EXPLODING MARKOV OPERATORS

BARTOSZ FREJ

Abstract. A special class of doubly stochastic (Markov) operators is constructed. These operators come from measure preserving transformations and inherit some of their properties, namely ergodicity and positivity of entropy, yet they may have no pointwise factors.

1. Introduction

The subject of the current paper lies in the border zone between ergodic theory and operator theory. The main motivation of study was the desire to increase the number of examples of doubly stochastic operators, which escape the scope of classical ergodic theory (because they are not induced by measure preserving maps as their Koopman operators), but they still reveal a nontrivial dynamical behavior. By a doubly stochastic or a Markov operator we understand an operator $P : L^p(\mu) \to L^p(\nu)$, where $(X, \mu)$ and $(Y, \nu)$ are probability spaces, which fulfills the following conditions:

(i) $Pf$ is positive for every positive $f \in L^p(\mu)$,
(ii) $P1 = 1$ (where $1$ is the function constantly equal to 1),
(iii) $\int Pf \, d\nu = \int f \, d\mu$ for every $f \in L^p(\mu)$.

For example, the well-studied class of quasi-compact doubly stochastic operators on $L^2$ lies pretty far from the theory of measure preserving maps. But the domain of a quasi-compact operator decomposes into the direct sum of two reducing subspaces, called reversible and almost weakly stable parts, respectively, such that the first one is finite dimensional, while on the other one orbits of functions converge to zero in $L^2$ norm. The restriction of such an operator to the reversible part is Markov isomorphic to a rotation of a compact abelian group (which is finite in this case). The transition probability associated to the operator forces points of the underlying space to ramble periodically through finitely many sets of states (in a fixed order), randomly choosing the succeeding state from a set which is next in the queue. These operators are null, meaning that their sequence entropy is always zero (see [8] for details). As another example one may think of a convex combination of finitely many measure preserving maps, which leads to studying a rich class of iterated function systems. Unfortunately, such operators are hard to handle by the entropy theory as defined in [3]—e.g., it is possible that the combination of maps with positive entropy has entropy equal to zero. In the current paper another class of

2010 Mathematics Subject Classification. 28D05, 28D20, 37A30, 47A35.

Key words and phrases. Markov operator and doubly stochastic operator and measure-preserving transformation and entropy and ergodicity and disintegration of measure.

Research is supported from resources for science in years 2013-2018 as research project (NCN grant 2013/08/A/ST1/00275, Poland).
examples which stem from pointwise maps is proposed and some of their properties are investigated.

2. The definition

Let \((X, \Sigma, \mu)\) be a standard probability space and let \(T : X \to X\) be a measure preserving surjection. Let \((a_n)\) be a strictly decreasing sequence, such that \(\sum_{n=1}^{\infty} a_n = 1\). Define a probability distribution \(m = m(a_n)\) on \(\mathbb{N}\) by \(m(\{k\}) = a_k\).

Now let \(Y\) be the Cartesian product \(X \times \mathbb{N}\) with the product \(\sigma\)-algebra and let \(\nu = \mu \times m\). For a fixed \(k \in \mathbb{N}\) let \(\xi_k\) be a partition of \(X\) into sets \(T^{-k}\{T^kx\}\), \(x \in X\), and denote by \(\xi_k(x)\) an element of the partition \(\xi_k\) which contains \(x\). For every \(k\) this partition is measurable (see Appendix 1 in [1] for the precise definition), in other words, the quotient space \(X/\xi_k\) is countably separated, so \(X/\xi_k\) is also a standard probability space with measure transported by the map \(x \mapsto \xi_k(x)\). Indeed, let \(\{B_1, B_2, \ldots\}\) be a separating collection in \(X\). If \(\xi_k(x)\) and \(\xi_k(y)\) are disjoint then \(T^{-k}x \neq T^{-k}y\). Without loss of generality one may assume that \(T^{-k}x \in B_i\) and \(T^{-k}y \notin B_i\) for some \(i \in \mathbb{N}\). Then, \(\xi_k(z) \in T^{-k}B_i\) if \(T^{-k}z \in B_i\) and \(\xi_k(z) \notin T^{-k}B_i\) otherwise, so the collection \(\{T^{-k}B_i : i \in \mathbb{N}\}\) separates elements of \(\xi_k\). Since \(T\) is onto, every point in \(B_i\) can be represented as \(T^k z\) for some \(z \in X\), so each \(T^{-k}B_i = \bigcup_{z \in T^{-k}B_i} \xi_k(z)\) and the collection \(\{T^{-k}B_1, T^{-k}B_2, \ldots\}\) separates elements of the partition \(\xi_k\).

