For Mixing Transformations $\operatorname{rank}(T^k) = k \cdot \operatorname{rank}(T)$

Jonathan L. King*

University of Florida, Gainesville 32611-2082, USA squash@math.ufl.edu Webpage http://www.math.ufl.edu/∼squash/

Abstract. For the class of finite–rank transformations having no partial rigidity (this includes the class of mixing transformations) rank behaves like a logarithm on positive powers in that $\operatorname{rk}(T^k) = k \cdot \operatorname{rk}(T)$.

One of the vague but central questions in the area of finite rank mixing transformations is whether each such T can be built from some set of "basic" transformations (optimistically: transformations with minimal self-joinings) via some class of "reasonable" operations such as powers, roots, and finite extensions. A first step would be to explore how rank varies under these operations.

One case of how rank varies under powers is known: If T is rank-1 mixing then $\operatorname{rk}(T^k) = k$; this is a special case of a result in [2] and has apparently been known for some time although the author does not know where it appears in print. Some assumption on the randomness of T is necessary —for instance, weak mixing is not enough. One can construct, see [1], a weak mixing T such that $\operatorname{rk}(T^k) = 1$ for all non-zero k. This T will be rigid.

The purpose of this paper is to demonstrate the following theorem.

Theorem (Rank of Powers). *If T has zero rigidity, with $p \in \mathbb{N}$ and T^p ergodic, then* $\operatorname{rk}(T^p) = p \cdot \operatorname{rk}(T)$.

*Affiliation when the article was submitted: *Dept. of Mathematics, SUNY Albany, Albany, NY 12222, USA*

The requirement that T^p be ergodic is made to ease the exposition of the proof, since the applications in which we are interested arise when T is mixing. If T is mixing and $S^q = T^p$ for some transformation S and positive integers p and q, then, since the mixing property is closed under roots and (non-zero) powers,

$$\operatorname{rk}(S) = \frac{p}{q} \cdot \operatorname{rk}(T)\,.$$

Notations and Conventions. A ***transformation*** will mean a measure preserving transformation on a Lebesgue probability space. A transformation is invertible unless said otherwise. All measures are assumed non-atomic. Every subset that we mention is assumed to be measurable. If A and B are subsets, $A \setminus B$ denotes the set-theoretic difference $A \cap B^c$. As is customary, statements have a tacit "almost everywhere" attached, where appropriate.

A partition (alphabet) is a finite collection of atoms (letters). Consider a transformation $T\colon X \to X$, a generating partition P and a point $x \in X$. Let $x|_i$, for $i \in \mathbb{Z}$, denote the letter of P that $T^i(x)$ is in. Let $x|_{-\infty}^{\infty}$ denote the doubly infinite T,P-name of x and, for integers a and b with $a < b$, let $x|_a^b$ denote the substring

$$x|_a x|_{a+1} \cdots x|_{b-a}\,.$$

Generally we will identify the point x with its name, and so x is a synonym for $x|_{-\infty}^{\infty}$. For W a P-h-word, that is, a word of length h over the P alphabet, we index W from zero, i.e, $W = W|_0^h$. We use $\operatorname{len}(W)$ to denote the length, h, of W. So

$$\operatorname{len}\big(x|_a^b\big) = b - a\,.$$

We use $[a\,..\,b)$ to indicate the half-open "interval of integers" $[a, b) \cap \mathbb{Z}$, with the analogous symbol for closed and open intervals.

Israel J. Math., vol. 56 (1986), 102–122.

The symbols $\varepsilon, \delta, \sigma, \kappa$ denote numbers in $(0,1]$. We use "$\forall$ large" as a quantifier so that "$\forall$ large n" means "$\exists N$ such that $\forall n > N$". The expression "ε-percent" means "a fraction thereof, of size ε". Saying that at least ε-percent of the name $x|_0^\infty$ is covered by the disjoint substrings $\{x|_{i_m}^{i_m+h}\}_{m=1}^\infty$, where $i_m + h \leqslant i_{m+1}$, means that the lower density of the set $\bigcup_{m=1}^\infty [i_m \,..\, i_m{+}h)$ of natural numbers is bounded below by ε.

We use the symbol "$\coloneqq$" so that

$$\text{“}a \coloneqq b\text{”} \quad \text{or} \quad \text{“}B \eqqcolon a\text{”}$$

means the expression b defines the new symbol a. An expression such as $7(1-\delta) \coloneqq 6$ means to define the symbol δ so that equality holds.

§0. We define the rank of a transformation.

DEFINITION. We have $T\colon X \to X$ and a fixed generating partition P. Given a collection $\mathbf{C}$ of P-h-words, an $x \in X$ and positive integers $\{i_m\}_{m=1}^\infty$ with $i_m + h \leqslant i_{m+1}$, we say that **"the $\{i_m\}_{m=1}^\infty$ sequence $(1-\delta)$-*covers*** (the name $x|_0^\infty$) up to ε $\bar{d}$-error**"** if

(1) *the disjoint substrings* $\{x|_{i_m}^{i_m+h}\}_{m=1}^\infty$ *cover at least* $(1-\delta)$*-percent of* $x|_0^\infty$

and for each m there is an associated word $W \in \mathbf{C}$ such that

$$\bar{d}\big(x|_{i_m}^{i_m+h}, W\big) \leqslant \varepsilon\,. \tag{2}$$

The sequence of natural numbers $\{i_m\}_{m=1}^\infty$ will be called a $(1-\delta)$***-covering*** (of $x|_0^\infty$ by $\mathbf{C}$).

Let r denote the number of words, $|\mathbf{C}|$, in $\mathbf{C}$. Suppose we have, associated to $\mathbf{C}$, a set $B \subset X$ which is the base of a Rohlin stack of height h. Suppose the stack fills up at least $(1-\delta)$-percent of the space, i.e, $h \cdot \mu(B) \geqslant 1-\delta$. To each $W \in \mathbf{C}$ suppose we have an associated set B_W such that B is the disjoint union $\bigcup_{W \in \mathbf{C}} B_W$ and

(3) *For each* $y \in B_W$: $\bar{d}\big(y|_0^h, W\big) \leqslant \varepsilon$.

Then given an $x \in X$, we have a prescription for making a $(1-\delta)$-covering $\{i_m\}_1^\infty$ —simply let i_m be the mth hitting time of the orbit of x against the set B. That is, let i_m be the smallest i, with $i > i_{m-1}$, such that $T^i(x) \in B$; start the induction by $i_0 = 0$. Thinking of $\mathbf{C}$ as a collection of paints and each $W \in \mathbf{C}$ as a *color*, we can interpret (1) and (2) as a prescription for painting $h \cdot \mu(B)$-percent of $x|_0^\infty$ using r colors of paint.

We will call $\mathbf{C}$, along with its associated sets B_W and the number ε, a ***palette*** if (3) is satisfied. Agree to let

$$\operatorname{len}(\mathbf{C}) \quad \text{and} \quad \bar{d}\text{-err}(\mathbf{C})$$

denote the numbers h and ε, respectively. Let base$(\mathbf{C})$ denote the base set, B, of the stack. For each color $W \in \mathbf{C}$, let base(W) denote B_W. Finally, let $\mu(\mathbf{C})$ denote the number $h \cdot \mu(B)$ –the measure of the stack– and let the symbol $\mu(W)$ denote the product $h\mu(B_W)$, for each $W \in \mathbf{C}$.

Given a palette $\mathbf{C}$ and a point x, the above sequence $\{i_m\}_1^\infty$, where i_m is the mth hitting time of x against base$(\mathbf{C})$, will be called **"the covering of x** (by $\mathbf{C}$)**"**. If index i and color W are such that $T^i(x) \in B_W$, then we will call the substring $x|_i^{i+h}$ a ***swatch*** of color W and write $\operatorname{color}\big(x|_i^{i+h}\big) = W$.

