ISRAEL JOURNAL OF MATHEMATICS 80 (1992), 65–86

ERGODIC PROPERTIES WHERE ORDER 4 IMPLIES INFINITE ORDER

BY

JONATHAN L. KING*

Department of Mathematics, University of Florida
Gainesville, FL 32611, USA

ABSTRACT

For a family of dynamical properties, knowing that the condition holds for order 4 implies that it holds for all orders. Here we establish this for the properties **minimal self-joinings**, **simplicity** and for **cartesian-disjointness.**

An application of the first yields an analog to Kalikow's celebrated result that for rank-1 transformations, 2-fold mixing implies 3-fold mixing. Via a joining argument we show that for any rank-1 $\mathbb{Z}^D$-action, 4-fold mixing implies mixing of all orders. Indeed, the rank need only be sufficiently close to 1 for the implication to hold and so this result is new even when the acting group is $\mathbb{Z}$.

By means of limit-joinings, we settle affirmatively an old open question by establishing, for any M, that M-fold Rényi-mixing implies M-fold mixing.

Introduction

Recent years have shown arguments using joinings to be useful in approaching the "mixing $\Longrightarrow$ multiple mixing" problem of Rohlin. Joinings techniques, when they are applicable, are at once easier and more general than the traditional "coding arguments" of ergodic theory. Subsequent to the results described in this article, which date from 1987, Bernard Host proved a singular-spectrum independence result [Ho], and Shahar Mozes [M] recently established that for actions of certain non-Abelian Lie groups, mixing implies mixing of all orders.

---

* Partially supported by National Science Foundation Postdoctoral Research Fellowship.

Received April 25, 1991 and in revised form October 10, 1991

In this article we prove that for a family of zero-entropy $\mathbb{Z}^D$-actions, 4-fold mixing implies ∞-fold mixing. The Fundamental Lemma appears in §1. "Rényi-mixing implies mixing" commences §2, followed by our principal result, theorem 2.8 and its corollary.

GENERAL NOTATION Agree to let the expression $a \stackrel{\triangleleft}{=} b$ indicate that expression b *defines* the symbol a. For reals a and b let $[a\,..\,b)$ denote the half-open "interval of integers" $[a,b) \cap \mathbb{Z}$, with analogous notation for closed and open intervals. $\mathbb{N}$ is a synonym for $[0\,..\,\infty)$. Use $\#A$ for the cardinality of the set A. All sets and functions are tacitly measurable.

Our context is that of a (measure-preserving) transformation T of a Lebesgue probability space $(X, \mathcal{X}, \mu)$. To indicate that T acts on this space we may write $(T\colon X, \mathcal{X}, \mu)$ or $(T\colon X, \mu)$ or simply $(T\colon \mu)$. Let $C(T)$ denote the **centralizer** of T; the semigroup of transformations which commute with T.

Let **field** be a synonym for "sigma-algebra". As usual, if $\mathcal{A}$ and $\mathcal{B}$ are the fields of two spaces then the symbol $\mathcal{A} \times \mathcal{B}$ denotes the smallest field on the product space which contains all rectangles $A \times B$.

Suppose that $\{\mathcal{Z}_k\}_k$ are subfields of $\mathcal{X}$. They are (collectively) **independent** with respect to μ, written

$$\perp^{\mu}\{\mathcal{Z}_k\}_k \quad \text{or just} \quad \perp\{\mathcal{Z}_k\}_k$$

if, whenever $\{A_k\}_k$ are sets with $A_k \in \mathcal{Z}_k$, then $\mu(\bigcap_k A_k) = \prod_k \mu(A_k)$.

JOININGS NOMENCLATURE For the standard joinings definitions, see [J,R] or [K] or [K,T]; our notation is from the latter. Use $\mathbb{J}(T,S)$ for the space of joinings between T and S and let $\mathbb{J}_{\mathrm{Erg}}(T,S)$ be its subset of ergodic joinings. Given factor maps

$$(T\colon X, \mathcal{X}, \mu) \longrightarrow (R\colon Z, \mathcal{Z}, \zeta) \longleftarrow (S\colon Y, \mathcal{Y}, \nu),$$

let $\mu \times_{\mathcal{Z}} \nu \in \mathbb{J}(T,S)$ denote the relative independent joining of T with S relative to the two factor maps.

NOTATION FOR COUNTABLE JOININGS It is occasionally convenient to talk about joinings of countably many maps. Agree to let $\bigotimes$ denote cartesian product: of transformations, spaces, fields, and measures. For typographical reasons, we use $\lfloor$ to mean "restricted to"; thus $\xi|_{\mathcal{Y}}$ and $\xi\lfloor\mathcal{Y}$ are synonyms.

Fix a $K \in \mathbb{N} \cup \{\infty\}$ and consider a list of transformations $(T_k\colon X_k, \mathcal{X}_k, \mu_k)$ for $k \in [0\,..\,K)$. A joining $\xi \in \mathbb{J}(\{T_k\}_k)$ is a $(\bigotimes_k T_k)$-invariant probability measure on $(\bigotimes_k X_k, \bigotimes_k \mathcal{X}_k)$ such that each $\xi\lfloor\mathcal{X}_k = \mu_k$. Suppose $\xi \in \mathbb{J}(\{T_k\}_k)$ and I is some subset of the index set $[0 .. K)$, say, of cardinality M. Then $\xi|_I$ shall denote **the M-fold marginal** of ξ upon I. That is,

$$\xi|_I \stackrel{\triangleleft}{=} \xi\lfloor\bigvee_{k\in I}\mathcal{X}_k,$$

which is an element of $\mathbb{J}(\{T_k \mid k \in I\})$.

DIAGONAL JOININGS Agree to let $\Delta_{\mu,M}$ or just Δ_M denote the M-fold diagonal joining defined on rectangles by

$$\Delta_M\Big(\bigotimes_{m\in[0\,..\,M)} A_m\Big) \stackrel{\triangleleft}{=} \mu\Big(\bigcap_{m\in[0\,..\,M)} A_m\Big) \qquad \text{for } A_m \in \mathcal{X}.$$

Given a measurable map $f\colon X \to Y$ and a measure μ on $(X, \mathcal{X})$, let $[\![f]\!]\mu$ denote the measure on Y of $A \mapsto \mu(f^{-1}(A))$. As an example, if ξ is a joining of transformation T_1 with T_2 and if $S_i \in C(T_i)$, then $[\![S_1 \times S_2]\!]\xi$ is the joining of T_1 with T_2 defined on rectangles by

$$[\![S_1{\times}S_2]\!]\xi(A_1 \times A_2) \stackrel{\triangleleft}{=} \xi(S_1^{-1}A_1 \times S_2{}^{-1}A_2).$$

COPIES To speak of self-joinings of a map $(T\colon X, \mu)$ it will be convenient to use a subscript (occasionally a superscript) within angle-brackets to refer to a **copy** of a transformation, space, field, measure or joining. Thus one may use $\mathbb{J}(T,T,T)$ or $\mathbb{J}(\{T\}_{m=1}^3)$ or $\mathbb{J}(\{T_{\langle m\rangle}\}_{m=1}^3)$ to denote the collection of 3-fold self-joinings of T. For a $\xi \in \mathbb{J}(T,T,T)$, the field on which ξ lives might be written as $\mathcal{X}_{\langle 1\rangle} \times \mathcal{X}_{\langle 2\rangle} \times \mathcal{X}_{\langle 3\rangle}$ or –by viewing each $\mathcal{X}_{\langle m\rangle}$ as a subfield of the whole– as $\mathcal{X}_{\langle 1\rangle} \vee \mathcal{X}_{\langle 2\rangle} \vee \mathcal{X}_{\langle 3\rangle}$.

1. Dynamical properties arising from joinings

In this section we show that for three joinings notions –Rudolph's **minimal self-joinings** [R], Veech's "property S" [V], also called **simplicity**, and Furstenberg's notion [Fur] of **disjointness** of transformations– that if the property holds up to order four, it holds for all orders.

