MEASURES OF MAXIMAL RELATIVE ENTROPY

KARL PETERSEN, ANTHONY QUAS, AND SUJIN SHIN

Abstract. Given an irreducible subshift of finite type X, a subshift Y, a factor map $\pi : X \to Y$, and an ergodic invariant measure ν on Y, there can exist more than one ergodic measure on X which projects to ν and has maximal entropy among all measures in the fiber, but there is an explicit bound on the number of such maximal entropy preimages.

1. Introduction

It is a well-known result of Shannon and Parry [17, 12] that every irreducible subshift of finite type (SFT) X on a finite alphabet has a unique measure μ_X of maximal entropy for the shift transformation σ. The maximal measure is Markov, and its initial distribution and transition probabilities are given explicitly in terms of the maximum eigenvalue and corresponding eigenvectors of the 0,1 transition matrix for the subshift. We are interested in any possible relative version of this result: given an irreducible SFT X, a subshift Y, a factor map $\pi : X \to Y$, and an ergodic invariant measure ν on Y, how many ergodic invariant measures can there be on X that project under π to ν and have maximal entropy in the fiber $\pi^{-1}\{\nu\}$? We will show that there can be more than one such ergodic relatively maximal measure over a given ν, but there are only finitely many. In fact, if π is a 1-block map, there can be no more than the cardinality of the alphabet of X (see Corollary 1, below). Call a measure ν on Y *π-determinate* in case it has a unique preimage of maximal entropy. We provide some sufficient conditions for π-determinacy and give examples of situations in which relatively maximal measures can be constructed explicitly.

Throughout the paper, unless stated otherwise X will denote an irreducible SFT, Y a subshift on a finite alphabet, and $\pi : X \to Y$ a factor map (continuous, onto, shift-commuting map). By recoding if necessary, we may assume that X is a 1-step SFT, so that it consists of all (2-sided) sequences on a finite alphabet consistent with the allowed transitions described by a directed graph with vertex set equal to the alphabet, and that π is a 1-block map. In the following, "measure" means "Borel probability measure", $\mathcal{C}(X)$ denotes the set of continuous real-valued functions on X, $\mathcal{M}(X)$ the space of σ-invariant measures on X, and $\mathcal{E}(X) \subset \mathcal{M}(X)$ the set of ergodic measures on X.

Some of the interest of this problem arises from its connections (discussed in [14]) with information-compressing channels [11], non-Markov functions of Markov chains [1, 2, 3, 11], measures of maximal Hausdorff dimension and measures that maximize, for a given $\alpha > 0$, the weighted entropy functional

$$\phi_\alpha(\mu) = \frac{1}{\alpha+1}[h(\mu) + \alpha h(\pi\mu)] \tag{1}$$

[5, 18, 19], and relative pressure and relative equilibrium states [9, 20]. The theory of pressure and equilibrium states (see [16, 6, 7]), relative pressure and relative equilibrium states [8, 20], and compensation functions [2, 20] provides basic tools in this area. For a factor map $\pi : X \to Y$ between compact topological dynamical systems and potential function $V \in \mathcal{C}(X)$, Ledrappier and Walters [8] defined the *relative pressure* $P(\pi, V) : Y \to \mathbb{R}$ (a Borel measurable function) and proved a relative variational principle: For each $\nu \in \mathcal{M}(Y)$,

$$\int_Y P(\pi, V)\, d\nu = \sup\{h_\mu(X|Y) + \int_X V\, d\mu : \mu \in \pi^{-1}\nu\}. \tag{2}$$

Any measure μ that attains the supremum is called a *relative equilibrium state*. A consequence is that the ergodic measures μ that have maximal entropy among all measures in $\pi^{-1}\{\nu\}$ have relative entropy given by

$$h_\mu(X|Y) = \int_Y \lim_{n\to\infty} \frac{1}{n} \log|\pi^{-1}[y_0 \dots y_{n-1}]|\, d\nu(y). \tag{3}$$

($|\pi^{-1}[y_0 \dots y_{n-1}]|$ is the number of n-blocks in X that map under π to the n-block $y_0 \dots y_{n-1}$.) By the Subadditive Ergodic Theorem, the limit inside the integral exists a.e. with respect to each ergodic measure ν on Y, and it is constant a.e.. The quantity

$$P(\pi, 0)(y) = \limsup_{n\to\infty} \frac{1}{n} \log|\pi^{-1}[y_0 \dots y_{n-1}]| \tag{4}$$

is the *relative pressure* of the function 0 over $y \in Y$. The maximum possible relative entropy may be thought of as a "relative topological entropy over ν"; we denote it by $h_{\text{top}}(X|\nu)$.

To understand when a Markov measure on Y has a Markov measure on X in its preimage under π, Boyle and Tuncel introduced the idea of a compensation function [2], and the concept was developed further by Walters [20]. Given a factor map $\pi : X \to Y$ between topological dynamical systems, a *compensation function* is a continuous function $F : X \to \mathbb{R}$ such that

$$P_Y(V) = P_X(V \circ \pi + F) \quad \text{for all } V \in \mathcal{C}(Y). \tag{5}$$

The idea is that, because $\pi : \mathcal{M}(X) \to \mathcal{M}(Y)$ is many-to-one, we always have

$$P_Y(V) = \sup\{h_\nu(\sigma) + \int_Y V\, d\nu : \nu \in \mathcal{M}(Y)\} \tag{6}$$

$$\leq \sup\{h_\mu(\sigma) + \int_X V \circ \pi\, d\mu : \mu \in \mathcal{M}(X)\}, \tag{7}$$

and a compensation function F can take into account, for all potential functions V on Y at once, the extra freedom, information, or free energy that is available in X as compared to Y because of the ability to move around in fibers over points of Y. A compensation function of the form $G \circ \pi$ with $G \in \mathcal{C}(Y)$ is said to be *saturated*.

The machinery of relative equilibrium states and compensation functions is used to establish the following basic result about relatively maximal measures [18, 20]:

Suppose that $\nu \in \mathcal{E}(Y)$ and $\pi\mu = \nu$. Then μ is relatively maximal over ν if and only if there is $V \in \mathcal{C}(Y)$ such that μ is an equilibrium state of $V \circ \pi$.

Notice that if there is a *locally constant* saturated compensation function $G \circ \pi$, then every Markov measure on Y is π-determinate with Markov relatively maximal lift, because in [20] it is shown that if there is a saturated compensation function $G \circ \pi$, then the relatively maximal measures over an equilibrium state of $V \in \mathcal{C}(Y)$ are the equilibrium states of $V \circ \pi + G \circ \pi$.

Further, μ_X is the unique equilibrium state of the potential function 0 on X, the unique maximizing measure for ϕ_0; and the relatively maximal measures over μ_Y are the equilibrium states of $G \circ \pi$, which can be thought of as the maximizing measures for ϕ_∞.

2. Bounding the number of ergodic relatively maximal measures

Let $\pi : X \to Y$ be a 1-block factor map from a 1-step SFT X to a subshift Y and let ν be an ergodic invariant measure on Y. Let $\mu_1, \dots, \mu_n \in \mathcal{M}(X)$ with $\pi\mu_i = \nu$ for all i. Recall the definition of the *relatively independent joining* $\hat{\mu} = \mu_1 \otimes \cdots \otimes_\nu \mu_n$ of $\mu_1, \dots, \mu_n$ over ν: if $A_1, \dots, A_n$ are measurable subsets of X and $\mathcal{F}$ is the σ-algebra of Y, then

$$\hat{\mu}(A_1 \times \ldots \times A_n) = \int_Y \prod_{i=1}^{n} \mathbb{E}_{\mu_i}(\mathbf{1}_{A_i} | \pi^{-1}\mathcal{F}) \circ \pi^{-1} \, d\nu. \tag{8}$$

Writing p_i for the projection $X^n \to X$ onto the i'th coordinate, we note that for $\hat{\mu}$-almost every $\hat{x}$ in X^n, $\pi(p_i(\hat{x}))$ is independent of i; denote it by $\phi(\hat{x})$.

We define a number of σ-algebras on X^n. Denoting by $\mathcal{B}_X$ the σ-algebra of X and by $\mathcal{B}_Y$ the σ-algebra of Y, let $\mathcal{B}_0 = \phi^{-1}\mathcal{B}_Y$, $\mathcal{B}_i = p_i^{-1}\mathcal{B}_X$ for $i = 1, \dots, n$, $\mathcal{B}_X^-$ the σ-algebra generated by $x_n, n < 0$, and $\mathcal{B}_i^- = p_i^{-1}\mathcal{B}_X^-$ for each i. *Note*: later we will use the same symbols for corresponding sub-σ-algebras of a different space, $Z = X \times X \times R$.

