A DENSITY VERSION OF THE HALES–JEWETT THEOREM

By

H. FURSTENBERG AND Y. KATZNELSON†

I. Introduction

I.1. The theorem of van der Waerden on arithmetic progressions states that for given natural numbers r, k there is a constant $K(r, k)$ so that for any partition of an arithmetic progression of length $K \geq K(r, k)$ to r subsets, one of these contains an arithmetic progression of length k. This result is the prototype of a Ramsey theorem whereby a certain kind of structure is reproduced in small scale when a large scale model is partitioned arbitrarily to a fixed number of subsets. Van der Waerden's theorem is a special case of a general combinatorial theorem proved by Hales and Jewett. To formulate their result we make some definitions. Let A denote a finite set $\{a_1, a_2, \ldots, a_k\}$, and let $W_N(A)$ denote the words of length N with letters in A, $W_N(A) = A^N$. We think of the points of $W_N(A)$ as vectors in a "combinatorial" N-dimensional space. If A is a finite field, then $W_N(A)$ is indeed a vector space over A. This example motivates the following definition of a "line" in $W_N(A)$: if $k = \#(A)$ then k points $\{w^1, w^2, \ldots, w^k\}$ in $W_N(A)$ constitute a *combinatorial line* if there is a partition $\{1,2,\ldots,N\} = I \cup J$, $I \cap J = \varnothing$, $J \neq \varnothing$ and writing $w^h = (w_1^h, w_2^h, \ldots, w_N^h)$ we have $w_n^1 = w_n^2 = \cdots = w_n^k$ for $n \in I$, and $w_n^h = a_h$ for $n \in J$. We can also describe $\{w^1, w^2, \ldots, w^k\}$ as follows. Let x denote a variable and form words with the alphabet $A \cup x$. Suppose $w(x)$ is such a word in which the letter x occurs. Then $w(1)$, $w(2), \ldots, w(k)$ form a combinatorial line. Note that if A is a finite field then a combinatorial line is a line in the geometric sense in the vector space A^N. Moreover if $A = \{0,1,\ldots,k-1\}$ and we interpret $W_N(A)$ as integers $< k^N$, then a combinatorial line forms an arithmetic progression.

The Hales–Jewett theorem can now be formulated as follows:

Theorem A. *There is a function $M(r, k)$ defined for $r, k \in \mathbf{N}$, so that for $\#(A) = k$ and $N \geq M(r, k)$, if $W_N(A) = C_1 \cup C_2 \cup \cdots \cup C_r$ is any partition of $W_N(A)$ into r subsets, one of these subsets contains a combinatorial line.*

†Research supported by the National Science Foundation under Grant No. DMS86-05098.

Van der Waerden's theorem follows from Theorem A by setting $K(r,k) = k^{M(r,k)}$. We take $\{0,1,\dots,K-1\}$ as the typical arithmetic progression of length K. Expressing numbers to the base k we can identify $\{0,1,\dots,K(r,k)-1\}$ with $W_{M(r,k)}(\{0,1,\dots,k-1\})$.

Interpreting A as a finite field we also get the following theorem:

Theorem B. *There is a function $N(r,q)$ defined for $r \in \mathbf{N}$ and q a prime power, so that if F is a field with q elements and V is a vector space over F of dimension $\geq N(r,q)$ and if $V = C_1 \cup C_2 \cup \cdots \cup C_r$ is any partition of V into r sets, then one of these sets contains an affine line.*

Theorems A and B have multi-dimensional analogues. Theorem A also gives at once the multi-dimensional analogue of van der Waerden's theorem (proved by Grünbaum).

Now both van der Waerden's theorem and Theorem B have density versions which are considerably more powerful theorems. Namely, Szemerédi proved [SZ]:

Theorem C. *There is a function $P(\epsilon,k)$ defined for $\epsilon > 0$ and $k \in \mathbf{N}$, so that if $N \geq P(\epsilon,k)$ and $S \subset \{1,2,\dots,N\}$ with $\#(S) \geq \epsilon N$ then S contains a k-term arithmetic progression.*

Moreover we proved [FK2]:

Theorem D. *There is a function $Q(\epsilon,q)$ defined for $\epsilon > 0$ and q a prime power, so that if F is the field with q elements and V is a vector space over F of dimension $N \geq Q(\epsilon,q)$, and if $S \subset V$ is a subset with $\#(S) \geq \epsilon q^N$, then S contains an affine line.*

In view of this it is natural to ask whether the "master" coloring theorem, Theorem A, has a density theoretic analogue. The purpose of this paper is to answer this affirmatively with the following result:

Theorem E. *There is a function $R(\epsilon,k)$ defined for $\epsilon > 0$ and $k \in \mathbf{N}$, so that if A is a set with k elements, $W_N(A)$ consists of words in A of length N, and if $N \geq R(\epsilon,k)$, then any subset $S \subset W_N(A)$ with $\#(S) \geq \epsilon k^N$ contains a combinatorial line.*

Theorem E implies all the foregoing results. For example, Theorem A follows from it by setting

$$M(r,k) = R\left(\frac{1}{r+1}, k\right).$$

For Theorem C we would take $P(\epsilon,k) = k^{R(\epsilon,k)}$ and use again the fact that a combinatorial line becomes an arithmetic progression when integers are represented as words. Theorem D also follows by choosing $A = F$ and observing that a combinatorial line is, in this case, a special case of an affine line.

One case of Theorem E has been known for a long time. If A has two elements ($k = 2$) we can identify words in $W_N(A)$ with subsets of $\{1,2,\ldots,N\}$. It is not hard to see that for this case a combinatorial line which has two elements in it corresponds to a pair of sets S_1, S_2 in $\{1,2,\ldots,N\}$ with $S_1 \subset S_2$. Now Sperner's lemma [S] states that a maximal family of subsets of $\{1,2,\ldots,N\}$ with no one containing another has $\leq \binom{N}{[N/2]}$ numbers. Since this is $o(2^N)$, if $\epsilon > 0$, for large N we will have

$$\binom{N}{[N/2]} < \epsilon 2^N$$

and it follows that Theorem E is, in this case, a consequence of Sperner's lemma.

I.2. The proof of Theorem E that we will give makes considerable use of ergodic theoretic techniques. The relevance of ergodic theory to this branch of combinatorics was demonstrated in [F1], [F2], [FKO], and [FK1] where ergodic theoretic methods are used to obtain an alternative proof of Theorem C as well as a proof of the multi-dimensional version of Szemerédi's theorem. The notions from ergodic theory that enter in these papers are conventional notions: measure-preserving systems, ergodicity, weak mixing, and factors. In [FK2] where a more powerful result is obtained which implies the earlier ones as well as Theorem D, new ergodic theoretic ideas are introduced. The present paper represents a further evolution of these ideas. The overall structure remains the same however; just the setting changes and becomes more general at each stage.

In each of the Theorems C, D, and E, the notion of a subset of a given "space" of density bounded away from zero plays a key role and this suggests introducing measure spaces into the picture. In fact it is quite easy to show in each case that the combinatorial theorem in question implies a "recurrence" result for certain families of measure-preserving transformations. Our technique consists of reversing this implication in each case showing that the combinatorial theorem is a consequence of a recurrence theorem for the appropriate family of measure preserving transformations. We then proceed to prove the ergodic theoretic assertion.

For Theorem C the setting for the ergodic theoretic result is one of classical ergodic theory, and the result in question is a generalization of classical Poincaré recurrence to multiple recurrence:

Theorem C′. *If T is a measure-preserving transformation on a measure space $(X,\mathcal{B},\mu)$ and $A \in \mathcal{B}$ with $\mu(A) > 0$, then for any $k \in \mathbf{N}$ there is an $n > 0$ and a set A' of positive measure, satisfying*

$$A' \subset A,\ T^n A' \subset A, \ldots, T^{(k-1)n} A' \subset A. \tag{I.1}$$

One shows that $C' \Rightarrow C$ and then one proves C'. The advantage of the more sophisticated ergodic theoretic formulation is that the ergodic theoretic setup places at our disposal a richer structure that can be exploited. One first notices that if T is sufficiently mixing (weak mixing is sufficient) then one has

$$\text{(I.2)} \qquad \lim \frac{1}{N} \sum \mu(A \cap T^{-n}A \cap \cdots \cap T^{-(k-1)n}A) = \mu(A)^k$$

so that (I.1) is clear. Now the obstruction to weak mixing is the existence of eigenfunctions for the operator T, and these correspond to a certain component (or "factor") of the system being almost periodic (or representing group rotations). It is easy to see that the existence of eigenfunctions introduces a constraint on triples $\{x, T^n x, T^{2n}x\}$ on account of which (I.2) need not hold. Now the totality of constraints on k-tuples $\{x, T^n x, \ldots, T^{(k-1)n}x\}$ can be accounted for by a particular "factor", a particular space Z_k onto which X can be mapped in such a way that inverse images of the same point of Z_k map to inverse images of the same point of Z_k. This induces an action of T on Z_k. In this more general situation (I.2) is replaced by the result that if the lim inf of the expression in (I.2) is positive for the action of T on Z_k and any nontrivial $A \subset Z_k$, then the same is true for T acting on X. We are left with dealing with

$$\text{(I.3)} \qquad \liminf \frac{1}{N} \sum \mu(A \cap T^{-n}A \cap \cdots \cap T^{-(k-1)n}A)$$

for the special system (Z_k, T).

What characterizes the latter systems is that they admit a "composition series" of successive factors:

$$\text{(I.4)} \qquad Z_k \to Z_{k-1} \to \cdots \to Z_3 \to Z_2 = \text{one point space},$$

where each system is a rotation-like" extension of the next system. For example, Z_3 is a compact abelian group with T acting by group multiplication, and Z_4 could be a sphere bundle over Z_1 with T taking spheres to spheres by rotation, etc. The rigid nature of these extensions now makes it possible to lift the multiple recurrence property—namely, the positivity of (I.3)—step by step starting with the trivial system up to Z_k. This is a sketch of the proof of Theorem C'. In broad outline this scheme was also used in [FK2]. The difference was that instead of dealing with a classical system involving a group of measure preserving transformations, we needed to consider what we call "IP" systems of transformations, a notion that will be prominent in our present work as well. An IP-system consists of a sequence $\{T_1, T_2, \ldots T_n, \ldots\}$ as well as all products $T_{i_1}T_{i_2} \cdots T_{i_r}$ for monotone sequences of indices. (In [FK2] the operators commuted so the order did not matter.) Theorem C' is replaced by an analogous multiple recurrence result with $I, T^n, T^{2n}, \ldots, T^{(k-1)n}$ replaced by operators from k commuting IP-systems corresponding to the same multi-index $(i_1, i_2, \ldots, i_r)$. IP-systems have a partially multiplicative structure and one can talk about notions analogous to mixing and the obstruction to mixing, namely, rigidity. A novel aspect of this theory is that ergodic averages are replaced by "IP-limits" which do not involve averaging, but nonetheless produce projection operators. In the case of mixing systems one obtains an analogue of (I.2) with the ergodic average replaced by the appropriate IP-limit.

Once again the obstruction to mixing is the existence of a composition series as in (I.4) where "rotation-like" extensions are replaced by an appropriate notion of "rigid" extension. Here too the crux of the argument consists in showing that a certain multiple recurrence property lifts from a system to a rigid extension. We might note parenthetically that in this portion of the argument we found it necessary to have recourse to the Hales–Jewett theorem.

In our present treatment this type of combinatorial analysis will again enter at the corresponding point.

I.3. The foregoing scheme will be followed in the present treatment as well. After a preliminary section, §2 will be devoted to showing the equivalence of Theorem E with a recurrence statement for a system—to be called a $W(k)$-system—of measure-preserving transformations, and it is this statement that we proceed to prove. Matters will be complicated by the fact that the transformations in question need no longer commute; indeed the commutative case reduces exactly to the situation of [FK2]. Moreover, in the present context, the analogue of the ergodic average will be not a single limit, but a continuum of limits parametrized by a space $\Omega(k)$ of infinite sequences. In producing this and related definitions extensive use will be made of some infinite-dimensional generalizations of the Hales–Jewett theorem. A result of this type was proved by T. Carlson (extending an earlier joint result with S. G. Simpson) [C], [CS], but we have given an independent presentation of the necessary combinatorial background in [FK4]. These are summarized in §1.

From the point of view of ergodic averages we could pick once and for all a single point of $\Omega(k)$ and evaluate all the limits there. However a novel notion of factor intervenes and this compels us to deal with the continuum $\Omega(k)$. We are unable to define for a given $W(k)$-system on X a factor Z on which the given family induces transformations. This can only be done in some asymptotic sense. What one can do is to define a "field" of factors, or, more correctly, *factorizations*, $Z(\omega)$ for $\omega \in \Omega(k)$, so that the transformations in various IP-systems that can be formed from the given system act on the factorization corresponding to ω moving it to the factorization corresponding to some ω'. As we move out to infinity in an appropriate sense, ω' will be closer and closer to ω. We can also achieve a certain continuity in ω so that this will give $Z(\omega)$ asymptotic properties of a factor. The elaboration of these ideas is the purpose of §3–§6. Having provided all the background material, we will be in a position in §7 to give a more detailed overview of the proof of the main result. Briefly, the proof consists of three steps. In §8 we show how the general case can in effect be reduced to one for which one has a "relatively rigid composition series". §9 is devoted to a technical result showing how a relatively rigid IP-system of transformations can be assumed, in some restricted fashion, to act as the identity map on fibers of the extension. The effect of this will be to reduce the analysis of what happens on fibers relative to the factor to a situation in which the k families occurring in the definition of a $W(k)$-system are replaced by $k - 1$. Finally, in §10, this is used to show that if Theorem E is known for alphabets of $k - 1$ letters, then the recurrence property that we are aiming for will lift from an Ω-factorization to a rigid extension.

1. Terminology, notation and preliminaries

1.1. Denote by $W(k)$ the free semigroup generated by the k symbols $\{1,2,\dots,k\}$, i.e. the set of all finite words $w = w(1)w(2)\dots w(n)$ with each $w(j) \in \{1,2,\dots,k\}$ (and the operation is concatenation). We denote by $|w|$ the length of the word w, and by $W_n(k)$ the subset of $W(k)$ consisting of all the words of length n.

$W(k,l)$ denotes the free semigroup generated by (the set of finite words on) the $k + l$ symbols $\{1,2,\dots,k,t_1,\dots,t_l\}$. The reason that we write $W(k,l)$ and not $W(k+l)$, even though the two are algebraically the same, is the different role that we have in mind for the last l digits.

Denote by $\Xi(l)$ the space of all finite words on the l digits in $\{t_1,\dots,t_l\}$, with no runs longer than one (i.e. no consecutive multiple appearance of any digit), and by $\Xi_N(l)$ the set of words in $\Xi(l)$ whose length is less than or equal to N.

Define $\pi_l\colon W(k,l) \mapsto \Xi(l)$ to be the collapsing map which deletes all the digits not in $\{t_1,\dots,t_l\}$, shortens the runs to singletons, and removes all the empty spaces. For example, for $l = 1$, $\Xi(1)$ is the set containing the empty word and the singleton t_1, and π_1 just distinguishes between words that contain the digit t_1 and those that do not. We extend the notation to sequences of the form $j_1,\dots,j_m$ with $1 \le j \le k + l$ and write $\pi_l(j_1,\dots,j_m)$ as if the sequence were a word in $W(k,l)$. The notation

$$W^*(k,l) = W(k,l) \cap \pi_l^{-1}(t_1 t_2 \cdots t_l)$$

for the space consisting of the words in which all the letters t_j occur, with all the occurrences of t_j preceding those of t_{j+1} for every $j < l$, will be particularly useful.

1.2. Definition. A *parametric line* $w(x)$ is a word in $W(k,1)$, with at least one occurrence of the digit t_1, in which the digit t_1 is replaced by a variable x. A parametric line defines (is) a mapping from $\{1,\dots,k\}$, resp. $\{1,\dots,k,t_1,\dots,t_l\}$, into $W(k)$, resp. $W(k,l)$.

A *combinatorial line* in $W(k)$, resp. $W(k, l)$, is the range of a parametric line, i.e., the set of words obtained by substituting for the variable x all the digits $1, \ldots, k$, resp. $1, \ldots, k, t_1, \ldots, t_l$.

A *parametric d-dimensional space* is a concatenation of d parametric lines, which we write $\Sigma(x_1, \ldots, x_d) = w_1(x_1) \cdots w_d(x_d)$. The range of Σ, obtained by substitution of all the allowed digits (of $W(k)$ resp. $W(k, l)$) independently for the variables $\{x_j\}$, is referred to as a *combinatorial d-dimensional space* in $W(k)$, resp. $W(k, l)$. A d-dimensional space in $W(k)$ is obtained in a natural way from any $w \in W^*(k, d) = W(k, d) \cap \pi_d^{-1}(t_1 t_2 \cdots t_d)$ by considering the letters t_j as variables to be substituted independently by the digits $\{1, \ldots, k\}$. We refer to the corresponding combinatorial space as the one *determined* by w.

A *combinatorial subspace* of $W(k)$, resp. $W(k, l)$, is determined similarly by an infinite sequence $\{w_j(\cdot)\}$ as the range of the map $\Sigma = \Sigma(\{w_j\})$ defined by

$$\Sigma : j_1 \cdots j_m \mapsto w_1(j_1) \cdots w_m(j_m) \tag{1.1}$$

from $W(k)$, resp. $W(k, l)$, into itself. We shall denote such subspaces by $\Sigma W(k)$, $\Sigma W(k, l)$ etc.

1.3. If $\Sigma_1 W(k)$ and $\Sigma_2 W(k)$ are combinatorial subspaces and $\Sigma_1 W(k) \subset \Sigma_2 W(k)$, then $\Sigma_1 W(k)$ is a combinatorial subspace of $\Sigma_2 W(k)$, that is the image under Σ_2 of some combinatorial subspace $\Sigma_3 W(k)$. This means that there is a Σ_3 such that $\Sigma_1 = \Sigma_2 \Sigma_3$. Similarly for subspaces of $W(k, l)$.

1.4. We introduce a metric on $W(k)$ by

$$\rho(w_1, w_2) = \inf\{2^{-q} : w_1(j) = w_2(j) \text{ for } 1 \leq j \leq q\}. \tag{1.2}$$

Thus, two distinct words are close if they are both long and they match for a long time. The completion of $W(k)$ in this metric is (can be identified with) $W(k) \cup \Omega(k)$ where $\Omega(k)$ is the space of all infinite words on $\{1, 2, \ldots, k\}$.

For $W(k, l)$ we do not simply take the metric induced by that of $W(k + l)$; instead we take

$$\rho'(w_1, w_2) = \inf\{2^{-q} : w_1(j) = w_2(j) \in \{1, 2, \ldots, k\} \text{ for } 1 \leq j \leq q\}. \tag{1.3}$$

The completion of $W(k, l)$ relative to this metric is $W(k, l) \cup \Omega(k)$, its infinite words using only the digits $\{1, 2, \ldots, k\}$.

The completions of the combinatorial subspaces $\Sigma W(k)$ and $\Sigma W(k, l)$ are obtained by adding to them the "points at infinity", namely the infinite words of the form $w_1(j_1) \cdots w_m(j_m) \cdots$. It is the image $\Sigma\Omega(k)$ of $\Omega(k)$ under the obvious extension of Σ.

1.5. The following theorem (cf. [FK4]) will be essential in a number of crucial steps in the sequel.

Theorem. *Let C be a compact metric space, and $f\colon W(k,l) \mapsto C$ an arbitrary function. Then there exists a combinatorial subspace $\Sigma W(k,l) \subset W(k,l)$ such that the restriction of f to $\Sigma W(k,l) \cap \pi_l^{-1}\xi$ is uniformly continuous for every $\xi \in \Xi(l)$.*

For every $\xi \in \Xi(l)$ we thus obtain an extension by continuity f_ξ of f to $\Sigma\Omega(k)$. In our applications of the above we shall usually take the special $\xi^* = t_1 t_2 \cdots t_l$.

1.6. We shall have to deal with (countably) many functions, or mappings, of $W(k,l)$ into various compact spaces. By taking Cartesian products we can consider the given mappings as coordinates of a single map into a compact metric space, and Theorem 1.5 above gives a combinatorial subspace on which all the given functions are uniformly continuous. Another way to proceed is to go to smaller and smaller subspaces as we add more functions which we want continuous (taking care to have a non-trivial intersection) which has the advantage of not having to specify in advance all the mappings.

A convenient observation is that there exist natural maps of $W(k,2)$ onto the union $\bigcup_l W^*(k,l)$, for example the map sending the digits $\{1,.., k\}$ onto themselves, and the alternating blocks of the first and the second t onto "new" t, so that if the length of $\pi_2(w)$ is l, w is mapped onto a word in $W^*(k,l)$. This permits the lifting to $W(k,2)$ of maps defined on $\bigcup_l W^*(k,l)$. Applying Theorem 1.5 to the lifted map gives one Σ which is good for all values of l (see 2.3 for a concrete application).

1.7. The following notation will be useful. If w is a word in $W(k)$ of length $|w|$, and $\alpha \subset \{1,2,\ldots,|w|\}$ we denote by w_α the word in $W(k,1)$ of the same length, with

$$w_\alpha(n) = \begin{cases} w(n), & n \notin \alpha, \\ t_1, & n \in \alpha, \end{cases}$$

that is the word obtained by replacing the original digits appearing in w in places contained in α by the additional letter t_1.

Using this notation we shall write parametric lines also as w_α^x or simply as w_α.

More generally, we define $w_{\alpha_1,\alpha_2,\ldots,\alpha_l} \in W(k,l)$ for disjoint sets α_i contained in $\{1,\ldots,|w|\}$, by

$$w_{\alpha_1,\alpha_2,\ldots,\alpha_l}(n) = \begin{cases} w(n), & n \notin \cup\alpha_j, \\ t_j, & n \in \alpha_j. \end{cases} \tag{1.4}$$

On the other hand, suppose $v \in W(k,l)$; we denote by $v^{i_1,\ldots,i_l}$, the word in $W(k)$ obtained by substituting the digits i_j for t_j in v. In particular, for $v = w_{\alpha_1,\alpha_2,\ldots,\alpha_l}$ we have

$$w^{i_1,i_2,\ldots,i_l}_{\alpha_1,\alpha_2,\ldots,\alpha_l}(n) = \begin{cases} w(n), & n \notin \cup\alpha_j, \\ i_j, & n \in \alpha_j. \end{cases} \tag{1.5}$$

Finally if $u_1, u_2, \ldots u_t$ are words such that the length of u_j is equal to $|\alpha_j|$ (we shall sometimes use the expression "u_j is defined over α_j"), then we denote by

$$w^{u_1,u_2,\ldots,u_t}_{\alpha_1,\alpha_2,\ldots,\alpha_t} \tag{1.6}$$

the word which coincides with w for $n \notin \cup\alpha_j$, and if n is the rth element of α_j then it takes on the value $u_j(r)$.

Definitions (1.5) and (1.6) can be made for $\omega \in \Omega(k)$ as for $w \in W(k)$. For (1.4) we would have to introduce the notation $\Omega(k,l)$ which will signify the space of infinite words on the digits $1, \ldots, k, t_1, \ldots, t_l$ with the stipulation, however, that the occurrence of the t's is finite. We use the notation ω_α only within the context of *admissible functions*, i.e.:

Definition. A function $f(w; \alpha_1, \ldots, \alpha_l)$, defined for $\alpha_1 \cup \alpha_2 \cup \cdots \cup \alpha_l \subset [1, |w|]$, is *admissible* if it does not depend on the value of $w(n)$ whenever $n \in \alpha_1 \cup \cdots \cup \alpha_l$ and if in addition it does not depend on $w(n)$ for $n > \max\{m : m \in \cup\alpha_j\}$.

The first condition justifies writing admissible functions as $f(w_{\alpha_1,\alpha_2,\ldots,\alpha_l})$ so that it can be viewed as defined over $W(k,l)$ and Theorem 1.5 can be applied to it. The second condition extends the domain of definition to Ω and there is no ambiguity in defining $f(\omega_{\alpha_1,\alpha_2,\ldots,\alpha_l})$ as $f(w_{\alpha_1,\alpha_2,\ldots,\alpha_l})$ where w is a sufficiently long beginning of ω.

