



Ann. Funct. Anal. 10 (2019), no. 1, 144–156  
<https://doi.org/10.1215/20088752-2018-0015>  
ISSN: 2008-8752 (electronic)  
<http://projecteuclid.org/afa>

## DOUBLY STOCHASTIC OPERATORS WITH ZERO ENTROPY

BARTOSZ FREJ<sup>\*</sup> and DAWID HUCZEK

Communicated by S. Barza

ABSTRACT. We study doubly stochastic operators with zero entropy. We generalize three famous theorems: Rokhlin’s theorem on genericity of zero entropy, Kushnirenko’s theorem on equivalence of discrete spectrum and nullity, and Halmos–von Neumann’s theorem on representation of maps with discrete spectrum as group rotations.

### 1. Introduction

Let  $\mu$  be a probability measure on a measurable space  $(X, \Sigma)$ . By a *doubly stochastic operator* (also called *bistochastic* or *Markov operator*) we mean a linear operator  $T$  on the space  $L^1(\mu)$  of integrable functions which fulfills the following conditions:

- (i)  $Tf$  is positive for every positive  $f \in L^1(\mu)$ ,
- (ii)  $T\mathbb{1} = \mathbb{1}$  (where  $\mathbb{1}(x) = 1$  for all  $x \in X$ ),
- (iii)  $\int Tf \, d\mu = \int f \, d\mu$  for every  $f \in L^1(\mu)$ .

For each  $p > 1$ , the space  $L^p(\mu) \subset L^1(\mu)$  is invariant under the action of a doubly stochastic operator  $T$ . Conversely, if  $T : L^p(\mu) \rightarrow L^p(\mu)$  is doubly stochastic, then it can be uniquely extended to an operator on  $L^q(\mu)$  for any  $1 \leq q < p$ , with no harm to the above properties. Therefore, we can study doubly stochastic operators on the space  $L^p(\mu)$  for any  $p \geq 1$ , in particular, on the space  $L^2(\mu)$  of square integrable real functions. Note that such operators preserve conjugacy,

---

Copyright 2019 by the Tusi Mathematical Research Group.

Received Mar. 20, 2018; Accepted Jun. 22, 2018.

First published online Nov. 21, 2018.

<sup>\*</sup>Corresponding author.

2010 *Mathematics Subject Classification*. Primary 37A30; Secondary 28D20, 47A35.

*Keywords*. Markov operator, entropy, discrete spectrum, Kushnirenko’s theorem, Halmos–von Neumann’s theorem.

and hence transform real functions into real functions; conversely, the action of  $T$  is determined by its action on the space of real functions. The set of all doubly stochastic operators on  $L^1(\mu)$  will be denoted by  $\mathcal{DS}(\mu)$ .

A special role is played by operators which are induced by *transition probabilities* (or *probability kernels*), that is, operators defined by the formula

$$Tf(x) = \int f(y) P(x, dy),$$

where  $P : X \times \Sigma \rightarrow [0, 1]$  is a transition probability. It is known that on a standard probability space every doubly stochastic operator is of this form.

In this article, we generalize three famous theorems on dynamical systems with zero entropy to doubly stochastic operators with zero entropy. The first one is Rokhlin's theorem (see [8]) on genericity of systems with entropy zero in the weak topology on the set of all measure-preserving maps on  $(X, \Sigma, \mu)$ . Vershik [9] asked if the same holds for entropy of Markov operators, whatever the definition of entropy can be in this case. We prove that, in the strong operator topology, the set of doubly stochastic operators with zero entropy is residual in the set of all doubly stochastic operators on  $L^1(\mu)$ . Recall that the *strong operator topology* is just the topology of pointwise convergence; that is,  $T_n$  converges to  $T$  in the strong operator topology if  $\|T_n f - T f\|$  converges to zero for each  $f \in L^1(\mu)$ . It is known that in pointwise case, the weak topology on the space of transformations coincides with the strong operator topology on the set of their Koopman operators. Our result (Theorem 3.5) is thus a true analogue of Rokhlin's result.

The second aim is to give an analogue of Kushnirenko's theorem, which characterizes null transformations, that is, maps with zero sequence entropy along all sequences, as maps with discrete spectrum. In the operator case this equivalence is no longer valid, because null operators need not have a linearly dense set of eigenfunctions. Nevertheless, we present an appropriate modification using the Jacobs–de Leeuw–Glicksberg decomposition associated to an operator (see Theorem 5.4). The description of null operators naturally leads us to a generalization of the Halmos–von Neumann theorem (see Theorem 6.3), which in this formulation becomes the characterization of nullity (not the discrete spectrum).

In defining entropy, we will follow the ideas which first appeared in [2]. To be more specific, we put the finite collections of functions in the role that finite partitions played in the classic definition of entropy. First we define their join and the entropy of a single collection, then we calculate the entropy of  $T$  with respect to a collection as an appropriate limit, and finally we take the supremum over all such collections. The details are given in the forthcoming section.

## 2. Definition of entropy

By a *collection* of functions we mean a finite sequence of measurable functions with range in the unit interval. For a function  $f : X \rightarrow [0, 1]$ , let

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

and let  $\mathcal{A}_f$  be the partition of  $X \times [0, 1]$  consisting of  $A_f$  and its complement. For a collection  $\mathcal{F}$ , we define  $\mathcal{A}_{\mathcal{F}} = \bigvee_{f \in \mathcal{F}} \mathcal{A}_f$ . If  $\mathcal{F} \vee \mathcal{G}$  denotes a concatenation of collections  $\mathcal{F}$  and  $\mathcal{G}$ , then, clearly,  $\mathcal{A}_{\mathcal{F} \vee \mathcal{G}} = \mathcal{A}_{\mathcal{F}} \vee \mathcal{A}_{\mathcal{G}}$ . To shorten the notation, we write  $T^n \mathcal{F} = \{T^n f : f \in \mathcal{F}\}$  and  $\mathcal{F}_T^n$  for the concatenation of  $\mathcal{F}, T\mathcal{F}, \dots, T^{n-1}\mathcal{F}$ . For two collections of measurable functions  $\mathcal{F} = \{f_1, \dots, f_r\}$  and  $\mathcal{G} = \{g_1, \dots, g_{r'}\}$ ,  $r' \leq r$ , we define their  $L^1$ -distance  $\text{dist}(\mathcal{F}, \mathcal{G})$  by a formula

$$\text{dist}(\mathcal{F}, \mathcal{G}) = \min_{\pi} \left\{ \max_{1 \leq i \leq r} \int |f_i - g_{\pi(i)}| d\mu \right\},$$

where the minimum ranges over all permutations  $\pi$  of a set  $\{1, 2, \dots, r\}$  and where  $\mathcal{G}$  is considered an  $r$ -element collection by setting  $g_i \equiv 0$  for  $r' < i \leq r$ .

