

# SHEARER'S INEQUALITY AND INFIMUM RULE FOR SHANNON ENTROPY AND TOPOLOGICAL ENTROPY

TOMASZ DOWNAROWICZ, BARTOSZ FREJ AND PIERRE-PAUL ROMAGNOLI

ABSTRACT. We review subadditivity properties of Shannon entropy, in particular, from the Shearer's inequality we derive the "infimum rule" for actions of amenable groups. We briefly discuss applicability of the "infimum formula" to actions of other groups. Then we pass to topological entropy of a cover. We prove Shearer's inequality for disjoint covers and give counterexamples otherwise. We also prove that, for actions of amenable groups, the supremum over all open covers of the "infimum formula" gives correct value of topological entropy.

February 24, 2015

## 1. INTRODUCTION

This note is devoted to properties of entropy, both measure-theoretic and topological, treated as a function defined on subsets of the acting group. One such property, called subadditivity, is popularly known and used. It implies, in particular, that when evaluating the dynamical entropy (of a partition or of an open cover) in an action of  $\mathbb{Z}$ , we can exchangeably apply  $\limsup$ ,  $\liminf$ ,  $\lim$  or  $\inf$ , of the terms  $\frac{1}{n}H(\{1, 2, \dots, n\})$  (where  $H(\{1, 2, \dots, n\})$  is appropriately understood), simply because the sequence  $\frac{1}{n}H(\{1, 2, \dots, n\})$  converges to its infimum. Similar statement holds for actions of countable amenable groups, with  $\{1, 2, \dots, n\}$  replaced by elements  $F_n$  of a Følner sequence (and  $\frac{1}{n}$  replaced by  $\frac{1}{|F_n|}$ ). But the measure-theoretic entropy fulfills a stronger property, called strong subadditivity, which implies that the same value will be obtained when taking infimum of the terms  $\frac{1}{|F|}H(F)$  over all finite subsets  $F$  of the acting group. Notice that this "infimum rule" allows to define (and evaluate) the entropy of a measure-preserving action of an amenable group without referring to any Følner sequence. In other words, the simple form  $\inf_F \frac{1}{|F|}H(F)$  does not depend upon amenability of the group and can be used as a definition of entropy of processes in actions of arbitrary groups. A natural question arises: is this definition good (i.e., does it fulfill desirable postulates one expects from a reasonable notion of entropy)? We will briefly discuss this question in the section *Beyond amenability*.

For the moment, it is not known whether a similar "infimum rule" applies to topological entropy of an open cover (even for the actions of  $\mathbb{Z}$ ). As one of our

---

2010 *Mathematics Subject Classification.* 37A35, 37B40.

*Key words and phrases.* Subadditivity, strong subadditivity, Shearer's inequality, infimum rule, Shannon entropy, topological entropy.

The research of the first named two authors is supported by the NCN (National Science Center, Poland) grant 2013/08/A/ST1/00275. The third named author acknowledges the support of Programa Basal PFB 03, CMM, Universidad de Chile. Part of the research was conducted during the visit of the first named author at CMM, Universidad de Chile.

examples shows, the corresponding function is not strongly subadditive, but this does not determine that the infimum rule fails. In fact, this rule is implied by a property lying between subadditivity and strong subadditivity, called Shearer's inequality. We will show that if the open cover consists of disjoint sets, Shearer's inequality holds, hence, for actions of countable amenable groups, the infimum rule does work. For non-disjoint covers, we show that Shearer's inequality fails. Moreover, we and give an example of a  $\mathbb{Z}_3$ -action in which the infimum does not hold. We do not have an analogous example for  $\mathbb{Z}$ —this open problem seems to be difficult. On the other hand, we prove that for amenable groups the supremum over all open covers of the “infimum formula” does yield the correct value of topological entropy of the action.

We remark that all, presented in this note, results concerning measure-theoretic entropy are well known. Novel are only the results concerning topological entropy (in particular the examples).

## 2. SUBADDITIVITY AND RELATED NOTIONS

Let  $G$  be an abstract set and let  $\mathcal{F}(G)$  be the collection of all nonempty finite subsets of  $G$ . A  $k$ -cover of a set  $F \in \mathcal{F}(G)$  is a family  $\{K_1, K_2, \dots, K_r\}$  of elements of  $\mathcal{F}(G)$  (with possible repetitions) such that each element of  $F$  belongs to  $K_i$  for at least  $k$  indices  $i \in \{1, 2, \dots, r\}$ . With slight abuse of precision, we will say “belongs to at least  $k$  elements of  $\mathcal{K}$ ” and the sums and products over  $i \in \{1, 2, \dots, r\}$  will be indexed by  $K \in \mathcal{K}$  (we must remember that repeated terms  $K$  are counted separately).

Let  $H$  be a nonnegative real function with domain  $\mathcal{F}(G)$ .

**Definition 2.1.** *We say that:*

- (M)  $H$  is monotone if  $F \subset F'$  implies  $H(F) \leq H(F')$ ;
- (S)  $H$  is subadditive if for any  $F, F'$  it holds that

$$H(F \cup F') \leq H(F) + H(F');$$

- (Sh)  $H$  satisfies Shearer's inequality if for any  $F$  and any  $k$ -cover  $\mathcal{K}$  of  $F$ ,

$$H(F) \leq \frac{1}{k} \sum_{K \in \mathcal{K}} H(K);$$

- (SS)  $H$  is strongly subadditive if for any  $F, F' \subset G$ ,

$$H(F \cup F') \leq H(F) + H(F') - H(F \cap F').$$

Let us define the *conditional value* of  $H$  by the formula

$$(2.1) \quad H(F|F') = H(F \cup F') - H(F').$$

Using the conditional value we introduce two more notions:

- (MC)  $H$  is monotone wrt. the condition if  $F' \subset F''$  implies  $H(F|F') \geq H(F|F'')$ ,
- (CS)  $H$  is conditionally subadditive if for any  $F, F', F''$  it holds that

$$H(F \cup F'|F'') \leq H(F|F'') + H(F'|F'').$$

*Remark 2.2.* If we extend  $H$  by setting  $H(\emptyset) = 0$ , then  $H(F|\emptyset) = H(F)$  for every  $F \in \mathcal{F}(G)$  and monotonicity wrt. the condition includes that  $H(F|F') \leq H(F)$ .

