

A new proof of the density Hales-Jewett theorem

D. H. J. Polymath*

October 22, 2009

Abstract

The Hales–Jewett theorem asserts that for every r and every k there exists n such that every r -colouring of the n -dimensional grid $\{1, \dots, k\}^n$ contains a combinatorial line. This result is a generalization of van der Waerden’s theorem, and it is one of the fundamental results of Ramsey theory. The theorem of van der Waerden has a famous density version, conjectured by Erdős and Turán in 1936, proved by Szemerédi in 1975, and given a different proof by Furstenberg in 1977. The Hales–Jewett theorem has a density version as well, proved by Furstenberg and Katznelson in 1991 by means of a significant extension of the ergodic techniques that had been pioneered by Furstenberg in his proof of Szemerédi’s theorem. In this paper, we give the first elementary proof of the theorem of Furstenberg and Katznelson, and the first to provide a quantitative bound on how large n needs to be. In particular, we show that a subset of $\{1, 2, 3\}^n$ of density δ contains a combinatorial line if $n \geq 2 \uparrow\uparrow O(1/\delta^2)$. Our proof is reasonably simple: indeed, it gives what is arguably the simplest known proof of Szemerédi’s theorem.

1 Introduction

1.1 Basic theorem statements

The purpose of this paper is to give the first elementary proof of the density Hales–Jewett theorem. This theorem, first proved by Furstenberg and Katznelson [FK89, FK91], has the same relation to the Hales–Jewett theorem [HJ63] as Szemerédi’s theorem [Sze75] has to van der Waerden’s theorem [vdW27]. Before we go any further, let us state all four theorems. We shall use the notation $[k]$ to stand for the set $\{1, 2, \dots, k\}$. If X is a set and r is a positive integer, then an r -colouring of X will mean a function $\kappa: X \rightarrow [r]$. A subset Y of X is called *monochromatic* if $\kappa(y)$ is the same for every $y \in Y$.

First, let us state van der Waerden’s theorem and Szemerédi’s theorem.

van der Waerden’s Theorem. *For every pair of positive integers k and r there exists N such that for every r -colouring of $[N]$ there is a monochromatic arithmetic progression of length k .*

Szemerédi’s Theorem. *For every positive integer k and every $\delta > 0$ there exists N such that every subset $A \subseteq [N]$ of size at least δN contains an arithmetic progression of length k .*

It is usually better to focus on the *density* $|A|/N$ of a subset $A \subseteq [N]$ rather than on its cardinality, since this gives us a parameter that we can think of independently of N . Szemerédi’s theorem is often referred to as the *density version* of van der Waerden’s theorem.

To state the Hales–Jewett theorem, we need a little more terminology. The theorem is concerned with subsets of $[k]^n$, elements of which we refer to as *points* (or *strings*). Instead of looking for

*<http://michaelnielsen.org/polymath1/index.php?title=Polymath1>

arithmetic progressions, the Hales–Jewett theorem looks for *combinatorial lines*. A combinatorial line in $[k]^n$ is a set of k points $\{x^{(1)}, \dots, x^{(k)}\}$ formed as follows: Given a line *template*, which is a string $\lambda \in ([k] \cup \{\star\})^n$, the associated combinatorial line is formed by setting $x^{(i)}$ to be the point given by changing each “wildcard symbol” ‘ \star ’ in λ to symbol ‘ i ’. For instance, when $k = 3$, $n = 8$, and $\lambda = 13\star\star 221\star$, the associated combinatorial line is the following set of 3 points:

$$\{13112211, 13222212, 13332213\}.$$

(We exclude *degenerate* combinatorial lines, those that arise from templates with no \star ’s. More formal definitions are given in Section 2.1.)

We are now ready to state the Hales–Jewett theorem.

Hales–Jewett theorem. *For every pair of positive integers k and r there exists a positive number $\text{HJ}_k(r)$ such that for every r -colouring of the set $[k]^n$ there is a monochromatic combinatorial line, provided $n \geq \text{HJ}_k(r)$.*

As with van der Waerden’s theorem, we may consider the density version of the Hales–Jewett theorem, where the density of $A \subseteq [k]^n$ is $|A|/k^n$. The following theorem was first proved by Furstenberg and Katznelson [FK91].

Density Hales–Jewett theorem. *For every positive integer k and real $\delta > 0$ there exists a positive number $\text{DHJ}_k(\delta)$ such that every subset of $[k]^n$ of density at least δ contains a combinatorial line, provided $n \geq \text{DHJ}_k(\delta)$.*

We sometimes write “ DHJ_k ” to mean the k case of this theorem. The first nontrivial case, DHJ_2 , is a weak version of Sperner’s theorem [Spe28]; we discuss this further in Section 3. We also remark that the Hales–Jewett theorem immediately implies van der Waerden’s theorem, and likewise for the density versions. To see this, temporarily interpret $[m]$ as $\{0, 1, \dots, m-1\}$ rather than $\{1, 2, \dots, m\}$, and identify integers in $[N]$ with their base- k representation in $[k]^n$. It’s then easy to see that a combinatorial line in $[k]^n$ is a length- k arithmetic progression in $[N]$; specifically, if the line’s template is λ , with $S = \{i : \lambda_i = \star\}$, then the progression’s common difference is $\sum_{i \in S} k^{n-i}$.

In this paper, we give a new, elementary proof of the density Hales–Jewett theorem, achieving quantitative bounds:

Theorem 1.1. *In the density Hales–Jewett theorem, one may take $\text{DHJ}_3(\delta) = 2 \uparrow\uparrow O(1/\delta^2)$. For $k \geq 4$, the bound $\text{DHJ}_k(\delta)$ we achieve is of Ackermann type.*

Here we use the notation $x \uparrow y$ for x^y , $x \uparrow^{(\ell)} y$ for $x \uparrow x \uparrow \dots \uparrow x \uparrow y$ (with ℓ many \uparrow ’s, associating right-to-left), and $x \uparrow\uparrow y$ for $x \uparrow^{(y)} x$.

Another way of phrasing our result is in terms of the number $c_{n,3}$, the cardinality of the largest subset of $[3]^n$ without a combinatorial line. Theorem 1.1 states that $c_{n,3}/3^n \leq O(1/\sqrt{\log^* n})$. The only known lower bound appears in a work concurrent to this one by an overlapping set of authors [Pol09]: therein it is shown that $c_{n,3} = 2, 6, 18, 52, 150, 450$ for $n = 1, 2, 3, 4, 5, 6$, and for large n that $c_{n,3}/3^n \geq \exp(-O(\sqrt{\log n}))$. Generalizing to DHJ_k , the authors show that $c_{n,k}/k^n \geq \exp(-O(\log n)^{1/\lceil \log_2 k \rceil})$, using ideas from recent work on the construction of Behrend [Beh46].

As Furstenberg and Katznelson observed [FK91], given DHJ_k it is not hard to deduce the following extensions of DHJ_k :

- a *multidimensional* version, in which one finds higher-dimensional combinatorial subspaces in the subset (cf. the multidimensional Szemerédi Theorem [FK78]);
- a *probabilistic* version, in which one shows that a randomly chosen combinatorial line (from a suitable distribution) is in the subset with positive probability depending only on k and δ (cf. Varnavides’s extension [Var59] of Szemerédi’s Theorem);
- the combined probabilistic multidimensional version.

In fact, to prove Theorem 1.1 we found it necessary to obtain these extensions in passing. See Section 2 for more detailed statements.

1.2 Some discussion

Why is it interesting to give a new proof of the density Hales–Jewett theorem? There are two main reasons. The first is connected with the history of results and techniques in this area. One of the main benefits of Furstenberg’s proof of Szemerédi’s theorem was that it introduced a technique—ergodic methods—that could be developed in many directions, which did not seem to be the case with Szemerédi’s proof. As a result, several far-reaching generalizations of Szemerédi’s theorem were proved [BL96, FK78, Fur85, FK91], and for a long time nobody could prove them in any other way than by using Furstenberg’s methods. In the last few years that has changed, and a programme has developed to find new and finitary proofs of the results that were previously known only by infinitary ergodic methods; see, e.g., [RS04, NRS06, RS06, RS07b, RS07a, Gow06, Gow07, Tao06, Tao07]. Giving a non-ergodic proof of the density Hales–Jewett theorem was seen as a key goal for this programme, especially since Furstenberg and Katznelson’s ergodic proof seemed significantly harder than the ergodic proof of Szemerédi’s theorem. Having given a purely finitary proof, we are able to obtain explicit bounds for how large n needs to be as a function of δ and k in the density Hales–Jewett theorem. Such bounds could not be obtained via the ergodic methods even in principle, since these proofs rely on the Axiom of Choice. Admittedly, our explicit bounds are not particularly good: tower-type dependence for $k = 3$, and Ackermann-type dependence for $k \geq 4$. Still, Ackermann-type bounds are all that is known even for the multidimensional Szemerédi theorem [Gow07, NRS06], and it is our hope that our work may spur improvements in this direction.

A second reason that a new proof of the density Hales–Jewett theorem is interesting is that it immediately implies Szemerédi’s theorem, and finding a new proof of Szemerédi’s theorem seems always to be illuminating—or at least this has been the case for the four main approaches discovered so far (combinatorial [Sze75], ergodic [Fur77, FKO82], Fourier [Gow01], hypergraph removal [Gow06, Gow07, RS04, NRS06]). In fact, quite surprisingly, the new proof we have discovered is arguably *simpler* than the previous approaches to Szemerédi’s theorem; the most advanced notion we need is that of the total variation distance between discrete probability distributions. It seems that by looking at a more general problem we have removed some of the difficulty. Related to this is another surprise. We started out by trying to prove the first difficult case of the theorem, DHJ₃. The experience of all four of the earlier proofs of Szemerédi’s theorem has been that interesting ideas are needed to prove results about progressions of length 3, but significant extra difficulties arise when one tries to generalize an argument from the length-3 case to the general case. Unexpectedly, it turned out that once we had proved the case $k = 3$ of the density Hales–Jewett theorem, it was straightforward to generalize the argument to the $k \geq 4$ cases.

Before we start working towards the proof of the theorem, we would like briefly to mention that it was proved in a rather unusual “open source” way, which is why it is being published under a pseudonym. The work was carried out by several researchers, who wrote their thoughts, as they had

them, in the form of blog comments at <http://gowers.wordpress.com>. Anybody who wanted to could participate, and at all stages of the process the comments were fully open to anybody who was interested. (Indeed, taking some inspiration from a few of these blog comments, Austin provided *another* new (ergodic) proof of the density Hales–Jewett theorem [Aus09].) This open process was in complete contrast to the usual way that results are proved in private and presented in a finished form. The blog comments are still available, so although this paper is a polished account of the DHJ_k argument, it is possible to read a record of the entire thought process that led to the proof. The constructions of new *lower* bounds for the DHJ_k problem, mentioned in Section 1.1, are being published by a partially overlapping set of researchers [Pol09]. The participants in the project also created a wiki, <http://michaelnielsen.org/polymath1/>, which contains sketches of the arguments, links to the blog comments, and a great deal of related material.

1.3 Outline of the paper

Very briefly, our proof of DHJ_k follows the *density increment method*, as pioneered by Roth [Rot53] in his proof of the $k = 3$ case of Szemerédi’s theorem. Given $A \subseteq [k]^n$ of density δ , we show that either A contains a combinatorial line or else it has a “density increment” within a somewhat structured subset of $[k]^n$ (specifically, an “intersection of *ab*-insensitive sets”). To iterate this increment we need to replace the somewhat structured subset by a copy of $[k]^{n'}$. This is achieved by a second iteration argument, relying on DHJ_{k-1} , which shows that the somewhat structured subsets can essentially be partitioned into copies of $[k]^{n'}$. This double-iteration proof structure was previously used by Shkredov [Shk06a, Shk06b] to obtain strong bounds for the so-called “Corners Problem”, a simplified version of DHJ_3 .

The remainder of the paper is organized as follows:

- §2 Basic definitions, and formal statements of our theorems.
- §3 Proofs for DHJ_2 : Sperner’s theorem, and the Gunderson–Rödl–Sidorenko theorem.
- §4 A detailed outline of our proof of DHJ_3 .
- §5 Definitions of the *equal-slices* distributions, used in the probabilistic DHJ theorem.
- §6 Some technical calculations which will let us pass freely between probability distributions.
- §7 Straightforward deductions of the probabilistic/multidimensional DHJ_k from the basic DHJ_k .
- §8 A key lemma showing how to obtain a density increment on a somewhat structured subset.
- §9 A theorem showing that somewhat structured subsets can be partitioned into large subspaces.
- §10 The proof of our main theorem, deducing DHJ_k from DHJ_{k-1} .

2 Basic concepts and theorem statements

In this section we introduce some basic concepts around the density Hales–Jewett theorem, and also state formally the extensions of DHJ_k that we will prove.

2.1 Points, lines, and line templates

In the density Hales–Jewett theorem the arithmetic properties of $[k] = \{1, 2, \dots, k\}$ play no role. It is only important that this set has k elements; this dictates the length of a combinatorial line. Therefore we can equivalently replace $[k]$ by any other *alphabet* Ω of k *symbols*, treating the elements of Ω^n as *strings*. We will typically use letters a, b, \dots for symbols, x, y, \dots for strings/points, and i, j, \dots for coordinates; e.g., $x_j = b$ means the j th coordinate of string x is symbol b .

We have already seen that it was convenient to use a different alphabet Ω when we took $\Omega = \{0, 1, \dots, k-1\}$ to deduce van der Waerden's theorem from Hales–Jewett. A more interesting observation is that a combinatorial line template over $[k]^n$, i.e. a string in $([k] \cup \{\star\})^n$, is again just a string over an alphabet of size $k+1$. If we use the symbol ' $k+1$ ' in place of ' \star ', we see that *a point in $[k+1]^n$ can be interpreted as a line in $[k]^n$* . For example, the point $13332213 \in [3]^8$ corresponds to the length-2 line $\{11112211, 12222212\} \subset [2]^8$. We introduce the notation $x^{b \rightarrow a}$ for the string formed by changing all occurrences of symbol b to symbol a in string x . Thus a line template $\lambda \in ([3] \cup \{\star\})^n$ corresponds to the line $\{\lambda^{\star \rightarrow 1}, \lambda^{\star \rightarrow 2}, \lambda^{\star \rightarrow 3}\} \subseteq [3]^n$.

Let us give some formal definitions, taking care of the somewhat tedious possibility of degenerate line templates:

Definition 2.1. We say that $x \in \Omega^n$ is a *degenerate string* if it is missing at least one symbol from Ω . A *line template over Ω^n with wildcard symbol c* ($c \notin \Omega$) is a string $\lambda \in (\Omega \cup \{c\})^n$. We say z is a *degenerate line template* if z contains no ' c ' symbols. If λ is nondegenerate, we associate to it the *combinatorial line* $\{\lambda^{c \rightarrow a} : a \in \Omega\} \subseteq \Omega^n$, which has cardinality $|\Omega|$.

We will often blur the distinction between a line template and a line; however, please note that for us a “line” is always nondegenerate, whereas a line template may not be.

2.2 Subspaces

As mentioned, we will find it useful to think of lines over $[k]^n$ as points in $[k+1]^n$, with $k+1$ serving as the wildcard symbol. It's natural then to ask for an interpretation of a point in, say, $[k+2]^n$. This can be thought of as a combinatorial *plane* over $[k]^n$. For example, let's consider $k=3$ and replace $[k+2]$ with $[3] \cup \{\star_1, \star_2\}$. A string in $([3] \cup \{\star_1, \star_2\})^n$ such as

$$\sigma = 13\star_1\star_221\star_1\star_13\star_21$$

(here $n=11$) corresponds to the following 9-point combinatorial plane:

$$\begin{aligned} &\{13112111311, \quad 13212122311, \quad 13312133311, \\ &\quad 13122111321, \quad 13222122321, \quad 13322133321, \\ &\quad 13132111331, \quad 13232122331, \quad 13332133331\} \subset [3]^{11}. \end{aligned}$$

More generally:

Definition 2.2. For $d \geq 1$, a *d -dimensional subspace template over Ω^n with wildcard symbols c_1, \dots, c_d* (distinct symbols not in Ω) is a string $\sigma \in (\Omega \cup \{c_1, \dots, c_d\})^n$. We say σ is a *degenerate template* if σ is missing any of the c_i symbols. If σ is nondegenerate, we associate to it the *d -dimensional combinatorial subspace* $\{\sigma^{(c_1, \dots, c_d) \rightarrow (a_1, \dots, a_d)} : (a_1, \dots, a_d) \in \Omega^d\} \subseteq \Omega^n$, which has cardinality $|\Omega|^d$.

Here we have introduced the extended notation $x^{(c_1, \dots, c_d) \rightarrow (a_1, \dots, a_d)}$ for the string in Ω^n formed by changing all occurrences of symbol c_i to symbol a_i in string x , for all $i \in [d]$.