It will be useful to write $w \to \omega$ for a sequence of words of length $\to \infty$ denoted generically by w converging to an infinite word ω, with the meaning that for every w far along enough in the sequence, a long initial segment coincides with that of ω. With this notation we can define, for an admissible function f,

$$f(\omega_{\alpha,\alpha_2,\ldots,\alpha_l}) = \lim_{w \to \omega} f(w_{\alpha_1,\alpha_2,\ldots,\alpha_l}).$$

1.8. We end this section with

Theorem. *If S is a residual subset of $\Omega(k)$ then there exists a combinatorial subspace $\Sigma\Omega(k)$ such that $S \supset \Sigma\Omega(k)$.*

Proof. The reader should develop the following outline. S contains a dense G_δ, i.e. there exists a sequence $\{O_n\}$ of dense open sets, $O_n \subset O_{n-1}$, such that $S \supset \cap O_n$.

There exists a combinatorial line $\{w_1(j)\}_{j=1}^k$ such that all the words in $\Omega(k)$ which start with any $w_1(j)$, $j = 1, \ldots, k$, are in O_1.

There exists a line $\{w_2(j)\}_{j=1}^k$ such that all the words in $\Omega(k)$ which start with $w_1(j_1)w_2(j_2)$, any j_1, j_2, are in O_2; etc. The subspace $\Sigma\Omega(k)$, with $\Sigma = \{w_n(\cdot)\}$, is clearly contained in S. QED

Corollary. *Let $\mathcal{H}$ be a Hilbert space and Φ a weakly continuous $\mathcal{H}$-valued function on $\Omega(k)$. Then Φ is continuous in norm on a combinatorial subspace $\Sigma\Omega(k)$.*

Proof. Weak continuity of Φ implies semi-continuity of $\|\Phi\|$, hence continuity on a residual set, hence on a subspace, and for a $\mathcal{H}$-valued function continuity in norm is equivalent to the conjunction of weak continuity and continuity of the norm. QED

2. The various equivalent forms

Our goal is to prove the density version of the Hales–Jewett theorem, which is statement (a) in Proposition 2.1 below. We show that the theorem has several equivalent forms, including the "measure preserving system" version, stated in Proposition 2.7 below, which we shall use.

2.1. Proposition. *The following statements are equivalent*:

(a) *For every $\epsilon > 0$ there exists $n(\epsilon) = n(\epsilon, k)$ such that if $n > n(\epsilon)$, and $A \subset W_n(k)$, $|A| > \epsilon k^n$, then A contains a combinatorial line.*

(a*) *If $A \subset W(k)$ and* $\limsup_n k^{-n}|A \cap W_n(k)| > 0$, *then A contains a combinatorial line.*

(b) *For every $\epsilon > 0$ there exists $n(\epsilon) = n(\epsilon, k)$ such that if $n > n(\epsilon)$, and for every $w \in W_n(k)$ there is given a measurable set B_w in some fixed probability measure space $\{X, \mathcal{B}, \mu\}$, and $\mu(B_w) > \epsilon$, then there exists a combinatorial line $w(x)$ in $W_n(k)$ such that $\mu(\bigcap_{j=1}^k B_{w(j)}) > 0$.*

(b*) *If for every $w \in W(k)$ there is given a measurable set B_w in some fixed probability measure space $\{X, \mathcal{B}, \mu\}$, and $\mu(B_w) > \epsilon > 0$, then there exists a line $w(x)$ in $W(k)$ such that $\mu(\bigcap_{j=1}^k B_{w(j)}) > 0$.*

Proof. The implications (a) $\Leftrightarrow$ (a*) and (b) $\Leftrightarrow$ (b*) are clear, and what we propose to prove here is that (b) $\Leftrightarrow$ (a).

Assuming that (a) is valid, let $n(\epsilon)$ be as in (a) and assume that $n > n(\epsilon)$, and for every $w \in W_n(k)$ there is given a measurable set B_w in some fixed probability measure space $\{X, \mathcal{B}, \mu\}$, and $\mu(B_w) > \epsilon$. For every $x \in X$ denote $A(x) = \{w : w \in W_n(k), x \in B_w\}$. For a set of positive measure of x, $|A(x)| > \epsilon k^n$; by (a) there exists a line $l(x) \subset A(x)$, and at least one of these (finitely many) lines will be shared by a set of positive measure of x's. This proves (a) $\Rightarrow$ (b).

For the proof of (b) $\Rightarrow$ (a) observe that the conclusion of (a) is clearly valid for $\epsilon > 1 - 1/k$. In that range one can take $n(\epsilon) = 1$ and notice that, given $A \subset W_n(k)$, one can split A to the sets $A_j = \{w \in A : w(1) = j\}$, and the sets $A'_j \subset W_{n-1}$, which are the projections of A_j on the last $n - 1$ digits, have a nontrivial intersection if the measure (density) of A exceeds $1 - 1/k$.

Denote $\epsilon_0 = \inf \epsilon'$ the infimum over all ϵ' for which the conclusion of (a) is valid. We use (b) to prove that $\epsilon_0 = 0$. Otherwise, if $\epsilon_0 > 0$, take $m > n(\epsilon_0/2)$, $n(\epsilon)$ the function given by (b). Take $\epsilon_1 = \epsilon_0(1 - k^{-m-2})$ so that

$$\epsilon_2 = \epsilon_1 + \frac{\epsilon_0}{2} k^{-m} > \epsilon_0.$$

Let M be large enough so that the conclusion of (a) is valid for $n > M$ and sets of measure $> \epsilon_2$. We claim that the conclusion is still valid for sets of measure $> \epsilon_1$ and $n > m + M$ which contradicts the definition of ϵ_0. This contradiction will show that $\epsilon_0 = 0$ and prove (a).

To prove the claim, let $A \subset W_n(k)$ have measure $k^{-n}|A| > \epsilon_1$, and define for every $w \in W_m(k)$ the set $A'_w = \{u \in W_{n-m} : wu \in A\}$. If the measure of every A'_w is at least $\epsilon_0/2$ we can invoke (b) and obtain a combinatorial line $\{w_j\} \subset W_m(k)$ such that the corresponding A'_{w_j} have a nontrivial intersection. If u is in that intersection then $\{w_j u\}$ is a combinatorial line in A. On the other hand, if the measure of one A'_w is less than $\epsilon_0/2$, then, since the average of the measures of A'_w for $w \in W_m(k)$ is the measure of A, some $A'_{w'}$ has measure exceeding ϵ_2 and we have our HJ-sequence in it. QED

We refer to (any of) the statements of Proposition 2.1 as DHJ_k.

2.2 Before stating the (equivalent) form of DHJ_k which we shall actually prove, we need the following

Definition.

(a) The *joint distribution* of the data $\{B_w : w \in W(k)\}$ is the function m defined for all subsets $I \subset W_n(k)$, n arbitrary, by

$$m(I) = \mu\left(\bigcap_{w \in I} B_w\right).$$

(b) The process† (indexed by $W(k)$) $\{B_w\}$, or equivalently, its joint distribution m, is *stationary* if whenever I is a subset of $W_n(k)$ and $v \in W_l(k)$ then, writing $Iv = \{wu : w \in I\} \subset W_{n+l}(k)$, we have $m(I) = m(Iv)$.

†We list here events rather than random variables, which is the customary way of describing processes. The indicator functions of B_w form the "customary" process.

2.3. We now introduce a stronger version of stationarity. Recall that a subspace of $W(k)$ is (by definition) the range of a map $\Sigma : W(k) \mapsto W(k)$ (cf. (1.1)), thus the restriction of the process $\{B_w\}$ to $\Sigma(W(k))$ can be viewed as a process on $W(k)$, namely of $\{\Sigma B_w\} = \{B_{\Sigma(w)}\}$, and we denote its joint distribution by $\Sigma(m)$.

Definition. The process $\{B_w\}$ (or its distribution m) is *strongly stationary* if its restriction to any subspace has the same joint distribution, in other words if $\Sigma(m) = m$ for all Σ.

The joint distribution of a process is a function with values in [0,1] defined on $\mathfrak{F}\mathfrak{S} = \bigcup_n 2^{W_n(k)}$, i.e. an element of the compact metric space $\mathfrak{X} = [0,1]^{\mathfrak{F}\mathfrak{S}}$. Not all points in $\mathfrak{X}$ are obtained as joint distributions, but a moment's thought shows that the joint distributions form a closed subset of $\mathfrak{X}$ (one needs to check for $W_n(k)$, separately for each n).

If m is the joint distribution of a process, we denote by $\mathcal{L}(m)$ the closure of the set $\{\Sigma(m)\}$ of the restrictions of m to all the combinatorial subspaces of $W(k)$.

Proposition. *For any process $\{B_w\}$ with joint distribution m, $\mathcal{L}(m)$ contains strongly stationary distributions.*

Proof. For $w \in W^*(k,l)$ denote $\tilde{m}(w) =$ the joint distribution function on the l-dimensional combinatorial space determined by w (the value of $\tilde{m}(w)$ is in $[0,1]^{2^{W_l(k)}}$). Invoking Theorem 1.5 we obtain a subspace $\Sigma_l W(k,l)$ such that $\tilde{m}(w)$ is uniformly continuous on $\Sigma_l W(k,l) \cap \pi_l^{-1}(t_1 \cdots t_l) = \Sigma_l W^*(k,l)$. Repeating the same for all values of l and successively refining the subspaces obtained at each stage, or using §1.6, we obtain a subspace $\Sigma_0 W(k)$ such that for all l, $\tilde{m}(w)$ is uniformly continuous on $\Sigma_0 W(k,l) \cap \pi_l^{-1}(t_1 \cdots t_l)$. For any $\omega \in \Sigma_0(\Omega(k))$ the limit distribution (which for any dimension l is $\lim \tilde{m}(w)$, the limit for $w \to \omega$ in $\Sigma_0 W(k,l) \cap \pi_l^{-1}(t_1 \cdots t_l)$) is clearly strongly stationary. QED

2.4. In the statement of Proposition 2.1, part (b*), we can add now the additional condition of strong stationarity:

Proposition. *The following statement is equivalent to* DHJ$_k$,

(b**) *For any strongly stationary process $\{\hat{B}_w\}$ in some fixed probability measure space $\{X, \mathcal{B}, \mu\}$, $\mu(\hat{B}_w) > 0$, there exists a line $w(x)$ in $W(k)$ such that*

$$\mu\left(\bigcap_{j=1}^{k} \hat{B}_{w(j)}\right) > 0.$$

Remark. The statement above parallels that of Proposition 2.1. For a strongly stationary process $\{\hat{B}_w\}$ the measures of the sets $\hat{B}_w$ are all equal as are the measures of their intersections along any (combinatorial) line; thus the claim of existence above is in fact equivalent to the statement that the intersection along any line has positive measure.

Proof. Since (b*) clearly implies (b**), we only need to prove the converse. Given an arbitrary process $\{B_w\}$, denote its distribution by m, and applying Proposition 2.3 we obtain a strongly stationary process $\{\hat{B}_w\}$ whose distribution $\hat{m}$ is contained in $\mathcal{L}(m)$. The assumption that $\mu(B_w) > \epsilon > 0$ guarantees that $\mu(\hat{B}_w) > 0$ and, by (b**), we have $\mu(\bigcap_{j=1}^k \hat{B}_{w(j)}) > 0$ for some (any) line $w(x)$ in $W(k)$. As $\hat{m} \in \mathcal{L}(m)$, the joint distribution on any finite subspace in $\{\hat{B}_w\}$ is also seen for subspaces of $\{B_w\}$ (with arbitrarily small error) and (b*) follows.
QED

2.5. Once we can guarantee the existence of combinatorial lines we have combinatorial subspaces of arbitrary (finite) dimension (for the same value of k).

Theorem. *Assume* DHJ$_k$. *Then for any positive integer d, every $\epsilon > 0$, there exist $n(d,\epsilon) = n(d,\epsilon,k)$ and $\delta = \delta(d,\epsilon)$ such that*

(i) *If $n \geq n(d,\epsilon)$, and $A \subset W_n(k)$, $|A| > \epsilon k^n$, then A contains a d-dimensional combinatorial subspace.*
(ii) *If $n \geq n(d,\epsilon)$, and for every $w \in W_n(k)$ there is given a measurable set B_w in some fixed probability measure space $\{X,\mathcal{B},\mu\}$, and $\mu(B_w) > \epsilon$, then there exists a d-dimensional combinatorial subspace $V \subset W_n(k)$ such that $\mu(\bigcap_{w\in V} B_w) > \delta$.*

Proof. Use induction on d. The assumption in the statement of the theorem is the case $d = 1$, statement (i). The implication (i) $\Rightarrow$ (ii) is clear for all values of d by a simple bookkeeping argument: each point in the probability space defines a set of words, namely those w such that B_w contains the point. For a subset of points of probability $> \epsilon/2$ the density of the corresponding set in $W(k)$ exceeds $> \epsilon/2$, and if $n > n(\epsilon/2)$ each such subset will contain a d-dimensional combinatorial subspace. Since the number of those is bounded, one at least will be the same for a set of points of probability greater than $\epsilon/2$ divided by their number; this is a lower bound for δ.

Assume now the validity of (ii) for $d - 1$, and let $\epsilon > 0$ be given. We take $n > n(d-1,\epsilon/2) + n(1,\delta(d-1,\epsilon/2))$, $A \subset W_n(k)$ of density at least ϵ, and consider $W_n(k)$ again as a direct product of $W_m(k)$ and $W_{n-m}(k)$. For a word $w \in W_m(k)$ denote $B_w = \{v \in W_{n-m}(k) : wv \in A\}$. The set $A_1 \subset W_m(k)$ of words w for which the measure of B_w is bigger than $\epsilon/2$ has measure at least $\epsilon/2$. If $m > n(d-1,\epsilon/2)$ there is a $(d-1)$-dimensional subspace $V_{d-1} \subset W_m(k)$ such that $\mu(\bigcap_{w\in V_{d-1}} B_w) > \delta(d-1,\epsilon/2)$, and if $n - m > n(1,\delta(d-1,\epsilon/2))$, the intersection will contain a combinatorial line, and the span of V_{d-1} and that line is a d-dimensional subspace in A. QED

Remark. We can obtain part (ii) trivially for strongly stationary processes (using induction), and obtain the general case from Proposition 2.3.

2.6. Lemma. *Let $\{B_w\}$ be a process in $\{X,\mathcal{B},\mu\}$ with stationary distribution. Then there exist invertible measure-preserving transformations $\{T_j^{(i)}\}_{j=1,2,\ldots}^{i=1,\ldots,k}$, and a set $A \subset X$ of positive measure such that writing*

$$(2.1) \qquad T(w) = T_1^{(w(1))} \cdots T_n^{(w(n))} \qquad \textit{for } w = \{w(1),\ldots,w(n)\},$$

we have $B_w = T(w)^{-1}A$.

Proof. The stationarity condition is equivalent to the statement that for arbitrary j, and $1 \leq i \leq k$, the obvious correspondence between the algebras $\mathcal{B}_0$ spanned by $\{B_w : w \in W_{j-1}(k)\}$ and $\mathcal{B}_1$ spanned by $\{B_{wi} : w \in W_{j-1}(k)\}$ is measure preserving. Remembering that any measure-preserving isomorphism between finite subalgebras of measurable sets is induced by an invertible measure-preserving point transformation, we can denote one such transformation by T_j^i.

QED

There is a vast amount of freedom in selecting the transformations above, and there is even more considering the fact that the data given by the process is superfluous in putting sets whose indices have different lengths on the same space. The real information is given by the joint distribution which ignores the relative positions of such pairs and consists of the full information about the algebras $\mathcal{B}_n$ spanned by $\{B_w : w \in W_n(k)\}$. The transformations defined above map from $\mathcal{B}_{j-1}$ into $\mathcal{B}_j$, but what really matters is the way the k images of $\mathcal{B}_{j-1}$ sit in $\mathcal{B}_j$ relative to each other.

2.7. Definition. A *$W(k)$-system* consists of a measure space $(X,\mathcal{B},\mu)$ together with k sequences $\{T_n^{(1)}\}_{n=1}^{\infty},\ldots,\{T_n^{(k)}\}_{n=1}^{\infty}$ of invertible measure-preserving transformations of X to itself.

We shall write $\mathcal{T} = (X,\mathcal{B},\mu,\{T_n^{(1)},\ldots,T_n^{(k)}\})$.

Conforming with the notation of the foregoing paragraph we shall write

$$(2.1) \qquad T(w) = T_1^{(w(1))} \cdots T_n^{(w(n))}, \qquad \text{for } w = \{w(1),\ldots,w(n)\},$$

and often describe a $W(k)$-system by means of the family $\{T(w) : w \in W(k)\}$ rather than enumerate the "generators" $\{T_n^{(j)}\}_{n=1}^{\infty}$, $j = 1,\ldots,k$. For instance, given a system $\mathcal{T}$ and a subspace $\Sigma W(k)$ we define the *restriction of $\mathcal{T}$ to $\Sigma W(k)$* as the system defined on the same probability space by the data $\{T(\Sigma(w))\}$. The point is that one can regard this either as restricting the domain of $T(\cdot)$ or as the pullback to the entire $W(k)$ via the map $\Sigma : W(k) \mapsto W(k)$ and as such it has the right structure. We denote this restriction-pullback by $\Sigma\mathcal{T}$, and refer to it simply as a *subsystem* of $\mathcal{T}$.

Proposition. DHJ$_k$ *is equivalent to the following statement*:

(c) *If* $\mathcal{T} = (X, \mathcal{B}, \mu, \{T_n^{(1)}, \ldots, T_n^{(k)}\})$ *is a* $W(k)$*-system and* $A \subset X$ *has positive measure, then there exists a line* w_α^x *in* $W(k)$ *such that* $\mu(\bigcap_{1 \le j \le k} T(w_\alpha^j)^{-1} A) > 0$.

Proof. This is an immediate consequence of 2.6 and the fact that (b*) $\Rightarrow$ (c) $\Rightarrow$ (b**). QED

2.8. If T is any measure-preserving transformation of $(X, \mathcal{B}, \mu)$ we denote by T^{-1} the induced transformation on the various function spaces on $(X, \mathcal{B}, \mu)$; namely for a function f on X, we shall write $T^{-1}f$ for the function

$$T^{-1}f(x) = f(Tx).$$

This notation ensures consistency, i.e.,

$$(TS)^{-1}f = S^{-1}T^{-1}f.$$

T^{-1} is an isometry on all the $L^p(X, \mathcal{B}, \mu)$, and in particular it is a unitary operator on $L^2(X, \mathcal{B}, \mu)$. We shall work mainly in L^2. Using this notation we now state the theorem in the form in which we propose to prove DHJ$_k$; the following clearly implies statement (c) of Proposition 2.6, and hence all the various forms discussed earlier in this section.

Theorem. *Let* $(X, \mathcal{B}, \mu, \{T_n^{(1)}, \ldots, T_n^{(k)}\})$ *be a* $W(k)$*-system, let* $f \in L^\infty(X, \mathcal{B}, \mu)$ *with* $f \ge 0$ *and with* $\int f\, d\mu > 0$. *Then there exists a line* w_α^x *in* $W(k)$ *such that*

$$\int T(w_\alpha^1)^{-1} f\, T(w_\alpha^2)^{-1} f \cdots T(w_\alpha^k)^{-1} f\, d\mu > 0. \tag{2.2}$$

3. Factorization

3.1. Let $(X, \mathcal{B}, \mu)$ be a probability space. We take X to be a compact metric space and $\mathcal{B} \subset$ Borel sets. If $\mathcal{D}$ is a sub-σ-algebra of $\mathcal{B}$, $L^2(X, \mathcal{D}, \mu)$ is a subspace of $L^2(X, \mathcal{B}, \mu)$. Since subspaces of a separable space are separable, it follows that $\mathcal{D}$ is countably generated. Choosing a countable generating sequence $\{D_n\} \subset \mathcal{D}$, and mapping $x \in X$ to $\{1_{D_n}(x)\} \in \{0,1\}^{\mathbf{N}}$, we can attach to a sub-σ-algebra $\mathcal{D} \subset \mathcal{B}$ a space $(Y, \mathcal{D}_Y, \nu)$ and a measurable measure preserving map $\pi : X \to Y$, so that $\mathcal{D} = \pi^{-1}(\mathcal{D}_Y)$. Moreover $\mathcal{D}_Y$ separates points in Y. We refer to $(Y, \mathcal{D}_Y, \nu)$ as a *factor* of $(X, \mathcal{B}, \mu)$, and we sometimes abuse the terminology and refer to $\mathcal{D}$ as a factor. Another way of describing the space Y is to define an equivalence relation on X, two points being equivalent if they cannot be separated by any D_n, that is if they have the same image under π, and Y is then the factor of X relative to this equivalence. For $\mathcal{D}$-measurable functions, i.e., functions which depend only on $y = \pi(x)$, we use the same symbol: if $g(x)$ is $\mathfrak{D}$-measurable on X, we denote by $g(y)$ the corresponding function on Y.

The proof of the following is left to the reader:

Lemma. *For a closed subspace $\mathfrak{L} \subset L^2(X,\mathfrak{B},\mu)$, containing the constants, the following three conditions are equivalent:*

(a) *$\mathfrak{L}$ has the form $\mathfrak{L} = L^2(X,\mathfrak{D},\mu)$, with $\mathfrak{D}$ a sub-σ-algebra of $\mathfrak{B}$.*

(b) *$\mathfrak{L}$ is spanned by a set of bounded functions $\mathfrak{L}_0$, and if $f, g \in \mathfrak{L}_0$ then $fg \in \mathfrak{L}$.*

(c) *$\mathfrak{L}$ is a lattice: $f, g \in \mathfrak{L} \Rightarrow f \vee g, f \wedge g \in \mathfrak{L}$.*

Here and throughout our discussion we take L^2 to be a Hilbert space of real-valued functions.

For a sub-σ-algebra $\mathfrak{D}$ of $\mathfrak{B}$, we denote by $P_{\mathfrak{D}} = E(\cdot\,|\mathfrak{D})$ the projection (conditional expectation): $L^2(X,\mathfrak{B},\mu) \to L^2(X,\mathfrak{D},\mu)$. We denote by I_π the isometry $L^2(Y,\mathfrak{D}_Y,\nu) \to L^2(X,\mathfrak{D},\mu) \subset L^2(X,\mathfrak{B},\mu)$. Note that $P = I_\pi I_\pi^*$.

3.2. For the following theorem see [F1].

Theorem. *Let $\pi : X \to Y$ be as above. There is a measurable* 1-1 *map* $Y \to \mathcal{P}(X)$, *the space of probability measures on X, $y \to \mu_y$, satisfying*

(i) $I_\pi^* f(y) = \int f\, d\mu_y$ *a.e.,*

(ii) $f \in L^2(X,\mathfrak{D},\mu) \Rightarrow f$ *constant with respect to μ_y for a.e. y,*

(iii) $\mu_y\{\pi^{-1}(y)\} = 1$ *for a.e. y,*

(iv) $\int f\, d\mu = \int \{\int f\, d\mu_y\}\, d\nu(y)$.

We speak of the family $\{\mu_y\}$ as the *disintegration of* μ with respect to $\mathfrak{D}$.

In view of the one–one correspondence between the points of Y and Borel measures on X we may identify Y with a subset of $\mathcal{P}(X)$. If σ is a Borel measurable 1-1 map of $X \to X$, then σ acts on $\mathcal{P}(X)$ and the measures $\sigma(\mu_y)$ are defined. Identifying y with μ_y we may write $\sigma(y)$ for the measures $\sigma(\mu_y)$. If σ preserves the measure μ, then $\{\sigma\mu_y\}$ is the disintegration of μ with respect to $\sigma(\mathfrak{D})$.

