Exploding Markov operators

Bartosz Frej¹

Received: 28 December 2018 / Accepted: 5 February 2020 / Published online: 15 February 2020
© The Author(s) 2020

Abstract

A special class of doubly stochastic (Markov) operators is constructed. In a sense 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.

Keywords Markov operator · Doubly stochastic operator · Measure-preserving transformation · Entropy · Ergodicity · Koopman operator

Mathematics Subject Classification 28D20 · 37A30 · 47A35

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 $P : L^p(\mu) \rightarrow L^p(\nu)$, where (X, μ) and (Y, ν) are probability spaces, which fulfills the following conditions:

- (i) Pf is positive for every positive $f \in L^p(\mu)$,
- (ii) $P\mathbb{1} = \mathbb{1}$ (where $\mathbb{1}$ 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

Research is supported from resources for science in years 2013–2018 as research project (NCN Grant 2013/08/A/ST1/00275, Poland).

✉ Bartosz Frej
Bartosz.Frej@pwr.edu.pl

¹ Faculty of Pure and Applied Mathematics, Wrocław University of Science and Technology, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland

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 [7] 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 examples which stem from pointwise maps is proposed and some of their properties are investigated.

2 The definition

Let us start with 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 Sect. 3 for definitions).

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

For each finite block $B = (B_1, B_2, \dots, B_k) \in \{0, 1\}^k$ let us define the map $\sigma_B : X \rightarrow X$ 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:

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

In other words, the operator is obtained by integrating with respect to the transition probability defined by:

$$\begin{aligned} P_T((x, 1), A) &= \sum_{k=1}^{\infty} 2^{-k} \sum_{B \in \{0, 1\}^k} 2^{-k} \mathbb{1}_A(\sigma_B x, k) \\ P_T((x, k), A) &= \mathbb{1}_A(\sigma x, k-1) = \delta_{(\sigma x, k-1)}(A) \quad \text{for } k \geq 2, \end{aligned}$$

where δ_y be the Dirac measure concentrated in y , that is, $\delta_y(A) = \mathbb{1}_A(y)$.

One may visualize the action of P 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 shift 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 some set of cofinal points, which share the common future with x) with counters reset to k with probability b_k .

Aiming to generalize the above example we notice that its definition may easily be rewritten with use of the adjoint operator. Indeed, if $T_\sigma f = f \circ \sigma$ is the Koopman operator of the shift, then the adjoint is given by

$$T_\sigma^* f(x) = \frac{1}{2} \int f(0x) d\mu(x) + \frac{1}{2} \int f(1x) d\mu(x),$$

where ax denotes x shifted to the right and completed by symbol a at the first coordinate. Thus, writing $f_k(\cdot) = f(\cdot, k)$ we have

$$\sum_{B \in \{0,1\}^k} 2^{-k} f(\sigma_B x, k) = T_\sigma^{k+1} (T_\sigma^*)^k f_k(x).$$

We arrive at the following general definition.

Let $T : L^1(X, \mu) \rightarrow L^1(X, \mu)$ be a Markov embedding (or a formal Koopman operator), i.e., a doubly stochastic operator satisfying $|Tf| = T|f|$ for every $f \in L^1(\mu)$ (a lattice homomorphism)—in particular, it may be a Koopman operator of a measure preserving transformation. Having in mind that a doubly stochastic operator defined on $L^P(\mu)$ is uniquely extended to the whole $L^1(\mu)$ we denote by $T^* : L^1(\mu) \rightarrow L^1(\mu)$ the operator adjoint to T . Let:

$$\begin{aligned} P_k &= T^k T^{*k} \\ E_k &= P_k T = T P_{k-1} \end{aligned}$$

Clearly, P_k and E_k are doubly stochastic and $E_1 = T T^* T = T$. Let (a_n) be a strictly decreasing sequence, such that $\sum_{n=1}^{\infty} a_n = 1$. 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). We denote by $\mathcal{V}$ the direct sum $l^1 \oplus_{n=1}^{\infty} L^1(X, a_n \mu)$, i.e., the Cartesian product of countably many copies of $L^1(X, \mu)$ with norm of $f = (f_1, f_2, \dots)$ given by

