# ON A THEOREM OF FURSTENBERG AND THE
# STRUCTURE OF TOPOLOGICALLY ERGODIC MEASURES

LEWIS PAKULA AND ROBERT SINE

ABSTRACT. An almost everywhere convergence theorem for topologically ergodic measures stated by Furstenberg for homeomorphisms is extended to Markov operators on $C(X)$ with compact Hausdörff state space. A structure theorem for topologically ergodic measures is obtained in the compact metric case again in the more general setting of Markov operators.

1. **Introduction.** Let $X$ be a compact Hausdörff space and let $T$ be a bounded linear operator on $C(X)$ satisfying $T1 = 1$ and $f \geqslant 0$ implies $Tf \geqslant 0$. Such an operator is called a *Markov operator*; this class of operators is a natural and nontrivial extension of the class of operators induced by homeomorphisms of $X$. Many of the concepts of topological dynamics extend to Markov operators with no difficulty. Thus we say $\mu$ is an *invariant probability* if $T^*\mu = \mu$ (where $\mu$ is a regular Borel positive normalized measure on $X$). A closed set $D$ is called *invariant* if for each $x$ in $D$ we have supp $T^*\delta(x)$ contained in $D$. We will always use the word support to mean that unique closed set which is the complement of the union of all open sets annihilated by the probability. If a set $B$ is measurable and $\mu(B) = 1$, we will say $\mu$ is *carried* by $B$. The following properties are well known and not difficult (see [7]). The support of an invariant probability is an invariant set and each nonempty invariant set carries at least one invariant probability.

We will have need of the notion of the *invariant core* of a closed set. For this let $D$ be an arbitrary closed set. Then we define the core by

$$D^* = \left\{ x \in D: \text{supp } T^{*n}\delta(x) \subset D \text{ for all } n \right\}.$$

Then $D^*$ is a closed and invariant subset of $D$ but may very well be empty. For a discussion of invariant cores with the proofs of the above assertions see [8] or [5] (where the term "self-supporting part") is employed.

2. **Topologically ergodic probabilities.** Our definition is taken from Furstenberg [3, p. 21].

DEFINITION. An invariant probability $\lambda$ for $(X, T)$ is *topologically ergodic* if $\lambda(D) = 0$ or 1 for all closed invariant sets $D$.

$T$ extends to a bounded linear operator on the space $B$ of bounded

measurable functions on $X$. Recall that an invariant probability $\lambda$ is ergodic iff $Tf = f$ [a.e. $\lambda$] for $f$ in $B$ implies $f$ is constant [a.e. $\lambda$]. We have a similar characterization for topologically ergodic probabilities.

THEOREM 1. *An invariant probability $\lambda$ is topologically ergodic iff for $f$ bounded lower-semicontinuous, $Tf = f$ [a.e. $\lambda$] implies $f$ is constant [a.e. $\lambda$].*

PROOF. First, suppose $f$ is bounded l.s.c. with $Tf = f$ [a.e. $\lambda$] for the invariant probability $\lambda$. If $f$ is not essentially constant [a.e. $\lambda$] then there is a lower section set $D = \{x: f(x) \leqslant d\}$ so that $0 < \lambda(D) < 1$. Since $f$ was l.s.c. we know $D$ is a closed set. The function $g = f \wedge d$ is also l.s.c. and satisfies $Tg \leqslant g$ [a.e. $\lambda$]. But if we integrate this last inequality with respect to the invariant probability $\lambda$ we obtain $Tg = g$ [a.e. $\lambda$]. Now in general the core of a closed set may be empty, but we claim here that $\lambda(D) = \lambda(D^*)$. Let

$$Y = \{x: T^n f(x) = f(x), T^n g(x) = g(x) \text{ for all } n\}.$$

Then $\lambda(Y) = 1$ and $f > g$ on $D^c$, $f = g$ on $D$ and $\int (f - g)\, dT^{*n}\delta(x) = 0$ for $x$ in $D \cap Y$ imply that $T^{*n}\delta(x)(D^c) = 0$ for $x$ in $D \cap Y$. Hence $D \cap Y \subset D^* \subset D$ and $\lambda(D \cap Y) = \lambda(D)$ gives $\lambda(D) = \lambda(D^*)$. Thus since $D^*$ is closed and invariant with $0 < \lambda(D^*) < 1$, we can conclude that $\lambda$ is not topologically ergodic.

Conversely, suppose $\lambda$ is not topologically ergodic. Then there is a closed invariant set $D$ with $0 < \lambda(D) < 1$. Let $f$ be the indicator function of the open set $U = X \setminus D$. By invariance of $D$, $f$ is subinvariant ($Tf \leqslant f$), so integrating with respect to $\lambda$ we get $Tf = f$ [a.e. $\lambda$]. Thus $f$ is a bounded l.s.c. function which is $\lambda$-invariant but not $\lambda$-constant. This finishes the proof.

