FUNDAMENTA
MATHEMATICAE
213 (2011)

An integral formula for entropy of doubly stochastic operators

by

Bartosz Frej and **Paulina Frej** (Wrocław)

Abstract. A new formula for entropy of doubly stochastic operators is presented. It is also checked that this formula fulfills the axioms of the axiomatic definition of operator entropy, introduced in an earlier paper of Downarowicz and Frej. As an application of the formula the 'product rule' is obtained, i.e. it is shown that the entropy of a product is the sum of the entropies of the factors. Finally, the proof of continuity of the new 'static' entropy as a function of the measure is given.

1. Introduction. The classical notion of a dynamical system as a quadruple $(X, \mathscr{B}, \mu, T)$, where $(X, \mathscr{B}, \mu)$ is a probability space and $T\colon X \to X$ is a measure preserving map, gives rise to many generalizations. One of the possible directions is the concept of a *doubly stochastic operator* on $L^1(\mu)$, i.e. a linear (continuous) operator $T\colon L^1(\mu) \to L^1(\mu)$ satisfying the following conditions:

(i) Tf is a positive function if f is positive,
(ii) $T\mathbb{1} = \mathbb{1}$, where $\mathbb{1}$ is the constant function equal to 1 everywhere,
(iii) $\int_X Tf\,d\mu = \int_X f\,d\mu$.

The class of all doubly stochastic operators includes Koopman operators of all measure preserving transformations and, more generally, all operators associated to the stationary transition probabilities $P(\cdot,\cdot)$ by the formula $Tf(x) = \int f(y)\,P(x, dy)$. Since each space L^p, $p \geq 1$, is invariant under the action of a doubly stochastic operator and, on the other hand, an operator on L^p satisfying (i)–(iii) uniquely extends to a doubly stochastic operator, the domain of an operator may be any L^p space. Our main interest lies in transferring the concept of entropy to such operators.

2010 *Mathematics Subject Classification*: Primary 28D20; Secondary 47A35.
Key words and phrases: doubly stochastic operator, Markov operator, entropy, product rule.

DOI: 10.4064/fm213-3-6

In the literature one can find various generalizations of the notion of entropy (see e.g. [AF], [CNT], [GLW], [M], [MR] and [V]). Some of them were designed exclusively for doubly stochastic operators, while some concern more general cases, including doubly stochastic operators as a special case. A natural question is whether these notions coincide on doubly stochastic operators. A partial answer was given in [DF], where an axiomatic theory of operator entropy was established. The axiomatic theory assumes that entropy is constructed in the following steps:

(1) one specifies a T-invariant collection $\mathbf{F}$ of finite families $\mathcal{F}$ of measurable functions; a family is understood as a set or a finite sequence of functions; it is also convenient if $\mathbf{F}$ contains a distinguished trivial family $\mathcal{O}$ invariant under T;
(2) one defines an operation $\sqcup$ of *joining* families, so that $\mathcal{F}\sqcup\mathcal{G} \in \mathbf{F}$ whenever $\mathcal{F} \in \mathbf{F}$ and $\mathcal{G} \in \mathbf{F}$, and the cardinality of the join is bounded by a number depending on the cardinalities of the components; the operation is assumed to be associative and commutative in the sense that $\mathcal{F} \sqcup \mathcal{G}$ consists of the same functions as $\mathcal{G} \sqcup \mathcal{F}$, possibly enumerated in a different order; the trivial family satisfies $\mathcal{F} \sqcup \mathcal{O} = \mathcal{O} \sqcup \mathcal{F}$ for every $\mathcal{F} \in \mathbf{F}$;
(3) one defines the *static* (independent of the dynamics induced by an operator) entropy $H_\mu(\mathcal{F})$ of a family $\mathcal{F} \in \mathbf{F}$ with respect to μ; it is required that $H_\mu(\mathcal{F})$ does not depend on the possible enumeration of the elements of $\mathcal{F}$ and that $H_\mu(\mathcal{O}) = 0$;
(4) denoting
$$\mathcal{F}^n = \bigsqcup_{k=0}^{n-1} T^k \mathcal{F}$$
one then defines
$$h_\mu(T, \mathcal{F}) = \limsup_{n\to\infty} \frac{1}{n} H_\mu(\mathcal{F}^n);$$
(5) and eventually one sets
$$h_\mu(T) = \sup_{\mathcal{F}\in\mathbf{F}} h_\mu(T, \mathcal{F}).$$

The conditional entropy is given by
$$H_\mu(\mathcal{F} \mid \mathcal{G}) = H_\mu(\mathcal{F} \sqcup \mathcal{G}) - H_\mu(\mathcal{G}).$$
The entropies defined in [AF] (after restricting to doubly stochastic operators), [GLW], [M] and an explicit formula for entropy given in [DF] fit in the above scheme (but see Remark 2.6), though they differ in the choice of $\mathbf{F}$, in the definition of the join operation $\sqcup$ and in the definition of the static entropy $H_\mu(\mathcal{F})$. To prove that they give the same value of $h_\mu(T)$ one only

needs to verify four axioms concerning properties of $H_\mu(\mathcal{F})$: subadditivity, monotonicity, continuity with respect to the family $\mathcal{F}$ and compatibility with the Shannon entropy of a partition. It turns out that the same value of entropy of a doubly stochastic operator is obtained if one uses the definition given in [MR], but since in this approach the notion of the join is absent, the argument does not employ the axiomatic theory (see [F]). In the present paper the theory of entropy of doubly stochastic operators is developed in the spirit of [DF]. Relations to the entropies defined in [CNT] and [V] have not been studied yet and they will not be investigated in the current paper.

Below we state the axioms formulated in [DF]:

(A) Monotonicity and subadditivity axiom. For $\mathcal{F}$, $\mathcal{G}$ and $\mathcal{H}$ belonging to $\mathbf{F}$,

$$0 \le H_\mu(\mathcal{F}\,|\,\mathcal{H}) \le H_\mu(\mathcal{F}\sqcup\mathcal{G}\,|\,\mathcal{H}) \le H_\mu(\mathcal{F}\,|\,\mathcal{H}) + H_\mu(\mathcal{G}\,|\,\mathcal{H}),$$

where, by convention, $H_\mu(\mathcal{F}\,|\,\mathcal{O}) = H_\mu(\mathcal{F})$.

Definition 1.1. Let $r' \le r$. For two families of measurable functions $\mathcal{F}=\{f_1,\dots,f_r\}$ and $\mathcal{G}=\{g_1,\dots,g_{r'}\}$ the L^1-distance, denoted by $\operatorname{dist}(\mathcal{F},\mathcal{G})$, is defined by

$$\operatorname{dist}(\mathcal{F},\mathcal{G}) = \min_\pi\Big\{\max_{1\le i\le r}\int|f_i - g_{\pi(i)}|\,d\mu\Big\},$$

where the minimum ranges over all permutations π of $\{1,\dots,r\}$ and where $\mathcal{G}$ is considered an r-element family by setting $g_i \equiv 0$ for $r' < i \le r$.

(B) L^1-Continuity axiom. For every $r \ge 1$ and $\varepsilon > 0$ there is a $\delta > 0$ such that if $\mathcal{F}$, $\mathcal{G}$ and $\mathcal{H}$ have cardinalities at most r and $\operatorname{dist}(\mathcal{F},\mathcal{G}) < \delta$ then

$$\operatorname{dist}(\mathcal{F}\sqcup\mathcal{H},\mathcal{G}\sqcup\mathcal{H}) < \varepsilon, \qquad |H_\mu(\mathcal{F}) - H_\mu(\mathcal{G})| < \varepsilon.$$

(C) Partitions axiom. If Ξ is a measurable partition of X, let $\mathbb{1}_\Xi = \{\mathbb{1}_A : A \in \Xi\}$ denote the family of the corresponding characteristic functions. Then $\mathbf{F}$ contains $\mathbb{1}_\Xi$ for every measurable partition Ξ of X. The entropy H_μ coincides on partitions with the classical notion in the sense that

$$H_\mu(\mathbb{1}_\Xi) = H_\mu(\Xi) = -\sum_{A\in\Xi}\mu(A)\log\mu(A)$$

and

$$H_\mu\Big(\bigsqcup_{i=1}^n \mathbb{1}_{\Xi_i}\Big) = H_\mu\Big(\bigvee_{i=1}^n \Xi_i\Big).$$

(D) Domination axiom. For every $r \ge 1$ and $\varepsilon > 0$ there exists $\delta > 0$ such that for every family $\mathcal{F} = \{f_1,\dots,f_r\}$ and every partition α

of the unit interval into finitely many subintervals of lengths not exceeding δ,

$$H_\mu(\mathcal{F} \mid \mathbb{1}_{\bigvee_i f_i^{-1}(\alpha)} \sqcup \overline{\alpha}) < \varepsilon,$$

where $\overline{\alpha}$ is some finite family depending only on α and satisfying $\lim_n n^{-1} H_\mu(\bigsqcup_{k=1}^n \overline{\alpha}) = 0$ (usually $\overline{\alpha}$ is the empty family or a family of some constant functions).