MINIMAL SELF-JOININGS Suppose ξ is an N-fold self-joining of T. A marginal $\xi|_I$ is an **off-diagonal** if there exists $p\colon I \to \mathbb{Z}$ such that

$$\xi|_I = \Big[\!\!\Big[\bigotimes_{n\in I} T^{p(n)}\Big]\!\!\Big]\,\Delta_M$$

where M here denotes $\#I$. An N-fold self-joining ξ is a "product of off-diagonals" if the index set $[0 \mathinner{..} N)$ can be written as a disjoint union $I(1) \sqcup I(2) \sqcup \ldots$ such that the following holds.

(i) For each ℓ: $\xi|_{I(\ell)}$ is an off-diagonal.

(ii) ξ is the direct product of these marginals. In other words, the corresponding subfields are collectively independent: $\perp^{\xi} \{\bigvee_{n \in I(\ell)} \mathcal{X}_{\langle n \rangle} \mid \ell = 1, 2, \ldots\}$.

Products of off-diagonals are called **trivial** N-fold self-joinings and T is said to have **N-fold minimal self-joinings** if $\mathrm{J}_{\mathrm{Erg}}(\{T \mid n \in [0 \mathinner{..} N)\})$ consists only of trivial joinings.

SIMPLICITY Map T is **N-fold simple** if for each ξ in $\mathrm{J}_{\mathrm{Erg}}(\{T_{\langle n \rangle}\}_n)$ the index set $[0 \mathinner{..} N)$ may be written disjointly as $I(1) \sqcup I(2) \sqcup \ldots$ so that –in addition to (ii) above– the following holds. For each ℓ: All the $\mathcal{X}_{\langle n \rangle}$, $n \in I(\ell)$, are equal with respect to ξ. In other words, weaken (i) to

(i′) For each ℓ: $\xi|_{I(\ell)}$ is a **graph self-joining** of T ie., is of the form $[\![\bigotimes_{n \in I(\ell)} R_n]\!] \Delta_M$ where $M \stackrel{\triangle}{=} \#I(\ell)$ and each R_n is in the centralizer of T. Thus "N-fold minimal self-joinings" is merely "N-fold simple with trivial centralizer".

Definition: Given a collection of transformations $\{(T_k \colon X_k, \mu_k)\}_k$, a joining $\xi \in \mathrm{J}(\{T_k\}_k)$ is said to be a **pairwise independent joining** if $\mathcal{X}_j \perp^{\xi} \mathcal{X}_k$ for each pair $j \neq k$. Further, ξ is the independent joining if $\perp^{\xi} \{\mathcal{X}_k\}_k$.

A collection $\{T_k\}$ of transformations is **bi-independent*** if the only pairwise independent joining is the independent joining. For a single map T, say that it is N-fold bi-independent if the collection $\{T_{\langle n \rangle}\}_{n=0}^{N-1}$ is bi-independent. Maps T and S are **autonomous** if the collection

$$\{T_{\langle 0 \rangle}, S_{\langle 0 \rangle}, T_{\langle 1 \rangle}, S_{\langle 1 \rangle}\}$$

is bi-independent. Map T is **self-autonomous** if T and T are autonomous. ∎

For $k = 0, 1$, suppose that ξ_k is a joining of measure space (X_k, μ_k) with (Y, ν). If $\mathcal{X}_0 \perp \mathcal{Y}$ with respect to ξ_0, then trivially $\mathcal{X}_0$ and $\mathcal{X}_1$ are independent

* In general, given a 2-fold property "P", say that a list $\{T_k\}_k$ is "pairwise P" if for every pair $j \neq k$ the couple (T_j, T_k) has property P. By way of contrast, bi-independence is not *a priori* a property verifiable only by looking at pairs of transformations.

with respect to the relative independent joining $\xi_0 \times_{\mathcal{Y}} \xi_1$. The following partial converse appears in [J,R]. For completeness we include its brief proof.

PROPOSITION 1.1 (Jensen's Inequality): *Suppose ξ is a joining of (X,μ) with (Y,ν). Let*

$$\rho \stackrel{\triangleleft}{=} \xi_{\langle 0\rangle} \times_{\mathcal{Y}} \xi_{\langle 1\rangle}$$

be the relative independent (self-)joining of ξ over $\mathcal{Y}$. Then $\mathcal{X}_{\langle 0\rangle} \perp^{\rho} \mathcal{X}_{\langle 1\rangle}$ implies $\mathcal{Y} \perp^{\xi} \mathcal{X}$.

Proof: Decompose ξ into $\{\xi_y \mid y \in Y\}$, its fiber measures over $\mathcal{Y}$. For any $A \in \mathcal{X}$ this gives, by definition, the first equality below.

$$\left[\int_Y \xi_y(A)\, d\nu(y)\right]^2 = \mu(A)^2 = \int_Y \xi_y(A)^2\, d\nu(y).$$

The second equality follows from the independence of $\mathcal{X}_{\langle 0\rangle}$ and $\mathcal{X}_{\langle 1\rangle}$ with respect to ρ, since the righthand side is $\rho(A \times A)$. We thus have equality in Jensen's inequality and must conclude that the mapping $y \mapsto \xi_y(A)$ is constant ν-a.e. But this is true for every $A \in \mathcal{X}$; in particular, for a countable ν-dense collection of sets A. Thus, after discarding a ν-nullset of y, all the fiber measures $\{\xi_y \mid y \in Y\}$ are identical; hence $\xi = \mu \times \nu$. ∎

Here is the trick which lifts 4-fold independence to higher order independence.

FUNDAMENTAL LEMMA: *Suppose T_0 and T_1 are autonomous. Then for any transformation $(S\colon Y)$ the collection $\{S, T_0, T_1\}$ is bi-independent.*

Proof: Consider any $\xi \in \mathbb{J}(S, T_0, T_1)$ for which, with respect to ξ,

$$\mathcal{Y} \perp \mathcal{X}_0, \quad \mathcal{Y} \perp \mathcal{X}_1 \tag{1.2}$$

$$\mathcal{X}_0 \perp \mathcal{X}_1 \tag{1.3}$$

and let $\rho \stackrel{\triangleleft}{=} \xi^{\langle 0\rangle} \times_{\mathcal{Y}} \xi^{\langle 1\rangle}$. From (1.2) we have $\mathcal{X}_j^{\langle 0\rangle} \perp^{\rho} \mathcal{X}_k^{\langle 1\rangle}$ for any $j,k \in \{0,1\}$. Since (1.3) yields that $\mathcal{X}_0^{\langle n\rangle} \perp^{\rho} \mathcal{X}_1^{\langle n\rangle}$ for $n = 0,1$, we may conclude from autonomy that the four fields are collectively independent:

$$\perp^{\rho}\{\mathcal{X}_k^{\langle n\rangle} \mid n = 0,1 \text{ and } k = 0,1\}.$$

A fortiori, then, $[\mathcal{X}_0 \vee \mathcal{X}_1]^{\langle 0\rangle} \perp^{\rho} [\mathcal{X}_0 \vee \mathcal{X}_1]^{\langle 1\rangle}$ and so by the preceding proposition one may assert

$$\mathcal{Y} \perp^{\xi} [\mathcal{X}_0 \vee \mathcal{X}_1].$$

This, together with (1.3) yields that algebras $\mathcal{Y}$, $\mathcal{X}_0$ and $\mathcal{X}_1$ are collectively independent with respect to ξ, as desired. ∎

FOUR-FOLD THEOREM 1.4:

(a) *If the collection of maps $\{T_k \| k \in [0 .. K)\}$ is pairwise autonomous then the collection is bi-independent.*

In particular, a 4-fold bi-independent map (self-autonomous map) is ∞-fold bi-independent.

(b) *If T is 4-fold simple then it is simple of all orders.*

(c) *If T has 4-fold minimal self-joinings then it has minimal self-joinings of all orders.*

Proof: The $K = \infty$ case follows from the finite case; assume $K < \infty$.