DEFINITION. A ***sequence of palettes*** $\{\mathbf{C}_n\}_{n=1}^\infty$ will mean a sequence of palettes with the understanding that

$$\operatorname{len}(\mathbf{C}_n) \to \infty,\ \mu(\mathbf{C}_n) \to 1,\ \text{and } \bar{d}\text{-err}(\mathbf{C}_n) \to 0\,,$$

as $n \to \infty$. Define the rank of the process T,P to be the smallest positive integer r for which there exists a sequence of palettes $\{\mathbf{C}_n\}_1^\infty$ with each $|\mathbf{C}_n| = r$. If no such r exists, define the rank –written $\operatorname{rk}(T,P)$– to be infinite. Finally, define the rank of the *transformation* T to be the supremum of $\operatorname{rk}(T,P)$ as P ranges over all partitions of X.

Remark. The definition of rank given above is called by some authors ***even rank*** or ***uniform rank***. We remark in passing that there is a more general notion called ***non-uniform rank***. When contrasting the two, it is mnemonic to use $\overline{\operatorname{rk}}(\cdot)$ for uniform rank and $\widetilde{\operatorname{rk}}(\cdot)$ for non-uniform rank.

The definition of $\widetilde{\operatorname{rk}}(T,P)$ is the same as for uniform rank except that the words in a palette need not all have the same length. Consequently, the associated Rohlin stack is permitted to have columns of varying heights. A sequence of palettes must have $\operatorname{len}(\mathbf{C}_n) \to \infty$ where, now, $\operatorname{len}(\mathbf{C})$ denotes the length of the *shortest* word in $\mathbf{C}$. *A fortiori*, $\widetilde{\operatorname{rk}}(T) \leqslant \overline{\operatorname{rk}}(T)$. We will not have occasion to use non-uniform rank in this article other than for a question in the conclusion.

It will be useful to show that the rank of an *ergodic* process T,P can be computed from the name of any single point x. (Here, x is to be "sufficiently generic".)

PROPOSITION. *Suppose x is generic and C is a finite collection of P-h-words. Also suppose that $\{i_m\}_{m=1}^\infty$ is a $(1-\delta^2)$-covering of $x|_0^\infty$ up to ε $\bar{d}$-error. Then we can find a palette $\mathbf{C}$, whose collection of colors is C, such that $\bar{d}$-err$(\mathbf{C}) = \varepsilon$ and $\mu(\mathbf{C}) > 1 - 5\delta$.*

Remark. Hence, if one can cover $(1-\delta^2)$-percent of a particular generic name $x|_0^\infty$ by the words in $\mathbf{C}$, one can cover at least $(1-5\delta)$-percent of *each* generic name by the words in C.

SKETCH OF PROOF. We obtain a set $B \subset X$, to play the role of base$(\mathbf{C})$, by describing how to construct the sets $\{B_W\}_{W\in\mathbf{C}}$. We leave to the reader the explicit calculation which shows that B can be so chosen that $h\mu(B) > 1 - 5\delta$. Actually even this computation is unnecessary —we are really only interested in showing that $h\mu(B)$ can be made as close to 1 as desired by having chosen δ sufficiently tiny.

Let $E \subset X$ be the base of a Rohlin stack

(4) $$E,\ T(E),\ T^2(E),\ \ldots,\ T^{H-1}(E)$$

where base E and the stack height H were chosen so $H \gg h$ and the measure of the stack is very close to 1, that is, $H \cdot \mu(E) \approx 1$. Define integers $\{j_\ell\}_{\ell=1}^\infty$ inductively by $j_0 \coloneqq 0$ and j_ℓ is the smallest number j exceeding $j_{\ell-1}$ such that $T^j(x) \in E$. Courtesy of the ergodic theorem:

(5) The substrings $\left\{x|_{j_\ell}^{j_\ell+H}\right\}_{\ell=1}^\infty$ cover most of $x|_0^\infty$.

By hypothesis, the swatches $\{x|_{i_m}^{i_m+h}\}_{m=1}^\infty$ cover most of $x|_0^\infty$. (Here "most" is actually $1-\delta^2$; but since we are not doing an explicit calculation, we need not have an explicit name for it.) This, combined with (5) and $H \gg h$, yield that for most values of ℓ:

(6) *Most of* $x|_{j_\ell}^{j_\ell+H}$ *is covered by* $\left\{x|_{i_m}^{i_m+h}\right\}_{m=m'}^{m''}$

where m' and m'' are chosen smallest and largest, respectively, that $j_\ell \leqslant i_{m'} < i_{m''} + h \leqslant j_\ell + H$. Hence, by discarding the bad values of ℓ and renumbering, we may assume that (6) holds for *all* ℓ and that (5) continues to hold.

Returning to our Rohlin stack, we can columnate it into columns determined by the T,P-H-names of points in the base. Letting K denote the number of such columns, we can write the base E as the disjoint union $\bigcup_{k=1}^k E_k$, where $E_1, \ldots, E_k$ are the bases of the K columns. For each k, let $\ell(k)$ denote the smallest positive integer, should one exist, such that

$$T^{j_{\ell(k)}}(x) \in E_k\,.$$

Discard from E those sets E_k for which no such $\ell(k)$ exists. The set E is now a bit smaller than it was, but we can continue to assert that the stack (4) fills up most of the space; the ergodic theorem and (5) imply that the total measure of the discarded columns must be small. For each discarded E_k, discard k from $\{1,\ldots,K\}$ and renumber so that we may now say that $\ell(k)$ exists for every $k \in \{1,\ldots,K\}$.

We proceed to define a set B, the base of a Rohlin stack of height h, as the disjoint union of sets B_W as W ranges over the "colors" in C. For $k = 1,\ldots,K$, let $m(k)$ and $M(K)$ be the numbers m' and m'' of (6) when $\ell := \ell(k)$. For each color $W \in C$, let B_W be, roughly, the union of those column levels of our stack, (4), which commence a swatch of color W. Specifically, let B_W be the union

$$\bigcup_{k=1}^{K} \bigcup \left\{ T^{i_m - j_{\ell(k)}}(E_k) \;\middle|\; \begin{smallmatrix} m(k) \leqslant m \leqslant M(k) \text{ and} \\ \mathrm{color}\left(x|_{i_m}^{i_m+h}\right) = W \end{smallmatrix} \right\}.$$

The disjoint union $B := \bigcup_{W\in C} B_W$ forms the base of a Rohlin stack of height h. As a set, this stack $\bigcup_{i=0}^{h-1} T^i(B)$ is a subset of our other stack

$$\bigcup_{i=0}^{H-1} T^i(E)\,.$$

Moreover, as sets these stacks are practically equal, by (6). One can have chosen H sufficiently $\gg h$ and $H \cdot \mu(E)$ sufficiently ≈ 1 to insure that $h\mu(B) > 1 - 5\delta$. ♦

We get immediately the following corollary.

Lemma 1. *Suppose we have numbers $h_n \nearrow \infty$, $\delta_n \to 0$, and $\varepsilon_n \to 0$, as well as a positive integer r. Further, suppose we have a generic point x and a collection C_n of P-h_n-words which $(1-\delta_n)$-covers $x|_0^\infty$ up to ε_n $\bar{d}$-error. If each $|C_n| \leqslant r$, then $\mathrm{rk}(T,P) \leqslant r$.*

The definition of $\mathrm{rk}(T,P)$ is a "coding" definition in that it allows for some $\bar{d}$-error. An advantage of the coding definition is that it inherits under codes.

Lemma 2. *P and Q are partitions of X and P generates under T. Then $\mathrm{rk}(T,Q) \leqslant \mathrm{rk}(T,P)$.*

Proof. Pick a point $x \in X$ generic for processes T,Q and T,P. The T,Q-name of x can be approximated arbitrarily well by finite codings of the T,P-name of x. ♦

A consequence of the lemma is that if Q generates, then $\mathrm{rk}(T,Q) = \mathrm{rk}(T,P)$. The coding definition of rank thus has the convenient feature that the rank of a transformation T can be read from each generating partition P; for $\mathrm{rk}(T)$ simply equals $\mathrm{rk}(T,P)$.