The *multidimensional density Hales–Jewett theorem* states that dense subsets of $[k]^n$ contain not just lines but combinatorial subspaces:

Multidimensional density Hales–Jewett theorem. *For every real $\delta > 0$ and every pair of positive integers k and d , there exists a positive number $\text{MDHJ}_k(\delta, d)$ such that every subset of $[k]^n$ of density at least δ contains a d -dimensional combinatorial subspace, provided $n \geq \text{MDHJ}_k(\delta, d)$.*

Indeed, as Furstenberg and Katznelson [FK91] showed, it is easy to deduce the multidimensional DHJ_k from the basic DHJ_k:

Proposition 2.3. *For a given k , the multidimensional density Hales–Jewett theorem follows from the density Hales–Jewett theorem.*

Proof. By induction on d . Suppose we have proven the multidimensional DHJ_k theorem for $d - 1$, and let $A \subseteq [k]^n$ have density at least δ . Let $m = \text{MDHJ}_k(\delta/2, d - 1)$, and write a generic string $z \in [k]^n$ as (x, y) , where $x \in [k]^m$, $y \in [k]^{n-m}$. Call a string $y \in [k]^{n-m}$ “good” if $A_y = \{x \in [k]^m : (x, y) \in A\}$ has density at least $\delta/2$ within $[k]^m$. Let $G \subseteq [k]^{n-m}$ be the set of good y ’s. By a simple argument, the density of G within $[k]^{n-m}$ must be at least $\delta/2$; otherwise, A could not have density at least δ in $[k]^n$.

By induction, for any good y the set A_y contains a $(d - 1)$ -dimensional combinatorial subspace. There are at most $M = (k + d - 1)^m$ such subspaces (since each is defined by a template in $[k + (d - 1)]^m$). Hence there must be a particular subspace $\sigma \subseteq [k]^m$ such that the set

$$G_\sigma = \{y \in [k]^{n-m} : (x, y) \in A \ \forall x \in \sigma\}$$

has density at least $(\delta/2)/M$ within $[k]^{n-m}$. Finally, provided $n \geq m + \text{DHJ}_k(\delta/2M)$, we conclude from DHJ_k that G_σ contains a combinatorial line, λ . Now $\sigma \times \lambda \subseteq [k]^n$ is the desired d -dimensional subspace contained in A . \square

The argument in Proposition 2.3 gives a (seemingly) poor bound for large d . We note that in the $k = 2$ case, Gunderson, Rödl, and Sidorenko [GRS99] have established a much superior bound:

Gunderson–Rödl–Sidorenko theorem. *One may take $\text{MDHJ}_2(\delta, d) = (10d)^{d^{2d}} \cdot (1/\delta)^{2^d}$.*

(To see this from [GRS99], combine that paper’s Theorem 4.2 and equation 8, and note that its “ $c(d)$ ” is at most $(10d)^d$.) The authors also showed that their result is fairly tight: they proved that $\text{MDHJ}_2(\delta, d)$ needs to be at least $(1/\delta)^{\frac{2^{d+1}-2}{d}(1-o_\delta(1))}$. This significant improvement for $k = 2$ over the generic Proposition 2.3 is ultimately the reason our quantitative bounds for DHJ₃ are only tower-type, whereas our bounds for the general DHJ_k case are Ackermann-type. In Section 3.2 we slightly sharpen the Gunderson–Rödl–Sidorenko upper bound to $25 \cdot (1/\delta)^{2^d}$; this improvement makes no difference whatsoever to our bounds for DHJ₃, but including it has the virtue of making our overall argument self-contained.

2.2.1 Passing to a subspace

A useful aspect of combinatorial subspaces is that they “look like” the original space. More precisely, a d -dimensional subspace $\{\sigma^{(c_1, \dots, c_d) \rightarrow (a_1, \dots, a_d)} : (a_1, \dots, a_d) \in \Omega^d\} \subseteq \Omega^n$ is isomorphic to Ω^d , in the sense that d' -dimensional subspaces of σ map to d' -dimensional subspaces of Ω^n . This identification is easiest to see when the subspace template σ has exactly one copy of each wildcard symbol c_i . In this case, writing J for the set of d wildcard coordinates, we will call $\sigma \cong \Omega^J$ a *restriction* to the coordinates J (see Definition 4.1).

This isomorphism is useful because it allows us to “pass to subspaces” when looking for a combinatorial line (or subspace) in a subset $A \subseteq [k]^n$. By this we mean the following: Given a d -dimensional subspace σ of $[k]^n$, “passing to σ ” means considering $A' = A \cap \sigma$ as a subset of $[k]^d$. Note that if we manage to find a line (or subspace) contained in A' , this maps back to a line (subspace) contained in A . This is helpful if A' has some kind of improved “structure”; e.g., a higher density (within $[k]^d$). Indeed, finding density increments on subspaces is the basic strategy

for our proof of DHJ_k; see Section 4.2. We have also seen this technique of passing to a restriction already, in the proof of Proposition 2.3.

When passing to a subspace, we will often not make a distinction between A and A' , writing A for both.

2.3 Probabilistic DHJ

To show that a certain combinatorial object exists, the Probabilistic Method suggests choosing it at random. Could this work for the density Hales–Jewett theorem? A difficulty is that it is not immediately clear how one should choose a random combinatorial line. The most obvious way to choose a line over, say, $[3]^n$ is to simply choose a line template $\lambda \in [4]^n$ uniformly at random (taking care of the extremely unlikely chance of a degenerate template). Unfortunately this has no real chance of working because the uniform distribution on lines is not “compatible” with the uniform distribution on points.

Before clarifying this point, let us introduce some notation. We will be concerned with probability distributions on the set Ω^n , where Ω is a finite alphabet. We will use letters such as γ, ξ for generic such distributions, and if $A \subseteq \Omega^n$ we write $\gamma(A) = \Pr_{x \sim \gamma}[x \in A]$ for the γ -density of A . This coincides with our usual notion of a set’s density in the case that γ is the uniform probability distribution on Ω^n . Indeed, we will often be concerned with product distributions on Ω^n ; if π is a distribution on Ω we write $\pi^{\otimes n}$ for the associated product distribution on Ω^n . We reserve μ for the uniform distribution, writing $\mu_\Omega^{\otimes n}$ for the uniform distribution on Ω^n and simply $\mu_k^{\otimes n}$ when $\Omega = [k]$. We may abbreviate this further to simply μ_k or even μ if the context is clear.

Let us now return to the difficulty with choosing lines uniformly at random. To clarify, suppose we choose $\lambda \in [4]^n$ according to the uniform distribution $\mu_4^{\otimes n}$. Then the distribution on the point $\lambda^{4 \rightarrow 1} \in [3]^n$ will be a product distribution $\pi^{\otimes n}$ in which symbol 1 has probability 1/2 under π and symbols 2 and 3 have probability 1/4 each. However this distribution is very “far” from the uniform distribution on $[3]^n$, meaning that even if $\mu_3^{\otimes n}(A) \geq \delta$, the probability that $\lambda^{4 \rightarrow 1} \in [3]^n$ may be a negligibly small function of n .

Even for other natural distributions on lines, it seems inevitable that $\lambda^{4 \rightarrow 1}$ will tend to have more 1’s than 2’s or 3’s, and similarly for $\lambda^{4 \rightarrow 2}, \lambda^{4 \rightarrow 3}$. To evade this difficulty we take inspiration from the probabilistic proof of Sperner’s theorem (see, e.g., [Spe90]) and introduce a new non-uniform, non-product distribution on points in $[k]^n$. We call this distribution the *equal-slices* distribution and denote it by $\bar{\nu}_k^n$. We will discuss $\bar{\nu}_2^n$ in Section 3 on Sperner’s theorem, and its generalization $\bar{\nu}_k^n$ in Section 5. For now we simply explain the name. Thinking of $\Omega = \{0, 1\}$, one draws a string $x \in \{0, 1\}^n$ from $\bar{\nu}_2^n$ as follows: first choose an integer $s \in \{0, 1, \dots, n\}$ uniformly at random; then choose x uniformly at random from the s th “slice” of the discrete cube, i.e., the set of strings with exactly s many 1’s. The more general $\bar{\nu}_k^n$ involves first picking the symbol-counts for the final string $x \in [k]^n$ in a uniformly random manner, then choosing x uniformly at random from the strings with the given symbol-counts.

As we will demonstrate, equal-slices is the natural distribution to use for a probabilistic version of DHJ_k. There is a small catch, which is that a line template $\lambda \sim \bar{\nu}_k^n$ has a very tiny chance of being degenerate. To evade this, we introduce the very closely related distribution $\tilde{\nu}_k^n$ (“equal-nondegenerate-slices”), which is $\bar{\nu}_k^n$ conditioned on λ being a nondegenerate *string* (i.e., having at least one of each symbol in $[k]$). For more details, see Section 5. We then prove the following two theorems:

Theorem 2.4. (*Equal-slices density Hales–Jewett theorem.*¹) *For every positive integer k and real*

¹We write this instead of “equal-nondegenerate-slices density Hales–Jewett theorem” for brevity.

$\delta > 0$ there exists a positive number $\text{EDHJ}_k(\delta)$ such that every $A \subseteq [k]^n$ with $\tilde{\nu}_k^n(A) \geq \delta$ contains a combinatorial line, provided $n \geq \text{EDHJ}_k(\delta)$.

Theorem 2.5. (Probabilistic density Hales–Jewett theorem.) For every positive integer k and real $\delta > 0$ there exists a positive number $\text{PDHJ}_k(\delta)$ and a positive real $\epsilon_k(\delta)$ such that every $A \subseteq [k]^n$ with $\tilde{\nu}_k^n(A) \geq \delta$ satisfies

$$\Pr_{\lambda \sim \tilde{\nu}_{k+1}^n} [\lambda \subseteq A] \geq \epsilon_k(\delta),$$

provided $n \geq \text{PDHJ}_k(\delta)$. Here we are interpreting λ as both a (nondegenerate) line template and as a line.

Theorem 2.4 follows from the density Hales–Jewett theorem by playing around with probability distributions; see Proposition 7.1. Theorem 2.5 follows almost immediately from Theorem 2.4, thanks to some nice properties of the equal-nondegenerate-slices distribution; see Proposition 7.2. Finally, the same two deductions also apply to the multidimensional DHJ, yielding the “equal-slices multidimensional density Hales–Jewett theorem” (with parameter “ $\text{EMDHJ}_k(\delta, d)$ ”) and also the following:

Theorem 2.6. (Probabilistic multidimensional density Hales–Jewett theorem.) For every real $\delta > 0$ and every pair of positive integers k and d , there exists a positive number $\text{PMDHJ}_k(\delta, d)$ and a positive real $\epsilon_k(\delta, d)$ such that every $A \subseteq [k]^n$ with $\tilde{\nu}_k^n(A) \geq \delta$ satisfies

$$\Pr_{\sigma \sim \tilde{\nu}_{k+d}^n} [\sigma \subseteq A] \geq \epsilon_k(\delta, d),$$

provided $n \geq \text{PMDHJ}_k(\delta, d)$. Here we are interpreting σ as both a (nondegenerate) d -dimensional template in $[k+d]^n$ and also as a d -dimensional subspace of $[k]^n$.

3 $k = 2$: Sperner’s theorem and Gunderson–Rödl–Sidorenko

The first nontrivial case of Szemerédi’s theorem is the case of arithmetic progressions of length 3. However, for the density Hales–Jewett theorem even the case $k = 2$ is interesting. DHJ_2 follows from a basic result in extremal combinatorics: Sperner’s theorem. In this section we review a standard probabilistic proof of Sperner’s theorem. Besides suggesting the equal-slices distribution, this proof can also be extended to give the $k = 2$ case of the probabilistic density Hales–Jewett theorem, a key component in our proof of DHJ_3 .

To investigate DHJ_2 it is slightly more convenient to take the alphabet to be $\Omega = \{0, 1\}$. Then a combinatorial line in $\{0, 1\}^n$ is a pair of distinct binary strings x and y such that y can be obtained from x by changing some 0’s to 1’s. If we think of the strings x and y as the indicators of two subsets X and Y of $[n]$, then this is saying that X is a proper subset of Y . Therefore, when $k = 2$ we can formulate the density Hales–Jewett theorem as follows: there exists $\text{DHJ}_2(\delta)$ such that for $n \geq \text{DHJ}_2(\delta)$, if \mathcal{A} is a collection of at least $\delta 2^n$ subsets of $[n]$, then there must exist two distinct sets $X, Y \in \mathcal{A}$ with $X \subset Y$. In the language of combinatorics, this is saying that \mathcal{A} is *not* an antichain. (Recall that an *antichain* is a collection \mathcal{A} of sets such that no set in \mathcal{A} is a proper subset of any other.) Sperner’s theorem gives something somewhat stronger: a precise lower bound on the cardinality of any antichain.

Sperner’s theorem. For every positive integer n , the largest cardinality of any antichain of subsets of $[n]$ is $\binom{n}{\lfloor n/2 \rfloor}$.

As the bound suggests, the best possible example is the collection of all subsets of $[n]$ of size $\lfloor n/2 \rfloor$. (It can be shown that this example is essentially unique: the only other example is to take all sets of size $\lceil n/2 \rceil$, and even this is different only when n is odd.) It is well known that $\binom{n}{\lfloor n/2 \rfloor} 2^{-n} \geq 1/(2\sqrt{n})$ for all n ; hence Sperner's theorem implies that one may take $\text{DHJ}_2(\delta) = 4/\delta^2$. Let us present a standard probabilistic proof of Sperner's theorem (see, e.g., [Spe90]):

Proof. (Sperner's theorem.) Consider the following way of choosing a random subset of $[n]$. First, we choose, uniformly at random, a permutation τ of $[n]$. Next, we choose, uniformly at random and independently of τ , an integer s from the set $\{0, 1, \dots, n\}$. Finally, we set $X = \{\tau(1), \dots, \tau(s)\}$ (where this is interpreted as the empty set if $s = 0$).

Let \mathcal{A} be an antichain. Then the probability that a set X that is chosen randomly in the above manner belongs to \mathcal{A} is at most $1/(n+1)$, since whatever τ is at most one of the $n+1$ sets $\{\tau(1), \dots, \tau(s)\}$ can belong to \mathcal{A} .

However, what we are really interested in is the probability that $X \in \mathcal{A}$ if X is chosen *uniformly* from all subsets of $[n]$. Let us write $\bar{\nu}_2[X]$ for the probability that we choose X according to the distribution defined above, and $\mu_2[X]$ for the probability that we choose it uniformly. Then $\mu_2[X] = 2^{-n}$ for every X , whereas $\bar{\nu}_2[X] = \frac{1}{n+1} \binom{n}{|X|}^{-1}$, since there is a probability $1/(n+1)$ that $s = |X|$, and all sets of size $|X|$ are equally likely to be chosen. Therefore, the largest ratio of $\mu_2[X]$ to $\bar{\nu}_2[X]$ occurs when $|X| = \lfloor n/2 \rfloor$ or $\lceil n/2 \rceil$. In this case, the ratio is $(n+1) \binom{n}{\lfloor n/2 \rfloor} 2^{-n}$.

Since we showed $\bar{\nu}_2(\mathcal{A}) \leq 1/(n+1)$, it follows that $|\mathcal{A}| = \mu_2(\mathcal{A}) 2^n \leq \binom{n}{\lfloor n/2 \rfloor}$, which proves the theorem. \square

As one sees from the proof, it is very natural to consider two different probability distributions on the set of all subsets of $[n]$, or equivalently, on $\{0, 1\}^n$. The first is the uniform distribution μ_2 , which is forced on us by the way the question is phrased. The second is what we called $\bar{\nu}_2$; the reader will note that this is precisely the “equal-slices” distribution $\bar{\nu}_2^n$ described in Section 2.3. After seeing the above proof, one might take the attitude that the “correct” statement of Sperner's theorem is that if \mathcal{A} is an antichain, then $\bar{\nu}_2(\mathcal{A}) \leq 1/(n+1)$, and that the statement given above is a slightly artificial and strictly weaker consequence.

3.1 Probabilistic DHJ(2)

Indeed, what the preceding proof of Sperner's theorem (essentially) establishes is the “equal-slices DHJ₂ theorem”; i.e., that in Theorem 2.4 one may take $\text{EDHJ}_2(\delta) = (1/\delta) - 1$. We say “essentially” because of the small distinction between the equal-slices distribution $\bar{\nu}_2^n$ used in the proof and the equal-nondegenerate-slices distribution $\tilde{\nu}_2^n$ in the statement. It will be convenient in this introductory discussion of Sperner's theorem to casually ignore this. We will introduce $\tilde{\nu}_k^n$ and be more careful about its distinction with $\bar{\nu}_k^n$ in Section 5.

To further bolster the claim that $\bar{\nu}_2^n$ is natural in this context we will show an easy proof of the *probabilistic* DHJ₂ theorem. Looking at the statement of Theorem 2.5, the reader will see it requires defining $\bar{\nu}_3^n$ (or, more precisely, $\tilde{\nu}_3^n$). We will make this definition in the course of the proof.

Lemma 3.1. *For every real $\delta > 0$, every $A \subseteq \{0, 1\}^n$ with $\bar{\nu}_2^n$ -density at least δ satisfies*

$$\Pr_{\lambda \sim \bar{\nu}_3^n} [\lambda \subseteq A] \geq \delta^2.$$

Note that there is no lower bound necessary on n ; this is because a template $\lambda \sim \bar{\nu}_3^n$ may be degenerate.

Proof. As in our proof of Sperner's theorem, let us choose a permutation τ of $[n]$ uniformly at random. Suppose we now choose $s \in \{0, 1, \dots, n\}$ and also $t \in \{0, 1, \dots, n\}$ *independently*. Let $x(\tau, s) \in \{0, 1\}^n$ denote the string which has 1's in coordinates $\tau(1), \dots, \tau(s)$ and 0's in coordinates $\tau(s+1), \dots, \tau(n)$, and similarly define $x(\tau, t)$. These two strings both have the distribution $\bar{\nu}_2^n$, but are not independent.

