A new proof of the density Hales-Jewett theorem

D. H. J. Polymath*

June 25, 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 $[3]^n$ of density δ contains a combinatorial line if $n \geq 2 \uparrow\uparrow O(1/\delta^3)$. Our proof is surprisingly[["reasonably", maybe]] simple: indeed, it gives what is probably 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 $\kappa\colon X \to [r]$. 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/

arithmetic progressions, the Hales–Jewett theorem looks for *combinatorial lines*. A combinatorial line in $[k]^n$ is a set of k points $\{x^{(1)}, \ldots, x^{(k)}\}$ 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 $k = 3$ and $\lambda = 13\star\star221\star$, the associated combinatorial line is the following set of 3 points:

$$\{13\mathbf{1}1221\mathbf{1}, 13\mathbf{2}\mathbf{2}221\mathbf{2}, 1\mathbf{3}\mathbf{3}3221\mathbf{3}\}.$$

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

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 integer $\mathrm{HJ}_k(r)$ such that for every r-colouring of the set $[k]^n$ there is a monochromatic combinatorial line, provided $n \geq \mathrm{HJ}_k(r)$.*

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.

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

We sometimes write "DHJ(k)" to mean the k case of this theorem.[[A bit confusing, since $\mathrm{DHJ}_k(\delta)$ is a number in the above theorem. Hmm.]] 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 $\{0, 1, \ldots, m-1\}$ rather than $\{1, 2, \ldots, m\}$, and identify integers in $[N]$ with their base-k representation in $[k]^n$.[[(we may assume N is a power of k)]] 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 $\mathrm{DHJ}_3(\delta) = 2 \uparrow\uparrow O(1/\delta^3)$. For $k \geq 4$, the bound $\mathrm{DHJ}_k(\delta)$ we achieve is of Ackermann type.[[cop-out]]*

Here we use the notation $x \uparrow y$ for x^y, $x \uparrow^{(\ell)} y$ for $x \uparrow x \uparrow \cdots \uparrow x \uparrow y$ (with ℓ many $\uparrow$'s), and $x \uparrow\uparrow y$ for $x \uparrow^{(y)} x$.

We add that it is not too hard to obtain the following extensions to this result: (i) a *multidimensional* version, in which one finds higher-dimensional combinatorial subspaces;* (ii) 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;[[name-check Varnavides?]] (iii) the combined probabilistic multidimensional version. Indeed, to prove Theorem 1.1 we found it necessary to obtain these extensions in passing. See Section 2 for more detailed statements. *ignored part*

1.2 Some discussion

Why is it interesting to give a new proof of this result? There are two main main reasons. The first is connected with the history of results 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, many far-reaching generalizations of Szemerédi's theorem were proved, 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. [[Citations would be good here; which are the far-reaching generalisations? Also, Tim's paper on Szemerédi is maybe 9 years old now... does that count as "few years"?]] Giving a non-ergodic proof of the density Hales–Jewett theorem was seen as a key goal for this programme: on the ergodic side it appeared to be significantly harder than Szemerédi's theorem, and this seemed to be reflected by certain significant combinatorial difficulties that we shall discuss later in this paper.[[Which difficulties does this refer to?]] Having given a purely combinatorial 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 (although admittedly the bounds for $k \geq 4$ are terrible and even the $k = 3$ bound is not especially good). Such bounds could not be obtained via the ergodic methods even in principle, since these proofs rely on the Axiom of Choice[[I just heard someone say this; is it true?]].[[maybe expand slightly on this, and/or cite the sister paper]]

A second reason is that the density Hales–Jewett theorem 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.[[Citations here.]] In retrospect, one can add to this the fact that the proof we have discovered is, against all our expectations[[a strong phrase...]], *simpler* than the previous approaches to Szemerédi's theorem; the most advanced notion needed 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, which is when $k = 3$. 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. However, 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 general case. We still do not have a convincing explanation of why our proof should differ from all the others in this respect.[[Perhaps we should think of one :)]]

* *ignored part*

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 done by several mathematicians, who wrote their thoughts, as they had them, in the form of blog comments.[[I would like to at least consider adding a link or statement that it was on Tim's blog. Otherwise, the only identifying part of the paper is the wiki's URL, which might make it seem like Michael Nielsen was the mastermind. Okay with you, Tim?]] Anybody who wanted to could participate, and at all stages of the process the comments were fully open to anybody who was interested. This 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 somewhat polished account of the argument, it is possible to read a record of the entire thought process that led to the proof. The participants also created a wiki, which contains sketches of the argument, links to the blog comments, and a great deal of related material. The wiki's URL is `http://michaelnielsen.org/polymath1/`.

*

*Tim: **Somewhere we should mention Tim Austin's work and how it relates to ours. Perhaps Terry is best placed to do this.**

1.3 Outline of the paper

XXX "table of contents" paragraph to be filled in XXX

2 Basic concepts

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 $[k] = \{1, 2, \ldots, k\}$ 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, ... for symbols, x, y, ... for strings/points, and i, j, ... for coordinates; e.g., $x_j = b$ means the jth coordinate of string x is symbol b.

We have already seen that it was convenient to use a different alphabet Ω when we took $\Omega = \{0, 1, \ldots, k-1\}$ 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 $\{1\mathbf{111}221\mathbf{1}, 1\mathbf{222}221\mathbf{2}\} \subset [2]^8$. We introduce the notation $x^{b \to a}$ 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 $\{\lambda^{\star \to 1}, \lambda^{\star \to 2}, \lambda^{\star \to 3}\} \subseteq [3]^n$.

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 ($\notin \Omega$) 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* $\{\lambda^{c \to a} : a \in \Omega\} \subseteq \Omega^n$, 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 simplicity, 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

$$\sigma = 13\star_1\star_2 21\star_1\star_1 3\star_2 1$$

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

$$\begin{aligned}\{&13\mathbf{11}2\mathbf{111}3\mathbf{1}1, \quad 13\mathbf{21}2\mathbf{12}2\mathbf{3}1, \quad 13\mathbf{31}2\mathbf{133}3\mathbf{1}1,\\ &13\mathbf{12}2\mathbf{111}3\mathbf{2}1, \quad 13\mathbf{22}2\mathbf{12}2\mathbf{3}2\mathbf{1}, \quad 13\mathbf{32}2\mathbf{133}3\mathbf{2}1,\\ &13\mathbf{13}2\mathbf{111}3\mathbf{3}1, \quad 13\mathbf{23}2\mathbf{12}2\mathbf{3}3\mathbf{1}, \quad 13\mathbf{33}2\mathbf{133}3\mathbf{3}1\} \subset [3]^{11}.\end{aligned}$$

More generally:

Definition 2.2. For $d \geq 1$, a *d-dimensional subspace template over Ω^n with wildcard symbols* $c_1, \ldots, c_d$ (distinct symbols not in Ω) is a string $\sigma \in (\Omega \cup \{c_1, \ldots, c_d\})^n$. 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* $\{\sigma^{(c_1,\ldots,c_d)\to(a_1,\ldots,a_d)} : (a_1, \ldots, a_d) \in \Omega^d\} \subseteq \Omega^n$, which has cardinality $|\Omega|^d$.

Here we have introduced the extended notation $x^{(c_1,\ldots,c_d)\to(a_1,\ldots,a_d)}$ 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 (also proved by Furstenberg and Katznelson [FK91]) 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 integer $\mathrm{MDHJ}_k(\delta, d)$ such that every subset of $[k]^n$ of density at least δ contains a d-dimensional combinatorial subspace, provided $n \geq \mathrm{MDHJ}_k(\delta, d)$.*

Indeed, it is easy to deduce the multidimensional $\mathrm{DHJ}(k)$ from the basic $\mathrm{DHJ}(k)$:

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 $\mathrm{DHJ}(k)$ theorem for $d-1$, and let $A \subseteq [k]^n$ have density at least δ. Let $m = \mathrm{MDHJ}_k(\delta/2, d-1)$, 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 counting argument, the density of G within $[k]^{n-m}$ must be at least $\delta/2$; otherwise, A would not have density at least δ in $[k]^{n-m}$.

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 $n \geq m + \mathrm{DHJ}_k(\delta/2M)$, we conclude from $\mathrm{DHJ}(k)$ that G_σ contains a combinatorial line, λ. Now $\sigma \times \lambda \subseteq [k]^n$ is the desired d-dimensional subspace contained in A. □

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* $\mathrm{MDHJ}_2(\delta, d) = (10d)^{d2^d} \cdot (1/\delta)^{2^d}$.