Ergodic theory is often considered on compact spaces. A continuous map of X is then a source of various metric dynamical systems, which arise after fixing an invariant measure. If $C(X)$ denotes the space of all real valued continuous functions on X, then a positive linear operator $T : C(X) \to C(X)$ which preserves constants is called a *Markov operator*. It is well known that on a metrizable space every Markov operator T is generated by the transition probability $P(x,\cdot\,) = T^*\delta_x$, where δ_x is the point mass at x and T^* is the operator adjoint to T, acting on the dual (to $C(X)$) space of signed Radon measures on X. Such a transition probability is called *Feller*; it is a continuous map from X into the set of probability measures with the weak* topology. The set of all T^*-invariant (i.e. satisfying $\int Tf\,d\mu = \int f\,d\mu$ for every continuous f) Radon probability measures μ is a non-empty convex set, compact in the weak* topology. For every such measure μ the operator T becomes a doubly stochastic operator.

In the classical (non-operator) case a measurable space X may be interpreted as the phase space of a physical system, where a finite r-element partition models an experiment performed on that system, giving r possible outcomes. The apparatus used to measure the outcomes of the experiment is assumed to be faultless, i.e. in each state of the system (point of a phase space) it yields an outcome unambiguously assigned to this state. Doubly stochastic operators may be used to deal with situations in which the measurement is disturbed or unclear; in each state the machinery gives outcomes according to some probability distribution. The action of an operator models a change in settings of the measuring tool or a flow of time. One would expect that the entropy of the family $\mathcal{F}$ satisfies the following conditions:

(i) if $\mathcal{F}$ consists solely of constant functions then its entropy is zero, because an experiment modeled by such family yields the same results regardless of the state of the system, providing no information about the actual state;
(ii) $H_\mu(\mathcal{F} \mid \mathcal{F}) = 0$ for every family $\mathcal{F}$, because copying the results of an experiment performed before does not give any new information.

The axioms from [DF] do not guarantee that these properties are satisfied and, in fact, none of the above mentioned versions of entropy has both these properties at the same time. Indeed, the entropies introduced in [AF], [GLW] and [M] violate the first condition, because they use pointwise multiplication of functions in the role of $\sqcup$. This makes the family $\mathcal{F} \sqcup \mathcal{F}$ essentially bigger than $\mathcal{F}$ and may cause $H_\mu(\mathcal{F} \sqcup \mathcal{F})$ to be strictly greater than $H_\mu(\mathcal{F})$. The [DF]-entropy avoids this problem, but here the entropy of a family of constant functions is strictly positive whenever at least one of these constants is neither 0 nor 1 (see Section 2 for the definition of [DF]-entropy). The goal of the current paper is to introduce a new formula for static entropy, compatible with the axioms and having both the desirable properties. Thus the paper refines the theory of operator entropy, but it does not introduce any new definition of the *dynamical* entropy $h_\mu(T)$.

In Section 2 we present the formula and check that it satisfies the axioms, which are rephrased in Lemmas 2.3, 2.5, 2.7 and 2.8. In Section 3 we prove that the limit in the definition of $h_\mu(T, \mathcal{F})$ exists (see step (4) above). In Section 4 we use the new formula to show that, similarly to the classical case, the entropy of a product system is equal to the sum of the entropies of the factors. Finally, the topological case is studied, namely, we show that the entropy $H_\mu(\mathcal{F})$ is continuous when considered as a function of the measure, with $\mathcal{F}$ being a fixed family of continuous functions (Section 5).

2. A formula for static entropy. Let $(X, \mathscr{B}, \mu)$ be a probability space. Denote $\eta(x) = -x \log x$ for $x \in (0,1]$ and $\eta(0) = 0$ (log means logarithm to base 2). For a function $f \colon X \to [0,1]$ let $A_f = \{(x,t) \in X \times [0,1] \colon t \le f(x)\}$ and denote by $\mathscr{A}_f$ the partition of $X \times [0,1]$ consisting of A_f and its complement. For a collection $\mathcal{F}$ of measurable functions we define $\mathscr{A}_{\mathcal{F}} = \bigvee_{f \in \mathcal{F}} \mathscr{A}_f$. Denote by A^t the t-section of a set $A \subset X \times [0,1]$ at t, i.e. $A^t = \{x \in X : (x,t) \in A\}$, and by $\mathscr{A}^t_{\mathcal{F}}$ the partition of X consisting of the t-sections A^t, where $A \in \mathscr{A}_{\mathcal{F}}$. Following the scheme of the introduction we declare that $\mathbf{F}$ is the collection of all finite sequences of measurable functions from X into the unit interval and the join will be the concatenation of such sequences (1). We assume that $\mathbf{F}$ also contains an empty sequence $\mathcal{O}$ and that $\mathscr{A}_{\mathcal{O}} = \{X\}$. Clearly, $\mathscr{A}_{\mathcal{F} \sqcup \mathcal{G}} = \mathscr{A}_{\mathcal{F}} \vee \mathscr{A}_{\mathcal{G}}$ and $(\mathscr{A}_{\mathcal{F} \sqcup \mathcal{G}})^t = (\mathscr{A}_{\mathcal{F}} \vee \mathscr{A}_{\mathcal{G}})^t = \mathscr{A}^t_{\mathcal{F}} \vee \mathscr{A}^t_{\mathcal{G}}$. We define the entropy of the collection $\mathcal{F}$ by the formula

$$\overline{H}(\mathcal{F}) = \int_0^1 H_\mu(\mathscr{A}^t_{\mathcal{F}})\, d\lambda(t), \tag{2.1}$$

where $H_\mu(\alpha)$ is the classical Shannon entropy of the partition α. It is obvious that this quantity has both properties mentioned in the introduction: the entropy of a collection of constant functions is zero, as is the conditional entropy $\overline{H}(\mathcal{F} \,|\, \mathcal{F})$. Recall that in [DF] the static entropy of $\mathcal{F}$ (we denote it

(1) In [DF] and [F] the join of $\mathcal{F}$ and $\mathcal{G}$ was denoted by $\mathcal{F} \cup \mathcal{G}$ and it was claimed that it is just the set-theoretic union. We comment on it in Remark 2.6.

by $H^{\mathrm{DF}}(\mathcal{F})$) was defined as the Shannon entropy of $\mathscr{A}_{\mathcal{F}}$ with respect to the product of μ and the Lebesgue measure,

$$H^{\mathrm{DF}}(\mathcal{F}) = H_{\mu\times\lambda}(\mathscr{A}_{\mathcal{F}}) = \sum_{A\in\mathscr{A}_{\mathcal{F}}} \eta((\mu\times\lambda)(A)).$$

Before we check that (2.1) satisfies the axioms, we make the following easy observation.

Proposition 2.1. $\overline{H}(\mathcal{F}) \le H^{\mathrm{DF}}(\mathcal{F})$.

