Ergod. Th. & Dynam. Sys. (2005), **25**, 455–481
doi:10.1017/S014338570400032X

Measure-theoretic and topological entropy of operators on function spaces

TOMASZ DOWNAROWICZ and BARTOSZ FREJ
Institute of Mathematics, Wrocław University of Technology,
Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
(e-mail: downar@im.pwr.wroc.pl, frej@im.pwr.wroc.pl)

(*Received* 16 *May* 2003 *and accepted in revised form* 23 *February* 2004)

Abstract. We study the entropy of actions on function spaces with the focus on doubly stochastic operators on probability spaces and Markov operators on compact spaces. Using an axiomatic approach to entropy we prove that there is basically only one reasonable measure-theoretic entropy notion on doubly stochastic operators. By 'reasonable' we mean extending the Kolmogorov–Sinai entropy on measure-preserving transformations and satisfying some obvious continuity conditions for H_μ. In particular, this establishes equality on such operators between the entropy notion introduced by R. Alicki, J. Andries, M. Fannes and P. Tuyls (a version of which was also studied by I. I. Makarov), another notion of entropy introduced by E. Ghys, R. Langevin and P. Walczak, and our new definition introduced later in this paper. The key tool in proving this uniqueness is the discovery of a very general property of all doubly stochastic operators, which we call *asymptotic lattice stability*. Unlike the other explicit definitions of entropy mentioned above, ours satisfies many natural requirements already on the level of the function H_μ, and we prove that the limit defining h_μ exists. The proof uses an integral representation of a stochastic operator obtained many years ago by A. Iwanik. In the topological part of the paper we introduce three natural definitions of topological entropy for Markov operators on $C(X)$. Then we prove that all three are equal. Finally, we establish the partial variational principle: the topological entropy of a Markov operator majorizes the measure-theoretic entropy of this operator with respect to any of its invariant probability measures.

1. *Introduction*

Measure-theoretic and topological entropy has been thoroughly studied in the context of measure-preserving transformations or continuous maps. Entropy quantifies the complexity of the dynamics by means of exponential growth in time of information obtained by observing the processes on finite partitions or covers.

A natural generalization of a deterministic dynamical system is a stochastic process in which every point x moves to an *a priori* undetermined location according to a certain probability distribution $P(x, \cdot)$ (called *transition probability*) associated with this point. Such dynamics is best understood by studying the evolution of functions under the generated *stochastic operator* $Tf(x) = \int f(y)P(x, dy)$. In this work we extend the notion of both measure-theoretic and topological entropy to a wide variety of actions on function spaces including those generated by transition probabilities and other stochastic operators.

We operate on three major levels of generality. The key definition of measure-theoretic entropy applies to any (not even linear) action $T : L \to L$, where L is any collection of measurable functions on a measure space (X, Σ, μ), with range contained in the unit interval $[0, 1]$. Most of our results on properties of measure-theoretic entropy concern *doubly stochastic operators*—analogs of measure-preserving transformations. Topological entropy and its relation with measure-theoretic entropy is studied for Markov operators—analogs of continuous transformations.

Let (X, Σ) be a measurable space. A *stochastic operator* is a linear map T defined at least on bounded measurable functions satisfying the conditions

(i) Tf is a positive function if f is positive,

(ii) $T\mathbf{1} = \mathbf{1}$,

where $\mathbf{1}$ is the constant function equal everywhere to 1. This includes operators defined by transition probabilities, though not every stochastic operator is such. In our theory general stochastic operators play only a technical role and do not appear in assumptions of the results.

For a fixed measure μ on Σ, a linear operator T is called *doubly stochastic* (with respect to μ) if, in addition to (i) and (ii), it satisfies

(iii) $\int_X Tf\, d\mu = \int_X f\, d\mu$,

for every f. We will usually consider doubly stochastic operators on $L^\infty(\mu)$, but it is quite clear that such operators extend to a doubly stochastic contraction on $L^1(\mu)$.

If X is a compact Hausdorff space and $C(X)$ denotes the space of all real-valued continuous functions on X, then a linear operator $T : C(X) \to C(X)$ satisfying conditions (i) and (ii) 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 $C(X)$, the 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 T^*-invariant (i.e. satisfying (iii) for every continuous f) Radon probability measures is a non-empty convex set, compact in the weak-* topology. For every such measure μ, the operator T becomes a doubly stochastic operator acting on $L^\infty(\mu)$ by the formula

$$Tf(x) = \int f(y)P(x, dy). \tag{1.0.1}$$

Every measurable (measure-preserving, continuous) transformation $S : X \to X$ induces a stochastic (doubly stochastic, Markov) operator T on the relevant function space by the formula $Tf = f \circ S$. Such operators are called *pointwise generated*. One of the most important requirements concerning operator entropy is the demand that for pointwise generated operators one obtains the classical entropy of a generating transformation in both the measure-theoretic and topological cases.

In the literature one can find several attempts to define entropy for operators. In the measure-theoretic case, the most deeply investigated notion is the quantum-dynamical entropy introduced by R. Alicki, J. Andries, M. Fannes and P. Tuyls (see [**AF**] for full details on this type of entropy), based on von Neumann's definition of entropy of a density matrix. A similar definition, formulated for doubly stochastic operators (called there Markov operators) on the space of integrable functions was given by I. I. Makarov in [**M**]. A quite different approach was presented by E. Ghys, R. Langevin and P. Walczak in [**GLW**]. The only topological entropy of a Markov operator known to us was defined by Langevin and Walczak in [**LW**]. All these definitions extend the classical cases, i.e. for pointwise generated doubly stochastic or Markov operators they are equal to the Kolmogorov–Sinai or topological entropy of the generating pointwise map. No other relations between the above approaches to operator entropy have been established.

In this work we achieve progress in understanding the phenomenon of entropy observed on functions. Section 2 presents an axiomatic approach to measure-theoretic entropy. Accepting certain basic properties of entropy of measure-preserving transformations as indispensable, we formulate five construction steps and four axioms, which a general entropy of an action on functions should follow. Our main result asserts that all entropy notions satisfying the axioms coincide on doubly stochastic operators, establishing in particular the equality, on such operators, between all the above mentioned measure-theoretic entropies introduced by other authors.

In §3 we introduce a new definition of measure-theoretic entropy, also fulfilling our axioms. It uses a very natural and effective way of quantifying the exponential growth of 'information content' in an evolution of a finite family of functions by tracing the partitions of $X \times [0, 1]$ determined by graphs of the functions. In addition to the axioms, this entropy satisfies some other desirable properties such as quasi-subadditivity, and the existence of certain limits. We present an example revealing that the entropy of a doubly stochastic operator captures essentially more than just the dynamics of factors behaving as pointwise transformations.

For Markov operators we propose, in §4, three natural methods of defining topological entropy. The fact that all three lead to the same quantity confirms that these notions are suitable. Two of these definitions exploit open covers, while the third one uses separated sets. Our notion coincides on continuous maps with the classic Adler–Konheim–McAndrew topological entropy. We provide simple examples identifying two other possible definitions, via the pointwise entropy of the shift on trajectories and via the action of T^* on probability measures, as not being satisfactory (too large). We do not investigate the relation between our topological entropy and that introduced by [**LW**].

Finally, in §5 we are able to prove the analog of the famous Goodwyn theorem (see e.g. [**DGS**, Theorem (18.4)], or [**W**, Theorem 8.6], the '$\geq$' part): the topological entropy of a Markov operator dominates its measure-theoretic entropy with respect to each invariant Radon probability measure. The converse inequality completing the operator variational principle remains an open question.

2. *The axiomatic measure-theoretic definition*

2.1. *The axioms.* For a measurable space (X, Σ) consider an operator (not even necessarily linear) $T : L \to L$ on some set L of measurable functions with range contained in [0,1], and let μ be a probability measure on Σ. Any reasonable way to define the *entropy* $h_\mu(T)$ of T with respect to μ would have to follow the major steps listed below:

(1) one needs to specify $\mathbf{F}$, a collection of selected finite families $\mathcal{F}$ of functions from L; this collection should be T-invariant, so that $\mathcal{F} \in \mathbf{F}$ implies $T\mathcal{F} \in \mathbf{F}$, where $T\mathcal{F}$ stands for $\{Tf : f \in \mathcal{F}\}$;

(2) one has to specify an associative and commutative operation $\sqcup$ of *joining* these families, so that $\mathcal{F} \sqcup \mathcal{G} \in \mathbf{F}$ whenever $\mathcal{F} \in \mathbf{F}$ and $\mathcal{G} \in \mathbf{F}$, and with the cardinality of the joint family bounded by a number depending on the cardinalities of the components;

(3) one needs to define the entropy $H_\mu(\mathcal{F})$ of a family $\mathcal{F} \in \mathbf{F}$ with respect to μ;

(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
$$h_\mu(T) = \sup_{\mathcal{F}\in\mathbf{F}} h_\mu(T, \mathcal{F}).$$

For example, the classical Kolmogorov–Sinai entropy for measurable maps uses $\mathbf{F}$ defined as families of characteristic functions corresponding to measurable partitions, joining is obtained by pointwise multiplication or equivalently by the application of pointwise infima. Some other more general definitions (see [**AF, GLW**]) use for $\mathbf{F}$ the measurable *partitions of unity*, $\mathcal{F} = \{f_i : 1 \le i \le r\}$ with each f_i non-negative and with $\sum_i f_i = 1$ (actually, $\sum_i f_i^2 = 1$ in [**AF**]). In both cases joinings are done via pointwise multiplication. In our (new) definition (see §3) we let $\mathbf{F}$ be all finite families of functions with range in [0, 1] and we use the ordinary set union for joining.

This standard construction is usually accompanied with some definition of a conditional entropy—a tool useful in verifying properties of entropy. At this level of generality we define this quantity by the formula
$$H_\mu(\mathcal{F}|\mathcal{G}) = H_\mu(\mathcal{F} \sqcup \mathcal{G}) - H_\mu(\mathcal{G}).$$

A notion of entropy should possess some elementary 'nice' properties, known for the Kolmogorov–Sinai entropy of measurable maps, and it should coincide with this classical notion on transformations. This leads to formulation of several conditions which we call the *axioms of entropy*.

Axiom (A). (Monotonicity and subadditivity axiom) For $\mathcal{F}$, $\mathcal{G}$ and $\mathcal{H}$ belonging to $\mathbf{F}$ we require that
$$0 \le H_\mu(\mathcal{F}|\mathcal{H}) \le H_\mu(\mathcal{F} \sqcup \mathcal{G}|\mathcal{H}) \le H_\mu(\mathcal{F}|\mathcal{H}) + H_\mu(\mathcal{G}|\mathcal{H}),$$
where, by convention, $H_\mu(\mathcal{F}|\mathcal{H}) = H_\mu(\mathcal{F})$ for $\mathcal{H}$ being the empty family.

https://doi.org/10.1017/S014338570400032X Published online by Cambridge University Press

We point out that this axiom alone implies the following, the proof of which is standard in entropy theory and will be omitted:

$$H_\mu\bigg(\bigsqcup_{k=1}^{n}\mathcal{F}_k\bigg|\bigsqcup_{k=1}^{n}\mathcal{G}_k\bigg) \le \sum_{k=1}^{n} H_\mu(\mathcal{F}_k|\mathcal{G}_k). \tag{2.1.1}$$

It also easily implies that, for every $n \ge 1$,

$$h_\mu(T, T^n\mathcal{F}) = h_\mu(T, \mathcal{F}). \tag{2.1.2}$$

Axiom (B). (L^1-continuity axiom) For two families of functions $\mathcal{F} = \{f_i : 1 \le i \le r\}$ and $\mathcal{G} = \{g_i : 1 \le i \le r'\}$, $r' \le r$, we define their L^1 distance as

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

where the minimum ranges over all permutations π of the set $\{1, 2, \ldots, r\}$, and where $\mathcal{G}$ is considered an r-element family by setting $g_i \equiv 0$ for $r' < i \le r$. In this axiom we require that for every $r \ge 1$ and $\varepsilon > 0$ there is a $\delta_\varepsilon > 0$ such that if $\mathcal{F}$, $\mathcal{G}$ and $\mathcal{H}$ have cardinalities at most r and $\operatorname{dist}(\mathcal{F}, \mathcal{G}) < \delta_\varepsilon$ then

$$\operatorname{dist}(\mathcal{F} \sqcup \mathcal{H}, \mathcal{G} \sqcup \mathcal{H}) < \varepsilon$$

and

$$|H_\mu(\mathcal{F}) - H_\mu(\mathcal{G})| < \varepsilon. \tag{2.1.3}$$

Combining both parts of the continuity axiom we obtain, in particular, that for every r and ε there exists a δ such that $\operatorname{dist}(\mathcal{F}, \mathcal{G}) < \delta$ implies

$$|H_\mu(\mathcal{H}|\mathcal{F}) - H_\mu(\mathcal{H}|\mathcal{G})| < \varepsilon, \tag{2.1.4}$$

whenever all families involved have at most r elements.