A key observation is that $\{x(\tau, s), x(\tau, t)\}$ is a combinatorial line in $\{0, 1\}^n$, unless $s = t$ in which case the two strings are equal. The associated line template is $\lambda \in \{0, 1, \star\}^n$, with

$$\lambda_i = \begin{cases} 1 & \text{if } i \leq \min\{s, t\}, \\ \star & \text{if } \min\{s, t\} < i \leq \max\{s, t\}, \\ 0 & \text{if } i > \max\{s, t\}. \end{cases}$$

This gives the *definition* of how to draw $\lambda \sim \bar{\nu}_3^n$ (with alphabet $\{0, 1, \star\}$). We remark that λ is a degenerate template with probability only $1/(n+1)$.

Assuming $\Pr_{x \sim \bar{\nu}_2^n}[x \in A] \geq \delta$, our goal is to show that $\Pr[x(\tau, s), x(\tau, t) \in A] \geq \delta^2$. But

$$\begin{aligned} \Pr[x(\tau, s), x(\tau, t) \in A] &= \mathbf{E}_{\tau} \left[\Pr_{s,t} [x(\tau, s), x(\tau, t) \in A] \right] \\ &= \mathbf{E}_{\tau} \left[\Pr_s [x(\tau, s) \in A] \Pr_t [x(\tau, s) \in A] \right] && \text{(independence of } s, t) \\ &= \mathbf{E}_{\tau} \left[\Pr_s [x(\tau, s) \in A]^2 \right] \\ &\geq \mathbf{E}_{\tau} \left[\Pr_s [x(\tau, s) \in A] \right]^2 && \text{(Cauchy-Schwarz)} \\ &= \Pr_{\tau, s} [x(\tau, s) \in A]^2, \end{aligned}$$

and this equals δ^2 since $x(\tau, s)$ has the distribution $\bar{\nu}_2^n$, completing the proof. \square

Having proved the probabilistic DHJ₂ theorem rather easily, an obvious question is whether we can generalize the proof to $k = 3$. The answer seems to be no; there is no obvious way to generate random length-3 lines in which the points are independent, or even partially independent as in the previous proof. Nevertheless, the equal-slices distribution remains important for our proof of the general case of DHJ_k; in Section 5 we shall introduce both $\bar{\nu}_k^n$ and $\tilde{\nu}_k^n$ and prove some basic facts about them.

3.2 Multidimensional DHJ(2)

As described in Section 2.2, the multidimensional DHJ₂ theorem follows generically from the basic DHJ₂ (i.e., Sperner), using Proposition 2.3. However this technique yields a very poor bound, of the shape $2 \uparrow^{(d-1)} O(1/\delta^2)$. The Gundersen–Rödl–Sidorenko theorem greatly improves on this bound. To conclude this section of the paper, we give a streamlined version of their proof using the ideas in the preceding Lemma 3.1. Our theorem improves their bound slightly, but the improvement makes no difference to the bounds we obtain for DHJ₃. Thus the reader may skip the result with no loss of continuity.

Theorem 3.2. *We may take $\text{MDHJ}_2(\delta, d) = 25 \cdot (1/\delta)^{2^d}$ for $\delta \leq 1/2$. In other words, suppose $A \subseteq \{0, 1\}^n$ has positive uniform density $\mu_2^{\otimes n}(A) = \delta \leq 1/2$, and d is a positive integer. Then A contains a d -dimensional combinatorial subspace, provided $n \geq 25 \cdot (1/\delta)^{2^d}$.*

Proof. Partition $[n] = J_1 \cup J_2 \cup \dots \cup J_{d-1} \cup E$, where $|J_i| = \lfloor n/4^{d-i} \rfloor$; it follows that $|E| \geq (2/3)n$. For a set $J = J_1 \cup J_2 \cup \dots \cup J_i$, we will use the notation \bar{J} for $[n] \setminus J$. We will also write simply μ for the uniform distribution on a set; the set in question will be clear from context.

Let τ be a random permutation of $[n]$ as in Lemma 3.1, but now let s and t be independent Binomial($|J_1|, 1/2$) random variables. Recalling the notation $x(\tau, s)$ from Lemma 3.1, we have that $x(\tau, s)$ and $x(\tau, t)$ are uniformly distributed strings in $\{0, 1\}^{J_1}$. They are not independent, though; in particular, $\{x(\tau, s), x(\tau, t)\}$ is always a combinatorial line in $\{0, 1\}^{J_1}$ unless $s = t$. Let $A_{x(\tau, s)}$ denote $\{y \in \{0, 1\}^{\bar{J}_1} : (x(\tau, s), y) \in A\}$, and similarly for t . If z denotes a uniformly random string in $\{0, 1\}^{\bar{J}_1}$, then

$$\begin{aligned} \mathbf{E}_{\tau, s, t} [\mu(A_{x(\tau, s)} \cap A_{x(\tau, t)})] &= \mathbf{Pr}_{\tau, s, t, z} [(x(\tau, s), z), (x(\tau, t), z) \in A] = \mathbf{E}_{\tau, z} \left[\mathbf{Pr}_{s, t} [(x(\tau, s), z), (x(\tau, t), z) \in A] \right] \\ &= \mathbf{E}_{\tau, z} \left[\mathbf{Pr}_s [(x(\tau, s), z) \in A]^2 \right] \geq \mathbf{E}_{\tau, z} \left[\mathbf{Pr}_s [(x(\tau, s), z) \in A] \right]^2 = \mathbf{Pr}_{\tau, s, z} [(x(\tau, s), z) \in A]^2 = \mathbf{Pr}_{x \sim \mu} [x \in A] = \delta^2, \end{aligned}$$

where on the second line we used the fact that s and t are independent and identically distributed, and we then used Cauchy-Schwarz. The line $\{x(\tau, s), x(\tau, t)\}$ degenerates with probability

$$\mathbf{Pr}[s = t] = \mathbf{Pr}[\text{Binomial}(2|J_1|, 1/2) = |J_1|] \leq \frac{1}{\sqrt{n/4^{d-1}}} = \frac{2^{d-1}}{\sqrt{n}},$$

where the first equality is an easy observation and the inequality uses the estimate

$$\mathbf{Pr}[\text{Binomial}(2[x], 1/2) = [x]] \leq \frac{1}{\sqrt{x}},$$

which is valid for all $x > 0$. Hence

$$\mathbf{E}_{\tau, s, t} [\mu(A_{x(\tau, s)} \cap A_{x(\tau, t)}) \mid s \neq t] \geq \delta^2 - \frac{2^{d-1}}{\sqrt{n}}.$$

We may therefore fix a particular (nondegenerate) combinatorial line $\{x_1^{(0)}, x_1^{(1)}\} \subseteq \{0, 1\}^{J_1}$ and pass to a subset $A' \subseteq \{0, 1\}^{\bar{J}_1}$ such that $(x_1^{(i)}, z) \in A$ for all $i = 0, 1$ and $z \in A'$, and such that $\mu(A') \geq \delta^2 - 2^{d-1}/\sqrt{n}$.

We now repeat this argument for the index set J_2 . This will give us another combinatorial line $\{x_2^{(0)}, x_2^{(1)}\} \subseteq \{0, 1\}^{J_2}$, and let us pass to a subset $A' \subseteq \{0, 1\}^{\bar{J}_1 \cup \bar{J}_2}$ with

$$\mu(A') \geq \left(\delta^2 - \frac{2^{d-1}}{\sqrt{n}} \right)^2 - \frac{2^{d-2}}{\sqrt{n}} \geq \delta^4 - 2\delta^2 \cdot \frac{2^{d-1}}{\sqrt{n}} - \frac{2^{d-2}}{\sqrt{n}} \geq \delta^4 - \frac{2^{d-1}}{\sqrt{n}},$$

where we used $\delta \leq 1/2$. Continuing the argument yields the line $\{x_3^{(0)}, x_3^{(1)}\} \subseteq \{0, 1\}^{J_3}$ and

$$\mu(A') \geq \left(\delta^4 - \frac{2^{d-1}}{\sqrt{n}} \right)^2 - \frac{2^{d-3}}{\sqrt{n}} \geq \delta^8 - 2\delta^4 \cdot \frac{2^{d-1}}{\sqrt{n}} - \frac{2^{d-3}}{\sqrt{n}} \geq \delta^8 - \frac{2^{d-2}}{\sqrt{n}}.$$

Etc. Having repeated the argument $d-1$ times, we obtain a $(d-1)$ -dimensional subspace

$$\sigma = \{x_1^{(0)}, x_1^{(1)}\} \times \{x_2^{(0)}, x_2^{(1)}\} \times \dots \times \{x_{d-1}^{(0)}, x_{d-1}^{(1)}\} \subseteq \{0, 1\}^{J_1 \cup \dots \cup J_{d-1}}$$

along with a subset $A' \subseteq \{0, 1\}^E$, such that $(x, y) \in A$ for all $x \in \sigma$, $y \in A'$, and such that $\mu(A') \geq \delta^{2^{d-1}} - 4/\sqrt{n}$. Now if this last quantity is at least $1/(2\sqrt{|E|})$ then A' contains a combinatorial line

by Sperner’s theorem, and we consequently obtained the required d -dimensional subspace in A . But

$$\frac{1}{2\sqrt{|E|}} \leq \frac{1}{2\sqrt{(2/3)n}} \leq \frac{1}{\sqrt{n}},$$

and even this quantity is at most $\delta^{2^{d-1}} - 4/\sqrt{n}$ provided $n \geq 5^2/\delta^{2^d}$. \square

4 Outline of the proof

In this section we sketch our proof of DHJ₃, which involves reducing it to DHJ₂. The general deduction of DHJ _{$k+1$} from DHJ _{k} is not much more difficult.

4.1 Passing between distributions, via subspaces

According to the statement of DHJ₃, we are given a subset $A \subseteq [3]^n$ with uniform density $\mu_3(A) \geq \delta$, and would like to find a combinatorial line contained in A , assuming n is large enough. It is helpful to think of δ as a fixed explicit constant such as 1% and to think of n as enormous, so that any function $o_n(1)$ is negligible compared with δ .

As we saw with Sperner’s theorem, it is often more natural to think of the DHJ problem under the equal-(nondegenerate)-slices distribution. Our first task therefore is to reduce to proving the equal-slices DHJ₃ Theorem 2.4, by “embedding” the equal-nondegenerate-slices distribution $\tilde{\nu}_3$ in the uniform distribution $\mu_3^{\otimes n}$. To describe this embedding we will need to define more explicitly the notion of *restrictions*:

Definition 4.1. A *restriction* on $[k]^n$ is a pair $(J, x_{\bar{J}})$, where $J \subseteq [n]$ and $x_{\bar{J}} \in [k]^{\bar{J}}$, where $\bar{J} = [n] \setminus J$. It is thought of as a “partial string”, where the \bar{J} -coordinates are “fixed” according to $x_{\bar{J}}$ and the J -coordinates are still “free”. Given a fixing $y_J \in [k]^J$, we write $(x_{\bar{J}}, y_J)$ for the complete composite string in $[k]^n$.

Definition 4.2. Given a set $A \subseteq [k]^n$ and a restriction $(J, x_{\bar{J}})$, we sometimes write $A_{x_{\bar{J}}}$ for the subset of $[k]^J$ defined by $A_{x_{\bar{J}}} = \{y \in [k]^J : (x_{\bar{J}}, y_J) \in A\}$.

Remark 4.3. As mentioned in Section 2.2.1, a restriction $(J, x_{\bar{J}})$ can be identified with a particular kind of combinatorial subspace. Specifically, it corresponds to the $|J|$ -dimensional subspace whose template is $(x_{\bar{J}}, w_J)$, where w_J consists of $|J|$ distinct wildcard symbols. The set $A_{x_{\bar{J}}}$ is renamed “ A ” when we “pass to” this subspace.

Having defined restrictions, let’s warm up by seeing how to embed the uniform distribution in the uniform distribution. (This was implicitly done in our deduction of multidimensional DHJ _{k} from DHJ _{k} , Proposition 2.3.) Suppose we have $\mu_3^{\otimes n}(A) \geq \delta$. Fix an arbitrary subset of the coordinates $J \subset [n]$ with $|J| = r$. Since $\mu_3^{\otimes n}$ is a product distribution, if we draw $x_{\bar{J}} \sim \mu_3^{\otimes \bar{J}}$ and $y_J \sim \mu_3^{\otimes J}$ independently then the composite string $z = (x_{\bar{J}}, y_J)$ is distributed as $\mu_3^{\otimes n}$. Hence

$$\delta \leq \mu_3^{\otimes n}(A) = \Pr_z[z \in A] = \Pr_{x_{\bar{J}}, y_J}[(x_{\bar{J}}, y_J) \in A] = \mathbf{E}_{x_{\bar{J}}} \left[\Pr_{y_J}[(x_{\bar{J}}, y_J) \in A] \right] = \mathbf{E}_{x_{\bar{J}}} \left[\mu_3^{\otimes J}(A_{x_{\bar{J}}}) \right].$$

It follows that there must exist a *particular* substring $x_{\bar{J}}$ for which $\mu_3^{\otimes J}(A_{x_{\bar{J}}}) \geq \delta$ as well. We can then pass to the subspace defined by $(J, x_{\bar{J}})$, and we have constructed a new instance of the original problem: “ n ” has decreased to $r = |J|$, and “ A ” (i.e., $A_{x_{\bar{J}}}$) still has $\mu_3(A) \geq \delta$.

Of course this hasn’t gained us anything, but it illustrates the basic technique of passing to a subspace. We can now explain how to use this technique to pass to a different *distribution*;

namely, equal-nondegenerate-slices. Suppose we construct strings $z \in [3]^n$ as follows: we choose r coordinates $J \subset [n]$ *randomly*, we choose the substring $x_{\bar{J}}$ according to the uniform distribution $\mu_3^{\otimes J}$, and we choose the substring y_J according to equal-nondegenerate-slices $\tilde{\nu}_3^J$. Write $\tilde{\mu}_3^n$ for the resulting distribution on the composite string $z = (x_{\bar{J}}, y_J)$. Now the count of each symbol in a draw from $\mu_3^{\otimes n}$ tends to fluctuate within a range of size roughly \sqrt{n} ; thus if $r \ll \sqrt{n}$, we may hope that the “corrupted distribution” $\tilde{\mu}_3^n$ is still close to the original distribution μ_3^n . In Section 6 we do some technical calculations to show that this is indeed the case; essentially, the total variation distance between distributions $\mu_3^{\otimes n}$ and $\tilde{\mu}_3^n$ is at most $O(r/\sqrt{n})$. Thus by taking, say, $r \approx n^{1/4}$, we may conclude

$$|\mu_3^{\otimes n}(A) - \tilde{\mu}_3^n(A)| \leq O(1/n^{1/4}) \ll \delta \quad \Rightarrow \quad \tilde{\mu}_3^n(A) \gtrsim \delta.$$

Hence

$$\delta \lesssim \tilde{\mu}_3^n(A) = \mathbf{Pr}_{J, x_{\bar{J}}, y_J} [(x_{\bar{J}}, y_J) \in A] = \mathbf{E}_{J, x_{\bar{J}}} \left[\mathbf{Pr}_{y_J} [(x_{\bar{J}}, y_J) \in A] \right] = \mathbf{E}_{J, x_{\bar{J}}} [\tilde{\nu}_3^J(A_{x_{\bar{J}}})],$$

and so again there must exist a particular J and $x_{\bar{J}}$ such that $\tilde{\nu}_3^J(A_{x_{\bar{J}}}) \gtrsim \delta$. If we pass to the associated subspace, we get a set A with *equal-nondegenerate-slices* density at least δ (essentially), and “ n ” has only decreased to $r \approx n^{1/4}$, which is rather affordable (especially in light of the quantitative bounds we ultimately have in Theorem 1.1). Now we may try to find a combinatorial line in this new A .

4.2 The density increment strategy

As we have seen, given a set A of density at least δ , it is not difficult to pass to a subspace and maintain density essentially δ (and we may even switch between probability distributions, if we wish). Our overall strategy for proving DHJ₃ involves passing to subspaces on which A ’s density *increases* by a noticeable amount; specifically, from δ to $\delta + \Omega(\delta^3)$. (Here δ^3 represents δ times the quantity $\delta^2 = \epsilon_2(\delta)$ from Lemma 3.1.) More precisely, we will show the following result:

Theorem 4.4. *Assume $n \geq 2 \uparrow^{(7)}(1000/\delta)$. If $\tilde{\nu}_3^n(A) \geq \delta$, then either A contains a combinatorial line, or we may pass to a subspace of dimension at least $\log^{(7)} n$ on which $\tilde{\nu}_3(A) \geq \delta + \Omega(\delta^3)$.*

Here and throughout $\log n$ denotes $\max\{2, \log_2 n\}$; $\log^{(7)}$ is this function iterated 7 times.

