

The wildcard set's size in the density Hales-Jewett theorem

Randall McCutcheon

Department of Mathematical Sciences
University of Memphis, Memphis, Tennessee, USA

rmcctchn@memphis.edu

Submitted: Feb 1, 2011; Accepted: May 3, 2011; Published: May 16, 2011

Mathematics Subject Classifications: 05D10, 11B25

Abstract

In this paper prove results concerning restrictions on the cardinality of the wildcard set in the density Hales-Jewett theorem, establishing in particular that for general k one may choose this cardinality from any IP set and that for $k = 2$ it may be chosen to be a square, thus providing an abstract extension of Sárközy's theorem on square differences in sets of positive upper density.

1. Introduction.

Let $k, N \in \mathbf{N}$. We view members of $\{0, 1, \dots, k-1\}^N$ as words of length N on the alphabet $\{0, 1, \dots, k-1\}$. A *variable word* is a word $w_1 w_2 \cdots w_N$ on the alphabet $\{0, 1, \dots, k-1, x\}$ in which the letter x (the *variable*) occurs. Indices i for which $w_i = x$ will be called *wildcards*, and $\{i : w_i = x\}$ will be called the *wildcard set*. We denote variable words by $w(x)$, e.g. $w(x) = 02x1x3210x$ is a variable word. If $w(x)$ is a variable word and $i \in \{0, 1, \dots, k-1\}$, we denote by $w(i)$ the word that results when all instances of “ x ” in $w(x)$ are replaced by “ i ”. E.g. $w(2) = 0221232102$ for the variable word $w(x)$ considered above.

In [HJ], A. Hales and R. Jewett proved the following theorem.

THEOREM 1.1. Let $k, r \in \mathbf{N}$. There exists $N = N(k, r)$ having the property that for any r -cell partition $\{0, 1, \dots, k-1\}^N = \bigcup_{i=1}^r C_i$, there are j , $1 \leq j \leq r$, and a variable word $w(x)$ such that $\{w(i) : i \in \{0, 1, \dots, k-1\}\} \subset C_j$.

In [FK2], H. Furstenberg and Y. Katznelson proved a density version of the theorem.

THEOREM 1.2. Let $\epsilon > 0$, $k \in \mathbf{N}$. There exists $M = M(\epsilon, k)$ having the property that if $E \subset \{0, 1, \dots, k-1\}^M$ with $|E| \geq \epsilon k^M$ then there exists a variable word $w(x)$ such that $\{w(t) : t \in \{0, 1, \dots, k-1\}\} \subset E$.

We now change our perspective slightly, viewing members of $\{0, 1, \dots, k-1\}^{N^2}$ as $N \times N$ matrices whose entries come from $\{0, 1, \dots, k-1\}$. A *variable matrix* is a matrix on the alphabet $\{0, 1, \dots, k-1, x\}$ in which the letter x occurs. We denote variable matrices by $m(x)$. If $m(x) = (m_{ij})_{i,j=1}^N$, the *wildcard set* of $m(x)$ is the set of

pairs (i, j) for which $m_{ij} = x$. If the wildcard set of $m(x)$ is equal to $\alpha \times \alpha$ for some $\alpha \subset \{1, 2, \dots, N\}$, we say that $m(x)$ is a *square variable matrix*.

In [BL], V. Bergelson and A. Leibman proved a “polynomial Hales-Jewett theorem”. Here is a special case.

THEOREM 1.3. Let $k, r \in \mathbf{N}$. There exists $N = N(k, r)$ having the property that for any r -cell partition $\{0, 1, \dots, k-1\}^{N^2} = \bigcup_{i=1}^r C_i$, there are j , $1 \leq j \leq r$, and a square variable matrix $m(x)$ such that $\{m(i) : i \in \{0, 1, \dots, k-1\}\} \subset C_j$.

It is natural to ask whether Theorem 1.3 admits of a density version.

CONJECTURE 1.4. Let $\epsilon > 0$, $k \in \mathbf{N}$. There exists $M = M(\epsilon, k)$ having the property that if $E \subset \{0, 1, \dots, k-1\}^{M^2}$ with $|E| \geq \epsilon k^{M^2}$ then there exists a square variable matrix $m(x)$ such that $\{m(i) : i \in \{0, 1, \dots, k-1\}\} \subset E$.

This question was first asked perhaps fifteen years ago, by V. Bergelson. Though a few of its would-be consequences have been established (see, e.g., [BLM], [M], [BM]), these results pay a high price for their polynomiality as none is strong enough to recapture the density Hales-Jewett theorem itself. It’s a good time for renewed interest in the matter; a recent online collaboration initiated by T. Gowers, *Polymath 1*, resulted in the discovery of a beautiful new proof of Theorem 1.2; see [P]. At around the same time, T. Austin found yet another proof; see [A]. Despite these positive results, however, Conjecture 1.4 has remained recalcitrant, and is open even for $k = 2$.

In the meantime, we seek to popularize here a somewhat weaker polynomial extension of Theorem 1.2 (Conjecture 1.6 below), which is nevertheless satisfying, natural and hopefully more amenable to attack. In support of this hope, we shall give two proofs of the initial case $k = 2$. The first is a simple density increment proof using the following theorem of A. Sárközy’s as a lemma.

THEOREM 1.5 ([S]). Let $\epsilon > 0$. There exists $S \in \mathbf{N}$ such that every $E \subset \{1, 2, \dots, N\}$ with $|E| \geq \epsilon S$ contains a configuration $\{a, a + n^2\}$, where $n \geq 0$.

This first proof for $k = 2$, which of course is not, in virtue of its use of Theorem 1.5, self-contained, is unlikely to generalize to cases $k > 2$. On the other hand our somewhat lengthy albeit fully self-contained second proof (given in Section 3) develops tools and a structure theory intended as a possibly viable first step in an attempt to prove the conjecture in full. Here now is the formulation.