Let \(\mu_{\mathcal{C}} : C \in \xi_k\) be the disintegration of \(\mu\) over \(X/\xi_k\), that is, there is a map \(C \mapsto \mu_{\mathcal{C}}\) defined on \(X/\xi_k\) with range in the space of all probability measures on \(X\), such that each measure \(\mu_{\mathcal{C}}\) satisfies \(\mu_{\mathcal{C}}(X \setminus C) = 0\) and there is a measure \(\hat{\mu}_k\) on \(X/\xi_k\) with the property that for any measurable function \(f \in L^1(\mu)\),

\[
\int f \, d\mu = \int_{X/\xi_k} \left( \int_C f(x) d\mu_{\mathcal{C}}(x) \right) d\hat{\mu}_k(C),
\]

(see [3]). In addition, the map \(C \mapsto \mu_{\mathcal{C}}\) is measurable, when the space of probability measures is endowed with the Borel \(\sigma\)-algebra for the weak* topology in the space of probability measures on \(X\). An operator \(E_k : L^1(X, \mu) \to L^1(X, \mu)\) given by the formula \(E_k f(x) = f \circ T d\mu_{\xi_k(x)}\) is doubly stochastic. Indeed, it is clear that it is positive and preserves constant functions. Moreover, the function \(x \mapsto \mu_{\xi_k(x)}\) is constant on atoms of \(\xi_k\), so \(E_k f(x) = \int_C f \circ T \, d\mu_{\mathcal{C}}\) for \(x \in C\). Therefore, for every \(f \in L^1(X, \mu)\) it holds that

\[
\int E_k f \, d\mu = \int_{X/\xi_k} \left( \int_C E_k f \, d\mu_{\mathcal{C}} \right) d\hat{\mu}_k(C) = \int_{X/\xi_k} \left( \int_C f \circ T \, d\mu_{\mathcal{C}} \right) d\hat{\mu}_k(C) = \int f \circ T \, d\mu = \int f \, d\mu.
\]

Define a sequence \((b_k)\) of positive numbers by

\[
b_k = \frac{a_k - a_{k+1}}{a_1}.
\]

As the simplest example one may consider \(a_k = b_k = \frac{1}{2^k}\) or, more generally, \(a_k = b_k = (1 - a) a^{k-1}\) \((0 < a < 1)\), but other choices are also possible (though if \(a_k = b_k\) then one automatically obtains a geometric sequence). Let \(\delta_y\) be the Dirac measure concentrated in \(y\), that is, \(\delta_y(A) = \mathbb{1}_A(y)\), and let sections of a set \(A \subset Y\) and of a
function $f$ on $Y$ be denoted by
\[ A|_k = \{ x \in X : (x, k) \in A \} \quad \text{and} \quad f|_k(x) = f(x, k). \]

**Definition 2.1.** An exploding operator induced by $T$ is a Markov integral operator on $L^1(Y, \nu)$ defined by
\[
\Upsilon_T f(y) = \int f(u)P_T(y, du),
\]
where the probability kernel $P_T$ is given by:
\[
P_T((x, 1), A) = \sum_{k=1}^{\infty} b_k \mu_{\xi_k}(x)(T^{-1}A|_k) \quad \text{for } k \geq 2
\]
\[
P_T((x, k), A) = \delta_{(Tx, k-1)}(A) \quad \text{for } k \geq 2
\]
In other words,
\[
\Upsilon_T f(x, 1) = \sum_{k=1}^{\infty} b_k \int f|_k \circ T \, d\mu_{\xi_k}(x) = \sum_{k=1}^{\infty} b_k E_k f|_k(x)
\]
\[
\Upsilon_T f(x, k) = f(Tx, k-1) \quad \text{for } k \geq 2
\]

One may visualize the action of $\Upsilon_T$ via the transition probability $P_T$ in the following way. Each point of $Y$ is a pair consisting of some $x \in X$ and a positive integer $k$. The integer coordinate represents the indication of a clock, which counts down time to an explosion. As long as this indication is greater than 1, the point is mapped according to the action of the pointwise transformation $T$ and the counter goes down by one. When the counter is to be reduced from 1 to 0, the point $x$ explodes and its images are spread over the space (more precisely, over the set of points which would share the common future with $x$, if one considered the evolution by $T$) with counters reset to $k$ with probability $b_k$.

This class of operators is a generalization of the following example described in [3]. It has positive entropy, yet it is strictly non-pointwise, meaning that the only pointwise factor of it is the trivial one (see section 3 for definitions).

