ISRAEL JOURNAL OF MATHEMATICS 110 (1999), 359–369

THE VARIATIONAL PRINCIPLE FOR THE DEFECT OF FACTOR MAPS

BY

KLAUS THOMSEN

Department of Mathematics, University of Aarhus
Ny Munkegade, 8000 Aarhus C, Denmark
e-mail: matkt@imf.au.dk

ABSTRACT

This note is a successor of [T], where the author introduced the concept of defect for factor maps out of dynamical systems on totally disconnected spaces. The purpose here is to prove the variational principle for the defect and derive some consequences of it.

1. Introduction

The defect of a factor map, π: $(Y,\psi) \to (X,\varphi)$, between dynamical systems with Y totally disconnected, is a number $D(\pi) \in [0,\infty]$ which gives a numerical indication of how far π is from being a conjugacy. It was defined, in [T], in the following way. For each finite partition $\mathcal{P} = \{P_i\}_{i\in I}$ of Y consisting of closed and open sets, let $q_k(\varphi,\pi(\mathcal{P}))$ denote the maximal number of elements in a subset $J \subset I^k$ such that

$$\bigcap_{(i_1,i_2,\dots,i_k)\in J} \pi(P_{i_1}) \cap \varphi^{-1}(\pi(P_{i_2})) \cap \cdots \cap \varphi^{-k+1}(\pi(P_{i_k}))$$

is not empty. Since $q_{k+n}(\varphi,\pi(\mathcal{P})) \leq q_k(\varphi,\pi(\mathcal{P}))q_n(\varphi,\pi(\mathcal{P}))$ for all k,n, we can consider the quantity

$$Q(\varphi,\pi(\mathcal{P})) = \lim_{n\to\infty} \frac{1}{n} \log q_n(\varphi,\pi(\mathcal{P})).$$

Then

$$D(\pi) = \sup_{\mathcal{P}} Q(\varphi,\pi(\mathcal{P})).$$

Received February 5, 1998

The defect is zero when π is a conjugacy, but non-zero as soon as φ has a periodic point with more than one pre-image under π. In [T] some general properties of the defect were deduced and the defect was calculated in a series of examples, including the factor maps out of finite type subshifts resulting from Markov partitions of expansive homeomorphisms and expansive endomorphisms. Furthermore, it was shown (when Y is metrizable) that

$$D(\pi) \geq \sup_{\mu} \int_X \log \#\pi^{-1}(x) d\mu(x), \tag{1.1}$$

where μ varies over all φ-invariant Borel probability measures on X, and that equality holds in (1.1) when there is only one μ to consider, i.e. when φ is uniquely ergodic. The main purpose of the present note is to remove this assumption on φ and prove that equality always holds in (1.1). This identity can be considered as a variational principle for the defect, in analogy with the celebrated variational principle for the topological entropy.

In addition to proving the variational principle for the defect, we derive here three consequences of it. Firstly, we show that when ψ and φ are both homeomorphisms, the defect must be the logarithm of a natural number or infinite, and when it is finite, say equal to $\log\ k$, there is a φ-ergodic measure μ such that $\#\pi^{-1}(x) = k$ for μ-almost all x. (In contrast, when the dynamical systems are not homeomorphisms, but merely endomorphisms, the defect can take on any value in $[0, \infty]$; see [T].) The second consequence of the variational principle which we present is also concerned only with invertible dynamical systems, and shows that the defect is subadditive with respect to composition of factor maps; in symbols,

$$D(\pi_2 \circ \pi_1) \leq D(\pi_2) + D(\pi_1).$$

See Theorem 3.3 for the precise statement. This subadditivity fails for factor maps between general (non-invertible) dynamical systems as we show by example. Finally, we use the variational principle to prove that the defect is finite only when the topological entropies of the dynamical systems involved are the same.

2. The variational principle

In the following a **dynamical system**, (X, φ), will mean a compact metric space X and a continuous map $\varphi: X \to X$. The dynamical system is called **invertible** when φ is a homeomorphism. We adopt the notation from [T].

