ISRAEL JOURNAL OF MATHEMATICS, Vol. 39, Nos. 1-2, 1981

CONDITIONAL PRESSURE AND CODING

BY
SELIM TUNCEL

ABSTRACT

We use pressure to obtain invariants for bounded-to-one block homomorphisms between Markov shifts. These invariants enable us to show that if there is a bounded-to-one block homomorphism between Bernoulli shifts given by probability vectors p and q then q may be obtained from p by a permutation. The invariants may be viewed as conditional pressures; a convergence theorem for eigenmeasures of Ruelle operators motivates the definition of conditional pressure and helps establish our invariants for regular isomorphism of Markov shifts. It follows that Bernoulli shifts given by probability vectors p and q are regularly isomorphic iff q is a permutation of p. We employ our invariants also in the context of a finite equivalence. Finally we indicate that ratio variational principles yield further invariants.

§0. Introduction

Since the powerful results of Ornstein [7] and Friedman and Ornstein [3] on isomorphisms of Bernoulli and of Markov shifts, various new types of isomorphism have been introduced and investigated. These isomorphisms, in addition to preserving measure, respect some extra structure and are of greater interest from the point of view of coding theory. One important example is the concept of finitary isomorphism for which Keane and Smorodinsky have recently obtained complete results ([4], [5]). In this note we consider block codes and regular isomorphisms.

We start by listing definitions, notation and some properties of topological Markov chains, pressure and Ruelle operators on one-sided topological Markov chains. Then we use pressure to obtain invariants of block isomorphism and show that two Bernoulli shifts are block isomorphic if and only if there is a trivial block isomorphism between them. In fact, block homomorphisms exist between two Bernoulli shifts of the same entropy only when one of the spaces can be obtained from the other by a permutation of states.

Received August 5, 1980 and in revised form September 22, 1980

In the third section we prove, for Ruelle measures (eigenmeasures of Ruelle operators), a convergence theorem which suggests a way of defining a conditional pressure. The type of limit considered in this theorem is used in the fourth section to establish our invariants (for block isomorphism) as regular isomorphism invariants, showing that two Bernoulli shifts are regularly isomorphic if and only if there is a trivial (regular) isomorphism between them. In the final section we use our invariants in the context of a measure theoretic-topological finite equivalence derived from the purely topological equivalence relations of [1] and [8]. We also prove a ratio variational principle and indicate how it may yield invariants.

I am grateful to my research supervisor Professor William Parry for his invaluable guidance and encouragement.

§1. Topological Markov chains, pressure and Ruelle operators

Let A be an $n \times n$ irreducible 0-1 matrix. Give $\{1, 2, \cdots, n\}$ the discrete topology and $\Sigma = \Pi_{-\infty}^{\infty}\{1, \cdots, n\}$ the product topology. Consider the subspace X of Σ defined

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

The shift T is defined by $(Tx)_i = x_{i+1}$ for $x = (x_i)$. T is a homeomorphism of the compact, metrizable space X. On X we use the metric $d(x, y) = 1/(k+1)$ when k is the greatest integer with $x_i = y_i$ $\forall -k < i < k$. (X, T) is the *topological Markov chain* (*or subshift of finite type*) given by A. If P is an $n \times n$ stochastic matrix compatible with A (i.e. $P(i, j) = 0$ iff $A(i, j) = 0$), the (shift invariant) Markov measure defined by P has as its support the topological Markov chain given by A. We will always regard Markov measures as being defined on their supporting topological Markov chains. The sets

$$[i_0 \cdots i_l]^m = \{(x_i) \in X : x_m = i_0, x_{m+1} = i_1, \cdots, x_{m+l} = i_l\} \quad (l, m \in Z, l \geqq 0)$$

are cylinders. They form a base for the topology of X. Write $[i_0 \cdots i_l]$ for $[i_0 \cdots i_l]^0$. The *state partition* consists of $[i]$, $1 \leqq i \leqq n$.

Given A as above, we define $\Sigma^+ = \Pi_0^{\infty}\{1, \cdots, n\}$ and

$$X^+ = \{(x_i) \in \Sigma^+ : A(x_i, x_{i+1}) = 1 \ \forall i \geqq 0\}.$$

The shift $(Tx)_i = x_{i+1}$, $i \geqq 0$, is now a continuous surjection and (X^+, T) is called the *one-sided subshift of finite type* (*topological Markov chain*) defined by A. We put on the compact space X^+ the metric $d(x, y) = 1/(k+1)$ when k is the greatest integer with $x_i = y_i$ $\forall 0 \leqq i < k$. Cylinders and the state partition are similar to the two-sided case. We denote by $M(X^+)$ ($M(X)$ in the two-sided case) the set of Borel probability measures on the space.