(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 this bound is fairly close to being sharp.) * *ignored part*

2.2.1 Passing to a subspace

A useful aspect of subspaces is that they "look like" the original space.[[I plagiarized this phrase from Walters.]] More precisely, a d-dimensional subspace $\{\sigma^{(c_1,\ldots,c_d)\to(a_1,\ldots,a_d)} : (a_1, \ldots, a_d) \in \Omega^d\} \subseteq \Omega^n$ 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 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 a 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 $\gamma(A) = \mathbf{Pr}_{x \sim \gamma}[x \in A]$ for the γ-density of A. 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 $\lambda^{4 \to 1} \in [3]^n$ will be a product distribution $\pi^{\otimes n}$ in which symbol 1 has probability 1/2 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 $\lambda^{4 \to 1} \in [3]^n$ may be a negligibly small function of n.

Even for other natural distributions on lines, it seems inevitable that $\lambda^{4 \to 1}$ will "have more 1's than 2's or 3's", and similarly for $\lambda^{4 \to 2}$, $\lambda^{4 \to 3}$. To evade this difficulty we take inspiration from the probabilistic proof Sperner's theorem (see, e.g., [Spe90])[[earliest citation I could find; note that Lubell's proof is a bit different]] 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 $\overline{\nu}_k^n$. We will discuss $\overline{\nu}_2^n$ in Section 3 on Sperner's theorem, and its generalization $\overline{\nu}_k^n$ 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 $\overline{\nu}_2^n$ as follows: first choose an integer $s \in \{0, 1, \ldots, n\}$ uniformly; then choose x uniformly at random from the sth "slice" of the discrete cube, i.e., the set of strings with exactly s many 1's. The more general $\overline{\nu}_k^n$ involves choosing the count of each symbol from $[k]$ in a certain uniform manner, then drawing a uniformly random string with the chosen symbol counts.[[awkward]]

As we will see, 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 $\lambda \sim \overline{\nu}_k^n$ has a very tiny chance of being being degenerate. To compensate for this, we introduce the very closely related distribution $\widetilde{\nu}_k^n$ ("equal-nondegenerate-slices"), which is $\overline{\nu}_k^n$ 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.) For every positive integer k and real $\delta > 0$ there exists a positive integer $\mathrm{EDHJ}_k(\delta)$ such that every $A \subseteq [k]^n$ with $\widetilde{\nu}_k^n(A) \geq \delta$ contains a combinatorial line, provided $n \geq \mathrm{EDHJ}_k(\delta)$.*

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

$$\mathop{\mathbf{Pr}}_{\lambda \sim \widetilde{\nu}_{k+1}^n}[\lambda \subseteq A] \geq \epsilon_k(\delta),$$

provided $n \geq \mathrm{PDHJ}_k(\delta)$. 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 give us multidimensional versions; the "equal-slices multidimensional density Hales–Jewett theorem" (with parameter "$\mathrm{EMDHJ}_k(\delta, d)$"), and 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 integer $\mathrm{PMDHJ}_k(\delta, d)$ and a positive real $\epsilon_k(\delta, d)$ such that every $A \subseteq [k]^n$ with $\widetilde{\nu}_k^n(A) \geq \delta$ satisfies*

$$\mathop{\mathbf{Pr}}_{\sigma \sim \widetilde{\nu}_{k+d}^n}[\sigma \subseteq A] \geq \epsilon_k(\delta, d),$$

provided $n \geq \mathrm{PMDHJ}_k(\delta, d)$.

3 Sperner's theorem

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 easily gives the $k = 2$ case of the probabilistic Density Hales–Jewett theorem, a key component in our proof of DHJ(3). *ignored part*

To investigate DHJ(2) it 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 to obtain y from x one changes 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 $\mathrm{DHJ}_2(\delta)$ such that for $n \geq \mathrm{DHJ}_2(\delta)$, 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 slightly 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 $\mathrm{DHJ}_2(\delta) = 4/\delta^2$.[[The constant can be sharpened to $\pi/2$ for small δ, of course. Also, $4/\delta^2$ is technically not an integer.]] 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 $\{0, 1, \ldots, n\}$. Finally, we set $X = \{\tau(1), \ldots, \tau(s)\}$ (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 $\{\tau(1), \ldots, \tau(s)\}$ 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 $\overline{\nu}_2[X]$ 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 $\overline{\nu}_2[X] = \frac{1}{n+1}\binom{n}{|X|}^{-1}$, 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 $\overline{\nu}_2[X]$ occurs when $|X| = \lfloor n/2 \rfloor$ or $\lceil n/2 \rceil$.[[by unimodality of binomial coefficients]] In this case, the ratio is $(n+1)\binom{n}{\lfloor n/2 \rfloor} 2^{-n}$.

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

As one sees from the proof, it is very natural to consider different probability distributions on $\{0,1\}^n$, or equivalently on the set of all subsets of $[n]$. 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 $\overline{\nu}_2$; the reader may check that this is precisely the "equal-slices" distribution $\overline{\nu}_2^n$ 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 $\overline{\nu}_2(\mathcal{A}) \leq 1/(n+1)$, and that the statement given above is a slightly artificial and strictly weaker consequence.

3.1 Probabilistic DHJ(2)

Indeed, what the proof (essentially) establishes is the "equal-slices DHJ(2) theorem"; i.e., that in Theorem 2.4 one may take $\mathrm{EDHJ}_2(\delta) = 1/\delta$.[[minus 1, even]] We say "essentially" because of the small distinction between the distribution $\overline{\nu}_2^n$ used in the proof and the distribution $\widetilde{\nu}_2^n$ in the statement. It will be convenient in this introductory discussion of Sperner's theorem to casually ignore this. We will introduce $\widetilde{\nu}_k^n$ and be more careful about its distinction with $\overline{\nu}_k^n$ in Section 5.

To further bolster the claim that $\overline{\nu}_2^n$ is natural in this context we will show an easy proof of the *probabilistic* DHJ(2) theorem. Looking at the statement of Theorem 2.5, the reader will see it requires defining $\overline{\nu}_3^n$; 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 $\overline{\nu}_2^n$-density at least δ satisfies*

$$\mathop{\mathbf{Pr}}_{\lambda \sim \overline{\nu}_3^n}[\lambda \subseteq A] \geq \delta^2.$$

Note that there is no lower bound necessary on n; this is because a template $\lambda \sim \overline{\nu}_3^n$ 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 $s \in \{0, 1, \ldots, n\}$ and also $t \in \{0, 1, \ldots, n\}$ *independently*. Let $x(\tau, s) \in \{0,1\}^n$ denote the string which has 1's in coordinates $\tau(1), \ldots, \tau(s)$ and 0's in coordinates $\tau(s+1), \ldots, \tau(n)$, and similarly define $x(\tau, t)$. These two strings both have the distribution $\overline{\nu}_2^n$, 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

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

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

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

$$\begin{aligned} \mathbf{Pr}[x(\tau,s), x(\tau,t) \in A] &= \mathop{\mathbf{E}}_{\tau}\Big[\mathop{\mathbf{Pr}}_{s,t}[x(\tau,s), x(\tau,t) \in A]\Big] \\ &= \mathop{\mathbf{E}}_{\tau}\Big[\mathop{\mathbf{Pr}}_{s}[x(\tau,s) \in A] \mathop{\mathbf{Pr}}_{t}[x(\tau,s) \in A]\Big] && \text{(independence of } s,\, t) \\ &= \mathop{\mathbf{E}}_{\tau}\Big[\mathop{\mathbf{Pr}}_{s}[x(\tau,s) \in A]^2\Big] \\ &\ge \mathop{\mathbf{E}}_{\tau}\Big[\mathop{\mathbf{Pr}}_{s}[x(\tau,s) \in A]\Big]^2 && \text{(Cauchy-Schwarz)} \\ &= \mathop{\mathbf{Pr}}_{\tau,s}[x(\tau,s) \in A]^2, \end{aligned}$$

and $x(\tau, s)$ has the distribution $\overline{\nu}_2^n$, completing the proof. □

Having proved the probabilistic DHJ(2) 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 $\overline{\nu}_k^n$ and $\widetilde{\nu}_k^n$ and prove some basic facts about them.

4 Outline of the proof

In this section we sketch our proof of DHJ(3), which involves reducing it to DHJ(2). 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 DHJ(3), we are a given a subset $A \subseteq [3]^n$ with uniform density $\mu_3(A) \ge \delta$, 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 Theorem 2.4, by "embedding" the equal-nondegenerate-slices distribution $\widetilde{\nu}_3$ in the uniform distribution $\mu_3^{\otimes n}$. To do this embedding, we define more explicitly the notion of *restrictions*:

Definition 4.1. A *restriction* on $[k]^n$ is a pair $(J, x_{\overline{J}})$, where $J \subseteq [n]$ and $x_{\overline{J}} \in [k]^{\overline{J}}$, where $\overline{J} = [n] \setminus J$. It is thought of as a "partial string", where the $\overline{J}$-coordinates are "fixed" according to $x_{\overline{J}}$ and the J-coordinates are still "free". Given a fixing $y_J \in [k]^J$, we write $(x_{\overline{J}}, y_J)$ for the complete composite string in $[k]^n$.[[Although this notation looks awkward, I believe it's useful.]]