**Lemma 2.3.** *If  $H$  is monotone then the conditions (SS), (MC) and (CS) are equivalent.*

*Proof.* (SS)  $\implies$  (MC): For  $F' \subset F''$ ,

$$\begin{aligned} H(F|F'') &= H(F \cup F'') - H(F'') = H(F \cup F' \cup F'') - H(F'') \\ &\leq H(F \cup F') + H(F'') - H((F \cup F') \cap F'') - H(F'') \\ &\leq H(F \cup F') - H(F') = H(F|F'). \end{aligned}$$

(MC)  $\implies$  (CS) (monotonicity is not used):

$$\begin{aligned} H(F \cup F'|F'') &= H(F \cup F' \cup F'') - H(F'') + H(F' \cup F'') - H(F' \cup F'') \\ &= H(F|F' \cup F'') + H(F'|F'') \leq H(F|F'') + H(F'|F''). \end{aligned}$$

(CS)  $\implies$  (SS) (monotonicity is not used):

$$\begin{aligned} H(F \cup F') &= H(F \cup F' \cup (F \cap F')) = H(F \cup F'|F \cap F') + H(F \cap F') \\ &\leq H(F|F \cap F') + H(F'|F \cap F') + H(F \cap F') = \\ &H(F) - H(F \cap F') + H(F') - H(F \cap F') + H(F \cap F') = H(F) + H(F') - H(F \cap F'). \end{aligned}$$

□

**Proposition 2.4.** *If  $H$  is monotone then (SS)  $\implies$  (Sh)  $\implies$  (S) and none of the implications may be reversed.*

*Proof.* Assume that  $H$  is strongly subadditive. Let  $F = \{f_1, f_2, \dots, f_m\}$  and let  $\mathcal{K}$  be a  $k$ -cover of  $F$ . Using (2.1) we can write

$$H(F) = H(\{f_1\}) + H(\{f_2\}|\{f_1\}) + H(\{f_3\}|\{f_1, f_2\}) + \dots + H(\{f_m\}|\{f_1, \dots, f_{m-1}\})$$

and similarly, for each  $K \in \mathcal{K}$ ,

$$H(K) = \sum_{\{j: f_j \in K\}} H(\{f_j\}|\{f_i \in K : i < j\}).$$

By the preceding lemma,  $H$  fulfills (MC), hence  $H(\{f_j\}|\{f_i \in K : i < j\}) \geq H(\{f_j\}|\{f_1, \dots, f_{j-1}\})$ . Since each  $f_j$  belongs to at least  $k$  elements of  $\mathcal{K}$ , summing over  $\mathcal{K}$ , we obtain

$$\sum_{K \in \mathcal{K}} H(K) \geq \sum_{j=1}^m k H(\{f_j\}|\{f_1, \dots, f_{j-1}\}) = k H(F).$$

For the proof of (Sh)  $\implies$  (S) note that  $\{F, F' \setminus F\}$  is a 1-cover of  $F \cup F'$ , hence  $H(F \cup F') \leq H(F) + H(F' \setminus F) \leq H(F) + H(F')$ .

For counterexamples we ask the reader to see section 6. In Example 6.4 we show that topological entropy of the standard time-zero partition (which is also a cover) in the golden mean shift is not strongly subadditive. On the other hand, it satisfies Shearer's inequality, because the cover is disjoint (see Proposition 6.2).

In example 6.5 we present a  $\mathbb{Z}_3$ -action such that topological entropy of a certain cover does not satisfy Shearer's inequality (it does not even satisfy the infimum rule, see below). On the other hand, it is known that topological entropy of a cover is subadditive. □

## 3. THE INFIMUM RULE

Let  $G$  be an amenable group and let  $(F_n)$  be a selected Følner sequence. By  $|F|$  we will denote the cardinality of  $F$ .

**Definition 3.1.** *We will say that a nonnegative function  $H$  on  $\mathcal{F}(G)$  satisfies the infimum rule if*

$$\limsup_{n \rightarrow \infty} \frac{1}{|F_n|} H(F_n) = \inf_{F \in \mathcal{F}(G)} \frac{1}{|F|} H(F).$$

**Definition 3.2.** *We say that  $H$  is  $G$ -invariant if for any  $g \in G$  it holds that  $H(Fg) = H(F)$ .*

**Proposition 3.3.** *If a nonnegative and  $G$ -invariant function  $H$  on  $\mathcal{F}(G)$  satisfies Shearer's inequality then it also obeys the infimum rule.*

*Proof.* Clearly,  $\limsup_{n \rightarrow \infty} \frac{1}{|F_n|} H(F_n) \geq \inf_{F \in \mathcal{F}(G)} \frac{1}{|F|} H(F)$ .

For the converse inequality, fix an  $F \in \mathcal{F}(G)$  and  $\epsilon > 0$ . For  $n$  large enough the Følner set  $F_n$  is  $(F^{-1}, \epsilon)$ -invariant, i.e., it satisfies  $\frac{|F_n \Delta F^{-1}F_n|}{|F_n|} < \epsilon$ . The family

$$\mathcal{K} = \{Fg : g \in G, Fg \cap F_n \neq \emptyset\}$$

is a  $k$ -cover of  $F_n$  with  $k = |F|$  (for  $g \neq g'$ , we treat the sets  $Kg$  and  $Kg'$  as different elements of the  $k$ -cover, even if they are equal as sets). Indeed, for  $f \in F_n$ , the condition  $f \in Fg$  can be written as  $g \in F^{-1}f$  so, it is fulfilled for exactly  $k$  elements  $g$ . By the same calculation,  $Fg \cap F_n \neq \emptyset$  if and only if  $g \in F^{-1}F_n$ , so the cardinality of  $\mathcal{K}$  equals that of  $F^{-1}F_n$ , i.e., it is not more than  $|F_n|(1 + \epsilon)$ . By invariance of  $H$ ,  $H(K) = H(F)$  for every  $K \in \mathcal{K}$ . The Shearer's inequality now reads

$$H(F_n) \leq \frac{1}{k} \sum_{K \in \mathcal{K}} H(K) \leq \frac{1}{|F|} |F_n|(1 + \epsilon)H(F),$$

which, after dividing by  $|F_n|$  and passing with  $n$  to infinity, ends the proof.  $\square$