For $f \in C(X)$, (X, T) subshift of finite type, write $S_n f$ for $\Sigma_{i=0}^{n-1} f \circ T^i$. Put $P_n f = \Sigma_{x_0 \cdots x_{n-1}} (\sup_{y \in [x_0 \cdots x_{n-1}]} \exp S_n f(y))$ where the sum is over all allowed words $x_0 \cdots x_{n-1}$ (i.e. all words $x_0 \cdots x_{n-1}$ with $[x_0 \cdots x_{n-1}] \neq \emptyset$). $P(f) = \lim_{n \to \infty} (1/n) \log P_n(f)$ exists and is the *pressure of f*. $P : C(X) \to R$ is a continuous map and

$$(*) \qquad P(f) = \sup\{h(m) + \textstyle\int f dm : m \in M(X) \text{ is } T\text{-invariant}\},$$

where $h(m)$ is the entropy of T with respect to m.

The above definitions and statements hold when X is replaced by X^+. In fact Walters' paper, [16], contains definitions of pressure for $f \in C(Y)$ where Y is a compact metric space with continuous $T : Y \to Y$. See [16] also for the properties of pressure and a proof of (*). We will use the following:

$$P(f + g \circ T - g) = P(f) \qquad \text{for } f, g \in C(X).$$

If $\pi : X_1 \to X_2$ is a bounded-to-one continuous surjective map between subshifts of finite type (X_1, T_1) and (X_2, T_2) and $f \in C(X_2)$, then $P(f \circ \pi) = P(f)$.

Let (X^+, T) be a one-sided subshift of finite type given by the matrix A. (X^+, T) is topologically mixing iff A is aperiodic (i.e. iff $A^M > 0$ for some $M > 0$). For $\phi \in C(X^+)$ the *Ruelle operator* $\mathscr{L}_\phi : C(X^+) \to C(X^+)$ is defined

$$(\mathscr{L}_\phi f)(x) = \sum_{y \in T^{-1}x} e^{\phi(y)} f(y).$$

$\mathscr{L}_\phi$ is positive, linear and continuous. Denote its spectral radius by $r(\mathscr{L}_\phi)$. Put $\operatorname{var}_n \phi = \sup\{|\phi(x) - \phi(y)| : x_0 = y_0, x_1 = y_1, \cdots, x_{n-1} = y_{n-1}\}$. In [15] Walters combines results by Ruelle and Keane to give a proof of:

RUELLE'S OPERATOR THEOREM (RUELLE'S PERRON–FROBENIUS THEOREM). *Let (X^+, T) be a topologically mixing one-sided subshift of finite type. Let $\phi \in C(X^+)$ satisfy $\Sigma_{n=1}^\infty \operatorname{var}_n \phi < \infty$. There exist $\lambda > 0$, $h \in C(X^+)$ and $\nu \in M(X^+)$ such that $h > 0$, $\nu(h) = 1$, $\mathscr{L}_\phi h = \lambda h$, $\mathscr{L}_\phi^* \nu = \lambda \nu$ and, for $f \in C(X^+)$, $\lambda^{-n} \mathscr{L}_\phi^n f \to \nu(f) h$ in $C(X^+)$. λ, h and ν are uniquely defined by these properties and $\lambda = e^{P(\phi)} = r(\mathscr{L}_\phi)$. Moreover $\mu \in M(X^+)$ defined $\mu(f) = \nu(hf)$, $f \in C(X^+)$, is the only T-invariant probability with $h(\mu) + \int \phi d\mu = P(\phi)$.*

For any $\phi \in C(X^+)$, there is $\lambda > 0$ and $\nu \in M(X^+)$ with $\mathscr{L}_\phi^* \nu = \lambda \nu$, by the

Schauder–Tychonov fixed point theorem. We shall see that there is precisely one such λ, $\lambda = e^{P(\phi)}$. Call $\nu \in M(X^+)$ satisfying $\mathcal{L}_\phi^* \nu = \lambda \nu$ a *Ruelle measure* for ϕ.

For references on the above, and an excellent account, see [15].

§2. Block codes and pressure

Let (X_i, T_i, m_i) be finite state processes with state partitions α_i $(i = 1, 2)$. A homomorphism $\phi : X_1 \to X_2$ (i.e. a measure-preserving map defined a.e. and satisfying $\phi T_1 = T_2 \phi$ a.e.) is a *block homomorphism* (*or a block code*) if there exists $p \in N$ such that each $\phi^{-1}A$, $A \in \alpha_2$, may be written as a union of sets in the (refined) partition $\bigvee_{i=-p}^{p} T^i \alpha_1$. An isomorphism (a.e. invertible homomorphism) $\phi : X_1 \to X_2$ is a *block isomorphism* if both ϕ and ϕ^{-1} are block codes.