**Example 2.2.** Let $(X, \Sigma, \mu)$ consist of the set $X = \{0, 1\}^N$ of one-sided 0-1 sequences with the product $\sigma$-algebra and with the uniform product measure $\mu = (\frac{1}{2}, \frac{1}{2})^N$. Let $m$ be the geometric distribution on natural numbers $\mathbb{N}$ given by $m(\{k\}) = 2^{-k}$.

For each finite block $B = (B_1, B_2, \ldots, B_k) \in \{0, 1\}^k$ let us define the map $\sigma_B : X \to X$ by
\[
(\sigma_B x)_n = \begin{cases} 
B_n & \text{for } n \leq k \\
x_{n+1} & \text{for } n > k
\end{cases}
\]
and define the operator $P$ on $L^1(\mu \times \nu)$ as follows:
\[
P f(x, k) = f(\sigma x, k-1) \quad \text{if } k > 1,
\]
\[
P f(x, 1) = \sum_{k=1}^{\infty} 2^{-k} \sum_{B \in \{0, 1\}^k} 2^{-k} f(\sigma_B x, k).
\]

This example is an instance of our construction if $T$ is the $(\frac{1}{2}, \frac{1}{2})$-Bernoulli shift. We will restore some of its features in the more general case.
Let us first prove that $\Upsilon_T(f)$ defined by (1) is Markovian. Clearly, $\Upsilon_T f \geq 0$ for $f \geq 0$ and $\Upsilon_T$ preserves constants, so the only thing left to show is the fact that $\Upsilon_T$ preserves measure. This is checked in the following calculation.

$$
\int_Y \Upsilon_T f d\nu = \int_{X \times \{1\}} \Upsilon_T f d\nu + \sum_{k=2}^{\infty} \int_{X \times \{k\}} \Upsilon_T f d\nu
$$

$$
= a_1 \int_X (\Upsilon_T f)_{|1} d\mu + \sum_{k=2}^{\infty} a_k \int_X (\Upsilon_T f)_{|k} d\mu
$$

$$
= \sum_{k=1}^{\infty} a_1 b_k \int_X E_k f_{|k} d\mu + \sum_{k=2}^{\infty} a_k \int_X f_{|k-1} \circ T d\mu
$$

$$
= \sum_{k=1}^{\infty} a_1 b_k \int_X f_{|k} d\mu + \sum_{k=1}^{\infty} a_{k+1} \int_X f_{|k} \circ T d\mu
$$

$$
= \sum_{k=1}^{\infty} \left( a_1 b_k + a_{k+1} \right) \frac{a_k}{a_k} \int_X f_{|k} d\mu = \int_Y f d\nu
$$

Remark 2.3. The above construction is particularly simple for invertible (injective) maps. In this case, $\Upsilon_T$ is closely related to the Koopman operator of $T$, precisely, $\Upsilon_T f(x, 1) = \sum_{k=1}^{\infty} b_k f(Tx, k)$ or, in other words, $P_T$ sends the point $(x, 1)$ to $(Tx, k)$ with probability $b_k$.

3. Pointwise factors

We begin with several definitions which can be found in [5].

**Definition 3.1.** A unital sublattice of $L^1(X, \mu)$ is a closed linear subspace of $L^1(X, \mu)$ which contains the constant 1 and together with each $f$ it contains its conjugate $\bar{f}$ and its absolute value $|f|$.

**Definition 3.2** ([5], Definition 12.9). Let $(X_1, \mu_1)$, $(X_2, \mu_2)$ be a probability space and let $U : L^1(\mu_2) \rightarrow L^1(\mu_1)$ be a Markov operator.

1. $U$ is a Markov embedding if it is a lattice homomorphism (i.e., $|Uf| = |Uf|$ for every $f \in L^1(\mu_2)$) or, equivalently, there is a Markov operator $S$ such that $SU$ is an identity.

2. $U$ is a Markov isomorphism if it is a surjective Markov embedding.

**Definition 3.3** ([5], Definition 13.26). A factor of a doubly stochastic operator $P$ on $L^1(X, \mu)$ is a unital sublattice of $L^1(X, \mu)$, which is invariant under the action of $P$.