For part (a), $K = 2$ is a tautology. Proceeding by induction, suppose we have $K + 1$ maps $T_0, \ldots, T_K$ for which every subcollection consisting of K maps is bi-independent. In particular, the bi-independence of $\{T_2, \ldots, T_K, T_0\}$ and of $\{T_2, \ldots, T_K, T_1\}$ forces any pairwise independent joining of $\{T_k\}_{k=0}^{K}$ to satisfy

$$\mathcal{X}_i \perp \bigvee_{k=2}^{K} \mathcal{X}_k, \qquad \text{for } i = 0, 1.$$

Thus this joining can be viewed as a pairwise independent joining of $\{S, T_0, T_1\}$, where S denotes $T_2 \times \cdots \times T_K$. But such a joining must actually *be* independent, by the preceding lemma.

Proof of (b,c): Evidently, 4-fold simplicity implies that T is 4-fold bi-independent. Hence T is bi-independent of all orders.

Take N minimal such that T fails to be N-fold simple. Then there exists ξ, an ergodic N-fold self-joining of T, which is not a product of graphs. Since N is minimal, no two of the N fields $\mathcal{X}_{\langle n \rangle}$ could be identified under ξ. Two-fold simplicity, then, implies that ξ is a pairwise independent joining. By bi-independence, ξ is the independent joining. ∎

DISJOINTNESS In his 1967 paper, [Fur], Furstenberg defines maps T and S to be **disjoint** if $\mathbb{J}(T, S)$ consists of only one point, product measure. One motivation for this was the question: When can the transmitted information be extracted from a signal which has been corrupted by noise? If the information comes from one kind of process, and the noise from another, then the information can be recovered if the two processes are disjoint in the above sense. There are three well-known pairs of classes of transformations which are disjoint:

$$\begin{array}{rl} \text{K-AUTOMORPHISM} & \text{ZEROENTROPY} \\ \text{MILDMIXING} & \text{RIGID} \\ \text{WEAKMIXING} & \text{ROTATION} \end{array} \tag{1.5}$$

The first line means that any K-automorphism is disjoint from any map with zero-entropy.

His second motivation was seeing how far the notion of "disjointness" for transformations emulates "co-primeness" for integers. In particular, PROBLEM B from [Fur, p.7]

If T is disjoint from S_1 and from S_2, must it be disjoint from their product $S_1 \times S_2$?

remains open. (The non-trivial case is when three maps are weak-mixing. For rotations it is false simply by taking three irrational rotation numbers, pairwise rationally-independent, but with the triple rationally-dependent.) Furstenberg's question remains open even if one symmetrizes the question and assumes that the three maps are pairwise disjoint. However, the application below shows that for a mildly stronger form of disjointness –the one which seems to arise in practice– pairwise implies collective.

Say that maps T and S are **cartesian-disjoint** if their countable cartesian powers $T^{\times\mathbb{N}}$ and $S^{\times\mathbb{N}}$ are disjoint. The motivation for this definition is that all six classes of (1.5) are closed under countable cartesian power; "ROTATION" meaning the class of compact-group rotations, not necessarily ergodic. Consequently, the three pairs of (1.5) are cartesian-disjoint.

DISJOINTNESS THEOREM 1.6:

(a) *If $\{T_k\}_k$ are pairwise disjoint and pairwise autonomous then the $\{T_k\}_k$ are collectively disjoint.*

(b) *Pairwise cartesian-disjointness $\Longrightarrow$ collective cartesian-disjointness.*

Proof of (b): Part (a) follows from (1.4a). To justify part (*b*), suppose T and S are cartesian-disjoint. Take ξ a pairwise independent joining of

$$T^{\times\mathbb{N}}_{\langle 1\rangle}, S^{\times\mathbb{N}}_{\langle 1\rangle}, T^{\times\mathbb{N}}_{\langle 2\rangle}, S^{\times\mathbb{N}}_{\langle 2\rangle}. \tag{1.7}$$

Then we can view ξ as a joining of $T^{\times\mathbb{N}} \times T^{\times\mathbb{N}}$ with $S^{\times\mathbb{N}} \times S^{\times\mathbb{N}}$. But these transformations are isomorphic to $T^{\times\mathbb{N}}$ and $S^{\times\mathbb{N}}$ respectively. Since the latter pair are disjoint, ξ is the independent joining of the four maps of (1.7). Thus $T^{\times\mathbb{N}}$ and $S^{\times\mathbb{N}}$ are autonomous.

Given a pairwise cartesian-disjoint collection $\{T_k\}_k$, we conclude that $\{T_k^{\times\mathbb{N}}\}_k$ are pairwise autonomous. By (a) then, $\{T_k^{\times\mathbb{N}}\}_k$ are collectively disjoint. ∎

Remark: This theorem is applicable in situations not included in (1.5). For instance, it is not difficult to build pairs of rigid weak-mixing maps which are cartesian-disjoint. Say that $\{n_k\}_1^\infty$ is a "mixing sequence" for T if $\mu(A \cap T^{-n_k} A) \to \mu(A)^2$ for all sets A. Call it a "rigidity sequence" for S if $\nu(B \triangle S^{-n_k} B) \to 0$ for all B. A consequence of the Mean Ergodic Theorem is the following.

(1.8) If there exists a sequence $\{n_k\}_1^\infty$ which is mixing for T while being rigid for S, then T and S are disjoint.

Hence they are cartesian-disjoint, since $\{n_k\}_1^\infty$ applies equally well to $T^{\times\mathbb{N}}$ and $S^{\times\mathbb{N}}$. Examples of transformations satisfying (1.8) can be built by simultaneous cutting & stacking. ∎

Remark: If the direct product $T \times S$ of weak-mixing maps is rank-1 then $\{T, S\}$ satisfies a type of disjointness stronger than cartesian-disjointness: Any joining of (countably many) powers of T is disjoint from any joining of powers of S, once one of these joinings is ergodic; see [F,G,K;§0]. Via simultaneous cutting&stacking it is possible to fabricate three weak-mixing maps $\{T_k\}_{k=1}^3$ with $T_j \times T_k$ rank-1 for j and k distinct, for which the Disjointness Theorem provides the only proof known that $\{T_1, T_2, T_3\}$ is triply disjoint. ∎

QUESTION The issue of whether pairwise disjointness of three weak-mixing transformations implies triple disjointness, remains open. If a counterexample exists, one exists with T_1, T_2 and T_3 each of zero entropy*. Also, in light of the Fundamental Lemma, no pair of the T_k is autonomous. Symmetrizing the argument of (1.4a) in the $K = 4$ case shows: If $\{T_1, T_2\}$ is autonomous and $\{S_1, S_2\}$ is autonomous, then $\{T_1, T_2, S_1, S_2\}$ is bi-independent. *A fortiori*

(1.9) If maps T and S are each self-autonomous then $\{T, S\}$ is autonomous.

---

* This follows from the existence of a maximal zero-entropy factor [Pinsker's theorem] and –for a positive entropy map– of Bernoulli factors [Sinai's theorem].

So if $\{T_1, T_2, T_3\}$ are pairwise but not collectively disjoint then some T_k fails to be self-autonomous. Which demands the following question.

Is every weak-mixing zero-entropy map self-autonomous?

The hope is to put ourselves out of business with an affirmative answer. Then autonomy is vacuously satisfied and 2-fold minimal self-joinings, simplicity, and disjointness would imply their infinite order counterparts.

2. Mixing and rank-1

The goal of this section is to build a context where the trick of §1 gives "4-fold $\Longrightarrow$ M-fold" for the property of mixing.