$$\|f\| = \sum_{k=1}^{\infty} a_n \|f_k\|_1.$$

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 σ -algebra and let $\nu = \mu \times m$. Clearly, the space $\mathcal{V}$ is isomorphic to $L^1(X \times \mathbb{N}, \mu \times m)$. We will freely use both the representation of elements of $\mathcal{V}$ as a sequence (f_n) and the element f of $L^1(X \times \mathbb{N}, \mu \times m)$, identifying $f_n(\cdot)$ with $f(\cdot, n)$. It is convenient to define an operator J embedding $L^1(X)$ in $\mathcal{V}$ by

$$Jf = (f, f, f, \dots).$$

We will also say that an operator P on the direct sum is doubly stochastic if: $(Pf)_n \geq 0$ if all coordinates of $f = (f_1, f_2, \dots)$ are positive, $P(\mathbb{1}, \mathbb{1}, \dots) = (\mathbb{1}, \mathbb{1}, \dots)$ and $\sum_{k=1}^{\infty} a_n \int (Pf)_k d\mu = \sum_{k=1}^{\infty} a_n \int f_k d\mu$ —this is consistent with the definition for operators on $L^1(X \times \mathbb{N})$.

Definition 1 An *exploding operator induced by T* is an operator $\mathcal{Y}_T : \mathcal{V} \rightarrow \mathcal{V}$ defined for every $f = (f_n)$ by the formula

$$\mathcal{Y}_T f = (V_T f, T f_1, T f_2, \dots), \quad (1)$$

where $V_T : \mathcal{V} \rightarrow L^1(X, \mu)$ is given by

$$V_T f = \sum_{k=1}^{\infty} b_k E_k f_k.$$

We notice that $\mathcal{Y}_T = \hat{T}U$, where $U(f_1, f_2, \dots) = (\sum_{k=1}^{\infty} b_k P_{k-1} f_k, f_1, f_2, \dots)$ and $\hat{T}(f_1, f_2, \dots) = (T f_1, T f_2, \dots)$. By a straightforward calculation one easily verifies that both these operators are doubly stochastic, and, consequently, so is $\mathcal{Y}_T$.

Remark 1 If T is a Koopman operator of a measure preserving map, then T^* is the Perron-Frobenius operator of this map.

Remark 2 (Topological model) By passing to a faithful topological model, e.g., the Stone model (see [5, Chapter 12]), one may without loss of generality assume that X is a compact Hausdorff space, μ is a regular Borel probability measure with full support and T is induced by a continuous surjective map $\varphi : X \rightarrow X$, which preserves the measure μ . Each element of $L^1(\mu)$ contains a unique continuous function, so both T and its adjoint T^* may be treated as operators on $C(X)$. Therefore, the Riesz representation theorem allows to define for each $k \in \mathbb{N}$ and $y \in X$ a conditional measure μ_u^k on X by the formula

$$\int f d\mu_y^k = (T^{*k} f)(y).$$

It is easy to verify that the measure is supported on the closed set $\varphi^{-k}(\{y\})$. We have

$$\int f d\mu_{\varphi^k(x)}^k = (T^{*k}f)(\varphi^k(x)) = (T^k T^{*k}f)(x) = P_k f(x)$$

$$\text{and } E_k f(x) = \int f d\mu_{\varphi^k(x)}^{k-1} = \int T f d\mu_{\varphi^k(x)}^k.$$

The space $Y = X \times \mathbb{N}$ is a locally compact Hausdorff space and $C_0(X)$ is isomorphic to a direct sum $c_0\text{-}\bigoplus_{k=1}^{\infty} C(X)$. We then have

$$\gamma_T f(x, n) = \begin{cases} \sum_{k=1}^{\infty} b_k E_k f_k(x) = \sum_{k=1}^{\infty} b_k \int T f_k d\mu_{\varphi^k(x)}^k & \text{for } n = 1 \\ T f_{n-1}(x) = f_{n-1}(\varphi(x)) & \text{for } n \geq 2 \end{cases}$$

for $f = (f_n)$. This means that γ_T is induced by a transition kernel given by

$$P_T((x, n), \cdot) = \begin{cases} \sum_{k=1}^{\infty} b_k \varphi^* \mu_{\varphi^k(x)}^k & \text{for } n = 1 \\ \delta_{(\varphi(x), n-1)} & \text{for } n \geq 2 \end{cases},$$

where φ^* is the map adjoint to φ , acting on regular Borel measures.

