*Ergod. Th. & Dynam. Sys.* (2011), **31**, 1889–1899 © *Cambridge University Press, 2010*
doi:10.1017/S0143385710000581

# Measures of maximal relative entropy with full support

JISANG YOO

*Department of Mathematical Sciences,*
*Korea Advanced Institute of Science and Technology, Daejeon 305-701, Korea*
*(e-mail: jisangy@kaist.ac.kr)*

(*Received* 28 *April* 2010 *and accepted in revised form* 16 *August* 2010)

*Abstract*. Let $\pi$ be a factor map from an irreducible shift of finite type $X$ to a shift space $Y$. Let $\nu$ be an invariant probability measure on $Y$ with full support. We show that every measure on $X$ of maximal relative entropy over $\nu$ is fully supported. As a result, given any invariant probability measure $\nu$ on $Y$ with full support, there is an invariant probability measure $\mu$ on $X$ with full support that maps to $\nu$ under $\pi$. If $\nu$ is ergodic, $\mu$ can be chosen to be ergodic. These results can be generalized to the case of sofic shifts. We demonstrate that the results do not extend to general shift spaces by providing counterexamples.

## 1. *Introduction*

Throughout the work, every measure is assumed to be a Borel probability measure.

It is well known in symbolic dynamics [**3**] that every irreducible shift of finite type (SFT) has a unique measure of maximal entropy called the Parry measure. If $\mu$ is the Parry measure of the SFT $\mathsf{X}_A$ defined by an irreducible 0-1 matrix $A$, then the measure of each word $u = u_1 \cdots u_n$ is given by

$$\mu(u) = A(u_1, u_2) \cdots A(u_{n-1}, u_n)\ell(u_1)r(u_n)/\lambda^{n-1},$$

where $\ell$ and $r$ are the left and right eigenvectors of $A$, respectively, corresponding to the largest eigenvalue $\lambda$ and satisfying $\ell \cdot r = 1$. The measure $\mu$ is a fully supported Markov measure on $\mathsf{X}_A$.

When studying factor maps between shift spaces, it is natural to be interested in measures of maximal relative entropy, which are relative versions of maximal measures. But they are not as much understood as maximal measures are.

*Definition 1.1.* Let $\pi : X \to Y$ be a factor map between shift spaces. Let $\nu$ be an invariant measure on $Y$. An invariant measure $\mu$ on $X$ is said to be *relatively maximal* over $\nu$ or to be a *measure of maximal relative entropy* over $\nu$ if $\pi(\mu) = \nu$ and $h(\mu) \geq h(\mu')$ for all $\mu' \in \pi^{-1}(\nu)$†.

† The relative entropy of $\mu$ is given by $h_\mu(X|Y) = h(\mu) - h(\nu)$.

https://doi.org/10.1017/S0143385710000581 Published online by Cambridge University Press

Such measures always exist, but in general they need not be unique, unlike the case of maximal measures, as an example can be found in [**6**].

A relatively maximal measure is a special case of a relative equilibrium state. For a factor map $\pi : X \to Y$ between topological dynamical systems and a potential function $V \in \mathcal{C}(X)$, the *relative pressure* $P(\pi, V) : Y \to \mathbb{R}$ is defined, which is the relative version of the pressure [**2**]. The relative variational principle [**2**] states that for each invariant measure $\nu$ on $Y$,

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

where $h_\mu(X|Y)$ denotes the relative entropy of $\mu$ under $\pi : X \to Y$. Any measure $\mu$ attaining the supremum is called a *relative equilibrium state*. The relative equilibrium states of the potential function $V = 0$ are exactly the relatively maximal measures. Walters [**6**] used the concept of compensation functions to further study properties of relative equilibrium states. It turns out that the concept of relative equilibrium states is useful in studying lifts of measures under a factor map on shift spaces. Similarly, as we will show later, relatively maximal measures are useful when we lift a fully supported ergodic measure to another such measure in one of our main theorems.

It was shown that given a factor map from an irreducible SFT $X$ to a shift space $Y$, for any ergodic measure on $Y$, there is an explicit bound on the number of ergodic measures on $X$ with maximal relative entropy that project to $\nu$ [**5**]. In the same paper, it was also shown that such ergodic measures with maximal relative entropy are relatively Markov in some sense, which is a relative version of the fact that the maximal measure on an irreducible SFT is always a Markov measure. Our main result is the relative version of the fact that maximal measures on irreducible SFTs are fully supported.

*Definition 1.2.* A measure $\mu$ on a compact metric space $X$ is said to have *full support* or to be *fully supported* if its support is the entire space $X$ or, equivalently, $\mu(U) > 0$ for every non-empty open subset $U \subseteq X$†.

THEOREM 1.1. *Let $X$ be an irreducible SFT, $Y$ a shift space, $\pi : X \to Y$ a factor map, and $\nu$ a fully supported invariant measure on $Y$. Then any measure of maximal relative entropy among the measures in $\pi^{-1}(\nu)$ is fully supported.*

We prove this result by bringing additional randomness to a measure $\mu$ without changing its image $\nu$, resulting in another measure $\hat{\mu}$ with an increased entropy. We show that this is possible if $\mu$ is not fully supported.

We take a randomly chosen point $x \in X$ (such that the probability distribution for this random point is $\mu$). Then we replace some of the appearances of a specific word $u$ with $\mu(u) > 0$ in $x$ with another specific word $v$ with $\mu(v) = 0$. We randomly determine which appearances will be replaced and therefore bring additional entropy. The probability distribution of the resulting random point is the new measure $\hat{\mu}$. In order to show that the new measure $\hat{\mu}$ indeed has higher entropy, we appeal to Pinsker's formula, which allows us to compare the entropies of the new and old measures. It is worth noting that the authors

† Fully supported measures are the same as strictly positive measures.

https://doi.org/10.1017/S0143385710000581 Published online by Cambridge University Press

of [**5**] also use interleaving to construct a measure with higher entropy in order to prove their result on the number of relatively maximal measures.