THEOREM 2.1: *Let π: $(Y,\psi) \to (X,\varphi)$ be a factor map between dynamical systems with Y totally disconnected. Then*

$$D(\pi) = \sup_{\mu} \int_X \log \#\pi^{-1}(x) d\mu(x),$$

where the supremum is taken over all φ-invariant Borel probability measures. In fact, it suffices to take the supremum over all φ-ergodic invariant Borel probability measures.

Proof: By Proposition 5.1 in [T] it suffices, for the proof of the first statement, to show that

$$D(\pi) \le \sup_{\mu} \int_X \log \#\pi^{-1}(x) d\mu(x).$$

Let $t \in \mathbb{R}$ be a number strictly less than $D(\pi)$ and let $\epsilon > 0$. We must produce a φ-ergodic invariant probability measure μ such that $\int_X \log \#\pi^{-1}(x)\, d\mu(x) \ge t - 2\epsilon$. Choose a finite partition, $\mathcal{P} = \{P_i\}_{i\in I}$, of Y by open and closed sets such that $Q(\varphi, \pi(\mathcal{P})) > t$. To simplify notation we set $F_i = \pi(P_i)$ and $\mathcal{F} = \pi(\mathcal{P}) = \{F_i\}_{i\in I}$. For each $m \in \mathbb{N}$, choose $x_m \in X$ such that

$$\frac{1}{m} \log q_m(\varphi, \mathcal{F}) - \epsilon \le \frac{1}{m} \log q_m(x_m, \mathcal{F}).$$

Set

$$\mu_m = \frac{1}{m} \sum_{i=0}^{m-1} \delta_{\varphi^i(x_m)},$$

where δ_y denotes the Dirac measure at y. Let μ be a weak*-condensation point for the sequence $\{\mu_m\}$. Then μ is clearly φ-invariant. For each $k \in \{1, 2, \cdots, \#I\}$, set $H_k = \bigcup_J \bigcap_{j\in J} F_j$ where we take the union over all subsets J of I of cardinality $\le k$. Since H_k is closed we can find a decreasing sequence $f_k^1 \ge f_k^2 \ge \cdots$ of continuous functions on X such that $f_k^n(X) \subset [0,1]$ and $f_k^n(x) = 1,\ x \in H_k$, for all $n \in \mathbb{N}$ and $\lim_{n\to\infty} f_k^n(x) = 1_{H_k}(x)$ for all $x \in X$. Fix $n \in \mathbb{N}$. There is then an m such that

$$\int_X \log(\sum_{k=1}^{\#I} f_k^n(x)) d\mu_m(x) \le \int_X \log(\sum_{k=1}^{\#I} f_k^n(x)) d\mu(x) + \epsilon.$$

Since $q_m(x_m, \mathcal{F}) \le \prod_{i=0}^{m-1} \sum_{k=1}^{\#I} f_k^n(\varphi^i(x_m))$ we have that

$$\begin{aligned} \frac{1}{m} \log q_m(x_m, \mathcal{F}) &\le \frac{1}{m} \sum_{i=0}^{m-1} \log(\sum_{k=1}^{\#I} f_k^n(\varphi^i(x_m))) \\ &= \int_X \log(\sum_{k=1}^{\#I} f_k^n(x)) d\mu_m(x) \le \int_X \log(\sum_{k=1}^{\#I} f_k^n(x)) d\mu(x) + \epsilon. \end{aligned}$$

