

## Fiber entropy and conditional variational principles in compact non-metrizable spaces

by

Tomasz Downarowicz and Jacek Serafin (Wrocław)

**Abstract.** We consider a pair of topological dynamical systems on compact Hausdorff (not necessarily metrizable) spaces, one being a factor of the other. Measure-theoretic and topological notions of fiber entropy and conditional entropy are defined and studied. Abramov and Rokhlin's definition of fiber entropy is extended, using disintegration. We prove three variational principles of conditional nature, partly generalizing some results known before in metric spaces: (1) the topological conditional entropy equals the supremum of the topological fiber entropy over the factor, which also equals the supremum of the topological fiber entropy given a measure over all invariant measures on the factor, (2) the topological fiber entropy given a measure equals the supremum of the measure-theoretic conditional entropy over all invariant measures on the larger system projecting to the given one. Combining the above, we get (3) the topological conditional entropy equals the supremum of the measure-theoretic conditional entropy over all invariant measures. A tail entropy of a measure is introduced in totally disconnected spaces. As an application of our variational principles it is proved that the tail entropy estimates from below the "defect of upper semicontinuity" of the entropy function.

**Introduction.** In this note we address some problems concerning the relationship between topological and measure-theoretic entropy, a subject which has gained a lot of attention in the study of dynamical systems. We begin with a short survey of the relevant results.

In 1969, Goodwyn [Gw] proved, assuming that the underlying phase space  $X$  is compact and metric, the inequality  $h_\nu(T) \leq h(T)$  between measure-theoretic and topological entropy. Later Goodman [Gm] showed that  $\sup_\nu h_\nu(T) \geq h(T)$  (the supremum is taken over all invariant measures), completing the variational principle.

---

2000 *Mathematics Subject Classification*: 37A35, 37B40.

*Key words and phrases*: conditional entropy, fiber entropy, topological conditional entropy, topological fiber entropy, conditional variational principle.

Supported by the KBN grant 2 P03A 04622 [2002-5 r.]

In the meantime Bowen [B] considered a factor map  $\pi$  between two compact metric dynamical systems  $(X, T)$  and  $(Y, S)$ , and formulated a version of topological conditional variational principle, proving that

$$h(T) \leq h(S) + \sup_{y \in Y} h(T, \pi^{-1}(y)),$$

where  $h(T, K)$  is the entropy of a compact subset of  $X$ .

In 1973 Ruelle [R] introduced the notion of pressure  $P(T, f)$  of a real-valued continuous function, which for  $f \equiv 0$  coincides with the notion of topological entropy. He formulated and proved a variational principle for pressure:  $\sup_{\nu} (h_{\nu}(T) + \int f d\nu) = P(T, f)$ , for  $T$  being an expansive homeomorphism satisfying a so-called specification property. Later, Walters [W1] proved the same statement under the sole assumption that  $T$  be continuous, extending the classical variational principle.

In 1976 Misiurewicz [M1] gave an elementary proof of the variational principle of Goodwyn and Goodman without the metrizable assumption, and versions of his proof (usually metric) became standard in many ergodic theory textbooks. In another paper [M2] Misiurewicz introduced the notions of topological conditional entropy with respect to a cover and the topological conditional entropy of a topological system, and connected those quantities with the “defect of upper semicontinuity” of the measure-theoretic entropy function defined on the space of invariant measures.

In 1977 Ledrappier and Walters [L-W] worked in the setting of Bowen and proved a relative variational principle in the form

$$\sup_{\nu} h_{\nu}(T) = h_{\mu}(S) + \int_Y h(T, \pi^{-1}(y)) d\mu(y),$$

where the supremum is taken over all invariant measures  $\nu$  projecting via  $\pi$  to  $\mu$ . Actually, they formulated a relative variational principle for the pressure with the above formula being a special case.

We should also mention more recent results: in 1997 Blanchard, Glasner and Host [B-G-H] proved a version of the variational principle valid for a fixed cover. Lately, some of their concepts have been applied in [D], where a close relation is shown between measure-theoretic conditional entropy and a topological notion of the so-called residual entropy introduced by M. Boyle. A forthcoming paper [B-F-F] with further connections between measure-theoretic and topological entropies has been announced to the authors during the preparation of this article.

The purpose of this note is to gather, enrich and unify the knowledge about the relations between several notions (some of them newly introduced) of measure-theoretic and topological fiber and conditional entropy. Both in measure-theoretic and topological setup, we consider a pair of dynamical systems, one being a factor of the other. Most attention is paid to the topo-

logical case, where both structures come into play. In our considerations we do not assume metrizable of the underlying spaces. Important examples of actions of a homeomorphism on nonmetrizable compact spaces arise naturally e.g. in the study of Ellis semigroups, universal flows on  $\beta\mathbb{Z}$ , etc. Nevertheless, most of our results seem to be new even in the metric context (see below).

In Section 1 we work exclusively in the measure-theoretic setup. Basing on the classical work of Abramov and Rokhlin [A-R] we introduce a new notion of fiber entropy of a partition (which turns out to be a nonconstant function on the factor space; we provide a simple example), and we prove that its integral is equal to the classical conditional entropy of that partition given the factor. This generalizes the result in [A-R] obtained for skew products and where the supremum over all partitions is taken.

In Section 2 (and later on) we turn to the topological situation. We define topological fiber entropy as a function on the factor space. This notion is similar to the integrand in the quoted result of [L-W]. It turns out that our function is constant almost everywhere on the supports of ergodic measures (earlier shown in [W2] in the metrizable case), which suggests that it should be more properly viewed as a function defined on the space of invariant measures on the factor. This leads to the notion of a topological fiber entropy given a measure, an analog of the integral in the result of [L-W].

Next, we define the topological conditional entropy of a system with respect to its factor. This is a new notion, which in some situations coincides with Misiurewicz's conditional entropy with respect to a cover.

In Section 3 we formulate and prove three variational principles. In the *outer variational principle* the topological conditional entropy is realized as the supremum over all invariant measures on the factor of the topological fiber entropies given these measures, or as the supremum over the factor space of the topological fiber entropy function, a term appearing in Bowen's inequality quoted above. We are not aware of any analog of this result in the metric case. We remark that one of the equalities in our assertion has already been applied in the study of relative Pinsker factors [Le-S].

The *inner variational principle* states that the topological fiber entropy given a measure  $\nu$  on the factor is equal to the supremum of the measure-theoretic conditional entropies over all invariant measures on the larger system projecting to  $\nu$ . This generalizes the above-mentioned result of Ledrappier and Walters to the nonmetrizable case.

Combining both inner and outer principles results in a conditional variational principle connecting directly the topological and measure-theoretic conditional entropies. Bowen's quoted inequality follows as a corollary. The authors of [B-F-F] prove a metric version of the conditional variational principle with an additional assumption of the factor system being asymptot-

ically  $h$ -expansive; their notion of the topological conditional entropy of a factor map is slightly different from our notion of the topological conditional entropy given a factor.

Due to a conflict in terminology, we call Misiurewicz's "topological conditional entropy of the system" the tail entropy. In Section 4 we discuss the case of a totally disconnected dynamical system, where, applying the outer variational principle, that notion can be expressed in terms of our topological fiber entropy. This leads naturally to a new concept of the tail entropy of an invariant measure. While the tail entropy of the system bounds globally from above the "defect of upper semicontinuity" of the entropy function (see [M2]), we easily prove, using the inner variational principle, that the tail entropy of a measure provides a pointwise lower bound.

For the convenience of the reader we include an Appendix in which we summarize all the necessary facts concerning upper semicontinuous functions, invariant Radon probability measures, and the like, which might not seem obvious without the metrizability assumption. We also give a straightforward and elementary proof of a version of Kingman's ergodic theorem for subadditive processes, which is used in Section 2. Some formulas of that proof are also applied in Section 3.

Because there are many different entropy functions appearing side by side in our note, we decided to slightly modify the classical notation. In order to easily distinguish between topological and measure-theoretic notions of entropy, the first are always denoted by boldface  $\mathbf{H}$  or  $\mathbf{h}$ , while we use upright  $H$  or  $h$  for the latter. To reduce the number of parameters, we never indicate the transformation, which remains fixed on a given space. It suffices to remember that capital  $H$  (in both settings) is an absolute entropy not referring to the action, while lower case  $h$  involves averaging along the orbits of sets. The measure, classically appearing as a subscript, is here most of the time treated as a variable, hence finds its place inside parentheses. A vertical bar inside parentheses separates objects referring to the larger system on its left from objects referring to the factor on its right. The only exception to this rule occurs at the beginning of Section 1, before the factor is introduced.

We thank Mariusz Lemańczyk for an inspiring discussion, and Zbigniew Lipecki for his comments regarding measure theory. We also thank the referee for valuable remarks.

**1. Measure-theoretic conditional and fiber entropy.** The first part of this section contains a short repetition on measure-theoretic entropy in dynamical systems, in which we partly acquaint the reader with our modified notation. Apart from recalling some classical definitions, we also prove a few more or less standard lemmas which we failed to find in the literature in the suitable form.

Let  $T$  denote a measurable transformation of a measurable space  $(X, \Sigma)$  to itself. The  $\sigma$ -field  $\Sigma$  will remain fixed and we will skip it in the notation of related objects. Let  $\mathcal{P}(X)$  denote the convex set of all probability measures on  $X$ . If  $\mu \in \mathcal{P}(X)$  then  $T\mu$  is given by the formula  $T\mu(A) = \mu(T^{-1}A)$  ( $A \in \Sigma$ ). By  $\mathcal{P}_T(X)$  we denote the convex subset of  $\mathcal{P}(X)$  consisting of all  $T$ -invariant measures. If  $\mu \in \mathcal{P}_T(X)$  then the triple  $(X, \mu, T)$  is called a *measure-theoretic dynamical system*.

Let  $\alpha$  and  $\beta$  denote finite measurable partitions of  $X$  (in what follows, we will skip the words “finite” and “measurable” since no other partitions will be considered). It is customary to write

$$\alpha \vee \beta := \{C = A \cap B : A \in \alpha, B \in \beta, C \neq \emptyset\}.$$

We say that  $\alpha$  is a *refinement* of  $\beta$  (and write  $\alpha \succ \beta$ ) if  $\alpha \vee \beta = \alpha$ . For a partition  $\alpha$  and an  $n \geq 1$  we set

$$\alpha^n := \bigvee_{i=0}^{n-1} T^{-i}\alpha.$$

Throughout this paper we write:

$$H(\mu, \alpha) := - \sum_{A \in \alpha} \mu(A) \log(\mu(A)),$$

$$H(\mu, \alpha | \beta) := H(\mu, \alpha \vee \beta) - H(\mu, \beta) = \sum_{B \in \beta} \mu(B) H(\mu_B, \alpha),$$

where  $\mu \in \mathcal{P}(X)$ , and  $\mu_B$  denotes the conditional measure induced by  $\mu$  on  $B$  (or zero if  $\mu(B) = 0$ ), the latter equality being well known. Using this equality and the fact that the real function  $\phi(x) := -x \log x$  is concave one easily shows that the function  $H(\cdot, \alpha | \beta)$  is concave on  $\mathcal{P}(X)$ . Another standard fact states that  $H(\mu, \alpha | \beta)$  increases with respect to  $\alpha$  and decreases with respect to  $\beta$ .

It is also known (see also Lemma 1 below) that if  $\mu$  is invariant then  $H(\mu, \alpha^n)$  is a subadditive sequence, and hence, when divided by  $n$  it converges to its infimum. The limit

$$h(\mu, \alpha) := \lim_{n \rightarrow \infty} \frac{1}{n} H(\mu, \alpha^n) = \inf_n \frac{1}{n} H(\mu, \alpha^n)$$

is the *entropy* of  $\mu$  with respect to  $\alpha$  (classically denoted by  $h_\mu(\alpha, T)$  and called “the entropy of  $\alpha$  with respect to the transformation  $T$ ”). Finally, the *entropy* of the measure  $\mu$  (usually called “the entropy of the measure-preserving transformation  $T$ ”) and denoted by  $h_\mu(T)$  is defined as

$$h(\mu) := \sup_{\alpha} h(\mu, \alpha),$$

where  $\alpha$  ranges over all (finite) partitions of  $X$ .