Let  $\lambda$  be the Lebesgue measure on the unit interval. We define

$$H(\mathcal{F}) = H_{\mu \times \lambda}(\mathcal{A}_{\mathcal{F}}) = - \sum_{A \in \mathcal{A}_{\mathcal{F}}} (\mu \times \lambda)(A) \cdot \log(\mu \times \lambda)(A),$$

$$h(T, \mathcal{F}) = \lim_{n \rightarrow \infty} \frac{1}{n} H(\mathcal{F}_T^n),$$

$$h(T) = \sup_{\mathcal{F}} h(T, \mathcal{F})$$

with the supremum ranging over all finite collections of measurable functions from  $X$  to  $[0, 1]$  (existence of the limit was proved in [2]). For the conditional entropy of a collection  $\mathcal{F}$  with respect to the collection  $\mathcal{G}$ , we put

$$H(\mathcal{F} | \mathcal{G}) = H(\mathcal{F} \vee \mathcal{G}) - H(\mathcal{G}) = H_{\mu \times \lambda}(\mathcal{A}_{\mathcal{F}} | \mathcal{A}_{\mathcal{G}}).$$

The following continuity assertion is true: for every  $r \geq 1$  and  $\varepsilon > 0$  there is a  $\delta > 0$  such that if  $\mathcal{F}$  and  $\mathcal{G}$  have cardinalities at most  $r$  and  $\text{dist}(\mathcal{F}, \mathcal{G}) < \delta$ , then  $|H(\mathcal{F} | \mathcal{G})| < \varepsilon$ .

It is known that the preceding procedure leads to a common value of entropy  $h(T)$  of a doubly stochastic operator for many reasonable choices of  $H(\mathcal{F})$  (see [1] or [2] for details on axiomatic theory of metric entropy of doubly stochastic operators). We find the above version convenient for our purposes. Moreover, we see that since the definition depends only on bounded functions, in fact only on functions  $X \rightarrow [0, 1]$ , the choice of the domain  $L^p(\mu)$  of an operator does not affect the value of entropy.

The *sequence entropy* of a measure-preserving map  $T$  with respect to a partition  $\xi$  along a subsequence  $A = (i_n)_{n \in \mathbb{N}}$  of  $\{0, 1, 2, \dots\}$  is defined by

$$h_A(T, \xi) = \limsup_{n \rightarrow \infty} \frac{1}{n} H\left(\bigvee_{k=1}^n T^{-i_k} \xi\right),$$

and the *sequence entropy* of  $T$  is

$$h_A(T) = \sup_{\xi} h_A(T, \xi),$$

where the supremum ranges over all finite partitions of  $X$ . Similarly, in the operator case we formulate the following definition.

*Definition 2.1.* The *sequence entropy* of a doubly stochastic operator  $T$  with respect to a collection of functions  $\mathcal{F}$  along a sequence  $A = (i_n)_{n \in \mathbb{N}}$  is defined as

$$h_A(T, \mathcal{F}) = \limsup_{n \rightarrow \infty} \frac{1}{n} H\left(\bigvee_{k=1}^n T^{i_k} \mathcal{F}\right),$$

and the *sequence entropy* of the operator  $T$  along a sequence  $A$  is given by

$$h_A(T) = \sup_{\mathcal{F}} h_A(T, \mathcal{F}).$$

By a brief inspection of arguments in [2], we check that the value of  $h_A(T)$  does not depend on the choice of the static entropy  $H(\mathcal{F})$  as long as the latter fulfills requirements of the axiomatic definition of entropy.

### 3. Genericity of entropy zero

**Lemma 3.1.** *Let  $T$  be a doubly stochastic operator on  $L^p(\mu)$ . Let  $Sf = \int f d\mu$ . Then  $(1 - \frac{1}{n})T + \frac{1}{n}S$  converges to  $T$  in operator norm, hence also in strong operator topology.*

*Proof.* We have

$$\begin{aligned} \left\| \left( \left(1 - \frac{1}{n}\right)T + \frac{1}{n}S \right) f - Tf \right\|_p &= \frac{1}{n} \|Tf - Sf\|_p \\ &\leq \frac{1}{n} (\|Tf\|_p + \|Sf\|_p) \leq \frac{2}{n} \|f\|_p. \end{aligned}$$

This ends the proof for strong operator topology. For norm topology, it is enough to take the supremum over  $f \in L^p(\mu)$  with  $\|f\|_p = 1$ .  $\square$

**Corollary 3.2.** *The set of doubly stochastic operators with zero entropy is dense in the strong operator topology and in the norm operator topology.*

*Proof.* For every  $\alpha \in (0, 1)$ , the operator  $(1 - \alpha)T + \alpha S$  has zero entropy, because  $((1 - \alpha)T + \alpha S)^n f$  converges to a constant function for every  $f$ .  $\square$

**Lemma 3.3.** *The set of doubly stochastic operators with zero entropy is a  $G_\delta$  set in the strong operator topology in  $L^1(\mu)$ .*

*Proof.* Fix  $\varepsilon > 0$ , a positive integer  $n$ , and a finite collection  $\mathcal{F}$  of  $r$  measurable functions  $X \rightarrow [0, 1]$ . We will show that the set

$$U(\varepsilon, n, \mathcal{F}) = \left\{ T \in \mathcal{DS}(\mu) : \frac{1}{n} H(\mathcal{F}^n) < \varepsilon \right\}$$

is open. Let  $T \in U(\varepsilon, n, \mathcal{F})$ , and let  $\frac{1}{n} H(\mathcal{F}^n) = \gamma < \varepsilon$ . Choose  $\delta > 0$  so that  $\text{dist}(\mathcal{G}, \mathcal{G}') < n\delta$  guarantees that  $H(\mathcal{G} \mid \mathcal{G}') + H(\mathcal{G}' \mid \mathcal{G}) < \varepsilon - \gamma$  for all collections  $\mathcal{G}, \mathcal{G}'$  containing  $r$  functions. Consider an open set

$$C = \left\{ P \in \mathcal{DS}(\mu) : \forall k = 0, \dots, n-1 \forall f \in \mathcal{F} \|PT^{k-1}f - T^k f\|_1 < \delta \right\}.$$