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

Remark 4.3. As mentioned in Section 2.2.1, a restriction $(J, x_{\overline{J}})$ can be identified with a particular kind of combinatorial subspace. Specifically, it corresponds to the $|J|$-dimensional subspace whose template is $(x_{\overline{J}}, w_J)$, where w_J consists of $|J|$ distinct wildcard symbols. The set $A_{x_{\overline{J}}}$ 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 $x_{\overline{J}} \sim \mu_3^{\otimes \overline{J}}$ and $y_J \sim \mu_3^{\otimes J}$ independently, then the composite string $z = (x_{\overline{J}}, y_J)$ is distributed as $\mu_3^{\otimes n}$. Hence

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

It follow that there must exist a *particular* substring $x_{\overline{J}}$ for which $\mu_3^{\otimes J}(A_{x_{\overline{J}}}) \geq \delta$ as well. We can then pass to the subspace defined by $(J, x_{\overline{J}})$, and we have constructed as new instance of the original problem: "n" has decreased to $r = |J|$, and "A" (i.e., $A_{x_{\overline{J}}}$) 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 $x_{\overline{J}}$ according to the uniform distribution $\mu_3^{\otimes \overline{J}}$, and we choose the substring y_J according to equal-nondegenerate-slices $\widetilde{\nu}_3^J$. Write $\widetilde{\mu}_3^n$ for the resulting distribution on the composite string $z = (x_{\overline{J}}, y_J)$. 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" $\widetilde{\mu}_3^n$ 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 $\widetilde{\mu}_3^n$ is at most $O(r/\sqrt{n})$. Thus by taking, say, $r \approx n^{1/4}$, we may conclude

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

Hence

$$\delta \lessapprox \widetilde{\mu}_3^n(A) = \mathop{\mathbf{Pr}}_{J, x_{\overline{J}}, y_J}[(x_{\overline{J}}, y_J) \in A] = \mathop{\mathbf{E}}_{J, x_{\overline{J}}}\Big[\mathop{\mathbf{Pr}}_{y_J}[(x_{\overline{J}}, y_J) \in A]\Big] = \mathop{\mathbf{E}}_{J, x_{\overline{J}}}\left[\widetilde{\nu}_3^J(A_{x_{\overline{J}}})\right],$$

and so again there must exist a particular J and $x_{\overline{J}}$ such that $\widetilde{\nu}_3^J(A_{x_{\overline{J}}}) \gtrapprox \delta$. 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 DHJ(3) 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 $n \geq 2 \uparrow^{(7)} (1/\delta)$.[[check this]] If $\widetilde{\nu}_3^n(A) \geq \delta$, then either A contains a combinatorial line, or we may pass to a subspace of dimension at least $\log^{(7)} n$ on which $\widetilde{\nu}_3(A) \geq \delta + \Omega(\delta^3)$.*

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[[references]], and yields the equal-slices DHJ(3) 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 $C/7\delta^3$ times, for C a sufficiently large absolute constant, we either find a combinatorial line somewhere along the way, or pass to a $\log^{(C/\delta^3)}(n)$-dimensional subspace on which A has density at least 2δ. (All the while we are assuming the dimension is still "sufficiently large".) Schematically,

$$\delta \xrightarrow{C/\delta^3 \text{ many logs}} 2\delta.$$

Iterating this scheme, we get:

$$\delta \xrightarrow{C/\delta^3 \text{ many logs}} 2\delta \xrightarrow{C/(2\delta)^3 \text{ many logs}} 4\delta \xrightarrow{C/(4\delta)^3 \text{ many logs}} 8\delta \longrightarrow \cdots$$

Of course the density cannot exceed 1, so eventually we *must* find a combinatorial line, after at most $C/\delta^3 + C/8\delta^3 + C/64\delta^3 + \cdots \leq 2C/\delta^3$ 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, $\log^{(2C/\delta^3)} n$, is at least $2 \uparrow^{(7)} (1/\delta)$. And this holds presuming the initial dimension, n, is at least $2 \uparrow\uparrow O(1/\delta^3)$. 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 DHJ(3) to equal-slices DHJ(3) is indeed negligible).

4.3 Extending length-2 lines to length-3 lines

We now look to establish the density increment Theorem 4.4. In this section we will describe our basic strategy for finding lines in a subset $A \subseteq [3]^n$ with $\widetilde{\nu}_3(A) = \delta$. We will show that if the strategy does not work, we can obtain the density increment stated in Theorem 4.4. The idea is to begin with what we already know: probabilistic DHJ(2), 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.

To begin with, though, we will need to be able to apply probabilistic DHJ(2), i.e. Lemma 3.1, to our A. It's not immediately clear how to do this: since $\widetilde{\nu}_3^n$ is supported on nondegenerate strings, we might have $\widetilde{\nu}_3(A) \geq \delta$ and yet $\widetilde{\nu}_2^n(A) = 0$. To evade this we can embed $\widetilde{\nu}_2$ into $\widetilde{\nu}_3^n$ on $n^{1/4}$ random coordinates into $\widetilde{\nu}_3^n$, as in Section 4.1. This will let us pass to a subspace on which $\widetilde{\nu}_2(A) \gtrapprox \delta$. But this is not enough; if $\widetilde{\nu}_3(A)$ becomes tiny, we'll have no hope of extending a length-2 line in A into a length-3 line.[[note explained satisfactorily]]

Looking at the argument in Section 4.1, we see that for a random restriction $(J, x_{\overline{J}})$ to some $n^{1/4}$ coordinates J, the expected $\widetilde{\nu}_2$-density of A is essentially δ and so is the expected $\widetilde{\nu}_3$-density. What we would like is for both densities to be $\gtrapprox \delta$ *simultaneously.* This can be arranged with a density increment trick. The idea is to look at the *variance* of the $\widetilde{\nu}_3$-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 wherein the $\widetilde{\nu}_3$-density

of A increases a little above its mean, to at least $\delta + \delta^{C-1}$. In this case, we have a density increment which is almost good enough for Theorem 4.4 (we can repeat the present argument some 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 $(J, x_{\overline{J}})$, the $\widetilde{\nu}_3$-density of A is very close to δ. Since we know that the expected $\widetilde{\nu}_2$-density of A is δ, it follows that for a noticeable fraction of restrictions the $\widetilde{\nu}_2$-density is $\gtrapprox \delta$. Thus we can find a restriction with both $\widetilde{\nu}_2(A) \gtrapprox \delta$ and $\widetilde{\nu}_3(A) \approx \delta$, as desired. [[not explained very well]].

Having arranged for both $\widetilde{\nu}_2(A), \widetilde{\nu}_3(A) \gtrapprox \delta$, we now come to a key idea in the proof. Suppose we pick a random $z \sim \widetilde{\nu}_3^n$. 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', and the probabilistic DHJ(2) 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*:* $\{z^{3\to 1}, z^{3\to 2}, z\}$.

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. If $L \cap A \neq \emptyset$ then we have found the desired line in A. Unfortunately, all we know is that $\widetilde{\nu}_3(L) \geq \delta^2$ and $\widetilde{\nu}_3(A) \geq \delta$, 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

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

where $\overline{L}$ denotes $[3]^n \setminus L$. In other words, A has a "relative" density increment on the set $\overline{L}$. Unfortunately, $\overline{L}$ is not a combinatorial subspace of $[3]^n$, so we have not established the density increment Theorem 4.4. Still, as we will see, $\overline{L}$ 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

Recall that

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

and hence

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

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 $x^{a\to b} \in C$.

Remark 4.6. This 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 $\overline{C}$ 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 $\overline{L} = B_2$. (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; our last task is 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 the key point is that B_2 is 23-insensitive; hence $z^{2\to 3} \in B_2$ for all $z \in \sigma$, and we conclude that 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 DHJ(2) 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 $|J| = m = \mathrm{MDHJ}_2(\delta^3/100, d)$ 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 $y \in [3]^{\overline{J}}$. These sets $\sigma \times \{y\}$ are disjoint d-dimensional subspaces that we may remove from B and place into the collection $\mathcal{S}$.