Axiom (C). (Partitions axiom) If $\mathcal{A}$ is a measurable partition of X then $\mathbf{1}_\mathcal{A} = \{\mathbf{1}_A : A \in \mathcal{A}\}$ denotes the family of the corresponding characteristic functions. We require that characteristic functions of measurable sets belong to L and $\mathbf{F}$ contains $\mathbf{1}_\mathcal{A}$ for every measurable partition $\mathcal{A}$ of X. The joinings and entropy H_μ should coincide on partitions with the classic notions:

$$\mathbf{1}_\mathcal{A} \sqcup \mathbf{1}_\mathcal{B} = \mathbf{1}_{\mathcal{A}\vee\mathcal{B}}, \quad H_\mu(\mathbf{1}_\mathcal{A}) = -\sum_{A\in\mathcal{A}} \mu(A)\log\mu(A).$$

In the proofs that a given entropy notion coincides on maps with the classic one, the entropy of each family $\mathcal{F}$ must be majorized by the entropy of some partition. Usually such a partition is obtained by preimages via the functions in $\mathcal{F}$ of a sufficiently fine partition of the range. The axiom below along with axiom (C) is easily seen to guarantee coincidence with the Kolmogorov–Sinai entropy for measure-preserving transformations.

Axiom (D). (Domination axiom) For every $r \ge 1$ and $\varepsilon > 0$ there exists $\gamma > 0$ such that for every family $\mathcal{F} = \{f_i : 1 \le i \le r\}$ and every partition α of the unit interval into finitely many subintervals of lengths not exceeding γ,

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

https://doi.org/10.1017/S014338570400032X Published online by Cambridge University Press

where $\overline{\alpha}$ is some finite family depending only on α and satisfying $\lim_n(1/n) H_\mu\big(\bigsqcup_{k=1}^n \overline{\alpha}\big) = 0$ (usually $\overline{\alpha}$ is the empty family, meaning that it should be skipped, or it is a family of some constant functions).

By a brief inspection we verify that all the mentioned examples of notions of entropy satisfy the above axioms (see [**AF**, Lemmas 11.1 and 11.4] for Axiom (A) with regard to the quantum entropy, the Axiom (D) is implicitly included in the proof of Theorem 11.2 there; see [**KSL**] for the Axiom (A) with regard to the [**GLW**] entropy, and the proof of the main theorem in [**GLW**] for Axiom (D); the rest is either obvious or explicit). In what follows we prove that all notions of entropy satisfying the axioms coincide not only on transformations but also on all doubly stochastic operators; in other words, we prove the following theorem.

THEOREM 2.1. *If T is a doubly stochastic operator on $L^\infty(\mu)$ then the Axioms (A)–(D) (along with the construction steps (1)–(5)) completely determine the value of $h_\mu(T)$.*

The proof will be based on two observations concerning a doubly stochastic operator: it eventually nearly preserves lattice operations, and, as a consequence, it sends certain characteristic functions to nearly characteristic functions.

2.2. *Asymptotic lattice stability.* Throughout the rest of this section we assume that T is a doubly stochastic operator on $L^\infty(\mu)$.

LEMMA 2.2. *Let f, g be two bounded measurable functions on X. For every $\delta > 0$ there exists $N \in \mathbb{N}$ such that for every $k \in \mathbb{N}$, $l \geq N$ we have*

$$\int |T^k(T^l f \vee T^l g) - (T^{k+l} f \vee T^{k+l} g)|\, d\mu < \delta$$

and

$$\int |T^k(T^l f \wedge T^l g) - (T^{k+l} f \wedge T^{k+l} g)|\, d\mu < \delta.$$

Proof. Clearly, we have $T(f \vee g) \geq Tf$ and $T(f \vee g) \geq Tg$, implying

$$T(f \vee g) \geq Tf \vee Tg \quad \text{(and analogously } T(f \wedge g) \leq Tf \wedge Tg). \tag{2.2.1}$$

Since T preserves the measure, for each n we obtain

$$\int T^n f \vee T^n g\, d\mu = \int T(T^n f \vee T^n g)\, d\mu \geq \int T^{n+1} f \vee T^{n+1} g\, d\mu,$$

leading to a decreasing and bounded (hence convergent) sequence of integrals. Given $\delta > 0$ one can find N so large that for every $l \geq N$ and every k

$$0 \leq \int T^l f \vee T^l g\, d\mu - \int T^{k+l} f \vee T^{k+l} g\, d\mu \leq \delta. \tag{2.2.2}$$

Since T^k preserves the measure and the pointwise inequality (2.2.1) between $T^k(T^l f \vee T^l g)$ and $T^{k+l} f \vee T^{k+l} g$ holds, the above difference represents the desired L^1-distance. The proof for infima is analogous. □

https://doi.org/10.1017/S014338570400032X Published online by Cambridge University Press

Definition. For a function f and two constants $a < b$ denote $f_a^b = (f \vee a) \wedge b$. 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 every pair of constants $a < b$.

LEMMA 2.3. *For every bounded function f and every $\delta > 0$ there exists an integer l such that $T^l f$ has property* CZ(δ).

Proof. Temporarily fix the constants a, b. By Lemma 2.2 there exists l' such that for any n the L^1-distance between $T^n((T^{l'} f) \vee a)$ and $(T^{n+l'} f) \vee a$ is smaller than $\delta/9$. The same result applied to $g = (T^{l'} f) \vee a$ allows us to find l'' such that for every n

$$\int |T^n((T^{l''} g) \wedge b) - (T^{n+l''} g) \wedge b| \, d\mu < \frac{\delta}{9}.$$

By the choice of g, and because T is an L^1-contraction and so is the application of an infimum with a constant, we obtain

$$\int |T^n((T^{l'+l''} f)_a^b) - (T^{n+l'+l''} f)_a^b| \, d\mu < \frac{\delta}{3}.$$

Let l be the largest of the integers $l' + l''$ obtained for finitely many choices of constants a, b distributed in $\delta/3$ distances over the range of f. Since changing the constants by less than $\delta/3$ moves the function f_a^b by at most $\delta/3$ in L^1, we can see that the last inequality holds for l replacing $l' + l''$, δ replacing $\delta/3$, and any a, b, as required. □

The same technique allows us to prove another result, which is of independent interest. By a *lattice polynomial* on r functions we shall mean any r-argument formal expression involving the lattice operations $\vee$ and $\wedge$ (and brackets). An example of a lattice polynomial on three functions is $\theta(\cdot, \cdot, \cdot)$ defined by $\theta(f, g, h) = (f \vee g) \wedge h$. When applying a lattice polynomial to a family $\mathcal{F}$ we implicitly fix a certain order in $\mathcal{F}$.

Definition. We say that a family $\mathcal{F} = \{f_i : 1 \leq i \leq r\}$ is *lattice δ-stable under T* if for any lattice polynomial θ on r functions and any $n \geq 1$ we have

$$\int |T^n(\theta(\mathcal{F})) - \theta(T^n \mathcal{F})| \, d\mu < \delta.$$

PROPOSITION 2.4. *For every finite family $\mathcal{F}$ of bounded functions and $\delta > 0$ there exists an integer $l \geq 0$ such that $T^l \mathcal{F}$ is lattice δ-stable.*

Proof. For two-element families the assertion coincides with Lemma 2.2. For larger families (and more complex lattice polynomials) the proof follows by induction: suppose

$$\theta(\mathcal{F} \cup \mathcal{G}) = \theta_1(\mathcal{F}) \vee \theta_2(\mathcal{G}).$$

First we find m' such that both $T^{m'} \mathcal{F}$ and $T^{m'} \mathcal{G}$ are lattice $(\delta/6)$-stable. Then, Lemma 2.2 applied to the functions $\theta_1(T^{m'} \mathcal{F})$ and $\theta_2(T^{m'} \mathcal{G})$ and $\delta/3$ produces an integer m''. The sum $m = m' + m''$ provides δ-stability for the polynomial θ on $T^m(\mathcal{F} \cup \mathcal{G})$. Since there are only finitely many distinct lattice polynomials on r functions, we will be able to find the desired integer l. □

https://doi.org/10.1017/S014338570400032X Published online by Cambridge University Press

2.3. *Behavior of special partitions.*

LEMMA 2.5. *Let* $\mathcal{F} \subset \mathbf{F}$ *consist of* r *functions with ranges in* $[0, 1]$*, satisfying the condition* $\mathrm{CZ}(\delta^3)$*. Then for every* n *and* $f \in \mathcal{F}$

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

for certain values ξ*, not depending on* n *or* $f \in \mathcal{F}$*, distributed in distances not exceeding* $2r\delta$*.*

Proof. Cover $[0, 1]$ by at most $(1/2\delta^2) + 1$ equal intervals of length $2\delta^2$, which we call 'pieces'. Obviously, at most $r((1/\delta) - 1)$ of them may have a preimage by some $f \in \mathcal{F}$ of measure larger than δ. This implies that each interval of length $2r\delta$ contains the center ξ of a piece, whose preimage by any function $f \in \mathcal{F}$ has measure at most δ. We can write it as

$$\mu\{x : \xi - \delta^2 \leq f(x) < \xi + \delta^2\} \leq \delta.$$

Notice the following:

$$\frac{f_\xi^{\xi+\delta^2} - \xi}{\delta^2} \leq \mathbf{1}_{\{x:f(x)\geq\xi\}} \leq \frac{f_{\xi-\delta^2}^{\xi} - (\xi - \delta^2)}{\delta^2}.$$

Denote the outer functions by F_f and G_f. Note also that

$$0 \leq G_f - F_f \leq \mathbf{1}_{\{x:\xi-\delta^2\leq f(x)<\xi+\delta^2\}}.$$

Thus, $\int |G_f - F_f|\,d\mu \leq \delta$. Applying T^n we get

$$T^n F_f \leq T^n(\mathbf{1}_{\{x:f(x)\geq\xi\}}) \leq T^n G_f,$$

and, again, the L^1-distances are smaller than δ, because T is an $L^1(\mu)$-contraction. On the other hand, we have

$$F_{T^n f} = \frac{(T^n f)_\xi^{\xi+\delta^2} - \xi}{\delta^2} \leq \mathbf{1}_{\{x:T^n f(x)\geq\xi\}} \leq \frac{(T^n f)_{\xi-\delta^2}^{\xi} - (\xi - \delta^2)}{\delta^2} = G_{T^n f}.$$

However, we cannot repeat the estimate for the L^1-distance. But the property $\mathrm{CZ}(\delta^3)$ yields that the L^1-distance between $F_{T^n f}$ and $T^n F_f$ is less than $\delta^3/\delta^2 = \delta$, and similarly for G_f. This easily implies that the L_1-distance between $\mathbf{1}_{\{x:T^n f(x)\geq\xi\}}$ and $T^n(\mathbf{1}_{\{x:f(x)\geq\xi\}})$ is smaller than 4δ. □