## 4. SHANNON ENTROPY

Let  $(X, \Sigma, \mu)$  be a probability space. A *partition* of  $X$  is a finite collection  $\mathcal{P}$  of pairwise disjoint measurable sets such that  $\bigcup_{A \in \mathcal{P}} A = X$ . By a *join* (or a *common refinement*) of partitions  $\mathcal{P}$  and  $\mathcal{P}'$  we mean the partition  $\mathcal{P} \vee \mathcal{P}' = \{A \cap B : A \in \mathcal{P}, B \in \mathcal{P}'\}$ . Now, let  $G$  be an amenable group acting on  $X$  via measurable maps, which preserve the measure  $\mu$ . If  $F$  is a finite subset of  $G$  we write  $\mathcal{P}^F$  for the common refinement  $\bigvee_{g \in F} g^{-1}\mathcal{P}$ , where  $g^{-1}\mathcal{P} = \{g^{-1}A : A \in \mathcal{P}\}$ . Recall that the *Shannon entropy* of a partition  $\mathcal{P}$  is defined by

$$H_\mu(\mathcal{P}) = - \sum_{A \in \mathcal{P}} \mu(A) \log \mu(A)$$

and *entropy of the action of  $G$  with respect to a partition  $\mathcal{P}$*  is defined as

$$h_\mu(G, \mathcal{P}) = \limsup_{n \rightarrow \infty} \frac{1}{|F_n|} H_\mu(\mathcal{P}^{F_n})$$

( $(F_n)$  is a Følner sequence in  $G$ ).

**Proposition 4.1.** *The function  $H(F) = H_\mu(\mathcal{P}^F)$  is*

- (1) *nonnegative,*

- (2)  $G$ -invariant,
- (3) monotone,
- (4) strongly subadditive.

These statements follow from standard properties of the entropy of a partition—for the proofs we refer the reader to any handbook on ergodic theory, ((SS) is usually replaced by (MC) or (CS), see e.g. [3, (1.6.7) and (1.6.9)]). In view of these facts, the following theorem follows from Propositions 2.4 and 3.3.

**Theorem 4.2.** *Measure-theoretic entropy obeys the infimum rule, i.e.*

$$h_\mu(G, \mathcal{P}) = \inf_{F \in \mathcal{F}(G)} \frac{1}{|F|} H_\mu(\mathcal{P}^F).$$

for every  $G$ -invariant measure  $\mu$  and every partition  $\mathcal{P}$ .

*Remark 4.3.* The above formula can be found, e.g. in [5], where it is attributed to Kolmogorov.

## 5. BEYOND AMENABILITY

The “mindblowingly” simple formula  $\inf_{F \in \mathcal{F}(G)} \frac{1}{|F|} H_\mu(\mathcal{P}^F)$  can be applied to processes under actions of any countable groups. (It can be applied to uncountable groups as well, however, it will typically yield zero; such is the case of flows.) To distinguish from other existing notions, we will denote it by  $h_\mu^*(G, \mathcal{P})$ . How good is this formula for countable non-amenable groups? The answer depends on the properties we expect from a good notion of dynamical entropy.

The notion  $h_\mu^*(G, \mathcal{P})$  has the following advantages:

- It is completely universal, can be defined for arbitrary groups.
- It is extremely simple, requires no details of the group (for instance in amenable groups it is formulated without referring to any Følner sequence).
- It satisfies *the Bernoulli shifts postulate*: Bernoulli shifts have “full” entropy (equal to the Shannon entropy of the independent generator).
- It has a very convincing interpretation for other processes (entropy is lost in finite-dimensional dependencies and *all such losses matter*).

Disadvantages can be detected by examining the action of the free group  $F_2$  with two generators, and they include:

- It fails *the factors postulate*: it can increase when passing to a factor.
- It fails *the invariance postulate*: it can change with change of a generator (hence is not an isomorphism invariant).

Before the examples, we recall the notions of the shift action and of a subshift. Suppose  $G$  is a group and  $\Lambda$  is a finite set with the discrete topology. By the *full shift* we understand the set  $\Lambda^G$  (whose elements are  $x = (x_g)_{g \in G}$ ) equipped with the product topology, on which  $G$  acts by shifts:  $(gx)_h = x_{hg}$ . A *subshift* is any closed  $G$ -invariant subset  $X \subset \Lambda^G$ . The full shift, as well as any subshift, admits a natural partition  $\mathcal{P}_\Lambda = \{[a] : a \in \Lambda\}$  by closed-and-open cylinder sets determined by the symbol “at zero”:  $[a] = \{x : x_e = a\}$  ( $e$  is the unity of  $G$ ). We call it the *time-zero partition* (or *time-zero cover*, depending on the context). The term *Bernoulli measure* is synonymous with a product measure  $\nu^G$  on  $\Lambda^G$ , where  $\nu$  is a probability measure on  $\Lambda$ .