[12] contains a discussion of the problem of block isomorphism of Markov chains.

The information cocycle of a finite state process (X, T, m) with state partition α is $I_T = I(\alpha \mid \alpha^-) = -\sum_{A \in \alpha} \chi_A \log m(A \mid \alpha^-)$ where $\alpha^- = \bigvee_1^\infty T^{-i}\alpha$ denotes the smallest σ-algebra containing the partitions $T^{-i}\alpha$, $i \geqq 1$.

We need the following two results:

1. Proposition [11]. *Let $\phi : X_1 \to X_2$ be a block homomorphism* (resp. *isomorphism*) *of Markov shifts (X_i, T_i, m_i), $i = 1, 2$. There exists a measure-preserving continuous surjection* (resp. *homeomorphism*) *$\phi' : X_1 \to X_2$ such that $\phi' T_1 = T_2 \phi'$ and $\phi' = \phi$* a.e.

2. Proposition [11]. *Let (x_i, T_i, m_i) be Markov shifts for $i = 1, 2$. If $\phi : X_1 \to X_2$ is a bounded-to-one measure-preserving continuous surjection satisfying $\phi T_1 = T_2 \phi$ then*

$$I_{T_1} = I_{T_2} \circ \phi + g \circ T_1 - g \qquad \textit{for some } g \in C(X_1).$$

1 reduces the study of block codes between Markov shifts to that of continuous measure-preserving surjections.

Let (X_i, T_i, m_i), $i = 1, 2$, be Markov shifts of the same entropy and let $\phi : X_1 \to X_2$ be a continuous, measure-preserving surjection. ϕ is then bounded to one (see [8]) and, applying 2, we have

$$tI_{T_1} = tI_{T_2} \circ \phi + tg \circ T_1 - tg$$

where all functions are continuous and $t \in R$. Applying pressure to both sides of this equation we obtain:

3. Proposition. *If $\phi : X_1 \to X_2$ is a block code between Markov shifts (X_i, T_i, m_i), $i = 1, 2$, of the same entropy then $P(tI_{T_1}) = P(tI_{T_2})$ for all $t \in R$.*

4. COROLLARY. *If the Markov shifts* (X_i, T_i, m_i), $i = 1, 2$, *are block isomorphic then* $P(tI_{T_1}) = P(tI_{T_2})$ *for all* $t \in R$.

For the calculation of $P(tI_T)$ we need:

5. THEOREM [6]. *Let* (X, T) *be a topological Markov chain and let* f *be a function of two coordinates*, $f(x) = f(x_0, x_1)$ *for* $x = (x_n) \in X$. *Then there is a unique* T*-invariant probability* m *such that* $P(f) = h(m) + \int f dm$. m *is Markov.*

In the notation of 5, m and $P(f)$ are obtained as follows:

Suppose (X, T) is given by the $n \times n$ irreducible 0-1 matrix A. Consider the $n \times n$ matrix $M = (e^{f(i,j)})$ where $M(i,j) = 0$ when $A(i,j) = 0$. Let $\beta > 0$ be the maximum eigenvalue of M with corresponding strictly positive right eigenvector $v = (v(1), \ldots, v(n))^{\mathrm{tr}}$ given by the Perron–Frobenius theorem (see [14]). Then $P(f) = \log \beta$ and m is given by the stochastic matrix $((v(j)/\beta v(i))e^{f(i,j)})$ compatible with A.

If (X, T, m) is a Bernoulli process given by the probability vector $p = (p_1, \ldots, p_n)$, $I_T = -\sum_{i=1}^{n} \chi_{[i]} \log p_i$ so that $P(tI_T)$ is obtained by considering the matrix M with identical columns $v = (p_1^{-t}, \ldots, p_n^{-t})^{\mathrm{tr}}$. Clearly, $Mv = \beta v$ where $\beta = \sum_{i=1}^{n} p_i^{-t}$. Using the characterization of the maximal eigenvalue given by the Perron–Frobenius theorem as the only eigenvalue with a strictly positive eigenvector, $P(tI_T) = \log \beta = \log(\sum_{i=1}^{n} p_i^{-t})$. This, combined with 3, shows that if there is a block code between two Bernoulli shifts given by probability vectors $p = (p_1, \ldots, p_n)$ and $q = (q_1, \ldots, q_l)$ of the same entropy, then $l = n$ and q may be obtained from p by a permutation. We summarize this as

6. THEOREM. *Between two Bernoulli shifts of the same entropy, block codes exist iff there is a trivial block isomorphism. In particular two Bernoulli shifts are block isomorphic iff there is a trivial block isomorphism between them.*

At the Durham symposium on Ergodic Theory held in June 1980, I learned that A. del Junco, M. Keane, B. Kitchens, B. Marcus and L. Swanson also have a proof of 6.

