugating a one-sided subshift of finite type to another. (A similar statement holds for left resolving maps.) Openness on the one-sided level does not imply openness on the two-sided level. These few remarks do little to clarify the general structure of semi-conjugating maps. They may be enough, however, to indicate that there is indeed a problem. The special choice of resolving maps in this paper (as in [1] and [7]) is dictated by the availability of relevant matrix theory, in particular the ideas centering on the results of Perron and Frobenius.

1. *Preliminaries*

Let S be an $n \times n$ irreducible 0–1 matrix. Give the set

$$A(S) = \{1, 2, \ldots, n\}$$

the discrete topology and

$$\Sigma = \prod_{-\infty}^{\infty} A(S)$$

the product topology. Consider the subspace X of Σ defined by

$$X = \{x = (x_i) \in \Sigma : S(x_i, x_{i+1}) = 1, \forall i \in \mathbb{Z}\}.$$

The shift S is defined by

$$(Sx)_i = x_{i+1} \quad \text{for } x = (x_i).$$

S is a homeomorphism of the compact, metrizable space X. (X, S) is called the *topological Markov chain (or subshift of finite type)* defined by the matrix S. The fact that we denote the defining matrix and the shift by the same symbol should not lead to any confusion. $A(S)$ is called the *state space (or the alphabet)* of (X, S). If P is an $n \times n$ stochastic matrix compatible with the matrix S (i.e. $P(i, j) = 0$ iff $S(i, j) = 0$), then the (shift invariant) Markov measure defined by P has as its support the topological Markov chain defined by S. We shall always regard Markov measures as being defined on their supporting topological Markov chains. The sets

$$[i_0 \cdots i_l]^m = \{(x_i) \in X : x_m = i_0, x_{m+1} = i_1, \ldots, x_{m+l} = i_l\}$$

$(l, m \in \mathbb{Z}, l \geq 0, i_0, \ldots, i_l \in A(S))$ are *cylinders*. They are closed–open and they form a base for the topology of X. Moreover, finite unions of cylinders give all closed–open subsets of X. Write $[i_0 \cdots i_l]$ for $[i_0 \cdots i_l]^0$. The *state partition* consists of the cylinders $[i]$, $i \in A(S)$.

A topological Markov chain (X, S) is said to be a *finite extension* of another, (Y, T) (and (Y, T) a *finite factor* of (X, S)) if there exists a bounded-to-one continuous surjection $\phi : X \to Y$ with $\phi S = T\phi$. Two topological Markov chains (Y_1, T_1) and (Y_2, T_2) are said to be *finitely equivalent* if they have a common finite extension.

Let (X, S) and (Y, T) be topological Markov chains and let $\phi : X \to Y$ be a continuous map with $\phi S = T\phi$. Then for each cylinder $[j]$, $j \in A(T)$, $\phi^{-1}([j])$ is closed–open, and therefore a finite union of cylinders. This allows us to find integers k, l with $0 \le l < k$ and a map $\phi_0 : A(S)^k \to A(T)$ such that

$$\phi(x)_{n+l} = \phi_0(x_n, x_{n+1}, \ldots, x_{n+k-1})$$

for all $n \in \mathbb{Z}$ whenever $x = (x_n) \in X$. (By composing ϕ with a power of the shift we may, and shall, assume that $l = 0$.) In other words we may find an integer k such that the image of any point is determined by looking at its blocks of length k – such a map is called a *k-block map*. If (Y_1, T_1) and (Y_2, T_2) are finitely equivalent topological Markov chains, we may choose their common extension so that the two factor maps are 1-block maps. (See [3] and [5].)

Let $\phi : X \to Y$ be a 1-block map between topological Markov chains (X, S) and (Y, T). Suppose ϕ is determined by the map $\phi_0 : A(S) \to A(T)$. ϕ is called *right resolving* if, given $(j_0, j_1) \in A(T)^2$ with $T(j_0, j_1) = 1$ and $i_0 \in \phi_0^{-1}(j_0)$, there exists a unique $i_1 \in \phi_0^{-1}(j_1)$ with $S(i_0, i_1) = 1$. ϕ is called *left resolving* if, given $(j_0, j_1) \in A(T)^2$ with $T(j_0, j_1) = 1$ and $i_1 \in \phi_0^{-1}(j_1)$, there exists a unique $i_0 \in \phi_0^{-1}(j_0)$ with $S(i_0, i_1) = 1$. (See [1].)

A Markov chain (X, S, m) is said to be a *finite extension* of another, (Y, T, p) (and (Y, T, p) a *finite factor* of (X, S, m)), if there exists a bounded-to-one continuous measure preserving surjection $\phi : X \to Y$ with $\phi S = T\phi$. Two Markov chains (Y_1, T_1, p_1) and (Y_2, T_2, p_2) are said to be *finitely equivalent* if they have a common Markov finite extension, (X, S, m) say. This may be pictured as in figure 3.
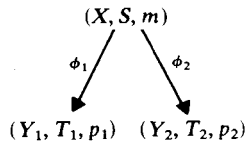

FIGURE 3

Finite equivalence of topological Markov chains was introduced in [7], where topological entropy was shown to be a complete invariant for this equivalence relation. (This proof will be given in § 3.) A refinement of finite equivalence was investigated by Adler & Marcus [1]. A direct proof that finite equivalence of topological Markov chains is an equivalence relation may be found in [1]. Finite equivalence of Markov chains was proved in [14] to be an equivalence relation – the proof is a combination of the topological proof of [1] with the following lemma.

1. LEMMA [14]. *Let (Y, T, p) be a Markov chain and (X, S) a topological Markov chain. If $\phi : X \to Y$ is a bounded-to-one continuous surjection satisfying $\phi S = T\phi$*

then on (X, S) *there is a unique invariant probability* m *which makes* ϕ *measure preserving.* m *is multiple Markov. If* ϕ *is a k-block map then* m *has memory at most* k.

The *information cocycle* of a Markov chain (X, S, m) with state partition α is defined to be

$$I_S = I(\alpha|S^{-1}\alpha) = -\sum_{i \in A(S)} \chi_{[i]} \log m([i]|S^{-1}\alpha).$$

If (X, S, m) is defined by the stochastic matrix P with left invariant probability vector p, $pP = p$, then

$$I_S = -\sum_{i,j \in A(S)} \chi_{[ij]} \log \left(\frac{p(i)P(i,j)}{p(j)}\right).$$

The main result connecting information cocycles and finite factor maps is:

2. THEOREM [8]. *Let* (X, S, m) *and* (Y, T, p) *be Markov chains. If* (Y, T, p) *is a finite factor of* (X, S, m) *by* $\phi : X \to Y$ *then*

$$I_S = I_T \circ \phi + g \circ S - g$$

for some continuous $g : X \to \mathbb{R}$. *If* ϕ *is a k-block map then* g *is a function of (at most)* k *coordinates, i.e.*

$$g(x) = g(x_0, \ldots, x_{k-1}) \quad \text{for } x = (x_n) \in X.$$

2. *The β-function*

Let $P = (P(i, j))$ be an irreducible stochastic matrix. For each $t \in \mathbb{R}$ consider the matrix P^t whose (i, j) entry is $P(i, j)^t$ when $P(i, j) > 0$ and is zero otherwise, and let $\beta_P(t)$ be the maximum eigenvalue of P^t given by the Perron–Frobenius theorem (see [13]). Thus β_P is a positive function $\beta_P : \mathbb{R} \to \mathbb{R}^+$. Let (X, S, m) be the Markov chain defined by P. The *β-function* of (X, S, m) is $\beta_S = \beta_P$. Write $\Sigma_n(I_S)$ for $\sum_{i=0}^{n-1} I_S \circ S^i$.

3. LEMMA. *Let* (X, S, m) *be a Markov chain and let* $t \in \mathbb{R}$. *Then*

$$\lim_{n \to \infty} \frac{1}{n} \log \int \exp \{t\Sigma_n(I_S)\} dm = \log \beta_S(1 - t).$$

Proof. Suppose (X, S, m) is defined by the matrix P with left invariant vector p, $pP = p$. Then

$$\Sigma_n(I_S) = -\sum_{i_0,\ldots,i_n \in A(S)} \chi_{[i_0\cdots i_n]} \log \left(\frac{p(i_0)P(i_0, i_1)P(i_1, i_2) \cdots P(i_{n-1}, i_n)}{p(i_n)}\right)$$

and

$$\int \exp \{t\Sigma_n(I_S)\} dm = \int \sum_{i_0,\ldots,i_n} \chi_{[i_0\cdots i_n]} \left(\frac{p(i_0)P(i_0, i_1) \cdots P(i_{n-1}, i_n)}{p(i_n)}\right)^{-t} dm$$

$$= \sum_{i_0,\ldots,i_n} p(i_0)^{1-t} p(i_n)^t (P(i_0, i_1) \cdots P(i_{n-1}, i_n))^{1-t}.$$

There exist constants k, K such that

$$0 < k \le p(i_0)^{1-t} p(i_n)^t \le K$$

for all $(i_0, i_n) \in A(S)^2$ hence

$$k \sum_{i_0,\dots,i_n} (P(i_0, i_1) \cdots P(i_{n-1}, i_n))^{1-t} \leq \int \exp\{t\Sigma_n(I_S)\}\, dm$$

$$\leq K \sum_{i_0,\dots,i_n} (P(i_0, i_1) \cdots P(i_{n-1}, i_n))^{1-t},$$

and

$$\lim_{n\to\infty} \frac{1}{n} \log \int \exp\{t\Sigma_n(I_S)\}\, dm = \lim_{n\to\infty} \frac{1}{n} \log \sum_{i_0,\dots,i_n} (P(i_0, i_1) \cdots P(i_{n-1}, i_n))^{1-t},$$

provided the last limit exists.