Remark 3 For a standard probability space (X, Σ, μ) and a measure preserving surjection $T : X \rightarrow X$ the operator may also be defined via the notion of a disintegration of a measure. For a fixed $k \in \mathbb{N}$ let ξ_k be a partition of X into sets $T^{-k}\{T^k x\}$, $x \in X$, and denote by $\xi_k(x)$ an element of the partition ξ_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/ξ_k is countably separated, so X/ξ_k is also a standard probability space with measure transported by the map $x \mapsto \xi_k(x)$. Indeed, let $\{B_1, B_2, \dots\}$ 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, $\xi_k(z) \subset T^{-k}B_i$ if $T^k z \in B_i$ and $\xi_k(z) \cap T^{-k}B_i = \emptyset$ otherwise, so the collection $\{T^{-k}B_i : i \in \mathbb{N}\}$ separates elements of ξ_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 $\{T^{-k}B_1, T^{-k}B_2, \dots\}$ separates elements of the partition ξ_k .

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

$$\int f d\mu = \int_{X/\xi_k} \left(\int_C f(x) d\mu_C^k(x) \right) d\hat{\mu}_k(C)$$

(see [8]). In addition, the map $C \mapsto \mu_C^k$ is measurable in the sense that for each measurable A the map $x \mapsto \mu_{\xi_k(x)}^k(A)$ is measurable. Let sections of a set $A \subset Y$ and of a function f on Y be denoted by

$$A_k = \{x \in X : (x, k) \in A\} \quad \text{and} \quad f_k(x) = f(x, k).$$

We have

$$E_k f(x) = \int f \circ T d\mu_{\xi_k(x)}^k$$

and

$$\mathcal{Y}_T f(x, k) = \begin{cases} \sum_{k=1}^{\infty} b_k E_k f_k(x) & \text{for } k = 1 \\ f(Tx, k-1) & \text{for } k \geq 2 \end{cases}$$

or, in other words, $\mathcal{Y}_T f(y) = \int f(u) P_T(y, du)$, where the probability kernel P_T is given by:

$$\begin{aligned} P_T((x, 1), A) &= \sum_{k=1}^{\infty} b_k \mu_{\xi_k(x)}^k(T^{-1}A_k) \\ P_T((x, k), A) &= \delta_{(Tx, k-1)}(A) \quad \text{for } k \geq 2. \end{aligned}$$

Remark 4 The construction is particularly simple for Koopman operators of invertible maps, say $S : X \rightarrow X$. Then, $V_T f = \sum_{k=1}^{\infty} b_k T f_k$ or, in other words, if $\mathcal{Y}_T$ is given by a transition kernel P_T , then P_T sends the point $(x, 1)$ to (Sx, k) with probability b_k .

3 Pointwise factors

We begin with the definition of a factor of a doubly stochastic operator.

- Definition 2** 1. A *unital sublattice* of $L^1(X, \mu)$ is a closed linear subspace of $L^1(X, \mu)$ which contains the constant $\mathbb{1}$ and together with each f it contains its conjugate $\bar{f}$ and its absolute value $|f|$.
2. 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 (see [5], Definition 13.26).

This definition identifies factors as certain subspaces of the domain. It can be shown these subspaces have the form $L^1(X, \Sigma_F, \mu)$, where Σ_F is a sub- σ -algebra of the σ -algebra of all measurable subsets of X , which is invariant in this sense that $P\mathbb{1}_A$ is Σ_F -measurable for each $A \in \Sigma_F$. Moreover, the representation is unique if one assumes that Σ_F is complete with respect to the measure μ . 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 2—the appropriate sublattice is obtained as the image $U(L^1(\mu_2))$.