Definition. We say that two measures $\mu_1, \mu_2 \in \mathcal{E}(X)$ with $\pi\mu_1 = \pi\mu_2 = \nu$ are *relatively orthogonal* (over ν) and write $\mu_1 \perp_\nu \mu_2$ if

$$(\mu_1 \otimes_\nu \mu_2)\{(u, v) \in X \times X : u_0 = v_0\} = 0. \tag{9}$$

Theorem 1. *For each ergodic ν on Y, any two distinct ergodic measures on X of maximal entropy in the fiber $\pi^{-1}\{\nu\}$ are relatively orthogonal.*

Since π is a 1-block factor map, for each symbol b in the alphabet of Y, $\pi^{-1}[b]$ consists of a union of 1-block cylinder sets in X. Let $N_\nu(\pi)$ denote the minimum number of cylinders in the union as b runs over the symbols in the alphabet of Y for which $\nu[b] > 0$.

Corollary 1. *Let X be a 1-step SFT, Y a subshift on a finite alphabet, and $\pi : X \to Y$ a 1-block factor map. For any ergodic ν on Y, the number of ergodic invariant measures of maximal entropy in the fiber $\pi^{-1}\{\nu\}$ is at most $N_\nu(\pi)$.*

Proof. Suppose that we have $n > N_\nu(\pi)$ ergodic measures $\mu_1, \dots, \mu_n$ on X, each projecting to ν and each of maximal entropy in the fiber

$\pi^{-1}\{\nu\}$. Form the relatively independent joining $\hat{\mu}$ on X^n of the measures μ_i as above. Let b be a symbol in the alphabet of Y such that b has $N_\nu(\pi)$ preimages $a_1, \dots, a_{N_\nu(\pi)}$ under the block map π. Since $n > N_\nu(\pi)$, for every $\hat{x} \in \phi^{-1}[b]$ there are $i \neq j$ with $(p_i\hat{x})_0 = (p_j\hat{x})_0$. At least one of the sets $S_{i,j} = \{\hat{x} \in X^n : (p_i\hat{x})_0 = (p_j\hat{x})_0\}$ must have positive $\hat{\mu}$-measure, and then also $(\mu_i \otimes_\nu \mu_j)\{(u,v) \in X \times X : \pi u = \pi v, u_0 = v_0\} > 0$, contradicting Theorem 1. □

Corollary 2. *Suppose that $\pi : X \to Y$ has a singleton clump: there is a symbol a of Y whose inverse image is a singleton, which we also denote by a. Then every ergodic measure on Y which assigns positive measure to $[a]$ is π-determinate.*

Before giving the proof of Theorem 1, we recall some facts about conditional independence of σ-algebras (see [10, p. 17]) and prove a key lemma.

Lemma 1. *Let $(X, \mathcal{B}, \mu)$ be a probability space. For sub-σ-algebras $\mathcal{B}_0, \mathcal{B}_1, \mathcal{B}_2$ of $\mathcal{B}$, the following are equivalent:*

1. *$\mathcal{B}_1 \perp_{\mathcal{B}_0} \mathcal{B}_2$, which is defined by the condition that for every $\mathcal{B}_1$-measurable f_1 and $\mathcal{B}_2$-measurable f_2, $\mathbb{E}(f_1 f_2|\mathcal{B}_0) = \mathbb{E}(f_1|\mathcal{B}_0)\mathbb{E}(f_2|\mathcal{B}_0)$;*
2. *for every $\mathcal{B}_2$-measurable f_2, $\mathbb{E}(f_2|\mathcal{B}_1 \vee \mathcal{B}_0) = \mathbb{E}(f_2|\mathcal{B}_0)$;*
3. *for every $\mathcal{B}_1$-measurable f_1, $\mathbb{E}(f_1|\mathcal{B}_2 \vee \mathcal{B}_0) = \mathbb{E}(f_1|\mathcal{B}_0)$.*

Lemma 2. *Let $(X, \mathcal{B}, \mu)$ be a probability space and let $\mathcal{B}_1, \mathcal{B}_2, \mathcal{C}_1, \mathcal{C}_2$ be sub-σ-algebras of $\mathcal{B}$. If $\mathcal{B}_1 \perp_{\mathcal{B}_0} \mathcal{B}_2, \mathcal{C}_1 \subset \mathcal{B}_1, \mathcal{C}_2 \subset \mathcal{B}_2$, then for every $\mathcal{B}_1$-measurable f_1,*

$$\mathbb{E}(f_1|\mathcal{B}_0 \vee \mathcal{C}_1 \vee \mathcal{C}_2) = \mathbb{E}(f_1|\mathcal{B}_0 \vee \mathcal{C}_1). \tag{10}$$

Proof. First note that $\mathcal{B}_1 \perp_{\mathcal{B}_0 \vee \mathcal{C}_2} \mathcal{B}_2$, since for $\mathcal{B}_1$-measurable f_1 we have $\mathbb{E}(f_1|(\mathcal{B}_0 \vee \mathcal{C}_2) \vee \mathcal{B}_2) = \mathbb{E}(f_1|\mathcal{B}_0 \vee \mathcal{B}_2) = \mathbb{E}(f_1|\mathcal{B}_0) = \mathbb{E}(f_1|\mathcal{B}_0 \vee \mathcal{C}_2)$. Similarly, $\mathcal{B}_1 \perp_{\mathcal{B}_0 \vee \mathcal{C}_1} \mathcal{B}_2$ and $\mathcal{B}_1 \perp_{\mathcal{B}_0 \vee \mathcal{C}_1} \mathcal{C}_2$. Thus for any f_1 that is $\mathcal{B}_1$-measurable, $\mathbb{E}(f_1|(\mathcal{B}_0 \vee \mathcal{C}_1) \vee \mathcal{C}_2) = \mathbb{E}(f_1|\mathcal{B}_0 \vee \mathcal{C}_1)$. □

Lemma 3. *Let $\pi : X \to Y$ be a 1-block factor map from a 1-step SFT X to a subshift Y. Let ν be an ergodic measure on Y and let μ_1 and μ_2 be ergodic members of $\pi^{-1}\{\nu\}$. Let $\hat{\mu}$ be their relatively independent joining. If $S = \{(u,v) \in X \times X : u_{-1} = v_{-1}\}$ has positive measure with respect to $\hat{\mu}$ and for every symbol j in the alphabet of X*

$$\mathbb{E}_{\hat{\mu}}(1_{[j]} \circ p_1|\mathcal{B}_1^- \vee \mathcal{B}_0) = \mathbb{E}_{\hat{\mu}}(1_{[j]} \circ p_2|\mathcal{B}_2^- \vee \mathcal{B}_0) \quad \text{a.e. on } S, \tag{11}$$

then $\mu_1 = \mu_2$.

Proof. Write $[i]_k$ for the set of points in X whose k'th symbol is i and $[i]_k^{(j)}$ for $p_j^{-1}[i]_k$. Write $1_{[i]_k^{(j)}}$ for the indicator function of this set. Define $g_i^{(j)} = \mathbb{E}(1_{[i]_0^{(j)}}|\mathcal{B}_0 \vee \mathcal{B}_j^-)$ and set $s_k = \sum_i 1_{[i]_k^{(1)}} 1_{[i]_k^{(2)}} = 1_{\{(u,v):u_k=v_k\}}$. Note that $s_{-1} = 1_S$.

Let $\mathcal{P}$ denote the time-0 partition of X into 1-block cylinder sets, $\mathcal{P}_i = p_i^{-1}\mathcal{P}$ $(i = 1, 2)$ the corresponding partitions of $X \times X$, and $T = \sigma \times \sigma$.

By assumption, we have $s_{-1}g_i^{(1)} = s_{-1}g_i^{(2)}$ for all symbols i in the alphabet of X. Taking expectations with respect to $\mathcal{B}_1 \vee T\mathcal{P}_2$, since $s_{-1}g_i^{(1)}$ is $\mathcal{B}_1 \vee T\mathcal{P}_2$-measurable, we see that

$$
\begin{aligned}
s_{-1}g_i^{(1)} &= s_{-1}\mathbb{E}(g_i^{(2)}|\mathcal{B}_1 \vee T\mathcal{P}_2) \\
&= s_{-1}\sum_j \frac{\mathbb{E}(g_i^{(2)}1_{[j]_{-1}^{(2)}}|\mathcal{B}_1)}{\mathbb{E}(1_{[j]_{-1}^{(2)}}|\mathcal{B}_1)} 1_{[j]_{-1}^{(2)}} \\
&= s_{-1}\sum_j \frac{\mathbb{E}(g_i^{(2)}1_{[j]_{-1}^{(2)}}|\mathcal{B}_0)}{\mathbb{E}(1_{[j]_{-1}^{(2)}}|\mathcal{B}_0)} 1_{[j]_{-1}^{(2)}},
\end{aligned}
\tag{12}
$$

where the last equality follows from Lemma 1, noting that $\mathcal{B}_0 \subset \mathcal{B}_1$. Observe that the terms in the final expression are all measurable with respect to $\mathcal{B}_0 \vee T\mathcal{P}_1 \vee T\mathcal{P}_2$.