Our result gives a positive answer to part of the following problem: is it true that for every factor map $\pi : X \to Y$ between irreducible SFTs, every fully supported Markov measure on $Y$ has a unique relatively maximal measure that maps to it? Is this measure also fully supported? (See [**1**].)

We also show the following result on lifting a fully supported ergodic measure.

THEOREM 1.2. *Let $X$ be an irreducible SFT, $Y$ a shift space, and $\pi : X \to Y$ a factor map. Then for any fully supported invariant measure $\nu$ on $Y$, there is a fully supported invariant measure $\mu$ on $X$ that projects to $\nu$. If $\nu$ is ergodic, $\mu$ can be chosen to be ergodic.*

Conclusions of Theorems 1.1 and 1.2 also hold when $X$ is only an irreducible sofic shift (see §4).

For more on relatively maximal measures, see [**1**, §3]. For results on relative equilibrium states and their implications, consult [**6**].

## 2. *Conditional entropy over $\sigma$-algebras and subsets*

Throughout the section, let $(\Omega, \mathcal{F}, m)$ denote a probability space and $\mathcal{B}$ be a sub-$\sigma$-algebra of the $\sigma$-algebra $\mathcal{F}$. Every measurable partition $\mathcal{Q}$ of $\Omega$ is assumed to satisfy $m(Q) > 0$ for all $Q \in \mathcal{Q}$. Let $\mathcal{P}$ be a measurable partition of $\Omega$. Let $A \in \mathcal{F}$ and $m(A) > 0$. Denote the restriction of $\mathcal{F}$ to the set $A$ by

$$\mathcal{F}_A = \{F \cap A | F \in \mathcal{F}\};$$

similarly for $\mathcal{B}_A$ and $\mathcal{P}_A$.

Define the measure $m_A$ on $A$ by

$$m_A(B) = \frac{m(B)}{m(A)}$$

for all $B \in \mathcal{F}_A$. Then $m_A(B)$ is the *conditional probability* of the event $B$ given the event $A$. We denote the expectation (the conditional expectation, respectively) with respect to $m$ by $E(\cdot)$ (by $E(\cdot|\cdot)$, respectively) and the one with respect to $m_A$ by $E_A(\cdot)$ (by $E_A(\cdot|\cdot)$, respectively).

The *conditional information function* of the partition $\mathcal{P}$ over $\mathcal{B}$ is defined to be

$$I(\mathcal{P}|\mathcal{B}) = -\sum_{P \in \mathcal{P}} 1_P \log E(1_P|\mathcal{B}).$$

It is a $\mathcal{P} \vee \mathcal{B}$-measurable function on $\Omega$. The *conditional entropy* of $\mathcal{P}$ over $\mathcal{B}$ is defined to be

$$H(\mathcal{P}|\mathcal{B}) = E(I(\mathcal{P}|\mathcal{B})).$$

The *conditional information function* of $\mathcal{P}$ over $\mathcal{B}$ and $A$ is defined to be

$$I_A(\mathcal{P}|\mathcal{B}) = -\sum_{P \in \mathcal{P}} 1_{P \cap A} \log E_A(1_{P \cap A}|\mathcal{B}_A).$$

This is the conditional information function of $\mathcal{P}_A$ over $\mathcal{B}_A$ (with respect to $m_A$). The *conditional entropy* of $\mathcal{P}$ over $\mathcal{B}$ and $A$ is defined to be

$$H_A(\mathcal{P}|\mathcal{B}) = E_A(I_A(\mathcal{P}|\mathcal{B})).$$

It is the conditional entropy of $\mathcal{P}_A$ over $\mathcal{B}_A$ (with respect to $m_A$).

https://doi.org/10.1017/S0143385710000581 Published online by Cambridge University Press

LEMMA 2.1. *Let $f$ be a measurable function on $\Omega$ and $\mathcal{Q}$ be a measurable partition of $\Omega$. Then*

$$E(f|\mathcal{B}\vee\mathcal{Q})=\sum_{Q\in\mathcal{Q}}1_Q E_Q(f|\mathcal{B}_Q).$$

*Proof.* Put $g=\sum_{Q\in\mathcal{Q}}1_Q E_Q(f|\mathcal{B}_Q)$. It is easy to check that $g$ is $\mathcal{B}\vee\mathcal{Q}$-measurable. Let $B\in\mathcal{B}$ and $Q\in\mathcal{Q}$. Then

$$\begin{aligned}E(1_{B\cap Q}\cdot g)&=E(1_{B\cap Q}\cdot E_Q(f|\mathcal{B}_Q))\\&=E(1_Q\cdot E_Q(1_{B\cap Q}\cdot f|\mathcal{B}_Q))\\&=P(Q)E_Q(E_Q(1_{B\cap Q}\cdot f|\mathcal{B}_Q))\\&=P(Q)E_Q(1_{B\cap Q}\cdot f)\\&=E(1_{B\cap Q}\cdot f).\end{aligned}$$

By summing up, we have

$$E(1_D\cdot g)=E(1_D\cdot f)$$

for all $D\in\mathcal{B}\vee\mathcal{Q}$. The result follows. □

LEMMA 2.2. *Let $\mathcal{P}$ and $\mathcal{Q}$ be measurable partitions of $\Omega$. Then the following hold.*

(1) $I(\mathcal{P}|\mathcal{B}\vee\mathcal{Q})=\sum_{Q\in\mathcal{Q}}1_Q I_Q(\mathcal{P}|\mathcal{B})$.