Note that if  $P$  lies in  $C$ , then we have

$$\begin{aligned} \|P^k f - T^k f\|_1 &\leq \|P^k f - P^{k-1}Tf\|_1 + \|P^{k-1}Tf - P^{k-2}T^2f\|_1 + \dots \\ &\quad + \dots + \|PT^{k-1}f - T^k f\|_1 < k\delta \end{aligned}$$

for every  $f \in \mathcal{F}$  and  $k = 0, 1, \dots, n-1$ , because  $P$  is a contraction. Then  $H(T^k F \mid P^k F) < \varepsilon$  implying that

$$H\left(\bigvee_{k=0}^{n-1} P^k F \mid \bigvee_{k=0}^{n-1} T^k F\right) \leq \sum_{k=0}^{n-1} (H(P^k F \mid T^k F)) < n(\varepsilon - \gamma)$$

and

$$\frac{1}{n} H\left(\bigvee_{k=0}^{n-1} P^k F\right) \leq \frac{1}{n} H\left(\bigvee_{k=0}^{n-1} T^k F\right) + \frac{1}{n} H\left(\bigvee_{k=0}^{n-1} P^k F \mid \bigvee_{k=0}^{n-1} T^k F\right) < \varepsilon.$$

Consequently,  $U(\varepsilon, n, \mathcal{F})$  is open.

Since  $L^1$  is separable, we can choose a countable dense subset  $\mathbb{F}$  of the set of all integrable functions with range in  $[0, 1]$ . Now

$$\mathcal{U} = \bigcap_m \bigcap_r \bigcap_{\mathcal{F} \in \mathbb{F}^r} \bigcap_N \bigcup_{n \geq N} U\left(\frac{1}{m}, n, \mathcal{F}\right)$$

is a  $G_\delta$  set. We will prove that this is exactly the set of operators having zero entropy.

If  $T$  has entropy zero, then it clearly belongs to this set. On the other hand, again let  $\varepsilon > 0$ , let  $n \in \mathbb{N}$ , and let  $\mathcal{F}$  be a collection of  $r$  functions. For  $T \in \mathcal{U}$ , choose  $\tilde{\mathcal{F}} \in \mathbb{F}^r$  so that  $\text{dist}(\mathcal{F}, \tilde{\mathcal{F}})$  is small enough to ensure that  $H(\mathcal{F} \mid \tilde{\mathcal{F}}) < \varepsilon/2$ . Then also  $H(T^k \mathcal{F} \mid T^k \tilde{\mathcal{F}}) < \varepsilon/2$  for  $k \in \mathbb{N}$  and, consequently,

$$H(\mathcal{F}_T^n \mid \tilde{\mathcal{F}}_T^n) \leq \sum_{k=0}^{n-1} H(T^k \mathcal{F} \mid T^k \tilde{\mathcal{F}}) < n\varepsilon/2.$$

Therefore,

$$H(\mathcal{F}_T^n) \leq H(\tilde{\mathcal{F}}_T^n) + H(\mathcal{F}_T^n \mid \tilde{\mathcal{F}}_T^n) < H(\tilde{\mathcal{F}}_T^n) + n\varepsilon/2.$$

For any  $m$  we can find arbitrarily large  $n$  for which  $\frac{1}{n} H(\tilde{\mathcal{F}}_T^n) < \frac{1}{m}$ , so taking  $m > \varepsilon/2$  we obtain the existence of a large  $n$  such that  $\frac{1}{n} H(\mathcal{F}_T^n) < \varepsilon$ . Finally, we obtain a subsequence of  $\frac{1}{n} H(\mathcal{F}_T^n)$  tending to zero, but since the whole sequence is convergent, it converges to zero. The collection  $\mathcal{F}$  was chosen freely, so the entropy of  $T$  is zero.  $\square$

*Remark 3.4.* Note that it follows that the set is also a  $G_\delta$  set in the norm topology, as it has fewer open sets than the strong operator topology.

From Lemma 3.3 and Corollary 3.2, we have the following result.

**Theorem 3.5.** *The set of doubly stochastic operators with zero entropy is residual in the strong operator topology and in the norm topology of  $L^1(\mu)$ .*

#### 4. Decompositions of the domain

We recall some basic facts of the Jacobs–de Leeuw–Glicksberg (JdLG-) decomposition and Nagy–Foias (NF-) decomposition. Our sources of information are [3] in case of JdLG-decomposition and [7] for NF-decomposition. We say that a subspace  $W$  of the space  $L^p(\mu)$  is *invariant* with respect to a doubly stochastic operator  $T : L^p(\mu) \rightarrow L^p(\mu)$  if  $TW \subset W$ .

*Definition 4.1.* A *JdLG-decomposition* of the space  $L^p(\mu)$  associated with a doubly stochastic operator  $T$  is a pair of  $T$ -invariant subspaces  $E_{\text{rev}}$  and  $E_{\text{aws}}$  such that  $L^p(\mu) = E_{\text{rev}} \oplus E_{\text{aws}}$ , where  $E_{\text{rev}}$  is the range and  $E_{\text{aws}}$  is the kernel of a unique minimal idempotent of the weak operator closure of the semigroup  $\{I, T, T^2, \dots\}$ . The spaces  $E_{\text{rev}}$  and  $E_{\text{aws}}$  are called the *reversible part* and the *almost weakly stable part* of the space, respectively.

The existence of a unique minimal idempotent follows from Ellis's famous theorem proved in [4, Lemma 1] and such a decomposition can be obtained even in more general situations. It is clear from the definition that this decomposition is unique. For our purposes, the following characterizations will be of greatest importance.