It then follows that

$$
s_{-1}g_i^{(1)} = \mathbb{E}(s_{-1}g_i^{(1)}|\mathcal{B}_0 \vee T\mathcal{P}_1 \vee T\mathcal{P}_2) = s_{-1}\mathbb{E}(g_i^{(1)}|\mathcal{B}_0 \vee T\mathcal{P}_1 \vee T\mathcal{P}_2). \tag{13}
$$

Since $g_i^{(1)}$ is $\mathcal{B}_1$-measurable and $\mathcal{B}_1$ and $\mathcal{B}_2$ are relatively independent over $\mathcal{B}_0$, by Lemma 2 the right side is equal to $s_{-1}\mathbb{E}(g_i^{(1)}|\mathcal{B}_0 \vee T\mathcal{P}_1)$. We have thus established the equation

$$
s_{-1}\mathbb{E}(g_i^{(1)}|\mathcal{B}_0 \vee T\mathcal{P}_1) = s_{-1}g_i^{(1)} = s_{-1}g_i^{(2)} = s_{-1}\mathbb{E}(g_i^{(2)}|\mathcal{B}_0 \vee T\mathcal{P}_2). \tag{14}
$$

Starting from the equation $s_{-1}g_i^{(1)} = s_{-1}\mathbb{E}(g_i^{(2)}|\mathcal{B}_0 \vee T\mathcal{P}_2)$, we take conditional expectations with respect to $\mathcal{B}_1$ to get

$$
\mathbb{E}(s_{-1}|\mathcal{B}_1)g_i^{(1)} = \mathbb{E}(s_{-1}\mathbb{E}(g_i^{(2)}|\mathcal{B}_0 \vee T\mathcal{P}_2)|\mathcal{B}_1). \tag{15}
$$

We have

$$
\mathbb{E}(g_i^{(2)}|\mathcal{B}_0 \vee T\mathcal{P}_2) = \sum_k \frac{\mathbb{E}(g_i^{(2)}1_{[k]_{-1}^{(2)}}|\mathcal{B}_0)}{\mathbb{E}(1_{[k]_{-1}^{(2)}}|\mathcal{B}_0)} 1_{[k]_{-1}^{(2)}}. \tag{16}
$$

Hence

$$s_{-1}\mathbb{E}(g_i^{(2)}|\mathcal{B}_0 \vee T\mathcal{P}_2) = \sum_k \frac{\mathbb{E}(g_i^{(2)} 1_{[k]_{-1}^{(2)}}|\mathcal{B}_0)}{\mathbb{E}(1_{[k]_{-1}^{(2)}}|\mathcal{B}_0)} 1_{[k]_{-1}^{(1)}} 1_{[k]_{-1}^{(2)}}. \tag{17}$$

Substituting this in (15) and again using relative independence, we see that

$$\begin{aligned} \mathbb{E}(s_{-1}|\mathcal{B}_1)g_i^{(1)} &= \sum_k \frac{\mathbb{E}(g_i^{(2)} 1_{[k]_{-1}^{(2)}}|\mathcal{B}_0)}{\mathbb{E}(1_{[k]_{-1}^{(2)}}|\mathcal{B}_0)} 1_{[k]_{-1}^{(1)}} \mathbb{E}(1_{[k]_{-1}^{(2)}}|\mathcal{B}_1) \\ &= \sum_k \mathbb{E}(g_i^{(2)} 1_{[k]_{-1}^{(2)}}|\mathcal{B}_0) 1_{[k]_{-1}^{(1)}}. \end{aligned} \tag{18}$$

We observe that the right-hand side and also $\mathbb{E}(s_{-1}|\mathcal{B}_1)$ are $\mathcal{B}_0 \vee T\mathcal{P}_1$-measurable (using the definition of s_{-1} and relative independence). Hence provided that $\mathbb{E}(s_{-1}|\mathcal{B}_1) > 0$ a.e., we will have that $g_i^{(1)}$ is $\mathcal{B}_0 \vee T\mathcal{P}_1$-measurable, and similarly $g_i^{(2)}$ is $\mathcal{B}_0 \vee T\mathcal{P}_2$-measurable.

We now demonstrate that $\mathbb{E}(s_{-1}|\mathcal{B}_1) > 0$ on a set of full measure. To prove this, we note that $\mathbb{E}(s_{-1}|\mathcal{B}_1)$ is of the form $f \circ p_1$ for f a function on X. Thus if we can show that $\mathbb{E}(s_{-1}|\mathcal{B}_1)(x) > 0$ implies $\mathbb{E}(s_{-1}|\mathcal{B}_1)(Tx) > 0$, it will follow that the set where f is positive is invariant and hence of measure 0 or 1 by ergodicity of μ_1. Since the integral of the function is positive (being equal to $(\mu_1 \otimes_\nu \mu_2)\{(u,v) : u_{-1} = v_{-1}\}$), to show that the function is positive on a set of full measure it is enough to establish the above invariance.

Now

$$\begin{aligned} \mathbb{E}(s_{-1}|\mathcal{B}_1)(Tx) &= \mathbb{E}(s_0|\mathcal{B}_1)(x) \\ &= \sum_i \mathbb{E}(1_{[i]_0^{(1)}} 1_{[i]_0^{(2)}}|\mathcal{B}_1) \\ &= \sum_i 1_{[i]_0^{(1)}} \mathbb{E}(1_{[i]_0^{(2)}}|\mathcal{B}_1) \\ &\geq \sum_i 1_{[i]_0^{(1)}} \mathbb{E}(s_{-1} 1_{[i]_0^{(2)}}|\mathcal{B}_1) \\ &= \sum_i 1_{[i]_0^{(1)}} \mathbb{E}(\mathbb{E}(s_{-1} 1_{[i]_0^{(2)}}|\mathcal{B}_1 \vee T\mathcal{P}_2)|\mathcal{B}_1). \end{aligned} \tag{19}$$

Using Lemma 2, this equals

$$
\begin{aligned}
&\sum_i 1_{[i]_0^{(1)}}\mathbb{E}(s_{-1}\mathbb{E}(1_{[i]_0^{(2)}}|\mathcal{B}_1 \vee T\mathcal{P}_2)|\mathcal{B}_1)\\
&= \sum_i 1_{[i]_0^{(1)}}\mathbb{E}(s_{-1}\mathbb{E}(1_{[i]_0^{(2)}}|\mathcal{B}_0 \vee T\mathcal{P}_2)|\mathcal{B}_1)\\
&= \sum_i 1_{[i]_0^{(1)}}\mathbb{E}(s_{-1}g_i^{(2)}|\mathcal{B}_1) \text{ (from (14))}\\
&= \sum_i 1_{[i]_0^{(1)}}\mathbb{E}(s_{-1}g_i^{(1)}|\mathcal{B}_1)\\
&= \sum_i g_i^{(1)} 1_{[i]_0^{(1)}}\mathbb{E}(s_{-1}|\mathcal{B}_1)\\
&= \mathbb{E}(s_{-1}|\mathcal{B}_1)\sum_i 1_{[i]_0^{(1)}}\mathbb{E}(1_{[i]_0^{(1)}}|\mathcal{B}_0 \vee \mathcal{B}_1^-).
\end{aligned} \tag{20}
$$

For x in a set of full measure, $1_D(x) > 0$ implies $\mathbb{E}(1_D|\mathcal{F})(x) > 0$ (consider integrating the conditional expectation over the set where it takes the value 0), so the sum on the right-hand side of the above is positive almost everywhere. Since the first factor is positive by assumption, the conclusion that $\mathbb{E}(s_0|\mathcal{B}_1) > 0$ follows, allowing us to deduce that $g_i^{(j)}$ is $\mathcal{B}_0 \vee T\mathcal{P}_j$-measurable.

Now we may write $g_i^{(j)}$ as

$$g_i^{(j)} = \sum_k 1_{[k]_{-1}^{(j)}} h_{k,i}^{(j)}, \tag{21}$$

where the $h_{k,i}^{(j)}$ are $\mathcal{B}_0$-measurable. Writing out the equation $s_{-1}g_i^{(1)} = s_{-1}g_i^{(2)}$, we have

$$\sum_k 1_{[k]_{-1}^{(1)}} 1_{[k]_{-1}^{(2)}} h_{k,i}^{(1)} = \sum_k 1_{[k]_{-1}^{(1)}} 1_{[k]_{-1}^{(2)}} h_{k,i}^{(2)}. \tag{22}$$

Since for distinct k, the terms are disjointly supported, we have for each k,

$$1_{[k]_{-1}^{(1)}} 1_{[k]_{-1}^{(2)}} h_{k,i}^{(1)} = 1_{[k]_{-1}^{(1)}} 1_{[k]_{-1}^{(2)}} h_{k,i}^{(2)}. \tag{23}$$

Taking conditional expectations of both sides with respect to $\mathcal{B}_0$ and using Lemma 1, we deduce

$$\mathbb{E}(1_{[k]_{-1}^{(1)}}|\mathcal{B}_0)\mathbb{E}(1_{[k]_{-1}^{(2)}}|\mathcal{B}_0)(h_{k,i}^{(1)} - h_{k,i}^{(2)}) = 0 \quad \text{a.e.} \tag{24}$$

