

## Gowers's Weblog

Mathematics related discussions

« [Questions of procedure](#)  
[Why this particular problem?](#) »

### A combinatorial approach to density Hales-Jewett

By gowers

Here then is the project that I hope it might be possible to carry out by means of a large collaboration in which no single person has to work all that hard (except perhaps when it comes to writing up). Let me begin by repeating a number of qualifications, just so that it is clear what the aim is.

1. It is *not* the case that the aim of the project is to find a combinatorial proof of the density Hales-Jewett theorem when  $k = 3$ . I would love it if that was the result, but the actual aim is more modest: it is *either* to prove that a certain approach to that theorem (which I shall soon explain) works, *or* to give a very convincing argument that that approach cannot work. (I shall have a few remarks later about what such a convincing argument might conceivably look like.)

2. I think that the chances of success even for this more modest aim are substantially less than 100%. So let me be less fussy still. I will regard the experiment as a success if it leads to anything that could count as genuine progress towards an understanding of the problem. However, that success has a quantitative aspect to it: what I am really interested in is progress that results from multiple collaboration. If what actually happens is that several people make intelligent remarks, but without building on each other's remarks, then what will have been gained will still be worthwhile but it will be less interesting from the point of view of this experiment.

3. I've thought of one more ground rule, and also a general guideline that clarifies a couple of the existing rules. The extra rule is that there needs to be some possibility for the experiment to be declared to have finished. What I suggest here is that if a consensus emerges that no further progress is likely, I (or perhaps somebody else) will write a summary of the main conclusions that have been reached, and it will then become public property in the way that a conventional mathematics preprint would be. That is, others would be welcome to make use of it in their own publications if they wanted to. Probably the summary would be posted on the ArXiv (after people had had a chance to suggest changes) but not conventionally published. The guideline, which is designed to stop people going away and doing a lot of work in private — the whole point is that we should all display our thought processes — is that you should not write anything that requires you to go away and do some calculations on a piece of paper. As I said in the other guidelines, if all goes well there will eventually be a place for detailed working out of ideas, but people should do that only after first establishing that that would be welcomed.

In order to encourage the style I'm hoping for, I shall practise what I preach, in a sense anyway, by giving my initial thoughts on the problem in small chunks. That is, I'll write them as though they were a lot of comments that could in theory have been made by several different people instead of just one. I'll number them with letters of the alphabet (so that they don't get confused with the numbered comments on this post).

A. Might it be possible to prove the density Hales-Jewett theorem for  $k = 3$  by imitating the triangle-removal proof that dense subsets of  $[n]^2$  contain corners?

- B. If one were going to do that, what would the natural analogue of the tripartite graph associated with a dense subset be? To put that question more precisely, if we've got a dense subset  $A$  of  $[3]^n$ , then how might we define a tripartite graph  $G$  in such a way that triangles in  $G$  correspond to (possibly degenerate) combinatorial lines?
- C. It's perhaps worth pointing out that there is a natural definition of a degenerate combinatorial line: it's one where the set of coordinates that vary from 1 to 3 is empty.
- D. If we just look at the vertex sets  $X$  and  $Y$  in the case of subsets  $A \subset [n]^2$ , we have a very simple definition for an edge: it's a pair  $(x, y)$  that belongs to  $A$ .
- E. Does  $[3]^n$  split up nicely as a Cartesian product?
- F. Well, you can write it as  $[3]^r \times [3]^s$ , but that doesn't help much.
- G. *Why* doesn't it help?
- H. Because if we represent a typical point as  $(x, y)$  with  $x \in [3]^r$  and  $y \in [3]^s$ , then a Hales-Jewett line is not a triple of the form  $(x, y), (x + d, y), (x, y + d)$ . That's partly because we can't even make sense of adding  $d$ , but more fundamentally because even if we could invent a useful definition of addition in this context, we still wouldn't get a Hales-Jewett line.
- I. It was clear in advance that that was not going to work, because the splitting up as a Cartesian product was completely non-canonical.
- J. By the way, there's a nice way of deducing the corners result from density Hales-Jewett. Let me quickly explain it, omitting a few details. Suppose we take a set  $A \subset [3]^n$  with the special property that whether or not a sequence  $x$  belongs to  $A$  depends only on the numbers of 1s, 2s and 3s in  $x$ . That is, let  $T$  be the set of all triples  $(r, s, t)$  of non-negative integers such that  $r + s + t = n$ , let  $S$  be some subset of  $T$ , and let  $A$  be the set of sequences  $x$  such that  $(n_1(x), n_2(x), n_3(x)) \in S$ , where  $n_i(x)$  stands for the number of  $i$ s in  $x$ . Now suppose you have a combinatorial line  $L$ . If  $d$  is the number of variable coordinates in  $L$ , then  $S$  must contain a triple of points of the form  $(r + d, s, t), (r, s + d, t), (r, s, t + d)$  (where  $r, s$  and  $t$  are the numbers of 1s, 2s and 3s amongst the fixed coordinates of  $L$ ).
- Now  $T$  is a large triangular subset of a triangular grid, and one can (if one really feels like it) shear it so that it becomes a right-angled triangle instead of an isosceles triangle. In this way it is not hard to show that the assertion that every dense subset of  $T$  contains the vertices of an equilateral triangle of the form  $(r + d, s, t), (r, s + d, t), (r, s, t + d)$  is equivalent to the corners result.
- Unfortunately it's not the case that dense subsets of  $T$  correspond to dense subsets of  $[3]^n$  (because almost every point in  $[3]^n$  has similar numbers of 1s, 2s and 3s) but there are dodges for getting round that.
- K. That's very helpful, because we almost certainly want whatever definition we come up with for the graph  $G$  to give rise to the definition that we used for  $[n]^2$  when the subset of  $[3]^n$  is of the special kind that depends only on the numbers of coordinates with the three possible values.
- L. All right then. That tells us that for subsets of  $[3]^n$  that just depend on numbers of different types of coordinates, we want edges of the  $XY$  graph to correspond in some sensible way to points in  $A$  that have a particular number of 1s and a particular number of 2s.
- M. That suggests that in the general case we want some data about the 1s and 2s in  $x$  to determine  $x$ , just as the number of 1s and 2s determines the numbers of all three types of coordinates. And of course this is trivial: instead of looking at the *numbers* of the three types, look at the *sets where they occur*.
- N. Indeed, there is an obvious tripartite graph we can define. Let its vertex sets, which I'll call  $X, Y$  and  $Z$ , all be copies of the power set of  $[n]$ : i.e., in each vertex set the vertices are subsets of  $[n]$ . Given two disjoint sets  $U \in X$  and  $V \in Y$ , there is a unique point  $x \in [3]^n$  that takes the value 1 in  $U$ , the value 2 in  $V$ , and the value 3 everywhere else. Join  $U$  to  $V$  if that

point  $x$  belongs to the given subset  $A \subset [3]^n$ . And do the obviously corresponding things for the  $YZ$  and  $XZ$  graphs.

O. Does that work? That is, if we have a triangle in that graph, does it correspond to a combinatorial line?

P. Yes. Here's a quick proof. Suppose we have  $U \in X$ ,  $V \in Y$  and  $W \in Z$ , and suppose that they form a triangle in the graph. By the way we defined the edges, the sets  $U$ ,  $V$  and  $W$  are disjoint. For  $i = 1, 2, 3$  let's write  $\phi_i(x)$  for the set of  $j$  such that  $x_j = i$ . Then in the set  $A$  we can find a point  $a$  such that  $\phi_1(a) = U$  and  $\phi_2(a) = V$ , and also a point  $b$  in  $A$  with  $\phi_2(b) = V$  and  $\phi_3(b) = W$ , and finally a point  $c$  in  $A$  with  $\phi_1(c) = U$  and  $\phi_3(c) = W$ . Now let  $T = [n] \setminus (U \cup V \cup W)$ . Since for every point  $x$  we have  $\phi_1(x) \cup \phi_2(x) \cup \phi_3(x) = [n]$ , we find that  $\phi_3(a) = W \cup T$ ,  $\phi_1(b) = V \cup T$  and  $\phi_2(c) = U \cup T$ . In other words, the three sequences  $a$ ,  $b$  and  $c$ , which all lie in  $A$ , all take values 1 on  $U$ , 2 on  $V$  and 3 on  $W$ , and  $b$  is 1 everywhere on  $T$ ,  $c$  is 2 everywhere on  $T$  and  $a$  is 3 everywhere on  $T$ . This gives us a combinatorial line except in the degenerate case  $T = \emptyset$ .

Q. Wow. That means we're done by triangle removal doesn't it?

R. Unfortunately it doesn't. The problem is that that tripartite graph is not dense. In fact, it's not even close to dense. To see why not, just pick a random pair of subsets  $U$  and  $V$ . They cannot be joined in the  $XY$  graph unless they are disjoint, and the probability of that happening is absolutely tiny.

S. That sounds a bit bad, but aren't there sparse versions of Szemerédi's regularity lemma?

T. I'm not sure it's all that likely that there's a version that applies when the density is *exponentially* small.

U. That's a misleading way of putting it when the graph itself is exponentially large. If we let  $N = 2^n$ , so that there are  $N$  vertices in each of the three vertex sets, then the density of the graph is going to be of the form  $N^{-\alpha}$  for some constant  $\alpha$ . But that's still pretty sparse, it has to be said.

V. Here's another problem. The existing sparse regularity lemmas tend to assume that you're sitting inside a highly random sparse graph, but are dense relative to that graph. In our case, the natural graph  $\Gamma$  that we're sitting inside is the one where you join two sets if and only if they are disjoint and ignore whether they define a sequence that belongs to  $A$ . It's possible that the graph defined above will turn out to be dense relative to this graph. However,  $\Gamma$  is not by any stretch of the imagination random, or quasirandom. For example, if you look at the  $XY$  part, you can define  $X'$  to be the subset of  $X$  where all sets have size at most  $n - C\sqrt{n}$ , and the same for  $Y'$ , and it's not hard to check that the density of the part of the graph from  $X'$  to  $Y'$  is substantially greater. (Basically, conditioning on sets being a bit smaller than average makes them quite a lot more likely to be disjoint.)

W. Yes, but at least the graph  $\Gamma$  is very concretely defined. Perhaps one could do some kind of regularity argument relative to  $\Gamma$  that exploited the fact that it has a nice simple definition.

X. Here's a small and slightly encouraging observation. What does a typical  $XY$  edge look like? Well, there's a one-to-one correspondence between  $XY$  edges and sequences in  $[3]^n$ . Since a typical sequence in  $[3]^n$  has roughly equal numbers of 1s, 2s and 3s, a typical edge in the  $XY$  graph consists of two disjoint sets of size about  $n/3$ .

Y. Unfortunately it's still very sparse if we somehow weight our sets so that almost all of them have size about  $n/3$ , since two sets of size  $n/3$  have a tiny probability of being disjoint.

Z. Aren't there tricks we could do to find dense bits of the graph  $\Gamma$ ? For example, we know that if  $U$  and  $V$  are random *small* sets (of size  $n^{1/3}$ , say) then with high probability they are disjoint. Might there be some way of restricting to "the small part of  $\Gamma$ "?

AA. How could one do that when almost all sequences have roughly equal numbers of 1s, 2s and 3s? The dense set  $A$  might consist just of well-balanced sequences.

BB. It's not as impossible as it sounds, because we can use an averaging trick. Suppose we want to restrict  $X$ ,  $Y$  and  $Z$  to

small sets. What we can do is this. First we randomly choose a very small subset  $P$  of  $[n]$ . Then for every  $m \in [n] \setminus P$  we randomly choose an element of  $\{1, 2, 3\}$ . Finally, we do everything we want inside  $P$ . That is, we define a graph by joining two subsets  $U$  and  $V$  of  $P$  if they are disjoint, and define associated sequences by letting  $U$  and  $V$  determine the values inside  $P$  and the random choice of values outside  $P$  give us the rest. This gives us a structure of the kind we are talking about, and the average density of points in  $A$  in this structure will be close to the density of  $A$  if  $P$  is small enough, because random sequences that are unbalanced inside a random small set  $P$  will not be especially unbalanced globally.

CC. That's an interesting thought, but I think it leads to a problem. Suppose our  $XY$  graph consists of pairs  $(U, V)$  of disjoint small subsets of  $P$ , and similarly for our  $YZ$  graph and our  $XZ$  graph. Where are the degenerate triangles? You can't have a sequence with very few 1s in  $P$ , very few 2s in  $P$  and very few 3s in  $P$  all at the same time.

DD. That's a serious objection, but I still think the averaging trick could be quite useful for this problem.

EE. Here's a different idea. The big problem is not so much that  $\Gamma$  is sparse as that it is not quasirandom. But perhaps we could *make* it quasirandom by putting weights on the edges (and perhaps if necessary the vertices as well, since the averaging trick above will enable us to get sets of positive density even with funny vertex weights). The basic idea would be to penalize edges that joined small sets together, because somehow those sets were "cheating" by being small. Is there some natural set of weights that one can just write down?

*[The history of this idea is that earlier this afternoon I had got up to Y when I had to go and fetch my daughter from school. While walking there I couldn't help thinking about the problem and this idea, which I have not explored, occurred to me. So even the effort of writing down my thoughts has been quite stimulating. Because I haven't yet thought about this idea properly, I'm still in that delicious state of thinking it might be the key to the whole problem. But I've also been in this game long enough to know that the chances of that are very small. It's tempting to investigate it, but I'm not going to as this is also an ideal opportunity to throw out an idea and wait for somebody else to destroy it (or, just perhaps, to develop it) instead.]*

FF. A natural way to find such weights would be to start by trying to make every vertex have constant (weighted) degree.

GG. It's easy to see that you can't make  $\Gamma$  quasirandom by weighting the edges. For example, suppose you let  $\mathcal{A} \subset X$  and  $\mathcal{B} \subset Y$  both be the set of all subsets of  $[n]$  that contain 1. These two sets have density  $1/2$  and yet there is no edge of  $\Gamma$  between them.

*[The history of this idea is that although I resolved not to think about the idea put forward in EE., I was on a car journey and I just couldn't help it. That goes for the next couple of responses to it too, which I could help even less because at that point I was starting to get worried that the whole thing was indeed hopeless.]*

HH. That sounds pretty bad.

II. Is it really as bad as all that? I can see that it's a problem if you have very dense parts of the graph, because then you don't have a global upper bound for the density of subgraphs when you are trying to prove a regularity lemma. But if you have parts of the graph that are very sparse, or even entirely free of edges, it's not so obviously a problem because there could still be an upper bound. In other words, perhaps it's enough for  $\Gamma$  to be "upper quasirandom" rather than quasirandom.

JJ. Another idea is that you just try to do everything relative to  $\Gamma$ . That is, you define the density of an induced subgraph of  $G$  to be the number of edges in that subgraph divided by the number of edges in the corresponding induced subgraph of  $\Gamma$ . Could that work?

KK. Going back to vertex weights, here are two remarks. First, the averaging trick shows that for more or less any weighting on the vertices, if you can prove the theorem with that weighting (which affects which sets count as dense) then you can prove the theorem with the uniform weighting. I could write a formal lemma to this effect if anybody wanted, but for now I think the informal statement is enough.

The second remark is that a natural weighting on sets is to make the weight of a set  $U$  equal to  $(1/3)^{|U|}(2/3)^{n-|U|}$ . Then the size of a typical set is  $n/3$ , and the probability distribution on the collection of all subsets of  $[n]$  is the same as the



probability distribution of the set of places where a random sequence in  $[3]^n$  takes any given value.

LL. In response to II, I don't think it's possible to make  $\Gamma$  even *upper* quasirandom by introducing edge weights. The idea to do so was roughly that one should penalize edges that join small sets because those were more likely to be disjoint. But making sets small is not the only way to make them more likely to be disjoint. For example, you could make  $\mathcal{A}$  consist of all sets that are slightly biased towards the first half of  $[n]$  (meaning that their intersection with the first  $n/2$  integers is slightly bigger than their intersection with the second  $n/2$  integers) and  $\mathcal{B}$  could consist of all sets that are slightly biased towards the second half of  $[n]$ . Of course, one could legislate against that particular example, but the same trick can be played for any partition of  $[n]$  into two equal-sized subsets. It seems to be impossible to legislate against this, since, for trivial reasons, *every* set will be biased with respect to some partitions.

*[The history of that idea was that I was sitting in the car waiting for a bus to arrive that had my daughter in it. Once again I just couldn't help my mind wandering in a Hales-Jewett direction.]*

### How might one completely kill off the above approach?

The tripartite graph that we form out of  $\mathcal{A}$  has few triangles in the following sense: every edge  $(U, V)$  in the  $XY$  graph is contained in just one triangle (the other vertex of which is  $[n] \setminus (U \cup V)$ ). If one is to prove the theorem by means of an adaptation of the triangle-removal lemma, one would want the following graph-theoretic statement to be true: let  $G$  be any subgraph of  $\Gamma$  such that each edge is contained in at most one triangle. Then it is possible to remove  $o(3^n)$  edges from  $G$  to make it triangle free. A counterexample to this more general statement would almost certainly make the approach hopeless. (I say "almost certainly" because one might just be able to use some extra structural property of the graph derived from  $\mathcal{A}$ . But that would change the nature of the approach quite a bit.)

Perhaps there could be convincing arguments that described the most general possible iterative approach to proving regularity lemmas and then proved that no iterative approach could give a regularity lemma that had any chance of helping with this problem. The seeming impossibility of making  $\Gamma$  quasirandom would undoubtedly be a part of such an argument.

That's all I have to suggest at the moment.

This entry was posted on February 1, 2009 at 1:43 pm and is filed under [polymath1](#). You can follow any responses to this entry through the [RSS 2.0](#) feed. You can [leave a response](#), or [trackback](#) from your own site.

## 194 Responses to "A combinatorial approach to density Hales-Jewett"

1. [A massively collaborative mathematical project « What's new](#) Says:

[February 1, 2009 at 8:47 pm](#) | [Reply](#)

[...] collaborative mathematical project" over at his blog. The project is entitled "A combinatorial approach to density Hales-Jewett", and the aim is to see if progress can be made on this problem by many small contributions [...]

2.  [gowers](#) Says:

[February 1, 2009 at 8:59 pm](#) | [Reply](#)

1. A quick question. Furstenberg and Katznelson used the Carlson-Simpson theorem in their proof. Does anyone know that proof well enough to know whether the Carlson-Simpson theorem might play a role here? If so, I could add it to the background-knowledge post. (But I'm sort of hoping it won't be needed.)

3.  [jozsef](#) Says:

[February 1, 2009 at 9:08 pm](#) | [Reply](#)

2. In this note I will try to argue that we should consider a variant of the original problem first. If the removal technique doesn't work here, then it won't work in the more difficult setting. If it works, then we have a nice result! Consider the

Cartesian product of an  $IP_d$  set. (An  $IP_d$  set is generated by  $d$  numbers by taking all the  $2^d$  possible sums. So, if the  $n$  numbers are independent then the size of the  $IP_d$  set is  $2^d$ . In the following statements we will suppose that our  $IP_d$  sets have size  $2^n$ .)

Prove that for any  $c > 0$  there is a  $d$ , such that any  $c$ -dense subset of the Cartesian product of an  $IP_d$  set (it is a two dimensional pointset) has a corner.

The statement is true. One can even prove that the dense subset of a Cartesian product contains a square, by using the density HJ for  $k=4$ . (I will sketch the simple proof later) What is promising here is that one can build a not-very-large tripartite graph where we can try to prove a removal lemma. The vertex sets are the vertical, horizontal, and slope -1 lines, having intersection with the Cartesian product. Two vertices are connected by an edge if the corresponding lines meet in a point of our  $c$ -dense subset. Every point defines a triangle, and if you can find another, non-degenerate, triangle then we are done. This graph is still sparse, but maybe it is well-structured for a removal lemma.

Finally, let me prove that there is square if  $d$  is large enough compare to  $c$ . Every point of the Cartesian product has two coordinates, a 0,1 sequence of length  $d$ . It has a one to one mapping to  $[4]^d$ ; Given a point  $((x_1, \dots, x_d), (y_1, \dots, y_d))$  where  $x_i, y_i$  are 0 or 1, it maps to  $(z_1, \dots, z_d)$ , where  $z_i=0$  if  $x_i=y_i=0$ ,  $z_i=1$  if  $x_i=1$  and  $y_i=0$ ,  $z_i=2$  if  $x_i=0$  and  $y_i=1$ , and finally  $z_i=3$  if  $x_i=y_i=1$ . Any combinatorial line in  $[4]^d$  defines a square in the Cartesian product, so the density HJ implies the statement.

(I should learn how to use Latex in blogs...)

4.  [Jason Dyer](#) Says:

[February 1, 2009 at 9:23 pm](#) | [Reply](#)

3. I find it reassuring the first thing I thought of when I read comment D was what turned out to be comment N.

Would it help to have a graph where each node was a set consisting of a combinatorial line, and the edges are the sets that share a sequence in common? That is,  $\{(1,2,3,3), (2,2,3,3), (3,2,3,3)\}$  would link to  $\{(1,2,3,3), (1,2,3,2), (1,2,3,1)\}$ .

I really do wonder if there's a more direct approach to the problem, simply working out for  $[3]^n$  what the minimal number of nodes one can remove before no combinatorial lines remain and then applying induction. Just fiddling with actual examples by hand a greedy approach works well. I'm sure this is horribly naive given the past solution had to apply ergodic theory.

5.  [Terence Tao](#) Says:

[February 1, 2009 at 9:26 pm](#) | [Reply](#)

4. As Gil pointed out in his post on this project, the  $k=2$  case of density Hales-Jewett is Sperner's theorem,

[http://en.wikipedia.org/wiki/Sperner\\_family](http://en.wikipedia.org/wiki/Sperner_family)

Now, of course, this theorem already has a combinatorial proof, but this proof has a rather different flavour than the triangle removal lemma (or its  $k=2$  counterpart, which is the trivial statement that if a subset of a set  $V$  of vertices is small, then it can be removed entirely by removing a small number of vertices.) But perhaps a subproblem would be to find an alternate combinatorial proof of Sperner's theorem.

As pointed out to me by my student, Le Thai Hoang, another difference between DHJ and, say, the corners problem, is that one does not expect a positive density of all combinatorial lines to lie in the set. For instance, in the  $k=2$  case, and identifying  $[2]^n$  with the set of subsets of  $\{1, 2, \dots, n\}$ , Hoang observed that the set of subsets of  $\{1, \dots, n\}$  of cardinality  $n/2 + O(\sqrt{n})$  has positive density, but only about  $(2 + o(1))^n$  of the  $3^n$  combinatorial lines here actually lie in the set. This seems to fit well with your observations that some weighting of the vertices etc. is needed.

At the ergodic theory level, there is also a distinction between the ergodic theory proofs of, say, the corners theorem,

and of density Hales-Jewett, which may be an indication that something different happens at the combinatorial level too. Namely, in the ergodic proof of corners (or Roth, or Szemerédi, etc.) one has a single dynamical system (with the action of a single group  $G$ ), and one works with that system (and its factors) throughout the proof. But for DHJ, one has a system of measurable sets indexed by words, and one repeatedly refines the set of words to much smaller subsets in order to obtain good “stationarity” properties. (My student, Tim Austin, could explain this much better than I could.) It is only once one has all this stationarity that the proof can really begin in earnest. This is also consistent with proofs of the colouring Hales-Jewett theorem, which also massively restrict the space of lines being considered, and with the above observation that we do not expect a positive density of the full space of lines to be contained in the set. So perhaps some preliminary “regularisation” of the problem is needed before doing anything triangle-removal-like.



6. [Terence Tao](#) Says:

[February 1, 2009 at 9:30 pm](#) | [Reply](#)

5. Incidentally, I only learned in the process of writing up my own post on this problem that Furstenberg-Katznelson have `_two_` papers on Density Hales-Jewett: the “big one”,

<http://www.ams.org/mathscinet-getitem?mr=1191743>

that does the full  $k$  case, but also an earlier paper,

<http://www.ams.org/mathscinet-getitem?mr=1001397>

that just does the  $k=3$  case. I’ve only looked at the former.

An unrelated thing: there is an annoying wordpress latex bug concerning latex that begins with a bracket `[`, which may cause some difficulty in this context. One can fix it by prefacing any latex string that begins with a bracket with some filler (I use `{ }`).



7. [Tim Austin](#) Says:

[February 1, 2009 at 10:16 pm](#) | [Reply](#)

6. Terry’s description of the difference between the ‘ergodic theoretic’ approaches to the corners theorem and DHJ is well-put: at no point does an action of an easy-to-work-with `_group_` of measure-preserving transformations emerge during the DHJ proof. Furstenberg and Katznelson introduce an enormous family of measure-preserving transformations indexed by all finite-length words in an alphabet of size  $k$  (and having some algebraic structure in terms of that family of words) in order to encode their notion of stationarity, but they neither use nor prove anything about the group generated by these transformations. Rather, they identify within their collection a great many ‘IP systems’ of transformations, a rather weaker sort of algebraic structure, and exploit that. In order to reach this point, they first obtain a kind of ‘stationarity’ for a family of positive-measure sets indexed by these words before introducing the transformations, and then expressing that stationarity in terms of these transformations in effect locks that notion of stationarity in place, and also gives a way to transport various multi-correlations between functions around on the underlying probability space.

In this connexion, it’s in obtaining this notion of stationarity that Furstenberg and Katznelson first use the Carlson-Simpson Theorem, and they then use it repeatedly throughout the paper in order to recover some similar kind of symmetry among a collection of objects indexed by the finite-length words. At this point, I can’t see where the same need would arise in the approach to DHJ under discussion here.



8. [gowers](#) Says:


[February 2, 2009 at 1:25 am](#) | [Reply](#)

7. With reference to Jozsef’s comment, if we suppose that the  $d$  numbers used to generate the set are indeed

independent, then it's natural to label a typical point of the Cartesian product as  $(\epsilon, \eta)$ , where each of  $\epsilon$  and  $\eta$  is a 01-sequence of length  $d$ . Then a corner is a triple of the form  $(\epsilon, \eta), (\epsilon, \eta + \delta), (\epsilon + \delta, \eta)$ , where  $\delta$  is a  $\{-1, 0, 1\}$ -valued sequence of length  $d$  with the property that both  $\epsilon + \delta$  and  $\eta + \delta$  are 01-sequences. So the question is whether corners exist in every dense subset of the original Cartesian product.

This is simpler than the density Hales-Jewett problem in at least one respect: it involves 01-sequences rather than 012-sequences. But that simplicity may be slightly misleading because we are looking for corners in the Cartesian product. A possible disadvantage is that in this formulation we lose the symmetry of the corners: the horizontal and vertical lines will intersect this set in a different way from how the lines of slope -1 do.

I feel that this is a promising avenue to explore, but I would also like a little more justification of the suggestion that this variant is likely to be simpler.


9.  [Terence Tao](#) Says:  
[February 2, 2009 at 1:31 am](#) | [Reply](#)

8. The following questions are perhaps somewhat tangential to the main project, but have the advantage of requiring less combinatorial expertise than the main question, and so might possibly take better advantage of this format.

Given any  $n$ , let  $c_n$  be the largest subset of  $[3]^n$  that does not contain a combinatorial line. Thus for instance  $c_0 = 1$ ,  $c_1 = 2$ ,  $c_2 = 6$  (for the latter, consider deleting the antidiagonal  $(3, 1), (2, 2), (1, 3)$  from  $[3]^2$ ). The  $k=3$  density Hales-Jewett theorem is the statement that  $c_n = o(3^n)$  as  $n \rightarrow \infty$ .


The first question, which seems very amenable to this medium, is to compute  $c_3$  exactly, and to get good upper and lower bounds on  $c_4$  (it seems unlikely that  $c_4$  will be easy to compute on the nose). This of course does not say too much about the asymptotic question, but one might at least try one's luck with the OEIS once one has enough data (note that this would have worked with the  $k=2$  theory).

The other question is to get as good a *lower* bound as one can for  $c_n$ . Of course, the trivial example  $[2]^n$  gives the lower bound  $c_n \geq 2^n$ , but one can clearly do better than this. Writing the Behrend example in base 3 gives something like  $c_n \geq 3^{n-O(\sqrt{n})}$ . But presumably one can do better still.

10.  [gowers](#) Says:  
[February 2, 2009 at 1:36 am](#) | [Reply](#)

9. Terry's comment about Le Thai Hoang's observation seems to me to be an important one, so let me spell out why, even though it's basically there in what Terry says. If you follow the triangle-removal proof of the corners result, you find that you can prove something stronger: in a set of density  $\delta$  you don't just get one corner, but you get  $c(\delta)n^2$  corners. (The proof is that if there are very few corners then there are very few triangles in the associated tripartite graph, and then you can apply triangle removal just as before to get a contradiction.) If one is going to find an analogous proof here, it is essential that it should not try to prove a lemma that has false consequences. In our case, we must be careful that we don't accidentally try to develop an approach that would also imply that a positive density of all possible combinatorial lines lie in the set.

I haven't checked, but I think that weighting the vertices as suggested in KK avoids this pitfall, because now the analogous set to the one used by Le Thai Hoang is concentrated around sets of size  $n/3$ . Probably this is exactly what Terry meant at the end of his second paragraph.

11.  [gowers](#) Says:  
[February 2, 2009 at 1:45 am](#) | [Reply](#)

10. In response to Terry and Tim's comments about regularization and the ergodic-theory proof, my feeling is that we

should bear in mind that something like this could well turn out to be necessary (on the grounds that genuine difficulties for one approach often have counterparts in other approaches), but should perhaps not try to guess in advance what form any regularization might take. (In case anyone's wondering what I mean by "regularization", a rough description would be that you have a subset  $A$  of a structure  $X$  and you find a subset  $Y$  of  $X$  of a similar form, such that  $A \cap Y$  is an easier subset of  $Y$  to handle than  $A$  is of  $X$ .) Instead, we should press ahead, try to prove some kind of triangle-removal statement, find that it fails for some reason, try to work out a nice property of  $A$  that would get round the problem, prove that one can regularize to obtain that property, etc. etc. But this process will still involve guesswork, and people who know the ergodic-theoretic proof are likely to be able to make better guesses.

12.  [Jason Dyer](#) Says:

[February 2, 2009 at 2:17 am](#) | [Reply](#)

11. *This of course does not say too much about the asymptotic question, but one might at least try one's luck with the OEIS once one has enough data (note that this would have worked with the  $k=2$  theory).*

Terry, has anyone tried to collect said data for  $k=3$ ?

13.  [Terence Tao](#) Says:

[February 2, 2009 at 2:27 am](#) | [Reply](#)

12. Dear Jason, I don't know, but I just emailed someone who I think would. My guess, though, is that density Hales-Jewett is probably not as well known as its cousins such as van der Waerden's theorem to have much data already worked out. If so, then presumably we can start collecting data here.

For instance, it is not hard to see that  $c_{n+1} \leq 3c_n$  and that  $c_{n+m} \geq c_n c_m$  for any  $n, m$ , so this already gives the bounds  $12 \leq c_3 \leq 18$  for  $c_3$ . Presumably one can do better.

14.  [Terence Tao](#) Says:

[February 2, 2009 at 4:08 am](#) | [Reply](#)

13. Perhaps in order to understand the issue coming from Hoang's observation a bit better, one can pose the following subproblem: what is the natural analogue of Varnavides theorem for DHJ? Varnavides used Roth's theorem as a black box to show that any subset of  $[n]$  of density  $\delta$  contained  $\geq c(\delta)n^2$  3-term arithmetic progressions when  $n$  was large; a similar argument using the Ajtai-Szemerédi theorem as a black box also shows that any subset of  $[n]^2$  of density  $\delta$  contains  $\geq c(\delta)n^3$  corners. So, using DHJ as a black box, what is the natural implication as to how "numerous" combinatorial lines are in a dense set? It seems that direct cardinality of such lines may not be the best measure of being "numerous", but there should presumably still be some Varnavides-like theorem out there.

15.  [josef](#) Says:


[February 2, 2009 at 5:26 am](#) | [Reply](#)

14. Regarding Terry's last comment about a Varnavides-like theorem, I'm a bit skeptical. Even for the  $k=2$  case (Sperner's thm) if we take points of the  $d$  dimensional cube where the difference between the 0-s and 1-s is at most  $\sqrt{d}$  then we have a positive fraction of the elements and there are less than  $\binom{d/2}{\sqrt{d}} 2^d$  combinatorial lines. Probably this should be the right magnitude for  $k=2$  as a Varnavides-like result and it shouldn't be difficult to prove. (But I don't have the proof ...)

16.  [Gil](#) Says:


[February 2, 2009 at 7:32 am](#) | [Reply](#)

15. Regarding lower bounds, fixing the numbers of '1' '2' and '3' coordinates or even just fixing the number of '1' coordinates gives better lower bounds, right? (Just like for Sperner's theorem.)

17.  [Terence Tao](#) Says:  
[February 2, 2009 at 8:00 am](#) | [Reply](#)


16. Hmm, Gil, you're right; so we get a bound of  $c_n \gg 3^n / \sqrt{n}$  asymptotically.

Dear Jozsef, yes, we will have to normalise things appropriately to get a Varnavides type theorem. For the  $k=3$  problem, for instance, one might try to shoot for a lower bound for the number of combinatorial lines which is something like  $(3+o(1))^n$  rather than  $4^n$ . Alternatively, we could somehow weight each line, perhaps by some factor depending on how many wildcards it contains. (Note that any set of density  $\delta$  in a large  $[3]^n$  will contain at least one combinatorial line in which the number of wildcards is at most  $C(\delta)$ , since one can partition  $[3]^n$  into copies of  $[3]^{n_0}$ , where  $n_0$  is the first index for which DHJ holds at density  $\delta/2$ , and observe that the original set will have density at least  $\delta/2$  in at least  $\delta/2$  of these copies. So maybe one has to weight things to heavily favour lines with very few wildcards.)

18.  [Terence Tao](#) Says:  
[February 2, 2009 at 8:46 am](#) | [Reply](#)


17. It just occurred to me that if a subset of  $[3]^n$  has positive density, then by the central limit theorem we may restrict attention to those strings which have  $n/3 + O(\sqrt{n})$  1's, 2's, and 3's, and still have a positive density of strings remaining. So without loss of generality we may assume that all strings in the set are "balanced" in this manner. Hence, the only combinatorial lines which really matter are those which have  $O(\sqrt{n})$  wildcards in them, rather than  $O(n)$ . So the number of combinatorial lines that are genuinely "in play" is just  $(3 + o(1))^n$  rather than  $4^n$ .

I wonder if DHJ implies the existence of a combinatorial line with  $\Theta(\sqrt{n})$  wildcards. In the  $k=2$  case it seems that a double counting argument should be able to establish something like this, though I haven't checked it thoroughly. If so, then there seems to be a good shot at getting the "right" Varnavides type theorem.

19.  [Terence Tao](#) Says:  
[February 2, 2009 at 9:01 am](#) | [Reply](#)

18. Two more thoughts... firstly, a sufficiently good Varnavides type theorem for DHJ may have a separate application from the one in this project, namely to obtain a "relative" DHJ for dense subsets of a sufficiently pseudorandom subset of  $[3]^n$ , much as I did with Ben Green for the primes (and which now has a significantly simpler proof by Gowers and by Reingold-Trevisan-Tulsiani-Vadhan). There are other obstacles though to that task (e.g. understanding the analogue of "dual functions" for Hales-Jewett), and so this is probably a bit off-topic.

Another thought is the following. For any  $a, b, c$  adding up to  $n$ , let  $\Gamma_{a,b,c}$  be the subset of  $[3]^n$  consisting of those strings with  $a$  1s,  $b$  2s, and  $c$  3s. If  $A$  is a dense subset of  $[3]^n$ , then  $A \cap \Gamma_{a,b,c}$  should be a dense subset of  $\Gamma_{a,b,c}$  for many  $a, b, c$ ; call a triple  $(a, b, c)$  *rich* if this is the case. By applying the corners theorem (!) we should then be able to find  $(a+r, b, c)$ ,  $(a, b+r, c)$ ,  $(a, b, c+r)$  which are simultaneously rich. But now this might be a good place with which to try a triangle removal argument, converting the lines of  $r$  wildcards connecting  $\Gamma_{a+r,b,c}$ ,  $\Gamma_{a,b+r,c}$ ,  $\Gamma_{a,b,c+r}$  into triangles as Tim suggests in the main post?

20.  [gowers](#) Says:  
[February 2, 2009 at 9:13 am](#) | [Reply](#)

19. Terry, I once tried something like what you suggest in the second paragraph of your last comment. There I was trying to do something different: come up with a definition of "quasirandom" that would guarantee that if you had



quasirandom subsets of  $\Gamma_{a+r,b,c}$ ,  $\Gamma_{a,b+r,c}$  and  $\Gamma_{a,b,c+r}$  then you would get a combinatorial line. I failed because I kept thinking of more and more obstructions and it seemed hard to find a definition that would deal with them all. This might be another project worth pursuing.

A problem here, which I feel is loosely related, is that you can't apply triangle removal if you restrict to  $\Gamma_{a+r,b,c}$ ,  $\Gamma_{a,b+r,c}$  and  $\Gamma_{a,b,c+r}$  because you don't get the degenerate triangles. This is similar to the problem that I had in comment CC above. (Incidentally, can anyone think of a better way of referring to comments than saying things like, "In Terry's third last comment, paragraph 2"? Added later: I've gone back and numbered the comments.)

21.  [gowers](#) Says:

[February 2, 2009 at 9:17 am](#) | [Reply](#)

20. A related observation is that if you've got three rich triples in the right configuration, you aren't guaranteed a combinatorial line. For example, in two of the sets the first coordinate could always be 1 and in the third it could always be 2.

22.  [gowers](#) Says:

[February 2, 2009 at 9:46 am](#) | [Reply](#)

21. This is a response to the first paragraph of Terry's first comment (comment number 4), in which he asks for another proof of Sperner's theorem. I like this idea: in one sense one could say that the difficulty with density Hales-Jewett is that the proof of Sperner's theorem doesn't generalize in an obvious way. (I don't necessarily mean directly, but one might try to find an argument that has the same relationship to the corners problem as a proof of Sperner's theorem has to the trivial one-dimensional analogue of the corners theorem, which says that a dense subset of  $[n]$  contains a pair of distinct points. But even this doesn't seem to be at all easy.)

A minor point of disagreement with what Terry says (again in comment 4). I think the correct analogue of triangle removal (in the tripartite case) one level down is the following statement: given two subsets  $A \subset [n]$  and  $B \subset [n]$  such that the number of pairs  $(a, b)$  with  $a \in A$  and  $b \in B$  is very small, then it is possible to remove a small number of elements from  $A \cup B$  and end up with no pairs. The proof of course is that one of  $A$  and  $B$  must be small. The reason I call this the analogue is that I think the analogue of a tripartite graph is a pair of subsets of the vertex sets, and the analogue of a triangle (a triple of edges) is a pair of vertices.

Now let me use this "pair-removal lemma" to prove the "one-dimensional corners" theorem. Suppose, then, that we have a subset  $A \subset [n]$  of size at least  $cn$ . I'm going to define two vertex sets  $X$  and  $Y$  as follows. I let  $X$  be the set of all ordered singletons  $x$  such that  $x \in A$ , and I let  $Y$  be the set of all ordered singletons  $y$  such that the sum of all the terms equals an element of  $A$ . (In other words, both  $X$  and  $Y$  are equal to  $A$ , but I want to demonstrate that I'm doing exactly what one does in higher dimensions.) If we can find a pair  $(x, y) \in X \times Y$ , then we are done (that is, we've found a configuration of the form  $a, a + d$  in the set  $A$ ), except in the degenerate case that  $x = y$ . So if we can't find a one-dimensional corner, then the only pairs in  $X \times Y$  are the degenerate ones, which implies that there are only  $cn$  pairs. But then we can apply the pair-removal lemma and remove a few elements from  $X \cup Y$  and end up with *no* pairs. But the degenerate pairs are vertex-disjoint, so we have to remove at least  $cn$  elements to get rid of all those, so we have reached a contradiction.

The next question might be whether we can use the "pair-removal lemma" to give a different proof of Sperner's theorem. Or rather, we don't actually want to prove the full strength of Sperner's theorem but just the weaker statement that if you have at least  $c2^n$  subsets of  $[n]$  then you can find one that's a proper subset of another. Up to now what I've done is silly, but this project would not be silly: it might tell us exactly what the difficulty is with this approach to density Hales-Jewett, but in a much simpler context.

23.  [gowers](#) Says:

[February 2, 2009 at 10:13 am](#) | [Reply](#)

22. Continuing the same theme, there is a natural pair of vertex sets to define in the Sperner case. For all higher dimensions, one takes the vertex sets to be copies of the power set of  $[n]$ , so one should surely do so here. Now in a certain sense the  $j$ th vertex set stands for the set of places where a sequence equals  $j$ . So here we would let  $X$  and  $Y$  be two copies of the power set of  $[n]$ , with sets in  $X$  representing the set of places where a sequence equals 0 and sets in  $Y$  representing the set of places where a sequence equals 1. Then we take as our subset  $\mathcal{A} \subset X$  the set of all sets  $U$  such that the sequence that's 0 on  $U$  and 1 on the complement of  $U$  belongs to  $A$ , and we take  $\mathcal{B} \subset Y$  to be the set of all sets  $V$  such that the sequence that's 1 on  $V$  and 0 on the complement of  $V$  belongs to  $A$ . In set theoretic terms,  $\mathcal{A}$  consists of sets that belong to  $A$  (which itself, just to be clear, is our given collection of  $2^{2^n}$  subsets of  $[n]$ ), and  $\mathcal{B}$  is consists of all *complements* of sets that belong to  $A$ .

At this point it is clear that what we are looking for is a pair  $(U, V)$  with  $U \in \mathcal{A}$ ,  $V \in \mathcal{B}$ , and  $U \cap V = \emptyset$ . This will be a degenerate pair if  $U \cup V = [n]$ , but otherwise will imply that  $A$  contains two sets with one a proper subset of the other. So this is the natural analogue of the triangles discussed in comments N and P.

Note that in this situation we seem to be forced to put a graph structure on the two vertex sets  $X$  and  $Y$ , since we cannot just take any old pair  $(U, V)$  but have to make them disjoint. At first this looks bad, but I actually think it's OK because the same disjointness condition occurs in all dimensions. In other words, you get a graph in higher dimensions and not a hypergraph.

The next step, I think, would be to try to find a suitable "pair-removal lemma" for this more complicated situation where the pairs have to be disjoint sets. Is the following true (or even the right question)? Suppose you have two collections  $\mathcal{A}$  and  $\mathcal{B}$  of subsets of  $[n]$ , and there are very few pairs  $(A, B)$  such that  $A \in \mathcal{A}$ ,  $B \in \mathcal{B}$ , and  $A \cap B = \emptyset$ . Then you can remove a small fraction of the sets from  $\mathcal{A} \cup \mathcal{B}$  and end up with *no* such pairs.

24.  [gowers](#) Says:

[February 2, 2009 at 11:02 am](#) | [Reply](#)

23. This comment is a further response to Jozsef Solymosi's comment (number 2). A slight variant of the problem you propose is this. Let's take as our ground set the set of all pairs  $(U, V)$  of subsets of  $[n]$ , and let's take as our definition of a corner a triple of the form  $(U, V), (U \cup D, V), (U, V \cup D)$ , where both the unions must be disjoint unions. This is asking for more than you asked for because I insist that the difference  $D$  is positive, so to speak. It seems to be a nice combination of Sperner's theorem and the usual corners result. But perhaps it would be more sensible not to insist on that positivity and instead ask for a triple of the form  $(U, V), ((U \cup D) \setminus C, V), (U, (V \cup D) \setminus C)$ , where  $D$  is disjoint from both  $U$  and  $V$  and  $C$  is contained in both  $U$  and  $V$ . That is your original problem I think.

I think I now understand better why your problem could be a good toy problem to look at first. Let's quickly work out what triangle-removal statement would be needed to solve it. (You've already done that, so I just want to reformulate it in set-theoretic language, which I find easier to understand.) We let all of  $X, Y$  and  $Z$  equal the power set of  $[n]$ . We join  $U \in X$  to  $V \in Y$  if  $(U, V) \in A$ .

Ah, I see now that there's a problem with what I'm suggesting, which is that in the normal corners problem we say that  $(x, y + d)$  and  $(x + d, y)$  lie in a line because both points have the same coordinate sum. When should we say that  $(U, V \cup D)$  and  $(U \cup D, V)$  lie in a line? It looks to me as though we have to treat the sets as 01-sequences and take the sum again. So it's not really a set-theoretic reformulation after all.

That suggests a wild question, which might have an easy counterexample (or an easy proof using density Hales-Jewett, which would be useful just to see whether it is true). If you have a dense subset  $A$  of the set of all pairs  $(U, V)$  of subsets of  $[n]$ , can you find a triple  $(U, V), (U, V \cup D)$  and  $(U \cup D, V)$  in  $A$  such that the three sets  $U, V$  and  $D$  are all disjoint?



25.  [gowers](#) Says:

[February 2, 2009 at 12:45 pm](#) | [Reply](#)

24. I want to continue the discussion in 13, 14, 16, 17 and 18 concerning the appropriate Varnavides-type theorem for density Hales-Jewett.

The basic difficulty appears to be that if you know that a point lies in a combinatorial line, then it affects the shape of the point. Let's see this first in a very simple context, where we just put the uniform distribution on the set of all points and the uniform distribution on the set of all lines. We can represent a combinatorial line as a point in the space  $\{1, 2, 3, *\}^n$ , where the asterisks represent the varying coordinates. In this way we see easily that there are  $4^n$  combinatorial lines and that the set of fixed coordinates that take a given one of the values 1, 2 or 3 will typically have size about  $n/4$ . (For anyone who isn't familiar with this kind of argument, if you choose a random combinatorial line, regarded as a sequence that consists of 1s, 2s, 3s and asterisks, then the set of places where you choose 1 is a random subset of  $[n]$  where each point is chosen with probability  $1/4$ . With very high probability such a set has size close to  $n/4$ .) So we see that while a typical point has roughly equal numbers of 1s, 2s and 3s, a typical combinatorial line contains points that have around  $n/4$  of two values and around  $n/2$  of the third value.

Suppose we try to remedy this by putting a different distribution on to the set of lines. Is there some natural way to do this that encourages the set of variable coordinates to be much smaller than the sets of fixed coordinates? A rather unpleasant approach would be simply to put some upper bound on the size of the set of variable coordinates and otherwise choose uniformly. The most natural thing I can think of to do is not all that much nicer, but at least it smooths the cutoff. It's to choose the variable set of coordinates by picking each point independently with probability  $p$ , where  $p$  is something like  $n^{-1/2}$ , and then choose the fixed coordinates randomly to be 1, 2 or 3.

Actually, I can already see that this doesn't work. The size of the resulting set of variable coordinates will be strongly concentrated around  $n^{1/2}$ . But it's not hard to choose a dense set of sequences such that if you choose any two of them,  $x$  and  $y$ , then the number of 1s in  $x$  and the number of 1s in  $y$  do not differ by approximately  $n^{1/2}$ . The number of combinatorial lines in such a set would not be dense.

The fact that the natural measure on the set of sequences does not project properly to the natural measure on the set of triples  $(r, s, t)$  of integers such that  $r + s + t = n$  seems to be a fundamental difficulty.

Another thought one might have is to put a different measure on the set of points, so as to encourage points where we get two values roughly  $n/4$  times and one value roughly  $n/2$  times. But the trouble with this is that it leads to the no-degenerate-triangles problem. To be precise, it would lead to three sets of sequences (according to whether 1, 2 or 3 was the preferred value) and no sequence would belong to all three sets. So it really does seem essential to force the variable set to be small somehow. And all I have to suggest so far is to do something very unnatural like first randomly choosing the cardinality according to some distribution that isn't too concentrated about any value — the most natural one I can think of is a gentle exponential decay — then choosing a set of that cardinality, and then choosing the fixed coordinates randomly. But I have no reason to suppose that that leads to a Varnavides-type theorem.

26.  [gowers](#) Says:

[February 2, 2009 at 1:44 pm](#) | [Reply](#)

25. I see now that there's quite a substantial overlap between my last comment and Terry's comment 16. But one can view my last comment as an elaboration of his, perhaps.

Here I just want to make the quick suggestion that one way to work out what the appropriate Varnavides theorem should say is to prove it first. How would this strategy work? Well, we assume the density Hales-Jewett theorem in the following form: for every  $\delta > 0$  there exists  $k$  such that every subset of  $[3]^k$  of size at least  $(\delta/2)3^k$  contains a combinatorial line. (The  $\delta/2$  is convenient in just a moment.)

Now let  $A$  be a subset of  $[3]^n$  of size at least  $\delta 3^n$ . Here,  $n$  is an integer that is much bigger than  $k$ . The set  $[3]^n$  can be covered uniformly by structures that are isomorphic to  $[3]^k$ , and a positive proportion (in fact,  $\delta/2$  at least) of these structures intersect  $A$  in a subset of density at least  $\delta/2$ . So a positive proportion of these structures contain a combinatorial line. And now a double-counting argument gives us lots of combinatorial lines.

Now there are several choices to make if one wants to pursue this line of attack. The main one is what set of  $k$ -dimensional substructures of  $[3]^n$  to average over. For the benefit of people who haven't thought too hard about Hales-Jewett, let me give a pretty general class of such structures. You choose  $k$  disjoint sets  $V_1, V_2, \dots, V_k$  and you choose fixed values in  $\{1, 2, 3\}$  for all elements of  $[n]$  that do not belong to one of the  $V_i$ . The  $k$ -dimensional structure consists of the  $3^k$  sequences that take the fixed values outside the  $V_i$  and are constant on each  $V_i$ . What can we deduce from the fact that there exists  $k$  that depends on  $\delta$  only such that if  $A$  intersects any  $k$ -dimensional set of this kind in at least  $(\delta/2)3^k$  places then that  $k$ -dimensional set contains a combinatorial line in  $A$ ?

The choice one has is this: over what set of structures of this kind (or with what weighting on the structures) do we do an averaging argument? And what can we expect to get out of it at the end? Is it likely to give anything useful?

One other question: is there a reasonably precise argument to the effect that if the triangle-removal approach worked, then it would imply a stronger Varnavides-type result? And if so, what would that result be? This could be another way to get a handle on the problem.

27.  [gowers](#) Says:

[February 2, 2009 at 3:18 pm](#) | [Reply](#)

26. Another small thought in response to Terry's comment number 18, paragraph 2, is this. Even if that idea runs into the difficulty I mentioned in comment 19, there might be a chance of getting something out of a small variant of the idea. Terry suggested using the corners theorem to find a triple  $\Gamma_{a+r,b,c}, \Gamma_{a,b+r,c}, \Gamma_{a,b,r+c}$  in each part of which  $A$  is "rich". The problem was that if one then used just these sets to form a tripartite graph, one would end up with no degenerate triangles. But what if instead one used the full 2-dimensional Szemerédi theorem (which can be proved using the simplex-removal lemma in sufficiently high dimension, so it doesn't depend on ergodic theory) to obtain a big network of sets  $\Gamma_{a,b,c}$  in which  $A$  was rich? They could be organized into a triangular lattice: they would consist of all sets of the form  $\Gamma_{a+ir,b+jr,c+kr}$  such that  $i, j$  and  $k$  are non-negative integers that sum to  $m$  for some reasonably large  $m$ . It's not obvious that this would be any help, but it's also not obvious that it wouldn't be any help.

28.  [gowers](#) Says:

[February 2, 2009 at 3:33 pm](#) | [Reply](#)

27. I was rather pleased with the "conjecture" in the final paragraph of comment 22, but have just noticed that it is completely false, at least if you interpret it in the most obvious way. Indeed, if you take both  $\mathcal{A}$  and  $\mathcal{B}$  to consist of *all* subsets of  $[n]$ , then you find that almost every set in  $\mathcal{A}$  intersects almost every set in  $\mathcal{B}$ . And it's clear that if you just remove a small fraction of the sets from  $\mathcal{A}$  and  $\mathcal{B}$  you aren't going to be able to ensure that *every* set left in  $\mathcal{A}$  intersects *every* set left in  $\mathcal{B}$ . (Proof: for each pair  $(U, U^c)$  you will either have to remove  $U$  from  $\mathcal{A}$  or  $U^c$  from  $\mathcal{B}$ .)

This leaves me with two questions.

(i) Can this pair-removal approach to Sperner be rescued somehow?

(ii) If not, can one use its failure as the basis of a rigorous proof that no triangle-removal approach can work for density Hales-Jewett?

29.  [gowers](#) Says:

[February 2, 2009 at 3:47 pm](#) | [Reply](#)

28. Perhaps the answer to (i) is rather simple. First, we see how many disjoint pairs there are in the case where both  $\mathcal{A}$  and  $\mathcal{B}$  are the full power set of  $[n]$ . The answer is  $3^n$ , because for each element of  $[n]$  you get to choose whether it belongs to the first set, the second set, or neither. So now we would deem a pair of set systems  $(\mathcal{A}, \mathcal{B})$  to have “few” disjoint pairs if the number of pairs  $(U, V)$  with  $U \in \mathcal{A}$ ,  $V \in \mathcal{B}$  and  $U \cap V = \emptyset$  is at most  $c3^n$  for a tiny positive constant  $c$ . Under that assumption, does it follow that we can remove a small number of sets and end up with no disjoint pairs? And if so, would that prove Sperner’s theorem (in the weak form where we assume that a set system has positive density)?



30. [Gil Kalai](#) Says:

[February 2, 2009 at 4:16 pm](#) | [Reply](#)

29. Several remarks which are not directly towards the specific project but related.

1. With all due respect to Sperner’s theorem it is hard to believe that the  $3^n / \sqrt{n}$  lower bound cannot be improved. Right? Maybe as a side project we should come up with a  $3^n / o(\sqrt{n})$  example for  $k=3$  density HJ or ask around if this is known?

2. A sort of generic attack one can try with Sperner is to look at  $f = 1_A$  and express using the Fourier expansion of  $f$  the expression  $\int f(x)f(y)1_{x < y}$  where  $x < y$  is the partial order (=containment) for 0-1 vectors. Then one may hope that if  $f$  does not have a large Fourier coefficient then the expression above is similar to what we get when  $A$  is random and otherwise we can raise the density for subspaces.

(OK, you can try it directly for the  $k=3$  density HJ problem too but Sperner would be easier;)

This is not unrelated to the regularity philosophy.

3. Reductions: the works of Furstenberg and Katznelson went gradually from multidimensional Szemerédi to density theorems for affine subspaces to density HJ. (Maybe there were more stops on the way)

Beside the obvious reduction - combinatorial lines imply long AP (and also high dimension Szemerédi if I remember correctly) reductions in the opposite directions are not known (to me, at least).

It is tempting to think that the when we find high dimensionall affine space perhaps in a much much much larger alphabet we can somehow get a little combinatorial line for the small alphabet.

4. There is an analogous for Sperner but with high dimensional combinatorial spaces instead of “lines” but I do not remember the details (Kleitman(?) Katona(?)) those are usual suspects.)



31. [Ryan O'Donnell](#) Says:

[February 2, 2009 at 4:35 pm](#) | [Reply](#)

30. As a person who writes the phrase “choose the sequence  $x$  at random; then form  $y$  by rerandomizing each coordinate of  $x$  with probability  $p$ ” in nearly half his papers, I’m naturally drawn to the 16-17-24-etc. comment thread.

In #24, Tim suggests (and I agree) that the most natural way to pick a line is to form three sequences  $(x, y, z)$  by choosing

$(x_i, y_i, z_i)$  uniformly from  $[3]^3$ , with probability  $1-p$ ,

$(x_i, y_i, z_i) = (1, 2, 3)$  with probability  $p$ ,

independently for each  $i = 1 \dots n$ , with  $p$  taken to be perhaps  $1/\sqrt{n}$ . But Tim also makes the good — and worrisome — point that this distribution has the “wrong” marginal on  $x$  (and on  $y$  and on  $z$ ). For example, the marginal on  $x$  is the  $(1/3+2p/3, 1/3-p/3, 1/3-p/3)$  product distribution on  $[3]^n$ , whereas it “ought to be” the uniform distribution, because we are measuring the density of  $A$  with respect to the uniform distribution.

It's not completely clear to me that this is an insuperable problem though. For example, in #24 Tim wrote, "But it's not hard to choose a dense set of sequences such that if you choose any two of them,  $x$  and  $y$ , then the number of 1s in  $x$  and the number of 1s in  $y$  do not differ by approximately  $n^{1/2}$ ."

I don't quite see it... Isn't it the case, as Terry pointed out in #17, that almost all strings in  $A$  have  $1/3 \pm O(\sqrt{n})$  1's? In fact, shouldn't a random pair  $x, y$  in  $A$  have a 1-count differing by  $\Theta(\sqrt{n})$  with high probability? [All hidden constants here depend on  $\delta$ .]

Anyway, I agree it's worrisome. To check whether it is indeed insuperable, I wonder if we could try "practising" this idea in the weak-Sperner's setting Tim suggested at the end of #21. The analogous question there would be:

Suppose  $A$  contains at least a  $\delta$ -fraction of the strings in  $\{0,1\}^n$  (w.r.t. the uniform distribution). Suppose we choose two strings  $(x,y)$  as follows:

$(x_i, y_i)$  is uniform from  $\{0,1\}^2$  with probability  $1 - 1/\sqrt{n}$

$(x_i, y_i)$  is  $(0,1)$  with probability  $1/\sqrt{n}$ ,

independently for each  $i = 1 \dots n$ . Is there a positive probability that both  $x$  and  $y$  are in  $A$ ?

32.  [gowers](#) Says:

[February 2, 2009 at 4:45 pm](#) | [Reply](#)

31. Gil, a quick remark about Fourier expansions and the  $k = 3$  case. I want to explain why I got stuck several years ago when I was trying to develop some kind of Fourier approach. Maybe with your deep knowledge of this kind of thing you can get me unstuck again.

The problem was that the natural Fourier basis in  $[3]^n$  was the basis you get by thinking of  $[3]^n$  as the group  $\mathbb{Z}_3^n$ . And if that's what you do, then there appear to be examples that do not behave quasirandomly, but which do not have large Fourier coefficients either. For example, suppose that  $n$  is a multiple of 7, and you look at the set  $A$  of all sequences where the numbers of 1s, 2s and 3s are all multiples of 7. If two such sequences lie in a combinatorial line, then the set of variable coordinates for that line must have cardinality that's a multiple of 7, from which it follows that the third point automatically lies in the line. So this set  $A$  has too many combinatorial lines. But I'm fairly sure — perhaps you can confirm this — that  $A$  has no large Fourier coefficient.

You can use this idea to produce lots more examples. Obviously you can replace 7 by some other small number. But you can also pick some arbitrary subset  $W$  of  $[n]$  and just ask that the numbers of 0s, 1s and 2s inside  $W$  are multiples of 7.

If these observations are correct, then they suggest another very interesting subproblem: describe all obstructions to randomness. That is, find a nice explicit set  $\mathcal{F}$  of functions defined on  $[3]^n$  such that (the characteristic function of) any set with the wrong number of combinatorial lines correlates with a function in  $\mathcal{F}$ . Ideally, correlation with a function in  $\mathcal{F}$  would imply a density increase on some subset.

Of course, this would be aiming for a Roth-type proof rather than a Ruzsa-Szemerédi-Solymosi-type proof.

33.  [Ryan O'Donnell](#) Says:

[February 2, 2009 at 4:45 pm](#) | [Reply](#)

32. [Calling Gil's post 29 and my earlier one 30.]

Sorry for the lengthiness of the posts, but I wanted to throw out one more idea and one more question.

Suppose the issue discussed in my #30 is indeed a problem. Here's a possible (perhaps far-fetched) way around it:

First, find the simultaneously “rich” triples  $(a+r, b, c)$ ,  $(a, b+r, c)$ ,  $(a, b, c+r)$  as in Terry’s post #18. Now take a new distribution on combinatorial lines: Uniform on lines which have exactly  $r$  wildcards.

This may improve things because its marginals on  $x$ ,  $y$ , and  $z$  are precisely the uniform distributions on  $\text{Gamma}_{\{a+r, b, c\}}$ ,  $\text{Gamma}_{\{a, b+r, c\}}$ , and  $\text{Gamma}_{\{a, b, c+r\}}$ . And we know  $A$  is dense under these distributions, by the definition of richness.

Working “slicewise” in this way has some intuitive appeal to me because it seems compatible with the prospect of getting the “correct” Sperner by these ideas, as opposed to just the “weak” Sperner.

34.  [gowers](#) Says:

[February 2, 2009 at 4:52 pm](#) | [Reply](#)

33. In response to Ryan’s question in 30, I see that I didn’t explain myself clearly enough. Here’s what I meant.

Suppose you have a distribution on combinatorial lines with the property that the size of the variable set has a very high probability of being very close to  $\sqrt{n}$ . (NB, I don’t just mean order  $\sqrt{n}$  but between  $(1 - \epsilon)\sqrt{n}$  and  $(1 + \epsilon)\sqrt{n}$ . This is I think the point of misunderstanding.) What I was claiming was that it was quite easy to choose a dense set of sequences such that if  $x$  and  $y$  are two of your sequences then the number of 1s in  $x$  and the number of 1s in  $y$  never differ by almost exactly  $\sqrt{n}$ .

35.  Ryan O'Donnell Says:

[February 2, 2009 at 4:55 pm](#) | [Reply](#)

34. Re Tim’s #33: Ah, I see. Right.

So perhaps the thing to do is not to choose the number of wildcards to be  $\text{Binomial}(n, 1/\sqrt{n})$ , but, uh, I dunno, perhaps  $\text{Poisson}(\sqrt{n})$ ??

36.  [Jason Dyer](#) Says:

[February 2, 2009 at 4:56 pm](#) | [Reply](#)

We’ve already got a number of threads developing. Here is my best effort at sorting them out.

Discussion of original Furstenberg-Katzelson proof (#1, #5, #6, #10)

$IP_d$  set problem (#2, #7, #23)

Finding lower and upper bounds (#3, #8, #11, #12, #15, #16)

Sperner’s theorem (#4, #21, #22, #27, #28, Gil #29)

Creating a quasirandom definition (#18, #19, #26)

Positive density (#4, #9, #17)

Analogue of Varnavides theorem (#13, #14, #16, #17, #18, #24, #25)

The latter two threads might be combinable.

*[I haven't numbered this comment as it's a metacontribution rather than a contribution -- but thanks for it.]*

37.  Ryan O'Donnell Says:

[February 2, 2009 at 4:58 pm](#) | [Reply](#)

35. Now the question, following up on the separate thread initiated by Jozsef (specifically, #2 and Tim’s #7), just to confirm I have the question right...

There is a dense subset  $A$  of  $\{0,1\}^n \times \{0,1\}^n$ . Is it true that it must contain three nonidentical strings  $(x, x'), (y, y'), (z, z')$  such that for each  $i = 1 \dots n$ , the 6 bits

$[x_i, x'_i]$   
 $[y_i, y'_i]$   
 $[z_i, z'_i]$

are equal to one of the following:

$[0,0], [0,0], [0,1], [1,0], [1,1], [1,1]$   
 $[0,0], [0,1], [0,1], [1,0], [1,0], [1,1]$   
 $[0,0], [1,0], [0,1], [1,0], [0,1], [1,1]$

?

38.  [gowers](#) Says:

[February 2, 2009 at 4:59 pm](#) | [Reply](#)

36. (Responding to 34) Unfortunately, Poisson( $\sqrt{n}$ ) is just as concentrated around  $\sqrt{n}$  as Binomial( $n, 1/\sqrt{n}$ ) (as one would expect since binomial tends to Poisson). That's what slightly bothers me — the most natural distributions have that awkward property.

39.  [Ryan O'Donnell](#) Says:

[February 2, 2009 at 5:11 pm](#) | [Reply](#)

37. (Re 36.) Hmm. But there has to be *some* correct distribution for the wildcards.

After all, what *is* the distribution on the difference in 1-counts between independent and uniform  $x$  and  $y$ ? It's distributed as the sum of  $n$  i.i.d. r.v.'s which are  $+1$  with probability  $2/9$ ,  $-1$  with probability  $2/9$ , and  $0$  with probability  $5/9$ . So it's like a Gaussian with standard deviation  $\sigma = 2/3\sqrt{n}$ .

So we could put in a number of wildcards equal to  $\text{round}(|N(0, 4/9n)|) \dots$

This looks weird enough that it's probably a wrong idea, but I still feel there may be a "natural" probabilistic way to do it.

Or maybe I revert to the suggestion of thinking about distributions on fixed numbers of 1's, 2's, 3's as in #32... 😊

40.  [gowers](#) Says:

[February 2, 2009 at 5:29 pm](#) | [Reply](#)

38. Here's an attempt to throw the spanner in the works of #32. Let's define  $\mathcal{A}$  to be the set of all sequences  $x$  with the following property: if the number of 1s in  $x$  is even then  $x_n = 2$ , and if the number of 1s in  $x$  is odd then  $x_n = 3$ .

I first claim that virtually all slices are rich. Indeed, take a typical slice defined by the triple  $r + s + t = n$ . Then just choose  $r$  1s, taking care not to choose  $x_n$ , choose the appropriate value of  $x_n$  according to whether  $r$  is even or odd, and then fill in the rest with the appropriate number of 2s and 3s. The only thing that can go wrong is that one of  $s$  and  $t$  is equal to 0, but for all normal slices we're OK.

Now I claim that we can't have a combinatorial line with an odd number of wildcards. The reason is that when the wildcards change from 1 to 2, the final coordinate  $x_n$  is forced to change from 2 to 3, which is not allowed.

So it looks to me as though it would be disastrous to take the uniform distribution over lines with some fixed number of



wildcards (unless, perhaps, one had done some more preprocessing to get a stronger property than mere richness).

41.  [Jason Dyer](#) Says:

[February 2, 2009 at 5:30 pm](#) | [Reply](#)

(Another metacomment.)

Tim, now may be the right time to create multiple posts so people can sort into the discussion they want to focus on. Given the rate posts are piling up it will get too hard to do it later. In fact, by the time I hit “submit” I’m guessing 3 more comments will have come up. 😊

#30, #32, #33, #34, #36, #37 spawned off Varnavides theorem thread, although tilting over to general randomness  
#31 (off of #29) Sperner, sort of — Fourier approach might spawn its own thread  
#35 IP\_d problem thread

*[Jason---I'm going to answer this question over on the Questions of Procedure post. Or rather, I'm going to ask what others think. I agree that it would, for instance, make it easier for a newcomer if they had a more organized structure of comments.]*

42.  [Terence Tao](#) Says:

[February 2, 2009 at 5:56 pm](#) | [Reply](#)

[Incidentally, in other comment threads, it is often customary to refer to a specific comment by a timestamp, e.g. "Gowers Feb 2 3:18 pm". But I will adhere to the numbering convention also...]

39. Wow, this thing is evolving rapidly ... I will need some time to catch up, but I wanted to contribute two thoughts I had this morning... with regard to Gil's 29 (Feb 2 4:16 pm), observe that the union of all  $\Gamma_{a,b,c}$  where  $b+2c$  lies in a set that has no length 3 progressions, will have no combinatorial lines. So by plugging in the Behrend example it now seems that we have the lower bound  $c_n \geq 3^{n-O(\sqrt{\log n})}$  (if I haven't made a miscomputation somewhere).

Secondly, if one applies the corners argument locally instead of globally, one can make the  $r$  in the rich triple  $(a+r, b, c), (a, b+r, c), (a, b, c+r)$  to be of bounded size,  $r=O(1)$ . I don't know if this is helpful though. But it begins to remind me of the isoperimetric inequality for the cube (which would correspond to the case  $k=2, r=1$ ).

*[Morning Terry! Yes, that time stamp idea would be good, and would save me having to do a lot of editing to number comments. (Even if everyone remembers to number their comments, they don't always know what number their comment will actually turn out to be.)]*

43.  [Terence Tao](#) Says:

[February 2, 2009 at 6:28 pm](#) | [Reply](#)

40. Two more thoughts. Firstly, I can now show that  $c_3 = 18$ , by taking the set  $\{(x, y, z) \in \mathbb{Z}_3^3 : x + y + z \not\equiv 0 \pmod{3}\}$ , or equivalently the union of all the  $\Gamma_{a,b,c}$  such that  $a + 2b$  is not divisible by 3. (The upper bound  $c_3 \leq 18$  follows from the general inequality  $c_{n+1} \leq 3c_n$ .) So maybe  $c_4$  is computable also, for what it's worth.

Secondly, one may need some sort of “hereditary” richness, in which one considers not only the global counts  $a, b, c$  of 1s, 2s, and 3s in  $[n]$ , but also considers all subsets of  $[n]$  and looks for richness there too. If there is one subset of  $[n]$  in which the richness fluctuates much more than in the original index set  $[n]$ , then we pass to that set (incrementing a certain “richness energy” along the way) and keep going until one obtains “richness regularity” (by a standard “energy increment argument”), which should mean something like that the “richness statistics” of any “non-negligible” subindex set of the index set should be close to the statistics of the original index set. This may help avoid the kind of

counterexamples that appear for instance in Gowers 20 (9.17am), or Gowers 38 (5.29 pm). It also reminds me of the stationarity reduction used in the ergodic theory world (Terry 4 9.26pm, Tim Austin 16 10:16pm).



44. [Terence Tao](#) Says:

[February 2, 2009 at 6:32 pm](#) | [Reply](#)

41. Incidentally, given that we are trying to find triangles in sparse (but now constant degree) graphs, we may eventually have to use the type of regularity lemmas associated with bounded degree graphs, as for instance studied by Elek and Lippner,

<http://arxiv.org/abs/0809.2879>



45. [Jason Dyer](#) Says:

[February 2, 2009 at 7:49 pm](#) | [Reply](#)

42. I need to break out the compiler, but for what it's worth, here's my brute force data structure I'm going to use to try to get (at the least lower bound, and if they match upper bound exact) values of  $c_n$ :

One set of nodes (call them "red") correspond to the elements of  $[3]^n$

Another set of nodes (call them "blue") correspond to the specific combinatorial lines. They each have degree 3 and are linked to by the blue nodes.

Greedy algorithm:

- 1.) Look at the red nodes; out of the set of red nodes with highest degree, pick one at random.
- 2.) Remove that node, and all blue nodes linking to that node, and all edges corresponding to the red node and the blue nodes.
- 3.) If there are any blue nodes remaining, return to 1. Otherwise, all combinatorial lines have been removed.



46. [Jason Dyer](#) Says:

[February 2, 2009 at 7:50 pm](#) | [Reply](#)

(42 fix) blue nodes are linked to by the red nodes.



47. [Christian Elsholtz](#) Says:

[February 2, 2009 at 8:27 pm](#) | [Reply](#)

(43)

Let me consider a related problem, which can be thought of as a generalization:

we consider sets of  $[3]^n$  where all lines, not just combinatorial lines, are forbidden. This problem is also known as Moser's cube problem. It is known that the extremal configurations have, 2,6,16,43 points in dimensions  $n=1,2,3,4$  respectively.

The bounds in this problem were  $3^{n/\sqrt{n}}$  and  $o(3^n)$  respectively, as well.

These general lines can be modeled by two different wild cards.

Say  $(1,x,3,y,x)$  is a line, if  $x$  runs from 1 to 3, and  $y$  runs from 3 to 1.

Here is a sketch of an improvement of the lower bound, in the spirit of Terry's idea (#39).



The sets  $\Gamma_{\{a,b,c\}}$  do not contain any line, so taking unions of several  $\Gamma$ s, one would need to study (with  $r$  wild cards  $x$  and  $s$  wild cards  $y$ )  $\Gamma$ s with indices  $(a+r, b, c+s)$ ,  $(a, b+r+s, c)$ ,  $(a+s, b, c+r)$ . Then take those  $\Gamma_{\{a',b',c'\}}$ , where  $a'+2b'+3c'$  is taken from a set without a 3 progression, and one gets a lower bound of  $3^n / \exp(c \sqrt{\log n})$

Since the Hales-Jewett theorem is on combinatorial lines:

some of the ideas discussed before might work for general lines, i.e. Moser's cube problem as well. Is there any general lower bound construction that works for combinatorial lines only? (Terry's example with 18 points in dimension 3 has lines, so what is the right generalization?).

What is genuinely different for combinatorial lines?

48. [Boris](#) Says:

[February 2, 2009 at 8:43 pm](#) | [Reply](#)

44. As far as I know describing obstructions to pseudorandomness is open even for Sperner's theorem or Kruskal-Katona theorem (the latter corresponds to dealing with sets  $\Gamma_{a,b,c}$ ). Consider the following simple problem:

Let  $A$  be a dense subset of  $\binom{[n]}{n/2}$  and  $B$  be a dense subset of  $\binom{[n]}{n/2+1}$ . When is the number of pairs  $(a, b) \in A \times B$  such that  $a \subset b$  close to expected?

Besides the obvious obstruction when all elements  $a \in A$  contain  $i \in [n]$ , but none of elements  $b \in B$  do, there is also a more subtle obstruction. Let  $s$  be any subset of size  $n/2$  in  $[n]$ . Choose any  $4 \leq t = o(\sqrt{n})$ . Let  $a$  belong to  $A$  iff  $|a \cap s| \bmod t$  is in the interval  $[0, t/2]$  and  $b$  belong to  $B$  iff  $|a \cap s| \bmod t$  is in the interval  $[(t/2 + 2) \dots t - 1]$ . The sets  $A$  and  $B$  are of density close to  $1/2$ , but their union is an antichain.

Of course, such examples can be made more complicated by taking several sets  $s$  for different moduli  $t$ .