**Theorem 4.2.** *If  $T$  is a doubly stochastic operator (or just a contraction) on  $L^p(\mu)$  and  $E_{\text{rev}} \oplus E_{\text{aws}}$  is its JdLG-decomposition, then*

- (1)  $E_{\text{rev}} = \overline{\text{lin}}\{f \in L^p(\mu) : \exists \lambda \in \mathbb{C}, |\lambda| = 1 \text{ s.t. } Tf = \lambda f\}$  (see [3, Theorem 16.33]);
- (2) for every  $f \in E_{\text{aws}}$  either  $\{T^n f : n \in \mathbb{N}\}$  is not precompact in  $L^p$ -norm topology or  $\inf_{n \in \mathbb{N}} \|T^n x\| = 0$  (see [3, Theorem 16.29]).

Now assume that  $\mathcal{H}$  is a Hilbert space and that  $T$  is a contraction of  $\mathcal{H}$ . We say that  $W$  *reduces*  $\mathcal{H}$  if both  $W$  and  $W^\perp$  are invariant (or, equivalently,  $W$  is invariant with respect both to  $T$  and  $T^*$ ). An operator  $T$  on a Hilbert space  $\mathcal{H}$  is *completely nonunitary* (c.n.u.) if there is no reducing subspace on which it is unitary.

*Definition 4.3.* Let  $T$  be a contraction of a Hilbert space  $\mathcal{H}$ . An *NF-decomposition* is a decomposition  $\mathcal{H} = \mathcal{H}_{\text{uni}} \oplus \mathcal{H}_{\text{cnu}}$  into an orthogonal sum of two subspaces reducing  $T$  such that  $T|_{\mathcal{H}_{\text{uni}}}$  is unitary and  $T|_{\mathcal{H}_{\text{cnu}}}$  is c.n.u. The spaces  $\mathcal{H}_{\text{uni}}$  and  $\mathcal{H}_{\text{cnu}}$  are the *unitary part* and the *completely nonunitary part* of  $\mathcal{H}$ , respectively, and the restrictions of  $T$  to these subspaces are the *unitary part* and the *completely nonunitary part* of  $T$  (cf. [7, Theorem 3.2]).

Just as in the previous case, the decomposition is unique. Note that for a unitary operator the c.n.u. part is trivial, while for a c.n.u. operator the unitary part is trivial. A doubly stochastic operator on  $L^2(\mu)$  is never c.n.u., because the subspace consisting of constants is a reducing subspace on which  $T$  is unitary. Roughly speaking, the unitary part captures all the invertibility of a doubly stochastic operator in the sense that if there is an invertible pointwise factor, then there is a closed invariant subspace of  $L^2(\mu)$ . Moreover, if  $T$  is a Koopman operator of an invertible map, then  $L^2(\mu)_{\text{cnu}} = \{0\}$ . In other cases the c.n.u.

part is nontrivial and it can be the whole  $L_0^2(\mu)$ —the orthogonal complement of constants—even for pointwise operators, for example, for a one-sided shift. It is easy to see that if  $\mathcal{H}_{\text{rev}} \oplus \mathcal{H}_{\text{aws}}$  is a JdLG-decomposition and  $\mathcal{H}_{\text{uni}} \oplus \mathcal{H}_{\text{cnu}}$  is an NF-decomposition of a Hilbert space  $\mathcal{H}$ , then

$$\mathcal{H}_{\text{rev}} \subset \mathcal{H}_{\text{uni}} \quad \text{and} \quad \mathcal{H}_{\text{aws}} \supset \mathcal{H}_{\text{cnu}}.$$

## 5. Discrete spectrum versus nullity

One of the basic definitions in ergodic theory is that of a discrete spectrum of a measure-preserving transformation. In a more general setup, we have the following.

*Definition 5.1.* A power bounded operator  $T$  on a Banach space  $V$  has *discrete spectrum* if

$$V = \overline{\text{lin}}\{v \in V : \exists \lambda \in \mathbb{C}, |\lambda| = 1, \text{ s.t. } Tv = \lambda v\}.$$

Clearly, it is applicable to the case of doubly stochastic operators because  $\|T^n\|_p = 1$  for all  $n \in \mathbb{N}$ . Similarly to the case of pointwise maps, we have the following.

*Definition 5.2.* A doubly stochastic operator  $T$  is *null* if  $h_A(T) = 0$  for every sequence  $A = \{t_1 < t_2 < t_3 < \dots\}$  of positive integers.

We now state the theorem which we aim to generalize.

**Theorem 5.3** (Kushnirenko [6, Theorem 4]). *A transformation  $T$  has discrete spectrum if and only if  $T$  is null.*

This equivalence fails for doubly stochastic operators. For example,  $Tf = \int f d\mu$  does not have discrete spectrum, but  $h_A(T) = 0$  for every sequence  $A$ , which shows that nullity does not imply discrete spectrum. On the other hand, a contraction may have a linearly dense set of eigenvectors, though its unitary part is reduced to constants. Indeed, let  $R$  be an irrational rotation on the circle  $S^1$ , let  $Rz = \alpha z$ , and let  $Tf = \frac{1}{2}f \circ R + \frac{1}{2} \int f d\lambda$ , where  $\lambda$  is the Haar probability measure. Clearly, each  $e_n(z) = z^n$  is a “c.n.u eigenfunction” for eigenvalue  $\frac{1}{2}\alpha^n$ . These examples show that for a general doubly stochastic operator null sequential entropy does not imply anything about existence or density of eigenvectors.

**Theorem 5.4.** *The following conditions are equivalent for a doubly stochastic operator  $T : L^p(\mu) \rightarrow L^p(\mu)$ .*

- (1)  $T$  is null.
- (2)  $L^p(\mu) = V \oplus W$ , where  $T$  has discrete spectrum on  $V$  and  $\lim_{n \rightarrow \infty} \|T^n f\|_p = 0$  for every  $f \in W$ .
- (3) If  $E_{\text{rev}} \oplus E_{\text{aws}}$  is a JdLG-decomposition of  $L^p(\mu)$  associated with  $T$ , then  $\lim_{n \rightarrow \infty} \|T^n f\|_p = 0$  for every  $f \in E_{\text{aws}}$ .

If  $p = 2$ , then each of the above conditions is equivalent to the following.