THE TOPOLOGY OF JOINING-SPACE The tool we will use is the notion of a "limit-joining" and for this we will need to topologize the space of joinings. For probability spaces (X, μ) and (Y, ν), let $\mathbb{M}(\mu, \nu)$ be the space of joinings of the two measures μ and ν. There is a canonical topology on $\mathbb{M}(\mu, \nu)$ defined by this notion of convergence: $\xi_n \to \xi_\infty$ if and only if

$$\lim_{n\to\infty} \left|\xi_n(A \times B) - \xi_\infty(A \times B)\right| = 0$$

for each rectangle $A \times B$. Since our spaces are Lebesgue, this is a metric topology under which $\mathbb{M}(\mu, \nu)$ is compact. Letting $\{A_j\}_1^\infty$ and $\{B_k\}_1^\infty$ be measure-theoretically dense collections inside of $\mathcal{X}$ and $\mathcal{Y}$ respectively, the (non-canonical) metric

$$\mathrm{dist}(\xi, \gamma) \stackrel{\triangle}{=} \sum_{j,k=1}^{\infty} \frac{1}{2^{j+k}} \left|\xi(A_j \times B_k) - \gamma(A_j \times B_k)\right|$$

realizes the topology. This definition generalizes to topologize the space $\mathbb{M}(\mu_1, \mu_2, \dots)$ of countable-fold joinings, making it compact –which yields the following. Agree to use the symbol "$3 \bullet \mu$" to mean the measure $A \mapsto 3\mu(A)$. Given two finite measures λ and ξ on the same space, let "$\lambda \geq \xi$" mean that $\lambda(A) \geq \xi(A)$ for all sets A.

OBSERVATION 2.1: *Fix a joining ξ in $\mathbb{M} = \mathbb{M}(\mu_1, \dots, \mu_K)$ and let $\mathbb{F}$ be a closed subset of $\mathbb{M}$. Then, the set*

$$\left\{\kappa \in [0,1] \,\middle|\, \exists \lambda \in \mathbb{F} \text{ with } \lambda \geq \kappa \bullet \xi\right\}$$

is closed.

An example of a closed subset of $\mathbb{M}(\mu,\nu)$ is $\mathbb{J}(T,S)$, where T preserves μ and S preserves ν. For a K-vector $\mathbf{v} \stackrel{\triangleleft}{=} (v[1],\ldots,v[K])$ in $\mathbb{Z}^{\times K}$, let $\Delta_K^{\mathbf{v}}$ denote the K-fold off-diagonal $[\![T^{v[1]}\times\ldots\times T^{v[K]}]\!]\Delta_K$. In the space $\mathbb{J}(T,\stackrel{K}{\ldots},T)$ the set of off-diagonal joinings is not closed. Its closure, the set of **limit-joinings**

$$\mathbb{J}_{\mathrm{Lim}}(T,\stackrel{K}{\ldots},T) \stackrel{\triangleleft}{=} \mathrm{Closure}\left(\left\{\Delta_K^{\mathbf{v}} \mid \mathbf{v}\in\mathbb{Z}^{\times K}\right\}\right)$$

will play an useful role in the following proof, which serves as a practice run for limit-joining arguments.

MIXING AND RÉNYI-MIXING Recall that T is M-fold mixing if for any choice of M sets $\{A_m\}$

$$\lim_{\|\mathbf{v}\|\to\infty} \mu\big(T^{-v[1]}A_1\cap\cdots\cap T^{-v[M]}A_M\big) = \mu(A_1)\cdot\ldots\cdot\mu(A_M) \tag{2.2}$$

where $\mathbf{v}$ ranges over all M-vectors of integers and $\|\mathbf{v}\|$ denotes the minimum of $\big|v[i]-v[j]\big|$ taken over all distinct i and j.

Rényi, in an elegant Hilbert space argument, showed that the property

$$\forall A\colon\ \mu(A\cap T^{-n}A)\to\mu(A)^2$$

implies the apparently stronger property of 2-fold mixing. With this as inspiration, say that T is **M-fold Rényi-mixing** if (2.2) holds whenever $A_1=\cdots=A_M$.

THE LIMIT-JOINING CHARACTERIZATION OF MIXING The Hilbert space argument does not seem to work immediately for $M>2$. However, by viewing mixing as a statement about limit-joinings, a brief argument demonstrates that M-fold Rényi-mixing implies M-fold mixing.

For a sequence of M-vectors $\mathbf{v}_1,\mathbf{v}_2,\dots\in\mathbb{Z}^{\times M}$ whose corresponding off-diagonal limit $\lambda \stackrel{\triangleleft}{=} \lim_n \Delta_M^{\mathbf{v}_n}$ exists, say that λ is a "non-trivial" limit-joining if $\|\mathbf{v}_n\|\to\infty$. This gives the following characterization.

> *A transformation T is M-fold mixing if and only if every non-trivial limit-joining is M-fold product measure.*

The "if" direction follows from the compactness of $\mathbb{J}(T,\stackrel{M}{\ldots},T)$.

THEOREM, 2.3: "Any joining which looks like product measure on cartesian power sets is product measure." *Fix a weak-mixing T and a natural number M. If a joining $\lambda \in \mathbb{J}(\{T_{\langle m\rangle} \mid 0 \leq m < M\})$ is such that*

$$\lambda(A \times \overset{M}{\cdots} \times A) = [\mu(A)]^M, \qquad \text{for each set } A, \tag{2.4}$$

then λ is M-fold product measure, $\mu^{\times M}$.

Proof: If false, then λ must fail to be absolutely continuous with respect to product measure, since the latter is ergodic. Thus there can exist no bound K such that $\lambda(\mathbf{B}) \leq K \cdot \mu^{\times M}(\mathbf{B})$ for all sets $\mathbf{B} \in \mathcal{X}^{\times M}$; indeed, this inequality must fail for some rectangle. Setting $K \stackrel{\triangleleft}{=} M^M$, then, there must exist sets $A_1, \ldots, A_M \in \mathcal{X}$ such that

$$\lambda(A_1 \times \cdots \times A_M) > M^M \cdot \mu^{\times M}(A_1 \times \cdots \times A_M).$$

Pick ε sufficiently small that if we discard from each A_m any set of mass less than ε, then the above inequality persists. Thus, we may assume that for each m the value $\mu(A_m)$ is a multiple of ε.

Now partition each A_m into disjoint pieces of mass ε. Evidently we can choose a piece from each A_m and call it A_m so as to have the foregoing inequality persist. The improvement is that, now, $\mu(A_1) = \cdots = \mu(A_M)$. So we can set $E \stackrel{\triangleleft}{=} \bigcup_{m=1}^M A_m$ and compute

$$\begin{aligned}\lambda(\underbrace{E \times \cdots \times E}_{M \text{ times}}) \geq \lambda(A_1 \times \cdots \times A_M) &> M^M \cdot \mu(A_1) \cdots \mu(A_M) \\ &\geq M^M \cdot \frac{\mu(E)}{M} \cdots \frac{\mu(E)}{M} = \mu(E)^M.\end{aligned}$$

But (2.4) flatly contradicts this. ∎

COROLLARY 2.5: *If T is M-fold Rényi-mixing then it is M-fold mixing.*

The proof of (2.3) actually establishes the following inequality between joinings of measures.

LEMMA 2.3′: *Suppose $\lambda \in \mathbb{M}(\mu, \overset{M}{\ldots}, \mu)$ and set*

$$B_0 \stackrel{\triangleleft}{=} \sup_A \lambda(A \times \cdots \times A)/\mu(A)^M,$$

where the supremum is taken over all sets A with positive mass. Then

$$\lambda \leq B_0 M^M \bullet \mu^{\times M}.$$

HIGHER ORDER MIXING FROM LOWER ORDER In [Ka] Steve Kalikow proved that for rank-1 $\mathbb{Z}$-actions, 2-fold mixing implies 3-fold mixing. He informs me that the same method, plus elbow grease, yields mixing of all orders.

We hope to prove a related theorem for $\mathbb{Z}^D$-actions (Corollary 2.9) by combining the results of §1 with the methods of [K,§3]. Since our methods need 4-fold mixing to get started, our results are weaker than Kalikow's—however, they are not contained in his theorem even in the $D = 1$ case. Concisely, but imprecisely: *Suppose* $\mathbf{T}$ *is a* $\mathbb{Z}^D$*-action which is (sufficiently close to being) rank-1. If* $\mathbf{T}$ *is 4-fold mixing then* $\mathbf{T}$ *is mixing of all orders.*

This holds also for $\mathbb{R}^D$-actions and should work for actions of a large class of amenable groups $\mathbb{G}$. At issue is what "rank-1" should mean. That $\mathbf{T}$ has a generating sequence of Rohlin stacks indexed by Følner sets is too permissive—indeed, that definition should be the amenable group version of Jean-Paul Thouvenot's notion "funny rank-1", [Fe], which is qualitatively more general than rank-1.