This definition identifies factors as certain subspaces of the domain. It can be shown these subspaces have form $L^1(X, \Sigma_F, \mu)$, where $\Sigma_F$ is a sub-$\sigma$-algebra of the $\sigma$-algebra of all measurable subsets of $X$, which is invariant in this sense that $P_{|A}$ is $\Sigma_F$-measurable for each $A \in \Sigma_F$. Moreover, the representation is unique if one assumes that $\Sigma_F$ is complete with respect to the measure $\mu$. Motivated by the theory of classical dynamical systems one may say that if $P_1$ is a doubly stochastic operator on $L^1(\mu_1)$ and $P_2$ is a doubly stochastic operator on $L^1(\mu_2)$ then $P_2$ is a factor of $P_1$ if there is a Markov embedding $U : L^1(\mu_2) \rightarrow L^1(\mu_1)$ such
that $UP_2 = P_1U$. This definition is consistent with Definition\textsuperscript{3.3} the appropriate sublattice is obtained as the image $U(L^1(\mu_2))$.

In case of standard Borel spaces an operator $P_2$ is a factor of $P_1$ if and only if there is a measure-preserving surjection $\pi: X_1 \to X_2$ satisfying, for every $f \in L^1(\mu_2)$, the condition $(P_2f) \circ \pi = P_1(f \circ \pi)$. Indeed, a Markov embedding sends characteristic functions of sets to characteristic functions of other sets, hence it defines a homomorphism between measure algebras. In case of standard Borel spaces such homomorphism $\Pi$ is always induced by a pointwise measure preserving map $\pi : X_1 \to X_2$ by the formula $\Pi = \pi^{-1}$, so the general definition boils down to the pointwise one. Furthermore, if a measure preserving map $T_2$ is a factor of a map $T_1$ (in a classical sense) then the Koopman operator of $T_2$ is a factor of the Koopman operator of $T_1$.

A factor of an operator $P$ is pointwise if it is a Koopman operator of a measure preserving map. This may be understood in two equivalent ways: either there is a sublattice of the form $L^1(X, \Sigma_F, \mu)$, where $\Sigma_F$ is a complete $\sigma$-algebra, which satisfies

$$A \in \Sigma_F \Rightarrow \exists B \in \Sigma_F \; P1_A = 1_B,$$

or there is a dynamical system $(Z, \lambda, S)$, where $S : Z \to Z$ is a measure preserving map, and a Markov embedding $U : L^1(\lambda) \to L^1(\mu)$ with $PUf = U(f \circ S)$ for every $f \in L^1(\lambda)$.

A factor is trivial if the measure is concentrated in one point. Obviously, such factors are pointwise and the factor of $P$ is trivial if and only if it can be represented as a subspace $L^1(X, \Sigma', \mu)$, where $\Sigma'$ consists solely of sets of measure zero or one.

Given a map $T$ we define the following relation on $X$:

$$x \sim x' \iff \exists n \in \mathbb{N} \; T^n(x) = T^n(x').$$

Equivalently, one can replace the condition above by $\xi_n(x) = \xi_n(x')$ for some $n$ or by

$$\forall i, j \in \mathbb{N} \exists z \in X \; x, x' \in \xi_i(z) \text{ and } x' \in \xi_j(z).$$

Clearly, this is an equivalence relation so it decomposes the space $X$ into disjoint equivalence classes. The equivalence class of $x \in X$ will be denoted by $[x]$ and the corresponding quotient space by $\tilde{X}$. All such classes are both measurable in $X$ and $\xi_k$-measurable for each $k$.

Consider the space $(\tilde{X}, \tilde{\Sigma}, \tilde{\mu})$, where $\tilde{\Sigma}$ and $\tilde{\mu}$ are the $\sigma$-algebra and the measure transported from $(X, \mu)$ by the canonical map $x \mapsto [x]$. There is a natural pointwise action $\tilde{T}$ on $(\tilde{X}, \tilde{\Sigma}, \tilde{\mu})$ given by $\tilde{T}[x] = [Tx]$. It is obvious that this is a factor of $T$, but by a straightforward calculation one verifies that it also is a factor of $\Upsilon_T$.

**Lemma 3.4.** Let $L^1(Y, \Sigma_F, \mu)$ be a pointwise factor of $\Upsilon_T$. Then for every $A \in \Sigma_F$ the function $\Upsilon_T 1_A$ is a characteristic function of a set being a union of some equivalence classes of $\sim$.