But

$$Q(\varphi, \mathcal{F}) = \inf_{k\in\mathbb{N}} \frac{1}{k} \log q_k(\varphi, \mathcal{F}) \leq \frac{1}{m} \log q_m(\varphi, \mathcal{F}) \leq \frac{1}{m} \log q_m(x_m, \mathcal{F}) + \epsilon,$$

so we find that $Q(\varphi, \mathcal{F}) \leq \int_X \log\left(\sum_{k=1}^{\#I} f_k^n(x)\right) d\mu(x) + 2\epsilon$. By taking the limit over n this shows that

$$Q(\varphi, \mathcal{F}) \leq \int_X \log A_{\mathcal{F}}(x) d\mu(x) + 2\epsilon, \tag{2.1}$$

where $A_{\mathcal{F}}(x) = \#\{i \in I : x \in F_i\}$. Let M denote the compact convex set of φ-invariant Borel probability measures on X. Define an affine function ψ: $M \to [0, \infty[$ by

$$\psi(\nu) = \int_X \log A_{\mathcal{F}}(x) d\nu(x).$$

Since $\psi(\nu) = \inf_n \int_X \log\left(\sum_{k=1}^{\#I} f_k^n(x)\right) d\nu(x)$ we see that ψ is upper semi-continuous, so by Bauers maximum principle ψ attains its maximum value at an extreme point. This extreme point, ν_0, is a φ-ergodic invariant Borel probability measure and, by (2.1),

$$Q(\varphi, \mathcal{F}) - 2\epsilon \leq \int_X \log A_{\mathcal{F}}(x) d\nu_0(x).$$

Since $A_{\mathcal{F}}(x) \leq \#\pi^{-1}(x)$, we find that

$$t - 2\epsilon < Q(\varphi, \mathcal{F}) - 2\epsilon \leq \int_X \log A_{\mathcal{F}}(x) d\nu_0(x) \leq \int_X \log \#\pi^{-1}(x) d\nu_0(x). \qquad \blacksquare$$

3. Consequences

THEOREM 3.1: *Let π: $(Y, \psi) \to (X, \varphi)$ be a factor map between invertible dynamical systems with Y totally disconnected. Assume that $D(\pi) < \infty$. There is then a $k \in \mathbb{N}$ and a φ-ergodic probability measure μ on X such that $\#\pi^{-1}(x) = k$ for μ-almost all x, and*

$$D(\pi) = \log k.$$

Proof: Let k be the natural number such that $\log(k-1) < D(\pi) \leq \log k$. By Theorem 2.1 there is some φ-ergodic invariant Borel probability measure μ such that

$$\mu(\{x \in X : \#\pi^{-1}(x) > k-1\}) = \mu(\{x \in X : \#\pi^{-1}(x) \geq k\}) > 0.$$

Since $x \mapsto \#\pi^{-1}(x)$ is φ-invariant, the ergodicity implies that

$$\mu(\{x \in X : \#\pi^{-1}(x) = k\}) = 1. \qquad \blacksquare$$

In contrast, when considering the defect between general (non-invertible) dynamical systems, the defect can take on all values in $[0, \infty]$; cf. [T] .

Remark 3.2: As this result shows, the calculations described in Example 4.11 of [T] are wrong. The defect of the factor map $\pi(k,q)$ is not $\frac{2k}{Q}\log 2$ as asserted, but $2k \log 2$. The number $\frac{2k}{Q}\log 2$ is only the quantity $Q(\varphi, \pi(k,q)(\mathcal{P}))$, where $\mathcal{P}$ is the cover of $\Sigma_{k,q}$ obtained by fixing the first coordinate. To get the correct larger value one must consider the partitions obtained by fixing, at least, Q successive coordinates. $\blacksquare$

We turn to other consequences of the variational principle for the defect.

THEOREM 3.3: (Subadditivity of the defect) *Let (X, φ), (Y, ψ) and (Z, λ) be invertible dynamical systems with X and Y totally disconnected. Let $\pi_1: (X, \varphi) \to (Y, \psi)$ and $\pi_2: (Y, \psi) \to (Z, \lambda)$ be factor maps. It follows that*