LEMMA 2.6. *Let* $\mathcal{F} = \{f_i : 1 \leq i \leq r\}$ *consist of functions with ranges in* $[0, 1]$*, all having the property* $\mathrm{CZ}(\delta^3)$*, and let* α *be a partition of* $[0, 1]$ *into* m *pieces* $A_0 = [0, \xi_1)$*,* $A_j = [\xi_j, \xi_{j+1})$ $(j = 1, \ldots, m-2)$ *and* $A_{m-1} = [\xi_{m-1}, 1]$*, where the points* ξ_j *all satisfy the assertion of Lemma 2.5. Then*

$$\mathrm{dist}(T^n(\mathbf{1}_{\bigvee_i f_i^{-1}(\alpha)}), \mathbf{1}_{\bigvee_i T^n f_i^{-1}(\alpha)}) < 8rm^r\delta,$$

for every $n \geq 0$*.*

https://doi.org/10.1017/S014338570400032X Published online by Cambridge University Press

Proof. The family $\mathbf{1}_{\bigvee_i f_i^{-1}(\alpha)}$ consists of functions of the form

$$\mathbf{1}_{\bigcap_i \{x: f_i(x) \in A_{j_i}\}} = \inf_i \mathbf{1}_{\{x: f_i(x) \in A_{j_i}\}},$$

and their total sum (over $1 \le i \le r$ and all sequences (j_i) belonging to the set $\{0, 1, \ldots, m-1\}^r$) is the constant function $\mathbf{1}$. By (2.2.1), for each n the images must satisfy

$$T^n(\mathbf{1}_{\bigcap_i \{x: f_i(x) \in A_{j_i}\}}) \le \inf_i T^n(\mathbf{1}_{\{x: f_i(x) \in A_{j_i}\}}), \tag{2.3.1}$$

and the total sum of the functions on the left must also be $\mathbf{1}$. Since

$$\mathbf{1}_{\{x: f_i(x) \in A_{j_i}\}} = \mathbf{1}_{\{x: f_i(x) \ge \xi_{j_i}\}} - \mathbf{1}_{\{x: f_i(x) \ge \xi_{j_i+1}\}},$$

by Lemma 2.5 and the choice of the points ξ_j, the L^1-distance between

$$T^n(\mathbf{1}_{\{x: f_i(x) \in A_{j_i}\}}) \quad \text{and} \quad \mathbf{1}_{\{x: T^n f_i(x) \in A_{j_i}\}}$$

is smaller than 8δ. Thus,

$$\int |\inf_i T^n(\mathbf{1}_{\{x: f_i(x) \in A_{j_i}\}}) - \inf_i \mathbf{1}_{\{x: T^n f_i(x) \in A_{j_i}\}}| \, d\mu < 8r\delta. \tag{2.3.2}$$

Note that the last function can be written as $\mathbf{1}_{\bigcap_i \{x: T^n f_i(x) \in A_{j_i}\}}$. These functions constitute the family $\mathbf{1}_{\bigvee_i T^n f_i^{-1}(\alpha)}$ and their sum is $\mathbf{1}$. Now, combining (2.3.1) and (2.3.2), we see that

$$T^n(\mathbf{1}_{\bigcap_i \{x: f_i(x) \in A_{j_i}\}}) \le \mathbf{1}_{\bigcap_i \{x: T^n f_i(x) \in A_{j_i}\}} + g,$$

where g is some positive function with integral not exceeding $8r\delta$. Because we are now comparing two families with the same sum, the largest difference of the form

$$|\mathbf{1}_{\bigcap_i \{x: T^n f_i(x) \in A_{j_i}\}} - T^n(\mathbf{1}_{\bigcap_i \{x: f_i(x) \in A_{j_i}\}})|$$

may have integral at most $8r\delta$ times the cardinality of the family of the functions $\mathbf{1}_{\bigcap_i \{x: T^n f_i(x) \in A_{j_i}\}}$, which is less than or equal to m^r. The proof is now complete. □

2.4. *Main proof.* We are now in a position to prove Theorem 2.1.

Proof. Consider an entropy definition following steps (1)–(5) and satisfying axioms (A)–(D). Fix $\varepsilon > 0$ and let $\mathcal{F} = \{f_i : 0 \le i \le r\} \in \mathbf{F}$ be such that

$$h_\mu(T) < h_\mu(T, \mathcal{F}) + \varepsilon$$

(see step (5)). Let γ be as specified in the domination Axiom (D) for the cardinality r of $\mathcal{F}$ and ε, and choose m between $1/\gamma$ and $2/\gamma$. Pick δ such that the inequality (2.1.4) is satisfied for m^r and ε. Replacing $\mathcal{F}$ by $T^l\mathcal{F}$ (see (2.1.2)) with an appropriate l as specified in Lemma 2.3, we can assume that every $f_i \in \mathcal{F}$ has property $\mathrm{CZ}((\delta/8rm^r)^3)$. The number $\delta/4m^r$ majorizes the distances between the points ξ in Lemma 2.5 (with δ replaced by $\delta/8rm^r$) and, because it is much smaller than γ (in particular, smaller than $\gamma/2$), we can pick $m-1$ of them, creating a partition α of $[0, 1]$ into m intervals of lengths smaller than γ. For each k, the domination axiom applied to $T^k\mathcal{F}$ yields

$$H_\mu(T^k\mathcal{F} | \mathbf{1}_{\bigvee_i T^k f_i^{-1}(\alpha)} \sqcup \overline{\alpha}) < \varepsilon.$$

https://doi.org/10.1017/S014338570400032X Published online by Cambridge University Press

Then, as an application of (2.1.1), we obtain for each n

$$H_\mu\left(\mathcal{F}^n\,\middle|\,\bigsqcup_{k=0}^{n-1}(\mathbf{1}_{\bigvee_i T^k f_i^{-1}(\alpha)}\sqcup\overline{\alpha})\right)<n\varepsilon.$$

Since $\lim_n(1/n)H_\mu(\bigsqcup_{k=1}^n\overline{\alpha})=0$, Axiom (A) and step (4) easily imply

$$h_\mu(T,\mathcal{F})\le\limsup_{n\to\infty}\frac{1}{n}H_\mu\left(\bigsqcup_{k=0}^{n-1}\mathbf{1}_{\bigvee_i T^k f_i^{-1}(\alpha)}\right)+\varepsilon.$$

Denote the above lim sup expression by $h_\mu(T,\mathcal{F},\alpha)$ and notice that it involves joinings and entropies exclusively of partitions, hence is completely determined by the partitions Axiom (C). Lemma 2.6 yields

$$\operatorname{dist}(\mathbf{1}_{\bigvee_i T^k f_i^{-1}(\alpha)},T^k(\mathbf{1}_{\bigvee_i f_i^{-1}(\alpha)}))<\delta,$$

which, by (2.1.4), implies

$$H_\mu(\mathbf{1}_{\bigvee_i T^k f_i^{-1}(\alpha)}|T^k(\mathbf{1}_{\bigvee_i f_i^{-1}(\alpha)}))<H_\mu(\mathbf{1}_{\bigvee_i T^k f_i^{-1}(\alpha)}|\mathbf{1}_{\bigvee_i T^k f_i^{-1}(\alpha)})+\varepsilon.$$

By the partitions Axiom (C) the right-hand side is just ε. Using (2.1.1) and step (4) again, we deduce that

$$h_\mu(T,\mathcal{F},\alpha)\le h_\mu(T,\mathbf{1}_{\bigvee_i f_i^{-1}(\alpha)})+\varepsilon\le h_\mu(T)+\varepsilon.$$

We have proved that

$$h_\mu(T,\mathcal{F},\alpha)-\varepsilon\le h_\mu(T)\le h_\mu(T,\mathcal{F},\alpha)+2\varepsilon.$$

In this manner $h_\mu(T)$ is completely determined by the axioms. □

3. *The explicit measure-theoretic definition*

3.1. *The functional definition.* A possible definition of entropy for doubly stochastic operators emerges from the previous proof by means of $h_\mu(T,\mathcal{F},\alpha)$. However, the technical restrictions on α make this definition rather inconvenient. As shown in Example 3.2 at the end of the section, dropping these conditions essentially affects the outcome notion. Instead, we propose a much more natural approach, which can be applied to literally any mapping $T: L\to L$. (We recall that L is a set of measurable functions with range in $[0,1]$.) We follow steps (1)–(5) of the axiomatic definition.

(1°) We let $\mathbf{F}$ be the collection of all finite subsets $\mathcal{F}$ of L.

(2°) $\mathcal{F}\sqcup\mathcal{G}$ is defined as the ordinary union $\mathcal{F}\cup\mathcal{G}$.

(3°) For a function f let

$$A_f=\{(x,t)\in X\times[0,1]:t\le f(x)\}$$

and denote by $\mathcal{A}_f$ a partition of $X\times[0,1]$ consisting of A_f and its complement A_f^c. For a collection $\mathcal{F}$ we define

$$\mathcal{A}_\mathcal{F}=\bigvee_{f\in\mathcal{F}}\mathcal{A}_f.$$

We now set

$$H_\mu(\mathcal{F})=H_{\widehat{\mu}}(\mathcal{A}_\mathcal{F})=-\sum_{A\in\mathcal{A}_\mathcal{F}}\widehat{\mu}(A)\log\widehat{\mu}(A),$$

where $\widehat{\mu}$ is the product of μ with the Lebesgue measure on the unit interval.

https://doi.org/10.1017/S014338570400032X Published online by Cambridge University Press

(4°) $h_\mu(T, \mathcal{F}) = \limsup_{n\to\infty} 1/n H_\mu(\mathcal{F}^n)$.
(5°) $h_\mu(T) = \sup_{\mathcal{F}\in\mathbf{F}} h_\mu(T, \mathcal{F})$.

Remark 3.1.
(i) If $\mathcal{F}, \mathcal{G} \in \mathbf{F}$ then $\mathcal{A}_{\mathcal{F}\cup\mathcal{G}} = \mathcal{A}_\mathcal{F} \vee \mathcal{A}_\mathcal{G}$.
(ii) In the definition of H_μ every family $\mathcal{F}$ can be replaced by an increasing family $\Theta(\mathcal{F})$ generating $\mathcal{A}_{\Theta(\mathcal{F})} = \mathcal{A}_\mathcal{F}$. This family is obtained by means of lattice polynomials as described below.
For r denoting the cardinality of $\mathcal{F}$, we introduce a linear order of the set $\{0, 1\}^r$ in the following way: $\alpha = (\alpha_i) \preccurlyeq \beta = (\beta_i)$ if either $\sum_i \alpha_i \geq \sum_i \beta_i$ or the sums are equal and $\alpha_i = 1$ at the first position i satisfying $\alpha_i \neq \beta_i$. Defining

$$\theta_\beta(\mathcal{F}) = \sup_{\alpha\preccurlyeq\beta} \inf\{f_i : \alpha_i = 1\}$$

we obtain an increasing finite sequence of functions $\Theta(\mathcal{F}) = \{\theta_\beta(\mathcal{F})\}_\beta$, to which we enclose the constants $\mathbf{0}$ (as the first element) and $\mathbf{1}$ (as the last one). With this notation

$$H_\mu(\mathcal{F}) = H_\mu(\Theta(\mathcal{F})) = -\sum_g \left(\int g\, d\mu\right) \cdot \log\left(\int g\, d\mu\right), \tag{3.1.1}$$

where g ranges over all differences between pairs of consecutive functions in $\Theta(\mathcal{F})$.

We remark that the definition of $\Theta(\mathcal{F})$ depends on the order in $\mathcal{F}$. It is interesting to note that the functions g form a partition of unity, an object used quite differently by other authors. Because $\Theta(T(\mathcal{F})) \neq T(\Theta(\mathcal{F}))$, we cannot replace $\mathcal{F}$ by $\Theta(\mathcal{F})$ in the definition of $h_\mu(T, \mathcal{F})$.