**Proof.** If $A \in \Sigma_F$ then

$$\Upsilon_T 1_A = \sum_{k=1}^{\infty} b_k \mu_{\xi_k(z)}(T^{-1}A|_k) = 1_B|_1$$

for some set $B \in \Sigma_F$. Hence given $x$ one has $\mu_{\xi_k(x)}(T^{-1}A|_k) = 0$ for all $k$ or $\mu_{\xi_k(x)}(T^{-1}A|_k) = 1$ for all $k$. The value of $\mu_{\xi_k(x)}(A)$ is constant on each element.
of $\xi_k$. But if $x \sim x'$ then $x$ and $x'$ belong to the same atom of $\xi_i$ for some $i$, so for every $k$ the function $x \mapsto \mu_{\xi_k(x)}(A)$ is also constant on each equivalence class of $\sim$. Thus, $B|_1$ is a union of equivalence classes on which $\mu_{\xi_k(x)}(A) = 1$. □ □

**Theorem 3.5.** The exploding operator $\Upsilon_T$ has no non-trivial pointwise factor if and only if the $\sigma$-algebra $\Sigma$ consists solely of sets of measure zero or one.

**Proof.** If $\tilde{\Sigma}$ contains a set with measure in $(0, 1)$ then $(\tilde{X}, \tilde{\Sigma}, \tilde{\mu}, \tilde{T})$ is a non-trivial pointwise factor. If not and if $L^1(Y, \Sigma_F, \mu)$ is a pointwise factor of $\Upsilon_T$ then by the above lemma for every $A \in \Sigma_F$ it holds that $\Upsilon_T 1_A|_1$ is either equal to zero $\mu$-a.e. or it is equal to one $\mu$-a.e.

$$\mu(A|_1) = \nu(A|_1 \times \mathbb{N}) = \int 1_{A|_1 \times \mathbb{N}}d\nu = \int \Upsilon_T 1_{A|_1 \times \mathbb{N}}d\nu$$

$$= \begin{cases} \sum_{k=2}^{\infty} a_k \mu(T^{-1}(A|_1)) & \text{if } \Upsilon_T 1_A|_1 = 0 \mu\text{-a.e.} \\ a_1 + \sum_{k=2}^{\infty} a_k \mu(T^{-1}(A|_1)) & \text{if } \Upsilon_T 1_A|_1 = 1 \mu\text{-a.e.} \end{cases}$$

which is not equal to $\mu(A|_1)$, unless $\mu(A|_1) = 0$ or $\mu(A|_1) = 1$. In the first case,

$$\sum_{k=1}^{\infty} a_k \mu(A|_k) = \nu(A) = \int \Upsilon_T 1_A d\nu = \int_{\times \{2, 3, \ldots\}} \Upsilon_T 1_A d\nu$$

$$= \sum_{k=2}^{\infty} a_k \mu(T^{-1}(A|_{k-1})) + \sum_{k=1}^{\infty} a_{k+1} \mu(A|_k).$$

Thus, $\sum_{k=1}^{\infty} (a_k - a_{k+1}) \mu(A|_k) = 0$, so the fact that the sequence $(a_k)$ is strictly decreasing implies that $\mu(A|_k) = 0$ for all $k$. Performing similar calculations for the case $\mu(A|_1) = 1$, we obtain

$$\sum_{k=1}^{\infty} b_k \mu(A|_k) = \sum_{k=1}^{\infty} \frac{a_k - a_{k+1}}{a_1} \mu(A|_k) = 1,$$

which happens only if $\mu(A|_k) = 1$ for all $k$. Hence, $\nu(A) = 0$ or $\nu(A) = 1$ for all $A \in \Sigma_F$. □ □

**Corollary 3.6.** If $T$ is a one-sided (noninvertible) Bernoulli shift then $\Upsilon_T$ is strictly non-pointwise.

**Proof.** The equivalence relation identifies all points, which are sequences having the same tail. It is not hard to show that any set which belongs to the $\sigma$-algebra defined by such relation is a member of the tail $\sigma$-algebra, therefore it has measure zero or one, by the Kolmogorov’s zero-one law. □ □

4. Ergodicity

A doubly stochastic operator is **ergodic** if constant functions are the only functions invariant under the action of the operator.

Let us start with the following operator-theoretic restatement of a classical equivalent definition of ergodicity. Though it is probably well known to the experts, we include the proof.

**Theorem 4.1.** A doubly stochastic operator $P$ is ergodic if and only if for any non-negative function $f$, which is not equal almost everywhere to zero, the sum $\sum_{k=1}^{\infty} P^k f$ is positive almost everywhere.
Proof. Assume first that $P$ is ergodic. Let $f$ be a nonnegative function. By the Chacon-Ornstein theorem, the averages $\frac{1}{n} \sum_{k=1}^{n} P^k f$ converge almost everywhere to an invariant function. If $\sum_{k=1}^{\infty} P^k f = 0$ on a set of positive measure then the limit function is zero on this set, hence by ergodicity it must be equal to zero almost everywhere. Since $f$ and the limit function have the same integral, also $f$ must be zero almost everywhere.