Now we turn to another useful construct, namely the hitting probability or *Brunel function*. Given the indicator function $f$ of a Borel set $A$ (or more generally, given a bounded Borel function $f$) the Brunel function $i(f)$ is defined to be the minimal subinvariant majorant of $f$. Given a bounded Borel function $f$ then $i(f)$ exists and is Borel. If $f$ is l.s.c. then $i(f)$ is also l.s.c. When $f = 1_A$ then we use the notation $i(A)$. We have

$i(A)(x) = $ probability of the process ever hitting $A$ given that it starts at $x$.

It is important to point out that this includes the hitting of $A$ at time $n = 0$. Thus $i(A) \geqslant 1_A$ for all Borel sets $A$. These envelope functions were introduced by Brunel in his work on $L_1$ contraction ergodic theorem (see [4]). Detailed proofs of existence and other useful properties of $i(A)$ together with applications when the state space is a topological space can be found in [2] and [8].

A probability measure $\mu$ on $X$ is called *diffuse* if the support of $\mu$ is $X$. (Note $\mu$ may be atomic and still be diffuse.) If $\mu$ is an invariant probability then $Y = \text{supp } \mu$ is a closed invariant set and the process may be restricted to new state space $Y$. Since there is nothing we can say about the process with respect to $\mu$ off of the support of $\mu$ we make the assumption that the normalization of restricting the process to supp $\mu$ is always made. This

convenient restriction in what follows to diffuse probabilities entails no loss of generality.

THEOREM 2. *Let* $\lambda$ *be a diffuse invariant probability. Then* $\lambda$ *is topologically ergodic iff for each nonempty open set* $U$ *we have*

(*)                        $\lambda\{x: i(U)(x) = 1\} = 1.$

PROOF. Suppose first that $\lambda$ is topologically ergodic diffuse and $U$ a nonempty open set. Then $i(U)$ is subinvariant, so (by integrating with respect to $\lambda$) we, in fact, have

$$Ti(U) = i(U) \quad [\text{a.e. } \lambda].$$

But $i(U)$ is bounded l.s.c. so this $\lambda$-invariant function must be $\lambda$-constant. But $1_U \leqslant i(U) \leqslant 1$ implies that $i(U)$ is 1 on $U$. Since $\lambda$ is diffuse (so $\lambda(U) > 0$) we conclude $i(U) = 1$ [a.e. $\lambda$]. Conversely, if $\lambda$ is not topologically ergodic we select a closed invariant set $D$ with $0 < \lambda(D) < 1$. Then we have for $U = X \setminus D$, $i(U) = 1_U$ at all points of $X$. But then equation (*) fails so we are done.

Let $P_z$ be the Kolmogorov probability measure (on $\Omega$, the sample path space for $(X, T)$) for the Markov process $X_0, X_1, \ldots$ determined by the transition $p(x, A) = T^*\delta(x)(A)$ and initial distribution $X_0 = z$.

If $X$ is second countable, in particular if $X$ is metrizable, then Theorem 2 easily implies

$$\lambda\{z: P_z(\text{path } X_0, X_1, \ldots \text{ is dense in } X) = 1\} = 1$$

if $\lambda$ is a diffuse topologically ergodic probability. For if $\{W_i\}$ is a countable open base then

$$\bigcap_i \{z: i(W_i)(z) = 1\} = \{z: P_z(\text{path enters } W_i) = 1 \text{ for all } i\}$$

$$= \{z: P_z(\text{path enters } W_i \text{ for all } i) = 1\}$$

since the $P_z$ probability of the intersection of countably many sets with $P_z$ probability 1 is again 1. Similarly the $\lambda$ measure of the intersection on the left is 1 by Theorem 2.

The generalization of Furstenberg's theorem [3, p. 22] holds without additional assumptions on $X$:

THEOREM. *Suppose* $\lambda$ *is a diffuse topologically ergodic probability and* $f \geqslant 0$ *is continuous. If* $Z$ *is the set of points z in X such that*

$$\max\{f(X_0), f(X_1), \ldots, f(X_n)\} \to \|f\|_\infty$$

*holds almost surely* $(P_z)$ *then* $\lambda(Z) = 1$.

To prove this we note that the set on which a continuous function attains its maximum is a $G_\delta$ set and the result follows as above from Theorem 2 and the observation that the intersection of a countable collection of sets of maximal measure is again of maximal measure.

3. **Structure of topologically ergodic probabilities.** In the previous section we consider results for topologically ergodic probabilities which were analogues of results for ergodic probabilities. One of the most elegant results for ergodic probabilities is the characterization as the extreme points of the compact convex set of all invariant probabilities. It is to the analogue of this result that we turn now. To gain some intuitive feeling for the structure we consider the following. Suppose $\{\sigma_n\}$ are ergodic probabilities with support $(\sigma_n) = X$ for all $n = 1, 2, \ldots$. If $a_n > 0$ with $\Sigma a_n = 1$, then it is not difficult to show that $\lambda = \Sigma a_n \sigma_n$ is a diffuse topologically ergodic probability. The intuitive picture is that the diffuse topologically ergodic probabilities are members of a facet of diffuse ergodic probabilities. This picture is almost technically correct; the technical requirement is that $\lambda$ have a Choquet representation carried by the diffuse ergodic measures. It is not at all clear a priori that the collection of diffuse ergodic probabilities is nonempty when there is a diffuse topologically ergodic probability $X$.