§3. Ruelle measures

Let (X^+, T) be a topologically mixing one-sided subshift of finite type, given by the 0-1 matrix A with $A^M > 0$. We prove a convergence theorem for Ruelle measures. We also show that for any $\phi \in C(X^+)$, $r(\mathcal{L}_\phi) = e^{P(\phi)}$ and that all Ruelle measures for ϕ correspond to this eigenvalue.

A finite subset $F \subset X^+$ is (n, ε) spanning if for $x \in X^+$ there exists $y \in F$ such that $d(T^ix, T^iy) < \varepsilon$ $\forall 0 \leqq i < n$. It is easy to see that T has the property that for $x \neq y$ there is $i \geqq 0$ with $d(T^ix, T^iy) > \frac{1}{2}$ (i.e. T is expansive with expansive constant $\frac{1}{2}$). This observation and a definition of pressure given in [16] show that if, for $\phi \in C(X^+)$, we put

$$Q_n(\phi) = \inf\left\{ \sum_{x \in F} e^{(S_n\phi)(x)} : F \text{ is } (n, \tfrac{1}{2}) \text{ spanning} \right\}$$

then $P(\phi) = \limsup_{n\to\infty} (1/n) \log Q_n(\phi)$.

7. Lemma. *Let $\phi \in C(X^+)$. For $n > M$,*

$$e^{-M\|\phi\|} Q_{n-M}(\phi) \leqq \mathscr{L}_\phi^n 1 \leqq P_n(\phi).$$

Proof. $(\mathscr{L}_\phi^n 1)(x) = \Sigma'_{y_0\cdots y_{n-1}} e^{(S_n\phi)(y_0\cdots y_{n-1}x)}$ where the sum Σ' is over all (allowed) words $y_0 \cdots y_{n-1}$ preceding x. Hence

$$(\mathscr{L}_\phi^n 1)(x) \leqq \sum_{y_0\cdots y_{n-1}}{}' \left(\sup_{z \in [y_0 \cdots y_{n-1}]} e^{(S_n\phi)(z)} \right) \leqq P_n(\phi).$$

For the other inequality note that, since $A^M > 0$, for any x, $T^{-(M+n)}x$ (in fact $T^{-(M+n-1)}x$) is an $(n, \frac{1}{2})$ spanning set. Now, for $n > 0$,

$$\begin{aligned}(\mathscr{L}_\phi^{(M+n)} 1)(x) &= \sum_{y \in T^{-(M+n)}x} \exp(\phi(y) + \cdots + \phi(T^{n-1}y) \\ &\qquad\qquad + \phi(T^n y) + \cdots + \phi(T^{n+M-1}y)) \\ &\geqq e^{-M\|\phi\|} \sum_{y \in T^{-(M+n)}x} e^{(S_n\phi)(y)} \\ &\geqq e^{-M\|\phi\|} Q_n(\phi).\end{aligned}$$

8. Theorem. *Let $\phi \in C(X^+)$. Then $r(\mathscr{L}_\phi) = e^{P(\phi)}$ and, for any probability μ, $(1/n)\log \mu(\mathscr{L}_\phi^n 1) \to P(\phi)$. All Ruelle measures for ϕ correspond to the eigenvalue $r(\mathscr{L}_\phi)$.*

Proof. By 7, $e^{-M\|\phi\|} Q_{n-M}(\phi) \leqq \|\mathscr{L}_\phi^n 1\| \leqq P_n(\phi)$ so

$$\begin{aligned}\log r(\mathscr{L}_\phi) &= \lim_{n\to\infty} \frac{1}{n} \log \|\mathscr{L}_\phi^n 1\| \\ &= \lim_{n\to\infty} \frac{1}{n} \log P_n(\phi) = \limsup_{n\to\infty} \frac{1}{n} \log Q_n(\phi) = P(\phi).\end{aligned}$$

Similarly, for a probability μ, 7 gives

$$e^{-M\|\phi\|} Q_{n-M}(\phi) \leqq \mu(\mathscr{L}_\phi^n 1) \leqq P_n(\phi)$$

and we get $P(\phi) = \lim (1/n) \log \mu(\mathscr{L}_\phi^n 1)$.