Conversely, assume that $f$ is a nonconstant invariant function for $P$. Then $f < \int f \, d\mu$ on a set of positive measure. If $g$ is not equal almost everywhere to zero. If $f$ is not equal almost everywhere to zero. If $g$ is not equal almost everywhere to zero. If $g$ is not equal almost everywhere to zero. If $g$ is not equal almost everywhere to zero. If $g$ is not equal almost everywhere to zero. If $g$ is not equal almost everywhere to zero.

Theorem 4.2. The operator $\Upsilon_T$ is ergodic if and only if the map $T$ is ergodic.

Proof. Assume that $T$ is not ergodic, so there is $A \subset X$ such that $0 < \mu(A) < 1$ and $T^{-1}A = A$. Clearly, for $k > 1$:
$$\Upsilon_T 1_A \times \{x, k\} = 1_A (T x, k - 1) = 1_A (T x) = 1_A (x, k).$$

Furthermore, if $x \in A$ then $\xi_k(x) \subset A$, so
$$\Upsilon_T 1_A \times \{x, 1\} = \sum_{k=1}^{\infty} b_k \mu(\xi_k(x) (T^{-1}A)) = \sum_{k=1}^{\infty} b_k \mu(\xi_k(x) (A)) = \sum_{k=1}^{\infty} b_k = 1.$$

Therefore, $0 \leq \Upsilon_T 1_A \times \{x, k\} \leq 1_A (x, k)$. Both these functions have equal integrals, so they must be equal and $1_A \times \{x, k\}$ is a non-constant function, which is invariant under the action of $\Upsilon_T$.

For the converse statement, let $T$ be ergodic and let $f$ be a non-negative function, which is not equal almost everywhere to zero. If $f|_{T^{-1}}$ is strictly positive on a set $A \subset X$ with $\mu(A) > 0$ then, since $f|_{T^{-1}}$ is constant on $\xi_1(x)$, it holds that
$$\Upsilon_T f(x, 1) \geq b_1 \int f|_{T^{-1}} d\mu_{\xi_1(x)} = b_1 f(T x, 1) > 0$$

for $x \in T^{-1}A$. Inductively, $\Upsilon_T f(x, 1) \geq b_1 \Upsilon_T f(x, 1) > 0$ for $x \in T^{-k}A$. Ergodicity of $T$ implies that $\bigcup_{n=1}^{\infty} T^{-n}A$ has measure equal to one, so the sum $\sum_{n=1}^{\infty} \Upsilon_T f(x, 1)$ is positive $\mu$-almost everywhere. But then it also holds that
$$\sum_{n=1}^{\infty} \Upsilon_T f(x, k) \geq \sum_{n=k}^{\infty} \Upsilon_T f(x, k) > 0$$

for almost every $x$ and every $k > 1$.

If $f$ equals zero almost everywhere on $X \times \{1\}$ then $f(x, k) > 0$ on a set $A \subset X$ of positive measure $\mu$ for some $k > 1$. By definition,
$$\Upsilon_T f(x, 1) \geq b_k \int f|_k d\mu_{\xi_k(x)}$$
and the right hand side is positive on a set of positive measure $\mu$. Indeed,

$$0 < \int f|_k \, d\mu = \int f|_k \circ T \, d\mu = \int_{X/\mathcal{C}} \int_C f|_k \circ T \, d\mu_C \, d\mu_k(C),$$

so $\int_C f|_k \circ T \, d\mu_C$ is positive on a set of positive measure $\mu$. Hence, $\Upsilon_T f(x,1)$ is positive on a set of positive measure $\mu$ and the hypothesis follows as before. □ □

5. Entropy

For the sequence $(b_n)$ let us denote its $i$th partial sum by $S(i) = \sum_{k=1}^{i} b_k$ and its $i$th tail by $R(i) = \sum_{k=i+1}^{\infty} b_k$. We will prove that if $\sum_i R(i)$ converges then the entropy of $\Upsilon_T$ is bounded from below by the entropy of $T$. This assumption is satisfied for example by geometric sequences (but not only for them). Let $\|\cdot\|_\infty$ denote the norm in $L^\infty(Y,\nu)$.

Lemma 5.1. If $R(i)$ is a summable sequence, then for every measurable set $A \subset X$ the sequence

$$\sup_{n \in \mathbb{N}} \| (\Upsilon_T)^n 1_{T^{-i}A \times X} - 1_{T^{-(i+n)}A \times X} \|_\infty$$

converges to 0, when $i$ goes to infinity.