(2) $H(\mathcal{P}|\mathcal{B}\vee\mathcal{Q})=\sum_{Q\in\mathcal{Q}}m(Q)H_Q(\mathcal{P}|\mathcal{B})$.

*Proof.* It follows from Lemma 2.1 that

$$\begin{aligned}I(\mathcal{P}|\mathcal{B}\vee\mathcal{Q})&=-\sum_{P\in\mathcal{P}}1_P\sum_{Q\in\mathcal{Q}}1_Q\log E_Q(1_{P\cap Q}|\mathcal{B}_Q)\\&=-\sum_{Q\in\mathcal{Q}}1_Q\sum_{P\in\mathcal{P}}1_{P\cap Q}\log E_Q(1_{P\cap Q}|\mathcal{B}_Q)\\&=\sum_{Q\in\mathcal{Q}}1_Q I_Q(\mathcal{P}|\mathcal{B}),\end{aligned}$$

which proves (1). Integrating both sides of (1) verifies (2). □

The following corollary lets us localize an entropy to a subset $A$.

COROLLARY 2.1. *Let $\mathcal{P}$ be a measurable partition of $\Omega$. Let $A\in\mathcal{F}$ and $m(A)>0$. Then*

$$H(\mathcal{P}|\mathcal{B})\geq m(A)H_A(\mathcal{P}|\mathcal{B}).$$

*Proof.* We may assume that $m(A)<1$. Let $\mathcal{Q}$ be the measurable partition $\{A,\ \Omega\backslash A\}$ of $\Omega$. It follows from Lemma 2.2 that

$$\begin{aligned}H(\mathcal{P}|\mathcal{B})&\geq H(\mathcal{P}|\mathcal{B}\vee\mathcal{Q})=\sum_{Q\in\mathcal{Q}}m(Q)H_Q(\mathcal{P}|\mathcal{B})\\&\geq m(A)H_A(\mathcal{P}|\mathcal{B}),\end{aligned}$$

which completes the proof. □

https://doi.org/10.1017/S0143385710000581 Published online by Cambridge University Press

LEMMA 2.3. *Let $\mathcal{P}$ be a measurable partition of $\Omega$ and $\mathcal{C}$ a sub-$\sigma$-algebra of $\mathcal{F}$ such that $\mathcal{C}$ and $\mathcal{B} \vee \mathcal{P}$ are independent. Let $C \in \mathcal{C}$ and $m(C) > 0$. Then the following hold.*
(1) *$I_C(\mathcal{P}|\mathcal{B} \vee \mathcal{C})(\omega) = I(\mathcal{P}|\mathcal{B})(\omega)$ for almost every $\omega \in C$.*
(2) *$H_C(\mathcal{P}|\mathcal{B} \vee \mathcal{C}) = H(\mathcal{P}|\mathcal{B})$.*

*Proof.* (1) Fix $\omega \in C$ and let $f = 1_P$ for some $P \in \mathcal{P}$. We may assume that $m(C) < 1$. Let $\mathcal{Q}$ be the measurable partition $\{C, \Omega \backslash C\}$ of $\Omega$. By Lemma 2.1,

$$\begin{aligned} E_C(f|(\mathcal{B} \vee \mathcal{C})_C)(\omega) &= E(f|(\mathcal{B} \vee \mathcal{C}) \vee \mathcal{Q})(\omega) \\ &= E(f|\mathcal{B} \vee \mathcal{C})(\omega) = E(f|\mathcal{B})(\omega). \end{aligned}$$

The last equality holds, since $\mathcal{C}$ and $\mathcal{B} \vee \mathcal{P}$ are independent. Thus,

$$\begin{aligned} I_C(\mathcal{P}|\mathcal{B} \vee \mathcal{C})(\omega) &= -\sum_{P \in \mathcal{P}} 1_{P \cap C}(\omega) \log E_C(1_{P \cap C}|(\mathcal{B} \vee \mathcal{C})_C)(\omega) \\ &= -\sum_{P \in \mathcal{P}} 1_P(\omega) \log E(1_P|\mathcal{B})(\omega) = I(\mathcal{P}|\mathcal{B})(\omega), \end{aligned}$$

which completes the proof.

(2) Integrating both sides of (1) over $C$ with respect to $m_C$, we have

$$\begin{aligned} H_C(\mathcal{P}|\mathcal{B} \vee \mathcal{C}) &= E_C(I(\mathcal{P}|\mathcal{B})) \\ &= \frac{1}{m(C)} E(1_C \cdot I(\mathcal{P}|\mathcal{B})) \\ &= \frac{1}{m(C)} E(1_C) E(I(\mathcal{P}|\mathcal{B})) \\ &= H(\mathcal{P}|\mathcal{B}), \end{aligned}$$

which completes the proof. □

## 3. *Proofs of the main theorems*

In this section we prove Theorems 1.1 and 1.2.

For a shift space $Z$, let $\mathcal{L}(Z)$ be the language of $Z$. Let $\mathsf{X}_G$ denote the edge shift presented by a graph $G$. If $G$ is irreducible, then $\mathcal{L}(\mathsf{X}_G)$ is exactly the set of paths in $G$. For a 1-block factor map $\pi$ from $\mathsf{X}_G$ to a shift space $Y$, we use the same notation $\pi$ to denote the 1-block map from $\mathcal{L}(\mathsf{X}_G)$ to $\mathcal{L}(Y)$ which induces $\pi$.

*Definition 3.1.* Two paths $u = u_1 \cdots u_n$ and $v = v_1 \cdots v_n$ in a graph $G$ are said to be *mutually separated* if $u_i$ and $v_i$ do not share the same initial vertex for all $i = 1, \ldots, n$ and $u_n$ and $v_n$ do not share the same terminal vertex.