Moreover, in the proof of our Proposition B, the geometry of individual Følner sets will turn out to be important. It is for this reason –to avoid stating our result with complicated hypotheses on the shapes of Følner sets– that we use $\mathbb{G} = \mathbb{Z}^D$, where the standard notion of rank-1 uses well-understood Følner sets: rectangles. Nonetheless, those arguments which work just as easily for general Abelian groups will be stated that way. In particular, we take pains to avoid using the Pointwise Ergodic Theorem which, unlike the Mean Ergodic Theorem, does not hold for arbitrary Følner sequences.

GROUP ACTIONS Suppose $(\mathbb{G}, \mathcal{G}, \eta)$ is an Abelian locally-compact topological group with $\eta(\cdot)$ the invariant Haar measure, $\mathcal{G}$ the Borel field, and $\mathbf{0}$ the identity element. A **$\mathcal{G}$-action** on probability space $(X, \mathcal{X}, \mu)$ is a measurable map

$$\mathbf{T}\colon \mathbb{G} \times X \to X$$

which, agreeing to write $\mathbf{T}^v(\cdot)$ for $\mathbf{T}(v, \cdot)$, satisfies: $\mathbf{T}^v$ *is a measure-preserving bijection of* $(X, \mathcal{X}, \mu)$ *and* $\mathbf{T}^v \circ \mathbf{T}^w = \mathbf{T}^{v+w}$ *for all elements* $v, w \in \mathbb{G}$. In particular, $\mathbf{T}^{\mathbf{0}}$ is the identity transformation.

ERGODICITY For a $\mathbb{G}$-action, saying that a set, function, subfield or joining is "$\mathbf{T}$-invariant" shall mean that it is $\mathbf{T}^v$-invariant for each $v \in \mathbb{G}$. The ergodic decomposition theorem holds; see theorem 1.7 of [J,R]. The corollary we need from this concerns invariant (probability) measures: *If* μ_1 *and* μ_2 *are* $\mathbf{T}$*-ergodic measures on* $(X, \mathcal{X})$ *which are unequal, then they are mutually singular.*

JOININGS The vocabulary of joinings carries over unchanged to general group actions. The centralizer $C(\mathbf{T})$ is the set of transformations S such that $\mathbf{T}^v S = S\mathbf{T}^v$ for every $v \in \mathbb{G}$; if each such S is some $\mathbf{T}^v$ then $\mathbf{T}$ has **trivial centralizer.**

The collection $\mathbb{J}(\mathbf{T},\mathbf{S})$ of joinings of two $\mathbb{G}$-actions is the set of $\xi \in \mathbb{M}(\mu,\nu)$ such that the equality $[\![\mathbf{T}^v \times \mathbf{S}^v]\!]\xi = \xi$ holds for every v. The aforementioned ergodic decomposition theorem implies, $\mathbf{T}$ and $\mathbf{S}$ being ergodic $\mathbb{G}$-actions, that the ergodic components of a joining $\xi \in \mathbb{J}(\mathbf{T},\mathbf{S})$ are themselves joinings. Thus, every element of $\mathbb{J}(\mathbf{T},\mathbf{S})$ is an average of members of $\mathbb{J}_{\mathrm{Erg}}(\mathbf{T},\mathbf{S})$.

MIXING FOR $\mathbb{G}$-ACTIONS For $\mathbf{T}$ to be weak-mixing means that $\mathbf{T}\times\mathbf{T}$ is ergodic. Given a sequence $\{v_n\}_{n=1}^{\infty}$ of elements of $\mathbb{G}$, let "$v_n \to \infty$" mean that for each compact $\mathrm{E} \subset \mathbb{G}$ there exists N such that $v_n \in \mathrm{E}^c$ for all n exceeding N. Let $\alpha_K(\mathbf{T})$ denote the degree of **partial-mixing** of $\mathbf{T}$; the maximum $\alpha \in [0,1]$ so that for all K-fold rectangles

$$\liminf_{\|\mathbf{v}\|\to\infty} \mu\big(T^{-v[1]}A_1 \cap \cdots \cap T^{-v[K]}A_K\big) \geq \alpha \cdot \mu(A_1)\cdot \ldots \cdot \mu(A_K).$$

The expression "$\|\mathbf{v}\| \to \infty$" is a shorthand which means: For i and j distinct, the difference $v[i]-v[j]$ is eventually outside of any given compact set. $\mathbf{T}$ is K-fold mixing if $\alpha_K = 1$; equivalently, if the above liminf is a limit and equality holds with $\alpha = 1$.

ALL PREVIOUSLY DEVELOPED JOINING RESULTS HOLD FOR $\mathbb{G}$-ACTIONS Replacing the word "transformation" by "$\mathbb{G}$-action", the results of FUNDAMENTAL LEMMA, FOUR-FOLD THEOREM, DISJOINTNESS THEOREM as well as the Rényi-mixing result, remain true. All the arguments heretofore are abstract and use nothing about the acting group. Indeed, all that one needs for $\mathbb{G}$-actions is that the relative independent joining over invariant fields is invariant and that the ergodic decomposition theorem holds.

Theorem 2.8 is the goal of the next two sections, whose proof consists of two limit-joining arguments Propositions A and B below. Note that an off-diagonal joining $\Delta_K^{\mathbf{v}}$ is unchanged by adding a constant vector $(c,\ldots,c)$ to $\mathbf{v}$ and so the joining is completely specified by knowing its "difference function"

$$e(i,j) \stackrel{\triangleleft}{=} v[i]-v[j] \qquad \text{for } i,j \in [1\mathinner{..}K].$$

Conversely, given a function $e(\cdot,\cdot)$ such that $e(i,j)+e(j,k)+e(k,i)$ is always zero, and given a subset $I \subset [1\mathinner{..}K]$, let $\Delta_{\mathrm{on}\ I}^{e(\cdot,\cdot)}$ denote the off-diagonal joining $[\![\bigotimes_{i\in I} T^{v[i]}]\!]\Delta_{\#I}$.

PROPOSITION A: *For any limit-joining* $\lambda \in \mathbb{J}_{\mathrm{Lim}}(\mathbf{T}, \overset{K}{\ldots}, \mathbf{T})$ *there exists a trivial* K*-fold joining* τ *such that*

$$\alpha_K(\mathbf{T}) \bullet \tau \leq \lambda.$$

Proof: Suppose λ is the limit-joining $\Delta_K^{\vec{\mathbf{s}}} \stackrel{\triangleleft}{=} \lim_{n\to\infty} \Delta_K^{\mathbf{s}_n}$, where $\vec{\mathbf{s}} = \{\mathbf{s}_n\}_1^\infty$ and each $\mathbf{s}_n$ is a K-vector with components $s_n[i] \in \mathbb{G}$. By the local compactness of $\mathbb{G}$, we can subsequence on n to arrange that now the $\{\mathbf{s}_n\}_1^\infty$ "converges" coordinatewise. That is, we can write the index set $[1 \,..\, K]$ as a disjoint union $\bigsqcup \mathcal{U}$ so that: For each $I, J \in \mathcal{U}$, for each $i \in I$ and $j \in J$,

$$\begin{array}{ll} \text{If } I = J: & e(i,j) \stackrel{\triangleleft}{=} \lim\limits_{n\to\infty} \big(s_n[i] - s_n[j]\big) \text{ exists;} \\ \text{If } I \neq J: & \lim\limits_{n\to\infty} \big(s_n[i] - s_n[j]\big) = \infty. \end{array}$$

The upper line says that the marginal $\Delta^{\vec{\mathbf{s}}}|_I$ equals $\Delta_{\text{on } I}^{e(\cdot,\cdot)}$. The lower line and a short argument yield

$$\Delta^{\vec{\mathbf{s}}} \geq \alpha_{\#\mathcal{U}}(\mathbf{T}) \bullet \tau, \qquad \text{where } \tau = \Big(\bigotimes_{I:\, I \in \mathcal{U}} \Delta_{\text{on } I}^{e(\cdot,\cdot)} \Big).$$