$$D(\pi_2 \circ \pi_1) \leq D(\pi_1) + D(\pi_2).$$

Proof: Let μ be a λ-invariant Borel probability measure. There is a ψ-invariant Borel probability measure ν such that $\mu = \nu \circ \pi_2^{-1}$. Let $\mathcal{B}$ and $\mathcal{C}$ denote the completions of the Borel sets in Z, resp. Y, with respect to the measure μ, resp. ν. If there is a $\mathcal{B}$-measurable set $A \subset Z$ of positive measure such that $\#\pi_2^{-1}(z) = \infty$ for $z \in A$, we have that $\int_Z \log \#\pi_2^{-1}(z) d\mu(z) = \infty$ and hence $D(\pi_2) = \infty$. In this case the inequality we are proving is trivial, so we may assume that $\#\pi_2^{-1}(z) < \infty$ for μ-almost all $z \in Z$. Similarly, we may assume that $\#\pi_1^{-1}(y) < \infty$ for ν-almost all $y \in Y$. Consider

$$Y_1 = \{y \in Y : \#\pi_1^{-1}(y) \geq \#\pi_1^{-1}(x),\ x \in \pi_2^{-1}(\pi_2(y))\}.$$

By using results of Rohlin, cf. [R], one can show that Y_1 is a $\mathcal{C}$-measurable subset of Y. Note that Y_1 is ψ-invariant since φ and ψ are invertible. Set $\pi = \pi_2|_{Y_1}$. By ignoring nullsets we can write Z as a disjoint union, $Z = Z_1 \cup Z_2 \cup Z_3 \cup \cdots$, where $\#\pi^{-1}(z) = k$, $z \in Z_k$. Furthermore, by [R] we can choose measurable functions $\kappa_i^k: Z_k \to Y_1$, $i = 1, 2, \ldots, k$, such that $\pi^{-1}(z) = \{\kappa_1^k(z), \kappa_2^k(z), \cdots, \kappa_k^k(z)\}$ for all $z \in Z_k$. There is then a Borel probability measure ν_1 on Y such that

$$\int_Y f(y) d\nu_1(y) = \sum_k \int_{Z_k} \frac{1}{k} \sum_{j=1}^{k} f \circ \kappa_j^k(z) d\mu(z), \quad f \in C(Y).$$

Since $\{\psi \circ \kappa_1^k(z), \psi \circ \kappa_2^k(z), \cdots, \psi \circ \kappa_k^k(z)\} = \{\kappa_1^k(\lambda(z)), \kappa_2^k(\lambda(z)), \cdots, \kappa_k^k(\lambda(z))\}$ for all $z \in Z_k$ we find that

$$\begin{aligned}\int_Y f \circ \psi(y) d\nu_1(y) &= \sum_k \int_{Z_k} \frac{1}{k} \sum_{j=1}^k f \circ \psi \circ \kappa_j^k(z) d\mu(z) \\ &= \sum_k \int_{Z_k} \frac{1}{k} \sum_{j=1}^k f \circ \kappa_j^k \circ \lambda(z) d\mu(z) \\ &= \sum_k \int_{Z_k} \frac{1}{k} \sum_{j=1}^k f \circ \kappa_j^k(z) d\mu(z) \\ &= \int_Y f(y) d\nu_1(y)\end{aligned}$$

for all $f \in C(Y)$, proving that ν_1 is ψ-invariant. Note that

$$\begin{aligned}\log \#(\pi_2 \circ \pi_1)^{-1}(z) &= \log \left(\sum_{y \in \pi_2^{-1}(z)} \#\pi_1^{-1}(y) \right) \\ &= \log \#\pi_2^{-1}(z) + \log \left(\frac{1}{\#\pi_2^{-1}(z)} \sum_{y \in \pi_2^{-1}(z)} \#\pi_1^{-1}(y) \right) \\ &\leq \log \#\pi_2^{-1}(z) + \frac{1}{k} \sum_{j=1}^k \log \#\pi_1^{-1}(\kappa_j^k(z))\end{aligned}$$

for all $z \in Z_k$, $k \in \mathbb{N}$. Hence

$$\begin{aligned}&\int_Z \log \#(\pi_2 \circ \pi_1)^{-1}(z) d\mu(z) \\ &\leq \int_Z \log \#\pi_2^{-1}(z) d\mu(z) + \sum_k \int_{Z_k} \frac{1}{k} \sum_{j=1}^k \log \#\pi_1^{-1}(\kappa_j^k(z)) d\mu(z) \\ &= \int_Z \log \#\pi_2^{-1}(z) d\mu(z) + \int_Y \log \#\pi_1^{-1}(y) d\nu_1(y) \\ &\leq D(\pi_2) + D(\pi_1).\end{aligned}$$