Before we continue, we introduce some abstract terminology. If  $T : X \rightarrow X$  is a map of any abstract set to itself then by a *process* we shall mean an arbitrary sequence  $(H_n)_{n \in \mathbb{N}}$  of nonnegative functions defined on  $X$ . A process is called *subadditive* if, for every  $x \in X$  and any natural  $m$  and  $n$ ,

$$H_{m+n}(x) \leq H_n(x) + H_m(T^n x).$$

LEMMA 1. *The process  $H(\mu, \alpha^n | \beta^n)$  defined on  $\mathcal{P}(X)$  is subadditive (with respect to  $T$  defined on measures).*

*Proof.* We have

$$\begin{aligned} H(\mu, \alpha^{m+n} | \beta^{m+n}) &= H(\mu, \alpha^{m+n} \vee \beta^{m+n}) - H(\mu, \beta^{m+n}) \\ &= H(\mu, \alpha^n \vee \beta^n | T^{-n} \alpha^m \vee T^{-n} \beta^m) \\ &\quad + H(\mu, T^{-n} \alpha^m \vee T^{-n} \beta^m) - H(\mu, \beta^n \vee T^{-n} \beta^m). \end{aligned}$$

We add and subtract  $H(\mu, T^{-n} \beta^m)$  in the last expression. The first term of this expression is clearly not larger than  $H(\mu, \alpha^n \vee \beta^n | T^{-n} \beta^m)$ , which combined with  $+H(\mu, T^{-n} \beta^m)$  produces

$$H(\mu, \alpha^n \vee \beta^n \vee T^{-n} \beta^m).$$

The second term combined with  $-H(\mu, T^{-n} \beta^m)$  produces

$$H(\mu, T^{-n} \alpha^m | T^{-n} \beta^m),$$

which we can write as  $H(T^n \mu, \alpha^m | \beta^m)$ . We thus obtain

$$\begin{aligned} H(\mu, \alpha^{m+n} | \beta^{m+n}) &\leq H(\mu, \alpha^n \vee \beta^n \vee T^{-n} \beta^m) + H(T^n \mu, \alpha^m | \beta^m) - H(\mu, \beta^n \vee T^{-n} \beta^m) \\ &= H(\mu, \alpha^n | \beta^n \vee T^{-n} \beta^m) + H(T^n \mu, \alpha^m | \beta^m) \\ &\leq H(\mu, \alpha^n | \beta^n) + H(T^n \mu, \alpha^m | \beta^m), \end{aligned}$$

and the subadditivity is proved. ■

As an immediate consequence of Lemma 1 and the relevant definitions we obtain:

COROLLARY 1. *If  $\mu \in \mathcal{P}_T(X)$  then  $H(\mu, \alpha^n | \beta^n)$  is a subadditive sequence and hence*

$$\inf_n \frac{1}{n} H(\mu, \alpha^n | \beta^n) = \lim_{n \rightarrow \infty} \frac{1}{n} H(\mu, \alpha^n | \beta^n) = h(\mu, \alpha \vee \beta) - h(\mu, \beta). \quad \blacksquare$$

Now let  $Y = (Y, \Theta)$  be another measurable space with a measurable map  $S : Y \rightarrow Y$ . We will assume that there exists a measurable surjection  $\pi : X \rightarrow Y$  such that  $\pi T = S \pi$ . In that case,  $\pi$  induces a map from  $\mathcal{P}(X)$  into  $\mathcal{P}(Y)$  (denoted by the same letter) by the formula  $(\pi \mu)(B) = \mu(\pi^{-1}(B))$  ( $B \in \Theta$ ). Clearly  $\pi(\mathcal{P}_T(X)) \subset \mathcal{P}_S(Y)$ . If  $\mu \in \mathcal{P}_T(X)$  and  $\nu = \pi \mu$  then we

say that  $(Y, \nu, S)$  is a measure-theoretic factor of  $(X, \mu, T)$  via the factor map  $\pi$ .

A partition  $\beta$  of  $X$  is said to be  $Y$ -measurable if  $\beta = \pi^{-1}\gamma$ , where  $\gamma$  is a partition of  $Y$ . Clearly, for any  $\mu \in \mathcal{P}(X)$  and  $\nu = \pi\mu$ , we have  $H(\mu, \beta) = H(\nu, \gamma)$ .

DEFINITION 1. Let  $(Y, \nu, S)$  be a measure-theoretic factor of  $(X, \mu, T)$ . For a partition  $\alpha$  of  $X$  we set

$$h(\mu, \alpha | \nu) := \inf_{\beta} \inf_n \frac{1}{n} H(\mu, \alpha^n | \beta),$$

where  $\beta$  ranges over all  $Y$ -measurable partitions of  $X$ . Furthermore, we set

$$h(\mu | \nu) := \sup_{\alpha} h(\mu, \alpha | \nu),$$

with  $\alpha$  ranging over all partitions of  $X$ . The latter is called the *conditional entropy of the system  $(X, \mu, T)$  given the factor  $(Y, \nu, S)$* .

LEMMA 2. We have

$$h(\mu | \nu) = \sup_{\alpha} \inf_n \inf_{\beta} \frac{1}{n} H(\mu, \alpha^n | \beta^n) = h(\mu) - h(\nu)$$

(for the last equality we assume that  $h(\nu) < \infty$ ).

*Proof.* We have

$$\begin{aligned} h(\mu | \nu) &= \sup_{\alpha} \inf_{\beta} \inf_n \frac{1}{n} H(\mu, \alpha^n | \beta) = \sup_{\alpha} \inf_n \inf_{\beta} \frac{1}{n} H(\mu, \alpha^n | \beta) \\ &= \sup_{\alpha} \inf_n \inf_{\beta} \frac{1}{n} H(\mu, \alpha^n | \beta^n) = \sup_{\alpha} \inf_{\beta} \inf_n \frac{1}{n} H(\mu, \alpha^n | \beta^n) \\ &= \sup_{\alpha} \inf_{\beta} (h(\mu, \alpha \vee \beta) - h(\mu, \beta)), \end{aligned}$$

with Corollary 1 applied in the last step. Next,

$$\sup_{\alpha} \inf_{\beta} (h(\mu, \alpha \vee \beta) - h(\mu, \beta)) \geq \sup_{\alpha} \inf_{\gamma} (h(\mu, \alpha) - h(\nu, \gamma)) = h(\mu) - h(\nu).$$

On the other hand, exchanging trivially supremum and infimum we get

$$\begin{aligned} \sup_{\alpha} \inf_{\beta} (h(\mu, \alpha \vee \beta) - h(\mu, \beta)) &\leq \inf_{\beta} \sup_{\alpha} (h(\mu, \alpha \vee \beta) - h(\mu, \beta)) \\ &= \inf_{\beta} (\sup_{\alpha} h(\mu, \alpha \vee \beta) - h(\mu, \beta)) \\ &= \inf_{\gamma} (\sup_{\alpha} h(\mu, \alpha) - h(\nu, \gamma)) = h(\mu) - h(\nu). \blacksquare \end{aligned}$$

The remaining part of this section is devoted to defining measure-theoretic fiber entropy as a function on the factor space  $Y$ . The original “fiber entropy” was introduced by Abramov and Rokhlin in skew products, and it was defined as a number, not a function. But if we drop the supremum over all partitions and the integral in their definition, we obtain a quite

interesting (nonconstant even in the ergodic case) function on the factor (see Example 1 below; we soon provide the details in a more general setup).

There are at least two ways of extending the notion of fiber entropy to a general dynamical system  $(X, \mu, T)$  and its factor  $(Y, \nu, S)$ . Firstly, we can try to isomorphically represent the system  $(X, \mu, T)$  as a skew product with  $(Y, \nu, S)$  in the base. This is possible in standard spaces if both  $T$  and  $S$  are ergodic and invertible (see [Ko]; the statement in [A-R] without invertibility is false). The second approach (which we shall exploit) is to define fiber entropy using a disintegration of a measure:

DEFINITION 2. Let  $\pi : (X, \Sigma, \mu) \rightarrow (Y, \Theta, \nu)$  be a homomorphism of measure spaces. By a *disintegration of  $\mu$  with respect to  $\nu$*  we mean a family of measures  $\mu_y$  defined for  $\nu$ -a.e.  $y \in Y$ , each supported by  $\pi^{-1}y$ , such that, for every bounded measurable function  $f$  on  $X$ , the function  $\bar{f}(y) := \int f(x) d\mu_y(x)$  is measurable, and

$$\int \bar{f}(y) d\nu(y) = \int f(x) d\mu(x).$$

Additionally, if  $\pi$  is a factor map between dynamical systems  $(X, \mu, T)$  and  $(Y, \nu, S)$  and  $\mu_{Sy} = T\mu_y$   $\nu$ -a.e. then we say that the disintegration is *equivariant*.

In the next theorem we will assume that there exists an equivariant disintegration of  $\mu$  with respect to  $\nu$ . Obviously, any skew product satisfies this assumption. In general, for the existence of a disintegration we need the measure space  $X$  to be compact in the sense of Marczewski (see [P]); this includes the case of Radon measures on compact Hausdorff spaces, on which most of our attention will be focused. If in addition the action of  $S$  on  $Y$  is invertible (this restriction does not apply to the next sections of this work), then the disintegration is equivariant (compare Proposition 5.9 in [F] for the metrizable case). Without the invertibility assumption the existence of an equivariant disintegration is not guaranteed.

The following definition is a direct generalization of a function appearing in [A-R] in the definition of the fiber entropy in skew products:

DEFINITION 3. Let  $\alpha$  be a finite partition of  $X$ . By the *fiber entropy of  $\alpha$  with respect to the measure  $\mu$*  we mean the function  $h(\mu, \alpha | \cdot)$  defined  $\nu$ -almost everywhere on  $Y$  by the formula

$$h(\mu, \alpha | y) := \lim_n H(\mu_y, \alpha | T^{-1}\alpha^n).$$

Note that for fixed  $y$  the above sequence is nonincreasing, hence its limit is the same as the infimum.

Translating to our notation, Abramov and Rokhlin have proved that (in case  $h(\nu) < \infty$ )

$$\begin{aligned} \sup_{\alpha} \int h(\mu, \alpha | y) d\nu(y) &= h(\mu) - h(\nu) \\ &= (\sup_{\alpha} h(\mu, \alpha | \nu) \text{ by our Lemma 2 and Definition 1}). \end{aligned}$$

Below we strengthen the Abramov–Rokhlin theorem by showing that it also holds for a fixed partition  $\alpha$  (i.e., that  $\sup_{\alpha}$  can be dropped on both sides).

**THEOREM 1.** *If  $(Y, \nu, S)$  is a factor of  $(X, \mu, T)$  via a map  $\pi$  such that there exists an equivariant disintegration of  $\mu$  with respect to  $\nu$ , then, for every partition  $\alpha$  of  $X$ ,*

$$\int h(\mu, \alpha | y) d\nu(y) = h(\mu, \alpha | \nu).$$

*Proof.* We have

$$\begin{aligned} \int h(\mu, \alpha | y) d\nu(y) &= \int \lim_n H(\mu_y, \alpha | T^{-1}\alpha^n) d\nu(y) \\ &= \lim_n \int [H(\mu_y, \alpha^{n+1}) - H(\mu_y, T^{-1}\alpha^n)] d\nu(y) \\ &= \lim_n \left[ \int H(\mu_y, \alpha^{n+1}) d\nu(y) - \int H(\mu_y, \alpha^n) d\nu(y) \right], \end{aligned}$$

using the monotone convergence theorem, equivariance of the disintegration and invariance of the measure  $\nu$ . The limit of the last sequence is the same as the limit of its averages, which, after cancellation, reads

$$\lim_n \frac{1}{n} \int H(\mu_y, \alpha^n) d\nu(y).$$

By Lemma 1, equivariance and invariance, the above integrals form a sub-additive sequence, hence we can replace the limit by the infimum. It remains to show that

$$\int H(\mu_y, \alpha^n) d\nu(y) = \inf_{\beta} H(\mu, \alpha^n | \beta).$$

Each partition  $\beta$  is obtained as the preimage of a partition  $\gamma$  of  $Y$ . After applying all relevant definitions and integrating on each set  $D \in \gamma$  with respect to the conditional measure  $\nu_D$  induced by  $\nu$  on  $D$ , the left hand side becomes

$$- \sum_{A \in \alpha^n} \sum_{D \in \gamma} \nu(D) \int \mu_y(A) \log(\mu_y(A)) d\nu_D(y),$$

which we write as

$$\sum_{A \in \alpha^n} \sum_{D \in \gamma} \nu(D) \int \phi(\mu_y(A)) d\nu_D(y),$$

where  $\phi(x) := -x \log x$  (a concave function). Viewing the last integral as an average value of  $\phi$ , we see that it is not larger than the value of  $\phi$  at the average of the arguments. But this latter average equals

$$\begin{aligned} \int \mu_y(A) d\nu_D(y) &= \frac{1}{\nu(D)} \int_Y \int_{\pi^{-1}(y)} \mathbf{1}_{A \cap B}(x) d\mu_y(x) d\nu(y) \\ &= \frac{\mu(A \cap B)}{\mu(B)} = \mu_B(A), \end{aligned}$$

by the definition of the disintegration, with  $B := \pi^{-1}(D) \in \beta$  (hence  $\mu(B) = \nu(D)$ ). Consequently, the last double sum is not larger than

$$\sum_{A \in \alpha^n} \sum_{B \in \beta} \mu(B) \phi(\mu_B(A)),$$

i.e., than  $H(\mu, \alpha^n | \beta)$ .