The following lemma is true without the stated assumption of ergodicity. However, we use a name argument for the proof as preparation for future name arguments.

Lemma 3. *For a given $p \in \mathbb{N}$ suppose that T^p is ergodic. Then*

$$p \cdot \mathrm{rk}(T) \geqslant \mathrm{rk}\big(T^p\big) \geqslant \mathrm{rk}(T). \tag{7}$$

Proof. Set $r := \mathrm{rk}(T)$. The right-hand inequality of (7) succumbs to the same kind of proof as does the left-hand inequality. Also, the case $p = 2$ gives the idea. So we content ourselves to show $2r \geqslant \mathrm{rk}\big(T^2\big)$.

Let P be a generating partition for T. Then

$$Q := P \vee T(P)$$

generates for T^2. A Q "letter" is an ordered pair $(w\,w')$ where w and w' are letters of P. So a Q-word is a sequence of pairs P-letters.

Given ε and δ, choose some palette $\mathbf{C}$ for the T,P process which has r colors, satisfies

$$\bar{d}\text{-err}(\mathbf{C}) < \frac{\varepsilon}{2} \quad \text{and} \quad \mu(\mathbf{C}) > 1-\delta.$$

Furthermore, it has $h' := \operatorname{len}(\mathbf{C})$ large compared with $1/\frac{\varepsilon}{2}$ and $1/\delta$. For notational convenience assume that h' is odd and write $2h+1 := h'$. Fix some $x \in X$ which is generic for the processes T,P and T^2,Q. Writing its forward T,P name as $x|_0^\infty = x_0x_1x_2\cdots$, the forward T^2,Q name of x is

$$(8) \qquad (x_0\,x_1)\,(x_2\,x_3)\,(x_4\,x_5)\cdots.$$

We show that $\operatorname{rk}(T^2) \leqslant 2r$ by constructing a collection of Q-h-words, $2r$ many of them, which $(1-\delta)$-covers (8), up to ε $\bar{d}$-error.

Given a $W \in \mathbf{C}$, write it as a sequence of P-letters $w_0w_1\cdots w_{2h}$. From W we create two Q-h-words:

$$(w_0\,w_1)\,(w_2\,w_3)\,\cdots(w_{2h-2}\,w_{2h-1}),$$
$$(w_1\,w_2)\,(w_3\,w_4)\,\cdots(w_{2h-1}\,w_{2h}).$$

As W ranges over $\mathbf{C}$, we so obtain a collection of $2r$ many Q-h-words. If h' was chosen sufficiently large, this collection will $(1-\delta)$-cover (8) up to ε $\bar{d}$-error. ♦

§**1.** We now turn to some of the interactions between rigidity of a transformation and finite rank.

Definition. Define the ***rigidity number***, $\rho(T)$, of a transformation T to be the supremum of numbers $\rho \in [0,1]$ for which there exists a sequence of integers $\{s_n\}_1^\infty$ going to infinity, such that

$$(9) \quad \forall A: \liminf_{n\to\infty} \mu(A \cap T^{s_n}(A)) \geqslant \rho\cdot\mu(A).$$

By splicing sequences, there exists a sequence $\{s_n\}_n$ for which the above holds with ρ replaced by $\rho(T)$.

Lemma 4. *In the preceding definition, rather than require $|s_n| \to \infty$ it suffices that $s_n \neq 0$.*

Proof. Suppose $\{s_n\}_n$ is a non-zero sequence satisfying (9), for some $\rho > 0$. If $|s_n| \not\to \infty$ then we can drop to a subsequence of $\{s_n\}_n$ which is a non-zero constant —say, 17. Thus

$$(9') \qquad \forall A: \mu\big(A \cap T^{17}A\big) \geqslant \rho\mu(A).$$

In particular this holds for every set A of the form $B \smallsetminus T^{17}(B)$. Since this A is disjoint from $T^{17}(A)$, inequality (9′) forces $\mu(A) = 0$. So $B = T^{17}(B)$ for each set B. Hence $\rho(T) = 1$. ♦

One says T has ***zero rigidity*** if $\rho(T) = 0$, ***partial rigidity*** if $\rho(T) > 0$, and is ***rigid*** if $\rho(T) = 1$. The following is easily checked.

Proposition. *If T is mixing then $\rho(T) = 0$.*

Hence the Rank-of-Powers theorem will apply to each finite rank mixing transformation.

It will be convenient to have a criterion, in terms of names, for recognizing partial rigidity. We will need a consequence of the ergodic theorem and the definition below. For the rest of this section, we have a fixed ergodic process T,P.

Suppose A is a T,P cylinder set of length k. Let $A|_0^k$ denote the P-k-word defining the set A. For each word $W|_0^\ell$, with $\ell > k$, define the frequency

$$\operatorname{freq}\big(A|_0^k \text{ in } W|_0^\ell\big)$$

to be $\frac{1}{\ell-(k-1)}$ times the cardinality of

$$\left\{ i \;\middle|\; 0 \leqslant i < i+k \leqslant \ell \text{ and } W|_i^{i+k} = A|_0^k \right\}.$$

The following well known result is a standard corollary of the ergodic theorem.

STANDARD CODING LEMMA. (In the statement of the lemma, the word "density" may be consistently replaced by either "lower density" or "upper density", in the set of natural numbers.) *For a.e T,P-name x, every ε, and each cylinder set $A|_0^k$, the following holds $\forall$ large ℓ. Suppose $\{x|_{i_m}^{i_m+\ell}\}_{m=1}^{\infty}$ is a sequence of disjoint blocks with $i_m + \ell \leqslant i_{m+1}$. Let κ denote the density of these blocks in $x|_0^\infty$, i.e,*

$$\kappa \coloneqq \text{density}\Big(\bigcup_{m=1}^{\infty} [i_m \mathinner{..} i_m + \ell)\Big) .$$

Then, the density in $x|_0^\infty$ of those blocks satisfying

$$\Big|\text{freq}\big(A|_0^k \text{ in } x|_{i_m}^{i_m+\ell}\big) - \mu(A)\Big| \;<\; \varepsilon \tag{10}$$

exceeds $\kappa - \varepsilon$.

Remark. In the sequel, a name $x|_0^\infty$ will be called ***generic*** for T,P if –in addition to seeing each cylinder set with the correct limiting frequency– it is in the set of full measure of the lemma.

PROOF. Say that the mth block is ***good*** if (10) holds. To prove the lemma in either the upper or lower density case, it suffices to show that for all large M:

$$\begin{aligned} &\text{density}\Big\{j \;\Big|\; \begin{matrix}\exists \text{ a good } m \text{ with} \\ i_m \leqslant j < i_{m+\ell}\end{matrix}\Big\} \\ &\quad > \text{density}\Big\{j \;\Big|\; \begin{matrix}\exists m \text{ with} \\ i_m \leqslant j < i_{m+\ell}\end{matrix}\Big\} - \varepsilon . \end{aligned} \tag{11}$$

Here m ranges over $\{1,\ldots,M\}$ and j ranges over the interval of integers $[0 \mathinner{..} i_M + \ell)$; "density" is computed relative to this interval of integers.

It will be convenient notationally, and there is no essential loss of generality, in assuming that $k = 1$ and that the set A can be regarded as a letter in the alphabet P. When so regarded, we write it as "A". By the ergodic theorem, we can choose an N large enough that $\mu(E) > 1 - \varepsilon$, where E denotes the set of points $y \in X$ such that

$$\begin{aligned} &\forall U,L \geqslant N : \\ &\quad \Big|\text{freq}\big(\text{“A” in } y|_{-L}^{U}\big) - \mu(A)\Big| < \varepsilon . \end{aligned} \tag{12}$$

Now choose ℓ large enough that $\mu(E) > 1 - \varepsilon + \frac{2N\varepsilon}{\ell}$. Take x to be any point whose orbit hits E with the correct frequency.