Choose a strictly positive vector v such that

$$P^{1-t}v = \beta_S(1-t)v.$$

This is possible by the Perron–Frobenius theorem. By an argument similar to the above,

$$\lim_{n\to\infty} \frac{1}{n} \log \sum_{i_0,\dots,i_n} (P(i_0, i_1) \cdots P(i_{n-1}, i_n))^{1-t}$$

$$= \lim_{n\to\infty} \frac{1}{n} \log \sum_{i_0,\dots,i_n} P(i_0, i_1)^{1-t} \cdots P(i_{n-1}, i_n)^{1-t} v(i_n)$$

$$= \lim_{n\to\infty} \frac{1}{n} \log \sum_{i_0,\dots,i_{n-1}} P(i_0, i_1)^{1-t} \cdots P(i_{n-2}, i_{n-1})^{1-t}\left(\sum_{i_n} P(i_{n-1}, i_n)^{1-t} v(i_n)\right)$$

$$= \lim_{n\to\infty} \frac{1}{n} \log \sum_{i_0,\dots,i_{n-1}} P(i_0, i_1)^{1-t} \cdots P(i_{n-2}, i_{n-1})^{1-t} v(i_{n-1}) \beta_S(1-t)$$

$$= \lim_{n\to\infty} \frac{1}{n} \log \left(\beta_S(1-t)^n \sum_{i_0} v(i_0)\right) \quad \text{(by repeating the last step } n \text{ times)}$$

$$= \log \beta_S(1-t). \qquad \qed$$

4. PROPOSITION [14]. *Let (X, S, m) and (Y, T, p) be Markov chains. If (Y, T, p) is a finite factor of (X, S, m) then $\beta_S = \beta_T$.*

Proof. Let $\phi : X \to Y$ be the finite factor map. By (2),

$$I_S = I_T \circ \phi + g \circ S - g$$

for some continuous $g : X \to \mathbb{R}$. Note that g is bounded, as X is compact. Using (3) we have, for any $t \in \mathbb{R}$,

$$\log \beta_S(1-t) = \lim_{n\to\infty} \frac{1}{n} \log \int \exp\{t\Sigma_n(I_S)\}\, dm$$

$$= \lim_{n\to\infty} \frac{1}{n} \log \int \exp\{t\Sigma_n(I_T \circ \phi)\} \cdot \exp\{t(g \circ S^n - g)\}\, dm$$

$$= \lim_{n\to\infty} \frac{1}{n} \log \int \exp\{t\Sigma_n(I_T) \circ \phi\}\, dm \quad \text{(since } g \text{ is bounded)}$$

$$= \lim_{n\to\infty} \frac{1}{n} \log \int \exp\{t\Sigma_n(I_T)\}\, dp = \log \beta_T(1-t). \qquad \qed$$

5. COROLLARY [14]. *If the Markov chains* (Y_1, T_1, p_1) *and* (Y_2, T_2, p_2) *are finitely equivalent, then* $\beta_{T_1} = \beta_{T_2}$.

Fix a Markov chain (X, S, m). We remark that

$$\log \beta_S(t) = \mathscr{P}(-tI_S),$$

the pressure of the function $-tI_S$. This is the form in which the invariant was presented in [14]. We have given a direct proof of (3). The result can also be deduced from Ruelle's operator theorem (see [12] and [14]). Next we shall discuss the properties of β_S, and it will be apparent that the function β_S contains all invariants of finite equivalence known to us. In fact we know only two other invariants:

$$h(S) = \int I_S \, dm$$

is the *entropy* of the Markov chain (X, S, m). (2) implies that entropy is an invariant of finite equivalence.

Let v be a real-valued function depending on only two coordinates of X and such that

$$\int v \, dm = 0.$$

Then

$$\sigma^2(v) = \lim_{n \to \infty} \frac{1}{n} \int (\Sigma_n v)^2 \, dm$$

exists (see the next theorem).

$$\sigma^2(S) = \sigma^2(I_S - h(S)) = \lim_{n \to \infty} \frac{1}{n} \int (\Sigma_n I_S - n h(S))^2 \, dm$$

is called the *information variance* of (X, S, m). Information variance was introduced as an invariant in [4]; that it is an invariant of finite equivalence follows from (2) by the results of [4]. For the existence and characterization of information variance we need:

6. THEOREM [9]. *Suppose the Markov chain* (X, S, m) *is defined by an aperiodic matrix* P *with invariant probability vector* p, $pP = p$. *Let* v *be a real-valued function depending on only two coordinates of* X *and such that*

$$\int v \, dm = 0.$$

Then

$$\sigma^2(v) = \int v^2 \, dm + 2 \sum_{i,j} p(i) c(j) P(i,j) v(i,j),$$

where the vector c *is uniquely determined by the properties* $P^n c \to 0$ *(as* $n \to \infty$*) and*

$$(I - P)c = b, \quad b(i) = \sum_j P(i,j) v(i,j).$$

7. THEOREM. *Let the Markov chain* (X, S, m) *be defined by the irreducible stochastic matrix* P *and* u *be a real-valued function depending on two coordinates of* X. *Define*

a positive function $\gamma : \mathbb{R} \to \mathbb{R}^+$ *by taking* $\gamma(t) = \gamma_u(t)$ *to be the maximum eigenvalue of the matrix*

$$(P(i, j) \exp \{tu(i, j)\})$$

given by the Perron–Frobenius theorem. Then γ *is an analytic function,*

$$\gamma(0) = 1, \quad \gamma'(0) = \int u \, dm$$

and

$$\gamma''(0) = \sigma^2\left(u - \int u\right) + \left(\int u\right)^2.$$

Proof. Let P be $k \times k$. Use the Perron–Frobenius theorem to find for each $t \in \mathbb{R}$ a unique probability vector

$$p(t) = (p_1(t), \ldots, p_k(t))$$

such that

$$\sum_i p_i(t)P(i, j) \exp \{tu(i, j)\} = \gamma(t)p_j(t) \quad \text{for all } 1 \le j \le k. \tag{*}$$

That γ is an analytic function follows, for example, from the work of Bhatia & Mukherjea [2]. It is not hard to show that $p(t)$ is the normalization of the first row of the matrix

$$\text{adj} \, (\gamma(t)I - (P(i, j) \exp \{tu(i, j)\}))$$

(see [13]). Since $\gamma(t)$ is analytic, it follows that $p(t)$ is also analytic. $\gamma(0) = 1$ since P is stochastic. Differentiating (*) with respect to t we have

$$\gamma'(t)p_j(t) + \gamma(t)p_j'(t) = \sum_i p_i'(t)P(i, j) \exp \{tu(i, j)\} + \sum_i p_i(t)P(i, j)u(i, j) \exp \{tu(i, j)\}. \tag{**}$$

Put $t = 0$ in (**), sum over j, and observe that, since $\sum_j p_j(t) \equiv 1$, its derivative $\sum_j p_j'(t) \equiv 0$ to obtain

$$\gamma'(0) = \sum_{i,j} p_i'(0)P(i, j) + \sum_{i,j} p_i(0)P(i, j)u(i, j)$$

$$= \sum_{i,j} p_i(0)P(i, j)u(i, j) = \int u \, dm.$$

To prove

$$\gamma''(0) = \sigma^2\left(u - \int u\right) + \left(\int u\right)^2,$$

we assume (for simplicity, the periodic case is similar) that P is aperiodic. Put

$$h = \int u \, dm = \sum_{i,j} p_i(0)P(i, j)u(i, j)$$

and

$$v(i, j) = u(i, j) - h.$$

Take

$$b_i = \sum_j P(i, j)v(i, j)$$

and use (6) to write

$$(I - P)c = b,$$

where c is uniquely determined by $P^n c \to 0$ (as $n \to \infty$). Clearly

$$\sum_i p_i(0)b_i = \int v = 0.$$

Since P is aperiodic

$$P^n \to \begin{pmatrix} p(0) \\ \vdots \\ p(0) \end{pmatrix},$$

the matrix with identical rows $p(0)$. Hence

$$P^n c \to \left(\sum_i p_i(0)c_i \right) \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = 0$$

i.e.

$$\sum_i p_i(0)c_i = 0.$$

Now using (**) we have

$$\sum_j p_j'(0)c_j = \sum_j \gamma(0)p_j'(0)c_j + \gamma'(0) \sum_j p_j(0)c_j$$

$$= \sum_{i,j} p_i'(0)P(i, j)c_j + \sum_{i,j} p_i(0)P(i, j)u(i, j)c_j,$$

i.e.

$$\sum_{i,j} p_i(0)P(i, j)u(i, j)c_j = \langle p'(0), (I - P)c \rangle = \langle p'(0), b \rangle. \qquad (***)$$

Differentiate (**) with respect to t, put $t = 0$, sum over j, and observe that $\sum_j p_j''(t) \equiv 0$ also to obtain

$$\gamma''(0) = 2 \sum_{i,j} p_i'(0)P(i, j)u(i, j) + \sum_{i,j} p_i(0)P(i, j)u(i, j)^2$$

$$= 2 \sum_i p_i'(0)(b_i + h) + \int u^2 \quad \text{(by the definition of } b_i\text{)}$$

$$= 2 \sum_i p_i'(0)b_i + \int u^2$$

$$= 2 \sum_{i,j} p_i(0)P(i, j)u(i, j)c_j + \int u^2 \quad \text{(by (***))}$$

$$= 2 \sum_{i,j} p_i(0)P(i, j)(v(i, j) + h)c_j + \int (v + h)^2$$

$$= 2 \sum_{i,j} p_i(0)P(i, j)v(i, j)c_j + 2 \sum_{i,j} p_i(0)P(i, j)c_j h + \int v^2 + h^2$$

$$= \sigma^2(v) + h^2 + 2h \sum_j p_j(0)c_j \quad \text{(by (6))}$$

$$= \sigma^2(v) + h^2. \qquad \square$$

8. COROLLARY. *Let* (X, S, m) *be a Markov chain.* β_S *is an analytic function,*

$$\beta_S(1) = 1, \quad \beta_S'(1) = -h(S)$$

and

$$\beta_S''(1) = \sigma^2(S) + h(S)^2.$$

If X does not consist of a single orbit, $\beta_S(t)$ is strictly decreasing to zero as $t \to \infty$.

Proof. Take $u(i, j) = -\log P(i, j)$ in (7) so that

$$\gamma(t) = \beta_S(1 - t),$$

and apply the theorem. For the last part, let $t_0 < t_1$. Then

$$P^{t_0} \geq P^{t_1}$$

and, if X does not consist of a single orbit, equality does not hold. That

$$\beta_S(t_0) > \beta_S(t_1)$$

is then a consequence of the Perron–Frobenius theorem (see [13]). We omit the simple proof that $\beta_S(t) \to 0$ as $t \to \infty$.   □

We remark that in (7)

$$\log \gamma(t) = P(-I_S + tu),$$

the pressure of the function $-I_S + tu$. Using this observation, (7) and (8) may also be deduced from the work of Ruelle [12].

3. *Finite equivalence of topological Markov chains*

As we have already remarked, it was proved in [7] that topological entropy is a complete invariant for finite equivalence of topological Markov chains. In the next section we shall refine the techniques of [7] to (stochastic) Markov chains. We now study the construction of the (topological) finite equivalences in [7].

The main argument of [7] relies on a lemma of H. Furstenberg. We give a new proof of this lemma. The proof, which is algebraic, will be extended later. (See also [1] for an alternative proof.)

9. LEMMA (Furstenberg). *Let S, T be irreducible non-negative integral matrices. S, T have the same maximum eigenvalue iff there exists a strictly positive integral matrix U with $US = TU$.*

Proof. It is easy to see by using the Perron–Frobenius theorem that if $US = TU$ for some strictly positive integral matrix U then S, T have the same maximum eigenvalue. Conversely, suppose S and T have the same maximum eigenvalue, β. Let χ_S, χ_T be the characteristic polynomials of S, T:

$$\chi_S(\lambda) = \det(\lambda I - S), \quad \chi_T(\lambda) = \det(\lambda I - T).$$

Let $s(\lambda)$ be the first row of adj $(\lambda I - S)$, and let $\tau(\lambda)$ be the first column of adj $(\lambda I - T)$. The entries of $\tau(\lambda)$, $s(\lambda)$ and $\tau(\lambda) \cdot s(\lambda)$ belong to $\mathbb{Z}[\lambda]$, the ring of polynomials in λ with integer coefficients. Moreover, $\tau(\beta)$ and $s(\beta)$ are strictly positive vectors (see [13]). Since

$$\text{adj} (\lambda I - S)S = \lambda \text{ adj} (\lambda I - S) - \chi_S(\lambda)I,$$
$$T \text{ adj} (\lambda I - T) = \lambda \text{ adj} (\lambda I - T) - \chi_T(\lambda)I,$$

we have

$$\tau(\lambda)s(\lambda)S - T\tau(\lambda)s(\lambda) = \chi_T(\lambda)B(\lambda) - \chi_S(\lambda)A(\lambda), \qquad (*)$$

where the matrices A, B are over $\mathbb{Z}[\lambda]$.