For the opposite inequality it suffices to notice that, by uniform continuity of  $\phi$  on the interval  $[0, 1]$ , the average value of  $\phi$  does not differ from the value at the average by more than  $\varepsilon/\#\alpha^n$  if the arguments stay in an interval of an appropriately chosen length  $\delta$ . Thus we need to find a finite partition  $\gamma$  of  $Y$  such that, for every  $D \in \gamma$  and  $A \in \alpha^n$ , the values of  $\mu_y(A)$  with  $y$  ranging over  $D$  do not vary by more than  $\delta$ . But such a partition is easily constructed as

$$\gamma := \bigvee_{A \in \alpha^n} \gamma_A,$$

where  $\gamma_A$  is the preimage under the measurable function  $\bar{\mathbf{I}}_A : y \mapsto \mu_y(A)$  of the partition of  $[0, 1]$  into intervals of length  $\delta$ . We now have

$$\begin{aligned} H(\mu, \alpha^n | \beta) - \int H(\mu_y, \alpha^n) d\nu(y) &= \sum_{A \in \alpha^n} \sum_{D \in \gamma} \nu(D) \left[ \phi(\mu_B(A)) - \int \phi(\mu_y(A)) d\nu_D(y) \right] \\ &\leq \sum_{A \in \alpha^n} \sum_{D \in \gamma} \nu(D) \frac{\varepsilon}{\#\alpha^n} \leq \varepsilon. \quad \blacksquare \end{aligned}$$

EXAMPLE 1. Let  $(Y, \nu, S)$  be the two-point periodic cycle on  $Y = \{a, b\}$  (here  $\nu(\{a\}) = \nu(\{b\}) = 1/2$ ), and  $(Z, \lambda, S)$  be the  $(1/2, 1/2)$  Bernoulli shift on two symbols  $\{0, 1\}$ . Define  $(X, \mu, T)$  as the skew product  $T(a, z) = (b, z)$ ,  $T(b, z) = (a, Sz)$ , with  $\mu = \nu \times \lambda$ . Let  $\alpha$  be a three-set generating partition:  $\{a\} \times Z$ ,  $\{b\} \times Z_0$  and  $\{b\} \times Z_1$ , where  $Z_0$  and  $Z_1$  denote the 0-coordinate cylinders in  $Z$ . Obviously  $h(\mu | \nu) = h(\mu) = (\log 2)/2$ . Clearly,  $\nu$  is an ergodic measure, thus any invariant function on  $Y$  is constant. On the other hand, it is immediately seen from the definition that the fiber entropy function is not constant on  $Y$ :  $h(\mu, \alpha | a) = 0$  and  $h(\mu, \alpha | b) = \log 2$ .

**2. Topological fiber entropy and conditional entropy.** Our goal is to introduce topological analogs of the notions appearing in the previous section. We choose to identify topological fiber entropy with what is known

as the “topological entropy of a fiber” in the sense of Bowen. The function on the factor space  $Y$  defined in this way is constant almost everywhere on the supports of ergodic measures (see Theorem 2 below), and in this sense it is not a perfect analog of the measure-theoretic fiber entropy of a partition. Attempts to improve the notion have given rather artificial results. It seems more appropriate to keep the definition, but view the topological fiber entropy as a function defined on the set of measures on the factor space. In this setting it becomes an analog of the measure-theoretic conditional entropy, and in the next sections we will prove strong relations between these notions. On the other hand, topological conditional entropy has yet another meaning, as a value not depending on the measures. Later, we will also show how that constant is related to topological fiber entropy (and through this to measure-theoretic conditional entropy).

STANDING ASSUMPTION. In the remainder of this note we shall be interested in the following case:  $(X, \Sigma)$  and  $(Y, \Theta)$  are compact Hausdorff spaces with  $\sigma$ -fields of Borel sets, both maps  $T : X \rightarrow X$  and  $S : Y \rightarrow Y$  are continuous (not necessarily onto). The factor map (surjection)  $\pi : X \rightarrow Y$  is also continuous. We then say that  $(X, T)$  and  $(Y, S)$  are *topological dynamical systems* and that  $(Y, S)$  is a *topological factor of  $(X, T)$  via the factor map  $\pi$* .

In such a case we restrict the meaning of  $\mathcal{P}(X)$ ,  $\mathcal{P}(Y)$ ,  $\mathcal{P}_T(X)$ , and  $\mathcal{P}_S(Y)$  to Radon (Borel regular) probability measures. Endowed with the weak\* topology, all these sets become nonempty compact Hausdorff spaces, and the maps  $T$ ,  $S$  and  $\pi$  defined appropriately on measures become continuous. It is important that  $\pi(\mathcal{P}_T(X)) = \mathcal{P}_S(Y)$  (use e.g. the Hahn–Banach theorem and an averaging technique).

From now on, the letters  $A, A', A_n$ , etc. will denote open sets in  $X$ , while script letters  $\mathcal{A}, \mathcal{A}'$ , etc. will be used for finite open covers of  $X$ . For covers we will skip the words “finite”, “open”, and “of  $X$ ”, as no other covers will be considered.

We say that a cover  $\mathcal{A}'$  is *inscribed* in  $\mathcal{A}$  (and write  $\mathcal{A}' \preceq \mathcal{A}$ ) if each  $A' \in \mathcal{A}'$  is contained in some  $A \in \mathcal{A}$ . The joinings of covers,  $\mathcal{A} \vee \mathcal{A}'$  and  $\mathcal{A}^n$ , are defined by identical formulas as for partitions. Clearly,  $\mathcal{A}^{m+n} = \mathcal{A}^m \vee T^{-m}\mathcal{A}^n$  for any  $m, n \in \mathbb{N}$ .

If  $\mathcal{F}$  is any family of sets then  $\bigcup \mathcal{F}$  denotes the union of all members of  $\mathcal{F}$ .

For a cover  $\mathcal{A}$  and a set  $K \subset X$ , we define

$$\mathbf{N}(\mathcal{A} | K) := \min\{\#\mathcal{F} : \mathcal{F} \subset \mathcal{A}, \bigcup \mathcal{F} \supset K\}.$$

(compare  $\mathbf{N}(B, \mathcal{A})$  in [M2] for an open set  $B$ ).

With a cover  $\mathcal{A}$  temporarily fixed we define a series of quantities (conditional in nature) leading eventually to the notion of topological fiber entropy.

DEFINITION 4. For  $y \in Y$  set

$$\mathbf{N}(\mathcal{A} | y) := \mathbf{N}(\mathcal{A} | \pi^{-1}y) \quad \text{and} \quad \mathbf{H}(\mathcal{A} | y) := \log \mathbf{N}(\mathcal{A} | y).$$

Since the partition of  $X$  into preimages of points is *upper semicontinuous* (u.s.c.; see Appendix, (A1)), it is not hard to see that the above are u.s.c. functions on  $Y$ . Clearly, if  $\mathcal{A}' \succcurlyeq \mathcal{A}$  then  $\mathbf{H}(\mathcal{A}' | y) \geq \mathbf{H}(\mathcal{A} | y)$ . We also easily verify that  $\mathbf{H}(T^{-1}\mathcal{A} | y) \leq \mathbf{H}(\mathcal{A} | Sy)$  (compare inequality (1.5) in [M2]).

LEMMA 3. For any two covers  $\mathcal{A}, \mathcal{B}$  we have  $\mathbf{H}(\mathcal{A} \vee \mathcal{B} | y) \leq \mathbf{H}(\mathcal{A} | y) + \mathbf{H}(\mathcal{B} | y)$ . The sequence  $(\mathbf{H}(\mathcal{A}^n | \cdot))_n$  of functions defined on  $Y$  is a *subadditive process*.

*Proof.* Let  $\mathcal{F}_1$  and  $\mathcal{F}_2$  be minimal cardinality covers of  $\pi^{-1}y$  by members of  $\mathcal{A}$  and  $\mathcal{B}$ , respectively. Clearly,  $\mathcal{F} = \mathcal{F}_1 \vee \mathcal{F}_2$  is a cover of  $\pi^{-1}y$  by members of  $\mathcal{A} \vee \mathcal{B}$ , and its cardinality is not larger than  $\#\mathcal{F}_1\#\mathcal{F}_2$ . Hence

$$\mathbf{N}(\mathcal{A} \vee \mathcal{B} | y) \leq \#\mathcal{F}_1\#\mathcal{F}_2 = \mathbf{N}(\mathcal{A} | y)\mathbf{N}(\mathcal{B} | y),$$

and the first assertion follows. To prove the second statement we note that  $\mathcal{A}^{m+n} = \mathcal{A}^m \vee T^{-m}\mathcal{A}^n$ , then apply the first statement of the lemma and the inequality  $\mathbf{H}(T^{-m}\mathcal{A}^n | y) \leq \mathbf{H}(\mathcal{A}^n | S^m y)$ . ■

We can continue with further definitions.

DEFINITION 5. If  $\nu$  is a Radon probability measure on  $Y$  then we set

$$\mathbf{H}(\mathcal{A} | \nu) := \int \mathbf{H}(\mathcal{A} | y) d\nu.$$

The function  $\mathbf{H}(\mathcal{A} | \cdot)$  defined as above is upper semicontinuous on the set  $\mathcal{P}(Y)$  (see Appendix, (A8)). Note that

$$0 \leq \frac{1}{n}\mathbf{H}(\mathcal{A}^n | \cdot) \leq a,$$

where  $a := \log \#\mathcal{A}$  (this notation will be used throughout the paper).

DEFINITION 6. The *topological fiber entropy of the cover  $\mathcal{A}$  given  $y$*  equals

$$\mathbf{h}(\mathcal{A} | y) := \limsup_{n \rightarrow \infty} \frac{1}{n}\mathbf{H}(\mathcal{A}^n | y)$$

(compare  $h(T, \pi^{-1}y, \mathcal{A})$  in [M2]).

From our next theorem (see Theorem 2 below) it will follow that at “most” of the points  $y$  in the above definition we have in fact convergence. The following statement is an easy consequence of the definition of  $\mathbf{h}(\mathcal{A} | y)$  and the proof will be omitted.

LEMMA 4. If  $\mathcal{A}, \mathcal{A}'$  are covers such that  $\mathcal{A}^n \succcurlyeq \mathcal{A}'$  for some  $n \geq 1$ , then

$$\mathbf{h}(\mathcal{A} | y) \geq \mathbf{h}(\mathcal{A}' | y). \quad \blacksquare$$

By Lemma 3, for any invariant measure  $\nu$  on  $Y$ , the sequence  $\mathbf{H}(\mathcal{A}^n | \nu)$  is subadditive. Thus  $n^{-1}\mathbf{H}(\mathcal{A}^n | \nu)$  converges to its infimum.

DEFINITION 7. If  $\nu$  is a probability measure on  $Y$  then we define

$$\mathbf{h}(\mathcal{A} | \nu) := \inf_n \frac{1}{n} \mathbf{H}(\mathcal{A}^n | \nu) \quad (= \lim_n \frac{1}{n} \mathbf{H}(\mathcal{A}^n | \nu) \text{ for } \nu \text{ invariant}).$$

Note that  $\mathbf{h}(\mathcal{A} | \nu)$  is still a u.s.c. function on  $\mathcal{P}(Y)$  (see Appendix, (A5)). Although  $\mathbf{H}(\mathcal{A}^n | y) = \mathbf{H}(\mathcal{A}^n | \delta_y)$  for any  $y \in Y$  and  $n \in \mathbb{N}$  (here  $\delta_y$  denotes the Dirac measure at  $y$ ), we must not confuse  $\mathbf{h}(\mathcal{A} | y)$  with  $\mathbf{h}(\mathcal{A} | \delta_y)$ , of which the latter may be strictly smaller.

REMARK 1. It follows easily from Definitions 5 and 7 that, for any positive integer  $m$ , the quantity analogous to  $\mathbf{h}(\mathcal{A} | \nu)$  evaluated for the transformation  $T^m$  is equal to  $m\mathbf{h}(\mathcal{A} | \nu)$ .