We now demonstrate (11). Fix some M large enough that

$$\text{density}(\mathcal{D}) \;>\; 1 - \varepsilon + \frac{2N}{\ell}\varepsilon ,$$

where $\mathcal{D}$ denotes the set of integers j in $[0 \mathinner{..} i_M + \ell)$ such that $T^j(x) \in E$. Given an $m \in \{1,\ldots,M\}$, suppose that the intersection

$$\mathcal{D} \cap \big[i_m + N \mathinner{..} (i_m + \ell) - N\big)$$

is non-empty. Letting j be a value in the intersection and setting $U \coloneqq (i_m + \ell) - j$ and $L \coloneqq j - i_m$, we conclude that (12) forces m to be good. Hence, for bad m, the intersection must be empty. Consequently, letting BAD be the union of $[i_m \mathinner{..} i_m + \ell)$ over all bad m,

$$\begin{aligned} \big[1 - \tfrac{2N}{\ell}\big] \cdot \text{density}\big(\text{BAD}\big) \;&<\; 1 - \text{density}(\mathcal{D}) \\ &<\; \big[1 - \tfrac{2N}{\ell}\big]\varepsilon , \end{aligned}$$

The inequality resulting from dividing both sides by $1 - \frac{2N}{\ell}$, combined with the disjointedness of the blocks, implies the desired (11). ♦

Remark. The above proof is more efficient than the standard proof thanks to the neat idea, due to Ornstein, of defining E bidirectionally in (12).

RIGIDITY CRITERION. *We have a $\kappa > 0$ and a generic point x. Suppose that for all ε we can find an arbitrarily large ℓ and a non-zero s for which: At least*

κ-percent of $x|_0^\infty$ can be covered by disjoint blocks $\{x|_{i_m}^{i_m+\ell}\}_{m=1}^\infty$ such that

$$\bar{d}\big(x|_{i_m}^{i_m+\ell},\, x|_{i_m+s}^{i_m+s+\ell}\big) \;\leqslant\; \varepsilon\,. \tag{13}$$

Then $\rho(T) \geqslant \kappa$.

PROOF. Fix sequences $\{\varepsilon_n\}_1^\infty$, $\{\ell_n\}_1^\infty$, and $\{s_n\}_1^\infty$ such that $\varepsilon_n \to 0$, $\ell_n \to \infty$, and $s_n \neq 0$, for which the hypotheses of the lemma hold with the roles of ε, ℓ and s, played by ε_n, ℓ_n and s_n. It suffices to show that for each T,P cylinder set A:

$$\liminf_{n\to\infty} \mu(A \cap T^{s_n} A) \geqslant \kappa \cdot \mu(A).$$

We argue for some fixed A, with $\mu(A) > 0$. As in the previous proof, there is no essential loss of generality in assuming that A is a cylinder set of length 1. When regarded as a letter of the alphabet, it will be written "A".

Choose a large n so that $\varepsilon_n \ll \mu(A)$ and let ℓ, s and ε, in (13), abbreviate ℓ_n, s_n and ε_n. By our choice of n we may have made ℓ sufficiently large for the Standard Coding lemma to work for the set A. Consequently, as least $(\kappa - \varepsilon)$-percent of $x|_0^\infty$ is covered by those blocks $x|_{i_m}^{i_m+\ell}$ such that

$$\big|\operatorname{freq}\big(\text{“A” in } x|_{i_m}^{i_m+\ell}\big) \;-\; \mu(A)\big| \;\leqslant\; \varepsilon\,.$$

We can have chosen ε $(=\varepsilon_n)$ as small as desired compared with $\mu(A)$. So we may harmlessly pretend, in the above inequality and (13), that ε equals zero. Hence, at least $(\kappa-\varepsilon)\cdot\mu(A)$-percent of indices $j \in |_0^\infty$ are such that

$$x|_j = \text{“A”} = x|_{j+s}\,.$$

So, by the ergodic theorem,

$$\mu(A \cap T^{s_n} A) \;\geqslant\; [\kappa - \varepsilon_n]\mu(A)\,,$$

where we have rematerialized the subscript n. Sending $n \to \infty$ sends $\varepsilon_n \to 0$ and completes the proof. ♦

Remark. (The upper density form of the Rigidity Criterion.) The Rigidity Criterion, just as stated, will be used repeatedly. However, to avoid making less intuitive its proof, we forbade stating the criterion in the triflingly stronger form which will be needed at the end of §3.

The conclusion, that $\rho(T) \geqslant \kappa$, persists under a weaker condition on the manner in which $x|_0^\infty$ is covered by the disjoint blocks $\{x|_{i_m}^{i_m+\ell}\}_{m=1}^\infty$. Namely, we only need that the *upper* density on $x|_0^\infty$ of these blocks be greater than or equal to κ. In fact, we need change neither the statement nor the proof of the criterion, if we agree to interpret each phrase

"... at least (such-and-such)-percent of so-and-so..."

as asserting only that the *upper* density of so-and-so exceeds such-and-such. Using the upper density version of the Standard Coding lemma, we will be able to conclude only that the upper density of $j \in |_0^\infty$ such that

$$x|_j \;=\; \text{“A”} \;=\; x|_{j+s}$$

exceeds $[\kappa - \varepsilon]\mu(A)$. Still, the ergodic theorem will imply that $\mu(A \cap T^s A) \geqslant [\kappa - \varepsilon]\mu(A)$ and complete the proof.

The reader may check the following corollary of the Rigidity Criterion, stated in terms of a sequence $\{\mathbf{C}_n\}_1^\infty$ of palettes, with h_n denoting $\operatorname{len}(\mathbf{C}_n)$.

COROLLARY. *Suppose $\exists \kappa, \sigma > 0$ such that $\forall \varepsilon$ and for arbitrarily large n, the following holds: There exists a $W \in \mathbf{C}_n$ with $\mu(W) \geqslant \kappa$, and integers $0 \leqslant j < j + s < j + s + \ell \leqslant h_n$, such that*

$$\bar{d}\big(W|_j^{j+\ell}, W|_{j+s}^{j+s+\ell}\big) \;\leqslant\; \varepsilon$$

and $\ell/h_n \geqslant \sigma$. Then $\rho(T) \geqslant \kappa\sigma > 0$.

§**2.** Before starting the proof of the Rank of Powers theorem we need a lemma.

LEMMA 5. *$R\colon X \to X$ a transformation with generating partition Q, and $\{\mathbf{C}'_n\}_1^\infty$ a sequence of palettes for R, Q such that $|C'_n| = r \coloneqq \operatorname{rk}(R)$. Then $\exists\kappa > 0$ such that $\forall$ large n:*

$$\mu(W') > \kappa, \text{ for each } W' \in \mathbf{C}'_n.$$

PROOF. If not, then there exist infinitely many n –by dropping to this subsequence we may say "for all n"– for which there exists a color $W'_n \in \mathbf{C}'_n$ such that $\mu(W'_n) \to 0$. But let $\mathbf{C}_n$ be the palette defined by deleting the color W'_n from $\mathbf{C}'_n$. Then $\mu(\mathbf{C}_n) = \mu(\mathbf{C}'_n) - \mu(W'_n)$, which goes to 1 as $n \to \infty$. Thus $\{\mathbf{C}_n\}_1^\infty$ is a sequence of palettes for R, Q and hence implies the contradiction that $\operatorname{rk}(R) \leqslant r - 1$. ♦