*Example 5.1.* Let  $F_2$  denote the free group with two generators  $a$  and  $b$  and unity  $e$ , and consider  $X = \{-1, 1\}^{F_2}$  with the shift action, the Bernoulli  $(\frac{1}{2}, \frac{1}{2})$ -measure, and the zero-coordinate partition  $\mathcal{P} = \{-1, 1\}$ . Clearly,  $H(\mathcal{P}) = \log 2$  and  $h_\mu^*(F_2, \mathcal{P}) = \log 2$ . Next, consider the mapping  $\psi : X \rightarrow \{-1, 1\} \times \{-1, 1\}$  given by

$$\psi(x) = (x(e)x(a), x(e)x(b))$$

and the associated four-element partition  $\mathcal{R}$ . It is not hard to see that the process generated by  $\mathcal{R}$  is the  $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ -Bernoulli shift: the one-dimensional distributions are independent. So,  $H_\mu(\mathcal{R}) = \log 4$  and so equals  $h_\mu^*(F_2, \mathcal{R})$ . On the other hand, the process generated by  $\mathcal{R}$  is clearly a factor of that generated by  $\mathcal{P}$ .<sup>1</sup>

Now let  $E = \{e, a, b\} \subset F_2$  and consider  $\mathcal{Q} = \mathcal{P}^E$ . Clearly, this partition is another generator of the process generated by  $\mathcal{P}$  (the generated processes are isomorphic). For any finite set  $F \subset F_2$  we have  $H(\mathcal{Q}^F) = H(\mathcal{P}^{EF}) = |EF| \log 2$ . However, the ratio  $\frac{|EF|}{|F|}$  does not drop below 2 (and can be arbitrarily close to 2). Hence  $h_\mu^*(F^2, \mathcal{Q}) = 2h_\mu^*(F^2, \mathcal{P}) = \log 4$ .<sup>2</sup>

So, either we accept  $h_\mu^*(G, \mathcal{P})$  as a parameter associated with a concrete *process*, maintaining its simplicity and interpretation, or we try to force it to become an isomorphism invariant. As an attempt in this direction we propose two invariants, both equal to  $h_\mu^*(G, \mathcal{P})$  for actions of amenable groups. Unfortunately, we are unable to verify whether these new notions fulfill the Bernoulli shift condition in a more general case.

**Definition 5.2.**

$$h_\mu^{**}(X, G) = \inf\{h_\mu^*(G, \mathcal{P}) : \mathcal{P} \text{ is a generator}\},$$

$$h_\mu^{***}(X, G) = \inf\{H_\mu(\mathcal{P}) : \mathcal{P} \text{ is a generator}\}.$$

Note that the latter notion has nothing to do with  $h_\mu^*(G, \mathcal{P})$ , we were driven to it just by analogy to  $h_\mu^{**}(X, G)$ . For actions of amenable groups we know that  $h_\mu^{**}(X, G)$  equals  $h_\mu(X, G)$ . We also have  $h_\mu^{***}(X, G) = h_\mu(X, G)$  (see [7, Corollary 2.7]).

In the general case it is obvious that  $h^{**}(X, G) \leq h^{***}(X, G)$ . B. Seward [6] can prove the opposite inequality for free actions (i.e., such that for  $g \neq e$ , the set of points fixed by  $g$  has measure 0). As we already mentioned, we still do not know whether any of these notions satisfies the Bernoulli shifts postulate.<sup>3</sup>

---

<sup>1</sup>The above example shows that the failure of the factors postulate is inevitable for any entropy notion satisfying the Bernoulli shifts postulate (in particular, for *sofic entropy* [2]). For this reason it is commonly agreed to give up the factors postulate in the search for a universal (i.e., valid for a range of acting groups) notion of entropy.

<sup>2</sup>The strength of the notion of *sofic entropy* is that it behaves better in this aspect; it does not depend on the partition as long as it generates the whole process. So, sofic entropy can be viewed as a parameter associated to measure-preserving actions, and becomes an isomorphism invariant. On the other hand, sofic entropy has its disadvantages: it has a very complicated definition, and it applies to actions of sofic groups only.

<sup>3</sup>B. Weiss ([9]) can prove the Bernoulli shifts postulate for  $h^{***}(X, G)$  in actions of sofic groups, so  $h^{***}(X, G)$  becomes a serious competition for sofic entropy. It is unknown whether these two notions coincide for actions of sofic groups.

## 6. TOPOLOGICAL ENTROPY

In the present section we enter the world of topological dynamical systems. We assume that  $X$  is a compact metric space and  $G$  acts by homeomorphisms on  $X$ . Similarly to the measure-theoretic case, for an open cover  $\mathcal{U}$  and a finite  $F \subset \mathbb{N}$  we write  $\mathcal{U}^F$  for the refinement  $\bigvee_{f \in F} f^{-1}\mathcal{U}$ , where  $f^{-1}\mathcal{U} = \{f^{-1}U : U \in \mathcal{U}\}$ . We recall that  $N(\mathcal{U})$  is the smallest cardinality of a subcover chosen from a cover  $\mathcal{U}$  and that topological entropy of a cover  $\mathcal{U}$  is defined by  $H_{\text{top}}(\mathcal{U}) = \log(N(\mathcal{U}))$ . Topological entropy of the action is defined in two steps:

$$h_{\text{top}}(T, \mathcal{U}) = \limsup_{n \rightarrow \infty} \frac{1}{|F_n|} H_{\text{top}}(\mathcal{U}^{F_n}),$$

$$h_{\text{top}}(T) = \sup_{\mathcal{U}} h_{\text{top}}(T, \mathcal{U}),$$

where  $(F_n)$  is a Følner sequence and the supremum is taken over all open covers of  $X$ .

We want to study the nonnegative function on  $\mathcal{F}(G)$  obtained by fixing an open cover  $\mathcal{U}$  and abbreviating  $H_{\text{top}}(\mathcal{U}^F)$  as  $H_{\text{top}}(F)$ . It is obvious that  $H_{\text{top}}$  is monotone and  $G$ -invariant. It is also commonly known (and easily verified) that this function is subadditive (see e.g. [3, (6.3.8)]). The natural next step is the verification of Shearer's inequality. We begin with a discrete version of [1, Thm. 2]. The proof is almost literally copied.