The conclusion follows by taking the supremum over all μ. ∎

Remark 3.4: The subadditivity of the defect fails when we consider general (non-invertible) dynamical systems. To see this it would suffice to consider Example 2.5 of [T], but it is better to put that example in the following, slightly wider perspective. Let π_1: $(F_1, \varphi_1) \to (F_2, \varphi_2)$ be a factor map where both F_1 and F_2 are finite sets. By identifying F_i with

$$\{(x, \varphi_i(x), \varphi_i^2(x), \varphi_i^3(x), \dots) \in F_i^{\mathbb{N}} \ : \ x \in F_i\}$$

we can consider (F_i, φ_i) as a (rather trivial) one-sided subshift of finite type. In this picture π_1 becomes a factor map which is full in the sense of Definition 2.1 of [T], so we may apply Theorem 2.4 of [T] to conclude that the defect $D(\pi_1)$ is

$$\max\Big\{ \frac{1}{p}\sum_{i=1}^{p} \log \#\pi_1^{-1}(\varphi_2^i(x)) \ : x \in F_2 \text{ has period } p \text{ under } \varphi_2, \\ p \in \{1, 2, \dots, \#F_2\} \Big\}.$$

In particular, when φ_2 is transitive and hence conjugate to a cyclic permutation as in Example 2.5 of [T], the defect is

$$\frac{1}{m}\sum_{i=1}^{m} \log d_i$$

where $m = \#F_2$, $F_2 = \{x_1, x_2, \dots, x_m\}$ and $d_i = \#\pi_1^{-1}(x_i)$, $i = 1, 2, \dots, m$. Let π_2: $F_2 \to \{0\}$ be the map which identifies all points in F_2. Then $D(\pi_2) = \log\, m$ and, in general, $D(\pi_2 \circ \pi_1) = \log \#F_1 = \log(\sum_{i=1}^m d_i) \not\leq \log\, m + (\sum_{i=1}^m \log\, d_i)/m$. So the subadditivity of the defect fails in these examples. ∎

The next goal will be to prove the following

THEOREM 3.5: *Let π: $(Y, \psi) \to (X, \varphi)$ be a factor map between dynamical systems with Y totally disconnected. Then*

$$D(\pi) < \infty \Rightarrow h(\psi) = h(\varphi).$$

This result seems highly probable in view of Theorem 2.1, but I haven't been able to find a short proof based on 2.1 alone. The strategy of the proof will be to prove it first when both dynamical systems are invertible, and then reduce the general case to that case by use of the 'natural extension'.

LEMMA 3.6: *Let π: $(Y, \psi) \to (X, \varphi)$ be a factor map between dynamical systems with Y totally disconnected. Then*

$$h(\psi) \leq h(\varphi) + D(\pi).$$

Proof: Let $\mathcal{P} = \{P_i\}_{i\in I}$ be a finite partition of Y consisting of open and closed sets. Then the entropy, $h(\psi, \mathcal{P})$, of ψ with respect to $\mathcal{P}$ is $\lim_{m\to\infty} \frac{1}{m} \log\ M_m$ where

$$M_m = \#\{(i_1, i_2, \dots, i_m) \in I^m \ :\ P_{i_1} \cap \psi^{-1}(P_{i_2}) \cap \cdots \cap \psi^{-m+1}(P_{i_m}) \neq \emptyset\}.$$