*Definition 3.2.* A word $w = w_1 \cdots w_n$ in a shift space is said to be *unbordered* if there is no $i$ with $1 \le i < n$ such that $w_1 \cdots w_i = w_{n-i+1} \cdots w_n$.

LEMMA 3.1. *Let $\mathsf{X}_G$ be an irreducible edge shift with $h_{\mathrm{top}}(\mathsf{X}_G) > 0$. Then it contains arbitrarily long unbordered words.*

*Proof.* Since $h_{\mathrm{top}}(\mathsf{X}_G) > 0$, there exist two distinct paths $u$ and $v$ in $G$ such that $u$ and $v$ both start and end at a vertex $I$ but do not pass through $I$ in between. Given $n \in \mathbb{N}$, the path $uv^n$ is an unbordered word of $\mathsf{X}_G$. □

https://doi.org/10.1017/S0143385710000581 Published online by Cambridge University Press

PROPOSITION 3.1. *Let $\pi$ be a 1-block factor map from an irreducible edge shift $\mathsf{X}_G$ to a shift space $Y$ and $h_{\mathrm{top}}(\mathsf{X}_G) > 0$. Let $Z$ be a proper shift space of $\mathsf{X}_G$ such that $\pi(Z) = Y$. Then there exist two paths $u$ and $v$ in $G$ with the following properties:*

(i) $\pi(u) = \pi(v)$;

(ii) *$u$ and $v$ share the same initial vertex and the same terminal vertex;*

(iii) $u \in \mathcal{L}(Z)$ *and* $v \notin \mathcal{L}(Z)$;

(iv) *for any paths $\gamma$, $\delta$ in $G$ with $\gamma u \delta \in \mathcal{L}(Z)$, the path $v$ occurs exactly once in the path $\gamma v \delta$ in $G$.*

*A pair of paths $u$ and $v$ in $G$ are said to be* $Z$-related *if the paths satisfy the above conditions (i)–(iv).*

*Proof.* We may assume that $G$ is irreducible. For each $n \in \mathbb{N}$, we say that the condition $P(n)$ holds if the following is true:

> *Given a path $\omega$ in $G$, there exist paths $u^{(1)}, \ldots, u^{(n)}$ in $\mathcal{L}(Z)$ such that $u^{(1)}, \ldots, u^{(n)}$ and $\omega$ are mutually separated and have the same label, i.e.,* $\pi(u^{(1)}) = \cdots = \pi(u^{(n)}) = \pi(\sigma)$

Obviously, $P(n)$ does not hold for $n \geq |\mathcal{V}(G)|$, where $\mathcal{V}(G)$ is the set of vertices in $G$. Suppose that there is no $Z$-related pair in $G$. We will inductively show that $P(n)$ holds for all $n \in \mathbb{N}$, which would complete the proof. Choose a path $\xi$ in $G$ that is not a word of $Z$.

*Step 1.* $P(1)$ holds.

Let $\omega$ be a path in $G$. Since $G$ is irreducible, there exist paths $\kappa_1$, $\kappa_4$ in $G$ such that $\omega\kappa_1\xi$ and $\xi\kappa_4\omega$ are paths in $G$. By Lemma 3.1, there is an unbordered word $\alpha$ in $\mathsf{X}_G$ with

$$|\alpha| > \max\{|\omega\kappa_1\xi|, |\xi\kappa_4\omega|\}.$$

Similarly, there exist paths $\kappa_2$, $\kappa_3$ in $G$ such that

$$\bar{v} = \omega\kappa_1\xi\kappa_2\alpha\kappa_3\xi\kappa_4\omega$$

is a path in $G$. Since $\pi(Z) = Y$, there is a path $\bar{u}$ in $\mathcal{L}(Z)$ with $\pi(\bar{u}) = \pi(\bar{v})$. Let $|\omega| = q$ and $|\bar{v}| = |\bar{u}| = p$. Write $\bar{v} = \bar{v}_1 \cdots \bar{v}_p$ and $\bar{u} = \bar{u}_1 \cdots \bar{u}_p$. Note that $p - q > q$. If $\bar{u}_i$ and $\bar{v}_i$ have the same initial vertex for some $i$ with $1 \leq i \leq q + 1$ and if $\bar{u}_j$ and $\bar{v}_j$ have the same terminal vertex for some $j$ with $p - q \leq j \leq p$, then the subpaths $u = \bar{u}_i \cdots \bar{u}_j$ and $v = \bar{v}_i \cdots \bar{v}_j$ in $G$ are $Z$-related, which is a contradiction. In fact, it is clear that $u$ and $v$ satisfy the properties (i) and (ii). The property (iii) holds, since $\xi \notin \mathcal{L}(Z)$ and so $v \notin \mathcal{L}(Z)$. The property (iv) holds, since $\xi \notin \mathcal{L}(Z)$ and $\alpha$ is unbordered. Hence, either the subpaths $\bar{u}_1 \cdots \bar{u}_q$ and $\bar{v}_1 \cdots \bar{v}_q$ $(= \omega)$ are mutually separated, or the subpaths $\bar{u}_{p-q+1} \cdots \bar{u}_p$ and $\bar{v}_{p-q+1} \cdots \bar{v}_p$ $(= \omega)$ are mutually separated. Set $u^{(1)} = \bar{u}_1 \cdots \bar{u}_q$ in the former case and $u^{(1)} = \bar{u}_{p-q+1} \cdots \bar{u}_p$ in the latter case. It follows that $u^{(1)}$ and $\omega$ are mutually separated and have the same label. Thus, the condition $P(1)$ holds.

*Step 2.* $P(n)$ implies $P(n + 1)$.