In the 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 $\pi : X_1 \rightarrow X_2$ satisfying, for every $f \in L^1(\mu_2)$, the condition $(P_2 f) \circ \pi = P_1(f \circ \pi)$. 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 Π is always induced by a pointwise measure preserving map $\pi : X_1 \rightarrow X_2$ 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 $F \subset L^1(X, \mu)$ of an operator P is *pointwise* if the restriction $P|_F$ is a Markov embedding. If $F = L^1(X, \Sigma_F, \mu)$ and Σ_F is a complete σ -algebra then

$$A \in \Sigma_F \Rightarrow \exists B \in \Sigma_F \ P \mathbb{1}_A = \mathbb{1}_B.$$

A factor is *trivial* if it consists only of constant functions or, in other words, if it can be represented as a subspace $L^1(X, \Sigma_F, \mu)$, where Σ_F consists solely of sets of measure zero or one. Obviously, a trivial factor is pointwise.

For a Markov embedding $T : L^1(X, \mu) \rightarrow L^1(X, \mu)$ let us consider a unital sublattice

$$R_T = \bigcap_{n \in \mathbb{N}} \text{ran}(T^n).$$

Note that T is an isometry, so $\text{ran}(T^n)$ is closed for every n . The sequence $(\text{ran}(T^n))$ is descending, so R_T is invariant, thus it is a factor of T . The restriction $T|_{R_T}$ is invertible, the restriction of the adjoint being the inverse. Moreover, R_T is also a factor of $\mathcal{Y}_T$ in the sense that $JTf = \mathcal{Y}_T Jf$ for every $f \in R_T$, so $J(R_T)$ is a unital sublattice of $\mathcal{V} \cong L^1(Y)$, invariant under the action of $\mathcal{Y}_T$. Since T is a Markov embedding, it is a pointwise factor of $\mathcal{Y}_T$.

Lemma 1 *If F is a pointwise factor of $\mathcal{Y}_T$ then $V_T f \in R_T$ for every $f \in F$.*

Proof Let $f = \mathbb{1}_A \in F$ and let $(\mathbb{1}_{A_1}, \mathbb{1}_{A_2}, \dots)$ be its direct sum representation. Then

$$V_T \mathbb{1}_A = \sum_{k=1}^{\infty} b_k E_k \mathbb{1}_{A_k} = \mathbb{1}_B$$

for some measurable $B \subset X$. Since $0 \leq E_k \mathbb{1}_{A_k} \leq 1$, while b_k are positive and sum up to 1, $E_k \mathbb{1}_{A_k} = \mathbb{1}_B$ for each k . But then $V_T \mathbb{1}_A = T^k (T^*)^{k-1} \mathbb{1}_{A_k}$, so $V_T \mathbb{1}_A \in \text{ran}(T^k)$ for each $k \in \mathbb{N}$. By linearity and continuity of V_T , one has $V_T f \in R_T$ for all $f \in F$. $\square$

Theorem 1 *The exploding operator $\mathcal{Y}_T$ has no non-trivial pointwise factor if and only if R_T is trivial.*

Proof If R_T contains a non-constant function f then it is a non-trivial pointwise factor.

On the other hand, let F be a pointwise factor of $\mathcal{Y}_T$ and let $\mathbb{1}_A \in F$. If R_T is trivial then $V_T \mathbb{1}_A$ is constant, by the above lemma. But then it is equal to zero or to one,

because the restriction $\gamma_T|_F$ is a Markov embedding. If it is zero, representing $\mathbb{1}_A$ by $(\mathbb{1}_{A_1}, \mathbb{1}_{A_2}, \dots)$ we obtain

$$\begin{aligned} \sum_{k=1}^{\infty} a_k \int \mathbb{1}_{A_k} d\mu &= \int \mathbb{1}_A d\nu = \int \gamma_T \mathbb{1}_A d\nu = \int_{X \times \{2,3,\dots\}} \gamma_T \mathbb{1}_A d\nu \\ &= \sum_{k=2}^{\infty} a_k \int T \mathbb{1}_{A_{k-1}} d\mu = \sum_{k=1}^{\infty} a_{k+1} \int \mathbb{1}_{A_k} d\mu. \end{aligned}$$