- (4) If  $\mathcal{H}_{\text{uni}} \oplus \mathcal{H}_{\text{cnu}}$  is the NF-decomposition of  $L^2(\mu)$  with respect to  $T$ , then  $T$  has discrete spectrum on  $\mathcal{H}_{\text{uni}}$  and  $\|T^n f\|_2$  converges to zero for every  $f \in \mathcal{H}_{\text{cnu}}$ .

Before starting the proof of this theorem, we need to state several lemmas. Walking in Kushnirenko's footsteps, we translate a part of his proof to the operator case.

**Lemma 5.5.** *Let  $A = \{t_1 < t_2 < t_3 < \dots\}$  be a sequence of positive integers. Then  $h_A(T) = 0$  if and only if  $h_A(T, \{f\}) = 0$  for every  $f$ .*

*Proof.* By the definition of entropy,  $h_A(T) \geq \sup_f h_A(T, \{f\})$ , where the supremum ranges over the set of all measurable functions  $f : X \rightarrow [0, 1]$ . On the other hand, it follows easily from the subadditivity of entropy that if  $h_A(T, \{f\}) = 0$  for every  $f$ , then  $h_A(T, \mathcal{F}) = 0$  for every collection  $\mathcal{F}$  of measurable functions.  $\square$

**Lemma 5.6.** *Let  $f : X \rightarrow [0, 1]$  be a measurable function. Then the closed orbit  $\bar{O}_T(f) = \{\overline{T^n f : n = 0, 1, 2, \dots}\}$  is compact if and only if  $h_A(T, \{f\}) = 0$  for all sequences  $A = \{t_1 < t_2 < t_3 < \dots\}$ .*

*Proof.* Fix  $\varepsilon > 0$ , and let  $A = \{t_n : n \in \mathbb{N}\}$ . Let us abbreviate  $\mathcal{A}_{T^n f}$  by  $\xi_n$ . Let  $\mathcal{Z}$  be the set of all countable measurable partitions of  $X$  with  $H_\mu(\xi) < \infty$ . We recall that in the context of entropy, the natural distance between two partitions  $\xi$  and  $\zeta$  belonging to  $\mathcal{Z}$  is given by the Rokhlin metric:

$$\rho(\xi, \zeta) = H_\mu(\xi \mid \zeta) + H_\mu(\zeta \mid \xi).$$

If  $\bar{O}_T(f)$  is compact in  $L^p(\mu)$ , then  $\{\mathcal{A}_g : g \in \bar{O}_T(f)\}$  is compact in  $\mathcal{Z}$ . Indeed,  $\lim_{n \rightarrow \infty} \|g_n - g\|_p = 0$  implies that  $\lim_{n \rightarrow \infty} \mu(A_{g_n} \Delta A_g) = 0$ , so a convergent subsequence of  $(g_n)_{n \in \mathbb{N}}$  induces a convergent subsequence of  $(\mathcal{A}_{g_n})_{n \in \mathbb{N}}$ . The closure of  $\{\xi_{t_n} : n \in \mathbb{N}\}$  is also compact, and hence it contains an  $\varepsilon$ -dense set  $\{\xi_{t_n} : 1 \leq n \leq N\}$ . Therefore,

$$\begin{aligned} h_A(T, \{f\}) &= \limsup_{n \rightarrow \infty} \frac{1}{n} H\left(\bigvee_{i=1}^n \xi_{t_i}\right) \\ &\leq \limsup_{n \rightarrow \infty} \frac{1}{n} \left( H\left(\bigvee_{i=1}^N \xi_{t_i}\right) + \sum_{j=N+1}^n H\left(\xi_{t_j} \mid \bigvee_{i=1}^N \xi_{t_i}\right) \right) \leq \varepsilon. \end{aligned}$$

Conversely, let  $\bar{O}_T(f)$  be noncompact; that is, for some  $\gamma > 0$  there is a sequence  $g_n \in \bar{O}_T(f)$  satisfying  $\|g_m - g_n\|_p \geq \gamma$  for any  $m \neq n$ . Then  $\|g_m - g_n\|_1 \geq \gamma^p$ , implying that  $\mu(A_{g_m} \Delta A_{g_n})$  are separated from zero for all pairs  $m \neq n$ . Therefore,  $\rho(\mathcal{A}_{g_m}, \mathcal{A}_{g_n}) > \varepsilon$  for some  $\varepsilon > 0$  and all  $m \neq n$ . We now inductively choose a sequence  $\xi_n$  satisfying  $\limsup_{n \rightarrow \infty} \frac{1}{n} H\left(\bigvee_{i=1}^n \xi_i\right) \geq \delta$  for some  $\delta > 0$ . It suffices to ensure that  $H(\xi_n \mid \bigvee_{i=1}^{n-1} \xi_i) \geq \delta$  for all  $n > 1$ . The appropriate number  $\delta > 0$  is chosen using the following lemma. This is a slight modification of the one proved in [6], so we leave the proof as an exercise (hint: if  $g \in \bar{O}_T(f)$ , then  $\int g d\mu = \int f d\mu$ ).

**Lemma 5.7.** *For every  $\varepsilon > 0$ , there exists a  $\delta > 0$  such that for any finite partition  $\xi$  of  $X \times [0, 1]$  and any partition  $\zeta = \mathcal{A}_g$  with  $g \in \bar{O}_T(f)$  the inequality  $H(\zeta \mid \xi) < \delta$  implies the existence of  $\xi' \prec \xi$  such that  $\rho(\zeta, \xi') < \varepsilon/2$ .*

Let  $\xi_1 = \mathcal{A}_{g_1}$ . Note that there are only two partitions coarser than  $\xi_1$ , namely, the trivial partition and  $\xi_1$  itself. By the above lemma, if  $H(\mathcal{A}_{g_m} \mid \xi_1) < \delta$



for all  $m$ , then each  $\mathcal{A}_{g_m}$  is  $\varepsilon/2$ -close (in Rokhlin metric) to one of these two partitions, which contradicts the fact that all partitions  $\mathcal{A}_{g_m}$  are  $\varepsilon$ -separated. Hence, there is a partition  $\xi_2$  of the form  $\mathcal{A}_{g_m}$  such that  $H(\xi_2 \mid \xi_1) \geq \delta$ . For the induction step, suppose that we have already found partitions  $\xi_1, \xi_2, \dots, \xi_{n-1}$ , chosen from  $(\mathcal{A}_{g_m})_m$ . Let  $\xi = \bigvee_{i=1}^{n-1} \xi_i$ . Similarly as before, there are only finitely many partitions  $\xi'$  coarser than  $\xi$ , and each  $\mathcal{A}_{g_m}$  satisfying  $H(\mathcal{A}_{g_m} \mid \xi) < \delta$  is  $\varepsilon/2$ -close to some  $\xi'$  (by Lemma 5.7). Since all partitions  $\mathcal{A}_{g_m}$  are  $\varepsilon$ -separated, the tail of the sequence  $(\mathcal{A}_{g_m})_m$  contains some element  $\xi_n$  satisfying  $H(\xi_n \mid \xi) \geq \delta$ . (Note that it is irrelevant that  $\xi'$  need not be of the form  $\mathcal{A}_g$  for any  $g$ .)  $\square$

Putting these pieces together, we obtain the following.