Proof. By Jensen's inequality,

$$\begin{aligned}\overline{H}(\mathcal{F}) &= \sum_{A\in\mathscr{A}_{\mathcal{F}}} \int_0^1 \eta(\mu(A^t))\,d\lambda(t) \le \sum_{A\in\mathscr{A}_{\mathcal{F}}} \eta\Big(\int_0^1 \mu(A^t)d\lambda(t)\Big)\\ &= \sum_{A\in\mathscr{A}_{\mathcal{F}}} \eta((\mu\times\lambda)(A)) = H^{\mathrm{DF}}(\mathcal{F}). \ \blacksquare\end{aligned}$$

Remark 2.2. Notice that the inequality above can be strict. Consider $X = [0,1]$ with the Lebesgue measure λ and the collection $\mathcal{F} = \{f\}$, where $f(x) = x$. Since $\mathscr{A}_{\mathcal{F}}$ consists of two sets of equal product measure, we have $H^{\mathrm{DF}}(\mathcal{F}) = 1$ (note that we use logarithm to base 2). On the other hand, integrating by parts we have

$$\overline{H}(\mathcal{F}) = \int_0^1 (-t\log t - (1-t)\log(1-t))\,dt = -2\int_0^1 t\log t\,dt < 1.$$

Lemma 2.3 (Monotonicity and subadditivity axiom). *For any finite families $\mathcal{F}$, $\mathcal{G}$ and $\mathcal{H}$,*

$$0 \le \overline{H}(\mathcal{F}\,|\,\mathcal{H}) \le \overline{H}(\mathcal{F}\sqcup\mathcal{G}\,|\,\mathcal{H}) \le \overline{H}(\mathcal{F}\,|\,\mathcal{H}) + \overline{H}(\mathcal{G}\,|\,\mathcal{H}).$$

Proof. The conclusion follows easily from (2.1) and the properties of the Shannon entropy of a partition. For instance, to justify the last inequality one needs to show that $\overline{H}(\mathcal{F}\,|\,\mathcal{H}) + \overline{H}(\mathcal{G}\,|\,\mathcal{H}) - \overline{H}(\mathcal{F}\sqcup\mathcal{G}|\mathcal{H})$ is nonnegative. But this expression is equal to

$$\begin{aligned}&\int_0^1 [H_\mu(\mathscr{A}^t_{\mathcal{F}}\vee\mathscr{A}^t_{\mathcal{H}}) + H_\mu(\mathscr{A}^t_{\mathcal{G}}\vee\mathscr{A}^t_{\mathcal{H}}) - H_\mu(\mathscr{A}^t_{\mathcal{F}}\vee\mathscr{A}^t_{\mathcal{G}}\vee\mathscr{A}^t_{\mathcal{H}}) - H_\mu(\mathscr{A}^t_{\mathcal{H}})]\,d\lambda(t)\\ &\qquad = \int_0^1 [H_\mu(\mathscr{A}^t_{\mathcal{G}}\,|\,\mathscr{A}^t_{\mathcal{H}}) - H_\mu(\mathscr{A}^t_{\mathcal{G}}\,|\,\mathscr{A}^t_{\mathcal{F}}\vee\mathscr{A}^t_{\mathcal{H}})]\,d\lambda(t) \ge 0. \ \blacksquare\end{aligned}$$

We recall that the above lemma suffices to prove that

$$\overline{H}\Big(\bigsqcup_{i=1}^n \mathcal{F}_i\ \Big|\ \bigsqcup_{i=1}^n \mathcal{G}_i\Big) \le \sum_{i=1}^n \overline{H}(\mathcal{F}_i\,|\,\mathcal{G}_i)$$

for arbitrary families $\mathcal{F}_1,\ldots,\mathcal{F}_n$ and $\mathcal{G}_1,\ldots,\mathcal{G}_n$.

The next lemma will help us prove that the entropy $\overline{H}$ satisfies the L^1-continuity axiom (Lemma 2.5) and the domination axiom (Lemma 2.8). We abbreviate $\overline{H}(\{f\}\,|\,\{g\})$ by $\overline{H}(f\,|\,g)$ and denote symmetric difference by $\triangle$.

LEMMA 2.4. *Let $f, g\colon X \to [0,1]$. For every $\varepsilon > 0$ there exists $\delta > 0$ such that if $\|f-g\|_1 < \delta$ then $\overline{H}(f\,|\,g) < \varepsilon$.*

Proof. Let $\mathscr{A}_{\{f,g\}} = \{A, B, C, D\}$, where

$$A = \{(x,t) : t < \min\{f(x), g(x)\}\}, \qquad C = \{(x,t) : f(x) \le t < g(x)\},$$
$$B = \{(x,t) : t \ge \max\{f(x), g(x)\}\}, \qquad D = \{(x,t) : g(x) \le t < f(x)\},$$

and recall that $\mathscr{A}_g = \{A_g, A_g^c\}$. We have

$$\begin{aligned}\overline{H}(f\,|\,g) \le & \int_0^1 |\eta(\mu(A^t)) - \eta(\mu(A_g^t))|\, d\lambda(t) \\ & + \int_0^1 |\eta(\mu(B^t)) - \eta(\mu((A_g^c)^t))|\, d\lambda(t) \\ & + \int_0^1 \eta(\mu(C^t))\, d\lambda(t) + \int_0^1 \eta(\mu(D^t))\, d\lambda(t).\end{aligned}$$

We now show that the last two summands are small. Let δ' be small enough to have $\eta(x) < \varepsilon/8$ for $x < \delta'$, and let $N = \max_{x\in[0,1]} \eta(x)$. Then

$$\begin{aligned}\int_0^1 \eta(\mu(C^t))\, d\lambda(t) &= \int_{\{t\colon \mu(C^t)<\delta'\}} \eta(\mu(C^t))\, d\lambda(t) + \int_{\{t\colon \mu(C^t)\ge\delta'\}} \eta(\mu(C^t))\, d\lambda(t) \\ &\le \frac{\varepsilon}{8}\cdot \lambda\{t\colon \mu(C^t) < \delta'\} + N\cdot\lambda\{t\colon \mu(C^t) \ge \delta'\}.\end{aligned}$$

The first summand is not larger than $\varepsilon/8$, while the second is estimated by

$$\lambda\{t\colon \mu(C^t) \ge \delta'\} \le \frac{1}{\delta'}\int_0^1 \mu(C^t)\, d\lambda = \frac{1}{\delta'}(\mu\times\lambda)(C).$$

Since $(\mu\times\lambda)(C) \le \|f-g\|_1 < \delta$ and δ' depends only on ε, choosing δ suitably small we obtain

$$\int_0^1 \eta(\mu(C^t))\, d\lambda(t) < \frac{\varepsilon}{8} + \frac{N}{\delta'}\cdot\delta < \frac{\varepsilon}{4}.$$

By the same reasoning $\int_0^1 \eta(\mu(D^t))\, d\lambda(t) < \varepsilon/4$.

Similarly, we estimate the first two summands. Let $\delta'' > 0$ be chosen so that $|\eta(x) - \eta(y)| < \varepsilon/8$ if $|x-y| < \delta''$. Decomposing the unit interval into $\{t\colon |\mu(A^t) - \mu(A_g^t)| < \delta''\}$ and its complement, and integrating separately over these sets, we obtain

$$\int_0^1 |\eta(\mu(A^t)) - \eta(\mu(A_g^t))|\, d\lambda(t) \le \frac{\varepsilon}{8} + N \cdot \lambda\{t\colon |\mu(A^t) - \mu(A_g^t)| \ge \delta''\}$$

and again

$$\begin{aligned}\lambda\{t\colon |\mu(A^t) - \mu(A_g^t)| \ge \delta''\} &\le \frac{1}{\delta''}\int_0^1 |\mu(A^t) - \mu(A_g^t)|\, d\lambda(t)\\ &\le \frac{1}{\delta''}(\mu \times \lambda)(A \triangle A_g) \le \frac{1}{\delta''}\|f-g\|_1 \le \frac{\delta}{\delta''},\end{aligned}$$

which can be made smaller than $\varepsilon/8N$ by an appropriate choice of δ. Finally, we get $\overline{H}(f\,|\,g) < \varepsilon$. ∎

Lemma 2.5 (L^1-continuity axiom). *For every $r \ge 1$ and $\varepsilon > 0$ there exists a constant $\delta > 0$ such that if $\mathcal{F}$, $\mathcal{G}$ and $\mathcal{H}$ have cardinalities at most r and* $\operatorname{dist}(\mathcal{F},\mathcal{G}) < \delta$ *then*

$$\operatorname{dist}(\mathcal{F}\sqcup\mathcal{H}, \mathcal{G}\sqcup\mathcal{H}) < \varepsilon \quad \textit{and} \quad |\overline{H}(\mathcal{F}) - \overline{H}(\mathcal{G})| < \varepsilon.$$

In fact, the latter may be replaced by $\overline{H}(\mathcal{F}\,|\,\mathcal{G}) < \varepsilon$.