**Lemma 6.1.** *Let  $\mathfrak{X}$  be a subset of  $\Lambda^n$ , where  $\Lambda$  is a finite set and  $n \in \mathbb{N}$ . Let  $\mathcal{K}$  be a  $k$ -cover of the set of coordinates  $\{1, 2, \dots, n\}$  (the elements of  $\mathcal{K}$  are nonempty subsets of  $\{1, 2, \dots, n\}$ , we admit repeated elements in  $\mathcal{K}$ , and each coordinate belongs to at least  $k$  elements of  $\mathcal{K}$ ). For  $K \in \mathcal{K}$  let  $\mathfrak{X}_K$  denote the projection of  $\mathfrak{X}$  onto the coordinates belonging to  $K$ . Then*

$$|\mathfrak{X}| \leq \prod_{K \in \mathcal{K}} |\mathfrak{X}_K|^{\frac{1}{k}}.$$

*Proof.* For  $n = 1$  the statement is obvious: each  $K \in \mathcal{K}$  equals  $\{1\}$ , each  $\mathfrak{X}_K$  equals  $\mathfrak{X}$  and the cardinality of  $\mathcal{K}$  is at least  $k$ . We proceed by induction. Consider an  $n \geq 2$  and suppose the statement holds for subsets of  $\Lambda^{n-1}$ . The set  $\mathfrak{X} \subset \Lambda^n$  splits into disjoint sets  $\mathfrak{X}^a$ ,  $a \in \Lambda$ , depending on the value at the last coordinate:

$$\mathfrak{X}^a = \{(x_1, x_2, \dots, x_n) \in \mathfrak{X} : x_n = a\}.$$

For  $K \in \mathcal{K}$  let  $K^\circ = K \setminus \{n\}$  and let  $\mathcal{K}^\circ = \{K^\circ : K \in \mathcal{K}\}$ . Clearly,  $\mathcal{K}^\circ$  is a  $k$ -cover of  $\{1, 2, \dots, n-1\}$ . For every  $a \in \Lambda$ ,  $\mathfrak{X}^a$  can be viewed as a subset of  $\Lambda^{n-1}$  (with the symbol  $a$  appended to each element), hence, by the inductive assumption, we have

$$|\mathfrak{X}^a| \leq \prod_{K \in \mathcal{K}^\circ} |\mathfrak{X}_{K^\circ}^a|^{\frac{1}{k}}.$$

Because in  $\mathfrak{X}^a$  the symbol at the last coordinate is determined, we have  $|\mathfrak{X}_{K^\circ}^a| = |\mathfrak{X}_K^a|$  for every  $K$ , and we get

$$|\mathfrak{X}^a| \leq \prod_{K \in \mathcal{K}} |\mathfrak{X}_K^a|^{\frac{1}{k}}.$$

Further,

$$|\mathfrak{X}| = \sum_{a \in \Lambda} |\mathfrak{X}^a|, \quad \text{and} \quad |\mathfrak{X}_K| = \sum_{a \in \Lambda} |\mathfrak{X}_K^a|,$$

for every  $K \in \mathcal{K}$  such that  $n \in K$ . For  $K$  not containing  $n$  we will apply the estimate  $|\mathfrak{X}_K^a| \leq |\mathfrak{X}_K|$  (regardless of  $a \in \Lambda$ ). And so, we have

$$(6.1) \quad |\mathfrak{X}| = \sum_{a \in \Lambda} |\mathfrak{X}^a| \leq \sum_{a \in \Lambda} \prod_{K \in \mathcal{K}} |\mathfrak{X}_K^a|^{\frac{1}{k}} = \sum_{a \in \Lambda} \left( \prod_{K \ni n} |\mathfrak{X}_K^a|^{\frac{1}{k}} \cdot \prod_{K \not\ni n} |\mathfrak{X}_K^a|^{\frac{1}{k}} \right) \\ \leq \prod_{K \not\ni n} |\mathfrak{X}_K|^{\frac{1}{k}} \cdot \sum_{a \in \Lambda} \prod_{K \ni n} |\mathfrak{X}_K^a|^{\frac{1}{k}}.$$

For each  $K$  containing  $n$ , on  $\Lambda$  we define a function  $f_K$ , by  $f_K(a) = |\mathfrak{X}_K^a|^{\frac{1}{k}}$ , and then we apply the generalized Hölder inequality:

$$\left\| \prod_{K \ni n} f_K \right\|_p \leq \prod_{K \ni n} \|f_K\|_k,$$

where  $\frac{1}{p} = \sum_{K \ni n} \frac{1}{k}$ . Because  $\mathcal{K}$  is a  $k$ -cover, this sum has at least  $k$  terms, hence  $p \leq 1$  (if  $p < 1$ , formally,  $\|\cdot\|_p$  is not a norm, but it does not matter). Since for a fixed finite-dimensional vector  $f$ , the term  $\|f\|_p$  is a decreasing function of  $p > 0$ , the above inequality holds also for  $p = 1$ , and then it reads:

$$\sum_{a \in \Lambda} \prod_{K \ni n} |\mathfrak{X}_K^a|^{\frac{1}{k}} \leq \prod_{K \ni n} \left( \sum_{a \in \Lambda} |\mathfrak{X}_K^a| \right)^{\frac{1}{k}} = \prod_{K \ni n} |\mathfrak{X}_K|^{\frac{1}{k}}.$$

Plugging this into (6.1) we end the proof.  $\square$

**Corollary 6.2.** *If  $\mathcal{U}$  consists of pairwise disjoint sets then the corresponding function  $H_{\text{top}}$  on  $\mathcal{F}(G)$  fulfills Shearer's inequality.*

*Proof.* Because the cover  $\mathcal{U}$  is by disjoint sets, for each  $F \in \mathcal{F}(G)$  we have  $N(\mathcal{U}^F) = |\mathcal{U}^F|$ , where  $\mathcal{U}^F$  is rid of empty elements. Labeling  $\mathcal{U}$  by elements of a finite alphabet  $\Lambda$  (of the same cardinality as  $\mathcal{U}$ ),  $\mathcal{U}^F$  can be identified with a set  $\mathfrak{X} \subset \Lambda^F$ , while for any  $K \subset F$  we have  $\mathcal{U}^K = \mathfrak{X}_K$ . Now,  $H_{\text{top}}(F) = \log |\mathfrak{X}|$  and  $H_{\text{top}}(K) = \log |\mathfrak{X}_K|$  and the assertion follows directly from Lemma 6.1.  $\square$

**Corollary 6.3.** *If  $(X, G)$  is a subshift and  $\mathcal{U} = \mathcal{P}_\Lambda$  then the infimum rule holds for  $H_{\text{top}}$ , i.e.,*

$$h_{\text{top}}(T) = h_{\text{top}}(T, \mathcal{U}) = \inf_{F \in \mathcal{F}(G)} \frac{1}{|F|} H_{\text{top}}(\mathcal{U}^F).$$

*Proof.* Follows immediately from Propositions 6.2 and 3.3.  $\square$

The following example shows that topological entropy is not in general strongly subadditive, even for subshifts. To make matters worse, in another example we will present a system and a (non-disjoint) cover such that both Shearer's inequality and infimum rule fail.