We claim that there exists a monic polynomial $\chi \in \mathbb{Z}[\lambda]$ such that $\chi(\beta) = 0$ and

$$\chi_S = u\chi, \quad \chi_T = v\chi$$

for some u, $v \in \mathbb{Z}[\lambda]$. Using $\chi_S(\beta) = \chi_T(\beta) = 0$ and the Euclidean properties of $\mathbb{Q}[\lambda]$ we can assert

$$a\chi_S = u'\chi, \quad b\chi_T = v'\chi,$$

where $a, b \in \mathbb{Z}$, $u', v', \chi \in \mathbb{Z}[\lambda]$, χ is primitive and $\chi(\beta) = 0$. Since $\mathbb{Z}[\lambda]$ is a unique factorization domain and χ is primitive, we see that $a \mid u'$ and $b \mid v'$ in $\mathbb{Z}[\lambda]$. Let $u, v \in \mathbb{Z}[\lambda]$ satisfy

$$au = u', \quad bv = v',$$

and note that χ_S is monic to complete the proof of our claim. Let $d \geq 1$ be the degree of χ. Use the fact that χ is monic to reduce the polynomials in $\tau(\lambda)s(\lambda)$ modulo χ to obtain a matrix $V(\lambda)$ over $\mathbb{Z}[\lambda]$. Since $\chi \mid \chi_S$ and $\chi \mid \chi_T$ in $\mathbb{Z}[\lambda]$, we obtain from (*) an equation

$$V(\lambda)S - TV(\lambda) = \chi(\lambda)C(\lambda),$$

where the matrix $C(\lambda)$ is also over $\mathbb{Z}[\lambda]$. Write

$$V(\lambda) = V_0 + \lambda V_1 + \cdots + \lambda^{d-1} V_{d-1},$$

V_i over $\mathbb{Z}$ for $i = 0, 1, \ldots, d-1$. Since $\chi(\lambda)$ has degree d, C must be the zero matrix and we have

$$V_i S = T V_i \quad \text{for } i = 0, 1, \ldots, d-1.$$

Let σ be a strictly positive vector such that $S\sigma = \beta\sigma$. There exists an i, $0 \leq i \leq d-1$, such that $V_i\sigma \neq 0$, for otherwise we would have $0 = \tau(\beta)s(\beta)\sigma$, a product of strictly positive matrices. The matrix U we require is obtained by modifying V_i. We assume that S, T are aperiodic; the periodic case is similar. By Perron–Frobenius theory, S^n/β^n tends to $\sigma s(\beta)$ as $n \to \infty$. Thus each row of $V_i S^n/\beta^n$ tends to a constant multiple of $s(\beta)$. Similarly, each column of

$$V_i S^n/\beta^n = T^n V_i/\beta^n$$

tends to a constant multiple of $\tau(\beta)$. It follows that either $V_i S^n/\beta^n$ tends to the zero matrix, or the entries of $V_i S^n$ are all non-zero and have the same sign for large n. But, since

$$V_i S^n \sigma/\beta^n = V_i\sigma \neq 0,$$

$V_i S^n/\beta^n$ cannot tend to the zero matrix. Hence we may find n such that $V_i S^n$ is strictly positive or strictly negative. Accordingly, take

$$U = V_i S^n \quad \text{or} \quad U = -V_i S^n. \qquad \square$$

As in [7], we shall use Williams' technique of splitting non-negative integral matrices into products of division and amalgamation matrices.

A rectangular 0–1 matrix is called a *division matrix* if its rows are non-trivial and each column contains exactly one non-zero entry. A 0–1 matrix is called an *amalgamation matrix* if its transpose is a division matrix.

10. PROPOSITION [**15**]. *If N is a non-trivial non-negative integral matrix, then it can be written as $N = DA$, where D is a division matrix and A is an amalgamation matrix. This decomposition into the product of a division matrix with an amalgamation matrix is essentially unique in the sense that, if $N = D'A'$ also, then $D' = DR$, $A' = R^{-1}A$ for some permutation matrix R.*

If (X, S) is a topological Markov chain, the topological entropy of (X, S) is given by $\log \beta$, where β is the maximum eigenvalue of the matrix S. We shall use the following lemma (see [**3**], [**7**] or [**8**]):

11. LEMMA. *If (X, S), (Z, W) are topological Markov chains and if $\phi : Z \to X$ is a continuous surjection with $\phi W = S\phi$ then ϕ is bounded-to-one iff the matrices S, W have the same maximum eigenvalue.*

12. PROPOSITION [**7**]. *Let (X, S) and (Z, W) be topological Markov chains. If their defining matrices satisfy $AS = WA$ for some amalgamation matrix A then (X, S) is a finite factor of (Z, W) by a right resolving 1-block map.*

Proof. Define $\phi_0 : A(W) \to A(S)$ by putting $\phi_0(z_0) = x_0$ whenever $A(z_0, x_0) = 1$. Since A is an amalgamation, for each $z_0 \in A(W)$ there exists a unique $x_0 \in A(S)$ such that $A(z_0, x_0) = 1$, and the map ϕ_0 is well defined. Note that since S is 0–1 and A an amalgamation, $AS = WA$ is 0–1. If $W(z_0, z_1) = 1$ then

$$W(z_0, z_1)A(z_1, \phi_0 z_1) = 1$$

so that

$$1 = (WA)(z_0, \phi_0 z_1) = (AS)(z_0, \phi_0 z_1)$$
$$= A(z_0, \phi_0 z_0)S(\phi_0 z_0, \phi_0 z_1).$$

Thus,

$$W(z_0, z_1) = 1 \quad \text{implies} \quad S(\phi_0 z_0, \phi_0 z_1) = 1$$

and the 1-block map $\phi : Z \to X$,

$$\phi(z) = \{\phi_0(z_i)\} \quad \text{for } z = \{z_i\},$$

is well defined. Clearly $\phi W = S\phi$. Given $x_0, x_1 \in A(S)$ with $S(x_0, x_1) = 1$ and $z_0 \in A(W)$ with $\phi_0 z_0 = x_0$ we have

$$1 = A(z_0, x_0)S(x_0, x_1) = (AS)(z_0, x_1)$$
$$= (WA)(z_0, x_1) = W(z_0, z_1)A(z_1, x_1)$$

for a uniquely defined $z_1 \in A(W)$. It follows that ϕ is right resolving. Surjectivity follows from the right resolving property. Let $x = (x_i) \in X$. For each $n \geq 0$ use the right resolving property to find $z^{(n)} \in Z$ such that

$$(\phi z^{(n)})_i = x_i \quad \text{for } i \geq -n.$$

Any limit point z of $z^{(n)}$ has $\phi z = x$. Finally, to see that ϕ is bounded-to-one note that $AS = WA$ implies that S, W have the same maximum eigenvalue, and use (11). $\qquad\square$

Interchanging the roles of rows and columns in the above proof we obtain:

13. PROPOSITION [**7**]. *Let* (Y, T) *and* (Z, W) *be topological Markov chains. If their defining matrices satisfy* $TD = DW$ *for some division matrix* D, *then* (Y, T) *is a finite factor of* (Z, W) *by a left resolving* 1-*block map.*

It is not hard to see that (12) gives all right resolving 1-block finite factor maps. Let (X, S), (Z, W) be topological Markov chains and let (X, S) be a finite factor of (Z, W) by a right resolving 1-block map $\phi : Z \to X$, given by $\phi_0 : A(W) \to A(S)$. Use ϕ_0 to define an amalgamation A_0 by putting

$$A_0(z_0, x_0) = 1 \quad \text{iff } x_0 = \phi_0 z_0.$$

Then $A_0 S = WA_0$ and ϕ is simply the map obtained from $A_0 S = WA_0$ by (12). Similarly, if the topological Markov chain (Y, T) is a finite factor of (Z, W) by a left resolving 1-block map ψ, then there exists a division matrix D_0 such that $TD_0 = D_0 W$ and ψ is the map obtained from $TD_0 = D_0 W$ by (13). Consequently, if (Y, T) and (X, S) have a common finite extension (Z, W) by 1-block maps $\psi : Z \to Y$ and $\phi : Z \to X$ which are, respectively, left and right resolving, then there exists a non-trivial non-negative integer matrix U satisfying $US = TU$. (Take $U = D_0 A_0$ above.)

14. THEOREM [**7**]. *Two topological Markov chains are finitely equivalent iff they have the same topological entropy* (*i.e. iff their defining matrices have the same maximum eigenvalue*).

Proof. Finitely equivalent topological Markov chains have the same topological entropy as a result of (11). For the converse, let (X, S), (Y, T) be topological Markov chains such that the matrices S, T have the same maximum eigenvalue. Let U be a non-trivial non-negative integral matrix satisfying $US = TU$. By (9), such matrices exist. We shall also insist that U is minimal in the following sense: if U' is a non-trivial integral matrix such that $0 \leq U' \leq U$ and $U'S = TU'$, then $U' = U$. Use (10) to decompose $U = DA$, $S = D_1 A_1$, $T = D_2 A_2$ into products of division matrices with amalgamation matrices and rewrite $AD_1 = D_3 A_3$, $A_2 D = D_4 A_4$ to obtain the commutative diagram in figure 4.
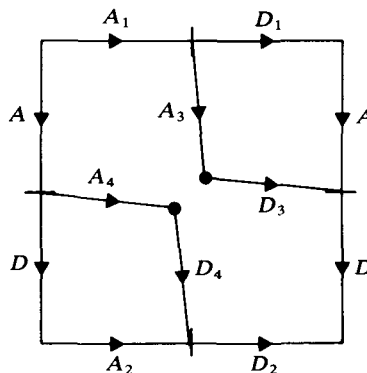

FIGURE 4

Now $DD_3A_3A_1$ and $D_2D_4A_4A$ are two decompositions of $US = TU$. (Note that
the product of two divisions (resp. amalgamations) is a division (resp. amalgama-
tion).) By (10), replacing D_4, A_4 by D_4R, $R^{-1}A_4$, where R is a suitable permutation
matrix, we may assume that

$$DD_3 = D_2D_4 \quad \text{and} \quad A_3A_1 = A_4A.$$

In other words we may assume that figure 5 is commutative. Defining $W = D_3A_4$,
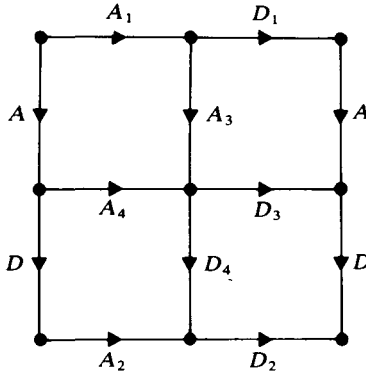we see that $AS = WA$ and $TD = DW$. Since A is an amalgamation and S is 0–1,