Assume that $P(n)$ holds. Let $\omega$ be a path in $G$. Since $G$ is irreducible, there exist paths $\kappa_1$, $\kappa_2$ in $G$ such that $\xi\kappa_1\omega\kappa_2\xi$ is a path in $G$. Choose an unbordered word $\alpha$ in $\mathsf{X}_G$ with

$$|\alpha| > \max\{|\omega\kappa_2\xi|, |\xi\kappa_1\omega|\}.$$

https://doi.org/10.1017/S0143385710000581 Published online by Cambridge University Press

Similarly, there exist paths $\kappa_3$, $\kappa_4$ in $G$ such that

$$\theta = \xi\kappa_1\omega\kappa_2\xi\kappa_3\alpha\kappa_4$$

is a path in $G$ and $\bar{v} = \theta^{n+1}$ is also a path in $G$. Since the condition $P(1)$ holds, there is a path $\bar{u}^{(0)}$ in $\mathcal{L}(Z)$ such that $\bar{u}^{(0)}$ and $\bar{v}$ are mutually separated and have the same label.

Also, since the condition $P(n)$ holds for the path $\bar{u}^{(0)}$, there exist paths $\bar{u}^{(1)}, \ldots, \bar{u}^{(n)}$ in $\mathcal{L}(Z)$ such that $\bar{u}^{(1)}, \ldots, \bar{u}^{(n)}$ and $\bar{u}^{(0)}$ are mutually separated and have the same label. Note that the $n+2$ paths $\bar{u}^{(0)}$, $\bar{u}^{(1)}, \ldots, \bar{u}^{(n)}$ and $\bar{v}$ have the same label. Let $|\omega| = q$, $|\theta| = p$ and $|\xi\kappa_1| = r$. Write $\bar{v} = \bar{v}_1 \cdots \bar{v}_{p(n+1)}$ and $\bar{u}^{(k)} = \bar{u}^{(k)}_1 \cdots \bar{u}^{(k)}_{p(n+1)}$ for $k = 0, 1, \ldots, n$. For each $i = 0, 1, \ldots, n$, let

$$I_i^+ = \{s \in \mathbb{N} \mid ip + r < s \leq ip + r + q + 1\},$$
$$I_i^- = \{s \in \mathbb{N} \mid ip + r \leq s \leq ip + r + q\},$$

and $I_i = I_i^+ \cap I_i^-$. Set $\mathcal{I} = \{I_i \mid 0 \leq i \leq n\}$. For each $I \in \mathcal{I}$, let

$$\bar{v}_I = \bar{v}_{ip+r+1} \cdots \bar{v}_{ip+r+q} \quad \text{and} \quad \bar{u}^{(k)}_I = \bar{u}^{(k)}_{ip+r+1} \cdots \bar{u}^{(k)}_{ip+r+q}$$

for $k = 0, 1, \ldots, n$, where $I = I_i$ with $0 \leq i \leq n$. Note that $\bar{v}_I = \omega$ for all $I \in \mathcal{I}$.

Suppose that there exist $1 \leq k \leq n$ and $i, j$ with $0 \leq i < j \leq n$ such that $\bar{u}^{(k)}_s$ and $\bar{v}_s$ have the same initial vertex for some $s \in I_i^+$ and such that $\bar{u}^{(k)}_t$ and $\bar{v}_t$ have the same terminal vertex for some $t \in I_j^-$. Then the subpaths $u = \bar{u}^{(k)}_s \cdots \bar{u}^{(k)}_t$ and $v = \bar{v}_s \cdots \bar{v}_t$ are $Z$-related, which is a contradiction. In fact, the properties (i) and (ii) clearly hold. The property (iii) holds, since $\xi \notin \mathcal{L}(Z)$ and so $v \notin \mathcal{L}(Z)$. The property (iv) holds, since $\xi \notin \mathcal{L}(Z)$ and $\alpha$ is unbordered. So, for each $k = 1, \ldots, n$, there is at most one interval in $\mathcal{I}$, say $I_{r_k}$, $0 \leq r_k \leq n$, such that $\bar{u}^{(k)}_s$ and $\bar{v}_s$ have the same initial vertex for some $s \in I^+_{r_k}$. Choose any interval $I \in \mathcal{I} \backslash \{I_{r_k} \mid 1 \leq k \leq n\}$. Then $\bar{u}^{(1)}_I, \ldots, \bar{u}^{(n)}_I$ and $\bar{v}_I$ are mutually separated. Set $u^{(k)} = \bar{u}^{(k)}_I$ for $k = 0, 1, \ldots, n$. Then $u^{(0)}, u^{(1)}, \ldots, u^{(n)}$ and $\omega$ are mutually separated and have the same label. Thus, $P(n+1)$ holds. □

In order to prove Theorem 1.1, we introduce some notation. Let $(\Omega, \mathcal{F}, m)$ be a probability space. Let $\hat{x} : \Omega \to M$ be a random variable with a finite range, i.e. $\hat{x}(\Omega)$ is finite, and $\hat{y} : \Omega \to N$ be a random variable where $M$ and $N$ are measurable spaces. Define

$$H(\hat{x}|\hat{y}) = H(\mathcal{P}|\mathcal{B}) \tag{1}$$

and

$$H_A(\hat{x}|\hat{y}) = H_A(\mathcal{P}|\mathcal{B}),$$

where $\mathcal{P}$ is the measurable partition of $\Omega$ generated by $\hat{x}$, $\mathcal{B}$ is the sub-$\sigma$-algebra of $\mathcal{F}$ generated by $\hat{y}$, and $A$ is a measurable subset of $\Omega$. We also use Pinsker's formula stated as follows: let $T$ be a measure-preserving transformation on $(\Omega, \mathcal{F}, m)$. Let $\alpha$ and $\beta$ be two measurable partitions of $\Omega$ such that $\beta$ is coarser than $\alpha$. Then

$$H\left(\alpha \,\middle|\, \bigvee_{i=1}^{\infty} T^i\alpha\right) = H\left(\beta \,\middle|\, \bigvee_{i=1}^{\infty} T^i\beta\right) + H\left(\alpha \,\middle|\, \left(\bigvee_{i=1}^{\infty} T^i\alpha\right) \bigvee \left(\bigvee_{i=-\infty}^{\infty} T^i\beta\right)\right) \tag{2}$$