The verification that the entropy defined in this section satisfies Axioms (A)–(C) is immediate. For (D), approximate each function $f_i \in \mathcal{F}$ by a simple function s_i with values at the breaking points ξ_j of the partition α. It easily follows from step (3°) of our definition and the formula (2.1.1) that $H_\mu(\mathcal{F}|\{s_i\})$ is small, while the partition associated with $\mathbf{1}_{\bigvee_i f_i^{-1}(\alpha)} \cup \overline{\alpha}$, where $\overline{\alpha}$ is the family of constant functions with the same values ξ_j, is finer than the one induced by $\{s_i\}$.

In addition to the axioms, the following condition, which seems to be natural in view of the information origins of entropy, is also true:

$$H_\mu(\mathcal{F}|\mathcal{F}) = 0.$$

Note that it is satisfied neither for [**GLW**] entropy (see [**KSL**, the corollary following Proposition 1]) nor for the quantum entropy (try the two point space with the measure $\{1/2, 1/2\}$, and the operational partition consisting of functions $(1/\sqrt{2}, 0)$ and $(1/\sqrt{2}, 1)$). The formula (2.1.4) along with the above equality implies that for every $\varepsilon > 0$ there is a $\delta > 0$ such that

$$\operatorname{dist}(\mathcal{F}, \mathcal{G}) < \delta \Longrightarrow H_\mu(\mathcal{F}|\mathcal{G}) < \varepsilon, \tag{3.1.2}$$

which, combined with (2.1.1), yields for a L^1-contraction T

$$\operatorname{dist}(\mathcal{F}, \mathcal{G}) < \delta \Longrightarrow |h_\mu(T, \mathcal{F}) - h_\mu(T, \mathcal{G})| < \varepsilon. \tag{3.1.3}$$

https://doi.org/10.1017/S014338570400032X Published online by Cambridge University Press

Also, our definition satisfies

$$h_\mu(T, \mathcal{F}_\phi) = 0 \tag{3.1.4}$$

for any family $\mathcal{F}_\phi$ of T-invariant functions.

Unlike in the pointwise case, H_μ need not be T-invariant, even for doubly stochastic operators. Moreover, it can increase under the action of T, which will be illustrated in a simple example below. Nevertheless, we are able to obtain an asymptotic invariance of H_μ (see Lemma 3.1).

Example 3.1. Let T be the doubly stochastic operator on the unit interval (equipped with the Lebesgue measure) defined by $Tf(x) = (1/2)(f(x) + f(1-x))$. Take $\mathcal{F}$ consisting of characteristic functions $\mathbf{1}_{[0,1/4]}$ and $\mathbf{1}_{(1/4,1]}$. Then $H_\mu(\mathcal{F}) = 2\log 2 - (3/4)\log 3 < \log 2$, while $H_\mu(T\mathcal{F}) = (3/2)\log 2$. Note that the pathology cannot be removed by the sometimes useful trick of refining $\mathcal{F}$ by some set of constants.

LEMMA 3.1. *Let T be doubly stochastic. For every $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that for every k*

$$|H_\mu(T^{k+N}\mathcal{F}) - H_\mu(T^N\mathcal{F})| < \varepsilon.$$

Proof. The assertion follows immediately from the lattice-based definition (3.1.1) of H_μ, asymptotic lattice stability of Proposition 2.4 and the continuity axiom (2.1.3). □

3.2. *The existence of the limit.* This section is devoted to proving the existence of the limit in the entropy construction step (4°), which is one of the advantages of our definition over the others. The key problem is the lack of subadditivity of the sequence $H_\mu(\mathcal{F}^n)$. By Axiom (A), we do have

$$H_\mu(\mathcal{F}^{n+m}) \leq H_\mu(\mathcal{F}^n) + H_\mu(T^n\mathcal{F}^m),$$

but we cannot drop T^n. The asymptotic T-invariance of Lemma 3.1 is insufficient. Increasing the number of functions in order to obtain $\mathcal{F}^m$, we lose control over N and ε.

A suitable strengthening is provided in Lemma 3.2. The main tool is the theorem on the integral representation of certain stochastic operators proved by A. Iwanik in [**I**]. It asserts that 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 with the Lebesgue measure and $(\omega, x) \mapsto \varphi_\omega(x)$ is a jointly measurable map from $\Omega \times X$ into X. On the product space $\Omega \times X$ consider the action of the map $(\omega, x) \mapsto (\omega, \varphi_\omega(x))$ and the associated pointwise generated operator Φ on bounded measurable functions of $\Omega \times X$. Denoting by $\overline{f}$ the function $(\omega, x) \mapsto f(x)$ we have

$$Tf(x) = \int \Phi\overline{f}(\omega, x)\, d\lambda(\omega). \tag{3.2.1}$$

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

$$\iint \Phi\overline{f}\, d\lambda\, d\mu = \int Tf\, d\mu = \int f\, d\mu. \tag{3.2.2}$$

https://doi.org/10.1017/S014338570400032X Published online by Cambridge University Press

Since every iterate of T is also induced by a transition probability, we may denote by Φ_k the pointwise generated operator corresponding to T^k. Note that, in general, Φ_k is not equal to the iterate Φ^k. (In fact, using a cocycle construction one can modify Φ so that the equality $\Phi_k = \Phi^k$ holds. Then, Φ becomes a dilation of T in the sense of [**FN**], where T is viewed as a contraction on $L^2(\mu)$. We skip the details.)

LEMMA 3.2. *Let T be a doubly stochastic operator on $L^\infty(\mu)$. For every $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that for every $k, m \in \mathbb{N}$, we have*

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

Proof. We first prove the lemma under the additional assumption that X is standard Borel and T is as required in Iwanik's theorem. Let ν denote the product measure $\lambda \times \mu$. We start by showing the following.

$$\text{For } f \in L \text{ and sufficiently large } l \text{ the functions } \overline{T^{k+l}f} \text{ and } \Phi_k\overline{T^l f} \text{ are close in the norm of } L^1(\nu) \text{ for every } k \geq 1. \tag{3.2.3}$$

Fix a positive integer M and $\delta > 0$. Since Φ_k is pointwise generated, for every bounded g we have

$$\int |\Phi_k\overline{g}|\,d\nu = \int \Phi_k|\overline{g}|\,d\nu \overset{(3.2.2)}{=} \int |g|\,d\mu,$$

which, combined again with (3.2.2), yields

$$\int (\Phi_k\overline{g})^+\,d\nu = \int g^+\,d\mu. \tag{3.2.4}$$

Since $(T^l f)^+ = T^l f \vee T^l 0$, by (2.2.2) one can find a positive integer N such that for every $l \geq N$ and $k \geq 1$

$$0 < \int (T^l f)^+\,d\mu - \int (T^{k+l}f)^+\,d\mu < \delta.$$

Hence, substituting $g = T^l f$ in (3.2.4) we obtain

$$0 \leq \int (\Phi_k\overline{T^l f})^+\,d\nu - \int (T^{k+l}f)^+\,d\mu < \delta. \tag{3.2.5}$$

To prove (3.2.3) it is enough to estimate the measure of the set of all points (ω, x) for which $|\Phi_k\overline{T^l f} - T^{k+l}f| \geq 2/M$. Denote

$$B(f) = \left\{(\omega, x) : \Phi_k\overline{T^l f}(\omega, x) > \frac{1}{M},\ T^{k+l}f(x) \leq 0\right\}.$$

https://doi.org/10.1017/S014338570400032X Published online by Cambridge University Press

In order to find the measure of $B(f)$ we proceed as follows.

$$\begin{aligned}\frac{1}{M}\cdot\nu(B(f)) &\le \int_{B(f)} \Phi_k\overline{T^l f}\,d\nu\\ &\le \int_{\{(\omega,x):T^{k+l}f(x)\le 0\}} (\Phi_k\overline{T^l f})^+\,d\nu\\ &= \int (\Phi_k\overline{T^l f})^+\,d\nu - \int_{\{(\omega,x):T^{k+l}f(x)>0\}} (\Phi_k\overline{T^l f})^+\,d\nu\\ &\le \int (\Phi_k\overline{T^l f})^+\,d\nu - \int_{\{(\omega,x):T^{k+l}f(x)>0\}} \Phi_k\overline{T^l f}\,d\nu\\ &\overset{(3.2.1)}{=} \int (\Phi_k\overline{T^l f})^+\,d\nu - \int_{\{(\omega,x):T^{k+l}f(x)>0\}} T^{k+l}f\,d\mu\\ &\overset{(3.2.5)}{<} \delta.\end{aligned}$$

Thus, $\nu(B(f)) < M\delta$. Clearly, we can improve N so that the above calculation remains valid for all functions of the form $f-(i/M)$, where $i=0,\ldots,M-1$. This yields the estimate

$$\nu\left(\left\{(\omega,x):\Phi_k\overline{T^l f}(\omega,x) - T^{k+l}f(x) > \frac{2}{M}\right\}\right) \le \sum_{i=0}^{M-1}\nu(B(f-(i/M))) < M^2\delta.$$

Taking $(-f)$ instead of f we also get

$$\nu\left(\left\{(\omega,x):\Phi_k\overline{T^l f}(\omega,x) - T^{k+l}f(x) < -\frac{2}{M}\right\}\right) < M^2\delta,$$

and hence

$$\int |\Phi_k\overline{T^l f} - \overline{T^{k+l}f}|\,d\mu < \frac{2}{M} + 2M^2\delta.$$

Because M and δ are arbitrary, (3.2.3) is proved.

To complete the proof in the case of Iwanik's theorem's validity, fix $\varepsilon > 0$ and choose N so large that for all r functions $f\in\mathcal{F}$, every $k\ge 1$ and $l\ge N$, the functions $\overline{T^{k+l}f}$ and $\Phi_k\overline{T^l f}$ are close enough (in $L^1(\nu)$) to ensure, according to (3.1.2), that both

$$H_\nu(\overline{T^{k+l}f}\,|\,\Phi_k\overline{T^l f}) < \frac{\varepsilon}{r} \quad\text{and}\quad H_\nu(\Phi_k\overline{T^l f}\,|\,\overline{T^{k+l}f}) < \frac{\varepsilon}{r}. \tag{3.2.6}$$

Pick $m\ge 1$. Denoting by $\overline{T^N\mathcal{F}^m}$ the collection of functions $\overline{T^l f}$, where $f\in\mathcal{F}$, $l=N,\ldots,N+m-1$, we can write

$$\begin{aligned}H_\mu(T^{k+N}\mathcal{F}^m) &= H_\nu(\overline{T^{k+N}\mathcal{F}^m})\\ &\le H_\nu(\Phi_k\overline{T^N\mathcal{F}^m}) + H_\nu(\overline{T^{k+N}\mathcal{F}^m}\,|\,\Phi_k\overline{T^N\mathcal{F}^m}).\end{aligned}$$

By Remark 3.1(ii),

$$\begin{aligned}H_\nu(\Phi_k\overline{T^N\mathcal{F}^m}) &= H_\nu(\Theta(\Phi_k\overline{T^N\mathcal{F}^m})) = H_\nu(\Phi_k\overline{\Theta(T^N\mathcal{F}^m)})\\ &\overset{(3.2.2)}{=} H_\mu(\Theta(T^N\mathcal{F}^m)) = H_\mu(T^N\mathcal{F}^m),\end{aligned}$$

https://doi.org/10.1017/S014338570400032X Published online by Cambridge University Press

and thus

$$H_\mu(T^{k+N}\mathcal{F}^m) \le H_\mu(T^N\mathcal{F}^m) + \sum_{l=N}^{N+m-1}\sum_{f\in\mathcal{F}} H_\nu(\overline{T^{k+l}f}|\Phi_k\overline{T^l f})$$
$$\overset{(3.2.6)}{<} H_\mu(T^N\mathcal{F}^m) + m\varepsilon.$$

Also,

$$\begin{aligned} H_\mu(T^N\mathcal{F}^m) &= H_\nu(\Phi_k\overline{T^N\mathcal{F}^m}) \\ &\le H_\nu(\overline{T^{k+N}\mathcal{F}^m}) + H_\nu(\Phi_k\overline{T^N\mathcal{F}^m}|\overline{T^{k+N}\mathcal{F}^m}) \\ &< H_\mu(T^{k+N}\mathcal{F}^m) + m\varepsilon \end{aligned}$$

and the assertion is proved.