**Corollary 5.8.** *A doubly stochastic operator  $T$  on  $L^p(\mu)$  is null if and only if the closed orbit  $\bar{O}_T(f)$  is compact for every  $f \in L^p(\mu)$ .*

**Lemma 5.9.** *If  $T$  is a doubly stochastic operator, then the closed orbit  $\bar{O}_T(f)$  is compact for any  $f \in E_{\text{rev}}$ .*

*Proof.* If  $\{e_n : n \in \mathbb{N}\}$  is a linearly dense set of eigenvectors, then the orbit of  $f = \sum_{n=1}^{\infty} a_n e_n$  is a subset of the compact set  $\{\sum_{n=1}^{\infty} b_n e_n : |b_n| \leq |a_n|\}$ .  $\square$

**Corollary 5.10.** *If  $T$  is a doubly stochastic operator with discrete spectrum, then  $T$  is null.*

*Proof of Theorem 5.4.* The third condition is just a reformulation of the second one, as  $E_{\text{rev}}$  is exactly the set spanned by the eigenvectors of  $T$ . We concentrate on proving that the first two statements are equivalent. We remark that for a doubly stochastic operator, we have  $\inf_{n \in \mathbb{N}} \|T^n f\|_p = \lim_{n \rightarrow \infty} \|T^n f\|_p$ .

Assume that  $T$  is null. Let  $V = E_{\text{rev}}$ ,  $W = E_{\text{aws}}$ , where  $E_{\text{rev}} \oplus E_{\text{aws}}$  is a JdLG-decomposition of  $L^p(\mu)$  associated with  $T$ . By definition of  $E_{\text{rev}}$ , an operator  $T$  has discrete spectrum on  $V$ . By Corollary 5.8, nullity is equivalent to all orbits being precompact, which in view of Theorem 4.2 forces all orbits in  $E_{\text{aws}}$  to approach zero. On the other hand, assuming the second condition, we obtain from Lemma 5.9 that all orbits of elements of  $V$  are precompact. Clearly, orbits of elements of  $W$  are also precompact (simply because they are convergent sequences), so all functions from  $V \oplus W$  have precompact orbits and it follows from Corollary 5.8 that  $T$  is null.

To prove the last statement, note that it is always true that  $E_{\text{rev}} \subset \mathcal{H}_{\text{uni}}$  and  $\mathcal{H}_{\text{cnu}} \subset E_{\text{aws}}$ . If  $T^n f$  converges to zero, then  $f \in \mathcal{H}_{\text{cnu}}$ , hence these two decompositions coincide. On the other hand, if  $T$  has discrete spectrum on  $\mathcal{H}_{\text{uni}}$ , then  $\mathcal{H}_{\text{uni}} = E_{\text{rev}}$ , so again the decompositions coincide.  $\square$

An entire class of well-studied operators is guaranteed to be null.

**Definition 5.11.** A doubly stochastic operator  $T$  on  $L^2(\mu)$  is *quasicompact* if there exists a direct sum decomposition  $L^2(\mu) = F \oplus H$  and  $r < 1$  such that

- (1)  $F$  and  $H$  are closed invariant subspaces of  $L^2(\mu)$ ,
- (2)  $\dim(F) < \infty$  and all eigenvalues of  $T|_F$  have modulus larger than  $r$ ,
- (3) the spectral radius of  $T|_H$  is smaller than  $r$ .

**Theorem 5.12.** *If  $T$  is a quasicompact doubly stochastic operator, then  $T$  is null.*

*Proof.* By Corollary 5.8, it suffices to show that the closed orbit of  $f$  is compact for every  $f \in L^2(\mu)$ . Let  $L^2(\mu) = F \oplus H$  be the decomposition from the definition of quasicompactness, and let  $f = g + h$ , where  $g \in F$ ,  $h \in H$ . Since  $F$  is a finite-dimensional invariant subspace, the closure of the set  $\{T^n g : n \in \mathbb{N}\}$  is compact. On the other hand, by the spectral radius formula, we know that for every  $\varepsilon > 0$  we have  $\|T^n h\|_2 / \|h\|_2 = O(e^{n\varepsilon} r^n)$ . If we take  $\varepsilon < \log \frac{1}{r}$ , this implies that  $\|T^n h\|_2$  converges to zero, and thus the orbit  $\bar{O}_T(f)$  is compact.  $\square$

*Remark 5.13.* The above argument can be repeated to prove the following fact: if  $T$  is a doubly stochastic operator and  $V \oplus W$  is the decomposition into unitary and c.n.u. parts such that  $T|_V$  has discrete spectrum and  $T|_W$  has spectral radius strictly smaller than 1, then  $T$  is null.

To illustrate that having the spectral radius equal to 1 does not guarantee nonnullity, we will show an example of a null c.n.u. operator which has spectral radius equal to 1 when considered on the orthogonal complement of constants. In addition, the operator has no nontrivial eigenfunctions.