FIGURE 5

$AS = WA$ and W must be 0–1 matrices. The minimality condition on U ensures
that W is irreducible. Suppose

$$W = \begin{pmatrix} W_a & 0 \\ B & W_b \end{pmatrix},$$

where W_a is irreducible. Accordingly, write

$$A = \begin{pmatrix} A_a \\ A_b \end{pmatrix}.$$

Then

$$A_aS = W_aA_a$$

so that W_a has the same maximum eigenvalue as S and T, β say. Now write

$$D = (D_a \quad D_b)$$

to obtain

$$D_aW_a + D_bB = TD_a,$$

and let v be a strictly positive vector such that

$$vT = \beta v.$$

Then

$$(vD_a)W_a + v(D_bB) = \beta(vD_a)$$

so that

$$(vD_a)W_a \le \beta(vD_a).$$

The Perron–Frobenius theorem shows that

$$(vD_a)W_a = \beta(vD_a)$$

i.e.

$$v(D_bB) = 0.$$

It follows that

$$B = 0 \quad \text{and} \quad D_aW_a = TD_a.$$

Now taking $U' = D_aA_a$ we obtain a contradiction to the minimality of U, unless $W = W_a$. Hence W is irreducible and we may apply (12) and (13) to conclude that the topological Markov chain it defines is a common finite extension of (X, S) and (Y, T). $\qquad\square$

Let (X, S), (Y, T) be topological Markov chains of the same topological entropy. The proof of (14) provides a finite equivalence (figure 6), where the map ϕ is right resolving and ψ is left resolving, i.e. we have a dichotomy. The next section is devoted to the measure-theoretic analogue of this situation.
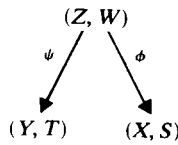

FIGURE 6

4. *Dichotomy of Markov chains*

Let (X, S, m), (Y, T, p) be Markov chains. Consider the following dichotomy:

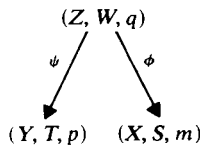(a) (X, S, m) and (Y, T, p) are finitely equivalent by a diagram (figure 7),

FIGURE 7

where ϕ, ψ are 1-block maps, ϕ is right resolving and ψ is left resolving.

In this section we develop coding to establish necessary and sufficient conditions for (a) to hold.

15. LEMMA. *Let (Y, T, p), (Z, W, q) be Markov chains defined by the stochastic matrices P, Q respectively. Let $\psi : Z \to Y$ be a bounded-to-one 1-block surjection with $\psi W = T\psi$. Suppose ψ is given by $\psi_0 : A(W) \to A(T)$. Define a matrix R compatible with Q by setting*

$$R(i, j) = \begin{cases} 0 & \text{if } Q(i, j) = 0, \\ P(\psi_0 i, \psi_0 j) & \text{if } Q(i, j) > 0. \end{cases}$$

Then R has maximum eigenvalue 1. Let ρ be a strictly positive right invariant vector, $R\rho = \rho$, and let P' be the stochastic matrix with

$$P'(i, j) = \frac{R(i, j)\rho(j)}{\rho(i)}.$$

ψ is measure preserving iff $Q = P'$.

Proof. By (1), on the topological Markov chain (Z, W) there is a unique Markov measure, given by the stochastic matrix Q_1, say, for which ψ is measure preserving. By (2), there is a function g depending on only one coordinate such that

$$-\log Q_1(i, j) = -\log P(\psi_0 i, \psi_0 j) + g(j) - g(i)$$

whenever $Q_1(i, j) > 0$. Exponentiate, note that Q and Q_1 are compatible matrices, and use the definition of R to obtain

$$Q_1(i, j) = \frac{R(i, j)\rho(j)}{\rho(i)}$$

for some strictly positive $\rho : A(W) \to \mathbb{R}$. Summing over j,

$$\rho(i) = \sum_j R(i, j)\rho(j),$$

so that R has maximum eigenvalue 1, ρ is a corresponding eigenvector and $Q_1 = P'$. By the uniqueness in lemma 1, as a map between the Markov chains (Y, T, p) and (Z, W, q), ψ is measure preserving iff $Q = Q_1$, i.e. iff $Q = P'$. □

If M is a non-negative matrix and $t \in \mathbb{R}$, M^t will denote the matrix whose entries are the t'th powers of the entries of M. More precisely, the (i, j) entry of M^t is $M(i, j)^t$ when $M(i, j) > 0$ and is zero otherwise. A rectangular non-negative matrix D is said to have *division shape* if the compatible 0–1 matrix D^0 is a division matrix. A rectangular non-negative matrix D is called a *stochastic division matrix* if it is stochastic and if it has division shape. Amalgamation matrices are stochastic by definition. Amalgamation and stochastic division matrices characterize all 1-block right resolving and left resolving finite factor maps between Markov chains:

16. LEMMA. *Let (X, S, m) and (Z, W, q) be Markov chains defined by the stochastic matrices M and Q respectively. If there exists an amalgamation matrix A such that*

$$AM^0 = Q^0 A \quad \text{and} \quad AM = QA$$

then (X, S, m) is a finite factor of (Z, W, q) by a right resolving 1-block map.

Proof. Define $\phi_0 : A(W) \to A(S)$ by taking

$$\phi_0(z_0) = x_0 \quad \text{when} \quad A(z_0, x_0) = 1.$$

According to (the proof of) (12), ϕ_0 gives a well-defined 1-block map ϕ which is a right resolving bounded-to-one surjection of the topological Markov chain (Z, W) onto (X, S). Let $(i, j) \in A(W)^2$. Since A is an amalgamation and $AM^0 = Q^0 A$ is 0–1, we have for all allowable (i, j)

$$Q(i, j) = Q(i, j)A(j, \phi_0 j)$$
$$= (QA)(i, \phi_0 j) = (AM)(i, \phi_0 j) = M(\phi_0 i, \phi_0 j),$$

and (15) shows that ϕ is measure preserving. □

17. LEMMA. *Let (Y, T, p) and (Z, W, q) be Markov chains defined by the stochastic matrices P, Q respectively. If there exists a stochastic division matrix D such that*

$$P^0 D^0 = D^0 Q^0 \quad and \quad PD = DQ$$

then (Y, T, p) is a finite factor of (Z, W, q) by a left resolving 1-block map.

Proof. Define $\psi_0 \colon A(W) \to A(T)$ by taking $\psi_0(z_0) = y_0$ when $D(y_0, z_0) > 0$. According to (13), ψ_0 gives a well-defined 1-block map ψ which is a left resolving bounded-to-one surjection of (Z, W) onto (Y, T). Let $(i, j) \in A(W)^2$. Since D^0 is a division matrix and P^0 is 0–1, $P^0 D^0 = D^0 Q^0$ is a 0–1 matrix. Using this fact,

$$D(\psi_0 i, i) Q(i, j) = (DQ)(\psi_0 i, j)$$
$$= (PD)(\psi_0 i, j) = P(\psi_0 i, \psi_0 j) D(\psi_0 j, j),$$

and, taking $f(i) = D(\psi_0 i, i)$, we see that for all allowable (i, j),

$$Q(i, j) = \frac{P(\psi_0 i, \psi_0 j) f(j)}{f(i)}.$$

It follows that f is a strictly positive eigenvector of the matrix

$$R(i, j) = P(\psi_0 i, \psi_0 j)$$

compatible with Q, and (15) shows that ψ is measure preserving. □

Companions (converses) to (16) and (17) are the following two lemmas:

18. LEMMA. *Let (X, S, m) and (Z, W, q) be Markov chains defined by the stochastic matrices M and Q. Suppose (X, S, m) is a finite factor of (Z, W, q) by a right resolving 1-block map ϕ. Then there exists an amalgamation matrix A such that*

$$AM^t = Q^t A$$

for all $t \in \mathbb{R}$ and ϕ is the map obtained from $AM^0 = Q^0 A$, $AM = QA$ by (16).

Proof. Suppose ϕ is given by the map $\phi_0 \colon A(W) \to A(S)$. Index by $A(W)$ and $A(S)$ the rows and columns respectively of an amalgamation matrix A by putting

$$A(i, j) = 1 \quad \text{iff } j = \phi_0 i.$$

Since ϕ is right resolving, the matrix R,

$$R(i_0, i_1) = \begin{cases} 0, & \text{if } Q(i_0, i_1) = 0, \\ M(\phi_0 i_0, \phi_0 i_1), & \text{otherwise,} \end{cases}$$

is stochastic and

$$Q(i_0, i_1) = M(\phi_0 i_0, \phi_0 i_1)$$

for all allowable $(i_0, i_1) \in A(W)^2$. But

$$(AM^t)(i, j) = M^t(\phi_0 i, j)$$

and

$$(Q^t A)(i, j) = \sum_{i_1} Q^t(i, i_1) A(i_1, j)$$
$$= Q^t(i, i_1)$$

for a unique i_1 with $\phi_0 i_1 = j$, and the result follows. □

19. LEMMA. *Let (Y, T, p) and (Z, W, q) be Markov chains defined by the stochastic matrices P and Q. Suppose (Y, T, p) is a finite factor of (Z, W, q) by a left resolving 1-block map ψ. Then there exists a stochastic division matrix D such that*

$$P'D' = D'Q'$$

for all $t \in \mathbb{R}$ and ψ is the map obtained from $P^0 D^0 = D^0 Q^0$, $PD = DQ$ by (17).

Proof. Suppose ψ is given by the map $\psi_0 : A(W) \to A(T)$. Let p and q be the left invariant probability vectors of P and Q, $pP = p$ and $qQ = q$. (Usage of the same symbol for the invariant vector and Markov measures defined by the matrices should not lead to any confusion.) Define D by putting

$$D(\psi_0 i, i) = \frac{q(i)}{p(\psi_0 i)}$$

and taking

$$D(j, i) = 0 \quad \text{when } j \neq \psi_0 i.$$

D is a stochastic division matrix, since ψ is measure preserving. We show $P'D' = D'Q'$ for all $t \in \mathbb{R}$.