3.3. Part (iii) of Theorem 3.2 says that the measures μ_y are mutually singular, each carried by the corresponding $\pi^{-1}(y)$. We can consider a function f on X as a map $y \mapsto f_y = f|_{\pi^{-1}(y)}$ and part (iv) of the same theorem, applied to f^2, shows that if $f \in L^2(X,\mathfrak{B},\mu)$ then $f_y \in L^2(X,\mathfrak{B},\mu_y)$ for almost all y, and

$$\|f\|^2_{L^2(\mu)} = \int \|f_y\|^2_{L^2(\mu_y)}\, d\nu(y). \tag{3.1}$$

On the other hand, we say that a map $y \mapsto \varphi_y \in L^2(\mu_y)$ is *measurable* if there exists a measurable function f on X such that $f_y = \varphi_y$ μ_y-a.e.; in that case we have $\|f\|^2_{L^2(\mu)} = \int \|\varphi_y\|^2\, d\nu(y)$ (both sides finite and equal or both infinite). We refer to the space of all measurable maps $y \mapsto \varphi_y$ with finite $\int \| \nu_y \|^2 \, d\nu(y)$ as the direct integral of the spaces $L^2(\mu_y)$, denote it $\int_\oplus L^2(X,\mathcal{B},\mu_y)\,d\nu(y)$, endow it with the natural norm, and obtain:

Theorem. *The map* $f \to f_y = f|_{\pi^{-1}(y)} \in L^2(X,\mathcal{B},\mu_y)$ *maps* $L^2(X,\mathcal{B},\mu)$ *onto* $\int_\oplus L^2(X,\mathcal{B},\mu_y)\,d\nu(y)$ *isometrically.*

3.4. $L^\infty(X,\mathfrak{D},\mu)$ is an algebra of functions, and a closed subspace $\mathfrak{M} \subset L^2(X,\mathcal{B},\mu)$ which is a module over $L^\infty(X,\mathfrak{D},\mu)$, in other words, invariant under multiplication by bounded $\mathfrak{D}$-measurable functions, will be referred to simply as a $\mathfrak{D}$-*module.*

Given a sub-σ-algebra $\mathfrak{D}$ of $\mathcal{B}$, and a $\mathfrak{D}$-module $\mathfrak{M}$, we can write the direct integral decomposition of $\mathfrak{M}$ as in 3.3,

$$\mathfrak{M} = \int_\oplus \mathfrak{M}_y \, d\nu(y). \tag{3.2}$$

The spaces $\mathfrak{M}_y$ are obtained as the images of $\mathfrak{M}$ under the restriction maps $f \mapsto f_y$, but a word of caution is in order. Since $\mathfrak{M}$ is uncountable and since elements of $L^2(\mu)$ are equivalence classes whose representatives can be modified on sets of measure zero, one can choose the representatives of the elements of a "small" module $\mathfrak{M}$ in such a way as to obtain $\mathfrak{M}_y = L^2(\mu_y)$ for all y. One therefore always takes a dense countable subset, in this case in $\mathfrak{M}$, and defines $\mathfrak{M}_y$ as the closure in $L^2(\mu_y)$ of the image of this countable set under the restriction map for each y. (3.2) should be understood with this convention as should the statement describing $\mathfrak{M}_y$. We refer to $\mathfrak{M}_y$ as the *fiber of* $\mathfrak{M}$ *over* y.

On the other hand, suppose that we are given, for every y, a closed subspace $\mathfrak{M}_y \subset L^2(X,\mathcal{B},\mu_y)$. The set $\mathfrak{M} \subset L^2(X,\mathcal{B},\mu)$ of all f, such that $f_y \in \mathfrak{M}_y$ a.e., is clearly a $\mathfrak{D}$-module. If the data $\{\mathfrak{M}_y\}$ is measurable in any reasonable sense then (3.2) is valid.

Definition. The module $\mathfrak{M} \subset L^2(X,\mathcal{B},\mu)$ is *of finite rank* or simply *finite* if there is a finite set $\{F_j\}_{j=1}^n \subset L^2(X,\mathcal{B},\mu)$ such that $\mathfrak{M} = \bigcup F_j L^\infty(X,\mathfrak{D},\mu)$. We say that $\{F_j\}_{j=1}^n$ span $\mathfrak{M}$. We say that $\mathfrak{M}$ is *bounded* if $\mathfrak{M} \cap L^\infty(X,\mathcal{B},\mu)$ is dense in $\mathfrak{M}$ (in the $\mathcal{L}^2$-norm).

It is clear that if $\{F_j\}_{j=1}^n$ span $\mathfrak{M}$, then their restrictions to any $\pi^{-1}y$ span $\mathfrak{M}_y$, and the dimension of $\mathfrak{M}_y$ is bounded by n for all y. On the other hand, if for a module $\mathfrak{M}$ the dimension of $\mathfrak{M}_y$ is uniformly bounded a.e., then $\mathfrak{M}$ is finite.

The Gram–Schmidt procedure can be applied to a sequence (finite or infinite) in a module $\mathfrak{M}$ to yield a fiberwise orthonormal sequence. The normalization is done by dividing the function in question, say F_1, by $P_\mathfrak{D}|F_1|^2$. The fiber norm of the function so obtained is one (except on fibers on which $F_1 = 0$). We shall say that a (necessarily spanning) sequence $\{\Phi_j\}$ is *a global orthonormal basis for* $\mathfrak{M}$ if for almost all y the non-zero elements in $\{\Phi_{j,y}\}$ form an orthonormal basis for $\mathfrak{M}_y$.

3.5. Definition. The measure $\tilde{\mu}_{\mathfrak{D}}$ is a measure on $(X \times X, \mathcal{B} \times \mathcal{B})$ defined by

(i) $\int f \, d\tilde{\mu}_{\mathfrak{D}} = \int [\int f \, d(\mu_y \times \mu_y)] \, d\nu(y)$, $f = f(x,x')$ measurable $\mathcal{B} \times \mathcal{B}$,

or, equivalently, by

(ii) $\int \varphi \otimes \psi \, d\tilde{\mu}_{\mathfrak{D}} = \int P_{\mathfrak{D}} \varphi P_{\mathfrak{D}} \psi \, d\mu$.

If we denote by $\tilde{\mathfrak{D}}$ the subalgebra of $\mathcal{B} \times \mathcal{B}$ spanned by the sets $\{A \times A : A \in \mathfrak{D}\}$ and remember that $\pi(x) = \pi(x')$ ($\tilde{\mu}$ a.e.), several things are clear: $\tilde{\mathfrak{D}}$ is a subalgebra of $\mathcal{B} \times \mathcal{B}$, isomorphic to $\mathfrak{D}$, and the corresponding space $\tilde{Y}$ can be identified with Y (we shall abuse the notation and identify $\mathfrak{D}$ with $\tilde{\mathfrak{D}}$ and Y with $\tilde{Y}$); finally, the definition above gives the disintegration of $\tilde{\mu}_{\mathfrak{D}}$ relative to $\mathfrak{D}$.

3.6. Definition. A *PDS (positive definite symmetric) kernel for* $\mathfrak{D}$ is a bounded function $H(x,x')$ on $X \times X$ defined a.e. with respect to $\tilde{\mu}_{\mathfrak{D}}$ and satisfying

(i) $H(x,x') = H(x',x)$,

(ii) $\int H(x,x')\varphi(x)\varphi(x') \, d\tilde{\mu}_{\mathfrak{D}}(x,x') \geq 0$ for all $\varphi \in L^\infty(X, \mathcal{B}, \mu)$.

Given $H \in L^2(\tilde{\mu}_{\mathfrak{D}})$ we can define the operator

$$\tilde{H}\varphi(x) = \int H(x,x')\varphi(x') \, d\mu_{\pi(x)}(x').$$

If H is PDS then the operator $\tilde{H}$ is positive-(semi)-definite, and self-adjoint, and, as we shall presently see, is the integral of compact (actually Hilbert–Schmidt) operators.

Lemma. *If H is a* PDS *kernel for $\mathfrak{D}$, then for a.e. y, $H(x,x')$ is defined a.e. with respect to $\mu_y \times \mu_y$ and the integral operator $\tilde{H}_y$ it defines on $L^2(\mu_y)$ is positive-(semi)-definite, self-adjoint, and compact (in fact Hilbert–Schmidt).*

Proof. The ingredients are Theorem 3.3 (for $\tilde{\mu}_{\mathfrak{D}}$), and the properties of PDS kernels as decreed in the definition above. Thus, $H_y = H|_{\pi^{-1}(y)}$ is bounded and symmetric on $\mu_y \times \mu_y$, and the only thing to prove is that it is positive-semi-definite for almost all y. Condition (ii) above can be written as

$$\int \left[\int H(x,x')\varphi(x)\varphi(x') d(\mu_y \times \mu_y) \right] d\nu(y) \geq 0$$

and the fact that this is true for all $\varphi \in L^\infty(X, \mathcal{B}, \mu)$ allows us to "localize", i.e. apply it to the products of a given φ by arbitrary indicator functions of sets in $\mathfrak{D}$, and obtain that for any such φ and ν-a.e. y,

$$\int H(x,x')\varphi(x)\varphi(x') d(\mu_y \times \mu_y) \geq 0,$$

and since we can take a countable collection of such φ's which is dense in $L^2(\mu_y)$ for a.e. y, the result follows. QED

3.7. The spectral theorem for compact self-adjoint operators gives the following picture for every $\tilde{H}_y$: The spectrum consists of a sequence of positive numbers and possibly (certainly if the dimension is infinite) of zero, and zero is the only (possible) limit point of this set; each $\lambda \neq 0$ in the spectrum is an eigenvalue with finite multiplicity, the eigenspace corresponding to $\lambda = 0$, i.e. the null-space of $\tilde{H}_y$ can be empty, finite or infinite dimensional. We can list the non-zero eigenvalues, repeating each according to its multiplicity, as $\lambda_1(y) \geq \lambda_2(y) \geq \cdots$, and denote a corresponding orthonormal sequence of eigenvectors $\{\varphi_{n,y}\}$, which means

$$\langle \varphi_{n,y}, \varphi_{m,y} \rangle_{L^2(\mu_y)} = \delta_{n,m}, \qquad \tilde{H}_y \varphi_{n,y} = \lambda_n(y) \varphi_{n,y}.$$

We then have

$$H_y(x, x') = \sum_n \lambda_n(y) \varphi_{n,y}(x) \varphi_{n,y}(x')$$

and

$$\sum_n (\lambda_n(y))^2 = \|H_y\|^2_{L^2(\mu_y \times \mu_y)} \leq \sup |H(x, x')|^2.$$

To get an orthonormal basis for $L^2(\mu_y)$ one may have to supplement $\{\varphi_{n,y}\}$ with an orthonormal basis $\{\varphi_{-n,y}\}$ of the null-space of $\tilde{H}_y$.

Remark. If the kernel H is bounded then all the eigenfunctions thereof are clearly bounded. We shall use this fact repeatedly.

Our next observation is that, as expected, this entire structure depends on y in a measurable way. Begin with

Lemma. *Let H be a compact, positive semi-definite, self-adjoint operator on a Hilber space $\mathcal{H}$. Write the spectrum of H, counting multiplicities, as $\{\lambda_j\}$ with $\lambda_j \geq \lambda_{j+1}$ and denote by $\{\varphi_j\}$ a corresponding orthonormal sequence of eigenvectors, i.e., $H\varphi_j = \lambda_j \varphi_j$. Then, for any orthonormal sequence $\{\psi_j\}_1^n \subset \mathcal{H}$*

$$\sum_{j,k=1}^{n} \langle H\psi_j, \psi_k \rangle^2 \leq \sum_1^n \lambda_j^2. \tag{3.3}$$

If $\lambda_{n+1} < \lambda_n$, we have equality in (3.3) *if, and only if, the span of $\{\varphi_j\}_{j=1}^n$ is the same as that of $\{\psi_j\}_{j=1}^n$. We have approximate equality if the distance of every ψ_j, $j = 1, \ldots, n$ to the span of $\{\varphi_j\}_1^n$ is very small.*

Proof. Complete the sequence $\{\varphi_j\}$, if necessary, to an orthonormal basis of $\mathcal{H}$ by adding an orthonormal basis of the null-space of H, say $\{\varphi_j\}_{j \leq 0}$. Write $\psi_k = \sum a_{k,j} \varphi_j$ so that $\sum_j a_{k,j}^2 = 1$ and, as the ψ_k's are orthogonal, also $\sum_{k=1}^n a_{k,j}^2 \leq 1$ for all j. Now $H\psi_k = \sum_j a_{k,j} \lambda_j \varphi_j$, and

(3.4) $$\sum_{j,k=1}^{n} \langle H\psi_j, \psi_k\rangle^2 \le \sum_1^n \|H\psi_k\|^2 = \sum_j \left(\sum_{k=1}^n a_{k,j}^2\right)\lambda_j^2 \le \sum_1^n \lambda_j^2.$$

If the spans of $\{\varphi_j\}_{j=1}^n$ and $\{\psi_j\}_{j=1}^n$ are the same, we clearly have equality, otherwise there is a loss of at least $(\lambda_n^2 - \lambda_{n+1}^2)(\sum_{j\le 0} + \sum_{j>n})a_{k,j}^2$. QED

Theorem. *The functions* $\lambda_j(y)$ *are measurable; moreover, there exists a sequence of functions* $\Phi_n \in L^2(\mu)$ *such that* $\langle \Phi_{n,y}, \Phi_{m,y}\rangle_{L^2(\mu_y)} = \delta_{n,m}$, *and*

$$\tilde{H}_y \Phi_{n,y} = \lambda_n(y)\Phi_{n,y}.$$

Proof. Denote by U the set $\{f : f \in L^2(X, \mathcal{B}, \mu),\ P_{\mathfrak{D}}|f|^2 \le 1 \text{ on } Y\}$. The lemma above implies that for all n,

(3.5) $$\sum_{j=1}^n \lambda_j^2(y) = \sup \sum_{j,k=1}^n (P_{\mathfrak{D}}(H(x,x')f_j(x)f_k(x')))^2,$$

the supremum being taken over all choices of n-tuples $\{f_j\} \subset U$ such that, for $j \ne k$, $P_{\mathfrak{D}} f_j f_k = 0$. This shows the measurability of the λ_j's.

It is equally clear that if $\{\Phi_j\} \subset U$ is such that $P_{\mathfrak{D}}\Phi_j\Phi_k = 0$ (for $j \ne k$) and for all n, $\int H(x,x')\Phi_n(x)\Phi_n(x')\, d\tilde{\mu}_{\mathfrak{D}}(x,x')$ is the maximum of

(3.6) $$\int H(x,x')f(x)f(x')\, d\tilde{\mu}_{\mathfrak{D}}(x,x')$$

under the conditions: $f \in U$, and $P_{\mathfrak{D}} f\Phi_j = 0$, $j = 1, \dots, n-1$, then in fact

$$\tilde{H}_y \Phi_{n,y} = \lambda_n(y)\Phi_{n,y}.$$

Thus all we have to do is show the existence of such $\{\Phi_j\}$.

We obtain the Φ's one term at a time, and the argument is the same at every stage and requires just the following observation: let $\{f_j\} \subset U$ be such that $\int H(x,x')f_j(x)f_j(x')\, d\tilde{\mu}_{\mathfrak{D}}(x,x')$ converges to its supremum. For j large enough, say $j \ge j_0$, $f_{j,y}$ is very close to an eigenvector of $\tilde{H}_y$ corresponding to $\lambda_1(y)$ for all but a very small set of y's. Assume, just to make the argument as simple as possible, that $\lambda_1(y)$ has multiplicity 1 a.e.; then, for $j > j_0$, there exists a $\mathfrak{D}$-measurable function g_j of modulus 1, such that $g_j f_j$ and f_{j_0} are close to the same $\lambda_1(y)$-eigenvector for most y. On most fibers $g_j f_j$ and f_{j_0} are therefore close and $\|g_j f_j - f_{j_0}\|_{L^2(\mu)}$ is small. On the other hand, the integral (3.6) for $g_j f_j$ and for f_j are the same. Thus, having fixed j_0, we can look for the remaining f_j's in the ball $\|f - f_{j_0}\|_{L^2(\mu)} \le \frac{1}{2}$ without affecting the supremum. If the multiplicity of $\lambda_1(y)$ is not always 1, the argument is similar except that instead of the one dimensional rotation, given by multiplication by ± 1, we use rotations in the appropriate finite dimensional subspace (of $L^2(\mu_y)$). Repeating with a (new) $j_1 > j_0$ such that the integral (3.6) for f_j, $j \ge j_1$ is much closer to the supremum, we can add the condition $\|f - f_{j_1}|_{L^2(\mu)} \le \frac{1}{4}$ for $j \ge j_1$ without affecting the supremum.

In other words there is no loss of generality in assuming that the maximizing sequence $\{f_j\}$ is a Cauchy sequence in $L^2(\mu)$ and its limit can be taken as Φ_1. The other Φ's are obtained similarly. QED

Remarks. (a) We clearly have $\lambda_j(y) = E(H(x,x')\Phi_j(x)\Phi_j(x') \mid \mathfrak{D})$, and

$$H(x,x') = \sum_n \lambda_n(y)\Phi_n(x)\Phi_n(x') \qquad \tilde{\mu}\text{-a.e.} \tag{3.7}$$

(b) In order to get an orthonormal basis for $L^2(\tilde{\mu})$ we typically have to add $\{\Psi_n\}$ which span the null-spaces of $\tilde{H}_y$ (the null-module of $\tilde{H}$). Since the dimensions may vary from point to point we have to allow the Φ's and Ψ's to be zero on some fibers.

3.8. The following (trivial) extension of Lemma 3.7 to the context of the theorem will be useful:

Lemma. *If H is a* PDS *kernel for $\mathfrak{D}$, and $\{\Psi_j\}_{j=1}^n$ is a fiberwise orthonormal sequence on a set $B \in \mathfrak{D}$ and zero outside B, then*

$$\sum_{j,k=1}^{n} \int \langle H(x,x')\Psi_j(x), \Psi_k(x')\rangle^2_{L^2(\mu_y)}\, d\nu \le \int_B \sum_1^n \lambda_j^2\, d\nu.$$

Conditions for equality (or approximate equality) are like in Lemma 3.7.

4. IP-systems and quasi-invariant factors

4.1. We denote by $\mathcal{F}$ the set of finite subsets of $\mathbf{N}$ (the positive integers) and refer to families indexed by $\mathcal{F}$ as $\mathcal{F}$-sequences. Given a family $\{\alpha_j\}$ of disjoint elements of $\mathcal{F}$, we can refer to the set $\mathcal{F}_0$ of all finite unions $\alpha_{j_1} \cup \cdots \cup \alpha_{j_s}$ as a *subring* of $\mathcal{F}$. An $\mathcal{F}$-subsequence is the restriction of an $\mathcal{F}$-sequence to a subring of $\mathcal{F}$.

We use the following partial order on $\mathcal{F}$: $\alpha < \beta$ will mean that $\max \alpha < \inf \beta$.

Definition. Let $\{x_\alpha\}$ be an $\mathcal{F}$-sequence in a topological space X. We say that the sequence *IP-converges* to x_0 and write† $x_0 = \text{IP-lim}\, x_\alpha$ if for every neighborhood V of x_0 there exists a $\beta \in \mathcal{F}$ such that if $\alpha > \beta$, then $x_\alpha \in V$.

We shall speak informally of $\alpha \to \infty$ meaning that $\min \alpha \to \infty$. We may then write $x_\alpha \to x_0$ as $\alpha \to \infty$ instead of $\text{IP-lim}\, x_\alpha = x_0$.

The following is equivalent to Hindman's theorem:

Theorem. *Every $\mathcal{F}$-sequence in a compact metric space has a convergent $\mathcal{F}$-subsequence.*

†We often omit the IP- initial outside of formulas.

Hindman's theorem is a special case of Theorem 1.5. To see this suppose we have an $\mathcal{F}$-sequence in X, a compact metric space. Regard this as a function g from $\mathcal{F}$ to X. Now choose any k and consider the map $h: W(k,1) \to \mathcal{F}$ in which a word $w(x)$ is mapped to the set of positions in which x appears. Then $f = g \circ h$ is a function to which Theorem 1.5 may be applied. The result can be reformulated as the assertion that an $\mathcal{F}$-subsequence converges. For details see [FK4].

4.2. Definition. An $\mathcal{F}$-sequence $\{\tau_\alpha\}$, with values in a semigroup, is an *IP-system* if for $\alpha < \beta$,

$$\tau_{\alpha\cup\beta} = \tau_\alpha \tau_\beta.$$

We say that $\{\tau_\alpha\}$ is a *reversed IP-system* if

$$\tau_{\alpha\cup\beta} = \tau_\beta \tau_\alpha.$$

IP-limits of systems in compact semigroups are necessarily idempotents. We shall make use of this in the following context:

Theorem. *Let $\{\tau_\alpha\}$ be an IP-system (or reversed IP-system) of unitary operators on a Hilbert space, and assume that* IP-lim $\tau_\alpha = P$ *in the weak operator topology. Then P is an orthogonal projection.*

For proof cf. [FK2], Theorem 1.7.

4.3. The following is Lemma 5.3 of [FK2]. It is inspired by van der Corput's classic proof of the uniform distribution of sequences with uniformly distributed differences and we sometimes refer to it as VDC.

Lemma. *Suppose $\{x_\alpha\}_{\alpha\in\mathcal{F}}$ is a bounded $\mathcal{F}$-sequence of vectors in a Hilbert space $\mathcal{H}$. If*

$$\operatorname*{IP-lim}_{\alpha\in\mathcal{F}} \operatorname*{IP-lim}_{\beta\in\mathcal{F}} \langle x_\beta, x_{\alpha\cup\beta}\rangle = 0$$

then for some subring $\mathcal{F}_0 \subset \mathcal{F}$

$$\operatorname*{IP-lim}_{\alpha\in\mathcal{F}_0} x_\alpha = 0$$

in the weak topology of $\mathcal{H}$. In particular, if $\text{IP-lim}_{\alpha\in\mathcal{F}} x_\alpha$ *is known to exist, it must vanish.*

Proof. The existence of a subring $\mathcal{F}_0 \subset \mathcal{F}$ along which the weak limit exists is a consequence of 4.1 and we may therefore assume that $\text{IP-lim}_{\alpha\in\mathcal{F}} x_\alpha = u$ and prove $u = 0$. Assume $u \neq 0$; if β is sufficiently far out, say $\beta \cap \gamma_0 = \varnothing$, then $\langle x_\beta, u\rangle > \delta = \frac{1}{2}\|u\|^2$. Choose $\epsilon > 0$. The condition of the lemma can be restated as follows. There exists $\gamma_1 \in \mathcal{F}$ such that if $\alpha \cap \gamma_1 = \varnothing$ there exist

$\gamma_2(\alpha) \in \mathfrak{F}$ such that $\beta \cap \gamma_2(\alpha) = \varnothing$ implies that $\langle x_\beta, x_{\beta\cup\alpha}\rangle < \epsilon$. Choose $\beta_1, \beta_2, \ldots, \beta_n$ inductively with $\beta_i \cap (\gamma_1 \cup \gamma_0) = \varnothing$ and with $\beta_j \cap \gamma_2(\beta_1 \cup \beta_2 \cup \cdots \cup \beta_i) = \varnothing$ for $j > i$; then $\langle x_{\beta_1\cup\beta_2\cup\cdots\cup\beta_i}, x_{\beta_1\cup\beta_2\cup\cdots\cup\beta_j}\rangle < \epsilon$ for $i < j$.

Write $y_i = x_{\beta_1\cup\cdots\cup\beta_i}$. Then $\langle y_i, u\rangle > \delta$ and at the same time $\langle y_i, y_j\rangle < \epsilon$ for $i \neq j$. But

$$\left\langle \frac{y_1 + \cdots + y_k}{k}, u\right\rangle > \delta \quad \text{implies} \quad \left\| \frac{y_i + \cdots + y_k}{k}\right\| > \delta\|u\|^{-1} = \frac{1}{2}\|u\|,$$

and, on the other hand,

$$\langle y_i, y_j\rangle < \epsilon \quad \text{implies} \quad \left\|\frac{y_1 + \cdots + y_k}{k}\right\|^2 < \frac{\max\{\|y_i\|^2\}}{k} + \epsilon.$$

Choosing $\epsilon > 0$ small and k large one is led to a contradiction which proves the lemma. QED

4.4. Let $\{\tau_\alpha\}_{\alpha\in\mathfrak{F}}$ be an IP-system of 1–1 measure preserving transformations of a space $(X, \mathcal{B}, \mu)$. The induced unitary operators on $L^2(X, \mathcal{B}, \mu)$ will be denoted $\{\tau_\alpha^{-1}\}$:

$$\tau_\alpha^{-1} f(x) = f(\tau_\alpha x).$$

Note that for $\alpha < \beta$, $\tau_\beta^{-1}\tau_\alpha^{-1} = \tau_{\alpha\cup\beta}^{-1}$, so that $\{\tau_\alpha^{-1}\}$ is a reversed IP-system.