CONJECTURE 1.6. Let $\delta > 0$, $k \in \mathbf{N}$. There exists $M = M(\delta, k)$ having the property that if $E \subset \{0, 1, \dots, k-1\}^M$ with $|E| \geq \delta k^M$ then there exist $n \in \mathbf{N}$ and a variable word $w(x)$ having n^2 occurrences of the letter x such that $\{w(t) : t \in \{0, 1, \dots, k-1\}\} \subset E$.

First proof for $k = 2$. Let δ_0 be the infimum of the set of δ for which the conclusion holds and assume for contradiction that $\delta_0 > 0$. Choose by Sárközy’s theorem m such that for any $A \subset \{1, 2, \dots, m\}$ with $|A| \geq \frac{\delta_0}{3}m$, A contains a configuration $\{a, a + n^2\}$, with $n > 0$. Let $\delta = \delta_0 - \frac{\delta_0}{4 \cdot 2^m}$ and put $M' = M(\delta_0 + \frac{\delta_0}{3 \cdot 2^m}, 2)$. Finally put $M = m + M'$. We claim M works as $M(\delta, 2)$. Suppose then that $E \subset \{0, 1\}^M$ with $|E| \geq \delta 2^M$.

Now, for each $v \in \{0, 1\}^m$, let $E_v = \{w \in \{0, 1\}^{M'} : vw \in E\}$. If $|E_v| > (\delta_0 + \frac{\delta_0}{3 \cdot 2^m})2^{M'}$ for some v we are done; E_v will by hypothesis contain $\{w(0), w(1)\}$ for some variable word w having n^2 wildcards for some n , so that $\{vw(0), vw(1)\} \subset E$. (Notice that $vw(x)$ is again a variable word having n^2 wildcards.)

We may therefore assume that $|E_v| \leq (\delta_0 + \frac{\delta_0}{3 \cdot 2^m})2^{M'}$ for every v . A simple calculation now shows that $|E_v| \geq \frac{\delta_0}{3}2^{M'}$ for all v (otherwise, E would be too small).

Now for $1 \leq i \leq m$, let v_i be the word consisting of i 0s followed by $(m-i)$ 1s. Since $\sum_{i=1}^m |E_{v_i}| \geq \frac{m\delta_0}{3}2^{M'}$, there must be some $u \in \{0, 1\}^{M'}$ with $|\{i : u \in E_{v_i}\}| \geq \frac{\delta_0}{3}m$. By choice of m , there are a and $n > 0$ such that $u \in E_{v_a} \cap E_{v_{a+n^2}}$. It follows that $\{v_a u, v_{a+n^2} u\} \subset E$. But this set plainly has the form $\{w(0), w(1)\}$ for a variable word $w(x)$ having n^2 wildcards. \square

2. Sets of word recurrence

Nothing about the set of squares beyond Sárközy's theorem was used in the previous section. In consequence, what holds for them should hold for more general sets of recurrence.

DEFINITION 2.1. Let $R \subset \mathbf{N}$. R is a *set of $(k-1)$ -recurrence* if for every $\epsilon > 0$ there exists $S \in \mathbf{N}$ such that every $E \subset \{1, 2, \dots, S\}$ with $|E| \geq \epsilon S$ contains a k -term arithmetic progression with common difference $r \in R$.

DEFINITION 2.2. Let $R \subset \mathbf{N}$. R is a *set of word $(k-1)$ -recurrence* if for every $\epsilon > 0$ there exists $M = M(\epsilon) \in \mathbf{N}$ having the property that if $E \subset \{0, 1, \dots, k-1\}^M$ with $|E| \geq \delta k^M$ then there exists a variable word $w(x)$ having $r \in R$ occurrences of the letter x such that $\{w(t) : t \in \{0, 1, \dots, k-1\}\} \subset E$.

A few remarks are in order. Sets of $(k-1)$ -recurrence are also known as sets of $(k-1)$ -density intersectivity. There is an analogous notion *sets of $(k-1)$ -chromatic intersectivity*, also known as *sets of topological $(k-1)$ -recurrence*; one could define “sets of chromatic word intersectivity” and inquire about them. Many variations are possible, e.g. the IP Szemerédi theorem [FK1] and IP van der Waerden theorems deal with set-valued parameters analogous to wildcard sets. Or, one could take the salient sets of recurrence to be families of finite subsets of \mathbf{N} from which one may always choose a suitable wildcard set, rather than sets of natural numbers from which one can always choose the cardinality of a suitable wildcard set. This brief discussion is intended as an introduction to these and other possibilities.

THEOREM 2.3. Let $R \subset \mathbf{N}$. If R is a set of word $(k-1)$ -recurrence then R is a set of $(k-1)$ -recurrence.

Proof. Let $\epsilon > 0$ and choose $M = M(\frac{\epsilon}{2})$ as in Definition 2.2. Let $J \gg (k-1)M$, let $E \subset \{1, 2, \dots, J\}$ with $|E| \geq \epsilon J$ and let X be a random variable uniformly distributed on $\{1, 2, \dots, J - (k-1)M\}$. Finally let $E' = \{w_1 w_2 \dots w_M \in \{0, 1, \dots, k-1\}^M : X + w_1 + w_2 + \dots + w_M \in E\}$. E' is a random subset of $\{0, 1, \dots, k-1\}^M$; since

each word is expected to be in E' with probability approaching ϵ as $J \rightarrow \infty$, by fixing J large enough we can ensure there is always a possible value of X for which $|E'| \geq \frac{\epsilon}{2}k^M$. Therefore, there is a variable word $w(x)$ having a wildcard set of size $r \in R$ for which $L = \{w(j) : j \in \{0, 1, \dots, k-1\}\} \subset E'$. But the image of L under the map $w_1w_2 \cdots w_M \rightarrow X + w_1 + w_2 + \cdots + w_M$ is an arithmetic progression contained in E and having common difference r . \square

QUESTION 2.4. Is every set of $(k-1)$ -recurrence a set of word $(k-1)$ -recurrence?