Define a diagonal matrix $\tilde{q}$ by putting

$$\tilde{q}(i, i) = q(i)$$

and let

$$Q^* = \tilde{q}^{-1} Q^{\mathrm{tr}} \tilde{q},$$

where Q^{tr} is the transpose of Q. Then Q^* is a stochastic square matrix and it defines a Markov chain (Z^*, W^*, q^*), which is the inverse (shift) of (Z, W, q). Similarly, define a diagonal matrix $\tilde{p}$ by putting

$$\tilde{p}(j, j) = p(j),$$

let

$$P^* = \tilde{p}^{-1} P^{\mathrm{tr}} \tilde{p}$$

and let (Y^*, T^*, p^*) be the Markov chain defined by P^*. Then (Y^*, T^*, p^*) is a finite factor of (Z^*, W^*, q^*) by a right resolving 1-block map defined by ψ_0. Applying (18) we see that the amalgamation A defined by setting

$$A(i, j) = 1 \quad \text{iff } j = \psi_0 i$$

satisfies

$$A(P^*)^t = (Q^*)^t A$$

for all $t \in \mathbb{R}$, i.e.

$$A\tilde{p}^{-t}(P^t)^{\mathrm{tr}}\tilde{p}^t = \tilde{q}^{-t}(Q^t)^{\mathrm{tr}}\tilde{q}^t A$$

for all $t \in \mathbb{R}$. Transposing, we see that

$$D = \tilde{p}^{-1} A^{\mathrm{tr}} \tilde{q}$$

satisfies

$$P'D' = D'Q'$$

for all $t \in \mathbb{R}$. □

Consider the product D^tA where D is a stochastic division matrix and A is an amalgamation matrix. D^tA is non-trivial and the entries of D^tA are non-negative integral combinations of exponentials, i.e. its non-zero entries are finite sums

$$\sum_k n_k \exp(a_k t), \quad a_k \in \mathbb{R}, \quad n_k \in \mathbb{Z}, \quad n_k > 0.$$

Let $U = U_t$ be a matrix whose entries are functions of $\mathbb{R}$. U is called a *non-negative exponential matrix* if all its entries are non-negative integral combinations of exponential functions.

Let M, P be non-negative matrices. Consider the statement:

(b) There exists a non-trivial non-negative exponential matrix $U = U_t$ such that $U_tM^t = P^tU_t$ for all $t \in \mathbb{R}$.

We may combine (18) and (19) to obtain:

20. PROPOSITION. *For Markov chains (X, S, m), (Y, T, p) defined respectively by the stochastic matrices M, P, (a) implies (b).*

In fact, we shall see that (a) and (b) are equivalent. Williams' decomposition result for non-negative integral matrices, (10), may be extended to non-negative exponential matrices:

21. PROPOSITION. *If $U = U_t$ is a non-trivial non-negative exponential matrix, then there exists an amalgamation matrix A and a non-negative matrix D with division shape such that $U_t = D^tA$ for all $t \in \mathbb{R}$. This decomposition is essentially unique in the sense that if $U_t = D''^tA'$ (for all $t \in \mathbb{R}$), where A' is an amalgamation and D' has division shape, then $D' = DR$, $A' = R^{-1}A$ for some permutation matrix R.*

Proof. First consider the non-trivial non-negative integer matrix U_0. Use (10) to write $U_0 = D_0A$, where D_0 is a division matrix and A is an amalgamation matrix. We shall define a non-negative matrix D such that $D^0 = D_0$ and $U_t = D^tA$ for all $t \in \mathbb{R}$. Take

$$D(i, k) = 0 \quad \text{if } D_0(i, k) = 0.$$

Suppose (i, j) is such that for some $n \geq 1$,

$$U_t(i, j) = \exp(a_1 t) + \cdots + \exp(a_n t),$$

where $a_1, \ldots, a_n \in \mathbb{R}$ are not necessarily distinct. Then

$$U_0(i, j) = n$$

and from

$$n = U_0(i, j) = \sum_k D_0(i, k)A(k, j)$$

we see that in the i'th row of D_0 there are precisely n places (i, k) with $D_0(i, k) = 1$ which are 'picked up' by the j'th column of A (i.e. such that $A(k, j) = 1$ also). Since A is an amalgamation, no other column of A 'picks up' the same places. Assign $\exp a_1, \ldots, \exp a_n$ one by one to these n places. Note that, again because A is an amalgamation, each place with $D_0(i, k) = 1$ is picked up by some column of A. It follows that when we have gone through all (i, j), the matrix D is fully and well defined. It is clear that $D^0 = D_0$ and that $U_t = D^tA$ for all $t \in \mathbb{R}$.

Now let $U_t = D''A'$ be another decomposition where A' is an amalgamation matrix and D' is a non-negative matrix of division shape. Putting $t = 0$ and using the uniqueness in (10) we see that, by supplementing with a permutation, we may assume $D'^0 = D^0$, $A' = A$. In other words, we may take $D''A$ to be the new decomposition and $D'^0 = D^0$. We shall show that there is a permutation matrix R such that $D' = DR$, $D^0 = D^0R$ and $A = R^{-1}A$.

Let (i, j) satisfy $U_0(i, j) = n > 0$. Using

$$U_t = D''A = D'A$$

and the fact that

$$D'^0 = D^0$$

we see that there are $k_1, \ldots, k_n$ such that

$$D^0(i, k_1) = \cdots = D^0(i, k_n) = 1 = A(k_1, j) = \cdots = A(k_n, j)$$

and

$$D(i, k_1)^t + \cdots + D(i, k_n)^t = D'(i, k_1)^t + \cdots + D'(i, k_n)^t$$

for all $t \in \mathbb{R}$. It follows that there exists a permutation Π of $\{1, \ldots, n\}$ such that

$$D'(i, k_{\Pi(l)}) = D(i, k_l)$$

for all $l \in \{1, \ldots, n\}$. Let R_{ij} be the matrix which permutes columns $k_1, \ldots, k_n$ of D' according to Π, and fixes all other columns. It is clear that

$$D^0 R_{ij} = D^0, \quad R_{ij}^{-1} A = A$$

and that

$$(D' R_{ij})(i, k_l) = D(i, k_l)$$

for all $l \in \{1, \ldots, n\}$. As D^0 is a division matrix and A an amalgamation, for any $l \in \{1, \ldots, n\}$

$$D^0(i', k_l) = A(k_l, j') = 1$$

implies

$$(i', j') = (i, j).$$

Thus none of the columns $k_1, \ldots, k_n$ are involved for pairs other than (i, j), and for $(i', j') \neq (i, j)$ the matrix $R_{i'j'}$ leaves the columns $k_1, \ldots, k_n$ fixed. Moreover, since A is an amalgamation matrix, for each pair (i, k) with $D^0(i, k) = 1$ there exists j such that $A(k, j) = 1$. It is now clear that if we let R be the product of all the permutation matrices R_{ij} as (i, j) runs through all possibilities with $U_0(i, j) > 0$, then

$$D'R = D, \quad R^{-1}A = A \quad \text{and} \quad D^0R = D^0. \qquad \square$$

Note that the product of two matrices of division shape again has division shape. Furthermore, if $D = D_1D_2$, where D_1, D_2 have division shape, then $D^t = D_1^t D_2^t$ for all real t. We may now prove:

22. THEOREM. *For Markov chains* (X, S, m), (Y, T, p) *defined respectively by the stochastic matrices* M, P, (a) *and* (b) *are equivalent conditions.*

Proof. In one direction the result is an immediate consequence of (18) and (19), and has already been given as proposition (20). For the converse, let $U = U_t$ be a

non-trivial non-negative exponential matrix such that

$$U_t M^t = P^t U_t$$

for all $t \in \mathbb{R}$. Taking $t = 1$ and using the fact that M is stochastic,

$$U_1 \mathbf{1} = P(U_1 \mathbf{1}),$$

where $\mathbf{1}$ is a column vector consisting of ones. Since P is stochastic, the Perron–Frobenius theorem shows that the entries of the vector $U_1 \mathbf{1}$ must all be the same number, c say. $c \neq 0$ as U_1 is non-trivial. Hence the rows of U_1 all have the same sum $c \neq 0$. Dividing U_t by c^t, we may assume that U_1 is stochastic. We shall also insist that $U = U_t$ is minimal in the following sense. If $U' = U'_t$ is a non-trivial non-negative exponential matrix such that

$$U'_t M^t = P^t U'_t$$

(for all real t) and $(U_t - U'_t)$ is also a non-negative exponential matrix, then

$$U' \equiv U.$$

Use (21) to decompose

$$U_t = D^t A, \quad M^t = D^t_1 A_1, \quad P^t = D^t_2 A_2$$

and to rewrite

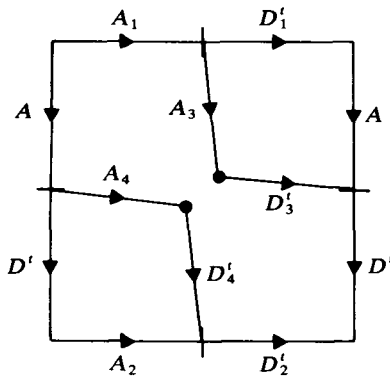$$A D^t_1 = D^t_3 A_3, \quad A_2 D^t = D^t_4 A_4,$$

FIGURE 8

where A, A_1, A_2, A_3, A_4 are amalgamation matrices and D, D_1, D_2, D_3, D_4 have division shape. We have the commutative diagram of figure 8. Now

$$D' D^t_3 A_3 A_1 = (D D_3)^t A_3 A_1$$

and

$$D^t_2 D^t_4 A_4 A = (D_2 D_4)^t A_4 A$$

are two decompositions of $U_t M^t = P^t U_t$. By (21), replacing D_4, A_4 by $D_4 R, R^{-1} A_4$, where R is a suitable permutation matrix, we may assume that

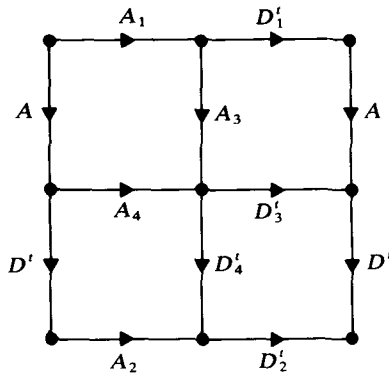$$D D_3 = D_2 D_4 \quad \text{and} \quad A_3 A_1 = A_4 A.$$

FIGURE 9

In other words, we may assume that figure 9 is commutative. Define $Q = D_3 A_4$. Since $AM^0 = (D_3^0 A_4)A$ and A is an amalgamation, $D_3^0 A_4$ is a 0–1 matrix. It follows that

$$Q^t = D_3^t A_4, \quad AM^t = Q^t A \quad \text{and} \quad P'D^t = D'Q^t$$

for all $t \in \mathbb{R}$. Q is a stochastic matrix as M, D_1 and D_3 are stochastic. D is stochastic since U_1 is. The minimality condition on $U = U_t$ ensures that Q is irreducible: Suppose

$$Q^t = \begin{pmatrix} Q_a^t & 0 \\ B^t & Q_b^t \end{pmatrix},$$

where Q_a is irreducible and accordingly write

$$A = \begin{pmatrix} A_a \\ A_b \end{pmatrix}, \quad D^t = (D_a^t \quad D_b^t).$$

Considering the case $t = 0$ and repeating the argument in the proof of (14), we see that B^0 is zero, i.e. B is a zero matrix. Hence

$$A_a M^t = Q_a^t A_a \quad \text{and} \quad P'D_a^t = D_a^t Q_a^t$$

for all real t. Taking

$$U^t = U_t' = D_a^t A_a$$

we obtain a contradiction to the minimality of $U = U_t$, unless $Q = Q_a$. Hence Q is also irreducible and we may apply (16) and (17) to conclude from $AM' = Q'A$, $P'D' = D'Q'$ that the Markov chain defined by Q is a common finite extension of (X, S, m) and (Y, T, p) by 1-block maps which are, respectively, right and left resolving.   □

It can be shown that the technique used in the proof of (22) works only for (matrices very closely related to) non-negative exponential matrices.