Since $\#\mathcal{U} \leq K$, certainly $\alpha_{\#\mathcal{U}}$ dominates α_K, completing the proof. ∎

RANK AND COVERING NUMBER Henceforth $\mathbb{G} = \mathbb{Z}^D$, for some fixed natural number D. A **rectangle** $\mathrm{R} \subset \mathbb{Z}^D$ is a set of the form

$$(2.6) \qquad [a_1 \,..\, b_1) \times \cdots \times [a_D \,..\, b_D), \qquad \text{with } 0 < \ell_d < \infty$$

where $\ell_d \stackrel{\triangleleft}{=} b_d - a_d$. Let len(R) denote the minimum of $\{\ell_1, \ldots, \ell_d\}$. A sequence $\vec{\mathrm{R}} = \{\mathrm{R}_n\}_1^\infty$ is a **Følner sequence** if $\mathrm{len}(\mathrm{R}_n) \nearrow \infty$. A set $\Xi \subset X$ is an **R-stack** if it is a disjoint union

$$\Xi = \bigsqcup_{v \in \mathrm{R}} \mathbf{T}^v(B)$$

with $\mu(B) > 0$. The set B is called the **base** of Ξ. Doing double duty, the symbol "Ξ" will also denote the pair (R, B). A (measurable) set $A \subset \Xi$ is "Ξ measurable" if for some set $\mathrm{E} \subset \mathrm{R}$

$$A \stackrel{\text{a.e.}}{=\!=} \bigsqcup_{v \in \mathrm{E}} \mathbf{T}^v(B), \qquad \text{where } B = \mathrm{Base}(\Xi).$$

Say that $A' \subset X$ is "ε-Ξ-measurable" if there exists a Ξ-measurable A with $\mu(A' \bigtriangleup A) \leq \varepsilon$.

A pair $(\vec{\mathrm{R}}, \vec{\Xi})$ is "good" for $\mathbf{T}$ if it satisfies: Ξ_n is an R_n-stack, $\vec{\mathrm{R}}$ is a Følner sequence, and the stacks **locally generate** in the sense that given any ε and $A \in \mathcal{X}$:

For all large n, the intersection $A \cap \Xi_n$
is ε-Ξ_n-measurable.

We can always drop to a subsequence so as to assume the limit $\mu(\vec{\Xi}) \stackrel{\triangleleft}{=} \lim_{n\to\infty} \mu(\Xi_n)$ exists. Note that when "dropping to a subsequence", we renumbered the indices as $n = 1, 2, 3, \ldots$. This renumbering will be done without announcement in the sequel.

The **covering number** of $\mathbf{T}$, a number in $[0,1]$ and written $\kappa(\mathbf{T})$, is the supremum of $\mu(\vec{\Xi})$ taken over all good pairs $(\vec{\mathrm{R}}, \vec{\Xi})$ for $\mathbf{T}$. By splicing sequences one sees that there exists a good pair for which $\kappa(\vec{\Xi}) = \kappa(\mathbf{T})$; such a pair will be called a **covering-pair.** Finally, $\mathbf{T}$ is **rank-1** if $\kappa(\mathbf{T}) = 1$. Any rank-1 action is ergodic.

Remark: By perturbation of Ornstein's random spacer technique of [O] one can make a mixing $\mathbb{Z}^D$-action with any prescribed covering number. ∎

In [K] appears a connection between covering number and a more familiar invariant: For an $\mathbb{Z}$-action T, its spectral multiplicity is dominated by $1/\kappa(T)$.

USING MINIMAL SELF-JOININGS TO LIFT MIXING The strategy for getting M-fold mixing from 4-fold is shown in diagram 2.7.

$$\begin{tikzcd}
\boxed{M\text{-fold mixing}} & \boxed{M\text{-fold MSJ}} \arrow[l, "\text{2-fold mixing}"'] \\
\boxed{\text{4-fold mixing}} \arrow[r, "\text{rank-1}"] & \boxed{\text{4-fold MSJ}} \arrow[u, "\text{Trick}"']
\end{tikzcd}$$

FIGURE 2.7. The upper implication is the observation that any non-trivial M-fold limit-joining is pairwise independent. By M-fold minimal self-joinings, each of its ergodic components is product measure—hence it itself is product measure.

The next several remarks, to be proved later, establish the lower implication.

PROPOSITION B: *Suppose positive integers M and K satisfy*

$$M \cdot \kappa(\mathbf{T}) > K - 1.$$

Then any $\xi \in \mathbb{J}_{\mathrm{Erg}}(\mathbf{T}, \overset{M}{\ldots}, \mathbf{T})$ *has a* K*-fold marginal –call it* $\widehat{\xi}$*– so that*

$$\left(\sqrt[D]{\kappa(\mathbf{T})} - \sqrt[D]{(K-1)/M} \right)^D \bullet \widehat{\xi} \le \lambda$$

for some limit-joining $\lambda \in \mathbb{J}_{\mathrm{Lim}}(\mathbf{T}, \overset{K}{\ldots}, \mathbf{T})$.

MAIN THEOREM 2.8: *Suppose* $M\kappa(\mathbf{T}) > K - 1$. *If*

$$\left(\sqrt[D]{\kappa(\mathbf{T})} - \sqrt[D]{(K-1)/M} \right)^D + \alpha_K(\mathbf{T}) > 1$$

then any ergodic M*-fold self-joining* ξ *has a trivial* K*-fold marginal. In particular, this conclusion holds as soon as* $\mathbf{T}$ *is* K*-fold mixing. So if, in addition to* K*-fold mixing,*

$$\kappa(\mathbf{T}) > 1 - \frac{1}{K}$$

then $\mathbf{T}$ *has* K*-fold minimal self-joinings.*

COROLLARY 2.9: *Suppose* $\mathbf{T}$ *is rank-1, or even just has* $\kappa(\mathbf{T}) > 3/4$. *If* $\mathbf{T}$ *4-fold mixes then* $\mathbf{T}$ *is mixing of all orders.*

Proof of Main Theorem: From Proposition B, then A, we obtain a K-fold marginal $\widehat{\xi}$, limit λ and trivial joining τ such that

$$\beta \bullet \widehat{\xi} \le \lambda$$
$$\alpha_K \bullet \tau \le \lambda$$

where β abbreviates the complicated constant of Proposition B. By hypothesis α_K is positive and so $\mathbf{T}$ is weak-mixing and all trivial self-joinings are ergodic. Hence ergodic joinings $\widehat{\xi}$ and τ are either mutually singular or are equal. But $\beta + \alpha_K$ strictly exceeds 1, so there is no room in the ergodic decomposition of λ for them to be mutually singular. ∎

Establishing Proposition B is the task of the next several lemmas.

PARTITIONS AND GENERIC POINTS Given a partition P on X, let $P(x)$ denote the atom of P containing x. The P-name of a point x is the mapping $v \mapsto x[v]$ where $x[v] \stackrel{\triangle}{=} P(\mathbf{T}^v x)$. Say that an R-stack Ξ is P**-monochromatic** if two conditions hold: Each **level** $\mathbf{T}^v(\mathrm{Base}\,\Xi)$, where $v \in \mathrm{R}$, is entirely contained in some atom $A_v \in P$. And the only points x with the same P-name as the stack, meaning $x[v] = A_v$ for all $v \in \mathrm{R}$, are those x in the base of Ξ.

It is not difficult to check the following (see [K,1.4]), whose purpose is to remove an "ε" from a later argument.

MONOCHROMATIC LEMMA: *If* $\mathbf{T}$ *ergodic then there exists a countable generating partition* P *and covering-pair* $(\vec{\mathrm{R}}, \vec{\Xi})$ *with each stack monochromatic.*

For Følner set R and function f let $\ddot{\mathrm{R}} f$ denote the average $\frac{1}{\eta(\mathrm{R})} \sum_{v \in \mathrm{R}} \mathbf{T}^v f$, where $\mathbf{T}^v f$ is the function $x \mapsto f(\mathbf{T}^v x)$.