The answer is *yes* for $k = 2$. To see this, simply note that in the proof of the $k = 2$ case of Conjecture 1.6, all that was used of the set of squares was Theorem 1.5, the analog of which for an arbitrary set of recurrence R is true by definition. We suspect the answer in general to be *no*.

The only (non-trivial) class of sets that we know to be sets of word $(k-1)$ -recurrence for all k are *IP sets*. (An IP set in \mathbf{N} consists of an infinite sequence (x_i) and its finite sums formed by adding terms with distinct indices, i.e. $\{\sum_{i \in \alpha} x_i : \alpha \subset \mathbf{N}, 0 < |\alpha| < \infty\}$.) This is the content of the following theorem, the proof of which requires the following notion: given an “ M -variable word” $w(x_1, x_2, \dots, x_M) = w_1w_2 \cdots w_J$, i.e. a word on the alphabet $\{0, 1, \dots, k-1\} \cup \{x_1, x_2, \dots, x_M\}$ in which each of the symbols x_i occurs at least once, the range of the map $\{0, 1, \dots, k-1\}^M \rightarrow \{0, 1, \dots, k-1\}^J$ defined by $a_1a_2 \cdots a_M \rightarrow w(a_1, a_2, \dots, a_M)$ is called an *M -dimensional subspace* of $\{0, 1, \dots, k-1\}^J$.

THEOREM 2.5. IP sets are sets of word $(k-1)$ -recurrence for all k .

Proof. Let (x_i) be an infinite sequence in \mathbf{N} and let $\epsilon > 0$, $k \in \mathbf{N}$. Let $M = M(\frac{\epsilon}{2}, k)$ as in Theorem 1.2 and let $J \gg x_1 + x_2 + \cdots + x_M$. Let now $E \subset \{0, 1, \dots, k-1\}^J$ with $|E| \geq \epsilon k^J$. Select a random M -dimensional subspace I of $\{0, 1, \dots, k-1\}^J$ as follows: choose disjoint sets $\alpha_i \subset \{1, 2, \dots, J\}$ with $|\alpha_i| = x_i$, $1 \leq i \leq M$ uniformly at random. Next, fix random letters at positions outside $\bigcup_{i=1}^M \alpha_i$. I consists of words having those fixed letters at positions outside $\bigcup_{i=1}^M \alpha_i$ that are also constant on each α_i .

I may be identified with $\{0, 1, \dots, k-1\}^M$ under a map that preserves combinatorial lines. Such lines in I are associated with variable words whose wildcard sets are unions of α_i s, so we will be done if $\frac{|I \cap E|}{k^M} \geq \frac{\epsilon}{2}$ with positive probability. Notice that if each word belonged to I with the same probability k^{M-J} this would be immediate. Such is not the case; for fixed J , $P(w \in I)$ is a function of the frequencies of occurrence of each letter of $\{0, 1, \dots, k-1\}$ in the word w , indeed is proportional to the probability that w_i is constant on each α_j . Constant words $jj \cdots j$ are most likely to belong to I (with probability $k^{x_1 + \cdots + x_M - J}$). However, as $J \rightarrow \infty$ the minimum over all words w of $P(w \in I)$ is asymptotically equivalent to the average value k^{M-J} . Indeed, $P(w \in I)$ is asymptotically equivalent to $k^{x_1 + \cdots + x_M - J} \prod_{i=1}^M \sum_{\lambda \in \{0, 1, \dots, k-1\}} f_\lambda^{x_i}$, where f_λ is the relative frequency of λ in w . The latter function is continuous in the variables f_λ and subject to the constraint $\sum_\lambda f_\lambda = 1$ its minimum value of k^{M-J} obtains at $f_\lambda = \frac{1}{k}$ for all λ (a calculus exercise). Choosing J large enough that $P(w \in I) > \frac{1}{2}k^{M-J}$ uniformly and summing over $w \in E$ yields an expectation for $\frac{|I \cap E|}{k^M}$ of at least $\frac{\epsilon}{2}$, as required. \square

3. A self-contained proof of Conjecture 1.6 for $k = 2$

We use a correspondence principle that recasts the problem as a recurrence question in ergodic theory (cf. [F]). Furstenberg and Katznelson developed such a principle for the density Hales-Jewett theorem in [FK2] via the Carlson-Simpson theorem [CS]. That approach is not useful here as it loses information about the size of the wildcard set. Therefore we use an alternate scheme proposed by T. Tao on the Polymath 1 blog [P].

Suppose to the contrary that there is an $\epsilon > 0$ such that for every n , there is a set $A^n \subset \{0, 1\}^n$ with $|A^n| \geq \epsilon 2^n$ containing no pair $\{w(0), w(1)\}$ where w is a variable word having r^2 wildcards for some r (we will call such sets “square line free”). Now for $0 \leq m \leq n$ we can form random square line free sets $A_m^n \subset \{0, 1\}^m$ by randomly embedding $\{0, 1\}^m$ in $\{0, 1\}^n$. More precisely,

1. Pick distinct x_1, \dots, x_m in $\{1, 2, \dots, n\}$ uniformly at random.
2. Pick a word $(y_i)_{i \notin \{x_1, x_2, \dots, x_m\}}$ in $\{0, 1\}^{n-m}$ to fill in the other positions, uniformly at random.
3. Put $w = (w_i)_{i=1}^m \in A_m^n$ if $(z_i)_{i=1}^n \in A^n$, where $z_{x_i} = w_i$, $1 \leq i \leq m$ and $z_j = y_j$, $j \notin \{x_1, x_2, \dots, x_m\}$.

Notice that for each $w \in \{0, 1\}^m$, $P(w \in A_m^n) \geq \epsilon$. By restricting n to a subsequence S , one may ensure that as $n \rightarrow \infty, n \in S$, the random sets stabilize in distribution for all m . Denote by μ_m the measure on $2^{\{0,1\}^m}$ giving the limiting distribution. Thus if I is a set of words of length m ,

$$\mu_m(\{I\}) = \lim_{n \rightarrow \infty, n \in S} P(A_m^n = I).$$

Since each A_m^n is square line free, $\mu_m(\{E\}) = 0$ for any E containing a square line.