I suspect that there might a Freiman-type theorem buried here. There is even a weak connection to sumset addition: a shadow of a set  $S$  is just  $S + e_1, \dots, e_n$  where  $e$ 's are the one-element sets, and the addition occurs in appropriate semigroup (it is easy to see that a few results from additive combinatorics do carry over to that setting).

49.  [Terence Tao](#) Says:

[February 2, 2009 at 8:47 pm](#) | [Reply](#)

(45) Not a major concern at the moment, but if at some point we write up the precise bounds obtained from Behrend set constructions, we may as well use the best known Behrend set bound, due to Elkin, see e.g. this exposition of Green and Wolf:

<http://arxiv.org/abs/0810.0732>

50. [William Gasarch](#) Says:

[February 2, 2009 at 9:42 pm](#) | [Reply](#)

45.5. In the paper

Some Lower Bounds for the Hales-Jewitt Numbers  
by Eric Tressler

he has the following:

$HJ(3,3) > 6$

$HJ(4,3) > 6$

51.  [Terence Tao](#) Says:

[February 2, 2009 at 9:45 pm](#) | [Reply](#)

(46) [Numbering of the previous two posts needs to be changed]

@Jason (42): in view of existing examples, it may be worth restricting one's search space to those sets that are unions of  $\Gamma_{a,b,c}$ 's, rather than arbitrary subsets of  $[3]^n$ . The problem is now some sort of weighted corner-avoidance problem, but one which may be more tractable to brute force computer search.

@Boris: The problem you describe sounds very close to the machinery of influence, hypercontractivity, isoperimetric inequality, etc., but Gil is of course the one to ask regarding these things.

52.  [jozsef](#) Says:

[February 2, 2009 at 9:52 pm](#) | [Reply](#)

Comment on the lower bound. (Gil 15, Terry 16, and Terry 39&44) Without losing anything, we can suppose that the number of 0-s, 1-s, and 2-s are  $d \pm \sqrt{d}$ . Then the range where  $a+2b$  (the sum of 0-s plus twice the sum of 1-s) is triangle free has length about  $\sqrt{d}$ . If we care about the details, like we want to use Elkin's improvement, then we should note this one as well. It changes the constant multiplier in the exponent.

53.  [jozsef](#) Says:

[February 2, 2009 at 10:09 pm](#) | [Reply](#)

(47) contd. I think that we should get a better lower bound by using bounds on corners than on 3-term arithmetic progressions, however I'm not sure what is the best known lower bound on corners.

54.  [Tyler Neylon](#) Says:


[February 2, 2009 at 10:39 pm](#) | [Reply](#)

(48)


@Jason (42): The structure you point out - a bipartite graph with one set of vertices representing combinatorial lines  $\mathcal{L}$ , and the other representing elements of  $[3]^n$  - can be used to algorithmically find  $c_n$  exactly (it sounds like you're going for a probabilistic approach), although the code is more involved, and a bit slow (but polynomial in the size of the graph).

I think we can reduce this to a min-cost flow problem. Basically, we want the smallest set of edges which covers  $\mathcal{L}$  — this corresponds to the smallest subset of  $[3]^n$  which can be removed to get a line-free set. Create the following directed graph: add vertex  $s$  with edges to everything in  $\mathcal{L}$ . Every line in  $\mathcal{L}$  has an edge to all three points it contains. Every point in  $[3]^n$  points to a new vertex  $t$ . Every edge has unit cost, and unit capacity, except the edges between  $\mathcal{L}$  and  $[3]^n$ , which have unlimited capacity. Then a min-cost flow from  $s$  to  $t$  with at least  $|\mathcal{L}|$  capacity is what we want. This can be found in time  $O(n^2 m^3 \log n)$  where  $n$  is number of nodes,  $m$  is number of edges [see Schrijver's Combinatorial Optimization, ch 12]. Not super fast, but maybe do-able for small values of our  $n$ . Also possibly of theoretical interest?


Side note: In case my latex notation doesn't turn out correctly, could someone mention the proper syntax for these comments? And in case it does look right (for others), I'm just surrounding the math parts in double dollar signs.

55.  [Tyler Neylon](#) Says:  
[February 2, 2009 at 10:40 pm](#) | [Reply](#)


oops, that cool-looking face is supposed to be a forty-eight ( 48 )

56.  [Jozsef](#) Says:  
[February 2, 2009 at 10:46 pm](#) | [Reply](#)

(48) To Christian (43) If  $r=s=1$  then you get 3,3,3 as an arithmetic progression. I don't see a simple way to resolve the problem of degenerate arithmetic progressions in the general setting.

57.  [Jozsef](#) Says:  
[February 2, 2009 at 10:48 pm](#) | [Reply](#)


oops twice! (as ( 48 ) was used twice)

58.  [Terence Tao](#) Says:  
[February 2, 2009 at 11:14 pm](#) | [Reply](#)

(50) [I don't envy Tim's job to clean up the numbering system... but one can of course use fractions or something...]

@ Jozsef (47), the best lower bound for corners is necessarily between that for 3-term progressions and for triangle removal, since triangle removal implies corners implies 3-term progressions. But ever since the original Ruzsa-Szemerédi paper, the best lower bounds for triangle removal come from... the Behrend construction for 3-term progressions. This is not a satisfactory state of affairs, but I don't think there has been much progress in this regard. (Although I could imagine that the Elkin refinement could be pushed a little further for corners instead of 3-term progressions.)


[Of course, Gowers has tower-type lower bounds for the regularity lemma used to prove triangle removal, but this does not directly seem to lead to any new lower bounds on the latter.]

59.  [Terence Tao](#) Says:  
[February 2, 2009 at 11:17 pm](#) | [Reply](#)


(51) Incidentally, to make latex comments, use “`< em > [Your LaTeX code] < /em >`”. There are some quirks to the code, see the bottom of my post

<http://terrytao.wordpress.com/about/>

for details.

60.  [Terence Tao](#) Says:  
[February 2, 2009 at 11:19 pm](#) | [Reply](#)

Well, that failed horribly ... I meant “`$\textit{[Your LaTeX code]}$`” (where I artificially italicised the first dollar sign).

61.  [Jason Dyer](#) Says:  
[February 3, 2009 at 1:08 am](#) | [Reply](#)

(52) Something that might help me out — could someone produce a counterexample where the greedy algorithm I

outlined above in (42) \*doesn't\* work?

62.  [Ryan O'Donnell](#) Says:

[February 3, 2009 at 1:15 am](#) | [Reply](#)

( 53 ). Hi Boris! My student Karl Wimmer and I have some recent results about obstructions to pseudorandomness for Kruskal-Katona. Karl's still writing them up for his thesis but he's currently sidetracked with job interviews etc.

The results say what you might guess: Roughly, if  $A$  is a subset of  $\{[n] \text{ choose } n/2\}$  with constant density  $\mu$ , then the density of the upper shadow of  $A$  is at least  $\mu + \text{const. } \log(n)/n \dots$  unless  $A$  is very noticeably correlated with a single coordinate. As Terry guesses, it's "influences", "hypercontractivity", "Gil stuff" etc.

Karl was planning to finish writing it and its applications in Learning Theory within two months. But if this polymath project veers too close I guess we'll have to try to "rush to print" 😊 I currently don't see connections to the analogous Sperner problem; however I've only thought about it for like 5 minutes...

63.  [Terence Tao](#) Says:

[February 3, 2009 at 1:33 am](#) | [Reply](#)

(54) Tim Austin had a small comment which may be a clue of some sort: the ergodic theory proof of DHJ is heavily reliant on the ordering of the index set  $[n]$ , even though the conclusion of DHJ is of course invariant under permutations of the index set. (For instance, it is convenient to divide  $[n]$  into  $[n-r]$  and  $[r]$  for some moderately large  $r$ , and re-interpret subsets of  $[3]^n$  instead as a family of subsets of  $[3]^r$ , indexed by elements of  $[3]^{n-r}$ .) The standard proof of the colouring HJ theorem also has this sort of feature, in that one must arbitrarily order  $[n]$ , or at least partition it into components, in some fashion. So one may have to "break symmetry" here and not rely on a purely graph-theoretic argument.

64.  [gowers](#) Says:

[February 3, 2009 at 1:52 am](#) | [Reply](#)

55. I've been busy this evening and am about to go to bed, so only time for a quick remark at this stage, which again may be implicit in previous remarks. Let me call a set of the form  $\{(a+x, b+y, c+z) : x, y, z \geq 0, x+y+z=m\}$  a *triangular grid* of sidelength  $m$ . The remark is that by means of an averaging argument, there is no problem in finding a triangular grid of sidelength  $n^{1/3}$ , say, such that for every element of the grid all of  $(a+x, b+y, c+z)$  is close to  $(n/3, n/3, n/3)$ , and such that a positive proportion of the sets  $\Gamma_{a+x, b+y, c+z}$  are rich in elements of  $A$ . If we restrict to such a set, then we can hope for a Varnavides-type theorem because now a typical combinatorial line will have the same typical size of variable set as the sidelength of a typical equilateral triangle in the grid. In other words, in this set we no longer have the problem that the natural weights differ from the natural weights for corners. I'll try to make this point more precisely tomorrow.

65.  [Terence Tao](#) Says:

[February 3, 2009 at 3:04 am](#) | [Reply](#)

56. I thought it might also be worth reporting here on some things I tried that *don't* seem to work, though perhaps I am missing something.

Firstly, I tried improving the lower bound we now have on  $c_n$  by not taking the global statistics  $a, b, c$  of 1s, 2s, and 3s (and their resulting sets  $\Gamma_{a,b,c}$ , but instead partitioning  $[n]$  into subsets, using the local statistics for each subsets to create local versions of the  $\Gamma$ 's, and then somehow gluing it all back together again. This failed to improve upon what we already have, and I think the reason is while the global  $\Gamma_{a,b,c}$  have no combinatorial lines, the same is not true of their local analogues (unless one takes a Cartesian product of several local  $\Gamma$ 's, but this seems to be rather lossy.) Indeed, I

still don't know of any good example of a really large set with no combinatorial lines that isn't somehow built out of the  $\Gamma$ 's.

The other thing I toyed with was trying to use the machinery of shifting and compression (in the spirit of Frankl, or Bollobas-Leader, or one of my papers with Ben Green) to try to get good answers to the  $k=2$  problems (e.g. finding a good "robust" version of Sperner's theorem, e.g. a Varnavides version). But this seems not to work too well. The problem seems to be that the extremisers for Sperner's theorem don't look much like downsets (they seem to be more like the boundary of a downset, though I don't see how to exploit this). In any event, these techniques are quite specific to  $k=2$  and would be unlikely to shed much light on  $k=3$  except perhaps by analogy.



66. *Ryan O'Donnell Says:*

[February 3, 2009 at 5:35 am](#) | [Reply](#)

( 57 ). On the topic of more analytic proofs of Sperner: I think I now see why Boris mentioned Kruskal-Katona in the same breath as Sperner. Let me try to sketch the connection I see. I'm sure it's well known.

Let  $n$  be odd for simplicity and write  $B_j$  for the set of strings in  $\{0, 1\}^n$  with exactly  $j$  1's. Let  $k = \lfloor n/2 \rfloor$ .

Let  $A_j$  denote  $A \cap B_j$  and let  $\mu_j$  denote  $A_j$ 's density within  $B_j$ .

One particular thing Sperner tells us is that if  $\mu_k + \mu_{k+1} > 1$  then  $A$  contains a "combinatorial line" (because the overall density of  $A$  is at least that of  $B_k$ 's density within  $\{0, 1\}^n$ ). I.e., there is  $x \in A_k, y \in A_{k+1}$  such that  $x \prec y$ .

Here is a way to see that with Kruskal-Katona: K-K implies that the upper-shadow  $\partial^+ A_k$  of  $A_k$  has density at least  $\mu_k$  within  $B_{k+1}$ . Now together  $\partial^+ A_k$  and  $A_{k+1}$  have density exceeding 1 within  $B_{k+1}$ , hence they overlap.

—

Indeed, I think you can get most of Sperner out of extending this argument. Let me be uncareful. Suppose that  $A$ 's overall density exceeds that of  $B_{k+1}$  within  $\{0, 1\}^n$ . As an example, perhaps  $A$  has 40% of the strings in  $B_k$ , 40% of the strings in  $B_{k+5}$  and 40% of the strings in  $B_{k+7}$ . Now K-K implies that  $A_k$ 's upper shadow is also at least 40% of  $B_{k+1}$ . And the upper shadow of that is at least 40% (basically) of  $B_{k+2}$ . Etc. When you get up to  $B_{k+5}$ , you either get a collision and are done, or you have at least 80% of  $B_{k+5}$ . Then you get at least 80% of  $B_{k+6}$ , and then you have 120% of  $B_{k+7}$  and you get a collision and are done.

—

The reason I kind of like this proof is that there is an "analytic" proof of Kruskal-Katona that seems amenable to "robustification".

Although to be honest, I can't quite remember how we got talking about this subject just now...


67. *Boris Says:*

[February 3, 2009 at 6:14 am](#) | [Reply](#)

(58) The reason I mentioned Kruskal-Katona is that Sperner's original proof of Sperner's theorem involved pushing shadows towards the center level like Ryan indicated. He used simple double-counting argument, of which Kruskal-Katona is the modern descendant.

The point of the example I gave is to show that even if the shadow grows by only  $1+o(1)$ , it does not imply correlation with a single coordinate or even a small set of them.


The version of Kruskal-Katona qualitatively similar to what Karl Wimmer and Ryan proved was proved recently by Peter Keevash <http://arxiv.org/abs/0806.2023>

68.  Sune Kristian Jakobsen Says:  
[February 3, 2009 at 11:48 am](#) | [Reply](#)

59.

It is possible to show that  $c_4 \geq 49$ . Look at the set

$$\{(x, y, z, t) \in \mathbb{Z}_3^4 : x+y+z+t \neq 0 \pmod{3}\} \setminus \{(1, 1, 1, 1), (1, 1, 1, 2), (1, 1, 2, 1), (1, 2, 1, 1), (2, 1, 1, 1)\}$$

69.  gowers Says:  
[February 3, 2009 at 11:50 am](#) | [Reply](#)

60. Thread title: Jozsef's alternative problem.


I'm going to start by adding a few small thoughts on things I was writing about yesterday, and after that I'll read more carefully the recent batch of comments and see if I have any reactions to them. Yesterday (comment 23) I suggested the following variant of Jozsef's alternative problem (comment 2). Suppose  $A$  is a dense subset of the set of all pairs  $(U, V)$  of subsets of  $[n]$ . Can you find three disjoint sets  $U, V$  and  $D$  such that  $(U, V)$ ,  $(U, V \cup D)$  and  $(U \cup D, V)$  all belong to  $A$ ?

This question can't have grabbed people's attention much, since the answer is a trivial no. Just let  $A$  be the set of all pairs  $(U, V)$  such that  $U \cap V \neq \emptyset$ . That's definitely almost all pairs.

However, can we make a sensible question out of this? A sensible question would be some density-type condition on  $A$  that would guarantee a triple of the given kind. An obvious candidate would be to insist from the start that  $A$  is a set of *disjoint* pairs of sets. Note that there are  $3^n$  of these, since each element can choose whether to go into  $U, V$  or neither.

Just for fun, let's represent each pair  $(U, V)$  of disjoint sets as a sequence of 1s, 2s and 3s. Then a configuration of the desired kind is ... a combinatorial line.

So this rather nice new question was in fact the question we started with. But at least this thought process clarifies the relationship between Jozsef's problem and density Hales-Jewett.

70.  gowers Says:  
[February 3, 2009 at 1:34 pm](#) | [Reply](#)

61. Thread title: Varnavides. (This is the thread devoted to thinking about how to find a suitable statement that says that a dense set doesn't just contain one combinatorial line but contains "a dense subset of all possible combinatorial lines". The problem is that the most natural interpretation of this is false. See Jason Dyer's comment just after 38 for numbers of other comments that belong to this thread.)

I want to continue with what I was talking about in 55. At the moment I am in the happy state of not knowing why we are not already done. Of course, I don't for a moment think that we *are* done, but being in this position is always nice because you know that by the time you see properly why you aren't done you are guaranteed to understand the problem better.

The grounds for this unrealistic optimism are that we now seem to have an easy solution to the Varnavides problem. Just to recap, the problem is that one would expect any triangle-removal argument to lead not just to the conclusion that a dense set contains one combinatorial line, but to the conclusion that the number of combinatorial lines it contains is within a constant of the trivial maximum (which is what you get when your set has density 1).

To recap further, this statement is false for density Hales-Jewett, because you can take as your set the set of all sequences  $x$  that take each value roughly  $n/3$  times.

Standard probabilistic arguments show that almost all sequences have this form, but the size of the set of variable

coordinates in any combinatorial line in this set is very small, and this can be used to show that the set of combinatorial lines has cardinality far smaller than  $4^n$ . (This is the number of combinatorial lines, because each point chooses whether to be 1, 2, 3 or variable.)

Let's regard this as a Varnavides-summary comment. I'll continue this line of thinking in my next comment.

71.  [gowers](#) Says:

[February 3, 2009 at 1:47 pm](#) | [Reply](#)

62. Thread title: Varnavides.

A problem we were grappling with yesterday was how to put a natural measure on the set of combinatorial lines in such a way that the variable sets were forced to be small. Here is a proposal for a solution to that problem. I've already hinted at it in 55, and I think Terry may have had it in the back of his mind at around 39 and 40. Briefly, the suggestion is to restrict to a *large triangular grid of slices*. Let me elaborate on what I mean by this.

First, a *slice* is a set of the form  $\Gamma_{r,s,t}$ , where that is the set of all  $x$  with  $r$  1s,  $s$  2s and  $t$  3s. (There's redundancy in the notation, since  $r + s + t = n$ , but the symmetry more than makes up for it.)

Secondly, a *triangular grid* is a subset  $T$  of the set  $T_n = \{(r, s, t) : r, s, t \geq 0, r + s + t = n\}$  of the form  $\{(a + r, b + s, c + t) : r, s, t \geq 0, r + s + t = m\}$  for some choice of  $a, b, c, m$  such that  $a, b, c \geq 0$  and  $a + b + c \leq n - m$ . Triangular grids are just translates of  $T_m$  that live inside  $T_n$ . (It would be possible also to consider dilated triangular grids where you look at points of the form  $(a + rd, b + sd, c + td)$  for some fixed  $d$ , but at the moment I have no need of these.)

A *triangular grid of slices* is then just a collection of slices indexed by a triangular grid.

Again, I think it makes sense to stop this comment here and continue in a new comment.

72.  [gowers](#) Says:

[February 3, 2009 at 2:03 pm](#) | [Reply](#)

63. Thread title: Varnavides.

Now I make the following claim. Suppose we choose a random point in  $[3]^n$  as follows. First we choose a random element of  $T_{n-m}$ , where the probability of choosing  $(a, b, c)$  is  $3^{-(n-m)} \binom{n-m}{r,s,t}$ . (Equivalently, we choose a random element of  $[3]^{n-m}$  and let  $a, b, c$  be the number of 1s, 2s and 3s.) Next, we choose a random element  $(r, s, t)$  of  $T_m$ , this time from the *uniform* distribution. Finally, we choose a random element of the slice  $\Gamma_{a+r,b+s,c+t}$ . The claim is that the resulting probability distribution on  $[3]^n$  is very close to uniform. (By this I mean that any event in one distribution has a probability that's within a very small additive constant of the probability according to the other distribution.) The reason for this is that when  $r, s, t$  are all smaller than  $m$  and  $a, b, c$  are all around  $n/3$  (as they almost always are), then the ratio of  $\binom{n-m}{a,b,c}$  to  $\binom{n}{a+r,b+s,c+t}$  is very close to 1.

An averaging argument tells us that there must exist  $(a, b, c)$  close to  $(n/3, n/3, n/3)$  such that a positive proportion of points in the triangular grid of slices indexed by  $(a, b, c) + T_m$  belong to  $A$ . So instead of looking at density Hales-Jewett in all of  $[3]^m$  we can try to prove that if  $A$  is a dense subset of a big triangular grid of slices, then  $A$  contains a combinatorial line. In my next comment I'll try to say why I think this is a good thing to do.

73.  [gowers](#) Says:

[February 3, 2009 at 2:25 pm](#) | [Reply](#)

64. Thread title: Varnavides.




Note that if we restrict to a large triangular grid of slices in this way, then we have various convenient properties. For example, all the points in all the slices have about  $n/3$  each of 1s, 2s and 3s. And any combinatorial line has a variable set of size at most  $m$ . I'd like to claim now that a Varnavides-type theorem applies to large triangular grids: that is, if you have a dense subset of a large triangular grid of slices, then not only do you get a combinatorial line, but you even get within a constant of the total number of combinatorial lines in the entire grid. The idea here, by the way, is not (yet) to try to prove this theorem directly, but just to check that it is true by showing that it follows from density Hales-Jewett.

Here's how density Hales-Jewett can be applied. Let's assume that  $A$  is a dense subset of the triangular grid of slices indexed by  $(a, b, c) + T_m$ , where  $(a, b, c)$  belongs to  $T_{n-m}$ . Suppose that  $h$  is a fixed constant chosen to be large enough that every subset of  $[3]^h$  of size at least  $(\delta/2)3^h$  contains a combinatorial line. Now for every  $(r, s, t) \in T_{m-h}$ , which means almost every  $(r, s, t) \in T_m$  since we're assuming that  $m$  is much much larger than  $h$ , let us randomly choose  $a + r$  places to put a 1,  $b + s$  places to put a 2 and  $c + t$  places to put a 3. That leaves  $h$  places unfilled. The set of ways of filling those places is a copy of  $[3]^h$ , so if it contains at least  $(\delta/2)3^h$  points from  $A$ , then it contains a Hales-Jewett line.

An averaging argument shows that that happens with ...

OK, I'm sorry to say that I do after all need to consider dilated triangular grids. Since that is the case, I'm going to abandon this attempt to be precise, for now at least, and simply say in the next comment what I think should happen. Then maybe others will tell me that they're not convinced, or that they are convinced, and perhaps we can zero in on something.


74.  *Ryan O'Donnell* Says:  
[February 3, 2009 at 2:48 pm](#) | [Reply](#)

65. If there's a "Sperner" thread I guess this would be it.

Boris, thanks for the reference to the Keevash paper. While you're handing out references 😊 ... I guess it would be relevant to ask, Is there a Sperner's Theorem known for different product distributions on  $\{0, 1\}^n$ ?

In the case relevant to us: What is the densest antichain in  $\{0, 1\}^n$  under the  $(1/3)$ -biased product distribution?

This must be known...

75.  *gowers* Says:  
[February 3, 2009 at 3:16 pm](#) | [Reply](#)

66. Thread title: Varnavides.

All I was really trying to get at in 63 was this. Choose a random translate and dilate of the triangular grid  $T_h$  in such a way that it is a subset of  $(a, b, c) + T_m$ . If you make the dilates you look at all have width at most  $\eta m$  for some smallish  $\eta$ , then nearly all aligned equilateral triangles in  $(a, b, c) + T_m$  will be included in one of these copies of  $T_h$  about the same number of times.

Now let  $S_h$  be a dilate and translate of  $T_h$ , where the constant of dilation is  $d$ . One can randomly choose a copy  $X_h$  of  $[3]^h$  inside  $[3]^n$  in such a way that the numbers of 1s, 2s and 3s is given by points in  $X_h$  is given by points in  $S_h$ . To do this, suppose that  $S_h = (a, b, c) + dT_h$ . Randomly fix  $a$  coordinates to be 1,  $b$  coordinates to be 2,  $c$  coordinates to be 3, and randomly partition the rest of  $[n]$  into  $h$  sets  $D_1, \dots, D_h$ , each of size  $d$ . Then our random copy  $X_h$  consists of all sequences that are constant on each  $D_i$  and take the given fixed values outside the  $D_i$ .

My claim is that a simple averaging argument should show that if a positive proportion of these  $X_h$  contain a Hales-Jewett line (as I claim they will if  $A$  has density  $\delta$ ) then the number of Hales-Jewett lines in  $A$  is within a constant of



the trivial maximum. I'm going to leave this for now as I feel that to do it properly I'd have to go away and write something down, and I'm not allowing myself to do that.

76.  [gowers](#) Says:

[February 3, 2009 at 3:26 pm](#) | [Reply](#)

67. Possible thread title: triangle removal.

Actually, I was trying to prove a Varnavides-type result not because we need it but because it was a sort of reality check: if we can't get such a result then it is unlikely that triangle removal will work. But since I'm now fairly confident that we can get such a result, I want to see what happens if one just goes ahead and tries to apply some kind of triangle-removal lemma in the case where  $A$  is a dense subset of a large triangular grid of slices.

So let's fix notation again. I'm going to suppose that  $A$  is a dense subset of the union of all slices  $\Gamma_{a+r, b+s, c+t}$ , where  $(a, b, c)$  is some fixed element of  $T_{n-m}$  and  $(r, s, t)$  ranges over  $T_m$ . Also, I'm assuming that  $a, b, c$  are all around  $n/3$  so that all slices have about the same size.

Now I want to apply the trick from comments N and P. So what is the graph now? Instead of taking  $X$  to be the full power set of  $[n]$  I want to take just those sets that could conceivably occur as the set of 1s in a sequence in  $A$ . And that's all sets of size  $a + r$ , where  $0 \leq r \leq m$ . Similarly,  $Y$  consists of all sets of size between  $b$  and  $b + m$ , and  $Z$  consists of all sets of size between  $c$  and  $c + m$ . The edges are as before: I join  $U \in X$  to  $V \in Y$  if the sequence that's 1 on  $U$  and 2 on  $V$  and 3 everywhere else belongs to  $A$ . And similarly for the other two bipartite graphs. And clearly a non-degenerate triangle in this graph gives rise to a combinatorial line.

I'll think about the implications of this in my next comment.

77.  [gowers](#) Says:

[February 3, 2009 at 3:44 pm](#) | [Reply](#)

68. Thread title: triangle removal.

I didn't mention it explicitly, but of course a necessary condition for joining  $U \in X$  to  $V \in Y$  is still that  $U$  and  $V$  should be disjoint. So the first reason that we are not yet done is that the graphs are still very sparse, since two random sets of size  $n/3$  have a tiny probability of being disjoint. So the question is, has anything been gained by restricting attention to a large triangular grid of slices? Something's been gained from the Varnavides point of view, I think, but has anything been gained from the triangle-removal point of view? That's what I'm asking.

At the moment it feels, rather disappointingly, as though *nothing* has been gained. It's still the case that the graph where you join two sets if they are disjoint is importantly non-quasirandom, since the remark made in comment LL carries over to this context too.

I would hazard a guess that there are two morals to draw from this.

- (i) It may still be quite convenient to restrict to a triangular grid of slices, in order to make certain distributions uniform.
- (ii) We are probably forced to consider obstructions to uniformity.

In fact, there's a different way of justifying (ii). At some point, if a triangle-removal argument is going to work, we will have to prove a statement to the effect that if you have three bipartite graphs that are all "regular" and "dense", then they must give rise to lots of triangles. In the normal dense-graphs case, "regular" just means quasirandom, and it even suffices for just one of the three bipartite graphs to be regular, provided there are plenty of vertices in the third set with high degree in both of the first two sets.

In our case, there is a one-to-one correspondence between edges of one of the bipartite graphs and points in  $[3]^n$ . So this suggests that at some point we are going to have to understand what a “regular” or “quasirandom” subset of  $[3]^n$  is like. To be continued in the next comment.

78.  [gowers](#) Says:

[February 3, 2009 at 3:54 pm](#) | [Reply](#)

69. Thread title: obstructions to uniformity.

What does it mean to “understand what a quasirandom subset of  $[3]^n$  is like? For those who don’t find the answer obvious, here’s an attempt to explain.

(i) We are looking for a *definition* of quasirandomness.

(ii) Experience with other problems suggests that this definition should have the following property: if  $A$  is a dense quasirandom subset of  $[3]^n$  and  $B$  is an arbitrary dense subset of  $[3]^n$ , then there is a combinatorial line  $(x, y, z)$  with  $x \in A, y \in B$  and  $z \in B$ . Indeed, one should have not just *one* such combinatorial line but a *dense set* of such combinatorial lines. (Here is a place where having a Varnavides-type statement is actually a big help and not just a mere reassurance.)

(iii) Experience with other problems also suggests that if  $A$  does *not* have the property in (ii) then there will be some simple reason for it. An example of such a reason is that  $x_1$  might be 1 for every  $x \in A$ . (This would fail to have the property in (ii) because we could take  $B$  to be the set of all sequences with  $x_1 = 2$ .)

(iv) If one can think of all possible “simple reasons” of the kind discussed in (iii), then it is often possible to pass to substructures or partitions of the ground set and obtain density or energy increases. (If you don’t know what this means, it’s not too important at this stage. The main point is that if you can do (iii) then there are standard techniques for using what you’ve done. In many other contexts, once you’ve got (iii) the problem is basically solved.)

In the next comment, I’m going to write down all the obstructions to uniformity that I can think of. (An *obstruction to uniformity* is basically what I called a “simple reason” in (iii).)

79.  [gowers](#) Says:

[February 3, 2009 at 4:15 pm](#) | [Reply](#)

70. Thread title: obstructions to uniformity.

Before I start this comment, I want to make a metacomment. If you are reading this and are not familiar with the kinds of techniques people are throwing around, and therefore feel you can’t really contribute, here is a place where you don’t necessarily need to know *anything* to contribute: we have a simply stated question, which may have a simple answer. (The question will be of the form, “Here are all the Xs I can think of off the top of my head. Can anyone think of any more?”)

For the purposes of this comment, I’ll define a set  $A \subset [3]^n$  of density  $\alpha$  to be *uniform* if for every set  $B \subset [3]^n$  of density  $\beta$  the number of combinatorial lines  $(x, y, z) \in A \times B^2$  is approximately  $\alpha\beta^2$  times the number of combinatorial lines in  $[3]^n$ . Similar definitions apply if we restrict to a triple of slices or a triangular grid of slices.

Obstruction 1. In any combinatorial line, a coordinate must either take all three values or just one. Therefore, a set  $A$  is not uniform unless the number of sequences  $x$  in  $A$  such that  $x_i = j$  is roughly  $|A|/3$  for every  $i \in [n]$  and  $j \in \{1, 2, 3\}$ . To prove this seems to be not quite trivial. An obvious set  $B$  to consider is to take  $j$  such that the number of points with  $x_i = j$  is significantly less than  $|A|/3$  and let  $B$  be the set of all sequences  $x$  such that  $x_i = j$ . Then the only points in  $A$  that can make lines with  $B$  are ones with  $x_i = j$ . Therefore, if we get the right number of combinatorial lines in  $A \times B^2$ , then we can drop to the subset  $A'$  that consists of those points only and we find that

we get more combinatorial lines than we would expect (given the densities of the restrictions of  $A$  and  $B$  to  $x_i = j$ ). I think one can mess around with this information and obtain a set  $B$  that proves that  $A$  is not uniform.

Given that that was slightly tricky (it might be nice to do it properly though), I'll restrict attention for now to what I might call *strong* obstructions to uniformity. I'll call a set  $A$  *strongly non-uniform* if I can find a dense set  $B$  such that there are *no* combinatorial lines  $(x, y, z)$  with  $x \in A$ ,  $y \in B$  and  $z \in B$ . So finding more strong obstructions really is a genuinely elementary problem (in the sense of not needing special knowledge).

Having come to this decision, let me start again in a new comment.

80.  [gowers](#) Says:

[February 3, 2009 at 4:49 pm](#) | [Reply](#)

71. Thread title: obstructions to uniformity.

Strong obstruction 1. If there exist  $i$  and  $j$  such that  $x_i = j$  for every  $x$  in  $A$ , then  $A$  is strongly non-uniform. Proof: let  $k \neq j$  and let  $B$  be the set of all sequences such that  $x_i = k$ . You can't have a combinatorial line in which the  $i$ th coordinate is  $j$  once and  $k$  twice.

Strong obstruction 2. If there is a positive integer  $m$  such that for every  $x \in A$  the number of 1s, 2s and 3s in  $A$  is always the same mod  $m$ , then  $A$  is strongly non-uniform. Proof: If  $y$  and  $z$  belong to a combinatorial line and have the same number of 1s, 2s and 3s mod  $m$ , then the size of the variable set must be a multiple of  $m$ . Therefore, we can choose  $B$  to be a set where the sequences again have fixed numbers of 1s, 2s and 3s mod  $m$ , but different numbers from those of  $A$ .

Strong obstruction 3. This generalizes the previous example. Suppose that for every  $x \in A$  we know that the number of 1s belongs to a set  $E_1$  of residues mod  $m$ , the number of 2s belongs to a set  $E_2$ , and the number of 3s belongs to a set  $E_3$ . Now let  $F_1, F_2, F_3$  be sets of residues mod  $m$ , and let  $B$  be the set of all sequences with the number of 1s in  $F_1$  etc. If  $y$  and  $z$  are points of  $B$  that belong to a combinatorial line, then let's suppose that to get from  $y$  to  $z$  you change some 2s to 3s. This tells us that the size of the variable set belongs to  $F_3 - F_2$  mod  $m$ . This in turn tells us that the number of 1s in the third point of the line belongs to  $F_1 + F_2 - F_3$  mod  $m$ . So if  $E_1$  is disjoint from any set of the form  $F_1 + F_2 - F_3$ , then  $A$  is strongly non-uniform. And similarly for the other variables. Of course, if  $E_1$  doesn't consist of all residues, then one might say that we could take  $F_1, F_2, F_3$  to be singletons. But that's true only if  $m$  is small, since otherwise  $B$  is not dense. This example works even when  $m$  is large — it can tend to infinity with  $n$ . (An example might be when we insist that for all sequences the numbers of 1s, 2s and 3s is between  $-cm$  and  $cm$  mod  $m$  for some small constant  $c$ .)

Strong obstruction 4. One can generalize further by passing to a subset. That is, instead of looking at the numbers of 1s, 2s and 3s in the whole sequence, one just looks at the numbers of 1s, 2s and 3s in some specified set  $X$  of coordinates. The proof is more or less the same.

I haven't tried hugely hard to find more obstructions. Does anyone know of any? Does Boris's comment 44 have any implications here?

81. [Boris](#) Says:

[February 3, 2009 at 4:52 pm](#) | [Reply](#)

72. Regarding the definition of triangular grid in #62 at 1:47 pm, to make sure I understood it properly, should it say  $a + b + c \leq n - m$  instead of  $a, b, c \leq n - m$ ?

[Thanks --- I've changed it now.]

Regarding the obstructions to the uniformity mentioned in #69 3:54 pm, doesn't it follow from discussion about

Varnavides that to talk about pseudorandom sets we want our sets to be dense subsets in a suitable union of  $\Gamma_{a,b,c}$  that live near the middle of the lattice?

Answer to Ryan's question from #65 at 2:48 pm: I can't think of a reference for the result you want, but it follows from LYM inequality that the level of largest weight is the largest antichain.

82. [Boris](#) Says:

[February 3, 2009 at 5:03 pm](#) | [Reply](#)

73. One can generalize obstructions from #71 at 4:49 pm further: instead of counting number of 1's 2's and 3', weight individual coordinates by small integer weights. Passing to a subset is a special case of that when some weights are zero.

*[Not sure what was going wrong, but I've deleted the other two copies of this comment.]*

83.  [gowers](#) Says:

[February 3, 2009 at 6:11 pm](#) | [Reply](#)

74. Thread title: obstructions to uniformity.

In answer to Boris's question (72, second paragraph), I agree that it looks as if any definition of pseudorandomness will have to restrict attention to something like a triangular grid of slices near the middle. But I was hoping that we could do things in two stages. Stage 1 would be to find strong obstructions to uniformity. Then in stage 2 one would weaken the definition back to what I first had: instead of getting no combinatorial lines in  $A \times B^2$  one just asks for a number that differs from what you would expect for a random  $A$ . At that point, one would have to ask, "What do we expect from a random  $A$ ?" and then it would become important to restrict to a triangular grid or some such structure.

84. [Michael Nielsen » The polymath project](#) Says:

[February 3, 2009 at 6:24 pm](#) | [Reply](#)

[...] Apparently, the original proof of the Density Hales-Jewett theorem used ergodic theory; Gowers' challenge is to find a purely combinatorial proof of the theorem. More background can be found here. Serious discussion of the problem starts here. [...]

85.  [gowers](#) Says:

[February 3, 2009 at 6:25 pm](#) | [Reply](#)

75. Thread title: obstructions to uniformity.

Let me check that I understand Boris's comment 73. The idea is to pick an integer  $m$  and a sequence of integers  $r_1, \dots, r_n$ , and to let  $A$  be the set of all sequences  $x$  such that  $\sum r_i x_i$  belongs to a specified set  $F$  of residues.

Now choose  $B$  to be the set of all sequences  $x$  such that  $\sum_i r_i x_i \in F \pmod m$ . Then if  $x, y$  and  $z$  all lie in a combinatorial line, and  $y$  and  $z$  belong to  $B$ , then the characteristic function  $D$  of the set of variable coordinates (or wildcards, as some are calling them) has the property that  $\sum_{i \in D} r_i$  belongs to  $F - F \pmod m$ . It follows that  $\sum_i r_i x_i \in F - (F - F) = F + F - F$ . So if we can find a dense set  $F$  of residues mod  $m$  such that  $E \cap (F + F - F) = \emptyset$  then  $A$  is strongly nonuniform.

That neatly covers a lot of examples. Can anyone think of any more? Or what about heuristic arguments that these are the only examples? (At the moment, my guess is that there are further examples waiting to be discovered.)

86.  [gowers](#) Says:

[February 3, 2009 at 6:39 pm](#) | [Reply](#)

76. Thread title: obstructions to uniformity.

I think I've got a further generalization of 75. Just replace the group of integers mod  $m$  by an arbitrary finite Abelian group  $G$ . So now  $r_1, \dots, r_n$  are elements of  $G$ , and  $E$  and  $F$  are dense subsets of  $G$ . Then repeat exactly what I wrote in 75 except that "residues" or "residues mod  $m$ " should be replaced by "elements of  $G$ ".

87.  [Terence Tao](#) Says:

[February 3, 2009 at 6:42 pm](#) | [Reply](#)

77. Thread title: Is triangle removal still necessary?

Greetings again from the Pacific Standard Time zone!

I had one thought about the bigger picture. The initial reasoning in the main post goes like this:

1. The standard proof of the corners theorem uses the triangle removal lemma.
2. The DHJ theorem generalises the corners theorem.
3. Hence, it is natural to look for a proof of DHJ that uses (some generalisation of) the triangle removal lemma.

But, now that we have *used* the corners theorem in our arguments (to locate a triple  $(a+r, b, c)$ ,  $(a, b+r, c)$ ,  $(a, b, r+c)$  of rich tuples), the above logic loses quite a bit of its force. Indeed, if we look at the standard way in which the corners theorem is embedded in the DHJ theorem, by representing a set  $A \subset [n]^2$  as the set  $\mathcal{F}(A) := \bigcup_{(a,b) \in A} \Gamma_{a,b,3n-a-b} \subset [3]^{3n}$ , we can see that any new proof of DHJ we have will not give any new proof of the corners theorem for a given dense set  $A$ , since we will need to use the corners theorem for *precisely that set*  $A$  in order to get the rich triples. So there is no *a priori* reason for the rest of the proof to involve a triangle removal argument at all. (Of course, triangle removal is still implicitly used in the invocation of the corners theorem.)

(Note that the sketch of Sperner via Kruskal-Katona in Ryan.57 has this flavour: one uses things like the pair removal theorem from Gowers.21 to locate a rich pair  $B_k, B_{k+r}$ , and then uses completely different arguments from then on.)

There is however one caveat that seems to suggest that either some vestige of triangle-removal still remains in the rest of the argument, or else some preliminary regularisation is needed. The reason for this is that there are other ways to embed the corners problem into the DHJ problem. For instance, one can embed a corner-free set  $A \subset [n]^2$  into a combinatorial line-free set  $\mathcal{F}(A) \times \mathcal{F}(A) \subset [3]^{6n}$ . The richness profile  $\{(a, b, c) : \mathcal{F}(A) \times \mathcal{F}(A) \cap \Gamma_{a,b,c} \text{ large}\}$  of this set is essentially  $A+A$ , and thus much more likely to contain corners than the original set  $A$ . So if we were to somehow prove DHJ on a rich triple without any further triangle removal, this would imply a "nearly-triangle-removal-lemma-free" proof of the corners theorem of  $A$ , reducing it to the substantially simpler-looking problem of finding corners in  $A+A$ .

But if we could find some way to "detect" this sort of "product structure" in the set  $\mathcal{F}(A) \times \mathcal{F}(A)$ , and pass to an appropriate sub-index set of  $[n]$  in order to isolate a more "essential" component of this set, such as  $\mathcal{F}(A)$ , then this objection may disappear. This of course ties in with the obstructions-to-uniformity thread, and also with the counterexamples mentioned in Gowers.20, Gowers.38, and the "hereditary richness" idea I proposed in Terry.40, and maybe also the stationarity in Terry.4, Austin.16.

A serious difficulty, though, would be if there are many other ways to embed corners into DHJ that are not easily "detectable" by inspecting things like the richness profile of the embedded set with respect to various index sets. If there are an enormous number of such embeddings then it seems unlikely that we can eradicate triangle removal completely from the remainder of our arguments.

88.  [Terence Tao](#) Says:

[February 3, 2009 at 6:46 pm](#) | [Reply](#)

78. Thread title:  $c_n$  for small  $n$

@Sune.59: The lower bound  $c_4 \geq 49$  you have compares pretty well with the upper bound  $c_4 \leq 3c_3 = 54$ . It may be possible to shave the upper bound down a notch, by trying to figure out what the extremal subsets of  $[3]^3$  are that have 18 elements but no combinatorial lines. Any 54-element subset of  $[3]^4$  with no combinatorial lines must be extremal on every three-dimensional slice, which looks unlikely if there are few extremals.

89.  [gowers](#) Says:

[February 3, 2009 at 6:49 pm](#) | [Reply](#)

79. Thread title: Sperner

As will be obvious to many people, given some of the discussion, the question I asked in 28 is not yet a good one. Indeed, if we take both  $\mathcal{A}$  and  $\mathcal{B}$  to be the set of all sets of size between  $n/2 - C\sqrt{n}$  and  $n/2 + C\sqrt{n}$ , then there will be far fewer than  $3^n$  disjoint pairs, but it will be very hard to remove all disjoint pairs.

However, in the light of the Varnavides discussion, there is a natural way of rescuing the question. No time to check whether the result is sensible. Let's just restrict attention to sets of size between  $n/2 - m$  and  $n/2 + m$  for some fairly small  $m$ . If  $\mathcal{A}$  and  $\mathcal{B}$  are two subsets of this "middle part" of the power set of  $[n]$  such that the number of disjoint pairs  $(U, V)$  with  $U \in \mathcal{A}$  and  $V \in \mathcal{B}$  is far smaller than the number of disjoint pairs of sets of sizes between  $n/2 - m$  and  $n/2 + m$ , does it follow that we can remove a small number of sets from  $\mathcal{A} \cup \mathcal{B}$  and end up with *no* disjoint pairs?

90.  [gowers](#) Says:

[February 3, 2009 at 6:53 pm](#) | [Reply](#)

80. Thread title: Sperner

Actually, one might as well just restrict to single layers rather than a whole cluster of layers around  $n/2$ . Choose  $m$  slightly less than  $n/2$ , say, and make  $\mathcal{A}$  and  $\mathcal{B}$  consist of subsets of the set of sets of size  $m$ .

91. [Boris](#) Says:

[February 3, 2009 at 7:07 pm](#) | [Reply](#)

81. I am not sure whether generalization of Gowers #76 at 6:39 pm is truly a generalization since if a set  $F$  is nonuniform (as it must be for  $F - F + F \neq G$  to hold), then it correlates with some character. That should imply correlation with some other set for which  $G$ =cyclic.

92.  [Ryan O'Donnell](#) Says:

[February 3, 2009 at 7:23 pm](#) | [Reply](#)

82. Re Sperner:

I have to run to class in a second so I can't say if this is actually be relevant to TimG.79,80, but...

See this paper (and its referencers/referencees) for the theory on structure vs. pseudorandomness for independent sets in Kneser graphs:

<http://www.ams.org/mathscinet-getitem?mr=2178966>

93.  [Tyler Neylon](#) Says:

[February 3, 2009 at 7:45 pm](#) | [Reply](#)




83. Thread title:  $c_n$  for small  $n$ .

In comment 40, the set  $D_n := \{x \in \mathbb{Z}_3^n \mid \sum x_i \not\equiv 0 \pmod{3}\}$  was introduced. This set only includes lines  $\ell$  with  $\#\{\ell_i = *\} \equiv 0 \pmod{3}$  and  $(\sum_{\ell_i \neq *} \ell_i) \not\equiv 0 \pmod{3}$ . So when  $n=4$ , we can just exclude (still thinking in  $\mathbb{Z}_3$ ) points  $\bar{1}$  and  $\bar{2}$ .

This shows that  $52 \leq c_4 \leq 54$ .

For  $n=5$ , we can keep  $\bar{1}, \bar{2}$  excluded, and omit the extra 15 points given by any choice of  $\{a, b\} \subset \mathbb{Z}_3$  plugged into each of  $(ababa), (babab), (abbaa), (aaaab), (baaaa)$ . This is decidedly less elegant than the  $n=4$  case, but we see that


$145 \leq c_5 \leq 162$ .

94.  [Terence Tao](#) Says:  
[February 3, 2009 at 8:12 pm](#) | [Reply](#)

84. Thread title:  $c_n$  for small  $n$ .

@Tyler.83: nice! It seems that just excluding those  $x$  for which  $\sum_i x_i = 0 \pmod{3}$  already knocks out a significant fraction of combinatorial lines, at least when  $n$  is small. It may be that some secondary set of a similar nature can then be removed to knock out quite a few more lines for medium  $n$  (e.g.  $n = 5-10$ ), at which point the rest can be picked off by ad hoc methods. This may lead the way to suggest a different asymptotic lower bound than the current Behrend-based lower bound of  $c_n \geq 3^{n-O(\sqrt{\log n})}$ . If so, this may provide an excellent set of examples with which to test the upper bound strategies (the Behrend example tends to be rather intractable to compute with in practice).

So it might be worth trying to push the lower bounds for  $c_6, c_7$ , etc. as best we can, though it's much less likely now that we will be able to compute these numbers exactly. (I don't see the upper bounds budging much from the trivial  $c_6 \leq 2 \times 3^5 = 486$ ,  $c_7 \leq 2 \times 3^6 = 1458$  - after all, density Hales-Jewett is a *hard* theorem - though I would imagine that one should be able to maybe knock one or two from the upper bound for, say  $c_7$  (after all, every 2-dimensional slice of an extremiser here has to have exactly 6 elements, and there are only three such slices available).)

95.  [Terence Tao](#) Says:  
[February 3, 2009 at 8:32 pm](#) | [Reply](#)

85. Thread title: A weak form of DHJ

As I mentioned in Terry.8, DHJ is the statement that  $c_n = o(3^n)$  as  $n \rightarrow \infty$ . This is a hard theorem. But one can at least give an easy proof of the weaker statement that the trivial bound  $c_{n+1} \leq 3c_n$  must be strict (i.e.  $c_{n+1} < 3c_n$ ) for infinitely many  $n$ , as follows.

Suppose that  $c_{n+1} = 3c_n$  for all but finitely many  $n$ , then there exists an  $m$ , and an  $n$  much larger than  $m$ , such that  $c_n = 3^{n-m}c_m$ . Thus there exists a combinatorial line-free subset  $A$  of  $[3]^n$  with exactly  $3^{n-m}c_m$  elements.

Now think of  $[3]^n$  as a whole bunch of parallel copies of  $[3]^m$ , each parameterised by an element from  $[3]^{n-m}$ . As  $A$  has exactly  $3^{n-m}c_m$  elements and is combinatorial line-free, each  $m$ -dimensional slice of  $A$  must have exactly  $c_m$  elements. We can assign each element of  $[3]^{n-m}$  a colour based on exactly what the associated  $m$ -dimensional slice of  $A$  looks like; there are at most  $\binom{3^m}{c_m}$  colours available.

Now we use the *colouring* Hales-Jewett theorem, and conclude that if  $n$  is large enough depending on  $m$ , there exists a monochromatic line in  $[3]^{n-m}$ . Unwinding all this, we get a combinatorial line in  $A$  (note that all the slices of  $A$  are non-empty).

[Incidentally, an argument like this appears in Szemerédi's original proof of Szemerédi's theorem, though of course

many many more things need to be done in that proof beyond this simple argument. A lot of Szemerédi's argument, by the way, applies to Hales-Jewett, except for a crucial step where he applies what is now known as the Szemerédi regularity lemma. In the Hales-Jewett setting, the graph that one would wish to apply this lemma to is far too sparse for the strategy to work, it seems.]

96. [Mathematics, Science, and Blogs « Combinatorics and more](#) Says:

[February 3, 2009 at 8:42 pm](#) | [Reply](#)

[...] is getting quickly off the ground. There are several new posts in Gowers's blog describing the project, its rules and motivation, and interesting discussions also over What's new and in In theory. [...]



97. [Terence Tao](#) Says:

[February 3, 2009 at 9:03 pm](#) | [Reply](#)

86. Szemerédi's proof of Roth's theorem

While on the subject of Szemerédi's arguments, Endre also has a simple combinatorial proof of Roth's theorem (avoiding Fourier analysis or the regularity lemma), instead relying on the Szemerédi cube lemma. This can for instance be found in Graham-Rothschild-Spencer's book, or Chapter 10.7 of my book with Van, or in Ernie's writeup

<http://www.math.gatech.edu/~ecroot/szemeredi.pdf>

the basic idea is to observe that if a set  $A$  in  $[N]$  is dense, then it must contain a high-dimensional cube  $Q$ . If  $A$  has no arithmetic progressions, then  $2 \cdot Q - A$  must be essentially disjoint from  $A$ . Expressing the  $d$ -dimensional cube  $Q$  as the final element of a sequence  $Q_0 \subset Q_1 \subset \dots \subset Q_d$  where each  $Q_i$  is one dimension larger than the preceding  $Q_{i-1}$ , and using the pigeonhole principle, we can find  $i$  such that  $2 \cdot Q_i - A$  has roughly the same size as  $2 \cdot Q_{i+1} - A$ . We can write  $Q_{i+1} = Q_i + \{0, v_i\}$  for some step  $v_i$ , and we thus see that  $2 \cdot Q_i - A$  is very stable under shifting by  $v_i$ , and thus both it and its complement can be partitioned into long arithmetic progressions of step  $v_i$ . Since  $A$  is disjoint from  $2 \cdot Q_i - A$ , it is squeezed into the other progressions, and on one of these we have a density increment. Now apply the usual density increment argument and we are done.

It's not clear whether this argument can be adapted to give the corners theorem, let alone the  $k=3$  case of DHJ, but perhaps it should be looked at.



98. [gowers](#) Says:

[February 3, 2009 at 10:55 pm](#) | [Reply](#)

87. Thread title: obstructions to uniformity. Reply to 82.

Boris, I was not sure it was a true generalization either, and I'm still not sure. Let me attempt to come up with a set where it might have a chance of being one.

Let us consider two functions defined on the set of all sequences in  $[3]^n$ . First of all, choose a largish  $m$  and a subset  $P$  of positive density in  $\mathbb{Z}_m^2$  with the following properties: (i)  $P + P - P$  is not much larger than  $P$ , and quite a bit smaller than  $\mathbb{Z}_m^2$ ; (ii) for every  $x$  there are about the same number of  $y$  such that  $(x, y) \in P$ , and vice versa.

Now let's choose some fairly random sequences of weights  $(r_1, \dots, r_n)$  and  $(s_1, \dots, s_n)$  and let  $A$  be the set of all sequences such that  $(\sum r_i x_i, \sum s_i x_i) \in P \pmod m$ .

OK that almost certainly doesn't work, for the reason you say: there will be some character that correlates in a big way with  $P$ , and that will give us  $u$  and  $v$  such that  $\sum (ur_i + vs_i)x_i$  has a tendency to lie in an unusual subset of  $\mathbb{Z}_m$ .

I'm becoming convinced that it isn't a generalization, but I can't quite rule out that there might be some further things



one could do.

99.  [gowers](#) Says:

[February 3, 2009 at 11:18 pm](#) | [Reply](#)

88. Thread title: is triangle removal still necessary?

This is a response to Terry's comment 77. Obviously what you ask is an important question, but I'd like to understand more precisely what you mean by it.

To begin with, let me distinguish carefully between two interpretations of the question. One would be "Is it possible to do without any analogue of the triangle removal lemma from this point on?" and another would be "Is it unnatural to try to use the triangle removal lemma from this point on?"

The background to the question is that if a positive proportion of the slices (for a definition of this word, see 62) is rich, then by the corners theorem one can find a triple of rich slices of the form  $\Gamma_{a+r,b,c}, \Gamma_{a,b+r,c}, \Gamma_{a,b,c+r}$ . So one might take the view that you use the triangle removal lemma (via the corners theorem) to get yourself to a triple of rich slices and then let other methods take over from there.

Now we haven't actually said what we mean by "rich". It has to mean more than "intersects densely with  $A$ " since it is easy to find examples where  $A$  intersects three such slices densely but there is no combinatorial line. (See e.g. comment 20.) And it seems unlikely that there is a definition of "rich" such that (i) every dense set  $A$  intersects a dense set of slices richly and (ii) given any triangle of slices intersected richly by  $A$  you must get a combinatorial line from that triangle of slices.

At this point it's worth going back and reading Terry's comment 40. The idea there is that if you don't get a dense set of rich slices, then perhaps you can pass to substructures (e.g. by fixing a few coordinates) where something like density or energy improves and try again. This would be a sort of mixed argument: first use density/energy-increase arguments to get to a situation where lots of slices are rich (which I see as some kind of quasirandomness property, but perhaps I shouldn't), then apply the corners theorem to get a triple of rich slices, and finally apply some rich-slices counting lemma to get a combinatorial line. (I'm not at all sure that that's what Terry had in mind, and would be interested to know.)

Anyhow, perhaps that suggests that triangle removal could be used just to get a triple of good slices, and other arguments could be used to do the rest. In my next comment I'll discuss the other interpretation of Terry's question.

100.  [gowers](#) Says:

[February 3, 2009 at 11:53 pm](#) | [Reply](#)

89. Thread title: is triangle removal still necessary?

Is it clearly wrong to try to prove the whole theorem via a triangle-removal lemma? I would present the reasoning in the main post slightly differently. Admittedly it starts out in the way that Terry describes at the beginning of his comment 77. However, I think that the argument gains considerably in force when you add the following fact.

4. It is possible to define a tripartite graph in such a way that nondegenerate triangles correspond to combinatorial lines in  $A$ . Moreover, the construction naturally generalizes the tripartite graph that you construct in order to deduce the corners theorem from the triangle removal lemma.

Here's a way one might justify this. For a long time it was known that the triangle removal lemma implied Roth's theorem, but until Jozsef came along, nobody had observed that it would also do corners. Before that one might have reasoned as follows.

Let us label the lines of slope 1 in  $[n]^2$  according to the value of  $x - y$ . If we have a corner, then the labels of the lines that it lies in must form an arithmetic progression of length 3. And indeed, since our dense subset of  $[n]^2$  might even be a union of lines of slope 1, we see that the corners theorem implies Roth's theorem. So perhaps we should use Roth's theorem to find a rich triple of diagonals and then use other arguments to obtain a corner.

Is this a fair analogy or does it break down somewhere?

Before giving up on triangle removal, I would want answers to the following questions.

(i) Is it possible to formulate a triangle-removal statement that isn't obviously false and that implies that the tripartite graph discussed in 67 contains lots of triangles (because it contains lots of edge-disjoint degenerate ones)?

(ii) If so, is there the slightest chance of proving the statement by imitating the proof of the usual triangle-removal lemma? Or does the sparseness and nonquasirandomness of the bipartite graph where we join two sets if they are disjoint make a project of this kind hopeless?

I could perhaps add a third question, of a slightly different kind.

(iii) Is it the case that in order to prove that a regular triple contained many triangles, one would have to develop tools (connected with obstructions to uniformity) that could be used to give an easier proof that didn't go via triangle removal, regularity lemmas etc.?

I feel reasonably optimistic about (i) if one restricts to a triangular grid of slices. I'm much less sure about (ii) and (iii).

101.  *Sune Kristian Jakobsen* Says:

[February 4, 2009 at 12:02 am](#) | [Reply](#)

90:  $c_n$  for small  $n$  (Re to 78 83 and 84)

$c_4 = 52$ :

In 2 dimensions there are only 4 such set with 6 elements:  $A_i = \{(x, y) \in \mathbb{Z}_3^2 : x + y \neq -i \pmod{3}\}$  for  $i=1,2,3$  and  $\{(x, y) \in \mathbb{Z}_3^2 : x \neq y\}$ .

In 3 dimensions there is only one such set  $S$  with 18 elements. Proof: Let  $S$  be such a set. Now let  $S_i = \{(x, y) \in \mathbb{Z}_3^2 : (x, y, i) \in S\}$ , for  $i=1,2,3$ . Now each  $S_i$  has to have 6 elements and be of the above form, and the intersection of the  $S_i$ 's has to be empty. Thus the  $S_i$ 's must be a permutation of the  $A_i$ 's. The only permutation, that does not give any combinatorial lines is  $S_i = A_i$  for  $i=1,2,3$ .


Now assume that  $c_4 \geq 53$  and let  $S$  be a set  $|S| \geq 53$  in 4 dimensions and with no combinatorial lines. Again let  $S_i = \{(x, y, z) \in \mathbb{Z}_3^3 : (x, y, z, i) \in S\}$ , for  $i=1,2,3$ . Now two of the  $S_i$ 's has to have 18 elements and thus be of the above form. By if two of them are identical, the third can at most have  $27-18=9$  elements, and  $S$  at most  $2 * 18 + 9 = 45$ . Contradiction.

102.  *gowers* Says:

[February 4, 2009 at 12:19 am](#) | [Reply](#)

91. Thread title: obstructions to uniformity.

I don't know whether this is genuinely different, but in the Hales-Jewett theorem one doesn't actually care that the set  $[3]$  is  $\{1, 2, 3\}$ . So perhaps we could take three other numbers, such as  $\{4, 7, 9\}$ , and take the set of all  $x$  such that  $\sum r_i x_i$  belongs to  $E \pmod{m}$ . I don't immediately see how to get this from Boris's construction, at least as I presented it in 75.

103.  *Ryan O'Donnell* Says:  
[February 4, 2009 at 12:44 am](#) | [Reply](#)

92. Thread title: obstructions to uniformity.

Let me try to sketch a another obstruction to strong uniformity. Hopefully I have understood the notion of strong non-uniformity correctly! For the remainder of this comment, let  $\log$  denote  $\log_3$ .

Partition the coordinates  $[n]$  into disjoint “blocks” of size roughly  $\log n - \log \log n$ . For a uniform random string  $x \in [3]^n$ , let  $\$F\$$  denote the event that  $x$  contains a block which is all-2’s. The probability a particular block is all-2’s is  $\log n/n$ ; hence the probability it is not-all-2’s is  $1 - \log n/n$ . Since there are also about  $n/\log n$  blocks, the probability that all blocks are not-all-2’s is basically  $1/e$ . So the probability of  $\$F\$$  is basically  $1 - 1/e$ . Call it  $.63$ .

So let  $A$  be the set of all strings which have an all-2’s block, which has density about  $.63$ .


On the other hand, suppose we have a combinatorial line  $(x, y, z)$  with  $x \in A$ . Then that block of 2’s stays fixed in  $y$  and  $z$ . So take  $B$  to be the set of strings which do \*not\* have a block of all-2’s. This is also plenty dense, having density about  $.37$ . But then we couldn’t have  $y, z \in B$ .

You might object that this only worked because we were insisting that it was  $x$  (i.e., the wildcard=1 point) that was in  $A$ , but this can be circumvented by making  $A$  equal the set of strings with a block of all-2’s \*and\* a block of all-1’s. (These events are practically disjoint, so this  $A$  would still have density about  $.63^2$ .)

This example does seem to me different from Boris and Tim’s examples. At first it looks similar because you are counting that the number of 2’s in a particular set of coordinates is equal to a particular number. But then things differ because you take an “or” across all blocks. You can make it even more different by taking the “block-of-all-2’s \*and\* block-of-all-1’s” example. Or you could instead let  $A$  be the set of strings with a “run” of 2’s of length about  $\log n$  somewhere within them.


Does this example work?

If so, I think my intuition agrees with Tim’s (expressed at the end of #75) that there may be “too many” obstructions...

104.  *jozsef* Says:  
[February 4, 2009 at 12:46 am](#) | [Reply](#)

93. Thread title: an alternative problem.

In 60. Tim suggested a problem where we consider certain pairs of sets. “A sensible question would be some density-type condition on  $A$  that would guarantee a triple of the given kind.” One candidate might be the following: Given a subset  $S$  of the  $2^n$  subsets of an  $n$ -element set, then two elements of  $S$  form a pair if their symmetric difference is also in  $S$ . I think, but I can’t prove it yet, that if  $S$  is dense,  $|S| \geq c2^n$  then the number of such pairs (triples) is at least  $c'2^{2n}$ . If so, then there is a direct application of the triangle removal lemma, however I have to think about the interpretation of the result.

105.  *Ryan O'Donnell* Says:  
[February 4, 2009 at 12:59 am](#) | [Reply](#)

94. Obstructions to strong uniformity.

Can’t you also do other goofy things like take any example  $A$  which is strongly-non-uniform and, say, tack 23 onto the end of each string (changing  $n$  to  $n + 2$ )? This hurts the density by only  $1/9$ . Now take the dense  $B$  which failed with

the original  $\mathcal{A}$  and tack, say, 12 on the end of all of its strings. This also only loses density  $1/9$ . But if  $\mathcal{A}$  and  $\mathcal{B}$  were a bad example before, they are still bad now. And  $\mathcal{A}$  is no longer of the weighted-sums-modulo- $m$  type, even if it was before.

106.  [gowers](#) Says:

[February 4, 2009 at 1:03 am](#) | [Reply](#)

95. Thread title: Sperner

I want to make a start on the question I eventually arrived at in 80, which I'll repeat here and make more precise. Let  $m$  be slightly smaller than  $n/2$ , write  $k$  for  $n - 2m$  and let  $\mathcal{A}$  and  $\mathcal{B}$  be two subsets of  $\binom{[n]}{m}$ . Is it the case that for every constant  $a > 0$  there exists a constant  $c > 0$  such that if the number of disjoint pairs  $(U, V)$  with  $U \in \mathcal{A}$  and  $V \in \mathcal{B}$  is at most  $c \binom{n}{m, m, k}$  then it is possible to remove at most  $a \binom{n}{m}$  sets from  $\mathcal{A} \cup \mathcal{B}$  and end up with two set systems  $\mathcal{A}'$  and  $\mathcal{B}'$  such that every  $U \in \mathcal{A}'$  intersects every  $V \in \mathcal{B}'$ ?

Actually, now that I've formulated the question (though I'm still not certain that it's a sensible one) I find that I have no idea where to start. All I have to say is that Boris's example in 44 is worth bearing in mind. But in a curious way that makes me feel the problem is worth pursuing. It seems that to prove Sperner via a generalization of the trivial pair removal lemma, one has to prove this non-trivial pair removal lemma. Working out how to do that therefore introduces an extra ingredient. Once one has that extra ingredient, one can see whether it generalizes to provide the extra ingredient needed to get from the usual triangle removal lemma to a more difficult Spernerish triangle removal lemma.

107.  [gowers](#) Says:

[February 4, 2009 at 1:13 am](#) | [Reply](#)

96. Thread title: obstructions to strong uniformity.

Ryan, I think we need to formulate more carefully what it means to have a new obstruction. I would want not to count the example you've given because it's contained a bigger example where  $\mathcal{A}$  and  $\mathcal{B}$  consist of all strings (before you do the tacking-on process). And that bigger example is covered by what I called strong obstruction 1 in comment 71.

108.  [gowers](#) Says:

[February 4, 2009 at 1:52 am](#) | [Reply](#)

97. Thread title: Sperner

Actually, I have a small idea where to start (see 95). If we replace sets in  $\mathcal{B}$  by their complements, then we obtain a set system that has to have a lower shadow that is very non-uniform. That is, if you count how many times a set is contained in the complement of a set in  $\mathcal{B}$  then the resulting function mustn't be close to constant. So it looks as though we do indeed have to understand obstructions to uniformity in Kruskal Katona.

109.  [Ryan O'Donnell](#) Says:

[February 4, 2009 at 3:10 am](#) | [Reply](#)

98. Obstructions.


Hi Tim, are you referring to the example from 92 or 94? I don't see how the #92 one fits into any of the example obstructions mentioned previously.

110.  [Terence Tao](#) Says:

[February 4, 2009 at 3:35 am](#) | [Reply](#)

99. Thread title: is triangle removal still necessary?


Tim, I guess my question was intended in the weak sense as in 88 ("can we now do without triangle removal?") as opposed to the strong sense as in 89 ("should we keep trying to mimic triangle removal?"). Basically, I think we should keep pursuing the triangle removal tack, but bear in mind the option to also utilise many other tools (e.g. Kruskal-Katona, or some ideas from Szemerédi's proofs of Roth's theorem, about which I will say more in my next comment), and to also entertain the possibility that triangle removal might in fact only play a minor role in the rest of the argument.

111.  [Terence Tao](#) Says:  
[February 4, 2009 at 3:43 am](#) | [Reply](#)

100! (do I get a prize?) Thread title: Density incrementation

One tool that seems to be underexploited so far is that of density incrementation. In the standard proof of Roth's theorem (as well as Szemerédi's proof), we may assume without loss of generality that the set  $A$  does not have significantly increased density on any reasonably long subprogression, since otherwise we could pass to that subprogression and apply some sort of induction hypothesis. We can do the same here: to prove DHJ for some set  $A \subset [3]^n$  with density  $\delta$ , we may assume without loss of generality that every slice of  $A$  of any reasonable size ( $m$ -dimensional, where  $m$  grows slowly with  $n$ ) has density at most  $\delta + o(1)$  (where I will be vague about exactly what  $o(1)$  means). By Markov's inequality, this also implies that almost all (i.e.  $1-o(1)$ ) of all such slices will have density at least  $\delta - o(1)$  as well. It is not clear to me how one can exploit such a fact, but it is something worth keeping in mind. (Admittedly, this structure is not used in the graph-theoretic proofs of Roth or corners, but as I said in my previous post, we are not bound to slavishly follow the graph-theoretic model.)

In a somewhat similar spirit, we can use the colouring Hales-Jewett theorem to stabilise various low-complexity quantities. Indeed, if we have some quantity  $X$  that depends on a medium number  $m$  of the coordinates of  $[3]^n$ , and takes a bounded number  $k$  of values, then (if  $m$  is sufficiently large depending on  $d$  and  $k$ ), we can reduce the  $m$  coordinates to  $d$ , passing to a slice, and force  $X$  to now be constant on these  $d$  coordinates. (I already used this argument in 85.) Again, I don't see an immediate way to utilise this freedom, but if we are encountering a problem that the "richness" of various slices varies too much with the slice, this is presumably the thing to do to fix that. And there is also the Carlson-Simpson theorem to hold in reserve to soup this up even further.

112.  [Terence Tao](#) Says:  
[February 4, 2009 at 3:52 am](#) | [Reply](#)


101.  $c_n$  for small  $n$ .

@Sune.90: Excellent... we now have the first five elements of  $c_n$ , namely 1, 2, 6, 18, 52. Consulting the OEIS,

<http://www.research.att.com/~njas/sequences/?q=1%2C2%2C6%2C18%2C52&language=english&go=Search>

there are a number of sequences here, but none that look likely to be a fit, especially given the bounds  $145 \leq c_5 \leq 3c_4 = 156$  we have (see Tyler.83).

Given that  $c_5$  does not look too close to being computable, perhaps it is time to submit what we have to the OEIS?

113.  [Ryan O'Donnell](#) Says:  
[February 4, 2009 at 3:58 am](#) | [Reply](#)

100. Kruskal-Katona obstructions.


Even in the case of Kruskal-Katona, I'm not 100% what the right question is. Let me work off Boris's #44 formulation of the problem.

Say the density of  $A$  is  $\mu_A$  and the density of  $B$  is  $\mu_B$ . When Boris asks, when is the number of pairs  $(a, b) \in A \times B$  with  $a \subset b$  close to expected, I assume by "expected" he means "what the number would be if  $A$  and  $B$  were random sets of the given densities". (Correct me if I've mistaken you here.)

Introduce a uniformly random pair  $(x, y) \in \binom{[n]}{n/2} \times \binom{[n]}{n/2+1}$  subject to  $x \subset y$ . Let  $L$  be the event that both  $x \in A$  and  $y \in B$ . Note that the marginal on both  $x$  and  $y$  is uniform. Hence if  $A$  and  $B$  were both random sets of the given densities, we'd expect the probability of  $L$  to be  $\mu_A \mu_B$ .

Now Boris gave two very different examples where  $\mu_A \approx \mu_B \approx 1/2$  and yet  $\Pr[L] = 0$ , much less than  $\mu_A \mu_B$  of course.

On the other hand, there are plenty of examples where  $\mu_A \approx \mu_B$  is a constant and yet  $\Pr[L] \approx \mu_A$ , much \*greater\* than  $\mu_A \mu_B$ . To achieve this, fix  $B$  to be the upper shadow of  $A$ . Now we just need to take  $A$  to be any constant-density set whose upper shadow has about the same size as  $A$ . And there is a world of such sets; too many, I think, to have a meaningful characterization by obstructions.

114.  [Terence Tao](#) Says:  
[February 4, 2009 at 4:05 am](#) | [Reply](#)

102. The HJ cube lemma.


As I mentioned in 86, there is a proof of Roth's theorem by Szemerédi which relies on the Szemerédi cube lemma, which basically asserts that any dense subset of  $[n]$  contains a large cube

$Q = \{a_0 + \omega_1 r_1 + \dots + \omega_d r_d : \omega_1, \dots, \omega_d \in \{0, 1\}\}$  (in practice one can take  $d$  to be something like  $\log \log n$ ).

There is a similar lemma for  $k=2$  Hales-Jewett: if  $n$  is sufficiently large depending on  $\delta$  and  $d$ , then any subset  $A$  of  $[2]^n$  of density at least  $\delta$  contains a  $d$ -dimensional cube, which basically is the same thing as a combinatorial line except that one now has  $d$  wildcards instead of 1.

The proof proceeds by induction on  $d$  and using the  $k=2$  DHJ (i.e. Sperner). The  $d=1$  case is of course exactly  $k=2$  DHJ. Now if  $d > 1$  and the claim is already proven for  $d-1$ , then we can foliate  $[3]^n$  into  $m$ -dimensional slices for some intermediate  $m$ . Many of these slices will be rich and thus contain a  $d-1$ -dimensional cube, by induction hypothesis. The placement of this cube can vary by slice, but the number of possible cubes is bounded by some constant depending on  $m$ , so by pigeonhole there is a positive density (depending on  $m$ ) of slices for which the cube position is the same. Now apply Sperner ( $k=2$  DHJ) one more time to get the result.

One can of course embed  $[2]^n$  into  $[3]^N$  in a variety of ways, and so this cube lemma is available for use in the  $k=3$  problem. I don't know yet whether Endre's proof (see 86) can carry over in this setting, but it might be worth taking a look at. Note that the colouring version of this cube lemma is used in the standard proof of  $k=3$  colouring Hales-Jewett.

115.  [gowers](#) Says:  
[February 4, 2009 at 9:38 am](#) | [Reply](#)

103. Obstructions to uniformity.

Ryan, re your comment 98, I'm sorry but for some reason I failed to notice your very interesting comment 92. In my comment 96 I was referring to what you said in 94.

Your example in 92 strikes me as a definite advance in the understanding of the problem. Let me try to explain why. One of the difficulties of the density Hales-Jewett theorem is its "combinatorial", as opposed to "group-theoretic",



nature. That is, we do not have a group structure on  $[3]^n$  (or at least, not one that seems to be relevant), and the notion of a combinatorial line is not one that can be expressed in purely group-theoretic terms.

On the other hand, the kinds of obstructions that Boris and I were coming up with did have a group-theoretic flavour to them, even if they weren't exactly group theoretic. I couldn't help thinking that it was too optimistic to hope that all obstructions were of this kind. And now you have demonstrated (assuming your example checks out, which it feels to me as though it does) that there are indeed some more purely combinatorial obstructions.

A quick thought at this stage. For each obstruction it is well worth thinking about the following question: is there an obvious way of restricting  $[3]^n$  so as to obtain a substructure where  $\mathcal{A}$  is denser? In your case I was about to say that I couldn't see one, but I think I can: we could do something like restricting to a random combinatorial subspace of the following kind. Take your partition of  $[n]$  into blocks and for each block choose a random subset (according to some sensible probability distribution — not sure what that is) to be a set of coordinates that vary simultaneously. But insist that at least once you ... the more I write the more this feels completely wrong. So instead let me ask: can one get a density increase on a combinatorial subspace for Ryan's set  $\mathcal{A}$  in 92?