If ν is a Ruelle measure for ϕ, i.e. if $\mathscr{L}_\phi^* \nu = \lambda \nu$, $\lambda = (\lambda^n(1))^{1/n} = \nu(\mathscr{L}_\phi^n 1)^{1/n} \to r(\mathscr{L}_\phi)$. Thus $\lambda = r(\mathscr{L}_\phi)$.

9. THEOREM. *Let $\phi \in C(X^+)$ and let ν be a Ruelle measure for ϕ. Then for $f \in C(X^+)$,*

$$\lim_{n\to\infty} \frac{1}{n} \log \nu(\exp S_n f) = P(f+\phi) - P(\phi).$$

PROOF. Observe that

$$\begin{aligned}(\mathscr{L}_{(f+\phi)}^n 1)(x) &= \sum_{y \in T^{-n}x} e^{(S_n(f+\phi))(y)} \\ &= \sum_{y \in T^{-n}x} e^{(S_n\phi)(y)} \cdot e^{(S_n f)(y)} \\ &= (\mathscr{L}_\phi^n e^{S_n f})(x).\end{aligned}$$

Put $\lambda = e^{P(\phi)}$ so that, by 8, $\mathscr{L}_\phi^* \nu = \lambda \nu$. We have

$$\begin{aligned}P(f+\phi) &= \lim_{n\to\infty} \frac{1}{n} \log \nu(\mathscr{L}_{(f+\phi)}^n 1) \quad (\text{by } 8) \\ &= \lim_{n\to\infty} \frac{1}{n} \log \nu(\mathscr{L}_\phi^n e^{S_n f}) \\ &= \lim_{n\to\infty} \frac{1}{n} \log(\lambda^n \nu(e^{S_n f})) \quad (\text{as } \mathscr{L}_\phi^* \nu = \lambda \nu) \\ &= P(\phi) + \lim_{n\to\infty} \frac{1}{n} \log \nu(\exp S_n f).\end{aligned}$$

If $\phi \in C(X^+)$ is such that $\sum_{n=1}^{\infty} \operatorname{var}_n \phi < \infty$, 9 follows easily from Ruelle's operator theorem, as is shown in [13]. 9 suggests the definition:

For $f, g \in C(X^+)$, $P(f/g) = P(f+g) - P(g)$ is the *conditional pressure of f given g.* 9 states that if ν is a Ruelle measure for g, $(1/n)\log \nu(\exp S_n f) \to P(f/g)$. The following list of properties of conditional pressure is derived from the properties of pressure.

10. PROPOSITION. *Let $f, g, h \in C(X^+)$ and $c \in R$. Then,*

(i) $P(f+g/h) = P(g/h) + P(f/g+h)$,

(ii) $P(f/h) \geqq P(g/h)$ *if* $f \geqq g$,

(iii) $P(f/0) = P(f) - h(T)$ *where $h(T)$ is the topological entropy,*

(iv) $P(f/h) = P(f + g \circ T - g/h)$,

(v) $P(f+c/h) = P(f/h) + c$.

§4. Regular isomorphisms

Let (X_i, T_i, m_i) be finite state processes with state partitions α_i $(i = 1, 2)$. For $i = 1, 2$ denote by α_i^- the "past" σ-algebra generated by the cylinders $[x_0 \cdots x_l]^m$ with $m \geqq 0$. α_i^- is the smallest σ-algebra containing the partitions $T_i^{-n}\alpha_i$, $n \geqq 0$, $\alpha_i^- = \bigvee_{n=0}^{\infty} T_i^{-n}\alpha_i$. An isomorphism $\phi : X_1 \to X_2$ of the processes (i.e. invertible measure-preserving ϕ with $\phi T_1 = T_2 \phi$) is a *regular isomorphism* if there exists $p \geqq 0$ such that $\phi^{-1}\alpha_2^- \subset T^p \alpha_1^-$ and $\phi\alpha_1^- \subset T^p \alpha_2^-$. We need

11. Proposition [11]. *Suppose $\phi : X_1 \to X_2$ is a regular isomorphism of the processes (X_i, T_i, m_i), $i = 1, 2$. Let $1 \leqq q \leqq \infty$. Then $I_{T_1} \in L^q(X_1)$ iff $I_{T_2} \in L^q(X_2)$ and, in this case,*

$$I_{T_1} = I_{T_2} \circ \phi + g \circ T_1 - g$$

for some $g \in L^q(X_1)$.

