TRANSACTIONS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 358, Number 5, Pages 2205–2216
S 0002-9947(05)03943-7
Article electronically published on December 20, 2005

RELATIVE ENTROPY FUNCTIONS FOR FACTOR MAPS BETWEEN SUBSHIFTS

SUJIN SHIN

Abstract. Let (X, S) and (Y, T) be topological dynamical systems and $\pi : X \to Y$ a factor map. A function $F \in C(X)$ is a compensation function for π if $P(F + \phi \circ \pi) = P(\phi)$ for all $\phi \in C(Y)$. For a factor code between subshifts of finite type, we analyze the associated relative entropy function and give a necessary condition for the existence of saturated compensation functions. Necessary and sufficient conditions for a map to be a saturated compensation function will be provided.

1. Introduction

Let $S : X \to X$ and $T : Y \to Y$ be continuous maps of compact metrizable spaces and let $\pi : X \to Y$ be a *factor map*, i.e., a continuous surjection with $\pi \circ S = T \circ \pi$. For a given compact subset K of X, for $n \geq 1$ and $\delta > 0$, denote by $\Delta_{n,\delta}(K)$ the set of (n, δ)-separated sets of X contained in K. Let $f \in C(X)$. Fix $\delta > 0$ and $n \geq 1$. For each $y \in Y$, let

$$P_n(\pi, f, \delta)(y) = \sup \left\{ \sum_{x \in E} \exp \sum_{i=0}^{n-1} f(S^i x) \middle| E \in \Delta_{n,\delta}\big(\pi^{-1}\{y\}\big) \right\}.$$

Define $P(\pi, f) : Y \to \mathbb{R}$ by

$$P(\pi, f)(y) = \lim_{\delta \to 0} \limsup_{n \to \infty} \frac{1}{n} \ln P_n(\pi, f, \delta)(y)$$

which we call the *relative pressure function* corresponding to f. In particular, in the case $f \equiv 0$, it is called the *relative entropy function*.

In [4], Ledrappier and Walters introduced relative pressure and presented the *relative variational principle*, generalizing the concept of *relative (metric) entropy* (see also [1]). As the variational principle (for entropy) provides a description of the maximal entropy (or the topological entropy), the relative variational principle gives information on the *maximal relative entropy* (see §2).

Relative pressure functions are connected with the notion of compensation functions [11]. A continuous function $F \in C(X)$ is called a *compensation function* for

Received by the editors September 12, 2003 and, in revised form, June 30, 2004.
2000 *Mathematics Subject Classification.* Primary 37B10, 28D99; Secondary 28D20.
Key words and phrases. Subshift, factor code, relative pressure, relative entropy, compensation function.
This work was supported by grant No. R04-2002-000-00060-0 from the Basic Research Program of the Korea Science and Engineering Foundation.

π if the pressure functions satisfy

$$P_X(S, F + \phi \circ \pi) = P_Y(T, \phi) \qquad \text{for all } \phi \in C(Y).$$

Boyle and Tuncel first considered the concept of compensation functions that are related to how T-invariant Borel probability measures on Y lift, by π, to S-invariant Borel probability measures on X [2], [11]. A compensation function of type $G \circ \pi \in C(X)$ with $G \in C(Y)$ is said to be *saturated*.

The analysis of factor maps in terms of their saturated compensation functions has played a significant role in the study of measures of maximal relative entropy and measures of maximal weighted entropy [8]. Not only can saturated compensation functions for factor maps enable one to find such measures, but they can also characterize the factor maps along with the concepts such as *relative* entropy, *relative* pressure and *relative* equilibrium states [11].

There always exists a compensation function when (X, σ_X) and (Y, σ_Y) are subshifts of finite type [11]. There is, however, an example of a factor map between subshifts of finite type for which no saturated compensation function exists [9]. In this work we study relative entropy functions corresponding to factor codes between subshifts of finite type and provide a necessary condition for the existence of saturated compensation functions. Necessary and sufficient conditions for a map to be a saturated compensation function will be presented in terms of relative entropy functions (Proposition 2.4 and Theorem 3.5). Also some observation on resolving maps will be given.

We assemble some notations and fundamental definitions used. If (X, S) is a topological dynamical system, i.e., X is a compact metric space with a homeomorphism $S : X \to X$, then $M(X, S)$ will denote the set of all S-invariant Borel probability measures on X. If $\mu \in M(X, S)$, then $h_S(\mu)$ denotes the measure-theoretic entropy of S relative to μ. We will omit the letter S if the map is clear from the context. For $x \in X$, let

$$\mu_x = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \delta_{S^i x} \in M(X, S)$$

if it exists, in which case we call x a *generic point*. Denote by Δ_X the set of all generic points in X.

A subshift X is a shift-invariant closed subset of $\{1, 2, \cdots, k\}^{\mathbb{Z}}$ for some $k \geq 1$, and σ_X (or simply σ) denotes the shift map on X. A metric ρ on X is given by $\rho(x, y) = 1/2^k$ if $k = \min\{|i| | x_i \neq y_i, i \in \mathbb{Z}\} \geq 0$ and $\rho(x, y) = 0$ otherwise. For each $n \geq 1$, $\mathcal{B}_n(X)$ denotes the set of n-blocks in X and $\mathcal{B}(X) = \bigcup_{n\geq 1} \mathcal{B}_n(X)$. Given $b_1 \cdots b_n \in \mathcal{B}_n(X)$, $n \geq 1$, we define $[b_1 \cdots b_n]$ to be a cylinder set in X with b_1 on the 0-th coordinate. A subshift X is *of finite type* if there is a finite forbidden block system $\mathcal{F}$ for X, i.e., $X = \{x \in \{1, \cdots, k\}^{\mathbb{Z}} |$ no block $\omega \in \mathcal{F}$ occurs in $x\}$. Throughout this work a subshift of finite type is assumed to be 1-step, i.e., there is a forbidden block system that only contains blocks of length ≤ 2 (without loss of generality). Also a subshift is accompanied by the shift map to represent a topological dynamical system. For more details, see [3], [10] and [5].