   Let $V = V_t$ be a non-trivial matrix whose entries are real-valued functions defined on a set $I \subset \mathbb{R}$, $0 \in I$, $I \neq \{0\}$. Let V_0 be non-negative integral. Suppose there exists a decomposition $V_t = D_t A$, where A is an amalgamation matrix, D_0 is a division
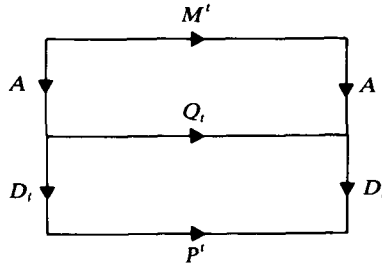
FIGURE 10

matrix and $D_t^0 = D_0$ for $t \in I$. Suppose also that there exists a matrix $Q = Q_t$ such that Q_0 is non-negative, irreducible and integral, $Q_t^0 = Q_0$ for $t \in I$ and figure 10 commutes. Then figure 10 may be used to show that (X, S, m) and (Y, T, p) satisfy (a). However, it can also be proved (by arguments similar to those used at the beginning of this section) that there is then a function $f: I \to \mathbb{R}$ and a non-negative exponential matrix $U = U_t$ such that

$$V_t = f(t)U_t \quad \text{for } t \in I$$

and

$$U_t M' = P' U_t \quad \text{for all } t \in \mathbb{R}.$$

The decomposition of (21) could have been given in the form $D\Delta'A$, where D is a division matrix, A is an amalgamation and Δ is a (real-valued) non-negative diagonal matrix. Then, in the proof of (22) we would have obtained a commutative diagram (figure 11), where the D's are division matrices, A's are amalgamations and the Δ's are non-negative diagonal matrices.
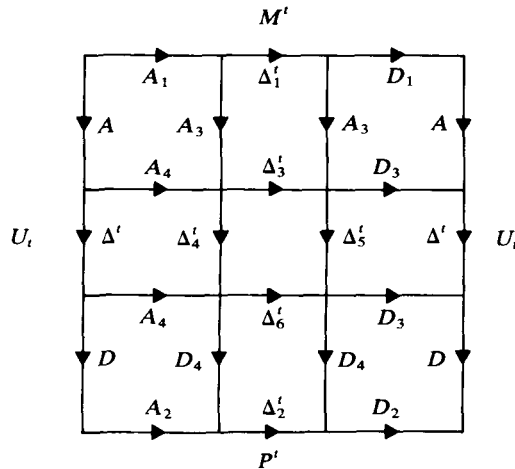

FIGURE 11

5. A necessary and sufficient condition for $\beta_M = \beta_P$

In § 2 we defined the function β_M for an irreducible stochastic matrix M. The same definition may be used if M is just an irreducible non-negative matrix (whether or

not it is stochastic): $\beta_M(t)$ is defined to be the maximum eigenvalue of M^t given by the Perron–Frobenius theorem. Again, β_M is an analytic function. In this section we fix two irreducible non-negative matrices M, P and show that a necessary and sufficient condition for $\beta_M = \beta_P$ is:

(c) There exists a matrix $U = U_t$, whose entries are integral combinations of exponential functions, such that U_t is strictly positive in a neighbourhood of 0 and $U_t M^t = P^t U_t$ for all $t \in \mathbb{R}$.

If $U = U_t$ satisfies $U_t M^t = P^t U_t$ and is strictly positive in a neighbourhood of 0 then $\beta_M(t) = \beta_P(t)$ in this neighbourhood of 0, by a standard argument using the Perron–Frobenius theorem. Since β_M, β_P are analytic, it follows that $\beta_M \equiv \beta_P$. Thus, (c) readily implies $\beta_M = \beta_P$. The rest of this section is devoted to the proof of the converse.

The ring $\mathbb{Z}[x_1, \ldots, x_k]$ of polynomials over $\mathbb{Z}$ in k indeterminates $x_1, \ldots, x_k$ is a unique factorization domain. (See [16].) In the following, a monomial shall mean a polynomial of the form

$$x_1^{n_1} x_2^{n_2} \cdots x_k^{n_k},$$

where $n_1 \geq 0, \ldots, n_k \geq 0$, $n_1, \ldots, n_k \in \mathbb{Z}$. The monomials N form a multiplicative system (see [16]) and the ring $\mathbb{Z}[x_1, \ldots, x_k]_N$ of elements of $\mathbb{Z}[x_1, \ldots, x_k]$ divided by elements of N is also a unique factorization domain. The units of $R = \mathbb{Z}[x_1, \ldots, x_k]_N$ are ratios of monomials.

23. LEMMA. *If $g_1, \ldots, g_k$ are integrally independent real numbers, then the ring of integral combinations of functions $\exp(ht)$ with h an integral combination of $g_1, \ldots, g_k$ is isomorphic to $\mathbb{Z}[x_1, \ldots, x_k]_N$.*

Proof. Define

$$x_i = \exp(g_i t) \quad (i = 1, \ldots, k)$$

so that if

$$h = n_1 g_1 + \cdots + n_k g_k,$$

$n_1, \ldots, n_k \in \mathbb{Z}$ we have

$$\exp(ht) = x_1^{n_1} x_2^{n_2} \cdots x_k^{n_k} = \mu/\nu.$$

for some $\mu, \nu \in N$. It is then clear that any integral combination of functions $\exp(ht)$ is a polynomial in $\mathbb{Z}[x_1, \ldots, x_k]$ divided by an element of N, i.e. takes the form of an element of $\mathbb{Z}[x_1, \ldots, x_k]_N$. To complete the proof it suffices to show that, if $m_1, \ldots, m_l \in \mathbb{Z}$ and distinct $h_1, \ldots, h_l$ satisfy

$$m_1 \exp(h_1 t) + \cdots + m_l \exp(h_l t) \equiv 0,$$

then

$$m_1 = m_2 = \cdots = m_l = 0.$$

This is well known. In fact, successive differentiation and evaluation at $t = 0$ yields a Vandermonde matrix

$$\begin{pmatrix} 1 & \cdots & 1 \\ h_1 & \cdots & h_l \\ \vdots & & \vdots \\ h_1^{l-1} & \cdots & h_l^{l-1} \end{pmatrix}$$

which has non-zero determinant.   □

24. COROLLARY. *If* $p \in \mathbb{Z}[x_1, \ldots, x_k]_N$ *satisfies*

$$p(\exp(g_1 t), \ldots, \exp(g_k t)) \equiv 0$$

then $p = 0$.

In the following we take $\beta(t)$ to be a real analytic function.

25. LEMMA. *If* χ_M, $\chi_P \in R[\lambda]$ *(where* $R = \mathbb{Z}[x_1, \ldots, x_k]_N$*) are monic polynomials such that*

$$\chi_M(\exp(g_1 t), \ldots, \exp(g_k t), \beta(t)) \equiv \chi_P(\exp(g_1 t), \ldots, \exp(g_k t), \beta(t)) \equiv 0,$$

then $\chi_M = u\chi$, $\chi_P = v\chi$, *where* χ, u, $v \in R[\lambda]$,

$$\chi(\exp(g_1 t), \ldots, \exp(g_k t), \beta(t)) \equiv 0$$

and χ *is monic.*

Proof. Using the Euclidean properties of the ring of polynomials over the quotient field of R we can assert

$$a\chi_M = u'\chi, \quad b\chi_P = v'\chi, \tag{†}$$

where $a, b \in R$, χ, u', $v' \in R[\lambda]$, χ is primitive and

$$c\chi = u''\chi_M + v''\chi_P$$

for some non-zero $c \in R$ and u'', $v'' \in R[\lambda]$.
   Substituting

$$x_1 = \exp(g_1 t), \ldots, x_k = \exp(g_k t), \quad \lambda = \beta(t)$$

we see that $c\chi$ vanishes for all $t \in \mathbb{R}$. If c vanishes for infinitely many t near zero then c vanishes for all t and, by (24), $c = 0$, which is not the case. Thus χ vanishes for infinitely many t near zero and therefore

$$\chi(\exp(g_1 t), \ldots, \exp(g_k t), \beta(t)) \equiv 0.$$

Since $R[\lambda]$ is a unique factorization domain and χ is primitive, we see from (†) that $a|u'$ and $b|v'$ in $R[\lambda]$. Thus, in fact,

$$\chi_M = u\chi, \quad \chi_P = v\chi$$

for some u, $v \in R[\lambda]$. Finally, as χ_M is monic, the leading coefficient of χ must be a unit and may be taken to be 1.   □

Now we return to the two irreducible non-negative matrices, M, P with the same β-function, $\beta(t)$. Let G be a finitely generated subgroup of $\mathbb{R}$ containing all $\log M(i, j)$, $\log P(l, n)$ $(M(i, j) \neq 0, P(l, n) \neq 0)$ and let $g_1, \ldots, g_k$ be a free basis for G. Put

$$x_1 = \exp(g_1 t), \ldots, x_k = \exp(g_k t).$$

Then $M(i, j)^t$, $P(l, n)^t$ are ratios of elements of N, the set of monomials in $x_1, \ldots, x_k$. In particular, $M(i, j)^t$, $P(l, n)^t$ belong to $R = \mathbb{Z}[x_1, \ldots, x_k]_N$.
   Let χ_M, χ_P be the characteristic polynomials of M, P:

$$\chi_M(\lambda) = \det(\lambda I - M^t), \quad \chi_P(\lambda) = \det(\lambda I - P^t).$$

Clearly χ_M, $\chi_P \in R[\lambda]$. Let $m(\lambda)$ be the first row of adj $(\lambda I - M^t)$ and let $\pi(\lambda)$ be the first column of adj $(\lambda I - P^t)$. The entries of $m(\lambda)$, $\pi(\lambda)$ and $\pi(\lambda) \cdot m(\lambda)$ belong to $R[\lambda]$. Moreover, $\pi(\beta(t))$ and $m(\beta(t))$ are strictly positive (see [13]). Since

$$\text{adj } (\lambda I - M^t)M^t = \lambda \text{ adj } (\lambda I - M^t) - \chi_M(\lambda)I,$$
$$P^t \text{ adj } (\lambda I - P^t) = \lambda \text{ adj } (\lambda I - P^t) - \chi_P(\lambda)I,$$

we have for any matrix L over $R[\lambda]$ (of appropriate dimension)

$$\text{adj } (\lambda I - P^t)L \text{ adj } (\lambda I - M^t)M^t - P^t \text{ adj } (\lambda I - P^t)L \text{ adj } (\lambda I - M^t) = \chi_P B - \chi_M A,$$

where A and B are suitable matrices over $R[\lambda]$. We take L to be the matrix with 1 in the top left-hand corner and zero elsewhere so that

$$\text{adj } (\lambda I - P^t)L \text{ adj } (\lambda I - M^t) = \pi(\lambda)m(\lambda),$$

and we have

$$\pi(\lambda)m(\lambda)M^t - P^t\pi(\lambda)m(\lambda) = \chi_P(\lambda)B(\lambda) - \chi_M(\lambda)A(\lambda). \qquad (*)$$