Using such a “density increment” theorem is a well-known strategy [Rot53, Gow01, Shk06a], and yields the equal-slices DHJ₃ theorem as follows: Assuming n is “sufficiently large” to begin with, we either find a combinatorial line or we pass to a $\log^{(7)}(n)$ -dimensional subspace on which A has density at least $\delta + \Omega(\delta^3)$. Repeating this $C/(7\delta^2)$ times, for C a sufficiently large absolute constant, we either find a combinatorial line somewhere along the way, or pass to a $\log^{(C/\delta^2)}(n)$ -dimensional subspace on which A has density at least 2δ . (All the while we are assuming the dimension is still “sufficiently large”.) Schematically,

$$\delta \xrightarrow{C/\delta^2 \text{ many logs}} 2\delta.$$

Iterating this scheme, we get:

$$\delta \xrightarrow{C/\delta^2 \text{ many logs}} 2\delta \xrightarrow{C/(2\delta)^2 \text{ many logs}} 4\delta \xrightarrow{C/(4\delta)^2 \text{ many logs}} 8\delta \longrightarrow \dots$$

Of course the density cannot exceed 1, so eventually we *must* find a combinatorial line, after at most $C/\delta^2 + C/8\delta^2 + C/64\delta^2 + \dots \leq 2C/\delta^2$ many logs. All the while we have assumed that the dimension is sufficiently large that the hypothesis of Theorem 4.4 holds; to ensure this, it certainly

suffices to ensure that the “last” dimension, $\log^{(2C/\delta^2)} n$, is at least $2 \uparrow^{(7)} (1000/\delta)$. And this holds presuming the initial dimension, n , is at least $2 \uparrow\uparrow O(1/\delta^2)$. Thus we obtain the quantitative bound for n given in our main Theorem 1.1 (noting that the fourth-root arising in the previous section’s reduction from DHJ₃ to equal-slices DHJ₃ is indeed negligible).

4.3 Extending length-2 lines to length-3 lines

We will now begin to describe how we establish the density increment Theorem 4.4. First, we will explain a basic strategy for finding lines in a subset $A \subseteq [3]^n$ with $\tilde{\nu}_3(A) = \delta$. We will argue that if the strategy does not work, we can obtain *some kind* of density increment, although not yet the one described in Theorem 4.4.

The strategy begins with what we already know: probabilistic DHJ₂, i.e., Lemma 3.1. At a high level, this can give us a large number of length-2 lines in A , one of which we hope to extend to a length-3 line. There is a difficulty in getting started with probabilistic DHJ₂, however: we only know that $\tilde{\nu}_3(A) = \delta$, whereas Lemma 3.1 requires $\tilde{\nu}_2(A)$ be large. Indeed, since $\tilde{\nu}_3^n$ is supported on nondegenerate strings, we might have $\tilde{\nu}_3(A) \geq \delta$ and yet $\tilde{\nu}_2^n(A) = 0$. To evade this difficulty we will embed $\tilde{\nu}_2$ into $\tilde{\nu}_3^n$ on $n^{1/4}$ random coordinates, in the manner described in Section 4.1. This will let us pass to a subspace on which $\tilde{\nu}_2(A) \gtrsim \delta$. But we need to take some extra care: if $\tilde{\nu}_3(A)$ becomes tiny after this embedding, we’ll have no hope of extending a length-2 line in A into a length-3 line.

Looking at the argument in Section 4.1, we see that for a random restriction $(J, x_{\bar{J}})$ to some $n^{1/4}$ coordinates J , the expected $\tilde{\nu}_2$ -density of A is essentially δ and so is the expected $\tilde{\nu}_3$ -density. What we would like is for both densities to be $\gtrsim \delta$ *simultaneously*. This can be arranged with a density increment trick. The idea is to look at the *variance* of the $\tilde{\nu}_3$ -density of A under the random restriction. There are two cases. First, suppose this variance is somewhat *high*, say at least δ^C for some large constant C . Then it must happen that there is some restriction under which the $\tilde{\nu}_3$ -density of A increases a little above its mean, to at least $\delta + \Omega(\delta^{C-1})$. In this case, we have a density increment which is almost good enough for Theorem 4.4 (we can repeat the present argument some $\text{poly}(1/\delta)$ more times to get the full density increment claimed). Otherwise, the variance is quite *low*. This means that for “almost every” restriction $(J, x_{\bar{J}})$, the $\tilde{\nu}_3$ -density of A is very close to δ . Since we know that the expected $\tilde{\nu}_2$ -density of A is δ , it follows that for a noticeable fraction of restrictions the $\tilde{\nu}_2$ -density is $\gtrsim \delta$. Thus we can find a restriction with both $\tilde{\nu}_2(A) \gtrsim \delta$ and $\tilde{\nu}_3(A) \approx \delta$, as desired.

Having essentially arranged for both $\tilde{\nu}_2(A), \tilde{\nu}_3(A) \geq \delta$, we are in a position to use probabilistic DHJ₂. The following is a key idea in the proof: Suppose we pick a random $z \sim \tilde{\nu}_3^n$. On one hand, we can think of z as a *point*, which will be in A with probability at least δ . On the other hand, we can think of z as a *line template over* $[2]^n$ with wildcard symbol ‘3’; then the probabilistic DHJ₂ tells us that the associated length-2 line is in A with probability at least δ^2 . *If both events happen, we have a length-3 line in A : $\{z^{3 \rightarrow 1}, z^{3 \rightarrow 2}, z\}$.*

To rephrase this, let $L \subseteq [3]^n$ be the set of line templates over $[2]^n$ for which the associated line is in A . Lemma 3.1 implies that $\tilde{\nu}_3(L) \geq \delta^2$. If $L \cap A \neq \emptyset$ then we have found a combinatorial line in A . Unfortunately, we only know $\tilde{\nu}_3(A) \geq \delta$, and there is plenty of room for a density- δ^2 set and a density- δ set to not intersect.

However, if A does not intersect L then we have that

$$\frac{\tilde{\nu}_3(A)}{\tilde{\nu}_3(\bar{L})} \gtrsim \frac{\delta}{1 - \delta^2} \geq \delta + \delta^3,$$

where \bar{L} denotes $[3]^n \setminus L$. In other words, A has a “relative” density increment on the set \bar{L} . Unfortunately, \bar{L} is not a combinatorial subspace of $[3]^n$, so we have not established the density

increment Theorem 4.4. Still, as we will see, \bar{L} is not a completely arbitrary subset; it has just enough structure that we will be able to convert A 's density increment on it to a density increment on a subspace.

4.4 ab -insensitive sets

Let us look more closely at the “shape” of the set \bar{L} . Recall that

$$L = \{z \in [3]^n : z^{3 \rightarrow 1} \in A, z^{3 \rightarrow 2} \in A\},$$

and hence

$$\bar{L} = B_1 \cup B_2, \quad \text{where } B_1 = \{z \in [3]^n : z^{3 \rightarrow 1} \notin A\}, \quad B_2 = \{z \in [3]^n : z^{3 \rightarrow 2} \notin A\}.$$

Consider, e.g., B_1 . It has an important structural property: it is “13-insensitive”.

Definition 4.5. Let $a, b \in \Omega$ be distinct symbols. We say that $C \subseteq \Omega^n$ is *ab-insensitive* if the following condition holds: $x \in C$ iff $x^{a \rightarrow b} \in C$.

The notion of ab -insensitivity also played an essential role in Shelah’s proof [She88] of the Hales–Jewett theorem.

Remark 4.6. The definition is symmetric in a and b . It is perhaps easier to understand the condition as follows: “altering some a ’s to b ’s and some b ’s to a ’s does not affect presence/absence in C ”.

Remark 4.7. Suppose C and C' are ab -insensitive subsets of Ω^n . Then so too are \bar{C} and $C \cap C'$. Further, if σ is a d -dimensional subspace of Ω^n , and we view $C \cap \sigma$ as a subset of $\sigma \cong \Omega^d$, then $C \cap \sigma$ is ab -insensitive.

It is clear from the definition that B_1 is a 13-insensitive set and B_2 is a 23-insensitive set. Let us temporarily pretend that there is no B_1 and we simply have $\bar{L} = B_2$. (This simplification will ultimately prove easy to patch up.) Thus we have a density increment of roughly δ^3 for A on the 23-insensitive set B_2 . To complete the proof of Theorem 4.4, our last task will be to convert this to an equally good density increment on a combinatorial subspace of dimension at least $\log^{(7)} n$. It turns out to be convenient to carry out this task under the uniform distribution, rather than under the equal-slices distribution; we can arrange for this by another distribution-embedding argument.

Sets which are ab -insensitive are very useful, for the following reason: for such sets, we can “boost” DHJ_k to DHJ_{k+1} , and similarly for the multidimensional version. Suppose the 23-insensitive set B_2 has $\mu_3(B_2) \geq \eta$. By “conditioning on the location of the 3’s”, it is easy to find a restriction under which $\mu_2(B_2) \geq \eta$ as well. Hence we can apply the multidimensional DHJ_2 theorem find a “length-2” d -dimensional combinatorial subspace σ contained in B_2 . This σ will only be an isomorphic copy of $\{1, 2\}^d$, not $\{1, 2, 3\}^d$. But now we may use the 23-insensitivity of B_2 to deduce that $z^{2 \rightarrow 3} \in B_2$ for all $z \in \sigma$. Hence the full “length-3” d -dimensional subspace (copy of $\{1, 2, 3\}^d$) is contained in B_2 as well.

4.5 Completing the proof: partitioning ab -insensitive sets

We now come to perhaps the key step in the proof: We show that a 23-insensitive set can be almost completely partitioned into a disjoint union \mathcal{S} of d -dimensional combinatorial subspaces, where $d \approx \log^{(7)} n$. Thus if A has a density increment on the 23-insensitive set B_2 , it must have an equally good density increment on one of the d -dimensional subspaces in \mathcal{S} .

Let us explain this slightly more carefully. Suppose we have some 23-insensitive set $B \subseteq [3]^n$ and we wish to partition it into a collection of disjoint d -dimensional subspaces \mathcal{S} , along with an “error” set E satisfying $\mu_3(E) \leq \delta^3/100$, say. If $\mu_3(B) \leq \delta^3/100$ already, we are done. Otherwise, using the density and 23-insensitivity of B , we use the multidimensional DHJ₂ to find a d -dimensional subspace contained in B . (We will describe later why we may take $d \approx \log^{(7)} n$.)

We can do even better than simply finding a single d -dimensional subspace in B . By an argument similar to Proposition 2.3, we can ensure that there is a set of coordinates $J \subseteq [n]$ with $|J| = m = \text{MDHJ}_2(\delta^3/100, d)$ and a particular d -dimensional subspace $\sigma \subseteq [3]^J$ such that $\sigma \times \{y\} \subseteq B$ for at least some $\tau = \tau(\delta, d) > 0$ fraction of all strings $y \in [3]^{\bar{J}}$. These sets $\sigma \times \{y\}$ are disjoint d -dimensional subspaces that we may remove from B and place into the collection \mathcal{S} . The total probability mass of these subspaces is crudely bounded below by $\tau(\delta, d)3^{-m}$.

We now wish to iterate this argument, but there is a catch; having removed from B the sets $\sigma \times \{y\}$, it is likely no longer 23-insensitive. Fortunately, B 's 23-insensitivity is only spoiled on the m coordinates J :

Definition 4.8. Let $a, b \in \Omega$ be distinct symbols. We say that $C \subseteq \Omega^n$ is *ab-insensitive on $I \subseteq [n]$* if the following condition holds: given a string x , altering some a 's to b 's and some b 's to a 's *within coordinates I* does not affect x 's presence/absence in C . If I is unspecified we take it to be all of $[n]$.

We can thus remove a collection of disjoint d -dimensional subspaces from B of probability mass at least $\tau(\delta, d)3^{-m}$, at the expense of making B 23-insensitive only on coordinates $[n] \setminus J$. We now iterate the argument. Since $|J|$ depends only on δ and d , we may assume that $n \gg |J|$. Thus B still has plenty of 23-insensitive coordinates, and we can find another d -dimensional subspace, localized to some new m coordinates J_2 , which has many extensions in B . Again, we can remove another collection of d -dimensional subspaces from B of mass at least $\tau(\delta, d)3^{-m}$, while only spoiling B 's 23-sensitivity on m more coordinates. By iterating this argument $T(\delta, d) = (3^m/\tau(\delta, d)) \cdot O(\log(1/\delta))$ times, B 's total density will drop below $\delta^3/100$, at which point we may call what remains the error set E , completing the partitioning argument. To ensure that we *can* iterate the argument $T(\delta, d)$ times, we must assume that $n \geq T(\delta, d)m$ so that B never “runs out” of 23-insensitive coordinates along the way.

It remains to calculate how large d can be so that the assumption $n \geq T(\delta, d)m$ is valid. According to the Gunderson-Rödl-Sidorenko theorem (or our version Theorem 3.2), $m = \text{MDHJ}_2(\delta^3/100, d)$ is, roughly speaking, doubly-exponential in d . It follows that $T(\delta, d)m \approx 3^m/\tau(\delta, d)$ is triply-exponential in d . In fact, we need to iterate this argument once more, leading to a six-fold exponential, because \bar{L} was not just the single 23-insensitive set B_2 but was rather $B_1 \cup B_2$. Finally, to cover the slack in various arguments, we ultimately require roughly $n \geq 2 \uparrow^{(7)} d$. Thus we are able to take $d = \log^{(7)} n$ as needed in Theorem 4.4.

5 Equal-slice distributions

In this section we describe the equal-slices distribution $\bar{\nu}_k^n$ and the equal-nondegenerate-slices distribution $\tilde{\nu}_k^n$. It will actually be more convenient to begin with the latter. Before proceeding, we briefly introduce some probability notions.

Definition 5.1. Given a probability distribution γ on a set Ω^n , we write $\min(\gamma) = \min_{x \in \Omega^n} \gamma[x]$.

Definition 5.2. Given probability distributions γ and ξ on Ω^n , their *total variation distance*

$d_{\text{TV}}(\gamma, \xi)$ is defined by

$$d_{\text{TV}}(\gamma, \xi) = \frac{1}{2} \sum_{x \in \Omega^n} |\gamma[x] - \xi[x]| = \max_{A \subseteq \Omega^n} |\gamma(A) - \xi(A)|.$$

If $d_{\text{TV}}(\gamma, \xi) \leq \tau$ we say that γ and ξ are τ -close.

5.1 Equal-nondegenerate-slices

Definition 5.3. Given a vector $s \in \mathbb{N}^k$ whose components sum to n , the associated *slice* of $[k]^n$ is the set

$$\{x \in [k]^n : x \text{ has } s_1 \text{ many 1's, } s_2 \text{ many 2's, } \dots, s_k \text{ many } k\text{'s}\}.$$

We call s the *slice vector*. A slice (vector) is *nondegenerate* when $s_a \neq 0$ for all $a \in [k]$.

In enumerative combinatorics, the slice vectors are known as the “weak k -compositions of n ”; the nondegenerate slice vectors are the “ k -compositions of n ”. It is a simple “dots and bars” combinatorics exercise to show that there are $\binom{n+k-1}{k-1}$ distinct slices, of which $\binom{n-1}{k-1}$ are nondegenerate; see, e.g., [Sta00]. It will be useful for us to see this in a slightly different way:

Definition 5.4. Let $2 \leq k \leq n$. The *equal-nondegenerate-slices* distribution on $[k]^n$, denoted $\tilde{\nu}_k^n$, generates a string $x \in [k]^n$ as follows: First, points q_1, \dots, q_n are arranged on a circle according to a uniformly random permutation. This also forms n “gaps” between the q_i 's. Second, a set of k gaps are chosen, uniformly from the $\binom{n}{k}$ possibilities; points r_1, \dots, r_k are put into the chosen gaps, in random order. This divides the q_i 's into “arcs”, with the “ r_a arc” consists of those q_i 's whose nearest r -point counterclockwise is r_a . Finally, x is defined by setting $x_i = a$ for all i such that q_i is in the r_a arc. See Figure 1 for an example.

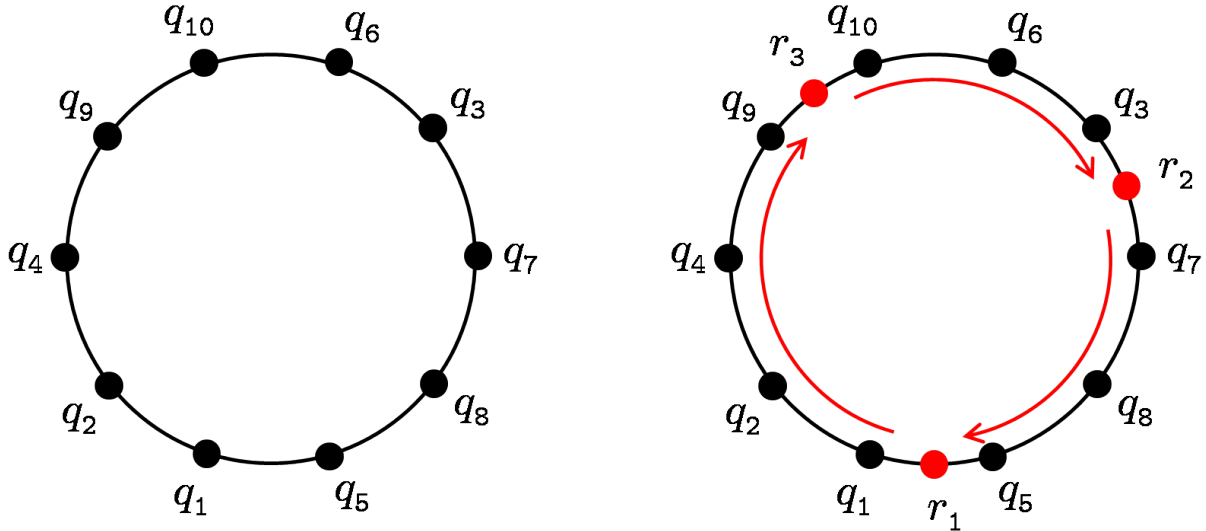


Figure 1: An example of the two stages in drawing from $\tilde{\nu}_3^{10}$. The resulting string is 1131232213.