We now wish to iterate this argument, but there is a catch; having removed the sets $\sigma \times \{y\}$ from B, it is likely no longer 23-insensitive. Fortunately, its 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 only 23-insensitive 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 contained in B, localized to some new m coordinates J_2. This lets us remove another collection of d-dimensional subspaces from B of mass at least $\tau(\delta, d)3^{-m}$. Since this quantity depends only on δ and d, we need only iterate some $T(\delta, d)$ many times before B's total density drops below $\delta^3/100$. And we may always continue so long as we still have m coordinates on which B is 23-insensitive; we ensure this by assuming $n \geq T(\delta, d)m$. Using the Gunderson-Rödl-Sidirenko theorem, roughly speaking $m = \mathrm{MDHJ}_2(\delta^3/100, d)$ doubly-exponential in d and so $\tau(\delta, d)3^{-m}$ is triply-exponential

in d. In fact, we need to iterate this argument once more, leading to a six-fold exponential, because $\overline{L}$ was not just the single 23-insensitive set B_2 but was rather $B_1 \cup B_2$. In the end, we are able to safely take $d \approx \log^{(7)} n$, as needed to prove Theorem 4.4.[[Ran out of steam here; exposition quality got low]]

5 Equal-slice distributions

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

Definition 5.1. Given a distribution π on Ω, we write $\min(\pi) = \min_{x \in \Omega} \pi[x]$.

A final important concept is the *distance* between two probability measures:

Definition 5.2. For γ and ξ probability distributions on Ω^n, the *total variation distance* $d_{\mathrm{TV}}(\gamma, \xi)$ is defined by

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

If $d_{\mathrm{TV}}(\gamma, \xi) \leq \tau$ 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

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

We abuse terminology by also referring to s as the slice. A slice is *nondegenerate* when $s_a \neq 0$ for all $a \in [k]$. There are $\binom{n+k-1}{k-1}$ different slices, of which $\binom{n-1}{k-1}$ are nondegenerate.[[Justify? Justify later?]]

Definition 5.4. Let $2 \leq k \leq n$. The *equal-nondegenerate-slices* distribution on $[k]^n$, denoted $\widetilde{\nu}_k^n$, generates a string x as follows: First, points $q_1, \ldots, q_n$ are arranged on a circle according to a uniformly random permutation. This also forms n "gaps" between the q_i's. Second, k gaps are chosen, uniformly from the $\binom{n}{k}$ possibilities; points $r_1, \ldots, r_k$ 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 in the r_a arc. See Figure 1 for an example. *ignored part*

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 $|\Omega| = k$.

* *ignored part*

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

Remark 5.7. It is not hard to see[[Honest, although we can clarify]] that drawing $x \sim \widetilde{\nu}_k^n$ is equivalent to first choosing a nondegenerate slice s uniformly from the $\binom{n-1}{k-1}$ possibilities, then choosing x uniformly from this slice.

The following lemma will be important; it essentially arose already in Lemma 3.1:

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

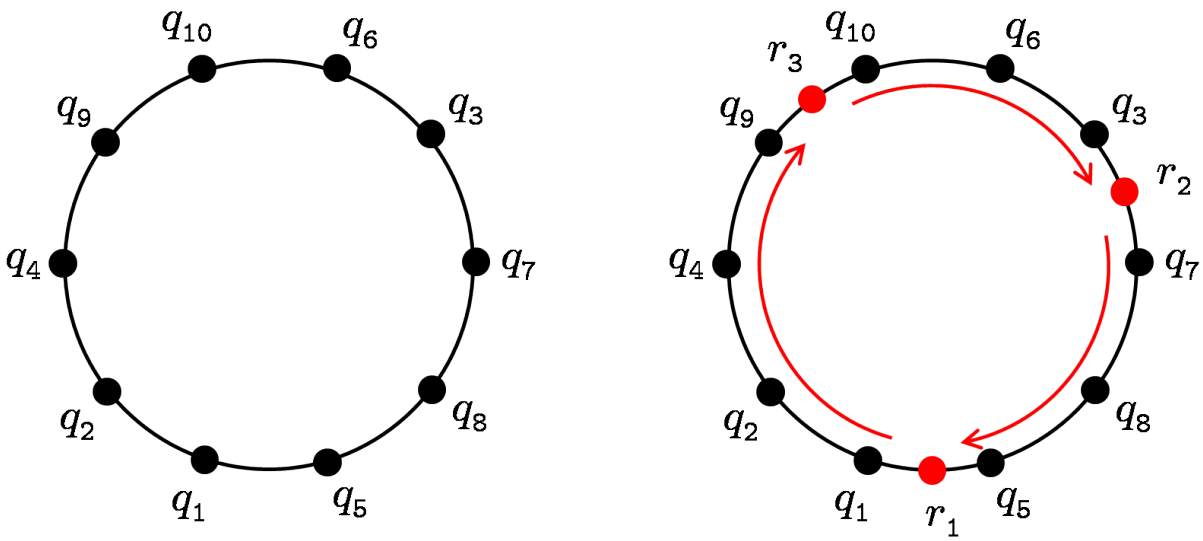

Figure 1: An example of the two stages in drawing from $\widetilde{\nu}_3^{10}$. The resulting string is 1131232213.

Proof. Suppose we are drawing $x \sim \widetilde{\nu}_k^n$ and have chosen the arrangement of $q_1, \ldots, q_n$ 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, $x^{k \to a}$ is obtained simply by "deleting" r_k and interpreting the result as a (correctly distributed) draw from $\widetilde{\nu}_{k-1}^n$.[[not put very eloquently]] □

We end our discussion of the equal-nondegenerate-slices distribution with a handy lemma (which is not true of the equal-slices distribution). 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 $x^{(1,\ldots,m) \to (y_1,\ldots,y_m)} \in [k]^n$. 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 $[n] \to \Omega$.

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

Proof. The effect of forming $y \circ x$ can be thought of as follows: First the points $q_1, \ldots, q_n$ and $r_1, \ldots, r_m$ 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 $s_1, \ldots, s_k$ into them in a random order. Then $z = y \circ x$ is given by setting $z_i = j$ for all i such that q_i is in the s_j "super-arc".

But since $\widetilde{\nu}_m^n$ is symmetric under permutations of $[m]$, the middle stage in this composite process—permuting the r_j arcs—is superfluous. Eliminating it, we see that the s_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 $\widetilde{\nu}_k^n$.[[Again, not put very eloquently]] □

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 $\overline{\nu}_k^n$, generates a string x in a manner similar to that of equal-nondegenerate-slices. The only distinction is that the points $r_1, \ldots, r_k$ 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. It is not hard to see[[honest]] that the process in Definition 5.11 yields a uniformly random (circular) ordering of the $n+k$ points. In particular, the definition of $\overline{\nu}_k^n$ is symmetric with respect ot the symbols $[k]$.

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

Proposition 5.13. *For $2 \leq k \leq n$ we have $d_{\mathrm{TV}}(\widetilde{\nu}_k^n, \overline{\nu}_k^n) \leq k(k-1)/n$.*

Proof. The distribution $\widetilde{\nu}_k^n$ is precisely $\overline{\nu}_k^n$ conditioned on no r_a's ending up adjacent; hence it suffices to show this event occurs with probability at most $k(k-1)/n$. In a draw from $\overline{\nu}_k^n$, the probability that r_a is placed adjacent to one of $r_1, \ldots, r_{a-1}$ 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. □

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

Definition 5.14. Given k points $r_1, \ldots, r_k$ on the the circle of unit circumference, dividing it into arcs, the associated *spacing* is the k-dimensional vector ν 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 spacing* if it is derived by choosing the points $r_1, \ldots, r_k$ independently and uniformly at random on the circle.

Remark 5.15. It is well known (see, e.g. [Rei89, Ch. 1])[[I think the traditional citation here is Pyke '65, but we can save a spot in the biblio by citing Reiss again (which is also a text, whereas Pyke is a bit painful to read)]] that a random spacing ν can equivalently be defined as choosing the vector of probabilities $(\nu[1], \ldots, \nu[k])$ uniformly from the simplex $\{(p_1, \ldots, p_k) : \sum p_a = 1, p_a \geq 0 \ \forall a \in [k]\}$. [[(Or, by taking the spacings of $k-1$ random points in a unit segment.) Since we don't use this fact, I think it's fair to not prove it.]]