2. Relative entropy and saturated compensation functions

Let (X, S) and (Y, T) be topological dynamical systems and $\pi : X \to Y$ a factor map. We also denote by π the naturally induced (onto) map (called the

extension of π) from $M(X)$ to $M(Y)$. For $\nu \in M(Y)$, let $M(\nu)$ denote the set of measures in $M(X)$ that project to ν. For $f \in C(X)$, the associated relative pressure function $P(\pi, f) : Y \to \mathbb{R}$ satisfies the *relative variational principle*, that is, for each $\nu \in M(Y)$,

$$\int P(\pi, f)\, d\nu = \sup \left\{ h(\mu) + \int f\, d\mu \middle| \mu \in M(\nu) \right\} - h(\nu).$$

Given $g \in C(Y)$, taking $f = g \circ \pi \in C(X)$, one can easily see that for $\nu \in M(Y)$,

$$\int P(\pi, g \circ \pi)\, d\nu = \sup_{\mu \in M(\nu)} h(\mu) - h(\nu) + \int g\, d\nu$$

and particularly, with $g \equiv 0$,

$$\int P(\pi, 0)\, d\nu = \sup_{\mu \in M(\nu)} h(\mu) - h(\nu).$$

Hence, given $\nu \in M(Y)$, by integrating the relative entropy function with respect to ν, one can evaluate the *maximal relative entropy* for ν, i.e., $\sup_{\mu \in M(\nu)} h(\mu)$. Furthermore, investigating the relative entropy function may help one to classify the *measures of maximal relative (weighted, respectively) entropy* over ν, that is, the measures in $M(X)$ that project to ν and attain the maximal relative (weighted, respectively) entropy for ν [11], [8].

For a notational convenience, define $\mathcal{T} : Y \to \mathbb{R}^-$ by

$$\mathcal{T}(y) = -P(\pi, 0)(y) \qquad \text{for } y \in Y$$

and deal with the (T-invariant Borel measurable) function $\mathcal{T}$, instead of $P(\pi, 0)$. Let $\widehat{M}(Y)$ be the set of all regular Borel measures on Y. Here, $\widehat{M}(Y) = C(Y)^*$, the dual space of $C(Y)$, will be equipped with the weak* topology. Define $\mathcal{L} : \widehat{M}(Y) \to \mathbb{R}$ by

$$\mathcal{L}(\nu) = \int \mathcal{T}\, d\nu \qquad \text{for } \nu \in \widehat{M}(Y)$$

so that $\mathcal{L}(\nu) = h(\nu) - \sup_{\mu \in M(\nu)} h(\mu)$ for $\nu \in M(Y)$. The map $\mathcal{R} : M(X) \to \mathbb{R}^+$ defined by $\mathcal{R}(\mu) = h(\mu) - h(\pi\mu)$ is upper semicontinuous in the weak* topology [4].

Lemma 2.1. *The affine map $\mathcal{L} : M(Y) \to \mathbb{R}^-$ is lower semicontinuous.*

Proof. Let $\nu, \nu_n \in M(Y)$, $n \geq 1$, and $\nu_n \to \nu$ as $n \to \infty$. Choose $\mu_* \in M(\nu)$ and $\mu_n \in M(\nu_n)$, $n \geq 1$, such that

$$h(\mu_*) = \max_{\mu \in M(\nu)} h(\mu) \quad \text{and} \quad h(\mu_n) = \max_{\mu \in M(\nu_n)} h(\mu)$$

for $n \geq 1$. Then there is $\{n_k\}_{k=1}^{\infty} \subset \mathbb{N}$ for which μ_{n_k} converges to a measure $\hat{\mu} \in M(X)$ as $n_k \to \infty$ and $\lim_{n_k \to \infty} \mathcal{R}(\mu_{n_k}) = \limsup_{n \to \infty} \mathcal{R}(\mu_n)$. Note that $\pi\hat{\mu} = \nu$. Thus

$$\begin{aligned} \liminf_{n \to \infty} \mathcal{L}(\nu_n) &= \liminf_{n \to \infty} \left[h(\nu_n) - h(\mu_n) \right] = -\limsup_{n \to \infty} \mathcal{R}(\mu_n) \\ &= -\lim_{n_k \to \infty} \mathcal{R}(\mu_{n_k}) \geq -\mathcal{R}(\hat{\mu}) \geq -[h(\mu_*) - h(\nu)] = \mathcal{L}(\nu). \end{aligned}$$

Therefore $\mathcal{L}$ is lower semicontinuous. □

The following relates relative entropy to saturated compensation functions and plays a key role in this work.

Theorem 2.2 ([11]). *Let X and Y be subshifts and let $\pi : X \to Y$ be a factor map. Let $g \in C(Y)$. Then $g \circ \pi \in C(X)$ is a saturated compensation function for π if and only if*

$$\int g\, d\nu = \int \mathcal{T}\, d\nu = \mathcal{L}(\nu) \qquad \text{for all } \nu \in M(Y). \tag{2.1}$$

Corollary 2.3. *Let X and Y be subshifts and $\pi : X \to Y$ a factor map. Then there is a saturated compensation function for π if and only if $\mathcal{L}$ is continuous on $\widehat{M}(Y)$.*

Proof. Let $g \in C(Y)$ and $g \circ \pi \in C(X)$ be a saturated compensation function. It follows from Theorem 2.2 that $\mathcal{L}(\nu) = \int g\, d\nu$ for all $\nu \in M(Y)$. It is not difficult to check that $\mathcal{L}(\nu) = \int g\, d\nu$ on $\widehat{M}(Y)$. Thus $\mathcal{L}$ is continuous on $\widehat{M}(Y)$. Conversely, suppose $\mathcal{L}$ is continuous on $\widehat{M}(Y)$, i.e., $\mathcal{L} \in (\widehat{M}(Y), \text{weak}^*)^*$. Since $C(Y)$ is locally convex, it follows that $(C(Y)^*, \text{weak}^*)^* = C(Y)$, hence $\mathcal{L} \in C(Y)$. Then there is $g \in C(Y)$ for which $\mathcal{L}(\nu) = \int g\, d\nu$ for all $\nu \in \widehat{M}(Y)$, in particular, for all $\nu \in M(Y)$. Again by Theorem 2.2, we obtain a saturated compensation function $g \circ \pi \in C(X)$. □

Consequently, when classifying factor maps in terms of the existence of saturated compensation functions, our task is to investigate whether $\mathcal{L}$ is continuous on $M(Y)$. If $\mathcal{T} \in C(Y)$, then automatically $\mathcal{L}$ is continuous and $\mathcal{T} \circ \pi \in C(X)$ is a saturated compensation function. It is the case when π is a *uniform map*, i.e., a factor map that maps the measure of maximal entropy of X to the measure of maximal entropy of Y. In general, however, $\mathcal{T}$ is not continuous. Finding a saturated compensation function amounts to finding $g \in C(Y)$ that satisfies (2.1). Our aim is to replace $M(Y)$ in (2.1) with a subset so that the statement is still valid.