Remark 5.5. The distribution is symmetric with respect to the k symbols in $[k]$, and thus we may naturally extend it to any alphabet Ω with $|\Omega| = k$.

Remark 5.6. If $x \sim \tilde{\nu}_k^n$ then x always belongs to a nondegenerate slice.

Remark 5.7. Drawing $x \sim \tilde{\nu}_k^n$ is equivalent to first choosing a nondegenerate slice vector s uniformly from the $\binom{n-1}{k-1}$ possibilities, then choosing x uniformly from the associated slice. To see this we view the choices made in Definition 5.4 in an alternate but equivalent way. First, we arrange n indistinguishable q -points on the circle without yet labeling them according to a random permutation. Next, we choose a random permutation ρ on $[k]$. We then choose one of the n gaps uniformly at random and place $r_{\rho(1)}$ there. Following this we choose a subset of $k-1$ of the remaining $n-1$ gaps, uniformly from among the $\binom{n-1}{k-1}$ possibilities. We place points $r_{\rho(2)}, \dots, r_{\rho(k)}$ in the chosen gaps, in clockwise order from the gap $r_{\rho(1)}$ sits in. Finally, we label the q -points according to a random permutation, and the string x is defined.

Consider this procedure just after $r_{\rho(1)}$ has been placed in the first-chosen gap. There is now an obvious bijection between the $\binom{n-1}{k-1}$ choices for the remaining gaps and the resulting slice vector of x . Further, once the remaining gaps are chosen and the slice is fixed, there is a one-to-one correspondence between the permutations of the q -points and the strings in the slice.

The idea of viewing Definition 5.4's steps in an alternate but equivalent way also leads to the following important lemma (which essentially arose already in Lemma 3.1):

Lemma 5.8. *For $k \geq 2$, let $x \sim \tilde{\nu}_k^n$, and let a be chosen uniformly at random from $[k-1]$. Then $x^{k \rightarrow a}$ is distributed as $\tilde{\nu}_{k-1}^n$.*

Proof. Suppose we are drawing $x \sim \tilde{\nu}_k^n$ and have chosen the arrangement of q_1, \dots, q_n and also the k gaps to be filled with r -points. To fill these gaps, we may do the following. First, choose one of the k gaps at random to contain r_k . Next, take the first gap counterclockwise from r_k and let it contain r_a , where $a \sim [k-1]$ is uniformly chosen. Finally, fill the remaining $k-2$ gaps according to a uniform ordering of $[k-1] \setminus \{a\}$. But having chosen x in this way, $x^{k \rightarrow a}$ is obtained simply by “deleting” r_k and interpreting the result as a (correctly distributed) draw from $\tilde{\nu}_{k-1}^n$. \square

We end our discussion of the equal-nondegenerate-slices distribution with a handy lemma (which is not true of “equal-slices” distribution we define next). First, one more piece of notation:

Definition 5.9. Let $x \in [m]^n$, $y \in [k]^m$. We write $y \circ x$ for the string $x^{(1, \dots, m) \rightarrow (y_1, \dots, y_m)} \in [k]^n$. For example, if $n = 6$, $m = 4$, $k = 3$, $x = 4124332$, $y = 3132$, then $y \circ x = 2312331$. This operation is indeed function composition if one thinks of a string in Ω^n as a map $[n] \rightarrow \Omega$.

Lemma 5.10. *Let $x \sim \tilde{\nu}_m^n$ and let $y \sim \tilde{\nu}_k^m$. Then $y \circ x \in [k]^n$ is distributed as $\tilde{\nu}_k^n$.*

Proof. The effect of forming $y \circ x$ can be thought of as follows: First the points q_1, \dots, q_n and r_1, \dots, r_m are randomly arranged as usual to form x . Next, we think of the arcs of q_i 's as being the m objects circularly permuted in the initial stage of drawing y ; we then choose k out of the m gaps between the arcs at random and place points r'_1, \dots, r'_k into them in a random order. Then $z = y \circ x$ is given by setting $z_i = a$ for all i such that q_i is in the r'_a “super-arc”.

But since $\tilde{\nu}_m^n$ is symmetric under permutations of $[m]$, the middle stage in this composite process—i.e., permuting the arcs—is superfluous. Eliminating it, we see that the r'_j 's are ultimately placed by first choosing m out of n gaps at random, and then further choosing k out of these m choices at random. This is equivalent to simply choosing k out of the n gaps at random; hence z is indeed distributed as $\tilde{\nu}_k^n$. \square

5.2 Equal-slices

We now define the closely related variant, the equal-slices distribution:

Definition 5.11. Let $k \geq 2$ and let n be a positive integer. The *equal-slices* distribution on $[k]^n$, denoted $\bar{\nu}_k^n$, generates a string x in a manner similar to that of equal-nondegenerate-slices. The only distinction is that the points r_1, \dots, r_k are placed into random gaps one-by-one, and once a point r_a is placed, it is considered to split the gap into two occupiable gaps. Thus when placing r_a , there are $n + a - 1$ gaps to choose from. As a consequence, some of the resulting r -arcs may be empty of q -points.

Remark 5.12. The process in the definition of $\tilde{\nu}_k^n$ is identical to the process in $\bar{\nu}_k^n$ conditioned on no r_a 's ending up adjacent.

Remark 5.13. It is not hard to see that the process in Definition 5.11 yields a uniformly random (circular) ordering of the $n + k$ points $\{q_1, \dots, q_n, r_1, \dots, r_k\}$. This observation shows that the definition of $\bar{\nu}_k^n$ is symmetric with respect to the symbols $[k]$.

Let's first show that the two distributions $\tilde{\nu}_k^n$ and $\bar{\nu}_k^n$ are very close:

Proposition 5.14. For $2 \leq k \leq n$ we have $d_{\text{TV}}(\tilde{\nu}_k^n, \bar{\nu}_k^n) \leq k(k-1)/n$.

Proof. As $\tilde{\nu}_k^n$ is equivalent to $\bar{\nu}_k^n$ conditioned on the event that no r_a 's end up adjacent, it suffices to show this event occurs with probability at most $k(k-1)/n$. In a draw from $\bar{\nu}_k^n$, the probability that r_a is placed adjacent to one of r_1, \dots, r_{a-1} is at most $2(a-1)/(n+a-1) \leq 2(a-1)/n$. Hence the probability of having any adjacent r_a 's is at most $\sum_{a=1}^k 2(a-1)/n = k(k-1)/n$, as desired. \square

It is also quite helpful to observe that equal-slices is a *mixture of product distributions*.

Definition 5.15. Given k points r_1, \dots, r_k on the the circle of unit circumference, dividing it into arcs, the associated *spacing* is the vector $\nu \in \mathbb{R}_{\geq 0}^k$ defined by setting $\nu[a]$ to be the length of the arc emanating clockwise from r_a . We identify a spacing ν with a probability distribution on $[k]$ in the natural way. We say that ν is a (*uniformly*) *random k -spacing* if it is derived by choosing the points r_1, \dots, r_k independently and uniformly at random on the circle.²

Proposition 5.16. The equal-slices distribution $\bar{\nu}_k^n$ can equivalently be defined as follows: First, draw a random k -spacing ν ; then, draw a string from the product distribution $\nu^{\otimes n}$ on $[k]^n$.

Proof. In the original definition of a draw $x \sim \bar{\nu}_k^n$, the string x only depends on the joint ordering of the $n + k$ points $\{q_1, \dots, q_n, r_1, \dots, r_k\}$ around the circle. As noted in Remark 5.13, this ordering is just a uniform (circular) permutation of $n + k$ objects. Hence we can also obtain it by choosing all $n + k$ points independently from the continuous uniform distribution on the circle of unit circumference (neglecting the probability-0 event that some points coincide). Further, in this continuous distribution, we may choose the r_a 's first, and then the q_i 's next, one-by-one. The choice of r_a 's gives a random k -spacing ν , and then the formation of x by choice of q_i 's is clearly equivalent to drawing x from the product distribution $\nu^{\otimes n}$. \square

²Strictly speaking our proof ceases to be entirely finitary at this point, as we are using continuous probability distributions. This is only to simplify subsequent probability calculations involving the discrete distributions $\bar{\nu}_k^n$ and $\tilde{\nu}_k^n$, and could be eliminated, albeit with additional inconvenience.

Remark 5.17. Drawing $x \sim \bar{\nu}_k^n$ is equivalent to first choosing a slice vector s uniformly from the $\binom{n+k-1}{k-1}$ possibilities (nondegenerate and degenerate) possibilities, then choosing x uniformly from the associated slice. We give a brief justification similar to the one in Remark 5.7: Start with $n+k$ indistinguishable points arranged on a circle. Let ρ be a random permutation on $[k]$, and choose one of the $n+k$ points randomly to be $r_{\rho(1)}$. Next, choose a subset of $k-1$ of the remaining $n+k-1$ points, uniformly at random, and define them be $r_{\rho(2)}, \dots, r_{\rho(k)}$, in clockwise order from the point $r_{\rho(1)}$. The $\binom{n+k-1}{k-1}$ choices here are in bijective correspondence with the possible slice vectors. Finally, labeling the remaining n points q_1, \dots, q_n in a random order yields a uniformly random string from the chosen slice.

It follows from this remark that for a particular string $x \in [k]^n$ with slice vector s , we have

$$\bar{\nu}_k^n[x] = 1 / \binom{n+k-1}{k-1, s_1, \dots, s_k}. \quad (1)$$

Finally, we will occasionally need the following technical bounds:

Proposition 5.18. *If ν is a random k -spacing as in Definition 5.15, then*

$$\Pr[\min(\nu) \leq \theta] \leq k(k-1)\theta.$$

Proof. By a union bound and symmetry, $\Pr[\min(\nu) \leq \theta] \leq k \Pr[\nu[1] \leq \theta]$. The event $\nu[1] \leq \theta$ occurs only if one of r_2, \dots, r_k falls within the arc of length θ clockwise from r_1 . By another union bound, this has probability at most $(k-1)\theta$. \square

Proposition 5.19. *Assume $k \geq 2$, $n \geq 2k^2$. Then we may crudely bound $\min(\bar{\nu}_k^n) \geq k^{-2n}$.*

Proof. Define $\alpha = k(k-1)/n$ and apply the multinomial theorem:

$$(\alpha + k)^{n+k-1} = (\alpha + 1 + \dots + 1)^{n+k-1} = \sum_{\substack{t_0+t_1+\dots+t_k \\ =n+k-1}} \binom{n+k-1}{t_0, t_1, \dots, t_k} \alpha^{t_0} \geq \binom{n+k-1}{k-1, s_1, \dots, s_k} \alpha^{k-1}.$$

Hence we can lower-bound the expression (1) by

$$\frac{\alpha^{k-1}}{(\alpha + k)^{n+k-1}} = \frac{(k-1)^{k-1}}{\left(1 + \frac{k-1}{n}\right)^{n+k-1} n^{k-1} k^n} \geq \frac{(k-1)^{k-1}}{\exp\left(k-1 + \frac{(k-1)^2}{n}\right) n^{k-1} k^n} \geq \frac{(k-1)^{k-1}}{\exp(k) n^{k-1} k^n}, \quad (2)$$

where the first step used the definition of α , the second used $1+x \leq \exp(x)$, and the last used $(k-1)^2/n \leq 1/2 \leq 1$, by assumption. For fixed k and large n we have $n^{k-1} \ll k^n$, and hence one expects the expression on the right in (2) to be at least k^{-2n} . It is a calculus exercise to verify that when $k \geq 2$ and $n \geq 2k^2$, this indeed holds. \square

Since $\tilde{\nu}_k^n$ is just $\bar{\nu}_k^n$ conditioned on nondegenerate strings, we conclude:

Corollary 5.20. *Assume $k \geq 2$, $n \geq 2k^2$. Then for each nondegenerate string $x \in [k]^n$ we have $\tilde{\nu}_k^n[x] \geq k^{-2n}$.*

6 Passing between probability measures

The goal of this section is to work out bounds for the error arising when passing back and forth between the uniform distribution μ_k and the equal-nondegenerate-slices distribution $\tilde{\nu}_k$, as described in Section 4.1. Lemma 6.3 below gives the bounds we need. The reader will not lose much by just reading its statement; the proof is just technical calculations.

Before stating Lemma 6.3 we need some definitions.

Definition 6.1. For $0 \leq p \leq 1$, we say that J is a p -random subset of $[n]$ if J is formed by including each coordinate $i \in [n]$ independently with probability p . Assuming $r \leq n/2$, we say that J is an $[r, 4r]$ -random subset of $[n]$ if J is a p -random subset of $[n]$, $p = 2r/n$, conditioned on $r \leq |J| \leq 4r$.

Definition 6.2. A *distribution family* $(\gamma^m)_{m \in \mathbb{N}}$ over $[k]$ is a sequence of probability distributions, where γ^m is a distribution on $[k]^m$. In this paper the families we consider will either be the equal-(nondegenerate-)slices families $\gamma^m = \tilde{\nu}_k^m$ or $\gamma^m = \bar{\nu}_k^m$, or will be the product distributions based on a single distribution π on $[k]$, $\gamma^m = \pi^{\otimes m}$.

Lemma 6.3 concerns passing from the distribution family (γ^m) to the distribution family (ξ^m) :

Lemma 6.3. *Let $2 \leq \ell \leq k$ and n be integers, and assume $2 \ln n \leq r \leq n/2$. Consider any of the following scenarios for the distribution families (γ^m) and (ξ^m) :*

1. (Uniform to equal-slices.) $\gamma^m = \mu_k^{\otimes m}$ and $\xi^m = \bar{\nu}_\ell^m$;
2. (Equal-slices to product.) $\gamma^m = \bar{\nu}_k^m$ and $\xi^m = \pi^{\otimes m}$, where π is a distribution on $[k]$;
3. (Equal-slices to equal-slices.) $\gamma^m = \bar{\nu}_k^m$ and $\xi^m = \bar{\nu}_\ell^m$.

Let J be an $[r, 4r]$ -random subset of $[n]$, let x be drawn from $[k]^J$ according to $\gamma^{|J|}$, and let y be drawn from $[k]^J$ according to $\xi^{|J|}$. Then the resulting distribution on the composite string $(x, y) \in [k]^n$ has total variation distance from γ^n which is at most $4k \cdot r/\sqrt{n}$.

(Although Lemma 6.3 mentions only the equal-slices distribution, one can replace this with equal-nondegenerate-slices if desired by using Proposition 5.14.) Since $\bar{\nu}_k^n$ is a mixture of product distributions (Proposition 5.16), the main work in proving Lemma 6.3 involves comparing product distributions.

6.1 Comparing product distributions

Definition 6.4. For γ and ξ probability distributions on Ω^n , the χ^2 distance $d_{\chi^2}(\pi, \nu)$ is defined by

$$d_{\chi^2}(\gamma, \xi) = \sqrt{\mathbf{Var}_{x \sim \gamma} \left[\frac{\xi[x]}{\gamma[x]} \right]}.$$

Note that $d_{\chi^2}(\gamma, \xi)$ is *not* symmetric in γ and ξ .

The χ^2 distance is introduced to help us prove the following fact:

Proposition 6.5. *Let π be a distribution on Ω with full support; i.e., $\min(\pi) \neq 0$. Suppose π is slightly mixed with ξ , forming $\hat{\pi}$; specifically, $\hat{\pi} = (1 - p)\pi + p\xi$. Then the associated product distributions $\pi^{\otimes n}, \hat{\pi}^{\otimes n}$ on Ω^n satisfy*

$$d_{\text{TV}}(\pi^{\otimes n}, \hat{\pi}^{\otimes n}) \leq d_{\chi^2}(\pi, \xi) \cdot p\sqrt{n}.$$

Proof. It is a straightforward consequence of Cauchy-Schwarz (see, e.g. [Rei89, p. 101]) that

$$d_{\text{TV}}(\pi^{\otimes n}, \hat{\pi}^{\otimes n}) \leq d_{\chi^2}(\pi, \hat{\pi}) \cdot \sqrt{n},$$

and the identity $d_{\chi^2}(\pi, \hat{\pi}) = p \cdot d_{\chi^2}(\pi, \xi)$ follows easily from the definitions. \square

This can be bounded independently of ξ , as follows:

Corollary 6.6. *In the setting of Proposition 6.5,*

$$d_{\text{TV}}(\pi^{\otimes n}, \hat{\pi}^{\otimes n}) \leq \sqrt{\frac{1}{\min(\pi)} - 1} \cdot p\sqrt{n},$$

Proof. It is easy to check that the distribution ξ maximizing $d_{\chi^2}(\pi, \xi)$ is the one putting all its mass on the x minimizing $\pi[x]$. In this case one calculates $d_{\chi^2}(\pi, \xi) = \sqrt{\frac{1}{\min(\pi)} - 1}$. \square

6.2 Proof of Lemma 6.3

Definition 6.7. Let $0 \leq p \leq 1$ and let (γ^m) , (ξ^m) be distribution families. Drawing from the (p, γ, ξ) -composite distribution on $[k]^n$ entails the following: J is taken to be a p -random subset of $[n]$; x is drawn from $[k]^J$ according to $\gamma^{|J|}$; and, y is drawn from $[k]^J$ according to $\xi^{|J|}$. We sometimes think of this distribution on J , x , and y as just being a distribution on composite strings $z = (x, y) \in [k]^n$.