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

Proof. In the original definition of a draw $x \sim \overline{\nu}_k^n$, the string x only depends on the joint ordering of the $n+k$ points q_i and r_a around the circle. As we earlier remarked, 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.[[Not worth pointing out that we're disregarding some probability-0 events here.]] 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 spacing ν, and then the formation of x by choice of q_i's is indeed equivalent to drawing x from the product distribution $\nu^{\otimes n}$.[[Not very eloquently put.]] □

Remark 5.17. [[proofs not necessary?]] Drawing $x \sim \overline{\nu}_k^n$ is equivalent to first choosing a slice s uniformly from the $\binom{n+k-1}{k-1}$ possibilities (nondegenerate and degenerate), then choosing x uniformly from this

slice. Hence $\widetilde{\nu}_k^n$ is precisely $\overline{\nu}_k^n$ conditioned on the string being nondegenerate. Equivalently, for each x in slice s we have

$$\overline{\nu}_k^n[x] = 1 \Big/ \binom{n+k-1}{k-1, s_1, \ldots, s_k}. \tag{1}$$

Finally, we will occasionally need the following technical bounds:

Proposition 5.18. *If ν is a random spacing on $[k]$ as in Definition 5.14, then*

$$\mathbf{Pr}[\min(\nu) \le \theta] \le k(k-1)\theta.$$

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

Proposition 5.19. *Assume $2k^2 \le n$. Then we may crudely bound $\min(\overline{\nu}_k^n) \ge k^{-2n}$.* *[[Can actually get something like $k^k n^{-k/2} k^{-n}$ here, but probably best just to get the typographically shortest thing here.]]*

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

$$(\alpha + 1 + \cdots + 1)^{n+k-1} = \sum_{t_0+t_1+\cdots t_k = n+k-1} \binom{n+k-1}{t_0, t_1, \ldots, t_k} \alpha^{t_0} \ge \binom{n+k-1}{k-1, s_1, \ldots, s_k} \alpha^{k-1}.$$

Hence we can lower-bound the expression (1) appearing in Remark 5.17 by

$$\frac{\alpha^{k-1}}{(\alpha+k)^{n+k-1}} = \frac{(k-1)^{k-1}}{(1+\frac{k-1}{n})^{n+k-1} n^{k-1} k^n} \ge \frac{(k-1)^{k-1}}{\exp(k-1+\frac{(k-1)^2}{n}) n^{k-1} k^n} \ge \frac{(k-1)^{k-1}}{\exp(k) n^{k-1} k^n},$$

where the first step used the definition of α, the second used $1 + x \le \exp(x)$, and the last used $(k-1)^2/n \le 1/2 \le 1$, by assumption. [[Here's where we waste a lot of factors.]] Now it's not hard to check[[Honest.]] that when $k \ge 2$ and $n \ge 2k^2$, the above quantity is at least 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 μ_k and $\widetilde{\nu}_k$, 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.

* *ignored part*

Definition 6.1. For $0 \le p \le 1$, 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 $r \le n/2$, we say that J is an *$[r, 4r]$-random subset* of $[n]$ if J is a p-random subset of $[n]$ conditioned on $r \le |J| \le 4r$, with $p = 2r/n$.

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 family $\gamma^m = \widetilde{\nu}_k^m$ or $\gamma^m = \overline{\nu}_k^m$, or will be the product distributions based on a single distribution π on $[k]$, $\gamma^m = \pi^{\otimes m}$.

Lemma 6.3. *Let (γ^m) and (ξ^m) be distribution families. Assume $2\ln n \le r \le n/2$. Let J be an $[r, 4r]$-random subset of $[n]$, let x be drawn from $[k]^{\overline{J}}$ according to $\gamma^{|\overline{J}|}$, and let y be drawn from $[k]^J$ according to $\xi^{|J|}$. The resulting distribution on the composite string $(x, y) \in [k]^n$ has total variation distance from γ^n which can be bounded as follows:*

1. *(Product to equal-slices.) If $\gamma^m = \pi^{\otimes m}$ and $\xi^m = \overline{\nu}_\ell^m$ for $\ell \le k$, the bound is* [[You know, we only need this result for the uniform distribution, in which case we can bound the below by the simpler $2k \cdot r/\sqrt{n}$.]]

$$(2\sqrt{\tfrac{1}{\min(\pi)} - 1}) + 2) \cdot r/\sqrt{n}.$$

2. *(Equal-slices to product.) If $\gamma^m = \overline{\nu}_k^m$ and $\xi^m = \pi^{\otimes m}$, the bound is $4k \cdot r/\sqrt{n}$, independent of π.*

3. *(Equal-slices to equal-slices.) If $\gamma^m = \overline{\nu}_k^m$ and $\xi^m = \overline{\nu}_\ell^m$ for $\ell \le k$, the bound is $4k \cdot r/\sqrt{n}$.*

Although Lemma 6.3 involves the equal-slices distribution, one can convert to equal-nondegenerate-slices if desired using Proposition 5.13.

Since $\overline{\nu}_k^n$ 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

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

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., $\min(\pi) \ne 0$. Suppose π is slightly mixed with ξ, forming $\widehat{\pi}$; specifically, $\widehat{\pi} = (1-p)\pi + p\xi$. Then the associated product distributions $\pi^{\otimes n}$, $\widehat{\pi}^{\otimes n}$ on Ω^n satisfy*

$$d_{\mathrm{TV}}(\pi^{\otimes n}, \widehat{\pi}^{\otimes n}) \le d_{\chi^2}(\pi, \xi) \cdot p\sqrt{n}.$$

Proof. It is a straightforward consequence of Cauchy-Schwarz (see, e.g. [Rei89, p. 101])[[This is the part using $\min(\pi) \ne 0$, by the way.]] that

$$d_{\mathrm{TV}}(\pi^{\otimes n}, \widehat{\pi}^{\otimes n}) \le d_{\chi^2}(\pi, \widehat{\pi}) \cdot \sqrt{n},$$

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

This can be bounded independently of ξ, as follows:

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

$$d_{\mathrm{TV}}(\pi^{\otimes n}, \widehat{\pi}^{\otimes n}) \le \sqrt{\tfrac{1}{\min(\pi)} - 1} \cdot p\sqrt{n},$$

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 $d_{\chi^2}(\pi, \xi) = \sqrt{\frac{1}{\min(\pi)} - 1}$. □

6.2 Proof of Lemma 6.3

Definition 6.7. Let $0 \le p \le 1$ 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 $[k]^{\overline{J}}$ according to $\gamma^{|\overline{J}|}$; and, y is drawn from $[k]^J$ according to $\xi^{|J|}$. We sometimes think of this distribution 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 (large-deviation) bound:[[Citation needed?]]

Fact 6.8. *If J is a p-random subset of $[n]$ with $p = 2r/n$ as in Definition 6.1, then $r \le |J| \le 4r$ 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 $\widehat{\pi}^{\otimes n}$, where $\widehat{\pi}$ 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 $\widetilde{\pi}$ for the (p, π, ξ)-composite distribution on strings in $[k]^n$, we have*

$$d_{\mathrm{TV}}(\pi^{\otimes n}, \widetilde{\pi}) \le \sqrt{\tfrac{1}{\min(\pi)} - 1} \cdot p\sqrt{n}.$$

Recall that for any $\ell \le k$, the equal-slices distribution $\overline{\nu}_\ell^m$ 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. *If $\widetilde{\pi}$ denotes the $(p, \pi, \overline{\nu}_\ell)$-composite distribution on strings in $[k]^n$, where $\ell \le k$, then we have*

$$d_{\mathrm{TV}}(\pi^{\otimes n}, \widetilde{\pi}) \le \sqrt{\tfrac{1}{\min(\pi)} - 1} \cdot p\sqrt{n}.$$

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 $\overline{\xi}$ denote the associated mixture distribution, given by drawing $\kappa \sim \varsigma$ and then drawing from ξ_κ. Then*

$$d_{\mathrm{TV}}(\gamma, \overline{\xi}) \le \mathop{\mathbf{E}}_{\kappa \sim \varsigma}[d_{\mathrm{TV}}(\gamma, \xi_\kappa)].$$

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 $[k]$. Writing γ for the $(p, \overline{\nu}_k, \xi)$-composite distribution on strings in $[k]^n$, we have*

$$d_{\mathrm{TV}}(\overline{\nu}_k^n, \gamma) \le (2k-1)p\sqrt{n}.$$

Proof. Thinking of $\overline{\nu}_k^m$ as the mixture of product distributions $\nu^{\otimes m}$, where ν is a random spacing on $[k]$, Fact 6.12 and Proposition 6.10 imply

$$d_{\mathrm{TV}}(\overline{\nu}_k^n, \gamma) \le \mathop{\mathbf{E}}_{\nu}\left[\sqrt{\tfrac{1}{\min(\nu)} - 1}\right] \cdot p\sqrt{n}.$$

We can upper-bound the expectation[[Undoubtedly someone has worked hard on this $-1/2$th moment of the least spacing before (Devroye '81 or '86 perhaps), but I think it's probably okay to do the following simple thing here]] by

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

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

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

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

$$d_{\mathrm{TV}}(\overline{\nu}_k^n, \gamma') \le (2k-1)p\sqrt{n}.$$

We can now obtain the proof of Lemma 6.3:

Proof. The three statements in Lemma 6.3 essentially follow from Propositions 6.11, 6.13, and 6.14, taking $p = 2r/n$. This would give bounds of $2\sqrt{\frac{1}{\min(\pi)} - 1} \cdot r/\sqrt{n}$, $(4k-2) \cdot r/\sqrt{n}$, and $(4k-2) \cdot r/\sqrt{n}$, respectively. However we need to account for conditioning on $r \le |J| \le 4r$. By Fact 6.8, this conditioning increases the total variation distance by at most $2\exp(-r/4)$. Using the lower bound $r \ge 2\ln n$ from the lemma's hypothesis, this quantity is at most $2r/\sqrt{n}$, completing the proof. □

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 $k \ge 3$, the DHJ theorem is equivalent to the equal-slices DHJ 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 5.13. The other direction has a very similar proof, using Lemma 6.3.2 and we omit it.