Set $F_i = \pi(P_i)$, $\mathcal{F} = \{F_i\}_{i\in I}$ and

$$N_m = \#\{(i_1, i_2, \cdots, i_m) \in I^m \ :\ F_{i_1} \cap \varphi^{-1}(F_{i_2}) \cap \cdots \cap \varphi^{-m+1}(F_{i_m}) \neq \emptyset\}.$$

Since $\pi(P_{i_1} \cap \psi^{-1}(P_{i_2}) \cap \cdots \cap \psi^{-m+1}(P_{i_m})) \subset F_{i_1} \cap \varphi^{-1}(F_{i_2}) \cap \cdots \cap \varphi^{-m+1}(F_{i_m})$, we find that $M_m \leq N_m$. Hence

$$h(\psi, \mathcal{P}) \leq \lim_{m\to\infty} \frac{1}{m} \log\ N_m. \tag{3.1}$$

We assert that

$$\lim_{m\to\infty} \frac{1}{m} \log\ N_m \leq h(\varphi) + Q(\varphi, \mathcal{F}). \tag{3.2}$$

To prove this, let d be a metric for the topology of X. Since $\mathcal{F}$ consists of closed sets, we can find, for each $x \in X$, an open neighbourhood $W(x)$ of x such that

$$F_i \cap W(x) \neq \emptyset \Rightarrow x \in F_i \tag{3.3}$$

for all $i \in I$. Let $k \in \mathbb{N}$ and set

$$U(x) = W(x) \cap \varphi^{-1}(W(\varphi(x))) \cap \varphi^{-2}(W(\varphi(x))) \cap \cdots \cap \varphi^{-k+1}(W(\varphi^{k-1}(x))).$$

Let $\mathcal{U}$ be a finite subcover of $\{U(x)\}_{x\in X}$ and let $\epsilon > 0$ be a Lebesgue number for $\mathcal{U}$, i.e. every ϵ-ball $B_\epsilon(z) = \{y \in X \ :\ d(y,z) < \epsilon\}$, $z \in X$, is contained in a member of $\mathcal{U}$. Let G: $X \to \mathcal{U}$ be a function such that $B_\epsilon(x) \subset G(x)$, $x \in X$. Let $m \in \mathbb{N}$ and let E be an (mk, ϵ)-spanning set of minimal cardinality; cf. [W], p. 168. Then $\#E = r_{km}(\epsilon, X)$ in the notation of [W]. Let S be the set of mk-tuples $(i_1, i_2, \dots, i_{mk}) \in I^{mk}$ for which there is an $e \in E$ such that

$$F_{i_{jk+1}} \cap \varphi^{-1}(F_{i_{jk+2}}) \cap \cdots \cap \varphi^{-k+1}(F_{i_{(j+1)k}}) \cap G(\varphi^{jk}(e)) \neq \emptyset, \tag{3.4}$$

$j = 0, 1, 2, \dots, m-1$. Fix some $e \in E$. There are elements $x_0, x_1, \dots, x_{m-1} \in X$ such that $G(\varphi^{jk}(e)) = U(x_j)$, $j = 0, 1, 2, \dots, m-1$. So if $(i_1, i_2, \dots, i_{mk}) \in I^{mk}$ satisfies (3.4) with respect to this e we have, because of (3.3), that

$$x_j \in F_{i_{jk+1}} \cap \varphi^{-1}(F_{i_{jk+2}}) \cap \cdots \cap \varphi^{-k+1}(F_{i_{(j+1)k}}),$$