Let $i \in \{0, 1\}$ and let \mathcal{J} be a family of sets of words of length m . Define a new family of words $\mathcal{J} * i$ of length $m + 1$ as follows: if B is a set of words of length $m + 1$, first throw away any member of B whose last letter is not i and truncate the remaining words to length m (i.e. knock off the final i). If (and only if) the set of words that remains is a member of \mathcal{J} , then $B \in \mathcal{J} * i$. Observe now the following stationarity condition: $\mu_{m+1}(\mathcal{J} * i) = \mu_m(\mathcal{J})$.

It is convenient to have the measures μ_m defined on the same space, so let $X = \prod_{m=1}^{\infty} 2^{\{0,1\}^m}$, and let \mathcal{B}_m be the algebra of sets $\{\prod_{r=1}^{\infty} E_r : E_r = 2^{\{0,1\}^r} \text{ if } r \neq m\}$; μ_m can be viewed as a measure on \mathcal{B}_m . Let \mathcal{B} be the σ -algebra generated by the \mathcal{B}_m and let μ be the product of the μ_m . We now require the following elementary lemma.

LEMMA 3.1. Let \mathcal{C} and \mathcal{D} be finite algebras of measurable sets in a probability space (X, \mathcal{B}, μ) . Assume there is a measure preserving isomorphism $U : \mathcal{C} \rightarrow \mathcal{D}$. There exists an invertible measure preserving transformation $T : X \rightarrow X$ with $U(C) = T^{-1}(C)$, $C \in \mathcal{C}$.

If $w \in \{0, 1\}^m$, put $B_w = \{E \in 2^{\{0,1\}^m} : w \in E\}$. Then $\mu_m(B_w) \geq \epsilon$. If $\{w(0), w(1)\}$ is a square line, then any $E \in B_{w(0)} \cap B_{w(1)}$ contains $\{w(0), w(1)\}$, which implies that $\mu_m(E) = 0$. Summing over all such E , we get $\mu_m(B_{w(0)} \cap B_{w(1)}) = 0$. The

sets B_w may of course be viewed as members of \mathcal{B}_m ; doing this we get $\mu(B_w) \geq \epsilon$ and $\mu(B_{w(0)} \cap B_{w(1)}) = 0$ for square lines $\{w(0), w(1)\}$.

If J is a set of words of length m , write $\lambda(J) = \mu_m(\bigcap_{w \in J} B_w)$. By stationarity, $\lambda(Ji) = \lambda(J)$ for $i \in \{0, 1\}$. (Note $Ji \subset F$ if and only if $F \in \{E \subset 2^{\{0,1\}^m} : J \subset E\} * i$.)

For $m \in \mathbf{N}$ and $i \in \{0, 1\}$, let \mathcal{C} be the algebra generated by $\{B_w : w \in \{0, 1\}^m\}$ and let \mathcal{D} be the algebra generated by $\{B_{wi} : w \in \{0, 1\}^m\}$. The stationarity just noted, i.e. $\lambda(Ji) = \lambda(J)$, implies that the map $\mathcal{C} \rightarrow \mathcal{D}$ induced by $B_w \rightarrow B_{wi}$ is an isomorphism. Moreover, it is easy to show that $\mu(B_w)$ is a constant across words of length 1, hence across all words. It follows that by picking an arbitrary set B_{nullword} having this same measure, we can consider the case $m = 0$ simultaneously.

We apply Lemma 3.1 to obtain measure preserving transformations R_{m+1} and S_{m+1} such that $B_{w0} = R_{m+1}^{-1}B_w$ and $B_{w1} = S_{m+1}^{-1}B_w$ for all $w \in \{0, 1\}^m$. It follows that if for a word $w = w_1w_2 \cdots w_m \in \{0, 1\}^m$ we write $Z_w = Z_1Z_2 \cdots Z_m$, where $Z_i = R_i$ if $w_i = 0$ and $Z_i = S_i$ if $w_i = 1$, then $B_w = Z_w^{-1}B$, where $B = B_{\text{nullword}}$.

If $w = w_1w_2 \cdots w_m$ is a fixed word in $\{0, 1\}^m$ and $\alpha \subset \{1, 2, \dots, m\}$, write $w^{(\alpha)}(x)$ for the word $u_1u_2 \cdots u_m$, where $u_i = x$ if $i \in \alpha$ and $u_i = w_i$ otherwise. Finally put $\rho_w(\alpha) = Z_{w^{(\alpha)}(0)}$ and $\sigma_w(\alpha) = Z_{w^{(\alpha)}(1)}$. If $|\alpha| = r^2$, then $\{w^{(\alpha)}(0), w^{(\alpha)}(1)\}$ is a square line and $\mu(B_{w^{(\alpha)}(0)} \cap B_{w^{(\alpha)}(1)}) = 0$. In other words,

$$\mu(Z_{w^{(\alpha)}(0)}^{-1}B \cap Z_{w^{(\alpha)}(1)}^{-1}B) = \mu(\rho_w(\alpha)^{-1}B \cap \sigma_w(\alpha)^{-1}B) = 0.$$

Thus, the proof will be complete if we can establish the following:

THEOREM 3.2. Let $\epsilon > 0$. There exist $m, r \in \mathbf{N}$, a word $w_1w_2 \cdots w_m \in \{0, 1\}^m$ and a set $\alpha \subset \{1, 2, \dots, m\}$ with $|\alpha| = r^2$ such that $\mu(\rho_w(\alpha)^{-1}B \cap \sigma_w(\alpha)^{-1}B) \geq \mu(B)^2 - \epsilon$.