As $\beta(t)$ is the common β-function of M and P,

$$\chi_M(\exp(g_1 t), \ldots, \exp(g_k t), \beta(t)) \equiv \chi_P(\exp(g_1 t), \ldots, \exp(g_k t), \beta(t)) \equiv 0.$$

Let $\chi \in R[\lambda]$ be the monic polynomial of (25) which divides both χ_M and χ_P in $R[\lambda]$ and satisfies

$$\chi(\exp(g_1 t), \ldots, \exp(g_k t), \beta(t)) \equiv 0.$$

Let $d \geq 1$ be the degree of χ. Use the fact that χ is monic to reduce the polynomials in $\pi(\lambda)m(\lambda)$ modulo χ to obtain a matrix $V(\lambda)$ over $R[\lambda]$,

$$V(\lambda) = V_0 + \lambda V_1 + \cdots + \lambda^{d-1}V_{d-1},$$

where $V_0, \ldots, V_{d-1}$ are matrices over R. From $(*)$ we obtain an equation

$$V(\lambda)M^t - P^t V(\lambda) = \chi(\lambda)C(\lambda),$$

where the matrix C is over $R[\lambda]$. Since χ is of degree d, C must be the zero matrix and we have

$$V_i M^t = P^t V_i \quad \text{for } i = 0, 1, \ldots, d-1.$$

We now assume that M and P are aperiodic; the periodic case is similar. Put $t = 0$ to obtain integral matrices. Repeating the argument at the end of the proof of (9), we see that there exists i, $0 \leq i \leq d-1$, and a non-negative integer n such that $V_i(M^t)^n$ or $-V_i(M^t)^n$ reduces to a strictly positive integral matrix at $t = 0$. Accordingly, take

$$U = V_i(M^t)^n \quad \text{or} \quad U = -V_i(M^t)^n.$$

(Of course, $(M^t)^n$ denotes the n'th power under matrix multiplication of M^t.) Clearly $U = U_t$ satisfies

$$U_t M^t = P^t U_t$$

for all real t. U is over R and strictly positive at $t = 0$. It follows that the entries of $U = U_t$ are integral combinations of exponential functions and U_t is strictly positive in a neighbourhood of zero. We have completed the proof of:

26. THEOREM. *For two irreducible non-negative matrices M and P, $\beta_M = \beta_P$ iff (c) holds.*

27. *Remark.* Above, {matrices U over $R: UM^t = P^tU$} is a finitely generated R-module. This follows from the fact that $\mathbb{Z}[x_1, \ldots, x_k]$ ideals are finitely generated (see [16]). In the proof of (26) we obtained an element

$$U = \pm V_i(M^t)^n$$

of this module which is positive at $t = 0$. It is not difficult to see that this U is an R-combination of $V_0, V_1, \ldots, V_{d-1}$. Indeed, we only have to check that for each $j, 0 \le j \le d-1$, $V_j M^t$ is an R-combination of $V_0, V_1, \ldots, V_{d-1}$. In fact,

$$\pi(\lambda)m(\lambda)M^t = \lambda\pi(\lambda)m(\lambda) \mod \chi,$$

so that

$$V(\lambda)M^t = \lambda V(\lambda) \mod \chi.$$

Letting

$$\chi(\lambda) = \lambda^d - (c_0 + c_1\lambda + \cdots + c_{d-1}\lambda^{d-1})$$

with $c_0, \ldots, c_{d-1} \in R$, we have

$$(V_0 + \lambda V_1 + \cdots + \lambda^{d-1}V_{d-1})M^t$$
$$= \lambda(V_0 + \lambda V_1 + \cdots + \lambda^{d-1}V_{d-1}) \mod \chi$$
$$= \lambda V_0 + \cdots + \lambda^{d-1}V_{d-2} + (c_0 + \cdots + c_{d-1}\lambda^{d-1})V_{d-1} \mod \chi$$
$$= c_0 V_{d-1} + \lambda(V_0 + c_1 V_{d-1}) + \cdots + \lambda^{d-1}(V_{d-2} + c_{d-1}V_{d-1}) \mod \chi.$$

Hence

$$V_0 M^t = c_0 V_{d-1}$$

and, for $1 \le j \le d-1$,

$$V_j M^t = V_{j-1} + c_j V_{d-1},$$

as we wished to prove.

6. *Does $\beta_M = \beta_P$ imply finite equivalence?*

Let (X, S, m), (Y, T, p) be Markov chains defined by the stochastic matrices M, P. Suppose $\beta_M = \beta_P$. Then, by (26),

$$U_t M^t = P^t U_t$$

for a matrix $U = U_t$ (which is positive in a neighbourhood of zero and) whose entries are integral combinations of exponential functions. From (22) we see that the Markov chains are finitely equivalent if the entries of U are *non-negative* integral combinations of exponentials. We are not able to prove this in general, but we conjecture that this is the case. In particular:

CONJECTURE. *If two Markov chains have the same β-function, then they are finitely equivalent.*

Recall that in the topological case topological entropy is a complete invariant of finite equivalence. In [1] Adler & Marcus obtained the stronger result that for

topological Markov chains topological entropy and period together characterize almost topological equivalence, i.e. finite equivalence with a.e. 1–1 maps. In view of this fact, one may wonder about the possibility of establishing for Markov chains finite equivalences with a.e. 1–1 maps when β-functions and periods coincide. We assert that this is not possible in general. The reader is referred to the group invariant introduced in [6]. This invariant would exclude, for example, the possibility of a finite equivalence with a.e. 1–1 maps between the Markov shifts defined by the matrices

$$\begin{pmatrix} p & q \\ p & q \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} p & q \\ q & p \end{pmatrix},$$

$0 < p < 1$, $p \neq \frac{1}{2}$, $q = 1 - p$. (These are aperiodic and have identical β-functions, $\beta(t) = p^t + q^t$.)

We devote the rest of the paper to examples, which will serve to illustrate our theory and the above conjecture.

Measures of maximal entropy. For a Markov measure of maximal entropy

$$\beta(t) = \gamma^{1-t},$$

where $\log \gamma$ is the topological entropy. Thus, for Markov measures of maximal entropy, the β-invariant reduces to topological entropy. The completeness of the invariant for this sub-class may easily be deduced from the purely topological classification (14).

We remark that a Markov shift and its inverse have identical β-functions. Other classes of Markov shifts with identical β-functions are:

Bernoulli-type shifts. Fix a probability vector

$$p = (p(1), \ldots, p(n)).$$

Consider all irreducible matrices M with the following property. In each row of M there are precisely n non-zero entries and the non-zero entries form a permutation of $\{p(1), \ldots, p(n)\}$. We call the Markov shifts defined by such matrices, and their inverses, *Bernoulli-type shifts* based on p. These were first studied by Rosenblatt [11]. In particular, the Bernoulli shift defined by p is in this class. All Bernoulli-type shifts based on $p = (p(1), \ldots, p(n))$ have the same β-function,

$$\beta(t) = p(1)^t + \cdots + p(n)^t.$$

We claim that all Bernoulli-type shifts based on the same probability vector $p = (p(1), \ldots, p(n))$ are finitely equivalent. Indeed, all elements of this class have the Bernoulli shift defined by p, (Y, T, p), as a finite factor.

Let M be a matrix as above, defining a Bernoulli-type shift (X, S, m) based on p. We give a 2-block finite factor map $\phi : X \to Y$. That (Y, T, p) is also a finite factor of the inverse of (X, S, m) follows from this on noting that Bernoulli shifts are their own inverses. In a row i of M there are exactly n non-zero entries and these may be ordered $k_1, \ldots, k_n$ so that

$$M(i, k_j) = p(j) \quad \text{for } j = 1, \ldots, n.$$

Define $\phi_0(i, k_j) = j$. Evidently, the map ϕ defined by ϕ_0 is right resolving. Given $(j_1, j_2) \in A(T)^2$ and $(i_1, i_2) \in A(S)^2$ with $M(i_1, i_2) > 0$ and $\phi_0(i_1, i_2) = j_1$, there exists a unique $i_3 \in A(S)$ such that

$$M(i_2, i_3) > 0 \quad \text{and} \quad \phi_0(i_2, i_3) = j_2.$$

It follows that ϕ is bounded-to-one and surjective (see the proof of (12)). ϕ is also measure preserving: it is sufficient to check this on cylinders of (Y, T, p). For a cylinder $[j_1 \cdots j_l]$, $\phi^{-1}[j_1 \cdots j_l]$ consists of card $(A(S))$ cylinders of length $l + 1$. For each $i_1 \in A(S)$ the right resolving property determines $i_2, \dots, i_{l+1} \in A(S)$ such that

$$m[i_1 \dots i_{l+1}] > 0$$

and

$$\phi_0(i_k, i_{k+1}) = j_k$$

for $k = 1, \dots, l$. Moreover,

$$M(i_k, i_{k+1}) = p(j_k)$$

so that

$$m[i_1 \cdots i_{l+1}] = m(i_1)p(j_1) \cdots p(j_l).$$

Hence

$$m(\phi^{-1}[j_1 \cdots j_l]) = \left(\sum_{i_1 \in A(S)} m(i_1) \right) p(j_1) \cdots p(j_l) = p(j_1) \cdots p(j_l),$$

as we wished to prove.

We have now exhausted the types of examples (we know) for which the β-function is readily computable. (1) and (15) may be used to obtain examples of Markov chains with identical β-functions from (topological) 1-block finite factor maps, but with these examples we have finite equivalence *ab initio*. However, a Markov chain and its inverse have the same β-function but we are not able to prove in general that they are finitely equivalent. Thus inverses (or, in terms of matrices, transposes) present good tests of our conjecture. We now turn to low dimensional examples $(2 \times 2$ and $3 \times 3)$ and use the techniques of § 5 to test if they satisfy condition (b) of § 4.

Consider a general 2×2 matrix

$$\begin{pmatrix} x & y \\ z & w \end{pmatrix}.$$

We think of the entries as exponential functions. Observe that

$$\begin{pmatrix} x & y \\ z & w \end{pmatrix} \begin{pmatrix} y & 0 \\ 0 & z \end{pmatrix} = \begin{pmatrix} y & 0 \\ 0 & z \end{pmatrix} \begin{pmatrix} x & z \\ y & w \end{pmatrix}.$$

Hence (b) is satisfied by 2×2 matrices and their transposes. (We should point out that 2×2 Markov chains are trivially finitely equivalent to their inverses.)