From this we see that if $\mathbb{E}(1_{[k]^{(1)}_{-1}}|\mathcal{B}_0) > 0$ and $\mathbb{E}(1_{[k]^{(2)}_{-1}}|\mathcal{B}_0) > 0$, then $h^{(1)}_{k,i} = h^{(2)}_{k,i}$. This allows us to make the following definition:

$$h_{k,i} = \begin{cases} h^{(1)}_{k,i} & \text{if } \mathbb{E}(1_{[k]^{(1)}_{-1}}|\mathcal{B}_0) > 0 \\ h^{(2)}_{k,i} & \text{if } \mathbb{E}(1_{[k]^{(2)}_{-1}}|\mathcal{B}_0) > 0 \ . \end{cases} \tag{25}$$

It follows that

$$g^{(j)}_i = \sum_k h_{k,i} 1_{[k]^{(j)}_{-1}} \quad \hat{\mu}\text{-a.e..} \tag{26}$$

We now show that the two measures agree. We will show by induction on the length of the cylinder set that for any $\mathcal{B}_0$-measurable function f and any cylinder set C in X,

$$\int 1_S 1_C \circ p_1 f \, d\hat{\mu} = \int 1_S 1_C \circ p_2 f \, d\hat{\mu}. \tag{27}$$

To start the induction, let C be the cylinder set $[i_0]$ in X. Then

$$\begin{aligned} \int 1_S 1_{[i_0]^{(j)}} f \, d\hat{\mu} &= \int 1_S f \mathbb{E}(1_{[i_0]^{(j)}}|\mathcal{B}_0 \vee \mathcal{B}_1^- \vee \mathcal{B}_2^-) \, d\hat{\mu} \\ &= \int 1_S f g^{(j)}_{i_0} \, d\hat{\mu}; \end{aligned} \tag{28}$$

but by assumption $1_S g^{(1)}_i = 1_S g^{(2)}_i$, showing the result in the case that C is a cylinder of length 1. Now suppose that the result holds for cylinders of length n and let $C = [i_0 \dots i_n]$. Write $D = [i_0 \dots i_{n-1}]$. Now

$$\begin{aligned} \int 1_S(1_C \circ p_j) f \, d\hat{\mu} &= \int 1_S(1_D \circ p_j) 1_{[i_n]^{(j)}_n} f \, d\hat{\mu} \\ &= \int 1_S(1_D \circ p_j) f \mathbb{E}(1_{[i_n]^{(j)}_n}|T^{-n}\mathcal{B}_1^- \vee T^{-n}\mathcal{B}_2^- \vee \mathcal{B}_0) \, d\hat{\mu} \\ &= \int 1_S(1_D \circ p_j) f g^{(j)}_{i_n} \circ T^n \, d\hat{\mu} \\ &= \int 1_S(1_D \circ p_j) f h_{i_{n-1},i_n} \circ T^n \, d\hat{\mu}. \end{aligned} \tag{29}$$

Since $h_{i_{n-1},i_n} \circ T^n$ is $\mathcal{B}_0$-measurable, it follows from the induction hypothesis that the integrals are equal for $j = 1$ and $j = 2$ as required.

In particular, taking f to be 1, we have $\hat{\mu}(S \cap p_1^{-1}C) = \hat{\mu}(S \cap p_2^{-1}C)$ for all C. Letting $\hat{\nu}(A) = \hat{\mu}(S \cap A)$, we see that $\hat{\nu} \circ p_1^{-1} = \hat{\nu} \circ p_2^{-1}$. Since

$\mu_i(A) \geq \hat{\nu} \circ p_i^{-1}(A)$ for all A and the measures μ_i are ergodic, it follows that μ_1 and μ_2 are not mutually singular and hence are equal. □

Proof of Theorem 1. Let μ_1 and μ_2 be two different ergodic relatively maximal measures over $\nu \in \mathcal{E}(Y)$ and suppose that they are not relatively orthogonal, so that $(\mu_1 \otimes_\nu \mu_2)\{(u,v) \in X \times X : u_0 = v_0\} > 0$. Let $\hat{\mu} = \mu_1 \otimes_\nu \mu_2$. We will construct a measure on X with strictly greater entropy than μ_1 or μ_2 by building a larger space from which the new measure will appear as a factor. (J. Steif reminded us that a similar interleaving of two processes is used in [4] for a different purpose.)

Let R denote the set $\{1,2\}^{\mathbb{Z}}$, and let β be the Bernoulli measure on R with probabilities $\frac{1}{2}, \frac{1}{2}$. Write $(r_n)_{n \in \mathbb{Z}}$ for a typical element of R. Form $Z = X^2 \times R$ with invariant measure $\eta = \hat{\mu} \times \beta$. We then define maps from Z to X as follows. Given a point $(u,v,r) \in Z$, set $\pi_1(u,v,r) = u$, $\pi_2(u,v,r) = v$ and write $N_k(u,v)$ for $\sup\{n < k \colon u_n = v_n\}$. Note that this quantity may be $-\infty$ if there are no coincidences. We will take $r_{-\infty}$ to be a further random variable taking the values 1 and 2 with equal probability for each $r \in R$. Define $\pi_3 \colon Z \to X$ by

$$\pi_3(u,v,r)_k = \begin{cases} u_k & \text{if } r_{N_k(u,v)} = 1 \\ v_k & \text{if } r_{N_k(u,v)} = 2\ . \end{cases} \tag{30}$$

To see that $\pi_3(u,v,r)$ is indeed a point of X, note that it consists of concatenations of parts of u and v, changing only at places where they agree. As a corollary, since $\pi(u) = \pi(v)$ for almost all $(u,v,r) \in Z$, it follows that $\pi(\pi_3(z)) = \pi(\pi_2(z)) = \pi(\pi_1(z))$ for η-almost every z in Z. Write Φ for the factor mapping $\pi \circ \pi_1$ from (Z,η) to (Y,ν).

By construction $\mu_1 = \eta \circ \pi_1^{-1}$ and $\mu_2 = \eta \circ \pi_2^{-1}$. Define $\mu_3 = \eta \circ \pi_3^{-1}$. We shall then demonstrate that $h_{\mu_3}(X) > h_{\mu_1}(X) = h_{\mu_2}(X)$.

We define σ-algebras on Z corresponding to those appearing above. Letting $\mathcal{B}_X$ be the Borel σ-algebra on X as before, we set for each $i = 1,2,3$, $\mathcal{B}_i = \pi_i^{-1}\mathcal{B}_X$. Write $\mathcal{B}_X^-$ for the σ-algebra generated by the cylinder sets in X depending on coordinates x_n for $n < 0$. These then give σ-algebras $\mathcal{B}_i^-$ on Z defined by $\mathcal{B}_i^- = \pi_i^{-1}\mathcal{B}_X^-$. We will require two further σ-algebras, $\mathcal{B}_0 = \Phi^{-1}\mathcal{B}_Y$ with $\mathcal{B}_0^-$ being defined analogously to the above. Note that $\mathcal{B}_i \supset \mathcal{B}_0$ for $i = 1,2,3$.

Again reusing previous notation in a slightly different context, continue to denote by $\mathcal{P}$ the partition of X into time 0 cylinders and write

$\mathcal{P}_i$ for $\pi_i^{-1}\mathcal{P}$, so that for $i = 1, 2, 3$, $\mathcal{P}_i$ is a partition of Z. Finally, write $\mathcal{Q} = \Phi^{-1}\{[j]\colon [j] \text{ is a cylinder set in } Y\}$.

It is useful to note the following property of (8): If $A_1 \in \mathcal{B}_1$ and $A_2 \in \mathcal{B}_2$, then

$$\eta(A_1 \cap A_2) = \int \mathbb{E}_\eta(1_{A_1}|\mathcal{B}_0)\mathbb{E}_\eta(1_{A_2}|\mathcal{B}_0)\, d\eta. \tag{31}$$

We will use the fact that if f is $\mathcal{B}_1$-measurable, then

$$\mathbb{E}_\eta(f|\mathcal{B}_2) = \mathbb{E}_\eta(f|\mathcal{B}_0), \tag{32}$$

a consequence of Lemma 1.