Proof. For $i \in \mathbf{N}$ let $w_{(i)}(x)$ be the variable word consisting of $(i - 1)$ 1s followed by an x . Let $T_i = T_{\{i\}} = \rho_{w_{(i)}}(\alpha)\sigma_{w_{(i)}}(\alpha)^{-1}$. We wish to take products of the T_i , and as they need not commute, order is important. Accordingly, we shall write \prod_i for a product taken in increasing order of i and \prod_i for product taken in decreasing order of i . For $\alpha \in \mathcal{F}$ let

$$T_\alpha = \prod_{i \in \alpha} T_i.$$

Next define unitary operators U_α on $L^2(X)$ by the rule $U_\alpha f(x) = f(T_\alpha x)$. Note that

$$U_\alpha = \prod_{i \in \alpha} U_i.$$

LEMMA 3.3. For $\alpha \in \mathcal{F}$ one has $T_\alpha = \rho_w(\alpha)\sigma_w(\alpha)^{-1}$, where w is a word of $\max \alpha = \{\max j : j \in \alpha\}$ 1s. For $\alpha, \beta \in \mathcal{F}$ with $\alpha < \beta$ one has $T_{\alpha \cup \beta} = T_\alpha T_\beta$ and $U_{\alpha \cup \beta} = U_\beta U_\alpha$.

Proof. Formal. □

Recall Ramsey's theorem [R]: for given $k \in \mathbf{N}$, if the k -element subsets of \mathbf{N} are partitioned into finitely many cells, there exists an infinite set $A \subset \mathbf{N}$, all of whose

k -element subsets belong to the same cell of the partition. A “compact version” (just mimic the proof of the Bolzano-Weierstrass theorem) is as follows: let $k \in \mathbf{N}$ and let $f : \{\alpha \subset \mathbf{N} : |\alpha| = k\} \rightarrow X$, where (X, d) is a compact metric space. One can find a sequence (n_i) along which f converges to some x in the sense that for every $\epsilon > 0$ there is M such that for $M < n_{i_1} < n_{i_2} < \dots < n_{i_k}$, $d(f(\{n_{i_1}, n_{i_2}, \dots, n_{i_k}\}), x) < \epsilon$.

Recall that if \mathcal{H} is a separable Hilbert space then the closed unit ball B_1 of \mathcal{H} is compact and metrizable in the weak topology. Choose by Ramsey’s theorem and the separability of $L^2(X)$ (via a diagonal argument, obtaining convergence for a dense set of functions) a sequence $i_1 < i_2 < \dots$ having the property that for every $k \in \mathbf{N}$,

$$\lim_{n_k > n_{k-1} > \dots > n_1 \rightarrow \infty} U_{\{i_{n_1}, i_{n_2}, \dots, i_{n_k}\}} = P_k$$

exists in the weak operator topology.

LEMMA 3.4. For $k, m \in \mathbf{N}$ one has $P_{k+m} = P_k P_m$.

Proof. Let $f \in B_1$. Using weak continuity of P_k , for any choice of $n_1 < n_2 < \dots < n_{m+k}$ with n_1 far enough out,

$$P_k P_m f \approx P_k U_{\{i_{n_1}, i_{n_2}, \dots, i_{n_m}\}} f$$

and

$$P_{k+m} f \approx U_{\{i_{n_1}, i_{n_2}, \dots, i_{n_{m+k}}\}} f,$$

where we use \approx to denote proximity in a metric for the weak topology on B_1 . Fix n_1, \dots, n_m . For $n_{m+1} < n_{m+2} < \dots < n_{m+k}$, with n_{m+1} far enough out,

$$P_k U_{\{i_{n_1}, i_{n_2}, \dots, i_{n_m}\}} f \approx U_{\{i_{n_{m+1}}, i_{n_{m+2}}, \dots, i_{n_{m+k}}\}} U_{\{i_{n_1}, i_{n_2}, \dots, i_{n_m}\}} f = U_{\{i_{n_1}, i_{n_2}, \dots, i_{n_{m+k}}\}} f.$$

The proof reduces therefore to the triangle inequality. \square

Next recall Hindman’s theorem [H]: let \mathcal{F} denote the family of all finite non-empty subsets of \mathbf{N} , and for $\alpha, \beta \in \mathcal{F}$ write $\alpha < \beta$ if $\max \alpha < \min \beta$. If (α_i) is a sequence in \mathcal{F} with $\alpha_i < \alpha_{i+1}$ then the set of finite unions of the α_i is an *IP ring*. Hindman’s theorem states that for any partition of an IP ring $\mathcal{F}^{(1)}$ into finitely many cells, some cell contains an IP ring $\mathcal{F}^{(2)}$. A compact version: let $g : \mathcal{F} \rightarrow X$ be a function, where (X, d) is compact metric. There exists an IP ring $\mathcal{F}^{(1)}$ and an $x \in X$ such that for any $\epsilon > 0$ there is an $M \in \mathbf{N}$ such that if $\alpha \in \mathcal{F}^{(1)}$ with $\min \alpha > M$ then $d(g(\alpha), x) < \epsilon$. We write in this case

$$\text{IP-lim}_{\alpha \in \mathcal{F}^{(1)}} g(\alpha) = x.$$

Now let $n : \mathcal{F} \rightarrow \mathbf{N}$ be any function satisfying $n(\alpha \cup \beta) = n(\alpha) + n(\beta)$ whenever $\alpha < \beta$ (such functions are called *IP systems*) and choose by compact Hindman an IP ring $\mathcal{F}^{(1)}$ such that

$$\text{IP-lim}_{\alpha \in \mathcal{F}^{(1)}} P_{n(\alpha)} = P \text{ and } \text{IP-lim}_{\alpha \in \mathcal{F}^{(1)}} P_{n(\alpha)^2} = Q$$

exist in the weak operator topology.

LEMMA 3.5. P is an orthogonal projection.