MEAN ERGODIC THEOREM: *Suppose* $\mathbf{T}$ *acts ergodically on* X. *Then for any Følner sequence* $\vec{\mathrm{R}}$ *and any function* $f \in L^2(\mu)$,

$$\ddot{\mathrm{R}}_n f \overset{\text{in } L^2}{\longrightarrow} \int_X f \, d\mu, \qquad \text{as } n \to \infty.$$

The mean ergodic theorem is enough to imply that a.e. x is "P-generic" in the following sense: Given a Følner sequence $\vec{\mathrm{R}}$, we can replace it by a sufficiently sparse subsequence so that a.e x is **$\vec{\mathrm{R}}$-generic for** μ: For every P-cylinder-set A, $\ddot{\mathrm{R}}_n \mathbf{1}_A(x) \to \mu(A)$ as $n \to \infty$.

INDEPENDENT SETS The following elementary lemma appears in [K, 3.3]. For a subset $S \subset \mathbb{N}$ let $\overline{\mathrm{Den}}(S)$ and $\underline{\mathrm{Den}}(S)$ denote its upper and lower density; the $\limsup_n$ and $\liminf_n$ of $\frac{1}{n}\left|S \cap [1..n]\right|$. Say that partitions P and Q are "independent upto δ", $P \perp^\delta Q$, if

$$\sum_{A,B} \left|\mu(A \cap B) - \mu(A)\mu(B)\right| \le \delta$$

with the summation taking place over all $A \in P$ and $B \in Q$. In assertions involving "$\perp^\delta$" it will be convenient to regard any subset $\Xi \subset X$ also as the two-set partition $\langle \Xi, \Xi^c \rangle$.

APPROXIMATE STRONG LAW OF LARGE NUMBERS: *In probability space* (X, μ) *suppose* $\{\Xi_n\}_1^\infty$ *is a sequence of sets satisfying the following for any* $\delta > 0$ *and* N_0.

$$\textit{For all large } N: \qquad \Xi_N \perp^\delta \bigvee_{n=1}^{N_0} \Xi_n.$$

Then we can drop to a subsequence of the $\{\Xi_n\}_n$ *and delete a nullset from* X *to obtain: For all* $x \in X$,

$$\underline{\mathrm{Den}}\{n \mid \Xi_n \ni x\} \ge \liminf_{n \to \infty} \mu(\Xi_n).$$

The Mean Ergodic Theorem implies the following.

LEMMA 2.10: $\mathbf{T}$ *an ergodic action and* $\kappa > 0$. *For any positive* δ *and set* A *there exists* L *so that: Whenever* Ξ *is an* R*-stack with* $\operatorname{len}(\mathrm{R}) \geq L$ *and* $\mu(\Xi) \geq \kappa$*, then* $\Xi \perp^{\delta} A$.

For R as in (2.6), let R^+ and R^- represent the disjoint (possibly empty) subrectangles

$$\mathrm{R}^+ \stackrel{\triangleleft}{=} \mathrm{R} \cap [0\,..\,\infty)^{\times D}$$
$$\mathrm{R}^- \stackrel{\triangleleft}{=} \mathrm{R} \cap (-\infty\,..\,0)^{\times D}$$

Let Ξ^+ abbreviate the union of levels $\bigsqcup_{v \in \mathrm{R}^+} \mathbf{T}^v(\operatorname{Base}\Xi)$, with the analogous meaning for Ξ^-.

SETTING UP THE ARGUMENT Fix, henceforth, a P and $(\vec{\mathrm{R}}, \vec{\Xi})$ as in the Monochromatic Lemma. A **word** W is a map from some rectangle E to the alphabet P. Given another word $\widetilde{W}\colon \widetilde{\mathrm{E}} \to P$, measure the coverage of $\widetilde{W}$ by W, written $\kappa(W \text{ on } \widetilde{W})$, by the maximum of

$$\eta\left(\bigsqcup\nolimits_{v \in \mathcal{V}} (\mathrm{E} + v)\right) \Big/ \eta(\widetilde{\mathrm{E}})$$

over all collections $\mathcal{V}$ with: The translated sets $\{\mathrm{E} + v\}_{v \in \mathcal{V}}$ are disjoint and contained in $\widetilde{\mathrm{E}}$, and each satisfies $\widetilde{W}|_{\mathrm{E}+v} = W$.

Proposition B asserts an inequality between two joinings. One way to prove an inequality between two measures is by means of generic points. The $\vec{\mathrm{R}}$ sequence will be used to study ξ. But we also need to examine μ, so fix a Følner sequence $\vec{\mathrm{B}} = \{\mathrm{B}_\ell\}_1^\infty$ for which, without loss of generality,

For all $z \in X$, z *is generic for* μ.

Here, and subsequently, an unadorned "generic" means $\vec{\mathrm{B}}$-generic for μ. Finally, viewing a name z as an infinite rectangle, define

$$\kappa(W \text{ on } z) \stackrel{\triangleleft}{=} \lim_{\ell \to \infty} \kappa(W \text{ on } z|_{\mathrm{B}_\ell})$$

A monochromatic R-stack Ξ has a word W **associated to** it,

$$W(v) \stackrel{\triangleleft}{=} P\big(\mathbf{T}^v(\operatorname{Base}\Xi)\big)$$

for $v \in \mathrm{R}$. By our second condition defining monochromaticity

$$\kappa(W \text{ on } z) = \mu(\Xi).$$

Proof of Proposition B: Henceforth κ will denote $\kappa(\mathbf{T})$. Take $\varepsilon \in (0,1)$ so that,

$$M\varepsilon^D \kappa > K-1. \tag{2.11}$$

Since our arguments will only use the Mean Ergodic theorem and not the pointwise theorem, we can freely replace each R_n by a convenient translation of itself: Arrange that any side $[a_d \mathinner{..} b_d)$ of a rectangle has approximately an ε fraction of its length negative, in $(-\infty, 0)$. So

$$\lim_{n\to\infty} \frac{\eta(\mathrm{R}_n^-)}{\eta(\mathrm{R}_n)} = \varepsilon^D \quad \text{and} \quad \lim_{n\to\infty} \frac{\eta(\mathrm{R}_n^+)}{\eta(\mathrm{R}_n)} = (1-\varepsilon)^D. \tag{2.12}$$

Since $(\mathbf{T}^{\times M} \colon \xi)$ is ergodic and $\vec{\mathrm{R}}^+ \stackrel{\triangleleft}{=} \{\mathrm{R}_n^+\}_{n=1}^\infty$ is itself a Følner sequence, we may drop to a subsequence so that

$$\text{For } \xi\text{-a.e. point } \mathbf{x} \in X^{\times M},\ \mathbf{x} \text{ is } \vec{\mathrm{R}}^+\text{-generic for } \xi$$

on $P^{\times M}$. Of course, this property persists under dropping to a further subsequence.

Since $\{\mathrm{R}_n^-\}_1^\infty$ is a Følner sequence we can, courtesy lemma 2.10, drop to a subsequence satisfying

$$\text{For all } \delta \text{ and } N_0, \text{ for all large } N: \qquad \Xi_N^- \perp^\delta \bigvee_{n=1}^{N_0} \Xi_n^-.$$

This entitles us to the conclusion of the Approximate Strong Law of Large Numbers. Since $\liminf_n \mu(\Xi_n^-)$ dominates $\liminf_n \varepsilon^D \mu(\Xi_n)$, we can delete a μ-nullset from X to obtain the following.

$$\text{For all } x \in X: \qquad \underline{\mathrm{Den}}\{n \mid x \in \Xi_n^-\} \geq \varepsilon^D \kappa.$$

Obtaining the marginal: Fix forevermore a point $\mathbf{x} = (x[1], \ldots, x[M])$ which is $\vec{\mathrm{R}}^+$- generic for ξ. There are "M choose K" distinct K-tuples in $[1 \mathinner{..} M]$. From the above inequality there exists a subset $I \subset [1 \mathinner{..} M]$ of cardinality K fulfilling

$$\overline{\mathrm{Den}}\{n \mid \forall i \in I:\ x[i] \in \Xi_n^-\} \geq \left[M\varepsilon^D\kappa - (K-1)\right] \Big/ \binom{M}{K}. \tag{2.13}$$

By (2.11) the righthand side is positive. This gives us a candidate K-fold marginal; set $\widehat{\xi} \stackrel{\triangleleft}{=} \xi|_I$.