Thus, $\sum_{k=1}^{\infty} (a_k - a_{k+1}) \int \mathbb{1}_{A_k} d\mu = 0$, so the fact that the sequence (a_k) is strictly decreasing implies that $\int \mathbb{1}_{A_k} d\mu = 0$ for all k , which means that $\mathbb{1}_A$ is constantly zero. Performing similar calculations for the second case, we obtain

$$1 = \int V_T \mathbb{1}_A d\mu = \sum_{k=1}^{\infty} b_k \int \mathbb{1}_{A_k} d\mu,$$

which happens only if $\int \mathbb{1}_{A_k} d\mu = 1$ for all k . Hence, $\mathbb{1}_A \equiv 1$ and, consequently, F consists only of constant functions. $\square$

Corollary 1 *If T is a Koopman operator of a one-sided (noninvertible) Bernoulli shift then γ_T is strictly non-pointwise.*

Proof Any characteristic function $\mathbb{1}_A$ which belongs to R_T is measurable with respect to the tail σ -algebra. Therefore, A has measure zero or one, by Kolmogorov's zero-one law, so R_T is trivial. $\square$

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 2 *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. For any functions g and h denote by $g \vee h$ the function being the pointwise maximum of g and h . Since

$$P(f \vee \int f d\mu) \geq Pf \vee \int f d\mu = f \vee \int f d\mu$$

and both functions have the same integrals, one has $P(f \vee \int f d\mu) = f \vee \int f d\mu$. The function $g = (f \vee \int f d\mu) - \int f d\mu$ is a positive invariant function, which is zero on a set of positive measure. But then $\sum_{k=1}^{\infty} P^k g$ is zero on the same set, which contradicts our assumption. $\square$

We will say that an element $(f_1, f_2, f_3, \dots)$ of $\mathcal{V}$ is constant if the corresponding function $f \in L^1(X \times \mathbb{N})$ is constant, i.e., $f_1 = f_2 = \dots = \text{const}$. With this convention we can study ergodicity of Υ_T using its direct sum representation.

Theorem 3 *The operator Υ_T is ergodic if and only if the Markov embedding T is ergodic.*

Proof Assume that T is not ergodic and let f be a non-constant invariant function. It is also invariant with respect to the adjoint T^* and, consequently, $E_k f = f$. Then,

$$V_T Jf = \sum_{k=1}^{\infty} b_k E_k f = f$$

and

$$\Upsilon_T Jf = (V_T Jf, Tf, Tf, \dots) = (f, f, f, \dots) = Jf,$$

so Jf is a non-constant Υ_T -invariant function.

For the converse statement, let T be ergodic and let f be a nonnegative function on $X \times \mathbb{N}$, which is not equal almost everywhere to zero. Representing f by $(f_1, f_2, \dots)$, we obtain that $(\Upsilon_T^k f)_1$, the first coordinate of $\Upsilon_T^k f$, is bounded from below by $b_1^k T^k(f_1)$ for every $k \in \mathbb{N}$. Indeed,

$$(\Upsilon_T f)_1 = V_T f \geq b_1 E_1 f_1 = b_1 T(f_1).$$

Assuming inductively that $(\Upsilon_T^k f)_1 \geq b_1^k T^k(f_1)$ we get

$$(\Upsilon_T^{k+1} f)_1 = V_T \Upsilon_T^k f \geq b_1 E_1 (\Upsilon_T^k f)_1 \geq b_1^{k+1} T^{k+1}(f_1).$$