Suppose then that we have proven the equal-slices DHJ(k). Let $\mu_k^{\otimes n}(A) \ge \delta$. We use Lemma 6.3, taking $r = n^{1/4}$, letting J be an $[r, 4r]$-random subset of $[n]$, $x \sim \mu_k^{\otimes \overline{J}}$, $y \sim \overline{\nu}_k^J$. (Note that $2\ln n \le r \le n/2$ provided n is at least a small universal constant.) According to the lemma, the resulting distribution on the composite string (x, y) is τ-close to $\mu_k^{\otimes n}$, where

$$\tau = (2\sqrt{k-1} + 2) \cdot r/\sqrt{n} \le 2k \cdot r/\sqrt{n} = 2k/n^{1/4}.$$

This is at most $\delta/3$ provided that $n \ge (6k/\delta)^4$. In that case we conclude

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

and so there must exist a restriction (J, x) with $|J| \geq r$ and $\overline{\nu}_k^J(A_x) \geq 2\delta/3$. By Proposition 5.13 the distributions $\overline{\nu}_k^J$ and $\widetilde{\nu}_k^J$ 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 $n \geq (3k(k-1)/\delta)^4$. In that case $\widetilde{\nu}_k^J(A_x) \geq 2\delta/3 - \delta/3 = \delta/3$. Finally, if $r \geq \mathrm{EDHJ}_k(\delta/3)$ then we may use equal-slices $\mathrm{DHJ}(k)$ to find a combinatorial line in A_x, and hence in A. This holds provided $n \geq \mathrm{EDHJ}_k(\delta/3)^4$. Presuming that $\mathrm{EDHJ}_k(\delta/3) \geq 3k(k-1)/\delta$, the last requirement on n is strictest. Hence we may take $\mathrm{DHJ}_k(()\delta) = \mathrm{EDHJ}_k(\delta/3)^4$.[[not the strongest of course, but it certainly doesn't matter for our bounds]] □

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

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

Proof. We will show that one can take $\mathrm{PMDHJ}_k(\delta, d) = \mathrm{EMDHJ}_k(\delta/2, d)$. Let us write m for this number. Let $x \sim \widetilde{\nu}_m^n$ and $y \sim \widetilde{\nu}_k^m$.[[Surely $m \geq k$.]] By Lemma 5.10, $y \circ x \sim \widetilde{\nu}_k^n$. Hence if $A \subseteq [k]^n$ satisfies $\widetilde{\nu}_k^n(A) \geq \delta$, we get that $\widetilde{\nu}_k^m(A_x) \geq \delta/2$ 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, for all good x we have that A_x contains a d-dimensional subspace. Thus by Proposition 5.19, a randomly drawn subspace $w \sim \widetilde{\nu}_{k+d}^m$ is in A_x with probability at least $(k+d)^{-2m}$. We conclude that with probability at least $\epsilon_k(\delta, d) = (\delta/2)(k+d)^{-2m}$ 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 $\widetilde{\nu}_{k+d}^n$ by Lemma 5.10. □

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 5.13 immediately[[pretty immediately]] yield the following:

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

* *ignored part*

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 $|I| \geq 2\mathrm{MDHJ}_{k-1}(\delta/2, d)$.*

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 $(I, x_{\overline{I}})$ 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 $n \geq 2\mathrm{MDHJ}_{k-1}(\delta/2, d)$.

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 $(J, x_{\overline{J}})$ where $x_{\overline{J}}$ is the all-k's substring, and finally draw $y \sim \mu_{k-1}^{\otimes J}$ and take $z = (x_{\overline{J}}, y)$. We have $\mu_k(B) = \mathbf{E}_J[\mu_{k-1}(B_{x_{\overline{J}}})]\delta$ (where we may think of $B_{x_{\overline{J}}} \subseteq [k-1]^J$). 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[[bother citing?]]. Thus there must exist a particular restriction $(J, x_{\overline{J}})$ where $|J| \geq n/2 \geq \mathrm{MDHJ}_{k-1}(\delta/2, d)$ and $\mu_{k-1}^{\otimes J}(B_{x_{\overline{J}}}) \geq \delta - \exp(-n/48) \geq \delta/2$.[[This uses some extremely mild assumption about $\mathrm{MDHJ}_{k-1}(\delta/2, d)$.]] By the multidimensional $\mathrm{DHJ}(k-1)$ theorem, $B_{x_{\overline{J}}}$ contains a nondegenerate d-dimensional subspace of $[k-1]^J$. Let us write the template for this subspace as $\sigma \in ([k-1] \cup \{k+1, \ldots, k+d\})^J$, meaning that $\sigma^{(k+t)\to\ell} \in B_{x_{\overline{J}}}$ 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 k's, and the resulting template has $\sigma^{(k+t)\to\ell} \in B$ for all $\ell \in [k-1]$. But B is ak-sensitive by assumption, so each string $\sigma^{(k+t)\to k}$ is in B because $\sigma^{(k+t)\to a}$ is. Thus σ is a nondegenerate d-dimensional subspace template in the usual sense with all its strings in B.[[not elegantly explained]] □

8 Line-free sets correlate with insensitive set intersections

In this section we make rigorous the argument described in Section 4.3; we show that a dense subset A either contains a combinatorial line, or it has slight correlation with a union of ab-insensitive sets.

[[I think it would be good to go through the paper and soften explicit constants to $O(\cdot)$'s, for readability.]]

First we reduce to the case where the set is dense under both $\overline{\nu}_k$ and $\overline{\nu}_{k-1}$.

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

$$10k\frac{\ln n}{\sqrt{n}} \leq \theta \leq \delta.$$

Then there exists a restriction $(J, x_{\overline{J}})$ with $|J| \geq (\theta/5k)\sqrt{n}$ such that one of the following two conditions holds:

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

Proof. Let $r = \lceil(\theta/5k)\sqrt{n}\rceil$. As in Lemma 6.3, let J be an $[r, 4r]$-random subset of $[n]$ and let x be drawn from $\overline{\nu}_k^{\overline{J}}$. If $\mathbf{E}_{J,x}[\overline{\nu}_k(A_x)^2] \geq (\delta+\theta)^2$, then there must exist some restriction (J, x) with $|J| \geq r$ and $\overline{\nu}_k^J(A_x) \geq \delta + \theta$. In this case, conclusion 1 above holds. We henceforth assume

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

and show that conclusion 2 holds.

* Since $\overline{\nu}_k(A) \geq \delta$, two applications of Lemma 6.3.3 yield *ignored part*

$$\mathop{\mathbf{E}}_{J,x}[\overline{\nu}_k(A_x)] \geq \delta - 4kr/\sqrt{n} \geq \delta - \theta \tag{3}$$

$$\mathop{\mathbf{E}}_{J,x}[\overline{\nu}_{k-1}(A_x)] \geq \delta - 4kr/\sqrt{n} \geq \delta - \theta, \tag{4}$$

where we've used our definition of r (and the hypothesis $\theta \geq (10k\ln n)/\sqrt{n}$ ensures that $r \geq 2\ln n$ as required for the lemma.) Combining (2), (3) gives

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

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

$$\overline{\nu}_k(A_x) \geq \delta - \theta - 2\theta^{1/4}\delta^{1/2} \geq \delta - 3\theta^{1/4}\delta^{1/2} \tag{5}$$

(using $\theta \leq \delta$). On the other hand, from (4) we immediately deduce that with probability at least $2\theta^{1/2}$ over the choice of (J, x) we have

$$\overline{\nu}_{k-1}(A_x) \geq \delta - \theta - 2\theta^{1/2} \geq \delta - 3\theta^{1/2}. \tag{6}$$

Thus with probability at least $2\theta^{1/2} - \theta^{1/2} > 0$ over the choice of (J, x), we have both (5), (6), establishing conclusion 2 of the lemma. □

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 $\widetilde{\nu}_k(A) \geq \delta_k$ and $\widetilde{\nu}_{k-1}(A) \geq \delta_{k-1}$. Finally, assume $n \geq \mathrm{PDHJ}_{k-1}(\delta_{k-1})$. Then either A contains a combinatorial line (of length k), or there exists $B \subseteq [k]^n$ with $\widetilde{\nu}_k(B) \geq \delta_k$ and*

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

such that

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

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

Proof. For $a \in [k-1]$, let $L_a = \{x \in [k]^n : x^{k \to a} \in A\}$, an ak-insensitive set. Let $L = L_1 \cap L_2 \cap \cdots \cap L_{k-1}$. 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 $\widetilde{\nu}_k(L \setminus L') = 0$.