116.  [jozsef](#) Says:

[February 4, 2009 at 9:51 am](#) | [Reply](#)

104. Thread title: Sperner

Re: Tim's question in 95. I just want to clarify, that the conjecture (?) is that in the Kneser graph  $K_{n,m}$  every induced bipartite graph with at most  $c \binom{n}{m,m,k}$  edges has a vertex cover of size at most  $a \binom{n}{m}$ . Or, equivalently, the largest matching is at most  $a \binom{n}{m}$ . It seems to me that the small number of edges in the induced graph implies that the number of vertices is small, but I don't remember what is the spectral gap in  $K_{n,m}$ . Anyways, if  $K_{n,m}$  is quasirandom then I don't understand the question, but it is almost 1 am, so I will review it tomorrow...

117.  [jozsef](#) Says:

[February 4, 2009 at 9:54 am](#) | [Reply](#)

Good morning Tim! Then mine is 104. ... *[Now changed]*

118.  [gowers](#) Says:

[February 4, 2009 at 10:00 am](#) | [Reply](#)

105. Obstructions to uniformity.

For some reason in 103 I was blind to the obvious example: take a subspace where all sequences are constant on Ryan's blocks, and it's actually contained in  $\mathcal{A}$ .

119.  [gowers](#) Says:

[February 4, 2009 at 11:00 am](#) | [Reply](#)

106. Sperner.

In response to Jozsef comment 104, I think your rephrasing of my question is indeed the same question (but I'm slightly shaky on the terminology you use). Just to check: the Kneser graph  $K_{n,m}$  is the graph consisting of all subsets of  $[n]$  of size  $m$ , with two of them joined if and only if they are disjoint. An induced bipartite graph just means that you pick two sets  $\mathcal{A}$  and  $\mathcal{B}$  of vertices in this graph and put in all the edges that you have from the graph: so you have two systems of sets of size  $m$  and join a set in  $\mathcal{A}$  to a set in  $\mathcal{B}$  if and only if they are disjoint. Finally, a vertex cover means a collection of vertices such that every edge contains a vertex in that collection. So if we have a small vertex cover

then we can remove it and end up with no edges.

It's not the case that a small number of edges implies that the vertex sets are small. For example,  $\mathcal{A}$  and  $\mathcal{B}$  could both be the set of all sets of size  $m$  that contain the element 1. Then there are no edges between  $\mathcal{A}$  and  $\mathcal{B}$ , but they both have density about  $1/2$ . Of course, this example doesn't have much bearing on the problem I asked.

120.  [gowers](#) Says:

[February 4, 2009 at 11:17 am](#) | [Reply](#)

*[This comment is deliberately not numbered.]*

If you'll forgive a metacomment, at some point I think we should consider trying to summarize everything we've learned so far (interesting examples, questions we haven't yet answered, data about small  $n$ , ideas for general approaches, etc.), write a new post, and start all over again, this time from camp 1 rather than base camp, so to speak. It would have the small advantage that people wouldn't have to scroll through over a hundred comments (my computer now takes quite a long time to load all the mathematical gifs — if that's the right word). Note that I'm not suggesting splitting up the threads: I've rather enjoyed having them all jumbled together, and I think this suggestion would keep the current atmosphere but keep things manageable. Any views? (I'm thinking of introducing a poll facility to this blog so that people can express opinions about this kind of question without having to make actual comments.)

*[To my amazement, I seem to have managed to produce a poll. So if you want to register an opinion, you have the chance to do so. It appears as a new and strange-looking post on the blog.]*

121.  [gowers](#) Says:

[February 4, 2009 at 11:31 am](#) | [Reply](#)

107. Thread title: big picture.

Terry has already described some possible ways that an eventual proof might go (see e.g. #77 and #100). In the next comment, I'll have a go at some proof sketches (not very detailed because I don't know the details!). Each one would ideally be the beginning of a top-down process that ends with a fully detailed proof. But the question in each case is whether the strategy is remotely realistic.

Strategy 1. Density incrementation.

This strategy is to model a proof on Roth's proof of Roth's theorem. The idea is to keep thinking about obstructions to uniformity until we think we've found all of them. At that point we try to prove the following: either a set  $A$  correlates with one of our obstructions to uniformity, in which case one can show that there is a combinatorial subspace inside which  $A$  has increased density; or  $A$  does not correlate with one of our obstructions to uniformity, in which case we can prove that  $A$  contains the expected number of combinatorial lines.

Two remarks about the strategy. First, as we have noted, the "expected" number of combinatorial lines is a troublesome concept. But it ought to be possible to deal with that by restricting to a triangular grid of slices near the centre. Secondly, a density-increment proof of density Hales-Jewett would presumably yield a density-increment proof of the corners theorem. Now such a proof has been obtained by Shkredov, but it is considerably more complicated than the density-increment proof of Roth's theorem. So we shouldn't expect this strategy to be easy. (Actually, this has given me an idea about obstructions to uniformity in DHJ. I'll explore it in a moment.)

Strategy 2. This is roughly the original strategy I suggested. We have a tripartite graph where we join two sets if and only if they are disjoint, and we try to prove a regularity lemma for subgraphs of this graph (or perhaps portions of this graph where we restrict the cardinalities of sets so that they become "typical"). This would be a relative regularity lemma.

Two remarks about this strategy. First, I think it ought to be easy to make at least some progress, either negative or positive, by trying to produce such a regularity lemma. Either it will rapidly become clear that the task is hopeless, or we'll manage it, or we'll generate a bunch of interesting questions that we can't yet answer. Secondly, in a subtle way the regularity approach to corners allows one not to understand too well what the obstructions to uniformity are for that problem: it just automatically seeks them out and deals with them. Might something like that happen for DHJ?

122.  [gowers](#) Says:

[February 4, 2009 at 11:56 am](#) | [Reply](#)

108. Obstructions to uniformity.

No idea whether this is going anywhere, but in Shkredov's proof of the corners theorem, he has to deal with examples of sets of the form  $X \times Y$ , where both  $X$  and  $Y$  are subsets of  $[n]$ . These sets have the feature that if they contain the two points  $(x, y + d)$  and  $(x + d, y)$  then they automatically contain the point  $(x, y)$ . It ought to be possible to exploit this to obtain an obstruction to uniformity in DHJ, but I don't yet know whether it will be a new one.

So let's choose an arbitrary pair of (reasonably dense and codense near  $n/3$ ) subsets  $X$  and  $Y$  of  $[n]$ , take  $B$  to be the set of all sequences such that the number of 1s belongs to  $X$  and the number of 2s belongs to  $Y$ , and take  $A$  to be the complement of  $B$ . Now let  $(x, y, z)$  be a combinatorial line (written in 1,2,3 order). The point  $x$  has the same number of 2s as  $z$ , and  $y$  has the same number of 1s as  $z$ . So if  $x$  and  $y$  belong to  $B$ , then  $z$  automatically belongs to  $B$ , so it can't belong to  $A$ .

This is open to the same objection that Ryan talked about in 92, namely that we're insisting that it is  $z$  that belongs to  $A$  rather than  $x$  or  $y$ , but the basic idea can be used to produce an example that gets round this objection. We just have to play the same game with 2s and 3s, and the same game again with 1s and 3s, and then intersect the three resulting sets. If you choose them reasonably randomly then you'll still get a dense  $A$  and a dense  $B$ .

(Just to clarify what I've just said, let's call the sets I constructed  $A_3$  and  $B_3$ , since 3 was the preferred value. Now you repeat the construction I gave twice over to produce  $A_1$  and  $B_1$  and  $A_2$  and  $B_2$ . Finally, you set  $A$  to be  $A_1 \cap A_2 \cap A_3$  and you set  $B$  to be  $B_1 \cap B_2 \cap B_3$ . Provided the sets  $X$  and  $Y$  are chosen reasonably randomly in each case, the sets  $A$  and  $B$  will be reasonably dense.)

This feels like a genuinely different obstruction because it is very nonlinear. In fact, I'm basically pretty sure it's different.

So going briefly back to the big picture, one might argue that the problem with strategy 1 is that it is already difficult to carry out for corners, and looks as though it could be very much more difficult still here. So if it were just a choice between strategies 1 and 2, there would be a good reason for continuing to pursue 2 (not that I feel like giving up on 1 just yet).

123.  [gowers](#) Says:

[February 4, 2009 at 1:20 pm](#) | [Reply](#)

109. Thread title: counting lemma.

I want to make a tentative appeal to procedural rule 6, the one that says that if you think you could work something out with the help of some private calculations, then you first see whether others think that would be a good thing to do. So here's what I think I could do reasonably easily privately and not so easily if I have to type everything in here.

It's clear that if any kind of regularity approach is going to work, then there needs to be a lemma that says that if you've got three suitably quasirandom subgraphs of the is-disjoint-from graph arranged in a triple, then you must have lots of triangles. But it's not clear with "suitably quasirandom" means. A natural way to attempt to get a handle on

this, if you've ever been involved with counting lemmas in sparse random settings, is to try to use the Cauchy-Schwarz inequality to prove that the number of triangles is close to what you expect unless something bad happens. And that something bad then turns into your definition of "is not quasirandom". I would like permission to go away, attempt some Cauchy-Schwarz calculations, and come back and report on the results.

124.  [gowers](#) Says:

[February 4, 2009 at 1:29 pm](#) | [Reply](#)

110. Thread title: regularity lemma.

Here, by the way, is the outline of how one might try to prove a regularity lemma. Almost certainly there will be an easy argument that says that this naive strategy has no chance of working.

Let  $G_0$  be the bipartite graph where you join two sets (fixed to have size around  $n/3$ ) if and only if they are disjoint. Let  $G$  be a dense subgraph of this graph. We want to partition the vertices into sets where  $G$  looks like a random subgraph of  $G_0$ . So why not argue as follows? Given partitions of the vertex sets, if there are lots of induced bipartite subgraphs where  $G$  doesn't look random in  $G_0$ , then you can refine the partitions in such a way that the ratio of the energy, or mean square density, of  $G$  to that of  $G_0$  goes up. The idea would be that although the mean square density of  $G_0$  could well be going up all the time (since  $G_0$  is not a quasirandom graph), the mean square density of  $G$  would be going up faster so the iteration would have to terminate.

As I say, this is probably a hopeless strategy, but it gives an idea of the kind of thing that I was hoping one might be able to do, and there might be ways of modifying it that made it more promising.

125.  [Gil Kalai](#) Says:

[February 4, 2009 at 3:02 pm](#) | [Reply](#)

A little thought about the issue Tim raised in 31. (Sorry if it duplicates some later comment or link.)

The question Tim ask was what conditions (Fourier or different) will ensure semirandomness for Sperner and for our combinatorial line question for  $k=3$ , namely, a condition for a set to contain approximately "the right number" of combinatorial lines.

One possibility to consider is to ask for density on a set  $X$  of solutions for arbitrary equation of the form  $a_1x_1 + a_2x_2 + \dots + a_nx_n \in T$ . ( $T$  can be any set but we can just think about a single number.) If the intersection of  $A$  and  $X$  is equal to its expected value up to a small additive error, for every  $X$ , does it suffices to imply that  $A$  has the right number of pairs  $x \subset y$ ? (for the Sperner case) or even the right number of combinatorial lines for  $k=3$ ?

Are you aware of counterexamples?

Talking about general solutions of linear equations may help also for cases where Fourier analysis gives a result but we would like to have a better one. (upper bounds on cup set? Influences of sets...) A difficulty with such a suggestion is that you need to argue recursively about dense sets inside sets which are the solution of a system of equations of the above kinds and maybe even more complicated. Also I do not see an argument that will show that if the number of solutions of linear equations is right then the number of combinatorial lines is right (for  $k \leq 3$ ).

Another possibility (or perhaps something forced by the above possibility if you try to prove things) is to consider for  $X$  sets which are described by even more general arithmetic or other type of circuits. Such sets may also be useful for lower bound examples; I find it yet hard to believe that the examples we know (for the various questions) are in the right ball park as reality.

126.  [Ryan O'Donnell](#) Says:

[February 4, 2009 at 3:46 pm](#) | [Reply](#)

112. Obstructions.

Tim, in your #105, I don't quite see what you're saying... And does it hold for the set  $A$  of all strings which contain a "run" of at least  $\log n - \log \log n$  2's somewhere within them? (This example has no fixed "blocks".)



127. *Ryan O'Donnell* Says:

[February 4, 2009 at 3:52 pm](#) | [Reply](#)

113. Obstructions: These examples from #92 were really just based on an example ("Tribes") from the  $[2]^n$  scenario. Like Jozsef#104, I haven't wrapped my head around the counting/normalization/distribution used in Tim's #95; this is why in my #100 I was writing about Boris's #44 setup instead.

I feel that before thinking too much about obstructions to combinatorial-line-uniformity in  $[3]^n$ , we should try to clarify things in the simpler scenario of obstructions-to-Sperner in  $[2]^n$ .

In #100, I tried to get across that I can't even think what this should mean.



128. *Ryan O'Donnell* Says:

[February 4, 2009 at 3:53 pm](#) | [Reply](#)

114. Sperner/Kruskal-Katona.

Actually, I'm not even sure "obstructions" is the right way to think about Sperner. I think that what Boris said in #58 — that Sperner's own proof by "pushing shadows around" (I like this phrase, Boris!) — may already be the "alternate combinatorial proof of Sperner" that Terry proposed to look for way back in #4!

Can we prove DHJ by "pushing shadows around"? 😊



129. *Jason Dyer* Says:

[February 4, 2009 at 4:40 pm](#) | [Reply](#)

@Tyler, Sune, Terry (on bounds)

Nice! I hope to have some code working by the end of the week;  $c_5$  ought to be within brute force range, but I'm not sure past that.

I've thought about the flow optimization algorithm, and I'm not convinced that will produce an improvement because of the symmetry of the original arrangement. I do still think there's some CS trick that will help knock this down to size and get exact values for  $c_6$  and beyond. For the moment I'm just going to Monte Carlo it so we can get some perspective.



130. *gowers* Says:

[February 4, 2009 at 5:21 pm](#) | [Reply](#)

115. Obstructions.

Ryan, 112, I won't bother trying to explain what I was saying because there's an even simpler example to show that your set yields a density increase on a subspace: you just fix a run of 1s and let everything else vary arbitrarily. That does both the blocks case and the fixed-length-of-run case.

131.  [gowers](#) Says:

[February 4, 2009 at 6:00 pm](#) | [Reply](#)

116. Big picture, density incrementation, obstructions

I've just realized that the example in 108 shows that strategy 1 in 107 doesn't work as stated, or at least not completely straightforwardly. For the benefit of people unfamiliar with Shkredov's work on corners, let me explain this in some detail.

Suppose we try to prove the corners result using a density-increment strategy. That is, we try to prove that if a subset of  $[n]^2$  doesn't have roughly the expected number of corners then there's a substructure inside which  $A$  is noticeably denser. What is the obvious substructure? Well, we want something that will support corners, so the obvious thing to take is a Cartesian product of two arithmetic progressions with the same common difference.

However, this doesn't work. Suppose we take  $A$  to be a Cartesian product of two random sets  $X$  and  $Y$ . Then there will be too many corners in  $A$  (because you just need  $(x, y + d)$  and  $(x, y + d)$  and you get  $(x, y)$  for free). But if  $P$  and  $Q$  are arithmetic progressions, then  $X \cap P$  and  $Y \cap Q$  will look like random subsets of  $P$  and  $Q$  so we won't get any density increase. (But see a qualification to this remark later on.)

This has an analogue in the Hales-Jewett setting. Suppose we take the example in 108, and we use random sets  $X$  and  $Y$  to define it. Then there won't be a combinatorial subspace contained in the resulting set  $A$  (or even intersecting it unusually densely) because the set of sizes of the 1-set of a vector in a combinatorial subspace will be a very special kind of set: the set of sums of some positive integers  $n_1, \dots, n_k$ , and a random set will intersect such sets randomly.

Both these conclusions are, however, false if we allow ourselves to pass to much smaller substructures. For instance, in the first case we could use Szemerédi's theorem to get ourselves an arithmetic progression contained in  $X \cap (Y + r)$  for some  $r$ , and in the second we could use a cube lemma to get  $n_1, \dots, n_k$  such that all sums belong to  $X$ . (I'm using these statements so that I don't actually use the fact that  $X$  and  $Y$  are random, which in general we are not going to be given.)

132.  [Ryan O'Donnell](#) Says:

[February 4, 2009 at 6:17 pm](#) | [Reply](#)

117. Obstructions.


Tim, re #115 — thanks! I understand what you mean. I wonder if you could suggest the “quantitatives” to think about for density increment arguments.

Let's work with a different strongly nonuniform set. Let  $A$  be the set of strings where the plurality digit in the first  $n/3$  coordinates is 1, the plurality digit in the second  $n/3$  coordinates is 2, and the plurality digit in the third  $n/3$  coordinates is 3. This has density  $1/27$  and is strongly nonuniform. (The “counterexample”  $B$  is the set of strings where the plurality digit in the first  $n/3$  coordinates is \*not\* 1, the plurality digit in the second  $n/3$  coordinates is \*not\* 2, and the plurality digit in the third  $n/3$  coordinates is \*not\* 3. This has density  $8/27$  and forms no lines with  $A$ .)

But now it's hard to improved  $A$ 's density by restricting to a smallish-codimension combinatorial subspace. Your best bet is (I assume) to put, say, lots of 1's into the first  $n/3$  coordinates. But so long as you fix only  $o(\sqrt{n})$  1's, you've only improved  $A$ 's density by  $o(1)$ .

So I'm wondering what “quantitative aspects” one should hope for. Perhaps  $o(1)$  increase is enough, or  $\Omega(\sqrt{n})$  codimension is “not too much”.



133.  [Terence Tao](#) Says:  
[February 4, 2009 at 6:29 pm](#) | [Reply](#)

118. Szemerédi's proof of Roth's theorem

Good morning again... this topic is perhaps orthogonal to the others, but I think I begin to understand Szemerédi's proof of Roth's theorem (see 86) better. For instance, I realised that it has the curious property of proving Roth in  $\mathbb{Z}/N\mathbb{Z}$  but not in  $(\mathbb{Z}/3\mathbb{Z})^n$  (at least, not without some further modification). I also begin to see how the "sparsity" of the corners or DHJ problem is a serious difficulty extending Endre's argument to this setting. I don't think this approach is totally hopeless, though, and might be worth pursuing in parallel with the "mainstream" obstructions-to-uniformity/triangle removal approach.


Basically, Endre's proof is based on three observations:

1. If  $A$  is dense, then it contains a structured set  $Q$  (specifically, a medium-dimensional cube).
2. If  $A$  has no 3-APs, and contains a structured set  $Q$ , then the complement of  $A$  contains a large structured set, namely  $2 \cdot Q - A$ .
3. If the complement of  $A$  contains a large structured set (or equivalently, if  $A$  is "squeezed" into the complement of a large structured set), then there is some subprogression of  $A$  of increased density.

Combining 1, 2, 3 and the density increment argument one obtains Roth's theorem in  $\mathbb{Z}/N\mathbb{Z}$ .

If we try to prove Roth in  $(\mathbb{Z}/3\mathbb{Z})^n$  rather than  $\mathbb{Z}/N\mathbb{Z}$ , then 1 and 2 still work fine, but I can only show 3 using Fourier analysis, but of course Roth in  $(\mathbb{Z}/3\mathbb{Z})^n$  is trivial with Fourier analysis anyway. (Specifically, the question is whether one can find dense subsets  $A, B$  of  $[3]^d$  such that  $A$  is in the complement of  $B + [2]^d$ , where  $d$  is a slowly growing function of  $n$ , and for which  $A, B$  have no density increments on a subspace. Easy with Fourier analysis, and I don't see how to do it otherwise.) Somehow the problem is that  $(\mathbb{Z}/3\mathbb{Z})^n$  is not as "connected" as  $\mathbb{Z}/N\mathbb{Z}$ .


If we try to prove corners or DHJ, then 1 still works (see 102), but 2 now breaks down, basically because Euclid's axiom "two points determine a line" begins to fail horribly. Instead of  $A$  being squeezed into the complement of a large structured set, it seems that  $A$  is only squeezed into a complement of a small structured set, which is saying practically nothing. (For instance, in the DHJ problem, colouring HJ already tells us that the complement of  $A$  must contain arbitrarily large combinatorial subspaces (for  $n$  large enough), but this seems to be a mostly useless piece of information, although it does at least imply the weak DHJ in 85.) It may be that by localising to a well-chosen piece of  $[n]^2$  or  $[3]^n$  one can mitigate this problem, but I don't see how yet. And even after problem 2. is solved, one still has to grapple with 3., which as we saw was already problematic even in the model  $(\mathbb{Z}/3\mathbb{Z})^n$  case.

134.  [gowers](#) Says:  
[February 4, 2009 at 6:33 pm](#) | [Reply](#)

119. Obstructions.

Ryan, by "plurality digit" do you mean "digit for which there is a long run" or do you mean "digit that occurs most often"? From your second paragraph I think you must mean the second.

In fact, I see that you definitely mean the second, the idea being that you'd expect that digit to occur something like  $n/9 + c\sqrt{n}$  times, on average anyway. But I'll leave this comment here in case it helps anyone else.

135.  [Terence Tao](#) Says:  
[February 4, 2009 at 6:38 pm](#) | [Reply](#)


## 120. Density incrementation

Re: Ryan.117 (as opposed to Terry.117 😊 [now corrected -- Tim], I can tell you what type of density incrementation you need to get in order to get a contradiction that finishes the proof.

From the trivial bound  $c_{n+1} \leq 3c_n$ , we know that the sequence  $c_n/3^n$  decreases monotonically to a limit  $\alpha$ . DHJ is precisely the claim that  $\alpha = 0$ , so suppose for contradiction that  $\alpha > 0$ . Then we can write  $c_n = (\alpha + o_{n \rightarrow \infty}(1))3^n$ , where  $o_{n \rightarrow \infty}(1)$  denotes a quantity that goes to zero as  $n \rightarrow \infty$ . Note that the decay rate  $o_{n \rightarrow \infty}(1)$  here is completely ineffective.

Thus, for  $n$  large, we can find a set  $A \subset [3]^n$  of density  $\alpha + o_{n \rightarrow \infty}(1)$  with no combinatorial lines. In particular, on any  $m$ -dimensional subspace, the density of  $A$  cannot exceed  $c_m/3^m = \alpha + o_{m \rightarrow \infty}(1)$ . So, in order to get a contradiction, one needs to find a medium-size dimensional subspace on which one has a density increment that is bounded below uniformly in the dimension  $m$  of that subspace.

A useful fact, by the way, is that upper bounds on density of all subspaces imply lower bounds on density of `_random_` subspaces. Indeed, consider a random subspace  $V$  of  $[3]^n$  with two properties: (a) the dimension of  $V$  is always at least  $m$ , and (b) any two subspaces which are “parallel” in the sense that their wildcard set is exactly the same will be drawn with equal probability. Then I claim that with probability  $1 - o_{m \rightarrow \infty}(1)$ , the relative density of  $A$  in  $V$  will be  $\alpha + o_{m \rightarrow \infty}(1)$ . Indeed, we can condition to fix the wildcard set of  $V$ , and observe that the remaining subspaces partition  $[3]^n$ . The mean density of the subspaces is thus the global density  $\alpha + o_{n \rightarrow \infty}(1)$  of  $A$ , while from the preceding discussion we know that the maximum density is at most  $\alpha + o_{m \rightarrow \infty}(1)$ , and the claim now follows from Markov's inequality.

136.  [Terence Tao](#) Says:  
[February 4, 2009 at 6:57 pm](#) | [Reply](#)

## 121. Szemerédi's uniformisation trick

In 120 I noted how an extremal set  $A$  tends to have density close to  $\alpha$  on most  $m$ -dimensional subspaces. In Endre's big paper (the proof of Szemerédi's theorem), he uses a very nice double counting argument to say a much stronger statement: given any specified dense subset  $\Omega$  of  $[3]^m$ ,  $A$  also has density close to  $\alpha$  on “most” copies of  $\Omega$ . (He eventually applies this fact to sets  $\Omega$  that are the cells of a Szemerédi partition of a certain graph he constructs on  $[3]^m$ .) This trick does not appear to be as well known as it should be (and is absolutely crucial to Endre's proof), so I am reproducing it here.

Here is the formal claim: let  $1 \ll k \ll d \ll m \ll n$  (where  $x \ll y$  means that “ $y$  is sufficiently large depending on  $x$ ”, and let  $\Omega_1, \dots, \Omega_k \subset [3]^m$ . Then one can find a copy of  $[3]^d \times [3]^m$  in  $A$  such that for every  $x$  in  $[3]^d$  and all  $1 \leq j \leq k$ , the relative density of  $A$  in  $x \times \Omega_j$  is  $\alpha + o_{d \rightarrow \infty}(1)$ .

Proof. We need some intermediate numbers  $d \ll r \ll l \ll m$ . Pick a random copy of  $[3]^l \times [3]^m$  in  $A$ . By the previous discussion, every  $x$  in  $[3]^l$  has a probability  $1 - o_{m \rightarrow \infty}(1)$  of being such that  $A$  has density  $\alpha + o_{m \rightarrow \infty}(1)$  in  $x \times [3]^m$ . By the union bound (assuming  $m$  is large depending on  $l$ ), we may thus select a copy of  $[3]^l \times [3]^m$  such that  $A$  has density  $\alpha + o_{m \rightarrow \infty}(1)$  in *every*  $x \times [3]^m$ .

For each  $j$ , let  $f_j(x)$  be the density of  $A$  in  $x \times \Omega_j$ , rounded to the nearest multiple of  $1/r$ . Applying coloring Hales-Jewett (because  $l$  is large compared to  $k, d, r$ ), we can pass from  $[3]^l$  to a subspace  $[3]^d$  such that the  $f_j$  are constant on this subspace for all  $j$ :  $f_j(x) = a_j$ . But by double counting the density of  $A$  on  $[3]^d \times \Omega_j$  we see that  $a_j \leq \alpha + o_{d \rightarrow \infty}(1)$ ; similarly by double counting the density of  $A$  on  $[3]^d \times ([3]^m \setminus \Omega_j)$  we have  $a_j \geq \alpha - o_{d \rightarrow \infty}(1)$ . The claim follows.

137.  [Ryan O'Donnell](#) Says:

[February 4, 2009 at 7:01 pm](#) | [Reply](#)

122. Terry, re your first #120, thanks! Er, would it be possible to explain it in the form, “The new dimension  $m$  needs to be at least  $f(n)$ , and the density needs to increase by at least  $g(m)$ .”?



138. [Ryan O'Donnell](#) Says:

[February 4, 2009 at 7:14 pm](#) | [Reply](#)

123. I guess one thing that worries me is that it already seems hard to give explicit (nonrandom) sets  $A$  — strongly nonuniform or not! — which don't have a modest density increment. The only examples I know are those of the “modular arithmetic” type discussed by Boris and Tim earlier.



139. [jozsef](#) Says:

[February 4, 2009 at 7:41 pm](#) | [Reply](#)

124 Sperner

Going back to Tim's removal lemma in Kneser graphs 106 ; It actually says that every large matching in  $K_{n,m}$  spans at least the expected number edges. This would be a Ruzsa-Szemerédi type theorem in a sparse graph. From a different viewpoint it reminds me to Bollobas' theorem, which states that any induced matching in  $K_{n,m}$  has size at most  $\binom{2m}{m}$ . There is an algebraic proof of Bollobas' theorem by Lovasz, it might give us some information about the number of edges spanned by a large matching.



140. [Gil](#) Says:

[February 4, 2009 at 7:48 pm](#) | [Reply](#)

125. ((follow on 111.) well, since the numeric values of the variables are not important for combinatorial lines this means that the equations (at least for  $k=3$ ) can be slightly more general.)



141. [Kevin O'Bryant](#) Says:

[February 4, 2009 at 8:19 pm](#) | [Reply](#)

126. Regarding Terry's post (45), I've extended the Green/Wolf version of Elkin's strengthened form of Behrend's construction (of sets of integers without 3-term APs) to longer APs. This parallels Rankin's extension of Behrend's construction.

<http://arxiv.org/abs/0811.3057>




142. [Terence Tao](#) Says:

[February 4, 2009 at 8:30 pm](#) | [Reply](#)

127. Density incrementation

@Ryan.122: This is true... where  $f(n)$  is any function you choose that goes to infinity, and  $g(m)$  is a function that you *don't* get to choose, and goes to zero very very slowly (and ineffectively) as  $m \rightarrow \infty$ . (But you are allowed to make  $f$  depend on  $g$ , for what it's worth.)


For sake of discussion, suppose the rate of convergence of  $c_n/3^n$  to  $\alpha$  was inverse tower-exponential... then the game is basically “get a density increment of at least  $1/\log_* m$  on a subspace of dimension  $m \geq \log_* \log_* n$ ”. Of course, we don't know in advance that we have an inverse tower-exponential rate - so you have to prove something like this but with  $\log_*$  replaced by some other unknown slowly growing function.

143.  [Terence Tao](#) Says:  
[February 4, 2009 at 9:04 pm](#) | [Reply](#)

128. Density incrementation


One reason, by the way, why it's so difficult to find dense subsets of  $[3]^n$  that don't have any significant density incrementation on a large subspace is that, thanks to the  $k=3$  DHJ theorem, they don't exist! (In particular, any dense subset has to have density 1 on a subspace of dimension  $f(n)$  for some  $f$  going to infinity, by the same argument used in Terry.102.)

Of course, we can't actually *use* this fact in our proof, as it would be circular.


144.  [gowers](#) Says:  
[February 4, 2009 at 10:55 pm](#) | [Reply](#)

129. Sperner

For people who aren't comfortable with induced matchings, here's a translation of what Jozsef says in 124 about Bollobás's theorem. An induced matching in a graph means a collection of edges  $x_i y_i$  such that no  $x_i$  is joined to any  $y_j$  with  $j \neq i$ . So the theorem is saying that in the graph where you join two sets of size  $m$  if and only if they are disjoint, the biggest collection of disjoint pairs  $(A_i, B_i)$  you can find with  $A_i$  always intersecting  $B_j$  if  $i \neq j$  is the one where you take all subsets  $A_i$  of a  $2m$ -element subset of  $[n]$  and let  $B_i = A_i^c$  in each case. Jozsef, there clearly needs to be a lower bound on  $m$  in terms of  $n$  here: can you remember what it is?

145.  [gowers](#) Says:  
[February 4, 2009 at 11:01 pm](#) | [Reply](#)

Metacomment: It would be useful if contributors to this discussion could look at the discussion over on the Questions of Procedure post. I've done a poll and there seems to be a small majority in favour of separate new posts for each thread. That would require summaries of the various threads, so I think we need to agree what they are and find out who would be prepared to write the summaries. Jason's given us some suggestions about what the threads might be, though I think things may have moved on since then. I'll discuss this over at the other post.


146.  [Terence Tao](#) Says:  
[February 5, 2009 at 4:07 am](#) | [Reply](#)

130. The  $k=2.5$  case of DHJ

I was trying to adapt Szemerédi's proof of Roth's theorem to the  $k=3$  DHJ case, somehow using the  $k=2$  DHJ (Sperner) as an initial step, when I discovered a very strange phenomenon for DHJ, not present for corners or for arithmetic progressions: I cannot logically relate  $k=2$  DHJ to  $k=3$  DHJ! Instead,  $k=3$  DHJ trivially implies the following statement:

$k=2.5$  DHJ: If  $n$  is sufficiently large depending on  $\delta$ , and  $A \subset [3]^n$  has density at least  $\delta$ , then there exists a combinatorial line whose first two points lie in  $A$ .


But I cannot deduce  $k=2.5$  DHJ from the  $k=2$  DHJ, or conversely, basically because  $[2]^n$  is an incredibly sparse subset of  $[3]^n$ . Actually, despite the Sperner-type appearance to  $k=2.5$  DHJ, I was not able to come up with a combinatorial proof of this statement at all! Presumably I must be missing something obvious. But in any case, this looks like an obvious test problem to work on, being strictly easier than  $k=3$  DHJ.

147.  [Terence Tao](#) Says:  
[February 5, 2009 at 4:12 am](#) | [Reply](#)

131. Szemerédi's proof of Roth's theorem

This is perhaps only for my own benefit, but I was at least able to extend Szemerédi's proof of Roth's theorem (see 86, 118) to the  $(\mathbb{Z}/3\mathbb{Z})^n$  setting. As in 118, we know that the complement of  $A$  contains  $2 \cdot Q - A$  for some medium-dimensional cube  $Q$ . We can represent  $Q$  as an increasing sequence  $Q_1 \subset Q_2 \subset \dots \subset Q_d = Q$  as in 86, and use the pigeonhole principle to find a place where  $2 \cdot Q_i - A$  stabilises, i.e. it has almost the same density as  $2 \cdot Q_{i+h} - A$  for some medium-sized  $h$ . Thus the set  $2 \cdot Q_i - A$  is almost invariant under the  $h$  steps  $v_i, \dots, v_{i+h}$ , and so is approximately equal to a union of cosets of the subspace  $V$  spanned by these  $h$  vectors. So  $A$  largely avoids a dense union of cosets of  $V$ , so by pigeonhole, it must have significantly increased density on another coset of  $V$ , and this is enough for the density increment argument to kick in.

But I do not see how to use this sort of argument to prove either DHJ or corners, the problem being that the analogue of  $2 \cdot Q - A$  is far too sparse in either setting.

148.  [Ryan O'Donnell](#) Says:  
[February 5, 2009 at 5:26 am](#) | [Reply](#)

132. The  $k=2.5$  case of DHJ.


Terry: how interesting! I sort of had thought of this  $k=2.5$  version at some point (it was in my mind when writing #65). My two thoughts on it were:

- a) it probably should follow from some Sperner-pushing-shadows-around argument;
- b) trying to extrapolate from it to the  $k=3$  result looked very difficult (due to the incredible sparsity you mention) but worth thinking about (this was sort of in my mind at the end of #114).


So I'm surprised and interested to hear that its combinatorial proof is not obvious. Seems indeed like a good problem to think about then.

149. [Boris](#) Says:  
[February 5, 2009 at 5:46 am](#) | [Reply](#)

132. To prove  $k=2.5$  in post Tao.#130 note that we can assume that the number of 1's 2's and 3's is  $n/3 + O(\sqrt{n})$  of each in all words of  $A$ . Then look at copies of  $[2]^n$  of the following type. For a word  $w$  in  $\{1, 2, 3, *\}$  of length  $n$  with say  $m$  stars, consider all the words obtained from  $w$  by replacing stars with 1's and 2's. Hence, for each  $w$  we get a copy of  $[2]^n$ . Density on each such copy is at most  $1/\sqrt{m}$ . If  $m = o(\sqrt{n})$ , double-counting argument gives DHJ  $k=2.5$ .

150.  [Ryan O'Donnell](#) Says:  
[February 5, 2009 at 5:59 am](#) | [Reply](#)

133. I was just going to say something similar. My version: Suppose  $A$  has density greater than  $\mu := \binom{n}{n/2} 2^{-n}$  in  $[3]^n$ . Then there has to be a particular way of fixing the positions of the 3's such that  $A$ 's density in the resulting copy of  $[2]^n$  is still greater than  $\mu$ . Now apply Sperner.

151.  [Ryan O'Donnell](#) Says:  
[February 5, 2009 at 6:05 am](#) | [Reply](#)

134. The "pushing shadows around" Sperner proof gives (I think) nice statements like, "If  $A$ 's density is a lot more

than  $\mu$  then it must contain lots of “half-combinatorial lines.” My super-duper-naive hope in #114 was that if we could find “enough” of these half-lines in  $A$  (since, after all,  $A$  has density  $\delta = \Omega(1) \gg 1/\sqrt{n} = \mu$ ), we might be able to find a third point in  $A$  to finish off one of the half-lines. But such a plan would have to be awfully sophisticated to have a shot at working, I’d guess.

152.  *jozsef* Says:

[February 5, 2009 at 7:30 am](#) | [Reply](#)

135.

The  $k=2.5$  case of DHJ

There is more true, namely there are subcubes in any dense subset of  $[3]^d$ . An  $m$ -dimensional subcube is a set of elements each has some fixed positions where all elements are the same (1,2, or 3) and  $m$  blocks where the elements in one block are all one or all two. There is an element for every possible 1-2 combination of the blocks so there are  $2^m$  elements in a subcube. I tried to use such subcubes for a Szemerédi type argument, but without success. It is also remarkable, that the shadow of such cube (elements where one block has 3-s) can’t be in our set or there would be a line. The shadow also contains an  $m-1$  dimensional - empty - subspace,  $[3]^{m-1}$ . So, we know that for our set is full with empty subspaces, however I don’t know if we have enough for pushing the density up in another subspace.

153.  *jozsef* Says:

[February 5, 2009 at 7:45 am](#) | [Reply](#)

Contd. There is a relevant paper of Gunderson, Rödl, and Sidorenko, where they give sharp bounds on subcubes I mentioned in the previous note. Title: “Extremal problems for sets forming Boolean algebras and complete partite hypergraphs, J. Combin. Th. Ser. A 88 (1999), 342-367. ”

You can find it at [http://home.cc.umanitoba.ca/~gunderso/published\\_work/paper7.pdf](http://home.cc.umanitoba.ca/~gunderso/published_work/paper7.pdf).

154.  *jozsef* Says:

[February 5, 2009 at 8:08 am](#) | [Reply](#)

136 Density increment

Let me try to make the previous observation clear. We show that if  $S \subset [3]^n$  is  $c$ -dense and line-free, then for every  $m$  there is a  $c' > 0$ ,  $c' = c'(c, m)$  such that  $c'$  - fraction of the  $m$ -dimensional subspaces ( $[3]^m$ ) have empty intersection with  $S$ .

155.  *jozsef* Says:

[February 5, 2009 at 8:34 am](#) | [Reply](#)

137. Sperner

On Tim’s question; I don’t think that there is a lower bound on  $m$ .

Also, I don’t see what would be the right question here. If we take a perfect matching, extremal for Bollobás’ theorem, with  $m=n/2-k$  ( $k$  is a small integer) then there are only a few edges, but you have to remove half of the sets to destroy all edges. Or, it is late night again, and I overlooked something.

156.  *gowers* Says:

[February 5, 2009 at 10:56 am](#) | [Reply](#)

138. Sperner




Ah yes, I was thinking that if  $n$  was much much larger than  $m$  then you would do better by choosing lots of completely disjoint pairs  $(A_i, B_i)$ , but I was overlooking the fact that  $A_i$  is not allowed to be disjoint from  $B_j$ .

Let me check the arithmetic in your last remark. If  $k$  is a small integer, then the number of sets in  $\binom{[2m]}{m}$  is about  $2^{-k}$  times the number of sets in  $\binom{[2m+k]}{m}$ , I think. Yes, that's right, since the ratio of  $\binom{n+1}{m}$  to  $\binom{n}{m}$  is  $(n+1)/(n+1-m)$ , which for us is about 2. So if  $k$  is a small integer, then it is indeed the case that choosing all subsets of  $[2m]$  gives us a positive fraction of all sets. But in this case, the only edges are the ones that join a set to its complement in  $[2m]$ .

The annoying thing (from the point of view of trying to formulate a decent conjecture) is that this makes the graph sparse. After all, the average degree in the graph, even when  $k = 1$ , tends to infinity with  $n$ . (It is  $\binom{m+k}{m}$  which is around  $(n/2)^k$ .)

However, I think that there is still hope for a less neat conjecture that concerns several layers around the middle. And I also think that there may be some counterpart of this false conjecture for  $\{0, 1, 2\}$ -sequences that is also false. I think once we identify this counterpart, we will see that we never expected or needed it to be true, and that will make us feel better about whatever  $\{0, 1\}$ -version we come up with. I'm busy for a few hours now but will try to make all this more precise later.


157.  [gowers](#) Says:  
[February 5, 2009 at 11:32 am](#) | [Reply](#)

139. Sperner

Just one tiny extra remark before I go. For now it seems that one way of rescuing the conjecture is to insist that  $k$  tends to infinity with  $n$ , so that  $2^{-k}$  is no longer a constant fraction. This feels like a natural restriction because it corresponds to insisting that the variable set has size that tends to infinity — which in general it will have to.

158. [Nielsen's posts tagged Science2.0 | The Daily Clique](#) Says:  
[February 5, 2009 at 5:04 pm](#) | [Reply](#)

[...] Weblog: Timothy Gowers updates the rules to his massively collaborative mathematics experiment. He posted the first project for this experiment a few days ago and has already received 145 [...]

159.  [gowers](#) Says:  
[February 5, 2009 at 5:10 pm](#) | [Reply](#)

140. Sperner

I think what I said in my previous two comments does indeed make sense. To see why, imagine that we tried in the DHJ case to a single triple of slices  $\Gamma_{a+r,b,c}$ ,  $\Gamma_{a,b+r,c}$  and  $\Gamma_{a,b,c+r}$ . Then the triangle removal lemma would say that if there were very few combinatorial lines coming from some subsets of these three slices, then one could remove a small proportion of points from the subsets and end up with no combinatorial lines. If  $r$  is a fixed positive integer, then this is false, but we don't expect to be able to produce combinatorial lines with a fixed size of variable set: the size will have to depend on the density of  $A$ .

If a triangle removal lemma is true for triples of slices like this when  $r$  is large, I'm not sure that it implies DHJ, because as I've already pointed out you don't get degenerate triples that involve three different slices — obviously enough. In fact, this is a general point to think about: if we found the most wonderful triangle removal lemma imaginable, would it actually lead to a proof of density Hales-Jewett? (I think there is an analogy with the situation with corners that suggests to me that the answer isn't obviously no.)

Going back to Sperner, my current pair-removal problem, not yet shown to be true and not yet shown to be badly formulated, is to show that if  $\mathcal{A}$  and  $\mathcal{B}$  are subsets of  $\binom{[n]}{m}$ , where  $n = 2m + k$  and  $k$  tends to infinity slowly with  $n$ , and if there are few disjoint pairs  $(U, V) \in \mathcal{A} \times \mathcal{B}$ , then it is possible to remove a small number of sets (compared with  $\binom{n}{m}$ ) from  $\mathcal{A} \cup \mathcal{B}$  and end up with no disjoint pairs.

160.  [gowers](#) Says:

[February 5, 2009 at 5:44 pm](#) | [Reply](#)

141. Sperner

In relation to 140, I tried to work out what the regularity lemma says for 1-uniform hypergraphs, but the resulting statement is even more ridiculous than pair removal in the dense case and not worth repeating here. This makes me think that there's a genuine chance that if we could prove an appropriate disjoint-pair removal lemma, that it might be the extra ingredient that's needed to mix with triangle-removal to get DHJ. I realize that I've sort of said this before.

161.  [gowers](#) Says:

[February 5, 2009 at 5:57 pm](#) | [Reply](#)

142. Sperner

Part of the Kneser-graphs philosophy is that they behave rather like the graph on the  $n$ -sphere where you join two points if they are almost antipodal to each other. So another question that might be worth investigating is whether there is any chance of an almost-antipodal-pair-removal lemma. It would say something like this. Let  $A$  and  $B$  be two subsets of  $S^n$  such that the measure of all pairs  $(x, y)$  such that  $x \in A, y \in B$  and  $\langle x, y \rangle \leq -1 + c$  is small. Then it is possible to remove sets of small measure from  $A$  and  $B$  and end up with no such pairs.

It may take some work to turn that into a decent question, if indeed it can be done. I think I know what it means for the set of disjoint pairs to have small measure. It should mean small compared with the measure of all such pairs in  $S^n \times S^n$  (which is equal to 1 times the measure of the spherical cap  $\{y : \langle x, y \rangle \leq -1 + c\}$ ).

In this continuous context, it might be necessary for the conclusion to say that after the removal what you no longer have is pairs that have inner product less than  $-1 + c'$  for some  $c' < c$ .

I'm slightly suspicious of this variant, because it feels to me as though if you've got two sets with *no* almost antipodal pairs then at least one of them must be very small, unless  $c$  is something like  $Cn^{-1/2}$ . In fact, that is indeed the case as it's saying that the distance from  $A$  to  $-B$  is large, so it comes from concentration of measure. I'm not yet sure what happens if you just assume that  $A \times B$  contains *few* almost antipodal pairs. If it is possible for both  $A$  and  $B$  to be large, then the result is false, and if it is impossible then it is in a certain sense trivial.

But I think this could be a useful observation: it shows that if there is any chance of a statement holding in the Kneser graph, then it must use some property of that graph that distinguishes it from the sphere graph.

162.  [Terence Tao](#) Says:

[February 5, 2009 at 6:30 pm](#) | [Reply](#)


143. Sperner

Tim, just a small correction to 139: the number  $k$  of wildcards can be taken to be bounded by a quantity depending only on the density  $\delta$ , either for Sperner or DHJ. Indeed, one can foliate  $[2]^n$  or  $[3]^n$  into slices of  $[2]^{n_0}$  or  $[3]^{n_0}$ , where  $n_0 = n_0(\delta)$  is the first integer for which Sperner or DHJ applies. By the pigeonhole principle, one of these slices has density at least  $\delta$ , and thus contains a combinatorial line with at most  $n_0$  wildcards.

Now, there should also be lines with wildcard length as high as  $\sqrt{n}$  or so, I think. I can reason by analogy with the corners setting: suppose one wanted to find corners in a dense subset of the strip  $\{(a, b) \in [n]^2 : |a - b| \ll \sqrt{n}\}$ . One can cover this strip by grids of spacing  $k$  and of length at least  $n_0$  (where  $n_0$  is similar to as in the previous paragraph), and it seems to me that one should get a corner of length equal to a bounded multiple of  $k$  for any specified  $k$  between 1 and a small multiple of  $\sqrt{n}$ . I think the same sort of argument should work in the DHJ case. But it seems to me that it is somehow easier to fish for lines in the small  $k$  end of the pool rather than the large  $k$  end, even though the graph is denser in the latter case.


Re: 132, 133, 135: thanks, it's clear now that 2.5 and 2 are equivalent. For some reason I was reluctant to partition the big cube  $[3]^n$  into variable-sized copies of  $[2]^m$ , but I see now that that is the way one should think about it.

Jozsef@135,136,137: this fits pretty well with what I said at 118. The key step in Endre's proof of Roth is that a small cube  $Q$  is suddenly amplified to a huge structured set  $2 \cdot Q - A$  in the complement which has really big density (at least  $\delta$ ) and still contains tons of cubes (and more to the point, *parallel* cubes). In our situation we have a much sparser set of cubes in the complement, and they are not all parallel to each other, and it doesn't look like this is enough to force a significant density increment anywhere. (Note also that one could also get a similar sparse collection of non-parallel cubes in the complement from the colouring HJ theorem.) I tried to combine all this with Endre's uniformisation trick (121) but without luck (again, the issue is sparsity, the  $\Omega_i$  in 121 need to be dense).


163.  Jozsef Says:  
[February 5, 2009 at 6:59 pm](#) | [Reply](#)

144 Sperner

I would like to see a removal lemma type argument for Kneser graphs very much and I agree that if  $k$  goes to infinity (Tim 138), then there should be a sensible one. Let me mention a light version of the triangle removal lemma, which is still strong enough for applications; If the number of triangles is  $o(n^2)$  in an  $n$ -vertex graph, then one can remove  $o(n^2)$  edges to destroy a positive fraction of the triangles. So, I don't think that in the Kneser graph problem we should go for removing all edges, it should be enough that with a few vertices we can destroy much more than expected edges.

164.  Jozsef Says:  
[February 5, 2009 at 7:02 pm](#) | [Reply](#)

144 Contd. I meant  $cn^2$  triangles ...

165.  Ryan O'Donnell Says:  
[February 5, 2009 at 7:09 pm](#) | [Reply](#)

145. Sperner.

Hi Tim, re your #142. It is known that for all  $c \leq 0$ , among sets  $A$  and  $B$  on the sphere of fixed measures, the ones which minimize the measure of pairs  $x \in A, y \in B$  with  $\langle x, y \rangle \leq c$  are the "concentric" caps. Note that in high dimensions, having inner product  $\leq c$  is very much similar to having inner product  $\approx c$ . The proof is by shifting. One such proof appears here: <http://www.wisdom.weizmann.ac.il/~feige/Approx/maxcut.ps>

Based on this it may be possible to calculate the answer to your question. I actually prefer to do these things not on the high-dimensional sphere but on the equally good surrogate, high-dimensional Gaussian space. It is known that for all  $\rho \leq 0$ , the subsets  $A$  and  $B$  of fixed Gaussian measure which minimize the probability of  $x \in A, y \in B$ , where  $(x, y)$  is a pair of correlated Gaussians with covariances  $\rho$ , are two "parallel" halfspaces. This is a theorem of Borell: <http://www.springerlink.com/content/n761353703r7p675/>

Just wanted to get those references out there; will try to think about the ensuing calculations...

166.  [jozsef](#) Says:

[February 5, 2009 at 7:21 pm](#) | [Reply](#)

146. Sperner

Terry @118 142, Yes, We need more structure to apply Endre's trick. It was easy to apply when (almost) every element is a center of symmetry in the following way; Choose an element of your set (center). Any other element has a unique mirror image and if both are in your set, then your set contains a 3-term arithmetic progression. With HJ, most pairs are not in line, so the "symmetry" is only local.

167.  [gowers](#) Says:

[February 5, 2009 at 7:56 pm](#) | [Reply](#)

147. Sperner

Terry — indeed. In fact, I came to a similar realization in the DHJ context at the end of the first paragraph of 140. But I suppose one could regard things in another way and say that we're trying to get the density to tend to zero (very very slowly) and therefore we need  $k$  to tend to infinity.

Ryan — at first I skimmed what you said and thought it was repeating what I had said, and then I realized that it wasn't at all. Although the sphere question seems genuinely distinct from the Kneser-graphs question, it is definitely good to know that something along these lines is known.

If  $A$  and  $B$  are concentric caps and are not of tiny measure, then they must reach almost all the way to the equator, at least, so my guess is that that will give them a lot of almost antipodal pairs. So at the moment my money is on few antipodal pairs implying that at least one of  $A$  and  $B$  has small measure. If so, then we would have the true-for-not-quite-the-kind-of-reason-we-hoped-for option rather than the false option.

168.  [gowers](#) Says:

[February 5, 2009 at 8:24 pm](#) | [Reply](#)

148. Sperner

As I was sitting through my daughter's piano lesson, my mind wandered back to pair removal in  $K_{n,m}$ , and I found myself wondering whether we were already done. Almost certainly not, but in the spirit of keeping thoughts public, let me attempt to deduce it from the theorem of Bollobás mentioned by Jozsef in 124.

First, here's the thought that's in the back of my mind. It's that the number of edges you need to remove to make a graph triangle free is within a constant of the largest possible number of edge-disjoint triangles in that graph. (This was pointed out to me by Tom Sanders.) The proof is that if you've got  $m$  edge-disjoint triangles then you need to remove at least  $m$  edges. Conversely, if you've got a maximal set of edge-disjoint triangles and there are  $m$  of them, then if you remove all edges from all those triangles you will be left with no triangles (because any triangle left would be edge-disjoint from the removed ones, contradicting maximality). So the two numbers are equal to within a factor of 3.

Now suppose we are trying to get rid of all disjoint pairs  $(U, V)$  from  $\mathcal{A} \times \mathcal{B}$ . If we can find  $m$  such pairs, with no two sharing a  $U$  or a  $V$ , then we have to remove at least  $m$  sets from  $\mathcal{A} \cup \mathcal{B}$ . Conversely, if we've got a maximal set of disjoint pairs and we remove all sets that are involved in those pairs, then we cannot have any disjoint pairs left.

I now see that Jozsef understood this way back in 104. And unfortunately, it's not covered by Bollobás's theorem because that concerns maximal *induced* matchings rather than maximal matchings.

However (and this too Jozsef clearly realized ages ago), this little argument allows us to rephrase the question: if

you've got a lot of disjoint pairs  $(A_i, B_i)$  (of sets of size  $m$  in  $[2m + k]$ ), where all the  $A_i$  are distinct and all the  $B_i$  are distinct, then are you forced to have a lot of disjoint pairs  $(A_i, B_j)$  with  $i \neq j$ . Bollobás's theorem states that you must have at least one such pair. I suppose an averaging argument would lift that up to something, but probably not to within a constant of the trivial maximum. But as Jozsef suggested, the *proof* of Bollobás's theorem might do something for us.

169.  [Terence Tao](#) Says:

[February 5, 2009 at 8:28 pm](#) | [Reply](#)

148. Obstructions to uniformity

It took me a while to catch up with this thread - this is perhaps one thread which is particularly ripe for a fresh start with a post summarising previous progress and listing the known examples.

Here are the examples of interesting “obstructing sets” to either DHJ, Sperner, or Kruskal-Katona that I can find in the above threads:

Gowers.38: A set  $A$  of words  $x$  whose last entry  $x_n$  depends on the parity of the number of 1s in all other entries.

Boris.44, Gowers.71, Boris.73, Gowers.76, Gowers.108, Ryan.117: A set  $A$  such that for all  $x$  in  $A$ , the number  $(a, b, c)$  of 1s, 2s, and 3s of  $x$  respectively in  $[n]$  (or some subset  $S$  of indices in  $[n]$ , or using some other weighting on  $[n]$ ) lie in a structured set.

Gowers.69, 70, 71: A set  $A$  such that the probability that  $x_i = j$  for some  $i \in [n]$ , some  $j \in \{1, 2, 3\}$ , and  $x$  drawn uniformly from  $A$ , deviates significantly from  $1/3$ .

Ryan.92:  $A$  consists of all the sets which contain a medium-sized block of 2s among some special collection of blocks.

Ryan.94: Take any existing example of a non-uniform set  $A$  and tack on some fixed string of digits (e.g. 23) at the end, increasing  $n$  to  $n+2$  (say) and decreasing the density by a factor of 9 (say).

It seems that one can unify all of the above examples so far. Given a set  $A$  and weighting  $w$  on  $[n]$  by integers, define the *richness profile* of  $A$  with respect to the weight  $w$  to be the function that assigns to each triple  $(a, b, c)$ , the number of elements of  $A$  whose weighted count of 1s, 2s, and 3s are equal to  $a, b, c$ , respectively. (In most examples,  $w$  is just the indicator function of some set  $S$  of indices in  $[n]$ , so the profile is counting how many 1s, 2s, and 3s there are in  $S$ .)

A set  $A$  has an *standard obstruction to uniformity* if there is a reasonably large subspace  $V$  on which  $A$  has equal or larger density, and a weight  $w$  such that the richness profile of  $A \cap V$  with respect to  $w$ , deviates in a structured way from what one would expect in the random case. (This is for instance the case if  $A$  has significantly large density on  $V$  than it does globally.) By “structured” I mean that the deviation is not pseudorandom in the sense of locating corners  $(a + r, b, c)$ ,  $(a, b + r, c)$ ,  $(a, b, c + r)$  in the richness profile.

This seems to cover all known examples, while still excluding the random example. (Incidentally: have we shown yet that the random set is quasirandom? This seems to be a necessary step if we are to have any hope at all here.)


170.  [gowers](#) Says:

[February 5, 2009 at 8:50 pm](#) | [Reply](#)

149. Obstructions to uniformity.

A couple of things I didn't quite understand there Terry. If you pull over the obstructions from  $[n]^2$  (as suggested in 116), do you call that a pseudorandom deviation? Where does that example fit into your scheme above? And what are you referring to when you talk about “the random example” in your last paragraph? Do you mean that your definition

of a standard obstruction does not apply to random sets? I think I don't really understand the last sentence of your penultimate paragraph.

171.  *Ryan O'Donnell* Says:

[February 5, 2009 at 9:44 pm](#) | [Reply](#)

150. Re the last sentence in Tim's #142: I mentioned this in #82 but perhaps I should have pointed to a later-on paper in the Irit Dinur oeuvre:

<http://www.cs.huji.ac.il/~dinuri/mypapers/intersecting.pdf>

Here she and Friedgut give a mostly complete characterization of large independent sets in Kneser graphs for  $m$  a little bit less than  $n/2$ . To be a large intersecting family, you have to nearly be a subset of a large intersecting family which is a "junta". (A set family  $\mathcal{F}$  is a "junta" if a set's absence or presence in  $\mathcal{F}$  depends only on its inclusion/exclusion of a constant number of coordinates from  $[n]$ .)

They also discuss the "cross-intersecting" question, which is basically when you have \*two\* set families — as we do. I would also guess that the improved characterization in their "Conjecture 1.3" may well be provable with "today's technology".

172.  *gowers* Says:

[February 5, 2009 at 10:54 pm](#) | [Reply](#)

151. Sperner

Ryan, I didn't manage to find an obviously relevant part of the paper you linked to in 82, but the reference you've just given does indeed look promising. Do you know enough about their argument to know whether it is robust? That is, if instead of assuming that you have an intersecting (or cross-intersecting) family you just assume that your sets intersect almost all the time, can you still say that there's some small set of coordinates that explains this phenomenon? I see that their methods of proof are pretty analytic, so maybe there's some hope here.

173.  *Ryan O'Donnell* Says:

[February 5, 2009 at 11:50 pm](#) | [Reply](#)

152. Hi Tim. Since the proof is analytic, it should almost surely be fine if the initial set(s) are almost always intersecting. Perhaps an even better reference is the proof of Thm. 1.1 in the last 4 pages of this intermediate paper:

<http://www.cs.huji.ac.il/~dinuri/mypapers/DFR.pdf>

174.  *gowers* Says:

[February 6, 2009 at 2:04 am](#) | [Reply](#)

153. Sperner.

I'm in the middle of writing a summary of the Sperner thread (as well as a few other related ones), to be used when this discussion splits up. At the moment the plan is to have three new discussions.

While writing the summary, I had the following thought. When proving the corners result, we can use a weaker triangle-removal lemma where the hypothesis is not that there are few triangles, but the stronger hypothesis that each edge is contained in at most one triangle. The pair-removal equivalent of this is that for each  $U$  the only disjoint pair is  $(U, U^c)$ . This starts to sound more Bollobás-like, though the sets  $U$  no longer all have the same cardinality.

So I have two questions. 1. Is there a variant of Bollobás's theorem where the sets don't all have to have the same size (though I don't mind insisting that the all have size close to  $n/2$ )? 2. If so, can we generalize its proof to a triangle-



removal lemma that uses as its hypothesis that the only triangles in the tripartite graph are degenerate?

This is a strategy that could in principle do the whole thing, so I'm expecting difficulties to arise very quickly ...

Unfortunately, I must go to bed now, so finishing the summary will have to wait till tomorrow.

175. [Upper and lower bounds for the density Hales-Jewett problem « What's new](#) Says:

[February 6, 2009 at 2:05 am](#) | [Reply](#)

[...] of several threads from Tim Gowers' massively collaborative mathematics project, "A combinatorial approach to density Hales-Jewett", which I discussed in my previous [...]

176.  [Terence Tao](#) Says:

[February 6, 2009 at 2:06 am](#) | [Reply](#)

Metacomment: All discussion on upper and lower bounds for  $c_n$ , and proof strategies related to Szemerédi's proofs of Roth's theorem, should go to the new thread on my blog.

<http://terrytao.wordpress.com/2009/02/05/upper-and-lower-bounds-for-the-density-hales-jewett-problem/>

177.  [Terence Tao](#) Says:

[February 6, 2009 at 2:18 am](#) | [Reply](#)

154: Obstructions to uniformity

@Tim.149: I'm thinking of a subset of  $[n]^2$  as being pseudorandom for the corners problem if it does not correlate with any Cartesian products (or if it has the right number of rectangles, or is small in what I like to call the "Gowers  $\square^2$  norm"). The example in 116, in particular, is not pseudorandom in this sense, being the Cartesian product of two random sets. So I will label this example (pulled back to  $[3]^n$  as a standard obstruction to uniformity (at least for one of the three indices 1, 2, 3 of a combinatorial line; for the other two indices one has to modify the notion of a Cartesian product slightly).

By the "random example", I take  $A$  to be a random subset of  $[3]^n$  of a fixed density  $\delta$ . This should have no noticeable density increment or other fluctuation in the richness profile for any subspace of dimension bigger than, say,  $\log^{10} n$ . But I haven't checked yet (presumably it is an easy entropy calculation) that  $A$  is quasirandom in the sense you want, i.e. the number of lines between  $A$ ,  $B$ , and  $B$  is the expected number for any large  $B$ . Obviously, this is something that has to be true if the obstructions to uniformity approach has any chance of working.

178. *r.moth* Says:

[February 6, 2009 at 6:08 am](#) | [Reply](#)

I have a suggestion - a blog is perhaps the wrong format for this type of experiment (or future experiments). Its sequential format obviously presents difficulties in following/retrieving information. What you want is a well structured webforum where users can create new threads on a whim, and all subposts to those threads can be easily followed - something with a directory tree structure where posts can be followed in a hierarchy.

You may even be able to find a forum that allows you to highlight threads spawned from posts in other threads - i know one exists...i just can't remember the name of it - though i might be thinking of news reader....but a news group is probably getting beyond the scope of this initial experiment.

When you start getting into hundreds, the blog is too one-dimensional to be efficient in following a line of thought. You most likely don't need a newsgroup (not a public one anyway) - just a private forum with enough goodies to make this thing more accessible, otherwise I think you're going to get swamped (think a thousand+ posts). just a thought.

179.  *josef* Says:

[February 6, 2009 at 6:48 am](#) | [Reply](#)

155. Sperner

There is a removal lemma for Kneser graphs! It holds in a certain range; “If a large subgraph has many but not too many edges, then one can remove a small fraction of the vertices destroying most of the edges.”

Some notations:  $G_N$  is an  $N$ -vertex subgraph of  $K_{n,m}$  the Kneser graph defined by the  $m$ -element subsets of the  $n$ -element base set. The graph has  $e$  edges and  $D$  denotes the largest degree and  $d=2e/N$  is the average degree.

If the vertices having degree at least  $Cd$  (for some large constant  $C$ ) are incident to most of the edges, then we are done; removing  $N/C$  vertices destroys most of the edges. So, we can suppose that  $D < Cd$  (after throwing away the large degree vertices) As we will see,  $D < Cd$  is not possible in a certain vertex (edge) range.

I will follow it in a new windows, I want to check what did I write.(latex)

180.  *Link Starbureiy* Says:

[February 6, 2009 at 6:53 am](#) | [Reply](#)

Tim, I'll publicize this on my blog, as well. This seems like a decent-enough collaborative scholastic.

181.  *josef* Says:

[February 6, 2009 at 7:12 am](#) | [Reply](#)

Contd. Well, I would change a few typos if I could, but it is OK.

We are going to use the “permutation trick” of Lubell. For any of the  $n!$  permutations we say that it separates the disjoint pair  $A, B$  if no element of  $B$  precedes any element of  $A$ . Any of the  $n!$  permutations can separate at most  $D^2$  pairs (edges) from  $G_N$ . Indeed, consider the set  $A$  which ends first, it would be separated from at least  $D+1$  sets if there were  $D^2 + 1$  separated pairs. Or, the rightmost set  $B$  (starting last in the permutation) would be separated from at least  $D+1$  sets, which would mean that  $D$  was not the largest degree in  $G_N$ . The number of permutations separating two elements is  $n! / \binom{2m}{m}$ . Collecting the two sides of this double counting we have  $e n! \binom{2m}{m}^{-1} \leq D^2 n!$  Using the initial bound on  $D$ , our final inequality is  $N^2 \leq C^2 \binom{2m}{m} e$ . Conclusion: If the previous inequality fails, then there are less than  $N/C$  vertices in  $G_N$  which are covering most of the edges.

182.  *josef* Says:

[February 6, 2009 at 8:40 am](#) | [Reply](#)

156 Sperner

contd. I don't think that the previous inequality is sharp, but I think it gives some nontrivial results. For example, if  $m \sim n - \sqrt{n}$  and we take a positive fraction of the  $m$ -element sets,  $N = c \binom{n}{m}$ , and the average degree is  $o(2^{\sqrt{n}})$  then  $o(N)$  vertices (sets) cover almost all edges.

183.  *josef* Says:

[February 6, 2009 at 8:46 am](#) | [Reply](#)

The previous short note is full with typos. I should go to sleep now...

I meant: the size of  $m$  is around  $n/2 - \sqrt{n}$  and that the aver. degree is  $o(2^{2\sqrt{n}})$  but not constant.

184.  *gowers* Says:

[February 6, 2009 at 9:08 am](#) | [Reply](#)

157. Sperner

Jozsef, I must try to digest what you say. It seems very promising.

185. [DHJ — the triangle-removal approach « Gowers's Weblog](#) Says:

[February 6, 2009 at 10:03 am](#) | [Reply](#)

[...] disjoint from precisely one  $\epsilon$ , then  $\epsilon$  must both be small. And Ryan (in comments 82, 145, 150 and 152) has drawn attention to existing papers on very closely related subjects. Very recently (at the [...])

186.  [gowers](#) Says:

[February 6, 2009 at 10:06 am](#) | [Reply](#)

OK, any thoughts about Sperner, triangle removal, regularity lemmas in tripartite Kneser-type graphs, etc., should now be given as comments on [a new post](#) that is meant to take over this thread of the discussion. If you would like me to modify the summary, then by all means suggest changes.

For now, the obstructions-to-uniformity thread stays here, but we should get a summary written for that as soon as we can.

*[Added 8/2/09 4.05pm GMT: there is now a [post on obstructions to uniformity](#), so all further discussion of this topic should continue there.]*

187. [DHJ — quasirandomness and obstructions to uniformity « Gowers's Weblog](#) Says:

[February 8, 2009 at 4:53 pm](#) | [Reply](#)

[...] By gowers This post is one of three threads that began in the comments after the post entitled A combinatorial approach to density Hales-Jewett. The other two are Upper and lower bounds for the density Hales-Jewett problem, hosted by Terence [...]

188. [Boris](#) Says:

[February 8, 2009 at 8:23 pm](#) | [Reply](#)

Metacomment: one really needs to have a preview function.

189. [yet another math2.0 proposal | neverendingbooks](#) Says:

[February 11, 2009 at 8:21 pm](#) | [Reply](#)

[...] were influential to Tim Gowers' dream of massively collaborative mathematics. He took an interesting problem, laid down a set of 12 rules-of-conduct and invited everyone to contribute. The project is still [...]

190. [A quick review of the polymath project « What Is Research?](#) Says:

[February 20, 2009 at 1:15 am](#) | [Reply](#)

[...] the next post, Gowers listed his ideas broken down into thirty-eight points. He also clarified the circumstances [...]

191. [Beginning of the Rest - Informal Online Research Collaboration](#) Says:

[March 19, 2009 at 4:30 pm](#) | [Reply](#)

[...] Cambridge University, used his blog for an experiment in massive online collaboration for solving a significant problem in math — combinatorial proof of the density Hales-Jewett theorem. Some six weeks (and nearly 1000 [...])

192. [Michael Nielsen » The Polymath project: scope of participation](#) Says:

[March 20, 2009 at 9:29 pm](#) | [Reply](#)

[...] an accomplished professional mathematician, initially thought that the chances of success were “substantially less than 100%”, even adopting a quite modest criterion for [...]

193. [A gentle introduction to the Polymath project « The Number Warrior](#) Says:  
[March 26, 2009 at 1:40 am](#) | [Reply](#)

[...] that he thought amenable to an online collaborative approach, then kicked things off with a blog post. Six weeks later, the main problem he proposed was declared (essentially) solved. However, the [...]

194. [Mythematics: a call for participation « Division by Zero](#) Says:  
[April 16, 2009 at 12:54 am](#) | [Reply](#)

[...] some high profile mathematicians have been using a blog to conduct a massively collaborative mathematics project. What a brilliant idea that could be done today, with our current [...]

## Leave a Reply

Logged in as [Ryan O'Donnell](#). [Logout »](#)

Submit Comment

☐ Notify me of follow-up comments via email.

---

[Blog at WordPress.com.](#)

[Entries \(RSS\)](#) and [Comments \(RSS\)](#).

⌵

# Gowers's Weblog

Mathematics related discussions

« [Quick question](#)

[A remarkable recent result in Banach space theory](#) »

## DHJ — the triangle-removal approach

By gowers

The post, [A combinatorial approach to density Hales-Jewett](#), is about one specific idea for coming up with a new proof for the density Hales-Jewett theorem in the case of an alphabet of size 3. (That is, one is looking for a combinatorial line in a dense subset of  $[3]^n$ .) In brief, the idea is to imitate a well-known proof that uses the so-called triangle-removal lemma to prove that a dense subset of  $[n]^2$  contains three points of the form  $(x, y)$ ,  $(x, y + d)$  and  $(x + d, y)$ . Such configurations are sometimes called *corners*, and we have been referring to the theorem that a dense subset of  $[n]^2$  contains a corner as the *corners theorem*.

The purpose of this post is to summarize those parts of the ensuing discussion that are most directly related to this initial proposal. Two other posts, [one hosted by Terence Tao on his blog](#), and the other here, take up alternative approaches that emerged during the discussion. (Terry's one is about more combinatorial approaches, and the one here will be about obstructions to uniformity and density-increment strategies.) I would be very happy to add to this summary if anyone thinks there are important omissions (which there could well be — I have written it fast and from my own perspective). If you think there are, then either comment about it or send me an email.

Two suggestions, [made by Terry in comment number 4](#) (the second originating with his student Le Thai Hoang), led to two of the main threads of the discussion relating to the triangle-removal approach. One was that since Sperner's theorem is the density Hales-Jewett theorem for alphabets of size 2, and since there is no obvious way of generalizing the existing proofs of Sperner's theorem to obtain the density Hales-Jewett theorem, then it would make sense to look for other proofs of Sperner's theorem. The other was that the existing methods of tackling density theorems tend to prove not just that you get one configuration of a desired type but a positive proportion of the number of configurations in the whole set. One can usually obtain a seemingly stronger result of this type by an easy averaging argument (a prototype of which was proved by Varnavides in the 1950s), but in the case of the density Hales-Jewett theorem the stronger statement is false: one can easily come up with dense subsets of  $[3]^n$  with far fewer than  $4^n$  combinatorial lines (the number that you get in  $[3]^n$  itself).

### The Varnavides thread.

This led to a search for a suitable Varnavides-like statement, and that search seems to have been successful — though we have not checked this carefully. (See [comments 61-64 and 66](#).) The basic idea is to restrict attention to the set  $T_m$  of all sequences  $x$  such that the numbers of 1s, 2s and 3s in  $x$  are all within  $m$  of  $n/3$ , where  $m$  is a large integer that is much smaller than  $n$ . An easy averaging argument shows that if you can prove that a dense subset of  $T_m$  contains a combinatorial line, then you have the density Hales-Jewett theorem. (It seems that one can take  $m$  to be anything from a very large constant that depends on the density only, to an integer that is a small multiple of  $\sqrt{n}$ .) The reason this helps is that it avoids the problem one has in  $[3]^n$  that the points in a random combinatorial line are very atypical: with high probability they contain about  $n/2$  of one

value and about  $n/4$  of the other two values.

This seems to have cleared up the Varnavides problem, so it is something to hold in reserve and use if we get to the stage where we have a definite approach to try out on the problem. For now, one can either just pretend that the Varnavides problem doesn't exist and think about  $[3]^n$ , or one can simply bear in mind that one is thinking about sequences with roughly equal numbers of 1s, 2s and 3s and combinatorial lines with small variable sets, or sets of wildcards as some of us are calling them.

### The Sperner thread.

Some definite progress has been made on trying to find alternative proofs of Sperner's theorem. One approach was initiated in [comments 21 and 22](#). The idea here was to try to find a statement that would have the same relationship to Sperner's theorem as the hoped-for (but not yet even precisely formulated) triangle-removal lemma had to density Hales-Jewett. An argument was given in comments 21 and 22 that the statement, if it existed, should have the following form.

Let  $\mathcal{A}$  and  $\mathcal{B}$  be collections of subsets of  $[n]$  such that there are very few pairs  $(U, V)$  with  $U \in \mathcal{A}$ ,  $V \in \mathcal{B}$  and  $U \cap V = \emptyset$ . Then it is possible to remove a small number of sets from  $\mathcal{A}$  and  $\mathcal{B}$  and end up with no such pairs.

We have been calling statements like this *pair-removal lemmas*, even though we have not yet proved any useful ones. A statement like this implies a weak form of Sperner's theorem as follows. Suppose you have a dense set system  $\mathcal{A}$  that contains no pair of distinct sets  $(U, W)$  with  $U \subset W$ . Then let  $\mathcal{B}$  consist of all complements of sets in  $\mathcal{A}$ . If we can find  $U \in \mathcal{A}$  and  $V \in \mathcal{B}$  with  $U \cap V = \emptyset$ , then  $U \subset V^c$ . Since  $V^c \in \mathcal{A}$ , this is a contradiction unless  $V^c = U$ . Therefore, the number of pairs  $(U, V)$  with  $U \in \mathcal{A}$ ,  $V \in \mathcal{B}$  and  $U \cap V = \emptyset$  is very small. The pair-removal statement then implies that one can remove small subsets from  $\mathcal{A}$  and  $\mathcal{B}$  in order to end up with no such pairs. But that contradicts the fact that  $\mathcal{A}$  is dense and for every  $U \in \mathcal{A}$  we have the pair  $(U, U^c)$ .

After some difficulty in formulating a precise pair-removal lemma that wasn't either trivial or false, we have arrived at the following test problem. (It's not actually the pair-removal lemma we're going to want, but it is cleaner, looks both true and non-trivial, and seems to involve all the right difficulties.)

*Tentative conjecture.* Let  $n = 2m + k$ , where  $k$  tends slowly to infinity with  $n$ . Let  $\mathcal{A}$  and  $\mathcal{B}$  be subsets of  $\binom{[n]}{m}$  such that the number of pairs  $(U, V)$  with  $U \in \mathcal{A}$ ,  $V \in \mathcal{B}$  and  $U \cap V = \emptyset$  is  $o(1) \binom{n}{m, m, k}$ . Then one can remove  $o(1) \binom{n}{m}$  sets from each of  $\mathcal{A}$  and  $\mathcal{B}$  and end up with no such pairs.