We develop the notation to be used in the next two sections. $T\colon X \to X$ is ergodic with $\rho(T) = 0$ and with generating partition P. Also, the transformation $R \coloneqq T^p$ is ergodic, where p is some fixed positive integer. Evidently, the partition $Q = \bigvee_{i=0}^{p-1} T^i P$ generates under R. By discarding a nullset, we may assume that every point of X is generic both for T,P and R, Q. Fix forevermore a point $x \in X$. Let r denote $\operatorname{rk}(R)$. Choose a sequence $\{\mathbf{C}'_n\}_1^\infty$ of palettes for R, Q such that $|\mathbf{C}'_n| = r$. Our goal is to show that $\operatorname{rk}(R) = p \cdot \operatorname{rk}(T)$. By a previous lemma it suffices to show that

$$\operatorname{rk}(R) \geqslant p \cdot \operatorname{rk}(T).$$

Constructing palettes $\{\mathbf{C}_n\}_1^\infty$ for the T,P process. To ease the notation, let us, for a moment, assume that $p = 3$. For our generic point $x \in X$, let $x|_{-\infty}^\infty$ denote its T,P name and $x'|_{-\infty}^\infty$ its R, Q name. Also abbreviate $x|_i$ and $x'|_i$ by x_i and x'_i, respectively. Each x_i is a P-letter and each x'_i a Q-letter, that is, a triple of P-letters. Specifically

$$x'_i = (x_{3i}\, x_{3i+1}\, x_{3i+2}),$$

since R is T^3.

Fix some n and let $\mathbf{C}'$ denote $\mathbf{C}'_n$ and h' denote $\operatorname{len}(\mathbf{C}')$. Let $1 - \delta \coloneqq \mu(\mathbf{C}')$ and $\varepsilon \coloneqq \bar{d}\text{-err}(\mathbf{C}')$. Each given $W' \in \mathbf{C}'$ is a Q-h'-word and may be written as a sequence of triples

$$(w_0^0\, w_1^0\, w_2^0)\,(w_0^1\, w_1^1\, w_2^1)\,\cdots\,(w_0^{h'-1}\, w_1^{h'-1}\, w_2^{h'-1}),$$

where each w_j^i is a P-letter. If we erase the parentheses and push the letters together, we can interpret W' as a P-$3h'$-word which we denote by W. Define $\mathbf{C}$ to be the set of such W as W' varies over the colors of $\mathbf{C}'$. Setting $h = 3h'$, this $\mathbf{C}$ is a collection of r many P-h-words.

In light of the lemma which says that $\operatorname{rk}(T, P)$ may be computed from the name of a single point, it will be useful to relax our definition of palette and regard $\mathbf{C}$ as a palette for T,P. For although we do not have a Rohlin stack associated to $\mathbf{C}$, we can use the colors of $\mathbf{C}$ to $(1-\delta)$-cover, up to ε $\bar{d}$-error, the forward name $x|_0^\infty$. In fact, there are three such coverings; one each inherited from the R, Q name of the points x, Tx and T^2x.

Let $\{i'_m\}_{m=1}^\infty$ be the covering by $\mathbf{C}'$ of the R, Q name of x. For each $i' \in \{i'_m\}_m$, letting W' be $\operatorname{color}(x'|_{i'}^{i'+h'})$, we have, by definition, the left-hand inequality below:

$$\varepsilon \;\geqslant\; \bar{d}\big(x'|_{i'}^{i'+h'}, W'\big) \;\geqslant\; \bar{d}\big(x|_i^{i+h}, W\big).$$

Putting $i \coloneqq 3i'$, the right-hand inequality follows from the construction of W from W'. We will therefore call $x|_i^{i+h}$ a swatch of color W. More precisely, we will call it a 0-***swatch*** since $i \equiv 0$, where the symbol $\equiv$ means congruence mod 3. Thus, setting $i_m \coloneqq 3i'_m$, we see that the 0-swatches $\{x|_{i_m}^{i_m+h}\}_{m=1}^{\infty}$ form a $(1-\delta)$-covering of $x|_0^\infty$ up to ε $\bar{d}$-error.

We will get a second such covering from the R, Q name of Tx. Denoting this name by $(Tx)'|_{-\infty}^{\infty}$ one has that

$$(Tx)'|_i \;=\; (x_{3i+1}\, x_{3i+2}\, x_{3i+3})\,.$$

Let $\{j'_m\}_{m=1}^{\infty}$ be the covering of $(Tx)'|_0^\infty$ by $\mathbf{C}'$. Set $j_m \coloneqq 3j'_m + 1$. Evidently $\{x|_{j_m}^{j_m+h}\}_1^\infty$ is a $(1-\delta)$-covering of $x|_0^\infty$ (up to ε $\bar{d}$-error). Agree to call each $x|_{j_m}^{j_m+h}$ a 1-***swatch***, since $j_m \equiv 1$.

Similarly, arising from the R, Q name of T^2x, we get a $(1-\delta)$-covering of $x|_0^\infty$ by 2-swatches.

The p different covering of $x|_0^\infty$ by $\mathbf{C}$. We recapitulate, in slightly different language, our definition of k-swatch, However, we will no longer be assuming that $p = 3$. The symbol $\equiv$ will accordingly henceforth mean congruence mod p.

Say that an index i commences a k-***swatch*** $x|_i^{i+h}$ of ***color*** W, where $k \in \{0, 1, \ldots, p-1\}$ and $W \in \mathbf{C}$, if

$$T'(x) \in \operatorname{base}(W') \text{ and } i \equiv k.$$

Set $\mu(\mathbf{C}) \coloneqq 1 - \delta$, $\bar{d}\text{-err}(\mathbf{C}) \coloneqq \varepsilon$ and, for each $W' \in \mathbf{C}'$, set $\mu(W) \coloneqq \mu(W')$. Then, for each value $k = 0, 1, \ldots, p-1$, the collection of k-swatches forms a covering of $x|_0^\infty$. We thus have p distinct coverings of $x|_0^\infty$ by the palette $\mathbf{C}$; we will call these the 0-covering, the 1-covering, …, up to the $(p-1)$-covering. Note that, in each of these different coverings, each color $W \in \mathbf{C}$ covers $\mu(W)$-percent of $x|_0^\infty$. Note also that the preceding lemma allows us to assert the existence of a positive constant κ such that

$$(14) \qquad \forall n \text{ and for each } W \in \mathbf{C}_n\colon\ \mu(W) > \kappa\,.$$

A convention. We can now dispense with palettes $\{\mathbf{C}'_n\}_n$. From now on our proofs will use facts about the $\{\mathbf{C}_n\}_n$: the p different coverings, (14), and $|\mathbf{C}_n| = \operatorname{rk}(T^p) = r$. The arguments will be of the form "choose an n sufficiently large …". At that point the n becomes implicit and $\mathbf{C}$ and h are automatically to mean $\mathbf{C}_n$ and $\operatorname{len}(\mathbf{C}_n)$, respectively. Also, for each integer ℓ (perhaps representing the length of some substring of some color) let $\ell\%$ denote the quantity ℓ/h.

The following theorem serves as a warmup for working with multiple coverings of a name $x|_0^\infty$. It is evidently a special case of the full Rank of Powers theorem since, when $\operatorname{rk}(T) = 1$, it reads $\operatorname{rk}(T^p) \geqslant p \cdot \operatorname{rk}(T)$, and hence this inequality is an equality.

THEOREM 6. $\operatorname{rk}(T^p) \geqslant p$.

PROOF. It suffices to show that for infinitely many n we have $|\mathbf{C}_n| \geqslant p$. Choose n so large that $\mu(\mathbf{C})$ is practically 1. Thus the name $x|_0^\infty$ is almost entirely covered by 0-swatches and, as we know, by k-swatches, for each and every $k \in \{0, \ldots, p-1\}$. Hence there is some integer I, with $I \equiv 0$, commencing a 0-swatch $x|_I^{I+h}$ which is mostly covered, for each k, by k-swatches. No harm will come if we assume perfection; that for each k, swatch $x|_I^{I+h}$ is completely covered by k-swatches (see Fig. 15). So for each $k = 1, \ldots, p-1$ there is an integer $j_k \in [I \mathinner{..} I+h)$, with $j_k \equiv k$, such that $x|_{j_k-h}^{j_k}$ and $x|_{j_k}^{j_k+h}$ are successive k-swatches.