When $\phi : X_1 \to X_2$ is a regular isomorphism of Markov shifts (X_i, T_i, m_i), $i = 1, 2$, 11 gives

$$I_{T_1} = I_{T_2} \circ \phi + g \circ T_1 - g$$

for some $g \in L^\infty(X_1)$, since the information cocycles of Markov shifts are bounded. Moreover, $P(-I_{T_1}) = P(-I_{T_2}) = 0$. Hence for $t \in R$ we have

$$\begin{aligned} P((t-1)I_{T_1}) &= \lim_{n\to\infty} \frac{1}{n} \log \int e^{tS_n I_{T_1}} dm_1 \\ &= \lim_{n\to\infty} \frac{1}{n} \log \int e^{tS_n I_{T_2} \circ \phi} e^{t(g \circ T_1^n - g)} dm_1 \\ &= \lim_{n\to\infty} \frac{1}{n} \log \int e^{tS_n I_{T_2} \circ \phi} dm_1 \qquad (\text{since } g \in L^\infty(X_1)) \\ &= \lim_{n\to\infty} \frac{1}{n} \log \int e^{tS_n I_{T_2}} dm_2 = P((t-1)I_{T_2}). \end{aligned}$$

Thus, the following analogue of 4 and 6 holds:

12. Theorem. *If the Markov shifts (X_i, T_i, m_i), $i = 1, 2$, are regularly isomorphic then $P(tI_{T_1}) = P(tI_{T_2})$ for all $t \in R$. Two Bernoulli shifts are regularly isomorphic iff there is a trivial regular isomorphism between them.*

It is interesting to note that (on taking $t = 0$ in 12) topological entropy is established as an invariant of regular isomorphism of Markov shifts.

In [4] Keane and Smorodinsky proved that entropy is a complete invariant for finitary isomorphism of Bernoulli shifts. In contrast to this, 12 shows that no finitary isomorphism of "distinct" Bernoulli shifts may have bounded future coding time for both the isomorphism and its inverse. For the intermediate notion of finitary isomorphism with finite expected (future) coding times, Parry has shown in [9] that entropy is not a complete invariant, but it is not known if any such isomorphisms exist between "distinct" Bernoulli shifts. Here we use "distinct" to mean that one process cannot be obtained from the other by a permutation of states.

I would like to thank Dr. Klaus Schmidt for suggesting the use of limits as above as a way of capturing our invariants for Bernoulli shifts.

§5. Finite equivalence, ratio variational principle

A Markov shift (Y, S, p) is said to be a *finite extension* of another, (X, T, m), (and (X, T, m) a *finite factor* of (Y, S, p)) if there exists a bounded-to-one continuous measure-preserving surjection $\phi : Y \to X$ with $\phi S = T\phi$. Such ϕ preserve entropy and $P(f \circ \phi) = P(f)$ for all $f \in C(X)$.

Two Markov shifts (X_1, T_1, m_1) and (X_2, T_2, m_2) are said to be *finitely equivalent* if they have a common Markov finite extension. This is the topological equivalence relations of [8] (also called finite equivalence) and [1] with Markov measures.

It is easy to see from 5 that if (X, T, m) is a Markov shift then $h(m) - \int I_T dm = P(-I_T) = 0$. It is also easy to see that 5 extends to functions of finitely many coordinates and multiple Markov measures. We will make use of these facts in the next proof.

13. Lemma. *Let (X, T, m) be a Markov shift and (Y, S) a topological Markov chain. If $\phi : Y \to X$ is a bounded-to-one continuous surjection satisfying $\phi S = T\phi$ then on (Y, S) there is a unique invariant probability p which makes ϕ measure-preserving. p is multiple Markov.*

Proof. Note that, as ϕ is bounded-to-one, $P(-I_T \circ \phi) = P(-I_T)$ and $h(q \circ \phi^{-1}) = h(q)$ for any invariant $q \in M(Y)$. Since ϕ is continuous $-I_T \circ \phi$ depends only on a finite number of coordinates of Y. Applying the extension of 5 we see that there is a unique invariant $p \in M(Y)$ satisfying $h(p) - \int I_T \circ \phi dp = P(-I_T \circ \phi)$ and that p is multiple Markov. Moreover

$$h(m) - \int I_T dm = P(-I_T) = P(-I_T \circ \phi)$$

$$= h(p) - \int I_T \circ \phi dp = h(p \circ \phi^{-1}) - \int I_T d(p \circ \phi^{-1})$$

and, by the uniqueness in 5 applied to $-I_T$, $m = p \circ \phi^{-1}$. To see that p is the only invariant probability which makes ϕ measure-preserving observe that if invariant $q \in M(Y)$ satisfies $m = q \circ \phi^{-1}$ then

$$h(q) - \int I_T \circ \phi dq = h(m) - \int I_T dm = P(-I_T) = P(-I_T \circ \phi).$$