Proof. Clearly, $\operatorname{dist}(\mathcal{F}\sqcup\mathcal{H}, \mathcal{G}\sqcup\mathcal{H}) \le \operatorname{dist}(\mathcal{F},\mathcal{G})$, so the first inequality is satisfied. Notice that

$$|\overline{H}(\mathcal{F}) - \overline{H}(\mathcal{G})| \le \overline{H}(\mathcal{F}\,|\,\mathcal{G}) + \overline{H}(\mathcal{G}\,|\,\mathcal{F}),$$

so it suffices to estimate $\overline{H}(\mathcal{F}\,|\,\mathcal{G})$. Let $r \ge r'$, and let $\mathcal{F} = \{f_1,\dots,f_r\}$ and $\mathcal{G} = \{g_1,\dots,g_{r'}\}$ satisfy $\operatorname{dist}(\mathcal{F},\mathcal{G}) < \delta$, where δ is obtained from the previous lemma for $\varepsilon/2r$ in place of ε. There is a correspondence $f_i \mapsto g_{\pi(i)}$, where $i = 1,\dots,r$ (possibly with some g_i's equal to zero if $r > r'$), such that $\max_{1\le i\le r}\|f_i - g_{\pi(i)}\|_1 < \delta$. Thus

$$\overline{H}(\mathcal{F}\,|\,\mathcal{G}) \le \sum_{i=1}^{r} \overline{H}(f_i\,|\,g_{\pi(i)}) < \varepsilon/2. \ ∎$$

Remark 2.6. In [DF] and [F] the join of $\mathcal{F}$ and $\mathcal{G}$ was denoted by $\mathcal{F}\cup\mathcal{G}$ and it was claimed that it is just the set-theoretic union. Unfortunately, if $\mathcal{F} = \{1/2\} = \mathcal{H}$ and $\mathcal{G} = \{1/2+\varepsilon\}$ (families of constant functions), then $\operatorname{dist}(\mathcal{F},\mathcal{G}) = \varepsilon$, while $\operatorname{dist}(\mathcal{F}\cup\mathcal{H}, \mathcal{G}\cup\mathcal{H}) = \operatorname{dist}(\mathcal{F}, \mathcal{G}\cup\mathcal{H}) = 1/2$ no matter how small ε is! This problem may be eliminated if we keep an additional copy of $1/2$ in the join of $\mathcal{F}$ and $\mathcal{H}$, i.e. if we concatenate the families rather than take their union. Notice that the partition $\mathscr{A}_{\mathcal{F}}$ is insensitive to any change in enumeration of functions and depends only on a set of distinguishable functions constituting $\mathcal{F}$, hence in the definition of $H^{\mathrm{DF}}(\mathcal{F})$ one can safely replace finite sets of functions by finite sequences of functions. Moreover, the formula $\mathscr{A}_{\mathcal{F}\sqcup\mathcal{G}} = \mathscr{A}_{\mathcal{F}} \vee \mathscr{A}_{\mathcal{G}}$ remains correct regardless of whether we use finite sets and their unions or finite sequences and concatenations as a basic tool. Thus, the values of $H^{\mathrm{DF}}(\mathcal{F})$, $h^{\mathrm{DF}}(T,\mathcal{F})$ and $h^{\mathrm{DF}}(T)$ remain unchanged if we pass from sets and unions to sequences and concatenations. A version using sequences satisfies the axioms with no exceptions, hence all theorems proved for dynamical entropy in [DF] and [F] are valid for any version of entropy compatible with the axioms.

Another approach to the problem, with a slight weakening of the axioms, was presented in a recent book [D] by Tomasz Downarowicz. We stress that our new formula is compatible with the axioms given there.

Lemma 2.7 (Partitions axiom). *$\overline{H}(\mathbb{1}_{\Xi}) = H_\mu(\Xi)$, where $H_\mu(\Xi)$ is the Shannon entropy of Ξ.*

Proof. Since every t-section of $\mathscr{A}_{\mathbb{1}_\Xi}$ is equal to Ξ we immediately get the assertion. ∎

If α is a partition of $[0,1]$, denote by $\mathcal{F}^{-1}(\alpha)$ the partition $\bigvee_{f\in\mathcal{F}} f^{-1}(\alpha)$ of X.

Lemma 2.8 (Domination axiom). *For every $r \geq 1$ and $\varepsilon > 0$ there exists $\delta > 0$ such that for every family $\mathcal{F}$ of at most r elements and every partition α of the unit interval into finitely many subintervals of lengths not exceeding δ,*

$$\overline{H}(\mathcal{F} \mid \mathbb{1}_{\mathcal{F}^{-1}(\alpha)}) < \varepsilon.$$

Proof. Fix $r \geq 1$ and $\varepsilon > 0$. For ε/r choose $\delta > 0$ according to Lemma 2.4. Let $\mathcal{F} = \{f_1, \ldots, f_r\}$ and let α denote a partition of the unit interval into intervals whose lengths do not exceed δ. Denote by $0 = a_0 < a_1 < \cdots < a_s = 1$ the endpoints of these intervals. For every $i = 1, \ldots, r$ define a simple function $s_i : X \to [0,1]$ by

$$s_i = \sum_{k=0}^{s-1} a_k \mathbb{1}_{B_{i,k}}, \quad \text{where} \quad B_{i,k} = \{x \colon a_k \leq f_i(x) < a_{k+1}\}.$$

Then $\|f_i - s_i\|_1 < \delta$ for every i. Let $\overline{\alpha}$ denote the family of constant functions with values $a_0, a_1, \ldots, a_s$. Since the entropy of a family of constants is zero, we have

$$\begin{aligned}\overline{H}(\mathcal{F} \mid \mathbb{1}_{\mathcal{F}^{-1}(\alpha)}) &= \overline{H}(\mathcal{F} \mid \mathbb{1}_{\mathcal{F}^{-1}(\alpha)}) - \overline{H}(\overline{\alpha}) \\ &\leq \overline{H}(\mathcal{F} \mid \mathbb{1}_{\mathcal{F}^{-1}(\alpha)} \sqcup \overline{\alpha}) \\ &= \int_0^1 H_\mu(\mathscr{A}^t_{\mathcal{F}} \mid \mathscr{A}^t_{\mathbb{1}_{\mathcal{F}^{-1}(\alpha)} \sqcup \overline{\alpha}})\, d\lambda(t).\end{aligned}$$

Since the partition of the product induced by $\mathbb{1}_{\mathcal{F}^{-1}(\alpha)} \sqcup \overline{\alpha}$ is finer than

$\mathscr{A}_{\{s_1,\dots,s_r\}}$, the right hand is not greater than

$$\int_0^1 H_\mu(\mathscr{A}^t_{\mathcal{F}} \mid \mathscr{A}^t_{\{s_1,\dots,s_r\}}) d\lambda(t) = \int_0^1 H_\mu\Big(\bigvee_{i=1}^r \mathscr{A}^t_{f_i} \Bigm| \bigvee_{i=1}^r \mathscr{A}^t_{s_i}\Big)\, d\lambda(t)$$
$$\le \int_0^1 \sum_{i=1}^r H_\mu(\mathscr{A}^t_{f_i} \mid \mathscr{A}^t_{s_i})\, d\lambda(t) = \sum_{i=1}^r \overline{H}(f_i \mid s_i) < \varepsilon. \ \blacksquare$$

The above lemmas together with Theorem 2.1 of [DF] imply the following theorem:

THEOREM 2.9. *The quantity $\overline{H}$ defined by* (2.1) *is a version of operator entropy in the sense of* [DF].

3. Dynamical entropy. In this section we prove that if the static entropy is defined as in the previous section then in the formula for $h(T,\mathcal{F})$ (step (4) of the general scheme) the upper limit may be replaced by a limit. To prove the convergence in the classical case one uses the subadditivity of $H_\mu(\bigvee_{i=0}^{n-1} T^{-i}\alpha)$, where α is a partition, which involves both subadditivity and invariance of the static entropy. Lemma 2.3 equips us with the first of these tools, but an easy example shows that we lack the second one. Indeed, let X be the unit interval with the Lebesgue measure and $Tf(x) = \frac{1}{2}f(x) + \frac{1}{2}f(1-x)$. The family $\mathcal{F} = \{\mathbb{1}_{[0,1/4]}, \mathbb{1}_{(1/4,1]}\}$ cuts every t-section of $X \times [0,1]$ into two disjoint sets $[0,1/4]$, $(1/4,1]$, while its image yields partitions into two sets of equal measure, so $H(\mathcal{F}) < H(T\mathcal{F})$. Therefore, below we reproduce the argument used in [DF] for the current version of static entropy.