Some promising suggestions have been made for how to go about proving this conjecture. Jozsef (in [comment 124](#)) has drawn attention to a theorem of Bollobás that implies that if each  $U \in \mathcal{A}$  is disjoint from precisely one  $V \in \mathcal{B}$ , then  $\mathcal{A}$  and  $\mathcal{B}$  must both be small. And Ryan (in comments [82](#), [145](#), [150](#) and [152](#)) has drawn attention to existing papers on very closely related subjects. Very recently (at the time of writing it is in the last two or three hours) [Jozsef has given us what appears to be a proof](#) of a pair-removal lemma. Assuming there aren't any irritating surprises, such as it turning out not to be quite the statement we really want, the obvious thing to do here seems to be to play around with Jozsef's argument and see whether it can be generalized from 01-sequences to 012-sequences.

A very quick word about what to expect of such a generalization. It would be "completing the square", where vertex 00 is the [trivial pair-removal lemma](#), vertex 01 is the pair-removal lemma in Kneser graphs, vertex 10 is the normal triangle-removal lemma (as described in the [background-information post](#)), and we are looking for vertex 11. Since getting from 00 to 10 is by no means a triviality, getting from 10 to 11 is unlikely to be a triviality either. But now we are in a stronger position than we were initially, because to get to 11 there are two potential routes. If we are really lucky, then we've got all the ingredients we need and just have to decide how to mix them together.

### Other approaches to Sperner.

The above approach to re-proving Sperner has a certain naturalness to it, at least in the context of the main problem, but it is not the only one to have been discussed. If generalizing it runs into difficulties, then we have other ideas to explore that have



definitely not yet been explored as far as they could have. They begin with [Ryan's comment 57](#) (in which he was responding to [a related comment of Boris](#)). This introduces the idea of what Boris calls “pushing shadows around”, which is apparently what Sperner himself did.

### Other ideas that are still to be explored.

If we are going to follow the triangle-removal proof of the corners theorem quite closely, then at some point we will need to find an analogue of Szemerédi's regularity lemma for subgraphs of the graph where you join two sets if they are disjoint. There are some obvious problems with doing this (discussed in the initial post in comment V, and comments EE to LL), and a very speculative suggestion for how a proof might look in [comment 110](#). There would also be a need for a counting lemma: an [offer to search for such a lemma](#) in a systematic way has not yet been taken up.

*[Remark added 7/2/09. It's important to understand that this post is how the situation seemed to be when it was written yesterday. If you read the comments you will see that some of the assertions that looked plausible (such as the precise form of the tentative conjecture) have now been discovered to be wrong.]*

### Possibly related posts: (automatically generated)

- [Background to a Polymath project](#)

This entry was posted on February 6, 2009 at 10:03 am and is filed under [polymath1](#). You can follow any responses to this entry through the [RSS 2.0](#) feed. You can [leave a response](#), or [trackback](#) from your own site.

## 70 Responses to “DHJ — the triangle-removal approach”

1.  [gowers](#) Says:

[February 6, 2009 at 11:16 am](#) | [Reply](#)

300. One of the main reasons for this comment is to establish the numbering. Comments on Terry's post will be numbered 200-299 (with a convention that from now on if a post receives 100 comments then it must be summarized and restarted), comments on this post will be 300-399, and comments on the obstructions-to-uniformity post, when it is written, will be 400-499.

To get the ball rolling for this post, I'd like to write out Jozsef's proof of pair removal in full detail, mainly for my own benefit so that I understand it properly, but also because if, as I hope, a major subproject for this thread will be to see if we can generalize Jozsef's argument, it will be convenient to have that argument here for easy referral.

The first thing to note is that the result Jozsef proves is not quite as strong as the pair-removal lemma we originally asked for, because what he shows is not that you can remove a small number of sets and end up with no disjoint pairs, but rather that you can remove a small number of sets and end up with removing *almost* all the disjoint pairs you started with. This kind of weakening of a removal lemma is good enough for many applications. For instance, if we applied it to the corners problem we would find that we could remove a small number of edges and get rid of almost all the degenerate triangles, which is just as much of a contradiction as the usual one. (In fact, even if we could show that we could get rid of one percent of the degenerate triangles we would have a contradiction.)

I'll get on to the actual proof in the next comment.

2.  [gowers](#) Says:

[February 6, 2009 at 11:29 am](#) | [Reply](#)

301. Pair removal in Kneser graphs.

The *Kneser graph*  $K_{n,m}$  is the graph whose vertices are all subsets of  $[n]$  of size  $m$ , with two sets  $U$  and  $V$  joined if and only if they are disjoint. We should think of  $n$  as being  $2m + k$  for some fairly small but nonconstant  $k$  but for now I will avoid trying to specify them.

Let  $\mathcal{A}$  and  $\mathcal{B}$  be collections of sets of size  $m$  (i.e., subsets of the vertex set of  $K_{n,m}$ ). We can form a bipartite graph by joining  $U \in \mathcal{A}$  to  $V \in \mathcal{B}$  if and only if  $U \cap V = \emptyset$ : this is not quite a bipartite subgraph of  $K_{n,m}$  (since  $\mathcal{A}$  and  $\mathcal{B}$  are not required to be disjoint), but it is an induced subgraph of what one might call the bipartite version of  $K_{n,m}$  (where each vertex set contains all  $m$ -element sets and you join two such sets if they are disjoint). From now on I will blur the distinction between this bipartite graph and  $K_{n,m}$  itself.

The maximum number of edges that could possibly join  $\mathcal{A}$  to  $\mathcal{B}$  is the number of edges in  $K_{n,m}$  (bipartite version), which is  $\binom{n}{m} \binom{m+k}{m} = \binom{n}{m, m, k}$ . So we shall assume that the total number of edges is at most  $c \binom{n}{m, m, k}$  for some very small constant  $c > 0$ . Our aim is to remove at most  $a \binom{n}{m}$  vertices and get rid of at least half these edges (or any other fixed fraction you might care to go for), where  $a$  tends to zero as  $c$  tends to zero.

I'll discuss how Jozsef does this in the next comment.

3.  [gowers](#) Says:

[February 6, 2009 at 11:45 am](#) | [Reply](#)

302. Pair removal in Kneser graphs.

What Jozsef ends up showing is that there is a very simple way of deciding which vertices to remove: you just get rid of all vertices of large degree. It turns out that if you do this then you either get rid of a significant fraction of all the edges, or you had almost no vertices to start with (in which case you can just remove all of them).

Now all the time we are hoping to remove a number of edges that is large as a fraction of the total number of edges, so “large degree” means “significantly larger than the average degree”. Let us suppose, then, that the average degree in our graph is  $d$ . (Here I am averaging over all vertices in  $K_{n,m}$ , including those that are not in  $\mathcal{A}$  or  $\mathcal{B}$ .) It follows from Markov's inequality that for any  $C$  the number of vertices of degree at least  $Cd$  is at most  $2C^{-1} \binom{n}{m}$ . So we will need to choose  $C$  so that  $2C^{-1}$  tends to zero as  $c$  tends to zero.

Suppose that removing these vertices doesn't get rid of half the edges. That means that after the removal we have a bipartite induced subgraph of  $K_{n,m}$  such that the average degree is at least  $d/2$  and the maximal degree is at most  $Cd$ . In the next comment I'll explain what Jozsef does in this case.

4.  [gowers](#) Says:

[February 6, 2009 at 12:26 pm](#) | [Reply](#)

303. Pair removal in Kneser graphs.

What he does is to use what he calls “the permutation trick of Lubell” to get an upper bound on the number of vertices that there can be in a graph with average degree at least  $d/2$  and maximal degree at most  $Cd$ . Take a random permutation  $\pi$  of  $[n]$ . Given two disjoint sets  $U$  and  $V$ , we shall say that  $\pi$  *separates*  $U$  and  $V$  if every element of  $\pi(U)$  is less than every element of  $\pi(V)$  or vice versa. What is the expected number of separated pairs with  $U \in \mathcal{A}$  or  $V \in \mathcal{B}$ ?

Let's suppose that the total number of vertices in the bipartite graph is  $N$ . Then the total number of disjoint pairs is at least  $(d/2)N$ . Each disjoint pair  $(U, V)$  has a probability of  $2 \binom{2m}{m}^{-1}$  of being separated (since there are  $\binom{2m}{m}$  ways of ordering  $U \cup V$  of which precisely two separate them), so the expected number of separated pairs is at least  $dN \binom{2m}{m}^{-1}$ .

In particular, there exists a permutation  $\pi$  that gives rise to at least  $dN \binom{2m}{m}^{-1}$  separated pairs. But how many separated pairs can there be for a single permutation? Well, let's suppose first that that permutation is the identity (without loss of generality) and that we have separated pairs  $(U_1, V_1), \dots, (U_r, V_r)$ , ordered in such a way that every element of  $U_i$  is less than every element of  $V_i$ . The crucial observation here is that if  $U_1$  is a set with the smallest maximal element out of any of the  $U_i$  (again without loss of generality), then  $(U_1, V_i)$  is a separated pair for all  $V_i$ . Therefore, there can be at most  $Cd$  distinct sets  $V_i$  (since we know that  $U_1$  has degree at most  $Cd$ ). Similarly, there can be at most  $Cd$  distinct  $U_i$ , so the total number of separated pairs is at most  $C^2 d^2$ .

It follows that  $dN \binom{2m}{m}^{-1} \leq C^2 d^2$ , and hence that  $N \leq C^2 d \binom{2m}{m} = C^2 d \binom{n-k}{m}$ . And having got to this point, I realize that I don't understand why the right-hand side is small. Does anyone see it, or do we wait till Jozsef wakes up? (I also have trouble understanding the significance of the final inequality in his comment 155).



5. [gowers](#) Says:

[February 6, 2009 at 12:36 pm](#) | [Reply](#)

304. Pair removal in Kneser graphs.

Maybe the point is that the right-hand side doesn't have to be small if all we know about  $d$  is that it's a small multiple of  $\binom{n+m}{m}$ , but if we assume that  $d$  is much smaller (it's possible that for applications to Sperner we may even be able to take  $d = 1$ ) and  $k$  is reasonably large, then the right-hand side is much smaller than  $\binom{n}{m}$ . So this would fit in with my [comment 153](#), where I suggested that using a stronger hypothesis might be helpful to prove a pair-removal lemma.



6. [gowers](#) Says:

[February 6, 2009 at 12:56 pm](#) | [Reply](#)

305. Triangle removal.

Here is a possible toy problem to think about, as an initial contribution to the project of generalizing Jozsef's proof. This time we let  $n = 3m + k$ . Our basic objects will be pairs  $(U, V)$  of disjoint sets. We are given three collections  $\mathcal{A}$ ,  $\mathcal{B}$  and  $\mathcal{C}$  of such pairs. A *triangle* is a triple of the form  $(U, V), (U, W), (V, W)$  with  $(U, V) \in \mathcal{A}$ ,  $(U, W) \in \mathcal{B}$  and  $(V, W) \in \mathcal{C}$ . We would like to prove that if for each pair  $(U, V) \in \mathcal{A}$  there is at most one  $W$  such that  $(U, W) \in \mathcal{B}$  and  $(V, W) \in \mathcal{C}$ , and similarly for the other two ways round, then it is possible to remove a small number of elements from  $\mathcal{A}$ ,  $\mathcal{B}$  and  $\mathcal{C}$  and end up removing a significant fraction of all the triangles.

There are no degenerate triangles here, so this wouldn't do DHJ straight off, but I think it would be a good way of trying to understand whether Jozsef's methods generalize. An alternative line of investigation would be to generalize Jozsef's methods to a statement about several central layers of the cube rather than just one.



7. [gowers](#) Says:

[February 6, 2009 at 1:16 pm](#) | [Reply](#)

306. Triangle removal.

An optimistic guess about how it might go: we just routinely try to generalize Jozsef's argument, and at a certain point a difficulty arises, which we turn out to be able to solve using the usual triangle-removal lemma.



8. [gowers](#) Says:

[February 6, 2009 at 1:38 pm](#) | [Reply](#)

307. Triangle removal.

At the very least, let's try to do the routine generalization and see what happens. So without any serious thought about what is sensible and what is not, let us say that a permutation  $\pi$  *separates* the three sets  $U$ ,  $V$  and  $W$  if every element of  $U$  comes before every element of  $V$ , which in turn comes before every element of  $W$ , or any one of the other five such statements you get by permuting the three sets. We'll also say that  $\pi$  separates a triangle if it separates the three sets used to build it.

Given any two disjoint sets  $U$  and  $V$ , let us define the  $\mathcal{A}$ -degree of  $(U, V)$  to be the number of  $W$  such that  $(U, V) \in \mathcal{A}$ ,  $(U, W) \in \mathcal{B}$  and  $(V, W) \in \mathcal{C}$ . (In this case, we say that  $U, V, W$  form a triangle.) We have similar definitions for the other two possible degrees.

Now let's suppose that the average degree (over all three bipartite graphs) is  $d$  and the maximal degree is  $Cd$ . What can we say?

Let's think about the expected number of separated triangles. The total number of triangles is  $d \binom{n}{m, m, n-2m} = d \binom{3m+k}{m, m, m+k}$  (up to some absolute constant depending on whether one orders the vertices etc.), and the probability that  $\pi$  separates any given triangle is  $6 \binom{3m}{m, m, m}^{-1}$ . So the expected number is  $6d \binom{3m}{m, m, m}^{-1} \binom{3m+k}{m, m, m+k}$ . Therefore, there must be some permutation  $\pi$  that separates at least this number of triangles.

Now let's think about how many triangles a permutation can separate. Our assumption is that no pair of sets forms a triangle with more than  $Cd$  other sets. It's at this point that, if our wildest dreams come true, we find that normal dense triangle removal is exactly what we need. But no reason to suppose that just yet. I'm going to continue in a new comment.



9. [gowers](#) Says:

[February 6, 2009 at 1:59 pm](#) | [Reply](#)

308. Triangle removal.

Let's suppose that  $(U_1, V_1, W_1), \dots, (U_m, V_m, W_m)$  are separated triangles. We would like an upper bound on  $m$ . If we choose  $i$  such that the maximal element of  $V_i$  is minimized, we see that the number of distinct sets  $W_j$  can be at most  $D = Cd$ . Similarly, the number of distinct  $U_i$  can be at most  $D$ . But each pair  $(U_i, W_j)$  is allowed at most  $D$  sets  $V_h$  in the middle, so we seem to have at most  $D^3$  triangles separated by  $\pi$ . This is slightly worrying as it came out a bit too simply for my liking.

If it's correct, it shows that  $6d \binom{3m}{m, m, m}^{-1} \binom{3m+k}{m, m, m+k} \leq Cd^3$ , and therefore that  $d^2 \geq 6C^{-1} \binom{3m}{m, m, m}^{-1} \binom{3m+k}{m, m, m+k}$ , which is exponential in  $k$ . So for very small  $d$ , such as we might expect to have in DHJ, we get a contradiction.

I've got to go now, so all I'll say at this point is that this argument feels too simple to be correct, but I can't see what's wrong with it. (If it's right, then the explanation must simply be that it is not what we need.)



10. [Ryan O'Donnell](#) Says:

[February 6, 2009 at 3:14 pm](#) | [Reply](#)

309. Pair removal in Kneser graphs.

Can someone help me with this dumb question?

Suppose  $A = B$  are the family of sets not including the last element  $n$ . Then  $A$  and  $B$  have density about  $1/2$  within  $KN_{n, n/2-k/2}$ . (We're thinking  $k(n) \rightarrow \infty$ ,  $k(n)/n \rightarrow 0$  here, right?) It seems that the fraction of Kneser graph edges which are in  $A \times B$  is about  $k/n$ , which is "negligible". So we should be able to delete any small constant fraction of vertices from  $A = B$  and make it an intersecting family. But  $A$  is 100% of the Kneser graph  $KN_{n-1, n/2-k/2}$ , and doesn't the largest intersecting family inside such a Kneser graph have density only around  $1/2$ ?

11.  [gowers](#) Says:

[February 6, 2009 at 4:49 pm](#) | [Reply](#)

310. Pair removal in Kneser graphs.

Ryan, all I can say is that I can't see a flaw in your reasoning. If it's correct, then it shows something rather interesting. I couldn't get Jozsef's argument to work unless I assumed that degrees were very small rather than just small, and your example suggests that it's actually false when you merely assume that the degrees are a negligible fraction of the maximum possible. In your example, the degree of a vertex in  $A$  is still large, even if it's proportionately small, so there's still some hope that Jozsef's argument is correct and adaptable to DHJ.

Now I must stare at the argument that I produced in 308 and try to work out what's going on.

12.  [Ryan O'Donnell](#) Says:

[February 6, 2009 at 5:10 pm](#) | [Reply](#)

311. I guess you can make the degree go down fairly rapidly compared to the density by requiring the last  $t$  coordinates to be 0. (Density goes down like  $2^{-t}$ , fraction of edges goes down like  $(k/n)^t$ .)

13.  [Jason Dyer](#) Says:

[February 6, 2009 at 5:11 pm](#) | [Reply](#)

312. Counting lemma

This is idle wondering, but given a random (not quasi-random) tripartite graph, do we know the probability  $P$  of  $n$  triangles appearing given a density  $D$ ? This seems like this sort of thing that ought to be just a textbook example.

14.  [gowers](#) Says:

[February 6, 2009 at 5:13 pm](#) | [Reply](#)


313. Triangle removal.

I'm going to try to do two things at once. The first is express myself more clearly when it comes to arguments such as the one in 308, and the second is to try to develop an argument that involves more than one slice, in the hope that degenerate triangles will come into play somehow.

Let me begin by trying to define the nicest possible [triangular grid of slices](#). I'll define  $T_m$  to be the union of all  $\Gamma_{a,b,c}$  such that  $(a, b, c)$  is a triple of integers that belongs to the convex hull of  $(n/3 + m, n/3 - m, n/3)$ , ... actually, cancel this. It seems to be more natural, if one is going for the nicest possible and most symmetrical set of slices, to go for a hexagon. So let's take all  $(a, b, c)$  that sum to  $n$  and that satisfy  $\max\{|a - n/3|, |b - n/3|, |c - n/3|\} \leq m$ . Here I'm imagining that  $n$  is a multiple of 3, not that it's all that important. (This is a hexagon because there are six linear inequalities, and the symmetry of the situation makes it a regular hexagon.) Let's write  $\Gamma$  for the union of all these  $\Gamma_{a,b,c}$ . We would like to prove that any dense subset  $A$  of  $\Gamma$  contains a combinatorial line.


Now I'm going to do roughly the same thing as before. I'll assume that there are no combinatorial lines in  $A$ . For each sequence  $x$  in  $A$  and each  $j \in \{1, 2, 3\}$ , let's define the  $j$ -set of  $x$  to be  $\{i : x_i = j\}$ , and we'll denote this by  $U_j(x)$ . As ever, we define  $\mathcal{A}$  to be the set of all pairs  $(U_1(x), U_2(x))$  with  $x \in A$ ,  $\mathcal{B}$  to be the set of all  $(U_1(x), U_3(x))$  and  $\mathcal{C}$  to be the set of all  $(U_2(x), U_3(x))$  (again, always with  $x \in A$ ).

The assumption that there are no combinatorial lines tells us that there are no triples  $(U, V, W)$  with  $(U, V) \in \mathcal{A}$ ,  $(U, W) \in \mathcal{B}$  and  $(V, W) \in \mathcal{C}$ , except in the degenerate case where  $U \cup V \cup W = [n]$ . In the next comment I'll try to hit this with Jozsef's permutation argument.

15.  [Jason Dyer](#) Says:  
[February 6, 2009 at 5:15 pm](#) | [Reply](#)

(Meta-comment)

Ryan, like the brackets problem you should toss a  $\{ \}$  before the 0 there if you want to render it in LaTeX in WordPress.

16.  [gowers](#) Says:  
[February 6, 2009 at 5:31 pm](#) | [Reply](#)


314. Triangle removal.

So now let's choose a random permutation and see what we can say about the number of separated triangles. To do this, my first task is to say how many *actual* triangles there are. And the answer is  $|A|$ , one degenerate triangle for each element of  $A$  and no others.

Next, I'd like some idea of the probability that a random permutation  $\pi$  separates a given triangle. Let's suppose that the point  $x$  of  $A$  that gives rise to the triangle belongs to  $\Gamma_{a,b,c}$ . Then the probability that  $\pi$  separates  $(U_1(x), U_2(x), U_3(x))$  is  $6 \binom{n}{a,b,c}^{-1}$ . So the expected number of separated triangles is  $6|A| \binom{n}{a,b,c}^{-1}$ .

Now suppose I just choose an arbitrary permutation — without loss of generality the identity. How many separated triangles can there be? I think this may be where the problems arise, but let's see. Suppose that  $(U_1, V_1, W_1), \dots, (U_s, V_s, W_s)$  are separated triangles. We know that for each  $(U_i, V_i) \dots$

I may come back to this, but I've now seen where I went wrong in 308, so I should think about that first, as it's a simpler problem.

17.  [gowers](#) Says:  
[February 6, 2009 at 5:49 pm](#) | [Reply](#)

315. Triangle removal: my mistake in 308.

I started 308 with the claim that if you choose  $(U_i, V_i)$  so as to minimize the maximum element of  $V_i$ , then you could see immediately that there were at most  $D$  distinct  $W_j$ . The thinking behind this was that each  $W_j$  would form a triangle.

But that was wrong: there is no reason for  $(U_i, W_j)$  to belong to  $\mathcal{B}$  or for  $(V_i, W_j)$  to belong to  $\mathcal{C}$ . So it's back to the drawing board here. (But this, I should stress, is *good* news as I wanted there to be a complication to give triangle-removal a chance to creep in.)

Here is a new proof strategy. First I'll copy the first part of Jozsef's argument to get to the stage where each *vertex* is contained in roughly the same number of triangles. (If I can't, I'll remove a few vertices and get rid of a significant fraction of the triangles.) Next, I'll argue as above that the expected number of separated triangles is reasonably large. Then I'll try to bound the number of separated triangles if no vertex is in too many triangles and no edge is in more than one triangle. The first step of that will be to pick ... actually, I still don't see it.

I was going to say that one should pick the  $U_i$  with smallest maximum, but I don't see what that gives us.

What we *can* say is this. For each  $U$  there will be some system of sets  $(V_i, W_i)$  such that  $(U, V_i, W_i)$  forms a triangle. If in this set system we choose  $i$  such that the maximum of  $V_i$  is minimized, then we can deduce that there are not too many distinct  $W_j$ . Except that I don't know why I even say *this* so confidently: what makes me think that  $(V_i, W_j)$



should lie in  $\mathcal{C}$ ? OK I'm floundering around a bit here so I'd better stop. But I'll leave these messy thoughts here in case anyone can get a clearer idea of what ought to be done. It feels as though it's in promising territory somehow.

18.  [gowers](#) Says:

[February 6, 2009 at 6:14 pm](#) | [Reply](#)

316. Triangle removal.

This is the kind of reason I find it promising. The graph where you join two sets if and only if one is completely to the left of the other is, under normal circumstances, much denser than the graph where you just join two sets if they are disjoint. So what Jozsef's permutation argument is cleverly achieving seems to be to get us from a sparse context to a dense context. (I don't yet have a fully precise formulation of this remark.)

Now let's suppose we have our separated triangles  $(U_i, V_i, W_i)$  above. Suppose also that we have managed to show somehow that many of the pairs  $(U_i, V_j)$  belong to  $\mathcal{A}$ , and so on. We know that there won't be many triangles, so this will imply, by the usual triangle-removal lemma, that we can remove a small number of edges and end up with no triangles. Could there be an argument along these lines that allows us to prove density Hales-Jewett by hitting it permutation by permutation? We're in that familiar state where there's a small probability  $p$  of being close to a solution and a large probability  $1 - p$  of having written a load of nonsense.

19.  [jozsef](#) Says:

[February 6, 2009 at 6:33 pm](#) | [Reply](#)

317. Pair removal in Kneser graphs.

Bad news - good news. Good morning! I hope that I didn't derail the project with my last calculation yesterday - that the "removal lemma" works when  $m \sim n - \sqrt{n}$ . The main inequality is correct, and we see that it tells us something non-trivial if  $m$  is small compare to  $n$ . As I see now, the average degree,  $d$ , grows as  $m = o(n - \sqrt{n})$ . I think that this is the range where we have a removal lemma so far. This is a removal lemma we get by simple double-counting, so I'm afraid that we can't expect too much out of it, but it supports Tim's conjecture.

20.  [gowers](#) Says:

[February 6, 2009 at 7:15 pm](#) | [Reply](#)

318. Pair/triangle removal.

Hi Jozsef. Two remarks. It seems to me (but I'd be reassured if I knew you agreed about this) that we have a removal lemma if the average degree is very small. And in interesting contexts it is. For example, if  $\mathcal{A}$  is a counterexample to Sperner's theorem and we let  $\mathcal{B}$  consist of all complements of sets in  $\mathcal{A}$ , then each set in  $\mathcal{A}$  is joined just to its complement, so the average degree is 1.

The second remark is that it's not impossible that a simple double counting gives us something good. The reason is that when going from 2 to 3 in the dense case you go from trivial to interesting, so when going from 2 to 3 here, perhaps you go from simple to very interesting.

21.  [Ryan O'Donnell](#) Says:

[February 6, 2009 at 7:25 pm](#) | [Reply](#)

319. Sperner.

Hi Jozsef, is that a typo?  $m$  needs to be less than  $n/2$ ; otherwise the  $KN_{n,m}$  Kneser graph is empty.

I'm not sure the scenario we should consider now... we have counterexamples to the Tentative Conjecture for  $k$  both constant and superconstant.

22.  *jozsef* Says:

[February 6, 2009 at 7:28 pm](#) | [Reply](#)

320 Pair removal lemma

Now I feel a bit dummy; I wanted to check where did the calculation go wrong yesterday in my last post, but I can't see. Tim, in 304 you asked why is the right hand side small. By checking  $\binom{n}{m} / \binom{2m}{m}$  what I see - again - is that it should be around  $2^{2\sqrt{n}}$ . I probably badly overlook the same thing again ...  
I'll go to have a morning coffee.

23.  *gowers* Says:

[February 6, 2009 at 7:34 pm](#) | [Reply](#)

321. Sperner

Ryan, you may be ahead of me here and already see why the suggestion I've been making is a bad one, but what about strengthening the hypothesis of the Tentative Conjecture from the number of disjoint pairs being  $o(1)\binom{n}{m,k}$  to each set belonging to at most one disjoint pair? (With this second hypothesis one tries to prove that the sets must be small.)

Actually, isn't that just Bollobás's theorem? I think it is because if I take  $d = C = 1$  in 303 then I get the correct bound of  $\binom{2m}{m}$ . Is that the standard proof of Bollobás's theorem?

24.  *Jason Dyer* Says:

[February 6, 2009 at 7:35 pm](#) | [Reply](#)

322. Pair removal in Kneser graphs.

Ok, I am trying to understand Jozsef's proof (I notice it uses the greedy algorithm, interesting) and I must be missing something elementary at 303:

The crucial observation here is that if  $U_1$  is a set with the smallest maximal element out of any of the  $U_i$  (again without loss of generality), then  $(U_1, V_i)$  is a separated pair for all  $V_i$ .

Could someone expand on this a little more? I'm not seeing why.

25.  *gowers* Says:

[February 6, 2009 at 7:37 pm](#) | [Reply](#)

323. Pair removal in Kneser graphs.

Jason, if  $U_1$  has the smallest maximal element and  $(U_i, V_i)$  is separated, then the largest element of  $U_1$  is at most the largest element of  $U_i$  is less than the smallest element of  $V_i$ . If this doesn't explain it then I don't understand what you're not understanding.

26.  *gowers* Says:

[February 6, 2009 at 8:08 pm](#) | [Reply](#)

324. Triangle removal

A quick technical question: I think it should be reasonably straightforward to answer. Suppose we have a dense set  $A$  in  $[3]^n$ . Suppose now that we take a random permutation  $\pi$  and just take from  $A$  the sequences that start with 1s, continue with 2s and finish with 3s (after you've used  $\pi$  to scramble the coordinates). And now suppose you form the tripartite graph from just those sequences. That is, you let  $\mathcal{A}$  consist of all pairs  $(U, V)$  that are equal to  $(U_1(x), U_2(x))$  for some  $x \in A$ , and so on. We can think of  $\mathcal{A}$  as a bipartite graph, whose vertices are subsets of  $[n]$ , where we join  $U$  to  $V$  if  $U$  finishes before  $V$  starts. What I'd like to know is whether this graph tends to be dense. I think the straightforward answer is no, so what I really mean is whether it can be made dense if you remove a smallish number of vertices and edges, or something like that. The motivation for this question is that I'd very much like to be able to define a dense tripartite graph with few triangles and see what happens when we apply the usual triangle-removal lemma.

27.  *jozsef* Says:

[February 6, 2009 at 8:30 pm](#) | [Reply](#)

325. Pair removals

Re: 319, yes there was a typo, I meant  $n/2$  there, thanks. But I still don't see if anything is bad with my calculation; Please help me, is it correct that if  $m = n/2 - \sqrt{n}$  then  $d$  should be at least  $c2^{2\sqrt{n}}$  or we have a removal lemma? After this I will try to digest Tim's triangle removal technique.

28.  *Ryan O'Donnell* Says:

[February 6, 2009 at 9:29 pm](#) | [Reply](#)

326. Kneser.

Hi Tim, in #321, if each set is involved in at most one Kneser edge, then I think the least number of sets you need to delete in order to eliminate all Kneser edges is with a factor of two of the number of sets (vertices), no?

29.  *gowers* Says:

[February 6, 2009 at 9:37 pm](#) | [Reply](#)

327. Kneser

Yes, so if you can prove that you can remove a small number of sets then you've proved that there weren't very many sets in the first place.

30.  *gowers* Says:

[February 6, 2009 at 9:43 pm](#) | [Reply](#)

328. Pair removal

Jozsef in 325, my calculation (which I checked and I'm pretty sure it works out to be the same as yours) suggests that you get something interesting as long as  $C$  is large and  $C^2 d \binom{2m}{m}$  is a lot smaller than  $\binom{2m+k}{m}$ . Since  $\binom{2m+k}{m}$  is about  $2^k \binom{2m}{m}$ , it follows that we need  $d$  to be small compared with  $2^k / C^2$  in order to get a removal lemma. I think this agrees with you, since you are taking  $m = n/2 - \sqrt{n}$  and  $k = 2\sqrt{n}$ . So it seems correct to you and it seems correct to me, which means that there's an  $\epsilon^{3/2}$  chance that it's incorrect.

31.  *jozsef* Says:

[February 6, 2009 at 9:53 pm](#) | [Reply](#)

329. Thank you Tim. I think I should apologize that, after reading 303, I didn't pay much attention to later posts from where I could have learned that things were OK with the calculation. In the next post I will try to understand and

extend your toy version of a triangle removal lemma. Unfortunately now I have to go to the university and probably I won't read the posts in the next 5-6 hours.

32.  [gowers](#) Says:

[February 6, 2009 at 11:11 pm](#) | [Reply](#)

### 330. Triangle removal

Jozsef, I should make it clear that I don't yet have an interesting new triangle removal lemma proved, so if you prove anything at all it counts as an extension of what I've done. But I might think just a little bit further about the vague idea of 324.

Suppose then that we have a dense set of sequences in  $[3]^n$ , and for simplicity let's assume that they're all reasonably well balanced (meaning that they all have roughly equal numbers of 1s, 2s and 3s). How many times do we expect a given set  $U$  to occur as  $U_1(x)$  for some  $x \in A$ ? The very rough answer is  $\binom{2n/3}{n/3}$ . Let's be even more rough and call that  $2^{2n/3}$ . Now  $A$  has size around  $\binom{n}{n/3, n/3, n/3} = \binom{n}{n/3} \binom{2n/3}{n/3}$ .

I'm going to interrupt this comment because I've had a different idea that I want to pursue urgently.

33.  [gowers](#) Says:

[February 6, 2009 at 11:52 pm](#) | [Reply](#)

### 331. Sperner and DHJ

The proof I know of Sperner goes like this. You take a random sequence of sets of the form  $\emptyset = A_0 \subset A_1 \subset A_2 \subset \dots \subset A_n = [n]$ , where  $A_i$  has cardinality  $i$ . If  $\mathcal{A}$  is an antichain, then it intersects each such chain in at most one set.

Now let's imagine we choose a set at random. What is the probability that it lies in  $\mathcal{A}$ ? Let's write  $\delta_i$  for the density of  $\mathcal{A}$  inside layer  $i$  (that is, the number of sets in  $\mathcal{A}$  of size  $i$  divided by  $\binom{n}{i}$ ). Then the expected size of the intersection of  $\mathcal{A}$  with our random chain is  $\delta_0 + \delta_1 + \dots + \delta_n$ . It follows that the  $\delta_i$  add up to at most 1 (since at least one random chain will have at least the expected number of elements of  $\mathcal{A}$  in it). Since the middle layer(s) is (are) biggest, it's clear that we maximize the size of  $\mathcal{A}$  by choosing a middle layer.

The idea that's just occurred to me is that we've just deduced Sperner from the one-dimensional corners result (which states that if you have a dense subset of  $[n]$  then you must be able to find  $x$  and  $x + d$  in it with  $d \neq 0$ ). What happens if we try to deduce density Hales-Jewett from 2-dimensional corners?

Well, first we need to decide what replaces a random chain. For reasons of symmetry I'd like to think of a random chain as follows: you pick a random permutation of  $[n]$  and then once you've scrambled the set you take all partitions  $(A, B)$  of  $[n]$  into two sets such that every element of  $A$  is less than every element of  $B$ .

At this point I'm just going to guess that it might be reasonable to choose for DHJ purposes a random permutation of  $[n]$  followed by all partitions  $(A, B, C)$  such that every element of  $A$  comes before every element of  $B$ , which comes before every element of  $C$ . This is at least a two-dimensional structure, since it's determined by the end points of  $A$  and  $B$ . I will also think of this as the set  $S$  of all sequences in  $[3]^n$  that start with some 1s, continue with some 2s, and end with some 3s.

Next question: if  $\mathcal{A}$  contains no combinatorial line, can we deduce that the intersection of  $\mathcal{A}$  with  $S$  is small? What I'd like to do is prove that it has no corners, in some useful sense of the word "corners".

How would we get a combinatorial line in this set? We can of course encode each element of  $S$  as a triple  $(a, b, c)$  where  $a + b + c = n$ . And at this point we have a problem:  $S$  contains no combinatorial lines.

Well, I suppose if that had worked it would have been discovered by now. But it suggests an amusing problem. Let  $T_m$  be the hexagonal portion of the triangular grid  $x + y + z = 0$  (a 2-dimensional subset of  $\mathbb{Z}^2$ ) defined by  $\max\{|x|, |y|, |z|\} \leq m$ . Let us call a function  $f : T_m \rightarrow [3]^n$  a *corner isomorphism* if it is an injection such that every aligned equilateral triangle in  $T_m$  maps to a combinatorial line. (An aligned equilateral triangle is a set of the form  $(a + r, b, c), (a, b + r, c), (a, b, c + r)$ .) If we could find a corner isomorphism for some fairly large  $m$ , then we could probably move the image about and prove density Hales-Jewett using an averaging argument. So can anyone either find a big corner isomorphism or (more likely) prove that no such function exists? I don't expect this problem to be very hard.

34.  [gowers](#) Says:

[February 7, 2009 at 1:48 am](#) | [Reply](#)

332. Sperner and DHJ

My last contribution for the day: I've just checked and there isn't even a corner isomorphism from  $T_2$  into  $[3]^n$ . The proof goes like this: you take an arithmetic progression of three points in a line and map them to 12333, 12233 and 12223. (I think I may also have wanted to insist that lines go to sets where the appropriate  $j$ -set is fixed.) Those three points determine a  $T_2$  and their images determine what the images of the other three points under a corner isomorphism would have to be, but those three images don't lie in a combinatorial line. (For what it's worth, they are 12133, 12113 and 12213.)

35.  [jozsef](#) Says:

[February 7, 2009 at 6:49 am](#) | [Reply](#)

333. Triangle removal

To understand the notations and arguments below, one should read 305-307-308 first. Tim's proof in 307-308 works nicely, if we change the initial conditions. Our base set  $\mathcal{S}$  is now just a collection of  $m$ -element subsets (where  $n=3m+k$ ) instead of  $\mathcal{A}, \mathcal{B}, \mathcal{C}$  and we are considering now every triangle formed by any three pairwise disjoint elements of  $\mathcal{S}$ . For this set of triangles, we do have the removal lemma as Tim has calculated it. Let me add two technical remarks which might be needed for the complete proof. A triangle is called "normal" if none of its edges is edge of more than  $Cd$  other triangles. Then, the argument gives bound on the number of normal triangles. If most of the triangles aren't normal, then we have a removal lemma. The conclusion is that if we have a small number of triangles in  $\mathcal{S}$  compare to the size of  $\mathcal{S}$ , then a small fraction of the edges covers most of the triangles. My second small remark is that when we bound the normal triangles separated by a permutation, then we consider the sets right from the leftmost (having the smallest last point) middle element and and the sets left from the rightmost (having the largest first point) element. These are both bounded by  $Cd$  and they span at most  $(Cd)^3$  separated triples, as Tim said. I'm just emphasizing the selection of the left size here, only a technical detail.

This theorem is more restricted than Tim's original statement, but I think it might be useful. In DHJ the dense subset is given and we will work with the induced substructure.

36.  [jozsef](#) Says:

[February 7, 2009 at 7:01 am](#) | [Reply](#)

334. Sperner

I was wondering what would be the proper Sperner-type theorem in  $[3]^n$  similarly useful as Sperner in  $[2]^n$ . Is there something before DHJ?

37.  [gowers](#) Says:

[February 7, 2009 at 10:22 am](#) | [Reply](#)

## 335. Sperner

Jozsef, I think I have the answer to your last question. Let's define the *Kneser product* of two set systems  $\mathcal{A}$  and  $\mathcal{B}$  to be the set of all pairs  $(U, V)$  such that  $U \in \mathcal{A}$ ,  $V \in \mathcal{B}$  and  $U \cap V = \emptyset$ . Then the theorem would be that every Kneser product that is dense in the set of all disjoint pairs (of which there are  $3^n$ ) contains a corner: that is, a configuration of the form  $(U, V)$ ,  $(U, V \cup D)$ ,  $(U \cup D, V)$  with all three of  $U$ ,  $V$  and  $D$  disjoint.

I've stated it like that to make it as similar as I can to Sperner. But a more symmetrical version is this. Suppose that  $\mathcal{S}$ ,  $\mathcal{T}$  and  $\mathcal{U}$  are set systems such that there are at least  $c3^n$  ways of partitioning  $[n]$  as  $S \cup T \cup U$  with  $S \in \mathcal{S}$  etc. Then the corresponding collection of sequences contains a combinatorial line.

To see that this is a natural extension of Sperner, consider the following silly way of stating (weak) Sperner. If  $\mathcal{A}$  and  $\mathcal{B}$  are two set systems such that there are at least  $c2^n$  ways of partitioning  $[n]$  as  $A \cup B$  with  $A \in \mathcal{A}$  and  $B \in \mathcal{B}$ , then the corresponding collection of 01-sequences contains a combinatorial line.

It looks to me as though there should be a two-dimensional family of density Hales-Jewett statements. That is, for each  $1 \leq j < k$  there should be a DHJ(j,k), where Sperner is DHJ(1,2), density Hales-Jewett in  $[k]^n$  is DHJ(k-1,k), and the theorem just discussed (for which it looks as though we may have a proof) is DHJ(1,3). The idea would be that for DHJ(j,k), whether or not a sequence belongs to the set depends on a j-uniform hypergraph with  $2^{[n]}$  as its vertex set.

We should see if there is any relation between this and [Terry's DHJ\(2.5\) problem](#). I suspect there isn't, in which case it's amusing to have come up with a different way of "interpolating" between Sperner and DHJ.



38. [gowers](#) Says:

[February 7, 2009 at 2:32 pm](#) | [Reply](#)

## 336. Sperner and DHJ.

A small observation. One could regard the previous comment as showing that at least one natural generalization of Sperner to  $[3]^n$  turns out not to be the statement we want. That is unfortunate, though by no means catastrophic. In this comment I want to point out that there is a different way of generalizing Sperner that leads (I think) to a deeper theorem, though it still isn't DHJ.

Let's go back to comment 331 and try to generalize the proof of Sperner and see what comes out, not worrying if it isn't DHJ. To do this, we associate with any triple  $a + b + c = n$  the three sets  $\{1, 2, \dots, a\}$ ,  $\{a + 1, \dots, a + b\}$  and  $\{a + b + 1, \dots, n\}$ . Let's call this the triple  $U_{a,b,c}$ . Now let's suppose we have a corner, by which in this context I mean a triple of triples of the form  $(a + r, b, c)$ ,  $(a, b + r, c)$  and  $(a, b, c + r)$  (with  $a + b + c + r = n$ ). What combinatorial configuration do we get from the corresponding triples  $U_{a+r,b,c}$ ,  $U_{a,b+r,c}$  and  $U_{a,b,c+r}$ ? The answer is that we partition  $[n]$  into sets  $A_1 = \{1, 2, \dots, a\}$ ,  $A_2 = \{a + 1, \dots, a + r\}$ ,  $A_3 = \{a + r + 1, \dots, a + b\}$ ,  $A_4 = \{a + b + 1, \dots, a + b + r\}$  and  $A_5 = \{a + b + r + 1, \dots, n\}$ . Then the sequence corresponding to  $U_{a+r,b,c}$  is 1 on  $A_1$  and  $A_2$ , 2 on  $A_3$  and  $A_4$ , and 3 on  $A_5$ . The sequence corresponding to  $U_{a,b+r,c}$  is 1 on  $A_1$ , 2 on  $A_2$ ,  $A_3$  and  $A_4$ , and 3 on  $A_5$ . Finally, the sequence corresponding to  $U_{a,b,c+r}$  is 1 on  $A_1$ , 2 on  $A_2$  and  $A_3$ , and 3 on  $A_4$  and  $A_5$ .

Thus, we have three fixed sets, namely  $A_1$ ,  $A_3$  and  $A_5$ , and two variable sets, namely  $A_2$  and  $A_4$ . If we identify each set to a point, we can think of the three sequences as 11223, 12223 and 12233. If we forget about the fixed sets, the restrictions to the variable sets are going 12, 22, 23. Also, and this is important, the two variable sets have the same size. (If they didn't, then the resulting theorem would be much easier.)

Given a dense set of triples that start with 1s, continue with 2s and end with 3s, then the corners theorem will tell you that you must have a configuration of the above kind. I haven't checked this, but I'm about 99% certain that if you imitate the proof of Sperner by averaging over a random permutation, then you will be able to deduce that a dense subset of  $[3]^n$  will contain a "peculiar combinatorial line" of the above form. (Just to be clear, a *peculiar combinatorial line* is a triple of sequences that are fixed outside two sets  $V$  and  $W$  that have the same size, and inside  $V$  and  $W$  they



go 12, 22, 23.)

The reason I call this a deeper theorem is that it relies on the corners theorem. Indeed, it implies the corners theorem if you restrict to sets that are unions of slices. (This is because we insist that  $V$  and  $W$  have the same size.)

Again, this is a slightly depressing observation because it shows that a reasonably natural generalization of the Sperner proof gives us the wrong theorem. However, again it doesn't actually tell us that we can't find a different way of generalizing Sperner that gives the right theorem.

39.  [gowers](#) Says:

[February 7, 2009 at 4:05 pm](#) | [Reply](#)

### 337. Sperner and DHJ

A tiny observation, following on from 332. The main difficulty seems to be as follows. Let's define the *order profile* of a sequence to be what you get when you collapse each block of equal coordinates to just one coordinate. For example, the order profile of 113323311113 is 132313. In  $[2]^n$  you can have a combinatorial line where every element has the same order profile: e.g. 11122 and 12222. But in  $[3]^n$  that is impossible. (Proof: if the last fixed coordinate before a variable string starts is  $i$  and the first fixed coordinate after it finishes is  $j$ , then the variable string can only take the values  $i$  and  $j$ .)

I've vaguely wondered about having a two-dimensional ground set, so that it's possible for three sets to meet at a point, but I haven't managed to define a nice two-parameter family of sequences.

40.  [gowers](#) Says:

[February 7, 2009 at 4:13 pm](#) | [Reply](#)

### 338. Sperner and DHJ

Actually, I can't resist just slightly expanding on the last paragraph of 337. Let's take not  $[3]^n$  but  $[3]^{n^2}$ , which we think of as the space of all functions from  $[n]^2$  to  $\{1, 2, 3\}$ . I'd like to regard such a function as "simple" if you've basically got three big simply connected patches, of 1s, 2s and 3s, and if the boundaries between these patches are not too wiggly. A combinatorial line of functions like this would have the desirable property that it wouldn't be obvious from just one of the functions what the variable set was: in some sense the "two-dimensional order profile" would be the same for all three points.

One could now ask whether the simpler structure here would make DHJ easier to prove. If so, then one could average up in the Sperner fashion. But it doesn't look easy to say exactly what this simpler structure is.

41.  [gowers](#) Says:

[February 7, 2009 at 4:58 pm](#) | [Reply](#)

### 339. Sperner and DHJ

I take that back. Here's an actual conjecture.

Let's take as our basic object the kind of thing one talks about in Sperner's proof of the Brouwer fixed point theorem (surely a coincidence though). Chop out a large hexagonal piece from a tiling of the plane by hexagons and colour two of its edges (next to each other) red, two green and two blue, and extend these colourings from the boundary into the inside of the hexagon until you've coloured it all, and do so in such a way that you have three simply connected regions that meet at some point. Colourings of this kind are the basic objects that we'll consider. Of course, "red", "green" and "blue" is another way of saying 1, 2 and 3.

The question is, if you have a dense set of such colourings, must you have a combinatorial line? A combinatorial line in this context will be like what I described except that the three regions meet not at a point but round the boundary of a fourth region, which is the set of variable coordinates.

At the moment I know pretty well nothing about this question. I don't know whether there's an easy counterexample, I don't know how to think about the measure on all those colourings (which feels like a fairly nasty object of a kind that the late and much lamented Oded Schramm would have been able to tell us about), and I don't know whether a positive answer follows from density Hales-Jewett. The one thing I feel fairly confident of is that this, if true would *imply* DHJ. Oh, and one other thing is that I don't have any feel for whether this is more or less equal to DHJ in difficulty, or whether there is a chance that it is easier.

Actually, there's another thing that I'm confident of — in fact, certain of — which is that it is a natural generalization of the trivial fact that underlies Sperner. (That is the fact that if you have a dense set of 2-colourings of a line of points, then two of those 2-colourings will form a combinatorial line.)

This conjecture is probably not to be taken too seriously, but I quite like the fact that it doesn't follow easily from DHJ.



42. *Sune Kristian Jakobsen* Says:

[February 7, 2009 at 7:09 pm](#) | [Reply](#)

340. Variant of Sperner

First I must admit, that I haven't read much in this thread, so I don't know if this has been said before. I got this idea while I tried to find some low values of  $c_n$ .

In 331 Gowers gives a proof that if  $\mathcal{A}$  is an antichain,  $\sum_{A \in \mathcal{A}} \frac{1}{\binom{n}{|A|}} \leq 1$ . Using the same idea it is possible to show that if  $\mathcal{A}$  is a family, that intersects every chain

$\emptyset = A_0 \subset A_1 \subset A_2 \subset \dots \subset A_n = [n]$ , then  $\sum_{A \in \mathcal{A}} \frac{1}{\binom{n}{|A|}} \geq 1$ . Notice that when the elements of  $\mathcal{A}$  is only allowed to have one of two given sizes, these two theorems are equivalent. The later theorem seems to be easier to generalize to the  $k=3$  case, and it might be useful, at least in the 200-thread.



43. *jozsef* Says:

[February 7, 2009 at 9:56 pm](#) | [Reply](#)

341. Sperner

I'm reading 335, and I'm trying to understand the conjecture. First, there is a coloring theorem for subsets that says that one can find three disjoint sets in the same colour class that their unions are also in that colour class. It is a special case of the Finite Union Theorem (Folkman's Theorem) which is equivalent to the finite sums theorem. It states that any finite colouring of the natural numbers contains arbitrary large monochromatic IP sets. The density statement is clearly false without extra conditions on the dense subset. It is obvious that the finite union theorem implies the finite sum theorem, however the other direction is difficult; one can use Van der Waerden for the reduction. Maybe we can also prove some density theorem for disjoint set unions using Szemerédi's thm. It is still not Tim's cross-product conjecture, but it seems to be quite promising. Let me also mention that the other proof of finite union theorem uses HJ! About the details: a possible reference is Graham, Rothschild, Spencer, "Ramsey Theory". Now it's family lunch-time, so I will be away for a while, but I'm quite excited about this possible "density finite union theorem" so I will come back a.s.a.p.



44. *gowers* Says:

[February 7, 2009 at 10:42 pm](#) | [Reply](#)

342. Sperner

Jozsef, What I wrote in 335 was meant to be more or less the same as what you suggested in 333 — a very slight generalization perhaps, but you can make it the same if you look at the special case  $\mathcal{S} = \mathcal{T} = \mathcal{U}$ . And I think that, as you say, it can be proved by the argument of 308. Or are we talking about different things?

45. [Boris](#) Says:

[February 7, 2009 at 11:37 pm](#) | [Reply](#)

343. In comment 335 what is meant in general by DHJ(j,k)?

46.  [gowers](#) Says:

[February 8, 2009 at 12:02 am](#) | [Reply](#)

344. Sperner/DHJ

Boris, I hadn't actually worked it out, but I think I can. Basically, DHJ(j,k) is density Hales-Jewett in  $[k]^n$  but with  $j$  placing a restriction on the complexity of the dense set  $\mathcal{A}$ . Let me give the case  $j = k - 1$  first. This is equivalent to the normal density Hales-Jewett theorem, but I'll state it slightly differently.

Actually, I think I may as well go for the whole thing right from the start.

Suppose that for each subset  $E$  of  $[k]$  of size  $j$  you have a set  $\mathcal{A}_E$  of functions from  $E$  to the power set of  $[n]$  such that the images of all the points in  $E$  are disjoint sets. Let's write  $(U_j : j \in E)$  for a typical element of  $\mathcal{A}_E$ . I now define a set  $\mathcal{A}$  of (ordered) partitions of  $[n]$  into  $k$  sets  $(U_1, \dots, U_k)$  by taking all partitions  $(U_1, \dots, U_k)$  such that for every  $E$  of size  $j$  we have that  $(U_j : j \in E)$  belongs to  $\mathcal{A}_E$ . Thus, in a certain sense the set of partitions (which are of course in one-to-one correspondence with sequences in  $[k]^n$ ) depends only on what collections of  $j$  sets there are in the partition. Then DHJ(j,k) is the claim that if your dense set  $\mathcal{A}$  has this special form, then it must contain a combinatorial line.

My suggestion, which I think I could justify completely if pushed to do so, is that if  $j = 1$  then one can prove DHJ(j,k) by a fairly simple adaptation of the proof of Sperner as given by Jozsef. Sperner, by the way, is DHJ(1,2), and the problem we have been considering is DHJ(2,3). In general, I suspect that DHJ(j,k) is of roughly equal difficulty to DHJ(j,j+1). It might be interesting to see if we can deduce DHJ(j,k) from DHJ(j,j+1), though at the moment my feeling is that we can prove DHJ(1,k) not by deducing it from DHJ(1,2) but by imitating the proof of DHJ(1,2), which is not quite the same thing.

47.  [jozsef](#) Says:

[February 8, 2009 at 12:03 am](#) | [Reply](#)

345. Triangle Removal in Kneser Graphs

I will go back to the "density finite union theorem" soon as I think it is relevant to our project, but first here is a general remark about the removal lemma. I just want to make it clear that in a certain range we have a real removal lemma. Given three families of sets  $A, B$ , and  $C$ . If the number of disjoint triples (one element from each) is smaller than  $X$  (something depending on the sizes  $|A|, |B|$ , and  $|C|$ ) then a small fraction of the disjoint pairs covers EVERY triangle. To see that, we just have to iterate the counting of "normal" triangles. (post 333.)

48. [Boris](#) Says:

[February 8, 2009 at 12:17 am](#) | [Reply](#)

346. I think I misunderstand comment 344. Consider DHJ(k-1,k). We have  $k$  sets  $\mathcal{A}_E$  each of which gives a rise to a subset of  $[k]^n$  which I call  $B_E$ . Then the subset  $B$  of  $[k]^n$  that corresponds to  $\mathcal{A}$  that is the intersection  $\bigcap_E B_E$ . However, nothing stops us from make  $B_E$  large, and their intersection empty.

49.  [gowers](#) Says:

[February 8, 2009 at 12:33 am](#) | [Reply](#)

347. Boris, I wasn't clear enough in what I meant. Of course, you are right in what you say, but the statement is that *if* the intersection is dense then you get a combinatorial line, and not the weaker assumption that each individual  $B_E$  is dense. Of course, in the case  $j = k - 1$  you can consider just a set system  $B_{[k-1]}$ , just as in Sperner you can consider just the set of points where your sequence is 0.

50.  [jozsef](#) Says:

[February 8, 2009 at 1:22 am](#) | [Reply](#)

348. Density Finite Union (DFU) Theorem

My previous remark was about a possible DFU thm. Let us see the simplest version. What would be a necessary density condition to guarantee a pair of disjoint sets  $A$  and  $B$  that  $A, B$ , and  $A \cup B$  are in this "dense" set? Considering the arithmetic analogue, if we have a set  $S \subset [n]$  with  $cn$  elements  $a$  that  $2a$  is also in  $S$ , then there is a nontrivial solution for  $x+y=z$ .

51.  [jozsef](#) Says:

[February 8, 2009 at 1:37 am](#) | [Reply](#)

349. DFU contd.

The first density condition would be that the number of disjoint pairs is at least  $c3^n$ . The second condition should be something related to sets having  $2/3n+k$  elements. I'm still not sure what could be a meaningful notation for density here.

52.  [jozsef](#) Says:

[February 8, 2009 at 2:03 am](#) | [Reply](#)

350. DFU

A possible statement could be the following: Given a family  $F$  of subsets of  $[n]$ . Let us suppose that the number of disjoint pairs in  $F$  is at least  $c3^n$ . We are going to define a graph on  $F$ . The vertices are the sets in  $F$  and two are connected iff they are disjoint and their union is also in  $F$ . Then, either one can remove  $o(|F|)$  vertices to destroy all edges, or there are  $c'3^n$  edges where  $c'$  depends on  $c$  only. I'm not sure that the statement is true, but something like this should be true.

53.  [jozsef](#) Says:

[February 8, 2009 at 2:14 am](#) | [Reply](#)

351. DFU (Density Finite Union) theorem.

Well, now I see that  $c'3^n$  is way too ambitious; let me change it to  $f(n)|F|$  where  $f(n)$  goes to infinity.

54.  [gowers](#) Says:

[February 8, 2009 at 2:30 am](#) | [Reply](#)

352. DFU

I've got to go to bed, but as a last comment, let me ask whether 350 is intended as a set-theoretic analogue of Ben

Green's theorem that if you have a set with few solutions of  $x + y = z$  then you can remove a small number of elements and end up with no solutions of  $x + y = z$ . It would be very nice to be able to prove something like that (and a generalization to larger finite unions).

If one were going to try to find a counterexample, one would have a large collection of sets of size around  $n/3$ , and one would partition them into disjoint pairs, and one would add to the collection the set of unions of these pairs, which would turn them into a matching in your graph. The question would then be whether one could do this without creating  $c3^n$  further edges.

Here's an attempt to do so. First we choose a random collection  $\mathcal{F}$  of sets of size  $2n/3$ , big enough that its  $n/3$  lower shadow is all of  $\binom{[n]}{n/3}$ . Next, we choose all sets of size at most  $n/3$ . Then the number of disjoint pairs is certainly at least  $c3^n$ . Also, the number of edges is at most  $2^{n/3}|\mathcal{F}|$ . Next, it seems to me that it will be very hard to remove all edges: the fact that  $\mathcal{F}$  is random and that every vertex is contained in an edge (and we could make  $\mathcal{F}$  slightly bigger so that each vertex is in several edges) ought to mean that we can choose a large matching. How big is  $\mathcal{F}$ ? Well, each set in  $\mathcal{F}$  covers  $2^{n/3}$  small sets, so by my calculations the number of edges is  $2^{2n/3}$ , which is a lot smaller than  $3^n$ . Sorry if some of this turns out to be nonsense — I must sleep now.



55. [gowers](#) Says:

[February 8, 2009 at 2:32 am](#) | [Reply](#)

353. DFU

Just to say that I wrote 352 before seeing your 351.



56. [jozsef](#) Says:

[February 8, 2009 at 8:36 am](#) | [Reply](#)

354. DFU (Density Finite Union conjecture)

Let me state here a conjecture. It might not be directly connected to our project, however I've been looking for the right formulation for a while and I think that this is the right one; First we have to find the proper notation of density since simple density wouldn't guarantee the result we are looking for. The "weighted density" of a set  $S \subset [n]$  is given by the sum of the (normalized) number of pairwise disjoint  $k$ -tuples in  $S$  for all  $1 \leq k \leq n$ . There are  $m$  pairwise disjoint sets,  $B_1, \dots, B_m$  that  $\cup_{i \in I} B_i \in S$  for any nonempty  $I \subset [m]$ . This form is clearly a density version of Folkman's theorem; If we colour the subsets of  $[n]$  by  $K$  colours, then at least one class has density  $> 1/K$ .



57. [jozsef](#) Says:

[February 8, 2009 at 8:45 am](#) | [Reply](#)

Somehow I lost the middle of the text between 10

355. DFU (Density Finite Union conjecture)

Let me state here a conjecture. It might not be directly connected to our project, however I've been looking for the right formulation for a while and I think that this is the right one; First we have to find the proper notation of density since simple density wouldn't guarantee the result we are looking for. The "weighted density" of a set  $S \subset 2^{[n]}$  is given by the sum of the (normalized) number of pairwise disjoint  $k$ -tuples in  $S$ .

$D(S) = \frac{1}{n-1} \sum_{k=2}^n |\{(A_1, \dots, A_k) : A_i \in S, |A_i \cap A_j| = 0\}| / (k+1)^n$ . The maximum number of pairwise disjoint  $k$ -tuples is  $(k+1)^n$  so  $D(S)$  is a number between 0 and 1. The conjecture is that for any number  $m$ , if  $S$  contains no  $m$  pairwise disjoint elements with all  $2^m - 1$  possible nonempty unions, then it is sparse.  $D(S)$  goes to 0 as  $n$  goes to infinity.

With different words, if  $S$  is dense (and large enough) then there are  $m$  pairwise disjoint sets,  $B_1, \dots, B_m$  that  $\bigcup_{i \in I} B_i \in S$  for any nonempty index set  $I \subset [m]$

58.  [gowers](#) Says:

[February 8, 2009 at 11:51 am](#) | [Reply](#)

356. DFU

Dear Jozsef, I've edited your last comment so that the formulae show up, and made a couple of other tiny changes (such as changing  $S \subset [n]$  to  $S \subset 2^{[n]}$ ). Please check that I haven't introduced any errors.

As it stands, I don't understand your conjecture. A typical disjoint  $k$ -tuple has all its sets of size around  $n/(k+1)$ , so it seems to me that the set of all sets of size around  $n/(k+1)$  is dense in your sense. And there is of course no hope of getting finite unions in this set. Am I misunderstanding something?

59. [DHJ — quasirandomness and obstructions to uniformity « Gowers's Weblog](#) Says:

[February 8, 2009 at 5:43 pm](#) | [Reply](#)

[...] and lower bounds for the density Hales-Jewett problem, hosted by Terence Tao on his blog, and DHJ — the triangle-removal approach, which is on this blog. These three threads are exploring different facets of the problem, though [...]

60.  [jozsef](#) Says:

[February 8, 2009 at 6:22 pm](#) | [Reply](#)

357. DFU

Tim, there is a  $1/(n-1)$  multiplier, so I thought that one layer ( $k$ ) will contribute a small fraction only. In particular, I think that rich layers form arithmetic progressions and it might be useful in a possible proof.

61.  [gowers](#) Says:

[February 8, 2009 at 6:37 pm](#) | [Reply](#)

358. DFU

Ah — I hadn't spotted that you were averaging over lots of layers. Now it starts to make more sense to me.

62. [Boris](#) Says:

[February 8, 2009 at 7:25 pm](#) | [Reply](#)

359. Averaging over multiple layers does not seem to help in comment 355. Consider the family  $S$  of all sets of odd size. It contains no disjoint  $B_1, B_2 \in S$  such that  $B_1 \cup B_2 \in S$ .

63. [Boris](#) Says:

[February 8, 2009 at 7:30 pm](#) | [Reply](#)

Oops...  $d(S) = 1/n$  for the example I gave.

64. [Boris](#) Says:

[February 8, 2009 at 8:05 pm](#) | [Reply](#)

361. I think Jozsef's conjecture is true, but not for an interesting reasons. I claim that for every fixed positive integer  $t$  the condition  $d(S) > 0$  implies that  $|S \cap \binom{[n]}{t}| \geq (1 - o(1)) \binom{n}{t}$ . Suppose not, let  $T = \binom{[n]}{t} \setminus S$ . Now let us generate



the partition of  $[n]$  into  $k$  parts as follows. Let  $P_1, \dots, P_k$  be the partition of  $[n]$  into  $k$  random parts. If  $k = cn$  with  $c > 0$ , then  $\Pr[P_i \in T] = p$  is positive, and is independent of  $i$ . Had  $P_1, \dots, P_k$  been independent, it would have implied that  $\Pr[\exists i P_i \in T] = p^k$  is exponentially small. Clearly,  $P$ 's are not independent, but they are nearly independent. More precisely, let  $R_i$  be the event that  $|P_i| \leq n^{1/4}$ . Then  $\Pr[P_i \cap P_j \neq \emptyset | R_i \wedge R_j] = 1 - O(n^{-1/2})$ . Hence, sampling  $n^{1/6}$  random sets by choosing each element with probability  $1/(k+1)$  is virtually indistinguishable from choosing a random  $k$ -tuple of disjoint sets, and then looking at first  $n^{1/6}$  of them. I believe I can write the appropriate string of inequalities to show that if asked.

Since density of  $S$  is nearly 1 on each  $\binom{[n]}{t}$  the existence of any given configuration follows by simple averaging.



65. [Jozsef](#) Says:

[February 8, 2009 at 9:18 pm](#) | [Reply](#)

362.

Thanks Boris! It seems that this is still not the right formulation for a DFU conjecture. Actually, I should have noticed that a random selection of sets with probability 1/2 already gives density zero. So, I'm still looking for the right density definition. Boris, do you have any suggestion?



66. [Terence Tao](#) Says:

[February 9, 2009 at 12:03 am](#) | [Reply](#)

363. DHJ(j,k)

I had to struggle a bit to understand what DHJ(j,k) meant - I had not followed this thread for a while - but when I finally got it, I can see how it fits nicely with the whole hypergraph regularity/hypergraph removal story, and so does look quite promising. Anyway, I thought I'd try to restate it here in case anyone else also wants to know about it...

In  $[3]^n$ , we can define the notion of a “(basic) complexity 1 set” to be a set  $A \subset [3]^n$  of strings whose 1-set lies in some fixed class  $\mathcal{A}_1 \subset 2^{[n]}$ , whose 2-set lies in some fixed class  $\mathcal{A}_2 \subset 2^{[n]}$ , and whose 3-set lies in some fixed class  $\mathcal{A}_3 \subset 2^{[n]}$ . A simple example here is  $\Gamma_{a,b,c}$ ; in this case  $\mathcal{A}_1$  is the set of subsets of  $[n]$  of size  $a$ , and so forth. As I just recently understood over at the obstructions to uniformity thread, basic complexity 1-sets are obstructions to uniformity, and so we definitely need to figure out how to find lines in these sets.

In analogy with hypergraph regularity (and also with ergodic theory), the next more complicated type of set might be “non-basic complexity 1 sets” - sets which are unions of a bounded number of basic complexity 1 sets. (In graph theory, basic complexity 1 corresponds to a complete bipartite weighted graph between two cells with some constant weight, and non-basic complexity 1 corresponds to a Szemerédi-type weighted graph between a bounded number of cells, with any pair of cells  $V_i, V_j$  having a constant weight  $d_{ij}$  on the edges connecting them.) But it seems fairly obvious that if dense basic complexity 1 sets have lines, then so do dense non-basic complexity 1 sets (though one may have to be careful with quantifiers, and use the usual Szemerédi double induction to organise the epsilons properly).

Then we have the basic complexity 2 sets in  $[3]^n$ , which are those sets  $A$  in which the 12-profile of  $A$  (i.e. the pair consisting of the 1-set and 2-set of  $A$ ; one can also identify this with an element of  $3^{[n]}$ ) lies in some fixed class  $\mathcal{A}_{12} \subset 3^{[n]}$ , the 23-profile of  $A$  lies in some fixed class  $\mathcal{A}_{23} \subset 3^{[n]}$ , and the 31-profile of  $A$  lies in some fixed class  $\mathcal{A}_{31} \subset 3^{[n]}$ . But of course when  $k=3$ , any one of these complexity 2 profiles determine the entire set  $A$ , so every set  $A$  is complexity 2, which is why DHJ(2,3) is the same as DHJ(3). [Incidentally, to answer a previous question, this hierarchy seems to be unrelated to my DHJ(2.5) - I was simplifying the lines rather than the sets.]

If we knew that sets  $A$  which were uniformly distributed wrt complexity 1 sets were quasirandom (a “counting lemma”), then we would be well on our way to reducing DHJ(2,3) to DHJ(1,3) by the usual regularity type arguments. (I'm sure Tim and the others here already know this, this is just me trying to get up to speed.)

67.  [gowers](#) Says:

[February 9, 2009 at 12:24 am](#) | [Reply](#)

### 364. Big picture

There is now a definite convergence of this thread and the obstructions-to-uniformity thread, which perhaps illustrates the advantages of not splitting the discussion up. At any rate, this last comment of Terry's is closely related to comment 411, and also to earlier comments on the 400-499 thread that attempt to establish some sort of counting lemma. I now get the feeling that the problem has been softened up, and that it is not ridiculous to hope that we might actually solve it. But it could be that certain difficulties that feel technical will turn out to be fundamental.

Unfortunately, I've got two big teaching days coming up, so my input is going to go right down for a while. But perhaps that's in the spirit of polymath — we get involved, then we go off and do other things, then we come back and catch up with what's going on, but all the while polymath continues churning away at the problem.

68.  [Terence Tao](#) Says:

[February 9, 2009 at 2:36 am](#) | [Reply](#)

Metacomment: Ideally, I suppose we should be working in an environment in which comments from different threads can somehow overlap, coalesce, or otherwise interlink to each other. But, failing that, I think this division into subthreads has been the least worst solution (e.g. the  $c_n$  for small  $n$  thread seems to have taken off once it was separated from the “mainstream” threads of the project). And revisiting the threads every 100 comments or so seems to be just about the right tempo, I think.

69.  [gowers](#) Says:

[February 9, 2009 at 10:49 am](#) | [Reply](#)

### 365. Technical clean-up idea

This is another comment that could go either here or on the obstructions-to-uniformity thread, but I think it is slightly more appropriate here. Terry, that's a good point about the 100-comments-per-post rule, and if the 300s and 400s threads really do seem to be converging, then they can merge into the 500s thread.

My comment is that I have a variant of DHJ that ought to be equivalent to DHJ itself but has the potential to tidy up a lot of the annoying technicalities to do with slices being of different sizes. (I don't know for certain that it will work, however.) It's motivated by the proof of Sperner that I gave in 331. There one shows the stronger result that if the slice densities add up to more than 1 then there is a “combinatorial line”. What should the analogue be for DHJ?

My contention is the following: define the slice density  $\delta_{a,b,c}$  to be  $|A \cap \Gamma_{a,b,c}|/|\Gamma_{a,b,c}|$ . Then I claim that if  $\sum_{a+b+c=n} \delta_{a,b,c} \geq cn^2$  (for fixed positive  $c$  and sufficiently large  $n$ ), then  $A$  contains a combinatorial line. This conjecture bears the same relation to corners that the strong form of Sperner bears to the trivial fact that if you have at least two points in  $[n]$  then you get a configuration of the form  $x, x+d$  with  $d \neq 0$ . Given that the averaging argument for Sperner naturally proves this stronger result, I think there is reason to hope that the numbers will all work out miraculously better if we go for this revised version of DHJ. I hope that way we can avoid boring details about being near the middle.

I haven't begun to check whether this miracle really does happen, but I think it's definitely worth investigating.

70.  [gowers](#) Says:

[February 9, 2009 at 11:07 am](#) | [Reply](#)

366. Technical clean-up idea.

Another way of looking at 365. All we're doing is choosing a random point of  $[3]^n$  not with the uniform measure but by first choosing a slice and then choosing a random point in that slice. It seems to me that this should have huge advantages. For instance, and most notably, the distribution of a random point used to be radically different if you conditioned on it belonging to a typical combinatorial line. Now it's different but not that different. (With the new measure the situation is on a par with the situation for corners, where the difference between the distributions of random point and random vertex of a random corner is not a serious problem.) But I think there may be all sorts of arguments that will now work out exactly. Perhaps even the quasirandomness calculations I was trying to do will become clean and nice. No time to check though!

## Leave a Reply

Logged in as [Ryan O'Donnell](#). [Logout »](#)

Submit Comment

☐ Notify me of follow-up comments via email.

---

[Blog at WordPress.com.](#)

[Entries \(RSS\)](#) and [Comments \(RSS\)](#).

⌵

# Gowers's Weblog

Mathematics related discussions

« [A remarkable recent result in Banach space theory](#)  
[DHJ — possible proof strategies](#) »

## DHJ — quasirandomness and obstructions to uniformity

By gowers

This post is one of three threads that began in the comments after the post entitled [A combinatorial approach to density Hales-Jewett](#). The other two are [Upper and lower bounds for the density Hales-Jewett problem](#), hosted by Terence Tao on his blog, and [DHJ — the triangle-removal approach](#), which is on this blog. These three threads are exploring different facets of the problem, though there are likely to be some relationships between them.

### Quasirandomness.

Many proofs of combinatorial density theorems rely in one way or another on an appropriate notion of quasirandomness. The idea is that, as we do, you have a dense subset of some structure and you want to prove that it must contain a substructure of a certain kind. You observe that a *random* subset of the given density will contain many substructures of that kind (with extremely high probability), and you then ask, “What properties do random sets have that cause them to contain all these substructures?” More formally, you look for some *deterministic* property of subsets of your structure that is sufficient to guarantee that a dense subset contains roughly the same number of substructures as a random subset of the same density. For example, suppose you are trying to find a triangle in a dense graph. (We can think of a graph with  $n$  vertices as a subset of the complete graph with  $n$  vertices. So the complete graph  $K_n$  is our “structure” in this case.) It turns out that a quasirandomness property that gives you the right number of triangles is this: a graph of density  $\alpha$  is quasirandom if the number of quadruples  $(x_1, x_2, x_3, x_4)$  such that  $x_1x_2, x_2x_3, x_3x_4$  and  $x_4x_1$  are all edges is approximately  $\alpha^4 n^4$ . That is, the number of labelled 4-cycles is roughly what it would be in a random graph. If this is the case, then it can be shown that it is automatically also the case that the number of triples  $(x_1, x_2, x_3)$  such that  $x_1x_2, x_2x_3$  and  $x_3x_1$  are edges is approximately  $\alpha^3 n^3$ , again the number that you would expect in a random graph.

This might seem a slightly disappointing answer to the question. We are trying to guarantee the expected number of one small subgraph and we assume as our quasirandomness property something that appears to be equally strong: that we have the expected number of *another* small subgraph. How can that possibly be a good thing to do?

A first answer is that if you have the right number of 4-cycles then it doesn't just give you triangles: it gives you all small subgraphs. But an answer that is more pertinent to us here is that we can draw very useful consequences if a graph does *not* have the right number of 4-cycles. For example, we can find two large sets of vertices  $A$  and  $B$  such that the number of pairs  $(x, y) \in A \times B$  such that  $xy$  is an edge is significantly larger than  $\alpha|A||B|$  (which is what we would get for a random graph). Note that this is a global conclusion (an over-dense bipartite subgraph) deduced from a local hypothesis (the wrong number of 4-cycles). We can turn this round and say that we have an alternative definition of quasirandomness: a graph of density  $\alpha$  is quasirandom if for any two large sets  $A$  and  $B$  of vertices, the number of pairs  $(x, y) \in A \times B$  such that  $xy$  is an edge is about  $\alpha|A||B|$ . This definition turns out to be equivalent to the previous definition.

This sort of local-to-global implication is extremely useful. The local property can be used to show that sets have substructures. And if the local property fails, then one has some global property of the set that can be exploited in a number of ways.

It is not too important for the purposes of this thread to understand what all these ways are. But one of them does stand out: a global property is useful if it allows you to find a substructure of the main structure such that the intersection of  $A$  with that substructure is denser than it would be for a random set. In the graphs case, our substructure will be a large complete bipartite graph. The obvious substructure to hope for in the density Hales-Jewett set-up is a combinatorial subspace (which is like a combinatorial line except that you have several different variable sets, with the number of these sets being called the *dimension* of the combinatorial subspace). However, as I shall argue in a comment after this summary, we will probably need to go for something more complicated than a combinatorial subspace.