Let $\mathcal{E}(Y)$ be the set of all ergodic measures in $M(Y)$. In the case where X is an irreducible subshift of finite type, denote by $\mathcal{M}(X)$ the set of all Markov measures in $M(X)$ (which are assumed to be fully supported). Since a Markov measure on an irreducible subshift of finite type is ergodic and so is its image, we have $\pi(\mathcal{M}(X)) \subset \mathcal{E}(Y)$.

Proposition 2.4. *Let X and Y be subshifts and $\pi : X \to Y$ a factor map. For $g \in C(Y)$, the following conditions are equivalent:*

(1) $\int g\, d\nu = \int \mathcal{T}\, d\nu$ *for all* $\nu \in M(Y)$.
(2) $\int g\, d\nu = \int \mathcal{T}\, d\nu$ *for all* $\nu \in \mathcal{E}(Y)$.

If X is an irreducible subshift of finite type, then (1) *or* (2) *is equivalent to*

(3) $\int g\, d\nu = \int \mathcal{T}\, d\nu$ *for all* $\nu \in \pi(\mathcal{M}(X))$.

Proof. Assume (2) holds and fix $\nu \in M(Y)$. By the ergodic decomposition, one can write

$$\nu = \int \rho\, d\omega(\rho)$$

for some (uniquely determined) ω whose support lies in $\mathcal{E}(Y)$. From the hypothesis,

$$\int g\, d\nu = \int \Big(\int g\, d\rho \Big) d\omega(\rho) = \int \Big(\int \mathcal{T}\, d\rho \Big) d\omega(\rho) = \int \mathcal{T}\, d\nu.$$

The last equality holds, since $\mathcal{L}$ is lower semicontinuous [3], [7].

Suppose now that X is an irreducible subshift of finite type and (3) holds. Given $\nu \in M(Y)$, choose $\mu_* \in M(\nu)$ such that $h(\mu_*) = \max_{\mu \in M(\nu)} h(\mu)$. Consider the

markovizations $\{\mu_n\}_{n=1}^{\infty}$ of μ_*, that is, each μ_n, $n \geq 1$, is the (unique) n-step Markov measure on X which coincides with μ_* on the set of all $(n+1)$-cylinder sets in X. Then $\mu_n \to \mu_*$ and $h(\mu_n) \searrow h(\mu_*)$ as $n \to \infty$ [12]. Note that $\pi\mu_n \to \nu$ as $n \to \infty$ and $\pi\mu_n \in \pi(\mathcal{M}(X))$, $n \geq 1$. By the hypothesis $\int \mathcal{T}\, d(\pi\mu_n) = \int g\, d(\pi\mu_n)$ for each $n \geq 1$. Since the mapping $\mu \mapsto h(\mu)$ is upper semicontinuous and $\mathcal{L}$ is lower semicontinuous, it follows that

$$\begin{aligned}\lim_{n\to\infty} \int g\, d(\pi\mu_n) &= \lim_{n\to\infty} \int \mathcal{T}\, d(\pi\mu_n) \leq \liminf_{n\to\infty} \big[h(\pi\mu_n) - h(\mu_n)\big] \\ &\leq h(\nu) - h(\mu_*) = \int \mathcal{T}\, d\nu \leq \lim_{n\to\infty} \int \mathcal{T}\, d(\pi\mu_n).\end{aligned}$$

Thus $\int g\, d\nu = \lim_{n\to\infty} \int g\, d(\pi\mu_n) = \int \mathcal{T}\, d\nu$. □

3. Relative entropy functions and periodic points

Throughout the section, let X and Y denote subshifts of finite type and $\pi : X \to Y$ a factor map. Let $g \in C(Y)$. By Theorem 2.2, a necessary condition for $g \circ \pi \in C(X)$ to be a saturated compensation function is that $\int g\, d\mu_y = \mathcal{T}(y)$ for all periodic points y of Y. We will show that if X is irreducible, then the condition is also sufficient (Theorem 3.5). That is, the set $M(Y)$ in (2.1) can be replaced by the set of measures with finite support.

Hereinafter, every factor map π between subshifts X and Y is assumed to be a *one-block code*, i.e., π is represented by a one-block map from $\mathcal{B}_1(X)$ to $\mathcal{B}_1(Y)$, which we denote again by π. For an n-block $u = u_1 \cdots u_n$ of X, $\pi(u)$ means the n-block $\pi(u_1) \cdots \pi(u_n)$ of Y; for an n-block v of Y, $\pi^{-1}(v)$ denotes the set of n-blocks of X that project to v by the block map π. Given $y \in Y$, for each $n \geq 1$, let $D_n(y)$ consist of one point from each nonempty set $\pi^{-1}(y) \cap [x_0 x_1 \cdots x_{n-1}]$. Then

$$\mathcal{T}(y) = -\limsup_{n\to\infty} \frac{1}{n} \ln \big|D_n(y)\big|$$

(although one may be able to choose $D_n(y)$ in various ways, the number of points in the set, $|D_n(y)|$, is uniquely determined, once y is fixed) [11].

Let $\mathcal{A}$ be the alphabet of X. Fix $y \in Y$. Given $b, c \in \mathcal{A}$ and $n \geq 1$, let

$$\Lambda_y^n(b,c) = \#\big\{u \in \mathcal{B}_{n-1}(X) \big| [buc] \cap \pi^{-1}(y) \neq \emptyset\big\}$$

(so that $\pi b = y_0$ and $\pi c = y_n$). Then

$$\big|D_{n+1}(y)\big| = \sum_{b,c\in\mathcal{A}} \Lambda_y^n(b,c).$$

For each $k \geq 1$, let $P_k(Y) = \{y \in Y | \sigma^k(y) = y\}$ and $P(Y) = \bigcup_{k\geq 1} P_k(Y)$.