Standard results of entropy theory tell us that $h_{\mu_i}(X) = H_\eta(\mathcal{P}_i|\mathcal{B}_i^-)$. Further, by Pinsker's Formula (see [13, Theorem 6.3, p. 67], applied with β coarser than α), this can be re-expressed as

$$h_{\mu_i}(X) = H_\eta(\mathcal{P}_i|\mathcal{B}_i^- \vee \mathcal{B}_0) + H_\eta(\mathcal{Q}|\mathcal{B}_0^-) = H_\eta(\mathcal{P}_i|\mathcal{B}_i^- \vee \mathcal{B}_0) + h_\nu(Y). \tag{33}$$

Since μ_1 and μ_2 were presumed to be measures of maximal entropy in the fiber, they have equal entropy and hence $H_\eta(\mathcal{P}_1|\mathcal{B}_1^- \vee \mathcal{B}_0) = H_\eta(\mathcal{P}_2|\mathcal{B}_2^- \vee \mathcal{B}_0)$. Our aim is to show that this leads to a contradiction by showing that $H_\eta(\mathcal{P}_3|\mathcal{B}_3^- \vee \mathcal{B}_0) > H_\eta(\mathcal{P}_1|\mathcal{B}_1^- \vee \mathcal{B}_0)$. By definition,

$$\begin{aligned} H_\eta(\mathcal{P}_i|\mathcal{B}_i^- \vee \mathcal{B}_0) &= \int -\sum_j (\mathbf{1}_{[j]} \circ \pi_i) \log \mathbb{E}(\mathbf{1}_{[j]} \circ \pi_i|\mathcal{B}_i^- \vee B_0)\, d\eta \\ &= \int -\sum_j \mathbb{E}(\mathbf{1}_{[j]} \circ \pi_i|\mathcal{B}_i^- \vee B_0) \log \mathbb{E}(\mathbf{1}_{[j]} \circ \pi_i|\mathcal{B}_i^- \vee B_0)\, d\eta \\ &= \int \sum_j \psi\left(\mathbb{E}(\mathbf{1}_{[j]} \circ \pi_i|\mathcal{B}_i^- \vee B_0)\right)\, d\eta, \end{aligned} \tag{34}$$

where ψ is the strictly concave function $[0,1] \to [0,1]$, $\psi(x) = -x\log x$ (with $\psi(0)$ defined to be 0).

The following claim is an essential point of the argument. We shall show that

$$\begin{gathered} \mathbb{E}_\eta(\mathbf{1}_{[j]} \circ \pi_3|\mathcal{B}_1^- \vee \mathcal{B}_2^- \vee \mathcal{B}_3^- \vee \mathcal{B}_0)(z) = \\ \mathbb{E}_\eta(\mathbf{1}_{[j]} \circ \pi_1|\mathcal{B}_1^- \vee \mathcal{B}_0) \quad \text{if } \pi_3(z)_{-1} = \pi_1(z)_{-1} \neq \pi_2(z)_{-1}; \\ \mathbb{E}_\eta(\mathbf{1}_{[j]} \circ \pi_2|\mathcal{B}_2^- \vee \mathcal{B}_0) \quad \pi_3(z)_{-1} = \pi_2(z)_{-1} \neq \pi_1(z)_{-1}; \\ \frac{1}{2}\mathbb{E}_\eta(\mathbf{1}_{[j]} \circ \pi_1|\mathcal{B}_1^- \vee \mathcal{B}_0) + \frac{1}{2}\mathbb{E}_\eta(\mathbf{1}_{[j]} \circ \pi_2|\mathcal{B}_2^- \vee \mathcal{B}_0) \quad \pi_3(z)_{-1} = \pi_1(z)_{-1} = \pi_2(z)_{-1}. \end{gathered} \tag{35}$$

Clearly, the right-hand side of the equation is measurable with respect to $\mathcal{B}_1^- \vee \mathcal{B}_2^- \vee \mathcal{B}_3^- \vee \mathcal{B}_0$. To verify the claim, it will be sufficient to integrate the right-hand side over the elements of a generating semi-algebra of $\mathcal{B}_1^- \vee \mathcal{B}_2^- \vee \mathcal{B}_3^- \vee \mathcal{B}_0$. Specifically, we will integrate over sets of the form $A \cap B \cap C \cap D$, where A, B and C are the preimages under the respective maps of cylinder sets in X of a common length (ending at time -1) and $D \in \mathcal{B}_0$.

Suppose A, B, and C are cylinders depending on the coordinates $-n$ to -1 of $\pi_1(z)$, $\pi_2(z)$, and $\pi_3(z)$ and that $A \cap B \cap C$ has positive measure. Then for $z \in A \cap B \cap C$, $\pi_3(z)_{-1}$ is equal to either $\pi_1(z)_{-1}$ or $\pi_2(z)_{-1}$ (or both) by definition of π_3. Further, $\pi_1(z)_{-1}$, $\pi_2(z)_{-1}$, and $\pi_3(z)_{-1}$ are constant over the intersection in question.

If on $A \cap B \cap C$, $\pi_3(z)_{-1} = \pi_1(z)_{-1} \neq \pi_2(z)_{-1}$, then we calculate

$$\begin{aligned} &\int_{A\cap B\cap C\cap D} \mathbb{E}_\eta(\mathbf{1}_{[j]} \circ \pi_1 | \mathcal{B}_1^- \vee \mathcal{B}_0)(z)\, d\eta = \\ &\int \mathbf{1}_B \mathbf{1}_C \mathbb{E}_\eta(\mathbf{1}_{[j]} \circ \pi_1 \mathbf{1}_A \mathbf{1}_D | \mathcal{B}_1^- \vee \mathcal{B}_0)\, d(\hat{\mu} \times \beta). \end{aligned} \tag{36}$$

Performing first the integration over R with respect to the measure β, we see that the only factor depending on the random part $r \in R$ is 1_C, the others being functions only of $(u, v) \in X^2$. The coordinates of $\pi_3(z)$ from $-n$ to -1 are concatenations of blocks of $\pi_1(z)$ and $\pi_2(z)$, the choice (between a block in u and a *different* block in v) being made according to the entries in r, hence with probabilities $1/2, 1/2$). If $k = k_{A,B}(u, v) = 1 + \operatorname{card}\{j : -n \leq j \leq -2, u_j = v_j, u_{j+1} \neq v_{j+1}\}$, then

$$\int_R \mathbf{1}_C(u, v, r)\, d\beta(r) = \frac{1}{2^{k_{A,B}(u,v)}}, \tag{37}$$

which is constant on $A \cap B$. The following calculation will be more readable if we write $\mathbb{E}^{\mathcal{B}} f$ for $\mathbb{E}(f|\mathcal{B})$. Since $B \in \mathcal{B}_2$ and $\mathcal{B}_2 \perp_{\mathcal{B}_0} \mathcal{B}_1^-$, we have $\mathbb{E}^{\mathcal{B}_1^- \vee \mathcal{B}_0}\mathbf{1}_B = \mathbb{E}^{\mathcal{B}_0}\mathbf{1}_B$. Consequently,

$$
\begin{aligned}
&\int_{A\cap B\cap C\cap D} \mathbb{E}_\eta(\mathbf{1}_{[j]} \circ \pi_1 | \mathcal{B}_1^- \vee \mathcal{B}_0)(z)\, d\eta \\
&= \int_{X^2} 2^{-k}\mathbf{1}_D\mathbf{1}_B\mathbf{1}_A\mathbb{E}^{\mathcal{B}_1^- \vee \mathcal{B}_0}(\mathbf{1}_{[j]} \circ \pi_1)\, d\hat{\mu}(u,v) \\
&= \int_{X^2} 2^{-k}\mathbf{1}_B\mathbb{E}^{\mathcal{B}_1^- \vee \mathcal{B}_0}(\mathbf{1}_D\mathbf{1}_A \cdot (\mathbf{1}_{[j]} \circ \pi_1))\, d\hat{\mu}(u,v) \\
&= \int_{X^2} 2^{-k}\mathbb{E}^{\mathcal{B}_1^- \vee \mathcal{B}_0}[\mathbf{1}_B\mathbb{E}^{\mathcal{B}_1^- \vee \mathcal{B}_0}(\mathbf{1}_D\mathbf{1}_A \cdot (\mathbf{1}_{[j]} \circ \pi_1))]\, d\hat{\mu}(u,v) \\
&= \int_{X^2} 2^{-k}[\mathbb{E}^{\mathcal{B}_1^- \vee \mathcal{B}_0}\mathbf{1}_B][\mathbb{E}^{\mathcal{B}_1^- \vee \mathcal{B}_0}(\mathbf{1}_D\mathbf{1}_A \cdot (\mathbf{1}_{[j]} \circ \pi_1))]\, d\hat{\mu}(u,v) \\
&= \int_{X^2} 2^{-k}[\mathbb{E}^{\mathcal{B}_0}\mathbf{1}_B][\mathbb{E}^{\mathcal{B}_1^- \vee \mathcal{B}_0}(\mathbf{1}_D\mathbf{1}_A \cdot (\mathbf{1}_{[j]} \circ \pi_1))]\, d\hat{\mu}(u,v) \\
&= \int_{X^2} 2^{-k}[\mathbb{E}^{\mathcal{B}_0}(\mathbf{1}_B\mathbf{1}_D)][\mathbb{E}^{\mathcal{B}_1^- \vee \mathcal{B}_0}(\mathbf{1}_A \cdot (\mathbf{1}_{[j]} \circ \pi_1))]\, d\hat{\mu}(u,v) \\
&= \int_{X^2} 2^{-k}\mathbb{E}^{\mathcal{B}_0}\{[\mathbb{E}^{\mathcal{B}_0}(\mathbf{1}_B\mathbf{1}_D)][\mathbb{E}^{\mathcal{B}_1^- \vee \mathcal{B}_0}(\mathbf{1}_A \cdot (\mathbf{1}_{[j]} \circ \pi_1))]\}\, d\hat{\mu}(u,v) \\
&= \int_{X^2} 2^{-k}[\mathbb{E}^{\mathcal{B}_0}(\mathbf{1}_B\mathbf{1}_D)][(\mathbb{E}^{\mathcal{B}_0}(\mathbf{1}_A \cdot (\mathbf{1}_{[j]} \circ \pi_1))]\, d\hat{\mu}(u,v) \\
&= \eta(A\cap B\cap C\cap D\cap \pi_1^{-1}[j]) = \eta(A\cap B\cap C\cap D\cap \pi_3^{-1}[j]),
\end{aligned}
\tag{38}
$$

by (31), since $B, D \in \mathcal{B}_2$ and $A, \pi_1^{-1}[j] \in \mathcal{B}_1$. This demonstrates the desired equality in the case $\pi_3(z)_{-1} = \pi_1(z)_{-1} \neq \pi_2(z)_{-1}$. The case $\pi_3(z)_{-1} = \pi_2(z)_{-1} \neq \pi_1(z)_{-1}$ is dealt with similarly.