Now, let T be an arbitrary doubly stochastic operator on $L^\infty(\mu)$ on any probability space. Note that the above argument involves only countably many functions obtained from the members of $\mathcal{F}$ via the iterates of T and also via some lattice polynomials. Denote by $\mathcal{L}$ the complex subalgebra of $L^\infty(\mu)$ generated by these functions. It is known that $\mathcal{L}$ is isometrically isomorphic to the algebra of (complex) continuous functions on a certain compact Hausdorff space $\mathcal{X}$, and the isomorphism τ preserves the involution; thus it sends real-valued functions to real-valued functions. Moreover, it can be shown that τ preserves also the lattice operations, hence the image of every positive element of $L^\infty(\mu)$ is a positive function. The measure μ, being a linear functional on $L^\infty(\mu)$, passes to a linear functional on the space of continuous functions, represented by a certain Borel probability measure on $\mathcal{X}$, which implies that the integrals of functions corresponding via the isomorphism are equal. By [**Z**, Corollary 10.10], in our case the space $\mathcal{X}$ is metric (hence standard Borel), because the algebra $\mathcal{L}$ has a denumerable set of generators. Clearly, the operator $\mathcal{T} = \tau T\tau^{-1}$ defined on the real subalgebra of all real continuous functions $C(\mathcal{X})$ is Markov; thus it is induced by a Feller transition probability. Hence, the assertion (for $\mathcal{T}$) follows from the first part of the proof. By Remark 3.1(ii), and since τ is a lattice isomorphism, 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 $C(\mathcal{X})$, which completes the proof. □

The following result is a substitute, valid for doubly stochastic operators, for the subadditivity property. We call it *quasi-subadditivity*.

PROPOSITION 3.3. *Let T be a doubly stochastic operator. For every $\varepsilon > 0$ there exists $N \in \mathbb{N}$ and a constant c such that for every $k \in \mathbb{N}$ and $m \ge N$ we have*

$$H_\mu(\mathcal{F}^{k+m}) \le H_\mu(\mathcal{F}^k) + H_\mu(\mathcal{F}^m) + c + m\varepsilon.$$

Proof. Fix $\varepsilon > 0$ and choose N according to Lemma 3.2. If $k > 0$, $m > N$, then, using the fact that $H_\mu(\mathcal{F}) \le r\log 2$, where r is the cardinality of $\mathcal{F}$, we obtain

$$\begin{aligned} H_\mu(\mathcal{F}^{k+m}) &\le H_\mu(\mathcal{F}^k) + H_\mu(T^k\mathcal{F}^N) + H_\mu(T^{k+N}\mathcal{F}^{m-N}) \\ &\le H_\mu(\mathcal{F}^k) + Nr\log 2 + H_\mu(T^N\mathcal{F}^{m-N}) + (m-N)\varepsilon \\ &\le H_\mu(\mathcal{F}^k) + H_\mu(\mathcal{F}^m) + Nr\log 2 + m\varepsilon. \end{aligned}$$ □

https://doi.org/10.1017/S014338570400032X Published online by Cambridge University Press

THEOREM 3.4. *If T is a doubly stochastic operator then the upper limit in* (4°) *is in fact a limit.*

Proof. Given the quasi-subadditivity of Proposition 3.3 the proof becomes a slightly modified copy of the classical lemma concerning the convergence of a_n/n for a subadditive sequence a_n (see [**DGS**, Proposition (10.7)] or [**W**, Theorem 4.9]). □

The above limit need not coincide with the corresponding infimum, as it holds in the pointwise case. Full subadditivity, required for such coincidence, is obtained if we modify the definition of entropy by including an additional step in the construction. We propose two possible such modifications.

Definition 3.1. $H'_\mu(\mathcal{F}) = \lim_{k\to\infty} H_\mu(T^k\mathcal{F})$ (existence of this limit is guaranteed by Lemma 3.1); $H''_\mu(\mathcal{F}) = \sup_{k\in\mathbb{N}} H_\mu(T^k\mathcal{F})$.

LEMMA 3.5. *The sequences $H'_\mu(\mathcal{F}^n)$ and $H''_\mu(\mathcal{F}^n)$ are subadditive.*

Proof. Clearly,

$$H'_\mu(T^m\mathcal{F}) = \lim_{k\to\infty} H_\mu(T^{m+k}\mathcal{F}) = H'_\mu(\mathcal{F}).$$

Thus,

$$H'_\mu(\mathcal{F}^{m+n}) \le H'_\mu(\mathcal{F}^m) + H'_\mu(T^m\mathcal{F}^n) = H'_\mu(\mathcal{F}^m) + H'_\mu(\mathcal{F}^n).$$

The proof for $H''_\mu(\mathcal{F}^n)$ is almost identical. Note only that $H''_\mu(T^m\mathcal{F})$ is smaller than or equal to $H''_\mu(\mathcal{F})$. □

Denote by $h'_\mu(T,\mathcal{F})$ and $h''_\mu(T,\mathcal{F})$ the quantities obtained by substituting H_μ in step (4°) of the construction of entropy by H'_μ or H''_μ respectively (according to the above lemma, the relevant limits exists).

PROPOSITION 3.6. *For a doubly stochastic operator T we have*

$$h_\mu(T,\mathcal{F}) = h'_\mu(T,\mathcal{F}) = h''_\mu(T,\mathcal{F}).$$

Proof. Clearly, $h_\mu(T,\mathcal{F}) \le h''_\mu(T,\mathcal{F})$ and $h'_\mu(T,\mathcal{F}) \le h''_\mu(T,\mathcal{F})$.

We show that $h'_\mu(T)$ dominates $h''_\mu(T)$. Fix a positive integer n. For the family $\mathcal{F}^n$ and $\varepsilon = 1$ choose N according to Lemma 3.1. Take m larger than $n+N$ and let p satisfy $(p-1)n \le m \le pn$. Then, find k such that the difference between $H_\mu(T^k\mathcal{F}^m)$ and $H''_\mu(\mathcal{F}^m)$ is smaller than 1. Let K be the maximum of N and k.

$$\begin{aligned} H''_\mu(\mathcal{F}^m) &\le H_\mu(T^k\mathcal{F}^m) + 1 \\ &\le H_\mu(\mathcal{F}^N) + H_\mu\left(\bigcup_{i=0}^{p-1} T^{K+in}\mathcal{F}^n\right) + 1 \\ &\le H_\mu(\mathcal{F}^N) + pH_\mu(T^N\mathcal{F}^n) + p + 1. \end{aligned}$$

Dividing the extreme sides of this inequality by m and letting m tend to infinity we obtain

$$h''_\mu(T,\mathcal{F}) \le \frac{1}{n}H_\mu(T^N\mathcal{F}^n) + \frac{1}{n}.$$

Now, taking the limit over N and afterwards over n leads to $h''_\mu(T,\mathcal{F}) \le h'_\mu(T,\mathcal{F})$.

In order to prove that $h_\mu(T,\mathcal{F}) \ge h'_\mu(T,\mathcal{F})$, by Lemma 3.2, given $\varepsilon > 0$ we can find N such that for arbitrary positive integers l, m we have

$$H_\mu(T^{l+N}\mathcal{F}^m) \le H_\mu(T^N\mathcal{F}^m) + m\varepsilon.$$

Letting $l \to \infty$ we obtain

$$H_\mu{}'(\mathcal{F}^m) \le H_\mu(T^N\mathcal{F}^m) + m\varepsilon,$$

hence $h_\mu{}'(T,\mathcal{F}) \le h_\mu(T, T^N\mathcal{F}) + \varepsilon \overset{(2.1.2)}{=} h_\mu(T,\mathcal{F}) + \varepsilon$ for arbitrary $\varepsilon > 0$. □

Unfortunately, the new quantities H'_μ and H''_μ fail the partitions axiom; the domination axiom is difficult to verify and most likely it also fails.

3.3. *Basic properties of entropy.* Below, we extend some properties of h_μ, known from the classical theory of dynamical systems. We begin with rephrasing the standard (pointwise) definitions.

Definition 3.2. Let T_1 and T_2 be doubly stochastic operators acting on $L^\infty(X_1,\mu_1)$ and $L^\infty(X_2,\mu_2)$, respectively.

(i) T_1 is a *factor* of T_2 if there is a measure-preserving surjection $\pi : X_2 \to X_1$ satisfying, for every $f \in L^\infty(X_1,\mu_1)$, the condition $(T_1 f)\circ\pi = T_2(f\circ\pi)$.

(ii) T_1 and T_2 are *isomorphic* if the map π defined above is invertible and the inverse is also measure-preserving.

PROPOSITION 3.7. *Let T_1, T_2 be doubly stochastic operators acting on $L^\infty(\mu_1)$ and $L^\infty(\mu_2)$, respectively.*

(i) *If T_1 is a factor of T_2 then $h_{\mu_1}(T_1) \le h_{\mu_2}(T_2)$.*

(ii) *If T_1 and T_2 are isomorphic then their entropies are equal.*

Proof. Let $\pi : X_2 \to X_1$ be a factor map. Let $\mathcal{F} \subset L^\infty(\mu_1)$ be a finite set of functions with ranges in $[0, 1]$ and denote

$$\pi\mathcal{F} = \{f\circ\pi : f \in \mathcal{F}\}.$$

Clearly, $h_{\mu_1}(T_1,\mathcal{F}) = h_{\mu_2}(T_2,\pi\mathcal{F})$; thus, (i) is now a consequence of the construction step (5°), and (ii) follows immediately. □

PROPOSITION 3.8. *If T is an operator on a set L of functions with range in* $[0, 1]$*, then for every $k \in \mathbb{N}$ we have*

$$h_\mu(T^k) = kh_\mu(T).$$

Proof. Because of Remark 3.1(i), the proof is obtained from the proof of the analogous classical property by making only minor modifications (see e.g. [**DGS**, Proposition (10.12)(c)] or [**W**, Theorem 4.13]). □

The following result concerns the nature of chaotic behavior recognized by our entropy. It establishes yet another natural property of our notion: the entropy vanishes for some sort of trivial dynamics.

https://doi.org/10.1017/S014338570400032X Published online by Cambridge University Press

PROPOSITION 3.9. *Let T be an operator on a set L of functions with range in* $[0, 1]$*, such that for every $f \in L$ the sequence $T^n f$ converges in L^1-norm to an invariant function ϕ_f. Then $h_\mu(T) = 0$.*

Proof. For every $\mathcal{F}$ and $\varepsilon > 0$ there exists a family $\mathcal{F}_\phi$ consisting of invariant functions, such that $\operatorname{dist}(T^N\mathcal{F}, \mathcal{F}_\phi)$ is small for some N. Then use (2.1.2), (3.1.3) and (3.1.4). □

We recall that some inconvenient technical assumptions regarding the choice of the partition α of the unit interval discouraged us from using $h_\mu(T, \mathcal{F}, \alpha)$ (see the proof of Theorem 2.1) as the explicit definition of entropy. We now provide the example, referred to at the beginning of this section, illustrating that skipping those assumptions may lead to an incorrect definition.