Lemma 3.1. *Let $y \in P_k(Y)$, $k \geq 1$, and $b \in \mathcal{A}$. Then*

$$\Lambda_y^k(b,b) = \#\big\{u \in \mathcal{B}_{k-1}(X) | bub \in \mathcal{B}(X) \text{ and } \pi(bub) = y_{[0,k]}\big\}. \tag{3.1}$$

Also, given $p \geq 1$,

$$\big[\Lambda_y^k(b,b)\big]^p \leq \Lambda_y^{pk}(b,b). \tag{3.2}$$

Proof. Let $N = \Lambda_y^k(b,b) \geq 1$. Then there exist distinct blocks $u^{(i)} \in \mathcal{B}_{k-1}(X)$, $1 \leq i \leq N$, such that $\pi(bu^{(i)}b) = y_{[0,k]}$ and $[bu^{(i)}b] \cap \pi^{-1}(y) \neq \emptyset$ for $i = 1, \cdots, N$. If $u \in \mathcal{B}_{k-1}(X)$ is given so that $bub \in \mathcal{B}(X)$ and $\pi(bub) = y_{[0,k]}$, then define $x \in P_k(X)$

by $x = \cdots bu.bubu\cdots$. Clearly $\pi(x) = y$, which implies that $u = u^{(i)}$ for some i, $1 \le i \le N$. Thus (3.1) follows.

Next fix $p \ge 1$. Given a string $\epsilon = i_1 \cdots i_p$ with $1 \le i_j \le N$ for $j = 1, \cdots, p$, set

$$u^\epsilon = bu^{(i_1)}bu^{(i_2)}b\cdots bu^{(i_p)} \in \mathcal{B}_{pk}(X)$$

and define $z^\epsilon \in P_{pk}(Y)$ by $z^\epsilon = \cdots u^\epsilon.u^\epsilon u^\epsilon \cdots$. This is well defined. Also $z_0^\epsilon = z_{pk}^\epsilon = b$ and $\pi(z^\epsilon) = y$. The number of such strings is N^p. Therefore $\Lambda_y^{pk}(b,b) \ge N^p = [\Lambda_y^k(b,b)]^p$. □

Lemma 3.2. *Let $y \in \Delta_Y$ and let $y^{(s)} \in P_{l_s}(Y)$, $l_s \ge s$, for each $s \ge 1$. If there is $N \ge 1$ such that $y^{(s)}_{[0,l_s-N]} = y_{[0,l_s-N]}$ for all s large enough, then $\mu_{y^{(s)}} \to \mu_y$ as $s \to \infty$.*

Proof. Fix $f \in C(Y)$. For each $s \ge 1$, define $\widehat{f}_s \in C(Y)$ by

$$\widehat{f}_s(z) = \frac{1}{l_s}\sum_{i=0}^{l_s-1} f(\sigma^i z)$$

so that $\widehat{f}_s(y^{(s)}) = \int f\, d\mu_{y^{(s)}}$. Since f is uniformly continuous, it follows that given $\epsilon > 0$, there is $q \ge 1$ such that $|f(x) - f(z)| < \epsilon$ whenever $\rho(x,z) < 2^{-q}$. For each $s \ge 1$, let $n_s = l_s - N - q$. Fix $s \ge 1$ so that $n_s \ge q$. Then for $i = q, q+1, \cdots, n_s$, $\sigma^i y^{(s)}$ and $\sigma^i y$ have the same centered $(2q+1)$-block, so that $\rho(\sigma^i y^{(s)}, \sigma^i y) < 2^{-q}$. Thus

$$\begin{aligned}
\left|\widehat{f}_s(y^{(s)}) - \widehat{f}_s(y)\right| &\le \frac{1}{l_s}\sum_{i=0}^{l_s-1}\left|f(\sigma^i y^{(s)}) - f(\sigma^i y)\right| \\
&\le \frac{1}{l_s}\left[\sum_{i=q}^{n_s}\left|f(\sigma^i y^{(s)}) - f(\sigma^i y)\right| + \sum_{\substack{0\le i<q \\ n_s<i<l_s}}\left|f(\sigma^i y^{(s)}) - f(\sigma^i y)\right|\right] \\
&\le \epsilon + \frac{N+2q}{l_s}\cdot 2\|f\|_\infty.
\end{aligned}$$

Since $y \in \Delta_Y$, we have $\widehat{f}_s(y)$ converge to $\int f\, d\mu_y$ as $s \to \infty$. Hence

$$\begin{aligned}
&\limsup_{s\to\infty}\left|\int f\, d\mu_{y^{(s)}} - \int f\, d\mu_y\right| \\
&\qquad \le \limsup_{s\to\infty}\left|\widehat{f}_s(y^{(s)}) - \widehat{f}_s y\right| + \limsup_{s\to\infty}\left|\widehat{f}_s(y) - \int f\, d\mu_y\right| \\
&\qquad \le \lim_{s\to\infty}\left[\epsilon + \frac{N+2q}{l_s}\cdot 2\|f\|_\infty\right] = \epsilon.
\end{aligned}$$

Since $\epsilon > 0$ is arbitrary, $\int f\, d\mu_{y^{(s)}} \to \int f\, d\mu_y$ as $s \to \infty$. Therefore $\mu_{y^{(s)}} \to \mu_y$ as $s \to \infty$. □

This result combined with Corollary 2.3 gives a necessary condition for the existence of saturated compensation functions.

Corollary 3.3. *Let X and Y be subshifts of finite type and $\pi : X \to Y$ a factor map for which a saturated compensation function exists. Let $y \in \Delta_Y$ and let $y^{(s)} \in P_{l_s}(Y)$, $l_s \ge s$, for each $s \ge 1$. If there is $N \ge 1$ such that $y^{(s)}_{[0,l_s-N]} = y_{[0,l_s-N]}$ for all s large enough, then $\mathcal{T}(y^{(s)})$ converges to $\int \mathcal{T}\, d\mu_y$ as s increases.*

Let Y be irreducible. For $y \in Y$, one can easily construct a sequence $\{y^{(s)}\}_{s=1}^{\infty}$ of periodic points that satisfies the condition in Lemma 3.2 (or Corollary 3.3), that is, $y^{(s)} \in P_{l_s}(Y)$ with $l_s \geq s$ for each $s \geq 1$ and there is $N \geq 1$ such that $y^{(s)}_{[0,l_s-N]} = y_{[0,l_s-N]}$ for all s large enough. It follows that $\mu_{y^{(s)}} \to \mu_y$ as $s \to \infty$. It is not always the case that $\mathcal{T}(y^{(s)}) \to \mathcal{T}(y)$ as s increases (if there is no saturated compensation function for π; see Example 3.1). It turns out, however, that for *almost every* $y \in Y$, one can construct such a sequence $\{y^{(s)}\}_{s=1}^{\infty} \subset P(Y)$ so that $\lim_{s\to\infty} \mathcal{T}(y^{(s)}) = \mathcal{T}(y)$.