Let $I \coloneqq i_0 < i_1 < \cdots < i_p = I + h$ be the $p+1$ integers $I, j_1, \ldots, j_{p-1}, I+h$, but arranged in ascending order. There must be at least one value $M \in \{0, \ldots, p-1\}$ such that $i_{M+1} - i_M \geqslant h/p$. Letting ℓ denote this difference $i_{M+1} - i_M$, we can restate this as

FIGURE 15. A swatch $x|_j^{j+h}$ is denoted pictorially by $[\!\!\longrightarrow\!\!)$.

$\ell\% \geqslant \sigma$, where we let σ be a synonym for the constant $1/p$. Let i denote i_M. We have arranged that, for each k, the fixed word $x|_i^{i+\ell}$ is a subword of one of the k-swatches overlapping $x|_I^{I+h}$. For $k = 0, \ldots, p-1$, let j'_k be the index commencing that particular k-swatch; so, for $k \neq 0$, this j'_k is the element of the pair $\{j_k - h, j_k\}$ such that

$$j'_k \leqslant i < i + \ell \leqslant j'_k + h\,.$$

We now argue that the number of colors, $|\mathbf{C}|$, must be at least p. If not, then there exist two distinct subscripts $k, K \in \{0, \ldots, p-1\}$ such that $x|_{j'_k}^{j'_k+h}$ and $x|_{j'_K}^{j'_K+h}$ are the same color, W. Since we can have chosen n arbitrarily large, we can think of the colors of $\mathbf{C}\,(=\mathbf{C}_n)$ as covering up to arbitrarily small $\bar{d}$-error. Hence there is no harm in assuming equality

$$x|_{j'_k}^{j'_k+h} = W = x|_{j'_K}^{j'_K+h}\,.$$

Now $j'_k \neq j'_K$, since $j'_k \equiv k \neq K \equiv j'_K$. For definiteness, assume $j'_k > j'_K$. Letting $s := j'_k - j'_K$ and $j := i - j'_k$ we may conclude that

$$W|_j^{j+\ell} = W|_{j+s}^{j+s+\ell}.$$

But this, by the corollary to the Rigidity Criterion, implies that T has positive rigidity, and hence gainsays an hypothesis of our theorem. ♦

DEFINITION. Suppose W and V are both words of length h. For each integer s, agree to regard $V \oplus s$ as the word V shifted left by s positions. It will be used only in the following context: Let the symbol $\bar{d}(W, V \oplus s)$ denote the frequency of those values i in $|_0^h$ such that the letter in the $(i+s)$th position of V either *does not exist* or *exists, but is unequal to* $W|_i$. This definition $\bar{d}(W, V \oplus s)$ can be expressed, for $s \geqslant 0$, by the formula

$$\frac{1}{h}\Big[s + [h-s]\cdot \bar{d}(W|_0^{h-s}, V|_s^h)\Big]\,,$$

which we agree to interpret as 1 if $s > h$. When s is negative, $\bar{d}(W, V \oplus s)$ can be defined by an analogous formula or, more succinctly, as $\bar{d}(V, W \oplus (-s))$. From its verbal description it is easy to see that

$$\bar{d}(W, V \oplus s) + \bar{d}(V, U \oplus s') \geqslant \bar{d}\big(W, U \oplus [s' + s]\big)$$

for h-words U, V, W and shifts s and s'. As a consequence of the triangle inequality above, note that the "shift distance" $\bar{s}(\cdot,\cdot)$

$$\bar{s}(W, V) := \operatorname*{infimum}_{s \in \mathbb{Z}} \bar{d}(W, V \oplus s) \tag{16}$$

is a metric on the set of words of a given length h.

Say that h-words W and V are ***ε-close*** if there exists a shift s for which $\bar{d}(W, V \oplus s) < \varepsilon$; note that this forces $|s| < \varepsilon h$. If we need to refer to the shift amount we will say that W and V are **"**ε-close via the shift s**"**.

LEMMA 7. *There is a positive constant ε_0 such that for all large n: There exists a subset of $\mathbf{D} \subset \mathbf{C}_n$ with $|\mathbf{D}| \geqslant \operatorname{rk}(T)$, such that no two distinct colors in $\mathbf{D}$ are ε_0-close.*

PROOF. From the assumption that the lemma is false, we will obtain the following impossibility: *For arbitrarily large integers h' and arbitrarily small ε, there exists a collection $\mathbf{D}'$ of h'-words which $(1-2\varepsilon)$-covers $x|_0^\infty$ up to 3ε $\bar{d}$-error. Furthermore, $|\mathbf{D}'| < \operatorname{rk}(T)$.*

Given ε, the presumed falsity of the lemma asserts we may choose n arbitrarily large for which there is subcollection $\mathbf{D}$ of $\mathbf{C}$, with $|\mathbf{D}| < \operatorname{rk}(T)$, and two maps $V(\cdot)$ and $s(\cdot)$,

$$V\colon \mathbf{C} \to \mathbf{D} \quad \text{and} \quad s\colon \mathbf{C} \to (-\varepsilon h, \varepsilon h) \cap \mathbb{Z}\,,$$

such that for each color $W \in \mathbf{C}$,

$$\bar{d}(W, V \oplus s) < \varepsilon\,. \tag{17}$$

Here, V and s are denoting, respectively, $V(W)$ and $s(W)$.

In addition, we may have chosen n sufficiently large that $1 - \mu(\mathbf{C}) \ll \varepsilon$ and $\bar{d}$-err$(\mathbf{C}) \ll \varepsilon$. Since this choice is made freely with ε known in advance, there is no danger if we simplify notation and assume perfection: $\mu(\mathbf{C}) = 1$ and $\bar{d}$-err$(\mathbf{C}) = 0$. Nor is it hazardous to pretend that the quantity εh is an integer. For by choosing n large, one can make $\varepsilon h\,(=\varepsilon h_n)$ as large as desired; hence we can alter ε by a small a percentage as desired and make εh an integer.

Define the quantities $s' := \varepsilon h - s$, $h' := h - 2\varepsilon h$, and the word $V' = V|_{\varepsilon h}^{h-\varepsilon h}$. Because $|s| < \varepsilon h$, one knows that $0 < s'$ and $s' + h' < h$. We claim that

$$\bar{d}\big(W|_{s'}^{s'+h'}, V'\big) < 3\varepsilon\,. \tag{17$'$}$$

We do the case $s > 0$. By the definitions involved

$$\begin{aligned}\tfrac{h-2\varepsilon h}{h-s} \cdot \bar{d}\big(W|_{\varepsilon h - s}^{h-\varepsilon h - s}, V|_{\varepsilon h}^{h-\varepsilon h}\big) \;&\leqslant\; \bar{d}\big(W|_0^{h-s}, V|_s^h\big)\\ &\leqslant\; \bar{d}(W, V \oplus s)\,.\end{aligned}$$

Since $\frac{h-2\varepsilon h}{h-s}$ is greater than $(1 - 2\varepsilon)$, which without loss of generality exceeds $\frac{1}{3}$, the above and (17) imply (17$'$). Finally, define the set of h'-words

$$\mathbf{D}' := \Big\{ V|_{\varepsilon h}^{h-\varepsilon h} \;\Big|\; V \in \mathbf{D} \Big\}\,.$$

This allows us to recapitulate and say that there are maps

$$V'\colon \mathbf{C} \to \mathbf{D}' \quad \text{and} \quad s'\colon \mathbf{C} \to (0, 2\varepsilon h) \cap \mathbb{Z}$$

such that (17$'$) holds for each $W \in \mathbf{C}$, where V' and s' are abbreviating $V'(W)$ and $s'(W)$.