Proof. Since $\|P\| \leq 1$, it suffices to show that $Pf = P^2f$ for f in the unit ball of $L^2(X)$. For all choices $\alpha, \beta \in \mathcal{F}^{(1)}$ with $\alpha < \beta$ and α sufficiently far out,

$$P^2f \approx P_{n(\alpha)}Pf$$

and

$$Pf \approx P_{n(\alpha \cup \beta)}f = P_{n(\alpha)+n(\beta)}f = P_{n(\alpha)}P_{n(\beta)}f.$$

Fix α . For all $\beta \in \mathcal{F}^{(1)}$ sufficiently far out,

$$P_{n(\alpha)}P_{n(\beta)}f \approx P_{n(\alpha)}Pf.$$

The proof reduces therefore to the triangle inequality. □

By the same token, for an appropriate IP ring (continue to call it $\mathcal{F}^{(1)}$),

$$\text{IP-lim}_{\alpha \in \mathcal{F}^{(1)}} P_{kn(\alpha)} = P^{(k)}$$

exists weakly and is an orthogonal projection for all $k \in \mathbf{N}$. Note now that if $P^{(r)}f = f$, so that $P_{rn(\alpha)}f \rightarrow f$ weakly, then since $\|P_{rn(\alpha)}\| \leq 1$, in fact $P_{rn(\alpha)}f \rightarrow f$ strongly as well. It follows now from the triangle inequality that $P_{krn(\alpha)}f \rightarrow f$, i.e. $P^{(kr)}f = f$, for every $k \in \mathbf{N}$. $(P^{(k!)})_{k=1}^\infty$ is, therefore, an increasing sequence of orthogonal projections. Denote by R the limit of this sequence. Note that R is an orthogonal projection, if $Rg = g$ then $\|P^{(k!)}g - g\| \rightarrow 0$ as $k \rightarrow \infty$, and if $Rh = 0$ then $P^{(k)}h = 0$ for all $k \in \mathbf{N}$.

LEMMA 3.6. Q is an orthogonal projection.

Proof. Again, it suffices to show that $Q^2f = Qf$ for f in the unit ball of $L^2(X)$. Fix f and write $g = Rf$, $h = f - g$ (so that $Rg = g$ and $Rh = 0$).

Claim 1: $Q^2g = Qg$. Choose a large k such that $\|P^{(k!)}g - g\| \approx 0$. Now for $\alpha, \beta \in \mathcal{F}^{(1)}$ with $\alpha < \beta$ and α sufficiently far out,

$$Qg \approx P_{n(\alpha \cup \beta)^2}g = P_{(n(\alpha)+n(\beta))^2}g = P_{n(\alpha)^2}P_{n(\beta)^2}P_{2n(\alpha)n(\beta)}g \tag{1}$$

and

$$Q^2g \approx P_{n(\alpha)^2}Qg.$$

Fix such α with the additional property that $k!|n(\alpha)$. (By Hindman's theorem, one may assume in passing to an IP-ring that $n(\alpha)$ is constant modulo $k!$; the additive property $n(\alpha_1 \cup \alpha_2) = n(\alpha_1) + n(\alpha_2)$ ensures that this constant value is idempotent, i.e. 0, under addition modulo $k!$.) Now we have

$$\|P^{(2n(\alpha))}g - g\| \approx 0.$$

Now for $\beta \in \mathcal{F}^{(1)}$ sufficiently far out,

$$P_{n(\alpha)^2} P_{n(\beta)^2} g \approx P_{n(\alpha)^2} Qg$$

and

$$\begin{aligned} & \|P_{2n(\alpha)n(\beta)}g - P^{(2n(\alpha))}g\| \\ & \leq \|P_{2n(\alpha)n(\beta)}g - P_{2n(\alpha)n(\beta)}P^{(2n(\alpha))}g\| + \|P_{2n(\alpha)n(\beta)}P^{(2n(\alpha))}g - P^{(2n(\alpha))}g\| \\ & \leq \|g - P^{(2n(\alpha))}g\| + \|g' - P_{2n(\alpha)n(\beta)}g'\| \approx 0. \end{aligned}$$

(Here $g' = P^{(2n(\alpha))}g'$, so the second summand goes to zero and the first was previously noted to be small.) It follows that

$$\|P_{2n(\alpha)n(\beta)}g - g\| \approx 0.$$

Combining this with (1) we get

$$Qg \approx P_{n(\alpha)^2} P_{n(\beta)^2} g.$$

Claim 1 now follows from the triangle inequality.

Claim 2: $Qh = 0$. Suppose not. We will reach a contradiction by showing that for any $T \in \mathbf{N}$ and $\lambda > 0$, it is possible to choose x_1, x_2, \dots, x_T from the orbit of h such that $\langle x_i, x_j \rangle < \lambda$ and $\langle x_i, Qh \rangle > \frac{\|Qh\|^2}{2}$, $1 \leq i \neq j \leq T$.

We adopt notation $Q_n = P_{n^2}$, so that $Q = \text{IP-lim}_{\alpha \in \mathcal{F}^{(1)}} Q_{n(\alpha)}$ as a weak limit. Note that $\langle Q_{n(\alpha)}h, Qh \rangle > \frac{\|Qh\|^2}{2}$ for all $\alpha \in \mathcal{F}^{(1)}$ sufficiently far out. Let $\alpha_1 < \alpha_2$ be from $\mathcal{F}^{(1)}$ and at least this far out and put $m_1 = n(\alpha_1)$, $m_2 = n(\alpha_2)$. Next choose $\alpha_3 > \alpha_2$ from $\mathcal{F}^{(1)}$ in such a way that letting $m_3 = n(\alpha_3)$ one has

$$\begin{aligned} \langle Q_{m_2}h, P_{2m_2m_3}^*h \rangle & \approx 0, \\ \langle P_{2m_1m_2}Q_{m_1}h, P_{2m_1m_3}^*h \rangle & \approx 0 \text{ and} \\ \langle Q_{m_1+m_2}h, P_{2(m_1+m_2)m_3}^*h \rangle & \approx 0. \end{aligned}$$