It can be shown that the set

$$Y_0 = \left\{ y \in Y \middle| \mathcal{T}(y) = -\lim_{n\to\infty} \frac{1}{n} \ln \left| D_n(y) \right| \right\}$$

is a total probability set, i.e., $\nu(Y_0) = 1$ for all $\nu \in M(Y)$ [11]. As a result $P(Y) \subset Y_0$ (see also the proof of Proposition 3.6).

Proposition 3.4. *Let X and Y be irreducible subshifts of finite type and $\pi : X \to Y$ a factor map. For $y \in \Delta_Y \cap Y_0$, there exist $y^{(s)} \in P(Y)$, $s \geq 1$, such that $\mu_{y^{(s)}} \to \mu_y$ as $s \to \infty$ and $\limsup_{s\to\infty} \mathcal{T}(y^{(s)}) \leq \mathcal{T}(y)$.*

Proof. Choose a symbol, say a, of Y that appears infinitely many times in $y^+ = y_{[0,\infty)}$. By shifting y, if necessary, we may assume that $y_0 = a$, since $\mu_{\sigma y} = \mu_y$ and $\mathcal{T}$ is shift-invariant. Then there is a strictly increasing sequence $\{n_k\}_{k=0}^{\infty}$ such that $y_i = a$ for $i \geq 0$ if and only if $i = n_k$ for some $k \geq 0$ ($n_0 = 0$). Let $\mathcal{C} = \pi^{-1}(a)$. For $k \geq 1$, choose $b_k, c_k \in \mathcal{C}$ so that

$$\Lambda_y^{n_k}(b_k, c_k) = \max_{b,c\in\mathcal{C}} \Lambda_y^{n_k}(b, c)$$

(so $\pi(b_k) = y_0$ and $\pi(c_k) = y_{n_k}$). Since

$$\left| D_{n_k+1}(y) \right| / |\mathcal{A}|^2 \leq \Lambda_y^{n_k}(b_k, c_k) \leq \left| D_{n_k+1}(y) \right|$$

for each $k \geq 1$, it follows that

$$\mathcal{T}(y) = -\lim_{n_k\to\infty} \frac{1}{n_k} \ln \left| D_{n_k+1}(y) \right| = -\lim_{n_k\to\infty} \frac{1}{n_k} \ln \left[\Lambda_y^{n_k}(b_k, c_k) \right]. \tag{3.3}$$

Note that there exist $b_*, c_* \in \mathcal{C}$ such that $b_k = b_*$ and $c_k = c_*$ for infinitely many k's, say k_s's, where $k_s \nearrow \infty$ as $s \to \infty$. Since X is irreducible, there is $w \in \mathcal{B}_{m-1}(X)$ for some $m \geq 1$ such that $c_* w b_* \in \mathcal{B}_{m+1}(X)$. Put $u = u_1 \cdots u_m = \pi(c_* w) \in \mathcal{B}_m(Y)$. Then $u_1 = \pi(c_*) = a$ and $u_m a \in \mathcal{B}(Y)$. Fix $s \geq 1$ and let $l_s = n_{k_s} + m$. Define $y^{(s)} \in P_{l_s}(Y)$ by

$$y^{(s)} = \cdots y_0 y_1 \cdots y_{n_{k_s}-1} u_1 \cdots u_m y_0 y_1 \cdots$$

(note that $y^{(s)}_{n_{k_s}} = u_1 = a = y_{n_{k_s}}$ and $u_m y_0 = u_m a \in \mathcal{B}(Y)$). Since $y^{(s)}_{[0,l_s-m]} = y_{[0,l_s-m]}$ and $l_s \geq n_{k_s} \geq k_s \geq s$, it follows from Lemma 3.2 that $\mu_{y^{(s)}} \to \mu_y$ as $s \to \infty$.

To show that $\limsup_{s\to\infty} \mathcal{T}(y^{(s)}) \leq \mathcal{T}(y)$, fix $s \geq 1$. For each $p \geq 1$,

$$\left| D_{p\cdot l_s+1}(y^{(s)}) \right| = \sum_{b,c\in\mathcal{A}} \Lambda_{y^{(s)}}^{p\cdot l_s}(b, c) \geq \Lambda_{y^{(s)}}^{p\cdot l_s}(b_*, b_*),$$

which combined with (3.2) implies that

$$\begin{aligned}\mathcal{T}(y^{(s)}) &= -\lim_{p\to\infty}\frac{1}{p\cdot l_s}\ln\left|D_{p\cdot l_s+1}(y^{(s)})\right| \le -\limsup_{p\to\infty}\frac{1}{p\cdot l_s}\ln\left[\Lambda^{p\cdot l_s}_{y^{(s)}}(b_*,b_*)\right]\\ &\le -\limsup_{p\to\infty}\frac{1}{l_s}\ln\left[\Lambda^{l_s}_{y^{(s)}}(b_*,b_*)\right] = -\frac{1}{l_s}\ln\left[\Lambda^{l_s}_{y^{(s)}}(b_*,b_*)\right].\end{aligned}$$