$j = 1, 2, \ldots, m-1$. Therefore the number of tuples $(i_1, i_2, \ldots, i_{mk}) \in I^{mk}$ for which (3.4) holds for this specific $e \in E$ cannot exceed $q_k(\varphi, \mathcal{F})$, i.e. we have that $\#S \leq (\#E) \cdot q_k(\varphi, \mathcal{F})^m = r_{km}(\epsilon, X) \cdot q_k(\varphi, \mathcal{F})^m$. Now if $(i_1, i_2, \ldots, i_{mk}) \in I^{mk}$ is a tuple such that $F_{i_1} \cap \varphi^{-1}(F_{i_2}) \cap \cdots \cap \varphi^{-mk+1}(F_{i_{km}})$ is not empty and x is an element in the intersection, then $\varphi^{jk}(x) \in F_{i_{jk+1}} \cap \varphi^{-1}(F_{i_{jk+2}}) \cap \cdots \cap \varphi^{-k+1}(F_{i_{(j+1)k}})$. On the other hand, there is an element $e \in E$ such that $d(\varphi^i(x), \varphi^i(e)) < \epsilon$, $i = 0, 1, 2, \ldots, km-1$, since E is (mk, ϵ)-spanning. Hence $\varphi^{jk}(x) \in B_\epsilon(\varphi^{jk}(e)) \subset G(\varphi^{jk}(e))$ and $\varphi^{jk}(x) \in F_{i_{jk+1}} \cap \varphi^{-1}(F_{i_{jk+2}}) \cap \cdots \cap \varphi^{-k+1}(F_{i_{(j+1)k}}) \cap G(\varphi^{jk}(e)) \neq \emptyset$ for all $j = 0, 1, 2, \ldots, m-1$. This shows that $N_{km} \leq \#S$ so that $N_{km} \leq r_{mk}(\epsilon, X) \cdot q_k(\varphi, \mathcal{F})^m$. Consequently

$$\lim_{m\to\infty} \frac{1}{m} \log\, N_m = \lim_{m\to\infty} \frac{1}{km} \log\, N_{km} \leq \limsup_n \frac{1}{n} \log\, r_n(\epsilon, X) + \frac{1}{k} \log\, q_k(\varphi, \mathcal{F}).$$

Since $\limsup_n \frac{1}{n} \log\, r_n(\epsilon, X) \leq h(\varphi)$, cf. [W], we obtain (3.2) by letting k tend to infinity. By combining (3.1) and (3.2) we find that $h(\psi, \mathcal{P}) \leq h(\varphi) + Q(\varphi, \pi(\mathcal{P}))$. The proposition follows by taking the supremum over all $\mathcal{P}$. ∎

The preceding proof is an elaboration of an argument involved in one of the possible proofs of the variational principle for the topological entropy; cf. Lemma 18.2 of [DGS].

LEMMA 3.7: *Let π: $(Y, \psi) \to (X, \varphi)$ be a factor map between invertible dynamical systems with Y totally disconnected. For any $k \in \mathbb{N}$ the defect of π: $(Y, \psi) \to (X, \varphi)$ is the same as the defect of π: $(Y, \psi^k) \to (X, \varphi^k)$.*

Proof: To distinguish between the two cases in the notation we let $\pi^{(k)}$ denote π when the map is considered as a factor map $(Y, \psi^k) \to (X, \varphi^k)$. Then, by Theorem 2.1,

$$D(\pi) = \sup_{\mu \in M} \int_X \log\, \#\pi^{-1}(x) d\mu(x)$$

and

$$D(\pi^{(k)}) = \sup_{\mu \in M_k} \int_X \log\, \#\pi^{-1}(x) d\mu(x)$$

where M and M_k denote the set of φ-invariant, resp. φ^k-invariant, Borel probability measures. Since $M \subset M_k$ we have that $D(\pi) \leq D(\pi^{(k)})$. For each $\nu \in M_k$,

$$\nu' = \frac{1}{k} \sum_{j=0}^{k-1} \nu \circ \varphi^{-j}$$

is in M and

$$\int_X \log \#\pi^{-1}(x)d\nu'(x) = \frac{1}{k}\sum_{j=0}^{k-1}\int_X \log \#\pi^{-1}(\varphi^j(x))d\nu(x)$$
$$= \int_X \log \#\pi^{-1}(x)d\nu(x)$$

since $\#\pi^{-1}(\varphi^j(x)) = \#\pi^{-1}(x)$ for all j, x. Hence $D(\pi^{(k)}) \leq D(\pi)$. ∎