If $\pi_3(z)_{-1} = \pi_1(z)_{-1} = \pi_2(z)_{-1}$, then the integrand is the average of the two previous integrands, so we see that

$$
\begin{aligned}
&\int_{A\cap B\cap C\cap D} \left(\tfrac{1}{2}\mathbb{E}_\eta(\mathbf{1}_{[j]} \circ \pi_1 | \mathcal{B}_1^- \vee \mathcal{B}_0) + \tfrac{1}{2}\mathbb{E}_\eta(\mathbf{1}_{[j]} \circ \pi_2 | \mathcal{B}_2^- \vee \mathcal{B}_0)\right)\, d\eta = \\
&\tfrac{1}{2}\eta(A\cap B\cap C\cap D\cap \pi_1^{-1}[j]) + \tfrac{1}{2}\eta(A\cap B\cap C\cap D\cap \pi_2^{-1}[j]) = \\
&\eta(A\cap B\cap C\cap D\cap \pi_3^{-1}[j]).
\end{aligned}
\tag{39}
$$

This completes the proof of equation (35).

Using (34), we have

$$H_\eta(\mathcal{P}_3|\mathcal{B}_3^- \vee \mathcal{B}_0) \geq H_\eta(\mathcal{P}_3|\mathcal{B}_1^- \vee \mathcal{B}_2^- \vee \mathcal{B}_3^- \vee \mathcal{B}_0) = \int \sum_j \psi(\mathbb{E}_\eta(\mathbf{1}_{[j]} \circ \pi_3|\mathcal{B}_1^- \vee \mathcal{B}_2^- \vee \mathcal{B}_3^- \vee \mathcal{B}_0))\, d\eta. \tag{40}$$

We separate the integral into parts according to whether $\pi_3(z)_{-1}$ is equal to $\pi_1(z)_{-1}$, $\pi_2(z)_{-1}$ or both. Let $S_1 = \{z\colon \pi_3(z)_{-1} = \pi_1(z)_{-1} \neq \pi_2(z)_{-1}\}$, $S_2 = \{z\colon \pi_3(z)_{-1} = \pi_2(z)_{-1} \neq \pi_1(z)_{-1}\}$ and $S_3 = \{z\colon \pi_3(z)_{-1} = \pi_1(z)_{-1} = \pi_2(z)_{-1}\}$. Let $A = \{z\colon \pi_1(z)_{-1} \neq \pi_2(z)_{-1}\}$ so that $A = S_1 \cup S_2$. Note that S_1 and S_2 have equal measure by definition of π_3.

By symmetry,

$$\int_{S_1} \sum_j \psi(\mathbb{E}_\eta(\mathbf{1}_{[j]} \circ \pi_1|\mathcal{B}_1^- \vee \mathcal{B}_0)\, d\eta = \int_{S_2} \sum_j \psi(\mathbb{E}_\eta(\mathbf{1}_{[j]} \circ \pi_1|\mathcal{B}_1^- \vee \mathcal{B}_0)\, d\eta, \tag{41}$$

so by (35),

$$\int_{S_1} \sum_j \psi(\mathbb{E}_\eta(\mathbf{1}_{[j]} \circ \pi_3|\mathcal{B}_1^- \vee \mathcal{B}_2^- \vee \mathcal{B}_3^- \vee \mathcal{B}_0))\, d\eta = \tfrac{1}{2} \int_A \textstyle\sum_j \psi(\mathbb{E}_\eta(\mathbf{1}_{[j]} \circ \pi_1|\mathcal{B}_1^- \vee \mathcal{B}_0)\, d\eta. \tag{42}$$

Similarly,

$$\int_{S_2} \sum_j \psi(\mathbb{E}_\eta(\mathbf{1}_{[j]} \circ \pi_3|\mathcal{B}_1^- \vee \mathcal{B}_2^- \vee \mathcal{B}_3^- \vee \mathcal{B}_0))\, d\eta = \tfrac{1}{2} \int_A \textstyle\sum_j \psi(\mathbb{E}_\eta(\mathbf{1}_{[j]} \circ \pi_2|\mathcal{B}_2^- \vee \mathcal{B}_0)\, d\eta. \tag{43}$$

Finally, integrating over S_3,

$$\begin{aligned} &\int_{S_3} \sum_j \psi(\mathbb{E}_\eta(\mathbf{1}_{[j]} \circ \pi_3|\mathcal{B}_1^- \vee \mathcal{B}_2^- \vee \mathcal{B}_3^- \vee \mathcal{B}_0))\, d\eta = \\ &\int_{A^c} \sum_j \psi\left(\tfrac{1}{2}(\mathbb{E}_\eta(\mathbf{1}_{[j]} \circ \pi_1|\mathcal{B}_1^- \vee \mathcal{B}_0) + \mathbb{E}_\eta(\mathbf{1}_{[j]} \circ \pi_2|\mathcal{B}_2^- \vee \mathcal{B}_0))\right)\, d\eta > \\ &\tfrac{1}{2} \int_{A_c} \textstyle\sum_j \left(\psi(\mathbb{E}_\eta(\mathbf{1}_{[j]} \circ \pi_1|\mathcal{B}_1^- \vee \mathcal{B}_0) + \psi(\mathbb{E}_\eta(\mathbf{1}_{[j]} \circ \pi_2|\mathcal{B}_2^- \vee \mathcal{B}_0)\right)\, d\eta. \end{aligned} \tag{44}$$

The strict inequality in the above arises since ψ is strictly concave and there exist a j in the alphabet of X and a set of points of positive measure in $A^c = \{(u, v, r) \in Z = X^2 \times R : u_{-1} = v_{-1}\}$ for which $\mathbb{E}_\eta(\mathbf{1}_{[j]} \circ \pi_1|\mathcal{B}_1^- \vee \mathcal{B}_0) \neq \mathbb{E}_\eta(\mathbf{1}_{[j]} \circ \pi_2|\mathcal{B}_2^- \vee \mathcal{B}_0)$— for, if not, Lemma 3 would imply that $\mu_1 = \mu_2$.

Now adding the preceding equalities, we see

$$H_\eta(\mathcal{P}_3|\mathcal{B}_3^- \vee \mathcal{B}_0) > \tag{45}$$
$$\frac{1}{2}\left(\int \sum_j \psi(\mathbb{E}_\eta(\mathbf{1}_{[j]} \circ \pi_1|\mathcal{B}_1^- \vee \mathcal{B}_0)\, d\eta + \int \sum_j \psi(\mathbb{E}_\eta(\mathbf{1}_{[j]} \circ \pi_2|\mathcal{B}_2^- \vee \mathcal{B}_0)\, d\eta\right)$$
$$= \tfrac{1}{2}(H(\mathcal{P}_1|\mathcal{B}_1^- \vee \mathcal{B}_0) + H(\mathcal{P}_2|\mathcal{B}_2^- \vee \mathcal{B}_0))$$
$$= h_{\mu_1}(X) - h_\nu(Y).$$

From (33), we see that $h_{\mu_3}(X) > h_{\mu_1}(X)$ as required. □

Remark. It would be desirable to have a proof of this result based on the Shannon-McMillan-Breiman Theorem, but so far we have not been able to construct one.

Definition. Let $(X, \mathcal{B}, \mu, T)$ and $(Y, \mathcal{C}, \nu, S)$ be measure-preserving systems, $\pi : X \to Y$ a factor map, and α a finite generating partition for X. We say that μ is *relatively Markov for* α *over* Y if it satisfies one of the following two equivalent conditions:

(1) $\alpha \perp_{T^{-1}\alpha \vee \pi^{-1}\mathcal{C}} \alpha_2^\infty$;
(2) $H_\mu(\alpha|\alpha_1^\infty \vee \pi^{-1}\mathcal{C}) = H_\mu(\alpha|T^{-1}\alpha \vee \pi^{-1}\mathcal{C})$.