Meanwhile, using the fact that $\pi(c_*wb_*) = y^{(s)}_{[n_{k_s},l_s]}$ and Lemma 3.1, one can see that

$$\Lambda^{n_{k_s}}_y(b_*,c_*) \le \Lambda^{l_s}_{y^{(s)}}(b_*,b_*).$$

Thus from (3.3),

$$\begin{aligned}\mathcal{T}(y) &= -\lim_{n_{k_s}\to\infty}\frac{1}{n_{k_s}}\ln\left[\Lambda^{n_{k_s}}_y(b_{k_s},c_{k_s})\right] = -\lim_{k_s\to\infty}\frac{1}{n_{k_s}}\ln\left[\Lambda^{n_{k_s}}_y(b_*,c_*)\right]\\ &\ge -\liminf_{k_s\to\infty}\frac{1}{n_{k_s}}\ln\left[\Lambda^{l_s}_{y^{(s)}}(b_*,b_*)\right] = -\liminf_{s\to\infty}\frac{1}{l_s}\ln\left[\Lambda^{l_s}_{y^{(s)}}(b_*,b_*)\right]\\ &= \limsup_{s\to\infty}\Big[-\frac{1}{l_s}\ln\left[\Lambda^{l_s}_{y^{(s)}}(b_*,b_*)\right]\Big] \ge \limsup_{s\to\infty}\mathcal{T}(y^{(s)}).\end{aligned}$$

Therefore $\limsup_{s\to\infty}\mathcal{T}(y^{(s)}) \le \mathcal{T}(y)$. □

Theorem 3.5. *Let X and Y be irreducible subshifts of finite type and $\pi : X \to Y$ a factor map. Let $g \in C(Y)$. Then $g\circ\pi \in C(X)$ is a saturated compensation function for π if and only if $\int g\,d\mu_y = \mathcal{T}(y)$ for all $y \in P(Y)$, or equivalently, for each $k \ge 1$,*

$$\frac{1}{k}\sum_{i=0}^{k-1} g(\sigma^i y) = \mathcal{T}(y) \qquad \text{for all } y \in P_k(Y). \tag{3.4}$$

Proof. Suppose $\int g\,d\mu_y = \mathcal{T}(y)$ for all $y \in P(Y)$. Let $\nu \in \mathcal{E}(Y)$. Then the set of all generic points $y \in \Delta_Y$ with $\mu_y = \nu$ is of full measure with respect to ν. Hence there is $y \in \Delta_Y \cap Y_0$ for which $\nu = \mu_y$. Moreover, one can choose such a y so that $\int \mathcal{T}\,d\nu = \int \mathcal{T}\,d\mu_y = \mathcal{T}(y)$. By Proposition 3.4, we obtain $y^{(s)} \in P(Y)$, $s \ge 1$, such that $\mu_{y^{(s)}} \to \mu_y$ and $\limsup_{s\to\infty}\mathcal{T}(y^{(s)}) \le \mathcal{T}(y)$. Meanwhile, since $\mathcal{L}$ is lower semicontinuous,

$$\mathcal{T}(y) = \mathcal{L}(\mu_y) \le \liminf_{s\to\infty}\mathcal{L}(\mu_{y^{(s)}}) = \liminf_{s\to\infty}\mathcal{T}(y^{(s)}).$$

Consequently $\lim_{s\to\infty}\mathcal{T}(y^{(s)}) = \mathcal{T}(y)$. By the assumption, $\int g\,d\mu_{y^{(s)}} = \mathcal{T}(y^{(s)})$ for all $s \ge 1$. Since $\int g\,d\mu_{y^{(s)}} \to \int g\,d\mu_y$ as $s \to \infty$, it follows that $\int g\,d\mu_y = \mathcal{T}(y)$, i.e., $\int g\,d\nu = \int \mathcal{T}\,d\nu$. Proposition 2.4 implies that $g\circ\pi \in C(X)$ is a saturated compensation function. □

Remark 3.1. Let $g \in C(Y)$. Theorem 3.5 implies that $\int g\,d\nu = \int \mathcal{T}\,d\nu$ for all $\nu \in M(Y)$ if and only if $\int g\,d\nu = \int \mathcal{T}\,d\nu$ for all $\nu \in M(Y)$ with finite support. That is, if the map $\mathcal{L}$ is continuous on the set of measures with finite support, then it is continuous on $M(Y)$. It is known that for an irreducible subshift of finite type Y, invariant measures with finite support (i.e., the periodic point measures) are ergodic and form a dense set in $M(Y)$ (see [3], [10]).

In an attempt to find a saturated compensation function, *theoretically*, one may construct $g \in C(Y)$ inductively so that it satisfies a *finite* number of equations presented in (3.4) for each $k \geq 1$, provided that $\mathcal{T}(y)$ is given for all $y \in P(Y)$.

We describe how to evaluate $\mathcal{T}$ on $P(Y)$. Fix $y \in P_k(Y)$, $k \geq 1$. To construct a graph G_y, set $\mathcal{A}_y = \pi^{-1}\{y_0 y_1 \cdots y_{k-1}\}$ and let $\pi^{-1}\{y_{k-1}\} = \{w_1, w_2, \cdots, w_s\}$, $s \geq 1$. Reordering w_i's (if necessary), we may assume that for some r, $1 \leq r \leq s$,

$$\{w_1, w_2, \cdots, w_r\} = \{u \in \mathcal{B}_1(X) | \text{ there is } v \in \mathcal{B}_{k-1}(X) \text{ such that } vu \in \mathcal{A}_y\}.$$

Then $w_1, w_2, \cdots, w_r$ form the set of vertices of G_y. Next fix j, $1 \leq j \leq r$. By the assumption, there is $u_0 u_1 \cdots u_{k-2} \in \mathcal{B}_{k-1}(X)$ such that $\pi(u_0 u_1 \cdots u_{k-2} w_j) = y_0 y_1 \cdots y_{k-1}$. If there is a $(k+1)$-block $w_i u_0 u_1 \cdots u_{k-2} w_j$ in X for some i, $1 \leq i \leq r$, then build an edge, say e, from the state w_i to the state w_j. We now label e with $\mathcal{L}_y(e) = u_0 u_1 \cdots u_{k-2} w_j$. Hence each block in $\mathcal{A}_y$ terminating at w_j generates incoming (labeled) edges at the state w_j. Letting $\mathcal{E}(\mathsf{G}_y)$ be the set of all such edges yields a labeled graph $\mathcal{G}_y = (\mathsf{G}_y, \mathcal{L}_y)$ with the labeling $\mathcal{L}_y : \mathcal{E}(\mathsf{G}_y) \to \mathcal{A}_y$. Observe that $\mathcal{G}_y$ is *right-resolving*, i.e., the outgoing edges at each vertex of G_y carry different labels. Thus $h(\mathsf{X}_{\mathcal{G}_y}) = h(\mathsf{X}_{G_y})$, i.e., the sofic shift $\mathsf{X}_{\mathcal{G}_y}$ and the edge shift X_{G_y} have the same topological entropy (for notations and more details, see [5]).