*Example 5.14.* For every  $n$ , let  $X_n = [0, 1]$ , let  $\mu_n$  be the Lebesgue measure, and let  $R_n$  be an operator defined on  $L^2(X_n)$  by the formula  $(R_n f)(x) = \frac{n-1}{n} f(x) + \frac{1}{n} \int f d\mu_n$ . Let  $X = X_1 \times X_2 \times \dots$ , and let  $\mu$  be the product measure. Define the operator  $R$  first on functions of the form  $f(x) = f_1(x_1) f_2(x_2) \cdots f_n(x_n)$ ,  $x = (x_k)_{k \in \mathbb{N}}$ ,  $f_k \in L^2(\mu_k)$ , as  $Rf = R_1 f_1 R_2 f_2 \cdots R_n f_n$ . Since the functions of this form are linearly dense in  $L^2(\mu)$ ,  $R$  extends to a doubly stochastic operator on all of  $L^2(\mu)$ . Let  $S$  be a doubly stochastic operator on  $L^2(\mu)$  induced by the left shift, that is,  $(Sf)(x_1, x_2, x_3, \dots) = f(x_2, x_3, \dots)$ . Finally, let  $T = SR$ . For any  $k$ , if we take any  $g \in L^2(\mu_k)$  and define  $f \in L^2_0(\mu)$  as  $f(x) = g(x_k)$ , then

$$\begin{aligned} T^m f(x) &= (R_{m+k-1} R_{m+k-2} \cdots R_k g)(x_{m+k-1}) \\ &= \frac{k-1}{m+k-1} g(x_{m+k-1}) + \frac{m}{m+k-1} \int g d\mu_k. \end{aligned} \quad (5.1)$$

Moreover,

$$T^m f(x) = \prod_{k=1}^n T^m f_k(x_{m+k-1})$$

for  $f(x) = f_1(x_1) f_2(x_2) \cdots f_n(x_n)$ ,  $x = (x_k)_{k \in \mathbb{N}}$ ,  $f_k \in L^2(\mu_k)$ . Below we verify the aforementioned properties of  $T$ .

- $T$  is null because the orbit of every function is precompact. To see this, first note that if  $f(x) = f_1(x_1) f_2(x_2) \cdots f_n(x_n)$ ,  $x = (x_k)_{k \in \mathbb{N}}$ , then  $T^m f$  converges to the constant function equal to  $\int f_1 d\mu_1 \cdots \int f_n d\mu_n = \int f d\mu$ . Since such functions are linearly dense in  $L^2(\mu)$ , the same convergence to a constant holds for every function in  $L^2(\mu)$ .

- If  $f$  is any function orthogonal to the constants, then the images of  $f$  under iterations of  $T$  converge to zero, which means that  $T$  is c.n.u. on the space of such functions.

- If  $f \in L^2_0(\mu)$  has the form  $f(x) = g(x_k)$  for  $g \in L^2(\mu_k)$ ,  $x = (x_k)_{k \in \mathbb{N}}$ , then by (5.1) we have  $\|T^m f\|_2 = \frac{k-1}{m+k-1} \|f\|_2$ . Since  $k$  was arbitrary, the norm of  $T^m$

is equal to 1. By the spectral radius formula, we conclude that  $P$  has spectral radius equal to 1.

• Finally, suppose that for some  $\lambda \neq 0$  there is a function  $f \in L_0^2(\mu)$  such that  $Tf = \lambda f$ . If  $g$  is any function dependent only on the first  $m$  coordinates, then since  $T^m(f)$  depends only on coordinates from  $m + 1$  onward,  $T^m f$  and  $g$  are orthogonal. As  $T^m f = \lambda^m f$ , this means that  $f$  is orthogonal to every function dependent only on the first  $m$  coordinates. As  $m$  was arbitrary, it follows that  $f$  is orthogonal to every function dependent only on finitely many coordinates, but since such functions are dense in  $L_0^2(\mu)$ , we conclude that  $f = 0$ .

## 6. Representation theorem

The celebrated theorem of Halmos and von Neumann states that every ergodic measure-preserving dynamical system with discrete spectrum is isomorphic to a Kronecker system, that is, a rotation of a compact Abelian group. Since by Kushnirenko's theorem the discrete spectrum is equivalent to being null, it is natural to search for an analogous representation theorem for null systems in case of doubly stochastic operators. Consider the following example.

*Example 6.1.* Let  $X$  be the annulus  $S^1 \times I$ ,  $I = [0, 1]$ , with the product of the Haar and Lebesgue measures. For  $\alpha \in [0, 1)$  define an operator by

$$Tf(z, x) = \int f(ze^{2\pi i\alpha}, x) d\lambda(x),$$

namely, a product of a rotation by  $\alpha$  and a trivial integral operator. Clearly, it is null with the reversible part  $E_{\text{rev}}$  of the JdLG-decomposition consisting of functions constant on fibers  $\{z\} \times I$ . It is essentially a rotation of a circle with each point of a circle pumped to the unit interval, and the transition probability between these intervals is given by the uniform distribution.

It is easy to modify the above example to see that we cannot expect a representation in a form of a product action. For instance, on the same space  $X = S^1 \times I$  let us define the measure

$$\nu(A) = \lambda \times \lambda((A \cap (N \times I)) + \lambda \times \delta_0(A \cap (N^c \times I)),$$

where  $N = \{z : \arg(z) \in [0, \pi)\}$ . An appropriate modification of  $T$ , namely,

$$Tf(z, x) = \int f(ze^{2\pi i\alpha}, x) d\lambda(x) \cdot \mathbb{1}_{R_\alpha^{-1}N}(z) + f(ze^{2\pi i\alpha}, 0) \cdot \mathbb{1}_{R_\alpha^{-1}N^c}(z),$$

again gives a null operator with the same reversible part  $E_{\text{rev}}$ .

Let  $\Sigma_{\text{rev}} = \{A \in \Sigma : \mathbb{1}_A \in E_{\text{rev}}\}$ . It is known that  $\Sigma_{\text{rev}}$  is a sub- $\sigma$ -algebra of  $\Sigma$  and  $E_{\text{rev}} = L^1(X, \Sigma_{\text{rev}}, \mu)$ .

*Definition 6.2.* Let  $T$  be a Markov operator on  $L^1(\mu)$ .

- $T$  is a *Markov embedding* if it is a lattice homomorphism (i.e.,  $|Tf| = T|f|$  for every  $f \in L^1(\mu)$ ) or, equivalently, there is a Markov operator  $S$  such that  $ST$  is an identity.
- $T$  is a *Markov isomorphism* if it is a surjective Markov embedding.