14. Proposition. *If two Markov shifts have a common Markov finite factor then they have a common Markov finite extension.*

Proof. Let (X_1, T_1, m_1) and (X_2, T_2, m_2) be Markov shifts with another Markov shift (Z, R, q) as a common finite factor, by π_1 and π_2 say. In [1] (2.14 and 3.36) it is shown that there is then a topological Markov chain (Y, S) and, for $i = 1, 2$, bounded-to-one continuous surjections $\phi_i : Y \twoheadrightarrow X_i$ such that $\phi_i S = T_i \phi_i$ and $\pi_1 \phi_1 = \pi_2 \phi_2$. It is sufficient to find a multiple Markov measure p on Y that makes both ϕ_1 and ϕ_2 measure-preserving; we may then pass to a conjugate topological Markov chain on which p gives a Markov measure (of memory one). According to 13 there are uniquely defined multiple Markov measures p_1, p_2 on Y that make ϕ_1, ϕ_2 measure-preserving, respectively. Both p_1 and p_2 make the map $\pi_1 \phi_1 = \pi_2 \phi_2$ measure-preserving, and on applying 13 to $\pi_1 \phi_1$ we see that $p_1 = p_2$.†

We may now easily see that finite equivalence of Markov shifts is an equivalence relation. Reflexivity and symmetry are clear. For transitivity suppose that, for $i = 1, 2$, (Z_i, R_i, q_i) and (Z, R, q) are finitely equivalent by a common finite extension (X_i, T_i, m_i). We then have a diagram of Markov shifts and bounded-to-one measure-preserving surjective maps

$$\begin{tikzcd}
 & & (Y, S, P) \arrow[dl, "\phi_1"'] \arrow[dr, "\phi_2"] & & \\
 & (X_1, T_1, m_1) \arrow[dl, "\pi_1'"'] \arrow[dr, "\pi_1"] & & (X_2, T_2, m_2) \arrow[dl, "\pi_2"'] \arrow[dr, "\pi_2'"] & \\
(Z_1, R_1, q_1) & & (Z, R, q) & & (Z_2, R_2, q_2)
\end{tikzcd}$$

where (Y, S, p) and ϕ_1, ϕ_2 are obtained from 14. Now (Z_1, R_1, q_1) and (Z_2, R_2, q_2) are finite factors of (Y, S, p), by $\pi_1' \circ \phi_1$ and $\pi_2' \circ \phi_2$ respectively, and transitivity is verified.

† I would like to thank the referee for his remark towards the simplification of this proof.

It follows from 3 that if (X_1, T_1, m_1) and (X_2, T_2, m_2) are finitely equivalent Markov shifts then $P(tI_{T_1}) = P(tI_{T_2})$ for all $t \in R$. In particular, Bernoulli shifts given by the probability vectors p and q are not finitely equivalent unless q can be obtained from p by a permutation.

Finally we prove a ratio variational principle (see also [2] and [10]).

15. THEOREM (Ratio Variational Principle). *Let (X^+, T) be a topologically mixing one-sided subshift of finite type given by the 0-1 matrix A with $A^M > 0$. Let $\phi \in C(X^+)$ be such that $\phi > 0$ and $\sum_{n=1}^{\infty} \mathrm{var}_n \phi < \infty$. Then there is a unique T-invariant probability m such that*

$$\frac{h(\mu)}{\mu(\phi)} \leqq \frac{h(m)}{m(\phi)}$$

for all T-invariant probabilities.

PROOF. Consider the Ruelle operators $\mathcal{L}_{-\phi/t}$ for $t \in R$, $t > 0$. By Ruelle's theorem, for each $t > 0$ there are unique $\lambda_t > 0$, $\nu_t, h_t > 0$ such that $\nu_t(h_t) = 1$, $\mathcal{L}_{-\phi/t} h_t = \lambda_t h_t$, $\mathcal{L}^*_{-\phi/t} \nu_t = \lambda_t \nu_t$ and, for $f \in C(X^+)$, $\lambda^{-n}(\mathcal{L}^n_{-\phi/t} f) \to \nu_t(f) h_t$ in $C(X^+)$.

Suppose $t_0 > 0$ is such that $\lambda_{t_0} = 1$. Since $\lambda_{t_0} = e^{P(-\phi/t_0)}$, this means $P(-\phi/t_0) = 0$. By Ruelle's theorem, the measure m defined $m(f) = \nu_{t_0}(h_{t_0} f)$, $f \in C(X^+)$, is the only T-invariant probability for which $h(m) - (1/t_0) m(\phi) = P(-\phi/t_0) = 0$. Thus $h(m)/m(\phi) = 1/t_0$ and, for T-invariant $\mu \neq m$, $h(\mu) - (1/t_0)\mu(\phi) < 0$, i.e. $h(\mu)/\mu(\phi) < 1/t_0$.