So one question that it would be very good to answer is this: what on earth do we mean by a quasirandom subset of  $[3]^n$ ? Or rather, what *local* property would guarantee roughly the expected number of combinatorial lines in a dense subset of  $[3]^n$ ? (Technical point: even this is probably not quite the right question, since we ought to work not in  $[3]^n$  but in a suitable triangular grid of slices, so that the uniform measure is a sensible one to use.) Of course, we would also like to understand what global properties would guarantee this, but that is a topic for the next section.

### Obstructions to uniformity.

If one wants to devise a proof that depends on a quasirandomness property, then one wants that property itself to have some properties. These are (i) that it should have a local definition that allows one to show that quasirandom sets contain many combinatorial lines; (ii) that every non-quasirandom set should be non-quasirandom for a global reason (such as intersecting a large structured set more densely than expected); and (iii) that every set that is non-quasirandom for a global reason like this should look “too dense” in some substructure that is similar to the main structure.

One of the main purposes of this thread is to understand (ii) in the case of the density Hales-Jewett problem. What are the global reasons for a set not to have roughly the expected number of combinatorial lines? These are what is meant by *obstructions to uniformity*. The dream is to be able to say, “It’s clear that a set with one of these global properties will not necessarily have the expected number of combinatorial lines. But in fact, the converse is true too: if a set does not have the expected number of combinatorial lines then it must have one of these global properties.”

How might we work out what the obstructions to uniformity are? There are two possible approaches, both of which are valuable. One is just to think of as many genuinely different ways as we can for a set to fail to have the expected number of combinatorial lines. Another is to devise a local quasirandomness property and to see what we can deduce if that local property does not hold. So far, all the discussion has centred on the first approach, though in the end it is the second approach that one would have to use (as it is in other contexts where the theory of quasirandomness is applied).

There was quite a bit of discussion about the first approach in the “obstructions to uniformity” thread on the [combinatorial approach to density Hales-Jewett post](#). It was summarized by Terence Tao in his [comment number 148](#), so I won’t give lots of links here (though I will give some). But let me briefly say what the obstructions we now know of are, and add one that has not yet been mentioned.

*Example 1.* One fairly general class of obstructions is obtained as follows. Let  $f$  be some fixed function from  $\{1, 2, 3\}$  to  $\mathbb{Z}$ , let  $m$  be a positive integer, let  $E$  be a structured set of integers mod  $m$  (I’ll say what “structured” means in a moment), let  $r_1, \dots, r_n$  be integers mod  $m$ , and let  $A$  be the set of all sequences  $x \in [3]^n$  such that  $\phi(x) = \sum_i r_i f(x_i) \in E$ .

Suppose now that  $(x, y, z)$  is a combinatorial line in  $A$ , with the variable coordinates being 1 for  $x$ , 2 for  $y$  and 3 for  $z$ . Let  $V$  be the variable set. Then  $\phi(y) - \phi(x) = \sum_{i \in V} r_i (f(2) - f(1))$  and  $\phi(z) - \phi(x) = \sum_{i \in V} r_i (f(3) - f(1))$ . Therefore,  $\phi(z) - \phi(x) = \frac{f(3)-f(1)}{f(2)-f(1)}(\phi(y) - \phi(x))$ . Now we already know that  $\phi(y) - \phi(x) \in E - E$ , so if there is a significant correlation between  $E$  and the dilation  $\frac{f(3)-f(1)}{f(2)-f(1)}(E - E)$  of  $E - E$  (let’s assume for simplicity that  $f(2) - f(1)$  is invertible mod  $m$ ), then knowing that  $x$  and  $y$  belong to  $A$  makes it more (or less if the correlation is negative) likely that  $z$  will belong

to  $A$ . Such correlation is possible if  $E - E$  is a set such as an arithmetic progression or a Bohr set, and the fraction  $\frac{f(3)-f(1)}{f(2)-f(1)}$  has reasonably small numerator and denominator when written in its lowest terms. (Incidentally, Boris Bukh is the expert on sums of sets with their dilates.)

The example given is what one might call an extreme obstruction. If  $A$  is an extreme obstruction constructed in that way, then a less extreme obstruction derived from  $A$  would be any set  $B$  such that  $|B \cap A|/|A|$  is significantly different from  $|B|/3^n$ . A simple example of such a set  $B$  is any set such that the number of sequences  $x \in B$  with  $x_1 = 1$  is significantly different from  $|B|/3$ . Here we are taking  $f(j) = j$  for  $j = 1, 2, 3$ ,  $r_1 = 1$ ,  $r_i = 0$  for every  $i > 1$ ,  $m = 3$ , and  $E = \{1\}$ . (These are not the unique choices that give the example.)

*Example 2.* The next class of obstructions is closely related to obstructions that occur in the corners problem. Suppose that  $X$  and  $Y$  are two subsets of  $[n]$ , with density  $\alpha$  and  $\beta$ , respectively. Then  $X \times Y$  has density  $\alpha\beta$  in  $[n]^2$ . Now suppose that we have two points in  $X \times Y$  of the form  $(x, y + d)$  and  $(x + d, y)$  for some  $d \neq 0$ . Then the point  $(x, y)$  also belongs to  $X \times Y$ . Therefore, if you have the two 45-degree vertices of a corner, you automatically get the 90-degree vertex. The effect of this is to give  $X \times Y$  more corners than a random subset of  $[n]^2$  of density  $\alpha\beta$ . This is an important example in the corners problem, because if  $S$  is a structured subset of  $[n]$  and the sets  $X$  and  $Y$  are random subsets of  $[n]^2$ , then  $X \times Y$  will have density roughly  $\alpha\beta$  inside  $S^2$  as well, unless  $S$  is very small. The original proof of Ajtai and Szemerédi accepted this problem and used Szemerédi's theorem to pass to a very small substructure with a density increase, and I think we should (initially at least) follow their lead, because it is likely to be much simpler. Shkredov, in his quantitative work on the corners problem, defined "structured set" to mean something like a set  $X \times Y$ , where both  $X$  and  $Y$  are highly quasirandom subsets of a structured subset of  $[n]$ .

These facts about corners naturally suggest the following class of obstructions in the DHJ problem. This class is a generalization that I should have spotted of a class that was suggested in [my comment 116](#). It also relates very interestingly to some of the discussion over on the Sperner thread about simpler forms of DHJ. (See for example [comment 335](#) and [comment 344](#).)

Let  $\mathcal{A}$ ,  $\mathcal{B}$  and  $\mathcal{C}$  be subsets of  $2^{[n]}$ . Then let  $A$  be the set of all sequences  $x$  such that the 1-set of  $x$  belongs to  $\mathcal{A}$ , the 2-set of  $x$  belongs to  $\mathcal{B}$  and the 3-set of  $x$  belongs to  $\mathcal{C}$ . Suppose that the resulting set  $A$  happens to be dense (which is not necessarily the case even if  $\mathcal{A}$ ,  $\mathcal{B}$  and  $\mathcal{C}$  are dense). Then  $A$  has a tendency to contain too many combinatorial lines, however random the sets  $\mathcal{A}$ ,  $\mathcal{B}$  and  $\mathcal{C}$  are, basically because if  $(x, y, z)$  is a combinatorial line then you know things like that if the 1-set of  $y$  lies in  $\mathcal{A}$ , then so does the 1-set of  $z$ .

The discussion over at the Sperner thread has almost certainly already given us the tools to deal with a set  $A$  of such a kind. The idea would be to find a partition of almost all of  $A$  into combinatorial subspaces with dimension tending to infinity. The proof would be a development of the proof of DHJ(1,3), which it looks as though we already have. This would be an imitation of the Ajtai-Szemerédi approach to corners, we will find ourselves wanting to use DHJ(1,3) as a lemma in the proof of DHJ(2,3). (Actually, as I write this, I begin to see that it may not be quite so straightforward. We would probably need a multidimensional Sperner theorem, and I now realize that it is thoughts of this kind that motivate Jozsef's attempt to find a density version of the finite-unions theorem—see [comment 341](#) and several subsequent comments.)

*Example 3.* This is an intriguing example introduced by Ryan O'Donnell in [comment 92](#). He considers (amongst other similar examples) the set  $A$  of all sequences  $x$  that contain a consecutive run of  $m$  1s, where  $m$  is chosen so that the density of  $A$  is between  $1/3$  and  $2/3$ , say. (A simple calculation shows that you want to take  $m$  to be around  $\log n - \log \log n$ .) Now if two points in a combinatorial line contain a run of  $m$  1s, then at least one, and hence both, of those points contains a run of  $m$  1s in its fixed coordinates, which forces the third point to contain a run of  $m$  1s as well. So once again we get the wrong number of combinatorial lines, because you only have to find two points in a combinatorial line to guarantee the third.

It is easy to see that a number of variants of this example can be defined. (For example, one can permute the ground set, or one can ask for a run whose length varies according to where you are, etc. etc.)

**What next?**



If we are going to try to prove that a non-quasirandom set must have an easily understood obstruction to uniformity, our task will be much easier if we can come up with a single description that applies to all the examples we know of. That is, we would like a single global property that is shared by all the above examples (and any other examples there might be) and that allows us to obtain a density increase on a substructure. In [comment 148](#), Terry did just that, defining a notion of a *standard obstruction to uniformity*. However, I do not yet understand his definition well enough to be able to explain it here, or to know whether it covers the more general version of Example 2 above. Terry, if you're reading this and would like to give a fully precise definition, it would be very helpful and I could update this post to include it.

The main questions for this thread, it seems to me, are these.

1. What would be a good definition of a quasirandom subset of  $[3]^n$ , or rather of some suitable union of slices in  $[3]^n$ ?
2. Have we thought of all the ways that a set can have too many combinatorial lines? (Ryan's example is troubling here: it seems to contain an idea that could be much more widely applicable. One might call this the "or" idea.)
3. Do we have a unified description of all the obstructions to uniformity that we have thought of so far, or, even better, a description that might conceivably apply to *all* obstructions to uniformity?

Comments on this post should begin at 400.

---

#### Possibly related posts: (automatically generated)

- [A combinatorial approach to density Hales-Jewett](#)
- [1.2 Modular Equivalences](#)
- [Preliminary Stuff: Introductory Set Theory and Relations](#)
- [The classical example of non-measurable set in the real line](#)

This entry was posted on February 8, 2009 at 4:53 pm and is filed under [polymath1](#). You can follow any responses to this entry through the [RSS 2.0](#) feed. You can [leave a response](#), or [trackback](#) from your own site.

## 116 Responses to "DHJ — quasirandomness and obstructions to uniformity"

1.  [gowers](#) Says:

[February 8, 2009 at 6:08 pm](#) | [Reply](#)

400. Quasirandomness.

In [comment number 109](#), I asked for the go-ahead to do certain calculations away from the computer. Nobody responded to the request, which suggests to me that making requests that demand blog comments for their replies is not a good idea. In future, if I have such a request I will do it via a poll. (If anybody else wants me to fix up a poll for them, I think I now understand how to get such a poll to appear in a comment, so I'd be happy to do it for them.) However, given that I have *not* been granted permission to do these calculations, I am going to do them in public instead. The result may be a mess, but I hope that if it is, then at least it will be an informative mess.

What I want to do is prove a counting lemma for triangles in a special class of tripartite graphs, where the vertices are subsets of  $[n]$ , and any edge has to join disjoint sets. (This class has already been much discussed.) Such a lemma would say that if all three parts of the graph are quasirandom, then the number of triangles is roughly what you expect it to be. At this stage, we do not know what "quasirandom" means, so we might seem to have a serious problem. However, this may well not be a problem, since the existing definitions of quasirandomness didn't just spring from nowhere and turn out to be useful: one can derive them by attempting to prove counting lemmas and seeing what hypotheses you need in order to get your attempts to work. That is what I plan to try to do here.

Let me give the definitions that I will need in order to state a counting lemma formally (with the exception of the

definition of “quasirandom”, which I’m hoping will emerge during the proof). I shall follow standard practice in additive combinatorics, and attempt to prove a result about  $[-1, 1]$ -valued functions rather than about sets. Let  $\mathcal{U}$ ,  $\mathcal{V}$  and  $\mathcal{W}$  be three copies of the set of all subsets of  $[n]$  of size between  $n/3 - m$  and  $n/3 + m$ . Let us write  $\mathcal{U} \times \mathcal{V}$  not for the usual Cartesian product, but rather for the set of all pairs  $(U, V)$  such that  $U \in \mathcal{U}$ ,  $V \in \mathcal{V}$  and  $U \cap V = \emptyset$ , and similarly for the other two possible products, and also for the product of all three. (I called this the *Kneser product* in another comment, and may do so again in this thread.) Here is the counting lemma.

**Lemma.** Let  $f, g$  and  $h$  be  $c$ -quasirandom functions from  $\mathcal{U} \times \mathcal{V}$ ,  $\mathcal{U} \times \mathcal{W}$  and  $\mathcal{V} \times \mathcal{W}$  to  $[-1, 1]$ , respectively. Then the absolute value of the expectation of  $f(U, V)g(U, W)h(V, W)$  over all triples  $(U, V, W) \in \mathcal{U} \times \mathcal{V} \times \mathcal{W}$  is at most  $c'$ , where  $c'$  tends to zero as  $c$  tends to zero.

No guarantee that this is a remotely sensible formulation, but it’s a first attempt. To be continued in my next comment.

2.  [gowers](#) Says:

[February 8, 2009 at 6:31 pm](#) | [Reply](#)

401. Quasirandomness.

To understand the lemma, it is helpful to bear in mind the example where each of  $f, g$  and  $h$  is a function that randomly takes the values 1 or -1.

Now let’s attempt to prove it, using standard Cauchy-Schwarz methods. To avoid overburdening the notation, I will adopt the following convention concerning whether it is to be understood that sets are disjoint. If I write something like  $\mathbb{E}_{U, V}$  then  $U$  and  $V$  are disjoint, whereas if I write  $\mathbb{E}_U \mathbb{E}_V$  then they are allowed to overlap. An exception is that if I use the same letter twice then overlap is allowed: so if I write  $\mathbb{E}_{W, W'}$  then  $W$  and  $W'$  do not have to be disjoint. I will also write  $D(U, V)$  for the set of  $W$  that are disjoint from both  $U$  and  $V$ , and so on. I hope it will be clear.

With the benefit of many similar arguments to base this one on, we try to find an upper bound for the fourth power of the quantity that ultimately interests us. Here goes:

$$\begin{aligned} & (\mathbb{E}_{U, V, W} f(U, V)g(U, W)h(V, W))^4 \\ &= (\mathbb{E}_{U, V} f(U, V)(\mathbb{E}_{W \in D(U, V)} g(U, W)h(V, W)))^4 \end{aligned}$$

Before I continue, I need to decide whether it is catastrophic to apply Cauchy-Schwarz at this point. The answer is no if and only if I have some reason to suppose that the function  $(U, V) \mapsto \mathbb{E}_{W \in D(U, V)} g(U, W)h(V, W)$  is reasonably constant (by which I mean that it doesn’t do something like being zero except for a very sparse set of  $U$ ). It seems OK, so let’s continue:

$$\leq (\mathbb{E}_{U, V} f(U, V)^2)^2 (\mathbb{E}_{U, V} (\mathbb{E}_{W \in D(U, V)} g(U, W)h(V, W))^2)^2$$

Now let’s use the fact that  $f$  is bounded above by 1 in modulus to get rid of it. (This looks ridiculous as we’re throwing away all that quasirandomness information about  $f$ . But experience with other problems suggests that we shall need just one of the three functions to be quasirandom.)

$$\leq (\mathbb{E}_{U, V} (\mathbb{E}_{W \in D(U, V)} g(U, W)h(V, W))^2)^2$$

Now we multiply out the inner bracket.

$$= (\mathbb{E}_{U, V} \mathbb{E}_{W, W' \in D(U, V)} g(U, W)h(V, W)g(U, W')h(V, W'))^2$$

Now interchange the order of summation.

$$= (\mathbb{E}_{W, W'} \mathbb{E}_{U, V \in D(W \cup W')} g(U, W)h(V, W)g(U, W')h(V, W'))^2$$

Hang on. That was wrong, and the first point where something happens that is interestingly different from what happens in other proofs of this kind. The problem is that there is a big difference between interchanging expectations and interchanging summations round here. In fact, even the very first step, which looked innocuous enough, was wrong. I'll pause here and reflect on this in another comment.



3. [Terence Tao](#) Says:

[February 8, 2009 at 6:41 pm](#) | [Reply](#)

402. Standard obstructions to uniformity

I'll try here to clarify my earlier comment 148. Let me start by reviewing the theory of corners  $(a, b), (a + r, b), (a, b + r)$  in  $[n]^2$ , or more precisely  $(\mathbb{Z}/n\mathbb{Z})^2$  to avoid some (extremely minor) technicalities.

Say that a dense set  $A \subset (\mathbb{Z}/n\mathbb{Z})^2$  is *\_uniform for the first corner\_* if for any other dense sets  $B, C$ , the number of corners  $(a, b), (a + r, b), (a, b + r)$  with  $(a, b) \in A, (a + r, b) \in B, (a, b + r) \in C$  is close to the expected number. It is standard that  $A$  is uniform for the first corner if and only if it does not have an anomalous density on a dense Cartesian product  $\{(a, b) : a \in X, b \in Y\}$ . There are similar results for the other corners, e.g.  $B$  is uniform on the second corner iff there is no anomalous density on a dense sheared Cartesian product  $\{(a, b) : a + b \in Z, b \in Y\}$ , etc.

One can of course replace sets by functions: a function  $f(a, b)$  is uniform for the first corner iff it does not have an anomalous average on dense Cartesian products, etc.

Now return to DHJ. Given a set  $A$  in  $[3]^n$ , we can define the function  $f(a, b)$  on  $[n]^2$  to be the number of elements of  $A$  with  $a$  1s and  $b$  2s (and thus  $c = n - a - b$  3s, which I will ignore here). Let me call this function  $f$  the *global richness profile* for  $A$ . We know that one good way to construct combinatorial line-free sets  $A$  is to ensure that  $f(a, b)$  is supported on a set with no corners (indeed, all the best examples in my own thread on this project come this way). More generally, it is pretty clear that  $A$  is not going to be quasirandom for the third point of combinatorial lines if its richness profile  $f$  is not uniform for the first corner. (I managed to interchange “first” and “third” in my notational conventions, but please ignore this for the sake of discussion.)

That's the “global” obstruction to uniformity. One can also do something similar using “local” obstructions. Consider a reasonably large subset  $S$  of  $[n]$ , and define the *local richness profile*  $f_S(a, b)$  of  $A$  to be the number of elements of  $A$  which have  $a$  1s in  $S$  and  $b$  2s in  $S$  (and thus  $|S| - a - b$  3s in  $S$ ). Now, if the local richness profile is supported on a corner-free set, then it is not quite true that  $A$  has no combinatorial lines; but the lines in  $A$  now must have all their wildcards outside of  $S$ , so this should cut down substantially on the number of lines in  $A$  (compared with the random set  $A$ ) if  $S$  is even moderately large. For similar reasons, I would expect  $A$  to fail to be quasirandom for the third point of a combinatorial line if  $f_S$  is not uniform for the first corner, and analogously for the other two points of the combinatorial line.

Boris suggested replacing the set  $S$  with a more general integer-valued weight  $w : [n] \rightarrow \mathbb{Z}$ , which should accomplish a similar effect.

One can localise these local obstructions even further, by working with slices of  $A$  instead of all of  $A$ , e.g. by fixing some moderate number of coordinates and then considering a local or global richness profile for the slice on the remaining coordinates (or some subset  $S$  thereof, or some weight  $w$ ). If this profile fails to be quasirandom for most slices, or its density deviates from the mean for most slices, then this should also be an obstruction to uniformity.

So I guess my formalisation of a “standard obstruction to uniformity” for the third point of a combinatorial line for a set  $A \subset [3]^n$  of density  $\delta$ , would consist of a small set  $T$  of fixed indices, then a disjoint large subset  $S$  of sampling indices, and dense sets  $X, Y$  of  $[n]$  (or more precisely  $[|S|]$ ) such that for most of the  $3^{|T|}$  choices  $\alpha$  of the fixed indices, the local richness profile  $f_{S, \alpha}(a, b)$ , defined as the number of elements of  $A$  with coordinates  $\alpha$  on  $T$ , and with  $a$  1s in  $S$  and  $b$  2s in  $S$ , has density significantly different from  $\delta$  on  $X \times Y$ . (Actually, as I write this it occurs to me that one may need  $X, Y$  to depend on  $\alpha$ .)

Anyway, that's my proposal for the standard obstruction... hopefully it makes some sense.

4.  [gowers](#) Says:

[February 8, 2009 at 6:54 pm](#) | [Reply](#)

403. Standard obstructions to uniformity.

Thanks very much for that Terry. I've just read it rather quickly, and I get the impression that because it focuses strongly on the cardinalities of sets, it may not have picked up the full (and new since your comment 148 ) generality of Example 2 in my post above. I'm not over-worried about that, as I have a strong feeling that that example can be dealt with. But if I am right that it isn't a standard obstruction to uniformity then it would be good to try to generalize the definition.

5.  [gowers](#) Says:

[February 8, 2009 at 7:45 pm](#) | [Reply](#)

404. Quasirandomness.

What, then was the problem with the first line of the calculations in 401? What I did was to replace  $\mathbb{E}_{U,V,W}$  by  $\mathbb{E}_{U,V} \mathbb{E}_{W \in D(U,V)}$ . Here, the first expectation was over disjoint triples  $(U, V, W)$  while the second took an expectation over disjoint pairs  $(U, V)$  of the expectation over all sets  $W$  that were disjoint from  $U \cup V$ . And there is a huge difference between the two, since in the second we count all disjoint pairs  $(U, V)$  equally, whereas in the first we count a pair  $(U, V)$  far more if the size of  $|U \cup V|$  is smaller, since then there are many more  $W$  that are disjoint from  $U$  and  $V$ . And since the sizes of the sets are not fixed, this is a serious concern.

For now I'm going to restrict attention to the simpler set-up where  $n = 3m + k$  for some smallish  $k$  and all our sets have size  $m$ . So now the sizes of the sets  $D(U, V)$  will always be  $\binom{m+k}{m}$ . However, while this corrects the first line, it doesn't get us out of the woods, because there is no trick like that that we can pull off to sort out the second interchanging of the order of summation. And that is because there is no getting away from the fact that  $W \cup W'$  can vary in size.

However, it now occurs to me that this may not be such a huge problem. The reason is that most of the time  $W \cup W'$  is of maximal size  $m + k$ , since the smallness of  $k$  makes it rather unlikely that a point will be in the complement of both  $W$  and  $W'$ . (What's more, the probability that the complement has size  $t$  goes down like  $(k/m)^t$  while the gain in the number of pairs  $(U, V)$  goes up like  $C^t$ , so the exceptions should be insignificant.)

I'm not sure that this reasoning is correct, but for now let us pretend that the problem doesn't exist and that that interchange of summation was after all legal, except that I should have written

$$\approx \left( \mathbb{E}_{|W \cup W'| = m+k} \mathbb{E}_{U,V \in D(W \cup W')} \right. \\ \left. g(U, W)h(V, W)g(U, W')h(V, W') \right)^2.$$

So the next stage is to apply Cauchy-Schwarz, which gives us

$$\leq \mathbb{E}_{|W \cup W'| = m+k} \left( \mathbb{E}_{U,V \in D(W \cup W')} \right. \\ \left. g(U, W)h(V, W)g(U, W')h(V, W') \right)^2.$$

Now I'm getting worried because fixing  $U \in D(W \cup W')$  determines  $V$ , and this feels wrong. So I'm going to start all over again, this time using sums rather than expectations.

6.  [Gil](#) Says:

[February 8, 2009 at 8:00 pm](#) | [Reply](#)

405. density-increasing-strategies & quasirandomness

Few remarks (which may be wrong so please correct me or hint if they are) Let me split them and start with one).

a) when we talk about quasi-randomness we often talk about “codimension-one-quasi-randomness” we start with the characteristic function  $f$  of the set, we show that if the number of desirable configuration is not what we expect then the function has some large correlation with some other function  $g$ ; then we use it to somehow show that the set has higher density on a small dimensional “space” (subcube of some kind.)

We can talk about conditions for quasirandomness that refer directly about low dimensional subspaces. i.e. look at all the various irregular examples that we have and ask: do we see there unusually high density for combinatorial subspace (or some more complicated gadget) of lower dimension - square root of  $n$ ,  $\log n$  etc.

(It is true that we do not have any proof-techniques that jumps directly to low dimensional subspaces or “varieties” without passing through a codimension 1 gadget.)

7.  [Gil](#) Says:

[February 8, 2009 at 8:14 pm](#) | [Reply](#)

Dear Tim, maybe rule 6 should be slightly relaxed and people could make off-line few-hour calculations without general consensus; (actually after just reading the rules, I would consider also a no-rules version of this endeavor.) In any case, I expect a wide consensus on letting you do the sum version on-line if you so prefer and perhaps granting you a wider mandate for related calculations. (I vote yes on both.)

8.  [gowers](#) Says:

[February 8, 2009 at 8:34 pm](#) | [Reply](#)

406. Quasirandomness.

This time I do everything with sums and try to interpret it later. The disjointness conventions are as before. I haven't repeated the remarks between lines.

$$\begin{aligned} & (\sum_{U,V,W} f(U,V)g(U,W)h(V,W))^4 \\ &= (\sum_{U,V} f(U,V)(\sum_{W \in D(U,V)} g(U,W)h(V,W)))^4 \\ &\leq (\sum_{U,V} f(U,V)^2)^2 (\sum_{U,V} (\sum_{W \in D(U,V)} g(U,W)h(V,W))^2)^2 \\ &\leq |\mathcal{U} \times \mathcal{V}|^2 (\sum_{U,V} (\sum_{W \in D(U,V)} g(U,W)h(V,W))^2)^2 \end{aligned}$$

Now we divide by the factor  $|\mathcal{U} \times \mathcal{V}|^2$ . Having done that we multiply out the inner bracket.

$$= (\sum_{U,V} \sum_{W,W' \in D(U,V)} g(U,W)h(V,W)g(U,W')h(V,W'))^2$$

Now interchange the order of summation.

$$= (\sum_{W,W'} \sum_{U,V \in D(W \cup W')} g(U,W)h(V,W)g(U,W')h(V,W'))^2.$$

Now the difficulty we face is much clearer. Normally one should apply Cauchy-Schwarz only if a function is broadly

speaking reasonably constant. But here the inner sum is far bigger if  $|W \cup W'|$  is small, and in fact it is zero for most  $W$  and  $W'$ . So let's try to think what a typical quadruple of sets  $(U, V, W, W')$  looks like if we know that all four sets have size  $m$ , and all six pairs of them are disjoint apart from the pair  $(W, W')$ . (I'm still assuming, for simplicity, that  $n = 3m + k$  for some smallish  $k$  and that all sets have size  $m$ .)

I now realize I made a mistake earlier. If  $|W \cup W'| = m + j$ , then the number of disjoint pairs  $(U, V)$  in  $D(W \cup W')$  is  $\binom{2m+j}{m, m, j} = \binom{2m+k-j}{m} \binom{m+k-j}{m}$ . If you decrease  $j$  by 1, the first of these binomial coefficients goes up by a factor of approximately 2, but the second goes up by more like  $m/(k-j)$ . As for the number of pairs  $(W, W')$  such that  $|W \cup W'| = m + j$ , it equals  $\binom{3m+k}{m} \binom{m}{m-j} \binom{2m+k}{j}$ . The second of these goes down by a factor of about  $m/j$  and the third goes down by a factor of about  $2m/j$ . So it still seems as though the sum is dominated by pairs with  $|W \cup W'| = m + k$ .

Ah, I'm forgetting that I should be thinking about the *square* of the inner sum, and now things look a lot healthier. So if you decrease  $j$  by 1, then the square of the inner sum goes up by about  $(2m/(k-j))^2$  and the number of  $(W, W')$  pairs goes down by around  $2(m/j)^2$ . So some kind of concentration should occur around the  $j$  that makes these two equal, which is  $k/(1 + \sqrt{2})$ . (I admit it: I cheated and used a piece of paper there.)

I need to stop this for a while. To be continued.

9.  [gowers](#) Says:

[February 8, 2009 at 10:48 pm](#) | [Reply](#)

407. Quasirandomness.

I think at this point it helps to bring in at least one expectation. Where I've written  $\sum_{U, V \in D(W \cup W')}$  I will now write  $d_2(W, W') \mathbb{E}_{U, V \in D(W \cup W')}$ , where  $d_2(W, W')$  is the number of disjoint pairs  $(U, V)$  in  $D(W \cup W')$ , and then apply Cauchy-Schwarz. That gives me

$$\begin{aligned} & \left( \sum_{W, W'} \sum_{U, V \in D(W \cup W')} g(U, W) h(V, W) g(U, W') h(V, W') \right)^2 \\ &= \left( \sum_{W, W'} d_2(W, W') \mathbb{E}_{U, V \in D(W \cup W')} \right. \\ & \quad \left. g(U, W) h(V, W) g(U, W') h(V, W') \right)^2 \\ &\leq \left( \sum_{W, W'} d_2(W, W') \right) \times \\ & \quad \sum_{W, W'} d_2(W, W') \left( \mathbb{E}_{U, V \in D(W \cup W')} \right. \\ & \quad \left. g(U, W) h(V, W) g(U, W') h(V, W') \right)^2. \end{aligned}$$

Now the first bracket is just the number of quadruples  $(U, V, W, W')$  of sets of size  $m$  such that all sets with different letters are disjoint. So let's drop it and try to estimate the second.

In this model, we are assuming that  $U$  and  $V$  both have size  $m$ . So I'm allowed to replace  $\mathbb{E}_{U, V \in D(W \cup W')}$  by  $\mathbb{E}_{U \in D(W \cup W')} \mathbb{E}_{V \in D(U, W \cup W')}$ . That gives

$$\begin{aligned} & \sum_{W, W'} d_2(W, W') \left( \mathbb{E}_{U \in D(W \cup W')} \mathbb{E}_{V \in D(U, W \cup W')} \right. \\ & \quad \left. g(U, W) h(V, W) g(U, W') h(V, W') \right)^2, \end{aligned}$$

which by Cauchy-Schwarz and the boundedness of  $g$  is at most




$$\sum_{W, W'} d_2(W, W') \mathbb{E}_{U \in D(W \cup W')} \left( \mathbb{E}_{V \in D(U, W \cup W')} h(V, W) h(V, W') \right)^2.$$

Expanding the bracket gives us


$$= \sum_{W, W'} d_2(W, W') \mathbb{E}_{U \in D(W \cup W')} \mathbb{E}_{V, V' \in D(U, W \cup W')} h(V, W) h(V, W') h(V', W) h(V', W').$$

This is a slight mess but it makes me hopeful, for reasons that I'll explain in my next comment.

10.  [Terence Tao](#) Says:  
[February 8, 2009 at 11:29 pm](#) | [Reply](#)


409. Standard obstructions to uniformity

Tim, you're right, Example 2 (where the 1-set, 2-set, and 3-set of  $A$  are independently constrained) is quite different from the class I had in mind. It may be that this is the more fundamental example, and I will try to think about whether the other examples we have can be reformulated to resemble this one. It may take some time before I will be able to look at it, though.

11.  [gowers](#) Says:  
[February 8, 2009 at 11:30 pm](#) | [Reply](#)

410. Quasirandomness.

The above calculation ends up with something a little bit strange, but here's what I hope might happen if it is tidied up. Given any  $U$  we can define a bipartite graph whose vertex sets are the set of all  $V$  that are disjoint from  $U$  and the set of all  $W$  that are disjoint from  $U$  and whose edges are pairs  $(V, W) \in \mathcal{V} \times \mathcal{W}$  (still the Kneser product). If we restrict  $h$  to this bipartite graph, we can "count 4-cycles", or rather evaluate the sum of  $h(V, W)h(V, W')h(V', W)h(V', W')$ , restricted to this bipartite graph. Could one perhaps hope that if this sum is on average small, then the function  $h$  is sufficiently quasirandom to make the original sum (at the beginning of 406) or expectation (at the beginning of 401) small? The calculations above suggest that something like that is true. But one needs to be wary since the intersection of the neighbourhoods of  $V$  and  $V'$  will get rapidly smaller as  $|V \cup V'|$  gets bigger, and this probably needs to be allowed for somehow. But instead of trying to sort that out, I want to speculate a bit, on the assumption that it can be made to work.

12.  [gowers](#) Says:  
[February 9, 2009 at 12:13 am](#) | [Reply](#)

411. Obstructions to uniformity.

As a result of thinking about what sort of global obstruction to uniformity might come out of the failure of quasirandomness of the type just discussed, it has just occurred to me to wonder whether in fact all the obstructions we have so far are special cases of Example 2. That is, if  $A$  is one of our basic examples of a set with the wrong number of combinatorial lines, can we always find collections of sets  $\mathcal{U}$ ,  $\mathcal{V}$  and  $\mathcal{W}$  such that  $A$  correlates unduly with the set of all sequences  $x$  with  $U_1(x) \in \mathcal{U}$ ,  $U_2(x) \in \mathcal{V}$  and  $U_3(x) \in \mathcal{W}$ , and this set of sequences is dense. (Recall that  $U_j(x)$  is the  $j$ -set of  $x$  is the set of coordinates  $i$  where  $x_i = j$ .)

Let's try Example 1. To keep things simple, let's go for the example of the sum of all the coordinates being a multiple of 7. What could we take as  $\mathcal{U}$ ,  $\mathcal{V}$  and  $\mathcal{W}$  in this case? Easy: we can take them all to consist of all sets for which the

number of elements is a multiple of 7. (If  $n$  isn't a multiple of 7, then we can find  $a, b, c$  such that  $a + b + c \equiv n$  and  $a + 2b + 3c \equiv 0 \pmod{7}$  and ask for sets in  $\mathcal{U}$  to have cardinality congruent to  $a$  etc.) It's clear that if we had different weights but still worked mod 7 we could do something similar, and I'm pretty sure that generalizing to structured sets in Abelian groups doesn't make things substantially harder. Of course, all this is to be expected because of Terry's comments about standard obstructions.

How about Ryan's example? There we could simply take  $\mathcal{U}$ ,  $\mathcal{V}$  and  $\mathcal{W}$  to consist of all sets that contain a run of  $m$  consecutive elements.

Based on that, here is a conjecture.

**Conjecture:** Let  $A$  be a subset of (the union of a suitable triangular grid of slices near the centre of)  $[3]^n$  that does not have the expected number of combinatorial lines. Then there are set systems  $\mathcal{U}$ ,  $\mathcal{V}$  and  $\mathcal{W}$  such that for a dense set of sequences  $x$  we have  $U_1(x) \in \mathcal{U}$ ,  $U_2(x) \in \mathcal{V}$  and  $U_3(x) \in \mathcal{W}$ , and such that, setting  $B$  to be the set of all such sequences, the density of  $A \cap B$  in  $\mathcal{B}$  is significantly greater than the density of  $A$ .

This conjecture is a direct analogue of a fact that is known to be true for corners. (This fact was basically mentioned by Terry in comment 402.) It's a slight generalization of what Terry was suggesting, and I think there's a definite chance that it could have dealt with question 3 in the main post.

If anyone can think of an obstruction to uniformity that is *not* of this form, then they will have demonstrated that DHJ is different in a very interesting way from the corners problem. Such an example would give a big new insight into the problem. But I'm an optimist and at the moment I am guessing that the conjecture is true.

*[Note: if this comment seems to ignore Terry's comment 409, it's because his comment appeared while I was writing it.]*

13. [Mathematics, Science, and Blogs « Combinatorics and more](#) Says:

[February 9, 2009 at 7:49 am](#) | [Reply](#)

[...] A third thread about obstructions to uniformity and density-increment strategies is forthcoming alive and kicking on Gowers's [...]

14. [Boris](#) Says:


[February 9, 2009 at 8:13 am](#) | [Reply](#)

412. The conjecture in 411 is, I am afraid, too optimistic. Observe that there are at most  $O(\sqrt{n})$  wildcards in every combinatorial line that is contained in a union of slices near the center of  $[3]^n$ . That is because the number of 1's does not vary by more than that.

Thinking of an element of  $[3]^n$  as a word of length  $n$  in  $\{1, 2, 3\}$ , let  $X_1, X_2$  be the number of occurrences of subwords 123, and 321 respectively ("subword" means a sequence of \_consecutive\_ letters). Let  $X = X_1 - X_2$ . Note that typically  $X$  varies in the interval of length  $\Theta(\sqrt{n})$  around origin. Thus, if  $m \ll O(\sqrt{n})$ , it is reasonable to suppose that  $X \bmod m$  is going to be distributed uniformly in  $\{0, \dots, m-1\}$ . Assume it is. Choose  $m = n^{1/3}$  for the sake of definiteness (we need it to be greater than  $n^{1/4}$  for what follows). Define  $\mathcal{A}$  by the condition  $X \bmod m \in \{0, 1, \dots, m/2\}$ . The density of  $\mathcal{A}$  is  $1/2$ , but I claim that the number of combinatorial lines is close to  $1/2$ .

Indeed in a typical combinatorial line, the letters adjacent to wildcards are not wildcards themselves. By symmetry, as change value of wildcards the expected change in  $X_1$  is equal to expected change in  $X_2$ . Since both  $X_1$  and  $X_2$  are concentrated in an interval of length  $\sqrt{n^{1/2}} = n^{1/4}$  (recall that the number of wildcards is  $n^{1/2}$ ), we infer that typically  $X$  does not change by more than  $n^{1/4}$ . Since  $m \gg n^{1/4}$ , if a combinatorial line contains a single element of  $\mathcal{A}$ , then it is very likely to be wholly contained in  $\mathcal{A}$ .

It seems intuitively clear to me that there is no density increment on  $\mathcal{A}$  of the kind desired in comment 411.

15.  [Terence Tao](#) Says:  
[February 9, 2009 at 8:56 am](#) | [Reply](#)

413. Correlation with “Cartesian semi-products”

I found a weird variant of the conjecture in 411 that I don’t know how to interpret. Let  $\mathcal{A}$  be a dense subset of  $[3]^n$  without combinatorial lines, or equivalently a collection of partitions  $[n] = A \cup B \cup C$  of  $[n]$  that does not contain a “corner”  $(A \cup D, B, C), (A, B \cup D, C), (A, B, C \cup D)$ .

Now pick a random subset  $C$  of  $[n]$  of size about  $n/3$  (up to  $\sqrt{n}$  errors) and look at the slice of  $\mathcal{A}$  with this  $C$  as the third set in the partition. This is basically a set system in  $[n] \setminus C$ , which we expect to be dense on the average. Now, following the proof of Sperner as in 331, let us pick a random maximal chain  $\emptyset = A_0 \subset \dots \subset A_{n-|C|} = [n] \setminus C$ . On the average, we expect this chain to be dense in the equator (the part of the chain with cardinality  $n/3 + O(\sqrt{n})$ ). Thus  $\mathcal{A}$  contains roughly  $\sqrt{n}$  triples of the form  $(A_j, [n] \setminus (A_j \cup C), C)$  where  $j = n/3 + O(\sqrt{n})$ . Let  $J$  be the set of all such  $j$ .

Since  $\mathcal{A}$  avoids all combinatorial lines, this means that  $\mathcal{A}$  must avoid the “Cartesian semi-product”

$$\{(A_j, [n] \setminus (A_k \cup C), C \cup (A_k \setminus A_j)) : j < k; j, k \in J\},$$


which implies that the set

$$\{(j, k) : j, k = n/3 + O(\sqrt{n}); j < k; (A_j, [n] \setminus (A_k \cup C), C \cup (A_k \setminus A_j)) \in \mathcal{A}\}$$

avoids the Cartesian product  $J \times J$ . In particular, it fails to be quasirandom with respect to such products, and so by two applications of Cauchy-Schwarz, it must have an anomalously large number of rectangles. Averaging over all choices of  $C$  and over all choices of maximal chain, this seems to imply that  $\mathcal{A}$  contains an anomalously large number of “rectangles”

$$(A_1, B_1, [n] \setminus (A_1 \cup B_1)), (A_2, B_1, [n] \setminus (A_2 \cup B_1)), (A_1, B_2, [n] \setminus (A_1 \cup B_2)), (A_2, B_2, [n] \setminus (A_2 \cup B_2))$$


where  $A_1 \subset A_2$  and  $B_1 \subset B_2$  all have cardinality about  $n/3 + O(\sqrt{n})$ . If it were not for the additional constraints  $A_1 \subset A_2, B_1 \subset B_2$ , this would start coming quite close to the conjecture in 411. As it is, I don’t know where to take this next.

16.  [Terence Tao](#) Says:  
[February 9, 2009 at 9:11 am](#) | [Reply](#)

414.

I was going to add that all the examples I mentioned in 148 happened to be of the right form for the conjecture in 411, but Boris in 412 seems to have created a totally new species of non-quasirandom set that needs to be understood better. This example is of course consistent with what I just discussed at 413, by the way: in a typical “Cartesian semi-product”, the 132 and 123 statistics don’t vary by much more than  $O(n^{1/4})$ , and so such semi-products are likely to either lie totally in  $\mathcal{A}$  or totally outside of  $\mathcal{A}$ .

(Incidentally, these Cartesian semi-products are closely related to the sets  $S$  discussed in Tim’s 331.)

17.  [gowers](#) Says:  
[February 9, 2009 at 10:24 am](#) | [Reply](#)

415. Obstructions to uniformity.

No time for serious mathematics at the moment, but I did want to give my gut reaction to the last few comments. First, my usual experience when making an overoptimistic conjecture is to go to bed, wake up, and see why it's nonsense (perhaps not immediately, but at some point when my mind wanders back to it and I'm in a more sceptical mood). With this collaboration it's different: I wake up and find a counterexample on the blog!

Having said that, I did very much like that 411 conjecture, so when I get the time I plan to do two things, unless someone has already done them. First is to try to define some set systems that do, contrary to expectation, give rise to a correlation with Boris's example. But at the moment I share his intuition that the roles of the coordinate values are somehow too intertwined to make this possible, so my main reason for doing this would be to try to get at a proof that it really is a new obstruction, or at least a sketch of a proof. (That may not be too difficult to achieve, so I'll check from time to time to see if someone else has already done it.)

Other things that one could do in response to the example are to try to formulate a yet more general notion of obstruction to uniformity that includes his example. I don't have any idea yet what that might look like, but it would be great if we could come up with something simple. It would also be good to find a combinatorial subspace in which his set is unusually dense. (Maybe Terry has done that above — I haven't yet followed his comments enough to be able to see.) And finally, I would like to think hard about what obstructions to uniformity would arise naturally from a set's not being quasirandom according to the definition that half-emerged from the calculations leading up to 410. (In 411 I just jumped to conclusions, and it now looks as though I did so prematurely.) Here I would take my cue from the way I proved one version of the regularity lemma for hypergraphs. I was in precisely this situation: there were a lot of bits of local non-uniformity (in small neighbourhoods) and one somehow had to add them up to obtain a global non-uniformity. It turned out to follow because the neighbourhoods with the local non-uniformities were nicely spread about. I think there's a good chance of pulling off a similar trick here. (An argument of the type I'm talking about appears in Section 8 of [my 3-uniform hypergraphs paper](#) in the run-up to the proof of Lemma 8.4, but our situation might be slightly simpler as we're talking about graphs rather than hypergraphs.)

Terry, I'm looking forward to understanding Cartesian semi-direct products better, but I'm coming to see the kind of thing you're doing. Actually, the more I think about it, the more I wonder if it is the kind of thing I was hoping to get to in the calculations at the beginning of this thread. If so, then the remarks towards the end of the previous paragraph of this comment could be helpful in taking things further.



18. [gowers](#) Says:

[February 9, 2009 at 10:26 am](#) | [Reply](#)

416. Obstructions to uniformity.

It also occurs to me that Boris's thought processes could be extremely helpful here. Did that example just jump out of thin air? If not, then were there some preliminary "philosophical thoughts" that led to it? If the latter, then it could be useful to know what they were so that we have a better chance of making correct guesses about how a proof might go.



19. [Gil](#) Says:

[February 9, 2009 at 11:45 am](#) | [Reply](#)

417. quasirandomness. (405 (point a) continued)

So a concrete question is this: suppose that a set of density  $c$  has density at most  $2c$  on all combinatorial subspaces of dimension  $100 \log n$  (or  $\log^2 n$ ; or  $n^{\{1/1000\}}$ ) does it necessarily have the right number of combinatorial lines? (Is any of our non quasi random example violate it?)


(Of course in the next procedural remark by "on-line" I meant "off-line".)

20. [Boris](#) Says:

[February 9, 2009 at 5:08 pm](#) | [Reply](#)


418. Answer to 416: My thought process was simple: First the analogue of Tim's conjecture for Sperner is false (if one thinks of Sperner as DHJ(2,1), then it is not true that obstructions come from DHJ(2,0) because there is no DHJ(2,0)), and hence the conjecture should be false. Second, one needs some example where coordinates are somehow linked together. Third, if we try to impose a condition on number of pairs 12, then the only condition I really know how to impose is a modular condition. It fails to give a counterexample since if one element of the combinatorial line has some number of these, it tells us next to nothing about the number of 12's in other elements of the combinatorial line because the number of 12's decreases by a lot when we change wildcard from 2 to 3. Well, then I thought about compensating for this huge change by forming some linear combination. My first construction was to count number of 12, 23, and 31 and add them weighted with third roots of unity. Finally, I simplified the construction.

Comment regarding to 417: If we take any known obstruction to the uniformity of density  $c$  and take a union with a random set of density  $c/2$  wouldn't we obtain an example Gil seeks?

21.  [Terence Tao](#) Says:  
[February 9, 2009 at 6:57 pm](#) | [Reply](#)

419. Just a quick comment, will respond at length later when I am less busy: Boris's 412 example has the property that all coordinates have extremely low influence: modifying a 1 to a 2, say, is unlikely to move you from A to not-A, or vice versa. As a consequence, it is easy to find lots of subspaces with a density increment, just by selecting a moderate number of indices and jiggling them from 1 to 2 to 3 independently.

Presumably one can more generally push the density increment argument to reduce to sets A in which all indices (and maybe sets of indices) have small influence, but I have not thought about this carefully.

22.  [gowers](#) Says:  
[February 9, 2009 at 7:02 pm](#) | [Reply](#)

420. Obstructions to uniformity.

Boris, many thanks for that — as I hoped, it was very illuminating. In particular, your initial step is thought-provoking. My first reaction to it is to try to prove you wrong by formulating DHJ(2,0), but with no particular hope of success. If indeed your reasoning is right, then it's good news in a way, because it suggests that we can get a better understanding of obstructions in DHJ(3,2) by thinking a bit harder about obstructions in DHJ(2,1).

Here's what DHJ(2,0) might say. (In order to do this I'm keeping open [comment 344](#) in a different tab and trying to extend the definition of DHJ(j,k) downwards to DHJ(0,k).) For every empty subset  $E \subset [k]$  you have a set  $\mathcal{A}_E$  of functions from  $E$  to the power set of  $[n]$  such that the images of all points in  $E$  are disjoint.

So far so good: there is precisely one such function, so  $\mathcal{A}_E$  is a singleton that contains just that trivial function.

Now we define  $\mathcal{A}$  to be the set of all ordered partitions of  $[n]$  into  $k$  sets  $U_1, \dots, U_k$  such that for every empty subset  $E$  of  $[k]$  the sequence  $(U_j : j \in E)$  belongs to  $\mathcal{A}_E$ . But that happens for every ordered partition of  $[n]$  into  $k$  sets, so  $\mathcal{A}$  corresponds to the set of all sequences.

So I claim that DHJ(0,k) is the assertion that  $[k]^n$  contains a combinatorial line!

This doesn't really affect the substance of what Boris is saying, but it allows us to reexpress it. The conjecture DHJ(0,k) (which, come to think of it, I think I see how to prove), suggests a class of 0-level obstructions: you may well not get the expected number of combinatorial lines if you have the wrong correlation with the constant function 1 — that is, if you have the wrong density. So the real point that Boris is making is that having the wrong density is not the only obstruction to Sperner (whereas it is to the trivial one-dimensional analogue of corners, since the number of one-dimensional corners is precisely determined by the density of the set).

Thus, Boris's thought processes do indeed contain a valuable new insight: that we should think hard about obstructions to Sperner, because these will feed into DHJ as an *additional* source of obstructions to DHJ(2,3), over and above the ones that come from DHJ(1,3). In this way, we see that there is something about DHJ(2,3) that genuinely goes beyond what happens for corners.

Unfortunately, it is still completely mysterious to me what that extra something is, but equally it is clear that what we must try to do is generalize Boris's obstruction as much as we possibly can and hope that we can reach a new and more sophisticated conjecture. A good start will be to see what we can do with obstructions to Sperner. (And perhaps there is some hope here, as people already seem to have thought about it. If anyone feels like giving a detailed summary of what is known, they will be doing me at least a big favour.)

23.  [Gil](#) Says:

[February 9, 2009 at 7:55 pm](#) | [Reply](#)

421. Boris, i do not think your example is a counter example. If you take the obstructions for pseudorandomness (at least the 2 I checked) you gets large combinatorial subspaces in the set. E.g. if the number of 1,2,3 coordinates is divisible by 7 or if there is a long runs.

OK i see, what you probably suggest is this: start with a obstruction with high density say  $7/8$  take a random subset of it which has density  $c$  and then even in the large combinatorial spaces the density will be below  $8/7$ . (You do not need to add anything random) Right.

Anyway, we can still ask: If the number combinatorial lines is more than 10 times or less than  $1/10$  times the expected number, must we have a combinatorial subspace of dimension  $100\log n$  or  $\log^2 n$  or  $n^{1/100}$  that the density on it is more than  $1.01c$ .

24.  [Ryan O'Donnell](#) Says:

[February 9, 2009 at 10:00 pm](#) | [Reply](#)

422. This is response to Terry's #413.

Let me just try to rephrase it in my own language. Let  $\mathcal{A}$  be a subset of  $[3]^n$  of density at least  $\delta$ .

Form a random "subcube" as follows: Pick a random string  $r \in \{*, 3\}^n$  by choosing each coordinate to be  $*$  with probability  $2/3$  and  $3$  with probability  $1/3$ . Let  $Q_r$  be the subcube  $\{1, 2\}^m$ , given by taking all ways of filling in  $*$  with 1 or 2. (Here  $m$  is the number of  $*$ 's in  $r$ .) For a given string  $x \in Q_r$ , let's abuse notation slightly by writing  $(x, r)$  for the associated string in  $[3]^n$ .


It is easy to see that with probability at least  $\delta/2$  over the choice of  $r$ , the density of  $\mathcal{A}$  on  $Q_r$  is at least  $\delta/2$ , a positive constant.

So we can apply some kind of souped-up Sperner. In particular, I believe it should be relatively easy to show that  $\mathcal{A}$  contains many "half-lines" on  $Q_r$ . I.e., if we look at pairs  $(x, y) \in Q_r$  such that  $x \prec y$  and  $(x, r), (y, r) \in \mathcal{A}$ , it should be that the set of  $x$ 's involved in such half-lines has positive density in  $Q_r$ , and the same for the  $y$ 's.

But since  $r$  was random from the right distribution, we conclude the following: Look at all "half-lines" in  $\mathcal{A}$ ; i.e., pairs  $u, v$  with the potential to be the 1- and 2-points of a combinatorial line. Then the  $u$ 's participating in these half-lines have positive density in  $[3]^n$ , and same for the  $v$ 's.


Now it almost seems like we're there, because we have a pretty dense set of half-lines. We just need to find one point in  $\mathcal{A}$  to cap off a half-line. I guess the trouble is that, while the set of points capping off a half-line has positive density in  $[3]^n$ , it still might completely miss  $\mathcal{A}$ . So perhaps the goal should be to show that either these capping points are *very* dense, or else we can see some structure in them out of Sperner.



25.  *Ryan O'Donnell* Says:  
[February 9, 2009 at 10:31 pm](#) | [Reply](#)

423. Re Boris's example #418:

One thing I do not quite understand: What is the definition of "typical combinatorial line"?

26.  *Ryan O'Donnell* Says:  
[February 9, 2009 at 10:55 pm](#) | [Reply](#)

424. Re Terry's #419.

I'm still a bit confused by this density-increment search. One finds counterexamples with the "wrong" number of combinatorial lines, but then one says, "Aha, but if I fix these coordinates this way, I get a large density increment."

Now Boris gives a counterexample in which each coordinate has low "influence". This is the only way to beat the above density-increment parry — if your example has coordinates with large influence, then you can of course fix these coordinates to get density increment.

But now in #419 Terry says, "Aha, if the coordinates all have \*low\* influence, we should be able to find a density increment."


It seems like we can always find a density increment! This is related to my comment #123.

27. *Boris* Says:  
[February 9, 2009 at 11:00 pm](#) | [Reply](#)

425. Answer to Ryan's question in #423: it is not important, as long as the definition is invariant under permutation of  $[n]$ . The only thing the construction uses is that no combinatorial line uses more than  $O(\sqrt{n})$  wildcards. Then if we take an orbit of any combinatorial under the action of symmetric group on  $[n]$  most elements of the orbits will have virtually no adjacent wildcards.


28. *Boris* Says:  
[February 9, 2009 at 11:03 pm](#) | [Reply](#)

426. I do not follow Ryan's argument in #424. If there are variables of large influence, it does not mean we can fix them to get density increment. Think of XOR function on the Boolean cube. All coordinates have large influence, but on every coordinate subcube the density is still  $1/2$ .

29.  *gowers* Says:  
[February 9, 2009 at 11:04 pm](#) | [Reply](#)

427. Re 424,419

Now I'm confused by Ryan's comment. If  $A$  is a random set, then doesn't every coordinate have large influence? And you can't get a density increase. Apologies if this shows that I don't properly understand what "influence" means.

30.  *gowers* Says:  
[February 9, 2009 at 11:08 pm](#) | [Reply](#)

428. Obstructions and possible technical simplification.

I would like to draw attention to [my comments 365 and 366](#) because they could be relevant here. Boris, if we use this

alternative measure on  $[3]^n$ , then does your example disappear?

31. [Boris](#) Says:

[February 9, 2009 at 11:27 pm](#) | [Reply](#)

429. Answer to #428: I do not think it disappears. It has density  $1/2$  on every  $\Gamma_{a,b,c}$ . This example seems to be rather hard to kill since as mentioned in #425 the only thing it really uses is invariance under permutations of  $[n]$ .

Actually, now thinking about it, there is a way to simplify/modify the example from #412 somewhat. I considered  $X = X_1 - X_2$ , but I could consider  $X_1$  or simply the number of subwords  $12$ . Yes, these quantities do vary by a lot, but this variation is predictable: it has some mean  $M$  that can be computed explicitly, and is concentrated in the interval around  $M$  of length  $n^{1/4}$ . Choose  $m$  so that  $M \bmod m$  is distance at least  $n^{1/3}$  from zero (large sieve should tell us that this can be done for most  $M$ , and maybe there is some elementary argument to do it for all  $M$  that I do not see). Then define  $\mathcal{A}$  by condition that  $(X_1 \bmod m) \bmod (M \bmod m) \in \{0, \dots, (M \bmod m)/2\}$ .

32.  [Terence Tao](#) Says:

[February 9, 2009 at 11:31 pm](#) | [Reply](#)

430. Influence vs. Bias

Ah, where is Gil when you need him 😊. There is a distinction between influence and bias for, say, a boolean function  $f(x_1, \dots, x_n)$  of  $n$  boolean inputs.

We say that the input  $x_i$  has large influence if, whenever we flip  $x_i$ , we have a high probability of flipping  $f(x_1, \dots, x_n)$ . Having low influence means that  $f$  is close to being invariant in the  $e_i$  direction, where  $e_1, \dots, e_n$  is the standard basis.

We say that there is large bias in the  $e_i$  direction (or more precisely, the basis vector  $e^i$  in the dual space, but never mind this), if  $f$  has a density increment on the hyperplane  $\{x_i = 0\}$  or on the hyperplane  $\{x_i = 1\}$ .

There is only a partial relationship between the two properties: low influence implies low bias, but not conversely; for instance, as pointed out above, a random boolean function has low bias but high influence for each variable.

High bias leads directly to a density increment and so we “win” in this case. We also “win” if there are many directions of low influence, because  $f$  is close to constant on subspaces generated by these directions and is therefore going to be unusually dense on some of them. But there is the great middle ground of low bias, high influence sets (such as the random set) in which we have to do something else.

33.  [gowers](#) Says:

[February 9, 2009 at 11:47 pm](#) | [Reply](#)

431. Obstructions.

Boris, in 425 you said you used the fact that the set of wildcards has size  $O(\sqrt{n})$ . With the measure where all slices are the same size, I think the size of the set of wildcards of a typical line becomes  $cn$  instead. So I allowed myself to be hopeful, because you needed  $m = O(\sqrt{n})$  and you also needed  $m$  to be bigger than the square root of the size of the set of wildcards. But I admit that I still don't feel as though I fully understand your example, so there may well be an obvious answer to this question.

34.  [Terence Tao](#) Says:

[February 9, 2009 at 11:52 pm](#) | [Reply](#)

431. Locality in the edit metric.

Something just occurred to me that, in retrospect, was extremely obvious. We know already that we may as well restrict our dense sets to have  $n/3 + O(\sqrt{n})$  1s, 2s, and 3s, which means we may as well restrict our combinatorial lines to have at most  $O(\sqrt{n})$  wildcards.

This implies that the points on our combinatorial lines differ by at most  $O(\sqrt{n})$  in the edit (Hamming) metric. In contrast, the diameter of the cube  $[3]^n$  is  $n$ .

Because of this, our obstructions to uniformity need to be at the scale of  $\sqrt{n}$  rather than  $n$  in the edit metric. Boris's example is a good example of this; it's basically a union of Hamming balls of radius  $m$  for some  $m \gg \sqrt{n}$ .

So maybe conjecture 411 can be recovered by somehow localising it to Hamming balls of size  $O(\sqrt{n})$ ? There is a major technical issue coming from the fact that the metric is hugely non-doubling at this scale, but perhaps if one is careful one can still localise properly.

35. [Boris](#) Says:

[February 9, 2009 at 11:53 pm](#) | [Reply](#)

432. Aha.... I missed the point about typical line having length  $cn$  with your measure. It does seem that the example disappears.



36. [Terence Tao](#) Says:

[February 9, 2009 at 11:56 pm](#) | [Reply](#)

433. Locality in the edit metric cont.

(The numbering on my previous comment should be 432.)

To illustrate what I mean by localising the obstruction, consider the problem of obstructions to length three progressions in  $\mathbb{Z}/N\mathbb{Z}$ . Fourier analysis tells us that the obstructions are given by plane waves  $e(x\xi/N)$  on the whole interval  $[N]$ .

But now suppose one is only interested in counting progressions of some much smaller length  $O(m)$  (e.g.  $m = \sqrt{N}$ ). Then the right kind of obstructions are given by *local* plane waves - waves which look like a plane wave  $e(x\xi_j/N)$  on each interval  $I_j$  of length  $m$ , but whose frequency  $\xi_j$  can vary arbitrarily with  $j$ .

Of course, one just zoom in on an interval of length  $m$  where the original set  $A$  is quite dense and get rid of this technicality; similarly here, we may have to zoom into a Hamming ball and work locally on that ball.

37. [Boris](#) Says:

[February 10, 2009 at 12:15 am](#) | [Reply](#)

434. We can assume that the density is uniform across all smallish Hamming ball. That is because if we get a density increment on a smallish Hamming ball, we get a density increment on a tiny combinatorial cube by double-counting analogous to comment #132 from the original thread. Thus, we can increment density on Hamming balls until we are stuck.




38. [Ryan O'Donnell](#) Says:

[February 10, 2009 at 12:29 am](#) | [Reply](#)


Re 426, 427, 430, distinction between influence and bias:

Right, my mistake. Most of the examples I've been thinking about have been "monotone", in which case the notions coincide.

39.  [Terence Tao](#) Says:  
[February 10, 2009 at 3:31 am](#) | [Reply](#)

#### 436. Quasirandomness

To answer Gil's question: if we manage to settle DHJ(1,3) (or more generally to find lots of combinatorial lines in our "obstructions to uniformity"), we should presumably be able to iterate and also find lots of combinatorial subspaces in all of the obstructions to uniformity. If we double count properly, this should (hopefully) tell us that any set which has anomalous density on an obstruction to uniformity has an anomalous density on at least one combinatorial subspace, and then we "win".


40.  [Terence Tao](#) Says:  
[February 10, 2009 at 3:37 am](#) | [Reply](#)

#### 437. Soft obstructions to uniformity

One possible route to keep in mind, by the way, is to back away from the "hard" goal of explicitly classifying all "obstructions to uniformity" in a very concrete manner, but instead settle for a "soft" classification which expresses these obstructions as something like a "dual function" from my primes in AP paper with Ben Green. In some cases, one can tools such as an enormous number of applications of the Cauchy-Schwarz inequality to deal with these sorts of dual functions.

Another example of a "soft" obstruction to uniformity, this time for length three progressions in  $[N]$ , is that of an almost periodic function: a function  $f : [N] \rightarrow \mathbb{C}$  whose shifts  $T^h f$  are "precompact in  $L^2$ " in some quantitative sense. One can show that these guys contain all the obstructions to uniformity by a soft argument based primarily on Cauchy-Schwarz (and does not use the Fourier transform). [Using the Fourier transform, of course, one can approximate almost periodic functions by quasiperiodic functions, e.g. bounded linear combinations of characters, thus moving from soft obstructions to hard obstructions]. It is also not hard to prove (again by "soft" means avoiding Fourier analysis) that almost periodic functions have lots of 3-term APs (and also the important strengthening that functions close to an almost periodic function also has lots of 3-term APs; there are some subtle issues with the hierarchy of epsilons here that I don't want to discuss here).

Anyway, it is another option to consider if the "hard" classification problem turns out to be too complicated.

41.  [Gil](#) Says:  
[February 10, 2009 at 7:36 am](#) | [Reply](#)

#### 438. density-increase-strategies, quasi randomness.

Can somebody remind me what is DHJ(1,3) DHJ(2,3) and DHJ(1,2)? thanks!

The point I was making is that our strategy for a proof for a density theorem is

1) Find obstruction to quasirandomness in terms of inner product with some function in a class of "obstruction" (This is what I referred to as a codim 1 condition)

2) deduce that this implies an increased density on lower dimensional subspace, and iterate.

It may be possible (although I am not aware how) to move directly from having too many or too few combinatorial lines to a statement about density on small dimensional spaces.

Also we can look at the example and see if the density increase on small spaces does not suggest "too good to be true" result. For example, if in all these obstructions we get a subspace of linear size with density-increase of  $(1+t)$  then this will imply a bound of the form  $3^n/n^\alpha$  which is too good.

BTW, there is a large density of people interested in the density HJ around here. The approach we are taking is a fixed  $k$  approach. Saharon Shelah gave the best bound for HJ using inductive argument which juggles different  $k$ s and different dimensions for the combinatorial subspaces. He was quite interested to see if approaches of similar nature can apply for DHJ.

Regarding Ryan's long runs/tribe-like examples; It looks that this may extend to sets defined by a large class of bounded depth circuits.

42.  [gowers](#) Says:

[February 10, 2009 at 9:11 am](#) | [Reply](#)

439. Local obstructions.

I've just woken up with a better grasp of what Terry is talking about in 431 and 433, so I'd like to rephrase it here (not very much actually, but perhaps just enough to be useful).

Terry points out that because, as we know, our example could live in a central area of slices of width around  $\sqrt{n}$  (meaning that the numbers of 1s, 2s and 3s are always within  $C\sqrt{n}$  of  $n/3$ ), for the simple reason that almost all sequences in  $[3]^n$  have this property, the set of wildcards in any combinatorial line can be assumed to have size at most  $C\sqrt{n}$ . This means that what happens at one sequence  $x$  should have nothing much to do with what happens at another sequence  $y$  if their Hamming distance (the number of coordinates where they differ) is substantially greater than  $\sqrt{n}$ .

This immediately suggests a generalization of the obstructions considered in Conjecture 411. The idea would be this: you would move a Hamming ball (that is, a ball in the Hamming metric) of radius  $m$  around in  $[3]^n$ , and in each translate you would have a different obstruction of the type in Conjecture 411.

The analogy Terry draws is with the case of  $\mathbb{Z}_n$ , where you could partition it into small intervals and have a different Fourier mode in each interval, and that would exhibit nonquasirandom behaviour for local phenomena such as arithmetic progressions of length 3 with small common difference.

Now if one is going to carry out such a programme, there is an issue that at first appears to be merely technical and then comes to seem more fundamental. It's that you can't do something different on each translate because the translates overlap. This is where the importance lies of Terry's remark that "the metric is hugely non-doubling on this scale", which I'll explain in a moment. But first, note that in  $\mathbb{Z}_n$  we don't have a huge problem: of course the intervals overlap, but you can just partition  $\mathbb{Z}_n$  into disjoint intervals and you're basically OK.


Back to Terry's remark. What does he mean? He means that if you double the radius of a Hamming ball, then its volume (or cardinality) hugely increases. This has the effect of making it extremely hard to pack Hamming balls efficiently. Now it isn't essential to use Hamming balls, but if we are going to partition  $[3]^n$  into sets  $H_i$  and use a different 411-style obstruction in each  $H_i$ , then we will need one crucial property: that if you modify an element of  $H_i$  by changing around  $C\sqrt{n}$  coordinates, then with non-negligible probability you remain inside  $H_i$ . (Otherwise, the supposed obstruction in  $H_i$  will not affect more than one point per typical combinatorial line.)

The isoperimetric inequality in the cube tells us that the sets where you have the best chance of staying inside when you modify an element by a given amount are Hamming balls, and something similar is true in  $[3]^n$ . And if you do the calculations for Hamming balls, you find that if you modify by  $C\sqrt{n}$  for some large constant  $C$ , then you hugely expand the set, unless the set is already quite large in which case you get everything.

In other words, you cannot do some partitioning argument like this unless you partition into huge Hamming balls, or sets of similar type, which would turn the "local" obstructions into global ones.

I think I've actually said a bit more than Terry now — I am claiming that Terry's "major technical issue" is in fact a fundamental issue that makes it impossible to devise local obstructions of the type he is talking about.

Where does that leave Boris's example? I don't have a full explanation for this, but I think it is that he somehow doesn't do completely unrelated things in each  $H_i$ . In other words, it is not a purely local obstruction and can't be massively generalized in the way suggested above. If anyone else can give a more precise explanation, I'd be very interested.

43.  [Terence Tao](#) Says:  
[February 10, 2009 at 9:17 am](#) | [Reply](#)

439. A potential strategy

I think I have a strategy for proving that dense sets  $\mathcal{A} \subset [3]^n$  "with no exploitable structure" contain combinatorial lines, basically by fooling around with the Cartesian semi-products from 413.

We have a density  $\delta$  and an extremely large  $n$ . I will also need an intermediate integer  $m = m(\delta)$  depending only on  $\delta$ . I am going to try to find a combinatorial line with at most  $m$  wildcards inside  $\mathcal{A}$ , and I think my strategy works unless  $\mathcal{A}$  has some explicit structure obstructing me, which looks exploitable by other methods.

Pick a random point  $x$  in  $[3]^n$ . We are going to look for a line inside the Hamming ball of radius  $m$  centred at  $x$ . (The discussion above has begun to persuade me that we should be working inside small Hamming balls rather than roaming across all of  $[3]^n$ .)

It is convenient to embed  $[3]^n = \{0, 1, 2\}^n$  in  $\mathbb{Z}^n$  and give the latter the standard basis  $e_1, \dots, e_n$ . Thus for instance  $x = \sum_{b \in B} e_b + 2 \sum_{c \in C} e_c$  where  $A, B, C$  are the 0-sets, 1-sets, and 2-sets of  $x$  (thus  $A, B, C$  partition  $[n]$ ). Generically, of course,  $A, B, C$  have size about  $n/3$ .

Now we pick  $m$  random elements  $a_1, \dots, a_m$  of  $A$ , which will generically be distinct. I'm going to run the Sperner argument used in 413. For  $i = 1, \dots, m$ , let's look at the points

$$x_i := x + e_{a_1} + \dots + e_{a_i} + 2e_{a_{i+1}} + \dots + 2e_{a_m},$$

thus  $x_i$  is the same string as  $x$  but with the coefficients at  $a_1, \dots, a_i$  increased from 0 to 1, and the coefficients at  $a_{i+1}, \dots, a_m$  increased from 0 to 2. Each  $x_i$  has a probability about  $\delta$  of lying in  $\mathcal{A}$ , so on average we expect to have about  $\delta m$  indices  $i$  with  $x_i \in \mathcal{A}$ . In fact there should be some sort of concentration of measure, unless  $\mathcal{A}$  has some blatant structure to it which should be exploitable, so I'm going to just assume that we do have  $|I| \sim \delta m$ , where  $I \subset [m]$  is the collection of indices  $i$  such that  $x_i \in \mathcal{A}$ .

Since  $\mathcal{A}$  avoids all combinatorial lines, we know that the points

$$x_{i,j} := x + e_{a_1} + \dots + e_{a_i} + 2e_{a_{j+1}} + \dots + 2e_{a_m} (*)$$

lie outside of  $\mathcal{A}$  whenever  $1 \leq i < j \leq m$  is such that  $i, j \in I$ . On the other hand, assuming concentration of measure I expect that about  $\delta m^2/2$  of the  $x_{i,j}$  with  $1 \leq i < j \leq m$  lie in  $\mathcal{A}$ . Applying Cauchy-Schwartz twice, I thus expect to get an anomalously large number of rectangles  $x_{i_1,j_1}, x_{i_1,j_2}, x_{i_2,j_1}, x_{i_2,j_2}$  in  $\mathcal{A}$ , where  $1 \leq i_1 < i_2 < j_1 < j_2 \leq m$ . (I'm too lazy to calculate the expected number of rectangles right now, it's something like  $\delta^4 m^4/4!$ .)

So far, I've mostly been rephrasing 413 in a different language. But now for the new idea: observe from (\*) that each  $x_{i,j}$  depends on a different subset of the random variables  $a_1, \dots, a_m$ . Because of this, it is likely that the events  $x_{i,j} \in \mathcal{A}$  exhibit a high degree of independence from each other (even after conditioning  $x$  to be fixed). Also, assuming concentration of measure again, these events should have probability about  $\delta$ . Assuming they behave like jointly independent variables, this should mean that the expected number of rectangles  $x_{i_1,j_1}, x_{i_1,j_2}, x_{i_2,j_1}, x_{i_2,j_2}$  in  $\mathcal{A}$  should be close to what is expected by random chance, contradicting the previous paragraph. (In fact, one does not need joint independence for this; independence of any four  $x_{i,j}$  arranged a rectangle will suffice, thanks to linearity of expectation.)

But what if joint independence fails? Then I think what happens is that there is some irregularity in one of the events



$x_{i,j} \in \mathcal{A}$ , in the sense that the  $i + (m - j)$ -uniform hypergraph

$$\{(a_1, \dots, a_i, a_{j+1}, \dots, a_m) : x_{i,j} \in \mathcal{A}\}$$

is non-uniform in the sense of hypergraph regularity (i.e. it has a large octahedral norm, in the notation of Tim's hypergraph regularity paper). [This is presumably the contrapositive of some nasty counting lemma that one would have to work out.] But I feel that this type of non-uniformity is a very usable type of structure on  $\mathcal{A}$  (it seems to be some sort of hypergraph generalisation of the concept of "low influence"). At any rate, now that we see some uniform hypergraphs enter the picture, it is tempting to break out the hypergraph counting, regularity, and removal machinery and see what one gets.

44. *Mathias* Says:

[February 10, 2009 at 12:02 pm](#) | [Reply](#)

440. Re: A potential strategy from Terry

I'd like to point out, that the size of the hyperedges of the hypergraph at the end of 439 could be of order  $\Theta(m)$ . I believe in order to use a removal lemma machinery here, one may have to chose  $m = m(\delta)$  to be an extremely slow growing function (like inverse Ackerman in  $1/\delta$  or worse). But this looks bad, as there are subsets of density  $3^{-m}$ , which contain no combinatorial line with at most  $m$  wildcards.

45.  *gowers* Says:

[February 10, 2009 at 2:08 pm](#) | [Reply](#)

441. General strategy

This is an idea that occurs to me after reading Terry's 439 and sort of half mixing it with my [comment 331](#). In that comment I tried, and miserably failed, to deduce DHJ from corners in exactly the way that Sperner follows from the trivial one-dimensional equivalent of corners. The problem, as I soon discovered, was that you can't embed grids into  $[3]^n$  in a nice way, analogous to the way you embed lines into  $[2]^n$  to prove Sperner. (The image of a line is a chain of sets such as empty set, 1, 12, 123, 1234, ....)

However, even if you can't get combinatorial lines to map to combinatorial lines, it does seem as though you can do at least something. Let's embed a point in the triangular grid  $\{(a, b, c) : a, b, c \geq 0, a + b + c = n\}$  into  $[3]^n$  by sending  $(a, b, c)$  to the point with  $a$  1s followed by  $b$  2s followed by  $c$  3s. In fact, I'll use some slightly more general notation:  $a1, b2, c3$  means what I've just said, but e.g.  $a1, b2, c3, d1, e2$  would mean  $a$  1s,  $b$  2s,  $c$  3s,  $d$  1s,  $e$  2s.

Then we can never have all three of  $(a1, b2, c3)$ ,  $((a+b)1, c3)$  and  $(a1, (b+c)3)$ . More generally, we can never have  $(a1, b2, c3, d2)$ ,  $((a+b)1, c3, d2)$  and  $(a1, (b+c)3, d2)$ .