Let $\mathfrak{D} \subset \mathcal{B}$ define a factor of $(X, \mathcal{B}, \mu)$ and let $P = P_{\mathfrak{D}}$ be the corresponding orthogonal projection.

Definition. $\mathfrak{D}$ or $P_{\mathfrak{D}}$ is *quasi-invariant with respect to* $\{\tau_\alpha\}$: *if*

$$\tau_\alpha P_{\mathfrak{D}} \tau_\alpha^{-1} \to P_{\mathfrak{D}} \tag{4.1}$$

as $\alpha \to \infty$.

The convergence is in either the weak or strong operator topology which coincide for projections. Notice that because of the strong convergence we have

$$P_{\mathfrak{D}}\tau_\alpha^{-1} f \sim \tau_\alpha^{-1} P_{\mathfrak{D}} f$$

in norm for all $f \in L^2(X, \mathcal{B}, \mu)$. This implies in particular that $P_{\mathfrak{D}}\tau_\alpha^{-1} f$ and $P_{\mathfrak{D}} f$ are close in distribution; thus if $B \in \mathfrak{D}$, then $\tau_\alpha^{-1} 1_B$ is, for α sufficiently far, close to an indicator function (which depends on α). Similarly, if $\{\varphi_j\}$ is fiber-wise orthonormal, that is if $P_{\mathfrak{D}}\varphi_j\varphi_k = \delta_{j,k}$, then so is, asymptotically, $\{\tau_\alpha^{-1}\varphi_j\}$. Finally, if $f \in L^2(X, \mathcal{B}, \mu)$ and $P_{\mathfrak{D}} f \leq 1$, then $P_{\mathfrak{D}}\tau_\alpha^{-1} f \leq 1 + o(1)$ but for an exceptional set whose measure goes to zero as $\alpha \to \infty$.

4.5. If $\mathfrak{L}$ is a subspace of $L^2(X,\mathcal{B},\mu)$, we write

$$\operatorname{dist}(f,\mathfrak{L}) = \min\{\|f-g\| : g \in \mathfrak{L}\}.$$

Definition. If $\mathfrak{D}$ is $\{\tau_\alpha\}$-quasi-invariant and $\mathfrak{M}$ is a $\mathfrak{D}$-module, we say that $\mathfrak{M}$ is *quasi-invariant* (with respect to $\{\tau_\alpha\}$) if for every $f \in \mathfrak{M}$

$$\operatorname{dist}(\tau_\alpha^{-1}f,\mathfrak{M}) \to 0.$$

Theorem. *Let $\mathfrak{D}$ be $\{\tau_\alpha\}$-quasi-invariant, $P = P_{\mathfrak{D}}$ the corresponding orthogonal projection, and assume that for all $\varphi, \psi \in L^\infty(X,\mathcal{B},\mu)$*

$$\Lambda(\varphi,\psi) = \lim_{\alpha\to\infty}\langle P(\varphi\tau_\alpha^{-1}\psi), \varphi\tau_\alpha^{-1}\psi\rangle \tag{4.2}$$

exists. Let $\mathfrak{M}$ be a finite $\mathfrak{D}$-module with orthonormal global basis $\varphi_1,\ldots,\varphi_r$ (see §3.4). *Then $\mathfrak{M}$ is quasi-invariant w.r.t. $\{\tau_\alpha\}$ if, and only if,*

$$\sum\Lambda(\varphi_i,\varphi_j) = \Sigma\|\varphi_i\|^2.$$

Proof. For any $f \in L^2(X,\mathcal{B},\mu)$ we have, by Parseval's inequality for each $L^2(\mu_y)$,

$$\sum_i P(\varphi_i\tau_\alpha^{-1}f)^2 \le P(\tau_\alpha^{-1}f^2).$$

Hence

$$\sum_i \int P(\varphi_i\tau_\alpha^{-1}f)^2\,d\mu \le \int P(\tau_\alpha^{-1}f^2)\,d\mu = \int f^2\,d\mu = \|f\|^2. \tag{4.3}$$

We are close to equality in (4.3) when for most y, $\tau_\alpha^{-1}f$ is close to a vector in $\mathfrak{M}_y$, and so $\tau_\alpha^{-1}f$ is close to $\mathfrak{M}$. The converse will clearly also be true. It follows that as $\alpha\to\infty$, $\tau_\alpha^{-1}\phi_i$ is close to $\mathfrak{M}$ for all i if, and only if,

$$\lim\sum_{i,j}\int P(\varphi_i\tau_\alpha^{-1}\varphi_j)^2\,d\mu = \sum\Lambda(\varphi_i,\varphi_j) = \Sigma\|\varphi_i\|^2.$$

This proves the theorem. QED

Corollary. *The same argument yields: Let $\mathfrak{M}$ be a finite rank module with global basis $\varphi_1,\ldots,\varphi_s$. For any $f \in L^2(X,\mathcal{B},\mu)$, if*

$$\sum\Lambda(\varphi_i,f) = \|f\|^2$$

then $\operatorname{dist}(\tau_\alpha^{-1}f,\mathfrak{M}) \to 0$ *as* $\alpha\to\infty$.

4.6. Definition. Assume that $\mathfrak{D}$ is $\{\tau_\alpha\}$-quasi-invariant. A function H in $L^2(\tilde\mu_{\mathfrak{D}})$ is said to be *quasi-invariant with respect to* $\{\tau_\alpha\}$ if for every φ, $\psi \in L^\infty(X,\mathcal{B},\mu)$

$$\int \tau_\alpha^{-1}\varphi \otimes \tau_\alpha^{-1}\psi H\, d\tilde{\mu}_{\mathfrak{D}} \to \int \varphi \otimes \psi H\, d\tilde{\mu}_{\mathfrak{D}}. \tag{4.4}$$

Note that if $H =$ constant, (4.4) reduces to†

$$\langle P\tau_\alpha^{-1}\varphi, P\tau_\alpha^{-1}\psi\rangle \to \langle P\varphi, P\psi\rangle. \tag{4.5}$$

This is valid since, by (4.1),

$$\langle P\tau_\alpha^{-1}\varphi, P\tau_\alpha^{-1}\psi\rangle = \langle P\tau_\alpha^{-1}\varphi, \tau_\alpha^{-1}\psi\rangle = \langle \tau_\alpha P\tau_\alpha^{-1}\varphi, \psi\rangle \to \langle P\varphi, \psi\rangle = \langle P\varphi, P\psi\rangle.$$

Another way of stating (4.5) is: as $\alpha \to \infty$,

$$\tau_\alpha \times \tau_\alpha \tilde{\mu}_{\mathfrak{D}} \to \tilde{\mu}_{\mathfrak{D}}.$$

Similarly, H is quasi-invariant if, and only if,

$$\tau_\alpha \times \tau_\alpha H\tilde{\mu}_{\mathfrak{D}} \to H\tilde{\mu}_{\mathfrak{D}}.$$

The convergence in both cases is in the weak topology on measures, determined by $L^\infty(X,\mathcal{B},\mu) \otimes L^\infty(X,\mathcal{B},\mu)$.

Example. If $\mathfrak{D}$ is $\{\tau_\alpha\}$-quasi-invariant, $\varphi \in L^\infty(X,\mathcal{B},\mu)$ and, assuming the existence of the limits, define a measure σ as the limit (for $\alpha \to \infty$) of $\tau_\alpha \times \tau_\alpha(\varphi \otimes \varphi\, \tilde{\mu}_{\mathfrak{D}})$, i.e., for bounded ϑ and ψ

$$\int \vartheta \otimes \psi\, d\sigma = \lim_\alpha \int \varphi \otimes \varphi(\tau_\alpha^{-1}\vartheta \otimes \tau_\alpha^{-1}\psi)\, d\tilde{\mu}_{\mathfrak{D}},$$

then $\sigma = H\tilde{\mu}_{\mathfrak{D}}$ with H quasi-invariant PDS.

The quasi-invariance is clear and the fact that σ is absolutely continuous with bounded Radon–Nikodym derivative follows from the positivity of $\tilde{\mu}_{\mathfrak{D}}$; in fact if $|\varphi| \le 1$ then $\tilde{\mu}_{\mathfrak{D}} - \sigma = \lim(\tau_\alpha \times \tau_\alpha)(1 - \varphi_1 \otimes \varphi_1)\tilde{\mu}_{\mathfrak{D}}$ is clearly a non-negative measure.

4.7. Definition. Assume that $\mathfrak{D}$ is $\{\tau_\alpha\}$-quasi-invariant. We say that the set $B \in \mathfrak{D}$ is $\{\tau_\alpha\}$-quasi-invariant if the measure of the symmetric difference of B and $\tau_\alpha^{-1}B$ tends to zero.

The collection of all quasi-invariant sets forms a subalgebra of $\mathfrak{D}$.

Theorem. *Assume that $H \in L^2(\tilde{\mu}_{\mathfrak{D}})$ is quasi-invariant with respect to $\{\tau_\alpha\}$. Then the functions $\lambda_j(y)$ are measurable with respect to the subalgebra of quasi-invariant sets.*

Proof. If $H \in L^2(\tilde{\mu}_{\mathfrak{D}})$ is a $\{\tau_\alpha\}$-quasi-invariant PDS with spectrum $\{\lambda_j(y)\}$, and if $B \in \mathfrak{D}$ is $\{\tau_\alpha\}$-quasi-invariant, then $1_B(x)1_B(x')H$ is $\{\tau_\alpha\}$-quasi-invariant

†$\langle\ ,\ \rangle$ without subscript means $\langle\ ,\ \rangle_{L^2(X,\mathcal{B},\mu)}$.

with spectrum $\{1_B(y)\lambda_j(y)\}$. On the other hand if $B = \{y : \lambda_1(y) \geq b\}$ for some $b > 0$ then $\| 1_B \Phi_1 \|^2_{L^2(\mu)} = \nu(B)$ and

$$\int H(x,x')\tau_\alpha^{-1}(1_B\Phi_1(x))\tau_\alpha^{-1}(1_B\Phi_1(x))\, d\tilde{\mu}_{\mathfrak{D}}$$

$$\leq \int \lambda_1(y) P\tau_\alpha^{-1}(1_B\Phi_1(x))^2\, d\nu \leq \int_B \lambda_1(y)\, d\nu + o(1).$$

Since $\int P\tau_\alpha^{-1}(1_B\Phi_1(x))^2\, d\nu = \|1_B\Phi_1\|^2_{L^2(\mu)} = \nu(B)$, and, as noted in 4.4,

$$\|\tau_\alpha^{-1}(1_B\Phi_1(x))\|_y^2 = P\tau_\alpha^{-1}(1_B\Phi_1(x))^2 \leq 1 + o(1)$$

on most of the space (as $\alpha \to \infty$), the integral is maximized by putting all the mass where λ_1 is biggest. As $\alpha \to \infty$ the left-hand side converges to the right-hand side which is possible only if $\tau_\alpha^{-1} 1_B \to 1_B$; thus B is quasi-invariant. Similarly, defining $B = \{y : \sum_1^n \lambda_j^2(y) \geq b\}$ for some $b > 0$ and applying Lemma 3.8 to $\{1_B\Phi_j\}_1^n$ we again obtain that B is quasi-invariant. QED

4.8. Theorem. *Let $\mathfrak{D}$ be $\{\tau_\alpha\}$-quasi-invariant, and H a $\{\tau_\alpha\}$-quasi-invariant PDS kernel for $\mathfrak{D}$, with spectrum $\{\lambda_j(y)\}$. Assume $l > 0$ is such that for a.e. y, all $\lambda_n(y) \neq l$. Let $\{\varphi_n(y)\}$ be a corresponding orthonormal eigenvector sequence. For each y, set*

$$\varphi'_{n,y} = \begin{cases} \varphi_{n,y} & \text{if } \lambda_n(y) > l, \\ 0 & \text{if } \lambda_n(y) < l. \end{cases}$$

Let $\mathfrak{M}_y \subset L^2(X, \mathfrak{B}, \mu_y)$ be the finite dimensional subspace spanned by $\{\varphi'_{n,y}\}_{n \geq 1}$, and let

$$\mathfrak{M} = \int_{\oplus} \mathfrak{M}_y\, d\nu(y).$$

Then $\mathfrak{M}$ is a finite module which is quasi-invariant with respect to $\{\tau_\alpha\}$.

Proof. By Theorem 4.7, the sets $\{y : \lambda_{N+1}(y) < l < \lambda_N(y)\}$ are quasi-invariant for every N, and, restricting to one of them, we may assume that $\varphi'_n = \varphi_n$ for $n \leq N$, and $\varphi'_n = 0$ for $n > N$. By the same theorem we may restrict to even smaller quasi-invariant sets defined by further restrictions on $\lambda_j(y)$, and assume with no loss of generality that the λ's are (close to) constants.

We have already noticed in §4.4 that for α sufficiently far, $\{\tau_\alpha^{-1}\varphi_n\}_{n=1}^N$ is almost orthogonal fiberwise, and assumption (4.4) implies

$$\int \langle H(x,x')\tau_\alpha^{-1}\varphi_n(x), \tau_\alpha^{-1}\varphi_m(x')\rangle_{L^2(\mu_y)}\, d\nu \to \begin{cases} \int \lambda_n(y)\, d\nu, & \text{if } n = m, \\ 0, & \text{if } n \neq m. \end{cases}$$

By convexity of the square, and since the λ's are essentially constants, we have for any $\epsilon > 0$ and α sufficiently large

$$\sum_{n,m=1}^{N} \int \langle H(x,x')\tau_\alpha^{-1}\varphi_n(x), \tau_\alpha^{-1}\varphi_m(x')\rangle^2_{L^2(\mu_y)}\, d\nu \geq \int \sum_{n=1}^{N} \lambda_n^2(y)\, d\nu - \epsilon$$

and this, combined with Lemma 3.8, completes the proof. QED

Corollary. *The functions in the range of $\tilde{H}$ can be approximated in norm by functions of finite rank (i.e., belonging to finite dimensional modules).*

5. $W(k)$-systems

5.1. We have defined $W(k)$-systems in 2.7 and stayed with them just long enough to state DHJ_k as Theorem 2.8. Having discussed some preliminary material in sections 2 and 3, we now return to lay the groundwork for the more detailed study of these systems in the following sections, needed for the proof of DHJ_k.

Let $\mathcal{T}$ be a $W(k)$-system with transformations $T(w)$.

Notation. For $w \in W(k)$ of length $l(w)$, if $\alpha_1, \alpha_2, \ldots, \alpha_l$ are disjoint subsets of the interval $[1,2,\ldots,|w|]$, and $u_1, u_2, \ldots, u_t; v_1, v_2, \ldots, v_t$ are words with $l(u_j) = l(v_j) = |\alpha_j|$, we define

$$(5.1) \qquad \tau_{v_1,v_2,\ldots,v_l}^{u_1,u_2,\ldots,u_l}(\alpha_1,\alpha_2,\ldots,\alpha_l; w) = T(w_{\alpha_1,\alpha_2,\ldots,\alpha_l}^{u_1,u_2,\ldots,u_l})T(w_{\alpha_1,\alpha_2,\ldots,\alpha_l}^{v_1,v_2,\ldots,v_l})^{-1}.$$

Note that the expression does not depend on $w(n)$ for $n > \max(\alpha_1 \cup \alpha_2 \cup \cdots \cup \alpha_l)$, i.e., this expression satisfies the condition of *admissibility* of Definition 1.7. In view of this we can also define, for $\omega \in \Omega(k)$,

$$(5.2) \qquad \tau_{v_1,v_2,\ldots,v_l}^{u_1,u_2,\ldots,u_l}(\alpha_1,\alpha_2,\ldots,\alpha_l;\omega)$$

by the expression in (5.1) for w = long initial segment of ω. When the words u_s are constant, i.e. the same digit i_s repeated $|u_s|$ times, we replace u_s by i_s in our notation. In particular we write $\tau_j^i(\alpha;\omega)$ instead of $\tau_{j,j,j,\ldots}^{i,i,i,\ldots}(\alpha;\omega)$.

The following lemma is fundamental.

Lemma.

$$(5.3) \qquad \tau_{v_1,v_2}^{u_1,u_2}(\alpha,\beta;\omega) = \tau_{v_1}^{u_1}(\alpha;\omega_\beta^{u_2})\tau_{v_2}^{u_2}(\beta;\omega_\alpha^{v_1}).$$

Proof. We take w = initial segment of ω so that α,β are contained in $[1,2,\ldots,|w|]$. We need to verify that:

$$T(w_{\alpha,\beta}^{u_1,u_2})T(w_{\alpha,\beta}^{v_1,v_2})^{-1} = T((w_\beta^{u_2})_\alpha^{u_1})T((w_\beta^{u_2})_\alpha^{v_1})^{-1}T((w_\alpha^{v_1})_\beta^{u_2})(T(w_\alpha^{v_1})_\beta^{v_2})^{-1}.$$

But the operations $w \to w_\gamma^u$ commute for disjoint sets γ and this immediately gives (5.3). QED

The following special case of (5.3) will be of importance.

Corollary. *If $\alpha < \beta$ and α is chosen from the positions for which $\omega(n) = j$ then*

(5.4) $$\tau_j^i(\alpha \cup \beta; \omega) = \tau_j^i(\alpha; \omega)\tau_j^i(\beta; \omega).$$

Proof. We apply (5.3) with u_1, u_2 consisting only of the entry i and v_1, v_2 consisting only of the entry j. Since $\alpha < \beta$

$$\tau_j^i(\alpha; \omega_\beta^i) = \tau_j^i(\alpha; \omega).$$

Moreover by our hypothesis $\omega_\alpha^j = \omega$. QED

5.2. Notation. If $\omega \in \Omega(k)$ and $l \leq j \leq k$, we write

(5.5) $$N_j(\omega) = \{n \in \mathbf{N} \mid \omega(n) = j\}.$$

We will be considering expressions of the form

$$\int T(w_{\alpha_1, \alpha_2, \dots, \alpha_l}^{i_1, i_2, \dots, i_l})^{-1} f T(w_{\alpha_1, \alpha_2, \dots, \alpha_l}^{j_1, j_2, \dots, j_l})^{-1} g \cdots T(w_{\alpha_1, \alpha_2, \dots, \alpha_l}^{r_1, r_2, \dots, r_l})^{-1} h \, d\mu.$$

Note that these expressions satisfy the condition of admissibility of Definition 1.7 and we shall write for $\omega \in \Omega$,

$$\int T(\omega_{\alpha_1, \alpha_2, \dots, \alpha_l}^{i_1, i_2, \dots, i_l})^{-1} f T(\omega_{\alpha_1, \alpha_2, \dots, \alpha_l}^{j_1, j_2, \dots, j_l})^{-1} g \cdots T(\omega_{\alpha_1, \alpha_2, \dots, \alpha_l}^{r_1, r_2, \dots, r_l})^{-1} h \, d\mu.$$

Corollary 5.1 implies that the transformations $\tau_j^i(\alpha; \omega)$ form an IP-system provided one restricts the sets α to $N_j(\omega)$. For that to be useful we need to assume that $N_j(\omega)$ is infinite and in subsequent considerations we shall frequently assume, implicitly or explicitly, that each of the symbols $1, 2, \dots, k$ occurs infinitely often in ω. The set of such ω's, referred to as *generic*, is residual in $\Omega(k)$ and in view of 1.8 we may, anytime that we prove something for all generic points, go to a subsystem $\Sigma\mathcal{T}$ for which the property in question is valid everywhere.

The fact that our main goal, DHJ_k, is not sensitive to restriction to subsystems, in other words, that its validity for a subsystem is as good as the claim itself, permits us to apply Theorem 1.5 again and again and to assume that practically all the expressions that we deal with which can be viewed as functions from $W(k, l)$ into a compact space are uniformly continuous and admit an extension by continuity to $\Omega(k)$. This includes bounded numerical functions, bounded Hilbert-space-valued functions (for the weak topology), operator valued functions (weak topology) and measures (weak topology determined by an appropriate collection of test functions). We can "restrict to a subspace" countably many times so that even upon introduction of new functions, by procedures that in themselves do not guarantee continuity, we may assume that all the limits that we consider exist. We shall refer to this observation as "the restriction principle", meaning that in the context in question, we can achieve a simplification by assuming as we may that we have restricted the variable words to an appropriate subspace of $W(k)$.

5.3. An important application of this principle relates to the following definition.

Definition. A $W(k)$-system $(X,\mathcal{B},\mu,\{T_n^{(1)},\dots,T_n^{(k)}\})$ *is coherent* if for each l and any choice of l-tuples $(i_1,i_2,\dots,i_l),(j_1,j_2,\dots,j_l),\dots,(r_1,r_2,\dots,r_l)$, and corresponding functions $f,g,\dots,h\in L^2(X,\mathcal{B},\mu)$ the limits

$$\lim_{\alpha_1,\alpha_2,\dots,\alpha_l\to\infty}\int T(\omega^{i_1,i_2,\dots,i_l}_{\alpha_1,\alpha_2,\dots,\alpha_l})^{-1}fT(\omega^{j_1,j_2,\dots,j_l}_{\alpha_1,\alpha_2,\dots,\alpha_l})^{-1}g\cdots T(\omega^{r_1,r_2,\dots,r_l}_{\alpha_1,\alpha_2,\dots,\alpha_l})^{-1}h\,d\mu$$

exist and define a continuous function of ω.

By the restriction principle described above, we may assume in our treatment of $W(k)$-systems that they are coherent, and we may replace Theorem 2.8 in the form in which we propose to prove it, namely:

Theorem. *Let $(X,\mathcal{B},\mu,\{T_n^{(1)},\dots,T_n^{(k)}\})$ be a coherent $W(k)$-system, and let f be in $L^\infty(X,\mathcal{B},\mu)$ with $f\geq 0$ and with $\int f\,d\mu>0$. Then the set of $\omega\in\Omega(k)$ for which there exist α with*

$$\int T(\omega^1_\alpha)^{-1}fT(\omega^2_\alpha)^{-1}f\cdots T(\omega^k_\alpha)^{-1}f\,d\mu>0 \tag{5.6}$$

or, alternatively,

$$\lim_{\alpha\to\infty}\int\tau^1_1(\alpha;\omega)^{-1}f\tau^2_1(\alpha;\omega)^{-1}f\cdots\tau^k_1(\alpha;\omega)^{-1}f\,d\mu>0, \tag{5.7}$$

is dense (and clearly open) in $\Omega(k)$.

We note that τ^1_1 is the identity and appears in (5.7) just for symmetry.

6. Ω-Factorization

6.1. We fix a $W(k)$-system $\mathcal{T}=(X,\mathcal{B},\mu,\{T_n^{(1)},T_n^{(2)},\dots,T_n^{(k)}\})$.

Definition. An *Ω-factorization of* $\mathcal{T}$ consists of the following data: to each $\omega\in\Omega(k)$ we assign a sub-σ-algebra $\mathfrak{D}_\omega\subset\mathcal{B}$ such that if P_ω denotes the corresponding projection in $L^2(X,\mathcal{B},\mu)$, $P_\omega:L^2(X,\mathcal{B},\mu)\to L^2(X,\mathfrak{D}_\omega,\mu)$, the following conditions are satisfied:

(i) For each $\alpha \in \mathcal{T}$, $\omega \in \Omega$, and words u, v with $l(u) = l(v) = |\alpha|$ we have

$$\tau_v^u(\alpha;\omega)P_{\omega_\alpha^v} = P_{\omega_\alpha^u}\tau_v^u(\alpha;\omega). \tag{6.1}$$

(ii) P_ω is continuous on $\Omega(k)$.

For projection valued operators, weak and strong convergence coincide, and this is the continuity referred to in (ii).