Markov embeddings preserve lattice operations of taking the maximum or minimum and transform characteristic functions into characteristic functions. Therefore, they are in a natural one-to-one correspondence with homomorphisms of measure algebras, and in the case of Lebesgue spaces, they are just Koopman operators of measure-preserving transformations. Hence, for a Lebesgue space a Markov isomorphism can always be replaced by a point isomorphism of dynamical systems.

**Theorem 6.3.** *An ergodic doubly stochastic operator  $T$  on  $L^1(X, \Sigma, \mu)$  is null if and only the following two conditions are satisfied:*

- (1) *the action of  $T$  on  $E_{\text{rev}} = L^1(X, \Sigma_{\text{rev}}, \mu)$  is Markov isomorphic to a rotation  $R$  of a compact Abelian group  $G$  with Haar measure  $\lambda$ ;*
- (2) *we have*

$$\lim_{n \rightarrow \infty} \|T^n f - T^n E(f | \Sigma_{\text{rev}})\|_1 = 0$$

*for every  $f$ .*

*Furthermore, if  $X$  is a Lebesgue space and  $P_T$  is a transition probability inducing  $T$ , then the Markov isomorphism becomes a point isomorphism of dynamical systems, there is a measure-preserving map  $\pi : X \rightarrow G$  satisfying  $T(g \circ \pi) = g \circ R \circ \pi$  for every  $g \in L^1(G, \lambda)$ , and  $P_T(x, \cdot)$  is supported on  $\pi^{-1}R\pi(x)$ .*

*Proof.* The fact that the action of  $T$  on  $E_{\text{rev}}$  is Markov isomorphic to a rotation of a compact Abelian group is an extension of the Halmos–von Neumann theorem formulated as Theorem 17.6 in [3]. Since  $E(f | \Sigma_{\text{rev}}) \in E_{\text{rev}}$ , the function  $f - E(f | \Sigma_{\text{rev}})$  belongs to  $E_{\text{aws}}$ . If  $h_A(T) = 0$  for every sequence  $A$ , then

$$\lim_{n \rightarrow \infty} \|T^n(f - E(f | \Sigma_{\text{rev}}))\|_1 = 0$$

by Theorem 5.4. Conversely, this condition guarantees that  $h_A(T) = 0$  for every  $A$ , because  $E(f | \Sigma_{\text{rev}}) = 0$  for  $f \in E_{\text{aws}}$ .

The existence of a factor map  $\pi$  in the case of a Lebesgue space follows from the fact that a natural injection  $\Sigma_{\text{rev}} \rightarrow \Sigma$  is a measure algebra homomorphism, hence it comes from a measure-preserving map. Finally, if  $X$  is a Lebesgue space, then for any  $g \in L^1(\mathcal{G}, \lambda)$ ,

$$\int g \circ \pi(y) P_T(x, dy) = T(g \circ \pi)(x) = g(R\pi(x)).$$

Taking  $g = \mathbb{1}_{\{R\pi(x)\}}$ , we obtain that  $P_T(x, \cdot)$  is supported on  $\pi^{-1}R\pi(x)$ .  $\square$

**Acknowledgment.** Both authors' work was partially supported from resources for science as a research project during the years 2013–2018 under Narodowe Centrum Nauki of Poland grant 2013/08/A/ST1/00275.

## References

1. T. Downarowicz, *Entropy in Dynamical Systems*, New Math. Monogr. **18**, Cambridge Univ. Press, Cambridge, 2011. [Zbl 1220.37001](#). [MR2809170](#). [DOI 10.1017/CBO9780511976155.146](#)

2. T. Downarowicz and B. Frej, *Measure-theoretic and topological entropy of operators on function spaces*, Ergodic Theory Dynam. Systems **25** (2005), no. 2, 455–481. [Zbl 1088.47006](#). [MR2129106](#). DOI [10.1017/S014338570400032X](#). [145](#), [146](#), [147](#)
3. T. Eisner, B. Farkas, M. Haasse, and R. Nagel, *Operator Theoretic Aspects of Ergodic Theory*, Grad. Texts in Math. **272**, Springer, Cham, 2015. [Zbl 1353.37002](#). [MR3410920](#). DOI [10.1007/978-3-319-16898-2](#). [149](#), [155](#)
4. R. Ellis, *Distal transformation groups*, Pacific J. Math. **8** (1958), 401–405. [Zbl 0092.39702](#). [MR0101283](#). DOI [10.2140/pjm.1958.8.401](#). [149](#)
5. B. Frej and P. Frej, *The Shannon-McMillan theorem for doubly stochastic operators*, Nonlinearity **25** (2012), no. 12, 3453–3467. [Zbl 1279.37005](#). [MR2997702](#). DOI [10.1088/0951-7715/25/12/3453](#).
6. A. G. Kushnirenko, *Metric invariants of entropy type* (in Russian), Uspekhi Mat. Nauk **22**, no. 5 (1967), 57–66; English translation in Russian Math. Surveys **22** (5) (1967), 53–61. [Zbl 0169.46101](#). [MR0217257](#). [150](#), [151](#)
7. B. Sz.-Nagy and C. Foias, *Harmonic Analysis of Operators on Hilbert Space*, North-Holland, Amsterdam 1970. [Zbl 0201.45003](#). [MR0275190](#). [149](#)
8. V. A. Rokhlin, *Entropy of metric automorphism*, Dokl. Akad. Nauk **124** (1959), 980–983. [Zbl 0096.31405](#). [MR0103258](#). [145](#)
9. A. M. Vershik, *What does a typical Markov operator look like?* (in Russian), Algebra i Analiz **17**, no. 5 (2005), 91–104; English translation in St. Petersburg Math. J. **17** (2006), 763–772. [Zbl 1173.47306](#). [MR2241424](#). DOI [10.1090/S1061-0022-06-00928-9](#). [145](#)

FACULTY OF PURE AND APPLIED MATHEMATICS, WROCLAW UNIVERSITY OF SCIENCE AND TECHNOLOGY, WYBRZEŻE WYSPIAŃSKIEGO 27, 50-370 WROCLAW, POLAND.

*E-mail address:* [Bartosz.Frej@pwr.edu.pl](mailto:Bartosz.Frej@pwr.edu.pl); [Dawid.Huczek@pwr.edu.pl](mailto:Dawid.Huczek@pwr.edu.pl)