*Example 6.4.* The golden mean shift is a subshift  $X \subset \{0, 1\}^{\mathbb{Z}}$  consisting of all sequences in which block 11 does not occur. Let  $\mathcal{U} = \mathcal{P}_\Lambda$  be the two-element time-zero cover. It was shown already that Shearer's inequality holds, but it can easily be checked that

$$H_{\text{top}}(\{-1, 0, 1\}) + H_{\text{top}}(\{0\}) = \log 5 + \log 2 \geq \log 3 + \log 3 \\ = H_{\text{top}}(\{-1, 0\}) + H_{\text{top}}(\{0, 1\}),$$

i.e., strong subadditivity does not hold.



The following example looks innocent, but it took us a lot of effort to find it.

*Example 6.5.* Let  $\Lambda = \{a, b, c\}$ . We consider the action of the group  $\mathbb{Z}_3 = \mathbb{Z}/3\mathbb{Z} = \{0, 1, 2\}$ , by shifts, on a space  $X \subset \Lambda^{\mathbb{Z}_3}$  (i.e., a subshift) defined by

$$X = \{(a, a, a), (b, b, b), (c, c, c), (a, b, c), (b, c, a), (c, a, b)\}.$$

(In fact,  $(c, c, c)$  can be dropped and the example with 5 elements will still work.) It is clear that  $X$  is closed and  $\mathbb{Z}_3$ -invariant. The collection  $\mathcal{V} = \{\{a, b\}, \{b, c\}, \{a, c\}\}$  is a (non-disjoint) cover of  $\Lambda$ . Let  $\bar{\mathcal{V}} = \{\bar{V} : V \in \mathcal{V}\}$ , where  $\bar{V} = V \times \Lambda \times \Lambda$ . Clearly,  $\bar{\mathcal{V}}$  is a cover of  $\Lambda^{\mathbb{Z}_3}$  hence also of  $X$ . The elements of  $\bar{\mathcal{V}}^{\mathbb{Z}_3}$  have the form  $V_1 \times V_2 \times V_3$ , where  $V_1, V_2, V_3$  are (not necessarily distinct) elements of  $\mathcal{V}$ . It is easy to check that  $X$  admits a subcover of  $\bar{\mathcal{V}}^{\mathbb{Z}_3}$  consisting of three sets, namely

$$\{a, b\} \times \{a, b\} \times \{a, b\}, \quad \{a, c\} \times \{b, c\} \times \{a, c\}, \quad \{b, c\} \times \{a, c\} \times \{a, b\}.$$

On the other hand, there is no subcover with two elements (one set of the form  $V_1 \times V_2 \times V_3$  may contain at most two elements of  $X$ , because in any three of them, on some coordinate there appear all three letters). Thus  $N(\bar{\mathcal{V}}^{\mathbb{Z}_3}) = 3$  and

$$h_{\text{top}}(\mathbb{Z}_3, \bar{\mathcal{V}}) = \frac{1}{|\mathbb{Z}_3|} H_{\text{top}}(\bar{\mathcal{V}}^{\mathbb{Z}_3}) = \frac{1}{3} \log 3$$

(in any finite group  $G$ , for any Følner sequence, eventually  $F_n = G$ ). However, the cover  $\bar{\mathcal{V}}^{\{0,1\}}$  has a minimal subcover consisting of only two sets (for example  $\{b, c\} \times \{a, c\}$  and  $\{a, b\} \times \{a, b\}$ ), and  $H_{\text{top}}(\bar{\mathcal{V}}^{\{0,1\}}) = \log 2$ . Since  $\frac{1}{2} \log 2 < \frac{1}{3} \log 3$ , the infimum rule does not hold. Consequently, by Proposition 3.3 (or by direct verification for the 2-cover of  $\mathbb{Z}_3$  by  $\{0, 1\}, \{1, 2\}, \{0, 2\}$ ), Shearer's inequality fails as well.

*Remark 6.6.* Shearer's inequality depends only vaguely on the acting group. The example (as a counterexample for Shearer's inequality) can be easily adapted to the action of any group with at least 3 elements, in particular of  $\mathbb{Z}$ . The infimum rule depends more heavily on the acting group (its proof uses only  $k$ -covers obtained by shifting one set  $F$  of cardinality  $k$ ), so the following question arises:

*Question 6.7.* Let  $G$  be an infinite countable amenable group acting on a compact space  $X$  and let  $\mathcal{U}$  be an open cover of  $X$ . Does the infimum rule hold for  $H_{\text{top}}(F) = H_{\text{top}}(\mathcal{U}^F)$ , i.e., is it true that  $h_{\text{top}}(G, \mathcal{U}) = h_{\text{top}}^*(G, \mathcal{U})$ ?

In spite of many efforts, we have not succeeded in answering this questions even for  $G = \mathbb{Z}$ . In fact, we do not even know toward which answer should we incline. Let us discuss the difficulties more extensively.