(see [**4**, p. 67, Theorem 6.3], applied to $\beta$ coarser than $\alpha$).

https://doi.org/10.1017/S0143385710000581 Published online by Cambridge University Press

*Proof of Theorem 1.1.* We may assume that $X$ is an irreducible edge shift $\mathsf{X}_G$ and $\pi$ is a 1-block factor map. We also assume that $h_{\text{top}}(X) > 0$. Otherwise, $X$ consists of a single cycle and the only invariant measure on $X$ is fully supported.

Let $\mu$ be a measure on $X$ of maximal relative entropy over $\nu$ and the support $Z$ of $\mu$ be a proper subset of $X$. Proposition 3.1 guarantees the existence of a $Z$-related pair of paths $u, v$ in $G$. Put $p = |u| = |v|$. Let $S$ be the shift space of $\{0, 1\}^{\mathbb{Z}}$ that consists of all bi-infinite concatenations of the words $10^{2p}$ and $10^{2p+1}$. Then $S$ is a mixing SFT with $h_{\text{top}}(S) > 0$. Denote by $\eta$ the unique measure on $S$ of maximal entropy, i.e. $h(\eta) = h_{\text{top}}(S)$. Using $\eta$, we will construct an invariant measure $\bar{\mu}$ on $X$ such that $\pi(\bar{\mu}) = \nu$ and $h(\bar{\mu}) > h(\mu)$, which would contradict that $\mu$ is a measure of maximal relative entropy over $\nu$.

Define a function $F$ from $X \times S$ to $X$ as follows: let $(x, s) \in X \times S$ and $N = \{k \in \mathbb{Z} \mid s_k = 1\}$. For each $k \in N$, replace the $k$th occurrence of $u$ in $x$ with $v$ if such $u$ exists. Here the 0th occurrence of $u$ in $x$ is defined to be the first occurrence of $u$ in $x_{[0,\infty)}$. The resulting point is denoted $F(x, s)$. Two occurrences of $u$ in $x$ may overlap. Meanwhile, any two consecutive occurrences of 1 in $s$ are separated by either $0^{2p}$ or $0^{2p+1}$ and $|u| < 2|p|$. Thus, $F(x, s)$ is well defined. Also, it is a point in $X$, since $u$ and $v$ have the same initial vertex and the same terminal vertex. It is not difficult to see that $F$ is continuous and hence is measurable on $X \times S$.

Define

$$\bar{\mu} = F(\mu \times \eta) = (\mu \times \eta) \circ F^{-1}.$$

Since $\pi(u) = \pi(v)$, it follows that $\pi(F(x, s)) = \pi(x)$ for all $(x, s) \in X \times S$, so

$$\pi(\bar{\mu}) = \pi(F(\mu \times \eta)) = \pi(\mu) = \nu.$$

Thus, $\bar{\mu}$ is a measure on $X$ that projects to $\nu$.

To see that $\bar{\mu}$ is $\sigma$-invariant, define a function $K$ from $X \times S$ to $S$ by

$$K(x, s) = \begin{cases} s & \text{if } x \notin [u], \\ \sigma(s) & \text{if } x \in [u], \end{cases}$$

where $[u] = {}_0[u]_{p-1} = \{z \in X \mid z_0 \cdots z_{p-1} = u\}$. For each $(x, s) \in X \times S$, let

$$T(x, s) = (\sigma(x), K(x, s)).$$

Then $T$ is a continuous transformation on $X \times S$. It is easy to see that $\sigma \circ F = F \circ T$.

Consider the $X$-valued random variable $\hat{x}$ on $(X \times S, \mu \times \eta)$ defined by $\hat{x}(x, s) = x$ and the $S$-valued random variable $\hat{s}$ on $(X \times S, \mu \times \eta)$ defined by $\hat{s}(x, s) = s$. It is easy to check that $\hat{x}$ and $\hat{s}$ are independent and that their probability distributions are $\mu$ and $\eta$, respectively. Also, the probability distribution of the $(X \times S)$-valued random variable $(\hat{x}, K(\hat{x}, \hat{s}))$ on $(X \times S, \mu \times \eta)$ is $\mu \times \eta$.

In fact, if $D$ and $E$ are measurable subsets of $X$ and $S$, respectively, then

$$\begin{aligned} &\Pr(\hat{x} \in D, K(\hat{x}, \hat{s}) \in E) \\ &\quad = \Pr(\hat{x} \in D \backslash [u], K(\hat{x}, \hat{s}) \in E) + \Pr(\hat{x} \in D \cap [u], K(\hat{x}, \hat{s}) \in E) \\ &\quad = \Pr(\hat{x} \in D \backslash [u], \hat{s} \in E) + \Pr(\hat{x} \in D \cap [u], \sigma(\hat{s}) \in E) \\ &\quad = \mu(D \backslash [u]) \cdot \eta(E) + \mu(D \cap [u]) \cdot \eta(E) \\ &\quad = \mu(D) \cdot \eta(E). \end{aligned}$$

https://doi.org/10.1017/S0143385710000581 Published online by Cambridge University Press

The second last equality is valid because $\hat{x}$ and $\hat{s}$ are independent and because $\eta$ is $\sigma$-invariant. We showed that the probability distribution of $T(\hat{x}, \hat{s})$ is $\mu \times \eta$ and therefore we have $T(\mu \times \eta) = \mu \times \eta$. Since $\sigma \circ F = F \circ T$, it follows that $\bar{\mu}$ is an invariant measure on $X$. Thus, $F$ is a measure-preserving homomorphism from $(X \times S, T, \mu \times \eta)$ to $(X, \sigma, \bar{\mu})$.