Now in 311 I was trying to use this kind of idea to find a grid with no corners, and thereby to deduce that  $\mathcal{A}$  was sparse in that grid, and hence (by a suitable averaging) that it was sparse everywhere. But what if instead of going for gold straight away, we merely try to find some kind of local atypical structure all over the place, which we then try to argue must have resulted from a global atypical structure?

In the last paragraph but one, I can associate the sequences I've written down with the points  $(a, b, c)$ ,  $(a+b, 0, c)$ ,  $(a, 0, b+c)$ . Now it's not very difficult to find dense sets of triples adding up to  $n$  that contain no configurations of that kind. But I think it is less easy to make them quasirandom. So it seems (and this is one of those situations where another contributor — in this case Terry — has understood what I am writing for some while and I am catching up) that we do have some kind of ubiquitous local non-quasirandomness. (The presence of the 0 is disturbing, but that's why it could be useful to add a  $d2$  on to the end so that the sequences look more typical.)

I've got more to say about this, but no time for several hours so I'll post the comment as it stands for now.

46.  [gowers](#) Says:

[February 10, 2009 at 8:11 pm](#) | [Reply](#)

442. General strategy.

Let me amplify on one small comment that I made in the penultimate paragraph of the previous comment. I said, “I think it is less easy to make them quasirandom”. Here’s what I mean. The possible pairs  $(a+b,0,c)$ ,  $(a,0,b+c)$  form a 1-parameter family, since  $a+b+c$  is required to equal  $n$ . If  $r < s$  and we have the pairs  $(r,0,n-r)$  and  $(s,0,n-s)$ , then we are forbidden to have the point  $(r,s-r,n-s)$ . Thus, if  $R$  is the set of  $r$  such that the line  $(r1,(n-r)3)$  belongs to  $A$ , we find that  $A$  is not allowed to contain any point of the form  $(r,s-r,n-s)$  with  $r$  and  $s$  in  $R$ . Since the outer two coordinates determine the middle one in a simple way, we obtain a forbidden set that is basically a Cartesian product of  $R$  with itself, or rather in a natural one-to-one correspondence with  $R \times R$ .

It seems to follow that either  $R$  is very small or we have a density increment on a different Cartesian product (of  $R$  with its complement or of the complement with itself). If  $R$  is often very small, then by considering the extra  $d_2$  above and averaging, one might hope to get some kind of global non-quasirandomness, whereas if  $R$  is often large we might hope to find global non-quasirandomness for the reason outlined in the above paragraph.

A key point in all this is how to argue that one can patch together all the little bits of non-quasirandomness in order to obtain a big bit. I’ll think about that in another comment.

47.  [Terence Tao](#) Says:

[February 10, 2009 at 8:24 pm](#) | [Reply](#)

443.

Dear Mathias,

Yes, we have to take  $m$  to be a rapidly growing function of  $1/\delta$  to avoid the problems you describe. But I don’t think this is necessarily a problem for two reasons. Firstly, the hypergraph is going to be on about  $n/3$  vertices, with  $n$  extremely large compared to both  $m$  and  $1/\delta$ , so hypergraph regularity is still going to say something nontrivial. Secondly, I am not planning to rely solely on hypergraph regularity to finish the job; I envision something along the lines of “the hypergraph counting lemma (combined with the Sperner argument) will produce a combinatorial line unless one of the hypergraphs here is not  $\varepsilon(m, \delta)$ -regular”, and then the hope is to use this irregularity (which is some sort of assertion of low influence, I think) to locate a density increment or some other structural control on the original set in  $[3]^n$ .

The point is that Sperner+Cauchy-Schwarz seems to eliminate the role of  $\{1, 2\}^m$  from the task of detecting combinatorial lines in  $\{0, 1, 2\}^m$ , leaving only the need to understand the distribution of  $\mathcal{A}$  in positions containing at least one zero. This is the type of situation in which hypergraph regularity/counting/removal theory kicks in - we want to understand solutions to a system of equations, each of which only involves a proper subset of the variables.

Tim: I think we are trying to do almost exactly the same thing, though instead of working with a random embedding of the big triangular grid  $\{(a, b, c) : a + b + c = n\}$ , I am working with a random embedding of the much smaller grid  $\{(a, b, c) : a + b + c = m\}$ , which I like to think of as pairs  $(i, j) = (a, a + b)$  with  $1 \leq i \leq j \leq m$ . To embed this grid into  $[3]^n$ , one has to pick a random base point  $x$ , which has 0s at some set  $A$  of size about  $n/3$ , and then pick  $m$  indices  $a_1, \dots, a_m$ . The pair  $(i, j)$  then embeds into the point  $x$  with the coordinates at  $a_1, \dots, a_i$  increased from 0 to 1, and the coordinates from  $a_{j+1}, \dots, a_m$  increased from 0 to 2, creating the point that I called  $x_{i,j}$  previously.

You observed that in the big triangular grid, that  $(r,0,n-r)$ ,  $(s,0,n-s)$ ,  $(r,s-r,n-s)$  cannot all embed into  $A$ . In my notation, if one lets  $E_{i,j}$  be the event that  $x_{i,j} \in A$ , this is the assertion that for any  $1 \leq i < j \leq m$ , the events  $E_{i,i}, E_{j,j}, E_{i,j}$  cannot all be true. On the other hand, we expect each of the events  $E_{i,j}$  to occur with probability  $\delta$ , which means that

with high probability, the  $E_{i,j}$  must avoid a dense Cartesian product, and thus the joint event  $E_{i_1,j_1} \wedge E_{i_2,j_1} \wedge E_{i_1,j_2} \wedge E_{i_2,j_2}$  for  $1 \leq i_1 < i_2 < j_1 < j_2 \leq m$  must occur with probability noticeably higher than  $\delta^4$ .

But the key is that each of the  $E_{i,j}$  only depend on a proper subset of the random variables  $a_1, \dots, a_m$ , and this is the place where the hypergraph counting lemma should step in and say something useful.

48.  [gowers](#) Says:

[February 10, 2009 at 9:19 pm](#) | [Reply](#)

444. Sperner and nonquasirandomness.

I want to do a little exercise. It goes like this. Suppose I take random chains as in the proof of Sperner, and I find that the average number of elements of  $\mathcal{A}$  in each chain is  $\delta n$ , but the average of the square of this number is not about  $\delta^2 n^2$ . What does that tell us about  $\mathcal{A}$ ?

One obvious thing it tells us is that the number of pairs  $A \subset B$  in  $\mathcal{A}$  is larger than expected—at least if one interprets “number of pairs” is interpreted appropriately. Let’s just quickly see what we actually get. If I pick a random pair  $A \subset B$  (not necessarily in  $\mathcal{A}$ ) by choosing a random maximal chain and then a random pair of elements of it, then what is the probability that I pick the particular pair  $A \subset B$ ? Let’s assume that  $|A| = r$  and  $|B| = s > r$ . Then the probability that the chain passes through  $A$  is  $\binom{n}{r}^{-1}$ , and the probability that it passes through  $B$  given that it passes through  $A$  is  $\binom{n-r}{s-r}^{-1}$ . Finally, the probability that I pick the pair  $(A, B)$  given that the chain passes through both  $A$  and  $B$  is  $\binom{n}{2}^{-1}$ . So the weighting of pairs is according to the following model: first pick a random pair of slices (according to the uniform distribution) and then pick a random pair  $A \subset B$  from that pair of slices.

Thus, the hypothesis that  $\mathcal{A}$  has too many pairs is the hypothesis that if you pick a random pair  $A \subset B$  as in the above paragraph, then there is a more than  $\delta^2$  chance that both  $A$  and  $B$  belong to  $\mathcal{A}^2$ .

Can we use this to get a global conclusion? In a non-quasirandom bipartite graph the trick is to choose a random edge and take the neighbourhoods of its vertices. Since there are too many 4-cycles, the resulting induced bipartite graph is on average too dense. What would be the analogue here?

The obvious one would be to choose a random point of  $\mathcal{A}$  and take the set of all pairs  $B$  such that  $B \subset A$  or  $A \subset B$ . This gives a structure (one might visualize it as a sort of “double cone”) that typically contains more points of  $\mathcal{A}$  than it should. But this doesn’t count as a global obstruction because it is absolutely tiny compared with  $[2]^n$ . So the next question is whether it is possible to have a set of density  $c$  (according to the slices-equal weighting of  $[2]^n$ ) such that (i) it intersects many of these “double cones” in density significantly different from  $c$  and (ii) one cannot obtain a density increase on a subspace in any simple way.

I have to bath my 15-month-old now, so will pursue this later if nobody else has.

49.  [jozsef](#) Says:

[February 11, 2009 at 12:22 am](#) | [Reply](#)

445. Tim, I’m just reading your exercise, It might be more relevant to us to consider sets of size  $n/2 \pm \sqrt{n}$ . In general, I don’t see why do you have  $\delta n$  and not  $\delta \sqrt{n}$  elements in a random chain.

50.  [Terence Tao](#) Says:

[February 11, 2009 at 12:22 am](#) | [Reply](#)

446. Sperner and nonquasirandomness

Tim, I think in order to get a meaningful answer one has to localise the chains somewhat. Currently, one can get a

ridiculously large number of pairs  $A \subset B$  from a very small  $\mathcal{A}$  by using the extreme ends of the chain, for instance by letting  $\mathcal{A}$  be those  $A$  for which  $|A| > 2n/3$  or  $|A| < n/3$ .

To move things closer to my proposals in 439, 443, I would instead do this: pick a random  $A$  in  $[n]$ , then pick  $m$  random points  $b_1, \dots, b_m$  in the complement of  $A$  (where  $m$  is bounded in  $n$ , e.g.  $m = \lfloor \delta^{-100} \rfloor$ ), and then look at the random variable  $X$ , defined as the number of  $1 \leq i \leq m$  such that  $A \cup \{b_1, \dots, b_i\}$  lies in  $\mathcal{A}$ . This variable  $X$  has expectation about  $\delta m$ ; what can we say if it has significant variance?

51.  [gowers](#) Says:

[February 11, 2009 at 12:30 am](#) | [Reply](#)

447. Obstructions to Sperner.

I feel that we haven't thought about this hard enough. Boris's remark in 418 that obstructions to Sperner (=DHJ(1,2)) don't follow from lower-level obstructions is something we still need to come to terms with.

It seems that one source of obstructions is “usually continuous” sets, by which I mean sets where if you pick a random sequence and change a random coordinate then you will usually not change whether that sequence belongs to the set. Except that as I write that I see that there is absolutely no reason to believe that it is true. As you go up a random chain, even if you only rarely switch from belonging to the set to not belonging, or the other way, you may well get the expected density by the end. (Note that all these discussions are very different if we take the uniform measure on  $[2]^n$ , but I am trying to push the slices-equal measure.)

If we let  $\mathcal{A}$  be the collection of all sets that contain 1, then what happens? The density is  $1/2$ , but this is not quite as trivial a fact as usual: it follows from the fact that the bijection from a set to its complement is measure preserving. If we pick a random pair  $A \subset B$  then there's obviously a correlation between  $A$  containing 1 and  $B$  containing 1 — indeed, it's about as extreme a correlation as one could hope for. What's more, we can think about it like this: if we pick a random chain, then we will add in the element 1 at a random moment between 1 and  $n$ , so the intersection of  $\mathcal{A}$  with the chain has a density that is uniformly distributed between 0 and 1, which means that the mean square density of  $\mathcal{A}$  in a random chain is  $1/3$ , rather than the  $1/4$  that it is supposed to be.

I now see that what I'm groping for is something like this: we want more or less any monotone set system that isn't too much like the collection of all sets of size at least  $n/2$ . What I mean by that is the following. If  $\mathcal{A}$  is a monotone set, then given any chain, it will start outside  $\mathcal{A}$  and then end up inside. So the density of the intersection is determined by the point at which your set first joins  $\mathcal{A}$ .

So one method of building an obstruction would be to take a reasonably dense subset  $\mathcal{B}$  of  $\binom{[n]}{n/3}$ , say, and take all sets that contain a set in  $\mathcal{B}$ . But this isn't going to give a set of density  $1/2$  unless the sets in  $\mathcal{B}$  are very bunched together, or else their upper shadows are very soon going to be pretty well everything.

I know this has already been discussed, but it feels as though it will be hard to come up with examples that aren't based on heavy bias (or influence if we stick to these monotone examples, or similar). Boris had something for two layers, but what happens if we are forced to look at several layers, as we are with the slices-equal measure? It's tempting to look at Fourier expansions, as Gil suggested in 29.

52.  [gowers](#) Says:

[February 11, 2009 at 12:44 am](#) | [Reply](#)

448. Obstructions to Sperner.

My response to both Jozsef and Terry is the same. I'm very deliberately not localizing, because I'm hoping to use a different measure in DHJ, where you choose a point by first choosing a random slice according to the uniform

distribution and then choosing a random point in the slice.

There are several motivations for that. First of all, it is exactly what one does when proving Sperner via random chains: you get the stronger result that you just need the sum of the densities in all the layers to be greater than 1. In a sense, it is rather artificial to go on and say that the middle layer is the largest.

Secondly, if you do this then there is a real chance that the distribution of a random point, given that it belongs to a randomly chosen combinatorial line, is not horribly different from the distribution of a random point full stop. All this I said in my comment near the end of the 300s thread. (To be precise, it's [comments 365 and 366](#).)

Thirdly, proving DHJ for this measure is equivalent to proving it for the uniform measure, by the usual averaging trick. (I confess that I haven't actually checked this, but it is surely true.)

So when Terry says that the sets of size at least  $2n/3$  or at most  $n/3$  form a ridiculously small set system, I say no they don't: that set system has density  $2/3$  in the slices-equal measure.

Similarly, when Jozsef says that a set of density  $\delta$  might well intersect a random chain in only  $\delta\sqrt{n}$ , he is talking about the uniform measure. But in the equal-slices measure the density of a set is *defined* to be the expected density of the intersection with a random chain!

Having said all that, there may well be other objections to my proposals.

*[Off to bed now. I look forward to plenty of reading material when I wake up ...]*

53.



[Terence Tao](#) Says:

[February 11, 2009 at 4:25 am](#) | [Reply](#)

449. Obstructions to Sperner

Dear Tim, I'm leery of working globally because one is going to encounter things like Boris's counterexample 412 again. Also, while I suppose one could work with all  $n$  layers  $\Gamma_{a,n-a}$  of  $[2]^n$  as you are proposing, and even give them all equal weight, this seems very strange to me, since we know that we may without loss of generality restrict to the middle  $O(\sqrt{n})$  (or more precisely  $O(\sqrt{n \log \frac{1}{\delta}})$ ) layers, since all the other layers add up at most to a set of density  $\delta/2$  and thus can be discarded without loss of generality at the very first step (at the trifling cost of replacing the density  $\delta$  with  $\delta/2$ ).

Nevertheless, the idea of trying to work out what the consequences are of knowing that  $\mathcal{A}$  has "low influence" in the sense that flipping one digit does not change membership in  $\mathcal{A}$  as often as it should, looks like a promising one. I think my own arguments in 443 say that we are done unless the influence (or something like the influence) is somewhat less than the influence of the random set, though the influence is not so low that we automatically get huge subspaces.

More thoughts on this later...

54.



[Terence Tao](#) Says:

[February 11, 2009 at 5:08 am](#) | [Reply](#)

450 DHJ(j,k)

To answer Gil's question in 438, here is what I understand DHJ(j,k) to mean:

DHJ(j,k): Every dense subset of  $[k]^n$  which has "complexity at most j" contains a combinatorial line (of length k).

Now, what does "complexity at most j" mean? We're still negotiating a precise definition, I think, but here was one

proposal. We can think of an element  $x$  of  $[k]^n$  as a partition of  $[n]$  into  $k$  sets  $A_1, \dots, A_k$ . Now suppose we only considered  $j$  of these sets, say the first  $j$   $A_1, \dots, A_j$  - call this a  $j$ -shadow of  $x$ , so that there are  $\binom{k}{j}$  different  $j$ -shadows of  $x$ . When  $j=k-1$ , each  $j$ -shadow of  $x$  can be used to recover  $x$  completely, but for smaller values of  $j$ , each  $j$ -shadow only gives partial information as to what  $x$  is.


Now, let us call a set  $\mathcal{A} \subset [k]^n$  *sub-basic of complexity at most  $j$*  if membership of a point  $x \in [k]^n$  in  $\mathcal{A}$  can be defined purely in terms of a single  $j$ -shadow of  $x$  (e.g. the first  $j$  sets  $A_1, \dots, A_j$ .) Example: the set  $\mathcal{A}$  of all strings with exactly  $a$  3s is a sub-basic set of complexity 1. the set of all strings with exactly  $b$  occurrences of the substring 31 is a sub-basic set of complexity 2.

Observe that every subset of  $[k]^n$  is a sub-basic set of complexity  $k-1$ .

Define a *basic set of complexity at most  $j$*  to be the intersection of sub-basic sets of complexity  $j$ . Example: the set  $\Gamma_{a,b,c}$  of strings in  $[3]^n$  with  $a$  1s,  $b$  2s, and  $c$  3s is a basic set of complexity 1. (My notation here is coming from point set topology, in particular the notion of a sub-base and base of a topology.)

Finally (and continuing the topological analogy) one could define a set of complexity  $j$  to be the union of a bounded number of basic sets of complexity  $j$ .

With these conventions,  $\text{DHJ}(k-1, k)$  is the same as  $\text{DHJ}(k)$ .  $\text{DHJ}(1, 3)$  is  $\text{DHJ}(3)$  specialised to those sets  $\mathcal{A}$  of the form  $\mathcal{A} = \mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{A}_3$ , where membership of a string  $x \in [3]^n$  in  $\mathcal{A}_i$  is determined entirely by the  $i$ -set of  $x$ .

55.  [Terence Tao](#) Says:  
[February 11, 2009 at 5:31 am](#) | [Reply](#)

451 01-insensitivity

One key insight in Shelah's proof of Hales-Jewett is that if you have a colouring of 012 strings which is "01"-insensitive in the sense that flipping a 0 to a 1 or vice versa does not affect the colour of the string, then the  $k=3$  HJ theorem collapses to the  $k=2$  HJ theorem.

In a similar spirit, if a dense subset  $\mathcal{A}$  of  $[3]^n$  is mostly 01-insensitive in the sense that flipping a random 0 entry of a random element of  $[3]^n$  to a 1 (or flipping a random 1 entry to 0) is almost certain not to change the membership status of that string in  $\mathcal{A}$ , then (by looking at a randomly chosen medium-dimensional slice of  $\mathcal{A}$ ) we can reduce this instance of  $\text{DHJ}(3)$  to  $\text{DHJ}(2)$ . This notion of "mostly 01-insensitive" is close to Tim's notion of "usually continuous" from 447 or the notion of "low (average) influence" we have been discussing previously, but note that it is only about 0 and 1; it doesn't say anything about what happens if a 0 is flipped to a 2, for instance.

After talking with Tim Austin a bit about how the Furstenberg-Katznelson proof proceeds, I am beginning to think that instead of just incrementing density, we may also have to increment "01-insensitivity". I think I can push my arguments in 413, 439 to say something along the lines of "either we have enough quasirandomness to find a combinatorial line, or else there is a little bit of 01-insensitivity present". By a "little bit" of insensitivity, I mean that if we take a random string  $x$ , and then flip one of the 0s of  $x$  randomly to a 1 to create a random string  $y$ , then there is a small amount of correlation (something like  $\delta^{O(1)}$ ) between the events  $x \in \mathcal{A}$  and  $y \in \mathcal{A}$ . [I am oversimplifying here; I need a more generalised notion of insensitivity, but this is the flavour of what I think is possible.]

At present I don't see how to iterate this observation (though I can imagine bringing in existing Ramsey theorems, and perhaps also multidimensional Szemerédi, to assist here), to try to increase the 01-insensitivity through repeated passage to sub-objects until we get to the point where Shelah's observation lets us drop from  $k=3$  to  $k=2$ . Tim Austin tells me that this is broadly how the ergodic theory proof of DHJ proceeds; at some point I should go read that paper more thoroughly and see this for myself.

56. [Boris](#) Says:



[February 11, 2009 at 5:53 am](#) | [Reply](#)

450. In #444 Tim asks what can we conclude if the average number of elements in the chain is  $\delta n$ , but the average of the square is much different from  $\delta^2 n^2$ . My hunch is that it should imply that there is a pair  $(i, j) \in \binom{[n]}{2}$  such that swapping the order of  $i$  and  $j$ . The intuition is that if it is not the case, then some kind of Azuma-type or isoperimetric inequality should establish concentration. This is probably very similar to what Terry suggests in #449.

My experience working on forbidden subposet generalization of Sperner suggests Tim is probably right in wanting to weight everything. The reason for this is that even if we working with a few levels near the middle level, then though the vertices (=the sets) are are weighted approximately the same in all of them, the edges (=the pairs  $A \subset B$ ) are weighted very differently according to the distance between the levels that contain them. The weight scales like  $n^{|A|-|B|}$ .

The reason why weighting edges is unavoidable is that if suppose we know by density considerations that there are pairs  $A \subset B$  in our set family, but we want to give a bound on the number of pairs  $A \subset B$ . Then the optimal bound will depend strongly on how far the levels on which the family lives are from one another.

So, even if we localize to a stack of layers of size  $\sqrt{n}$ , we probably will have to weight edges anyways.

Unfortunately, I cannot see anything resembling chains for  $[3]^n$ .



57. [Terence Tao](#) Says:

[February 11, 2009 at 6:11 am](#) | [Reply](#)

451.

Boris, I think the random model I have in 439 (in which one is looking at random embeddings of  $[3]^m$  into  $[3]^n$ , and then analyses correlations between elements of the resulting random subset of  $[3]^m$ , rather than deal directly with the original set in  $[3]^n$ ) seems to do the edge weighting you want automatically. (It also fits well with the ergodic theory approach to these sorts of problems.)

The analogue of the chains for  $[3]^n$  (or for the embedded copy of  $[3]^m$ ) seem to be the strings consisting of a connected string of 1s, 0s, and 2s. For instance we get triangles such as this one in  $[3]^4$ :

```
0000 0002 0022 0222 2222
1000 1002 1022 1222
1100 1102 1122
1110 1112
1111
```

note that every right-angled triangle with the hypotenuse on the diagonal of the above grid is a combinatorial line. My strategy in 439 is to look at all the random embeddings of this type of picture into  $[3]^n$ , and then use Cauchy-Schwarz to eliminate the role of the diagonal and only focus on the interior (in which all strings have at least one 0 in them, which sets one up for hypergraph counting lemmas). Tim Austin tells me that this is related to the van der Corput type lemmas used in the ergodic proof of DHJ, which is an encouraging sign for me.



58. [gowers](#) Says:

[February 11, 2009 at 11:30 am](#) | [Reply](#)

452. Slices-equal weights.

I'm not wedded to my suggestion of putting a different measure on  $[3]^n$  (or  $[2]^n$  for that matter) but I think the idea has enough going for it that I don't want to drop it unless I see a truly convincing argument for doing so. And at the moment I don't.

First of all, no argument that is given *from the perspective of the uniform measure* can be an argument against using a different measure. One has to step outside both. So this informal argument of Terry doesn’t persuade me:

*Also, while I suppose one could work with all layers of as you are proposing, and even give them all equal weight, this seems very strange to me, since we know that we may without loss of generality restrict to the middle  $O(\sqrt{n})$  (or more precisely  $O(\sqrt{\log(\delta^{-1})n})$  layers, since all the other layers add up at most to a set of density  $\delta/2$  and thus can be discarded without loss of generality at the very first step (at the trifling cost of replacing the density  $\delta$  with  $\delta/2$ ).*

Of course if you are working with the uniform measure you can discard everything that isn’t near the middle, but if you are working with the slices-equal measure then you can’t. So the equal-slices measure is different, but that doesn’t make it “very strange”, any more than it’s very strange to use the uniform measure because WLOG you can restrict to a set (if you’re a slices-equal person) that keeps away from the middle layer. To summarize: no argument that basically boils down to saying that the slices-equal measure is different from the uniform measure carries any weight: it’s supposed to be different!

A more serious point raised by Terry is the general one that global arguments are vulnerable to Boris-type counterexamples. But, as discussed in 428, 429, 431, 432, it is not completely obvious that that is the case with the slices-equal measure. The reason is that then the wildcard sets become (I’m fairly sure, but must check) of linear size. What’s more, I think the standard deviation of the size of the wildcard set is linear too. So it seems to me that the phenomena that Boris was exploiting don’t arise with the slices-equal measure.

I was therefore allowing myself to hope that conjecture 411 might actually be true for the slices-equal measure.

However, I am still disturbed by Boris’s point that the analogue of this one level down seems to be false (since the number of  $A \subset B$  pairs in Sperner doesn’t just depend on the density). That is why I wanted to think hard about what the obstructions to uniformity are for Sperner with the slices-equal measure.

I just want to add, once again, that the slices-equal measure has one very strong argument going for it, from the point of view of what is natural, which is that all (I think) the proofs of Sperner prove something about the slices-equal measure and only at the very end convert that into a statement about the uniform measure.

One reason for resisting the slices-equal measure is that it is less symmetric than the uniform measure. However, the formulations of Sperner/DHJ are not symmetric either, or rather the symmetries they do have are also symmetries of the slices-equal measure.

Another argument for its naturalness is that it seems rather strange to look only for lines with very small wildcard sets, when almost all lines have linear-sized wildcard sets. With the uniform measure, this is forced, but that is because of the unfortunate fact that the measure of a typical point on a typical line is radically different from the measure of a typical point. Note that this is an argument for the unnaturalness of the uniform measure that is taken from an *external* perspective: we would like a measure that doesn’t get completely distorted when you condition on belonging to a random combinatorial line, we then look at the two measures, and we find that the slices-equal measure does better in that respect.

I’m not at all dismissing Terry’s very interesting suggestions for dealing with things locally. It’s more that I would love to do something global if it turned out to be possible, so I’d like to invite people to attempt to find reasons for its not being possible. For example, if there is an obstruction to uniformity that’s not of type 411 then I may well be forced to abandon my hopes of a global argument. If no such reason is forthcoming, then I think we should keep global arguments in play.

If it turns out that the only obstructions to Sperner (in the slices-equal set-up) are ones that depend on a tiny number of variables, then we might be lucky and find that Conjecture 411 is true. At first that seems strange, when it’s false one level down. But the point is that the assumption of disjointness that we have all over the place concerns *pairs* of sets, at whatever level one is, so it could be that Sperner is exceptional because it occurs at a level below the level of the

disjointness restriction. In other words, examples that depend on just a few variables affect you at all levels, but when you get to DHJ they can be absorbed into Conjecture 411, whereas for Sperner they can't.

This is, as usual, an optimistic suggestion, and as usual I welcome a cold blast of realism if anyone can supply it.

59.  [Gil Kalai](#) Says:

[February 11, 2009 at 2:42 pm](#) | [Reply](#)

453. (Clarification to 452) In particular, do you conjecture that a dense set in the slices-equal measure must contain a combinatorial line? This is stronger than DHJ, right? (Can you prove it as a consequence of DHJ?)

(Just to make sure I understood correctly: slice equal measure is this: you give equal weight to all solutions of  $a+b+c=n$  and then split this weight uniformly for all sequences of  $a$  '0's  $b$  '1's and  $c$  '2's. Indeed this natural measure appears in various places.)

60.  [Ryan O'Donnell](#) Says:

[February 11, 2009 at 3:04 pm](#) | [Reply](#)

454. Re Sperner.

It was suggested in #446 to pick a random  $x_0 \in \{0, 1\}^n$ , form  $x_1, \dots, x_m$  by taking a random "upwards" walk of length  $m = 1/\delta^{100}$ , and then look at the random variable  $X = \#\{i : x_i \in \mathcal{A}\}$ , where  $\mathcal{A} \subseteq \{0, 1\}^n$  has density  $\delta$ . Although the  $x_i$ 's are not uniformly distributed, they are very close to being so, since  $m$  is small. Hence the expectation of  $X$  should be very nearly  $\delta m$ .

Consider the case that  $\mathcal{A}$  is a monotone subset ("up-set"). Then  $X$  will be very nonconcentrated. Indeed, conditioned on  $x_0 \in \mathcal{A}$  (probability  $\delta$ ), so are all the  $x_i$ 's. I.e., in this case,  $X$  is  $m$  with probability  $\delta$  and therefore nearly 0 with probability  $1 - \delta$ .

On the other hand, it's not clear that this is an immediate problem for the plan; in #439, one may well be even happier to have  $|I| = m$  with  $\delta$  probability.

61.  [Gil Kalai](#) Says:

[February 11, 2009 at 3:32 pm](#) | [Reply](#)

455. If we want a ridiculously-hyper-optimistic statement extending LYM to 3-letters alphabet we can take this: Consider all blocks  $B$  of slices  $\Gamma_{a,b,c}$  not containing a combinatorial line. Let  $S$  be a subset of  $\{0, 1, 2\}^n = \Gamma$  without a combinatorial line, then the sum of densities of  $S$  in some weighted collections of blocks which form a fractional covering of  $\Gamma$  is at most 1.

62.  [Gil](#) Says:

[February 11, 2009 at 4:03 pm](#) | [Reply](#)

456. There is another finer notion of slices which may be a righter extensions from the  $k=2$  case. Here you grade the elements of  $\Gamma$  according to words in '0' and '1' that you obtain by erasing the '2's. It would be nice to see (estimate) what is the maximum size of unions of such fine slices without a combinatorial line.

63.  [Terence Tao](#) Says:

[February 11, 2009 at 4:18 pm](#) | [Reply](#)

457. Slices-equal measure

It occurs to me that there may in fact be essentially no difference of opinion between Tim's "global, slices-equal measure" viewpoint and my "local, uniform-measure" viewpoint, because in order to get from a slices-equal measure DHJ to a uniform measure DHJ, one has to average on local subspaces anyway. (Note that one can easily construct sets of large uniform measure with small slices-equal measure, e.g. the strings  $x \in [2]^n$  with  $n/2 + O(\sqrt{n})$  1s, and conversely, so localisation is necessary here.)

64.  [Ryan O'Donnell](#) Says:

[February 11, 2009 at 5:20 pm](#) | [Reply](#)

458. Sorry for the double-post above. *[Now sorted out -- Tim.]* I guess it was already clear from the first para of Terry's #451 that if  $\mathcal{A}$  is, say, "01-monotone", meaning that changing a 1 to a 0 cannot take you out of  $\mathcal{A}$ , then DHJ holds for  $\mathcal{A}$ , and even with the lower density  $\approx 1/\sqrt{n}$  of Sperner. This already covers a number of my previous obstruction examples; e.g., having runs of 0's or having plurality digit 0.

65.  [gowers](#) Says:

[February 11, 2009 at 5:40 pm](#) | [Reply](#)

459. Slices-equal measure

Terry, I was going to say something similar. It seems to me that localization and slices-equal measure are different ways of forcing typical lines to contain typical points. But the difference is more of a technical one than a fundamental one. I am hoping that slices-equal measure will turn out to be technically rather convenient, but I haven't done enough calculations with it to be certain that it will be.

Gil, I'm pretty sure that DHJ for the slices-equal measure is an easy consequence of DHJ, and vice versa. Here's a sketch of how to prove that DHJ for any reasonably continuous measure (meaning that the weight of a point is usually very similar to the weight of a neighbouring point) is equivalent to the usual DHJ. You just pick some  $m$  such that  $1 \ll m \ll n$  and pick a random subspace by randomly fixing  $n - m$  coordinates and letting the remaining  $m$  vary. Suppose  $\mathcal{A}$  has density  $\delta$  in measure 1. Then the average density in one of these subspaces is roughly  $\delta$ . Moreover, by hypothesis the measure in almost all of the subspaces is almost uniform, so we can apply the normal DHJ to get it for the funny measure. And in the reverse direction, if we regard each tiny subspace as a copy of  $[3]^m$  and apply the funny measure to it, then the average of all these funny measures will be roughly uniform, so once again we can find a subspace where  $\mathcal{A}$  is dense with respect to the funny measure. (Note that I needed continuity of the funny measure only in one direction of this equivalence.)

Gil, I'm not sure I understand your hyper-optimistic conjecture, but let me give one of my own. Apologies if it's the same. Let  $\alpha(n)$  be the minimal density that guarantees a non-trivial aligned equilateral triangle in the set of non-negative  $(a,b,c)$  such that  $a+b+c=n$ . Then if the slices-equal measure of  $\mathcal{A}$  is greater than  $\alpha$ ,  $\mathcal{A}$  must contain a combinatorial line. This conjecture, if true, is trivially best possible since a corner-free union of slices contains no combinatorial line. It is also a direct generalization of Sperner. I wonder if there's some simple way of seeing that it is false.

66.  [Terence Tao](#) Says:

[February 11, 2009 at 6:16 pm](#) | [Reply](#)

460. The ergodic perspective

I think I now see more clearly the way to put my approach and Tim's approach on a common footing, and that is using the lessons drawn from the ergodic theory approach to density Ramsey problems that proceeds via the Furstenberg correspondence principle.

Now before everyone runs away screaming, let me try to explain the Furstenberg philosophy without actually using the words "ergodic", "limit", "infinite", or "measure space" (though I will need "random" as a substitute for "measure

space”).

In a nutshell, the philosophy is this:

\* Study a global dense object by looking at a random local subobject. This trades a globally dense deterministic object  $\mathcal{A}$  for a *pointwise* dense random object  $\mathcal{A}'$ . Furthermore, this random object obeys various symmetry (or more precisely, *stationarity*) properties.

So, one gets much more structure on the object  $\mathcal{A}'$  being studied, provided that one lets go of the idea of thinking of this object deterministically, and instead embracing its random nature.

Let me give some concrete examples of this philosophy. Firstly, Roth's theorem in  $\mathbb{Z}/N\mathbb{Z}$ , with  $N$  assumed prime for simplicity. To locate length three progressions in a set  $A \subset \mathbb{Z}/N\mathbb{Z}$ , what one can do is pick a medium size number  $m$ , and look at a random arithmetic progression  $a, a+r, \dots, a+(m-1)r$  of length  $m$  in  $\mathbb{Z}/N\mathbb{Z}$ . One can then pull back the deterministic set  $A$  by this progression to create a random subset  $A' := \{j \in [m] : a+jr \in A\}$  of  $[m]$ . If we can find a progression of length 3 in  $A'$  with positive probability, then we get a progression of length 3 in  $A$  (in fact, we get lots of such progressions; this is essentially the proof of Varnavides theorem).

The original deterministic set  $A$  had global density  $\delta$ . Because of this, the random set  $A'$  has a better property: it has a *pointwise* density  $\delta$ , in the sense that every  $j$  in  $[m]$  belongs to  $A'$  with probability  $\delta$ . From linearity of expectation this implies that the expected density of  $A'$  with respect to uniform measure of  $[m]$  is  $\delta$ , but it is far stronger - indeed, it implies that the expected density of  $A'$  with respect to *any* probability measure on  $[m]$  is  $\delta$ .

Similarly, the deterministic set  $A$  has no symmetry properties:  $A+1$ , for instance, may well be a completely different set from  $A$ . However, the random set  $A'$  has an important symmetry property: the random set  $A'+1$ , while being different from  $A'$ , *has the same probability distribution* as  $A'$  (ignoring some trivial issues at the boundary of  $[m]$ ; in practice one eliminates this by taking the limit  $m \rightarrow \infty$ ). In the language of probability theory, we say that the probability distribution of  $A'$  is *stationary*; in the language of ergodic theory, we say that the transformation  $T : A' \rightarrow A' + 1$  is a *measure-preserving transformation* with respect to the probability distribution  $\mu$  of  $A'$ .

Now, we know that the events  $0 \in A', 1 \in A', 2 \in A'$  each occur with probability  $\delta$ . If they were independent, then we would have  $0, 1, 2 \in A'$  with probability  $\delta^3$  and we would get Roth's theorem. But of course they could be correlated - but one can try to quantify this correlation by a variety of tools, e.g. using the spectral theory of the measure-preserving transformation  $T$  (this is the analogue of using Fourier analysis in the original combinatorial setting). This is the starting point for the ergodic theory method for solving density Ramsey theory problems.

Now we return to Hales-Jewett. Similarly to above, if we wish to study a subset  $\mathcal{A}$  of  $[3]^n$  which has density  $\delta$  wrt uniform measure, we can take a random  $m$ -dimensional subspace of  $[3]^n$  (where I will be a little vague for now as to exactly what “random” means; there are a variety of options here) and pull back the deterministic  $\mathcal{A}$  to get a random subset  $\mathcal{A}'$  of  $[3]^m$ . Again, finding lines in  $\mathcal{A}'$  will generate lines in  $\mathcal{A}$  (and there will be an appropriate Varnavides statement).

Also as before, the global density of  $\mathcal{A}$  ensures that the random subset  $\mathcal{A}'$  has *pointwise* density very close to  $\delta$  (there is a slight distortion here, as pointed out in Ryan.454, but it is minor). Thus, the expected density of  $\mathcal{A}'$  with respect to *any* particular measure on  $[3]^m$  - be it uniform, slices-equal, or anything else - is basically  $\delta$ .

It is at this point that the three approaches to the problem (ergodic approach, Tim's approach, and my approach) begin to diverge.

The ergodic theory approach would then proceed by studying correlations between events such as “0000  $\in \mathcal{A}'$ , 1111  $\in \mathcal{A}'$ , 2222  $\in \mathcal{A}'$ . If these events are independent, we are done; otherwise, we hope to discern some usable structural information on  $\mathcal{A}'$  out of this. The original set  $\mathcal{A}$  is not used at all in the ergodic theory approach; it's only legacies are the pointwise density and stationary properties it bestows on  $\mathcal{A}'$ .

Tim's approach is proceeding by using the first moment method to locate a specific *deterministic* instance of  $\mathcal{A}'$  which

is dense wrt slices-equal measure, and then proceeding from there. As with the ergodic theory method, the original set  $\mathcal{A}$  (which was dense in the uniform measure, rather than slices-equal) is now discarded, and one is now working entirely with the slices-equal-dense  $\mathcal{A}'$ . As I understand it, Tim is then trying to show that lack of combinatorial lines in  $\mathcal{A}'$  will force some new structure on  $\mathcal{A}'$ .

My approach starts off similarly to the ergodic theory approach - studying correlations between events such as  $ijkl \in \mathcal{A}'$  - but the difference is that once a correlation is detected, I try to see what this tells me about the original set  $\mathcal{A}$ , in particular to show that it has some “low-influence”, “insensitivity”, “mostly continuous”, or “monotone” like properties. I have a vague idea of trying to iterate this procedure until we have enough structure on  $\mathcal{A}$  to either pass to a denser subspace or find combinatorial lines directly, but I have not worked this part out yet.



67. [Terence Tao](#) Says:

[February 11, 2009 at 6:28 pm](#) | [Reply](#)

461. Ergodic perspective cont.

Perhaps one more remark to illustrate the connection between my approach and Tim's approach. For sake of illustration I will take  $m=4$ , since I can then use the triangular grid  $\Delta_4$  from 451.

The construction in 460 gives a random subset  $\mathcal{A}'$  of  $[3]^4$ , with pointwise density essentially  $\delta$ . In particular, the equal-slices density of  $\mathcal{A}'$  in  $[3]^4$  has expectation about  $\delta$ , and similarly the uniform density of  $\mathcal{A}'$  on  $\Delta_4$  has expectation about  $\delta$ .

But because a permutation of a cube is still a cube, we see that applying any of the  $4!$  permutations of the indices to  $\mathcal{A}'$  will give a new random set with the same probability distribution as  $\mathcal{A}'$ ; thus  $\mathcal{A}'$  is stationary with respect to the action of  $S_4$  on  $[3]^4$ .

But observe that equal-slices measure is nothing more than the uniform measure on  $\Delta_4$  averaged by the action of  $S_4$  ( $\Delta_4$  intersects each slice in exactly one point). So the expected density of  $\mathcal{A}'$  wrt equal slices measure is necessarily equal to the expected density of  $\mathcal{A}'$  wrt uniform measure  $\Delta_4$ . But the fact that both of these densities are  $\delta$  is weaker than the statement that the pointwise density of  $\mathcal{A}'$  is  $\delta$ , which is what I think we should be using.



68. [Jason Dyer](#) Says:

[February 11, 2009 at 6:44 pm](#) | [Reply](#)

460. Taxonomy of slices

This is more a general-organization comment, but at some point could we get together a “taxonomy of slices”? I'm starting to lose track of all the different methods people are proposing.

I believe it also might be helpful to apply the slices side-by-side to a single example to get an impression of how the sets are different.



69. [gowers](#) Says:

[February 11, 2009 at 6:52 pm](#) | [Reply](#)

462. Ergodic perspective.

Terry, thanks for that useful summary. I think it's helped me to understand the ergodic side of things better. Let me test that, chiefly for my own benefit, by trying to put part of it in my own words. The idea would be that you take a random copy of  $[3]^m$  in  $[3]^n$  and intersect it with  $\mathcal{A}$ . If you do that, then instead of a function from  $[3]^m$  to  $\{0, 1\}$  you've got a Bernoulli random variable at each point of  $[3]^m$ , which is 1 with probability  $\delta$  (the density of  $\mathcal{A}$ ) and 0




with probability  $1 - \delta$ .

At first it seems as though one has gained precisely nothing by doing this. For example, here's a way of obtaining some random variables: you take a dense set  $B \subset [3]^m$ , and for each permutation  $\pi$  of  $[m]$  you let  $B_\pi$  be the set of sequences you get by taking a sequence in  $B$  and using  $\pi$  to permute its coordinates. Then the probability that the random set  $B_\pi$  contains a combinatorial line is trivially the same as the probability that  $B$  contains a combinatorial line (and moreover the expectation of the characteristic function of  $B_\pi$  is  $\delta$ ).

However, the point is that if you know that the random set  $B \subset [3]^m$  comes from random restrictions of a much larger set  $A \subset [3]^n$ , then you will be very surprised if with probability 1 it is always obtained from some fixed set by a simple permutation of the coordinates. A phenomenon like that would have the potential to show that  $A$  has some very strange and exploitable structure. More generally, if the elements of the random set exhibit significant dependences, this should come from structure in  $A$ , and if they are independent then as you point out you get a combinatorial line.

If that sketch is not too wide of the mark, then I see that extreme localization of the kind you propose does potentially give you quite a lot of extra resources. And given that the ergodic-theory proof is known to work it is always encouraging if one's argument has formal similarities with it. My feeling about slices-equal is that it's a bit of a long shot, but it might be possible to get away without that extra power and obtain a stronger result as a consequence. But if that doesn't work out, then we know that we've got localization to fall back on. Or we could have two parallel threads, one exploring a global approach and one a local approach. This one is already getting to the point where it takes a boring amount of time to load on to my computer, and the 300s thread seems to have come to a standstill for now.


Incidentally, it occurs to me that we as a collective are doing what I as an individual mathematician do all the time: have an idea that leads to an interesting avenue to explore, get diverted by some temporarily more exciting idea, and forget about the first one. I think we should probably go through the various threads and collect together all the unsolved questions we can find (even if they are vague ones like, "Can an approach of the following kind work?") and write them up in a single post. If this were a more massive collaboration, then we could work on the various questions in parallel, and update the post if they got answered, or reformulated, or if new questions arose.

70.  [Randall](#) Says:  
[February 11, 2009 at 7:01 pm](#) | [Reply](#)

463. Speaking of ergodic theory:

If anyone's interested in reading the ergodic proof for  $k=3$ , we've been discussing it in seminar in Memphis; the proof can be found on my webpage. (You need the 1989 paper of Furstenberg and Katznelson, in which the proof is set up but not given, too; that's on Katznelson's webpage.) It's far less intimidating than the general proof as given in their 1991 paper.

Is there a proof (of the kind you guys are looking for) of the IP Szemerédi theorem for  $k=3$ ? That would seem to be logically prior.

71.  [gowers](#) Says:  
[February 11, 2009 at 7:23 pm](#) | [Reply](#)

464. Varnavides.

Something I've been meaning to do for quite a while is prove, or at least sketch a proof, that a Varnavides-type statement holds for the slices-equal measure. I don't think it's going to be hard, but I want to check it to be sure. (It is supposed to be another argument in favour of slices-equal, though I can see that the local approach gives a local Varnavides statement, so it's not an argument that slices-equal is more natural than localization, but it is an argument that slices-equal is more natural than a global uniform approach.)

Let  $\delta > 0$ , and let  $m$  be large enough that every subset of  $[3]^m$  of density at least  $\eta = \eta(\delta)$  contains a combinatorial line. (Here,  $\eta$  is a constant to be chosen later.) Now choose a small constant  $c > 0$  (depending on  $m$ ) and embed  $[3]^m$  randomly into  $[3]^n$  as follows. First pick  $m$  integers  $r_1, \dots, r_m$  independently, randomly and uniformly from  $\{1, 2, \dots, cn\}$ . Next, randomly pick disjoint sets  $V_1, \dots, V_m$ , with  $|V_i| = r_i$ . Finally, randomly assign 1s, 2s and 3s to the complement of the union of the  $V_i$ . Sorry, I don't mean that. I mean that you randomly choose  $a + b + c = n - \sum r_i$  and then you randomly assign  $a$  1s,  $b$  2s and  $c$  3s to the coordinates in the complement of the union of the  $V_i$ . And now define a  $m$ -dimensional combinatorial subspace by fixing the coordinates outside the  $V_i$  according to the random assignment of 1s, 2s and 3s, and taking all possible  $3^m$  assignments to the rest of the coordinates that are constant on the  $V_i$ . (Since  $m$  is constant, I don't think it matters if one does this uniformly, but I would prefer to do even this in a slices-equal way. That is, I choose  $d + e + f = m$  randomly and then I randomly choose  $d$  of the sets  $V_i$  to get a 1, etc.)

There are a few statements that I don't immediately see how to argue for without going away and doing calculations. I claim that there is some constant  $\eta = \eta(\delta) > 0$  such that the proportion of copies that intersect  $\mathcal{A}$  in a density of at least  $\eta$  is at least  $\eta$ . I also claim that no combinatorial line is significantly more likely than average to belong to a copy. And from that I claim that a double-count gives that a random combinatorial line has a probability depending on  $\delta$  only of belonging to  $\mathcal{A}$ .

I might try to think about justifying those statements a bit better.

72.  [gowers](#) Says:

[February 11, 2009 at 7:56 pm](#) | [Reply](#)

465. Hyper-optimistic conjecture

Terry, re last paragraph of 459, it occurs to me that with your upper-and-lower-bounds thread you may already know that the hyper-optimistic conjecture is false. Do you know of any example that beats the best example that's constant on slices?

73.  [Terence Tao](#) Says:

[February 11, 2009 at 9:13 pm](#) | [Reply](#)

466. Various


Randall, thanks for the pointer to your notes (which, incidentally, are at <http://www.msci.memphis.edu/~randall/preprints/HJk3.pdf>). I will try to digest them.

It looks like it is indeed soon time to close this thread and open some new ones. I can host a thread on ergodic-flavoured combinatorial approaches, focusing on my localisation approach and also on trying to explicate the Furstenberg-Katznelson proof of DHJ (I am imagining some internet version of a "reading seminar"). Then I guess Tim could try to summarise the other main active thread (based on obstructions to Sperner, etc.; my approach, incidentally, doesn't seem to directly need to understand Sperner obstructions, though it does need to understand the closely related concept of low influence).

I will also try to translate the hyper-optimistic conjectures 455, 459 to something that the 200-299 thread can chew on. Gil, can you explain your conjecture in 455 in more detail? I'm afraid I didn't really get it (e.g. I don't know what a

“block” or a “fractional covering” is).

As for all the “background” of discarded questions (e.g. the IP-Szemerédi thing Randall mentioned was already brought up all the way back in József’s comment 2)... here I think this is a job that requires more manpower than we have right now, since the “foreground” threads are already keeping all of us very busy. It also feels like this stuff would be better placed in something like a wiki than in a blog thread. I would say for now that we just continue the “active” threads for now and leave the trawling through the remainder for a later point in time if our current active threads get stuck.


74.  [Terence Tao](#) Says:  
[February 11, 2009 at 9:18 pm](#) | [Reply](#)

467. Various cont.

Actually, what I might do is set up a Google Document to hold random threads, questions, etc. that can be collaboratively edited. This has worked pretty well for the 200-299 thread so far.

Here’s one little tidbit from that thread which might be useful here. Currently, the best  $\delta$  for which we can combinatorially prove DHJ(3) is 0.633. Thus, for instance, we don’t know without ergodic theory that if one takes 60% of  $[3]^n$  for large  $n$ , that one gets a combinatorial line. So, if one wants to play with a quantitative value of  $\delta$ : take 0.6.


We also have an example of a subset of  $[3]^{99}$  of density  $> 1/3$  that still has no combinatorial line. So even with moderately enormous dimension, it is not possible to handle  $\delta = 1/3$ .

75.  [Terence Tao](#) Says:  
[February 11, 2009 at 10:04 pm](#) | [Reply](#)

468. Ergodic perspective

Re Tim’s 462: yes, this is one way of thinking about the ergodic perspective, as providing a inter-related (and presumably correlated) network of Bernoulli random variables (aka “events”), and the game is to get enough control on how these events overlap each other that one can eventually find a line of three events which are simultaneously satisfied with positive probability.

By the way, another distinction between the three approaches lies in the range of  $m$  to pick, a point which has already been alluded to by Tim (when referring to my approach as “extreme localization”). In my approach I am picking  $m = m(\delta)$  to just depend on  $\delta$  (in fact I think I may be able to take a very concrete  $m$ , e.g.  $m = \lfloor \delta^{-100} \rfloor$ ). In the ergodic approach,  $m = m(n)$  is a very slowly growing function of  $n$  (so in the limit,  $m$  is infinite). For Tim’s approach, one can take  $m$  close to  $\sqrt{n}$  (or anything smaller, so long as it is still sufficiently large depending on  $\delta$ ).

76.  [Randall](#) Says:  
[February 11, 2009 at 10:30 pm](#) | [Reply](#)

469. IP Roth:

Just to be clear on the formulation I had in mind (with apologies for the unprocessed code) [*now processed -- Tim*]: for every  $\delta > 0$  there is an  $n$  such that any  $E \subset [n]^{[n]} \times [n]^{[n]}$  having relative density at least  $\delta$  contains a corner of the form  $\{a, a + (\sum_{i \in \alpha} e_i, 0), a + (0, \sum_{i \in \alpha} e_i)\}$ . Here  $(e_i)$  is the coordinate basis for  $[n]^{[n]}$ , i.e.  $e_i(j) = \delta_{ij}$ .

Presumably, this should be (perhaps much) simpler than DHJ,  $k=3$ .

77.  [Terence Tao](#) Says:

[February 11, 2009 at 10:45 pm](#) | [Reply](#)

469. Hyper-optimistic conjecture

Ah, I see now that Gil's conjecture is a weakening of Tim's conjecture in which one takes a different measure on the space of slices than the uniform measure. As Tim's is more concrete, that's the one I've posted on the 200-299 thread to be tested for small  $n$ .

78.  [Terence Tao](#) Says:

[February 11, 2009 at 11:27 pm](#) | [Reply](#)

470. Google document

OK, I've set up a repository for unsolved questions for Polymath1 at

[http://docs.google.com/Doc?id=dhs78pth\\_15dfw8jhz4](http://docs.google.com/Doc?id=dhs78pth_15dfw8jhz4)

Unlike the spreadsheet over at 200-299, I can't make this editable by the general public (I suppose there are security risks in documents that are not present in spreadsheets), but I'll add Tim as a "collaborator", and anyone else who asks can have permission to edit it also.

79.  [Ryan O'Donnell](#) Says:

[February 11, 2009 at 11:54 pm](#) | [Reply](#)

471. Re Terry's strategy (#439).

Hi Terry, could you help me understand the strategy a bit in the case that  $\mathcal{A}$  is, say, the set of strings in  $\{0, 1, 2\}^n$  whose 0-count is congruent to  $0 \bmod 1/\delta$ ? (Assume  $1/\delta$  is an integer.)

I'm worried about the first concentration of measure statement. Or does this set already have some exploitable blatant structure — perhaps being a 1-sub-basic complexity set?

Thanks!

80.  [Terence Tao](#) Says:

[February 12, 2009 at 12:23 am](#) | [Reply](#)

472. Terry's strategy 439

It's true, the first concentration of measure statement breaks down; either all the  $x_i$  lie in  $\mathcal{A}$  (when the number of 0s in  $x$  is equal to  $m \bmod 1/\delta$ ) or none of them do.

But it turns out we don't really need that particular concentration, it's enough to have a lot of  $x_i$  in  $\mathcal{A}$  with positive probability, which is true from the first moment method.

A more necessary concentration of measure is the assertion that the density of  $x_{i,j}$  that lie in  $\mathcal{A}$  is close to  $\delta$ . In this particular case, this is true: the set of pairs  $(i, j)$  with  $x_{i,j} \in \mathcal{A}$  is given by  $\{(i, j) : i + j = c \bmod 1/\delta\}$ , where  $c$  is some constant depending on  $m$  and  $x$ .

The second moment method tells us that we can get some concentration of measure (enough for our purposes, in fact) as soon as we have small pairwise correlations between the events  $E_{ij}$ , which are the events that  $x_{ij} \in \mathcal{A}$ . Now, in your example, there are extreme pair anti-correlations: for instance,  $E_{00}$  and  $E_{10}$  are never simultaneously true. To put it another way, if we take an element in  $\mathcal{A}$  and flip a 0 to 1, we *always* leave  $\mathcal{A}$  as a consequence.

This is certainly some sort of structure. But can we exploit it? Well, one thing I was thinking of is that in the setup in 439, the  $m$  different generators  $e_{a_1}, \dots, e_{a_m}$  of the embedded  $m$ -dimensional cube are just single basis vectors; to put it another way, I'm playing with a string with  $m$  distinct wildcards, with each wildcard being used exactly once. But one could instead have each of the  $m$  wildcards appear  $r$  times (or a Poisson number of times, or whatever), e.g. picking  $e_{a_{1,1}}, \dots, e_{a_{m,r}}$  at random and defining

$$x_{i,j} = x + \sum_{k=1}^i \sum_{j=1}^r e_{a_{i,j}} + 2 \sum_{k=j+1}^m \sum_{j=1}^r e_{a_{i,j}}.$$

Then the correlation between, say,  $E_{0,0}$  and  $E_{0,1}$  relates to what happens when one flips  $r$  0s to 1s, rather than just a single 0 to 1. This continues to be negatively correlated until  $r$  hits  $1/\delta$  (or a multiple thereof), at which point the correlation shoots up to be massively positive and in fact we find an abundance of combinatorial lines.

So we probably have to average over the “dilation” scale  $r$  rather than just set it equal to 1 as is done currently. (Much like with the corners problem: it's not enough to look for corners of size 1, or any other fixed size, but we must average over the length  $r$  of the corner. Actually, one should probably let each of the  $m$  wildcard symbols appear a different number of times, e.g. assign a Poisson process to each.) In the current example, averaging  $r$  over any range significantly larger than  $1/\delta$  will eliminate pair correlations. In fact, if one allows each wildcard symbol to have an independent number of occurrences, triple correlations such as between  $E_{0,0}$ ,  $E_{0,1}$  and  $E_{1,1}$  also seem to vanish, which incidentally yields combinatorial lines already (since  $x_{0,0}$ ,  $x_{0,1}$ ,  $x_{1,1}$  form a line).

81.  [gowers](#) Says:

[February 12, 2009 at 12:47 am](#) | [Reply](#)

Metacomment.

Just briefly on the subject of new threads, I'm not completely convinced that the multiple-threads idea has worked. Or rather I think it has partially worked. It seems that the upper-and-lower bounds thread was sufficiently separate that it has thrived on its own. But the experience of the other two threads was that once the 400 thread was opened the 300 thread died pretty rapidly. (Actually, at some point I wouldn't mind revisiting it, but that's not my main point.) So in effect we've just had one linear thread, but at any given moment we've been reasonably focused and not had too many different discussions going on at once.

As a result, I'm not sure that splitting up into a global thread and a local thread is a good idea. I think that many of the ideas for one approach are likely to be relevant to the other, and I also think it is possible for the two discussions to coexist. I think the one thing I might suggest (but I'd be interested in other people's opinion on this and everything else) is that we should restrict the number of comments further. An advantage of the multiple threads was that we had multiple summaries: I think it would be quite good to be forced to summarize more often. Also, a number of people have commented that it is difficult to catch up with the discussion if you come to it late. We should try to start each thread in a way that makes it as self-contained as possible. (I think it is legitimate to expect people to read the posts, but we should try not to force them to wade through hundreds of comments.)

82.  [gowers](#) Says:

[February 12, 2009 at 1:08 am](#) | [Reply](#)


473. Obstructions.

Something I want to think about, but won't be able to think about tonight as I'm going to have to stop this soon, is whether it is possible to put together a lot of local obstructions to get a global one. What I mean is this. Let us regard a point in  $[3]^n$  as a pair of disjoint sets  $(A,B)$  (which are the 1-set and 2-set of a sequence). Suppose that an argument of the Cartesian semi-products variety (see Terry 413) can be used to prove that for a large number of sets  $H$  (in the slices-equal measure) you get unexpectedly many 4-cycles, by which I mean quadruples  $(A,B), (A,B'), (A',B'), (A',B)$  of points with all four sets contained in  $H$ . (Here  $H$  would in fact arise as the complement of some set  $C$ .) Just to make

my question a bit more concrete, let me go for an extreme statement and suppose that what actually happens is that for many  $H$  you have two systems  $\mathcal{A}_H$  and  $\mathcal{B}_H$  of subsets of  $H$  such that there are many disjoint pairs  $(A, B)$  and they are all points in the original dense set. (In previous terminology, I want to suppose that many sets  $H$  contain a large Kneser product.) Can we somehow piece together these set systems  $\mathcal{A}_H$  and  $\mathcal{B}_H$  to obtain set systems  $\mathcal{A}$  and  $\mathcal{B}$  such that the set of disjoint pairs  $(A, B) \in \mathcal{A} \times \mathcal{B}$  correlates with the original dense subset of  $[3]^m$ ?

If the answer is no, I feel it ought to be obviously no: you would just more or less randomly choose your  $\mathcal{A}_H$  and  $\mathcal{B}_H$ . But if for some reason that turns out to be difficult (because there is too much overlap and in order to get compatibility on the overlap you are forced to relate the various  $\mathcal{A}_H$  to each other and similarly the  $\mathcal{B}_H$ ) then perhaps you can get from local to global in the way I would like.


I'm not sure how easy it would be to prove a lemma of the desired kind, but I'm hoping it will be easyish to decide whether anything like that has a chance of being true. And now it's bedtime.

83.  [Terence Tao](#) Says:  
[February 12, 2009 at 1:21 am](#) | [Reply](#)

Metacomment

I guess the number of active threads is, in practice, bounded by the number of distinct groups of active participants. The participants in the 200 thread are largely disjoint from those in the 300 and 400 threads, but the 300 and 400 participants have been basically the same, so the 400 thread has basically served as an update to the 300 thread.


Perhaps we can still have two threads: one thread (the 500 thread?) would be the local and global approaches to the problem, and another thread, which I could host, would be a slower-paced “reading course” on the ergodic proof of  $k=3$  DHJ, presumably using Randall's notes as a guide. This would probably proceed by different rules than the existing threads, since the aim is now to understand a paper rather than to solve a problem (I will have to think about what the ground rules of such a reading course would be). But I would imagine that the readership here would be a bit disjoint from the one in the “mainstream” threads - in particular, I expect more ergodic theorists to participate - and so should be able to operate independently.

84.  [gowers](#) Says:  
[February 12, 2009 at 1:25 am](#) | [Reply](#)

474.

Re Terry's metacomment — that all sounds pretty sensible to me, so let's do it unless someone comes up with an objection to the plan in the near future.

Re 473, I just wanted to add (in case there was any misunderstanding) that when I counted quadruples above, it was with slices-equal measure, so you couldn't say things like that almost all the sets  $A$  and  $B$  had roughly the same size, etc. etc.

85.  [gowers](#) Says:  
[February 12, 2009 at 1:41 am](#) | [Reply](#)

475. Obstructions.

Thinking about it a bit further, I now think 473 is true and not that hard. More details tomorrow if I wake up in a similar frame of mind.

86.  [Ryan O'Donnell](#) Says:



[February 12, 2009 at 1:51 am](#) | [Reply](#)

476. A Fourier approach to Sperner. (This would probably make more sense in the 300's thread, but I concur with Tim's #472.5.)

I was thinking about Terry's approach and his #472. (Thanks for that, btw; I agree that the second concentration of measure is the important one and that it holds in my simple example.) More specifically, I was trying to think about the "structure" of the pairs  $x_i, x_j \in \mathcal{A}$  (half-lines, as I was calling them in #134). This led me to think again about a Fourier approach to Sperner. Not sure if it has any relevance at all to DHJ(3), but I thought I'd throw it out there.

One difficulty with such an approach (discussed already in comments #29–#37) is that it's not immediately clear how to come up with a good distribution on pairs  $(x, y)$  with  $x \prec y$ . But I think Terry's idea of localising quite severely — to "constant" distances like  $m = 1/\delta^{100}$  — is very promising.

Since things are always better in Fourier-land when one uses a product distribution, I suggest the following:

Let  $\epsilon$  denote  $m/n$ . Let  $f : \{-1, 1\}^n \rightarrow \{0, 1\}$  be the indicator of a subset  $\mathcal{A} \subseteq [2]^n$  of density  $\delta$ . Jointly choose a pair of strings  $x, y \in \{-1, 1\}^n$  as follows:

For each coordinate  $i$ , independently:

- . with probability  $1 - \epsilon$ , set  $x_i = y_i =$  a random  $\pm$  bit.
- . with probability  $\epsilon$ , set  $x_i = -1, y_i = 1$ .

Finally, reverse the roles of  $x$  and  $y$  with probability  $1/2$ .

Under this distribution, certainly  $x \leq y$  or vice versa always, with the distance between them being distributed roughly as Poisson( $m$ ). The strings  $x$  and  $y$  are not uniformly distributed (which is annoying) but they are nearly so assuming  $m$  small.

To be continued in the next post. (PS: sorry, I guess I'm violating the "Don't calculate on scratch paper" rule.)



87. Ryan O'Donnell Says:

[February 12, 2009 at 1:52 am](#) | [Reply](#)

477. A Fourier approach to Sperner. (cont'd)

If one does the usual Fourier analysis thing, one gets that the probability of  $x, y$  both in  $\mathcal{A}$  equals

$$\sum_{S, T \subseteq [n]} (1 - 2\epsilon)^{|S \cap T|} (\epsilon)^{|S \Delta T|} c(S, T),$$

where  $c(S, T)$  is 0 if  $|S \Delta T|$  is odd, and is  $\pm 1$  otherwise, depending on the parity of  $|S \setminus T|$ .

I believe (and feel free to press me on this) that with some linear/matrix algebra, one can show that this expression is approximately


$$(1/2) \mathbb{S}_{1-2\epsilon}^{(\epsilon)}(f) + (1/2) \mathbb{S}_{1-2\epsilon}^{(-\epsilon)}(f),$$

where  $\mathbb{S}_\rho^{(\xi)}(f)$  is the "noise stability of  $f$  at  $\rho$  under the  $\xi$ -biased distribution".

This post is getting long, so let's say this is roughly something like  $\mathbb{E}[f(u)f(v)]$ , where  $u$  is a random string and  $v$  is formed by flipping each coordinate of  $u$  with probability  $\epsilon$ . Rather like the original distribution, but note that  $u$  and  $v$  are likely incomparable.

Even if you buy my calculations, you may ask what this has bought. I guess it's that this noise stability is a fairly well understood quantity, so we pretty much know exactly what kind of  $f$  have this quantity near  $\delta$ , what kind of  $f$  have

this quantity near  $\delta^2$ , etc.


88.  [Terence Tao](#) Says:  
[February 12, 2009 at 3:56 am](#) | [Reply](#)

478. Reading seminar

I've begun a "reading seminar" on the ergodic proof of DHJ at

<http://terrytao.wordpress.com/2009/02/11/a-reading-seminar-on-density-hales-jewett/>


No idea how lively it will be, but I will try to "read aloud" my way through paper #1 at least, and see if anyone jumps in to help out. I expect the pace to be slower than on these threads.

89.  [Ryan O'Donnell](#) Says:  
[February 12, 2009 at 5:26 am](#) | [Reply](#)

479. Followup to #472.

Hmm, what if the set  $\mathcal{A}$  is just the strings whose plurality digit is 0? This will very likely have either all or none of the  $x_{ij}$ 's in  $\mathcal{A}$ , if I'm not mistaken. Of course, being "monotone" this set has plenty of structure; I'm just wondering what we're seeking.

Indeed, there are Boolean functions (see "The Influence of Large Coalitions" by Ajtai and Linial) where every subset of  $n/\log^3 n$  coordinates has negligible "influence" — meaning, almost surely the function is determined as soon as you fix the other  $n - n/\log^3 n$  coordinates randomly. These should be easy to adapt to the  $[3]^n$  setting.


90.  [Terence Tao](#) Says:  
[February 12, 2009 at 6:28 am](#) | [Reply](#)

480. Reply to #479

Yeah, that's a pretty blatant violation of concentration of measure. But the plurality set  $\mathcal{A}$  you give has very low influence, and as such it is very easy to create combinatorial lines (or even moderately large combinatorial subspaces) - just pick a random element in  $\mathcal{A}$  and jiggle one or more of the coordinates, and with high probability you've created a combinatorial line in  $\mathcal{A}$ .

Low influence, as we've said before, leads to an "instant win" for us. Almost as instant a win is near-total "01-insensitivity", which is like low influence but restricted to swapping 0 and 1, rather than the full range of swapping 0, 1, and 2. I discussed this in 451. In terms of the  $x_{i,j}$ , this type of insensitivity is basically the same as saying that the event  $x_{i,j} \in \mathcal{A}$  is almost maximally positively correlated with  $x_{i+1,j} \in \mathcal{A}$ .

What I don't understand well yet, though, is what happens when there is a smaller amount of positive correlation; this gives a little bit of insensitivity, but not so much that the argument in 451 immediately applies. A typical example here might be a random subset of, say, the set in 479. Maybe some sort of "regularity lemma" describes all moderately-correlated sets as unions of random subsets of very highly correlated (and thus very structured and low-influence or insensitive) sets?

91.  [Terence Tao](#) Says:  
[February 12, 2009 at 6:31 am](#) | [Reply](#)

Metacomment.

Michael Nielsen has kindly transplanted the google document above to a Wiki format, where it really can be edited by everyone, and which allows for LaTeX support, creation of subpages, etc:

[http://michaelnielsen.org/polymath1/index.php?title=Main\\_Page](http://michaelnielsen.org/polymath1/index.php?title=Main_Page)

So I guess this can serve as the “background” repository for the project, to complement the “foreground” threads here and on my blog

(We’re also running out of numbers on this thread... time to start the 500 thread soon!)

92.  [Randall](#) Says:

[February 12, 2009 at 7:03 am](#) | [Reply](#)

480. Z\_k Roth:

I’d like to propose a further simplification of the problem, motivated by Gowers’ observations in post 31. Here it is: For all  $\delta > 0$ , there is an  $n$  such that any  $E \subset \{0, 1, 2\}^n$  having relative density at least  $\delta$  contains a configuration  $(a, a + \sum_{i \in \alpha} e_i, a + 2 \sum_{i \in \alpha} e_i)$ , where  $e_i = (0, 0, \dots, 0, 1, 0, \dots, 0)$ , the “1” occurring in the  $i$ th place. (Addition is modulo 3.)

I will assume no decent quantitative version of this is known. (If I’m mistaken, the following may be ignored.)

This is stronger than just getting  $(a, a+b, a+2b)$  for some  $b \in \{0, 1, 2\}^n$ , which, if I understand things properly, one could obtain via Roth’s density increment method. And, indeed, the set  $E$  mentioned in post 31, the one which has sequences where the number of 0s, 1s and 2s all multiples of 7, will contain too many of these configurations.

On the other hand, the ergodic story suggests (heuristically, at least) that restrictions may come into play under “quadratic” Fourier analysis (perhaps modulo some normalization of some kind). The basis for this heuristic is the existence of an ergodic proof of the result (never written down) in which a compact extension of the Kronecker factor is shown to be characteristic. (This is dramatically different from DHJ or even IP Roth, where I should think no distal factor would be characteristic, suggesting that perhaps no degree of  $\alpha$ -uniformity would yield constraints on the number of configurations.)

This is only a heuristic, but the implication seems to be that perhaps the above mentioned result is at the approximate level of “depth” of existence of progressions of length 4 in sets of positive density in  $\{0, 1, 2\}^n$ . And, if all this is right, that it’s the sort of problem that might be attacked by familiar methods.

93.  [Gil](#) Says:

[February 12, 2009 at 9:49 am](#) | [Reply](#)

481. Various types of slices and ultra-optimism

My UO conjecture meant to be an extension of the LYM inequality which is stronger than just saying that maximum antichain is the maximum slice. Here we have various slices and we can conjecture that the maximum size of a line-free set is attained at the union of slices (which do not contain a line). We can also conjecture (as Tim’s) that the maximum measure (when every slice has the same overall measure) for line-free set is attained with union of slides. (This reduces to LYM for the sperner case.) I think that my conjecture (or at least what I meant) is a bit stronger. Unlike the Sperner case there are many ways to cover or fractionally cover all slices by line-free union of slices.

Now there are several meanings for “slices”. The coarser is  $\Gamma_{a,b,c}$  but there are finer ones. One finer one is to associate to each word  $w$  in  $\{0, 1\}^n$  the slice  $\Gamma_w$  of all words of length  $n$  in  $\{0, 1, 2\}^n$  that if you erase the 2s you get  $w$ .

We can ask if the ultra optimistic conjecture for line-free sets is true for union of slices  $\Gamma_w$ . (This

seems a feasible problem.)

We can also talk together about such slices in all 3 directions.

There is an even finer division to slices. Consider all triples of words  $w_1, w_2$  in  $0-2$  and  $1-2$  respectively. Then let  $\Gamma_{w_1, w_2}$  be all words of length  $n$  in  $012$  s.t. when you delete  $2$  you get  $w_1$  and when you delete  $1$  you get  $w_2$ . We can ask if blocks in finer slices gives us better examples than blocks in coarser slices.

(If yes, we can update our ultra optimistic conj.) (Again you can take slices in all 3 possible “directions”).

94.  [Terence Tao](#) Says:

[February 12, 2009 at 10:03 am](#) | [Reply](#)

482.

Randall, your question sounds a bit like that of looking for progressions of length 3 whose difference is one less than a prime, which was tackled by Bergelson-Host-Kra using the nilsequences machinery and the Gowers norms. My guess is that a similar approach would work for your problem, though I didn't check it.

Gil, I am still having trouble understanding the precise formulation of your conjecture. For instance, what does “fractionally cover” mean? And what is a “block”? Also, what does “erase” or “delete” mean - are you compressing the deleted string, so that for instance deleting 1 from  $012$  gives you  $02$ ? It all sounds quite interesting, but I am afraid I may have to see a formal statement of the conjecture in order to understand it properly.

95.  [Gil](#) Says:

[February 12, 2009 at 10:03 am](#) | [Reply](#)

Regarding influences which were mentioned in several comments. It looks that indeed examples with unusual number of lines has unusual (compared to random examples) influence, but it does not look clear how such a property can be exploited to lead to higher density subspaces. So I am not sure if influence would be a useful notion in attacking the problem at hand. But there may be some useful analogies between problems about influences and about finding lines of various types.

When we talk about influence we look at a projections: we look at a subspace describe by a subset of the coordinates (but maybe we can also talk about different notions of subspace for some appropriate notion of “projection”) and consider the projection of our set to this subspace. For our problem we are more interested in the intersection of our set with various subspaces.

96.  [Gil](#) Says:

[February 12, 2009 at 10:11 am](#) | [Reply](#)

484. clarification for slices, also seconding Ryan

Dear Terry,

yes deleting 1 from  $012$  gives you  $02$  and from  $011220212012$  gives you  $0220202$ ,

a block of slices is the union of different slices (in our original notion as in the refined ones when you talk only about one direction the slices are disjoint)

(We are only interested in line-free blocks)

We say that a collection of block of slices covers  $\Gamma$  if every point in  $\Gamma$  is contained in one (or more) of them.

A fractional collection of blocks is assigning a nonnegative number to every block.

A fractional cover is a fractional collection where for every  $x \in \Gamma$  the sum of weights for blocks containing it is at least 1.

Also, I second Ryan that giving analogs to Ajtai-Linial examples may be useful for our problem and perhaps also for the cap set problem. (I will try to explain why sometime later.)

97.  [gowers](#) Says:

[February 12, 2009 at 1:23 pm](#) | [Reply](#)

485. Uniformity norms

This is an attempt to push forward in a small way a strategy for a global density-increment strategy. I would like to find a norm on functions from  $[3]^n$  to  $\mathbb{R}$  with the property that if  $\|f\|$  is small then the expectation of  $f(x)f(y)f(z)$  over all combinatorial lines is small, and if  $\|f\|$  is large then one can find a structured set on which the average of  $f$  is positive and bounded away from 0. If we've got these two properties, then it should be pretty easy to finish the proof.

Here is my imprecise proposal for such a norm. In fact, it's imprecise to the extent that it may well not even be a norm: the idea would be to massage the details until it is. I'll just draw it out of a hat, but of course it hasn't sprung out of nowhere. It comes from a mixture of various sources: analogies with corners, and also the kinds of things that were coming out of my calculations at the beginning of the 400s, and also Terry's Cartesian semi-product thoughts.