Proof. Let $i$ be a positive integer. For $k < i$ the set $T^{-(i+1)}A \cap \xi_k(x)$ is nonempty if and only if $T^k x \in T^{-(i-k+1)}A$, i.e., $x \in T^{-(i+1)}A$. In this case, if $y \in \xi_k(x)$ then $T^k y = T^k x \in T^{-(i-k+1)}A$, so $y \in T^{-(i+1)}A$. Therefore, $\xi_k(x) \subset T^{-(i+1)}A$, implying

$$\mu \xi_k(x) \left( T^{-(i+1)}A \right) = 1_{T^{-(i+1)}A}(x) \quad \text{for } k \leq i.$$

Consequently,

$$S(i) \cdot 1_{T^{-1}A}(x) \leq \Upsilon T 1_{T^{-1}A}(x,1) = \sum_{k=1}^{\infty} b_k \mu \xi_k(x) \left( T^{i-1}A \right) \leq S(i) \cdot 1_{T^{-1}A}(x) + R(i),$$

hence

$$\left| \Upsilon T 1_{T^{-1}A}(x,1) - 1_{T^{-(i+1)}A}(x,1) \right| \leq R(i).$$

Since $\Upsilon T 1_{T^{-1}A}(x,k) = 1_{T^{-i+1}A}(x)$ for $k \geq 2$, using the above inequality one gets

$$\| \Upsilon T 1_{T^{-1}A} - 1_{T^{-(i+1)}A} \|_\infty \leq R(i)$$

for all $i$.

Assume inductively that

$$\| \Upsilon T 1_{T^{-1}A} - 1_{T^{-(i+n)}A} \|_\infty \leq R(i) + \ldots + R(i + n - 1)$$

for some $n$. Since $\Upsilon T$ is a $L^\infty$ contraction, it holds that

$$\| \Upsilon T 1_{T^{-1}A} - 1_{T^{-(i+n-1)}A} \|_\infty \leq \| \Upsilon T 1_{T^{-1}A} - 1_{T^{-(i+n-1)}A} \|_\infty + \| 1_{T^{-(i+n-1)}A} \|_\infty \leq R(i) + \ldots + R(i + n - 1) + R(i + n).$$

Hence, $\| \Upsilon T 1_{T^{-1}A} - 1_{T^{-(i+n)}A} \|_\infty \leq \sum_{k=1}^{\infty} R(k)$ for every $n$, which ends the proof □ □
Theorem 5.2. If \( h_\mu(P,F) \) is a summable sequence, then \( h_\mu(T) \leq h_\nu(T) \).
Proof. Fix a partition $\xi$ of $X$ and $\varepsilon > 0$. Denote by $\text{Id} : X \to X$ the identity map on $X$. Given $i \in \mathbb{N}$ one calculates:

$$H_{\mu} \left( \bigvee_{n=0}^{N} T^{-i-n} \xi \right) = H_{\mu} \left( \bigvee_{n=0}^{N} T^{n} \mathbb{1}_{T^{-i} \xi} \right) = H_{\nu} \left( \bigvee_{n=0}^{N} (T \times \text{Id})^{n} \mathbb{1}_{T^{-i} \xi \times \mathbb{N}} \right) \leq H_{\nu} \left( \bigvee_{n=0}^{N} \mathbb{Y}^{n}_{T} \mathbb{1}_{T^{-i} \xi \times \mathbb{N}} \right) + H_{\nu} \left( \bigvee_{n=0}^{N} (T \times \text{Id})^{n} \mathbb{1}_{T^{-i} \xi \times \mathbb{N}} \bigg| \bigvee_{n=0}^{N} \mathbb{Y}^{n}_{T} \mathbb{1}_{T^{-i} \xi \times \mathbb{N}} \right) \leq H_{\nu} \left( \bigvee_{n=0}^{N} \mathbb{Y}^{n}_{T} \mathbb{1}_{T^{-i} \xi \times \mathbb{N}} \right) + \sum_{n=0}^{N} H_{\nu} \left( (T \times \text{Id})^{n} \mathbb{1}_{T^{-i} \xi \times \mathbb{N}} \bigg| \mathbb{Y}^{n}_{T} \mathbb{1}_{T^{-i} \xi \times \mathbb{N}} \right).$$

By the preceding lemma and the continuity of entropy with respect to a collection of functions, the expression under the sum is smaller than $\varepsilon$ if only $i$ is big enough. Moreover, for every $i$ one has $h_{\mu}(T, \xi) = h_{\mu}(T, T^{-i} \xi)$. Therefore,

$$h_{\mu}(T, \xi) \leq h_{\nu}(\mathbb{Y}_{T}, \mathbb{1}_{T^{-i} \xi \times \mathbb{N}}) + \varepsilon \leq h_{\nu}(\mathbb{Y}_{T}) + \varepsilon$$

and the inequality follows by taking supremum over $\xi$ and infimum over $\varepsilon$. \hfill \Box \hfill \Box

Final remarks.