We now assume the entries of

$$P = \begin{pmatrix} x & y \\ z & w \end{pmatrix}$$

are algebraically independent and look for 3×3 matrices whose β-functions coincide with that of P. Remembering that we are concerned only with irreducible matrices and noting that evaluation at 1 must give a 0–1 matrix with maximum eigenvalue 2, one quickly concludes that either row sums must equal 2 or column sums must equal 2. Without loss of generality, we consider the former case (this covers the latter by transposing). We require the characteristic polynomial of P, namely

$$\lambda^2 - (w + x)\lambda + wx - yz,$$

to divide the characteristic polynomial of the 3×3 matrix whose entries are (ratios of) monomials. For certain forms of matrices (namely when the determinant is a positive form, i.e. a sum of monomials) this is impossible. We arrive at

$$\begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$$

as the only possible 'shape' (up to permutation of the indexing set $\{1, 2, 3\}$).

A matrix

$$\begin{pmatrix} 0 & a & b \\ c & d & 0 \\ 0 & e & f \end{pmatrix}$$

has characteristic polynomial

$$\lambda[\lambda^2 - (d + f)\lambda + (df - ac)] + c(af - be).$$

For divisibility by

$$\lambda^2 - (w + x)\lambda + wx - yz,$$

we require $af = be$ $(c \neq 0$ as none of a, b, c, d, e, f can be zero). Therefore

$$d + f = w + x \quad \text{and} \quad df - ac = wx - yz.$$

It follows that $d = w, f = x$ (or $d = x, f = w$) and $ac = yz$ so that the matrix must be

$$M = \begin{pmatrix} 0 & \dfrac{yz}{v} & \dfrac{xyz}{uv} \\ v & w & 0 \\ 0 & u & x \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} 0 & \dfrac{yz}{v} & \dfrac{wyz}{uv} \\ v & x & 0 \\ 0 & u & w \end{pmatrix}.$$

We concentrate on the former; the latter may be treated similarly. Adopting the notation of the proof of (26), λ is defined by

$$\chi(\lambda) = \lambda^2 - (w + x)\lambda + wx - yz = 0.$$

The first row of adj $(\lambda I - M)$ is

$$m(\lambda) = \left((w - \lambda)(x - \lambda), \frac{yz}{v}\lambda, -(w - \lambda)\frac{xyz}{uv} \right)$$

$$= \left(yz, \frac{yz}{v}\lambda, (\lambda - w)\frac{xyz}{uv} \right),$$

and the first column of adj $(\lambda I - P)$ is

$$\pi(\lambda) = \begin{pmatrix} w - \lambda \\ -z \end{pmatrix}.$$

Hence (modulo χ) we have

$$\pi(\lambda)m(\lambda) = V_0 + \lambda V_1,$$

where

$$V_0 = \begin{pmatrix} wyz & \dfrac{yz}{v}[wx - yz] & -\dfrac{xyz}{uv}[w^2 + yz - wx] \\ -yz^2 & 0 & \dfrac{wxyz^2}{uv} \end{pmatrix},$$

$$V_1 = \begin{pmatrix} -yz & -\dfrac{xyz}{v} & -\dfrac{xyz}{uv}[x - w] \\ 0 & -\dfrac{yz^2}{v} & -\dfrac{xyz^2}{uv} \end{pmatrix}.$$

Now

$$V_0 + wV_1 = \begin{pmatrix} 0 & -\dfrac{y^2z^2}{v} & -\dfrac{xy^2z^2}{uv} \\ -yz^2 & -\dfrac{wyz^2}{v} & 0 \end{pmatrix}.$$

Multiplying this by $-(uv)/(yz^2)$, we obtain

$$U = \begin{pmatrix} 0 & uy & xy \\ uv & uw & 0 \end{pmatrix},$$

which satisfies $UM = PU$.

Now we consider a 3×3 matrix $P = (P_{ij})$ and its transpose P^{tr}. Observing that

$$UP = P^{\mathrm{tr}}U \quad \text{implies} \quad (U + U^{\mathrm{tr}})P = P^{\mathrm{tr}}(U + U^{\mathrm{tr}})$$

for the symmetric matrix $U + U^{\mathrm{tr}}$, we seek a symmetric solution U of $UP = P^{\mathrm{tr}}U$ such that the entries of U are non-negative integral combinations of monomials. Let

$$U = \begin{pmatrix} x & a & b \\ a & y & c \\ b & c & z \end{pmatrix}.$$

By symmetry, we need only check $UP = P^{\mathrm{tr}}U$ in the three top right-hand positions $(1, 2)$, $(1, 3)$ and $(2, 3)$. Working out UP and $P^{\mathrm{tr}}U$ in these three positions we see that

$$\mu = \begin{pmatrix} x \\ y \\ z \\ a \\ b \\ c \end{pmatrix}$$

must satisfy the equation

$$(S \quad \xi \quad \eta \quad \zeta)\mu = 0, \tag{*}$$

where the block matrix

$$(S \quad \xi \quad \eta \quad \zeta) = \begin{pmatrix} P_{12} & -P_{21} & 0 & \vdots & (P_{22}-P_{11}) & \vdots & P_{32} & \vdots & -P_{31} \\ P_{13} & 0 & -P_{31} & \vdots & P_{23} & \vdots & (P_{33}-P_{11}) & \vdots & -P_{21} \\ 0 & P_{23} & -P_{32} & \vdots & P_{13} & \vdots & -P_{12} & \vdots & (P_{33}-P_{22}) \end{pmatrix}.$$

If

$$\det S = P_{12}P_{23}P_{31} - P_{21}P_{32}P_{13} = 0,$$

then we may take

$$a = b = c = 0, \quad x = 1, \quad y = P_{12}/P_{21} \quad \text{and} \quad z = P_{13}/P_{31}$$

to obtain a 'non-negative' solution. If $\det S \neq 0$, defining

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = (\text{adj } S)\xi$$

we have for the vector

$$u = \begin{pmatrix} (\text{adj } S)\xi \\ -\det S \\ 0 \\ 0 \end{pmatrix},$$

$$(S \quad \xi \quad \eta \quad \zeta)u = S(\text{adj } S)\xi - (\det S)\xi = 0.$$

Similarly,

$$v = \begin{pmatrix} (\text{adj } S)\eta \\ 0 \\ -\det S \\ 0 \end{pmatrix} \quad \text{and} \quad w = \begin{pmatrix} (\text{adj } S)\zeta \\ 0 \\ 0 \\ -\det S \end{pmatrix}$$

satisfy (∗). Moreover u, v, w form a basis for solutions of (∗). On computing adj S we obtain:

$$(\text{adj } S)\xi = \begin{pmatrix} P_{23}P_{31}(P_{22}-P_{11}) - P_{21}P_{32}P_{23} + P_{21}P_{31}P_{13} \\ P_{13}P_{32}(P_{22}-P_{11}) - P_{12}P_{32}P_{23} + P_{12}P_{31}P_{13} \\ P_{13}P_{23}(P_{22}-P_{11}) - P_{12}P_{23}P_{23} + P_{21}P_{13}P_{13} \end{pmatrix},$$

$$(\text{adj } S)\eta = \begin{pmatrix} P_{23}P_{31}P_{32} - P_{21}P_{32}(P_{33}-P_{11}) - P_{21}P_{31}P_{12} \\ P_{13}P_{32}P_{32} - P_{12}P_{32}(P_{33}-P_{11}) - P_{12}P_{31}P_{12} \\ P_{13}P_{23}P_{32} - P_{12}P_{23}(P_{33}-P_{11}) - P_{21}P_{13}P_{12} \end{pmatrix},$$

$$(\text{adj } S)\zeta = \begin{pmatrix} -P_{23}P_{31}P_{31} + P_{21}P_{32}P_{21} + P_{21}P_{31}(P_{33}-P_{22}) \\ -P_{13}P_{32}P_{31} + P_{12}P_{32}P_{21} + P_{12}P_{31}(P_{33}-P_{22}) \\ -P_{13}P_{23}P_{31} + P_{12}P_{23}P_{21} + P_{21}P_{13}(P_{33}-P_{22}) \end{pmatrix}.$$

Recalling that rows, 4, 5, 6 of u, v, w are either 0 or

$$-\det S = P_{21}P_{32}P_{13} - P_{12}P_{23}P_{31} \neq 0,$$

one is led to seek non-negative integral combinations k, l, m of (ratios of) monomials such that

$$V = (\text{adj } S)(k\xi + l\eta + m\zeta)$$

is divisible by $\det S$ and the entries of $-(1/\det S)V$ are non-negative integral combinations. Without attempting to do this, we observe that a large number of special cases have particularly simple solutions. For instance, if $P_{23} = 0$ and if one of P_{11} or $P_{22} - P_{11}$ is zero, then taking $k = 1$, $l = m = 0$ suffices.

REFERENCES

[1] R. L. Adler & B. Marcus. Topological entropy and equivalence of dynamical systems. *Mem. A.M.S.* **219** (1979).
[2] R. Bhatia & K. K. Mukherjea. On the rate of change of spectra of operators. *Linear Algebra and its Appl.* **27** (1979), 147–157.
[3] E. M. Coven & M. E. Paul. Endomorphisms of irreducible subshifts of finite type. *Math. Syst. Theory* **8** (1974), 167–175.
[4] R. Fellgett & W. Parry. Endomorphisms of a Lebesgue space II. *Bull. L.M.S.* **7** (1975), 151–158.
[5] G. A. Hedlund. Endomorphisms and automorphisms of the shift dynamical system. *Math. Syst. Theory* **3** (1969), 320–375.
[6] W. Parry. Endomorphisms of a Lebesgue space III. *Israel J. Math.* **21** (1975), 167–172.
[7] W. Parry. A finitary classification of topological Markov chains and sofic systems. *Bull L.M.S.* **9** (1977), 86–92.
[8] W. Parry. University of Warwick lecture notes.
[9] W. Parry & K. Schmidt. A note on cocycles of unitary representations. *Proc. A.M.S.* **55** (1976), 185–190.
[10] W. Parry & R. F. Williams. Block coding and a zeta function for finite Markov chains. *Proc. L.M.S.* **35** (1977), 483–495.
[11] M. Rosenblatt. Independence and dependence. *Proc. 4th Berkeley Symp. Math. Stats. and Prob.* Vol. II, pp. 431–443.
[12] D. Ruelle. *Thermodynamic Formalism.* Addison-Wesley: Reading, Mass., 1978.
[13] E. Seneta. *Non-negative Matrices.* Allen and Unwin: London, 1973.
[14] S. Tuncel. Conditional pressure and coding. *Israel J. Math.* **39** (1981), 101–112.
[15] R. F. Williams. Classification of subshifts of finite type. *Ann. Math.* **98** (1973), 120–153; Errata. *Ann. Math.* **99** (1974), 380–381.
[16] O. Zariski & P. Samuel. *Commutative Algebra*, Vol. 1. Springer-Verlag: Berlin, 1958.