Since T is ergodic, $\sum_{k=1}^{\infty} T^k(f_1)$ is strictly positive if only f_1 is not identically equal to zero, hence so is $\sum_{k=1}^{\infty} b_1^k T^k(f_1)$, implying that $\sum_{k=1}^{\infty} (\Upsilon_T^k f)_1$ is strictly positive. For $n \geq 2$ and $k \geq n-1$ we have

$$(\Upsilon_T^k f)_n = T(\Upsilon_T^{k-1} f)_{n-1} = T^{n-1}(\Upsilon_T^{k-n+1} f)_1,$$

so $\sum_{k=1}^{\infty} (\mathcal{Y}_T^k f)_n \geq \sum_{k=n}^{\infty} (\mathcal{Y}_T^k f)_n = T^{n-1} \sum_{k=1}^{\infty} (\mathcal{Y}_T^k f)_1$, which is positive, because $(\mathcal{Y}_T^k f)_1$ is not identically equal to zero and T is a Markov embedding. Hence, $\sum_{k=1}^{\infty} \mathcal{Y}_T^k f$ is positive.

If f_1 is identically equal to zero, but f is not, there exists k such that f_k does not vanish. Then $(\mathcal{Y}_T f)_1 = V_T f \geq b_k E_k f_k$ and the last function does not vanish. By the first part of the argument we obtain $\sum_{k=1}^{\infty} \mathcal{Y}_T^k f \geq \sum_{k=1}^{\infty} \mathcal{Y}_T^k \mathcal{Y}_T f > 0$, so $\mathcal{Y}_T$ is ergodic. $\square$

5 Entropy

We will prove that the entropy of $\mathcal{Y}_T$ is bounded from below by the entropy of T . The definition of entropy of a doubly stochastic operator is not widely known, so I will devote the next few lines for a short introduction to the subject—a detailed exposition may be found in [2] or [3] and an alternative approach in [6]. Similarly to the classical case of the Kolmogorov-Sinai invariant, the entropy of a doubly stochastic operator on $L^1(Y, \nu)$ is defined in several steps. First, the entropy $H_\nu(\mathcal{F})$ of a finite collection $\mathcal{F}$ of measurable functions with range contained in $[0, 1]$ is defined (such collections replace partitions in operator-theoretic definition). Furthermore, an operation of joining such collections is introduced. For instance, one can define the join of collections $\mathcal{F}$ and $\mathcal{G}$ as a union of $\mathcal{F}$ and $\mathcal{G}$ (or as their concatenation if collections are interpreted as finite sequences of sets). Then, the entropy $h_\nu(P, \mathcal{F})$ of an operator P with respect to a collection $\mathcal{F}$ is obtained as an upper limit (or a limit, if it exists) $\limsup_{n \rightarrow \infty} \frac{1}{n} H_\nu(\mathcal{F}^n)$, where $\mathcal{F}^n$ stands for the join of $\mathcal{F}, P\mathcal{F}, \dots, P^{n-1}\mathcal{F}$ and $P^k \mathcal{F} = \{P^k f : f \in \mathcal{F}\}$. Finally, $h_\nu(\mathcal{F})$ is the supremum $\sup_{\mathcal{F}} h_\nu(P, \mathcal{F})$ over all collections under consideration. It was proved in [3] that any specification of the joining operation and the ‘static’ entropy $H_\nu(\mathcal{F})$, which satisfies certain set of axioms, leads to a common value of the final notion $h_\nu(P)$. In addition, the conditional entropy of a collection $\mathcal{F}$ with respect to $\mathcal{G}$ is defined as

$$H_\nu(\mathcal{F}|\mathcal{G}) = H_\nu(\mathcal{F} \vee \mathcal{G}) - H_\nu(\mathcal{G}).$$

The explicit formula for $H_\nu(\mathcal{F})$ will not be needed in the current paper, but I will recall some of the properties of operator entropy, which I use in the proof of Theorem 4. Below, there is a list of some of the properties of entropy which can be found in [3].