REMARK 2. If  $\nu \in \mathcal{P}_S(Y)$  then  $\mathbf{h}(T^{-1}\mathcal{A} | \nu) = \mathbf{h}(\mathcal{A} | \nu)$ . Indeed, from the inequality  $\mathbf{H}(T^{-1}\mathcal{A} | y) \leq \mathbf{H}(\mathcal{A} | Sy)$  and Lemma 3 we deduce that

$$\begin{aligned} \mathbf{H}(T^{-1}\mathcal{A}^n | y) &\leq \mathbf{H}(\mathcal{A}^n | Sy) = \mathbf{H}(\mathcal{A} \vee T^{-1}\mathcal{A}^{n-1} | Sy) \\ &\leq \mathbf{H}(\mathcal{A} | Sy) + \mathbf{H}(T^{-1}\mathcal{A}^{n-1} | Sy). \end{aligned}$$

Integrating with respect to an invariant measure  $\nu$  on  $Y$ , dividing by  $n$ , and letting  $n \rightarrow \infty$  proves the assertion.

As an application of Lemma 3 and the subadditive ergodic theorem (see e.g. [Kr], Theorem 5.3; see also Appendix, (A13), for a direct proof) we obtain the following

THEOREM 2 (cf. Corollary 1.2 in [W2]). *If  $\tilde{\nu}$  is an ergodic measure on  $Y$  then*

$$\mathbf{h}(\mathcal{A} | y) = \lim_n \frac{1}{n} \mathbf{H}(\mathcal{A}^n | y) = \mathbf{h}(\mathcal{A} | \tilde{\nu}) \quad \tilde{\nu}\text{-a.e. } \blacksquare$$

Let  $Y'$  denote the set of points  $y \in Y$  such that the limit  $\lim_n n^{-1}\mathbf{H}(\mathcal{A}^n | y)$  exists and belongs to the closure of the set of values  $\mathbf{h}(\mathcal{A} | \tilde{\nu})$  for all ergodic measures  $\tilde{\nu}$ . Then  $Y'$  is a Borel set, and  $\nu(Y') = 1$  for any invariant measure  $\nu$  (see Appendix, (A12)). Thus, by the bounded convergence theorem, for any such  $\nu$  we have

COROLLARY 2.

$$\mathbf{h}(\mathcal{A} | \nu) = \int \mathbf{h}(\mathcal{A} | y) d\nu.$$

*In particular,  $\mathbf{h}(\mathcal{A} | \cdot)$  is an affine function on  $\mathcal{P}_S(Y)$ .  $\blacksquare$*

So far the cover  $\mathcal{A}$  was fixed. We will now pass to a limit over  $\mathcal{A}$  becoming more and more fine.

DEFINITION 8. We set

$$\mathbf{h}(X | y) := \sup_{\mathcal{A}} \mathbf{h}(\mathcal{A} | y), \quad \mathbf{h}(X | \nu) := \sup_{\mathcal{A}} \mathbf{h}(\mathcal{A} | \nu),$$

where  $\mathcal{A}$  ranges over all finite covers, and  $\nu$  is a probability measure on  $Y$ . The above quantities will be called the *topological fiber entropy of  $X$  given  $y$* , and *given  $\nu$* , respectively. Of course, infinity is admitted in both cases.

REMARK 3. By Remark 1, the quantities analogous to  $\mathbf{h}(X|y)$  and  $\mathbf{h}(X|\nu)$  evaluated for the transformation  $T^m$  ( $m \in \mathbb{N}$ ) are equal to  $m\mathbf{h}(X|y)$  and  $m\mathbf{h}(X|\nu)$ , respectively.

As a consequence of Lemma 4 we note that in metric spaces the above suprema can be realized as monotone limits over a fixed countable increasing sequence  $(\mathcal{A}_i)_{i \in \mathbb{N}}$  of covers, thus the function  $\mathbf{h}(X|\cdot)$  is Borel-measurable (II Baire class) both on  $Y$  and on  $\mathcal{P}(Y)$ , and

$$\mathbf{h}(X|\nu) = \int \mathbf{h}(X|y) d\nu$$

for any invariant measure  $\nu$ , while

$$\mathbf{h}(X|y) = \mathbf{h}(X|\tilde{\nu})$$

$\tilde{\nu}$ -a.e. for an ergodic measure  $\tilde{\nu}$ .

Warning: In nonmetrizable spaces the above functions may not be measurable (hence the above integral is not defined). Also, we no longer control the size of the set where the last equality holds. Nevertheless, we will be able to prove variational principles concerning these functions.

Recall that a cover  $\mathcal{A}$  is called a *generator* of  $(X, T)$  if  $\bigcup_n \mathcal{A}^n$  is a basis for the topology. It also follows from Lemma 4 that if a cover  $\mathcal{A}$  is a generator of  $(X, T)$  then it realizes the suprema in Definition 8.

We now proceed with defining the topological conditional entropy.

DEFINITION 9. Let

$$\mathbf{N}(\mathcal{A}|Y) := \sup_{y \in Y} \mathbf{N}(\mathcal{A}|y), \quad \mathbf{H}(\mathcal{A}|Y) := \log \mathbf{N}(\mathcal{A}|Y) = \sup_{y \in Y} \mathbf{H}(\mathcal{A}|y).$$

From Lemma 3 it follows that  $\mathbf{H}(\mathcal{A}^n|Y)$  is a subadditive sequence. Thus we can introduce the following:

DEFINITION 10. Set

$$\mathbf{h}(\mathcal{A}|Y) := \inf_n \frac{1}{n} \mathbf{H}(\mathcal{A}^n|Y) = \lim_{n \rightarrow \infty} \frac{1}{n} \mathbf{H}(\mathcal{A}^n|Y).$$

DEFINITION 11. The *topological conditional entropy of  $(X, T)$  given the factor  $(Y, S)$*  is defined by

$$\mathbf{h}(X|Y) := \sup_{\mathcal{A}} \mathbf{h}(\mathcal{A}|Y).$$

REMARK 4. We recall that M. Misiurewicz [M2] defines the topological conditional entropy of a cover  $\mathcal{A}$  and of the system given a cover  $\mathcal{B}$  by  $\mathbf{h}(\mathcal{A}|\mathcal{B}) := \inf_n n^{-1} \log \mathbf{N}(\mathcal{A}^n|\mathcal{B}^n)$  and  $\mathbf{h}(X|\mathcal{B}) := \sup_{\mathcal{A}} \mathbf{h}(\mathcal{A}|\mathcal{B})$ , respectively, where  $\mathbf{N}(\mathcal{A}|\mathcal{B}) := \sup_{B \in \mathcal{B}} \mathbf{N}(\mathcal{A}|B)$ . It is elementary to see that if  $\mathcal{B}$

is a sufficiently fine  $Y$ -measurable cover (preimage under  $\pi$  of a cover of  $Y$ ), then our  $\mathbf{N}(\mathcal{A} | Y)$  is equal to Misiurewicz's  $\mathbf{N}(\mathcal{A} | \mathcal{B})$ . Thus

$$\begin{aligned} \mathbf{h}(\mathcal{A} | Y) &= \inf_n \frac{1}{n} \mathbf{H}(\mathcal{A}^n | Y) = \inf_n \frac{1}{n} \inf_{\mathcal{B}} \log \mathbf{N}(\mathcal{A}^n | \mathcal{B}) \\ &= \inf_n \frac{1}{n} \inf_{\mathcal{B}} \log \mathbf{N}(\mathcal{A}^n | \mathcal{B}^n) = \inf_{\mathcal{B}} \mathbf{h}(\mathcal{A} | \mathcal{B}), \end{aligned}$$

where  $\mathcal{B}$  ranges over all  $Y$ -measurable covers. Taking the supremum over all covers  $\mathcal{A}$ , we obtain

$$\mathbf{h}(X | Y) = \sup_{\mathcal{A}} \inf_{\mathcal{B}} \mathbf{h}(\mathcal{A} | \mathcal{B}).$$

If  $\mathcal{B}$  is a generator for  $(Y, S)$  then it realizes the infimum, hence

$$\mathbf{h}(X | Y) = \mathbf{h}(X | \mathcal{B}).$$

REMARK 5. In [B-F-F] a different notion is considered:

$$\mathbf{e}^*(\pi) := \inf_{\mathcal{B}} \sup_{\mathcal{A}} \mathbf{h}(\mathcal{A} | \mathcal{B}).$$

The relation between that notion and our topological conditional entropy is the following:

$$\mathbf{h}(X | Y) \leq \mathbf{e}^*(\pi) \leq \mathbf{h}(X | Y) + \mathbf{h}^*(Y),$$

where  $\mathbf{h}^*(Y)$  denotes Misiurewicz's tail entropy of the factor system (see Introduction for terminology). The first inequality is obvious, and the second follows from the standard inequality  $\mathbf{h}(\mathcal{A} | \mathcal{B}) \leq \mathbf{h}(\mathcal{A} | \mathcal{B}') + \mathbf{h}(\mathcal{B}' | \mathcal{B})$  (see [M2], (1.17)) by passing to the limit with  $\mathcal{B}'$ ,  $\mathcal{A}$ , and  $\mathcal{B}$ , consecutively.

**3. Conditional variational principles.** This section contains the main results of our paper. We will connect the notions of topological fiber entropy, topological conditional entropy and measure-theoretic conditional entropy.

THEOREM 3 (Outer variational principle). *Let  $(Y, S)$  be a topological factor of  $(X, T)$ . Then*

$$\mathbf{h}(X | Y) = \sup_{y \in Y} \mathbf{h}(X | y) = \sup_{\nu} \mathbf{h}(X | \nu) = \sup_{\tilde{\nu}} \mathbf{h}(X | \tilde{\nu}),$$

where the last two suprema are taken over all invariant and all ergodic Radon probability measures on  $Y$ , respectively.

*Proof.* This proof relies largely on exchanging suprema and infima.

For the first inequality “ $\geq$ ” we need to show that

$$\sup_{\mathcal{A}} \inf_n \sup_y \frac{1}{n} \mathbf{H}(\mathcal{A}^n | y) \geq \sup_y \sup_{\mathcal{A}} \limsup_{n \rightarrow \infty} \frac{1}{n} \mathbf{H}(\mathcal{A}^n | y).$$

Since the infimum on the left is in fact a limit (hence also lim sup), the above follows trivially from basic properties of suprema. The next inequality “ $\geq$ ” is also elementary: by Corollary 2,

$$\sup_y \mathbf{h}(\mathcal{A} | y) \geq \sup_\nu \mathbf{h}(\mathcal{A} | \nu)$$

for every cover  $\mathcal{A}$ , so we can apply  $\sup_{\mathcal{A}}$  on both sides, and then exchange the suprema.

Next, we show that  $\sup_\nu \mathbf{h}(X | \nu) \geq \mathbf{h}(X | Y)$ , i.e.

$$\sup_\nu \sup_{\mathcal{A}} \inf_n \frac{1}{n} \mathbf{H}(\mathcal{A}^n | \nu) \geq \sup_{\mathcal{A}} \inf_n \sup_y \frac{1}{n} \mathbf{H}(\mathcal{A}^n | y)$$

with the first supremum taken over all invariant measures. Since we can exchange the suprema on the left hand side, it remains to show that, for a fixed cover  $\mathcal{A}$ ,

$$\sup_\nu \inf_n \frac{1}{n} \mathbf{H}(\mathcal{A}^n | \nu) \geq \inf_n \sup_y \frac{1}{n} \mathbf{H}(\mathcal{A}^n | y) =: Q.$$

Suppose that the left hand side of the last inequality is smaller than  $Q$  minus some  $\varepsilon$ . Recall that on the compact set  $\mathcal{P}_S(Y)$  we have

$$\inf_n \frac{1}{n} \mathbf{H}(\mathcal{A}^n | \nu) = \mathbf{h}(\mathcal{A} | \nu),$$

and, by subadditivity, this infimum is realized as a monotone limit of a subsequence indexed by  $(n_k)$  with each  $n_{k+1}$  being a multiple of  $n_k$ . Since each  $\mathbf{H}(\mathcal{A}^n | \cdot)$  is a u.s.c. function, we conclude that

$$\frac{1}{n} \mathbf{H}(\mathcal{A}^n | \nu) < Q - \varepsilon$$

for some positive integer  $n$  and all invariant measures  $\nu$  (see Appendix, (A6)). Further, by upper semicontinuity of  $\mathbf{H}(\mathcal{A}^n | \cdot)$  on the set  $\mathcal{P}(Y)$  of all probability measures, the same holds on some neighborhood  $U$  of  $\mathcal{P}_S(Y)$ . Thus there exists an  $m_0$  such that the above inequality is valid for all measures of the form

$$\nu = M_m(\delta_y) := \frac{1}{m} \sum_{i=0}^{m-1} \delta_{S^i y},$$

for any  $m$  being a multiple of  $m_0$  and  $y \in Y$  (see Appendix, (A9)). We choose the  $m$  so large that  $2an/(m+n) \leq \varepsilon/2$ . From the inequality (A15) in the Appendix, we deduce that

$$\frac{1}{m+n} \mathbf{H}(\mathcal{A}^{m+n} | y) \leq \frac{1}{n} \mathbf{H}(\mathcal{A}^n | M_m(\delta_y)) + \frac{\varepsilon}{2} \leq Q - \frac{\varepsilon}{2}$$

for any  $y$ , that is, we can apply  $\sup_y$  on the left hand side. But by definition,  $Q \leq \sup_y (m+n)^{-1} \mathbf{H}(\mathcal{A}^{m+n} | y)$ , a contradiction.