(As usual, $\alpha_i^j = \bigvee_{k=i}^j T^{-k}\alpha$.)

Corollary 3. *If X is a 1-step SFT, Y is a subshift, $\pi : X \to Y$ is a 1-block factor map, ν is an ergodic measure on Y, and μ is an ergodic relatively maximal measure over ν, then μ is relatively Markov for the time-0 partition of X over Y.*

Proof. We apply the first half of the proof of Lemma 3 with $\mu_1 = \mu_2 = \mu$. Note that then $\hat{\mu}(S) > 0$. If $s_{-1}g_i^{(1)} = s_{-1}g_i^{(2)}$ for all symbols i in the alphabet of X, the proof proceeds as before to show that the information function with respect to μ of the time-0 partition $\mathcal{P}$ of X given $\mathcal{P}_1^\infty \vee \pi^{-1}\mathcal{B}_Y$ is measurable with respect to $\mathcal{P} \vee \sigma^{-1}\mathcal{P} \vee \pi^{-1}\mathcal{B}_Y$, and hence μ is a 1-step relatively Markov measure.

If there is a symbol i in the alphabet of X for which $s_{-1}g_i^{(1)} \neq s_{-1}g_i^{(2)}$, then the construction in the proof of Theorem 1, by interleaving strings according to another random process, will again produce a measure projecting to ν which will have entropy greater than $h(\mu)$. □

3. Examples

Example 1. In case π has a singleton clump a and ν is Markov on Y, we can construct the unique relatively maximal measure above ν explicitly. Denote the cylinder sets $[a]$ in X and in Y by X_a and Y_a, respectively. If ν is (1-step) Markov on Y, then the first-return map $\sigma_a : Y_a \to Y_a$ is countable-state Bernoulli with respect to the restricted and normalized measure $\nu_a = \nu/\nu[a]$: the states are all the loops or return blocks aC^i with $aC^i a = ac_1^i \dots c_{r_i}^i a$ appearing in Y and no $c_j^i = a$.

Under π^{-1}, the return blocks to $[a]$ expand into bands $aB^{i,j}$, with $aB^{i,j}a$ appearing in X and $\pi B^{i,j} = C^i$ for all i, j. Topologically, (X_a, σ_a) is a countable-state full shift on these symbols $aB^{i,j}$. We define μ_a to be the countable-state Bernoulli measure on (X_a, σ_a) which equidistributes the measure of each loop (state) of Y_a over its preimage band:

$$\mu_a[aB^{i,j}] = \frac{\nu_a[aC^i a]}{|\pi^{-1}[aC^i a]|} \quad \text{for all } i, j. \tag{46}$$

We show now that this choice of μ_a is relatively maximal over ν_a. Let λ_a be any probability measure on X_a which maps under π to ν_a. Then the countable-state Bernoulli measure on X_a which agrees with λ_a on all the 1-blocks $aB^{i,j}$ (its "Bernoullization") has entropy no less than that of λ_a and still projects to the Bernoulli measure ν_a, so we may as well assume that λ_a is countable-state Bernoulli. If $\lambda_a[aB^{i,j}] = q^{i,j}$ and $|\pi^{-1}(aC^i a)| = J_i$ for all i, j, then

$$h(X_a, \sigma_a, \lambda_a) = \sum_{i=1}^{\infty} \sum_{j=1}^{J_i} q^{i,j} \log q^{i,j}. \tag{47}$$

Note that for each i

$$\sum_{j=1}^{J_i} q^{i,j} = \nu_a[aC^i a] \tag{48}$$

is fixed at the same value for all λ_a. Thus for each i,

$$\sum_{j=1}^{J_i} q^{i,j} \log q^{i,j} \tag{49}$$

is maximized by putting all the $q^{i,j}$ equal to one another.

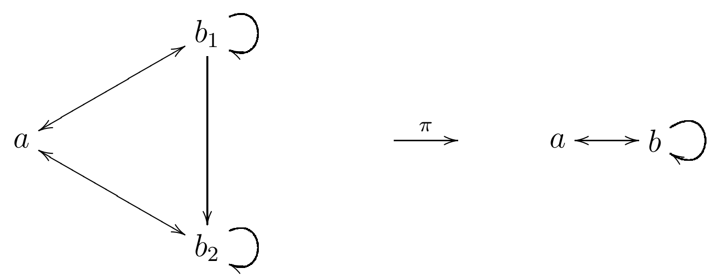
Finally, this unique relatively maximal μ_a over ν_a determines the unique relatively maximal μ on X over ν on Y, since according to

Abramov's formula

$$h(X,\sigma,\mu) = \mu[a]\, h(X_a,\sigma_a,\mu_a), \tag{50}$$

and $\mu[a] = \nu[a]$.

We show how this calculation of the unique relatively maximal measure over a Markov measure in the case of a singleton clump works out in a particular case. It was shown in [18, 19] that for the following factor map there is a saturated compensation function $G \circ \pi$ with $G \in \mathcal{C}(Y)$ but no such compensation function with $G \in \mathcal{F}(Y)$. There is a singleton clump, a.



For each $k \geq 1$ the block $ab^k a$ in Y has $k+1$ preimages, depending on when the subscript on b switches from 1 to 2. Let ν be Markov on Y. To each preimage $aB_1aB_2a\dots aB_r$ of $ab^{k_1}ab^{k_2}\dots ab^{k_r}$ the optimal measure μ_a assigns measure

$$\mu_a[aB_1aB_2a\dots aB_r] = \frac{1}{k_1+1}\cdots\frac{1}{k_r+1}\nu_a[ab^{k_1}ab^{k_2}\dots ab^{k_r}]. \tag{51}$$

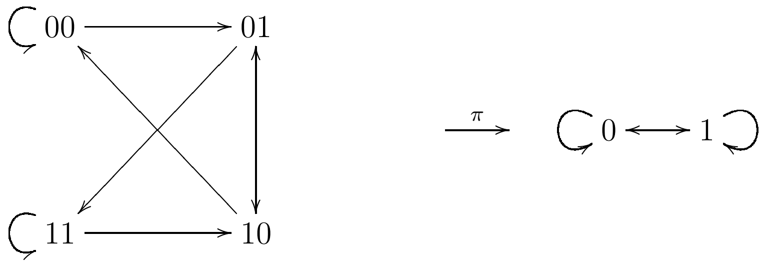
The unique relatively maximal measure over ν_a can be described in terms of fiber measures as follows. Given $y = ab^{k_1}ab^{k_2}\dots ab^{k_r}\dots \in Y_a$, $\mu_{a,y}$ chooses the preimages of each b^{k_i} with equal probabilities and independently of the choice of preimage of any other b^{k_j}. Then

$$\mu_a[aB_1aB_2a\dots aB_r] = \int_{Y_a} \mu_{a,y}[aB_1aB_2a\dots aB_r]\, d\nu_a(y). \tag{52}$$

Example 2. The relatively maximal measures over an ergodic measure ν on Y which is supported on the orbit $\mathcal{O}(y)$ of a periodic point $y = CCC\dots \in Y$ can be found by analyzing the SFT $X_y = \pi^{-1}\mathcal{O}(y)$. The relatively maximal measures over ν are determined by the maximal (Shannon-Parry) measures on the irreducible components of X_y. Consequently, if X_y is irreducible, then the discrete invariant measure on the orbit of y is π-determinate.

Example 3. *Failure of π-determinacy for a fully-supported measure.* In the preceding example, along with others discussed in [14], failure of π-determinism can be blamed on lack of communication among fibers. An example suggested by Walters (see [20]) also shows that there can be *fully supported* ν on Y which are not π-determinate. For such examples there are potential functions $V \in \mathcal{C}(Y)$ such that $V \circ \pi$ has two equilibrium states which project to the *same* ergodic measure on Y.

In this example, $X = Y = \Sigma_2 =$ full 2-shift, and $\pi(x)_0 = x_0 + x_1$ mod 2 is a simple cellular automaton 2-block map. If we replace X by its 2-block recoding, so that π becomes a 1-block map, we obtain the following diagram:



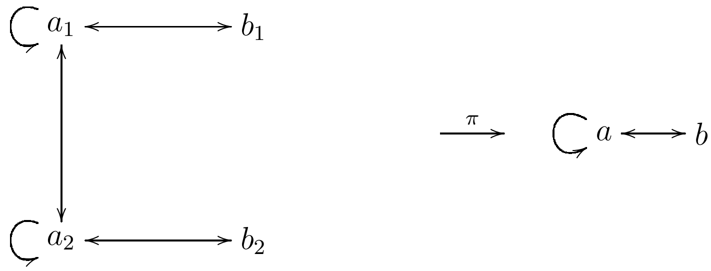
This is a finite-to-one map and hence is Markovian—for example, the Bernoulli $1/2, 1/2$ measure on Σ_2 is mapped to itself. The constant function 0 is a compensation function. Thus every Markov measure on Y is π-determinate: the equilibrium state μ_V of a locally constant V on Y lifts to the equilibrium state of $V \circ \pi$, which is the unique relatively maximal measure over μ_V (in fact it's the only measure in $\pi^{-1}\{\mu_V\}$).