Some aspects of the results of this section hold under weaker assumptions but we will assume throughout this section that $X$ is a *compact metric* space.

Let $K$ be a compact set of probabilities on $X$ and let $K_0$ be the diffuse probabilities in $K$. We remark that $K_0$ is a $G_\delta$ set since if $\{W_i\}$ is a global basis for the topology of $X$ and $K_i = \{\mu \in K: \mu(W_i) = 0\}$, then $K_i$ is compact and we have $K_0 = K \setminus \bigcup K_i$.

For the structure theorem we must modify this decomposition slightly. Note first that if $\lambda$ is a diffuse topologically ergodic probability and $D$ is a proper closed invariant set then $\lambda(D) = 0$.

Again select a global base $\{W_i\}$ and set $D_i = X \setminus W_i$. Let $E_i$ be the invariant core of $D_i$. Now for each ergodic invariant probability, $\mu$, which is not diffuse we select one set $W_i$ with $W_i \cap \text{supp } \mu = \varnothing$. Then $E_i$ is not empty and indeed $E_i$ carries $\mu$. Now we set $K_i = \{\mu \in \text{ext } K: \text{supp } \mu \subset \overline{E_i}\}$ and

$$K_d = \text{ext } K \setminus \bigcup K_i.$$

THEOREM 4. $\lambda$ *is a diffuse topologicall ergodic probability iff* $\lambda$ *has a Choquet representation* $dR_\lambda$,

(∗)
$$(\lambda, f) = \int \hat{f}(\mu) \, dR_\lambda(\mu),$$

*which is carried by* $K_d$, *the set of diffuse ergodic probabilities. In particular, if* $(X, T)$ *has a diffuse topologically ergodic probability then* $K_d$ *is not empty.*

PROOF. First we observe that the Choquet representation (∗) above holds not only for the continuous affine functions on $K$ but for affine functions of the first Baire class as well. Thus we may apply (∗) for functions $f = 1_D$ where $D$ is closed. (See [1] and [6].)

Suppose now that $\lambda$ is an invariant probability with a Choquet representation $R_\lambda$ carried by $K_d$. If $\text{supp}(\lambda) = D$ is proper in $X$ then $\mu(D) = 0$ for each diffuse ergodic probability. Thus

$$1 = \lambda(D) = \int \hat{1}_D(\mu) \, dR_\lambda(\mu) = 0$$

and this contradiction shows support $\lambda = X$. Next suppose that $D$ is a closed invariant proper subset of $X$. Then again $\mu(D) = 0$ for all $\mu$ in $K_d$ so $\lambda(D) = 0$. Thus $\lambda$ is topologically ergodic.

Conversely, suppose $\lambda$ is diffuse topologically ergodic. Then $R_\lambda(\text{ext } K) = R_\lambda(K_d) + R_\lambda(\cup K_i)$. Suppose $R_\lambda(K_d) < 1$. Then $R_\lambda(K_i) > 0$ for some $i$. Then

$$\lambda(E_i) = \int \hat{1}_{E_i}(\mu) \, dR_\lambda(\mu) > 0.$$

We conclude that either $\lambda(E_i) = 1$ so that $\lambda$ is not diffuse or $0 < \lambda(E_i) < 1$ so that $\lambda$ is not topologically ergodic. In either case we have a contradiction which finishes the proof.

It is of some interest to compare the results here with the negative results of B. Weiss [9]. Suppose that $\phi$ is a homeomorphism of a compact metric space and $Tf = f \circ \phi$. If there is at least one point which has dense orbit and there is a diffuse invariant probability, it is natural to ask if there then exists a diffuse ergodic invariant probability. The example constructed by Weiss shows that the answer is no. On the other hand Theorems 2 and 4 imply the following: If $\lambda$ is a diffuse invariant probability and almost all points with respect to $\lambda$ have dense orbits, then there is a diffuse ergodic probability.

### REFERENCES

1. E. M. Alfsen, *Compact convex sets and boundary integrals*, Springer-Verlag, Berlin and New York, 1971.

2. S. R. Foguel, *The ergodic theory of positive operators on continuous functions*, Ann. Scuola Norm. Sup. Pisa. **27** (1973), 19–51.

3. H. Furstenberg, *Stationary processes and prediction theory*, Princeton Univ. Press, Princeton, N.J., 1960.

4. A. Garsia, *Topics in almost everywhere convergence*, Markham, Chicago, Ill., 1970.

5. B. Jamison and R. C. Sine, *Sample path convergence for stable Markov operators*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **28** (1974), 173–177.

6. R. R. Phelps, *Lectures on Choquet's theorem*, Van Nostrand, Princeton, N.J., 1966.

7. R. C. Sine, *Geometric theory of a single Markov operator*, Pacific J. Math. **27** (1968), 155–166.

8. _____, *Sample path convergence of stable Markov processes*. II, Indiana Univ. Math. J. **25** (1976), 23–44.

9. B. Weiss, *Topological transitivity and ergodic measures*, Math. Systems Theory **5** (1971), 71–75.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF RHODE ISLAND, KINGSTON, RHODE ISLAND 02881