To see that the supremum over all ergodic measures  $\tilde{\nu}$  is not too small (it is obviously not too large) we use the following argument: Fix a cover  $\mathcal{A}$ . Let  $Y'$  be as defined before Corollary 2. Now, by that corollary, for any invariant measure  $\nu$ ,  $\mathbf{h}(\mathcal{A} | \nu)$  cannot be larger than  $\sup\{\mathbf{h}(\mathcal{A} | y) : y \in Y'\}$ , i.e., than  $\sup_{\tilde{\nu}} \mathbf{h}(\mathcal{A} | \tilde{\nu})$ . We have shown that

$$\sup_{\nu} \mathbf{h}(\mathcal{A} | \nu) \leq \sup_{\tilde{\nu}} \mathbf{h}(\mathcal{A} | \tilde{\nu}).$$

Applying the supremum over all covers  $\mathcal{A}$  on both sides, and exchanging the suprema, we conclude this argument. ■

**THEOREM 4** (Inner variational principle; cf. Theorem 2.1 in [L-W]). *Let  $\pi : X \rightarrow Y$  be a topological factor map between topological dynamical systems  $(X, T)$  and  $(Y, S)$ . If  $\nu$  is an invariant Radon probability measure on the factor  $(Y, S)$  then*

$$\mathbf{h}(X | \nu) = \sup\{\mathbf{h}(\mu | \nu) : \mu \in \mathcal{P}_T(X), \pi\mu = \nu\}.$$

*Proof.* In general, we follow Misiurewicz’s proof of the classical variational principle (see e.g. [W3]). Some details, however, are much more complicated, due to conditionality of the entropies.

For the inequality “ $\geq$ ” fix an invariant measure  $\mu$  on  $X$  such that  $\pi\mu = \nu$ . We will first show that  $\mathbf{h}(X | \nu) \geq \mathbf{h}(\mu | \nu) - \log 2 - 1$ , i.e. (by definitions)

$$\sup_{\mathcal{A}} \inf_n \frac{1}{n} \int \mathbf{H}(\mathcal{A}^n | y) d\nu \geq \sup_{\alpha} \inf_{\beta} \inf_n \frac{1}{n} \mathbf{H}(\mu, \alpha^n | \beta) - \log 2 - 1.$$

We do so by constructing, for each partition  $\alpha$ , a cover  $\mathcal{A}$  such that for every  $n$  there exists a  $Y$ -measurable partition  $\beta$  satisfying

$$\frac{1}{n} \int \mathbf{H}(\mathcal{A}^n | y) d\nu \geq \frac{1}{n} \mathbf{H}(\mu, \alpha^n | \beta) - \log 2 - 1.$$

Let  $\alpha = \{C_1, \dots, C_k\}$ . By regularity of  $\mu$ , we can find another partition  $\alpha' = \{C'_0, C'_1, \dots, C'_k\}$  with each  $C'_i$  closed and contained in  $C_i$  ( $i = 1, \dots, k$ ) and with  $\mu(C'_0) < 1/\log k$ . It is elementary to see that  $\mathbf{H}(\mu, \alpha | \alpha') < 1$ . We construct the cover  $\mathcal{A} = \{A_1, \dots, A_k\}$  by letting  $A_i = C'_0 \cup C'_i$  (every such  $A_i$  is open). From now on we fix the cover  $\mathcal{A}$ .

For a given  $n$  the cover  $\mathcal{A}^n$  consists of elements of the form

$$J(j) = \bigcap_{i=0}^{n-1} T^{-i}(C'_0 \cup C'_{j(i)}),$$

where  $j = (j(i)) \in \{1, \dots, k\}^n$ . For given  $j$  and a binary sequence  $c = (c(i)) \in \{0, 1\}^n$  we define a set

$$J_c(j) := \bigcap_{i=0}^{n-1} T^{-i}(C'_{c(i)j(i)}),$$

where  $c(i)j(i)$  denotes the ordinary product of numbers. Clearly, each non-empty set  $J_c(j)$  belongs to the partition  $(\alpha')^n$ , and

$$J(j) = \bigcup_{c \in \{0,1\}^n} J_c(j).$$

Fix some  $y \in Y$  and let  $\mathcal{F}$  be a family realizing the minimum in the definition of  $\mathbf{N}(\mathcal{A}^n | y)$ . Then the family  $\{J_c(j) : J(j) \in \mathcal{F}, c \in \{0,1\}^n\}$  covers not only  $\pi^{-1}y$  but the entire set  $\bigcup \mathcal{F}$ . Thus, if  $B$  is any Borel subset of  $\bigcup \mathcal{F}$  containing  $\pi^{-1}y$  then

$$\mathbf{N}(\mathcal{A}^n | y) \cdot 2^n \geq \mathbf{N}((\alpha')^n | B)$$

(the definition of  $\mathbf{N}(\cdot | B)$  for partitions is analogous to that for covers). If now  $\mu_B$  is any probability measure supported by  $B$ , then

$$\mathbf{H}(\mu_B, (\alpha')^n) \leq \log \mathbf{N}((\alpha')^n | B) \leq \mathbf{H}(\mathcal{A}^n | y) + n \log 2.$$

(The first inequality follows from the well known estimate of the entropy of a partition with respect to any measure by the entropy with respect to a measure assigning equal weights to all elements of the partition.)

Let  $\gamma$  be the partition of  $Y$  into the equivalence classes with respect to the following relation:  $y \sim y'$  whenever  $\pi^{-1}y$  and  $\pi^{-1}y'$  are covered by exactly the same families  $\mathcal{F} \subset \mathcal{A}^n$ . It is straightforward to verify (using e.g. upper semicontinuity of the partition into fibers of  $\pi$ ; see Appendix, (A1)) that each class is an intersection of an open set and a closed set, hence it is Borel. Set  $\beta := \pi^{-1}\gamma$  and let  $D$  be an element of  $\gamma$ . It is clear that if  $y \in D$  and  $\mathcal{F} \subset \mathcal{A}^n$  is a family which covers  $\pi^{-1}y$  then  $\pi^{-1}D \subset \bigcup \mathcal{F}$ , hence the last displayed estimate applies to  $B := \pi^{-1}D$ . Moreover, the function  $\mathbf{H}(\mathcal{A}^n | y)$  assumes a constant value, denoted by  $\mathbf{H}(\mathcal{A}^n | D)$ , on each element  $D$  of the partition  $\gamma$ . Thus we have

$$\begin{aligned} \int \mathbf{H}(\mathcal{A}^n | y) \, d\nu &= \sum_{D \in \gamma} \nu(D) \mathbf{H}(\mathcal{A}^n | D) \\ &\geq \sum_{B \in \beta} \mu(B) \mathbf{H}(\mu_B, (\alpha')^n) - n \log 2 \\ &= \mathbf{H}(\mu, (\alpha')^n | \beta) - n \log 2. \end{aligned}$$

Further, applying a standard estimate for conditional entropy (see [W3], Theorem 4.3(i)) and subadditivity of the sequence  $\mathbf{H}(\mu, \alpha^n | (\alpha')^n)$  (see Corollary 1), we obtain

$$\mathbf{H}(\mu, \alpha^n | \beta) \leq \mathbf{H}(\mu, (\alpha')^n | \beta) + \mathbf{H}(\mu, \alpha^n | (\alpha')^n) < \mathbf{H}(\mu, (\alpha')^n | \beta) + n,$$

which implies the desired inequality

$$\frac{1}{n} \int \mathbf{H}(\mathcal{A}^n | y) \, d\nu \geq \frac{1}{n} \mathbf{H}(\mu, \alpha^n | \beta) - \log 2 - 1.$$

We have proved that  $\mathbf{h}(X | \nu) \geq \mathbf{h}(\mu | \nu) - \log 2 - 1$ . Replacing  $T$  by  $T^m$ , and

$S$  by  $S^m$ , where  $m$  is an arbitrary positive integer, we can write

$$m\mathbf{h}(X | \nu) \geq m\mathbf{h}(\mu | \nu) - \log 2 - 1$$

(see Remark 3). Dividing both sides by  $m$  and letting  $m \rightarrow \infty$  completes this part of the proof.

Proving the other inequality in the assertion is the most complicated reasoning of this note. Recall that, by compactness, an open set  $A$  is an  $F_\sigma$ -set if and only if  $A = \{x : f(x) > 0\}$  for some continuous function  $f$  on  $X$  (see e.g. [G-J]). We begin by picking a cover  $\mathcal{A}$  consisting of open  $F_\sigma$ -sets, and an  $\varepsilon > 0$ , which will remain fixed for the rest of the proof.

For every pair  $(y, m)$  ( $y \in Y$  and  $m \in \mathbb{N}$  sufficiently large), we want to construct a probability measure  $\mu_{(y,m)}$  on  $X$  supported by  $\pi^{-1}y$ , with large value of  $\mathbf{H}(\mu_{(y,m)}, \alpha^m)$  for any sufficiently fine partition  $\alpha$ , chosen uniformly for all pairs  $(y, m)$ . In the most popular version of the proof of the classical variational principle, Bowen’s definition of entropy (using a metric) is strongly used. As a substitute of the metric we shall employ pseudometrics. We remark that in his original proof of the classical variational principle in nonmetrizable spaces, Misiurewicz uses the uniform structure.

For each set  $A \in \mathcal{A}$  we fix a nonnegative continuous function  $f_A$  such that  $A = \{x : f_A(x) > 0\}$  (this is possible since  $A$  is an open  $F_\sigma$ ). The formula

$$d(x, x') := \max\{|f_A(x) - f_A(x')| : A \in \mathcal{A}\}$$

defines a pseudometric on  $X$ . Denote by  $\varepsilon_{\mathcal{A}}$  the minimal value of the continuous function obtained as the pointwise maximum of all the functions  $f_A$  (we remark that  $\varepsilon_{\mathcal{A}}$  coincides with the Lebesgue number of the cover  $\mathcal{A}$  with respect to  $d$ ). Below we will use the convention that  $\sqrt{m}$  denotes the integer part of  $\sqrt{m}$ . A set  $E \subset X$  is said to be  $(m, \varepsilon_{\mathcal{A}})$ -separated if, for any  $x, x'$  in  $E$ , there exists a positive integer  $k \in [\sqrt{m}, m + \sqrt{m} - 1]$  such that  $d(T^k x, T^k x') \geq \varepsilon_{\mathcal{A}}$  (we have slightly modified the classical definition for purely technical reasons). By compactness of  $X$  and continuity of  $d$ , every such set is finite.

We define the measure  $\mu_{(y,m)}$  as follows: in  $\pi^{-1}y$  we select a maximal  $(m, \varepsilon_{\mathcal{A}})$ -separated set  $E_{(y,m)}$  (i.e., one realizing the maximal cardinality of all  $(m, \varepsilon_{\mathcal{A}})$ -separated subsets of  $\pi^{-1}y$ ), and we let  $\mu_{(y,m)}$  be the equidistributed probability measure on this finite set.