For every ergodic ν on Y, *all* of $\pi^{-1}\{\nu\}$ consists of relatively maximal measures over ν, all of them having the same entropy as ν.

If $p \neq 1/2$, the two measures on the SFT X that correspond to the Bernoulli measures $\mathcal{B}(p, 1-p)$ and $\mathcal{B}(1-p, p)$ both map to the same measure ν_p on Y. Thus ν_p, which is fully supported on Y, is *not* π-determinate. (An entropy-decreasing example is easily produced by forming the Cartesian product of X with another SFT.)

Moreover, ν_p is the unique equilibrium state of some continuous function V_p on Y [15]. Then the set of relatively maximal measures over ν_p, which is the entire set $\pi^{-1}\{\nu_p\}$, consists of the equilibrium states of $V_p \circ \pi + G \circ \pi = V_p \circ \pi$ [20], so this potential function $V_p \circ \pi$ has many equilibrium states.

Example 4. *Homogeneous clumps.* In the following example there is no singleton clump, but the clumps are homogeneous with respect to π so there is a locally constant compensation function (see [2, 18, 19]), and hence every Markov measure on Y is π-determinate and its unique relatively maximal lift is Markov.



In this case the return time to $[a]$ is bounded, so X_a is a finite-state SFT rather than the countable-state chain of the general case. There are six states, a_1a_1, $a_1b_1a_1$, a_1a_2, a_2a_2, $a_2b_2a_2$, and a_2a_1, according to the time 0 entries of $x \in X_a$ and $\sigma_a x$. Fix this order of the states for indexing purposes. It can be shown by direct calculation that for this example a stochastic matrix P determines a Markov measure on X_a that is relatively maximal over its image if and only if it is of the form

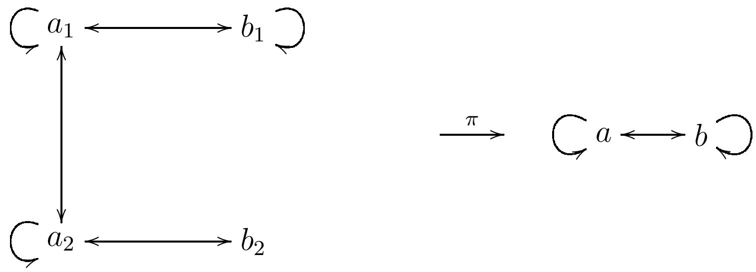
$$\begin{pmatrix} x & 1-2x & x & 0 & 0 & 0 \\ y & 1-2y & y & 0 & 0 & 0 \\ 0 & 0 & 0 & x & 1-2x & x \\ 0 & 0 & 0 & x & 1-2x & x \\ 0 & 0 & 0 & y & 1-2y & y \\ x & 1-2x & x & 0 & 0 & 0 \end{pmatrix}. \tag{53}$$

(In this case the image measure is also Markov.)
Here $0 < x, y < 1/2$ and the probability vector fixed by P is

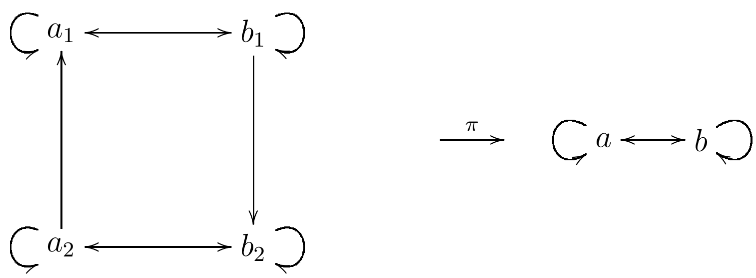
$$p = \frac{1}{4y + 2(1-2x)}(y, 1-2x, y, y, 1-2x, y). \tag{54}$$

Further, given a (1-step) Markov measure ν on Y, put $K = \nu[aa]/\nu[aba]$. Then a stochastic matrix of the form (53) with fixed vector p satisfies $p_1+p_3+p_4+p_6 = \nu[aa]$ and $p_2+p_5 = \nu[aba]$ (so that the Markov measure μ that it determines projects to ν) if and only if $x = y = K/(2K+2)$ (and then μ is relatively maximal over ν).

Example 5. *Singleton clump after recoding.* Make the preceding example a little bit more complicated by adding a loop at b_1, so that now the return time to $[a]$ is unbounded. It can be verified that now there is still a continuous saturated compensation function, but there is no locally constant compensation function, so the code is not Markovian. However, if we look at higher block presentations of X and Y, we can find singleton clumps, for example *abba*. Therefore again every Markov measure on Y is π-determinate.



Example 6. *No singleton clumps.* Complicating Example 5 a bit more, we can produce a situation in which there are no singleton clumps, not even for any higher block presentation.



For this example it can be shown that there is a continuous saturated compensation function $G \circ \pi$, but we do not know exactly which measures are π-determinate. Although the example appears simple, the question of how many fibers allow how much switching is complex.

Acknowledgment. An exposition of some of these results was included in [14]. We thank Mike Boyle, Xavier Méla, Jean-Paul Thouvenot, and Peter Walters for their very helpful questions, suggestions, and discussions and the referee for comments that improved the writing.

References

1. D. Blackwell, *The entropy of functions of finite state Markov chains*, Transactions First Prague Conference, 1957, pp. 13–20.

2. M. Boyle and S. Tuncel, *Infinite-to-one codes and Markov measures*, Trans. Amer. Math. Soc. **285** (1984), 657–683.
3. C. J. Burke and M. Rosenblatt, *A Markovian function of a Markov chain*, Ann.Math. Stat. **29** (1958), 1112–1122.
4. H. Furstenberg, Y. Peres, and B. Weiss, *Perfect filtering and double disjointness*, Ann. Inst. H. Poincaré Prob. Stat. **31** (1995), 453–465.
5. D. Gatzouras and Y. Peres, *The variational principle for Hausdorff dimension: A survey*, Ergodic Theory of $\mathbb{Z}^d$-Actions (M. Pollicott and K. Schmidt, eds.), LMS Lecture Notes, vol. 228, Cambridge Univ.Press, 1996, pp. 533–568.
6. R. B. Israel, *Convexity in the Theory of Lattice Gases*, Princeton University Press, 1979.
7. G. Keller, *Equilibrium States in Ergodic Theory*, Cambridge Univ. Press, 1998.
8. F. Ledrappier and P. Walters, *A relativised variational principle for continuous transformations*, J. London Math. Soc. **16** (1977), 568–576.
9. F. Ledrappier and L.-S. Young, *The metric entropy of diffeomorphisms II*, Ann. Math. **122** (1985), 540–574.
10. M. Loève, *Probability Theory*, fourth ed., vol. II, Springer-Verlag, 1963, 1978.
11. B. Marcus, K. Petersen, and S. Williams, *Transmission rates and factors of Markov chains*, Contemp. Math. **26** (1984), 279–293.
12. W. Parry, *Intrinsic markov chains*, Trans. Amer. Math. Soc. **112** (1964), 55–66.
13. ———, *Entropy and Generators in Ergodic Theoy*, W. A. Benjamin, Inc., 1969.
14. K. Petersen, *Information compression and retention in dynamical processes*, Preprint, notes of lecture course at Workshop on Dynamics and Randomness, University of Chile, 2000.
15. R. R. Phelps, *Unique equilibrium states*, Proceedings of the Workshop on Dynamics and Randomness, Santiago, Chile, December 2000, 2001, to appear.
16. D. Ruelle, *Statistical mechanics on a compact set with F action satisfying expansiveness and specification*, Trans. Amer. Math. Soc. **185** (1973), 237–251.
17. C. Shannon, *A mathematical theory of communication*, Bell System Tech. J. **27** (1948), 379–423, 623–656.
18. S. Shin, *Measures that maximize weighted entropy for factor maps between subshifts of finite type*, Ph.D. thesis, Univ. of N. Carolina, 1999.
19. ———, *Measures that maximize weighted entropy for factor maps between subshifts of finite type*, Erg. Th. Dyn. Sys. **21** (2001), 1249–1272.
20. P. Walters, *Relative pressure, relative equilibrium states, compensation functions and many-to-one codes between subshifts*, Trans. Amer. Math. Soc. **296** (1986), 1–31.

Department of Mathematics, CB 3250, Phillips Hall, University of North Carolina, Chapel Hill, NC 27599 USA

E-mail address: petersen@math.unc.edu

Department of Mathematical Sciences, University of Memphis, Memphis, TN 38152-6429

E-mail address: quasa@msci.memphis.edu

Department of Mathematics and Statistics, University of Victoria, Victoria, BC V8W 3P4, Canada. Department of Mathematics, Ajou University, Suwon 422-749, South Korea

Current address: Department of Mathematics, Korea Advanced Institute of Science and Technology, Daejon, 305-701, South Korea

E-mail address: sjs@math.kaist.ac.kr