Rename the K points $\{x[i] \mid i \in I\}$ as $\widehat{\mathbf{x}} \stackrel{\triangleleft}{=} (\widehat{x}[1], \ldots, \widehat{x}[K])$. Thus

$$\widehat{\mathbf{x}} \text{ is an } \vec{\mathrm{R}}^{+}\text{-generic point for } \widehat{\xi}.$$

Furthermore, since the set on the lefthand side of (2.13) is infinite, we can subsequence and assume

$$\text{For all } n \text{ and for all } k \in [1 \mathinner{..} K]: \qquad \widehat{x}[k] \in \Xi_n^{-}. \tag{2.14}$$

THE SHIFT VECTOR: For the moment, fix n and let the subscript n become implicit in symbols $\mathrm{R}_n, \mathrm{R}_n^{+}, \Xi_n$ etc. Define the kth **shift** $s[k]$ to be the element of R^{-} such that

$$\widehat{x}[k] \in \mathbf{T}^{s[k]}(\mathrm{Base}\,\Xi)$$

and set $\mathbf{s} \stackrel{\triangleleft}{=} \big(s[1], \ldots, s[K]\big)$. Let z be $\widehat{x}[1]$ and $\mathbf{z}$ be

$$\big(z, \mathbf{T}^{s[2]-s[1]}z, \ldots, \mathbf{T}^{s[K]-s[1]}z\big).$$

Since z is a generic point for μ,

$$\mathbf{z} \text{ is a generic point for } \Delta_K^{\mathbf{s}}.$$

Let W be the word associated to Ξ. Define $\mathbf{W}^{+} \stackrel{\triangleleft}{=} \widehat{\mathbf{x}}|_{\mathrm{R}^{+}}$. Notice from (2.14) that, for each shift, the translation $\mathrm{R} - s[k]$ contains R^{+}. By consequence, $\mathbf{z}|_{\mathrm{R}^{+}} = \mathbf{W}^{+}$ and

$$\kappa\big(\mathbf{W}^{+} \text{ on } \mathbf{z}\big) \geq \frac{\eta(\mathrm{R}^{+})}{\eta(\mathrm{R})}\kappa(W \text{ on } \mathrm{z}) = \frac{\eta(\mathrm{R}^{+})}{\eta(\mathrm{R})}\mu(\Xi) \;=\; \mu(\Xi^{+}).$$

Now fix a $P^{\times K}$-word $\mathbf{A}$. Then

$$\begin{aligned}\Delta_K^{\mathbf{s}}(\mathbf{A}) = \kappa(\mathbf{A} \text{ on } \mathbf{z}) &\geq \kappa(\mathbf{A} \text{ on } \mathbf{W}^{+}) \cdot \kappa(\mathbf{W}^{+} \text{ on } \mathbf{z}) \\ &\geq \kappa(\mathbf{A} \text{ on } \mathbf{W}^{+}) \cdot \mu(\Xi^{+}).\end{aligned}$$

SENDING n TO INFINITY: Rematerializing the subscript n we have

$$\Delta_K^{\mathbf{s}_n}(\mathbf{A}) \geq \kappa(\mathbf{A} \text{ on } \widehat{\mathbf{x}}|_{\mathrm{R}_n^{+}}) \cdot \mu(\Xi_n^{+}).$$

By subsequencing, $\lambda \stackrel{\triangleleft}{=} \lim_n \Delta_K^{\mathbf{s}_n}$ exists. Since $\widehat{\mathbf{x}}$ is $\vec{\mathrm{R}}^{+}$-generic for $\widehat{\xi}$,

$$\kappa(\mathbf{A} \text{ on } \widehat{\mathbf{x}}|_{\mathrm{R}_n^{+}}) \to \widehat{\xi}(\mathbf{A}), \qquad \text{as } n \to \infty,$$

and $\mu(\Xi_n^{+}) \to (1-\varepsilon)^D \cdot \kappa$, by (2.12). Thus $\lambda(\mathbf{A}) \geq \widehat{\xi}(\mathbf{A}) \cdot (1-\varepsilon)^D \kappa$, for each $\mathbf{A}$. The upshot is that

$$\lambda \;\geq\; (1-\varepsilon)^D \kappa \bullet \xi|_I. \tag{*}$$

FINAL STEP: The above reasoning holds for each ε satisfying (2.11); that is, with

$$\varepsilon > \sqrt[D]{(K-1)/M\kappa}.$$

For each such ε there a cardinality K subset $I \subset [1 \mathinner{..} M]$ and limit-joining λ for which $(*)$ holds.

Since there are but finitely many possible subsets I, we can take a sequence of $\varepsilon \searrow \sqrt[D]{(K-1)/M\kappa}$ with a constant value for I. Now, since the collection of limit-joinings is a closed set, (2.1) gives us a limit-joining λ fulfilling $(*)$ with ε actually equal to $\sqrt[D]{(K-1)/M\kappa}$. This establishes Proposition B. ∎

QUESTION "Rectangleness" of our Følner sets was used in (2.12) to control the Haar measure of R^+ and R^- relative to R; no conclusion results if the Følner sequence decays, $\eta(\mathrm{R}_n^+)/\eta(\mathrm{R}_n) \to 0$ or $\eta(\mathrm{R}_n^-)/\eta(\mathrm{R}_n) \to 0$. What geometric properties of a Følner sequence $\vec{\mathrm{R}}$, such as convexity conditions on the R_n, determine whether subset sequences $\vec{\mathrm{R}}^+$ and $\vec{\mathrm{R}}^-$ can be chosen which do not decay? That each R_n tiles the group is not sufficient, even when $\mathbb{G} = \mathbb{Z}$.

One context in which the proof goes through, *mutatis mutandis*, is when $\mathbb{G}$ is a finitely-generated group with polynomial growth and R_n consists of those elements which are products of at most n generators and their inverses.

References

[Fe] S. Ferenczi, *Systèmes de rang un gauche*, Annals Inst. Henri Poincaré **21(2)** (1985), 177–186.

[F,G,K] N. Friedman, P. Gabriel and J.L. King, *An invariant for rank-1 rigid transformations*, Ergodic Theory and Dynamical Systems **8** (1988), 53–72.

[Fur] H. Furstenberg, *Disjointness in ergodic theory, minimal sets, and a problem in diophantine approximation*, Mathematical Systems Theory **1** (1967), 1–49.

[Ho] B. Host, *Mixing of all orders and pairwise independent joinings of systems with singular spectrum*, preprint.

[J,R] A. del Junco and D.J. Rudolph, *On ergodic actions whose self-joinings are graphs*, Ergodic Theory and Dynamical Systems **7** (1987), 531–557.

[Ka] S. Kalikow, *Twofold mixing implies threefold mixing for rank one transformations*, Ergodic Theory and Dynamical Systems **4** (1984), 237–259.

[K] J.L. King, *Joining-rank and the structure of finite rank mixing transformations*, J. Analyse Math. **51** (1988), 182–227.

[K,T] J. L. King and J-P Thouvenot, *A canonical structure theorem for finite joining-rank maps*, J. Analyse Math. **56** (1991), 211–230.

[M] S. Mozes, *Mixing of all orders of Lie group actions*, preprint.

[O] D.S. Ornstein, *On the root problem in ergodic theory*, Proc. Sixth Berkeley Symp. Math. Stat. Prob. **II**, Univ. of California Press, 1967, pp. 347–356.

[R] D.J. Rudolph, *An example of a measure preserving map with minimal self-joinings, and applications*, J. Analyse Math. **35** (1979), 98–122.

[V] W.A. Veech, *A criterion for a process to be prime*, Monatshefte Math. **94** (1982), 373–409.