(1) It seems very unlikely that the entropy of $\mathbb{Y}_{T}$ could ever be strictly higher than the entropy of $T$, but, surprisingly, I was not able to prove the equality. However, my conjecture is that the equality holds at least if $T$ is a Bernoulli shift.

(2) Let $X$ be a compact or, more generally, Polish space. An operator $P : C(X) \to C(X)$ is Markov operator if it is positive and preserves constants (in non-compact case $C(X)$ is understood as the space of bounded continuous functions). For a continuous map $T$ our definition \ref{definition:markov} gives a Markov operator on $C(Y)$ if for every continuous $g \in C(X)$ the map $x \mapsto \int g d\mu_{\xi_{k}(x)}$ is everywhere defined and continuous, i.e., if $x \mapsto \mu_{\xi_{k}(x)}$ is continuous in the weak$^*$ topology. It seems reasonable to ask how restrictive are these demands. In \cite{9} one finds an interesting non-classical approach to the idea of disintegration, which yields the same result as the usual disintegration, if well-defined. In particular, theorem 7.1 there gives (together with preceding definitions and construction) a set of assumptions guaranteeing that our definition of $\mathbb{Y}_{T}$ is possible in topological setup. It states that if $X$ and $Z$ are both locally compact and $\sigma$-compact Hausdorff spaces with Radon measures $\mu$ and $\nu$, respectively, $t : X \to Z$ is a continuous map and $Z$ is the support of $\nu$, then the disintegration $\mu_{z}$ of $\mu$ along $t$ is defined for all $z \in Z$ and the map $z \to \mu_{z}$ is continuous. In our case, for a given $k$ we consider a partition of $X$ into closed sets $C_{x,k} = T^{-k}T^{k}(x)$ and the role
of $Z$ is played by the quotient space $X/\xi_k$. By the definition of identification topology in $X/\xi_k$, this space is a $T_1$-space and a canonical surjection $x \to C_{x,k}$ is continuous (e.g., see [4]). If this map was open, then compactness of $X$ would imply that $X/\xi_k$ is a Hausdorff space and compactness of $X/\xi_k$ would follow easily. It is indeed open if $X$ is a subshift—it is easy to see that the image of a cylinder under the identification map is a set in $X/\xi_k$, such that the union of its elements (treated as subsets of $X$) is also a cylinder. Any measure on $X$ with full support transports to a measure with full support on $X/\xi_k$. So the definition 2.1 makes sense in topological setup at least in the class of all subshifts having invariant measure with full support. However, to study entropy of this operator one either needs to extend entropy theory introduced in [3] beyond compact spaces or to define the operator on some compactification of $Y$.

References

[1] I.P.Cornfeld, S.V.Fomin, Ya.G.Sinai, *Ergodic Theory*, Springer-Verlag, 1982
[2] T.Downarowicz, *Entropy in dynamical systems*, Cambridge University Press, *New Mathematical Monographs* 18, Cambridge, 2011
[3] T.Downarowicz and B.Frej, *Measure-theoretic and topological entropy of operators on function spaces*. Ergod.Th.&Dynam.Sys. 25 (2005), no. 2, pp.455–481
[4] J.Dugundji, *Topology*. Allyn and Bacon Inc., 1966
[5] T.Eisner, B.Farkas, M.Haase, R.Nagel, *Operator Theoretic Aspects of Ergodic Theory*. Graduate Texts in Mathematics, Springer 2015
[6] E.Glasner, *Ergodic Theory via Joinings*. Mathematical Surveys and Monographs, vol. 101, American Mathematical Society, Providence, RI, 2003
[7] B.Frej and P.Frej, *An integral formula for entropy of doubly stochastic operators*. Fund.Math. 213, 2011, pp.271–89
[8] B.Frej and D.Huczek, *Doubly stochastic operators with zero entropy*. To appear in Annals of Functional Analysis.
[9] T.Tjur, *A constructive definition of conditional distributions*. Preprint 13. University of Copenhagen: Institute of Mathematical Statistics, 1975.

Faculty of Pure and Applied Mathematics, Wroclaw University of Science and Technology, Wybrzeże Wyspiańskiego 27, 50-370 Wroclaw, Poland.

https://orcid.org/0000-0002-6084-7292

E-mail address: Bartosz.Frej@pwr.edu.pl