In all the density-increment proofs we know where the obstructions are of a simplicity that we expect for DHJ(3), a key concept is something like a rectangle: for Roth's theorem it's a quadruple  $(x, x+a, x+b, x+a+b)$ , for Shkredov it's  $((x, y), (x+a, y), (x, y+b), (x+a, y+b))$ , and for bipartite graphs it's a 4-cycle  $(xy, x'y, x'y', xy')$ . What would be the obvious quadruples to look at here? Well, we know that we can think of a subset  $\mathcal{A} \subset [3]^n$  as a bipartite graph: its elements are pairs of sets  $(A, B)$  such that there exists  $x \in \mathcal{A}$  such that  $A$  is the 1-set of  $x$  and  $B$  is the 2-set of  $x$ . So a natural suggestion (if you don't think too hard about it) for a norm to put on a function  $f$  is

$$\|f\| = \mathbb{E} f(A, B) f(A, B') f(A', B') f(A, B'),$$

where the expectation is over all quadruples  $(A, B, A', B')$  such that the above four pairs are disjoint pairs. (We don't need complex  $f$  but if we did then we'd put bars over the second and fourth terms in the product.)

What are some immediate objections to this? Well, given such a quadruple, each element has seven possibilities: it can belong to up to two of the four sets  $A, B, A', B'$  but cannot belong to both an  $A$  and a  $B$ . Therefore, a random such quadruple is obtained by randomly partitioning  $[n]$  into seven parts and choosing the first one to be disjoint from all four sets, the second one to be in just  $A$ , and so on. In particular, the set that's disjoint from all of them has cardinality very close to  $n/7$  with very high probability, so it's got nothing much to do with combinatorial lines.

But not so fast. This is uniform-measure speak rather than slices-equal speak, and I'm a slices-equal person. If slices are equal, then the fact that  $A \cup A'$  is disjoint from  $B \cup B'$  has a much smaller bearing on their cardinalities. (It does have some effect, but it's just not a big deal.)

So we have no a priori reason to reject this norm. In the next comment I'll try to argue (as much to myself as to anyone else) that it's actually good for something.

98.  [gowers](#) Says:

[February 12, 2009 at 1:51 pm](#) | [Reply](#)

486. Uniformity norms.

Now for a small technical comment. Suppose we take the subset of  $\mathbb{Z}^2$  consisting of all points  $(a, b)$  such that  $a, b \geq 0$  and  $a + b \leq n$ . This is an isosceles right-angled-triangular chunk of the integer lattice. And now suppose we try to investigate functions on this set using a box norm (that is, averaging expressions like  $f(x, y)f(x + a, y)f(x, y + b)f(x + a, y + b)$ ). We will have a little bit of trouble with this because some points are contained in lots of boxes and some in hardly any. So what we might decide to do instead is a modest localization: choose some square patch of the triangle where the set we are interested in is dense.

We have essentially this problem with a uniformity norm for DHJ if we use the slices-equal measure. To combat it, I want to apologize to the number 3 and lose some symmetry. Recall (or prove as a very easy exercise) that DHJ(3) is equivalent to the following problem: if you have a dense set  $\mathcal{A}$  of pairs  $(A, B)$  of disjoint subsets of  $[n]$  then you can find disjoint sets  $A, B, D$  such that  $(A, B)$ ,  $(A \cup D, B)$  and  $(A, B \cup D)$  all belong to  $\mathcal{A}$ . I'm going to focus on this version of the statement, but with the added stipulation that the cardinalities of  $A$  and  $B$  are both at most  $n/2$ , so that I have a square grid of slices. All the time, I need hardly add, we use slices-equal measure. So a random point in the space is obtained by choosing two random numbers  $r$  and  $s$  that are at most  $n/2$  and then a random pair of disjoint sets  $A$  and  $B$  of sizes  $r$  and  $s$ .

Unfortunately I have to go while I've still got a lot to say, and won't have a chance to add to this for about nine hours, if not more. Anyhow, the next step is to try to do a sort of Sperner-averaging style lift to get from the usual proof that the box norm controls the number of corners to a proof that the above attempt at a uniformity norm controls the number of set-theoretic corners. And if that works, I'd like an inverse theorem for the uniformity norm to tell us that we've got a 411-style obstruction.



99. [gowers](#) Says:

[February 12, 2009 at 5:31 pm](#) | [Reply](#)

487. Uniformity norm.

Another technical point. In order to define a norm, we need that expression to be non-negative (and zero only if  $f$  is zero). To get that we can do the following. For convenience define  $f(A, B)$  to be zero if  $A$  and  $B$  intersect. Then define the norm as follows. First pick a random permutation. Then take the average of  $f(A, B)f(A, B')f(A', B')f(A', B)$  over all quadruples of intervals in the permuted  $[n]$ . This is then easily seen to have the positive definiteness that one wants. Does it matter that we are now averaging over lots of intersecting pairs of sets? Not really, because the intersecting pairs are given a lot less weight (because intervals, on average, intersect far less frequently than random sets of the same size).

So this is my revised candidate for a uniformity norm (perhaps with some restriction on the sizes of the sets involved — I'm not sure that matters much any more though).



100. [Ryan O'Donnell](#) Says:

[February 13, 2009 at 12:24 am](#) | [Reply](#)

488. Terry's strategy #439.

I'm still slowly plugging through that post... 😊

Perhaps it's not necessary to show that the second concentration of measure (for  $x_{ij} \in \mathcal{A}$ ) holds \*because\* of small pairwise correlations between the events. For example, it holds anyway in the case of the “mod  $1/\delta$ ” set (#472) and the “random subset of plurality” set (end of #480).

Except, wait: I suppose it doesn't hold in a variant of #480's example. Take the set of strings which have at least  $n/3 + c\sqrt{\log(1/\delta)}\sqrt{n}$  0's in them, where that deviation is cooked up so that the overall density is  $2\delta$ . Then take a random subset of half of these strings.



In this case, it seems like you could get too \*many\* of these  $x_{ij} \in \mathcal{A}$  events happening. Is this the problem? I must confess, I haven't gotten myself up to speed yet on the particulars of the second half of the plan in #439.

101.  [gowers](#) Says:

[February 13, 2009 at 12:55 am](#) | [Reply](#)

489. Uniformity norms.

Now I would very much like to be able to prove that if  $f$  has a small uniformity norm in the sense of 486/7, then the expectation of  $f(A, B)f(A \cup D, B)f(A, B \cup D)$  is small. But let me go for the slightly easier task of showing that if  $f$  is a 01-valued function and the expectation of  $f(A, B)f(A \cup D, B)f(A, B \cup D)$  is almost zero, then the expectation of  $f(A, B)f(A, B')f(A', B')f(A', B)$  is larger than it should be. Incidentally, my convention now has to be that if the average of  $f$  is  $\delta$ , then  $f(A, B) = \delta$  whenever  $A$  and  $B$  intersect.

Actually, I want to revise the definition of the uniformity norm. I take a random permutation as in 486, but I insist that  $A$  and  $A'$  are initial segments (rather than just any old intervals) and that  $B$  and  $B'$  are final segments. But they're still allowed to overlap, so that the expression for the norm splits up as a sum of squares.

Now let's go back to Terry's Cartesian semi-products, as I discussed them in 442. First of all, if we take a random permutation of  $[n]$  and then take a random interval  $J$  in the permuted set, there will on average be several pairs of subintervals  $(A, B)$  that partition  $J$ . For fixed  $J$  each such pair is of course determined by  $A$  only. And if we have two such pairs  $(A, B)$  and  $(A', B')$  with  $A \subset A'$ , then the pair  $(A, B') = (A, J \setminus A')$  cannot lie in  $\mathcal{A}$ . This means that  $\mathcal{A}$  has to avoid the Cartesian product of the set of all the  $A$  such that  $(A, J \setminus A) \in \mathcal{A}$  with the set of all the  $B$  such that  $(J \setminus B, B) \in \mathcal{A}$ . But that gives us a nonuniformity in the function  $f$ . Averaging over all permutations  $\pi$  and all intervals  $J$  gives us (I think) that the norm in 486/7 is large.

What I *think* this sketch shows is that the first stage of the two-step programme is OK: if a set contains no combinatorial lines then its slices-equal uniformity norm is too big. I've got to go to bed now, but I'll write a very brief comment about the second stage.

102.  [gowers](#) Says:

[February 13, 2009 at 1:03 am](#) | [Reply](#)

490. Inverse theorem.

The second stage is to prove an inverse theorem for the uniformity norm: that is, to prove that if  $f$  has a large uniformity norm then it has positive correlation with an exploitable structure (meaning a structure that can be almost uniformly covered by subspaces).

The approach one would take to this in a dense bipartite graph would be to take a random edge and take the two neighbourhoods of its end vertices. The analogue here would be to take a random disjoint pair of sets  $(A', B')$  and then to take all disjoint pairs  $(A, B)$  such that  $(A, B')$  and  $(A', B)$  belong to  $\mathcal{A}$ . Note that this is a Kneser product of two sets, just as we want, and that on average we expect  $\mathcal{A}$  to be too dense in it.

I was about to say that the problem is that there are far too few such pairs, but it's just beginning to occur to me that perhaps that doesn't matter. I've got to go to bed now, but I'm not completely certain that the inverse theorem isn't almost as trivial here as it is for graphs. I'll go to bed on that note of hyperoptimism — it feels as though the whole problem could be basically solved, but with me that feeling is extremely unreliable, and all the more so as there's a giant claim in this paragraph that I haven't checked at all.

103.  [gowers](#) Says:

[February 13, 2009 at 1:35 am](#) | [Reply](#)

491. Inverse theorem.

Oops. Forget last paragraph of 490. It didn't survive the going-to-bed process. I'll think about a possible inverse theorem tomorrow.

104. [DHJ — possible proof strategies « Gowers's Weblog](#) Says:

[February 13, 2009 at 10:52 am](#) | [Reply](#)

[...] a proof on at least some of the ideas that come out of the proof of Furstenberg and Katznelson, see Terry's comment 439, as well as the first few comments after it. Another helpful comment of Terry's is his [...]

105.  [Gil](#) Says:

[February 13, 2009 at 11:15 am](#) | [Reply](#)

492. DHJ is equivalent for different measures.

Dear Tim,

"I'm pretty sure that DHJ for the slices-equal measure is an easy consequence of DHJ, and vice versa. Here's a sketch of how to prove that DHJ for any reasonably continuous measure (meaning that the weight of a point is usually very similar to the weight of a neighbouring point) is equivalent to the usual DHJ."

"...the measure in almost all of the subspaces is almost uniform, so we can apply the normal DHJ..."

I do not understand the argument. Is it ok?

106.  [gowers](#) Says:

[February 13, 2009 at 2:16 pm](#) | [Reply](#)

493. DHJ equivalence.

Dear Gil,

I see what you mean. I had in my mind the false idea that the function  $\binom{n}{k}$  is roughly continuous in  $k$ , when of course that is true only near the centre. I now think you are right and that DHJ for slices-equal measure is a stronger result than DHJ (and therefore, I imagine, not a known result at all). The reason I think that is that Sperner for slices-equal seems to be stronger than Sperner for the uniform measure.

Of course, the other direction is the one we actually need, so this doesn't mean that proving DHJ for slices-equal measure is a waste of time.

107.  [gowers](#) Says:

[February 13, 2009 at 3:28 pm](#) | [Reply](#)

494. DHJ equivalence.

Actually, perhaps there is a way of deducing slices-equal DHJ from uniform DHJ. Let  $A$  be a dense subset of  $[3]^n$  in the slices-equal measure. Find a small triangular grid of slices  $\Gamma_{a,b,c}$  where  $A$  is on average dense, where all  $(a, b, c)$  in the grid are close to  $(a_0, b_0, c_0)$  that are bounded away from 0 and  $n$ . Suppose, for the sake of example, that your slices are all very close to  $(n/4, n/4, n/2)$ . Now choose a random  $n/4$  coordinates and fix them to be 3. On average, your set will still be dense, and now you have a new  $n$  (it is the old  $3n/4$ ) and all slices are near the middle so that they all have roughly the same weight.

108.  [Gil Kalai](#) Says:

[February 13, 2009 at 10:56 pm](#) | [Reply](#)

495. DHJ equivalence.

Actually, on second thought, I do not think the DHJ for sliced-equal measure easily implies DHJ or is easily implied by DHJ. Since the two measures in questions are mutually singular it seems a priori that an easy reduction will be a little miracle.

For Sperner, the result that in the sliced-equal measure the maximum measure of an antichain is at most  $1/n$  indeed easily implies that the maximum measure for an antichain in the uniform measure is at most  $c/\sqrt{n}$ .

Now, the heuristic argument from 494 is appealing but suspicious. Will all the slices be sufficiently near the middle, or rather look near but actually be in the shallow boundary of the middle where “roughly the same” still be too “roughly”?

109.  [gowers](#) Says:

[February 14, 2009 at 12:58 am](#) | [Reply](#)

496. DHJ equivalence.

There's no problem getting from an arbitrary measure to the uniform measure. The proof is as follows. Suppose I know that every set that is dense in the measure  $\mu$  contains a combinatorial line. Now suppose I want to find a combinatorial line in a set  $A$  that has density  $\delta$  in the uniform measure on  $[3]^n$  if  $n$  is sufficiently large. Pick  $m$  such that every subset of  $[3]^m$  of  $\mu$ -density  $\delta$  contains a combinatorial line. Now randomly embed  $[3]^m$  into  $[3]^n$  by choosing  $m$  variable coordinates and fixing the rest. We may suppose that every point in  $A$  has  $n/3 + O(\sqrt{n})$  of each coordinate value and that  $m \ll \sqrt{n}$ . Therefore, changing  $m$  coordinates hardly changes the density of a slice. It follows that each point of  $A$  is in approximately the same number of these random subspaces. Therefore, by averaging, there is a random subspace inside which  $A$  has  $\mu$ -density at least  $\delta/2$ . (We could think of it Terry's way: as we move the random subspace around, what we effectively have is a bunch of random variables, each with mean approximately  $\delta$ , so by linearity of expectation we'll get  $\mu$ -density at least  $\delta/2$  at some point, whatever the measure  $\mu$  is.)

The other way round I am less sure of. I think I'll have to go away and actually try to write out the calculation rigorously.

110.  [Gil](#) Says:

[February 14, 2009 at 8:21 am](#) | [Reply](#)

Another general question (I think we discussed it but I forgot what was the conclusion.) How far can we bound from above the number of wild coordinates in the combinatorial lines so that DHJ (for  $k=3$ ) is still correct? For Sperner it is enough that the number of wild coordinates tends to infinity as slow as we want, is that so also for DHJ?

111.  [Terence Tao](#) Says:

[February 14, 2009 at 8:41 am](#) | [Reply](#)

Dear Gil,

Yes, we can in fact make the number of wildcards bounded by  $m = m(\delta)$  depending only on  $m$ . Indeed, one basically takes  $m$  to be the first index for which DHJ is true at this density. The pigeonhole principle then tells us that there is an  $m$ -dimensional slice of the big set which also has density  $\delta$ , and the claim follows.

112.  [gowers](#) Says:

[February 14, 2009 at 9:15 am](#) | [Reply](#)

Dear Gil,

In the end I didn't have to do a calculation to see that my argument for the reverse implication was wrong. If you condition on  $n/4$  particular coordinates equalling 3, then that hugely affects what slice you are likely to have chosen, so you no longer have slices-equal measure. This is one area where the uniform measure scores over the slices-equal measure.



113. [Gil](#) Says:

[February 14, 2009 at 8:17 pm](#) | [Reply](#)

497. DHJ for various measures

Thanks Tim and Terry; The implication from general measures to uniform measures is neat and simple. (it is even simpler when the problem is group-invariant like for cap sets.) I suppose that there is no simple implication in the other direction and this has the benefit of making the DHJ problem for the slice-equal measure an open problem.

I find the idea that changing the measure will help (e.g. difficulties described by some counter examples go away) quite appealing. (We talked in the threads about biased Bernoulli measures and about the slice-equal measure which is simply the simple average of all Bernoulli measures.) I wonder if there is a nice basis of orthogonal functions (like the Fourier transform) suitable for the slice-equal measure.



114. [Terence Tao](#) Says:

[February 15, 2009 at 4:41 am](#) | [Reply](#)

498. DHJ for various measures

(Getting squeezed for space in this thread!)

I think I can derive equal-slices DHJ from regular DHJ.

Suppose that  $A \subset [3]^n$  has density  $\delta$  in the equal-slices sense. By the first moment method, this means that  $A$  has density  $\gg \delta$  on  $\gg \delta$  on the slices.

Let  $m$  be a medium integer (much bigger than  $1/\delta$ , much less than  $n$ ).

Pick  $(a, b, c)$  at random that add up to  $n-m$ . By the first moment method, we see that with probability  $\gg \delta$ ,  $A$  will have density  $\gg \delta$  on the  $\gg \delta$  of the slices  $\Gamma_{(a', b', c')}$  with  $a' = a + m/3 + O(\sqrt{m})$ ,  $b' = b + m/3 + O(\sqrt{m})$ ,  $c' = c + m/3 + O(\sqrt{m})$ . This implies that  $A$  has expected density  $\gg \delta$  on a random  $m$ -dimensional subspace generated by  $a$  1s,  $b$  2s,  $c$  3s, and  $m$  independent wildcards. Applying regular DHJ to that random subspace we obtain the claim.



115. [Gil](#) Says:

[February 15, 2009 at 9:28 am](#) | [Reply](#)

499. Influence strategy of proof

There were several comments regarding the relation with influences, so let me mention a few connections and start with a "strategy" based on influence proofs. (probably related to Ryan's suggested approach to Sperner.)

a) 0-1

We have a subset  $A$  of  $\{0, 1\}^n$  [or  $\{0, 1, 2\}^n$ ].

Let  $\ell$  be a combinatorial line let  $R$  be the set of fixed coordinates and  $S$  be the set of wild coordinates.

Let's define the projection of  $A$  in the direction orthogonal to  $\ell$  as the vectors  $z$  in  $\{0, 1\}^R$  such that completing  $z$  by either the all 0 vector on the  $S$  coordinates or the all 1 vector [or the all 2 vector] gives you a vector in  $A$ .

We can define the influence as the measure of the projection minus the measure of  $A$ . if  $S$  is a single coordinate this is the usual influence.

If  $A$  has density  $c$  and  $A$  contains no lines then its influence in every direction is  $c$ . (We would like to show that the influence must be substantially smaller in some directions.)

So we would like to develop some strategy for showing that there is always a direction (maybe restrict our attention to directions for which  $R$  is small) so that the influence in this direction is smaller than  $c$ . (Or even smaller than  $c - c^2/100$ .)

(To immitate these proofs we do not need to adopt the influence/projection terminology and can just talk about combinatorial lines.)

If we try to immitate the influence proofs we need to do the following:

a) To give a Fourier expression for the influence (or equivalently the number of combinatorial lines) in direction with  $S$  as wild coordinates.

b) To note that (perhaps just when  $S$  is small) such an expression must obey some restrictions because it is the Fourier expansion of a set or some similar restraint function.

c) to reach a contradiction, somehow.

(Usually in the influence proofs b is applying some hypercontractive inequality; a major difference is that in these proofs we want to reach a contradiction from having too small influence and here from having too large influence.)

b) 0-1-2

We have a subset  $A$  of  $\{0, 1, 2\}^n$ .

Let  $\ell$  be a combinatorial line let  $R$  be the set of fixed coordinates and  $S$  be the set of wild coordinates.

We need to give a Fourier description for the numbers of vectors on  $\{0, 1, 2\}^R$  which are the projection of at least 1, and at least 2 vectors from  $A$ .

Having no combinatorial line amounts to untypical relations between these two quantities and we want to show that they cannot all hold for every  $S$  (or every small  $S$ ). Again this may be easier for DHJ(2.5) and alike.

In this vague plan, what seems clearly doable (but require calculations) is giving the Fourier expansion for the different projections and influences.

116.  [Gil](#) Says:

[February 15, 2009 at 9:57 am](#) | [Reply](#)

500. DHJ for different measurese  
hmm. Right!

*[We already have a comment 500 over on the other thread, so this looks like the time to declare this thread officially closed! (However, I have something to say about DHJ for different measures, so I'll do so over there.)]*

## Leave a Reply

Logged in as [Ryan O'Donnell](#). [Logout »](#)

Submit Comment

☐ Notify me of follow-up comments via email.

---

[Blog at WordPress.com](#).

[Entries \(RSS\)](#) and [Comments \(RSS\)](#).

⌵



# Gowers's Weblog

Mathematics related discussions

---

« [DHJ — quasirandomness and obstructions to uniformity](#)  
[Must an “explicitly defined” Banach space contain c 0 or ell p?](#) »

## DHJ — possible proof strategies

By gowers

I will give only a rather brief summary here, together with links to some comments that expand on what I say.

If we take our lead from known proofs of Roth's theorem and the corners theorem, then we can discern several possible approaches (each one attempting to imitate one of these known proofs).

### 1. Szemerédi's original proof.

Szemerédi proved a proof of a one-dimensional theorem. I do not know whether his proof has been modified to deal with the multidimensional case (this was not done by Szemerédi, but might perhaps be known to Terry, who recently digested and rewrote Szemerédi's proof). So it is not clear whether one can base an entire argument for DHJ on this argument, but it could well contain ideas that would be useful for DHJ. For a sketch of Szemerédi's proof, see this post of Terry's, and in particular his discussion of question V. He also has links to comments that expand on what he says.

Ajtai and Szemerédi found a clever proof of the corners theorem that used Szemerédi's theorem as a lemma. This gives us a direction to explore that we have hardly touched on. Very roughly, to prove the corners theorem you first use an averaging argument to find a dense diagonal (that is, subset of the form  $x + y = r$  that contains many points of your set  $A$ ). For any two such points  $(x, r - x)$  and  $(y, r - y)$  you know that the point  $(x, r - y)$  does not lie in  $A$  (if  $A$  is corner free), which gives you some kind of non-quasirandomness. (The Cartesian semi-product arguments discussed in the 400s thread are very similar to this observation.) Indeed, it allows you to find a large Cartesian product  $X \times Y$  in which  $A$  is a bit too dense. In order to exploit this, Ajtai and Szemerédi used Szemerédi's theorem to find long arithmetic progressions  $P \subset X$  and  $Q \subset Y$  of the same common difference such that  $A$  has a density increment in  $P \times Q$ . (I may have misremembered, but I think this must have been roughly what they did.) So we could think about what the analogue would be in the DHJ setting. Presumably it would be some kind of multidimensional Sperner statement of sufficient depth to imply Szemerédi's theorem. A naive suggestion would be that in a dense subset of  $[2]^n$  you can find a large-dimensional combinatorial subspace in which all the variable sets have the same size. If you apply this to a union of layers, then you find an arithmetic progression of layer-cardinalities. But this feels rather artificial, so here's a question we could think about.

**Question 1.** Can anyone think what the right common generalization of Szemerédi's theorem and Sperner's theorem ought to be? (Sperner's theorem itself would correspond to the case of progressions of length 2.)

### Density-increment strategies.

The idea here is to prove the result by means of the following two steps.

1. If  $A$  does not contain a combinatorial line, then it correlates with a set  $S$  with some kind of atypical structure that we can describe.
2. Every set  $S$  with that kind of structure can be partitioned into (or almost evenly covered by) a collection of combinatorial subspaces with dimension tending to infinity as  $n$  tends to infinity.

One then finishes the argument as follows. If  $A$  contains no combinatorial line, then find  $S$  such that  $A$  is a bit denser in  $S$  than it is in  $[3]^n$ . Cover  $S$  with subspaces. By averaging we find that the density of  $A$  is a bit too large in one of these subspaces. But now we are back where we started with a denser set  $A$  so we can repeat. The iteration must eventually terminate (since we cannot have a density greater than 1) so if the initial  $n$  was large enough then we must have had a combinatorial line.

One can also imagine splitting 1 up further as follows.

- 1a. If  $A$  contains no combinatorial line, then  $A$  contains too many of some other simple configuration.
- 1b. Any set that contains too many of those configurations must correlate with a structured set.

This is certainly how the argument goes in some of the proofs of related results.

### Triangle removal.

This was the initial proposal, and we have not been concentrating on it recently, so I will simply refer the reader to the original post, and add the remark that if we manage to obtain a complete and global description of obstructions to uniformity, then regularity and triangle removal could be an alternative way of using this information to prove the theorem. And in the case of corners, this is a somewhat simpler thing to do.

### Ergodic-inspired methods.

For a proposal to base a proof on at least some of the ideas that come out of the proof of Furstenberg and Katznelson, see [Terry's comment 439](#), as well as the first few comments after it. Another helpful comment of Terry's is his [comment 460](#), which again is responded to by other comments. Terry has also begun an [online reading seminar](#) on the Furstenberg-Katznelson proof, using not just the original paper of Furstenberg and Katznelson but also a more recent paper of Randall McCutcheon that explains how to finish off the Furstenberg-Katznelson argument.

**Suggestion.** I have tried to do a bit of further summary in my first couple of comments. It would be good if others did the same, with the aim that anyone who reads this should be able to get a reasonable idea of what is going on without having to read too much of the earlier discussion.

**Remark.** There is now a [wiki associated with this whole enterprise](#). It is in the very early stages of development but has some useful things on it already.

This entry was posted on February 13, 2009 at 8:27 am and is filed under [polymath1](#). You can follow any responses to this entry through the [RSS 2.0](#) feed. You can [leave a response](#), or [trackback](#) from your own site.

## 115 Responses to “DHJ — possible proof strategies”

1.  [gowers](#) Says:

[February 13, 2009 at 11:30 am](#) | [Reply](#)

500. Inverse theorem.

This is continuing where I left off in 490. Let me briefly say what the aim is. I am using the disjoint-pairs formulation of DHJ: given a dense set  $\mathcal{A}$  of pairs  $(A, B)$  of disjoint sets, one can find a triple of the form  $(A, B), (A \cup D, B), (A, B \cup D)$  in  $\mathcal{A}$  with all of  $A, B$  and  $D$  disjoint. Here “dense” is to be interpreted in the equal-slices measure: that is,

if you randomly choose two non-negative integers  $a, b$  with  $a + b \leq n$ , and then randomly choose disjoint sets  $A$  and  $B$  with cardinalities  $a$  and  $b$ , then the probability that  $(A, B) \in \mathcal{A}$  is at least  $\delta$ .

Now I define a norm on functions  $f : [2]^n \times [2]^n \rightarrow \mathbb{R}$  as follows. (Incidentally, this is the genuine Cartesian product here, but pairs of sets will not make too big a contribution if they intersect.) First choose a random permutation  $\pi$  of  $[n]$ . Then take the average of  $f(A, B)f(A, B')f(A', B)f(A', B')$  over all pairs  $A, A'$  of initial segments of  $\pi[n]$  and all pairs  $B, B'$  of final segments of  $\pi[n]$ . Finally, take the average over all  $\pi$ .

A few things to note about this definition. First, if you choose the initial and final segments randomly, then the probability that  $A \cup A'$  is disjoint from  $B \cup B'$  is at least  $1/16$  (and in fact slightly higher than this). Second, the marginal distributions of  $(A, B)$ ,  $(A', B)$ ,  $(A, B')$  and  $(A', B')$  are just the slices-equal measure. Third, for any fixed  $\pi$  the expectation we take can also be thought of as  $\mathbb{E}_{A, A'}(\mathbb{E}_B f(A, B)f(A', B))^2$ , so we get positivity. If we take the fourth root of this quantity and average over all  $\pi$  then we get a norm, by standard arguments. (I'm less sure what happens if we first average and then take the fourth root. I'm also not sure whether that's an important question or not.)

Given a set  $\mathcal{A}$  of density  $\delta$  (as a subset of the set of all disjoint pairs, with slices-equal measure), define its *balanced function*  $f : [2]^n \times [2]^n \rightarrow \mathbb{R}$  as follows. If  $A$  and  $B$  are disjoint, then  $f(A, B) = 1 - \delta$  if  $(A, B) \in \mathcal{A}$  and  $-\delta$  otherwise. If  $A$  and  $B$  intersect, then  $f(A, B) = 0$ . Then  $\mathbb{E}f = 0$  if we think about it suitably. And the way to think about it is this. To calculate  $\mathbb{E}f$  we first take a random permutation of  $[n]$ . Then we choose a random initial segment  $A$  and a random final segment  $B$ . Then we evaluate  $f(A, B)$ . This is how I am *defining* the expectation of  $f$ .

Though I haven't written out a formal proof (since I am not doing that — though I think a general rule is going to have to be that if a sketch starts to become rather detailed, and if nobody can see any reason for it not to be essentially correct, then the hard work of turning it into a formal argument would be done away from the computer), I am fairly sure that if a dense set  $\mathcal{A}$  contains no combinatorial lines and  $f$  is the balanced function of  $\mathcal{A}$ , then  $\|f\|$  is bounded away from 0.

That was just a mini-summary of where this uniformity-norms idea may possibly have got to. Because it feels OK I want to think about the part that feels potentially a lot less OK, which is obtaining some kind of usable structure where  $\|f\|$  has a large average, on the assumption that  $\|f\|$  is large.

2.  [gowers](#) Says:

[February 13, 2009 at 11:48 am](#) | [Reply](#)

501. Inverse theorem.

What I would very much like to be able to do is rescue a version of the conjecture put forward in [comment 411](#). The conjecture as stated seems to be false, but it may still be true if we use the slices-equal measure. In terms of functions, what I'd like to be able to prove is that if  $\|f\|$  is at least  $\delta$ , then there exist set systems  $\mathcal{U}$  and  $\mathcal{V}$  such that the density of disjoint pairs  $(A, B)$  with  $A \in \mathcal{U}$  and  $B \in \mathcal{V}$  is at least  $c(\delta) > 0$  and the average of  $f(A, B)$  over all such pairs is at least  $c'(\delta) > 0$ . (The first condition could be relaxed to positive density in some subspace if that helped.)

I'm fairly sure that easy arguments give us the following much weaker statement. If you choose a random pair  $\sigma = (A_0, B_0)$  according to equal-slices measure, then you can find set systems  $\mathcal{U}_\sigma$  and  $\mathcal{V}_\sigma$  with the following properties.

- (i) Every set in  $\mathcal{U}_\sigma$  is disjoint from  $B_0$  and every set in  $\mathcal{V}_\sigma$  is disjoint from  $A_0$ .
- (ii) The set of disjoint pairs  $(A, B)$  such that  $A \in \mathcal{U}_\sigma$  and  $B \in \mathcal{V}_\sigma$  is on average dense in the set of disjoint pairs  $(A, B)$  such that  $A \cap B_0 = B \cap A_0 = \emptyset$ .
- (iii) For each  $\sigma$  the expectation of  $f(A, B)$  over all disjoint pairs  $(A, B) \in \mathcal{U}_\sigma \times \mathcal{V}_\sigma$  is at least  $c'(\delta)$ .

If this is correct, then the question at hand is whether one can piece together the pairs  $(\mathcal{U}_\sigma, \mathcal{V}_\sigma)$  to form two set systems  $\mathcal{U}$  and  $\mathcal{V}$  such that the density of disjoint pairs  $(A, B)$  with  $A \in \mathcal{U}$  and  $B \in \mathcal{V}$  is at least  $c(\delta)$  and the average of  $f(A, B)$  over all such pairs is at least  $c''(\delta)$ .

3.  [gowers](#) Says:

[February 13, 2009 at 12:36 pm](#) | [Reply](#)

502. Inverse theorem.

In this comment I want to explain slightly better what I mean by “piecing together” local lack of uniformity to obtain global obstructions. To do so I’ll look at a few examples.

**Example 1** Terry has already mentioned this one. Let  $f$  be a function defined on  $\mathbb{Z}_N$ . Suppose we know that for every interval  $I$  of length  $m$  we can find a trigonometric function  $\tau_I$  such that  $\mathbb{E}_{x \in I} f(x) \tau_I(x) \geq \delta$ . Can we find a trigonometric function  $\tau$  defined on all of  $\mathbb{Z}_N$  such that  $\mathbb{E}_x f(x) \tau(x) \geq c(\delta)$ ? The answer is very definitely no. Basically, you can just partition  $\mathbb{Z}_N$  into intervals of length a bit bigger than  $m$  and define  $f$  to be a randomly chosen trigonometric function on each of these intervals. This will satisfy the hypotheses but be globally quasirandom.

**Example 2** Now let’s modify Example 1. This time we assume that  $f$  has trigonometric bias not just on intervals but on arithmetic progressions of length  $m$  (in the mod- $N$  sense of an arithmetic progression). I don’t know for sure what the answer is here, but I think we do now get a global correlation. Suppose for instance that  $m$  is around  $\sqrt{N}$ . Then if we choose random trigonometric functions on each interval of length  $m$ , they will give rise to random functions on arithmetic progressions of common difference  $m$ . So the frequencies are forced to relate to each other in a way that they weren’t before.

**Example 3.** Suppose that  $G$  is a graph of density  $\delta$  with vertex set  $X$  of size  $n$ . Suppose that  $m$  is some integer that’s much smaller than  $n$  and suppose that for a proportion of at least  $c$  of the choices  $Y$  of  $m$  vertices you can find a subset  $Z \subset Y$  of size at least  $cm$  such that the density of the subgraph induced by  $Z$  (inside  $Z$ ) is at least  $\delta + c$ . Can we find a subset  $W$  of  $X$  of size at least  $c'n$  such that the density of the subgraph induced by  $W$  is at least  $\delta + c'$ ?

The answer is yes, and it is an easy consequence of standard facts about quasirandomness. The hypothesis plus an averaging argument implies that  $G$  contains too many 4-cycles, which means that  $G$  is not quasirandom, which implies that there is a global density change. (Sorry, in this case I misstated it — you don’t actually get an increase, as the complete bipartite graph on two sets of size  $n/2$  illustrates, but you can get the density to differ substantially from  $\delta$ . If you want a density increase, then you need a similar statement about bipartite graphs.)

**Example 4.** Suppose that  $G$  is a random graph on  $n$  vertices with edge density  $n^{-1/100}$ . Let  $H$  be a subgraph of  $G$  of relative density  $\delta$ . Then one would expect  $H$  to contain about  $(\delta n^{-1/100})^3 n^3$  labelled triangles. Suppose that it contains more triangles than this. Can we find two large subsets  $U$  and  $V$  of vertices such that the number of edges joining them is substantially more than  $\delta n^{-1/100} |U| |V|$ ?

Our hypothesis tells us that if we pick a random vertex in the graph, then its neighbourhood will on average have greater edge density than  $\delta n^{-1/100}$ . So we have a large collection of sets where  $H$  is too dense. However, these sets are all rather small. So can we put them together? The answer is yes but the proof is non-trivial and depends crucially on the fact that the neighbourhoods of the vertices are forced (by the randomness of  $G$ ) to spread themselves around.

**Question.** Going back to our Hales-Jewett situation, as described at the end of the previous comment, is it more like Example 1 (which would be bad news for this approach, but illuminating nevertheless), or is it more like the other examples where the non-uniformities are forced to relate to one another and combine to form a global obstruction?

4.  [gowers](#) Says:

[February 13, 2009 at 5:23 pm](#) | [Reply](#)

## 503. Combining obstructions.

I want to think as abstractly as possible about when obstructions in small subsets are forced to combine to form a single global obstruction. For simplicity of discussion I'll use a uniform measure. So let  $X$  be a set, let  $E_1, \dots, E_N$  be subsets of  $X$ , and let us assume that every  $x \in X$  is contained in exactly  $m$  of the  $E_i$ . Let us assume also that the  $E_i$  all have the same size (an assumption I'd eventually hope to relax) and that a typical intersection  $E_i \cap E_j$  is not *too* small, though I won't be precise about this for now.

For each  $E_i$  let  $g_i : E_i \rightarrow [-1, 1]$  be some function. Then one simple remark one can make (which doesn't require anything like all the assumptions I've just made) is that if we can find  $f$  that correlates with many of the  $g_i$ , then there must be lots of correlation between the  $g_i$  themselves. To see this, recall first that  $m^{-1} \sum_i \chi_{E_i}$  is the constant function 1. Therefore,  $g = m^{-1} \sum_i g_i$  is a function that takes values in  $[-1, 1]$ . Furthermore,  $\|g\|_2^2 = m^{-2} \sum_{i,j} \langle g_i, g_j \rangle$ . Since  $m^{-2} \sum_{i,j} \langle \chi_{E_i}, \chi_{E_j} \rangle = 1$ , and we can rewrite it as  $m^{-2} \sum_{i,j} |E_i \cap E_j|$ , we find that on average  $\langle g_i, g_j \rangle$  must be a substantial fraction of its trivial maximum  $|E_i \cap E_j|$ . If  $m$  is large, so that the diagonal terms are negligible, then this says that there are lots of correlations between the  $g_i$ .

In my next comment I will discuss a strategy for getting from this to a globally defined function, built out of the  $g_i$ , that correlates with  $f$  and is likely to have good properties under certain circumstances.



5. [jozsef](#) Says:

[February 13, 2009 at 7:13 pm](#) | [Reply](#)

## 504. Combining obstructions.

Tim - Before you define a general local-global uniformity, please let me ask a question. For extremal set problems it was very useful - and we used it before - to take a random permutation of the base set  $[n]$  and considering only sets which are "nice" in this ordering. For example, let us consider sets where the elements are consecutive. If the original set system was dense, then in this window, that we had by a random permutation, is a dense set of intervals. Here we can define/check uniformity easily. Disjoint sets are disjoint, and one can also analyze intersections. I didn't do any formal calculations, however this window might give us some information we want and it is easy to work with. (We won't find combinatorial lines there, so I don't see direct a proof for DHJ, but it's not a surprise ...)



6. [Randall](#) Says:

[February 13, 2009 at 9:06 pm](#) | [Reply](#)

## 505. Uniformity norms:

The ergodic proof of DHJ reduces it to an IP recurrence theorem. Now, for an IP system setup, we can easily define analogs of the (ergodic) uniformity norms. So, let  $(T_a)$  be a measure preserving IP system;  $a$  ranges over finite subsets of the naturals. Put  $\|f\|_1 = IP - \lim (\int f T_a f)^{\frac{1}{2}}$ . Now let  $P$  be the projection onto the factor that is asymptotically invariant under  $(T_a)$ ; we can write  $\|f\|_1 = \|Pf\|_{L^2}$  if we like. Now put  $\|f\|_2 = IP - \lim_b IP - \lim_a (\int f T_a f T_b f T_{a \cup b} f)^{\frac{1}{4}}$ . (This should look awfully familiar.) Presumably, if  $\|f\|_2$  is small, where  $f$  is the balanced version of the characteristic function of a set, then the set should have, asymptotically, the right number of arithmetic progressions of length 3 whose difference comes from some relevant IP set (or something like that). On the other hand, if  $\|f\|_2$  is big,  $f$  must correlate with....well, let's just say if  $\|f\|_1$  is big then  $f$  correlates with a rigid function, that much is easy. At any rate I think the above may be on the right track. I may try to translate from ergodic theory to a more recognizable form, but right now a baby is waking up behind me....

However, I will say this much...there is no conventional averaging in any proof of any ergodic IP recurrence theorem (IP limits serve that purpose in this context). Also, there are extremely strong analogies between the uniformity norms on  $\mathbb{Z}_N$  and the uniformity norms in ergodic theory (actually...to me it's not really an analogy; these are exactly the same norms). Although no norms are defined in ergodic IP theory, there could be...what does a norm stand in for? To

an ergodic theorist, the  $d$ th uniformity norm is a stand-in for  $d$  iterations of van der Corput. In the IP case, we do in fact use  $d$  iterations of a version of van der Corput. If the correlation holds to form, the above norms may well be the ones to be looking at.

7.  *jozsef* Says:

[February 13, 2009 at 9:08 pm](#) | [Reply](#)

505. Combining obstructions.

In my previous post I said that we won't see combinatorial lines in a random permutation. Now I'm not that sure about it. So, here is a simple statement which would imply DHJ (so it's false most likely, but I can't see a simple counterexample) If you take a dense subset of pairwise disjoint intervals in  $[n]$  then there are four numbers  $a, b, c, d$  in increasing order that  $[a, b], [c, d]$  and  $[a, c], [c, d]$  and  $[a, b], [b, d]$  are in your subset.

The rough numbers - counting the probability that there is no such configuration in a random permutation - didn't give me any hint. I have to work with a more precise form of Stirling's formula. I need time and I will work away from the computer. (I have to leave shortly anyways)

8.  *jozsef* Says:

[February 13, 2009 at 9:18 pm](#) | [Reply](#)

506.

The "simple statement" above is clearly false, just take pairs where one interval is between 1 and  $n/2$  and the second is between  $n/2$  and  $n$ . But I think this is not a general problem for our application where we might be able to avoid such "bipartite" cases.

9.  *bengreen* Says:

[February 14, 2009 at 1:07 am](#) | [Reply](#)

507.

Tim,

Regarding your Example 2 in 502 above, here is a cheat way to get a global obstruction. If you correlate with a linear phase on many progressions of length  $\sqrt{N}$  then you also have large  $U^2$  norm on each such progression. Summing over all progressions, this means that the sum of  $f$  over all parallelograms  $(x, x+h, x+k, x+h+k)$  with  $|k/h| \leq \sqrt{N}$ , or roughly that, is large.

But even if you only control such parallelograms with  $k = 2h$  this is the same as controlling the average of  $f$  over 4-term progressions, and one knows that is controlled by the \*global\*  $U^3$  norm and hence by a global quadratic object (nilsequence). That will allow you to "tie together" the frequencies of the trig polys that you correlated with at the beginning, though I haven't bothered to think exactly how. One just needs to analyse which 2-step nilsequences correlate with a linear phase on many progressions of length  $\sqrt{N}$ , and it might just be ones which are essentially 1-step, and that would imply that  $f$  correlates globally with a linear phase.

BY the way much the same argument shows that if you have large  $U^k$  norm on many progressions of length  $\sqrt{N}$  (say) then you have large global  $U^{\{2^k - 1\}}$  norm.

Ben

10.  *Randall* Says:

[February 14, 2009 at 1:13 am](#) | [Reply](#)



## 508. Ergodic-mimicking general proof strategy:

Okay, I will give my general idea for a proof. I'm pretty sure it's sound, though it may not be feasible in practice. On the other hand I may be badly mistaken about something. I will throw it out there for someone else to attempt, or say why it's nonsense, or perhaps ignore. I won't formulate it as a strategy to prove DHJ, but of what I've called IP Roth. If successful, one could possibly adapt it to the DHJ,  $k=3$  situation, but there would be complications that would obscure what was going on.

We work in  $X = [n]^{\times [n]}$ . For a real valued function  $f$  defined on  $X$ , define the first coordinate 2-norm by  $\|f\|_2^2 = (\lim_{b \rightarrow \infty} \lim_{a \rightarrow \infty} \frac{1}{|X|} \sum_{(x,y) \in X} f(x,y)f(x+a,y)f(x+b,y)f(x+a+b,y))^{1/4}$ . The second coordinate 2-norm is defined similarly (on the second coordinate, obviously). Now, let me explain what this means.  $A$  and  $B$  are subsets of  $[n]$ , and we identify  $A$  with the characteristic function of  $A$ , which is a member of  $[n]^{\times [n]}$ . (That is how we can add  $A$  to  $X$  inside, etc. Since  $[n]$  is a finite set, you can't really take limits, but if  $n$  is large, we can do something almost as good, namely ensure that whenever  $\max \alpha < \min \beta$ , the expression we are taking the limit of is close to something (Milliken Taylor ensures this, I think). Of course, you have to restrict  $A$  and  $B$  to a subspace. What is a subspace? You take a sequence  $A_i$  of subsets of  $[n]$  with  $\max A_i < \min A_{i+1}$  and then restrict to unions of the  $A_i$ .

Now here is the idea. Take a subset  $E$  of  $X$  and let  $f$  be its balanced indicator function. You first want to show that if \*either\* of the coordinate 2-norms of  $f$  is small, then  $E$  contains about the right number of corners  $\{(x,y), (x+a,y), (x,y+a)\}$ . Restricted to the subspace of course. What does that mean? Well, you treat each of the  $A_i$  as a single coordinate, moving them together. The other coordinates I'm not sure about. Maybe you can just fix them in the right way and have the norm that was small summing over all of  $X$  still come out small. At any rate, the real trick is to show that if \*both\* coordinate 2-norms are big, you get a density increment on a subspace. Here a subspace surely means that you find some  $A_i$ s, treat them as single coordinates, and fix the values on the other coordinates.

11.  [gowers](#) Says:


[February 14, 2009 at 1:19 am](#) | [Reply](#)

## 509. Quasirandomness.

Jozsef, re 504, what you describe is what I thought I was doing in the second paragraph of 500. The nice thing about taking a random permutation and then only looking at intervals is that you end up with dense graphs. The annoying thing is that these dense graphs don't instantly give you combinatorial lines. However, they do give correlation with a Cartesian product, so my hope was to put together all these small dense Cartesian products to get a big dense "Kneser product". At the moment I feel quite hopeful about this. I've got to go to bed very soon, but roughly my planned programme for doing it is this.

Suppose you have proved that there are a lot of pairwise correlations between the  $g_i$ . Then I want to think of these  $g_i$  as unit vectors in a Hilbert space and find some selection of them such that most pairs of them correlate. Here's a general dependent-random-selection procedure for vectors in a finite-dimensional Hilbert space. Suppose I have unit vectors  $x_1, \dots, x_n \in \mathbb{R}^N$ . Now I choose random unit vectors  $y_1, \dots, y_k$  and pick all those  $x_i$  such that  $\langle x_i, y_j \rangle > 0$  for every  $j \leq k$ . The probability of choosing any given  $x_i$  is obviously  $2^{-k}$ . Now the probability that  $\langle x, y_i \rangle$  and  $\langle x', y_i \rangle$  are both positive is equal to the proportion of the sphere in the intersection of the positive hemispheres defined by  $x$  and  $x'$ , which has measure  $1 - \theta/\pi$ , where  $\theta$  is the angle between  $x$  and  $x'$ . So the probability of picking both  $x$  and  $x'$  is  $(1 - \theta/\pi)^k$ . So if  $k \gg 1/\delta$  and we have lots of correlations of at least  $\delta$ , then we should end up with almost all pairs that we choose having good correlations. It is some choice like this that I am currently hoping will give rise to a global obstruction to uniformity.

The more I think about this, the more I am beginning to feel that it's got to the point where it's going to be hard to make progress without going away and doing calculations. I don't really see any other way of testing the plausibility of some of the ideas I am suggesting.

12.  [Terence Tao](#) Says:  
[February 14, 2009 at 8:37 am](#) | [Reply](#)

### 510. Stationarity

The reading seminar is already paying off for me, in that I now see that the random Bernoulli variables associated to a Cartesian half-product automatically enjoy the stationarity property, which will presumably be useful in showing that those half-products contain lines with positive probability. (I also plan to write up this strategy on the wiki at some point.)


Let's recall the setup. We have a set  $A$  of density  $\delta$  in  $[3]^n$  that we wish to find lines in. Pick a random  $x$  in  $[3]^n$ , then pick random  $a_1, \dots, a_m$  in the 0-set of  $x$ , where  $m = m(\delta)$  does not depend on  $n$ . This gives a random embedding of  $[3]^m$  into  $[3]^n$ , and in particular creates the strings  $x_{i,j}$  for  $0 \leq i \leq j \leq m$ , formed from  $x$  by flipping the  $a_1, \dots, a_i$  digits from 0 to 1, and the  $a_{j+1}, \dots, a_m$  digits from 0 to 2. I like to think of the  $x_{i,j}$  as a "Cartesian half-product", which in the  $m=4$  case is indexed as follows:

```
0000 0002 0022 0222 2222
1000 1002 1022 1222
1100 1102 1122
1110 1112
1111
```

Observe that  $x_{i,i}, x_{i,j}, x_{j,j}$  form a line whenever  $0 \leq i < j \leq m$ . We have the Bernoulli events  $E_{i,j} := \{x_{i,j} \in A\}$ .

We already observed that each of the  $E_{i,j}$  have probability about  $\delta$ , thus the one-point correlations of  $E_{i,j}$  are basically independent of  $i$  and  $j$ . More generally, it looks like the  $k$ -point correlations of the  $E_{i,j}$  are translation-invariant in  $i$  and  $j$  (so long as one stays inside the Cartesian half-product). For instance, when  $m = 4$ , the events  $E_{0,4} \wedge E_{0,3}$  and  $E_{1,4} \wedge E_{1,3}$  (which are the assertions that the embedded copies of 0000, 1000 and 0002, 1002 respectively both lie in  $A$ ) have essentially the same probability. This is obvious if you think about it, and comes from the fact that if you take a uniformly random string in  $[3]^n$  and flip a random 0-bit of that string to a 2, one essentially gets another uniformly random string in  $[3]^n$  (the error in the probability distribution is negligible in the total variation norm). Indeed, all the probabilities of  $E_{i,j+1} \wedge E_{i,j}$  are essentially equal for all  $i, j$  for which this makes sense. And similarly for other  $k$ -point correlations.

I don't yet know how to *use* stationarity, but I expect to learn as the reading seminar continues.

13.  [gowers](#) Says:  
[February 14, 2009 at 9:42 am](#) | [Reply](#)

### 511. Local-to-global

Re 502 Example 2 and 507, I think I now have a simple way of getting a global obstruction in this case, but it would have to be checked. Suppose that  $f$  is a function that correlates with a trigonometric function on a positive proportion of the APs in  $\mathbb{Z}_N$  of length  $m$ . Here  $m$  could be very small — depending on the size of the correlation only I think.

Now suppose I pick a random AP of length 3. Then it will be contained in several of these APs of length  $m$ , and if  $N$  is prime each AP of length 3 will be in the same number of APs of length  $m$ . At this stage there's a gap in the argument because lack of uniformity doesn't imply that you have the wrong sum over APs of length 3, but it sort of morally does, so I think it might well be possible to deal with this and get that the expectation of  $f(x)f(x+d)f(x+2d)$  is not small — perhaps by modifying  $f$  in some way or playing around with different  $m$ . And then we'd have a global obstruction by the  $U^2$  inverse theorem.

Alternatively, instead of using APs of length 3, it might be possible to use additive quadruples

$(x, x+a, x+b, x+a+b)$  but under the extra condition that there is a small linear relation between  $a$  and  $b$ . That would give a useful positivity. I haven't checked whether lack of uniformity on such additive quadruples implies that the  $U^2$  norm is large, but the fact that 3-APs give you big  $U^2$  is a promising sign.



14. [Ryan O'Donnell](#) Says:

[February 14, 2009 at 5:15 pm](#) | [Reply](#)

512. Stationarity.

Just a small note on Terry.#510: Is there any particular reason for choosing the indices  $a_i$  from the 0-set of the initial random string  $x$ ?

If not, the probabilistic analysis might be very slightly cleaner if we simply say, "Choose  $m$  random coordinates  $a_i$  and fill in the remaining  $n - m$  coordinates at random. Now consider the events  $E_{ij} \dots$ "



15. [Terence Tao](#) Says:

[February 14, 2009 at 9:57 pm](#) | [Reply](#)

513. Stationarity

Yeah, you're right; I was rather clumsily thinking of the base point  $x$  as being the lower left corner of the embedded space, when in fact one only needs to think of what  $x$  is doing outside of the variable indices. (I've updated the wiki to reflect this improvement.)



16. [Terence Tao](#) Says:

[February 15, 2009 at 8:06 am](#) | [Reply](#)

514. DHJ(2.6)

Back in 130 I asked a weaker statement than DHJ(3), dubbed DHJ(2.5): if  $A$  is a dense subset of  $[3]^n$ , does there exist a combinatorial line  $w^0, w^1, w^2$  whose first two positions only  $w^0, w^1$  are required to lie inside  $A$ ? It was quickly pointed out to me that this follows easily from DHJ(2).

From the reading seminar, I now have a new problem intermediate between DHJ(3) and DHJ(2.5), which I am dubbing DHJ(2.6): if  $A$  is a dense subset of  $[3]^n$ , does there exist an  $r$  such that for each  $ij=01, 12, 20$ , there exists a combinatorial line  $w^0, w^1, w^2$  with exactly  $r$  wildcards with  $w^i, w^j \in A$ ? So it's three copies of DHJ(2.5) with the extra constraint that the lines all have to have the same wildcard length. Clearly this would follow from DHJ(3) but is a weaker problem. At present I do not see a combinatorial proof of this weaker statement (which Furstenberg and Katznelson, by the way, spend an entire paper just to prove).



17. [gowers](#) Says:

[February 15, 2009 at 3:53 pm](#) | [Reply](#)

515. Different measures.

Terry, I'm not sure I understand your derivation of slices-equal DHJ from uniform DHJ. For convenience let me copy your argument and then comment on specific lines of it.

*Suppose that  $A \subset [3]^n$  has density  $\delta$  in the equal-slices sense. By the first moment method, this means that  $A$  has density  $\gg \delta$  on  $\gg \delta$  on the slices.*

Let  $m$  be a medium integer (much bigger than  $1/\delta$ , much less than  $n$ ).

Pick  $(a, b, c)$  at random that add up to  $n-m$ . By the first moment method, we see that with probability  $\gg \delta$ ,  $A$  will have density  $\gg \delta$  on the  $\gg \delta$  of the slices  $\Gamma_{(a', b', c')}$  with  $a' = a + m/3 + O(\sqrt{m})$ ,  $b' = b + m/3 + O(\sqrt{m})$ ,  $c' = c + m/3 + O(\sqrt{m})$ .

This implies that  $A$  has expected density  $\gg \delta$  on a random  $m$ -dimensional subspace generated by a 1s,  $b$  2s,  $c$  3s, and  $m$  independent wildcards.

Applying regular DHJ to that random subspace we obtain the claim.

One of my arguments for deducing slices-equal from uniform collapsed because sequences belonging to non-central slices are very much more heavily weighted than sequences that belong to central ones. I don't understand why that's not happening in your argument (though I haven't carefully done calculations). Consider for example what might happen if  $(a, b, c) = (n/6, n/3, n/2) - (m/3, m/3, m/3)$ . Then the weights of individual sequences in the slices of the form  $(a, b, c) + (r, s, t)$  go down exponentially as  $r$  increases and as  $t$  decreases, even if  $(r, s, t)$  is close to  $(m/3, m/3, m/3)$ . Or am I missing something?

In general, there seems to be a difficult phenomenon associated with the slices-equal measure, which is that unless you're lucky enough for  $A$  to contain bits near the middle, it will be very far from uniform on all combinatorial subspaces where it might have a chance to be dense. Worse still, it doesn't restrict to slices-equal measure on subspaces.

18.  [gowers](#) Says:

[February 15, 2009 at 5:28 pm](#) | [Reply](#)

516. Different measures.

On second thoughts, I do now follow your argument!

19.  [Terence Tao](#) Says:

[February 15, 2009 at 6:29 pm](#) | [Reply](#)

517. Different measures

I've put a summary of the different measures discussion on the wiki at

[http://michaelnielsen.org/polymath1/index.php?title=Equal-slices\\_measure](http://michaelnielsen.org/polymath1/index.php?title=Equal-slices_measure)


(More generally, I will probably be more silent on this thread than previously, but will be working "behind the scenes" by distilling some of the older discussion onto the wiki.)

20.  [Randall](#) Says:

[February 15, 2009 at 9:43 pm](#) | [Reply](#)


DHJ (2.6):

Here's what looks like a combinatorial proof, modulo reduction to sets and Graham-Rothschild being kosher (at least no infinitary theorems are used). Put everything in the sets formulation, where you have sets  $B_w$  of measure at least  $\delta$  in a prob. space. (No need for stationarity.) Now color combinatorial lines  $\{w(0), w(1), w(2)\}$  according to whether  $B_{w(0)} \cap B_{w(1)}$ ,  $B_{w(0)} \cap B_{w(2)}$  and  $B_{w(1)} \cap B_{w(2)}$  are empty or not (8 colors). By Graham-Rothschild, there is an  $n$ -dimensional subspace all of whose lines have the same color for this coloring. Just take  $n$  large enough for the sets version of DHJ (2.5).

21.  [Terence Tao](#) Says:  
[February 15, 2009 at 10:59 pm](#) | [Reply](#)

519.

Dear Randall, I like this proof, but I am a little worried that  $A$  is still dense on the subspace one passes to by Graham-Rothschild. If  $A$  is deterministic, it may well be that the space one passes to is completely empty; if instead  $A$  is random, then the Graham-Rothschild subspace will depend on  $A$  and it is no longer clear that all the events in that space have probability  $\delta$ .


22.  [Randall](#) Says:  
[February 16, 2009 at 12:40 am](#) | [Reply](#)

520.

Terry...the subspace doesn't depend on the random set  $A$ . What I am proving is the (b) formulation, not the (a) formulation. There seems to be some disconnect between the language of ergodic theory and the language of "random sets", so let me try to translate: one reduces to a subspace where, for each  $ij=01,12,02$ , one of the following two things happens:

1. for all variable words  $w$  over the subspace, the probability that both  $w(i)$  and  $w(j)$  are in  $A$  is zero.
2. for all variable words  $w$  over the subspace, the probability that both  $w(i)$  and  $w(j)$  are in  $A$  is positive.


If you still think there is an issue about a density decrement, just read my proof as a proof of the (b) formulation...there, you can't have a density decrement upon passing to a subspace because all the sets  $B_w$  have density at least  $\delta$ .

23.  [Terence Tao](#) Says:  
[February 16, 2009 at 1:50 am](#) | [Reply](#)

521.

Oh, I see now. That's a nice argument - and one that demonstrates an advantage in working in the ergodic/probabilistic setting (It would be interesting to see a combinatorial translation of the argument, though.)


I'll write it up on the wiki (on the [DHJ page](#)) now.

24.  [Ryan O'Donnell](#) Says:  
[February 16, 2009 at 5:35 am](#) | [Reply](#)

522. DHJ(2.6).


Hmm, I see how this is a bit trickier than DHJ(2.5). For example, for any of the three  $ij$  pairs you can probably show that if  $r \ll \text{Poisson}(m)$  for some moderate  $m$  then there is at least an  $\epsilon = \epsilon(\delta) > 0$  chance of having an " $ij$  pair" at distance  $r$ . But the trouble is, these could potentially be different measure  $\epsilon$  sets, and since they're small it's not so obvious how to show they're not disjoint.

As an example of the trickiness, let  $q = 1/\sqrt{\delta}$ , assumed an integer. Perhaps  $A$  is the set of strings whose 0-count is divisible by  $q$  and whose 1-count is divisible by  $q + 1$ . Now 02-pairs have to be at distance which is a multiple of  $q$ , and 12-pairs have to be at a distance which is a multiple of  $q + 1$ . So you have to be a bit clever to extract a common valid distance, a multiple of  $q(q + 1)$ .

25.  [Gil](#) Says:  
[February 16, 2009 at 8:01 am](#) | [Reply](#)

Ryan,


Can you explain more your approach in comment 477. ?

26.  [Terence Tao](#) Says:  
[February 16, 2009 at 8:38 am](#) | [Reply](#)


524. DHJ(2.6)

A small observation: one can tone the Ramsey theory in Randall's proof of DHJ(2.6) a notch, by replacing the Graham-Rothschild theorem by the simpler Folkman's theorem (this is basically because of the permutation symmetry of the random subcube ensemble). Indeed, given a dense set  $A$  in  $[3]^n$  we can colour  $[n]$  in eight colours, colouring a shift  $r$  depending on whether there exists a combinatorial line with  $r$  wildcards whose first two (or last two, or first and last) vertices lie in  $A$ . By Folkman's theorem, we can find a monochromatic  $m$ -dimensional pinned cube  $Q$  in  $[n]$  (actually it is convenient to place this cube in a much smaller set, e.g.  $[n^{0.1}]$ ). If the monochromatic colour is such that all three pairs of a combinatorial line with  $r$  wildcards can occur in  $A$ , we are done, so suppose that there is no line with  $r$  wildcards with  $r$  in  $Q$  with (say) the first two elements lying in  $A$ .

Now consider a random copy of  $[3]^m$  in  $[3]^n$ , using the  $m$  generators of  $Q$  to determine how many times each of the  $m$  wildcards that define this copy will get. The expected density of  $A$  in this cube is about  $\delta$ , so by DHJ(2.5) at least one of the copies is going to get a line whose first two elements lie in  $A$ , which gives the contradiction.

27.  [Terence Tao](#) Says:  
[February 16, 2009 at 8:59 am](#) | [Reply](#)

p.s. I have placed the statement of relevant Ramsey theorems (Graham-Rothschild, Folkman, Carlson-Simpson, etc.) on the wiki for reference.

28.  [Gil Kalai](#) Says:  
[February 16, 2009 at 2:22 pm](#) | [Reply](#)

525. Strategy based on influence proofs

Let me say a little more (vague) things about possible strategy based on mimicking influence proofs. We want to derive facts on the Fourier expansion of our set based on the fact that there are no lines and let me just consider  $k=2$  (Sperner). Suppose that your set  $A$  has density  $c$  and you know that there is no lines with one wild cards. This gives that

$$\sum f^2(S)|S| = cn/2. \text{ (In a random set we would expect } (c - c^2)n/2; \text{ anyway this is of course possible.)}$$


Next you want to add the fact that there are no lines with 2 wild cars.

The formula (in fact there are few and we have to choose a simple one) for the number of lines with two wild cards  $i$  and  $j$  and the conclusion for the Fourier coefficients are less clear to me. Maybe we can derive conclusions on the higher moments  $\sum f^2(S)|S|^2$  or some other indormations.

Here, we do not want to reach a subspace with higher density but rather a contradiction. In the usual influence proofs the tool for that were certain inequalities asserting that for sets with small support the Fourier coefficients are high. This particular conclusion will not help us but the inequalities can help in some other way.




If indeed we are getting from the no line assumptions that the Fourier coefficients are high and concentrated this may be in conflict with facts about codes. Namely that our  $A$  will be a better than possible code in terms of its distance distribution. Anyway the first thing to calculate would be a good Fourier expression for the number of lines with 2 wildcards.

29.  *Ryan O'Donnell* Says:  
[February 16, 2009 at 3:18 pm](#) | [Reply](#)

526. DHJ(2.6).

Thanks for the observation about using Folkman's Theorem, Terry, it looks quite nice. I have some other ideas about trying to give a Ramsey-less proof based on Sperner's Theorem; I'll try to write them soon.

30.  *Ryan O'Donnell* Says:  
[February 16, 2009 at 3:52 pm](#) | [Reply](#)

527. Fourier approach to Sperner.

Gil, here are a few more details re #477. First, I see I dropped the Fourier coefficients at the beginning of the post, whoops! What I meant to say is that if you do the usual Fourier thing, you get the probability of  $x, y \in \mathcal{A}$  is

$$\mathcal{B}(f) := \sum_{|S \Delta T| \text{ even}} (1 - 2\epsilon)^{|S \cap T|} \epsilon^{|S \Delta T|} (-1)^{|S \setminus T|} \hat{f}(S) \hat{f}(T).$$

The intuition here is that since  $\epsilon$  is tiny, hopefully all the terms involving a positive power of  $\epsilon$  are “negligible”. Obviously this is unlikely to precisely happen, but if it were true then we would get


$$\mathcal{B}(f) \approx \sum_S (1 - 2\epsilon)^{|S|} \hat{f}(S)^2,$$

a noise stability quantity. Now there are something like  $2^{2n}$  terms we want to drop, so it would be hard to do something simple like Cauchy-Schwarz. My idea is one from a recent paper with Wu. Think of  $\mathcal{B}(f)$  as a quadratic form over the Fourier coefficients of  $f$ . I.e., think of  $\hat{f}$  as a vector of length  $2^n$  and think of

$$\mathcal{B}(f) = \hat{f}^\top B \hat{f},$$

where  $B$  is a matrix whose  $(S, T)$  entry is the obvious expression from above with  $\epsilon$ 's and  $|S \Delta T|$ 's and so on. Our goal is to find two similar-looking matrices  $A, C$  such that  $A \preceq B \preceq C$ , meaning  $C - B, B - A$  are psd. Then we get  $\mathcal{A}(f) \leq \mathcal{B}(f) \leq \mathcal{C}(f)$ . (I see I've used  $\mathcal{A}$  for two different things here; whoops.)

To be continued...

31.  *Ryan O'Donnell* Says:  
[February 16, 2009 at 3:53 pm](#) | [Reply](#)

528. Fourier approach to Sperner.

Let me explain what I hope  $C$  will be; the matrix  $A$  will be similar.

Let  $C_1$  be the  $2 \times 2$  matrix with entries  $1, \epsilon, \epsilon, 1 - \epsilon$ . This is like a little “ $\epsilon$ -biased  $\epsilon$ -noise stability” matrix. Let  $C_n = C_1^{\otimes n}$ . Similarly, let  $C'_1$  be the  $2 \times 2$  matrix with entries  $1, -\epsilon, -\epsilon, 1 - \epsilon$  and define  $C'_n$ . Finally, let  $C = (C_n + C'_n)/2$ . The quadratic form associated to  $C$  should be that average of two noise stabilities I wrote in post #477, except with noise rate  $1 - \epsilon$ , not  $1 - 2\epsilon$ .

My unsubstantiated claim/hope is that indeed  $B \preceq C$ . Why do I believe this? Uh... believe it or not, because I checked

empirically for  $n$  up to 5 with Maple. 😊 Sorry, I guess this isn't how a real mathematician would do it. But I hope it can be proven.

To define the matrix  $A$ , do the same thing but set  $A_1$  to be the  $2 \times 2$  matrix with entries  $1, \epsilon, \epsilon, 1 - 3\epsilon$ , etc. So eventually I hope to sandwich  $\mathcal{B}(f)$  between the average of the  $\pm\epsilon$ -biased  $\epsilon$ -noise stabilities of  $f$  and the average of the  $\pm\epsilon$ -biased  $3\epsilon$ -noise stabilities of  $f$ .

PS: Although it may work for Sperner, I'm less optimistic ideas like this will work for DHJ. The trouble is, in a distribution on comparable pairs  $(x, y)$  for Sperner, there is \*imperfect\* correlation between  $x$  and  $y$ . Roughly, if you know  $x$ , you are still unsure what  $y$  is. On the other hand, in any distribution on combinatorial lines  $(x, y, z)$  for DHJ, there is \*perfect\* correlation between  $(x, y)$  and  $z$  (and the other two pairs). If you know  $x$  and  $y$ , then you know  $z$  with certainty. This seems to break a lot of hopes to use “Invariance”-type methods like those in Elchanan's follow-on to MOO. This is related to the fact that there is no hope to prove a generalization of DHJ of the form, “If  $\mathcal{A}, \mathcal{B}$  are two subsets of  $[3]^n$  of density  $\delta$ , then there is a combinatorial line  $(x, y, z)$  with  $x, y \in \mathcal{A}$  and  $z \in \mathcal{B}$ .”

32.  Ryan O'Donnell Says:

[February 16, 2009 at 5:10 pm](#) | [Reply](#)

529. Ramsey-less DHJ(2.6) plan.

Here is a potential angle for a Ramsey-free proof of DHJ(2.6). The idea would be to soup up the Sperner-based proof of DHJ(2.5) and show that the set  $D$  of (Hamming) distances where we can find a 01-half-line in  $\mathcal{A}$  is “large” and “structured”.

So again, let  $x$  be a random string in  $[3]^n$  and condition on the location of  $x$ 's 2's. There must be some set of locations such that  $\mathcal{A}$  has density at least  $\delta$  on the induced 01 hypercube (which will likely have dimension around  $(2/3)n$ ). So fix 2's into these locations and pass to the 01 hypercube. Our goal is now to show that if  $\mathcal{A}$  is a subset of  $\{0, 1\}^n$  of density  $\delta$  then the set  $D$  of Hamming distances where  $\mathcal{A}$  has Sperner pairs is large/structured. (I've written  $n$  here although it ought to be  $n' \approx (2/3)n$ .)

Almost all the action in  $\{0, 1\}^n$  is around the  $\Theta(\sqrt{n})$  middle slices. Let's simplify slightly (although this is not much of a cheat) by pretending that  $\mathcal{A}$  has density  $\delta$  on the union of the  $\sqrt{n}$  middle slices \*and\* that these slices have “equal weight”. Index the slices by  $i \in [\sqrt{n}]$ , with the  $i$ th slice actually being the  $n/2 - \sqrt{n}/2 + i/2$  slice.

Let  $\mathcal{A}_i$  denote the intersection of  $\mathcal{A}$  with the  $i$ th slice, and let  $\mu_i$  denote the relative density of  $\mathcal{A}_i$  within its slice.

To be continued...

33.  Ryan O'Donnell Says:

[February 16, 2009 at 5:10 pm](#) | [Reply](#)

530. Ramsey-less DHJ(2.6) plan continued.

Here is a slightly wasteful but simplifying step: Since the average of the  $\mu_i$ 's is  $\delta$ , a simple Markov-type argument shows that  $\mu_i \geq \delta/2$  for at least a  $\delta/2$  fraction of the  $i$ 's. Call such an  $i$  “marked”.

Now whenever we have, say,  $3/\delta$  marked  $i$ 's, it means we have a portion of  $\mathcal{A}$  with a number of points at least  $3/2$  times the number of points in a single middle slice. Hence Sperner's Theorem tells us we get two comparable points  $x \leq y$  from among these slices.

Thus we have reduced to the following setup:

There is a graph  $G = (V, E)$ , where  $V \subseteq [\sqrt{n}]$  (the “marked  $i$ 's”) has density at least  $\delta/2$  and the edge set  $E$  has the

following density property: For every collection  $V' \subset V$  with  $|V'| \geq 3/\delta$ , there is at least one edge among the vertices in  $V'$ .

Let  $D$  be the set of “distances” in this graph, where an edge  $(i, j)$  has “distance”  $|i - j|$ . Show that this set of distances must be large and/or have some useful arithmetic structure.

34.  [Gil](#) Says:

[February 16, 2009 at 9:57 pm](#) | [Reply](#)

I wonder if the following “ideology” can be promoted or justified: A set without a line with at most  $r$  wildcards “behaves like” an error correcting code with minimal distance  $r$  where  $r$  grows to infinity (possibly very slowly) when  $r$  does. for  $k=2$  of course we do not get an error correcting codes but when we have to split the set between many slices points from different slices are far apart so it is a little like a behavior of a code. for  $k=3$  this looks much more dubious but maybe has some truth in it.

35.  [Randall](#) Says:

[February 16, 2009 at 10:40 pm](#) | [Reply](#)

532. Need for Carlson’s theorem?

Terry’s reduction to a use of Folkman instead of Graham-Rothschild, as well as his ideas for increasing the symmetry available, got me thinking there might be a way to reduce to a situation where you are using Hindman’s theorem instead of from Carlson’s theorem in the entire proof. Here is my first attempt...perhaps others can check if this makes sense. Fix a big  $n$  and let  $A$  be your subset of  $[k]^n$  that you want to find a combinatorial line in. Let  $Y$  be the space  $2^{[k]^n}$ , that is, all sets of words of length  $n$ .  $U^i_j$  is the map on  $Y$  that takes a set  $E$  of words and interchanges 0s and js in the  $i$ th place to get a new set of words, namely the set  $U^i_j E$ . (Some of the words in  $E$  may be unaffected; in particular, if no word in  $E$  has a 0 or j in the  $i$ th place,  $E$  is a fixed point.) Now let  $X$  be the orbit of  $A$  under  $G$ , the group gen. by these maps. It seems that to find a combinatorial line in any member of  $X$  will give you a line in  $A$ . Let  $B = \{E : 00 \dots 00 \in E\} \subset X$ . In general, for  $U \subset X$ , define the measure  $\mu(U)$  to be the fraction of  $g$  in  $G$  such that  $gA \in U$ . This should give measure to  $B$  equal to that of the relative density of  $A$ , and I am thinking the measure is preserved by the  $U^i_j$  maps. If all of this is right, then things are simpler than FK imagined, because we have that  $U^i_j$  commutes with  $U^k_l$  for  $i \neq k$ , which should allow for the use of Hindman in all places where before Carlson was used. In particular, it allows for the use of Folkman in any finitary reduction (rather than Graham-Rothschild). And, in the case of Sperner’s Lemma, well, that becomes a simple consequence of the pigeonhole principle. (Because it’s just recurrence for a single IP system.) To be a bit more precise, in the  $k=3$  case, I want to prove that there is some  $\alpha \subset [n]$  such that  $\mu(B \cap \prod_{i \in \alpha} U^i_1 B \cap \prod_{i \in \alpha} U^i_2 B) > 0$  and conclude DHJ (3). The point is you get to IP systems much more easily with the extra commutativity assumption.

Sorry for the ergodic wording. Really at this state I am just hoping someone will check to see if everything I am saying is wrong for some simple reason I have overlooked. If not, we can write up a trivial proof of Sperner and a simplified version of  $k=3$  and go from there.

36.  [Randall](#) Says:

[February 16, 2009 at 11:10 pm](#) | [Reply](#)

533. Well it seems that was wrong...the maps don’t take lines to lines. Can anyone else think of a way to preserve some of the symmetry that Terry has gotten in the early stages?

37.  [jozsef](#) Says:

[February 17, 2009 at 12:08 am](#) | [Reply](#)

## 534. Folkman's theorem

Back in 341 and in some following comments I tried to popularize Folkman's theorem as a statement relevant to our project. I didn't find a meaningful density version yet, however Tim pointed out that statements like Ben Green's removal lemma for linear equations might work. It would be something like this; If the number of IP<sub>d</sub>-sets is much less than what is expected from the density, then removing a few elements one can destroy all IP<sub>d</sub> sets. In the next post I will say a few more words about the set variant of Folkman's theorem, however I'm not sure that this 500 blog series is the best forum for that.