Proposition 3.6. *Let X and Y be subshifts of finite type and $\pi : X \to Y$ a factor map. If $y \in P_k(Y)$, $k \geq 1$, and $\mathcal{G}_y = (\mathsf{G}_y, \mathcal{L}_y)$ is defined as above, then*

$$\mathcal{T}(y) = -\frac{h(\mathsf{X}_{\mathcal{G}_y})}{k} = -\frac{h(\mathsf{X}_{\mathsf{G}_y})}{k}.$$

Proof. Fix $n > k$. Set $p \geq 2$ to be the smallest integer so that $n \leq pk$. Note that

$$\left|\mathcal{B}_p(\mathsf{X}_{\mathsf{G}_y})\right| \geq \left|D_{pk}(y)\right| \geq \left|D_n(y)\right| \geq \left|D_{(p-1)k}(y)\right| \geq \left|\mathcal{B}_{p-1}(\mathsf{X}_{\mathcal{G}_y})\right|.$$

Hence

$$\frac{1}{p-1} \ln \left|\mathcal{B}_p(\mathsf{X}_{\mathsf{G}_y})\right| \geq \frac{k}{n} \ln \left|D_n(y)\right| \geq \frac{1}{p} \ln \left|\mathcal{B}_{p-1}(\mathsf{X}_{\mathcal{G}_y})\right|$$

from which it follows that

$$\begin{aligned} h(\mathsf{X}_{\mathsf{G}_y}) &= \lim_{p\to\infty} \frac{1}{p-1} \ln \left|\mathcal{B}_p(\mathsf{X}_{\mathsf{G}_y})\right| \geq \limsup_{n\to\infty} \frac{k}{n} \ln \left|D_n(y)\right| \\ &\geq \liminf_{n\to\infty} \frac{k}{n} \ln \left|D_n(y)\right| \geq \lim_{p\to\infty} \frac{1}{p} \ln \left|\mathcal{B}_{p-1}(\mathsf{X}_{\mathcal{G}_y})\right| = h(\mathsf{X}_{\mathcal{G}_y}). \end{aligned}$$

Thus

$$h(\mathsf{X}_{\mathsf{G}_y}) = k \cdot \lim_{n\to\infty} \frac{1}{n} \ln \left|D_n(y)\right| = h(\mathsf{X}_{\mathcal{G}_y}),$$

which completes the proof. □

Remark 3.2. Since $h(\mu_y) = 0$ for all $y \in P(Y)$, it follows that for $y \in P_k(Y)$, $k \geq 1$, the maximal relative entropy for $\mu_y \in M(Y)$ is $h(\mathsf{X}_{\mathsf{G}_y})/k$.

Example 3.1. Let X and Y be the subshifts of finite type determined by allowing the transitions marked in Figure 3.1 and the one-block code $\pi : X \to Y$ map 1 to 1, and $2, 3, 4, 5$ to 2.

For each $s \geq 1$, let

$$y^{(s)} = \cdots 12^s.12^s 12^s 1 \cdots \in P_{s+1}(Y).$$

Fix $s = 2m + 2 \geq 2$ for some $m \geq 0$ and put $y = y^{(s)}$. Under the notations as above, note that $v \in \mathcal{A}_y = \pi^{-1}(12^s)$ if and only if either $v = 12^s$ or $v =$

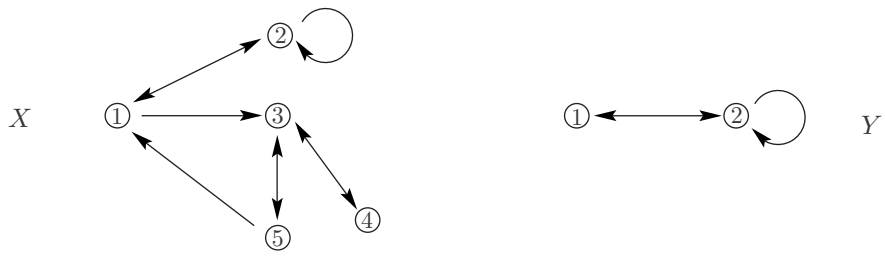

FIGURE 3.1.

$13\epsilon_1 3\epsilon_2 3\cdots 3\epsilon_m 3\epsilon_{m+1}$, where $\epsilon_i \in \{4,5\}$ for $i = 1,\cdots,m+1$. For $\epsilon = (\epsilon_1,\cdots,\epsilon_m) \in \{4,5\}^m$, define $b_\epsilon, c_\epsilon \in \mathcal{B}_{s+1}(X)$ by

$$b_\epsilon = 13\epsilon_1 3\epsilon_2 3\cdots 3\epsilon_m 34 \qquad \text{and} \qquad c_\epsilon = 13\epsilon_1 3\epsilon_2 3\cdots 3\epsilon_m 35.$$

Setting $a_1 = 12^s$, we have $\mathcal{A}_y = \{a_1, b_\epsilon, c_\epsilon | \epsilon \in \{4,5\}^m\}$ so that $|\mathcal{A}_y| = 2^{m+1}+1$. Thus the underlying graph G_y has three states, $2, 4$ and 5, and the labeled graph $\mathcal{G}_y = (\mathsf{G}_y, \mathcal{L}_y)$ is seen in Figure 3.2.

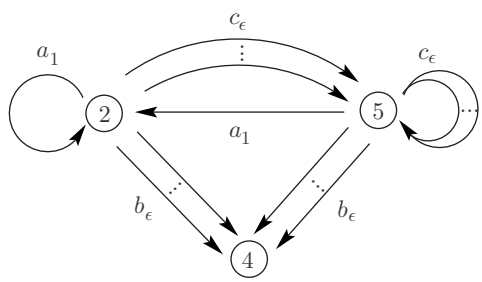

FIGURE 3.2.