Note that the distribution described in Lemma 6.3 is very similar to the (p, γ, ξ) -composite distribution, except that it uses an $[r, 4r]$ -random subset rather than a p -random subset. We can account for this difference with a standard Chernoff bound (see, e.g., [AS08, App. A]):

Fact 6.8. *If J is a p -random subset of $[n]$ with $p = 2r/n$ as in Definition 6.1, then $r \leq |J| \leq 4r$ holds except with probability at most $2 \exp(-r/4)$.*

The utility of using p -random subsets in Definition 6.7 is the following observation:

Fact 6.9. *If π and ξ are distributions on $[k]$, thought of also as product distribution families, then the (p, π, ξ) -composite distribution on $[k]^n$ is precisely the product distribution $\hat{\pi}^{\otimes n}$, where $\hat{\pi}$ is the mixture distribution $(1-p)\pi + p\xi$ on $[k]$.*

Because of this, we can use Corollary 6.6 to bound the total variation distance between $\pi^{\otimes n}$ and a composite distribution. We conclude:

Proposition 6.10. *Let π and ξ be any distributions on $[k]$, thought of also as product distribution families. Writing $\tilde{\pi}$ for the (p, π, ξ) -composite distribution on strings in $[k]^n$, we have*

$$d_{\text{TV}}(\pi^{\otimes n}, \tilde{\pi}) \leq \sqrt{\frac{1}{\min(\pi)} - 1} \cdot p\sqrt{n}.$$

Recall that for any $2 \leq \ell \leq k$, the equal-slices distribution $\bar{\nu}_\ell^n$ on m coordinates is a mixture of product distributions $\nu^{\otimes m}$ on $[k]^m$. We can therefore average Proposition 6.10 over ξ to obtain:

Proposition 6.11. *Let π be any distribution on $[k]$, thought of also as a product distribution family. Writing $\tilde{\pi}$ for the $(p, \pi, \bar{\nu}_\ell)$ -composite distribution on strings in $[k]^n$, where $\ell \leq k$, we have*

$$d_{\text{TV}}(\pi^{\otimes n}, \tilde{\pi}) \leq \sqrt{\frac{1}{\min(\pi)} - 1} \cdot p\sqrt{n}.$$

Here we have used the following basic bound, based on the triangle inequality:

Fact 6.12. *Let $(\xi_\kappa)_{\kappa \in K}$ be a family of distributions on Ω^n , let ς be a distribution on K , and let $\bar{\xi}$ denote the associated mixture distribution, given by drawing $\kappa \sim \varsigma$ and then drawing from ξ_κ . Then*

$$d_{\text{TV}}(\gamma, \bar{\xi}) \leq \mathbf{E}_{\kappa \sim \varsigma} [d_{\text{TV}}(\gamma, \xi_\kappa)].$$

If we instead use this fact to average Proposition 6.10 over π , we can obtain:

Proposition 6.13. *Let ξ be any distribution on $[k]$, thought of also as a product distribution family. Writing γ for the $(p, \bar{\nu}_k, \xi)$ -composite distribution on strings in $[k]^n$, we have*

$$d_{\text{TV}}(\bar{\nu}_k^n, \gamma) \leq (2k - 1)p\sqrt{n}.$$

Proof. Thinking of $\bar{\nu}_k^n$ as the mixture of product distributions $\nu^{\otimes n}$, where ν is a random k -spacing, Fact 6.12 and Proposition 6.10 imply

$$d_{\text{TV}}(\bar{\nu}_k^n, \gamma) \leq \mathbf{E}_\nu \left[\sqrt{\frac{1}{\min(\nu)} - 1} \right] \cdot p\sqrt{n}.$$

We can upper-bound the expectation above by

$$\begin{aligned} \mathbf{E}_\nu \left[\sqrt{\frac{1}{\min(\nu)}} \right] &= \int_0^\infty \mathbf{Pr}_\nu \left[\sqrt{\frac{1}{\min(\nu)}} \geq t \right] dt = \int_0^\infty \mathbf{Pr}_\nu [\min(\nu) \leq 1/t^2] dt \\ &\leq k + \int_k^\infty \mathbf{Pr}_\nu [\min(\nu) \leq 1/t^2] dt \leq k + \int_k^\infty (k(k-1)/t^2) dt = 2k - 1, \end{aligned}$$

where in the second-to-last step we used Proposition 5.18. \square

Averaging now once more in the second component, we obtain the following:

Proposition 6.14. *Let $2 \leq \ell \leq k$ and let γ' denote the $(p, \bar{\nu}_k, \bar{\nu}_\ell)$ -composite distribution on strings in $[k]^n$. Then*

$$d_{\text{TV}}(\bar{\nu}_k^n, \gamma') \leq (2k - 1)p\sqrt{n}.$$

We can now obtain the proof of Lemma 6.3:

Proof. The claimed bound of $4k \cdot r/\sqrt{n}$ for each of the three scenarios essentially follows from Propositions 6.11, 6.13, and 6.14, taking $p = 2r/n$. These propositions each imply a total variation bound of $(4k - 2) \cdot r/\sqrt{n}$. (In the case of the first scenario, uniform to equal-slices, Proposition 6.11 gives the bound

$$2\sqrt{\frac{1}{\min(\mu_k)} - 1} \cdot r/\sqrt{n} = 2\sqrt{k-1} \cdot r/\sqrt{n},$$

which is indeed at most $(4k - 2) \cdot r/\sqrt{n}$.) However we need to account for conditioning on $r \leq |J| \leq 4r$. By Fact 6.8, this conditioning increases the total variation distance by at most $2\exp(-r/4)$. Using the lower bound $r \geq 2\ln n \geq 1$ from the lemma's hypothesis, this increase is at most $2/\sqrt{n} \leq 2r/\sqrt{n}$, completing the proof. \square

7 Easy reductions

In this section we collect some easy reductions between various DHJ problems for use in our main proof.

Proposition 7.1. *For a given $k \geq 2$, the DHJ_k theorem is equivalent to the equal-slices DHJ_k theorem.*

Proof. The fact that DHJ follows from equal-slices DHJ was sketched in Section 4.1; we give the full proof here, using Lemma 6.3.1 and Proposition 5.14. The other direction has a virtually identical proof using Lemma 6.3.2 and we omit it. We do not bother to optimize the bounds in this reduction, as they have a negligible effect on the overall proof.

Suppose then that we have proven the equal-slices DHJ_k . Let $\mu_k^{\otimes n}(A) \geq \delta$. We use Lemma 6.3.1, taking $r = n^{1/4}$, letting J be an $[r, 4r]$ -random subset of $[n]$, $x \sim \mu_k^{\otimes \bar{J}}$, $y \sim \bar{\nu}_k^J$. (Note that $2 \ln n \leq r \leq n/2$ provided n is at least a certain universal constant.) According to the lemma, the resulting distribution on the composite string (x, y) is τ -close to $\mu_k^{\otimes n}$, where $\tau = 4k \cdot r / \sqrt{n} = 4k/n^{1/4}$. This is at most $\delta/3$ provided that $n \geq (12k/\delta)^4$. In that case we conclude

$$\mathbf{E}_{J,x}[\bar{\nu}_k^J(A_x)] \geq \delta - \delta/3$$

and so there must exist a restriction (J, x) with $|J| \geq r$ and $\bar{\nu}_k^J(A_x) \geq 2\delta/3$. By Proposition 5.14 the distributions $\bar{\nu}_k^J$ and $\tilde{\nu}_k^J$ have total variation distance at most $k(k-1)/r \leq k^2/n^{1/4}$, and this quantity is also at most $\delta/3$ provided $n \geq (3k^2/\delta)^4$. In that case $\tilde{\nu}_k^J(A_x) \geq 2\delta/3 - \delta/3 = \delta/3$. Finally, if $r \geq \text{EDHJ}_k(\delta/3)$ then we may use equal-slices DHJ_k to find a combinatorial line in A_x , and hence in A . This holds provided $n \geq \text{EDHJ}_k(\delta/3)^4$. Thus we may take $\text{DHJ}_k(\delta) = \max\{\text{EDHJ}_k(\delta/3)^4, O(k^8/\delta^4)\}$. \square

The deduction of probabilistic multidimensional DHJ from the equal-slices version uses the composition property of equal-nondegenerate-slices, Lemma 5.10; we remark that we only need the $d = 1$ case for our overall proof:

Proposition 7.2. *For a given k , the probabilistic multidimensional DHJ theorem follows from the equal-slices multidimensional DHJ theorem.*

Proof. Assuming the EMDHJ_k theorem, write $m = \max\{\lceil \text{EMDHJ}_k(\delta/2, d) \rceil, k\}$. We will show that one can take $\text{PMDHJ}_k(\delta, d) = m$ and $\epsilon_k(\delta, d) = (\delta/2)(k+d)^{-2m}$. Let $x \sim \tilde{\nu}_m^n$ and $y \sim \tilde{\nu}_k^m$. By Lemma 5.10, $y \circ x \sim \tilde{\nu}_k^n$. Hence if $A \subseteq [k]^n$ satisfies $\tilde{\nu}_k^n(A) \geq \delta$, we get that $\tilde{\nu}_k^m(A_x) \geq \delta/2$ with probability at least $\delta/2$ over x , where A_x denotes $\{z \in [k]^m : z \circ x \in A\}$. Call such x 's "good". By our choice of m , the EMDHJ_k theorem implies that for all good x we have that A_x contains a d -dimensional subspace. In fact, we may assume these subspaces have templates $\sigma \in [k+d]^m$ which are nondegenerate strings, meaning that they contain at least one copy of each symbol in $[k+d]$. (To ensure this one can remove from each A_x all degenerate strings; this does not change $\tilde{\nu}_k^m(A_x)$. Then any subspace found in A_x must have a template which is a nondegenerate string.) Thus by Corollary 5.19, a randomly drawn subspace $w \sim \tilde{\nu}_{k+d}^m$ is in A_x with probability at least $(k+d)^{-2m}$. We conclude that with probability at least $(\delta/2)(k+d)^{-2m} = \epsilon_k(\delta, d)$ over the choice of x and w , the d -dimensional subspace $w \circ x$ is in A . This completes the proof, since $w \circ x$ is distributed as $\tilde{\nu}_{k+d}^n$ by Lemma 5.10. \square

As we saw in Section 3.1, there is an elementary proof of probabilistic DHJ(2). For completeness, we observe here that Lemma 3.1 and Proposition 5.14 straightforwardly yield the following:

Lemma 7.3. *One may take $\text{PDHJ}_2(\delta) = 12/\delta^2$ and $\epsilon_2(\delta) = \delta^2/2$.*

The following observation was sketched in Section 4.4:

Proposition 7.4. *The multidimensional DHJ_k theorem for ab-insensitive sets follows from the multidimensional DHJ_{k-1} theorem. More precisely, let $k \geq 3$, $d \geq 1$, and assume we have established the multidimensional DHJ_{k-1} theorem. Suppose $B \subseteq [k]^n$ is ab-insensitive on $I \subseteq [n]$ and has $\mu_k^{\otimes n}(B) \geq \delta$. Then B contains a d -dimensional subspace, provided $|I| \geq 48\text{MDHJ}_{k-1}(\delta/2, d)$.*

Proof. We may assume without loss of generality that $b = k$, by renaming symbols. We may also assume $I = [n]$ without loss of generality; this is because we can always pass to a restriction $(I, x_{\bar{I}})$ on which B has uniform density at least δ . So from now on we assume B is an ak -insensitive subset of $[k]^n$, with $n \geq 48\text{MDHJ}_{k-1}(\delta/2, d)$.

Think of conditioning a draw from $\mu_k^{\otimes n}$ on the location of the symbols k . In other words, draw $z \sim \mu_k^{\otimes n}$ as follows: first choose a $(1 - 1/k)$ -random subset $J \subseteq [n]$, form the restriction $(J, x_{\bar{J}})$ where $x_{\bar{J}}$ is the all- k 's substring, and finally draw $y \sim \mu_{k-1}^{\otimes J}$ and take $z = (x_{\bar{J}}, y)$. We have $\mu_k(B) = \mathbf{E}_J[\mu_{k-1}(B_{x_{\bar{J}}})] \geq \delta$ (where we may think of $B_{x_{\bar{J}}} \subseteq [k-1]^J$). We also have $\mathbf{E}[|J|] = (1 - 1/k)n$, and $\mathbf{Pr}[|J| < n/2] \leq \exp(-n/48)$ by a standard Chernoff bound. Thus there must exist a particular restriction $(J, x_{\bar{J}})$ where $|J| \geq n/2 \geq \text{MDHJ}_{k-1}(\delta/2, d)$ and $\mu_{k-1}^{\otimes J}(B_{x_{\bar{J}}}) \geq \delta - \exp(-n/48) \geq \delta/2$. (Here we used $n/48 \geq \ln(2/\delta)$, which holds because $\text{MDHJ}_{k-1}(\delta/2, d) \geq \text{MDHJ}_2(\delta/2, 1) \geq (2/\delta)^2$, by Sperner's theorem.) By the multidimensional DHJ_{k-1} theorem, $B_{x_{\bar{J}}}$ contains a nondegenerate d -dimensional subspace of $[k-1]^J$. Let us write the template for this subspace as $\sigma \in ([k-1] \cup \{k+1, \dots, k+d\})^J$, meaning that $\sigma^{(k+t) \rightarrow \ell} \in B_{x_{\bar{J}}}$ for each $t \in [d]$ and $\ell \in [k-1]$.

We can extend σ from coordinates J to all of $[n]$ in the natural way, filling in the \bar{J} coordinates with k 's, and the resulting template has $\sigma^{(k+t) \rightarrow \ell} \in B$ for all $\ell \in [k-1]$. But B is ak -sensitive by assumption, so each string $\sigma^{(k+t) \rightarrow k}$ is in B because $\sigma^{(k+t) \rightarrow a}$ is. Thus σ is a nondegenerate d -dimensional subspace template in the usual sense, with all its strings in B . \square

8 Line-free sets correlate with insensitive set intersections

The essence of our proof of DHJ_k begins with this section. Herein we make rigorous the argument described in Section 4.3: we show that a dense set A either contains a combinatorial line, or it has slight correlation with an intersection of ab -insensitive sets. We begin with a reduction to the case where the set is dense under both $\bar{\nu}_k$ and $\bar{\nu}_{k-1}$.

Lemma 8.1. *Let $k \geq 3$ and let $A \subseteq [k]^n$ satisfy $\bar{\nu}_k(A) \geq \delta$. Supposing θ satisfies*

$$8k \frac{\ln n}{\sqrt{n}} \leq \theta \leq \delta,$$

there exists a restriction $(J, x_{\bar{J}})$ with $|J| \geq (\theta/4k)\sqrt{n}$ such that one of the following two conditions holds:

1. $\bar{\nu}_k^J(A_{x_{\bar{J}}}) \geq \delta + \theta$; or,
2. both $\bar{\nu}_k^J(A_{x_{\bar{J}}}) \geq \delta - 3\theta^{1/4}\delta^{1/2}$ and $\bar{\nu}_{k-1}^J(A_{x_{\bar{J}}}) \geq \delta - 3\theta^{1/2}$.

Proof. Let $r = (\theta/4k)\sqrt{n}$. As in Lemma 6.3, let J be an $[r, 4r]$ -random subset of $[n]$ and let x be drawn from $\bar{\nu}_k^J$. If $\mathbf{E}_{J,x}[\bar{\nu}_k(A_x)^2] \geq (\delta + \theta)^2$, then there must exist some restriction (J, x) with $|J| \geq r$ and $\bar{\nu}_k^J(A_x) \geq \delta + \theta$. In this case, conclusion 1 above holds. We henceforth assume

$$\mathbf{E}_{J,x}[\bar{\nu}_k(A_x)^2] < (\delta + \theta)^2 \tag{3}$$

and show that conclusion 2 holds.

Since $\bar{\nu}_k(A) \geq \delta$, two applications of Lemma 6.3.3 yield

$$\mathbf{E}_{J,x}[\bar{\nu}_k(A_x)] \geq \delta - 4k \cdot r/\sqrt{n} \geq \delta - \theta, \quad (4)$$

$$\mathbf{E}_{J,x}[\bar{\nu}_{k-1}(A_x)] \geq \delta - 4k \cdot r/\sqrt{n} \geq \delta - \theta, \quad (5)$$

where we've used our definition of r (and the hypothesis $\theta \geq (8k \ln n)/\sqrt{n}$ ensures that $r \geq 2 \ln n$ as required for the lemma). Combining (3), (4) gives

$$\mathbf{Var}_{J,x}[\bar{\nu}_k(A_x)] < (\delta + \theta)^2 - (\delta - \theta)^2 = 4\theta\delta.$$

Thus Chebyshev's inequality implies that except with probability at most $\theta^{1/2}$ over the choice of (J, x) , we have that $\bar{\nu}_k(A_x)$ is within $\theta^{-1/4} \cdot (4\theta\delta)^{1/2} = 2\theta^{1/4}\delta^{1/2}$ of its expectation. When this happens,

$$\bar{\nu}_k(A_x) \geq \delta - \theta - 2\theta^{1/4}\delta^{1/2} \geq \delta - 3\theta^{1/4}\delta^{1/2} \quad (6)$$

(using $\theta \leq \delta$). On the other hand, from (5) we immediately deduce that with probability at least $2\theta^{1/2}$ over the choice of (J, x) we have

$$\bar{\nu}_{k-1}(A_x) \geq \delta - \theta - 2\theta^{1/2} \geq \delta - 3\theta^{1/2}. \quad (7)$$

Thus with probability at least $2\theta^{1/2} - \theta^{1/2} > 0$ over the choice of (J, x) , we have both (6), (7), establishing conclusion 2 of the lemma. \square

We now come to the main theorem in this section, deducing a key step in the proof of DHJ_k from the probabilistic DHJ_{k-1} theorem.