(Regarding the first of these, note that α_3 may be chosen so that $\langle Q_{m_2}h, P_{2m_2m_3}^*h \rangle = \langle P_{2m_2m_3}Q_{m_2}h, h \rangle \approx \langle P^{(2m_2)}Q_{m_2}h, h \rangle = \langle Q_{m_2}h, P^{(2m_2)}h \rangle = 0$. The others are similar.) Note the following:

$$\langle Q_{m_1+m_2+m_3}h, Qh \rangle > \frac{\|Qh\|^2}{2}, \langle Q_{m_2+m_3}h, Qh \rangle > \frac{\|Qh\|^2}{2} \text{ and } \langle Q_{m_3}h, Qh \rangle > \frac{\|Qh\|^2}{2}.$$

We now map $\mathbf{N} \times \mathbf{N}$ onto the sequence (i_n) as follows. Let $\pi(1, 1) = i_1$, $\pi(2, 1) = i_2$, $\pi(1, 2) = i_3$, $\pi(3, 1) = i_4$, $\pi(2, 2) = i_5$, $\pi(1, 3) = i_6$, $\pi(4, 1) = i_7$, etc. Write U_{ij} for $U_{\pi(i,j)}$ and for $\alpha \in \mathcal{F}$ define

$$V(\alpha) = \prod_{(i,j) \in (\alpha \times \alpha)} U_{ij}.$$

Let \otimes denote symmetric product, i.e. $\alpha \otimes \beta = (\alpha \times \beta) \cup (\beta \times \alpha)$. For $\alpha < \beta$, we write

$$D_\alpha V(\beta) = \prod_{(i,j) \in (\alpha \otimes \beta)} U_{ij}.$$

Notice that if $\min \beta > 2 \max \alpha$, one has $V(\alpha \cup \beta) = V(\beta) D_\alpha V(\beta) V(\alpha)$. We will write $\alpha \ll \beta$ when this condition is met.

Fix a large number R_0 having the property that if $\{R_0\} < \alpha_1 < \alpha_2 < \alpha_3$ for some $\alpha_i \in \mathcal{F}$ (it is instructive to notice that we do not require $\alpha \in \mathcal{F}^{(1)}$ here) with $|\alpha_i| = m_i$ then

$$\begin{aligned} \langle V(\alpha_1 \cup \alpha_2 \cup \alpha_3)h, Qh \rangle &\approx \langle Q_{m_1+m_2+m_3}h, Qh \rangle, \\ \langle V(\alpha_2 \cup \alpha_3)h, Qh \rangle &\approx \langle Q_{m_2+m_3}h, Qh \rangle, \\ \langle V(\alpha_3)h, Qh \rangle &\approx \langle Q_{m_3}h, Qh \rangle \text{ and} \\ \langle V(\alpha_1 \cup \alpha_2)h, P_{2(m_1+m_2)m_3}^* h \rangle &\approx \langle Q_{m_1+m_2}h, P_{2(m_1+m_2)m_3}^* h \rangle. \end{aligned}$$

Choose $\alpha_1 > \{R_0\}$ with $|\alpha_1| = m_1$ and

$$\langle V(\alpha_1)h, P_{2m_1m_2}^* P_{2m_1m_3}^* h \rangle \approx \langle Q_{m_1}h, P_{2m_1m_2}^* P_{2m_1m_3}^* h \rangle.$$

Now pick $\alpha_2 \gg \alpha_1$ with $|\alpha_2| = m_2$,

$$\begin{aligned} \langle D_{\alpha_1} V(\alpha_2) V(\alpha_1)h, P_{2m_1m_3}^* h \rangle &\approx \langle P_{2m_1m_2} V(\alpha_1)h, P_{2m_1m_3}^* h \rangle \text{ and} \\ \langle V(\alpha_2)h, P_{2m_2m_3}^* h \rangle &\approx \langle Q_{m_2}h, P_{2m_2m_3}^* h \rangle. \end{aligned}$$

Finally, pick $\alpha_3 \gg \alpha_2$ with $|\alpha_3| = m_3$,

$$\begin{aligned} \langle D_{\alpha_1 \cup \alpha_2} V(\alpha_3) V(\alpha_1 \cup \alpha_2)h, h \rangle &\approx \langle P_{2(m_1+m_2)m_3} V(\alpha_1 \cup \alpha_2)h, h \rangle, \\ \langle D_{\alpha_1} V(\alpha_3) D_{\alpha_1} V(\alpha_2) V(\alpha_1)h, h \rangle &\approx \langle P_{2m_1m_3} D_{\alpha_1} V(\alpha_2) V(\alpha_1)h, h \rangle \text{ and} \\ \langle D_{\alpha_2} V(\alpha_3) V(\alpha_2)h, h \rangle &\approx \langle P_{2m_2m_3} V(\alpha_2)h, h \rangle. \end{aligned}$$

Note now the following:

$$\begin{aligned} \langle V(\alpha_1 \cup \alpha_2 \cup \alpha_3)h, V(\alpha_3)h \rangle &= \langle D_{\alpha_1 \cup \alpha_2} V(\alpha_3) V(\alpha_1 \cup \alpha_2)h, h \rangle \\ &\approx \langle P_{2(m_1+m_2)m_3} V(\alpha_1 \cup \alpha_2)h, h \rangle \\ &\approx \langle P_{2(m_1+m_2)m_3} Q_{m_1+m_2}h, h \rangle \\ &= \langle Q_{m_1+m_2}h, P_{2(m_1+m_2)m_3}^* h \rangle \approx 0, \end{aligned}$$

$$\begin{aligned} \langle V(\alpha_1 \cup \alpha_2 \cup \alpha_3)h, V(\alpha_2 \cup \alpha_3)h \rangle &= \langle D_{\alpha_1} V(\alpha_3) D_{\alpha_1} V(\alpha_2) V(\alpha_1)h, h \rangle \\ &\approx \langle P_{2m_1m_3} D_{\alpha_1} V(\alpha_2) V(\alpha_1)h, h \rangle \\ &\approx \langle P_{2m_1m_3} P_{2m_1m_2} V(\alpha_1)h, h \rangle \\ &\approx \langle P_{2m_1m_3} P_{2m_1m_2} Q_{m_1}h, h \rangle \\ &= \langle P_{2m_1m_2} Q_{m_1}h, P_{2m_1m_3}^* h \rangle \approx 0, \end{aligned}$$

$$\begin{aligned}
\langle V(\alpha_2 \cup \alpha_3)h, V(\alpha_3)h \rangle &= \langle D_{\alpha_2} V(\alpha_3) V(\alpha_2)h, h \rangle \\
&\approx \langle P_{2m_2 m_3} V(\alpha_2)h, h \rangle \\
&\approx \langle P_{2m_2 m_3} Q_{m_2} h, h \rangle \\
&= \langle Q_{m_2} h, P_{2m_2 m_3}^* h \rangle \approx 0.
\end{aligned}$$