We now show that the cardinality of  $E_{(y,m)}$  is at least  $\mathbf{N}(T^{-\sqrt{m}}\mathcal{A}^m | y)$ . Namely, by the definition of  $\varepsilon_{\mathcal{A}}$ , for each  $x \in X$  and  $k \geq 0$  there exists an  $A \in \mathcal{A}$  such that  $f_A(T^k x) \geq \varepsilon_{\mathcal{A}}$ . We denote this  $A$  by  $A_{x,k}$ . Then we define

$$A_x := \bigcap_{k=\sqrt{m}}^{m+\sqrt{m}-1} T^{-k} A_{x,k}.$$

Clearly,  $x \in A_x \in T^{-\sqrt{m}}\mathcal{A}^m$ . It suffices to show that the family  $\mathcal{F} := \{A_x : x \in E_{(y,m)}\}$  covers  $\pi^{-1}y$ . Suppose there exists a point  $x' \in \pi^{-1}y$  not covered by  $\mathcal{F}$ . This means that for every  $x \in E_{(y,m)}$  there exists a  $k \in [\sqrt{m}, m + \sqrt{m} - 1]$  with  $x' \notin T^{-k}A_{x,k}$ , i.e., such that  $f_{A_{x,k}}(T^k x') = 0$ . Since  $f_{A_{x,k}}(T^k x) \geq \varepsilon_{\mathcal{A}}$ , we obtain  $d(T^k x, T^k x') \geq \varepsilon_{\mathcal{A}}$ . Thus  $E_{(y,m)} \cup \{x'\}$  is an  $(m, \varepsilon_{\mathcal{A}})$ -separated set, a contradiction to maximality of  $E_{(y,m)}$ .

Let  $\alpha$  be any partition of  $X$  into sets of  $d$ -diameter strictly smaller than  $\varepsilon_{\mathcal{A}}$  (this restriction does not depend on the pair  $(y, m)$ ). Then, for each  $0 \leq i \leq \sqrt{m} - 1$ , every element of  $T^{-i}\alpha^{m+\sqrt{m}}$  contains at most one element of any  $(m, \varepsilon_{\mathcal{A}})$ -separated set, in particular it contains at most one atom of the measure  $\mu_{(y,m)}$ . This implies that

$$\begin{aligned} \mathbf{H}(T^i(\mu_{(y,m)}), \alpha^{m+\sqrt{m}}) &= \mathbf{H}(\mu_{(y,m)}, T^{-i}\alpha^{m+\sqrt{m}}) = \log \#E_{(y,m)} \\ &\geq \mathbf{H}(T^{-\sqrt{m}}\mathcal{A}^m | y). \end{aligned}$$

If  $\beta$  is any  $Y$ -measurable partition of  $X$ , then, since  $T^i(\mu_{(y,m)})$  is supported by a single fiber, the left hand side of the last inequality is

$$\mathbf{H}(T^i(\mu_{(y,m)}), \alpha^{m+\sqrt{m}} | \beta^{m+\sqrt{m}}).$$

Fix a positive integer  $n \leq \sqrt{m}$ . Summing over  $0 \leq i \leq n - 1$ , we get

$$\sum_{i=0}^{n-1} \mathbf{H}(T^i(\mu_{(y,m)}), \alpha^{m+\sqrt{m}} | \beta^{m+\sqrt{m}}) \geq n\mathbf{H}(T^{-\sqrt{m}}\mathcal{A}^m | y).$$

Now, by Lemma 1 and the formula (A14) in the Appendix, the left hand side is not larger than

$$\sum_{i=0}^{m+\sqrt{m}-1} \mathbf{H}(T^i(\mu_{(y,m)}), \alpha^n | \beta^n) + n\sqrt{m}a.$$

Next, we divide both sides by  $m + \sqrt{m}$  and apply the concavity of the conditional entropy function  $\mathbf{H}(\cdot, \alpha^n | \beta^n)$  to further majorize the left hand side expression. Consequently, our inequality becomes

$$\mathbf{H}(M_{m+\sqrt{m}}(\mu_{(y,m)}), \alpha^n | \beta^n) \geq \frac{n}{m + \sqrt{m}}(\mathbf{H}(T^{-\sqrt{m}}\mathcal{A}^m | y) - \sqrt{m}a),$$

where, as before,  $M_{m+\sqrt{m}}$  denotes the averaging operator. Now, consider an ergodic measure  $\tilde{\nu}$  on  $Y$ . Let  $Y_{\tilde{\nu}}$  be the Borel set where the convergence of Theorem 2 holds for  $\tilde{\nu}$  and the cover  $\mathcal{A}$ . Let  $y \in Y_{\tilde{\nu}}$ . Then, by Lemma 3,

$$\frac{1}{m}\mathbf{H}(T^{-\sqrt{m}}\mathcal{A}^m | y) \geq \frac{1}{m}\mathbf{H}(\mathcal{A}^{m+\sqrt{m}} | y) - \frac{1}{m}\mathbf{H}(\mathcal{A}^{\sqrt{m}} | y),$$

the right hand side converging to  $\mathbf{h}(\mathcal{A} | \tilde{\nu})$ . Therefore,

$$\frac{1}{m} \mathbf{H}(T^{-\sqrt{m}} \mathcal{A}^m | y) \geq \mathbf{h}(\mathcal{A} | \tilde{\nu}) - \varepsilon$$

for  $m$  larger than some integer  $N(y, \varepsilon)$ . Summarizing, we have proved that, for every ergodic  $\tilde{\nu}$ , if  $y \in Y_{\tilde{\nu}}$  and  $m > N(y, \varepsilon)$  then

$$\mathbf{H}(M_{m+\sqrt{m}}(\mu_{(y,m)}), \alpha^n | \beta^n) \geq \frac{mn}{m + \sqrt{m}} \left( \mathbf{h}(\mathcal{A} | \tilde{\nu}) - \varepsilon - \frac{a}{\sqrt{m}} \right)$$

for every  $n < \sqrt{m}$ , sufficiently fine  $\alpha$ , and  $Y$ -measurable  $\beta$ .

As stated (with a proof) in (A10) of the Appendix, each ergodic measure  $\tilde{\nu}$  can be approximated by measures of the form  $M_m(\delta_y)$  with  $y \in Y_{\tilde{\nu}}$  and  $m$  sufficiently large. Thus, if  $\nu' = c_1 \tilde{\nu}_1 + \dots + c_k \tilde{\nu}_k$  is a convex combination of ergodic measures, it can be approximated by measures of the form  $M_{m'}(\Delta_{\nu'})$ , where  $\Delta_{\nu'} := c_1 \delta_{y_1} + \dots + c_k \delta_{y_k}$ ,  $y_i \in Y_{\tilde{\nu}_i}$ , for each  $1 \leq i \leq k$  and  $m' > \max_i N(y_i, \varepsilon)$ . For a given pair  $(\nu', m')$  we set  $\mu_{(\nu', m')} := c_1 \mu_{(y_1, m')} + \dots + c_k \mu_{(y_k, m')}$ . Obviously,  $\mu_{(\nu', m')}$  projects by  $\pi$  to  $\Delta_{\nu'}$ , and hence  $M_{m'}(\mu_{(\nu', m')})$  projects to  $M_{m'}(\Delta_{\nu'})$ . Applying concavity to the left hand side and the fact that  $\mathbf{h}(\mathcal{A} | \cdot)$  is affine on invariant measures (Corollary 2) to the right hand side of the last displayed inequality, we obtain

$$\mathbf{H}(M_{m'+\sqrt{m'}}(\mu_{(\nu', m')}), \alpha^n | \beta^n) \geq \frac{m'n}{m' + \sqrt{m'}} \left( \mathbf{h}(\mathcal{A} | \nu') - \varepsilon - \frac{a}{\sqrt{m'}} \right).$$

Let  $\nu$  be an arbitrary invariant measure on  $Y$ . We choose a net of convex combinations  $\nu'$  of ergodic measures converging to  $\nu$ . By the properties of the function  $\mathbf{h}(\mathcal{A} | \cdot)$  (see Corollary 2 and remark following Definition 7), we can do it so that  $\mathbf{h}(\mathcal{A} | \nu')$  converges to  $\mathbf{h}(\mathcal{A} | \nu)$  (see Appendix, (A11)). We approximate each  $\nu'$  in this net by a measure of the form  $M_{m'}(\Delta_{\nu'})$  with  $m' > \max_i N(y_i, \varepsilon)$  (depending on  $\nu'$ ) so that  $M_{m'}(\Delta_{\nu'}) \rightarrow \nu$ . Of course, we can do it so that the parameter  $m'$  grows to infinity with  $\nu'$ . By choosing a subnet we can assume that the corresponding measures  $M_{m'}(\mu_{(\nu', m')})$  converge to some  $\mu$  in  $\mathcal{P}(X)$ . Obviously,  $\mu$  is an invariant measure and  $\pi(\mu) = \nu$ . It is elementary to verify that also

$$M_{m'+\sqrt{m'}}(\mu_{(\nu', m')}) \rightarrow \mu.$$

Passing with  $\nu'$  to the limit in the last displayed estimate (so that  $m' \rightarrow \infty$ ), and then dividing both sides by  $n$ , we obtain

$$\frac{1}{n} \mathbf{H}(\mu, \alpha^n | \beta^n) \geq \mathbf{h}(\mathcal{A} | \nu) - \varepsilon,$$

provided that  $\mu$  is a continuity point of the function  $\mathbf{H}(\cdot, \alpha^n | \beta^n)$ .

Recall that our estimates hold for all partitions  $\alpha$  into sets of  $d$ -diameter smaller than  $\varepsilon_{\mathcal{A}}$  and all  $Y$ -measurable partitions  $\beta$ . The definition of  $\mu$  does

not depend on the choice of  $\alpha$  and  $\beta$ . The required continuity condition is satisfied whenever  $\alpha$  and  $\beta$  consist of sets with boundaries of  $\mu$  measure zero. As is not hard to see (using the functions  $f_A$ ), such an  $\alpha$  exists among all the partitions considered. We now fix such an  $\alpha$  (remembering that this choice does not depend on  $n$ ). As we soon explain, required partitions  $\beta$  also exist and suffice for our purposes.

We would like to apply the first equality of Lemma 2. Taking  $\inf_{\beta}$  would yield

$$\inf_{\beta} \frac{1}{n} \mathbf{H}(\mu, \alpha^n | \beta^n) \geq \mathbf{h}(\mathcal{A} | \nu) - \varepsilon.$$

The above holds if the infimum is taken over all  $Y$ -measurable partitions having zero boundaries with respect to  $\mu$  (for such partitions our previous estimation works). But, using regularity of  $\nu$  and an argument involving Urysohn functions, every  $Y$ -measurable partition  $\beta$  can be easily modified (by modifying the corresponding partition  $\gamma$  of  $Y$ ) to a partition with zero boundaries in such a manner that the value of  $\mathbf{H}(\mu, \alpha^n | \beta^n)$  is changed as little as we wish. Thus we can skip the zero boundaries restriction on  $\beta$ . Next, with no further obstacles, we apply the infimum over  $n$ . Finally, the supremum over  $\alpha$  is obviously not smaller than the value obtained for our fixed  $\alpha$ . We have proved that there exists an invariant measure  $\mu$  projecting to  $\nu$  such that

$$\mathbf{h}(\mu | \nu) \geq \mathbf{h}(\mathcal{A} | \nu) - \varepsilon.$$

It is well known (see e.g. [G-J], 16.7(a)) that in every cover one can inscribe a cover by open  $F_{\sigma}$ -sets, hence we can take the supremum of the right hand side over all covers. Since  $\varepsilon$  is arbitrary, the desired inequality is proved. ■

We conclude this section by formulating several obvious consequences of the outer and inner variational principles.

**THEOREM 5** (Conditional variational principle; cf. [B-F-F]). *Let  $\pi : X \rightarrow Y$  be a topological factor map between topological dynamical systems  $(X, T)$  and  $(Y, S)$ . Then*

$$\mathbf{h}(X | Y) = \sup_{\mu \in \mathcal{P}_T(X)} \mathbf{h}(\mu | \pi\mu). \quad \blacksquare$$

**COROLLARY 3** (cf. Theorem 17 in [B]). *If  $\mathbf{h}(Y) < \infty$  then*

$$\mathbf{h}(X | Y) \geq \mathbf{h}(X) - \mathbf{h}(Y). \quad \blacksquare$$

**COROLLARY 4.** *For every  $\nu \in \mathcal{P}_S(Y)$  with  $\mathbf{h}(\nu) < \infty$ ,*

$$\mathbf{h}(X | \nu) \leq \mathbf{h}(X) - \mathbf{h}(\nu). \quad \blacksquare$$

**4. Tail entropy in totally disconnected spaces.** M. Misiurewicz [M2] introduced the notion of “topological conditional entropy” of a dynamical system  $(X, T)$ :

$$h^*(X) := \inf_{\mathcal{B}} h(X | \mathcal{B}),$$

where  $\mathcal{B}$  ranges over all covers. We would rather call it the *tail entropy* of  $(X, T)$ , because in our paper, “topological conditional entropy” has a different meaning. One of the important properties of  $h^*(X)$  is that it estimates from above the possible “defect of upper semicontinuity” of the entropy function (see below for details).