Theorem 8.2. *Let $k \geq 3$ and suppose we have established the probabilistic DHJ_{k-1} theorem. Further suppose $A \subseteq [k]^n$ has $\tilde{\nu}_k(A) \geq \delta_k$ and $\tilde{\nu}_{k-1}(A) \geq \delta_{k-1}$. Finally, assume $n \geq \text{PDHJ}_{k-1}(\delta_{k-1})$. Then either A contains a combinatorial line (of length k), or there exists $B \subseteq [k]^n$ with $\tilde{\nu}_k(B) \geq \delta_k$ and*

$$\frac{\tilde{\nu}_k(A \cap B)}{\tilde{\nu}_k(B)} \geq \delta_k + \delta_k \cdot \epsilon_{k-1}(\delta_{k-1})$$

such that

$$B = B_1 \cup B_2 \cup \dots \cup B_{k-1},$$

where for each $a \in [k-1]$, B_a is ak -insensitive.

Proof. For $a \in [k-1]$, let $L_a = \{x \in [k]^n : x^{k-a} \in A\}$, an ak -insensitive set. Let $L = L_1 \cap L_2 \cap \dots \cap L_{k-1}$. In other words, L is the set line templates over $[k-1]^n$ (possibly degenerate) whose corresponding lines are entirely in A . Let $L' \subseteq L$ be the nondegenerate such line templates; i.e., $L' = \{x \in L : \exists j \in [n] \text{ s.t. } x_j = k\}$. Note that $\tilde{\nu}_k(L \setminus L') = 0$.

The key observation is that if $A \cap L' \neq \emptyset$, we have a (nondegenerate) combinatorial line of length k entirely in A . So assume otherwise; hence $A \subseteq \bar{L}' = [k]^n \setminus L'$. Since $\tilde{\nu}_{k-1}(A) \geq \delta_{k-1}$ and $n \geq \text{PDHJ}_{k-1}(\delta_{k-1})$, the probabilistic DHJ_{k-1} theorem implies that $\tilde{\nu}_k(L') = \tilde{\nu}_k(L) \geq \epsilon_{k-1}(\delta_{k-1})$. It follows that

$$\frac{\tilde{\nu}_k(A \cap \bar{L}')}{\tilde{\nu}_k(\bar{L}')} = \frac{\tilde{\nu}_k(A \cap \bar{L}')}{\tilde{\nu}_k(\bar{L}')} = \frac{\tilde{\nu}_k(A)}{\tilde{\nu}_k(L')} \geq \frac{\delta_k}{1 - \epsilon_{k-1}(\delta_{k-1})} \geq \delta_k + \delta_k \cdot \epsilon_{k-1}(\delta_{k-1}).$$

The proof is completed by taking $B = \bar{L}'$; this has $\tilde{\nu}_k(B) = \tilde{\nu}_k(\bar{L}') \geq \delta_k$ because $\bar{L}' \supseteq A$, and has the required form as a union of insensitive sets because each \bar{L}'_a is ak -insensitive. \square

We will prefer to obtain this relative density increment under the uniform distribution, rather than under equal-nondegenerate-slices:

Lemma 8.3. *Let $k \geq 2$ and $n \geq k^2$ be integers. Suppose $A \subseteq B \subseteq [k]^n$ satisfy $\tilde{\nu}_k(B) = \beta > 0$ and $\tilde{\nu}_k(A) \geq \delta \cdot \tilde{\nu}_k(B)$, where $0 < \delta < 1$. Let $0 < \eta < \delta$ be a parameter, define $r = (\beta\eta/15k)\sqrt{n}$, and assume $2 \ln n \leq r \leq n/2$. Then there exists a restriction $(J, x_{\overline{J}})$ with $|J| \geq r$ under which $\mu_k(B_{x_{\overline{J}}}) \geq (\eta/3)\beta$ and $\mu_k(A_{x_{\overline{J}}}) \geq (\delta - \eta)\mu_k(B_{x_{\overline{J}}})$.*

Proof. Let J be an $[r, 4r]$ -random subset of $[n]$, let $x \sim \overline{\nu}_k^J$, and let $y \sim \mu_k^J$. From Lemma 6.3.2, the distribution on the composite string (x, y) is $(4kr/\sqrt{n})$ -close to $\overline{\nu}_k^n$. Further, from Proposition 5.14 the total variation distance between $\overline{\nu}_k^n$ and $\tilde{\nu}_k^n$ is at most $k(k-1)/n \leq kr/\sqrt{n}$ (using $n \geq k^2$). By choice of r , the combined distance bound $5kr/\sqrt{n}$ is at most $(\eta/3)\beta$. We therefore have

$$\mathbf{E}_{J,x}[\mu_k(A_x)] \geq \delta\beta - (\eta/3)\beta, \quad \mathbf{E}_{J,x}[\mu_k(B_x)] \leq \beta + (\eta/3)\beta. \quad (8)$$

Let H be the event that $\mu_k(B_x) \geq (\eta/3)\beta$. Since we also have $\mathbf{E}_{J,x}[\mu_k(B_x)] \geq \beta - (\eta/3)\beta \geq (\eta/3)\beta$, it follows that $\mathbf{Pr}[H] > 0$. Now on one hand, since $A_x \subseteq B_x$ always holds we have

$$\mathbf{E}_{J,x}[\mu_k(A_x)] \leq \mathbf{E}_{J,x}[\mu_k(A_x) \mid H] \mathbf{Pr}[H] + (\eta/3)\beta.$$

On the other hand, we have

$$\mathbf{E}_{J,x}[\mu_k(B_x)] \geq \mathbf{E}_{J,x}[\mu_k(B_x) \mid H] \mathbf{Pr}[H].$$

Combining these two deductions with (8) yields

$$\mathbf{E}_{J,x}[\mu_k(B_x) \mid H] \mathbf{Pr}[H] \leq \beta(1 + \eta/3), \quad (\delta - 2\eta/3)\beta \leq \mathbf{E}_{J,x}[\mu_k(A_x) \mid H] \mathbf{Pr}[H].$$

Thus

$$\frac{\delta - 2\eta/3}{1 + \eta/3} \cdot \mathbf{E}_{J,x}[\mu_k(B_x) \mid H] \mathbf{Pr}[H] \leq \mathbf{E}_{J,x}[\mu_k(A_x) \mid H] \mathbf{Pr}[H],$$

whence

$$0 \leq \mathbf{E}_{J,x}[\mu_k(A_x) - (\delta - \eta)\mu_k(B_x) \mid H] \mathbf{Pr}[H],$$

where we used $(\delta - 2\eta/3)/(1 + \eta/3) \geq \delta - \eta$. As $\mathbf{Pr}[H] > 0$, it follows that there exists some (J, x) for which H occurs and also $\mu_k(A_x) \geq (\delta - \eta)\mu_k(B_x)$. This completes the proof. \square

Finally, the set B we get from Theorem 8.2 is a union of insensitive sets. We would prefer for this to be a *disjoint* union; we can achieve this if we settle for intersections of insensitive sets.

Lemma 8.4. *Suppose $B = B_1 \cup B_2 \cup \dots \cup B_{k-1} \subseteq [k]^n$, where B_a is ak -insensitive. Then B is a disjoint union of $k-1$ sets C_1, \dots, C_{k-1} , where C_b is an intersection of ak -insensitive sets, $a = 1 \dots b$.*

Proof. Take $C_b = B_b \setminus (B_1 \cup \dots \cup B_{b-1})$. \square

9 Partitioning ab -insensitive sets

The last tool we need is the result described in Section 4.5; given the multidimensional DHJ $_{k-1}$ theorem, we show that ab -insensitive subsets of $[k]^n$ can be almost completely partitioned into disjoint d -dimensional subspaces.

Theorem 9.1. *Let $k \geq 3$ and $d \geq 1$ be integers and let $0 < \eta < 1/2$. Suppose $C \subseteq [k]^n$ is ab -insensitive on $I \subseteq [n]$, where $a, b \in [k]$ are distinct symbols. Assume*

$$|I| \geq m(k(k+d))^m \ln(1/\eta),$$

where m is an integer satisfying

$$m \geq 48\text{MDHJ}_{k-1}(\eta/4, d).$$

Then C can be partitioned into a collection \mathcal{S} of disjoint d -dimensional subspaces, along with an error set E having uniform-distribution density $\mu_k^{\otimes n}(E) \leq \eta$.

Proof. The proof proceeds in “rounds”, $t = 1, 2, 3, \dots$. In each round, we remove some disjoint d -dimensional subspaces from C and put them into \mathcal{S} ; and, the set I shrinks by m coordinates. We will show that after the t th round,

$$\mu_k^{\otimes n}(C) \leq \left(1 - (k(k+d))^{-m}\right)^t. \quad (9)$$

We continue the process until $\mu_k^{\otimes n}(C) \leq \eta$, at which point we may stop and set $E = C$. Because of (9), the process stops after at most $T = (k(k+d))^m \ln(1/\eta)$ rounds. Since we insisted $|I| \geq mT$ initially, the set I never “runs out of coordinates”.

Suppose we are about to begin the t th round; hence, writing $\alpha = \mu_k^{\otimes n}(C)$ we have

$$\eta < \alpha \leq \left(1 - (k(k+d))^{-m}\right)^{t-1}.$$

The round begins by choosing an arbitrary $J \subseteq I$ with $|J| = m$. We have

$$\alpha = \Pr_{x \sim \mu_k^{\otimes n}} [x \in C] = \mathbf{E}_{y \sim \mu_k^{\otimes J}} [\mu_k^{\otimes J}(C_y)],$$

where we have written $C_y = \{z \in [k]^J : (y, z) \in C\}$. Hence $\mu_k^{\otimes J}(C_y) \geq \alpha/2$ for at least an $\alpha/2$ probability mass of y 's (under $\mu_k^{\otimes J}$); call these y 's “good”. Since $J \subseteq I$ and C is ab -insensitive on I , it follows that each C_y is ab -insensitive on J . Since

$$|J| = m \geq 48\text{MDHJ}_{k-1}((\eta/2)/2, d) \geq 48\text{MDHJ}_{k-1}((\alpha/2)/2, d),$$

it follows from Proposition 7.4 that for each good y there must exist a d -dimensional subspace $\rho \subseteq C_y$. Since the number of d -dimensional subspaces in $[k]^m$ is at most $(k+d)^m$, there must exist a fixed d -dimensional subspace $\rho_0 \subseteq [k]^J$ such that

$$\Pr_{y \sim \mu_k^{\otimes J}} [\rho_0 \subseteq C_y] \geq \frac{\alpha}{2(k+d)^m}. \quad (10)$$

Let $R \subseteq [k]^{\bar{J}}$ be the set of y 's with $\rho_0 \subseteq C_y$. Since C is ab -insensitive on I , it is easy to see that R is ab -insensitive on $I \setminus J$. Thus $R \times \rho_0$ is ab -insensitive on $I \setminus J$; hence so too is $C \setminus (R \times \rho_0)$.

We therefore complete the round by setting $I = I \setminus J$ and transferring $R \times \rho_0$ (a disjoint union of d -dimensional subspaces $\{y\} \times \rho_0$) from C into \mathcal{S} . This shrinks the number of coordinates in I by m , as promised. And since we can crudely bound $\mu_k^{\otimes J}(\rho_0) = k^{d-m} \geq 2k^{-m}$, we conclude

$$\mu_k^{\otimes n}(R \times \rho_0) \geq \frac{\alpha}{(k(k+d))^m},$$

as required to establish (9) inductively. \square

We conclude this section by simplifying parameters slightly, then deducing Theorem 9.1 for intersections of ab -insensitive sets.

Corollary 9.2. *Let $d \geq k \geq 3$, $0 < \eta < 1/2$, and define m as in Theorem 9.1. If $C \subseteq [k]^n$ is ab -insensitive and $n \geq d^{3m}$, then the conclusion of Theorem 9.1 holds.*

Proof. We only need check that $d^{3m} \geq m(k(k+d))^m \ln(1/\eta)$. We use the bounds $k \leq d$ and $m \geq 4/\eta$ (which is certainly necessary, just by Sperner's theorem), and hence must show $d^{3m} \geq m \ln(m/4)(2d^2)^m$. This is easily verified for every $d \geq 3$ and $m > 4$. \square

Corollary 9.3. *Let d, k, η, m be as in Corollary 9.2, and write $f(d) = d^{3m(d)}$, treating m as a function of d , with k and η fixed. Let $C \subseteq [k]^n$ be expressible as $C = C_1 \cap \dots \cap C_\ell$, where each C_s is $a_s b_s$ -insensitive for some pair of distinct symbols $a_s, b_s \in [k]$. If $n \geq f^{(\ell)}(d)$ (meaning the ℓ th iteration of f at d) then the conclusion of Theorem 9.1 holds with error bound $\ell\eta$.*

Proof. The proof is an induction on ℓ , with Corollary 9.2 serving as the base case. In the general case, since $n \geq f^{(\ell-1)}(f(d))$, by induction we can partition $C' = (C_1 \cap \dots \cap C_{\ell-1})$ into a collection \mathcal{S}' of disjoint nondegenerate $f(d)$ -dimensional subspaces, along with an error set E' satisfying $\mu_k^{\otimes n}(E') \leq (\ell-1)\eta$. For each $\sigma \in \mathcal{S}'$, define $D_\sigma = C_\ell \cap \sigma$. If we identify σ with $[k]^{f(d)}$ then we may think of D_σ as an $a_\ell b_\ell$ -insensitive subset of $[k]^{f(d)}$. Thus we may apply Corollary 9.2 and partition D_σ into a collection \mathcal{S}_σ of d -dimensional subspaces, along with an error set E_σ of small probability mass. The elements of \mathcal{S}_σ are d -dimensional subspaces of $\sigma \cong [k]^{f(d)}$; note that these are also d -dimensional subspaces of the original space $[k]^n$. As for E_σ , Corollary 9.2 guarantees it has uniform density at most η when viewed as a subset of $\sigma \cong [k]^{f(d)}$; thus in the original space $[k]^n$ we have $\mu_k^{\otimes n}(E_\sigma) \leq \eta \cdot \mu_k^{\otimes n}(\sigma)$. We may now complete the induction by taking \mathcal{S} to be $\{D_\sigma : \sigma \in \mathcal{S}'\}$ and E to be $E' \cup \bigcup \{E_\sigma : \sigma \in \mathcal{S}'\}$, observing that

$$\mu_k^{\otimes n}(E) \leq (\ell-1)\eta + \eta \left(\sum_{\sigma \in \mathcal{S}'} \mu_k^{\otimes n}(\sigma) \right) \leq (\ell-1)\eta + \eta = \ell\eta$$

as required. \square

10 Completing the proof

In this section, we show how to deduce DHJ_k from DHJ_{k-1} . We will treat $k = 3$ separately, since we have better bounds in this case thanks to the Gunderson-Rödl-Sidorenko theorem (and to a lesser extent the probabilistic Sperner theorem).

10.1 The $k = 3$ case

We establish the bound $\text{DHJ}_3(\delta) = 2 \uparrow\uparrow O(1/\delta^2)$ in Theorem 1.1 via the density increment argument described in Section 4.2. As explained in that section, once we have the reduction from DHJ_3 to the equal-slices version (Proposition 7.1), it remains to establish the following density increment theorem:

Theorem 4.4. *Assume $n \geq 2 \uparrow^{(7)} (1000/\delta)$. If $\tilde{\nu}_3^n(A) \geq \delta$, then either A contains a combinatorial line, or we may pass to a subspace of dimension at least $\log^{(7)} n$ on which $\tilde{\nu}_3(A) \geq \delta + \Omega(\delta^3)$.*

Proof. Throughout the proof we may assume $\delta \leq 2/3$; otherwise DHJ_3 is trivial, because $[3]^n$ can be partitioned into disjoint combinatorial lines. We will use the assumption $n \geq 2 \uparrow^{(7)} (1000/\delta)$ without comment in most of the proof, and we will not be especially economical when it comes to choosing parameters that do not affect the final bound.

By Proposition 5.14, we have $\bar{\nu}_3^n(A) \geq \delta - 6/n \geq \delta - \delta^{50}$. Let $\theta = 2\delta^{50}$. It is easy to check that the hypotheses of Lemma 8.1 are satisfied; hence we may pass to subspace of dimension at least $(\delta^{50}/6)\sqrt{n} \geq n^{1/3}$ on which either $\bar{\nu}_3(A) \geq \delta + \delta^{50}$, or both

$$\begin{aligned} \bar{\nu}_3(A) &\geq \delta - \delta^{50} - 3(2\delta^{50})^{1/4}(\delta - \delta^{50})^{1/2} \geq \delta - 4\delta^{13}, \\ \text{and } \bar{\nu}_2(A) &\geq \delta - \delta^{50} - 3(2\delta^{50})^{1/2} \geq \delta - 4\delta^{25}. \end{aligned}$$

We repeatedly use Lemma 8.1 until either the second conclusion obtains or until we have repeated $\lceil 1/\delta^{47} \rceil$ times. In either case, the dimension n becomes some n' satisfying $n' \geq n \uparrow 3 \uparrow (-1/\delta^{47})$. Having done this, we either have $\bar{\nu}_3(A) \geq \delta + \Omega(\delta^3)$ and may stop (as $n' \gg \log^{(7)} n$, by the assumption on n), or we have $\bar{\nu}_3(A), \bar{\nu}_2(A) \geq \delta - 4\delta^{13}$. With another application of Proposition 5.14 we may obtain $\tilde{\nu}_3(A), \tilde{\nu}_2(A) \geq \delta - 5\delta^{13}$.