We write $\tilde{P} = \{P_\omega\}$. We shall use the shortened notation P_ω for $P_{\mathfrak{D}_\omega}$ above, as well as $\tilde{\mu}_\omega$ for $\tilde{\mu}_{\mathfrak{D}_\omega}$ (cf. Definition 3.5). An Ω-factorization $\tilde{Q}$ is an *extension* of $\tilde{P}$ if $Q_\omega \geq P_\omega$ for all $\omega \in \Omega$.

Attached to each ω we have the corresponding factor space $Y(\omega)$. All these subsets should be thought of as subsets of $\mathcal{P}(X)$. The transformations $\tau_v^u(\alpha;\omega)$ then operate on $Y(\omega)$, and by the remark in §3.2, $\{\tau_v^u(\alpha;\omega)\mu_y : y \in Y(\omega)\}$ is the disintegration of μ with respect to $\tau_v^u(\alpha;\omega)\mathfrak{D}_\omega$. By (6.1) it follows that $\tau_v^u(\alpha;\omega)\mathfrak{D}_{\omega_\alpha^v} = \mathfrak{D}_{\omega_\alpha^u}$. Accordingly, we shall have

$$\tau_v^u(\alpha;\omega)Y(\omega_\alpha^v) = Y(\omega_\alpha^u). \tag{6.2}$$

Ω-factorizations are families of quasi-invariant factors in the context of §4:

Lemma. *Let $\omega_0 \in \Omega(\omega)$ with $\omega_0(n) = j$ infinitely often. If $\{P_\omega\}$ defines an Ω-factorization of $(X, \mathcal{B}, \mu)$, then P_{ω_0} is a quasi-invariant factor of $(X, \mathcal{B}, \mu)$ relative to the IP-system $\{\tau_j^i(\alpha;\omega_0) : \alpha \subset N_j(\omega_0)\}$.*

Proof. We have to show that

$$\tau_j^i(\alpha;\omega_0)P_{\omega_0}\tau_j^i(\alpha;\omega_0)^{-1} \to P_{\omega_0}$$

as $\alpha \to \infty$. But since $\alpha \subset N_j(\omega_0)$, $(\omega_0)_\alpha^j = \omega_0$. By (6.1)

$$\tau_j^i(\alpha;\omega_0)P_{\omega_0} = \tau_j^i(\alpha;\omega_0)P_{(\omega_0)_\alpha^j} = P_{(\omega_0)_\alpha^i}\tau_j^i(\alpha;\omega_0)$$

and, as $\alpha \to \infty$,

$$\tau_j^i(\alpha;\omega_0)P_{\omega_0}\tau_j^i(\alpha;\omega_0)^{-1} = P_{(\omega_0)_\alpha^i} \to P_{\omega_0}$$

by continuity. QED

6.2. Example. Assume $\mathcal{T}$ is a coherent $W(k)$-system. Fix i, j, $i \neq j$. Coherence implies that for all $f, g \in L^2(X, \mathcal{B}, \mu)$

$$\int fT(w_\alpha^i)T(w_\alpha^j)^{-1}g\,d\mu = \int T(w_\alpha^i)^{-1}fT(w_\alpha^j)^{-1}g\,d\mu$$

converges uniformly as $w \to \omega \in \Omega(k)$, $\alpha \to \infty$. Hence the operators $\tau_j^i(\alpha;\omega)$ converge uniformly (in the weak topology) as $\alpha \to \infty$ and define a continuous family of operators P_ω. By the remark in 5.2, if $\omega(n) = j$ infinitely often, then, restricting α to $N_j(\omega)$, we obtain an IP-system of transformations $\{\tau_j^i(\alpha;\omega)\}$ on

$L^2(X,\mathcal{B},\mu)$. This implies that P_ω is a self-adjoint projection for those ω with $\omega(n)=j$ infinitely often (and therefore, since $\tau_j^i(\alpha;\omega)=\tau_i^j(\alpha;\omega)^*$, also for ω with $\omega(n)=i$ infinitely often). As noted in 5.2, there is no loss of generality (restricting to a subspace $\Sigma W(k)$ such that every $\omega\in\Omega(k)_\Sigma$ has an i or j occurring infinitely often) in assuming that P_ω is a self-adjoint projection for all ω, and we shall show that $\{P_\omega\}$ defines an Ω-factorization.

We first show that $P_\omega(L^2(X,\mathcal{B},\mu))=L^2(X,\mathfrak{D}_\omega,\mu)$ for a σ-algebra $\mathfrak{D}_\omega$. Since $T(w_\alpha^i)T(w_\alpha^j)^{-1}\to P_\omega$ as $w\to\omega$, $\alpha\to\infty$, we see that $f\in P_\omega(L^2(X,\mathcal{B},\mu))\Leftrightarrow$

$$\|T(w_\alpha^i)^{-1}f-T(w_\alpha^j)^{-1}f\|\to 0.$$

But this condition clearly defines a lattice in $L^2(X)$, which defines $\mathfrak{D}_\omega$ as required.

We next check (i):

$$\tau_v^u(\alpha;\omega)\lim_{\beta\to\infty}\tau_j^i(\beta;\omega)=\lim_{\beta\to\infty}\tau_v^u(\alpha;\omega)\tau_j^i(\beta;\omega).$$

By Lemma 5.1,

$$\tau_{v,j}^{u,i}(\alpha,\beta;\omega)=\tau_v^u(\alpha;\omega_\beta^i)\tau_j^i(\beta;\omega_\alpha^v),$$

and since we can assume $\beta>\alpha$,

$$\tau_v^u(\alpha;\omega_\beta^i)=\tau_v^u(\alpha;\omega).$$

Thus

$$\tau_v^u(\alpha;\omega)P_{\omega_\alpha^v}=\lim_{\beta\to\infty}\tau_{v,j}^{u,i}(\alpha,\beta;\omega). \tag{6.3}$$

On the other hand, by Lemma 5.1,

$$\tau_{v,j}^{u,i}(\alpha,\beta;\omega)=\tau_j^i(\beta;\omega_\alpha^u)\tau_v^u(\alpha;\omega_\beta^j).$$

Again we have $\tau_v^u(\alpha;\omega_\beta^j)=\tau_v^u(\alpha;\omega)$. Combining this with (6.3) we obtain

$$\tau_v^u(\alpha;\omega)P_{\omega_\alpha^v}=P_{\omega_\alpha^u}\tau_v^u(\alpha;\omega).$$

Finally, the condition (ii) has been guaranteed by the uniform convergence of $\tau_j^i(\alpha;\omega)$.

6.3. We now describe a procedure, generalizing that of §6.2, for constructing examples of Ω-factorizations. We fix i,j, $i\neq j$, $1\le i,j\le k$, and we shall define the "τ_j^i-rigid extension" of a given Ω-factorization. Example 6.2 will turn out to be the τ_j^i-rigid extension of the trivial Ω-factorization.

Let $\{P_\omega\}$ define an Ω-factorization. We now consider the following function of the pair (w, β) (more precisely, of w_β), defined for $\varphi_1, \varphi_2 \in L^\infty(X, \mathcal{B}, \mu)$:

$$\Phi(\varphi_1, \varphi_2; w, \beta) = P_{\overline{w}_\beta^j}(\tau_j^i(\beta; w)\varphi_1 \cdot \varphi_2)\tau_j^i(\beta; w)\varphi_1$$

where $\overline{w} \in \Omega(k)$ is defined by

$$\overline{w}(n) = \begin{cases} w(n) & \text{for } n \leq l(w), \\ 1 & \text{for } n > l(w). \end{cases}$$

By the restriction principle (§5.2) we may assume that all these functions converge (in the weak topology of $L^2(X, \mathcal{B}, \mu)$) uniformly as $w \to \omega$, $\beta \to \infty$ (equivalently as $w_\beta \to \omega$; see 1.7 for the meaning of $w_\beta \to \omega$). We then obtain, for $\omega \in \Omega(k)$,

$$\Phi(\varphi_1, \varphi_2; \omega) = \lim_{\substack{w \to \omega \\ \beta \to \infty}} P_{\overline{w}_\beta^j}(\tau_j^i(\beta; w)\varphi_1 \cdot \varphi_2)\tau_j^i(\beta; w)\varphi_1. \tag{6.4}$$

These expressions are weakly continuous in ω and, by Corollary 1.8, there is no loss of generality assuming that they are norm continuous in $\Omega(k)$.

The limit in (6.4) exists no matter how $w \to \omega$ and $\beta \to \infty$. We can assume the convergence is such that $\tau_j^i(\beta; w) = \tau_j^i(\beta; \omega)$. We then have for the right-hand side of (6.4), using (6.1),

$$\lim \tau_j^i(\beta; \omega)\big(P_{\overline{w}_\beta^j}(\varphi_1 \cdot \tau_j^i(\beta; \omega)^{-1}\varphi_2)\varphi_1\big).$$

Assume now that $N_j(\omega)$ is infinite so that β can be restricted to be a subset of $N_j(\omega)$. We may then replace $\overline{w}_\beta^j$ by $\overline{w}$. Letting $w \to \omega$ for fixed β and using the continuity of P_ω, we obtain

$$\Phi(\varphi_1, \varphi_2; \omega) = \lim_{\beta \to \infty} \tau_j^i(\beta; \omega)\big(P_\omega(\varphi_1 \cdot \tau_j^i(\beta; \omega)^{-1}\varphi_2)\varphi_1\big). \tag{6.5}$$

Another way of describing $\Phi(\varphi_1, \varphi_2; \omega)$ is, for fixed ω, writing $\tau_\beta = \tau_j^i(\beta; \omega)$,

$$\begin{aligned}
\int \Phi(\varphi_1, \varphi_2; \omega)\psi\, d\mu &= \lim_{\beta \to \infty} \int \tau_\beta\{P_\omega(\varphi_1 \tau_\beta^{-1}\varphi_2)\varphi_1\}\psi\, d\mu \\
&= \lim_{\beta \to \infty} \int P_\omega(\varphi_1 \cdot \tau_\beta^{-1}\varphi_2)\varphi_1 \tau_\beta^{-1}\psi\, d\mu \\
&= \lim_{\beta \to \infty} \int (\varphi_1 \otimes \varphi_1)\tau_\beta^{-1} \times \tau_\beta^{-1}(\varphi_2 \otimes \psi)\, d\tilde{\mu}_\omega \\
&= \int \varphi_2 \otimes \psi H\, d\tilde{\mu}_\omega
\end{aligned}$$

where $H = H(\varphi_1, \omega) \in L^2(\tilde{\mu}_\omega)$ is a bounded quasi-invariant function, and the measure $\sigma(\varphi_1; \omega) = H\, d\tilde{\mu}_\omega$ is, as in Example 4.6, $\tau_\beta \times \tau_\beta$ quasi-invariant. The existence of the limits is assumed (without loss of generality) and the lines above just help identify them.

By 3.6 we have

$$\int \varphi_2 \otimes \psi H \, d\tilde{\mu}_\omega = \langle \tilde{H}\varphi_2, \psi \rangle$$

and we can write

$$\Phi(\varphi_1, \varphi_2; \omega) = \tilde{H}\varphi_2 \tag{6.6}$$

for a bounded quasi-invariant $H = H(\varphi_1, \omega)$. An immediate consequence of that is: if ϑ is $\mathfrak{D}_\omega$-measurable then

$$\Phi(\varphi_1, \vartheta\varphi_2; \omega) = \vartheta\Phi(\varphi_1, \varphi_2; \omega). \tag{6.7}$$

Notation. We denote by $\mathcal{L}_\omega$ the span in $L^2(X, \mathcal{B}, \mu)$ of $\{\Phi(\varphi_1, \varphi_2; \omega)\}$ where φ_1, φ_2 range over $L^\infty(X, \mathcal{B}, \mu)$, and by Q_ω the orthogonal projection onto $\mathcal{L}_\omega$.

Notice that if $f = P_\omega f$, then $\Phi(1, f; \omega) = f$, so that $Q_\omega \geq P_\omega$. Also, it follows from (6.7) that $\mathcal{L}_\omega$ is a $\mathfrak{D}_\omega$-module. Our next goal is to show that Q_ω defines an Ω-factorization extending the given P_ω. We check condition (i) of 6.1 in subsection 6.6, condition (ii) in subsection 6.5, and the fact that $\mathcal{L}_\omega = L^2(X, \mathcal{E}_\omega, \mu)$ for an appropriate σ-algebra $\mathcal{E}_\omega$ in 6.7.

6.4. Lemma. *If $\psi \in \mathcal{L}_{\omega_0}$ and $\omega \to \omega_0$ in $\Omega(k)$, then $Q_\omega\psi \to \psi$.*

Proof. It suffices to check this for $\psi = \Phi(\varphi_1, \varphi_2; \omega_0)$. We have $\Phi(\varphi_1, \varphi_2; \omega) \to \Phi(\varphi_1, \varphi_2; \omega_0)$ strongly so that

$$Q_\omega\Phi(\varphi_1, \varphi_2; \omega_0) - Q_\omega\Phi(\varphi_1, \varphi_2; \omega) \to 0.$$

But $Q_\omega\Phi(\varphi_1, \varphi_2; \omega) = \Phi(\varphi_1, \varphi_2; \omega) \to \Phi(\varphi_1, \varphi_2; \omega_0)$, hence $Q_\omega\psi \to \psi$. QED

6.5. Lemma. *For every $\psi \in L^2(X, \mathcal{B}, \mu)$ the function $\langle Q_\omega\psi, \psi\rangle$ is lower semi-continuous on $\Omega(k)$.*

Proof. Write $\psi = \psi_1 + \psi_2$ with $\psi_1 \in \mathcal{L}_{\omega_0}$ and $\psi_2 \perp \mathcal{L}_{\omega_0}$ and let $\omega \to \omega_0$. By the foregoing lemma $\langle Q_\omega\psi_1, \psi_1\rangle \to \langle Q_{\omega_0}\psi_1, \psi_1\rangle$. Also $\langle Q_\omega\psi_1, \psi_2\rangle = \langle Q_\omega\psi_2, \psi_1\rangle$, and since $Q_\omega\psi_1 \to \psi_1$ and $\psi_1 \perp \psi_2$, both $\langle Q_\omega\psi_1, \psi_2\rangle$ and $\langle Q_\omega\psi_2, \psi_1\rangle$ converge to 0. Finally $\langle Q_\omega\psi_2, \psi_2\rangle \geq 0$; hence

$$\liminf \langle Q_\omega\psi, \psi\rangle \geq \langle Q_{\omega_0}\psi, \psi\rangle = \langle \psi_1, \psi_1\rangle. \qquad \text{QED}$$

Since lower semi-continuous functions are continuous on residual sets and since we only need to check continuity for a countable collection, it follows from our restriction principle that we may assume that $\langle Q_\omega \psi, \psi \rangle$ is continuous for all $\psi \in L^2(X, \mathcal{B}, \mu)$ as a function of $\omega \in \Omega(k)$. By polarization

$$\langle Q_\omega \psi, \psi' \rangle = \tfrac{1}{2}\{\langle Q_\omega(\psi + \psi'), (\psi + \psi')\rangle - \langle Q_\omega \psi, \psi \rangle - \langle Q_\omega \psi', \psi \rangle\}$$

it follows that $Q_\omega \psi$ is weakly continuous, and as we have remarked before for self-adjoint projections this is equivalent to strong continuity. We have proved:

Proposition. *Without loss of generality we can assume that $\omega \to Q_\omega$ is strongly continuous.*

6.6. Lemma. *With the notation of Definition* 6.1(i) *we have*

$$\tau_v^u(\alpha;\omega) Q_{\omega_\alpha^v} = Q_{\omega_\alpha^u} \tau_v^u(\alpha;\omega).$$

Proof. This is equivalent to

$$\tau_v^u(\alpha;\omega)\mathcal{L}_{\omega_\alpha^v} = \mathcal{L}_{\omega_\alpha^u}.$$

We will prove

$$\tau_v^u(\alpha;\omega)\mathcal{L}_{\omega_\alpha^v} \subset \mathcal{L}_{\omega_\alpha^u} \tag{6.8}$$

and by interchanging u and v we obtain the reverse containment. We prove (6.8) by showing that

$$\tau_v^u(\alpha;\omega)\Phi(\varphi_1, \varphi_2; \omega_\alpha^v) = \Phi(\varphi_1', \varphi_2'; \omega_\alpha^u) \tag{6.9}$$

for appropriate φ_1', φ_2'. In fact the left side of (6.9) can be written as

$$\begin{aligned}
&\tau_v^u(\alpha;\omega)\lim_{\beta\to\infty} \tau_j^i(\beta;\omega_\alpha^v)\{P_{\omega_\alpha^v}(\varphi_1 \cdot \tau_i^j(\beta;\omega_\alpha^v)\varphi_2)\varphi_1\} \\
&= \lim_{\beta\to\infty} \tau_{v,j}^{u,i}(\alpha,\beta;\omega)\{P_{\omega_\alpha^v}(\varphi_1 \cdot \tau_i^j(\beta;\omega_\alpha^v)\varphi_2)\cdot\varphi_1\} \\
&= \lim_{\beta\to\infty} \tau_j^i(\beta;\omega_\alpha^u)\tau_v^u(\alpha;\omega)\{P_{\omega_\alpha^v}(\varphi_1 \cdot \tau_i^j(\beta;\omega_\alpha^v)\varphi_2)\cdot\varphi_1\} \\
&= \lim_{\beta\to\infty} \tau_j^i(\beta;\omega_\alpha^u)\{P_{\omega_\alpha^u}(\tau_v^u(\alpha;\omega)\varphi_1 \cdot \tau_v^u(\alpha;\omega)\tau_i^j(\beta;\omega_\alpha^v)\phi_2)\tau_v^u(\alpha;\omega)\varphi_1\} \\
&= \lim_{\beta\to\infty} \tau_j^i(\beta;\omega_\alpha^u)\{P_{\omega_\alpha^u}(\varphi_1' \cdot \tau_{v,i}^{u,j}(\alpha,\beta;\omega)\varphi_2)\varphi_1'\} \\
&= \lim_{\beta\to\infty} \tau_j^i(\beta;\omega_\alpha^u)\{P_{\omega_\alpha^u}(\varphi_1' \cdot \tau_i^j(\beta;\omega_\alpha^u)\varphi_2')\varphi_1'\} \\
&= \Phi(\varphi_1', \varphi_2'; \omega_\alpha^u)
\end{aligned}$$

where $\varphi_1' = \tau_v^u(\alpha;\omega)\varphi_1$, $\varphi_2' = \tau_v^u(\alpha;\omega)\varphi_2$. Here we use repeatedly Lemma 5.1 and the fact that $\alpha < \beta$. This proves the lemma. QED

6.7. Proposition 6.5 and Lemma 6.6 give two of the facts needed to show that Q_ω defines an Ω-factorization. We have remarked in 6.3 that $Q_\omega \geq P_\omega$ and that $\mathfrak{L}_\omega$ is a $\mathfrak{D}_\omega$-module, and the only remaining element is to show that $\mathfrak{L}_\omega = L^2(X,\mathcal{E}_\omega,\mu)$ for an appropriate σ-algebra $\mathcal{E}_\omega \supset \mathfrak{D}_\omega$. For this it will be enough to show that $\mathfrak{L}_\omega$ is generated by a family of bounded functions closed under multiplication.

We consider in this discussion a fixed $\omega \in \Omega(k)$ with $N_j(\omega)$ infinite. This is justified by our restriction principle. With this assumption the $\tau_j^i(\beta;\omega)$ can be taken as an IP-system (by restricting β to $N_j(\omega)$). We abbreviate $\tau_j^i(\beta;\omega)$ to τ_β. We also write $\tilde{\mu}_\omega$ for $\tilde{\mu}_{P_\omega}$.

Definition. A function in $L^2(X,\mathcal{B},\mu)$ is of *finite rank* relative to P_ω and the system $\{\tau_\beta\}$ if it belongs to a $\{\tau_\beta\}$-quasi-invariant bounded P_ω-module of finite rank.

Lemma. *Each $\Phi(\varphi_1,\varphi_2;\omega)$ can be approximated by functions of finite rank.*

Proof. This is an immediate consequence of (6.6), and Corollary 4.8. We just need to remark that the finite rank module constructed in Theorem 4.8 is generated by (fiberwise) eigenfunctions of $\tilde{H}$ (which are again in $\mathfrak{L}_\omega$ by (6.6)), and since H is bounded, the corresponding eigenfunctions are bounded. QED

Now if $\mathfrak{M}_1$ and $\mathfrak{M}_2$ are modules of finite rank spanned by bounded functions, then $\mathfrak{M}_1\mathfrak{M}_2$ is a module of finite rank in $L^2(X,\mathcal{B},\mu)$, and if we knew that this module is contained in $\mathfrak{L}_\omega$, it would follow that $\mathfrak{L}_\omega$ is spanned by an algebra of bounded functions, which in turn implies that $\mathfrak{L}_\omega = L^2(X,\mathcal{E}_\omega,\mu)$. Thus the only missing element is

Proposition. *$\mathfrak{L}_\omega$ contains all functions of finite rank (relative to P_ω and $\{\tau_\beta\}$).*

Proof. Let $\mathfrak{M}$ be a bounded $\{\tau_\beta\}$-quasi-invariant P_ω-module of finite rank. Then $\mathfrak{M}$ has a basis $\{\varphi_s\}$ such that $P_\omega\{\varphi_s\varphi_t\} = 0$ for $s \neq t$, $P_\omega\{\varphi_s^2\} = 1$ for $s \leq r(x)$, and $P_\omega\{\varphi_s^2\} = 0$ for $s > r(x)$, where $r(x) \leq r$ is $\mathfrak{D}_\omega$-measurable. We shall prove that

$$\varphi_s = \sum_t \Phi(\varphi_t,\varphi_s;\omega) \tag{6.10}$$

and, since $\mathfrak{L}_\omega$ is a $\mathfrak{D}_\omega$-module, this implies $\mathfrak{M} \subset \mathfrak{L}_\omega$.

To prove (6.10) write the right-hand side as

$$\sum_t \lim_{\beta\to\infty} \tau_\beta\{P_\omega(\varphi_t\tau_\beta^{-1}\varphi_s)\cdot\varphi_t\}. \tag{6.11}$$

By the quasi-invariance of $\mathfrak{M}$, we can write

$$\tau_\beta^{-1}\varphi_s \sim \sum_{t'} u^\beta_{st'}\varphi_{t'}$$

in norm, with $u^\beta_{st'}$ measurable with respect to $\mathfrak{D}_\omega$. Substituting in (6.11) we obtain as an approximation

$$\begin{aligned}\lim_{\beta\to\infty}\tau_\beta\Big\{\sum_{t,t'}P_\omega(\varphi_t u^\beta_{st'}\varphi_{t'})\cdot\varphi_t\Big\} &= \lim_{\beta\to\infty}\tau_\beta\Big\{\sum_{t,t'}u^\beta_{st'}P_\omega(\varphi_t\varphi_{t'})\cdot\varphi_t\Big\}\\ &= \lim_{\beta\to\infty}\tau_\beta\sum u^\beta_{st}\varphi_t\\ &= \varphi_s\end{aligned}$$

where we see, *a posteriori*, that the limits in question exist. QED

Definition. The Ω-factorization given by $\{Q_\omega\}$ is the foregoing discussion will be called the *τ^i_j-rigid extension* of $\{P_\omega\}$.

6.8. We state the following theorem without proof, just to complete somewhat the picture. Lemma 6.10, which is essential later on and which we do prove in detail, could be obtained as a consequence thereof.

Theorem. *Let $\mathfrak{M}$ be a $\{\tau^i_j(\beta;\omega_0)\}$-quasi-invariant P_{ω_0}-module of finite rank. Then there exists a neighborhood V of ω_0 and a continuous family $\{\mathfrak{M}_\omega : \omega\in V\}$ of $\{\tau^i_j(\beta;\omega)\}$-quasi-invariant P_ω-modules, such that $\mathfrak{M}_{\omega_0}=\mathfrak{M}$.*

6.9. We state a simple lemma regarding approximately orthonormal systems in finite-dimensional subspaces of Hilbert space. This lemma will be applied to the finite-dimensional subspaces $\mathfrak{M}_y$ induced by $\mathfrak{M}$ in $L^2(\mu_y)$. The proof is straightforward.