Finally, we will show that $h(\bar{\mu}) > h(\mu)$. Using (1), we have

$$\begin{aligned} h(\bar{\mu}) &= H(F(\hat{x}, \hat{s})_0 | F(\hat{x}, \hat{s})_{(-\infty,-1]}), \\ h(\mu) &= H(\hat{x}_0 | \hat{x}_{(-\infty,-1]}), \\ h(\eta) &= H(\hat{s}_0 | \hat{s}_{(-\infty,-1]}). \end{aligned}$$

Due to the property (iv) in Proposition 3.1, given any $x \in Z$ and $s \in S$, the only occurrences of $v$ in $F(x, s)$ are those obtained from replacing some occurrences of $u$ in $x$ with $v$, since any two occurrences of 1 in $s$ are separated by at least $2p$ zeros. So, by replacing every instance of $v$ in $F(x, s)$ with $u$, one recovers $x$. For each $k \in \mathbb{Z}$, the symbol $x_k$ is determined by $F(x, s)_{[k-p+1,k+p-1]}$ for all $(x, s) \in Z \times S$, and hence for $(\mu \times \eta)$-almost every $(x, s) \in X \times S$. Let $\alpha$ be the partition of $X \times S$ generated by $F(\hat{x}, \hat{s})_{[-p+1,p-1]}$ and $\beta$ be the partition of $X \times S$ generated by $\hat{x}_0$. Applying equation (2), we have

$$\begin{aligned} h(\bar{\mu}) &= H(F(\hat{x}, \hat{s})_{[-p+1,p-1]} | F(\hat{x}, \hat{s})_{(-\infty,p-2]}) \\ &= H(\hat{x}_0 | \hat{x}_{(-\infty,-1]}) + H(F(\hat{x}, \hat{s})_{[-p+1,p-1]} | (F(\hat{x}, \hat{s})_{(-\infty,p-2]}, \hat{x})) \\ &= h(\mu) + H(F(\hat{x}, \hat{s})_0 | (F(\hat{x}, \hat{s})_{(-\infty,-1]}, \hat{x})) \\ &= h(\mu) + \widehat{H}, \end{aligned}$$

where

$$\widehat{H} = H(F(\hat{x}, \hat{s})_0 | (F(\hat{x}, \hat{s})_{(-\infty,-1]}, \hat{x})).$$

Note that $\widehat{H}$ can be estimated as

$$\begin{aligned} \widehat{H} &= \frac{1}{p} \cdot H(F(\hat{x}, \hat{s})_{[0,p-1]} | (F(\hat{x}, \hat{s})_{(-\infty,-1]}, \hat{x})) \\ &\geq \frac{\mu([u])}{p} \cdot H_{[u]\times S}(F(\hat{x}, \hat{s})_{[0,p-1]} | (F(\hat{x}, \hat{s})_{(-\infty,-1]}, \hat{x})). \end{aligned} \tag{3}$$

The inequality follows from Corollary 2.1.

If $x \in [u]$, then $F(x, s)_{[0,p-1]}$ is either $u$ or $v$ depending on the value of $s_0$. Hence, the last term in equation (3) is equal to

$$\widetilde{H} := \frac{\mu([u])}{p} \cdot H_{[u]\times S}(\hat{s}_0 | (F(\hat{x}, \hat{s})_{(-\infty,-1]}, \hat{x})).$$

If $x \in [u]$, then $x$ and $s_{(-\infty,-1]}$ determine $F(x, s)_{(-\infty,-1]}$. Thus,

$$\begin{aligned} \widetilde{H} &\geq \frac{\mu([u])}{p} \cdot H_{[u]\times S}(\hat{s}_0 | (\hat{s}_{(-\infty,-1]}, \hat{x})) \\ &= \frac{\mu([u])}{p} \cdot H(\hat{s}_0 | \hat{s}_{(-\infty,-1]}) \\ &= \frac{\mu([u])}{p} \cdot h(\eta) > 0. \end{aligned}$$

https://doi.org/10.1017/S0143385710000581 Published online by Cambridge University Press

The first equality is valid because of Lemma 2.3(2). So, $\widehat{H} \geq \widetilde{H} > 0$, which implies that $h(\bar{\mu}) > h(\mu)$, as desired. Consequently, $\bar{\mu}$ is an invariant measure on $X$ which maps to $\nu$ and satisfies $h(\bar{\mu}) > h(\mu)$. This is a contradiction. Thus, $\mu$ must be fully supported. This completes the proof. □

Now we are ready to prove Theorem 1.2. For a shift space $Z$, denote by $\mathcal{M}(Z)$ the set of invariant measures on $Z$.