Next, we apply Theorem 8.2, taking $\delta_k = \delta_{k-1} = \delta - 5\delta^{13}$. Note that we have proved the probabilistic DHJ_2 theorem in Lemma 7.3. We certainly have $n' \geq \text{PDHJ}_2(\delta_{k-1}) = \Theta(1/\delta^2)$, and thus we either obtain a combinatorial line in A , or a set $B \subseteq [3]^{n'}$ with $\tilde{\nu}_3(B) \geq \delta - 5\delta^{13}$ and

$$\frac{\tilde{\nu}_3(A \cap B)}{\tilde{\nu}_3(B)} \geq \delta - 5\delta^{13} + (\delta - 5\delta^{13})(\delta - 5\delta^{13})^2/2 \geq \delta + \delta^3/2 - 12.5\delta^{13} \geq \delta + \delta^3/4,$$

where we used $\epsilon_2(\delta_{k-1}) = \delta_{k-1}^2/2$. Also, B may be written as $B_1 \cup B_2$, where B_a is $a3$ -insensitive. We henceforth write A instead of $A \cap B$.

We now return to the uniform distribution by applying Lemma 8.3. Its β satisfies $\beta \geq \delta - 5\delta^{13}$, and its δ needs (crucially) to be replaced by $\delta + \delta^3/4$. We choose its $\eta = \delta^3/8$. It is easy to check that the resulting $r = (\beta\eta/15k)\sqrt{n'}$ indeed satisfies $2 \ln n' \leq r \leq n'/2$, using $n' \geq n \uparrow 3 \uparrow (-1/\delta^{47})$. We now pass to a restriction on some

$$n'' \geq r \geq n \uparrow 3 \uparrow (-1/\delta^{48}) \tag{11}$$

coordinates on which we have, say,

$$\mu_3(B) \geq \delta^4/50, \quad \mu_3(A) \geq (\delta + \delta^3/8)\mu_3(B). \tag{12}$$

Crucially, since B was the union of a 13-insensitive set and a 23-insensitive set before, it remains so in the restriction. We also now apply Lemma 8.4 to partition B into sets C_1 and C_2 , where C_1 is 13-insensitive and C_2 is an intersection of 13- and 23-insensitive sets.

Nearing the end, we now apply Corollary 9.3 twice; once with $C = C_1$ and once with $C = C_2$. We choose $\eta = \delta^7/4000$, and for simplicity we take $\ell = 2$ both times. We will select the parameter d

later, at which point we will verify that n'' is large enough for our choice of d . We thereby partition each of C_1, C_2 into disjoint d -dimensional combinatorial subspaces, along with error sets E_1, E_2 . Writing $E = E_1 \cup E_2$ we have

$$\mu_3(E) \leq 2\eta + 2\eta \leq \frac{\delta^7}{1000} \leq \frac{\delta^3}{20} \mu_3(B), \quad (13)$$

using the first inequality (12). The second inequality in (12) implies

$$\frac{\mu_3(A \setminus E)}{\mu_3(B \setminus E)} \geq \frac{\mu_3(A) - \mu_3(E)}{\mu_3(B)} \geq \delta + \delta^3/8 - \frac{\mu_3(E)}{\mu_3(B)} \geq \delta + \delta^3/8 - \delta^3/20 \geq \delta + \delta^3/20,$$

where we also used (13). Thus if we simply discard the strings in E from A and B , we still have $\mu_3(A)/\mu_3(B) \geq \delta + \delta^3/20$, and now B is perfectly partitioned into disjoint d -dimensional subspaces. It follows that we can pass to at least one d -dimensional subspace on which $\mu_3(A) \geq \delta + \delta^3/20$.

We come now to the final setting of parameters. To use the partitioning argument, Corollary 9.3, we require that

$$n'' \geq f(f(d)), \quad \text{where } f(d) = 3^{m(d)} \quad (14)$$

and $m(d) \geq 48\text{MDHJ}_2(\delta^7/16000, d)$. By Theorem 3.2, our improvement to Gunderson-Rödl-Sidorenko theorem, we may take

$$m(d) = \lceil 48 \cdot 25 \cdot (16000/\delta^7)^{2^d} \rceil \leq (10^6/\delta^7)^{2^d}.$$

We will ensure that

$$d \geq 1/\delta^{48} \geq 10^6/\delta^7, \quad (15)$$

so that we may bound $m(d) \leq d^{2^d}$. This yields

$$f(d) = 3^{m(d)} \leq 3 \uparrow d \uparrow 2 \uparrow d \leq 3 \uparrow^{(3)} d \quad \Rightarrow \quad f(f(d)) \leq 3 \uparrow^{(6)} d.$$

Combining our lower bound (11) on n'' with the assumption in (15), we have $n'' \geq n^{1/3^d}$; thus the main constraint (14) is satisfied so long as

$$n^{1/3^d} \geq 3 \uparrow^{(6)} d, \quad \text{which is implied by } n \geq 4 \uparrow^{(6)} d. \quad (16)$$

We may therefore set

$$d \approx (\log^{(7)} n)^{50};$$

this satisfies both (15) and (16), using the fact that $n \geq 2 \uparrow^{(7)} (1000/\delta)$.

It remains to note that, having passed to a $(\log^{(7)} n)^{50}$ -dimensional subspace on which $\mu_3(A) \geq \delta + \delta^3/20$, we can pass to a further subspace of dimension at least the claimed $\log^{(7)} n$ on which $\tilde{\nu}_3(A) \geq \delta + \delta^3/40$; we need only use one more embedding argument of the type described in Section 4.1 and justified by Lemma 6.3.1 and Proposition 5.14. \square

10.2 The general k case

For general $k \geq 4$ we will need to use the inductive deduction of the multidimensional DHJ_{k-1} theorem from the DHJ_{k-1} theorem, given in Proposition 2.3. Because the overall argument then becomes a double induction, the resulting bounds for DHJ_k are of Ackermann-type, and we do not work them out precisely.

Theorem 10.1. *For $k \geq 4$, the DHJ_k theorem follows from the DHJ_{k-1} theorem.*

Proof. (Sketch.) As DHJ_k is equivalent to the equal-slices version (Proposition 7.1), we may begin with a subset $A \subseteq [k]^n$ satisfying $\tilde{\nu}_k(A) \geq \delta$. Assuming DHJ_{k-1} , we deduce the multidimensional DHJ_{k-1} via Proposition 2.3, the equal-slices DHJ_{k-1} via Proposition 7.1, and the probabilistic DHJ_{k-1} via Proposition 7.2. We henceforth think of δ and k as “constants”, and by contrast we think of $n \rightarrow \infty$. Hence even, say, $\text{PDHJ}_{k-1}(\delta)$ and $1/\epsilon_{k-1}(\delta)$ are to be thought of as “constants” (albeit very large ones), depending only on δ and k .

The overall goal for proving DHJ_k is, as in the $k = 3$ case, to show that for n sufficiently large as a function of δ and k , we can either find a line in A or can obtain a density increment of $\Omega(\epsilon)$ on a combinatorial subspace of dimension d' . Here $\epsilon = \delta \cdot \epsilon_{k-1}(\delta/2)$ is a positive “constant” depending only on δ and k , whereas $d' = d'(n)$ is a function of n alone satisfying $d'(n) \rightarrow \infty$ as $n \rightarrow \infty$.

Beginning with $\tilde{\nu}_k(A) \geq \delta$, we choose $\theta = 2\epsilon^C$ for some large universal constant C ; note that θ only depends on δ and k . We change from $\tilde{\nu}_k(A)$ to $\bar{\nu}_k(A)$ via Proposition 5.14, incurring density loss only ϵ^C by assuming n large enough. We next apply Lemma 8.1, passing to a subspace of dimension roughly $n^{1/3}$, say, assuming n large enough. This gives us either a density increment of ϵ^C , or achieves $\bar{\nu}_k(A), \bar{\nu}_{k-1}(A) \geq \delta - O(\epsilon^c)$ for some smaller but still large universal constant c . We can repeat this application of Lemma 8.1 $\lfloor 1/\epsilon^C \rfloor$ many times: if the first outcome always happens we obtain the desired density increment of $\Omega(\epsilon)$; otherwise, we achieve $\bar{\nu}_k(A), \bar{\nu}_{k-1}(A) \geq \delta - O(\epsilon^c)$. This brings the dimension n down to some fractional power n' , but the fraction depends only on ϵ^C and hence on δ and k . Thus we can assume in future arguments that n' is sufficiently large as a function of δ and k . With another application of Proposition 5.14 we ensure that $\tilde{\nu}_k(A), \tilde{\nu}_{k-1}(A) \geq \delta - O(\epsilon^c)$.

We now apply Theorem 8.2, taking $\delta_k = \tilde{\nu}_k(A)$ and $\delta_{k-1} = \delta/2$ for simplicity (without loss of generality, $\delta - O(\epsilon^c) \geq \delta/2$). This requires n' to be larger than $\text{PDHJ}_{k-1}(\delta/2)$, a “constant” depending only on δ and k . Having applied the theorem, we either get a combinatorial line, or a relative density increment of $\delta_k \cdot \epsilon_{k-1}(\delta/2) = \Omega(\epsilon)$ on some set B (where we use the fact that $\epsilon \gg O(\epsilon^c)$). The set B has $\tilde{\nu}_k(B) \geq \delta/2$ and is a union of ak -insensitive sets, as in the theorem statement. We can preserve this relative density increment under the uniform distribution by passing to another subspace according to Lemma 8.3 with $\eta = \epsilon^C$. The new dimension n'' is still just some fractional power of n , where the fraction depends only on the constants δ and k . As in the DHJ_3 proof, B still has the special structure as a union of ak -insensitive sets. We again apply Lemma 8.4 to write it as a disjoint union of k sets, each of which is an intersection of up to k many ak -insensitive sets.

We now apply Corollary 9.3 to each of these sets, taking $\eta = \epsilon^C$ again and $\ell = k$ each time for simplicity. It suffices now to analyze what parameter d we may choose; the remainder of the proof is essentially the same as in the DHJ_3 case. The function $m(d)$ arising in the corollary is $O(\text{MDHJ}_{k-1}(\eta/4, d))$, which depends on the “constant” η as well as on d . As we will eventually take $d = d(n)$ to be a function of n alone, by assuming n sufficiently large we can ensure d is larger than any constant function. Hence we can upper-bound $m(d)$ by a function of d alone. With this upper bound, $f(d) = d^{3m(d)}$ and also $f^{(\ell)}(d) = f^{(k)}(d)$ are some other (huge) functions of d alone.

Now Corollary 9.3 demands $n'' \geq f^{(k)}(d)$; we may satisfy this demand by taking $d = d(n)$ to be some function of n alone, with $d(n) \rightarrow \infty$ as $n \rightarrow \infty$. Notice this allows us to make $d(n)$ larger than any function of the constants δ and k , by requiring n sufficiently large. Thus we get the required density increment of $\Omega(\epsilon)$ under the uniform distribution, on a subspace of dimension $d(n)$. As in the proof of DHJ_3 , we convert this to a density increment under $\tilde{\nu}_k$ by passing to a further subspace of dimension $d' = d^{\Theta(1)}$, requiring once more that d is a sufficiently large function of δ and k . \square

References

- [AS08] Noga Alon and Joel H. Spencer. *The Probabilistic Method*, 3rd ed. Wiley–Interscience, 2008.
- [Aus09] Tim Austin. Deducing the Density Hales–Jewett Theorem from an infinitary removal lemma. <http://arxiv.org/abs/0903.1633>, 2009.
- [Beh46] Felix A. Behrend. On sets of integers which contain no three terms in arithmetical progression. *Proceedings of the National Academy of Sciences*, 32(12):331–332, 1946.
- [BL96] Vitaly Bergelson and Alexander Leibman. Polynomial extensions of van der Waerden’s and Szemerédi’s theorems. *Journal of the American Mathematical Society*, 9(3):725–753, 1996.
- [FK78] Hillel Furstenberg and Yitzhak Katznelson. An ergodic Szemerédi theorem for commuting transformations. *Journal d’Analyse Mathématique*, 34(1):275–291, 1978.
- [FK89] Hillel Furstenberg and Yitzhak Katznelson. A density version of the Hales–Jewett Theorem for $k = 3$. *Discrete Mathematics*, 75:227–241, 1989.
- [FK91] Hillel Furstenberg and Yitzhak Katznelson. A density version of the Hales–Jewett Theorem. *Journal d’Analyse Mathématique*, 57:64–119, 1991.
- [FKO82] Hillel Furstenberg, Yitzhak Katznelson, and Donald Ornstein. The ergodic theoretical proof of szemerédi’s theorem. *Bulletin of the American Mathematical Society*, 7(3):527–552, 1982.
- [Fur77] Hillel Furstenberg. Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *Journal d’Analyse Mathématique*, 31(1):204–256, 1977.
- [Fur85] Hillel Furstenberg. An ergodic Szemerédi theorem for IP-systems and combinatorial theory. *Journal d’Analyse Mathématique*, 45(1):117–168, 1985.
- [Gow01] W. Timothy Gowers. A new proof of Szemerédi’s theorem. *Geometric and Functional Analysis*, 11(3):465–588, 2001.
- [Gow06] W. Timothy Gowers. Quasirandomness, counting and regularity for 3-uniform hypergraphs. *Combinatorics, Probability and Computing*, 15(1–2):143–184, 2006.
- [Gow07] W. Timothy Gowers. Hypergraph regularity and the multidimensional Szemerédi theorem. *Annals of Mathematics*, 166(3):897–946, 2007.
- [GRS99] David S. Gunderson, Vojtěch Rödl, and Alexander Sidorenko. Extremal problems for sets forming Boolean algebras and complete partite hypergraphs. *Journal of Combinatorial Theory, Series A*, 88(2):342–367, 1999.
- [HJ63] Alfred W. Hales and Robert I. Jewett. Regularity and positional games. *Transactions of the American Mathematical Society*, 106(2):222–229, 1963.
- [NRS06] Brendan Nagle, Vojtěch Rödl, and Mathias Schacht. The counting lemma for regular k -uniform hypergraphs. *Random Structures and Algorithms*, 28(2):113–179, 2006.

- [Pol09] D. H. J. Polymath. Density Hales–Jewett and Moser numbers in low dimensions. Unpublished, <http://michaelnielsen.org/polymath1/>, 2009.
- [Rei89] Rolf-Dieter Reiss. *Approximate distributions of order statistics*. Springer-Verlag, 1989.
- [Rot53] Klaus F. Roth. On certain sets of integers. *Journal of the London Mathematical Society*, 28(1):104–109, 1953.
- [RS04] Vojtěch Rödl and Jozef Skokan. Regularity lemma for k -uniform hypergraphs. *Random Structures and Algorithms*, 25(1):1–42, 2004.
- [RS06] Vojtěch Rödl and Jozef Skokan. Applications of the regularity lemma for uniform hypergraphs. *Random Structures and Algorithms*, 28(2):180–194, 2006.
- [RS07a] Vojtěch Rödl and Mathias Schacht. Regular partitions of hypergraphs: Counting lemmas. *Combinatorics, Probability and Computing*, 16(6):887–901, 2007.
- [RS07b] Vojtěch Rödl and Mathias Schacht. Regular partitions of hypergraphs: regularity lemmas. *Combinatorics, Probability and Computing*, 16(6):833–885, 2007.
- [She88] Saharon Shelah. Primitive recursive bounds for van der Waerden numbers. *Journal of the American Mathematical Society*, 1(3):683–697, 1988.
- [Shk06a] Ilya D. Shkredov. On a generalization of Szemerédi’s theorem. *Proceedings of the London Mathematical Society*, 93(3):723–760, 2006.
- [Shk06b] Ilya D. Shkredov. On a problem of Gowers. *Izvestiya: Mathematics*, 70(2):385–425, 2006.
- [Spe28] Emanuel Sperner. Ein Satz über Untermengen einer endlichen Menge. *Mathematische Zeitschrift*, 27(1):544–548, 1928.
- [Spe90] Joel Spencer. The probabilistic lens: Sperner, Turán and Bregman revisited. In *A tribute to Paul Erdős*, pages 391–396. Cambridge University Press, 1990.
- [Sta00] Richard P. Stanley. *Enumerative combinatorics, volume 1*. Cambridge University Press, 2000.
- [Sze75] Endre Szemerédi. On sets of integers containing no k elements in arithmetic progression. *Acta Arithmetica*, 27:199–245, 1975.
- [Tao06] Terence Tao. A quantitative ergodic theory proof of Szemerédi’s theorem. *The Electronic Journal of Combinatorics*, 13(1), 2006.
- [Tao07] Terence Tao. A correspondence principle between (hyper)graph theory and probability theory, and the (hyper)graph removal lemma. *Journal d’Analyse Mathématique*, 103(1):1–45, 2007.
- [Var59] Panayiotis Varnavides. On certain sets of positive density. *Journal of the London Mathematical Society*, 1(3):358–360, 1959.
- [vdW27] Bartel L. van der Waerden. Beweis einer Baudetschen Vermutung. *Nieuw Archief voor Wiskunde*, 15:212–216, 1927.