The adjacency matrix is given by

$$\mathsf{A}(\mathsf{G}_y) = \begin{bmatrix} 1 & 2^m & 2^m \\ 0 & 0 & 0 \\ 1 & 2^m & 2^m \end{bmatrix}$$

which implies that $h(\mathsf{X}_{\mathsf{G}_y}) = \ln(2^m+1)$. By Proposition 3.6,

$$\mathcal{T}(y) = \mathcal{T}(y^{(s)}) = -\frac{1}{s+1}\ln(2^{s/2-1}+1). \tag{3.5}$$

It was shown in [9] that there is no saturated compensation function for π. This can be verified by the arguments developed in this section. Let $y^* = \cdots 22.122\cdots \in Y$, i.e., $y^*_i = 1$ if $i = 0$ and $y^*_i = 2$ if $i \neq 0$. Note that $y^* \in \Delta_Y$ and $y^{(s)}_{[0,s]} = y^*_{[0,s]}$ for each $s \geq 1$. Thus y_* and $y^{(s)}$, $s \geq 1$, satisfy the conditions in Corollary 3.3. It is a simple observation that if s is odd, then $|D_n(y^{(s)})| = 1$ for all $n \geq 1$, so that $\mathcal{T}(y^{(s)}) = 0$. Fix $s = 2m+2$, $m \geq 0$. An easy computation shows that $|D_s(y^{(s)})| = 2^m+1$. Using the fact that $\pi^{-1}\{1\} = \{1\}$, we obtain that for $p \geq 1$,

$$\left|D_{p(s+1)}(y^{(s)})\right| = \left|D_s(y^{(s)})\right|^p = \left(2^{s/2-1}+1\right)^p.$$

Thus

$$\begin{aligned}\mathcal{T}(y^{(s)}) &= -\lim_{p\to\infty}\frac{1}{p(s+1)}\ln\left|D_{p(s+1)}(y^{(s)})\right| \\ &= -\lim_{p\to\infty}\frac{1}{s+1}\ln\left(2^{s/2-1}+1\right) = -\frac{1}{s+1}\ln\left(2^{s/2-1}+1\right)\end{aligned}$$

(this coincides with (3.5)). Consequently $\mathcal{T}(y^{(2m)}) \to -\ln\sqrt{2}$ as $m\to\infty$, so that $\mathcal{T}(y^{(s)})$ does *not* converge as s increases. By Corollary 3.3 no saturated compensation function exists.

The property that $\pi^{-1}\{1\} = \{1\}$ characterizes the symbol 1 as a singleton clump for π. In general, a factor map $\pi : X \to Y$ between two subshifts X and Y is said to have a *singleton clump* if there is a symbol a of Y whose inverse image (by the one-block map π) is a singleton. We briefly discuss right-resolving factor maps with a singleton clump regarding the existence of saturated compensation functions.

Definition 3.1. A factor map $\pi : X \to Y$ between subshifts X, Y is called a *right-resolving map* if whenever w is a symbol of X, $\pi(w) = i$, and ij is a 2-block in Y, there is a follower symbol u of w that maps to j.

This definition extends the bounded-to-one right-resolving codes used in [6] to the infinite-to-one case (see [5]). It is not difficult to see the following. We will omit the proof.

Lemma 3.7. *Let X and Y be irreducible subshifts of finite type such that $Y \subset \{u,v\}^{\mathbb{Z}}$. Let $\pi : X \to Y$ be a right-resolving factor code. If $uv, vv \in \mathcal{B}(Y)$, then*

$$\lim_{n\to\infty}\frac{1}{n}\ln\left|\pi^{-1}(v^n)\right| = \lim_{n\to\infty}\frac{1}{n}\ln\left|\pi^{-1}(uv^nu)\right|. \tag{3.6}$$

It was shown that if u is a singleton clump for π, then (3.6) guarantees the existence of a saturated compensation function [9]. The following is immediate.

Theorem 3.8. *Let X be an irreducible subshift of finite type and let $\pi : X \to Y$ be a factor map with a singleton clump, where Y is a subshift of the full 2-shift. If π is right-resolving, then there is a saturated compensation function for π.*

References

1. R. Bowen, Entropy for group endomorphisms and homogeneous spaces, *Trans. Amer. Math. Soc.* **153** (1971), 401–414. MR0274707 (43:469)
2. M. Boyle and S. Tuncel, Infinite-to-one codes and Markov measures, *Trans. Amer. Math. Soc.* **285** (1984), 657–683. MR0752497 (86b:28024)
3. M. Denker, C. Grillenberger and K. Sigmund, *Ergodic Theory on Compact Spaces*, Lecture Notes in Math., vol. **527**, Springer-Verlag, New York, 1976. MR0457675 (56:15879)
4. F. Ledrappier and P. Walters, A relativised variational principle for continuous transformations, *J. London Math. Soc.* **16** (1977), 568–576. MR0476995 (57:16540)
5. D. Lind and B. Marcus, *An Introduction to Symbolic Dynamics and Coding*, Cambridge Univ. Press, 1995. MR1369092 (97a:58050)
6. W. Parry and S. Tuncel, On the classification of Markov chains by finite equivalence, *Erg. Th. and Dyn. Sys.* **1** (1981), 303–335. MR0662472 (83j:28020)
7. D. Ruelle, *Thermodynamic Formalism*, Encyclopedia of Math., vol. **5**, Addison-Wesley, Reading, Mass., 1978. MR0511655 (80g:82017)
8. S. Shin, Measures that maximize weighted entropy for factor maps between subshifts of finite type, *Erg. Th. and Dyn. Sys.* **21** (2001), 1249–1272. MR1849609 (2002i:37009)
9. _____, An example of a factor map without a saturated compensation function, *Erg. Th. and Dyn. Sys.* **21** (2001), 1855–1866. MR1869074 (2002h:37020)

10. P. Walters, *An Introduction to Ergodic Theory*, Graduate Texts in Math., no. **79**, Springer-Verlag, Berlin, 1982. MR0648108 (84e:28017)
11. ———, Relative pressure, relative equilibrium states, compensation functions and many-to-one codes between subshifts, *Trans. Amer. Math. Soc.* **296** (1986), 1–31. MR0837796 (87j:28028)
12. B. Weiss, Subshifts of finite type and sofic systems, *Monats, Math.* **77** (1973), 462–474. MR0340556 (49:5308)

Department of Mathematics, Korea Advanced Institute of Science and Technology, Daejeon 305-701, South Korea

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