Example 3.2. Let (X, Σ, μ) be the set of one-sided 0–1 sequences $X = \{0, 1\}^{\mathbb{N}}$ with the product σ-algebra and with the uniform product measure $\mu = \{1/2, 1/2\}^{\mathbb{N}}$. Define the doubly stochastic operator T by

$$Tf(x) = \frac{1}{2}\left(f(\sigma x) + \int_X f(x)\,d\mu\right),$$

where σ denotes the shift $(x_n) \mapsto (x_{n+1})$ on X. Let $\mathcal{F}$ contain only one function, namely $f(x) = x_0$, and let $\alpha = \{[0, 1/2], (1/2, 1]\}$. Notice that f is the indicator of the cylinder set

$$C = \{x \in X : x_0 = 1\}.$$

Then

$$T^n f(x) = \frac{1}{2} + \frac{1}{2^n}\left(\mathbf{1}_{\sigma^{-n}(C)}(x) - \frac{1}{2}\right) = \begin{cases} \dfrac{1}{2} - \dfrac{1}{2^{n+1}} & \text{if } x_n = 0, \\[2mm] \dfrac{1}{2} + \dfrac{1}{2^{n+1}} & \text{if } x_n = 1, \end{cases}$$

converges in L^1-norm to the constant function $1/2$. It follows from Proposition 3.9 that $h_\mu(T) = 0$. But $H_\mu\big(\bigsqcup_{k=0}^{n-1} \mathbf{1}_{T^k f^{-1}(\alpha)}\big)$ is the entropy of a partition consisting of cylinder sets of length n hence, $h_\mu(T, \mathcal{F}, \alpha)$ is equal to $\log 2$.

It is interesting to observe that the discussed operator entropy extends the Kolmogorov–Sinai notion in an essential way: in addition to pointwise-type dynamics, it also captures some pure ‘operator dynamics’. In the example below an operator has positive entropy without admitting non-trivial pointwise generated factors at all.

Example 3.3. Let (X, Σ, μ) be the same as in the previous example. Let ν be the geometric distribution on natural numbers $\mathbb{N}$ given by $\nu(k) = 2^{-k}$. The element (x, k) of the product space $X \times \mathbb{N}$ can be visualized as the 0–1-valued sequence $(x_1, x_2, \ldots, x_k^*, x_{k+1}, \ldots)$ with a marker (star) over position k.

For each finite block $B = (b_1, b_2, \ldots, b_k) \in \{0, 1\}^k$ define the map $\sigma_B : X \to X$ by

$$(\sigma_B x)_n = \begin{cases} b_n & \text{for } n \le k, \\ x_{n+1} & \text{for } n > k. \end{cases}$$

https://doi.org/10.1017/S014338570400032X Published online by Cambridge University Press

Next, we define the operator T on $L^\infty(\mu \times \nu)$ as follows:

$$Tf(x,k) = f(\sigma x, k-1) \quad \text{if } k > 1,$$
$$Tf(x,1) = \sum_{k=1}^{\infty} 2^{-k} \sum_{B\in\{0,1\}^k} 2^{-k} f(\sigma_B x, k).$$

The corresponding transition probability $P((x,k),\cdot)$ can be described as the shift map (also shifting the position of the marker) on points with marker further to the right, while points with the marker over the first position are shifted and then the initial block of length k (chosen according to the geometric distribution) is replaced by a random block of length k. A marker is added over the last position of the replaced block. Our first claim is that T is doubly stochastic with respect to the product measure $\mu \times \nu$. To see this, consider the characteristic function $f = \mathbf{1}_C$ of a cylinder of the form

$$C = C(y_1, y_2, \dots, y_k^*, y_{k+1}, \dots y_n) = \{(x,k) : x_i = y_i,\ i = 1, \dots, n\}.$$

Clearly, the integral of f is $2^{-(k+n)}$. All points in both cylinders $C(x_0, y_1, y_2, \dots, y_k^*, y_{k+1}, \dots, y_n)$ ($x_0 \in \{0,1\}$) are sent deterministically into C. So, $Tf = 1$ on these two cylinders. The integral of this part is $2^{-(k+1)} \cdot 2 \cdot 2^{-(n+1)} = 2^{-(k+n+1)}$ (the marker appears at the position $k+1$). Also, each point in any cylinder $C(x_0^*, x_1, \dots, x_k, y_{k+1}, \dots, y_n)$ (with any choice of $x_0, \dots, x_k$) contributes with probability 2^{-2k} to C (via the map σ_B, where $B = (y_1, y_2, \dots, y_k)$), so f at such points is 2^{-2k}. The integral of this part is $2^{-2k} \cdot 2^{k+1} \cdot 2^{-1} \cdot 2^{-(n+1)} = 2^{-(k+n+1)}$. The sum of both parts equals the initial integral, so T is doubly stochastic with respect to $\mu \times \nu$.

Now, suppose T has a pointwise generated factor. This factor corresponds to a sub-σ-algebra Σ' with the property $P((x,k), A) = 0$ or 1 for each $A \in \Sigma'$ and almost every x. Consider a pair of cofinal points (x,k), (x',k'), i.e. such that $x_n = x'_n$ for n larger than some k''. We can assume $k'' > k$ and $k'' > k'$. Consider also a point (x'', k'') which coincides with both x and x' above the index k''. Both points (x'', k'') and (x,k) are accessible with positive transition probabilities from the point $(0x, 1) = 0^*x$ (which has 0^* at the first position and then looks like x shifted to the right). This means (unless 0^*x falls in some zero-measure set) that the factor identifies (x'', k'') with (x,k). Similarly, it identifies (x'', k'') with (x', k'), and hence (x,k) with (x',k'). We have shown that the factor identifies (almost surely) cofinal points regardless to positioning of the marker. This implies that Σ' is (up to measure μ) contained in the tail σ-algebra of $\{0,1\}^{\mathbb{Z}}$. By the Kolmogorov 0–1 law, such a σ-algebra is trivial with respect to the measure μ. So, T admits no non-trivial pointwise (measure-theoretic) factors.

Finally, we will show that $h_\mu(T) \geq \log 2$. Consider the 0–1-valued function $f_i(x,k) = x_i$. We will show that

$$T^n f_i(x,k) = \alpha \tfrac{1}{2} + (1-\alpha) x_{i+n},$$

with $\alpha \leq 2^{-(i-1)}$. This is obvious if $n < k$ (then $\alpha = 0$), because such T^n is deterministic on (x,k). Now consider $n = k$. This is the first time the point (x,k) is actually spread. Notice that with respect to $P^n((x,k),\cdot)$ the position of the marker has the geometric distribution ν. In the next step $n+1$, with probability 1/2 the markers will be shifted

https://doi.org/10.1017/S014338570400032X Published online by Cambridge University Press

(creating half of the same geometric distribution) and with probability 1/2—spread again (with the same geometric distribution). As a result, the distribution of the marker's position with respect to $P^{n+1}((x,k),\cdot)$ remains the same. This applies to all further steps. We have proved that for $n \geq k$ the position of the marker has the geometric distribution with respect to $P^n((x,k),\cdot)$. Now, the points in the support of $P^n((x,k),\cdot)$ whose markers fall below coordinate i have x_{i+n} at i, because this position has never been altered (only shifted). This happens with probability

$$\sum_{j=1}^{i-1} 2^{-j} = 1 - 2^{-(i-1)},$$

contributing a factor $(1-2^{-(i-1)})x_{i+n}$ to $T^n f_i(x,k)$. Otherwise, the value at position i is 0 or 1 with equal chances, contributing the factor $2^{-(i-1)}/2$. We have proved the required formula for $n \geq k$ with α equal to $2^{-(i-1)}$.

From what we have derived it is seen that the images of f_i behave almost as the functions f_{i+n}, except that instead of oscillating with amplitude 1 they oscillate with (non-constant) amplitude not smaller than $1-2^{-(i-1)}$. For large i, however, such functions generate entropy arbitrarily close to $\log 2$.

4. *Topological entropy of a Markov operator*

Let $C_I(X)$ be the set of all continuous functions $f : X \to [0,1]$. Throughout this section X is a compact Hausdorff space and T denotes a Markov operator acting on $C(X)$. Our first definition of topological entropy for T uses the notion of measure-theoretic entropy of a stochastic operator introduced in the preceding section. The covers $\mathcal{U}^\varepsilon_\mathcal{F}$ are obtained by 'thickening' the sets in $\mathcal{A}_\mathcal{F}$. The second definition uses continuity of functions in $\mathcal{F}$ to transport open covers from the unit interval to X. In the third one we make use of a certain pseudometric on X induced by a finite collection of functions. This leads us to a definition similar to Bowen's definition of entropy.

For a continuous function f let us define

$$U^\varepsilon_{<f} = \{(x,t) \in X \times [0,1] : t < f(x) + \varepsilon\},$$
$$U^\varepsilon_{>f} = \{(x,t) \in X \times [0,1] : t > f(x) - \varepsilon\},$$
$$\mathcal{U}^\varepsilon_f = \{U^\varepsilon_{<f}, U^\varepsilon_{>f}\}.$$

Given a finite collection $\mathcal{F} \subset C_I(X)$ we obtain a finite open cover of $X \times [0,1]$ by the formula

$$\mathcal{U}^\varepsilon_\mathcal{F} = \bigvee_{f \in \mathcal{F}} \mathcal{U}^\varepsilon_f.$$

If $\mathcal{V}$ is a finite open cover of the unit interval then we let

$$\mathcal{F}^{-1}(\mathcal{V}) = \bigvee_{f \in \mathcal{F}} f^{-1}(\mathcal{V}).$$

We leave the easy proof of the following lemma to the reader.

LEMMA 4.1. *Let $\mathcal{F}$, $\mathcal{G}$ be finite subsets of $C_I(X)$, $\mathcal{V}$ a finite open cover of the unit interval and ε a positive number. Then:*

(i) $\mathcal{U}^{\varepsilon}_{\mathcal{F}\cup\mathcal{G}} = \mathcal{U}^{\varepsilon}_{\mathcal{F}} \vee \mathcal{U}^{\varepsilon}_{\mathcal{G}}$;
(ii) $(\mathcal{F}\cup\mathcal{G})^{-1}(\mathcal{V}) = \mathcal{F}^{-1}(\mathcal{V}) \vee \mathcal{G}^{-1}(\mathcal{V})$;
(iii) *if* $Tf = f \circ S$, *where* $S : X \to X$ *is a continuous transformation, then*

$$\mathcal{U}^{\varepsilon}_{\mathcal{F}^n} = \bigvee_{i=0}^{n-1} (S \times \mathrm{Id})^{-i}\left(\mathcal{U}^{\varepsilon}_{\mathcal{F}}\right)$$

and

$$(\mathcal{F}^n)^{-1}(\mathcal{V}) = \bigvee_{i=0}^{n-1} S^{-i}(\mathcal{F}^{-1}(\mathcal{V})).$$

Recall that for any open cover $\mathcal{U}$ the symbol $N(\mathcal{U})$ denotes the minimal cardinality of a subcover chosen from $\mathcal{U}$.

Definition. Let $\mathcal{F} \subset C_I(X)$ be a finite collection of functions and let ε be a positive number. We define:
(i) $H_1(\mathcal{F}, \varepsilon) = \log N(\mathcal{U}^{\varepsilon}_{\mathcal{F}})$;
(ii) $h_1(T, \mathcal{F}, \varepsilon) = \limsup_{n\to\infty} (1/n) H_1(\mathcal{F}^n, \varepsilon)$;
(iii) $h_1(T) = \sup_{\mathcal{F}} \sup_{\varepsilon} h_1(T, \mathcal{F}, \varepsilon)$.

Definition. Let $\mathcal{V}$ be a cover of [0, 1].
(i) $H_2(\mathcal{F}, \mathcal{V}) = \log N(\mathcal{F}^{-1}(\mathcal{V}))$;
(ii) $h_2(T, \mathcal{F}, \mathcal{V}) = \limsup_{n\to\infty} (1/n) H_2(\mathcal{F}^n, \mathcal{V})$;
(iii) $h_2(T) = \sup_{\mathcal{F}} \sup_{\mathcal{V}} h_2(T, \mathcal{F}, \mathcal{V})$.

Given $\mathcal{F}$ we define a pseudometric on X by

$$d_{\mathcal{F}}(x, y) = \sup_{f\in\mathcal{F}} |f(x) - f(y)|.$$

We say that a subset of X is $(d_{\mathcal{F}}, \varepsilon)$-separated if it is ε-separated in the pseudometric $d_{\mathcal{F}}$. Since the space X is compact, there exists a finite $(d_{\mathcal{F}}, \varepsilon)$-separated subset of maximal cardinality in X. We denote the number of elements of this subset by $s(d_{\mathcal{F}}, \varepsilon)$.