Our example for  $\mathbb{Z}_3$  works only because the two-element set  $\{0, 1\}$  and its shifts form a *non-splitting* 2-cover of  $\mathbb{Z}_3$ . In  $\mathbb{Z}$ , an analogous 2-cover is splitting (i.e., it splits as a union of two 1-covers) and it is easy to show, using plain subadditivity, that for a splitting  $k$ -cover of  $G$ , obtained by shifting one set  $F$ , the Shearer's inequality nearly holds (up to a small error) on large elements of the Følner sequence. This suffices to prove the infimum rule. The simplest finite subset of  $\mathbb{Z}$  whose shifts produce a non-splitting 3-cover is  $\{0, 1, 3\}$ , but we failed to find a counterexample for Shearer's inequality, based on any 3-cover. The point is that the key inequality  $\frac{1}{k} \log k \leq \frac{1}{k+1} \log(k+1)$  does not hold for  $k > 2$ .

Nevertheless, we are able to formulate a positive result. In its proof we benefit from the fact that the infimum rule works for the Shannon entropy and we apply

the variational principle (for amenable groups it was first proved in [8]). Let us recall: if  $\mathcal{M}_G(X)$  is the collection of all  $G$ -invariant probability measures on  $X$  then  $h_{\text{top}}(G) = \sup_{\mu \in \mathcal{M}_G(X)} h_\mu(G)$ . It suffices to take the supremum over ergodic measures.

**Theorem 6.8.** *Let  $X$  be a compact metric space and  $G$  a countable amenable group acting on  $X$  by continuous maps. Define*

$$h_{\text{top}}^*(G, \mathcal{U}) = \inf_F \frac{1}{|F|} H_{\text{top}}(\mathcal{U}^F)$$

$$h_{\text{top}}^*(G) = \sup_{\mathcal{U}} h_{\text{top}}^*(G, \mathcal{U})$$

Then

$$h_{\text{top}}^*(G) = h_{\text{top}}(G)$$

*Proof.* Clearly,  $h_{\text{top}}^*(G) \leq h_{\text{top}}(G)$ .

To obtain the converse inequality, consider an ergodic measure  $\mu$  on  $X$ , a finite partition  $\mathcal{P} = \{A_1, \dots, A_p\}$  of  $X$ , and fix an  $\epsilon > 0$ . For any  $\delta > 0$  one can choose compact sets  $B_i \subset A_i$ ,  $i = 1, \dots, p$ , so that  $\mu(A_i \setminus B_i) < \delta/p$ . Additionally, let  $B_0 = X \setminus \bigcup_{i=1}^p B_i$  and define  $U_i = B_0 \cup B_i$ ,  $i = 1, \dots, p$ . Then  $\mathcal{U} = \{U_1, \dots, U_p\}$  is a cover with  $U_i \supset A_i$ ,  $\mu(U_i \setminus A_i) < \delta$ . The family  $\xi = \{B_0, B_1, \dots, B_p\}$  is a partition of  $X$ . For any  $F \in \mathcal{F}(G)$ , we have

$$H_\mu(\mathcal{P}^F) \leq H_\mu(\mathcal{P}^F \vee \xi^F) = H_\mu(\mathcal{P}^F | \xi^F) + H_\mu(\xi^F) \leq |F| H_\mu(\mathcal{P} | \xi) + H_\mu(\xi^F)$$

(we have used subadditivity of the function  $H(F) = H_\mu(\mathcal{P}^F | \xi^F)$ , see e.g. [3, (1.6.33)]). Choosing  $\delta$  appropriately small, we may demand that  $H_\mu(\mathcal{P} | \xi) < \epsilon$ , so that

$$(6.2) \quad H_\mu(\mathcal{P}^F) \leq H_\mu(\xi^F) + |F|\epsilon.$$

Let  $(F_n)$  be a *tempered* Følner sequence (we skip the definition, every Følner sequence has a tempered subsequence, the ergodic theorem holds along any tempered Følner sequence, see [4]). By the ergodic theorem, for almost all  $x$  it holds that

$$\lim_{n \rightarrow \infty} \frac{1}{|F_n|} |\{g \in F_n : gx \in B_0\}| = \mu(B_0) < \epsilon.$$

The same is true for  $f^{-1}B_0$  ( $f \in F$ ), which implies that for almost all  $x$ ,

$$\lim_{n \rightarrow \infty} \frac{1}{|F_n|} |\{g \in F_n : fgx \in B_0\}| < \epsilon$$

for all  $f \in F$ . Therefore, we can choose  $N \in \mathbb{N}$  such that the set

$$X_{\epsilon, N} = \bigcap_{n \geq N} \bigcap_{f \in F} \left\{ x \in X : \frac{1}{|F_n|} |\{g \in F_n : fgx \in B_0\}| < \epsilon \right\}$$

has positive measure. For  $x \in X_{\epsilon, N}$  and  $n \geq N$  we have

$$|\{(g, f) : f \in F, g \in F_n, fgx \in B_0\}| < |F_n| \cdot |F| \cdot \epsilon,$$

which implies

$$(6.3) \quad |\{g \in F_n : |\{f \in F : fgx \in B_0\}| > |F| \sqrt{\epsilon}\}| \leq |F_n| \sqrt{\epsilon}.$$

Consider a collection of all sets of the form  $\bigcap_{f \in F} f^{-1}C_f$ , where  $C_f = B_0$  or  $C_f = X \setminus B_0$ , such that the first case occurs for at most  $|F| \sqrt{\epsilon}$  indices  $f$ . Let  $Y$  be the union of all such sets. Note that  $Y$  is a union of elements of the partition

$\xi^F$ . The ratios  $\frac{1}{|F_n|} |\{g \in F_n : gx \in Y\}|$  converge to  $\mu(Y)$  for  $\mu$ -almost every  $x$ . On the other hand,  $gx \in Y$  is equivalent to the fact that  $fgx \in B_0$  for at most  $|F| \sqrt{\epsilon}$  elements  $f \in F$ . If, in addition, we pick  $x \in X_{\epsilon, N}$  then (6.3) implies that for every  $n \geq N$ ,

$$\frac{1}{|F_n|} |\{g \in F_n : gx \in Y\}| \geq 1 - \sqrt{\epsilon}.$$

Since  $X_{\epsilon, N}$  has positive measure, it contains a point which fulfills the ergodic theorem for  $Y$ , implying that  $\mu(Y) \geq 1 - \sqrt{\epsilon}$ .