We can now give the

Proof of Theorem 3.5: Assume $D(\pi) < \infty$, and consider first the case where both ψ and φ are homeomorphisms. If $h(\psi) = \infty$ it follows from Lemma 3.6 that $h(\varphi) = \infty$. So assume that $h(\psi) < \infty$. Since $h(\psi^k) = kh(\psi)$ and $h(\varphi^k) = kh(\varphi)$ for all $k \in \mathbb{N}$, we get from Lemma 3.6 and Lemma 3.7 that

$$k(h(\psi) - h(\varphi)) \leq D(\pi) < \infty$$

for all $k \in \mathbb{N}$. This implies that $h(\psi) = h(\varphi)$. This argument does not extend to the non-invertible case because Lemma 3.7 fails in the general case, but we can extend the conclusion in the following way. Let $(\Sigma^\psi, \sigma_\psi)$ denote **the natural extension** of ψ, i.e.

$$\Sigma^\psi = \{(y_0, y_1, y_2, \dots) \in Y^{\mathbb{N}} : \ \psi(y_{n+1}) = y_n, n \geq 0\},$$

and σ_ψ: $\Sigma^\psi \to \Sigma^\psi$ is given by $\sigma_\psi((y_i)) = (\psi(y_i))$. The natural extension $(\Sigma^\varphi, \sigma_\varphi)$ is defined similarly. We can then define a factor map $\tilde{\pi}$: $(\Sigma^\psi, \sigma_\psi) \to (\Sigma^\varphi, \sigma_\varphi)$ by $\tilde{\pi}((y_i)) = (\pi(y_i))$. (The fact that $\tilde{\pi}$ is surjective follows from the following little compactness argument: Let $x = (x_i) \in \Sigma^\varphi$. For each $k \in \mathbb{N}$ choose $z_k \in Y$ such that $\pi(z_k) = x_k$. Let a be any point in Y and define $y^k \in Y^{\mathbb{N}}$ by $y_i^k = \psi^{k-i}(z_k), i \leq k, y_i^k = a, i > k$. Let y be a condensation point for the sequence $\{y_k\}$ in the compact metric space $Y^{\mathbb{N}}$. Then $y \in \Sigma^\psi$ and $\tilde{\pi}(y) = x$.) It follows from [T], Remark 1.10, that $D(\tilde{\pi}) \leq D(\pi)$. Hence $D(\tilde{\pi}) < \infty$, since we assume that $D(\pi) < \infty$; σ_ψ and σ_φ are homeomorphisms and we conclude from the first part of the proof that $h(\sigma_\psi) = h(\sigma_\varphi)$. By Proposition 5.2 of [B], $h(\psi) = h(\sigma_\psi) = h(\sigma_\varphi) = h(\varphi)$. ∎

The moral of Theorem 3.5 is that the defect is finite, and hence interesting, only when the topological entropy can't tell the involved dynamical systems apart. Very simple examples show that the reverse implication, $h(\psi) = h(\varphi) \Rightarrow D(\pi) < \infty$, is false in general. But in many interesting cases (e.g. when (Y, ψ) is an irreducible sofic system), the reversed implication is actually true. So in such cases the defect is interesting, as an invariant for factor maps, precisely when the entropy of the dynamical systems agree.

References

[B] R. Bowen, *Topological Entropy and Axiom A*, Proceedings of Symposia in Pure Mathematics **14** (1970), 43–50.

[DGS] M. Denker, C. Grillenberger and K. Sigmund, *Ergodic Theory on Compact Spaces*, Lecture Notes in Mathematics **527**, Springer-Verlag, Berlin, Heidelberg, New York, 1976.

[R] V. A. Rohlin, *On the fundamental ideas of measure theories*, Translations of the American Mathematical Society (Series 1) **10** (1962), 1–54.

[T] K. Thomsen, *The defect of factor maps*, Ergodic Theory and Dynamical Systems **17** (1997), 1233–1256.

[W] P. Walters, *An Introduction to Ergodic Theory*, Springer-Verlag, New York, Heidelberg, Berlin, 1982.