- i. Let ξ and ξ' be partitions of Y and let $\mathbb{1}_\xi = \{\mathbb{1}_A : A \in \xi\}$, $\mathbb{1}_{\xi'} = \{\mathbb{1}_A : A \in \xi'\}$. Then $H_\nu(\xi) = H_\nu(\mathbb{1}_\xi)$ and $H_\nu(\xi \vee \xi') = H_\nu(\mathbb{1}_\xi \vee \mathbb{1}_{\xi'})$, where the expressions on the left hand sides denote Shannon entropies of partitions.
- ii. For any finite collections $\mathcal{F}$ and $\mathcal{G}$ one has $H_\nu(\mathcal{F} \vee \mathcal{G}) \leq H_\nu(\mathcal{F}) + H_\nu(\mathcal{G})$.
- iii. For any finite collections $\mathcal{F}_1, \dots, \mathcal{F}_n$ and $\mathcal{G}_1, \dots, \mathcal{G}_n$ one has

$$H_\nu\left(\bigvee_{i=1}^n \mathcal{F}_i \mid \bigvee_{i=1}^n \mathcal{G}_i\right) \leq \sum_{i=1}^n H_\nu(\mathcal{F}_i | \mathcal{G}_i)$$

- iv. Entropy $H_v(\mathcal{F})$ is continuous with respect to $\mathcal{F}$ in the following sense. For two collections $\mathcal{F}$ and $\mathcal{G}$ both having r elements, $\mathcal{F} = \{f_1, \dots, f_r\}$, $\mathcal{G} = \{g_1, \dots, g_r\}$, one defines their L^1 -distance $\text{dist}(\mathcal{F}, \mathcal{G})$ by the 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 π of a set $\{1, 2, \dots, r\}$. For every $\varepsilon > 0$ and $r \in \mathbb{N}$ there is $\delta > 0$ such that if $\mathcal{F}$ and $\mathcal{G}$ have cardinalities at most r and $\text{dist}(\mathcal{F}, \mathcal{G}) < \delta$ then $|H_v(\mathcal{F}|\mathcal{G})| < \varepsilon$.

We turn to the proof of the aforementioned result.

Lemma 2 *For every measurable function $f \in L^1(X, \mu)$ the sequence*

$$\sup_{n \in \mathbb{N}} \left\| \gamma_T^n(JT^i f) - JT^{i+n} f \right\|_1$$

converges to 0, when i goes to infinity.

Proof Let i be a positive integer. First, we evaluate the above norm for $n = 1$. Note that if $1 \leq k \leq i$ then $E_k T^i f = T^k (T^*)^{k-1} T^i f = T^{i+1} f$, hence

$$\begin{aligned} \left\| V_T(T^i f) - T^{i+1} f \right\|_1 &= \left\| \sum_{k=1}^{\infty} b_k E_k T^i f - T^{i+1} f \right\|_1 \\ &= \left\| \sum_{k=i+1}^{\infty} b_k (E_k T^i f - T^{i+1} f) \right\|_1 \leq \sum_{k=i+1}^{\infty} b_k \cdot 2 \|f\|_1 = \frac{2a_{i+1}}{a_1} \|f\|_1. \end{aligned}$$

Since $(\gamma_T J T^i f)_k = T^{i+1} f$ for $k \geq 2$, using the above inequality one gets

$$\left\| \gamma_T J T^i f - J T^{i+1} f \right\|_1 \leq \frac{2a_{i+1}}{a_1} \|f\|_1$$

for all i . Assume inductively that

$$\left\| \gamma_T^n J T^i f - J T^{i+n} f \right\|_1 \leq \frac{2\|f\|_1}{a_1} \sum_{j=1}^n a_{i+j}$$

for some n . Since Υ_T is doubly stochastic, one has

$$\begin{aligned} & \left\| \Upsilon_T^{n+1} J T^i f - J T^{i+n+1} f \right\|_1 \\ & \leq \left\| \Upsilon_T^{n+1} J T^i f - \Upsilon_T J T^{i+n} f \right\|_1 + \left\| \Upsilon_T J T^{i+n} f - J T^{i+n+1} f \right\|_1 \\ & \leq \left\| \Upsilon_T^n J T^i f - J T^{i+n} f \right\|_1 + \frac{2a_{i+n+1}}{a_1} \|f\|_1 \\ & \leq \frac{2\|f\|_1}{a_1} \sum_{j=1}^{n+1} a_{i+j}. \end{aligned}$$