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 $A \subseteq \overline{L'} = [k]^n \setminus L'$. Since $\widetilde{\nu}_{k-1}(A) \geq \delta_{k-1}$ and $n \geq \mathrm{PDHJ}_{k-1}(\delta_{k-1})$, the probabilistic DHJ($k-1$) theorem implies that $\widetilde{\nu}_k(L') = \widetilde{\nu}_k(L) \geq \epsilon_{k-1}(\delta_{k-1})$. It follows that

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

The proof is completed by taking $B = \overline{L}$; this has $\widetilde{\nu}_k(B) = \widetilde{\nu}_k(\overline{L'}) \geq \delta_k$ because $\overline{L'} \supseteq A$. □

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

Lemma 8.3. *Suppose $A \subseteq B \subseteq [k]^n$ satisfy $\widetilde{\nu}_k(B) = \beta$ and $\widetilde{\nu}_k(A) \geq \delta \cdot \widetilde{\nu}_k(B)$. Let $0 < \eta < 3\delta$ 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 $\mu_k(A_{x_{\overline{J}}}) \geq (\delta - \eta)\mu_k(B_{x_{\overline{J}}})$.* *[[And if B was a union of ij-insensitive sets before...]]*

Proof. Let J be an $[r, 4r]$-random subset of $[n]$, let $x \sim \overline{\nu}_k^{\overline{J}}$, and let $y \sim \mu_k^J$. From Lemma 6.3, the distribution on (x, y) is $(4kr/\sqrt{n})$-close to $\overline{\nu}_k^n$. Further, from (5.13) the total variation distance between $\overline{\nu}_k^n$ and $\widetilde{\nu}_k^n$ is at most $k(k-1)/n \leq kr/\sqrt{n}$[[uses $k \leq \sqrt{n}$ or so]]. By choice of r, the combined distance bound $5kr/\sqrt{n}$ is at most $(\eta/3)\beta$. We therefore have

$$\mathop{\mathbf{E}}_{J,x}[\mu_k(A_x)] \geq \delta\beta - (\eta/3)\beta, \qquad \mathop{\mathbf{E}}_{J,x}[\mu_k(B_x)] \leq \beta + (\eta/3)\beta. \tag{7}$$

Let H be the event that $\mu_k(B_x) \geq (\eta/3)\beta$. On one hand, since $A_x \subseteq B_x$ always we have

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

On the other hand, we have

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

Combining these two deductions with (7) yields

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

Thus

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

whence

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

where we used $(\delta - 2\eta/3)/(1+\eta/3) \ge \delta - \eta$. Thus there exist some (J, x) where H occurs and also $\mu_k(A_x) \ge (\delta-\eta)\mu_k(B_x)$, completing the proof. □

Finally, the set B we get from Theorem 8.2 is a union of insensitive sets; we would prefer an intersection:

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

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

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 $k \ge 3$, $d \ge 1$, $0 < \eta < 1/2$. Let $C \subseteq [k]^n$ be ab-insensitive on I, and assume*

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

where

$$m = 2\mathrm{MDHJ}_{k-1}(\eta/4, d).$$

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

Proof. [[Maybe change all the $\mu_k^{\otimes n}$'s just to μ's, for typographical simplicity.]] The proof proceeds in "rounds", $t = 1, 2, 3, \ldots$. In each round, we remove some disjoint subspaces from C and put them into $\mathcal{S}$; and, the set I shrinks by m coordinates. We will show that after the tth round,

$$\mu_k^{\otimes n}(C) \le \Big(1 - \big(k(k+d)\big)^{-m}\Big)^t. \tag{8}$$

We continue the process until $\mu_k^{\otimes n}(C) \le \eta$, at which point we may stop and set $E = C$. Because of (8), the process stops after at most $T = \big(k(k+d)\big)^m \ln(1/\eta)$ rounds. Since we insisted $|I| \ge mT$ initially, the set I never "runs out of coordinates".

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

$$\eta < \alpha \le \Big(1 - \big(k(k+d)\big)^{-m}\Big)^{t-1}.$$

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

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

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 a $\alpha/2$ probability mass of y's (under $\mu_k^{\otimes \overline{J}}$); 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 $|J| = m = 2\mathrm{MDHJ}_{k-1}((\eta/2)/2, d) \geq 2\mathrm{MDHJ}_{k-1}((\alpha/2)/2, d)$, 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

$$\mathop{\mathbf{Pr}}_{y \sim \mu_k^{\otimes \overline{J}}}[\rho_0 \subseteq C_y] \geq \frac{\alpha}{2(k+d)^m}. \tag{9}$$

Let $R \subseteq [k]^{\overline{J}}$ be the set of y's with $\rho_0 \subseteq C_y$. Since C is ab-insensitive on I, it is easy to see[[Honest.]] 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 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

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

as required to establish (8) inductively. □

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 $d^{3m} \geq m\big(k(k+d)\big)^m \ln(1/\eta)$. We use the bounds $k \leq d$ and $m \geq 4/\eta$ (which is certainly necessary), and hence must show $d^{3m} \geq m \ln(m/4)(2d^2)^m$. This holds for every $d \geq 3$ and $m > 4$.[[I checked]] □

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 $C = C_1 \cap \cdots \cap C_\ell$, where each C_s is $a_s b_s$-insensitive. If $n \geq f^{(\ell)}(d)$[[that's f iterated ℓ times]] 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 $C' = (C_1 \cap \cdots \cap C_{\ell-1})$ 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 $[k]^{f(d)}$[[justify?]]. Thus we may apply Corollary 9.2 and partition D_σ into a collection $\mathcal{S}_\sigma$ of nondegenerate d-dimensional subspaces, along with an error set E_σ satisfying "$\mu_k^\sigma(E_\sigma)$" $\leq \eta$.[[bad invented notation]] Note that in the original space $[k]^n$ we have $\mu_k^{\otimes n}(E_\sigma) \leq \eta \cdot \mu_k^{\otimes n}(\sigma)$.[[$= \eta \cdot k^{f(d)-n}$. There's a slight subtlety here.]] Hence we may complete the induction by taking $\mathcal{S}$ to be $\{D_\sigma : \sigma \in \mathcal{S}'\}$[[Point out that a d-dim subspace of a subspace is a d-dim subspace?]] and E to be $E' \cup \bigcup\{E_\sigma : \sigma \in \mathcal{S}'\}$, observing that $\mu_k^{\otimes n}(E) \leq (\ell-1)\eta + \eta(\sum_{\sigma \in \mathcal{S}'} \mu_k^{\otimes n}(\sigma)) \leq \ell\eta$ as required. □

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 due to the Gunderson-Rödl-Sidirenko theorem (and the probabilistic Sperner theorem).

10.1 The $k=3$ case

As described in Section 4.2, to establish DHJ(3) it suffices to establish the density increment theorem stated therein. (One also needs the rigorous reduction from DHJ(3) to the equal-slices version, Proposition 7.1.) For the reader's convenience, we restate the density increment theorem:

Theorem 4.4. *Assume $n \geq 2 \uparrow^{(7)} (1/\delta)$. If $\widetilde{\nu}_3^n(A) \geq \delta$, then either A contains a combinatorial line, or we may pass to a subspace of dimension at least $\log^{(7)} n$ on which $\widetilde{\nu}_3(A) \geq \delta + \Omega(\delta^3)$.*

Proof. Throughout the proof we may assume $\delta \leq 2/3$, as otherwise DHJ(3) is easy. By Proposition 5.13, we have $\overline{\nu}_3^n(A) \geq \delta - 6/n \geq \delta - \delta^{50}$, by our assumption on n.