It remains to show that there exists $t_0 > 0$ with $\lambda_{t_0} = 1$. We have

$$(\mathcal{L}_{-\phi/t} 1)(x) = \sum_{y \in T^{-1}x} e^{-\phi(y)/t}$$

so $\mathcal{L}_{-\phi/t} 1 \to 0$ in $C(X^+)$ as $t \to 0$. It follows that, as $t \to 0$, $\lambda_t = \nu_t(\mathcal{L}_{-\phi/t} 1) \to 0$. Now take $0 < \varepsilon < \frac{1}{2}$ and $\delta > 0$ such that $|e^s - 1| < \varepsilon$ whenever $|s| < \delta$. Let $t > 0$ be such that $\|\phi/t\| < \delta/M$. Then,

$$(\mathcal{L}^M_{-\phi/t} 1)(x) = \sum_{y \in T^{-M}x} e^{-(1/t)(\phi(y) + \phi(Ty) + \cdots + \phi(T^{M-1}y))}$$

$$\geqq \sum_{y \in T^{-M}x} (1 - \varepsilon) \geqq 2(1 - \varepsilon) > 1 \qquad \text{as } A^M > 0.$$

(The theorem holds when X^+ consists of a point. We assume X^+ is not a point so that there are at least two states.) Now

$$\lambda_t^M = \nu_t(\mathcal{L}^M_{-\phi/t} 1) > 1 \quad \text{and} \quad \lambda_t > 1.$$

Finally, note that, as pressure is continuous, the function $t \to \lambda_t = e^{P(-\phi/t)}$, $t > 0$, is continuous.

For $\phi \in C(X^+)$ such that $\phi > 0$, $\Sigma_{n=1}^{\infty} \mathrm{var}_n \phi < \infty$ put $R(\phi) = h(m)/m(\phi)$ where m is given by 15. We see from 2 that for $s \not\in [-\max I_T, -\min I_T]$, $R(I_T + s)$ are invariants of block isomorphism and of finite equivalence of Markov shifts. These invariants, however, are more difficult to compute than the ones we have used.

References

1. R. L. Adler and B. Marcus, *Topological entropy and equivalence of dynamical systems*, Mem. Amer. Math. Soc. **219** (1979).
2. R. Bowen and D. Ruelle, *The ergodic theory of Axiom A flows*, Invent. Math. **29** (1975), 181–202.
3. N. Friedman and D. Ornstein, *On the isomorphism of weak Bernoulli transformations*, Advances in Math. **5** (1970), 365–394.
4. M. Keane and M. Smorodinsky, *Bernoulli schemes of the same entropy are finitarily isomorphic*, Ann. of Math. **109** (1979), 397–406.
5. M. Keane and M. Smorodinsky, *Finitary isomorphism of irreducible Markov shifts*, Israel J. Math. **34** (1979), 281–286.
6. O. E. Lanford and D. Ruelle, *Observables at infinity and states with short range correlations in Statistical Mechanics*, Comm. Math. Phys. **13** (1969), 194–215.
7. D. S. Ornstein, *Bernoulli shifts with the same entropy are isomorphic*, Advances in Math. **4** (1970), 337–352.
8. W. Parry, *A finitary classification of topological Markov chains and sofic systems*, Bull. London Math. Soc. **9** (1977), 86–92.
9. W. Parry, *Finitary isomorphisms with finite expected code lengths*, Bull. London Math. Soc. **11** (1979), 170–176.
10. W. Parry, *Topological Markov chains and suspensions*, University of Warwick notes.
11. W. Parry, University of Warwick lecture notes.
12. W. Parry and R. F. Williams, *Block coding and a zeta function for finite Markov chains*, Proc. London Math. Soc. **35** (1977), 483–495.
13. D. Ruelle, *Thermodynamic Formalism*, Addison-Wesley, Reading, Mass., 1978.
14. E. Senata, *Non-negative Matrices*, Allen and Unwin, London, 1973.
15. P. Walters, *Ruelle's operator theorem and g-measures*, Trans. Amer. Math. Soc. **214** (1975), 375–387.
16. P. Walters, *A variational principle for the pressure of continuous transformations*, Amer. J. Math. **97** (1976), 937–971.

MATHEMATICS INSTITUTE
UNIVERSITY OF WARWICK
COVENTRY CV4 7AL, ENGLAND