What we have shown is that for any $\lambda > 0$ there are some $\alpha_1 \ll \alpha_2 \ll \alpha_3$ with

$$\begin{aligned}
\langle V(\alpha_1 \cup \alpha_2 \cup \alpha_3)h, V(\alpha_3)h \rangle &< \lambda, \\
\langle V(\alpha_1 \cup \alpha_2 \cup \alpha_3)h, V(\alpha_2 \cup \alpha_3)h \rangle &< \lambda, \\
\langle V(\alpha_2 \cup \alpha_3)h, V(\alpha_3)h \rangle &< \lambda, \\
\langle V(\alpha_1 \cup \alpha_2 \cup \alpha_3)h, Qh \rangle &> \frac{\|Qh\|^2}{2}, \\
\langle V(\alpha_2 \cup \alpha_3)h, Qh \rangle &> \frac{\|Qh\|^2}{2} \text{ and} \\
\langle V(\alpha_3)h, Qh \rangle &> \frac{\|Qh\|^2}{2}.
\end{aligned}$$

By an elaboration of the same method, one can show, as promised, that for any $T \in \mathbf{N}$ and $\lambda > 0$, it is possible to choose $\alpha_1 \ll \alpha_2 \ll \dots \ll \alpha_T$ such that, letting $x_i = V(\alpha_i \cup \alpha_{i+1} \cup \dots \cup \alpha_T)h$, $1 \leq i \leq T$, $\langle x_i, x_j \rangle < \lambda$ and $\langle x_i, Qh \rangle > \frac{\|Qh\|^2}{2}$, $1 \leq i \neq j \leq T$. As mentioned at the outset, choosing λ small and T large leads to a contradiction. \square

With Lemma 3.6 in hand the proof of Theorem 3.2 is almost complete. Let $\epsilon > 0$. One has

$$\text{IP-lim}_{\alpha \in \mathcal{F}^{(1)}} \int 1_B P_{n(\alpha)} 1_B d\mu = \langle 1_B, Q1_B \rangle = \|Q1_B\|^2.$$

Fix n with $\langle 1_B, P_{n^2} 1_B \rangle > \mu(B)^2 - \frac{\epsilon}{2}$. Let w be a word of necessary length consisting of all 1s. Now for $n_1 < n_2 < \dots < n_{n^2}$ with n_1 far enough out, letting $\alpha = \{i_{n_1}, i_{n_2}, \dots, i_{n_{n^2}}\}$,

$$\begin{aligned}
\mu(\sigma_w(\alpha)^{-1} B \cap \rho_w(\alpha)^{-1} B) &= \mu(\rho_w(\alpha) \sigma_w(\alpha)^{-1} B \cap B) \\
&= \int 1_B \rho_w(\alpha) \sigma_w(\alpha)^{-1} 1_B d\mu \\
&= \langle 1_B, U_\alpha 1_B \rangle \geq \langle 1_B, P_{n^2} 1_B \rangle - \frac{\epsilon}{2} > \mu(B)^2 - \epsilon.
\end{aligned}$$

The proof of Theorem 3.2, and hence of the $k = 2$ case of Conjecture 1.6, is thus complete. \square

References

- [A] T. Austin, Deducing the density Hales-Jewett theorem from an infinitary removal lemma. Available at <http://arxiv.org/abs/0903.1633>
- [BL] V. Bergelson and A. Leibman, Set polynomials and a polynomial extension of the Hales-Jewett theorem, *Ann. Math.* **150** (1999).
- [BLM] V. Bergelson, A. Leibman and R. McCutcheon, Polynomial Szemerédi theorem for countable modules over integral domains and finite fields, *J. d'Analyse Math.* **95** (2005), 243-296.
- [BM] V. Bergelson and R. McCutcheon, Idempotent ultrafilters, multiple weak mixing and Szemerédi's theorem for generalized polynomials, *J. d'Analyse Math.* To appear.
- [F] H. Furstenberg, Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions, *J. d'Analyse Math.* **31** (1977), 204-256.
- [FK1] H. Furstenberg and Y. Katznelson, An IP Szemerédi theorem for computing IP systems and combinatorial theory, *J. d'Analyse Math.* **45** (1985), 117-168.
- [FK2] H. Furstenberg and Y. Katznelson, A density version of the Hales-Jewett theorem, *J. d'Analyse Math.* **57** (1991), 64-119.
- [HJ] A. Hales and R. Jewett, Regularity and positional games, *Trans. Amer. Math. Soc.* **106** (1963), 222-229.
- [H] N. Hindman, Finite sums from sequences within cells of a partition of \mathbf{N} , *J. Combinatorial Theory* (Series A) **17** (1974) 1-11.
- [M] R. McCutcheon, FVIP systems and multiple recurrence, *Israel J. Math.* **146** (2005), 157-188.
- [P] Polymath1, A combinatorial approach to the density Hales-Jewett theorem. Online project at <http://michaelnielsen.org/polymath1/index.php?title=Polymath1>; also D.H.J. Polymath, A new proof of the density Hales-Jewett theorem. Preprint, available online at ArXiv.org:0910.3926.
- [R] F. Ramsey, On a problem of formal logic, *Proc. London Math. Soc.* (2) **30** (1930), 264-286.
- [S] A. Sárközy, On difference sets of integers III, *Acta. Math. Acad. Sci. Hungar.*, **31** (1978) 125-149.