We recall the theorem on the integral representation of stochastic operators proved by A. Iwanik in [I].

THEOREM. *If T is an operator on the set of bounded measurable functions of a standard Borel space and T is induced by a transition probability then*
$$Tf(x) = \int_\Omega f(\varphi_\omega(x))\, d\lambda(\omega),$$
where (Ω,λ) denotes the unit interval equipped with the Lebesgue measure and $(\omega,x) \mapsto \varphi_\omega(x)$ is a jointly measurable map from $\Omega \times X$ into X.

Suppose that T satisfies the assumptions of the above theorem and let Φ be the operator on bounded measurable functions on $\Omega \times X$ generated by a pointwise map $\phi(\omega,x) = (\omega,\varphi_\omega(x))$. Denoting by $\overline{f}$ the function $(\omega,x) \mapsto f(x)$ we have
$$Tf(x) = \int \Phi\overline{f}(\omega,x)\, d\lambda(\omega).$$

Though Φ need not preserve the product measure, using Fubini's theorem we do have

$$\iint \Phi\overline{f}\,d\lambda\,d\mu = \int Tf\,d\mu = \int f\,d\mu. \tag{3.1}$$

In particular, $(\lambda\times\mu)(\phi^{-1}(\Omega\times A)) = \mu(A)$ for all measurable $A\subset X$. Since every iterate of T is induced by a transition probability (because T is), we can denote by Φ_k the pointwise generated operator corresponding to T^k and by ϕ_k the map that generates Φ_k. Note that, in general, Φ_k is not equal to the iterate Φ^k. The proof of the following lemma can be found in [DF] (the statement (3.2.3) in the proof of Lemma 3.2).

Lemma 3.1. *Let $\mathcal{F}$ be a collection of measurable functions from X to $[0,1]$. For all $f\in\mathcal{F}$ and $\delta>0$ there exists $N\in\mathbb{N}$ such that for all $k\in\mathbb{N}$ and $l\ge N$,*

$$\|\overline{T^{k+l}f} - \Phi_k\overline{T^l f}\|_1 < \delta.$$

We say that a family $\mathcal{G}=\{g_1,\dots,g_r\}$ is *increasing* if $g_1\le\dots\le g_r$.

Lemma 3.2. *For an increasing family $\mathcal{G}=\{g_1,\dots,g_r\}$ of functions on X,*

$$\overline{H}_{\lambda\times\mu}(\Phi_k\overline{\mathcal{G}}) = \overline{H}_\mu(\mathcal{G}) \qquad \forall k\in\mathbb{N}.$$

Proof. Let $g_0\equiv 0$ and $g_{r+1}\equiv 1$ be functions defined on X. Notice that $\mathscr{A}_{\mathcal{G}} = \{A_0,A_1,\dots,A_r\}$, where $A_0=\{(x,t): t\le g_1(x)\}$ and $A_i=\{(x,t): g_i(x)<t\le g_{i+1}(x)\}$ for $i=1,\dots,r$. So for $\overline{\mathcal{G}}=\{\overline{g}_1,\dots,\overline{g}_r\}$ we have $\mathscr{A}_{\overline{\mathcal{G}}}=\{B_0,\dots,B_r\}$, where $B_i=\Omega\times A_i$. Thus we have $B^t=\Omega\times A^t$ and

$$\begin{aligned}(\lambda\times\mu)(B^t) &= \mu(A^t) = \int T^k\mathbb{1}_{A^t}(x)\,d\mu(x)\\ &= \iint \Phi_k\mathbb{1}_{B^t}(\omega,x)\,d\lambda(\omega)\,d\mu(x) = (\lambda\times\mu)(\phi_k^{-1}(B^t)).\end{aligned}$$

Denoting $\mathscr{A}_{\Phi_k\overline{\mathcal{G}}}=\{C_i: i=1,\dots,r\}$, where

$$C_i=\{(\omega,x,t): \Phi_k\overline{g}_i(\omega,x)<t\le\Phi_k\overline{g}_{i+1}(\omega,x)\},$$

we obtain $C_i^t=\phi_k^{-1}(B_i^t)$, where $B_i\in\mathscr{A}_{\overline{\mathcal{G}}}$. Hence $\mathscr{A}^t_{\Phi_k\overline{\mathcal{G}}}=\phi_k^{-1}(\mathscr{A}^t_{\overline{\mathcal{G}}})$ and

$$\begin{aligned}\overline{H}_{\lambda\times\mu}(\Phi_k\overline{\mathcal{G}}) &= \int_0^1 \sum_{B\in\mathscr{A}_{\overline{\mathcal{G}}}} \eta((\lambda\times\mu)(\phi_k^{-1}(B^t)))\,d\lambda(t)\\ &= \int_0^1 \sum_{A\in\mathscr{A}_{\mathcal{G}}} \eta(\mu(A^t))\,d\lambda(t) = \overline{H}_\mu(\mathcal{G}). \ \blacksquare\end{aligned}$$

An arbitrary family $\mathcal{F}=\{f_1,\dots,f_r\}$ of functions may be transformed into an increasing one in the following way. Denote by $\preccurlyeq$ the lexicographic order on the set of words $\{0,1\}^r$. For every $\beta \in \{0,1\}^r$ we define

$$\theta_\beta = \begin{cases} 1 & \text{for } \beta = 11\ldots1, \\ \sup_{\alpha \preccurlyeq \beta} \inf\{f_i \colon \alpha_i = 0\} & \text{otherwise.} \end{cases}$$

It is clear that the family $\{\theta_\beta\}$ is increasing with respect to the lexicographic order on $\{0,1\}^r$—the outcome depends, however, on the initial enumeration. Excluding from θ_β's spare copies of functions and including (for convenience) the function equal to zero everywhere we obtain an increasing collection which will be denoted by $\Theta(\mathcal{F})$. It is not hard to prove that $\mathscr{A}_{\Theta(\mathcal{F})} = \mathscr{A}_{\mathcal{F}}$, so $\mathscr{A}^t_{\mathcal{F}} = \mathscr{A}^t_{\Theta(\mathcal{F})}$ and $\overline{H}(\mathcal{F}) = \overline{H}(\Theta(\mathcal{F}))$. Moreover, $\Theta(\Phi_k\mathcal{F}) = \Phi_k(\Theta(\mathcal{F}))$ for every $k \in \mathbb{N}$, because Φ_k is pointwise generated (thus preserves the maximum and minimum operations).

Lemma 3.3. *For every $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that for all $k, m \in \mathbb{N}$,*

$$|\overline{H}_\mu(T^{k+N}\mathcal{F}^m) - \overline{H}_\mu(T^N\mathcal{F}^m)| < m\varepsilon.$$

Proof. Assume that X is standard Borel and T is as required in Iwanik's theorem. Fix $\varepsilon > 0$ and denote $\overline{\mathcal{F}} = \{\overline{f} : f \in \mathcal{F}\}$. Applying Lemma 3.1 to each function of $\mathcal{F}$ and then Lemma 2.5 we see that there exists $N \in \mathbb{N}$ such that for all $l \geq N$ and $k \in \mathbb{N}$,

$$\overline{H}_{\lambda\times\mu}(\overline{T^{k+N}\mathcal{F}^m} \mid \Phi_k\overline{T^N\mathcal{F}^m}) \leq \sum_{l=N}^{N+m-1} \overline{H}_{\lambda\times\mu}(\overline{T^{k+l}\mathcal{F}} \mid \Phi_k\overline{T^l\mathcal{F}}) < m\varepsilon.$$

Hence,

$$\begin{aligned} \overline{H}_\mu(T^{k+N}\mathcal{F}^m) &= \overline{H}_{\lambda\times\mu}(\overline{T^{k+N}\mathcal{F}^m}) \\ &\leq \overline{H}_{\lambda\times\mu}(\Phi_k\overline{T^N\mathcal{F}^m}) + \overline{H}_{\lambda\times\mu}(\overline{T^{k+N}\mathcal{F}^m} \mid \Phi_k\overline{T^N\mathcal{F}^m}) \\ &\leq \overline{H}_{\lambda\times\mu}(\Phi_k\overline{T^N\mathcal{F}^m}) + m\varepsilon. \end{aligned}$$