Hence, $\left\| \Upsilon_T^n J T^i f - J T^{i+n} f \right\|_1 \leq \frac{2\|f\|_1}{a_1} \sum_{j=i+1}^{\infty} a_j$ for every n , which ends the proof. $\square$

Theorem 4 *If T is a Markov embedding then*

$$h_\mu(T) \leq h_\nu(\Upsilon_T).$$

Proof Fix a collection $\mathcal{F}$ and $\varepsilon > 0$. Denote $J\mathcal{F} = \{Jf : f \in \mathcal{F}\}$. Given $i \in \mathbb{N}$ one calculates:

$$\begin{aligned} H_\mu \left(\bigvee_{n=0}^N T^{i+n} \mathcal{F} \right) &= H_\nu \left(\bigvee_{n=0}^N (T^{i+n} J\mathcal{F}) \right) \\ &\leq H_\nu \left(\bigvee_{n=0}^N \Upsilon_T^n (T^i J\mathcal{F}) \right) \\ &\quad + H_\nu \left(\bigvee_{n=0}^N (T^{i+n} J\mathcal{F}) \middle| \bigvee_{n=0}^N \Upsilon_T^n (T^i J\mathcal{F}) \right) \\ &\leq H_\nu \left(\bigvee_{n=0}^N \Upsilon_T^n (T^i J\mathcal{F}) \right) \\ &\quad + \sum_{n=0}^N H_\nu \left((T^{i+n} J\mathcal{F}) \middle| \Upsilon_T^n (T^i J\mathcal{F}) \right). \end{aligned}$$

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

$$h_\mu(T, \mathcal{F}) \leq h_\nu(\Upsilon_T, T^i J\mathcal{F}) + \varepsilon \leq h_\nu(\Upsilon_T) + \varepsilon$$

and the inequality follows by taking supremum over $\mathcal{F}$ and infimum over ε . $\square$

Final remarks

1. It seems very unlikely that the entropy of $\mathcal{Y}_T$ could ever be strictly higher than the entropy of T , but I was not able to prove the equality. However, my conjecture is that the equality holds at least if T is the Koopman operator of a Bernoulli shift.
2. By passing to a topological model one gains possibility to define exploding operators in a purely topological setup. However, the underlying space $Y = X \times \mathbb{N}$ is not compact, so to study the entropy of such operators one either needs to extend entropy theory introduced in [3] beyond compact spaces or to define the operator on some compactification of Y .

Acknowledgements I would like to thank the anonymous referee for his excellent comments on the previous version of the manuscript. According to his advice, the presentation changed from measure theoretic deliberations on the notion of disintegration of measure (as in Remark 3) to much more convenient and, in fact, more general language of operator theory.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Cornfeld, I.P., Fomin, S.V., Sinai, Ya.G.: Ergodic Theory. Springer, New York (1982)
2. Downarowicz, T.: Entropy in Dynamical Systems. New Mathematical Monographs, vol. 18. Cambridge University Press, Cambridge (2011)
3. Downarowicz, T., Frej, B.: Measure-theoretic and topological entropy of operators on function spaces. *Ergod. Theory Dyn. Syst.* **25**(2), 455–481 (2005)
4. Dugundji, J.: Topology. Allyn and Bacon Inc., Boston (1966)
5. Eisner, T., Farkas, B., Haase, M., Nagel, R.: Operator Theoretic Aspects of Ergodic Theory. Graduate Texts in Mathematics. Springer, New York (2015)
6. Frej, B., Frej, P.: An integral formula for entropy of doubly stochastic operators. *Fund. Math.* **213**, 271–89 (2011)
7. Frej, B., Huczek, D.: Doubly stochastic operators with zero entropy. *Ann. Funct. Anal.* **10**(1), 144–156 (2019)
8. Glasner, E.: Ergodic Theory via Joinings. Mathematical Surveys and Monographs, vol. 101. American Mathematical Society, Providence (2003)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.