Let $\theta = 2\delta^{50}$. (We will not be especially economical when it comes to parameters which do not matter.) Applying Lemma 8.1[[not bothering to point out that the hypothesis holds]], we pass to subspace of dimension at least $(\delta^{50}/7.5)\sqrt{n} \geq n^{1/3}$ on which either $\overline{\nu}_3(A) \geq \delta + \delta^{50}$, or both $\overline{\nu}_3(A) \geq \delta - 4\delta^{13}$ and $\overline{\nu}_2(A) \geq \delta - 4\delta^{25}$. We repeat this step until either the second conclusion obtains or until the step has been repeated $\lfloor 1/\delta^{47} \rfloor$ times. In either case, the dimension n becomes no smaller than $n' = n \uparrow 3 \uparrow (-1/\delta^{47})$. Having done this, we either have $\overline{\nu}_3(A) \geq \delta + \Omega(\delta^3)$ and may stop (as $n' \gg \log^{(7)} n$ by the assumption on n), or we have $\overline{\nu}_3(A), \overline{\nu}_2(A) \geq \delta - 4.5\delta^{13}$. [[4.5 is weird, I know, I messed up the constants earlier and was too lazy to readjust them]] With another application of Proposition 5.13 we may obtain $\widetilde{\nu}_3(A), \widetilde{\nu}_2(A) \geq \delta - 5\delta^{13}$, reducing n' to at most, say, $n \uparrow 3 \uparrow (-1/\delta^{48})$.

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 $n' \geq \epsilon_2(\delta_{k-1}) = \Theta(1/\delta^2)$, and thus we either obtain a combinatorial line in A, or a set $B \subseteq [3]^{n'}$ with $\widetilde{\nu}_3(B) \geq \delta - 5\delta^{13}$ and

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

where we used $\epsilon_2(\delta_{k-1}) = \delta_{k-1}^2/2$. [[check these calculations]] 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 $\eta = \delta^3/8$. Note that the resulting $r = (\beta\eta/15k)\sqrt{n'}$ indeed satisfies $r \gg 2\ln n$, by the initial assumption on n.[[I think I checked.]] We now pass to a restriction on some $n' \geq r \geq n \uparrow 3 \uparrow (-1/\delta^{49})$ coordinates on which we have, say,

$$\mu_3(B) \geq \delta^4/50, \quad \mu_3(A) \geq (\delta + \delta^3/8)\mu_3(B). \tag{10}$$

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 and check the hypothesis on n' later. We can therefore 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

$$\mu_3(E) \leq 2\eta + 2\eta \leq \frac{\delta^7}{1000} \leq \frac{\delta^3}{20}\mu_3(B). \tag{11}$$

We now simply discard the strings in E from A and B. Then B is perfectly partitioned into disjoint d-dimensional subspaces, and some arithmetic shows that, whereas we had $\mu_3(A)/\mu_3(B) \geq \delta + \delta^3/8$ in (10), we now still have

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

and the fraction on the right is easily seen to be at most $\delta^3/19$, using (11). Thus we have $\mu_3(A)/\mu_3(B) \geq \delta + \delta^3/20$, with B 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 execute the partitioning argument, we required that

$$n' \geq \left(d^{3m(d)}\right)^{3m(d^{3m(d)})}, \tag{12}$$

where

$$m(d) = 2\mathrm{MDHJ}_2(\delta^7/16000, d) = 2(10d)^{d2^d}(16000/\delta^7)^{2^d},$$

using the Gunderson-Rödl-Sidirenko theorem. We will ensure that

$$d \geq 800/\delta^7, \tag{13}$$

so that we may bound $m(d) \leq (20d)^{2d2^d} \leq 2 \uparrow 4 \uparrow d$. Then the right-hand side of (12) is at most $2 \uparrow^{(6)} (3d)$, and with $n' \geq n \uparrow 3 \uparrow (-1/\delta^{49})$ we get that (12) is satisfied so long as $d \ll \log^{(6)} n$, using also the initial assumption on n.[[Getting pretty out of control at this point.]] Thus we set, say,

$$d \approx (\log^{(7)} n)^{100}$$

and we've satisfied (13) using the initial assumption $n \geq 2 \uparrow^{(7)} (1/\delta)$.

It remains to note that, having passed to a $(\log^{(7)} n)^{100}$-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 $\mu_3(A) \geq \delta + \delta^3/40$ via the embedding argument sketched in Section 4.1, justified by Lemma 6.3. □

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[[I think]], the resulting bounds for DHJ(k) are of Ackermann-type [[I think]], and we do not work them out precisely.[[cop-out]].

Theorem 10.1. *The general k case of the density Hales–Jewett theorem follows from the $k-1$ case.*

Proof. (Sketch.) Having proved DHJ$(k-1)$, we deduce the multidimensional version via Proposition 2.3, the equal-slices version via Proposition 7.1, and the probabilistic version from Proposition 7.2 (only the $d = 1$ case is necessary). 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, we can either find a line in A or can obtain a density increment of $\Omega(\epsilon)$ on a combinatorial subspace of dimension d', where $\epsilon = \delta \cdot \epsilon_{k-1}(\delta/2)$ depends only on δ and k, and where $d' = \omega_1(n)$.

Beginning with $\widetilde{\nu}_k(A) \geq \delta$, we choose $\theta = 2\epsilon^C$ for some large constant C; note that θ only depends on δ and k. We pass from $\widetilde{\nu}_k(A)$ to $\overline{\nu}_k(A)$ via Proposition 5.13, incurring density loss only ϵ^c by assuming n large enough (n becomes slightly smaller here). We next apply Lemma 8.1. This gives us either a density increment of ϵ^c, or achieves $\overline{\nu}_k(A), \overline{\nu}_{k-1}(A) \geq \delta - \epsilon^c$ for some smaller but still

large constant c. We can repeat this step $1/\epsilon^c$ times, eventually achieving a density increment of $\Omega(\epsilon)$ and stopping if the first outcome keeps happening; otherwise achieving $\overline{\nu}_k(A), \overline{\nu}_{k-1}(A) \geq \delta - O(\epsilon^c)$. This brings n down to some fractional power n', but the fraction depends only on ϵ^c and hence on δ and k; thus we may continue to assume it is "large enough". [[guess I'm getting pretty sketchy here]] This lets us ensure that $\widetilde{\nu}_k(A), \widetilde{\nu}_{k-1}(A) \geq \delta - O(\epsilon^c) \geq \delta/2$.

We now apply Theorem 8.2 with $\delta_k = \delta_{k-1} = \delta/2$. This requires n' to be larger than $\mathrm{PDHJ}_{k-1}(\delta/2)$, which will be a very large but still "constant" quantity, depending only on δ and k. Having applied the theorem, we either get a combinatorial line, or a density increment of $\delta_k \cdot \epsilon_{k-1}(\delta/2) = \Omega(\epsilon)$ on some set B. Note that $\Omega(\epsilon) \gg O(\epsilon^c)$, so are able to obtain the desired relative density increment. The set B has $\widetilde{\nu}_k(B) \geq \delta/2$ 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 dimension n' continues to decrease to some r, but this is still $\omega_n(1)$ as we have only involved "constant" factors depending on 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 suffice now to analyze what parameter d we may choose; the remainder of the proof is essentially the same as in the DHJ(3) case. At this point, the original n has shrunk to some n' which is some fractional power of n depending on δ and k only, and this n' is required to be large enough compared to $\epsilon_{k-1}(\delta/2)$, a very large function of δ and k only. The function $m(d)$ arising in the corollary depends on η and hence on δ and k, and also on d; the dependence of $f(d) = d^{3m(d)}$ on d will be quite bad due to the inductive argument Proposition 2.3. Still, when we iterate it $\ell = k$ times, the function $f^{(k)}(d)$ is just another (huge) function of k, δ, and d. We will then have some demand on how large n' needs to be as a function of d. Crucially, we may take d to be some function $\omega_n(1)$ which satisfies this demand.

Finally, we need to take our uniform-distribution density increment of $\Omega(\epsilon)$ on a d-dimensional subspace, and produce an equal-slices density increment of $\Omega(\epsilon)$. This will require reducing d further to some d', a function of d and ϵ (hence δ and k), but this d' will still be $\omega_n(1)$. [[this is all poorly explained and probably wrongly explained in several places :) luckily it's right overall :)]] □

11 Discussion

[[Perhaps comparison with other proofs, discussion of how one might improve the argument, or at least discussion on how to improve the quantitatives of the present argument...]]

References

[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 Mathematique*, 57:64–119, 1991.

[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.

[Rei89] Rolf-Dieter Reiss. *Approximate distributions of order statistics.* Springer-Verlag, 1989.

[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 Erdös*, pages 391–396. Cambridge University Press, 1990.

[Sze75] Endre Szemerédi. On sets of integers containing no k elements in arithmetic progression. *Acta Arithmetica*, 27:199–245, 1975.

[vdW27] Bartel L. van der Waerden. Beweis einer Baudetschen Vermutung. *Nieuw Archief voor Wiskunde*, 15:212–216, 1927.