Using Lemma 3.2 we obtain

$$\begin{aligned} \overline{H}_{\lambda\times\mu}(\Phi_k\overline{T^N\mathcal{F}^m}) &= \overline{H}_{\lambda\times\mu}(\Theta(\Phi_k\overline{T^N\mathcal{F}^m})) = \overline{H}_{\lambda\times\mu}(\Phi_k\overline{\Theta(T^N\mathcal{F}^m)}) \\ &\overset{\text{(Lem. 3.2)}}{=} \overline{H}_\mu(\Theta(T^N\mathcal{F}^m)) = \overline{H}_\mu(T^N\mathcal{F}^m), \end{aligned}$$

and consequently

$$\overline{H}_\mu(T^{k+N}\mathcal{F}^m) < \overline{H}_\mu(T^N\mathcal{F}^m) + m\varepsilon.$$

A similar argument, but with the roles of $T^{k+N}\mathcal{F}^m$ and $T^N\mathcal{F}^m$ exchanged, yields

$$\overline{H}_\mu(T^N\mathcal{F}^m) < \overline{H}_\mu(T^{k+N}\mathcal{F}^m) + m\varepsilon.$$

In the general case the main idea is to pass to the complex subalgebra $\mathscr{L}$ of $L^\infty(\mu)$ generated by (countably many) functions used in the above argument. It is known that $\mathscr{L}$ is isometrically isomorphic to the algebra of complex continuous functions on a certain compact Hausdorff space $\mathscr{X}$. The isomorphism τ sends a measure μ to a certain Borel probability measure on $\mathscr{X}$ and the operator T to a Markov operator $\mathcal{T} = \tau T \tau^{-1}$ defined on the real algebra $C(\mathscr{X})$ of all continuous real functions. Since a Markov operator is always induced by a (Feller) transition probability the assertion (for $\mathcal{T}$) follows from the first part of the proof. The entropies of $T^N\mathcal{F}^m$ and $T^{k+N}\mathcal{F}^m$ are equal to the entropies of the corresponding families in $C(\mathscr{X})$, which completes the proof. We refer the reader to [DF] for more details. ∎

The above lemma allows one to replace subadditivity by the following quasi-subadditivity property. We omit the proofs of both the property and the consequent theorem, as they strongly resemble arguments used in the classical pointwise theory.

LEMMA 3.4. *For every $\varepsilon > 0$ there exists $N \in \mathbb{N}$ and a constant c such that for every $k \in \mathbb{N}$ and $m \geq N$,*

$$\overline{H}_\mu(\mathcal{F}^{k+m}) \leq \overline{H}_\mu(\mathcal{F}^k) + \overline{H}_\mu(\mathcal{F}^m) + c + m\varepsilon.$$

THEOREM 3.5. *If T is a doubly stochastic operator then*

$$h_\mu(T,\mathcal{F}) = \lim_{n\to\infty} \frac{1}{n}\overline{H}_\mu(\mathcal{F}^n).$$

We remark that the above limit need not coincide with the corresponding infimum.

4. The product rule. If T is a doubly stochastic operator on $L^1(\mu)$ and S a doubly stochastic operator on $L^1(\nu)$ then one can define an operator on functions of the form $f(x)\cdot g(y) \in L^1(\mu\times\nu)$ by $(T\times S)(fg)(x,y) = Tf(x)\cdot Sg(y)$. Extending the definition to linear combinations of such products and then, by density, to the whole $L^1(\mu\times\nu)$ we obtain a doubly stochastic operator which will be called the *product* of T and S and denoted by $T\times S$. In other words, if we represent T and S by their stochastic kernels (or less generally by transition probabilities) then the stochastic kernel of the product is the product of the kernels of T and S (or its transition probability is the product of the transition probabilities of T and S).

As an application of the new formula we will prove that, similarly to the classical case, the entropy of a product is the sum of the entropies of the factors. Notice that since Theorem 4.5 concerns the dynamical entropy of an operator it remains valid regardless of the choice of static entropy. We start by stating some crucial lemmas.

DEFINITION 4.1. For a function f and constants $a < b$ let $f_a^b = (f\vee a)\wedge b$, where $\vee$ and $\wedge$ denote maximum and minimum, respectively. We say that f

has *property* CZ(δ) if

$$\int |T^n(f_a^b) - (T^n f)_a^b|\, d\mu < \delta$$

for every $n \geq 0$ and any constants $a < b$.

LEMMA 4.2 ([DF, Lemma 2.3]). *If T is a doubly stochastic operator and f a bounded function then for every $\delta > 0$ there exists an integer l such that $T^l f$ has property* CZ(δ).

LEMMA 4.3 ([DF, Lemma 2.5]). *Let $\mathcal{F}$ consist of r functions with ranges in $[0,1]$, having property* CZ(δ^3). *Then there is a partition Ξ of $[0,1]$ into subintervals of lengths not exceeding $2r\delta$ such that for every n and $f \in \mathcal{F}$, and all breakpoints ξ of Ξ,*

$$\int |T^n \mathbb{1}_{\{x:\, f(x)\geq\xi\}} - \mathbb{1}_{\{x:\, T^n f(x)\geq\xi\}}|\, d\mu < 4\delta. \tag{4.1}$$

LEMMA 4.4 ([DF, Lemma 2.6]). *Let $\mathcal{F}$ consist of r functions with ranges in $[0,1]$, all having property* CZ(δ^3), *and let α be a partition of $[0,1]$ into m pieces $A_0 = [0,\xi_1)$, $A_j = [\xi_j, \xi_{j+1})$ $(j = 1,\dots,m-2)$ and $A_{m-1} = [\xi_{m-1}, 1]$, where the points ξ_j all satisfy condition* (4.1). *Then*

$$\operatorname{dist}(T^n(\mathbb{1}_{\mathcal{F}^{-1}(\alpha)}), \mathbb{1}_{(T^n\mathcal{F})^{-1}(\alpha)}) < 8rm^r\delta \quad \textit{for every } n \geq 0.$$

Noticing a special role of characteristic functions in the domination axiom (Lemma 2.8) we observe that if α is a partition of X and β is a partition of Y then

$$\overline{H}(\mathbb{1}_{\alpha\times Y} \sqcup \mathbb{1}_{X\times\beta}) = \overline{H}(\mathbb{1}_{\alpha\times\beta}) = \overline{H}(\mathbb{1}_\alpha) + \overline{H}(\mathbb{1}_\beta). \tag{4.2}$$

THEOREM 4.5.

$$\overline{h}(T\times S) = \overline{h}(T) + \overline{h}(S).$$

Proof. Let $U = T\times S$ and fix $\varepsilon > 0$. Let $\mathcal{E}$ be a family of measurable functions $X\times Y \to [0,1]$. Each of them can be approximated in L^1 norm by a sum $\sum_i f_i g_i$, where $f_i\colon X\to[0,1]$, $g_i\colon Y\to[0,1]$. Denote the collection of all these sums by $\mathcal{E}'$. Since the product operator U is a contraction, each $U^n e$ for $e\in\mathcal{E}$ is L^1-approximated without increasing the error by the nth image of a suitable element of $\mathcal{E}'$. Demanding that $\operatorname{dist}(\mathcal{E},\mathcal{E}')$ is appropriately small we obtain

$$\overline{H}(\mathcal{E}^N) \leq \overline{H}((\mathcal{E}')^N) + N\varepsilon.$$

If $\mathcal{F}_e$ is the family of all functions f_i used in sums $\sum_i f_i g_i$ to approximate $e\in\mathcal{E}$ and $\mathcal{G}_e$ is the family of all corresponding functions g_i then we let $\mathcal{F} = \bigsqcup_{e\in\mathcal{E}} \mathcal{F}_e$ and $\mathcal{G} = \bigsqcup_{e\in\mathcal{E}} \mathcal{G}_e$. Denote $r = \#\mathcal{F} = \#\mathcal{G}$.

Let δ be as specified in Lemma 2.5 (continuity axiom) for the cardinality $\#\mathcal{E}$ and $\varepsilon > 0$. Let $m = \lceil 8r/\delta\rceil$. Then choose $\delta' < \delta/4rm^r$ again from Lemma 2.5, but for the cardinality m^r and ε. Let L be large enough to ensure that every $T^L f$ for $f\in\mathcal{F}$ and every $S^L g$ for $g\in\mathcal{G}$ have property

$\mathrm{CZ}((\delta'/8rm^r)^3)$ for the actions of T and S, respectively. We use Lemma 4.3 for T and $T^L\mathcal{F}$ to obtain a suitable partition Ξ. The number $\delta'/4m^r$ majorizes the distances between the breakpoints ξ of Ξ and, because it is much smaller than $\delta/8r$, we can pick $m-1$ of them creating a partition α of $[0,1]$ into m intervals of lengths smaller than $\delta/4r$. The same procedure is applied to S and $S^L\mathcal{G}$, giving a partition β of $[0,1]$. We stress that the same parameters m, r, δ are valid for both α and β.