Suppose $\big\{x|_{i_m}^{i_m+h}\big\}_{m=1}^{\infty}$ is a covering of $x|_0^{\infty}$ by $\mathbf{C}$. Let W_m be the color of the mth swatch and set $s'_m := s'(W_m)$. Then (17$'$) implies that

$$\bar{d}\big(x|_{i_m+s'_m}^{i_m+s'_m+h'}, V'(W_m)\big) \;<\; 3\varepsilon\,.$$

Hence $\big\{x|_{i_m+s'_m}^{i_m+s'_m+h'}\big\}_{m=1}^{\infty}$ is a $(1 - 2\varepsilon)$-covering up to 3ε $\bar{d}$-error, by the collection $\mathbf{D}'$. And $|\mathbf{D}'|$ equals $|\mathbf{D}|$, which is strictly less than $\operatorname{rk}(T)$. ♦

Fundamental Lemma. *$\forall\varepsilon$, $\forall$ large n: For each color $W \in \mathbf{C}$ there is a color $V \in \mathbf{C}$ such that W and V are ε-close via a shift s satisfying $s \equiv 1$.*

Deferring the proof of the Fundamental Lemma until the next section, let us see how it implies the theorem.

Rank-of-Powers Theorem. $\operatorname{rk}\big(T^p\big) = p \cdot \operatorname{rk}(T)$.

Proof. Arguing that $\operatorname{rk}(T) \leqslant \operatorname{rk}\big(T^p\big)/p$ will suffice. Fix an ε less than the ε_0 of Lemma 7. Choose some n large enough that that lemma hands us a subcollection $\mathbf{D}$ of $\mathbf{C}\,(=\mathbf{C}_n)$, with $\operatorname{rk}(T) \leqslant |\mathbf{D}|$, such that no two colors in $\mathbf{D}$ are ε-close. Our goal is to prove that $|\mathbf{D}| \leqslant r/p$.

Define an equivalence relation $\sim$ on our set $\mathbf{C}$ of colors, as follows. Say that $W \sim V$ if there exists a finite sequence of colors (in $\mathbf{C}$),

$$W =: V_1, V_2, \ldots, V_{K-1}, V_K := V\,, \tag{18}$$

such that for each k, colors V_k and V_{k+1} are $\frac{\varepsilon}{r}$-close via some shift $s_k \equiv 1$. Evidently there is no need for a color ever to be repeated along a sequence (18). So when $W \sim V$ we can, in fact, find a sequence whose length K is no greater than r. Consequently W and V are ε-close (via the shift $\sum_{k=1}^{K-1} s_k$). We see thus that the relation $\sim$ partitions $\mathbf{C}$ into equivalence classes such that each two colors in the same class are ε-close. Thus $\mathbf{D}$ can have at most one member in each equivalence class. So it suffices to show that there are no more than r/p equivalence classes. We show this by demonstrating that each equivalence class has at least p members.

We can have chosen n sufficiently large that the Fundamental Lemma holds for "$\frac{\varepsilon}{r}$-close" replacing "ε-close". Now given a color, call it V_1, in some equivalence class, we can iteratively apply the Fundamental Lemma $p-1$ times to produce a sequence like (18), with $K := p$. Evidently, for every i and j with $1 \leqslant i < j \leqslant p$, the colors V_i and V_j are ε-close via the shift $\sum_{k=i}^{j-1} s_k$. But each $s_k \equiv 1$, so this sum is congruent mod p to $j-i$. Hence V_i and V_j are ε-close via a *non-zero* shift. Contingent on our having chosen ε small enough and n large enough, the Rigidity Criterion corollary implies that no color of $\mathbf{C}$ can be ε-close to a non-zero shift of itself. Hence V_i and V_j must be distinct colors. Thus, the equivalence class containing V_1 has at least p members. ♦

§**3.** This section arrives at a proof of the Fundamental Lemma. In the preceding section we proved the theorem $\mathrm{rk}(T^p) \geqslant p$, which implied the Rank of Powers theorem in the special case $\mathrm{rk}(T) = 1$. The proof of $\mathrm{rk}(T^p) \geqslant p$ used a local argument on the name $x|_0^\infty$. That is, beneath a single swatch $x|_I^{I+h}$ in the 0-covering we looked at the swatches –their colors and shifts– in the other coverings that overlapped this particular swatch $x|_I^{I+h}$. However to obtain the Rank of Powers theorem in the general case, that is, to prove the Fundamental Lemma, we will use "global" arguments along $x|_0^\infty$. We will contrast the colors and shifts of swatches in the other coverings which overlap *different* swatches of the *same* color in the 0-covering.

Proof of the Fundamental Lemma. We argue by contradiction. If the lemma is false then there exist positive constants ε_0 and σ_0 such that the following holds. For each n (by having dropped to a subsequence and renumbered) there exists a color $U \in \mathbf{C}_n$, call it ***red***, for which: Whenever a color $V \in \mathbf{C}_n$ and positive shift s, with $s \equiv 1$, are such that

$$\bar{d}\big(U|_0^{h-s}, V|_s^h\big) < \varepsilon_0 ,$$

then, of necessity, $s\% > \sigma_0$. In particular (having assumed n sufficiently large that $\bar{d}$-err$(\mathbf{C}_n) \ll \varepsilon_0$),

(19) *If* $x|_i^{i+h}$ *is a red* 0*-swatch and* $x|_j^{j+h}$ *a* 1*-swatch, then* $|j-i|\% > \sigma_0$.

♦

Fix a value n, to be specified later. Recalling the constant κ (from §2, (14)) we know that at least κ-percent of $x|_0^\infty$ is covered by red swatches of the 0-covering. Let $\{i_m\}_{m=1}^\infty$ be the 0-covering of $x|_0^\infty$. If we fix an integer constant $L > \frac{1}{\kappa/4}$, we may assert that along the 0-covering

(20) *Greater than $\frac{3}{4}$-percent of red swatches are such that: Within at most L swatches into the future occurs another red swatch.*

In other words, for at least $\frac{3}{4}$ of red indices i_m, there exists an m', with $L \geqslant m'-m > 0$, such that $x|_{i_{m'}}^{i_{m'}+h}$ is also red.

Now fix a positive number $\sigma < \sigma_0$. The following definitions and arguments can be made using swatches in general. However, we only need use 0-swatches and 1-swatches. Agree henceforth to let "swatch" refer only to swatches of the 0 or 1 coverings.

DEFINITION. Say that a swatch $x|_i^{i+h}$ ***significantly overlaps*** a swatch $x|_j^{j+h}$ if, letting s denote the *shift* $j-i$, we have that $0 \leqslant s < h$ and $s\% < 1-\sigma$. This latter condition says that the swatches must overlap by at least σ-percent. An example is shown in Fig. 23.

DEFINITION. Suppose $x|_i^{i+h}$ is a 0-swatch. Let $x|_{i_\ell}^{i_\ell+h}$ denote the ℓth 0-swatch after it in the 0-covering; so $i_0 = i$. Let j_0 be the largest index commencing a 1-swatch such that $j_0 < i$; let j_ℓ commence the ℓth 1-swatch after this j_0 swatch (see Fig. 21). Say that $x|_i^{i+h}$ commences a ***chain*** if, for $\ell = 1, \ldots, L$:

> *$x|_{i_{\ell-1}}^{i_{\ell-1}+h}$ significantly overlaps $x|_{j_\ell}^{j_\ell+h}$ and this latter significantly overlaps $x|_{i_\ell}^{i_\ell+h}$.*

FIGURE 21. Each swatch drawn above (below) the line significantly overlaps the next swatch in the future below (above) the line.

How to pick n. Note that if $\mu(\mathbf{C}_n) = 1$, then the 0-swatches are contiguous and so are the 1-swatches. Then (19) forces that each red 0-swatch commences a chain.