Lemma. *For $\epsilon>0$, $r\in\mathbf{N}$, there exists $\delta=\delta(\epsilon,r)>0$ so that if $e_1,e_2,\dots,e_r$ span a subspace V of a Hilbert space $\mathcal{H}$, and $u\in\mathcal{H}$, with $\|u\|\le 1$, then if*

(i) *for each $1\le i,j\le r$, $|\langle e_i,e_j\rangle-\delta_{ij}|<\delta(\epsilon,r)$, we will have*

$$\|u\|^2-\sum\langle u,e_i\rangle^2>-\epsilon,$$

and if in addition

(ii) $\|u\|^2-\sum\langle u,e_i\rangle^2<\delta(\epsilon,r)$, *we will have*

$$\|u-P_V u\|<\epsilon$$

where P_V denotes the orthogonal projection from $\mathcal{H}$ to V.

6.10. Returning to the context of 6.8, we write, as before, $\tau_\beta=\tau^k_1(\beta;\omega)$ (we take $i=1$ and $j=k$ so that the letters i and j can be freed for use as variables).

Let $\mathfrak{M}$ be a $\{\tau_\beta\}$-quasi-invariant P_{ω_0}-module of finite rank, with basis $\{\varphi_s\}$. For each $\omega \in \Omega$ let $\mathfrak{M}(\omega)$ be the P_ω-module spanned by $\{\varphi_1, \ldots, \varphi_r\}$. Note that the φ_i are approximately orthonormal at ω's close to ω_0 since $P_\omega(\varphi_i \bar{\varphi}_j) \to P_{\omega_0}(\varphi_i \bar{\varphi}_j)$ in $L^2(X, \mathcal{B}, \mu)$ as $\omega \to \omega_0$. This, as in all global approximations, means that but for a small set of $y \in Y(\omega)$

$$|\langle \varphi_i, \varphi_j \rangle_{L^2(\mu_y)} - \delta_{ij}|, \qquad 1 \le i, j \le r \tag{6.13}$$

are small.

We can use Lemma 6.9 to show that if $g \in \mathfrak{M} = \mathfrak{M}(\omega_0)$ then, for α far out,

$$\tau_1^k(\alpha; \omega)^{-1} g$$

is close to the module $\mathfrak{M}(\omega)$, provided ω is close to ω_0. We make this precise in the next lemma.

Lemma. *Given $\epsilon > 0$, $\exists N', N''$ so that if $\omega(n) = \omega_0(n)$ for $n \le N'$ and if $\alpha \subset \{N'', N'' + 1, N'' + 2, \ldots\}$ then*

$$\operatorname{dist}(\tau_1^k(\alpha; \omega)^{-1} g, \mathfrak{M}(\omega)) < \epsilon.$$

Proof. Recall the definition of the τ_1^k-rigid extension of $\tilde{P}$. The expressions

$$P_{\overline{w}_\alpha^k}(\tau_1^k(\alpha; w)\varphi \cdot \psi)\tau_1^k(\alpha; w)\varphi$$

are uniformly close (in the weak topology) to their limit for w close to ω and α sufficiently far out. So $\exists N''$ such that if $\alpha \subset \{N'', N'' + 1, \ldots\}$

$$\left| \int P_{\omega_\alpha^k}(\tau_1^k(\alpha; \omega)\varphi \cdot \psi)\tau_1^k(\alpha; \omega)\varphi \cdot \psi \, d\mu - \int \Phi(\varphi, \psi; \omega) \cdot \psi \, d\mu \right| < \epsilon'$$

for a preassigned ϵ'. The first integral can be rewritten (cf. (6.1)):

$$\int \tau_1^k(\alpha; \omega) P_\omega(\varphi \cdot \tau_1^k(\alpha; \omega)^{-1}\psi)\tau_1^k(\alpha; \omega)\varphi \cdot \psi \, d\mu = \int P_\omega(\varphi \cdot \tau_1^k(\alpha; \omega)^{-1}\psi)^2 \, d\mu$$

$$= \int \langle \tau_1^k(\alpha; \omega)^{-1}\psi, \varphi \rangle_{L^2(\mu_y)}^2 \, d\nu(y),$$

and the second integral is by definition $\Lambda(\varphi, \psi; \omega)$. Apply this to $\psi = g$, $\varphi = \varphi_i$, and sum:

$$\left| \int \sum_{i=1}^r \langle \tau_1^k(\alpha; \omega)^{-1} g, \varphi_i \rangle_{L^2(\mu_y)}^2 \, d\nu(y) - \sum \Lambda(\varphi_i, g; \omega) \right| < r\epsilon'. \tag{6.14}$$

Here we have adapted the notation $\Lambda(\;,\;)$ of §4 to the case of a factor that depends on ω.

Now apply Lemma 6.9 using the fact that for ω close to ω_0, the expressions in

(6.13) are small. In particular, we can find N' so that if $\omega(n) = \omega_0(n)$ for $n \leq N'$,

$$|\langle \varphi_i, \varphi_j \rangle_{L^2(\mu_y)} - \delta_{ij}| < \delta(\epsilon'', \Lambda)$$

where ϵ'' will be specified, for $y \in Y(\omega)$ outside of a set of measure $< \epsilon''$.

By Lemma 6.9 this gives

$$\sum \langle \tau_1^k(\alpha;\omega)^{-1} g, \varphi_i \rangle^2_{L^2(\mu_y)} \leq \| \tau_1^k(\alpha;\omega)^{-1} g \|^2_{L^2(\mu_y)} + \epsilon'' \tag{6.15}$$

outside of a set of measure $< \epsilon''$.

Now $\Lambda(\varphi_i, g; \omega)$ is a continuous function of ω and so suppose that N' is chosen so that if $\omega(n) = \omega_0(n)$ for $n \leq N'$ we have

$$\left| \sum \Lambda(\varphi_i, g; \omega) - \sum \Lambda(\varphi_i, g; \omega_0) \right| < \epsilon''', \tag{6.16}$$

the latter also to be specified.

We now have

$$\int \sum_{i=1}^{r} \langle \tau_1^k(\alpha;\omega)^{-1} g, \varphi_i \rangle^2_{L^2(\mu_y)} \, d\nu(y) \geq \sum_{i=1}^{r} \Lambda(\varphi_i, g; \omega_0) - r\epsilon' - \epsilon'''.$$

By the corollary to Theorem 4.5

$$\sum \Lambda(\varphi_i, g; \omega_0) = \|g\|^2 = \|\tau_1^k(\alpha;\omega)^{-1} g\|^2$$

whence

$$\begin{aligned} &\int \sum_{i=1}^{r} \langle \tau_i^k(\alpha;\omega)^{-1} g, \varphi_i \rangle^2_{L^2(\mu_y)} \, d\nu(y) \\ &\quad \geq \|\tau_1^k(\alpha;\omega)^{-1} g\|^2 - r\epsilon' - \epsilon''' \\ &\quad = \int \|\tau_1^k(\alpha;\omega)^{-1} g\|^2_{L^2(\mu_y)} \, d\nu(y) - r\epsilon' - \epsilon'''. \end{aligned} \tag{6.17}$$

Let $\eta > 0$. If we had

$$\sum_{i=1}^{r} \langle \tau_1^k(\alpha;\omega)^{-1} g, \varphi_i \rangle^2_{L^2(\mu_y)} \leq \|\tau_1^k(\alpha;\omega)^{-1} g\|^2_{L^2(\mu_y)} - \eta$$

for a set B of a measure $\geq \eta$, then

$$\int_B \sum \langle \tau_1^k(\alpha;\omega)^{-1} g, \varphi_i \rangle^2_{L^2(\mu_y)} \, d\nu(y) \leq \int_B \|\tau_1^k(\alpha;\omega)^{-1} g\|^2_{L^2(\mu_y)} \, d\nu(y) - \eta^2.$$

Using (6.15) and bearing in mind that for all y, the left-hand side of (6.15) is bounded by $r\|g\|^2_\infty$, we have

$$\int \sum \langle \tau_1^k(\alpha;\omega)^{-1}g, \varphi_i\rangle^2_{L^2(\mu_y)}\, d\nu(y)$$

$$\leq \int \|\tau_1^k(\alpha;\omega)^{-1}g\|^2_{L^2(\mu_y)}\, d\nu(y) + \epsilon''(1 + r\|g\|^2_\infty) - \eta^2.$$

This will contradict (6.17) if

$$(6.18) \qquad r\epsilon' + \epsilon''' + \epsilon''(1 + r\|g\|^2_\infty) < \eta^2.$$

Choosing for a prescribed $\eta > 0$, $\epsilon', \epsilon'', \epsilon'''$ so that (6.18) holds, we find that we must have

$$\sum \langle \tau_1^k(\alpha;\omega)^{-1}g, \varphi_i\rangle^2_{L^2(\mu_y)} > \|\tau_1^k(\alpha;\omega)^{-1}g\|^2_{L^2(\mu_y)} - \eta$$

outside a set of y of measure $< \eta$. If now $\eta < \delta(\epsilon_1, r)$ with $\epsilon'' < \epsilon_1$ we will have by the second part of Lemma 6.9, and for non-exceptional y,

$$\|\tau_1^k(\alpha;\omega)^{-1}g - P_{\mathfrak{M}(\omega)_y}\tau_1^k(\alpha;\omega)^{-1}g\|^2_{L^2(\mu_y)} < \epsilon_1^2.$$

This gives

$$\|\tau_1^k(\alpha;\omega)^{-1}g - P_{\mathfrak{M}(\omega)_y}\tau_1^k(\alpha;\omega)^{-1}g\|^2 < \epsilon_1^2 + \eta\|g\|^2_\infty.$$

Finally, with $\sqrt{\epsilon_1^2 + \eta\|g\|^2_\infty} \leq \epsilon$ we obtain the assertion of the lemma. QED

7. Outline of the proof

7.1. By now we have completed the background material needed for the proof of our main result, and before launching into the details we would like to give an overview of the argument. The first part of the proof of our density version of the Hales–Jewett theorem has already been carried out in §2 where we have shown that Theorem A of the introduction is equivalent to Theorem 2.8, which was reformulated in §5.3 as

Theorem. *Let $(X, \mathcal{B}, \mu, \{T_n^{(1)}, \ldots, T_n^{(k)}\})$ be a coherent $W(k)$-system, and let f be in $L^\infty(X, \mathcal{B}, \mu)$ with $f \geq 0$ and with $\int f\, d\mu > 0$. Then the set of $\omega \in \Omega(k)$ for which*

$$(7.1) \qquad \int T(\omega_\alpha^1)^{-1} f\, T(\omega_\alpha^2)^{-1} f \cdots T(\omega_\alpha^k)^{-1} f\, d\mu > 0$$

is dense (and clearly open) in $\Omega(k)$.

7.2. For the moment consider a fixed $\omega \in \Omega(k)$. An Ω-factorization $\tilde{Q}$ gives us at ω data consisting of a σ-algebra $\mathcal{E}_\omega$ and the projection Q_ω onto $L^2(X, \mathcal{E}_\omega, \mu)$. The first step in proving Theorem 7.1 is to show that we may restrict the function f appearing in the theorem to be measurable with respect to $\mathcal{E}_\omega$ for a particular Ω-factorization $\tilde{Q} = \tilde{Q}^{(k)}$ which will be defined in §8. This factor will be obtained by a series of rigid extensions from the trivial (one point) factor. The reduction to this case is accomplished by showing more generally that

$$\lim_{\alpha\to\infty} \int T(\omega_\alpha^1)^{-1} f_1 T(\omega_\alpha^2)^{-1} f_2 \cdots T(\omega_\alpha^k)^{-1} f_k \, d\mu$$
$$= \lim_{\alpha\to\infty} \int T(\omega_\alpha^1)^{-1} Q_\omega f_1 T(\omega_\alpha^2)^{-1} Q_\omega f_2 \cdots T(\omega_\alpha^k)^{-1} Q_\omega f_k \, d\mu, \tag{7.2}$$

thus we can replace f in (7.1) by $Q_\omega f$, which satisfies the same conditions but is measurable with respect to $\mathcal{E}_\omega$.

7.3. The assertion to be proved,

$$\lim_{\alpha\to\infty} \int T(\omega_\alpha^1)^{-1} f T(\omega_\alpha^2)^{-1} f \cdots T(\omega_\alpha^k)^{-1} f \, d\mu > 0 \tag{7.3}$$

for non-negative, Q_ω-measurable functions, is equivalent to the corresponding statement for indicator functions $f = 1_A$ for $A \in \mathcal{E}_\omega$. In this form (7.3) says that for w close to ω and α far out, the transformations $T(w_\alpha^1), T(w_\alpha^2), \ldots, T(w_\alpha^k)$ "coalesce" in the sense that for a substantial set of points x,

$$T(w_\alpha^1)x \in A, T(w_\alpha^2)x \in A, \ldots, T(w_\alpha^k)x \in A. \tag{7.4}$$

Now $\tilde{Q}^{(k)}$ is obtained by a series of rigid extensions corresponding to the different $r_j^i(\alpha;\omega)$ systems, and starting from the trivial factor. Our strategy will be to show that this type of property passes from a factor to an extension when the extension is rigid for some r_j^i. This will be carried out in §9 and §10.

The notion of coalescence will have to be strengthened, however, in order to be able to lift the property up the ladder of rigid extensions. If we think of (7.4) as asserting that for a combinatorial line of words w the corresponding $T(w)$ coalesce, then the strengthening we have in mind asserts that the same is true for an arbitrarily high-dimensional combinatorial subspace of words. The reason we need the stronger version is that in the inductive procedure described, the line will be chosen to satisfy certain conditions among which will be the monochromaticity of the words for a certain coloring. Having a large subspace at our disposal will enable us to obtain the line in question by applying the classical Hales-Jewett theorem.

7.4. The core of the argument is the induction step taking us from one Ω-factorization $\tilde{P}$ having the desired property to a rigid extension $\tilde{Q}$.

Let P_ω and $\mathfrak{D}_\omega$ be the data associated with $\tilde{P}$ and Q_ω, $\mathcal{E}_\omega$ the data associated with $\tilde{Q}$. To highlight the essential features of the argument, we will present a somewhat simplified picture and we shall ignore here the fact that the σ-algebra $\mathfrak{D}_\omega$ is only asymptotically invariant under $\tau_j^i(\alpha;\omega)$. If we denote by $Y(\omega)$ the factor space corresponding to $\mathfrak{D}_\omega$, the fibers $y \in Y(\omega)$ are actually carried under $\tau_j^i(\alpha;\omega)$ to fibers of $Y(\omega_j^i)$ (see §6.1) but it will be useful to think of the τ_j^i as acting on the space $Y(\omega)$ itself. We also illustrate the argument by showing how to obtain a one-dimensional coalescence property from a multidimensional coalescence property of $\tilde{P}$.

Write $x' = T(w_\alpha^1)x$. Then (7.4) is the same as

$$x' \in A, T(w_\alpha^2)T(w_\alpha^1)^{-1}x' \in A, \dots, T(w_\alpha^k)T(w_\alpha^1)^{-1}x' \in A$$

or, by (5.1),

$$x' = \tau_1^1(\alpha;\omega)x' \in A, \tau_1^2(\alpha;\omega)x' \in A, \dots, \tau_1^k(\alpha;\omega)x' \in A. \tag{7.5}$$

If π is the projection $\pi: X \mapsto Y(\omega)$, let $B = \pi(A)$. Thinking (imprecisely) of the $\tau_j^i(\alpha;\omega)$ as acting on $Y(\omega)$ and letting $y = \pi(x')$, then (7.5) will require that $\tau_1^i(\alpha;\omega)y \in B$ for $i = 1, \dots, k$. Now the induction hypothesis on $\tilde{P}$ enables us to find points y, and α far out, so that this is true. But let us use the stronger hypothesis regarding $\tilde{P}$. We can then arrange that for large l there will be $\alpha_1, \dots, \alpha_l \subset N_1(\omega)$ and points $y \in Y$ with

$$\tau_{1,1,\dots,1}^{i_1,i_2,\dots,i_l}(\alpha_1,\alpha_2,\dots,\alpha_l;\omega)y \in B \tag{7.6}$$

for all l-tuples of letters $i_1, i_2, \dots, i_l$ from $\{1,2,\dots,k\}$.

Then each set $\tau_{1,1,\dots,1}^{i_1,i_2,\dots,i_l}(\alpha_1,\alpha_2,\dots,\alpha_l;\omega)^{-1}A$ meets the fiber $\pi^{-1}(y)$. What is needed to obtain the point x' in question is to find a combinatorial line L among these l-tuples so that

$$\mu_y\left(\bigcap_{(i_1,i_2,\dots,i_l)\in L} \tau_{1,1,\dots,1}^{i_1,i_2,\dots,i_l}(\alpha_1,\alpha_2,\dots,\alpha_l;\omega)^{-1}A\right) > 0. \tag{7.7}$$

Actually we need this bounded away from 0 for a set of y of measure bounded away from 0. This will give us enough points x' satisfying (7.5) where now α is some union of atoms $\alpha_1, \alpha_2, \dots, \alpha_l$, and where, strictly speaking, ω is a small perturbation of the original ω.

Now we can assume that all the sets $\tau_{1,1,\dots,1}^{i_1,i_2,\dots,i_l}(\alpha_1,\alpha_2,\dots,\alpha_l;\omega)^{-1}A$, which by (7.7) meet $\pi^{-1}(y)$ non-trivially, actually have μ_y-measure larger than some fixed positive number. This is achieved by replacing B by a slightly smaller set. Notice that with this assumption the validity of (7.7) for some combinatorial line L would be a consequence of DHJ_k (Proposition 2.1, statement b), the sets in question being indexed by long words. Since DHJ_k is what we are trying to prove, it appears that we have gained nothing from the decomposition.

But now the rigidity of $\tilde{Q}$ relative to $\tilde{P}$ will play a role. Let us assume (without loss of generality) that $\tilde{Q}$ is a τ_1^k-rigid extension of $\tilde{P}$. Suppose for the moment that τ_1^k acted as the identity map in sets of $\mathcal{E}_\omega$ with respect to the measure μ_y—an extreme version of rigidity. This will mean, using Lemma 5.1, that the sets

$$\tau_{1,\ldots,1,\ldots,1}^{i_1,\ldots,k,\ldots,i_l}(\alpha_1, \alpha_2, \ldots, \alpha_l; \omega)^{-1}A \quad \text{and} \quad \tau_{1,\ldots,1,\ldots,1}^{i_1,\ldots,1,\ldots,i_l}(\alpha_1, \alpha_2, \ldots, \alpha_l; \omega)^{-1}A$$

are almost everywhere equal with respect to the measure μ_y. In this case the sets in question can be indexed by l-tuples from $\{1,2,\ldots,k-1\}$ since k can be replaced by 1. If we now assume inductively that DHJ_{k-1} is valid, we could then complete the proof.

7.5. Of course rigidity does not imply this invariance—even asymptotically—for the sets $A \subset \mathcal{E}_\omega$. One can think of typical rigidity as corresponding to a situation where the fibers $\pi^{-1}(y)$ are spheres and are rotated by the transformations $\tau_1^k(\alpha;\omega)$. What rigidity does imply is that the functions $\tau_1^k(\alpha;\omega)^{-1}1_A$ are asymptotically close to a finite rank module over $\mathfrak{D}_\omega$. We illustrate how this is used in the following simplified version of an argument presented in §9. Choose a long sequence $\alpha_1 < \alpha_2 < \cdots < \alpha_p$ far out. In $L^2(X,\mathcal{B},\mu_y)$ the functions $\tau_1^k(\alpha_h \cup \cdots \cup \alpha_p;\omega)^{-1}1_A$ are all close to some finite set of functions, and so two of them are close, say

$$\tau_1^k(\alpha_h \cup \cdots \cup \alpha_p;\omega)^{-1}1_A \quad \text{and} \quad \tau_1^k(\alpha_{h'} \cup \cdots \cup \alpha_p;\omega)^{-1}1_A \tag{7.8}$$

with $h' < h$. Now by (5.4) and the fact that $\alpha_{h'} \cup \cdots \cup \alpha_{h-1} < \alpha_h \cup \cdots \cup \alpha_p$ we have, provided the α are chosen from $N_1(\omega)$,

$$\tau_1^k(\alpha_{h'} \cup \cdots \cup \alpha_p;\omega) = \tau_1^k(\alpha_{h'} \cup \cdots \cup \alpha_{h-1};\omega)\tau_1^k(\alpha_h \cup \cdots \cup \alpha_p;\omega).$$

(7.8) then says that

$$\tau_1^k(\alpha_h \cup \cdots \cup \alpha_p;\omega)^{-1}A \quad \text{and}$$
$$\tau_1^k(\alpha_h \cup \cdots \cup \alpha_p;\omega)\tau_1^k(\alpha_{h'} \cup \cdots \cup \alpha_{h-1};\omega)^{-1}A$$

are close at μ_y, which means that A and $\tau_1^k(\alpha_{h'} \cup \cdots \cup \alpha_{h-1};\omega)^{-1}A$ are close at $\mu_{\bar{y}}$ where $\bar{y} = \tau_1^k(\alpha_h \cup \cdots \cup \alpha_p;\omega)y$. What we obtain is an alternate point $\bar{y}$ and an α for which A and $\tau_1^k(\alpha;\omega)^{-1}A$ are interchangeable.

Now this is not strong enough because we need a large array of possibilities where k can be replaced by 1. To achieve this we will use, in addition to the pigeonhole argument just given, also an application of the multidimensional Hales-Jewett (coloring) theorem. This will be done in detail in §9. The upshot of the combinatorial argument of §9 will be that for a τ_1^k-rigid extension, we can find long sequences $\alpha_1 < \alpha_2 < \cdots < \alpha_l$ so that, if α is composed of these, $\tau_1^k(\alpha;\omega)$

acts as close to the identity as is needed on sets $A \in \mathcal{E}_\omega$ (relative to μ_y) and so, in effect, the sets

$$\tau_{1,\ldots,1}^{i_1,\ldots,i_l}(\alpha_1,\ldots,\alpha_l;\omega)^{-1}A$$

can be treated as if they were indexed by words in $k-1$ letters.

The final section of the paper makes use of this possibility to complete the proof as outlined in 7.4. We should bear in mind that the present section is intended to clarify the broad structure of the proof. In carrying out the details we will have to deal with the fact that the $\tau_j^i(\alpha;\omega)$ leave $\mathfrak{D}_\omega$ invariant only asymptotically and that the rigidity properties are only valid asymptotically. In particular $\tau_1^k(\alpha;\omega)$ does not act on $Y(\omega)$, but rather maps $Y(\omega)$ to $Y(\omega_\alpha^k)$. We ultimately obtain the information desired not for ω but for a neighboring point. Since, however, the limit in (7.2) is a continuous function of ω, this will not invalidate the argument.

8. Reduction of the main theorem to the rigid case

8.1. We are now ready to embark on the proof of Theorem 5.3:

Theorem. *Let $\mathcal{T} = (X,\mathcal{B},\mu,\{T_n^{(1)},\ldots,T_n^{(k)}\})$ be a coherent $W(k)$-system, and let f be in $L^\infty(X,\mathcal{B},\mu)$ with $f \geq 0$ and with $\int f\,d\mu > 0$. Then the set of $\omega \in \Omega(k)$ for which*

$$\int T(\omega_\alpha^1)^{-1} f T(\omega_\alpha^2)^{-1} f \cdots T(\omega_\alpha^k)^{-1} f\,d\mu > 0$$

is dense (and clearly open) in $\Omega(k)$.

In the present section we expand on 7.2. Given a $W(k)$-system $\mathcal{T}$, we define a sequence of Ω-factorizations of our system.

If $\tilde{P}$ is any Ω-factorization we shall denote by $\tilde{P}_{ij}$ the τ_j^i-rigid extension of $\tilde{P}$. (The definite article subsumes an earlier use of our standard assumption, i.e., a possible selection of a subspace of $W(k)$ to guarantee the existence of the relevant limits.)