We would like to compare that notion with our topological fiber entropy. Because the latter relies on the existence of a factor, it will be convenient to assume that the space  $X$  is totally disconnected—this guarantees a rich structure of factors (or existence of a finite topological generator). Namely, in that case, in each cover we can inscribe a cover  $\mathcal{B}$  by disjoint (hence closed and open) sets. Such a cover generates a topological factor of  $(X, T)$  which we denote by  $(Y_{\mathcal{B}}, S_{\mathcal{B}})$ . The dynamical system  $(X, T)$  is then isomorphic to the inverse limit

$$(X, T) = \varprojlim_{\mathcal{B}} (Y_{\mathcal{B}}, S_{\mathcal{B}}),$$

where  $(\mathcal{B})$  is the net of all disjoint covers ordered by the relation “ $\supseteq$ ”.

REMARK 6. In the metrizable case, if the topological entropy is finite, we can replace the system  $(X, T)$  by its totally disconnected extension in a way that does not change any of the entropy properties. This is done via theorems by E. Lindenstrauss and B. Weiss ([Li-W], [Li]) which allow one to find a basis of the topology consisting of sets whose boundaries have measure zero for all invariant measures. In this manner, the facts stated below implicitly apply to all compact metrizable dynamical systems with finite topological entropy.

By Remark 4, in the totally disconnected case, Misiurewicz’s definition can be reformulated as follows:

$$h^*(X) := \inf_{\mathcal{B}} h(X | Y_{\mathcal{B}}) (= \lim_{\mathcal{B}} h(X | Y_{\mathcal{B}})),$$

where the infimum (limit) is taken over all covers by disjoint open and closed sets. By the outer variational principle, we obtain

$$h^*(X) = \inf_{\mathcal{B}} \sup_{\mu \in \mathcal{P}_T(X)} h(X | \nu_{\mathcal{B}}),$$

where  $\nu_{\mathcal{B}}$  denotes the projection of  $\mu$  onto the factor space  $Y_{\mathcal{B}}$ . The following definition sounds natural in this context:

DEFINITION 12. Let  $\mu \in \mathcal{P}_T(X)$ . The *tail entropy* of  $\mu$  is defined as

$$\mathbf{h}^*(\mu) := \inf_{\mathcal{B}} \mathbf{h}(X | \nu_{\mathcal{B}}).$$

Again, it is not hard to verify (using for example the inner variational principle) that the above infimum is of a monotone net, hence it can be replaced by the appropriate limit. Clearly, we have

$$\mathbf{h}^*(X) \geq \sup_{\mu \in \mathcal{P}_T(X)} \mathbf{h}^*(\mu).$$

Unfortunately, a “tail variational principle” (equality in the above statement) does not hold in general: the function  $\mathbf{h}(X | \nu_{\mathcal{B}})$  is usually not u.s.c., hence we cannot take infimum outside the supremum. We will show that the “defect of upper semicontinuity” of the entropy function locates in-between (see Theorem 6). Let

$$(\mathbf{h}^+ - \mathbf{h})(\mu) := \limsup_{\mu' \rightarrow \mu} \mathbf{h}(\mu') - \mathbf{h}(\mu)$$

( $\mu, \mu' \in \mathcal{P}_T(X)$ ). (In [M2], the above quantity is denoted by  $h_{\mu}^*(T)$ . We are not using that notation, since we reserve the star for the tail entropy.)

THEOREM 6. For every  $\mu \in \mathcal{P}_T(X)$  with  $\mathbf{h}(\mu) < \infty$ ,

$$\mathbf{h}^*(X) \geq (\mathbf{h}^+ - \mathbf{h})(\mu) \geq \mathbf{h}^*(\mu).$$

*Proof.* The first inequality is well known (see [M2], Theorem 4.2). By the definition of  $\mathbf{h}^*(\mu)$  and the inner variational principle, for every cover  $\mathcal{B}$  and  $\varepsilon > 0$  we can find an invariant measure  $\mu^{\mathcal{B}, \varepsilon}$  on  $X$  projecting to  $\nu_{\mathcal{B}}$  and such that

$$\mathbf{h}(\mu^{\mathcal{B}, \varepsilon}) - \mathbf{h}(\nu_{\mathcal{B}}) \geq \mathbf{h}^*(\mu) - \varepsilon$$

(if  $\mathbf{h}^*(\mu)$  is infinite then the right hand side can be made arbitrarily large). Since the covers  $\mathcal{B}$  eventually generate the topology, it is easily seen that the net of measures  $(\mu^{\mathcal{B}, \varepsilon})_{\mathcal{B}, \varepsilon}$  converges weak\* to  $\mu$ . This implies the second inequality. ■

We remark that none of the inequalities of Theorem 6 can be reversed (even with  $\sup_{\mu}$  applied to the middle and right hand side). In [D], Remark 5, we find an example with constant entropy function and  $\mathbf{h}^*(X)$  positive. Using techniques similar to those used in Example 1 of [D], an example with  $\mathbf{h}^*(\cdot) \equiv 0$  and non-u.s.c. entropy function can be constructed; one needs a jump down of  $\mathbf{h}(\cdot)$  realized exclusively by a net of measures projecting differently than the limit measure. Finally, one easily verifies that in Misiurwicz’s example of a system with  $\mathbf{h}^*(X) > 0$  (see also  $X''$  in Example 1 of [D]),  $\mathbf{h}^*(\mu) = \mathbf{h}(X) = \log 2$  for a certain measure  $\mu$ , which shows that all quantities in Theorem 6 can be equal and positive.

## Appendix

*Upper semicontinuous partitions and functions.* Let  $\alpha$  be a (usually infinite) partition of a compact Hausdorff space  $X$  into closed sets. Then

- (A1) there exists a continuous map  $\pi : X \rightarrow Y$  into another Hausdorff space such that  $\alpha$  coincides with the partition into fibers of  $\pi$  if and only if for every open set  $A \subset X$  the union of all elements of  $\alpha$  contained in  $A$  is open.

In that case we say that  $\alpha$  is an *upper semicontinuous* partition of  $X$ .

A real-valued function  $f$  defined on a compact Hausdorff space  $X$  is called *upper semicontinuous* (u.s.c.) if one of the following equivalent conditions holds:

- (A2)  $f = \inf_{\alpha} g_{\alpha}$  for some family  $\{g_{\alpha}\}$  of continuous functions,  
 (A3)  $f = \lim_{\alpha} g_{\alpha}$ , where  $(g_{\alpha})$  is a nonincreasing net of continuous functions,  
 (A4) for each  $r \in \mathbb{R}$  the set  $\{x : f(x) \geq r\}$  is closed.

Clearly, by (A2),

- (A5) the infimum of any family of u.s.c. functions is again a u.s.c. function.

By (A4), both the sum and supremum of finitely many u.s.c. functions are u.s.c. functions.

The following statement will be useful:

- (A6) If  $f = \lim f_{\alpha}$  is the limit of a nonincreasing net of u.s.c. functions defined on a compact space, and  $g$  is a continuous function strictly larger than  $f$ , then  $g$  is strictly larger than  $f_{\alpha}$  for some  $\alpha$ .

*Proof.* For each  $\alpha$  the function  $f_{\alpha} - g$  is u.s.c. Thus the set  $\{x : f_{\alpha}(x) \geq g(x)\}$  is closed. The above sets decrease with respect to  $\alpha$ , so if they all were nonempty they would have a nonempty intersection, on which  $f(x) \geq g(x)$ , a contradiction. ■

Assume that  $X$  is a compact Hausdorff space and  $\mu$  a Radon (Borel regular) probability measure on  $X$ . We can use neither bounded nor monotone convergence theorems for nets of functions. Nonetheless, the measure satisfies the so-called  *$\tau$ -smoothness condition*, in particular we do have

- (A7) if  $f = \lim_{\alpha} f_{\alpha}$  is the limit of a nonincreasing net of nonnegative continuous functions, then

$$\lim_{\alpha} \int f_{\alpha} d\mu = \int f d\mu.$$

*Proof.* First observe that  $f$  is u.s.c. and bounded, hence the integral on the right is defined. The inequality “ $\geq$ ” is obvious. For the opposite inequality, fix an  $\varepsilon > 0$  and for each  $n \in \mathbb{N}$  define  $F_n = \{x : f(x) \geq n\varepsilon\}$ . By boundedness of  $f$  the number  $m$  of nonempty such sets is finite, and since  $f$  is u.s.c., these sets are closed. By regularity, we can inductively find open sets  $U_n$  satisfying, for each  $1 \leq n \leq m$ ,

$$F_n \subset U_n, \quad \mu(U_n \setminus F_n) < \frac{1}{m(m+1)}, \quad U_{n+1} \subset U_n.$$

Let  $g_n$  denote an Urysohn function:  $0 \leq g_n \leq 1$ ,  $g_n \equiv 0$  on the complement of  $U_n$  and  $g_n \equiv 1$  on  $F_n$  (see e.g. [E]). It is easily seen that the continuous function

$$g = \varepsilon \left( 1 + \sum_{n=1}^m g_n \right) \leq (m+1)\varepsilon$$

satisfies  $g(x) > f(x)$  at each point of  $X$ , and  $\mu\{x : g(x) - f(x) > \varepsilon\} < 1/(m+1)$ . Thus  $\int g d\mu < \int f d\mu + 2\varepsilon$ , and, by (A6),  $g > f_\alpha$  for some  $\alpha$ , which implies  $\int g d\mu > \lim_\alpha \int f_\alpha d\mu$ . Since  $\varepsilon$  was arbitrary, this concludes the proof. ■

From the definition of the weak\* topology, and by (A3) and (A7), we conclude that

(A8) if  $f$  is a u.s.c. function on  $Y$  then  $\int f d\mu$  is a u.s.c. function on  $\mathcal{P}(X)$ .

*Invariant Radon probability measures.* The collection  $\mathcal{P}(X)$  of all Radon probability measures defined on  $X$  is well known to be weak\* compact. Let  $(X, T)$  be a topological dynamical system. Then  $T$  induces a continuous map on  $\mathcal{P}(X)$  by the formula  $T\mu(B) = \mu(T^{-1}B)$  ( $B$  is a Borel subset of  $X$ ). It is well known that the set  $\mathcal{P}_T(X)$  of all  $T$ -invariant Radon probability measures on  $X$  is nonempty, compact, convex, and that its extreme points are exactly the ergodic measures.

For  $n \in \mathbb{N}$  define a continuous map  $M_n : \mathcal{P}(X) \rightarrow \mathcal{P}(X)$  by

$$M_n(\mu) := \frac{1}{n} \sum_{i=0}^{n-1} T^i \mu.$$

Then, as is easily verified, for any subsequence  $(n_k)$  such that, for each  $k$ ,  $n_{k+1}$  is a multiple of  $n_k$ , the sets  $M_{n_k}(\mathcal{P}(X))$  decrease and their (nonempty) intersection is contained in  $\mathcal{P}_T(X)$ . In particular this implies that

(A9) if  $U \supset \mathcal{P}_T(X)$  is an open set in  $\mathcal{P}(X)$  then  $M_{n_k}(\mathcal{P}(X)) \subset U$  for sufficiently large  $k$ .

Let  $\tilde{\mu} \in \mathcal{P}_T(X)$  be an ergodic measure. In contrast to the metrizable case, the so-called *generic points* (i.e., points  $x \in X$  such that  $M_n(\delta_x) \rightarrow \tilde{\mu}$

in the weak\* topology as  $n \rightarrow \infty$ ) may not exist. But still, the following holds:

(A10) if  $X' \subset X$  satisfies  $\tilde{\mu}(X') = 1$  then  $\tilde{\mu}$  can be approximated in the weak\* topology by a net of the form  $M_{n_\iota}(\delta_{x_\iota})$  with  $x_\iota \in X'$  for every  $\iota$ . Moreover, the convergence still holds if each  $n_\iota$  is replaced by an arbitrary larger integer.

*Proof.* By the ergodic theorem, for every  $\varepsilon$  and every finite collection  $\mathcal{F}$  of continuous functions, there exists an  $x_{(\varepsilon, \mathcal{F})} \in X'$  and an  $n_{(\varepsilon, \mathcal{F})} \in \mathbb{N}$  such that if  $n \geq n_{(\varepsilon, \mathcal{F})}$  and  $f \in \mathcal{F}$  then  $\int f d(M_n(\delta_{x_\iota}))$  differs from  $\int f d\tilde{\mu}$  by less than  $\varepsilon$ . The pairs  $(\varepsilon, \mathcal{F})$  form a directed family, hence a net of the desired properties has been constructed. ■

Let  $\mathcal{P}$  be a compact convex subset of a locally convex linear space  $V$ . Then, by the Krein–Milman theorem,  $\mathcal{P}$  is equal to  $\overline{\text{conv}}(\text{ex } \mathcal{P})$ , i.e., to the closure of the convex hull of the extreme points of  $\mathcal{P}$ .