*Proof of Theorem 1.2.* Let $\mu$ be an invariant measure on $X$ that is of maximal relative entropy over $\nu$. Then, by Theorem 1.1, this measure is fully supported. Now suppose that $\nu$ is ergodic. If $\mu$ has the ergodic decomposition $\tau$, i.e. $\mu = \int_{\mathcal{M}(X)} \rho \, d\tau(\rho)$, then

$$\nu = \pi\mu = \int_{\mathcal{M}(X)} \pi\rho \, d\tau(\rho).$$

Since $\nu$ is ergodic, $\pi\rho = \nu$ for $\tau$-almost every $\rho \in \mathcal{M}(X)$. Hence, $h(\rho) \leq h(\mu)$ for $\tau$-almost every $\rho \in \mathcal{M}(X)$. Since $h(\mu) = \int_{\mathcal{M}(X)} h(\rho) \, d\tau(\rho)$, it follows that $h(\rho) = h(\mu)$ for $\tau$-almost every $\rho \in \mathcal{M}(X)$. Consequently, there is an ergodic measure on $X$ that maps to $\nu$ and is of maximal relative entropy over $\nu$. By Theorem 1.1, this measure is fully supported. This completes the proof. □

## 4. *Sofic case*

The two main theorems extend to the case where $X$ is only assumed to be an irreducible sofic shift. We only need to extend the first main theorem. Let $\pi$ be a factor map from an irreducible *sofic* shift $X$ to a shift space $Y$. Let $\nu$ be an invariant measure on $Y$ with full support. Let $\mu$ be an invariant measure on $X$ of maximal relative entropy over $\nu$. Since $X$ is sofic, we can take a right resolving factor map $\pi_R$ from an irreducible SFT $X_R$ to $X$. There exists an invariant measure $\mu_R$ on $X_R$ that projects to $\mu$ under $\pi_R$. If we can show that $\mu_R$ is a measure of maximal relative entropy over $\nu$ (with respect to $\pi \circ \pi_R : X_R \to Y$), then it follows from our main theorem that $\mu_R$ has full support and therefore so does its image $\mu$. To prove the maximality, let $\mu'_R$ be an invariant measure on $X_R$ that projects to $\nu$ under $\pi \circ \pi_R$ and let $\mu' = \pi_R(\mu'_R)$. By our assumption on $\mu$, we have $h(\mu) \geq h(\mu')$. Since $\pi_R$ is bounded-to-one, we also have $h(\mu_R) = h(\mu)$ and $h(\mu'_R) = h(\mu')$. Therefore, the desired inequality $h(\mu_R) \geq h(\mu'_R)$ holds.

We provide an example which shows that the conclusion of Theorem 1.1 need not be true if $X$ is not a sofic shift.

*Example 4.1.* Let $\{u^{(n)}\}$ be a sequence of words in $\{0, 1\}^{\mathbb{Z}}$ defined as follows: let $u^{(0)} = 1$ and, for each $n \geq 1$, let

$$u^{(n+1)} = u^{(n)} 0^{3^n} u^{(n)}.$$

Then $|u^{(n)}| = 3^n$ for all $n \in \mathbb{Z}^+$. Define a point $x = (x_i)_{i \in \mathbb{Z}}$ in $\{0, 1\}^{\mathbb{Z}}$, which is a discrete version of the Cantor set, as follows: put $x_i = 0$ for each $i < 0$. For $i > 0$, choose any $n \in \mathbb{N}$ with $3^n \geq i$ and put $x_i = u_i^{(n)}$. Let $X$ denote the orbit closure of $x$, that is, the closure of the set $\{\sigma^n x \mid n \in \mathbb{Z}\}$. Then $X$ is irreducible. Note that $|\mathcal{B}_{3^n}(X)| \leq 2|u^{(n)}| = 2 \cdot 3^n$ for all $n \in \mathbb{N}$. Hence, $h(X) = 0$. So, every invariant measure on $X$ has maximal entropy. In particular, the measure $\mu_\circ$ concentrated on the fixed point $0^\infty$ in $X$ is a measure of maximal

https://doi.org/10.1017/S0143385710000581 Published online by Cambridge University Press

entropy that is not fully supported. Consider the factor map $\pi$ from $X$ to $Y = \{0^\infty\}$ that sends every point in $X$ to $0^\infty$. If $\nu$ denotes the unique invariant measure on $Y$, then not every measure on $X$ of maximal relative entropy over $\nu$ is fully supported. Thus, Theorem 1.1 does not extend to all irreducible shift spaces.

It is not difficult to see that the frequencies of 1's in the blocks of the point $x$ go to 0 uniformly. More precisely, there is $C > 0$ such that the number of 1's in the block $x_{[i,i+3^n-1]}$ is less than $C \cdot 2^n$ for all $i \in \mathbb{Z}$ and $n \in \mathbb{N}$. Let $\mu$ be an ergodic measure on $X$. The point-wise ergodic theorem says that for $\mu$-almost every $z \in X$, the frequency of 1's in the block $z_{[0,n-1]}$ goes to $\mu(1)$ as $n \to \infty$. Hence, we must have $\mu(1) = 0$. By the ergodic decomposition, we conclude that the measure $\mu_\circ$ is the only invariant measure on $X$. This implies that the conclusion of Theorem 1.2 may not be valid if $X$ is not an SFT.

*Acknowledgement.* This work was supported by the second stage of the Brain Korea 21 Project, The Development of Human Resources in Mathematics, KAIST in 2010.

## REFERENCES

**[1]** M. Boyle and K. Petersen. Hidden Markov processes in the context of symbolic dynamics. *Preprint*, 2009; arXiv:0907.1858.
**[2]** F. Ledrappier and P. Walters. A relativised variational principle for continuous transformations. *J. London Math. Soc.* (2) **16**(3) (1977), 568–576.
**[3]** W. Parry. Intrinsic Markov chains. *Trans. Amer. Math. Soc.* **112** (1964), 55–66.
**[4]** W. Parry. *Entropy and Generators in Ergodic Theory*. W. A. Benjamin, New York, 1969.
**[5]** K. Petersen, A. Quas and S. Shin. Measures of maximal relative entropy. *Ergod. Th. & Dynam. Sys.* **23**(1) (2003), 207–223.
**[6]** P. Walters. Relative pressure, relative equilibrium states, compensation functions and many-to-one codes between subshifts. *Trans. Amer. Math. Soc.* **296**(1) (1986), 1–31.

https://doi.org/10.1017/S0143385710000581 Published online by Cambridge University Press