Definition. For $\mathcal{F}$ and ε define:
(i) $H_3(\mathcal{F}, \varepsilon) = \log s(d_{\mathcal{F}}, \varepsilon)$;
(ii) $h_3(T, \mathcal{F}, \varepsilon) = \limsup_{n\to\infty} (1/n) H_3(\mathcal{F}^n, \varepsilon)$;
(iii) $h_3(T) = \sup_{\mathcal{F}} \sup_{\varepsilon} h_3(T, \mathcal{F}, \varepsilon)$.

THEOREM 4.2. *For every Markov operator* T *we have*

$$h_1(T) = h_2(T) = h_3(T).$$

Proof. We begin by showing that $h_1(T) \le h_2(T)$. Choose $\varepsilon > 0$ and let $\mathcal{V}$ be a finite open cover of the unit interval consisting of sets having diameters not greater than ε. We claim that the cover defined by the formula

$$\mathcal{W}_n = \{U \times V : U \in (\mathcal{F}^n)^{-1}(\mathcal{V}), V \in \mathcal{V}\}$$

is inscribed in $\mathcal{U}^{\varepsilon}_{\mathcal{F}^n}$. Indeed, for each $U \times V \in \mathcal{W}_n$ we let

$$\mathcal{F}' = \{f \in \mathcal{F}^n : (\forall_{x\in U}) f(x) \ge \inf V\}.$$

https://doi.org/10.1017/S014338570400032X Published online by Cambridge University Press

It is not hard to verify that

$$U \times V \subset \bigcap_{f \in \mathcal{F}'} U^{\varepsilon}_{<f} \cap \bigcap_{f \in \mathcal{F}^n \setminus \mathcal{F}'} U^{\varepsilon}_{>f} \in \mathcal{U}^{\varepsilon}_{\mathcal{F}^n}.$$

Thus,

$$N(\mathcal{U}^{\varepsilon}_{\mathcal{F}^n}) \le N(\mathcal{W}_n) \le N((\mathcal{F}^n)^{-1}(\mathcal{V})) \cdot N(\mathcal{V})$$

and, since $N(\mathcal{V})$ is independent of n,

$$h_1(T, \mathcal{F}, \varepsilon) \le h_2(T, \mathcal{F}, \mathcal{V}).$$

The desired inequality follows by taking appropriate suprema.

Now, we prove that $h_2(T) \le h_3(T)$. Let $\mathcal{V}$ be a finite open cover of the unit interval. Denote its Lebesgue number by δ and let E be a maximal $(d_{\mathcal{F}^n}, \delta/2)$-separated set in X. It follows from the maximality of E that the collection $\{B^{\delta/2}(x) : x \in E\}$ of balls in the pseudometric $d_{\mathcal{F}^n}$ constitutes a finite open cover of X. For every $f \in \mathcal{F}^n$ and $x \in E$ the interval $(f(x) - \delta/2, f(x) + \delta/2)$ is contained in some element $V_f(x)$ of $\mathcal{V}$. Hence,

$$\begin{aligned} B^{\delta/2}(x) &= \bigcap_{f \in \mathcal{F}^n} f^{-1}\left(f(x) - \frac{\delta}{2}, f(x) + \frac{\delta}{2}\right) \\ &\subset \bigcap_{f \in \mathcal{F}^n} f^{-1}(V_f(x)) \in (\mathcal{F}^n)^{-1}(\mathcal{V}) \end{aligned}$$

and

$$N((\mathcal{F}^n)^{-1}(\mathcal{V})) \le \operatorname{card}\{B^{\delta/2} : x \in E\} = s(d_{\mathcal{F}^n}, \delta/2),$$

which implies $h_2(T) \le h_3(T)$.

We end the proof by showing that $h_3(T) \le h_1(T)$. Let $D \subset X$ be a $(d_{\mathcal{F}}, \varepsilon)$-separated set of maximal cardinality. Put $\gamma = \varepsilon/6$ and define

$$\widetilde{\mathcal{F}} = \left\{\frac{1}{2} f + i\gamma : f \in \mathcal{F}, i \in \mathbb{Z}, 0 \le i \le \frac{1}{2\gamma}\right\}.$$

We will show that the cover $\mathcal{U}^{\gamma}_{\widetilde{\mathcal{F}}}$ separates points of $D \times \{1/2\}$ in the sense that each element of the cover contains at most one point from $D \times \{1/2\}$. Consider two elements x, y of D. We can choose a function $f \in \mathcal{F}$ satisfying $|f(x) - f(y)| \ge \varepsilon$ and, since both x and y play the same role in the formula, we may assume that $f(x) + \varepsilon \le f(y)$. There exists an integer i such that $\tilde{f} = f/2 + i\gamma$ belongs to $\widetilde{\mathcal{F}}$ and

$$\gamma \le \tfrac{1}{2} - \tilde{f}(x) \le 2\gamma.$$

Then

$$\begin{aligned} \tilde{f}(y) - \frac{1}{2} &= \frac{1}{2} f(y) + i\gamma - \frac{1}{2} \\ &\ge \frac{1}{2} f(x) + \frac{\varepsilon}{2} + i\gamma - \frac{1}{2} \\ &= \tilde{f}(x) - \frac{1}{2} + \frac{\varepsilon}{2} \ge \gamma, \end{aligned}$$

https://doi.org/10.1017/S014338570400032X Published online by Cambridge University Press

which implies that $(y, 1/2)$ belongs to $U^\gamma_{<\tilde f}$ and not to $U^\gamma_{>\tilde f}$, while, on the contrary, $(x, 1/2)$ belongs to $U^\gamma_{>\tilde f}$ and not to $U^\gamma_{<\tilde f}$. Since every element of $\mathcal{U}^\gamma_{\widetilde{\mathcal{F}}}$ is contained either in $U^\gamma_{<\tilde f}$ or in $U^\gamma_{>\tilde f}$, this proves that the cover separates points of $D \times \{1/2\}$. Moreover, every subcover of $\mathcal{U}^\gamma_{\widetilde{\mathcal{F}}}$ has the same property, so that

$$s(d_{\mathcal{F}}, \varepsilon) \le N(\mathcal{U}^\gamma_{\widetilde{\mathcal{F}}}).$$

Recall that T, as a Markov operator, is linear and preserves constants. This implies that $\widetilde{\mathcal{F}^n} = (\widetilde{\mathcal{F}})^n$, so we can replace $\mathcal{F}$ with $\mathcal{F}^n$, obtaining

$$s(d_{\mathcal{F}^n}, \varepsilon) \le N(\mathcal{U}^\gamma_{(\widetilde{\mathcal{F}})^n}).$$

The proof is ended by taking upper limits and suprema. □

In the following we will use the symbol h_{top} to denote the common value of h_1, h_2 and h_3. According to the next result, the coincidence with the classical notation is reasonable.

THEOREM 4.3. *If $Tf = f \circ S$ is an operator generated by a continuous map $S : X \to X$, then $h_{\mathrm{top}}(T)$ is equal to the classical topological entropy of S.*

Proof. It is an obvious corollary from Lemma 4.1(iii) that $h_2(T)$ is less than or equal to the classical topological entropy of S.

For the opposite inequality we will exploit the definition of h_1. Let $\mathcal{W}$ be a finite open cover of X. Choose a minimal subcover $\{W_1, \dots, W_r\} \subset \mathcal{W}$. Using the Tietze–Urysohn theorem (see [**E**, Theorem 2.1.8]) one can construct a partition of unity $\mathcal{F} = \{f_1, \dots, f_r\}$ consisting of continuous functions on X and such that if $x \in {W_i}^c$ then $f_i(x) = 0$ $(i = 1, \dots, r)$. Consequently, if $x \in \bigcap_{j \ne i} ({W_j}^c)$, i.e. x is covered exclusively by W_i, then $f_i(x) = 1$. Fix $0 < \varepsilon < 1/2r$. Every member of $\mathcal{U}^\varepsilon_{\mathcal{F}^n}$ is of the form

$$\bigcap_{g \in \mathcal{F}^n} U_g = \bigcap_{k=0}^{n-1} \bigcap_{g \in T^k \mathcal{F}} U_g,$$

where $U_g \in \mathcal{U}^\varepsilon_g$. We will prove that each subcover $\mathcal{U}'$ chosen from $\mathcal{U}^\varepsilon_{\mathcal{F}^n}$ determines a subcover of $\mathcal{W}^n$ of the same or smaller cardinality. Suppose that an element of $\mathcal{U}'$ satisfies the following condition:

$$(\forall_{k<n})(\exists_{g_k \in T^k\mathcal{F}})\, U_{g_k} = U^\varepsilon_{<g_k}. \tag{4.0.1}$$

Denote by $\widetilde{W}_k$ the element of $T^{-k}\mathcal{W}$ such that g_k equals 1 on the set covered exclusively by $\widetilde{W}_k$ and vanishes on $\widetilde{W}^c_k$. The set $\bigcap_{k=0}^{n-1} \widetilde{W}_k$ belongs to $\mathcal{W}^n$. Pick any $x \in X$. The point $(x, 1/2r)$ does not belong to $\bigcap_{g \in T^k\mathcal{F}} U^\varepsilon_{>g}$ for any $k = 0, \dots, n-1$, since one can always find $g_k \in T^k\mathcal{F}$ such that $g_k(x)$ is greater than or equal to $1/r$. Thus, if an element U of $\mathcal{U}'$ contains $(x, 1/2r)$, it satisfies the condition (4.0.1) and determines some set $\bigcap_{k=0}^{n-1} \widetilde{W}_k \in \mathcal{W}^n$. This set contains x, because $g_k(x)$ is positive (at least $1/r$) for $k = 0, \dots, n-1$. Since x was arbitrary, sets of the form $\bigcap_{k=0}^{n-1} \widetilde{W}_k$ constitute a cover of X, and hence

$$N(\mathcal{W}^n) \le N(\mathcal{U}^\varepsilon_{\mathcal{F}^n}).$$ □

The following result is analogous to Proposition 3.9.

https://doi.org/10.1017/S014338570400032X Published online by Cambridge University Press

PROPOSITION 4.4. *If for every continuous f there exists an invariant function ϕ_f such that*

$$\lim_{n\to\infty} \sup_{x\in X} |T^n f(x) - \phi_f(x)| = 0$$

then $h_{\mathrm{top}}(T) = 0$.

Proof. For a given collection $\mathcal{F}$ of continuous functions let

$$\Phi(\mathcal{F}) = \{\phi_f : f \in \mathcal{F}\}.$$

Notice that for every ε and every finite collection $\mathcal{F}$ of continuous functions there exists N such that for every $f \in \mathcal{F}$

$$\sup_{x\in X} |T^n f(x) - \phi_f(x)| < \frac{\varepsilon}{3} \quad \text{for all } n \geq N.$$

Then, for every $x, y \in X$, we have

$$\begin{aligned} |T^n f(x) - T^n f(y)| &\leq |T^n f(x) - \phi_f(x)| + |\phi_f(x) - \phi_f(y)| \\ &\quad + |\phi_f(y) - T^n f(y)| \\ &< \frac{2\varepsilon}{3} + |\phi_f(x) - \phi_f(y)|, \end{aligned}$$

implying that

$$s(d_{\mathcal{F}^n}, \varepsilon) \leq s(d_{\mathcal{F}^N \cup \Phi(\mathcal{F})}, \varepsilon/3) \quad \text{for all } n \geq N,$$

where the right-hand side does not depend on n. □

We can define a topological factor of a Markov operator in the same way as was done in the measure-theoretic Definition 3.2, replacing only the term 'measure-preserving' by 'continuous'. Two operators are isomorphic if the relevant factor map is a homeomorphism. We will say that a compact set Y is *invariant* under T if $P_T(x, Y^c) = 0$ for every $x \in Y$ (P_T is a transition probability corresponding to T). Using the formula (1.0.1) we can define a Markov operator on $C(Y)$, which may be treated as a restriction of T (since Y is closed, every $f \in C(Y)$ extends to a continuous function on X). The proofs of the following statements concerning topological entropy are standard and will be omitted.