We put $\alpha_n = (T^n\mathcal{F})^{-1}(\alpha)$ and $\beta_n = (S^n\mathcal{G})^{-1}(\beta)$ for $n \in \mathbb{N}$. Note that α_n and β_n are partitions of X. By Lemma 4.4 we obtain

$$\text{(4.3)} \qquad \operatorname{dist}(T^n \mathbb{1}_{\alpha_L}, \mathbb{1}_{\alpha_{L+n}}) < \delta' \quad \text{and} \quad \operatorname{dist}(S^n \mathbb{1}_{\beta_L}, \mathbb{1}_{\beta_{L+n}}) < \delta'$$

for every $n \in \mathbb{N}$.

Each $T^L f \in T^L\mathcal{F}$ may be approximated with an error of at most $\delta/4r$ by a simple function $\sum_k a_k \mathbb{1}_{A_k}$, where $A_k \in \alpha_L$ and the values a_k are breakpoints of the partition α. Since each such simple function admits at most m values and the preimage (by $T^L f$) of each element of α is a union of at most m^{r-1} elements of α_L, the sum $\sum_k a_k \mathbb{1}_{A_k}$ has at most m^r elements. Then each $T^{L+n} f \in T^{L+n}\mathcal{F}$ is $\delta/4r$-approximated by $\sum_k a_k T^n \mathbb{1}_{A_k}$. Using (4.3) we can replace functions $T^n \mathbb{1}_{A_k}$ by certain characteristic functions $\mathbb{1}_{\tilde{A}_k}$, where $\tilde{A}_k \in \alpha_{L+n}$, which yields an error of δ' on each such function. Therefore,

$$\Big\| T^{L+n} f - \sum_k a_k \mathbb{1}_{\tilde{A}_k} \Big\| \le \Big\| T^{L+n} f - \sum_k a_k T^n \mathbb{1}_{A_k} \Big\| + \sum_k a_k \| T^n \mathbb{1}_{A_k} - \mathbb{1}_{\tilde{A}_k} \|$$
$$< \frac{\delta}{4r} + m^r \cdot \delta' < \frac{\delta}{2r}.$$

Similarly, given $S^L g \in S^L\mathcal{G}$ we find a simple function $\sum_k b_k \mathbb{1}_{\tilde{B}_k}$, where $\tilde{B}_k \in \beta_{L+n}$ and b_k are breakpoints of β, such that $\| S^{L+n} g - \sum_k b_k \mathbb{1}_{\tilde{B}_k} \| < \delta/2r$. Since every function $\sum_i f_i g_i \in \mathcal{E}'$ consists of at most r summands, an easy calculation proves that

$$\Big\| U^{L+n} \sum_i f_i g_i - \sum_i \sum_{k,l} a_k^{(i)} b_l^{(i)} \mathbb{1}_{\tilde{A}_k^{(i)} \cap \tilde{B}_l^{(i)}} \Big\| < \delta$$

for some $a_k^{(i)}$ and $b_l^{(i)}$ which are breakpoints of α and β, $\tilde{A}_k^{(i)} \in \alpha_{L+n}$ and $\tilde{B}_l^{(i)} \in \beta_{L+n}$. If $\mathcal{E}_n''$ denotes the set of all functions of the form

$$\sum_i \sum_{k,l} a_k^{(i)} b_l^{(i)} \mathbb{1}_{\tilde{A}_k^{(i)} \cap \tilde{B}_l^{(i)}}$$

chosen for elements of $U^{L+n}\mathcal{E}'$ then both $U^{L+n}\mathcal{E}'$ and $\mathcal{E}_n''$ have cardinality $\#\mathcal{E}$, so $\overline{H}(U^{L+n}\mathcal{E}' \mid \mathcal{E}''_n) < \varepsilon$, and consequently

$$\overline{H}((\mathcal{E}')^{L+N}) \le \overline{H}((\mathcal{E}')^{L}) + \overline{H}\Big(\bigsqcup_{n=0}^{N-1} \mathcal{E}''_n\Big) + N\varepsilon.$$

But $\bigsqcup_{n=0}^{N-1} \mathcal{E}''_n$ consists of simple functions, so its entropy is smaller than or equal to the entropy of the family of characteristic functions used to produce elements of $\bigsqcup_{n=0}^{N-1} \mathcal{E}''_n$. These characteristic functions represent (possibly not all) sets from the partitions $\alpha_{L+n} \times \beta_{L+n}$ (with $n < N$), which are coarser than $\bigvee_{n<N} \alpha_{L+n} \times \bigvee_{n<N} \beta_{L+n}$. Thus, using (4.2) we obtain

$$\begin{aligned}\overline{H}\Big(\bigsqcup_{n=0}^{N-1} \mathcal{E}''_n\Big) &\le \overline{H}(\mathbb{1}_{\bigvee_{n<N} \alpha_{L+n} \times \bigvee_{n<N} \beta_{L+n}}) \\ &= \overline{H}(\mathbb{1}_{\bigvee_{n<N} \alpha_{L+n}}) + \overline{H}(\mathbb{1}_{\bigvee_{n<N} \beta_{L+n}}) \\ &= \overline{H}\Big(\bigsqcup_{n<N} \mathbb{1}_{\alpha_{L+n}}\Big) + \overline{H}\Big(\bigsqcup_{n<N} \mathbb{1}_{\beta_{L+n}}\Big).\end{aligned}$$

By (4.3) and since the cardinalities of the families involved are smaller than m^r, we deduce from Lemma 2.5 that

$$\overline{H}\Big(\bigsqcup_{n<N} \mathbb{1}_{\alpha_{L+n}}\Big) \le \overline{H}((\mathbb{1}_{\alpha_L})^N) + N\varepsilon,$$
$$\overline{H}\Big(\bigsqcup_{n<N} \mathbb{1}_{\beta_{L+n}}\Big) \le \overline{H}((\mathbb{1}_{\beta_L})^N) + N\varepsilon.$$

So for any N we have

$$\overline{H}(\mathcal{E}^{L+N}) \le \overline{H}((\mathcal{E}')^{L}) + \overline{H}((\mathbb{1}_{\alpha_L})^N) + \overline{H}((\mathbb{1}_{\beta_L})^N) + (4N+L)\varepsilon,$$

which after dividing by N and taking the limit yields

$$\overline{h}(U, \mathcal{E}) \le \overline{h}(T, \mathbb{1}_{\alpha_L}) + \overline{h}(S, \mathbb{1}_{\beta_L}) + 4\varepsilon \le \overline{h}(T) + \overline{h}(S) + 4\varepsilon.$$

Since ε and $\mathcal{E}$ were chosen arbitrarily and independently, we get $\overline{h}(T \times S) \le \overline{h}(T) + \overline{h}(S)$.