Set $\tilde{Q}^{(1)}$ to be the trivial 1-point factor. Set $\tilde{Q}^{(2)} = \tilde{Q}^{(1)}_{(1,2)}$, $\tilde{Q}^{(3)} = (\tilde{Q}^{(2)}_{(1,3)})_{(2,3)}$, and in general, for $j = 2,3,\ldots,k$,

$$\tilde{Q}^{(j)} = (\cdots((\tilde{Q}^{(j-1)}_{(1,j)})_{(2,j)})\cdots)_{(j-1,j)}. \tag{8.1}$$

8.2. Lemma. *For $1 \leq j \leq k$, for $f_1, f_2, \ldots, f_j \in L^\infty(X,\mathcal{B},\mu)$ and for a generic ω*

$$\begin{aligned}&\lim_{\alpha\to\infty}\int T(\omega_\alpha^1)^{-1} f_1 \cdots T(\omega_\alpha^j)^{-1} f_j\,d\mu \\ &= \lim_{\alpha\to\infty}\int T(\omega_\alpha^1)^{-1} Q_\omega^{(j)} f_1 \cdots T(\omega_\alpha^j)^{-1} Q_\omega^{(j)} f_j\,d\mu.\end{aligned} \tag{8.2}$$

Proof. Induction on j. For $j = 1$ the result is clear. Assume it is true for $j-1$. In view of the identity

$$\prod_{m=1}^{j} a_m - \prod_{m=1}^{j} b_m = \sum_{i=1}^{j}\left(\prod_{m=1}^{i-1} a_m\right)(a_i - b_i)\left(\prod_{m=i+1}^{j} b_m\right)$$

it will suffice to prove (8.2) under the hypothesis that for some i, $Q_\omega^{(j)} f_i = 0$. We consider two cases.

I. Assume $i \neq j$. We use VDC (cf. §4.3) to show in this case that

$$\lim_{\alpha\to\infty} \tau_j^1(\alpha;\omega)^{-1} f_1 \cdots \tau_j^{j-1}(\alpha;\omega)^{-1} f_{j-1} = 0 \tag{8.3}$$

in the weak topology. (The limit exists by coherence.) Since ω is generic, the set $N_j(\omega) = \{n : \omega(n) = j\}$ is infinite and we can restrict α to this set. Then by Lemma 5.1, $\tau_j^i(\alpha;\omega)$ is an IP-system. By VDC (4.3), (8.3) will follow once it is shown that

$$\lim_{\alpha\to\infty}\lim_{\beta\to\infty}\int \tau_j^1(\beta;\omega)^{-1}\{f_1\tau_j^1(\alpha;\omega)^{-1}f_1\}\cdots$$
$$\tau_j^{j-1}(\beta;\omega)^{-1}\{f_{j-1}\tau_j^{j-1}(\alpha;\omega)^{-1}f_{j-1}\}\,d\mu = 0.$$

Rewrite the foregoing expression as

$$\lim_{\alpha\to\infty}\lim_{\beta\to\infty}\int T(\omega_\beta^1)^{-1}\{f_1\tau_j^1(\alpha;\omega)^{-1}f_1\}\cdots T(\omega_\beta^{j-1})^{-1}\{f_{j-1}\tau_j^{j-1}(\alpha;\omega)^{-1}f_{j-1}\}\,d\mu$$

and we obtain by the induction hypothesis that this equals

$$\lim_{\alpha\to\infty}\lim_{\beta\to\infty}\int T(\omega_\beta^1)^{-1}Q_\omega^{(j-1)}\{f_1\tau_j^1(\alpha;\omega)^{-1}f_1\}\cdots$$
$$T(\omega_\beta^{j-1})^{-1}Q_\omega^{(j-1)}\{f_{j-1}\tau_j^{j-1}(\alpha;\omega)^{-1}f_{j-1}\}\,d\mu.$$

This will vanish if any factor of the integrand converges to 0 in $L^2(X,\mathcal{B},\mu)$. But

$$\begin{aligned}&\|Q_\omega^{(j-1)}\{f_i\tau_j^i(\alpha;\omega)^{-1}f_i\}\|^2\\ &= \int (f_i\tau_j^i(\alpha;\omega)^{-1}f_i)Q_\omega^{(j-1)}(f_i\tau_j^i(\alpha;\omega)^{-1}f_i)\,d\mu \to \int f_i\Phi(f_i,f_i;\omega)\,d\mu\end{aligned} \tag{8.4}$$

and this will vanish if $f_i \perp \Phi_j^i(f_i,f_i;\omega)$. We have used the notation $\Phi(\ )$ occurring in the definition of the τ_j^i-rigid extension of $\tilde{Q}^{(j-1)}$ (see (6.5)). If $Q_\omega^{(j)}f_i = 0$, then in particular $(Q_{(i,j)}^{(j-1)})_\omega f_i = 0$ (since $\tilde{Q}_{(i,j)}^{(j-1)} \leq \tilde{Q}^{(j)}$) and indeed f_i is orthogonal to the τ_j^i-rigid extension of $\tilde{Q}^{(j-1)}$ at ω. Hence the limit in (8.4) is 0.

II. Assume $Q^{(j)}_\omega f_j = 0$. We already know from the case $i < j$ that

$$\lim_{\alpha\to\infty}\int T(\omega^1_\alpha)^{-1}f_1 \cdots T(\omega^j_\alpha)^{-1}f_j\,d\mu$$

$$(8.5)\qquad = \lim_{\alpha\to\infty}\int T(\omega^1_\alpha)^{-1}(Q^{(j)}_\omega f_1) \cdots T(\omega^{j-1}_\alpha)^{-1}(Q^{(j)}_\omega f_{j-1})T(\omega^j_\alpha)^{-1}f_j\,d\mu$$

$$= \lim_{\alpha\to\infty}\int (\tau^j_1(\alpha;\omega)Q^{(j)}_\omega f_1) \cdots (\tau^j_{j-1}(\alpha;\omega)Q^{(j)}_\omega f_{j-1})f_j\,d\mu.$$

By the (strong) continuity of the operators $Q^{(j)}_\omega$ in ω we can replace $\tau^j_i(\alpha;\omega)Q^{(j)}_\omega f_i$ by

$$\tau^j_i(\alpha;\omega)Q^{(j)}_{\omega^i_\alpha} f_i = Q^{(j)}_{\omega^j_\alpha}\tau^j_i(\alpha;\omega)f_i.$$

We obtain for the last line of (8.5)

$$\lim_{\alpha\to\infty}\int (Q^{(j)}_{\omega^j_\alpha}\tau^j_1(\alpha;\omega)f_1) \cdots (Q^{(j)}_{\omega^j_\alpha}\tau^j_{j-1}(\alpha;\omega)f_{j-1})f_j\,d\mu.$$

Now the product of functions in the image of projections $Q^{(j)}_{\omega^j_\alpha}$ is in this image and so we can replace f_j by $Q^{(j)}_{\omega^j_\alpha}f_j$. This gives

$$\lim_{\alpha\to\infty}\int T(\omega^1_\alpha)^{-1}f_1 \cdots T(\omega^j_\alpha)^{-1}f_j\,d\mu$$

$$= \lim_{\alpha\to\infty}\int (Q^{(j)}_{\omega^j_\alpha}\tau^j_1(\alpha;\omega)f_1) \cdots (Q^{(j)}_{\omega^j_\alpha}\tau^j_{j-1}(\alpha;\omega)f_{j-1})Q^{(j)}_{\omega^j_\alpha}f_j\,d\mu$$

$$= \lim_{\alpha\to\infty}\int (\tau^j_1(\alpha;\omega)Q^{(j)}_{\omega^1_\alpha}f_1) \cdots (\tau^j_{j-1}(\alpha;\omega)Q^{(j)}_{\omega^{j-1}_\alpha}f_{j-1})Q^{(j)}_{\omega^j_\alpha}f_j\,d\mu$$

$$= \lim_{\alpha\to\infty}\int (\tau^j_1(\alpha;\omega)Q^{(j)}_\omega f_1) \cdots (\tau^j_{j-1}(\alpha;\omega)Q^{(j)}_\omega f_{j-1})Q^{(j)}_\omega f_j\,d\mu$$

by continuity of $Q^{(j)}_\omega$. Multiplying the integrand (formally) by $T(\omega^j_\alpha)^{-1}$ we obtain the desired result. QED

9. A key property of rigid extensions

9.1. It now follows that in proving Theorem 8.1 for a generic point ω we can assume that $f = Q^{(k)}_\omega f$. To prove the resulting statement we formulate a stronger statement which we prove for each level of the "rigid-extension tower" which terminates in $\tilde{Q}^{(k)}$.

Definition. An Ω-factorization $\tilde{P}$ of a coherent system $(X,\mathcal{B},\mu,\{T_n^{(1)},\ldots,T_n^{(k)}\})$ *is said to be SZ* if for every generic $\omega\in\Omega$ and for every $l\in\mathbf{N}$ and every $f\in L^\infty(X,\mathcal{B},\mu)$ with $\int f\,d\mu>0$ and such that $f=P_\omega f$, we have

$$\text{(9.1)}\qquad \lim_{\alpha_1,\alpha_2,\ldots,\alpha_l\to\infty}\int\prod_{i_1,\ldots,i_l=1}^{k}T(\omega_{\alpha_1,\alpha_2,\ldots,\alpha_l}^{i_1,i_2,\ldots,i_l})^{-1}f\,d\mu>0.$$

The trivial Ω-factor is SZ and so our theorem will follow if we show that the SZ property is preserved under rigid extensions. So we assume that $\tilde{P}$ is an Ω-factor which is SZ and assume $\tilde{Q}$ is a τ_j^i-rigid extension of $\tilde{P}$. To fix matters let us take $i=k$, $j=1$ so that $\tilde{Q}$ is a τ_1^k-rigid extension of $\tilde{P}$.

Our proof will be based on the following proposition. In this proposition having fixed $\tilde{P}$ and $\tilde{Q}$ we denote by $Y(\omega)$ the space of the factor whose projection operator is P_ω. For $y\in Y(\omega)$, $\|g\|_y^2=\|g\|_{L^2(\mu_y)}^2=P_\omega(g^2)(y)$. In the ensuing discussion we use systematically the notation $\tau_1^u(\alpha;\omega)y$, where for $y\in Y(\omega)$ and $\alpha\subset N_1(\omega)$ we shall have in accordance with (6.2)

$$\tau_1^u(\alpha;\omega)y\in Y(\omega_\alpha^u).$$

Proposition. *Let $\tilde{Q}$ be the τ_1^k-rigid extension of $\tilde{P}$, and for a generic point $\omega_0\in\Omega$, let $\mathfrak{M}$ be a finite rank module for Q_{ω_0} over P_{ω_0} which is quasi-invariant with respect to the IP-system $\{\tau_1^k(\beta;\omega_0),\ \beta\subset N_1(\omega_0)\}$, and let g be a bounded function in $\mathfrak{M}$. Then for any $\epsilon,\eta>0$ and $q\in\mathbf{N}$ there exist numbers $L=L(\eta,q)$, $N'=N'(\epsilon,\eta,q)$, $N''=N''(\epsilon,\eta,q)$ such that if $\omega(n)=\omega_0(n)$ for $n\le N'$ and if $\delta_1<\delta_2<\cdots<\delta_L$ are subsets of $N_1(\omega_0)\cap[N'',\infty]$, then but for a set of $y\in Y(\omega)$ of measure $<\epsilon$ we can find $\alpha,\beta_1,\beta_2,\ldots,\beta_q$ which are disjoint subsets and each of which is a union of δ_j with $\beta_1<\beta_2<\cdots<\beta_q$ and a word v of length $|\alpha|$, so that for every pair of words u,u' defined over $\beta=\beta_1\cup\cdots\cup\beta_q$ which are constant on each β_j and which satisfy for each coordinate*

$$u(r)=u'(r)\quad\text{or}\quad u(r)=k\text{ and }u'(r)=1$$

we have

$$\text{(9.2)}\qquad \|\tau_1^u(\beta;\omega_\alpha^v)^{-1}g-\tau_1^{u'}(\beta;\omega_\alpha^v)^{-1}g\|_{\tau_1^v(\alpha;\omega)y}<\eta.$$

Remark. We shall need this proposition only for $\omega=\omega_0$ but we formulate it more generally because our proof, which will be by induction, is based on the stronger statement.

Proof. In the course of this proof expressions of the form $\tau_1^u(\alpha;\omega)$ occur frequently and we will omit the subscript 1. We also write $\|\ \|_y$ for $\|\ \|_{L^2(\mu_y)}$. We proceed by induction on q. For $q=1$ the assertion is that we can find N',N'',L so that, if $\omega(n)=\omega_0(n)$ for $n\le N'$ and $N''<\delta_1<\delta_2<\cdots<\delta_L$ with $\delta_j\subset$

$N_1(\omega)$, then outside a set of y of measure $< \epsilon$, there exist disjoint α and β, each a union of δ's, and some v of length $|\alpha|$, such that

$$\|\tau^k(\beta;\omega_\alpha^v)^{-1}g - g\|_{\tau^v(\alpha;\omega)y} < \eta. \tag{9.3}$$

Consider the module $\mathfrak{M}(\omega)$ for ω to close to ω_0 (see §6.10 for the definition of $\mathfrak{M}(\omega)$). Assume $\|g\|_\infty \leq 1$. $\mathfrak{M}(\omega)_y$ is a vector space of dimension $\leq r$ for each $y \in Y(\omega)$ and we can find an $\eta/3$-dense subset of its unit ball of cardinality $\leq L-1$ for some L. These vectors may be chosen to depend measurably on $y \in Y(\omega)$ so that we can assume the existence of $L-1$ functions $\{g_1, g_2, \ldots, g_{L-1}\}$ with $\|g_j\|_y \leq 1$ and for all y,

$$\inf_{1\leq j\leq L-1} \|h - g_j\|_y < \eta/3 \tag{9.4}$$

for each $h \in \mathfrak{M}(\omega)$ with $\|h\|_y \leq 1$.

Choose N', N'' in accordance with Lemma 6.10 with ϵ replaced by ϵ_1, which we will specify later. The inequality of Lemma 6.10 implies that, but for a set B_α of y of measure $< \epsilon_1$,

$$\inf_{\substack{\|h\|_y\leq 1\\ h\in\mathfrak{M}(\omega)}} \|\tau^k(\alpha;\omega)^{-1}g - h\|_y^2 < \epsilon_1$$

where g is the function occurring in the proposition and we assume $|g| \leq 1$. For each remaining $y \in Y(\omega)$ we will have by (9.4)

$$\inf_{1\leq j\leq L-1} \|\tau^k(\alpha;\omega)^{-1}g - g_j\|_y < \sqrt{\epsilon_1} + \eta/3. \tag{9.5}$$

Here we assume $\alpha \subset \{N'', N''+1, \ldots\}$. Set $\alpha_p = \{\delta_p \cup \delta_{p+1} \cup \cdots \cup \delta_L\}$. p ranges from 1 to L and, since there are only $L-1$ choices for g_j in (9.5), we will obtain for two values p_1, p_2 that

$$\|\tau^k(\alpha_{p_1};\omega)^{-1}g - \tau^k(\alpha_{p_2};\omega)^{-1}g\|_y < 2\sqrt{\epsilon_1} + 2\eta/3 < \eta \tag{9.6}$$

provided $2\sqrt{\epsilon_1} < \eta/3$ and for $y \notin \bigcup_{p=1}^L B_{\alpha_p}$. Suppose $p_2 > p_1$ and write $\alpha_{p_1} = \beta \cup \alpha_{p_2}$. Then $\tau^k(\alpha_{p_1};\omega) = \tau^k(\beta;\omega)\tau^k(\alpha_{p_2};\omega)$ (by Corollary 5.1) and (9.6) becomes

$$\|\tau^k(\alpha_{p_2};\omega)^{-1}\{\tau^k(\beta;\omega)^{-1}g - g\}\|_y < \eta$$

which implies

$$\|\tau^k(\beta;\omega)^{-1}g - g\|_{\tau^k(\alpha_{p_2};\omega)y} < \eta.$$

If we assume further that $\epsilon_1 < \epsilon/L$, then the exceptional set of y has measure $< \epsilon$ and this completes the proof for $q = 1$.

Now assume the proposition is true for q; we wish to extend it to $q+1$. Let

$N'_q, N''_q, L(\eta', q)$ be the corresponding numbers associated with ϵ', $\eta' = \eta/(q+1)$ and q where ϵ' will be determined momentarily. (Notice that $L = L(\eta', q)$ does not depend on ϵ' and is therefore available before the definition of the latter.) Let K be the cardinality of an $\eta/6$-dense set in the unit ball in r-dimensional Euclidean space. Set $H = K^{k^q} + 1$.

We now invoke the Hales-Jewett theorem to obtain a number $N(r, h, k)$ so that, if words of length $\geq N(r, h, k)$ from an alphabet of k letters are partitioned into r subsets, one of these contains an h-dimensional combinatorial subspace. Set

$$L = L_q + N(r, H, k) \qquad \text{with } r = 2^{(q+1)L_q} k^{L_q}, \tag{9.7}$$

where L_q is an abbreviation for $L(\eta', q)$. Note that L depends on q and η. Finally obtain $N' \geq N'_q$, $N'' \geq N''_q$ and so that Lemma 6.10 holds with ϵ replaced by $\epsilon_1(\epsilon, \eta, q)$, which we determine later.

Take $\omega \in \Omega$ with $\omega(n) = \omega_0(n)$ for $n \leq N'$ and suppose $N'' \leq \delta_1 < \delta_2 < \cdots < \delta_L$ with $\delta_j \subset N_1(\omega)$. (By $N'' < \delta_1$ we mean $N'' < \min \delta_1$.) Letting $R = \delta_1 \cup \delta_2 \cup \cdots \cup \delta_{L_q}$ we know that for $y \in Y(\omega)$ outside an exceptional set of measure ϵ', we can find $\alpha, \beta \subset R$, $\alpha \cap \beta = \varnothing$ and a word v defined over α so that

$$\| \tau^u(\beta; \omega^v_\alpha)^{-1} g - \tau^{u'}(\beta; \omega^v_\alpha)^{-1} g \|_{\tau^v(\alpha;\omega)y} < \eta/(q+1) \tag{9.8}$$

where we write β for $\beta_1 \cup \cdots \cup \beta_q$, and u, u' are words of "length" q as described in the proposition.

If $\rho = \delta_{L_q+1} \cup \cdots \cup \delta_L$ and $\omega' = \omega^w_\rho$ this last result will be true for ω' as well. For each ω' there may be an exceptional set of points in $Y(\omega')$ of measure $< \epsilon'$. We can write $Y(\omega') = \tau^w(\rho; \omega) Y(\omega)$ and the exceptional set in $Y(\omega')$ corresponds to a set in $Y(\omega)$ of measure $< \epsilon'$. If

$$\epsilon' < \epsilon / 2k^{L(\eta, q)}$$

then the totality of exceptional values of $y \in Y(\omega)$ comprises a set B of measure $< \epsilon/2$.

Now take y outside of B. Consider words w defined over $\delta_{L_q+1} \cup \cdots \cup \delta_L$ and constant on each δ_j. To each such word, (9.8) will be valid for ω^w_ρ for some choice of α, β and v. Partition the set of these words w in accordance with the triple α, β, v or, more precisely: $\alpha, \beta_1, \dots, \beta_q, v$ that corresponds to ω^w_ρ. The set of possibilities is less than $2^{L_q} 2^{qL_q} k^{L_q} = r$. We now invoke the Hales-Jewett theorem. Accordingly we can find disjoint subsets $\gamma, \bar\delta_1, \bar\delta_2, \dots, \bar\delta_H$ composed of the atoms $\delta_{L_q+1}, \dots, \delta_L$, with $\bar\delta_1 < \bar\delta_2 < \cdots < \bar\delta_H$ and a word w defined over γ so that, for every choice of $j_1, \dots, j_H$, the same choice α, β, v ($\beta = \beta_1 \cup \cdots \cup \beta_q$) will satisfy (9.8). We thus find

$$\| \tau^u(\beta; \omega^{v, w, j_1, \dots, j_H}_{\alpha, \gamma, \bar\delta_1, \dots, \bar\delta_H})^{-1} g - \tau^{u'}(\beta; \omega^{v, w, j_1, \dots, j_H}_{\alpha, \gamma, \bar\delta_1, \dots, \bar\delta_H})^{-1} g \|_{y^*} < \eta/2 \tag{9.9}$$

where

$$y^* = \tau^v(\alpha;\omega^{w,j_1,\dots,j_H}_{\gamma,\bar\delta_1,\dots,\bar\delta_H})\tau^{w,j_1,\dots,j_H}(\gamma,\bar\delta_1,\dots,\bar\delta_H;\omega)y$$
$$= \tau^{v,w,j_1,\dots,j_H}(\alpha,\gamma,\bar\delta_1,\dots,\bar\delta_H;\omega)y. \tag{9.9'}$$

Fix the data α,β,γ,v,w and let u be a word of "length" q. Now choose s with $1\le s\le H$ and consider $\delta_s'=\bar\delta_s\cup\bar\delta_{s+1}\cup\cdots\cup\bar\delta_H$. With our choice of N',N'' we will have, in applying Lemma 6.10,

$$\inf_{h\in\mathfrak{M}(\omega^{v,u,w}_{\alpha,\beta,\gamma})}\|\tau^k(\delta_s';\omega^{v,u,w}_{\alpha,\beta,\gamma})^{-1}g-h\|^2_{\tau^{v,u,w}(\alpha,\beta,\gamma;\omega)y}<\epsilon_1(\epsilon,\eta,q) \tag{9.10}$$

but for a set of $\tau^{v,u,w}(\alpha,\beta,\gamma;\omega)y$ of measure $<\epsilon_1(\epsilon,\eta,q)$. Let $G(L)$ bound the number of possibilities for α,β,γ and v,u,w. If $\epsilon_1(\epsilon,\eta,q)<\epsilon/2G(L)$ then, outside of a set of $y\in Y(\omega)$ of measure ϵ, we will have (9.9) for some α,β,γ, $\bar\delta_1,\dots,\bar\delta_H,v,w$ as well as (9.10) for each u over β.

Fix y outside of this exceptional set. Next fix in each of the spaces $\mathfrak{M}_{\omega^{v,u,w}_{\alpha,\beta,\gamma}}$ as u ranges over all k^q possibilities, an $\eta/6$-dense subset of the unit ball of cardinality K. Denote the set $\{g_u^{(1)},\dots,g_u^{(K)}\}$ where we have suppressed the remaining parameters. applying (9.10) we will find for each s and u an index $p(s,u)$ with

$$\|\tau^k(\delta_s';\omega^{v,u,w}_{\alpha,\beta,\gamma})^{-1}g-g_u^{(p(s,u))}\|_{\tau^{v,u,w}(\alpha,\beta,\gamma;\omega)y}<\sqrt{\epsilon_1(\epsilon,\eta,q)}+\eta/6. \tag{9.11}$$

If we impose the condition $\sqrt{\epsilon_1(\epsilon,\eta,q)}<\eta/12$, since $H>K^{k^q}$ we can find $1\le s<t\le H$ so that, for each u,

$$\|\tau^k(\delta_s';\omega^{v,u,w}_{\alpha,\beta,\gamma})^{-1}g-\tau^k(\delta_t';\omega^{v,u,w}_{\alpha,\beta,\gamma})^{-1}g\|_{\tau^{v,u,w}(\alpha,\beta,\gamma;\omega)y}<\eta/2. \tag{9.12}$$

We now set $\delta=\bar\delta_s\cup\bar\delta_{s+1}\cup\cdots\cup\bar\delta_{t-1}$ so that $\delta_s'=\delta\cup\delta_t'$, and rewrite (9.12) as

$$\begin{aligned}
&\|\tau^k(\delta_t';\omega^{v,u,w}_{\alpha,\beta,\gamma})^{-1}\{\tau^k(\delta;\omega^{v,u,w}_{\alpha,\beta,\gamma})^{-1}g-g\}\|_{\tau^{v,u,w}(\alpha,\beta,\gamma;\omega)y}\\
&=\|\tau^k(\delta;\omega^{v,u,w}_{\alpha,\beta,\gamma})^{-1}g-g\|_{\tau^k(\delta_t';\omega^{v,u,w}_{\alpha,\beta,\gamma})\tau^{v,u,w}(\alpha,\beta,\gamma;\omega)y}\\
&=\|\tau^k(\delta;\omega^{v,u,w}_{\alpha,\beta,\gamma})^{-1}g-g\|_{\tau^{k,v,u,w}(\delta_t',\alpha,\beta,\gamma;\omega)y}\\
&=\|\tau^k(\delta;\omega^{v,u,w}_{\alpha,\beta,\gamma})^{-1}g-g\|_{\tau^u(\beta;\omega^v_\alpha)\tau^{k,v,w}(\delta_t',\alpha,\gamma;\omega)y}\\
&=\|\tau^u(\beta;\omega^v_\alpha)^{-1}\tau^k(\delta;\omega^{v,u,w}_{\alpha,\beta,\gamma})^{-1}g-\tau^u(\beta;\omega^v_\alpha)^{-1}g\|_{\tau^{k,v,w}(\delta_t',\alpha,\gamma;\omega)y}\\
&=\|\tau^{u,k}(\beta,\delta;\omega^{v,w}_{\alpha,\gamma})^{-1}g-\tau^{u,1}(\beta,\delta;\omega^{v,w}_{\alpha,\gamma})^{-1}g\|_{\tau^{k,v,w}(\delta_t',\alpha,\gamma;\omega)y}<\eta/2.
\end{aligned} \tag{9.13}$$

Here we use (5.3) repeatedly bearing in mind that $\alpha,\beta<\gamma\cup\delta\cup\delta_t'$ and that, for $\vartheta_1<\vartheta_2$, we have $\tau^z(\vartheta_1;\omega^{z'}_{\vartheta_2})=\tau^z(\vartheta_1;\omega)$.