PROPOSITION 4.5.

(i) *The entropy of a factor of a Markov operator T is smaller than or equal to the entropy T.*

(ii) *Two isomorphic Markov operators have equal entropies.*

(iii) *If Y is a compact invariant subset of X then the entropy of the Markov operator T restricted to $C(Y)$ is smaller than or equal to the entropy of T on $C(X)$.*

(iv) *For every $k \in \mathbb{N}$, $h_{\mathrm{top}}(T^k) = k h_{\mathrm{top}}(T)$.*

We now discuss two quite different attempts to define topological entropy for Markov operators using directly the topological entropy of an associated pointwise generated action. There are at least two such natural actions of continuous transformations: the shift on the space of trajectories, and the adjoint operator on the space $\mathcal{M}(X)$ of all Radon probability measures. The examples below show that the entropy of neither one suits our expectations.

https://doi.org/10.1017/S014338570400032X Published online by Cambridge University Press

Example. Let X be a finite space consisting of n elements. Denote by p the probability measure uniformly distributed on X and let T be given by the formula

$$Tf(x) = \int_X f\,dp = \frac{1}{n}\sum_{y\in X} f(y),$$

i.e. T sends each function to a constant. Clearly, we expect the entropy of such a trivial (when observed on functions) action to be zero. Indeed, our Proposition 4.4 yields $h_{\text{top}}(T) = 0$. However, the corresponding action on trajectories is the full shift over the alphabet consisting of n symbols, and its topological entropy equals $\log n$.

Example. Consider the one-sided full shift (X, T) over the alphabet $\{0, 1\}$. Let $(a_n)_{n\in\mathbb{N}}$ be any sequence chosen from the unit interval. For every natural n define a measure $P_{a_n} = \{a_n, 1-a_n\}$ on $\{0, 1\}$ and let $\mu_{(a_n)} = \prod P_{a_n}$ be the product measure on X. It is easy to verify that

$$T^*\mu_{(a_n)} = \mu_{(a_{n+1})}.$$

Since $(a_n)_{n\in\mathbb{N}}$ was an arbitrary sequence from the unit interval, the system $(\mathcal{M}(X), T^*)$ contains a subsystem isomorphic to the full shift over the infinite alphabet $[0, 1]$ having infinite entropy. Thus, the topological entropy of T^* does not satisfy the main demand concerning equality with classical entropy in the pointwise case.

In fact, from the results obtained by E. Glasner and B. Weiss in [**GW**] it follows that if T is an ergodic map of positive entropy then the entropy of T^* is infinite (earlier the same authors proved that zero entropy is preserved). A natural question arises if the entropy of an arbitrary Markov operator T is equal to the entropy of the adjoint map T^* restricted to the smallest compact invariant subset containing all Dirac measures.

5. *Relation between measure-theoretic and topological entropy*

THEOREM 5.1. *Let X be a compact Hausdorff space and T a Markov operator acting on $C(X)$. For every invariant Radon probability measure μ on X*

$$h_\mu(T) \le h_{\text{top}}(T).$$

Proof. We will refer to h_μ as defined in §3. By (3.1.3), for calculations of measure-theoretic entropy it suffices to consider families $\mathcal{F}$ of continuous functions. Let $\mathcal{F}$ be a subset of $C_I(X)$ of cardinality r. Choose a positive number ε, such that $2r\varepsilon\log(2r\varepsilon) < 1/2^r$. For every $A \in \mathcal{A}_\mathcal{F}$, we denote by $\mathcal{F}_{\ge A}$ the set of all functions in $\mathcal{F}$, for which $f(x) \ge t$ whenever $(x, t) \in A$. Analogously, $\mathcal{F}_{<A}$ is the set of all functions from $\mathcal{F}$, such that $f(x) < t$ if $(x, t) \in A$. It is easy to see that the union of $\mathcal{F}_{<A}$ and $\mathcal{F}_{\ge A}$ is the whole of $\mathcal{F}$. We define a new partition $\mathcal{B}^\varepsilon_\mathcal{F}$ of $X \times [0, 1]$ consisting of compact sets:

$$\begin{aligned} B_A = B^\varepsilon_{A,\mathcal{F}} = \{(x, t) : (\forall_{f\in\mathcal{F}_{\ge A}}) t \le f(x) - \varepsilon\} \\ \cap \{(x, t) : (\forall_{f\in\mathcal{F}_{<A}})\, t \ge f(x) + \varepsilon\}, \end{aligned}$$

where A belongs to $\mathcal{A}_\mathcal{F}$, and the open set

$$\widetilde{B} = \widetilde{B}^\varepsilon_\mathcal{F} = \bigcap_{A\in\mathcal{A}_\mathcal{F}} {B_A}^c = \{(x, t) : (\exists_{f\in\mathcal{F}})|f(x) - t| < \varepsilon\}.$$

https://doi.org/10.1017/S014338570400032X Published online by Cambridge University Press

Notice that $A \cap B_{A'} = B_A$ if and only if $A = A'$. Otherwise, the intersection $A \cap B_{A'}$ is empty. Thus,

$$\begin{aligned} H_{\widehat{\mu}}(\mathcal{A}_{\mathcal{F}}|\mathcal{B}^{\varepsilon}_{\mathcal{F}}) &= H_{\widehat{\mu}}(\mathcal{A}_{\mathcal{F}} \vee \mathcal{B}^{\varepsilon}_{\mathcal{F}}) - H_{\widehat{\mu}}(\mathcal{B}^{\varepsilon}_{\mathcal{F}}) \\ &= -\sum_{A\in\mathcal{A}_{\mathcal{F}}} \widehat{\mu}(A \cap \widetilde{B}) \log(\widehat{\mu}(A \cap \widetilde{B})) \end{aligned}$$

(recall that $\widehat{\mu}$ is the product of μ with the Lebesgue measure). Since $\widehat{\mu}(\widetilde{B}) < 2r\varepsilon$ and $\mathcal{A}_{\mathcal{F}}$ has at most 2^r elements, we get

$$H_{\widehat{\mu}}(\mathcal{A}_{\mathcal{F}}|\mathcal{B}^{\varepsilon}_{\mathcal{F}}) < 1. \tag{5.0.1}$$

Similarly, for every natural number k

$$H_{\widehat{\mu}}(\mathcal{A}_{T^k\mathcal{F}}|\mathcal{B}^{\varepsilon}_{T^k\mathcal{F}}) < 1, \tag{5.0.2}$$

because ε was chosen according only to the cardinality of $\mathcal{F}$. We will abbreviate $\bigvee_{k=0}^{n-1} \mathcal{B}^{\varepsilon}_{T^k\mathcal{F}}$ by $\mathcal{B}^n$ (note that this is not equal to $\mathcal{B}^{\varepsilon}_{\mathcal{F}^n}$). Using the estimates (5.0.1) and (5.0.2) we derive

$$H_{\widehat{\mu}}(\mathcal{A}_{\mathcal{F}^n}|\mathcal{B}^n) \le \sum_{k=0}^{n-1} H_{\widehat{\mu}}(\mathcal{A}_{T^k\mathcal{F}}|\mathcal{B}^{\varepsilon}_{T^k\mathcal{F}}) < n.$$

Hence,

$$H_{\widehat{\mu}}(\mathcal{A}_{\mathcal{F}^n}) < H_{\hat{\mu}}(\mathcal{B}^n) + n.$$

Since the entropy of a partition is always less than or equal to the logarithm of its cardinality we need to estimate the number of sets in $\mathcal{B}^n$.

Let $\mathcal{U}'$ be a minimal subcover chosen from $\mathcal{U}^{\varepsilon}_{\mathcal{F}^n}$. Obviously,

$$\operatorname{card} \mathcal{B}^n \le \sum_{U\in\mathcal{U}'} \operatorname{card}\{B \in \mathcal{B}^n : B \cap U \ne \emptyset\}.$$

Every $U \in \mathcal{U}^{\varepsilon}_{\mathcal{F}^n}$ has the form $\bigcap_{k=0}^{n-1} U_k$, where $U_k \in \mathcal{U}^{\varepsilon}_{T^k\mathcal{F}}$. For each U_k there exists $A \in \mathcal{A}_{T^k\mathcal{F}}$ such that U_k is contained in the union of $B^{\varepsilon}_{A,T^k\mathcal{F}}$ and $\widetilde{B}^{\varepsilon}_{T^k\mathcal{F}}$. This implies that

$$\operatorname{card} \left\{B \in \mathcal{B}^n : B \cap U \ne \emptyset\right\} \le 2^n,$$

and hence

$$\operatorname{card} \mathcal{B}^n \le 2^n \cdot N(\mathcal{U}^{\varepsilon}_{\mathcal{F}^n}).$$

Thus, finally, we have

$$H_{\widehat{\mu}}(\mathcal{A}_{\mathcal{F}^n}) < \log N(\mathcal{U}^{\varepsilon}_{\mathcal{F}^n}) + n\log 2 + n,$$

implying

$$h_\mu(T) \le h_{\text{top}}(T) + \log 2 + 1.$$

Repeating the argument for T^k, replacing T and using Propositions 3.8 and 4.5(iv) we complete the proof by writing

$$h_\mu(T) = \frac{1}{k} h_\mu(T^k) \le \frac{1}{k}(h_{\text{top}}(T^k) + \log 2 + 1) \le h_{\text{top}}(T) + \frac{\log 2 + 1}{k}$$

for arbitrary k. □

https://doi.org/10.1017/S014338570400032X Published online by Cambridge University Press

Remark 5.2. The problems with proving the converse inequality arise from the lack of full subadditivity of the sequence $H_\mu(\mathcal{F}^n)$. In the classical proofs the invariant measure μ achieving large entropy is usually constructed as an accumulation point of measures μ_n attaining large values for $H_{\mu_n}(\mathcal{F}^n)$ for increasing n. Subadditivity guarantees that measures with large indices $m > n$ maintain large values of $H_{\mu_m}(\mathcal{F}^n)$. Then, continuity (or upper semicontinuity) of $H_\mu(\mathcal{F}^n)$ is used to yield large values for the entire sequence at μ. In our situation the index N, above which the sequence $H_{\mu_n}(\mathcal{F}^k)$ becomes 'nearly subadditive', depends on μ_n and may be much larger than n. The alternative functions $H'_\mu(\mathcal{F}^n)$ and $H''_\mu(\mathcal{F}^n)$ are subadditive in n, but, in turn, they need not be upper semicontinuous on measures.

References

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

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

[E] R. Engelking. *General Topology*. PWN, Warsaw, 1977.

[FN] B. Sz.-Nagy and C. Foiaş. *Harmonic Analysis of Operators on Hilbert Space*. North-Holland, Budapest, 1970.

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

[GW] E. Glasner and B. Weiss. Quasifactors of ergodic systems with positive entropy. *Israel J. Math.* **134** (2003), 363–380.

[I] A. Iwanik. Integral representations of stochastic kernels. *Aspects of Positivity in Functional Analysis (Proc. Conf. Occas. H. H. Schaefer's Birthday, Tübingen 1985) (North-Holland Mathematical Studies, 122)*. Elsevier, 1986, pp. 223–230 (in English).

[KSL] B. Kamiński and J. de Sam Lazaro. A note on the entropy of a doubly stochastic operator. *Colloq. Math.* **84/85**(1) (2000), 245–254.

[LW] R. Langevin and P. G. Walczak. Entropy, transverse entropy and partitions of unity. *Ergod. Th. & Dynam. Sys.* **14** (1994), 551–563.

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

[W] P. Walters. *An Introduction to Ergodic Theory*. Springer, Berlin 1982.

[Z] W. Żelazko. *Algebry Banacha* (in Polish), PWN, Warsaw, 1968.