(A11) If  $f$  is a nonnegative affine u.s.c. function defined on  $\mathcal{P}$ , then, for every  $\mu \in \mathcal{P}$ , we have

$$f(\mu) = \limsup_{\mu' \rightarrow \mu} f(\mu'),$$

where  $\mu' \in \text{conv}(\text{ex } \mathcal{P})$ .

*Proof.* By upper semicontinuity,  $f(\mu)$  cannot be smaller than the expression on the right. Suppose it is strictly larger. In the space  $V \times \mathbb{R}$  consider the compact convex set  $F = \overline{\text{conv}}\{(\tilde{\mu}, r) : \tilde{\mu} \in \text{ex } \mathcal{P}, 0 \leq r \leq f(\tilde{\mu})\}$ . By our assumption, the pair  $(\mu, f(\mu))$  does not belong to  $F$ . By the Hahn–Banach theorem, there exists a continuous affine function  $g$  on  $\mathcal{P}$  such that  $g(\tilde{\mu}) > f(\tilde{\mu})$  for every  $\tilde{\mu} \in \text{ex } \mathcal{P}$ , and  $g(\mu) < f(\mu)$ . The function  $h = f - g$  is affine, u.s.c., assumes negative values at all extreme points of  $\mathcal{P}$ , and is positive at  $\mu$ . We arrive at a contradiction by showing that any affine u.s.c. function  $h$  attains its maximal value at some extreme point. Indeed, the set  $\mathcal{M}$  where  $h$  attains its maximum  $Q$  is nonempty, closed, and convex. Let  $\mu$  be an extreme point of  $\mathcal{M}$ . If  $\mu \notin \text{ex } \mathcal{P}$ , then it admits a decomposition  $\mu = \frac{1}{2}(\mu_1 + \mu_2)$  with  $\mu_1 \neq \mu_2$ . But then  $h(\mu_1) = h(\mu_2) = Q$ , which implies that both  $\mu_1$  and  $\mu_2$  are in  $\mathcal{M}$ , a contradiction with extremality of  $\mu$  in  $\mathcal{M}$ . ■

We now make sure that the following statement holds without metrizable-ability of  $X$ :

(A12) Let  $A$  be a Borel set. If  $\tilde{\mu}(A) = 0$  for every ergodic measure  $\tilde{\mu}$  then  $\mu(A) = 0$  for every invariant measure  $\mu$ .

*Proof.* By regularity of  $\mu$ , it suffices to prove the above for closed sets  $A$ . For such  $A$ ,  $\mu(A)$  is a nonnegative affine u.s.c. function on  $\mathcal{P}_T(X)$  (see (A8)). The statement now follows directly from (A11). ■

*Subadditive ergodic theorem.* Let  $(H_n)_{n \geq 1}$  be a subadditive measurable process defined on a measure-theoretic dynamical system  $(X, \mu, T)$ , i.e., a sequence of measurable functions on  $X$  such that for every  $x \in X$  and any natural  $m$  and  $n$ ,

$$H_{m+n}(x) \leq H_n(x) + H_m(T^n x).$$

Assume that  $H_1$  is bounded by a constant  $a$ . Then obviously  $n^{-1}H_n$  is also bounded by  $a$ . The following *subadditive ergodic theorem* holds (see [Kr], Theorem 5.3 for a more general version):

(A13) If  $\tilde{\mu}$  is an ergodic measure on  $X$  then  $n^{-1}H_n(x)$  converges  $\tilde{\mu}$ -a.e. to the constant  $h = \inf_n n^{-1} \int H_n d\tilde{\mu}$ .

*Proof.* Fix two natural numbers  $m > n$  and write  $m = kn + l$  ( $l < n$ ). By subadditivity of  $H_n$ , we have

$$H_m(x) \leq \sum_{i=0}^{k-1} H_n(T^{ni} x) + la.$$

Replacing  $x$  by  $Tx, \dots, T^{n-1}x$  and adding, we obtain

(A14) 
$$\sum_{i=0}^{n-1} H_m(T^i x) \leq \sum_{i=0}^{m-1} H_n(T^i x) + nla.$$

On the other hand, applying subadditivity again, for each  $0 \leq i < n$  we have

$$H_{m+n}(x) \leq H_i(x) + H_m(T^i x) + H_{n-i}(T^{m+i} x),$$

where the sum of the first and last terms does not exceed  $na$ . Averaging over  $0 \leq i < n$  and applying the previous estimate we get

(A15) 
$$H_{m+n}(x) \leq \frac{1}{n} \sum_{i=0}^{n-1} H_m(T^i x) + na \leq \frac{1}{n} \sum_{i=0}^{m-1} H_n(T^i x) + 2na.$$

Dividing by  $m$  and letting  $m \rightarrow \infty$  we obtain, by the ergodic theorem,

$$\limsup_{m \rightarrow \infty} \frac{1}{m} H_m(x) \leq \frac{1}{n} \int H_n d\tilde{\mu}$$

for  $\tilde{\mu}$ -almost every  $x$ . Since  $n$  is arbitrary, we can write

$$\limsup_{m \rightarrow \infty} \frac{1}{m} H_m(x) \leq h \quad \tilde{\mu}\text{-a.e.}$$

It remains to prove the reverse inequality with  $\liminf$ . Suppose for contradiction that for some positive  $\varepsilon$ ,

$$\liminf_{n \rightarrow \infty} \frac{1}{n} H_n(x) < h - 3a\varepsilon =: c$$

on a positive measure set. By subadditivity,  $H_{n+1}(x) \leq a + H_n(Tx)$ , hence the function  $\liminf_n n^{-1} H_n$  is subinvariant, and thus, by ergodicity, constant  $\tilde{\mu}$ -a.e. This implies that the inequality  $\liminf_n n^{-1} H_n(x) < c$  holds in fact on a full measure set. In particular,

$$\tilde{\mu} \left( \bigcup_{n=1}^{\infty} \left\{ x : \frac{1}{n} H_n(x) < c \right\} \right) = 1.$$

Then there exists  $n_0$  such that the set

$$E := \bigcup_{n=1}^{n_0} \{x : H_n(x) < nc\}$$

has measure larger than  $1 - \varepsilon$ . By the von Neumann theorem applied to the characteristic function of  $E$ , there exists a positive integer  $m_0$  larger than  $n_0/\varepsilon$  such that the set

$$F := \left\{ x : \frac{1}{m_0} \#\{i : 0 \leq i < m_0, T^i x \in E\} > 1 - \varepsilon \right\}$$

has measure also larger than  $1 - \varepsilon$ . We have

$$m_0 h \leq \int H_{m_0} d\tilde{\mu} = \int_F H_{m_0} d\tilde{\mu} + \int_{Y \setminus F} H_{m_0} d\tilde{\mu}.$$

The second integral is smaller than  $m_0 a \varepsilon$ . We arrive at a contradiction by estimating the first integral by  $m_0(c + 2a\varepsilon)$ . This will be done by showing that  $H_{m_0}(x) \leq m_0(c + 2a\varepsilon)$  for every  $x \in F$ . Namely, for such  $x$  we proceed as follows:

We denote by  $i_1$  the smallest nonnegative integer with  $T^{i_1} x \in E$ , and we choose an  $n_1 \leq n_0$  with  $H_{n_1}(T^{i_1} x) < n_1 c$  (see definition of  $E$ ). Inductively, for each  $k > 1$  we let  $i_k$  be the smallest integer satisfying  $i_k \geq i_{k-1} + n_{k-1}$  and  $T^{i_k} x \in E$ , and we choose an  $n_k \leq n_0$  with  $H_{n_k}(T^{i_k} x) < n_k c$ . We call  $[i_k, i_k + n_k)$  a *good interval*. The number of positive integers smaller than  $m_0$  not contained in good intervals is at most  $m_0 \varepsilon$  (see definition of  $F$ ). The length of the last incomplete part of a good interval intersecting  $[0, m_0)$  (if any) is at most  $n_0$ , also smaller than  $m_0 \varepsilon$ . The sum  $n_1 + \dots + n_{k_0}$  representing the joint length of good intervals fully contained in  $[0, m_0)$  is thus larger than  $m_0(1 - 2\varepsilon)$ . A final application of subadditivity allows us to write

$$H_{m_0}(x) \leq \sum_{k=1}^{k_0} H_{n_k}(T^{i_k} x) + \left( m_0 - \sum_{k=1}^{k_0} n_k \right) a,$$

where the first sum comes from the good intervals, and the second estimates the rest. Replacing each  $H_{n_k}(T^{i_k}x)$  by  $n_k c$  we obtain

$$H_{m_0}(x) \leq m_0 c + 2m_0 \varepsilon a = m_0(c + 2a\varepsilon),$$

as claimed. ■

## References

- [A-R] L. M. Abramov and V. A. Rokhlin, *The entropy of a skew product of measure-preserving transformations*, Vestnik Leningrad. Univ. 17 (1962), 5–13 (in Russian).
- [B-G-H] F. Blanchard, E. Glasner and B. Host, *A variation on the variational principle and applications to entropy pairs*, Ergodic Theory Dynam. Systems 17 (1997), 29–43.
- [B] R. Bowen, *Entropy for group endomorphisms and homogeneous spaces*, Trans. Amer. Math. Soc. 153 (1971), 401–414.
- [B-F-F] M. Boyle, D. Fiebig and U. Fiebig, *Residual entropy, conditional entropy and subshift covers*, Forum Math., to appear.
- [D] T. Downarowicz, *Entropy of a symbolic extension of a totally disconnected dynamical system*, Ergodic Theory Dynam. Systems 21 (2001), 1051–1070.
- [E] R. Engelking, *General Topology*, Polish Sci. Publ., Warszawa, 1977.
- [F] H. Furstenberg, *Recurrence in Ergodic Theory and Combinatorial Number Theory*, Princeton Univ. Press, Princeton, NJ, 1981.
- [G-J] L. Gillman and M. Jerison, *Rings of Continuous Functions*, van Nostrand, Princeton, NJ, 1960.
- [Gm] T. N. T. Goodman, *Relating topological and measure entropy*, Bull. London Math. Soc. 3 (1971), 176–180.
- [Gw] L. W. Goodwyn, *Topological entropy bounds measure-theoretic entropy*, Proc. Amer. Math. Soc. 23 (1969), 679–688.
- [Ko] Z. Kowalski, *A generalized skew product*, Studia Math. 87 (1987), 215–222.
- [Kr] U. Krengel, *Ergodic Theorems*, de Gruyter Stud. in Math. 6, de Gruyter, Berlin, 1985.
- [L-W] F. Ledrappier and P. Walters, *A relativised variational principle for continuous transformations*, J. London Math. Soc. 16 (1977), 568–576.
- [Le-S] M. Lemańczyk and A. Siemaszko, *A note on the existence of a largest topological factor with zero entropy*, Proc. Amer. Math. Soc. 129 (2001), 475–482 (electronic).
- [Li] E. Lindenstrauss, *Mean dimension, small entropy factors and an imbedding theorem*, Publ. Math. I.H.E.S. 89 (1999), 227–262.
- [Li-W] E. Lindenstrauss and B. Weiss, *Mean topological dimension*, Israel J. Math. 115 (2000), 1–24.
- [M1] M. Misiurewicz, *A short proof of the variational principle for a  $Z_+^n$  action on a compact space*, Astérisque 40 (1976), 147–158.
- [M2] —, *Topological conditional entropy*, Studia Math. 55 (1976), 175–200.
- [P] J. Pachl, *Two classes of measures*, Colloq. Math. 42 (1979), 331–340.
- [R] D. Ruelle, *Statistical mechanics on a compact set with  $Z^V$  action satisfying expansiveness and specification*, Trans. Amer. Math. Soc. 185 (1973), 237–252.

- [W1] P. Walters, *A variational principle for the pressure of continuous transformations*, Amer. J. Math. 97 (1976), 937–971.
- [W2] —, *Relative pressure, relative equilibrium states, compensation functions and many-to-one codes between subshifts*, Trans. Amer. Math. Soc. 296 (1986), 1–31.
- [W3] —, *An Introduction to Ergodic Theory*, Springer, Berlin, 1982.

Institute of Mathematics  
Wrocław Technical University  
Wybrzeże Wyspiańskiego 27  
50-370 Wrocław, Poland  
E-mail: downar@im.pwr.wroc.pl  
serafin@im.pwr.wroc.pl

*Received 10 January 2001;  
in revised form 13 August 2001*