Let  $\mathcal{U}'$  be a subcover of  $\mathcal{U}^F$  with minimal cardinality. Denoting by  $\xi_Y^F$  the collection  $\{B \in \xi^F : B \cap Y \neq \emptyset\}$  (note that we can as well write  $B \subset Y$ , so  $\xi_Y^F$  is a partition of  $Y$ ), we obviously have

$$|\xi_Y^F| \leq \sum_{U \in \mathcal{U}'} |\{B \in \xi_Y^F : B \cap U \neq \emptyset\}|.$$

Fix some  $U \in \mathcal{U}'$ ,  $U = \bigcap_{f \in F} f^{-1}U_{j_f}$ ,  $1 \leq j_f \leq p$ . Consider a  $B \in \xi_Y^F$  satisfying  $B \cap U \neq \emptyset$ . Then  $B$  is a cylinder in  $\xi^F$ ,  $B = \bigcap_{f \in F} f^{-1}B_{k_f}$ ,  $0 \leq k_f \leq p$ , and because  $B$  is contained in  $Y$ ,  $k_f = 0$  may occur for not more than  $|F| \sqrt{\epsilon}$  indices  $f \in F$ . If  $k_f \neq 0$  then it must be equal to  $j_f$ , because otherwise  $B_{k_f}$  and  $U_{j_f}$  would be disjoint, implying  $B \cap U = \emptyset$ . Therefore, for given  $U$  the number of sets  $B \in \xi_Y^F$  intersecting  $U$  is estimated by the number of ways in which the (few) indices 0 can be distributed over the  $k_f$ 's, i.e.,

$$|\{B \in \xi_Y^F : B \cap U \neq \emptyset\}| \leq \sum_{i=0}^{\lfloor |F| \sqrt{\epsilon} \rfloor} \binom{|F|}{i}.$$

It is well known that  $\log \binom{n}{k} \leq nH(\frac{k}{n})$ , where  $H(\delta)$  is the binary entropy of the vector  $(\delta, 1 - \delta)$ , so we get

$$(6.4) \quad \log |\xi_Y^F| \leq H_{\text{top}}(\mathcal{U}^F) + |F| H(\sqrt{\epsilon}) + \log |F| \sqrt{\epsilon}.$$

We still need to compare the partition  $\xi_Y^F$  with  $\xi^F$ . Let  $\mathcal{R}$  be the partition  $\{Y, X \setminus Y\}$  and let  $\mu_Y$  be the normalized conditional measure induced by  $\mu$  on  $Y$ . We have

$$\begin{aligned} H_{\mu}(\xi^F) &\leq H_{\mu}(\xi^F \vee \mathcal{R}) = H_{\mu}(\xi^F | \mathcal{R}) + H(\mathcal{R}) \\ &= \mu(Y) H_{\mu_Y}(\xi_Y^F) + \mu(X \setminus Y) H_{\mu_{X \setminus Y}}(\xi^F) + H(\mathcal{R}) \\ &\leq H_{\mu_Y}(\xi_Y^F) + \sqrt{\epsilon} \cdot \log |\xi^F| + H(\sqrt{\epsilon}). \end{aligned}$$

Eventually, combining the above with (6.2) and (6.4) (and the standard estimate of entropy by means of cardinality), we get

$$H_{\mu}(\mathcal{P}^F) \leq H_{\text{top}}(\mathcal{U}^F) + |F| H(\sqrt{\epsilon}) + \log |F| \sqrt{\epsilon} + \sqrt{\epsilon} \cdot \log |\xi^F| + H(\sqrt{\epsilon}) + |F| \epsilon.$$

Dividing both sides by  $|F|$  and noticing that all but the first term on the right can be made arbitrarily small, regardless of  $F$ , by the choice of  $\epsilon$ , next taking infima over  $F$  on both sides, and supremum over  $\mathcal{U}$  on the right (which eliminates the small error terms), we arrive to

$$\inf_F \frac{1}{|F|} H_{\mu}(\mathcal{P}^F) \leq h_{\text{top}}^*(G).$$

Now Theorem 4.2 allows to replace the left hand side by  $h_\mu(G, \mathcal{P})$ . Since the inequality holds for any  $\mathcal{P}$  and any ergodic  $\mu$ , taking appropriate suprema (and applying the variational principle), we can further replace the left hand side by  $h_{\text{top}}(G)$ , concluding the proof.  $\square$

## REFERENCES

1. B. Bollobás and A. Thomason, *Projections of bodies and hereditary properties of hypergraphs*, Bull. London Math. Soc. **27** (1995), 417–424.
2. L. Bowen, *Measure conjugacy invariants for actions of countable sofic groups*, J. Amer. Math. Soc. **23** (2010), 217–245.
3. T. Downarowicz, *Entropy in dynamical systems*, Cambridge University Press, *New Mathematical Monographs 18*, Cambridge 2011.
4. E. Lindenstrauss, *Pointwise theorems for amenable groups*, Electronic Research Announcements of AMS, **5** (1999).
5. B. Seward, *Krieger’s finite generator theorem for ergodic actions of countable groups II*, preprint, <http://arxiv.org/pdf/1501.03367.pdf>
6. B. Seward, private communication.
7. B. Seward and R. Tucker-Drob, *Borel structurability on the 2-shift of a countable group*, preprint. <http://arxiv.org/pdf/1402.4184.pdf>
8. A. Stepin and A. Tagi-Zade, *Variational characterization of topological pressure of the amenable groups of transformations* (Russian), Dokl. Akad. Nauk SSSR **254** (1980), 545–549.
9. B. Weiss, private communication.

Institute of Mathematics, Polish Academy of Science, Sniadeckich 8, 00-656 Warsaw, Poland

downar@pwr.edu.pl

Department of Mathematics, Wrocław University of Technology, Wybrzeże Wyspińskiego 27, 50-370 Wrocław, Poland

frej@pwr.edu.pl

Departamento de Matemáticas, Facultad Ciencias Exactas, Universidad Andres Bello, República 252, Santiago, Chile

promagnoli@unab.cl