We return to (9.9) where we now choose

$$j_1=j_2=\cdots=j_{s-1}=1,\quad j_s=j_{s+1}=\cdots=j_{t-1}=j,\quad j_t=j_{t+1}=\cdots=j_H=k.$$

This gives

$$\| \tau^u(\beta; \omega^{v,w,j,k}_{\alpha,\gamma,\delta,\delta'_t})^{-1} g - \tau^{u'}(\beta; \omega^{v,w,j,k}_{\alpha,\gamma,\delta,\delta'_t})^{-1} g \|_{\tau^{v,w,j,k}(\alpha,\gamma,\delta,\delta'_t;\omega)y} < \eta/2.$$

Rewrite the subscript as

$$\tau^j(\delta; \omega^{v,w,k}_{\alpha,\gamma,\delta'_t}) \tau^{k,v,w}(\delta'_t, \alpha, \gamma; \omega) y$$

and this gives

$$\| \tau^{u,j}(\beta,\delta; \omega^{v,w,k}_{\alpha,\gamma,\delta'_t})^{-1} g - \tau^{u',j}(\beta,\delta; \omega^{v,w,k}_{\alpha,\gamma,\delta'_t}) \|_{\tau^{v,w,k}(\alpha,\gamma,\delta'_t;\omega)y} < \eta/2. \tag{9.14}$$

Now a word pair u, u' of length $q+1$ has the form $(u,j),(u',j)$ with u, u' a pair for length q or it has the form $(u,k),(u',1)$. Combining (9.14) with (9.13) we obtain the desired result. This completes the proof of the proposition. QED

10. Lifting *SZ*

10.1. Theorem 8.1 (= Theorem 5.3 ⇒ Theorem 2.8) will follow from the following result.

Theorem. *Let $\tilde{Q}$ be a τ^i_j-rigid extension of $\tilde{P}$ and assume $\tilde{P}$ is an SZ Ω-factorization. Then $\tilde{Q}$ is an SZ Ω-factorization.*

In proving the theorem we shall assume that the density version of the Hales-Jewett theorem has already been established for an alphabet of $k-1$ letters. This will appear in the form of the following lemma (based on the aforementioned induction hypothesis):

Lemma. *There exist functions $M(\delta, l) < \infty$ and $\theta(\delta, l, M) > 0$ so that if for some $M > M(\delta, l)$ and for every M-tuple $j_1, \ldots, j_M \in \{1,2,\ldots,k-1\}^M$ there is defined a measurable function $f_{j_1,\ldots,j_M}$, $0 \le f_{j_1,\ldots,j_M} \le 1$, on some measure space (S, σ) with*

$$\prod_{j_1,\ldots,j_M=1}^{k-1} \int_S f_{j_1,\ldots,j_M}\, d\sigma \ge \delta^{(k-1)^M}, \tag{10.1}$$

then there exists a word $w(t_1, \ldots, t_l) \in W(k-1, l)$ of length M, so that

$$\int \prod_{t_1,\ldots,t_l=1}^{k-1} (f_{w(t_1,\ldots,t_l)})^{2^l}\, d\sigma > \theta(\delta, l, M).$$

Proof. Let $M(\delta, l)$ be such that for $M > M(\delta, l)$, a subset of $W_M(k-1)$ of density $\ge \delta/4$ contains an l-dimensional combinatorial subspace. By (10.1) and Jensen's inequality we have

$$\frac{1}{(k-1)^M} \sum \int f_{j_1,\ldots,j_M}\, d\sigma > \delta$$

so for $\delta/2$ of the measure of S we will have

$$\frac{1}{(k-1)^M}\sum f_{j_1,\ldots,j_M}(s) > \frac{\delta}{2}$$

and hence for $\delta/4$ of these terms we must have $f_{j_1,\ldots,j_M}(s) > \delta/4$. Now use density Hales–Jewett for $k-1$ to find the word $w(t_1,\ldots,t_l)$ with $f_{w(t_1,\ldots,t_l)} > \delta/4$. The word depends on the point $s \in S$ but the number of possibilities depends on l and M; say $G(l,M)$. Then

$$\theta(\delta,l,M) = \frac{\delta}{2G(l,M)}\cdot\left(\frac{\delta}{4}\right)^{(k-1)^l\cdot 2^l}. \qquad \text{QED}$$

10.2. Proof of Theorem 10.1. As before we take $i=k, j=1$ and suppress the index 1 in τ_1^k. Let $\omega_0 \in \Omega$ be a generic point and assume $0 \le f \le 1$ with $f \in L^\infty(X,\mathcal{B},\mu)$ satisfying $Q_{\omega_0}f = f$. Fix a natural number l. We wish to show that

$$\lim_{\alpha_1,\alpha_2,\ldots,\alpha_l\to\infty}\int\prod_{i_1,\ldots,i_l=1}^{k} T(\omega^{i_1,i_2,\ldots,i_l}_{\alpha_1,\alpha_2,\ldots,\alpha_l})^{-1}f\,d\mu > 0$$

for $\omega = \omega_0$.

This will be rewritten as

$$\lim_{\alpha_1,\alpha_2,\ldots,\alpha_l\to\infty}\int\prod_{i_1,\ldots,i_l=1}^{k}\tau^{i_1,\ldots,i_l}(\alpha_1,\alpha_2,\ldots,\alpha_l;\omega)^{-1}f\,d\mu > 0. \tag{10.2}$$

By continuity of this limit as a function of ω (the existence of the limits as well as their continuity is ensured by our assumption of coherence), it will suffice to exhibit a number $\rho > 0$ so that, for some ω arbitrarily close to ω_0 and for $\alpha_1, \alpha_2, \ldots, \alpha_l$ arbitrarily far out, the integral in (10.2) is larger than ρ.

Note that if the product in (10.2) were taken for indices from 1 to $(k-1)$, then our result would follow from Lemma 10.1 (where the exponent 2^l would be superfluous). The idea of the proof is to consider products with indices from 1 to $k-1$ and their integrals over suitable "fibers", i.e. for μ_y, $y \in Y(\omega)$ and then to use Proposition 9.1 to compare the remaining expressions in which k occurs with the expression in which k is replaced by 1. Since each word without k can be compared with at most 2^l words with k, we have formulated Lemma 10.1 with the exponent 2^l. Now Proposition 9.1 is formulated for functions g belonging to finite rank modules over P_{ω_0}. We must therefore approximate f by such a function. In fact we find a function f', $0 \le f' \le f$ and a function g in a finite rank module over P_{ω_0} with f' close to g for each $L^2(\mu_y)$, $y \in Y(\omega_0)$. We prove the desired result for g, then for f', and it then follows for $f \ge f'$.

Consider $P_{\omega_0}f$. There exist $a, b > 0$ with $P_{\omega_0}f > a$ on a set of $y \in Y(\omega_0)$ of measure b.

Set $M = M(a/4, l)$, $\theta = \theta(a/4, l, M)$, and $\epsilon_1 \leq \frac{3}{20}\theta/k^l$ and also $\epsilon_1 < a/4$. Since f belongs to the τ_1^k-rigid extension of P_{ω_0}, there is a function g_1 in some finite rank module $\mathfrak{M}$, say of rank r, so that

$$\|f - g_1\|_y < \epsilon_1 \tag{10.3}$$

but for a set of y of measure $<b/4$. Let $f' = f$ on the fibers of those y for which (10.3) is valid, and let $f' = 0$ for the remainder of the space. Let $g_1' = g_1$ for those y where (10.3) is valid and $g_1' = 0$ for the remainder of the space. g_1' is again in the module $\mathfrak{M}$. Finally, let $g = (g_1' \vee 0) \wedge 1$. Clearly g is at least as close to f' as g_1'. We then have

$$\|f' - g\|_y < \epsilon_1 \tag{10.4}$$

for all $y \in Y(\omega_0)$.

We will now show that there exists ω arbitrarily close to ω_0 and $\alpha_1, \ldots, \alpha_l$ arbitrarily far out, $\alpha_j \subset N_1(\omega)$, so that

$$\int \prod \tau^{i_1, \ldots, i_l}(\alpha_1, \alpha_2, \ldots, \alpha_l; \omega)^{-1} g \, d\mu_y > \theta/2, \tag{10.5}$$

for a set of $y \in Y(\omega)$ of measure bounded away from 0. The bound, which we denote by $b(f, l)$, will be estimated later. (10.4) can be rewritten as

$$P_{\omega_0}(f' - g)^2(y) < \epsilon_1^2 \qquad \text{for all } y \in Y(\omega_0).$$

By L^2-continuity of $P_\omega(f' - g)^2$, if ω is sufficiently close to ω_0 we will have

$$\|f' - g\|_y^2 < (\tfrac{3}{2}\epsilon_1)^2$$

but for a set of $y \in Y(\omega)$ of measure $< b(f, l)/4k^l$. This implies that, but for a set of points of measure $< b(f, l)/4k^l$,

$$\tau^{i_1, \ldots, i_l}(\alpha_1, \ldots, \alpha_l; \omega)^{-1} P_\omega (f' - g)^2 < \left(\frac{3}{2}\epsilon_1\right)^2.$$

Now

$$\tau^{i_1, \ldots, i_l}(\alpha_1, \ldots, \alpha_l; \omega) P_\omega = P_{\omega^{i_1, \ldots, i_l}_{\alpha_1, \ldots, \alpha_l}} \tau^{i_1, \ldots, i_l}(\alpha_1, \ldots, \alpha_l; \omega)$$

whence

$$\tau^{i_1, \ldots, i_l}(\alpha_1, \ldots, \alpha_l; \omega)^{-1} P_{\omega^{i_1, \ldots, i_l}_{\alpha_1, \ldots, \alpha_l}} (f' - g)^2 = P_\omega \tau^{i_1, \ldots, i_l}(\alpha_1, \ldots, \alpha_l; \omega)^{-1} (f' - g)^2.$$

Since as $\alpha_1, \ldots, \alpha_l \to \infty$, $P_{\omega^{i_1,\ldots,i_l}_{\alpha_1,\ldots,\alpha_l}} \to P_\omega$ strongly, the functions

$$\tau^{i_1,\ldots,i_l}(\alpha_1, \ldots, \alpha_l; \omega)^{-1} P_\omega (f' - g)^2, \quad P_\omega \tau^{i_1,\ldots,i_l}(\alpha_1, \ldots, \alpha_l; \omega)^{-1}(f' - g)^2 \tag{10.6}$$

are as close as we like in $L^2(X, \mathcal{B}, \mu)$, provided the α's are chosen far out. We impose on N_0 the condition that $\alpha_1 > N_0$, $i = 1, \ldots, l$ implies that the two functions in (10.6) are close enough so that

$$P_\omega \tau^{i_1,\ldots,i_l}(\alpha_1, \ldots, \alpha_l; \omega)^{-1}(f' - g)^2 < (\tfrac{5}{3}\epsilon_1)^2 \tag{10.7}$$

but for a set of points $y \in Y(\omega)$ (or equivalently, a set of points in X) of measure $< b(f, l)/3k^l$.

We will now have simultaneously the inequalities (10.7) for all $(i_1, \ldots, i_l)$ in $\{1, \ldots, k-1\}^l$ and all y outside a set of measure $< \frac{1}{3}b(f, l)$. Since all our functions are bounded by 1, at each $y \in Y(\omega)$

$$\begin{aligned} &P_\omega\left(\left|\prod \tau^{i_1,\ldots,i_l}(\alpha_1, \ldots, \alpha_l; \omega)^{-1} g - \prod \tau^{i_1,\ldots,i_l}(\alpha_1, \ldots, \alpha_l; \omega)^{-1} f'\right|\right) \\ &\quad \le \sum P_\omega\{\tau^{i_1,\ldots,i_l}(\alpha_1, \ldots, \alpha_l; \omega)^{-1}|g - f'|\} \\ &\quad \le \sum \|\tau^{i_1,\ldots,i_l}(\alpha_1, \ldots, \alpha_l; \omega)^{-1}(g - f')\|_y \end{aligned}$$

and thus will be less than $\frac{5}{3}k^l\epsilon_1$ but for a set of y of measure $< b(f, l)/2$. Since $\frac{5}{3}k^l\epsilon_1 \le \theta/4$ we now have

$$\int \prod \tau^{i_1,\ldots,i_l}(\alpha_1, \alpha_2, \ldots, \alpha_l; \omega)^{-1} f' \, d\mu_y > \theta/4$$

on a set of y of measure $> b(f, l)/2$. Integrating over y and replacing f' by f we obtain, for $\alpha_1, \alpha_2, \ldots, \alpha_l > N_0$,

$$\int \prod \tau^{i_1,\ldots,i_l}(\alpha_1, \alpha_2, \ldots, \alpha_l; \omega)^{-1} f \, d\mu > \theta b(f, l)/8.$$

Thus ρ can be taken as $\theta b(f, l)/8$.

We are left with establishing (10.5) for ω close to ω_0 and $\alpha_1, \ldots, \alpha_l$ far out.

Recall that the function g_1' belongs to the finite rank module $\mathfrak{M}$ over P_{ω_0}. We shall apply Proposition 9.1 to g_1'. We first determine the η to be used. Namely, we take

$$\eta = \theta/2^{l+1}.$$

We next prescribe q by

$$q = M(a/4, l)$$

where $M(\delta, l)$ is the function defined in Lemma 10.1. This value of (η, q) determines L so that the $\alpha, \beta_1, \ldots, \beta_q$ of Proposition 9.1 are constructed from L atoms occurring "far out" in ω, where ω is near ω_0. We proceed to determine the parameter ϵ.

Recall that we have, by (10.3),

$$\|f - g_1\|_y < a/4$$

but for a set of y of measure $< b/4$. Since g is at least as close to f as g_1 is, we have

$$\|f - g\|_y < a/4$$

but for a set of y of measure $< b/4$. Since $P_{\omega_0} f > a$ on a set of measure b, we will have

$$P_{\omega_0} g > a/2$$

on a set of measure $> b/2$. Let F be defined by

$$F(x) = \begin{cases} 1 & \text{if } P_{\omega_0} g(x) > a/2, \\ 0 & \text{otherwise;} \end{cases}$$

then $\int F\, d\mu > 0$ and $F = P_{\omega_0} F$. Use the fact that $\tilde{P}$ is SZ to obtain for L defined above

$$\lim_{\gamma_1, \ldots, \gamma_L \to \infty} \int \prod_{j_1, \ldots, j_L = 1}^{k} \tau^{j_1, \ldots, j_L}(\gamma_1, \ldots, \gamma_L; \omega_0)^{-1} F\, d\mu = c > 0.$$

This implies that for $\gamma_1, \ldots, \gamma_L$ sufficiently far out

$$\tau^{j_1, \ldots, j_L}(\gamma_1, \ldots, \gamma_L; \omega_0)^{-1} P_{\omega_0} g > a/2 \tag{10.8}$$

simultaneously for all $(j_1, \ldots, j_L) \in \{1, \ldots, k\}^L$ on a set of measure $> c/2$. Use the continuity of P_ω and the commutation relations of 6.1 to write

$$P_{\omega_0} \tau^{j_1, \ldots, j_L}(\gamma_1, \ldots, \gamma_L; \omega_0)^{-1} g > a/4, \quad (j_1, \ldots, j_L) \in \{1, \ldots, k\}^L \tag{10.9}$$

on a set of measure $> c/4$. Set $\epsilon = c/8$.

Suppose now $\alpha, \beta_1, \ldots, \beta_q$ are disjoint sets built from $\gamma_1, \ldots, \gamma_L$, and let v be a word defined over α. We let u range over the words defined over $\beta_1 \cup \cdots \cup \beta_q$ that are constant on each β_j, so that u can be identified with q-tuples. By (10.9) we then have

$$P_{\omega_0} \tau^{v, u}(\alpha, \beta; \omega_0)^{-1} g > a/4$$

simultaneously on a set of measure $> c/4$, or

$$\int \tau^{v,u}(\alpha,\beta;\omega_0)^{-1} g \, d\mu_y > a/4 \tag{10.10}$$

for y in a set of measure $> c/4$ in $Y(\omega_0)$. We have $q = M(a/4, l)$ and Lemma 10.1 can be applied to the sequence of functions

$$\tau^{v,u}(\alpha,\beta;\omega_0)^{-1} g$$

indexed by u restricted to $\{1,\dots,k-1\}^q$ and for fixed v. Lemma 10.1 then provides for the existence of l sets $\delta_1 < \delta_2 < \cdots < \delta_l$ built from the β_i and $\bar{\alpha} \supset \alpha$ and $\bar{v}$ defined over $\bar{\alpha}$, extending v, with

$$\int \prod_{i_1,\dots,i_l=1}^{k-1} (\tau^{\bar{v},i_1,\dots,i_l}(\bar{\alpha},\delta_1,\dots,\delta_l;\omega_0)^{-1} g)^{2^l} \, d\mu_y > \theta \tag{10.11}$$

for y in a set of measure $> c/4$. We now invoke Proposition 9.1, applying it to the point $\omega = \omega_0$, and to the function g_1'. We let $\alpha, \beta_1, \dots, \beta_q$ and v satisfy (9.2). The latter can be rewritten

$$\| \tau^{v,u}(\alpha,\beta;\omega_0)^{-1} g_1' - \tau^{v,u'}(\alpha,\beta;\omega_0)^{-1} g_1' \|_y < \eta$$

but for a set of $y \in Y(\omega_0)$ of measures $< \epsilon$. Now g is g_1' truncated to lie between 0 and 1. The distance between truncated functions never exceeds the distance between the original functions; hence

$$\| \tau^{v,u}(\alpha,\beta;\omega_0)^{-1} g - \tau^{v,u'}(\alpha,\beta;\omega_0)^{-1} g \|_y < \eta$$

but for a set of y of measure $< \epsilon$. In particular if $(i_1',\dots,i_l') \in \{1,\dots,k-1\}^l$ and $(i_1,\dots,i_l) \in \{1,\dots,k\}^l$ with $i_j' = i_j$ unless $i_j' = 1$ and $i_j = k$, then

$$\| \tau^{\bar{v},i_1,\dots,i_l}(\bar{\alpha},\delta_1,\dots,\delta_l;\omega_0)^{-1} g - \tau^{\bar{v},i_1',\dots,i_l'}(\bar{\alpha},\delta_1,\dots,\delta_l;\omega_0)^{-1} g \|_y < \eta \tag{10.12}$$

with all these holding simultaneously but for a set of y of measure $< \epsilon$. Since $\epsilon = c/8$ we see that (10.11) and (10.12) hold simultaneously for a set of y of measure $\geq c/8$.

Now for each $(i_1,\dots,i_l) \in \{1,\dots,k\}^l$ let $(i_1',\dots,i_l')$ be the corresponding l-tuple in $\{1,\dots,k-1\}^l$. With this notation we have, bearing in mind that each l-tuple in $\{1,\dots,k-1\}^l$ occurs at most 2^l times,

$$\int \prod_{i_1,\dots,i_l=1}^{k} \tau^{\bar{v},i_1',\dots,i_l'}(\bar{\alpha},\delta_1,\dots,\delta_l;\omega_0)^{-1} g \, d\mu_y > \theta.$$

Combining this with (10.12) and using the fact that $2^l \eta = \theta/2$ we find

$$\int \prod_{i_1,\dots,i_l=1}^{k} \tau^{\bar{v},i_1,\dots,i_l}(\bar{\alpha},\delta_1,\dots,\delta_l;\omega_0)^{-1} g \, d\mu_y > \theta/2.$$

This can be rewritten

$$\int \tau^{\bar{v}}(\bar{\alpha};\omega_0)^{-1} \prod_{i_1,\dots,i_l=1}^{k} \tau^{i_1,\dots,i_l}(\delta_1,\dots,\delta_l;(\omega_0)^{\bar{v}}_{\alpha})^{-1} g\, d\mu_y$$

$$= \int \prod_{i_1,\dots,i_l=1}^{k} \tau^{i_1,\dots,i_l}(\delta_1,\dots,\delta_l;\omega)^{-1} g\, d\mu_{\zeta} > \theta/2$$

where $\zeta \in Y(\omega)$, $\omega = (\omega_0)^{\bar{v}}_{\alpha}$ and the latter holds for ζ in a set of measure $c/8$. This establishes (10.5) with $b(f,l) = c/8$ and completes the proof of Theorem 10.1. QED

REFERENCES

[C] T. J. Carlson, *Some unifying principles in Ramsey theory*, Discrete Math. **68** (1988), 117–169.

[CS] T. J. Carlson and S. G. Simpson, *A dual form of Ramsey's theorem*, Advances in Math. **53** (1984), 265–290.

[F1] H. Furstenberg, *Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions* J. Analyse Math. **31** (1977), 204–256.

[F2] H. Furstenberg, *Recurrence in Ergodic Theory and Combinatorial Number Theory*, Princeton Univ. Press, 1981.

[FK1] H. Furstenberg and Y. Katznelson, *An ergodic Szemerédi theorem for commuting transformations*, J. Analyse Math. **34** (1978), 275–291.

[FK2] H. Furstenberg and Y. Katznelson, *An ergodic Szemerédi theorem for IP-systems and combinatorial theory*, J. Analyse Math. **45** (1985), 117–168.

[FK3] H. Furstenberg and Y. Katznelson, *A density version of the Hales–Jewett theorem for* $k = 3$, Discrete Math. **75** (1989), 227–241.

[FK4] H. Furstenberg and Y. Katznelson, *Idempotents in compact semigroups and Ramsey theory*, Isr. J. Math. **68** (1990), 257–270.

[FKO] H. Furstenberg, Y. Katznelson and D. Ornstein, *The ergodic theoretic proof of Szemerédi's Theorem*, Bull. Am. Math. Soc. **7** (1982), 527–552.

[HJ] A. W. Hales and A. I. Jewett, *Regularity and positional games*, Trans. Am. Math. Soc. **106** (1963), 222–229.

[S] E. Sperner, *Ein Saltz über Untermengen einer Endlicher Menge*, Math. Z. **27** (1928), 544–548.

[SZ] E. Szemerédi, *On sets of integers containing no k elements in arithmetic progression*, Acta Arith. **27** (1975), 199–245.

H. Furstenberg
LANDAU CENTER FOR ANALYSIS
INSTITUTE OF MATHEMATICS
THE HEBREW UNIVERSITY OF JERUSALEM
JERUSALEM 91404, ISRAEL

Y. Katznelson
DEPARTMENT OF MATHEMATICS
STANFORD UNIVERSITY
STANFORD, CA 94305, USA

(Received March 20, 1990 and in revised form March 10, 1991)