To prove the opposite inequality take families $\mathcal{F}$, $\mathcal{G}$ of functions on X and Y, respectively, with cardinalities at most r. Fix $\varepsilon > 0$ and take δ according to Lemma 2.8 (domination axiom) for cardinality r. Pick $m > 1/\delta$. Suppose $\delta' < \delta$ satisfies the assertion of Lemma 2.5 for cardinality m^r and simultaneously $2\delta'$ satisfy the assertion of the same lemma for families of cardinality m^{2r}. Take L so that $T^L\mathcal{F}$ and $S^L\mathcal{G}$ have property $\mathrm{CZ}((\delta'/8rm^r)^3)$. In the same way as in the first part of the proof we can choose partitions α and β, and define $\alpha_n = (T^n\mathcal{F})^{-1}(\alpha)$ and $\beta_n = (S^n\mathcal{G})^{-1}(\beta)$, so that (4.3) holds. By Lemma 2.8,

$$\overline{H}(\mathcal{F}^{L+N}) + \overline{H}(\mathcal{G}^{L+N}) \le \overline{H}(\mathcal{F}^{L}) + \overline{H}\Big(\bigsqcup_{n=0}^{N-1} T^n \mathbb{1}_{\alpha_L}\Big) + \overline{H}(\mathcal{G}^{L}) + \overline{H}\Big(\bigsqcup_{n=0}^{N-1} S^n \mathbb{1}_{\beta_L}\Big) + 2N\varepsilon.$$

On the other hand, using (4.3) and then (4.2) we obtain

$$\begin{aligned}\overline{H}\Big(\bigsqcup_{n=0}^{N-1} T^n \mathbb{1}_{\alpha_L}\Big)+\overline{H}\Big(\bigsqcup_{n=0}^{N-1} S^n \mathbb{1}_{\beta_L}\Big) &\le \overline{H}\Big(\bigsqcup_{n=0}^{N-1} \mathbb{1}_{\alpha_{L+n}}\Big)+\overline{H}\Big(\bigsqcup_{n=0}^{N-1} \mathbb{1}_{\beta_{L+n}}\Big)+2N\varepsilon\\ &= \overline{H}(\mathbb{1}_{\bigvee_{n=0}^{N-1}\alpha_{L+n}\times\bigvee_{n=0}^{N-1}\beta_{L+n}}) + 2N\varepsilon\\ &= \overline{H}\Big(\bigsqcup_{n=0}^{N-1} \mathbb{1}_{\alpha_{L+n}\times\beta_{L+n}}\Big) + 2N\varepsilon.\end{aligned}$$

Every function from $\mathbb{1}_{\alpha_{L+n}\times\beta_{L+n}}$ is a product of characteristic functions, which by (4.3) are approximated by elements of $T^n\mathbb{1}_{\alpha_L}$ or $S^n\mathbb{1}_{\beta_L}$. Thus it is $2\delta'$-approximated by a function from $U^n\mathbb{1}_{\alpha_L\times\beta_L}$. Since both the last family and $\mathbb{1}_{\alpha_{L+n}\times\beta_{L+n}}$ have each at most m^{2r} elements, by Lemma 2.5 we have

$$\overline{H}\Big(\bigsqcup_{n=0}^{N-1} \mathbb{1}_{\alpha_{L+n}\times\beta_{L+n}}\Big) \le \overline{H}\Big(\bigsqcup_{n=0}^{N-1} U^n\mathbb{1}_{\alpha_L\times\beta_L}\Big) + N\varepsilon.$$

We have thus obtained

$$\overline{H}(\mathcal{F}^{L+N}) + \overline{H}(\mathcal{G}^{L+N}) \le \overline{H}(\mathcal{F}^{L}) + \overline{H}(\mathcal{G}^{L}) + \overline{H}\Big(\bigsqcup_{n=0}^{N-1} U^n\mathbb{1}_{\alpha_L\times\beta_L}\Big) + 5N\varepsilon.$$

After dividing both sides by N and taking the limit we get

$$\overline{h}(T,\mathcal{F}) + \overline{h}(S,\mathcal{G}) \le \overline{h}(T\times S, \mathbb{1}_{\alpha_L\times\beta_L}) + 5\varepsilon \le \overline{h}(T\times S) + 5\varepsilon. \ \blacksquare$$

5. Continuity of static entropy with respect to the measure. We end the paper by considering the case of a compact space X. The main aim of this section is to show that if a sequence μ_n of probability measures converges weakly* to a measure μ then $\lim_{n\to\infty}\overline{H}_{\mu_n}(\mathcal{F}) = \overline{H}_{\mu}(\mathcal{F})$ provided that $\mathcal{F}$ consists of continuous functions (we recall that $\overline{H}_{\mu}(\mathcal{F})$ is the static entropy defined in Section 2).

If f is continuous then the boundary of the set $A_f = \{(x,t)\colon t\le f(x)\}$ (and of A_f^c) is characterized by the condition $f(x)=t$. Let $\mathcal{F}=\{f_1,\dots,f_r\}$ be a family of continuous functions. The inclusion $\partial(E\cap F)\subset\partial E\cup\partial F$ implies that the boundary of every $A\in\mathscr{A}_{\mathcal{F}}$ is contained in the union of the graphs of functions $f_1,\dots,f_r$. Since the graph of a real measurable function has $\mu\times\lambda$ measure zero, we get $(\mu\times\lambda)(\partial A)=0$ for every $A\in\mathscr{A}_{\mathcal{F}}$.

Fix $A \in \mathscr{A}_{\mathcal{F}}$. We have $\partial(A^t) \subset (\partial A)^t$, so

$$0 = (\mu \times \lambda)(\partial A) = \int_0^1 \mu((\partial A)^t)\, d\lambda(t) \geq \int_0^1 \mu(\partial(A^t))\, d\lambda(t),$$

hence $\mu((\partial A)^t) = 0$ for λ-almost every t. But $\mathscr{A}_{\mathcal{F}}$ consists of finitely many elements, so $\mu((\partial A)^t) = 0$ for all A and almost all t.

Finally, let μ_n converge to μ in the weak* topology. If $\mu((\partial A)^t) = 0$ then the sequence $\mu_n(A^t)$ converges to $\mu(A^t)$, which in turn implies that $H_{\mu_n}(\mathscr{A}^t_{\mathcal{F}})$ tends to $H_\mu(\mathscr{A}^t_{\mathcal{F}})$ for almost all t. Since for every $n \in \mathbb{N}$ the function $t \mapsto H_{\mu_n}(\mathscr{A}^t_{\mathcal{F}})$ is bounded by a constant $\log \#\mathscr{A}_{\mathcal{F}}$, we can use the dominated convergence theorem to prove that

$$\overline{H}_{\mu_n}(\mathcal{F}) = \int_0^1 H_{\mu_n}(\mathscr{A}^t_{\mathcal{F}})\, d\lambda(t) \to \int_0^1 H_\mu(\mathscr{A}^t_{\mathcal{F}})\, d\lambda(t) = \overline{H}_\mu(\mathcal{F}).$$

We have thus obtained the following theorem.

THEOREM 5.1. *If X is a compact metric space and $\mathcal{F}$ is a finite family of continuous functions then the function mapping a probability measure μ to the entropy $\overline{H}_\mu(\mathcal{F})$ is (uniformly) continuous in the weak* topology on the space of all probability measures of X.*

Acknowledgements. This research was supported from resources for science in years 2009-2012 as a research project (grant MENiSW N N201 394537, Poland).

References

[AF] R. Alicki and M. Fannes, *Quantum Dynamical Systems*, Oxford Univ. Press, Oxford, 2001.

[CNT] A. Connes, H. Narnhofer and W. Thirring, *Dynamical entropy of C^*-algebras and von Neumann algebras*, Comm. Math. Phys.

[D] T. Downarowicz, *Entropy in Dynamical Systems*, Cambridge Univ. Press, 2011

[DF] T. Downarowicz and B. Frej, *Measure-theoretic and topological entropy of operators on function spaces*, Ergodic Theory Dynam. Systems 25 (2005), 455–481.

[F] B. Frej, *Maličky–Riečan's entropy as a version of operator entropy*, Fund. Math. 189 (2006), 185–193.

[GLW] E. Ghys, R. Langevin et P. G. Walczak, *Entropie mesurée et partitions de l'unité*, C. R. Acad. Sci. Paris Sér. I 303 (1986), 251–254.

[I] A. Iwanik, *Integral representations of stochastic kernels*, in: Aspects of Positivity in Functional Analysis (Tübingen, 1985), North-Holland, 1986, 223–230

[M] I. I. Makarov, *Dynamical entropy for Markov operators*, J. Dynam. Control Systems 6 (2000), 1–11.

[MR] P. Maličky and B. Riečan, *On the entropy of dynamical systems*, in: Proc. Conference on Ergodic Theory and Related Topics II (Georgenthal, 1986), Teubner-Texte Math. 94, Teubner, Leipzig, 1987, 135–138.

[V] D. Voiculescu, *Dynamical approximation entropies and topological entropy in operator algebras*, Comm. Math. Phys. 170 (1995), 249–281.

Bartosz Frej, Paulina Frej
Institute of Mathematics and Informatics
Wrocław University of Technology
Wybrzeże Wyspiańskiego 27
50-370 Wrocław, Poland
E-mail: Bartosz.Frej@pwr.wroc.pl
Paulina.Frej@pwr.wroc.pl

Received 24 January 2011;
in revised form 6 June 2011