Our first lower bound on n is that $1-\mu(\mathbf{C}_n)$ be so small that at least $\frac{3}{4}$-percent of the red 0-swatches are forced to commence chains. (We can accomplish this, in light of (19), because σ is strictly smaller than σ_0.) So, combined with (20),

> *Greater than $\frac{\kappa}{2}$-percent of $x|_0^\infty$ is covered by those red 0-swatches which commence a chain, and whose chain contains another red 0-swatch.*

Our second lower bound on n comes from the Rigidity Criterion. Its corollary implies the existence of an ε for which, $\forall$ large n:

(22) *If there exists $W \in \mathbf{C}$ and non-negative integers $0 \leqslant j < j+s+\ell \leqslant h$, with $\ell\% \geqslant \sigma$, such that $\bar{d}\big(W|_j^{j+\ell}, W|_{j+s}^{j+s+\ell}\big) < \varepsilon$, then $s = 0$.*

Now fix an n adequately large to exceed our lower bounds and with $\bar{d}$-err$(\mathbf{C}) < \sigma\varepsilon/4$. As discussed earlier, h can be made so large that we may harmlessly regard σh as an integer.

Recapping the order in which we picked our constants: Given κ, ε_0 and σ_0, we choose L and σ, then ε, and then n.

SHIFT-UNIQUENESS LEMMA. "For each pair of colors there is at most one possible shift value for their ovelap."

For each ordered pair (W, V) of colors, $\exists s$ such that: If $x|_i^{i+h}$ significantly overlaps $x|_j^{j+h}$, where these are any two swatches of color W and V, respectively, then $j - i$ must equal s.

PROOF. Before arguing uniqueness of the shift, s, let us see what an overlap –as in Fig. 23– implies about the colors W and V. By definition,

$$\bar{d}\big(x|_j^{j+h}, V\big) < \bar{d}\text{-err}(\mathbf{C})$$

and hence is less than $\sigma\varepsilon/4$. Consequently

$$\bar{d}\big(x|_j^{j+\sigma h}, V|_0^{\sigma h}\big) \;<\; \varepsilon/4.$$

FIGURE 23. Swatch $x|_i^{i+h}$ significantly overlapping $x|_j^{j+h}$ via a shift s. The length of their common overlap exceeds σ-percent of the length, h, of a swatch.

Similarly, $\bar{d}(W, x|_i^{i+h}) < \sigma\varepsilon/4$ and so

$$\bar{d}(W|_s^{s+\sigma h}, x|_{i+s}^{i+s+\sigma h}) < \varepsilon/4.$$

Since $i + s$ equals j, the triangle inequality yields the following:

Whenever a swatch of color W significantly overlaps a swatch of color V, by a shift of s, then

$$\bar{d}\big(W|_s^{s+\sigma h},\, V|_0^{\sigma h}\big) \;<\; \varepsilon/2\,. \tag{24}$$

Now suppose there were two possible shift values, s' and s'', for (a swatch of) color W to significantly overlap (a swatch of) color V. Then (24) would hold both for $s := s'$ and $s := s''$. By the triangle inequality

$$\bar{d}\big(W|_{s'}^{s'+\sigma h}, W|_{s''}^{s''+\sigma h}\big) \;<\; \varepsilon\,.$$

So by the Rigidity Criterion, in the form (22), we conclude that s' must equal s''. ♦

The Shift-Uniqueness Lemma implies the existence of a bound, independent of n, on the number of different kinds of chains.

To see this, suppose that $x|_{i_0}^{i_0+h}$ commences a chain as in Fig. 21. Let W_ℓ and V_ℓ denote the color of $x|_{i_\ell}^{i_\ell+h}$ and $x|_{j_\ell}^{j_\ell+h}$, respectively. By the preceding lemma, knowledge of the colors $W_{\ell-1}$ and V_ℓ determines the shift $j_\ell - i_{\ell-1}$. Also, V_ℓ and W_ℓ determine $i_\ell - j_\ell$. Consequently, knowledge of the tuple

$$\langle W_0, V_1, W_1, V_2, W_2, \ldots, V_L, W_L\rangle$$

of $2L+1$ colors, determines all the shifts $i_\ell - i_0$, for every ℓ. (That is, for $\ell = 1, \ldots, L$.) There are r^{2L+1} such tuples and so we may restate as follows. *There is a set $\mathcal{S}$ of "possible shifts" (positive integers), with*

$$|\mathcal{S}| \;\leqslant\; L \cdot r^{2L+1}\,,$$

such that for each chain (21), *each ℓ, the shift $i_\ell - i_0$ is an element of $\mathcal{S}$.* The significance of $\mathcal{S}$ is that, although the shifts in the Shift-Uniqueness Lemma depend on n, and hence so does the set $\mathcal{S}$, the *size* of $\mathcal{S}$ is (bounded by a constant) independent of n.

Recall that greater than $\frac{\kappa}{2}$-percent of $x|_0^\infty$ is covered by those red 0-swatches (which we agree to enumerate as $\{x|_{i'_k}^{i'_k+h}\}_{k=1}^\infty$) that commence a chain containing another red 0-swatch. Since $|\mathcal{S}| \leqslant Lr^{2L+1}$, there is some particular shift $s \in \mathcal{S}$ for which

$$\text{Upper Density}(\mathcal{D}) \;\geqslant\; 1/Lr^{2L+1}\,,$$

where $\mathcal{D}$ is the set of $k \in \mathbb{N}$ such that the index $i'_k + s$ commences a red swatch. Define $\{i_m\}_{m=1}^\infty$ to be set $\{i'_k\}_{k\in\mathcal{D}}$ enumerated in ascending order. The upper density of the swatches $\{x|_{i_m}^{i_m+h}\}_{m=1}^\infty$ on $x|_0^\infty$ will exceed the constant

$$\tfrac{1}{2}\kappa/Lr^{2L+1}\,.$$

For every m, each of i_m and $i_m + s$ commences a red swatch and so

$$\bar{d}\big(x|_{i_m}^{i_m+h}, x|_{i_m+s}^{i_m+s+h}\big) \;\leqslant\; 2\cdot\bar{d}\text{-err}(\mathbf{C})\,.$$

By having taken n sufficiently large, we could have made h as large, and $\bar{d}$-err$(\mathbf{C})$, as small as desired. Hence, by the upper density version of the Rigidity Criterion, we arrive at the contradiction

$$\rho(T) \;\geqslant\; \tfrac{1}{2}\kappa/Lr^{2L+1} \;>\; 0\,.$$

This completes the proof of the Fundamental Lemma and hence of the Rank-of-Powers theorem.

Remark. If T is mixing and S is a root of a positive power of T, i.e, $S^q = T^p$, then S is mixing and consequently

$$\text{rk}(S) \;=\; \frac{p}{q}\cdot\text{rk}(T)\,.$$

For T finite rank mixing, it turns out that every S in the commutant of T satisfies a relation $S^q = T^p$, where p and q are integers with q non-zero. This can be exploited to yield a structure theorem for the commutant group of T; this will appear separately.

Question. Does the Rank-of-Powers result hold for non-uniform rank, $\widetilde{\mathrm{rk}}$? For T rank-1 it does hold; an argument similar to that of Theorem 6 yields that $\widetilde{\mathrm{rk}}\big(T^p\big) \geqslant p$.

References

[1] J.L. King, *The commutant is the weak-closure of the powers, for rank-1 transformations*, Ergodic Theory and Dynamical Systems **6** (1986), 363–384.

[2] J.L. King, *Joining-rank and the structure of finite rank mixing transformations*, J. d'Analyse Math. **51** (1988), 182–227, (This was originally called "Remarks on the commutant and factors of finite rank mixing transformations", and was cited under this title. However, the title was changed before publication.).

Filename: `Article/03MixingPowers/mpowers.ams.tex`
As of: *Fri Aug 21, 1998* Typeset: *31May2009*