38. [Jozsef](#) Says:

[February 17, 2009 at 12:41 am](#) | [Reply](#)

## 535. Folkman's theorem

on second thought I decided not to talk about Folkman's theorem here, since this blog is about possible proof strategies for DHJ. I will try to find another forum to discuss some questions on Folkman's theorem.



39. [Terence Tao](#) Says:

[February 17, 2009 at 6:27 pm](#) | [Reply](#)

## 536.

Jozsef: I guess this is the best forum we have right now (we tried splitting up into more threads earlier, but found that all but one of them would die out quickly). I can always archive the discussion at the wiki to be revived at some later point.

I wanted to point out two thoughts here. Firstly, I think I have a Fourier-analytic proof of DHJ(2.6) avoiding Ramsey theory. Firstly, observe that the usual proof of Sperner shows that the set of wildcard-lengths  $r$  of the combinatorial lines in a dense subset of  $[2]^n$  contains a difference set of the form  $A-A$  where  $A$  is a dense subset of  $[\sqrt{n}]$ . Using the deduction of DHJ(2.5) from Sperner, we see that the set of wildcard lengths of combinatorial lines whose first two points lie in a dense subset of  $[3]^n$  also contains a similar difference set. Similarly for permutations. So it comes down to the claim that for three dense sets  $A, B, C$  of  $[\sqrt{n}]$ , that  $A-A, B-B, C-C$  have a non-trivial intersection outside of zero.

This can be proven either by the triangle removal lemma (try it!) or by the arithmetic regularity lemma of Green, finding a Bohr set on which  $A, B, C$  are dense and pseudorandom. There may also be a more direct proof of this.

Secondly, I think we may have a shot at a combinatorial proof of Moser(3) - that any dense subset  $A$  of  $[3]^n$  contains a *geometric* line rather than a combinatorial line. This is intermediate between DHJ(3) and Roth, but currently has no combinatorial proof. Here is a sketch of an idea: we use extreme localisation and look at a random subcube  $[3]^m$ . Let  $B$  be the portion of  $A$  on the corners  $\{1, 3\}^m$  of the subcube, and let  $C$  be the portion of  $A$  which is distance one away (in Hamming metric) from the centre  $2^m$  of this subcube. The point is that there are a lot of potential lines connecting one point of  $C$  with two nearly opposing points of  $B$ . I haven't done it yet, but it looks like the number of such lines has a nice representation in terms of the low Fourier coefficients of  $A$ , and should therefore tell us that in order to be free of geometric lines,  $A$  has to have large influence. I am not sure where to take this next, but possibly by boosting this fact with the Ramsey theory tricks we can get some sort of contradiction.




40. [Gil](#) Says:

[February 17, 2009 at 8:04 pm](#) | [Reply](#)

Here is a nice example I heard from Muli Safra. (It's origin is in the problem of testing monotonicity of Boolean functions.) Consider a Boolean function  $f(x_1, x_2, \dots, x_n) = g(x_1, x_2, \dots, x_m)$  where  $m = o(\sqrt{n})$  and  $g$  is a random Boolean function. When you consider strings  $x$  and  $y$  that correspond to two random sets  $S$  and  $T$  so that  $S$  is a subset

of  $T$  with probability close to one  $f(x) = f(y)$  because almost surely the variables where  $S$  and  $T$  differs miss the first  $m$  variables. We can have a similar example for an alphabet of three letters.

41.  *Ryan O'Donnell* Says:

[February 17, 2009 at 8:06 pm](#) | [Reply](#)

537. DHJ(2.6)

Hi Terry, re your #536: great! this is just what I was going for in #529 & #530... But one aspect of the deduction I didn't quite get: in my #530, you don't have the differences from all pairs in  $V$ , just the ones where you have an "edge". So do you produce your set  $A$  by filtering  $V$  somehow? Or perhaps I'm missing some simple deduction from the proof of Sperner.

42.  *Gil* Says:

[February 17, 2009 at 8:20 pm](#) | [Reply](#)

538. Fourier of line avoiding sets.

Thanks a lot for the details, Ryan, it is very interesting. I think that using Maple **is** the way a real mathematician will go about it!

As for my vague (related) suggestions: There is a nice Fourier expression of subsets of  $\{0, 1\}^n$  without a combinatorial line with one wild cards. But already when you assume there is no lines with one and two wildcards its get messy. (In particular these scare expressions  $\hat{f}(S)\hat{f}(R)$ .)

Maybe there would be a nice expression for sets in  $\{0, 1, 2\}^n$  without a combinatorial line with one wildcard. This can be nice.

Another little (perhaps unwise) question regarding avoiding special configurations of "small distance" vectors. Suppose you look at subsets of  $\{0, 1\}^n$  without two distinct sets  $S$  and  $T$  so that  $S \setminus T$  has precisely twice as many elements as  $T \setminus S$ . Does this implies that the size of the family is  $o(2^n)$ ? (Again you can take a slice.)

43.  *Ryan O'Donnell* Says:

[February 17, 2009 at 8:24 pm](#) | [Reply](#)

538. Moser(3).

For this problem I think it might be helpful to go all the way back to Tim.#70 and think about "strong obstructions to uniformity" of the following form: dense sets  $\mathcal{A}$  for which  $\{y : \exists x, z \in \mathcal{A} \text{ s.t. } (x, y, z) \text{ is a geom. line}\}$  is not almost everything

44.  *Terence Tao* Says:

[February 17, 2009 at 9:06 pm](#) | [Reply](#)

540.

Re: Sperner: if  $A$  is a dense subset of  $[2]^n$ , then a random chain in this set is going to hit  $A$  in a dense subset of its equator (which is the middle  $O(\sqrt{n})$  of the chain, which has length  $n$ ). If  $A$  hits this chain in the  $i^{\text{th}}$  and  $j^{\text{th}}$  positions, then we get a combinatorial line with  $j-i$  wildcards. This is why the set of  $r$  arising from lines in  $A$  contain a difference set of a dense subset of  $[\sqrt{n}]$ .

Incidentally, I withdraw my claim that the joint intersection of  $A$ - $A$ ,  $B$ - $B$ ,  $C$ - $C$  can be established from triangle removal

- but the arithmetic regularity lemma argument should still work.

Finally, for Moser, I agree with Ryan's 538 - except that I think one should restrict  $y$  to be very close to the centre  $2^n$  of the cube  $\{1, 2, 3\}^n$ , otherwise  $y$  is only going to interact with a small fraction of the points of the cube and I don't think this will be easily detectable by global obstructions. I still haven't worked out what goes on when  $y$  is just a single digit away from  $2^n$  but I do believe it will have a nice Fourier-analytic interpretation.

45.  [gowers](#) Says:

[February 17, 2009 at 9:07 pm](#) | [Reply](#)

Metacomment.

I'd just like to chip in here and say that I've been devoting most of my polymath1 energies to developing the associated wiki. If anyone else has any thoughts that are in a sufficiently organized state to make wiki articles (they don't have to be perfect of course) then it would be great. It will make it much easier to keep track of what we know, what we've asked, what we think we might be able to prove and roughly how, etc. So far, I think Terry and I are the only two contributors. The other thing I wanted to say is that I get the sense that other people are doing a certain amount of thinking away from the blog. I've got to the point where I want to do this too (and report back regularly if I come up with anything) but I don't want to do it if others are in fact not doing so. Anybody care to tell me? It somehow feels as though we've got to the stage where a certain amount of private thought is needed to push things further, but because I formulated rule 6 I have felt obliged to stick to it. (Gil, I know you are relaxed about all this.)

46.  [Randall](#) Says:

[February 17, 2009 at 9:20 pm](#) | [Reply](#)

541. Cave man methods:

Here's a curious low-tech trick to get difference sets to intersect: suppose you have sets  $A, B, C$  of density  $d$ . Let  $m > 1/d$  and use Ramsey's theorem to choose  $n$  so that for any 8-coloring of the 2-member sets of an  $n$  element set, you get an  $m$  element set whose 2-member sets are of one color. Now color  $\{x, y\}$ ,  $x \leq y$ , the  $x_{i+1} + A$  cannot all be disjoint, hence  $x_j - x_i$  is in  $A - A$  for some, hence all,  $i, j$  and similarly for  $B - B, C - C$ .

47.  [Randall](#) Says:

[February 17, 2009 at 9:24 pm](#) | [Reply](#)

541. Cave man again:

I see stuff in what are interpreted as comment brackets gets left out! Let's try this again.

Here's a curious low-tech trick to get difference sets to intersect: suppose you have sets  $A, B, C$  of density  $d$ . Let  $m > 1/d$  and use Ramsey's theorem to choose  $n$  so that for any 8-coloring of the 2-member sets of an  $n$  element set, you get an  $m$  element set whose 2-member sets are of one color. Now color  $\{x, y\}$ ,  $x < y$ , according to which of  $A - A, B - B, C - C$   $y - x$  is in. Pass to an  $m$  member set  $\{x_1, \dots, x_m\}$  whose 2-member subsets are of one color for this coloring. Since  $m$  is greater than  $1/d$ , the  $x_{i+1} + A$  cannot all be disjoint, hence  $x_j - x_i$  is in  $A - A$  for some, hence all,  $i, j$  and similarly for  $B - B, C - C$ .

48.  [Ryan O'Donnell](#) Says:


[February 17, 2009 at 9:28 pm](#) | [Reply](#)

542. Metacomment.

I too feel like trying out longer calculations off the blog from time to time. Tim, why don't you go ahead and try the




ones you're thinking of? Perhaps we can just report here how the calculations go...

49.  [Terence Tao](#) Says:  
[February 18, 2009 at 6:11 am](#) | [Reply](#)

Metacomment.


I think there may be some tension here between the objective of solving the problem by any means necessary, and the objective of trying to see whether a maximally collaborative approach to problem-solving can work. I would favour a relaxed approach at this point; it seems that we are already getting enough benefit from the collaborative forum here that we don't need to be utterly purist about keeping it that way. And I think things are moving to the point where, traditionally, one of us would simply sit down for an hour or two and work out a lot of details at once.

50.  [Terence Tao](#) Says:  
[February 18, 2009 at 6:19 am](#) | [Reply](#)


543. Moser(3)

Hmm, there do seem to be some obstructions to uniformity here that are annoying. For instance, if  $m$  is an odd number comparable to  $0.1n$ , and  $A$  is the set of points in  $\{1,3\}^n$  which have an even number of 3's in the first  $m$  positions, and  $B$  is the set of points formed by taking  $2^n$  and changing one of the last  $n-m$  positions to a 1 or a 3, then there are no geometric lines connecting  $A$ ,  $B$ , and  $A$  despite  $A$  having density 0.5 and  $B$  having density 0.9 in  $\{1,3\}^n$  and the radius 1 Hamming ball with centre  $2^n$  respectively. On the other hand,  $A$  has a lot of coordinates with low influence wrt swapping 1 and 3; if 1 and 3 were completely interchangeable then Moser(3) would collapse to DHJ(2), so perhaps there is something to exploit here.

Randall: I'm beginning to realise that if one wants to hold on to permutation symmetry on the cube, then one can't use Ramsey theorems such as Graham-Rothschild or Carlson, as the symmetry is broken when one passes to a combinatorial subspace (the wildcard lengths are unequal). So it may actually not be a good tradeoff.

51.  [Gil](#) Says:  
[February 18, 2009 at 8:15 am](#) | [Reply](#)

Metacomment. I also agree that we need some off-line calculations, and we can keep the spirit of point 6 (asking not to go away for weeks to study some avenue) by participants trying to document and describe what they try to do and what they do (even if unsuccessful) in short time intervals.

52.  [gowers](#) Says:  
[February 18, 2009 at 10:00 am](#) | [Reply](#)

Metacomment. I think we all basically agree about off-line calculations. At some point fairly soon I will try to do some, but I will be careful to follow Gil's suggestion. I will write a comment in advance to say what I am going to do, and I will report back frequently and in detail on what results from the calculations. (I may give less detail about the things that don't work, but even there I will try to give enough detail to save anyone else from needing to go off and repeat my failures.)

Briefly, the kinds of things I'd like to do are to write out a rigorous proof that a quasirandom set (in a sense that I have defined on the wiki) must contain many combinatorial lines, and to try to deduce something from a set's not being quasirandom. I'd also like to play around with Fourier coefficients on  $[2]^n$  with equal-slices measure and try to prove that the only obstruction to Sperner in that measure is lots of low-influence variables. I won't get down to either of these today, if anyone has any remarks that will potentially stop me or change how I go about it.

53.  [gowers](#) Says:

[February 18, 2009 at 1:47 pm](#) | [Reply](#)

544. Corners(1,3)

Over at the wiki I have written about a special case of the corners problem that is like DHJ(1,3). It's corners when your set  $A$  is of the form  $\{(x, y) : x \in U, y \in V, x + y \in W\}$ . Can anyone see an elementary proof that a dense set of this kind contains a corner? I can prove it fairly easily if I'm allowed to use Szemerédi's theorem, so at least the transition from one to two dimensions is elementary, but I'd rather a fully elementary proof. My account of the problem (and of DHJ(1,3)) can be found [here](#).

54.  [Terence Tao](#) Says:

[February 18, 2009 at 4:49 pm](#) | [Reply](#)

Metacomment.

I should point out that over at the 700 thread, things have already advanced to the point where we have already had several people run programs to get the latest upper and lower bounds (and writing programs is not something that one can really do collaboratively, at least not with the structures in place right now).

55.  [gowers](#) Says:

[February 18, 2009 at 11:12 pm](#) | [Reply](#)

545. Fourier analysis on equal-slices  $[2]^n$

This isn't the calculation I was talking about, as this one was possible to do in my head. I don't yet know where we could go with this, but it felt like a good idea to think about how Walsh functions behave with respect to equal-slices measure. (I used to call this slices-equal measure, which I feel is slightly more correct, but in the end I have had to concede that "equal-slices" trips off the tongue much better.) In particular, I wondered about orthogonality.

The calculation turns out to be rather pretty. Let's write  $w_A$  for the Walsh function associated with the set  $A$ . That is,  $w_A(x)$  is 1 if  $x$  has an even number of 1s in  $A$ , and -1 if it has an odd number of 1s in  $A$ . Then  $w_A(x)w_B(x) = w_{A \Delta B}(x)$ .

Therefore, the inner product  $\mathbb{E}_x w_A(x)w_B(x)$  is equal to  $\mathbb{E}_x w_{A \Delta B}(x)$ , so what we really care about is the expectation of individual Walsh functions. In the case of uniform measure, these expectations are all zero except when  $A$  is the empty set, and this implies the orthogonality of the Walsh functions.

It turns out to be quite easy to think about the value of  $\mathbb{E}_x w_A(x)$  when expectations are with respect to equal-slices measure. This is because we can rewrite the expectation over  $[2]^n$  as the expectation over all permutations  $\pi$  of  $[n]$  and all integers  $m$  between 0 and  $n$  (in both cases chosen uniformly) of the sequence  $x$  such that  $x_{\pi(i)}$  is 1 if and only if  $i \leq m$ . (In set terms, we randomly permute  $[n]$  and then pick a random initial segment.)

Suppose that  $A$  has cardinality  $k$ . Then the calculation we need to do is this: what is the probability that if you choose a random subset  $B \subset [n]$  of size  $k$  and take a random  $m \in \{0, 1, \dots, n\}$ , then  $B$  will have an even number of elements less than or equal to  $m$ ? (If the probability is  $p$ , then  $\mathbb{E}_x w_A(x)$  will be  $2p - 1$  for every set  $A$  of size  $k$ .)

Now if  $k$  is odd and  $n$  is even, then we can replace  $B$  by  $n + 1 - B$  and we can replace  $m$  by  $n - m$  and the number of elements of  $n + 1 - B$  that are at most  $n - m$  is the number of elements of  $B$  that are greater than  $m$ . Also,  $B$  cannot equal  $n + 1 - B$ . Therefore, there is a bijection between pairs  $(B, m)$  that give an odd intersection and pairs  $(B, m)$  that give an even intersection. So the probability is  $1/2$  in this case, and we find, as we want, that  $\mathbb{E}_x w_A(x) = 0$ .

From this we find that  $w_A$  is orthogonal to  $w_B$  whenever  $A$  and  $B$  have different parities (assuming that  $n$  is even, which we may as well).

If  $A$  has even cardinality, things do not work so well. Consider, for example, the case where  $A$  consists of two elements. Here we choose a random  $m$  and a random set  $\{x, y\}$  and ask for the probability that an even number of  $x$  and  $y$  are at most  $m$ . For fixed  $m = cn$  the probability is approximately  $c^2 + (1 - c)^2$ , which integrates to  $2/3$ . So we get  $\mathbb{E}_x w_A(x) = 4/3 - 1 = 1/3$ .

Another way to think of this is that  $x$  and  $y$  divide up  $[n]$  into three subintervals, each of expected size  $n/3$ , so the probability that  $m$  lands in the middle subinterval is on average  $1/3$ . This second way is quite useful, as it shows that for general even  $k = 2s$  we will get approximately  $(s + 1)/2s$  for the probability of an even intersection and therefore  $1/s$  for  $\mathbb{E}_x w_A(x)$ .

So we end up not with an orthonormal basis but with something fairly nice: a basis that splits into two sets  $B_0$  and  $B_1$  of vectors, one for even parity sets and one for odd parity sets, such that every vector in  $B_0$  is orthogonal to every vector in  $B_1$ , and vectors  $w_A$  and  $w_B$  in the same set are nearly orthogonal except if the symmetric difference of  $A$  and  $B$  is small.

56.  [gowers](#) Says:

[February 18, 2009 at 11:59 pm](#) | [Reply](#)

546. Fourier analysis and Sperner obstructions.

Now I want to try to do a similar calculation to see if we can say anything about obstructions to Sperner in the equal-slices measure. The model will be this. Let  $f$  be a bounded function defined on  $[2]^n$ . Now we pick a random pair of sets  $(U, V)$  by first picking a random permutation  $\pi$  of  $[n]$  and then picking two random initial segments of the permuted set. Then we take the expectation of  $f(U)f(V)$ . I am interested in what we can say if this expectation is not small. The dream would be to show that almost all the influence on  $f$  comes from just a small set.

Clearly the calculation we want to do is this: we pick two sets  $A$  and  $B$  and work out the expectation of  $w_A(U)w_B(V)$  over all pairs  $(U, V)$  chosen according to the distribution above.

I think it is fairly clear that this expectation will be small if either of  $A$  or  $B$  is large, by the arguments of the previous post. This is promising, but I don't know that I can see, without going off and calculating, whether it will be small enough to give us something useful (particularly as I'm not quite sure what substitute we will have, if any, for Parseval's identity). If  $A$  and  $B$  are both small, then there is quite a bit of dependence between  $w_A(U)$  and  $w_B(V)$ , or so it seems. For example, if  $A$  has size 2,  $B$  has size 4,  $A \subset B$ , and for a particular permutation the probability that an initial segment contains an even number of elements of  $A$  is close to 1, then that permutation bunches  $A$  up to one of the two ends of the interval, so  $B$  behaves like a set of size 2 and the probability of getting an even intersection with  $B$  is roughly  $2/3$ .

This is all suggesting to me that lack of uniformity in Sperner implies that a few variables have almost all the influence. If any influence experts think they can see where these thoughts would be likely to lead, then I'd be interested to hear about it. Otherwise, this is one of the private calculations I'd like to do. If it worked, then we would have some modest evidence for what I very much hope will turn out to be the case: that with equal-slices measure the only obstructions to uniformity for DHJ(3) come from sets of complexity 1 (which are defined in [the discussion of DHJ\(1,3\) on this wiki page](#)). This would be because of the fluke that Sperner-type obstructions happened to correlate with sets of complexity 1.

57.  [Ryan O'Donnell](#) Says:

[February 19, 2009 at 6:25 am](#) | [Reply](#)

547. Equal slices measure / Sperner.

Hi Tim. I'm still not 100% on board for equal-slices measure 😊 but...

1. If I'm not mistaken, it's the same as first choosing  $p \in [0, 1]$  uniformly, and then choosing from the  $p$ -biased product distribution on  $\{0, 1\}^n$ . Thinking of it this way might help with Fourier methods, since they tend to mate nicely with product distributions.

2. When you say "obstructions to ...", do you mean obstructions to having probability close to  $\delta$  for the event  $U, V \in \mathcal{A}$  or for probability close to  $\delta^2$ ? The former may be difficult, since any monotone  $\mathcal{A}$  will lead to probability close to  $\delta$ , and there are an awful lot of monotone functions.

If you are interested in the standard product measure (as opposed to equal-slices) then the Fourier calculations from #477 and company seem like they could lead to a pretty good understanding of density-Sperner for a small Poisson number of wildcards. But I'm not currently sure if understanding density-Sperner really helps at all for understanding HJ or even Moser. (I may say more on why I feel this way later; I'm feeling a bit gloomy about Fourier approaches right now...)

58. [Which Coalition? « Combinatorics and more](#) Says:

[February 19, 2009 at 8:12 am](#) | [Reply](#)

[...] between the uniform distribution and the equal-slice distribution plays a role in a recent project discusses on Gowers's [...]



59. [Gil](#) Says:

[February 19, 2009 at 8:36 am](#) | [Reply](#)

Yes, integrating over  $\mu_p$  is a nice way to think about equal-slices probability. It would be interesting to come up with a good orthonormal basis of functions w.r.t. the equal slices probability. (Indeed, we have such functions for every  $\mu_p$ .)

I would be very interested to see a good understanding of density Sperner with small number of Poisson number of wild-cards. (I tried to do some computation and I do not see how to proceed right now even with Sperner.)



60. [gowers](#) Says:

[February 19, 2009 at 11:35 am](#) | [Reply](#)

549. Sperner obstructions

Ryan, that is indeed a nice way of thinking about equal-slices measure, and thanks for reminding me about monotone functions, which, despite a number of your previous comments, had slipped out of my mind. I must try to think what the most general phenomenon is, of which monotone functions are a particularly nice manifestation. (For example, the intersection of an up-set and a down-set is likely to be highly non-quasirandom for Sperner.) Better still, it would be good to generalize that to subsets of  $[3]^n$ , where I don't find it completely obvious what one should mean by a "monotone" set. (Of course, one can come up with definitions, but I'd like a definition with a proper justification.) In general, however, it seems to me that anything even remotely monotone-like in  $[3]^n$  ought to correlate with a set of complexity 1.



61. [gowers](#) Says:

[February 19, 2009 at 1:38 pm](#) | [Reply](#)

550. Equal-slices orthonormal basis

Ryan, I completely understand your not being 100% convinced by equal slices probability. I myself am in the position

of (i) very much wanting it to do something good and (ii) feeling that it would be a bit of a miracle if it did.

I've been thinking about Gil's question of finding a nice orthonormal basis of functions with respect to equal-slices density, and have just come up with a guess. In the true polymath spirit I will present the guess and only after that see whether it actually works. I've got a good feeling about it though (because I scientifically inducted from 2 to the general case).

Here are my initial desiderata. I would like a set of functions  $w_A$  (sort of pseudo-Walsh functions), one for each subset  $A \subset [n]$ , such that, regarding the elements of  $[2]^n$  as subsets of  $[n]$ , the value of  $w_A(B)$  depends only on the cardinality of  $A \cap B$ . I would also like them to be orthogonal in the equal-slices measure. (I think this could be enough to nail the functions completely, but I'm just going to guess a solution. If it turns out to be rubbish then I'll go away and try to solve some simultaneous equations.)

There is more or less no choice about what to do with singletons: up to a constant multiple, I've basically got to define  $w_i(B)$  to be 1 if  $i \notin B$  and 0 otherwise. Since the equal-slices weight of  $B$  is equal to that of its complement, this function averages zero in the equal-slices measure, and is therefore orthogonal to the constant function 1 (which is what we have to choose for  $w_A$  when  $A$  is empty).

Things get more complicated with pairs. As I argued in an earlier comment, if you just choose the usual Walsh function  $w_A(B) = (-1)^{|A \cap B|}$  then with equal-slices measure you get that the average of the function is  $1/3$  (when  $|A| = 2$ ). And that was basically because if you randomly choose two points in  $[n]$  and then choose a random  $m$  between 0 and  $n$ , then the probability that exactly one of your two random points is less than  $m$  is approximately  $1/3$  instead of  $1/2$ . So the nicest thing to do seems to be this (always assuming that  $|A| = 2$  for this discussion): we define  $w_A(B)$  to be 1 if  $|A \cap B|$  is 0 or 2 and  $-2$  if  $|A \cap B| = 1$ .

Now let's jump straight to the guess. I won't bother normalizing anything so I'm aiming for orthogonality rather than orthonormality. If  $|A| = k$  then I'd like  $w_A(B)$  to be  $\binom{k}{j} (-1)^{|A \cap B|}$ .

OK, now let me begin the probably painful process of checking this guess. First of all, this set of functions is no longer closed under pointwise multiplication, but that was always going to be too much to ask, since otherwise the functions  $w_i$  would have determined everything and we'd have ended up with the usual Walsh functions. This has the annoying consequence that it's not enough just to check that  $w_A$  has average zero when  $A$  is non-empty. And I have to admit that at this stage I'm just hoping for a big slice of luck (no cheesy pun intended).

What, then, is the product  $w_A w_A$ ? Hmm, I think it's going to be beyond me to do this straight up on to the computer, so now I really am going to do an offline computation. I've got various things coming up so it may be an hour or two before I can report back. (Also, I want to check that the expectation of  $w_A$  really is zero. I'm satisfied that it is approximately zero, but it would be much nicer to have exact orthogonality.)

62.  [gowers](#) Says:

[February 19, 2009 at 3:52 pm](#) | [Reply](#)

551. Equal slices and orthonormality

Oh dear, just noticed, before getting round to any calculation, that this idea is dead before it's even got going. If for each  $A, B$  we want  $w_A(B)$  to depend on  $|A \cap B|$  only, then as I pointed out above we need  $w_A$  to be the usual Walsh function when  $A$  is a singleton. But for two such functions to be orthogonal we then need their pointwise product to average zero as well. But their pointwise product is the usual Walsh function associated with a set of size 2, and the whole motivation for this line of enquiry was that those Walsh functions did *not* average zero in the equal-slices measure.

This shows that my two desiderata above are inconsistent. In my next comment I'll give a suggestion for what to do.

63.  [gowers](#) Says:

[February 19, 2009 at 5:00 pm](#) | [Reply](#)

### 552. Equal slices and orthonormality

It seems to me that Ryan's way of looking at equal-slices density is too good to ignore, so instead of asking what the most natural orthonormal basis is with respect to equal-slices density, let us ask instead what the most natural thing to do is if (i) we want orthonormality and (ii) we want to think of equal-slices density as the average of the weighted densities on the cube.

As Gil points out, for each fixed  $p \in [0, 1]$  you have a natural orthonormal basis. Indeed, if we choose a random point of  $\{0, 1\}^n$  by letting the coordinates be independent Bernoulli random variables with probability  $p$  of being 1 and  $1 - p$  of being 0, then the function that is  $p - 1$  when  $x_1 = 1$  and  $p$  when  $x_1 = 0$  has mean 0. The expectation of the square of this function is  $p(1 - p)$ , so if we divide by  $p^{1/2}(1 - p)^{1/2}$  then we get a function  $w_{1,p}$  of mean 0 and variance 1. We can define  $w_{i,p}$  similarly for each  $i$ .

Now for a general set  $A$  define  $w_{A,p}$  to be  $\prod_{i \in A} w_{i,p}$ . Then the independence of the random variables  $w_{i,p}$  implies that  $\mathbb{E}_B w_{A,p}(B) = 0$  if  $A$  is non-empty (always assuming that the elements of  $B$  are chosen randomly and independently with probability  $p$ ). Therefore, the  $w_{A,p}$  are orthogonal. If  $p = 1/2$  then we get the usual Walsh functions.

I think that there is just one natural thing to do at this point. For each  $p \in [0, 1]$  let us write  $\mu_p f$  for the expectation of  $f(X_1, \dots, X_n)$ , where each  $X_i$  is Bernoulli with probability  $p$ . Then the equal-slices measure of  $f$  is  $\mathbb{E}_p \mu_p(f)$ .

How about inner products? Well,  $\mathbb{E}_x f(x)g(x) = \mathbb{E}_p \mu_p(fg)$ . If we want to understand this in a Fourier way, it seems crazy (with the benefit of hindsight) to try to find a single orthonormal basis that will do the job. Instead, for each  $p$  one should expand  $f$  and  $g$  in terms of the weighted Walsh basis ( $w_{A,p}$ ) and then take the average. For each individual  $p$  there is a nice Parseval identity, since the  $w_{A,p}$  are orthonormal. To spell it out, if we write  $\hat{f}_p(A)$  for  $\mu_p(f w_{A,p})$ , then  $\mathbb{E}_x f(x)g(x) = \mathbb{E}_p \sum_A \hat{f}_p(A) \hat{g}_p(A)$ .

This looks to me like a nice usable Parseval identity, even if it doesn't have the form I was originally looking for. So we're taking our Fourier coefficients to belong to  $C[0, 1]$  rather than to  $\mathbb{R}$ .

64.  [gowers](#) Says:

[February 19, 2009 at 5:48 pm](#) | [Reply](#)

### 553. Sperner obstructions.

The next thing I'd like is a good measure on pairs  $A \subset B$ . With the old way of thinking about equal-slices density this was easy: we took two initial segments of a randomly permuted  $[n]$ . But can we think of it in a Ryan-like way?

Here's a natural suggestion. First we randomly pick a pair  $(p, q)$  of real numbers with  $0 \leq p \leq q \leq 1$ . Next, we choose the set  $B$  by picking each element independently and randomly with probability  $q$ . And then we pick the set  $A \subset B$  by picking each element of  $B$  independently and randomly with probability  $p/q$ . Note that the elements of  $A$  end up being chosen independently with probability  $p$ , so the marginal distributions of  $A$  and  $B$  are sensible.

The next task is to try to sort out what we mean by  $\mathbb{E}_{A \subset B} f(A)g(B)$ . (We want to say that  $f$  is non-uniform if this expectation is not small.) Wait, that's not quite what I meant to say. It's clear what it means, since I've just specified the distribution on pairs  $(A, B)$ . What is not quite so clear is how to express it in terms of the vector-valued Fourier coefficients. I think I need a new comment for that.

65.  [gowers](#) Says:

[February 19, 2009 at 7:55 pm](#) | [Reply](#)



## 554. Sperner-related calculations.

Finally I've found something that I definitely couldn't do straight up. But with a bit of paper and a couple of false starts I think I've got it down, or at least have made some calculational progress in that direction. The aim is to express the quantity  $\mathbb{E}_{U \subset V} f(U) f(V)$  (defined in the previous comment—it is of course essential that we are talking about an equal-slices-related joint distribution here) in terms of the vector-valued Fourier coefficients of  $f$ . Also, I've changed  $A$  and  $B$  to  $U$  and  $V$  because I want to use  $A$  and  $B$  to label some Walsh functions.

For this calculation it will be more convenient to think in terms of Bernoulli random variables. Let  $Y_1, \dots, Y_n$  be Bernoulli with mean  $q$ , let  $Z_1, \dots, Z_n$  be Bernoulli with mean  $p/q$ , and let  $X_i = Y_i Z_i$ . All the  $Y_i$  and  $Z_i$  are of course independent, so the  $X_i$  are independent Bernoulli random variables with mean  $p$ . Later on we will average over  $0 \leq p \leq q \leq 1$  but for now let us regard them as fixed and see what happens. Later we will take  $U$  to be the set of  $i$  such that  $X_i = 1$  and  $V$  to be the set of  $i$  such that  $Y_i = 1$ . (What this will ultimately be doing is investigating the number of pairs  $U \subset V$  in  $\mathcal{A}$  with  $U$  of size roughly  $pn$  and  $V$  of size roughly  $qn$ .)

Now we can expand  $f$  in terms of the functions  $w_{A,p}$  and also in terms of the functions  $w_{A,q}$ . If we do so, then our main task is to try to understand the quantity  $\mathbb{E} w_{A,p}(X_1, \dots, X_n) w_{B,q}(Y_1, \dots, Y_n)$ , which we are rewriting as  $\mathbb{E} w_{A,p}(Y_1 Z_1, \dots, Y_n Z_n) w_{B,q}(Y_1, \dots, Y_n)$ .

Here we find that independence helps us a lot. Let us begin by conditioning on  $Y_1, \dots, Y_n$ . Since this fixes  $w_{B,q}(Y_1, \dots, Y_n)$  we are now interested in  $\mathbb{E} w_{A,p}(Y_1 Z_1, \dots, Y_n Z_n)$ . For every  $p$  let us write  $\phi_p$  for the function that takes 0 to  $-(1-p)^{-1/2} p^{1/2}$  and 1 to  $p^{-1/2} (1-p)^{1/2}$ . Then  $w_{i,p}(x) = \phi_p(x_i)$  and  $w_{A,p}(x) = \prod_{i \in A} \phi_p(x_i)$ . Therefore,  $\mathbb{E} w_{A,p}(Y_1 Z_1, \dots, Y_n Z_n)$  is equal to  $\mathbb{E} \prod_{i \in A} \phi_p(Y_i Z_i)$  which by independence is equal to  $\prod_{i \in A} \mathbb{E} \phi_p(Y_i Z_i)$ . I have to feed my son now so will continue with this calculation later.



66. [gowers](#) Says:

[February 19, 2009 at 8:30 pm](#) | [Reply](#)

## 555. Sperner-related calculations continued.

While walking to a local grocer's I realized that conditioning on the  $Y_i$  was not such a good idea. Instead, we should note that  $w_{A,p}(Y_1 Z_1, \dots, Y_n Z_n) w_{B,q}(Y_1, \dots, Y_n)$  splits up as a product of three products, which I will write separately. They are  $\prod_{i \in A \setminus B} \phi_p(Y_i Z_i)$ ,  $\prod_{i \in B \setminus A} \phi_q(Y_i)$  and  $\prod_{i \in A \cap B} \phi_p(Y_i Z_i) \phi_q(Y_i)$ . Therefore, because of independence, the expectation that we're interested in (or at least I am) also splits up as a product of three products, namely  $\prod_{i \in A \setminus B} \mathbb{E} \phi_p(Y_i Z_i)$ ,  $\prod_{i \in B \setminus A} \mathbb{E} \phi_q(Y_i)$  and  $\prod_{i \in A \cap B} \mathbb{E} \phi_p(Y_i Z_i) \mathbb{E} \phi_q(Y_i)$ . Thus our calculation becomes very simple, since each of these expectations in the products is a number that does not depend on  $i$ . Must go again.



67. [gowers](#) Says:

[February 19, 2009 at 9:59 pm](#) | [Reply](#)

## 556. Sperner-related calculations continued.

There are three very finite calculations to do next.

1. Since  $Y_i Z_i = X_i$  is Bernoulli with mean  $p$ ,  $\mathbb{E} \phi_p(Y_i Z_i) = 0$ , by the way we defined  $\phi_p$ . Wow, I wasn't expecting that.
2. Similarly,  $\mathbb{E} \phi_q(Y_i) = 0$ .
3. As for  $\mathbb{E} \phi_p(Y_i Z_i) \phi_q(Y_i)$ , it works out to be some number  $\lambda_{p,q}$ . I'll calculate it later as I have to go.

So the whole thing collapses to 0 unless  $A = B$  and then we get  $\lambda_{p,q}^{|A|}$ . That looks pretty nice to me.

68.  [gowers](#) Says:

[February 20, 2009 at 1:08 am](#) | [Reply](#)

557. Sperner-related calculations continued.

Right,  $(Y_i Z_i, Y_i)$  is  $(0, 0)$  with probability  $1 - q$ ,  $(0, 1)$  with probability  $q - p$  and  $(1, 1)$  with probability  $p$ . Recall that  $\phi_p(0) = (1 - p)^{1/2} p^{-1/2}$  and  $\phi_p(1) = -p^{1/2} (1 - p)^{-1/2}$ , and similarly for  $\phi_q$ . (I've changed these to minus what they were before, but this isn't important.) To tidy things up a bit I'm going to make the substitution  $p = \sin^2 \alpha$  and  $q = \sin^2 \beta$ . So  $\phi_p(0) = \cot \alpha$ ,  $\phi_p(1) = \tan \alpha$ , and similarly for  $q$  and  $\beta$ .

Therefore,  $\phi_p(Y_i Z_i) \phi_q(Y_i)$  is  $\cot \alpha \cot \beta$  with probability  $\cos^2 \beta$ ,  $\cot \alpha \tan \beta$  with probability  $\sin^2 \beta - \sin^2 \alpha$  and  $\tan \alpha \tan \beta$  with probability  $\sin^2 \alpha$ . Therefore, the expectation is  $\cos \alpha \cos^3 \beta / \sin \alpha \sin \beta + \dots$  hmm, I've just tried it on a piece of paper and it doesn't seem to simplify. That's annoying, as I would very much like to get some idea of the average size of  $\lambda_{p,q}^{|A|}$  as  $p$  and  $q$  vary.

69. [A quick review of the polymath project « What Is Research?](#) Says:

[February 20, 2009 at 1:16 am](#) | [Reply](#)

[...] few days later, Gowers announced at the end of this post that there was a wiki on the enterprise of solving the density Hales-Jewett theorem. In the same [...]

70.  [gowers](#) Says:

[February 20, 2009 at 1:43 am](#) | [Reply](#)

558. Sperner-related calculations continued.

Actually, I forgot the minus signs:  $\phi_p(1)$  should have been  $-\tan \alpha$ . So we get  $\cot \alpha \cot \beta \cos^2 \beta - \cot \alpha \tan \beta (\sin^2 \beta - \sin^2 \alpha) + \tan \alpha \tan \beta \sin^2 \alpha$ .

I'm going to have to do this a few times until I'm confident it's right. But at the moment it looks as though  $\lambda_{p,q}$  is sometimes bigger than 1, which I don't like, but also as though there is a range where it is less than 1, which might mean that we were OK after some very modest localization. The reason I'd like  $\lambda_{p,q} < 1$  is that it would then mean that the contribution to the sum from large  $A$  was fairly negligible, and I think we'd have a proof of at least some statement along the lines of incorrect number of  $U \subset V$  implying large Fourier coefficient at some small  $A$ . And if that worked, it would be hard to resist the temptation to try something similar for DHJ(3), though the thought of it is rather daunting.


71.  [jozsef](#) Says:

[February 20, 2009 at 3:28 am](#) | [Reply](#)

559. Removal lemma for DHJ

The main difficulty - at least for me - in applying a removal lemma to the set model (DHJ Version 4. in the Wiki) is the disjointness of sets as it makes the corresponding graph very sparse. Here I'll try to relax the disjointness condition in a slightly different model. Instead of working inside of one dense cube showing that it contains a line, here we will consider now several cubes; The "original" cube  $[3]^n$ ,  $n$  copies of  $[3]^{n-1}$  ...  $\binom{n}{k}$  copies of  $[3]^{n-k}$ , and so on. For any subset of  $[n]$  we have one cube. We can suppose that each cube is at least  $c$ -dense and line-free. For any pair of subsets of  $[n]$ , say  $A$  and  $B$  there is a unique point assigned; it is in the cube over  $[n](A \cap B)$  with 1-s in the  $A - B$  positions, 2-s in the  $B - A$  positions, and 0-s elsewhere. Every point represents a pair of subsets in  $[n]$ . This model allows us to state the following "corner-like" form of DHJ:

For every  $\delta > 0$  there exists  $n$  such that every collection of pairs  $(A, B)$  of subsets of  $[n]$  of cardinality at least  $\delta 4^n$  contains a "corner"  $(A, B), (A \cup D, B), (A, B \cup D)$ , where  $D$  is disjoint from  $A$  and  $B$ . ( $A, B, D$  are subsets of  $[n]$ )


72.  *Ryan O'Donnell* Says:  
[February 20, 2009 at 6:56 am](#) | [Reply](#)

560. Sperner.

Tim, regarding your problem beginning in #553 and fixing  $p$  and  $q$  first... Is it correct that if we fix  $p = 1/2 - \epsilon$  and  $q = 1/2 + \epsilon$ , then we recover the scenario from #476?


One reason I like to have  $p$  very close to  $q$  ("extreme localization") is that the  $p$ -biased and  $q$ -biased measures are practically disjoint if  $p$  is even a little bit less than  $q$ . So then how a set  $\mathcal{A}$  acts under the  $p$ -biased distribution has nothing to do with how it acts under the  $q$ -biased distribution.

(Of course, this is only if you \*fix\* distant  $p, q$  in advance. If you pick them at random, some bets are off.)

73.  *Ryan O'Donnell* Says:  
[February 20, 2009 at 7:00 am](#) | [Reply](#)


561. Moser.

Given that we feel a bit stuck on HJ, I feel tempted to think more about Moser instead. I mean, why not? — it's a strictly easier problem. Is there an "easier" ergodic-theory proof of Density-Moser?


74.  *jozsef* Says:  
[February 20, 2009 at 7:11 am](#) | [Reply](#)

560. Removal lemma for DHJ (contd.)

Now that a dense system is given, let's check how can we use a removal lemma. Given a collection of pairs of subsets of  $[n]$ , denoted by  $S$ , of cardinality  $\delta 4^n$ . The family of pairs having the same symmetric difference set form a matching in this graph. By Ruzsa-Szemerédi there will be many edges from the same matching (symmetric difference) that are connected by another edge. More precisely, there will be at least  $\epsilon 8^n$  quadruples  $A, B, C, D$  that  $(A, B)$ ,  $(C, D)$ , and  $(A, D)$  are from  $S$  and  $A \Delta B = C \Delta D$ . Unfortunately it is not a combinatorial line in general. However it is a combinatorial line if  $A \cap B = C \cup D = A \cup D$ , i.e. when the three pairs are representing points from the same cube.


75.  *jozsef* Says:  
[February 20, 2009 at 7:18 am](#) | [Reply](#)

oops, here is a typo in the previous post; the correct statement is  
 $A \cap B = C \cap D = A \cap D$

76.  *jozsef* Says:  
[February 20, 2009 at 7:32 am](#) | [Reply](#)

563. Moser

Re:561. Ryan, I don't think that there is a published ergodic proof for Moser's problem and I think that Moser  $k=6$  implies DHJ  $k=3$ . But I agree, there should be a simpler way to prove Moser  $k=3$ . Note that for any algebraic line  $L$  in  $(\mathbb{Z}/3\mathbb{Z})^n$  one can find an  $x$  such that  $x+L$  is a geometric line.

77.  *Gil* Says:  
[February 20, 2009 at 7:44 am](#) | [Reply](#)

Dear Jozsef, how do we get DHJ for  $k=3$  from Moser for  $k=6$ ? (Exploring these type of reductions may also be a nice avenue of some potential.)

78.  [josef](#) Says:

[February 20, 2009 at 8:18 am](#) | [Reply](#)

Dear Gil, We have to check, but my thought was that if you map  $3^n$  to  $6^n$  as 1 maps to 1 or 6, 2 maps to 2 or 5, and 3 maps to 3 or 4, then if DHJ  $k=3$  was combinatorial line free then Moser  $k=6$  is geometric line free. And if DHJ was dense, then Moser is dense too. I hope it's right.

79.  [gowers](#) Says:

[February 20, 2009 at 9:08 am](#) | [Reply](#)

566. Equal-slices measure.

Ryan, my answer to your point in 560 is that it is indeed true that if you fix  $p$  and  $q$  that are some way apart, then the measures  $\mu_p$  and  $\mu_q$  are approximately disjointly supported. However, this is quite natural once you average over  $p$  and  $q$ . It's a bit like saying that in order to understand Sperner we will look at pairs of layers and see what happens for those. Obviously, one cannot look at each pair in complete isolation if one wants to prove Sperner's theorem (since the weight inside each layer might be less than  $1/2$ , say) but it's not so clear that one can't get useful information about obstructions that way: for example, one might be able to understand when two layers don't make the right contribution to the number of subset pairs, and one might then be able to put that together to understand when it happens after averaging over all pairs of layers.

But whether or not you are convinced by this, your alternative way of thinking about equal-slices measure has been hugely useful!

80.  [gowers](#) Says:

[February 20, 2009 at 9:22 am](#) | [Reply](#)

567. Sperner calculations continued.

A quick remark re 558. I realized as I woke up this morning that  $\lambda_{p,q}$  must be at most 1 by Cauchy-Schwarz because it's the covariance of the random variables  $\phi_p(Y_i Z_i)$  and  $\phi_q(Y_i)$ , each of which has been carefully designed to have variance 1. (To put it another way, it follows easily from Cauchy-Schwarz.) So I think we're in business here. I made some small slips in the calculations earlier, but I *think* that we have the very nice expression

$\mathbb{E}_{p \leq q} \sum_{A \subset [n]} \lambda_{p,q}^{[A]} \hat{f}_p(A) \hat{f}_q(A)$  for the total weight  $\mathbb{E}_{U \subset V} f(U) f(V)$ , where each  $\lambda_{p,q}$  has modulus less than 1, with equality if and only if  $p = q$ . From this one can informally read off a few quite interesting things. For instance, the contribution from a set  $A$  of large cardinality will be tiny if  $p$  and  $q$  are not very close to each other, but that happens only rarely (Ryan won't like this bit), so the total contribution from large  $A$  appears to be small. This fits in nicely with one's intuition that by Kruskal-Katona the upper shadow of a layer of positive density will be huge if you go up a lot of layers, unless there is some simple explanation such as all sets containing 1. Perhaps at this point I'll try to tidy up these calculations and put them on the wiki.

81.  [gowers](#) Says:

[February 20, 2009 at 12:17 pm](#) | [Reply](#)

568. Fourier and DHJ(3).

Before I go off and do actual calculations, I want to see how much I can easily generalize of the Sperner calculations to DHJ(3).

The first step is to come up with a Ryan-style model for equal-slices measure. For points it is easy enough: first, pick  $p + q + r = 1$  and for each coordinate  $x_i$  let it be 1 with probability  $p$ , 2 with probability  $q$  and 3 with probability  $r$ . Then average the resulting density over all such  $p, q, r$ .

Now for  $p + q + r + s = 1$  I'd like to pick a random combinatorial line in such a way that its 1-set is chosen according to  $(p, q, r)$  measure, its 2-set is chosen according to  $(p, q, s, r)$  measure, and its 3-set is chosen according to  $(p, q, r, s)$  measure. I think this too is easy: for each  $i$  you make it 1 with probability  $p$ , 2 with probability  $q$ , 3 with probability  $r$ , and a wildcard with probability  $s$ .

The next task is to come up with analogues of the “biased Walsh functions”  $w_{A,p}$  that we used on  $[2]^n$ . So now we are looking for functions  $\tau_{A,p,q,r}$  (where the  $r$  is redundant but there for symmetry). We very much want to use independence, so we would like  $\tau_{A,p,q,r}$  to be  $\prod_{i \in A} \tau_{i,p,q,r}$ , where  $\tau_{i,p,q,r}(x)$  will be  $\phi_{p,q,r}(x_i)$ . Amongst the properties we will want of the function  $\phi_{p,q,r}$  is that its expectation (defined to be  $p\phi_{p,q,r}(1) + q\phi_{p,q,r}(2) + r\phi_{p,q,r}(3)$ ) is 0 and that its variance is 1. To achieve this, I plan to let  $\phi_{p,q,r}(1) = cp^{-1}\omega$ ,  $\phi_{p,q,r}(2) = cq^{-1}\omega^2$  and  $\phi_{p,q,r}(3) = cr^{-1}$ . Here  $\omega = \exp(2\pi i/3)$  and  $c$  is a normalizing factor to make the variance 1. This seems not to be the unique thing one could do but it does seem to be quite a natural thing to do. I'm hoping, however, that the expectation and variance will give me most of the information I care about.

In particular, the functions  $\tau_{A,p,q,r}$  defined this way do indeed form an orthonormal basis for (complex) functions defined on  $[3]^n$ .

Now let us think about the expression  $\mathbb{E}f(x)g(y)h(z)$ , where the expectation is over all combinatorial lines  $(x, y, z)$  chosen according to the  $(p, q, r, s)$  measure. To do this, we expand  $f$  as  $\sum_A \hat{f}_{p,q,r,s}(A) \tau_{A,p,q,r,s}$ , and similarly for  $g$  and  $h$ . So what we would like to understand is  $\mathbb{E} \tau_{A,p,q,r,s}(x) \tau_{B,p,q,r,s}(y) \tau_{C,p,q,r,s}(z)$  for fixed sets  $A, B, C$ . Again, the expectation is over all  $(p, q, r, s)$ -combinatorial lines. (Eventually we will average over  $p+q+r+s=1$ .)

I want to be able to see this expression, so will start a new comment.

82.  [gowers](#) Says:

[February 20, 2009 at 12:41 pm](#) | [Reply](#)

569. Fourier and DHJ(3).

Now let us choose a sequence of random variables  $U_1, \dots, U_n$ , where each  $U_i$  is 1 with probability  $p$ , 2 with probability  $q$ , 3 with probability  $r$ , and a wildcard with probability  $s$ . Then let  $X_i$  be  $U_i$  if  $U_i = 1, 2$  or  $3$ , and 1 if  $U_i$  is a wildcard, and define  $Y_i$  and  $Z_i$  similarly, so that the three sequences  $(X_1, \dots, X_n)$ ,  $(Y_1, \dots, Y_n)$  and  $(Z_1, \dots, Z_n)$  form a random combinatorial line  $(x, y, z)$  determined by  $(U_1, \dots, U_n)$ .

Now each term in the product  $\mathbb{E} \tau_{A,p,q,r,s}(x) \tau_{B,p,q,r,s}(y) \tau_{C,p,q,r,s}(z)$  itself splits up into a big product, the first over all  $i \in A$ , the second over all  $i \in B$  and the third over all  $i \in C$ . For example,

$$\tau_{A,p,q,r,s}(X_1, \dots, X_n) = \prod_{i \in A} \phi_{p,q,r}(X_i).$$

Let us try to understand when the expectation of this product has a chance of not being zero.

One case where it *is* zero is if there exists  $i$  that belongs to  $A$  but not to  $B$  or  $C$ . In that case, the big product will include  $\phi_{p,q,r}(X_i)$ , but it won't involve  $Y_i$  or  $Z_i$ . Therefore, by independence and the fact that  $\mathbb{E} \phi_{p,q,r}(X_i) = 0$ , the entire expectation is 0.

What about if  $i \in A \cap B$  but  $i \notin C$ ? Now we want to know what  $\mathbb{E} \phi_{p,q,r}(X_i) \phi_{p,q,r}(Y_i)$  is. We know that the probability that  $X_i = Y_i = 1, 2, 3$  is  $p, q, r$ , and the probability that  $X_i = 1$  and  $Y_i = 2$  is  $s$ . I would very much like the resulting expectation to be 0, but at the moment I don't see any reason for this. So I think I've got to the point where I need to do a back-of-envelope calculation. Here is where it may turn out that my guess for a good choice of functions  $\phi_{p,q,r}$  was not a good guess. (I should say that what I'm eventually hoping for is something reminiscent of

what one gets in  $\mathbb{Z}_3^n$ , namely a sum over cubes of Fourier coefficients. This may be asking a bit much though.)

83.  [gowers](#) Says:

[February 20, 2009 at 1:55 pm](#) | [Reply](#)

570. Fourier and DHJ(3)

A quick report back. The guess doesn't work, but I'm still not quite sure whether there might be another choice of functions  $\phi_{p,q,r}$  that do the job. If we forget normalizations then the only property that matters is  $p\phi_{p,q,r}(1) + q\phi_{p,q,r}(2) + r\phi_{p,q,r}(3) = 0$ . And then what we'd very much like is that if  $\phi_{p+s,q,r}(i) = z_i$  and  $\phi_{p,q,s,r}(i) = w_i$ , then  $pz_1w_1 + qz_2w_2 + rz_3w_3 + sz_1w_2 = 0$ . But I think that's not going to be achievable. If we set  $s = 0$ , then it suggests that we should also have  $p\phi_{p,q,r}(1)^2 + q\phi_{p,q,r}(2)^2 + r\phi_{p,q,r}(3)^2 = 0$ , which more or less pins down the function (I think it decides it up to a constant multiple and taking complex conjugates). But I'm getting a bit bogged down in algebra at the moment, nice though that condition is.

84.  [Ryan O'Donnell](#) Says:

[February 20, 2009 at 4:03 pm](#) | [Reply](#)

571. Sperner.

Tim, the expression in #567 is so elegant, how can I not be on board for equal-slices now? 😊 Here are some questions that occur to me:

1. Can we get a handle on the meaning of the first term? What does  $\mathbb{E}_{p \leq q}[\hat{f}_p(\emptyset)\hat{f}_q(\emptyset)]$  represent? Well, it represents the average of  $f$ 's  $p$ -biased measure times its  $q$ -biased measure under this way of choosing  $p, q$ , but what does that... mean?
2. Can we understand why this expression is miraculously always  $\delta$  whenever  $f$  is a monotone function of equal-slices-mean  $\delta$ ? Being "monotone" is a fairly "non-Fourier-y" property.
3. Is this quantity always at least  $\delta^2$ ?
4. What exactly might we be hoping to prove in this Sperner case? Something like, "The extent to which you don't have a  $\delta$  fraction of Sperner pairs (under Tim's distribution) is related to the extent to which you are not monotone."

85.  [gowers](#) Says:

[February 20, 2009 at 4:46 pm](#) | [Reply](#)

572. Sperner

Ryan, I don't have complete answers to your questions, needless to say, but they are interesting questions so let me give my first thoughts.

1. I haven't checked that it's exactly the same, but I'd have thought it would be something like what I was looking at before you came along with your  $p$ s and  $q$ s. That is, you randomly scramble  $[n]$ , then randomly choose two initial segments  $U$  and  $V$ , and calculate  $f(U)f(V)$ . And then you average over everything you've done. In the case where  $f$  is the characteristic function of a set  $\mathcal{A}$ , you are calculating the expectation of the square of the density of initial segments of a random permutation of  $[n]$  that belong to  $\mathcal{A}$ . So in that model at any rate it's always at least  $\delta^2$  since the average is  $\delta$ . So this may possibly answer 3.

2. I don't understand your question here. For example, if  $f$  is the characteristic function of all sets of size at least  $n/2$ , then the  $p$ -biased measure is basically 0 if  $p < 1/2$ . So the equal-slices mean is  $1/2$ . And if you choose a random pair



$(p, q)$  such that  $0 \leq p \leq q \leq 1$  then the probability that they both exceed  $1/2$  is  $1/4$ . And this works if you replace  $1/2$  by any  $\delta$ . In order to depart from  $\delta^2$ , you have to find a monotone function that doesn't look like a union of layers, of which the obvious example is the set of all sets that contain 1.

3. See 1.

4. At one point I was convinced by you that monotonicity was a key property. But I've gone back to thinking that what really matters is small sets with big bias. I think what I'd like to see (and I think it could be quite close) is a density-increment proof of Sperner. The proof would go something like this. By the expansion above, either get the expected number of pairs  $U \subset V$  or we find a small  $A$  such that  $\hat{f}_p(A)$  is large for a dense set of  $p$ , where  $f$  is the balanced function of  $A$  (suitably defined). In the latter case, we fix a small number of variables and give ourselves a density increase in a subspace of bounded codimension. The motivation for such an argument would be to find something that could be adapted to DHJ(3). There are at least two technical problems to sort out. The first is to work out  $\mathbb{E}_{p \leq q} \lambda_{p,q}^k$ . The second is that restricting to subspaces is problematic when one is dealing with equal-slices measure. Or at least, at some point I persuaded myself that it was.

86.  gowers Says:

[February 20, 2009 at 6:14 pm](#) | [Reply](#)

573. Fourier, DHJ(3), corners

It's just occurred to me that what I was trying to do in 568-570 is unlikely to work because we don't have a direct Fourier expansion of this kind for the corners problem. So before I go any further I want to think about corners.

Actually, I've now had a counterthought, which is that if by some miracle we ended up with an expression of the form  $\mathbb{E}_{p+q+r=1} \sum \lambda_{p,q,r,s}^{[u]} \hat{f}_{p+s,q,r}(u) \hat{f}_{p,q+s,r}(u) \hat{f}_{p,q,r+s}(u)$  for the number of combinatorial lines (I've also just realized that I was absent-mindedly trying to define a basis that consisted of only  $2^n$  vectors, so I've changed  $A$  to  $u$  and will go back and think about this), then we would not particularly expect to get a non-zero answer (if  $f$  was the characteristic function of our dense set) unless we could find a good selection  $p, q, r, s$  such that all the Fourier coefficients at  $u = 0$  were reasonably large. This would be asking for corners in the triangle  $p + q + r = 1$ . So it could conceivably be that the corners theorem would get us started and Fourier would take over from there—which is reminiscent of, though not quite the same as, what Terry suggested way back in [comment 77](#). Here, the idea would be that the corners theorem (suitably Varnavides-ized) would guarantee that there were many candidate triples of rich slices that could lead to combinatorial lines, and Fourier methods would show that you could differ from the expected number of combinatorial lines *given the slice-density function* only if something showed up in the Fourier expansions.

However, having written that optimistic sounding paragraph, I feel pessimism returning. In fact, I have a little test to apply. I think if one chooses a random set of complexity 1 (see [the discussion of DHJ\(1,3\) on this page](#) for a definition) then it may well not have any large Fourier coefficients, even though it will typically have the wrong number of combinatorial lines.

To end this comment on an optimistic note, perhaps this strange form of Fourier analysis could be used to prove DHJ(1,3), and perhaps it could come in at the point where Shkredov needs it for his corners proof.

87.  gowers Says:

[February 20, 2009 at 6:57 pm](#) | [Reply](#)

574. Fourier and DHJ(3)

Here's the test then. Suppose you take two random collections  $\mathcal{U}$  and  $\mathcal{V}$  of subsets of  $[n]$ . Now let  $\mathcal{A}$  be the set of all points  $x$  with 1-set in  $\mathcal{U}$  and 2-set in  $\mathcal{V}$ . Let  $\phi$  and  $\psi$  be two functions from  $[3]$  to  $\mathbb{C}$  such that  $\phi(1) + \phi(2) + \phi(3) = \psi(1) + \psi(2) = \psi(3) = 0$ . Let  $\tau$  be a product of functions  $x \mapsto \sigma_i(x_i)$  where each  $\sigma_i$  is either

1,  $\phi$  or  $\psi$ . The question now is whether  $\tau$  can possibly correlate with  $\mathcal{A}$ . And the answer seems to me to be so obviously no that I think I can get away without actually checking it carefully.

In fact, here's a sketch of a proof: there are only  $3^n - 1$  non-constant functions of the given type, the expectation of each one over the random set  $\mathcal{A}$  is 0, and there will be very very strong concentration. (I haven't checked this argument, but it convinces me.)

The next step is to see whether it is possible to make up for this slight disappointment (slight because we had no right to expect things to work better for DHJ(3) than they do for corners) by finding something comparably nice about bipartite graphs and singular values, for possible use in a regularity lemma. I'll start a fresh comment and then explain what I'm hoping for here.

88.  [gowers](#) Says:

[February 20, 2009 at 7:24 pm](#) | [Reply](#)

575. Global quasirandomness

Recall that associated with a subset  $\mathcal{A}$  of  $[3]^n$  are three bipartite graphs, each with two copies of the power set of  $[n]$  as its vertex sets, with edges joining some pairs of disjoint sets. For instance, the 12-graph joins  $U$  to  $V$  if the sequence  $x$  with 1-set equal to  $U$  and 2-set equal to  $V$  belongs to  $\mathcal{A}$ .

Now the averaging-over-permutations technique looked as though it would yield a definition of quasirandomness that was sufficient to guarantee the correct number of combinatorial lines. However, we did not get out of that the kind of global correlations that would be needed for a regularity lemma. Basically, the reason was that the definition that seemed to be emerging was of the number-of-4-cycles kind rather than the correlates-with-bipartite-subgraph kind. However, the thoughts that led to the Sperner expression feel as though they could also give rise to a useful definition of the latter kind. So that is what I want to try next.

89.  [gowers](#) Says:

[February 20, 2009 at 7:58 pm](#) | [Reply](#)

576. Global quasirandomness

First, let me give a Ryan-style equal-slices measure for the bipartite graph we're talking about, or more generally for functions  $f$  that take pairs  $(U, V)$  of disjoint sets to  $[-1, 1]$ . As usual, I'll discuss things for fixed  $p$  and  $q$  and then average later. So let  $p$  and  $q$  be positive real numbers that add up to less than 1. For each element of  $[n]$ , put it in  $U$  with probability  $p$  and in  $V$  with probability  $q$ , with all choices independent. Then we would like to find a natural decomposition of such functions, analogous to the singular value decomposition of a bipartite graph.

In the graphs case we can derive the SVD variationally. The analogue here would be to maximize  $\mathbb{E}_{U,V} f(U, V) a(U) b(V)$  over all functions  $a$  and  $b$  of  $L_2$  norm 1. Here, I imagine that we choose  $U$  and  $V$  according to the distribution specified above, and measure the  $L_2$  norms of  $a$  and  $b$  using the  $\mu_p$  and  $\mu_q$  measures, respectively. Then if  $d$  is a very small function orthogonal to  $b$ , then the above expectation must change by  $o(\|d\|)$ , which implies that the function  $c(V) = \mathbb{E}_{U,V} f(U, V) a(U)$  is proportional to  $b(V)$ .

Got to go. To be continued.

90.  [gowers](#) Says:

[February 21, 2009 at 12:26 am](#) | [Reply](#)

577. Global quasirandomness

My basic thought here is to try to do linear algebra but with a different model of matrix multiplication: when we write  $\mathbb{E}_{U,V} f(U, V) a(U)$  we are not allowing  $U$  and  $V$  to vary independently. Having decided on that, I would like to try to follow the usual arguments as similarly as possible. For example, I would like to be able to say that  $f$  is quasirandom if  $\mathbb{E}_{U,V} f(U, V) a(U)$  is approximately the constant function  $\mu_p(a)$  for every bounded  $a$ . And I'd like to have a counting lemma for quasirandom functions (this is likely to be less easy). With those two ingredients, there would be a serious chance of proving a triangle-removal lemma.

From the argument in the previous comment, we have seen that if  $\mathbb{E}_{U,V} f(U, V) a(U) b(V)$  is maximized over all unit vectors  $a$  and  $b$ , then  $\mathbb{E}_{U,V} f(U, V) a(U)$  is proportional to  $b(V)$ . Therefore, if  $\langle b, b' \rangle_q = 0$ , it follows that  $\mathbb{E}_{U,V} f(U, V) a(U) b'(V) = 0$ . By symmetry, if  $\langle a, a' \rangle_p = 0$ , it follows that  $\mathbb{E}_{U,V} f(U, V) a'(U) b(V) = 0$ . In other words, the linear map defined by  $f$  takes the space orthogonal to  $a$  to the space orthogonal to  $b$ , so we can find orthonormal bases  $(a_i)$  and  $(b_i)$  of  $[2]^n$ , one with respect to the measure  $\mu_p$  and one with respect to the measure  $\mu_q$ , such that for every  $i$ ,  $\mathbb{E}_{U,V} f(U, V) a_i(U)$  is proportional to  $b_i(V)$ .

Let's write  $\mathbb{E}_{U,V} f(U, V) a_i(U) = \lambda_i b_i(V)$ . (Everything here depends on  $p$  and  $q$  but we are thinking of those as fixed for now.) Then if  $a = \sum_i \theta_i a_i$ , then  $\mathbb{E}_{U,V} f(U, V) a(U) = \sum_i \lambda_i \theta_i b_i(V)$ , which has  $L_2$  norm  $\sum_i \lambda_i^2 \theta_i^2$ . For this not to be small, we need some  $\lambda_i$  to be large, I think, and this seems to give us  $a_i$  and  $b_i$  such that  $\mathbb{E}_{U,V} f(U, V) a_i(U) b_i(V)$  is large. Obviously this needs to be checked pretty carefully, but it looks on the face of it as though we find something like a complexity-1 function that correlates with  $f$ . (However, when  $p$  and  $q$  vary, so does the complexity-1 function.)

But I still have the local-to-global problem, but from the other side: this time there seems to be a global non-quasirandomness, but it's not clear to me that global quasirandomness leads to the right number of triangles.



91. [Terence Tao](#) Says:

[February 21, 2009 at 6:47 am](#) | [Reply](#)

578. A new proof of DHJ(2)

Hi; I've been busy with other things for a while, so am not caught up... but meanwhile, I think I found a genuinely different proof of DHJ(2) based on obstructions-to-uniformity (similar, incidentally, to my paper proving convergence for multiple commuting shifts) which may have some potential.

It goes like this. Call a bounded function  $f : [2]^n \rightarrow \mathbb{R}$  *uniform at scale m* if  $f(x)$  and  $f(y)$  are uncorrelated, whenever  $x$  is chosen randomly from  $[2]^n$  and  $y$  is formed from  $x$  by randomly flipping  $O(m)$  0s to 1s.

Call a bounded function  $f$  *basic anti-uniform at scale m* if its mean influence is  $O(1/m)$ , i.e. if  $x$  is chosen randomly from  $[2]^n$  and  $y$  is obtained from  $x$  by randomly flipping one 0 to a 1, then  $f(x) = f(y)$  with probability  $1 - O(1/m)$ .

**Lemma:** A bounded function  $f$  is either uniform at scale  $m$ , or has non-trivial correlation with a function  $g$  which is basic anti-uniform at scale  $m$ .

**Proof:** Take  $g(x) := \mathbb{E} f(y)$ , where  $y$  is formed from  $x$  by randomly flipping  $m$  0s to 1s.  $\square$

At this point we have two possible strategies.

The first is to observe that basic anti-uniform functions tend to be roughly constant on large subspaces. Because of this, any function  $f$  of mean zero which correlates with a basic anti-uniform function will have a density increment on some large subspace. If one then starts with  $f := 1_A - \delta$  and runs the density increment argument (and reduce  $m$  at each iteration), one can eventually get to the point where  $1_A - \delta$  is uniform at some scale  $m$ , at which point one can find lots of combinatorial lines with  $m$  wildcards.


The other strategy is to use energy-increment methods instead of density-increment ones. Indeed if we do the usual "energy-increment" iteration (as for instance appears in my first paper with Ben), and decreasing the parameter  $m$

appropriately with each iteration, we obtain

**Corollary:** Every bounded function  $f$  can be split into a function  $f_U$  uniform at scale  $m$ , plus a polynomial combination  $f_{U^\perp}$  of functions anti-uniform at scale  $m'$ , plus a small error, where  $m'$  is much larger than  $m$ . Also, if  $f$  is non-negative and has large mean, one can arrange matters so that  $f_{U^\perp}$  is also non-negative and has large mean.


Now on subspaces of size about  $m'$ ,  $f_{U^\perp}$  is close to constant, and on many subspaces, this quantity is going to be large. Meanwhile,  $f_U$  is uniform on most subspaces, and so most subspaces are going to contain lines (with about  $m$  wildcards).

The corresponding obstructions to uniformity for DHJ(3) are more complex. Roughly speaking, I'm facing things like this: take a random  $x$  in  $[3]^n$ , let  $y$  be formed from  $x$  by randomly flipping a bunch of 0s to 1s, and then let  $x'$ ,  $y'$  be formed from  $x$  and  $y$  by randomly flipping a bunch of 0s to 2s (with the digits being flipped for  $x$  being the same as those for  $y$ ). Then one needs to understand what it means if  $\mathbb{E}f(x)f(y)f(x')f(y')$  is large. There seems to be a sense that  $f$  is somehow correlating with a "local complexity 1" set - i.e. the type of expression we already saw in DHJ(1,3), but somehow localised to a much smaller scale than  $n$  - but I don't see clearly what it will be yet.

92.  [Terence Tao](#) Says:  
[February 21, 2009 at 5:36 pm](#) | [Reply](#)

579.

There looks like there is a good chance that the type of arguments in 578 can be combined with the basic strategy I sketched out (very vaguely) in 536 (as Ryan already suggested in 538). Basically, if there are many large slices of  $A$  which are not uniform (which means that their Fourier transform is concentrated in a Hamming ball of radius  $n/m$  or so centred at the origin) then that slice of  $A$  is correlating with an anti-uniform object of low influence; this object is roughly constant on medium-dimensional subslices and so we can locate a density increment. So we can assume that most slices of  $A$  are uniform, and I think this is getting close to what we need for a counting lemma for geometric lines in  $A$  where the middle point of the line is close to the centre  $(2,2,...,2)$ . I'll have to do some calculations to clarify all this.

93.  [gowers](#) Says:  
[February 21, 2009 at 5:40 pm](#) | [Reply](#)

580. Extreme localization and DHJ(3)

Although I am still very interested in trying to find a "global" proof of DHJ(3) (though cheating first and changing the measure to equal-slices), Terry's 578 prompts me to ask why we don't already have a proof of DHJ(3). I'm not saying that I think we do, but rather that I want to understand where the gap is.

In order to explain that, let me try to give some sort of proof sketch. Either it will look fairly complete, in which case we will need to examine its steps more carefully, or there will be some step that I don't yet see how to carry out, even roughly.

The broad plan is similar to things that both Terry and I have discussed in much earlier comments, and that Terry alludes to in his final paragraph above. I shall write my points as  $(U, V, W)$ , where these are the 1-set, 2-set and 3-set of a sequence in  $[3]^n$ . If  $\mathcal{A}$  is dense, then there will be many choices of  $W$  for which the set of  $(U, V)$  with  $U \cup V = [n] \setminus W$  and  $(U, V, W) \in \mathcal{A}$  is dense (in the set of  $(U, V)$  such that  $U \cup V = [n] \setminus W$ ). Therefore, by Sperner we can find a pretty large number of pairs  $(U, V), (U', V')$  of such pairs with  $U \subset U'$ . And for each such pair of pairs we know that the point  $(U, V', [n] \setminus (U \cup V'))$  does not belong to  $\mathcal{A}$ .

Now let's rephrase that. This time I'll write  $(U, V)$  for the point with 1-set equal to  $U$  and 2-set equal to  $V$ . Then for

each  $W$  we find a set-system  $\mathcal{U}_W$  such that  $(U, V) \notin \mathcal{A}$  whenever  $U \in \mathcal{U}_W$  and  $[n] \setminus (V \cup W) \in \mathcal{U}_W$ . Let us write  $\mathcal{V}_W$  for the set of all  $V$  with the second property. Sperner should tell us that on average the set of all disjoint pairs  $(U, V) \in \mathcal{U}_W \times \mathcal{V}_W$  is reasonably dense in the set of all disjoint pairs that are disjoint from  $W$ . So we have something that could be described as a “local complexity 1” set of points that do *not* belong to  $\mathcal{A}$ . Indeed, we have lots of such sets.


I’m going to gloss over a few technicalities such as whether we are thinking about equal-slices measure or insisting that our lines have very few wildcards or whatever. Or perhaps I’ll just go for the latter: let’s focus on lines that have at most  $m$  wildcards, where  $1 \ll m \ll n$ .

Now for the main question. Can we convert this local non-uniformity into a global non-uniformity by some random restriction? At the moment things look a bit strange because the lines we are considering are of the form  $(U, V)$ , where  $U$  and  $V$  are disjoint from each other and from a fixed set  $W$ , and all of  $U$ ,  $V$  and  $W$  have size about  $n/3$ . But if we restrict to the subspace of all sequences that are 3 on  $W$ , then in this subspace we are looking at points where the 1-set and 2-set use up about half the coordinates and the 3-set uses up only a tiny number of coordinates. So the plan would be to choose random sets  $U_0$  and  $V_0$  of size something like  $n/2 - m$  and restrict to the subspace of points that are 1 on  $U_0$  and 2 on  $V_0$ . Inside that subspace I think we’d still have correlation with a set of complexity 1 but now a typical point would have around  $m$  each of 1s, 2s and 3s (over and above the already fixed coordinates).

Obviously there would be some details to check there, but at the moment I don’t see why it doesn’t give us correlation with a set of complexity 1 after restricting to a small (but not all that small) subspace.


Where to proceed from there? Well, the set of complexity 1 is of a nice form in that whether or not you belong to it depends just on whether your 1-set belongs to  $\mathcal{U}_W$  and your 2-set belongs to  $\mathcal{V}_W$ . (In other words, there isn’t an additional condition about the 3-set. As shown on [section 2 on this wiki page](#), such sets are simpler to deal with than general complexity-1 sets.

I will be interested to see whether anyone can spot a flaw in the above sketched argument. If not, then attention should turn to a problem that up to now I have assumed is doable, which is to show that if you correlate with a complexity-1 set then you can get a density increase on a subspace. I’ll post this comment and then think about that problem in my next one.

94.  [Terence Tao](#) Says:  
[February 21, 2009 at 5:42 pm](#) | [Reply](#)

Metacomment:

Unrelated to this thread, but I thought I would like to report a success of this collaborative environment: at the 700 thread we have now produced two proofs (one computer-assisted, one not) that  $c_6 = 450$ , i.e. that the largest line-free subset of  $[3]^6$  contains exactly 450 points. Six dimensions is already well into the range where the “curse of dimensionality” really begins to bite, so I find this a significant milestone. At least six people made significant contributions to one of the two proofs.

95.  [gowers](#) Says:  
[February 21, 2009 at 6:03 pm](#) | [Reply](#)

581. Multidimensional Sperner.

I’ve just realized that multidimensional Sperner is easy. I must have been aware of this argument at some point in the past, and certainly I know lots of similar arguments, but this time round I had been blind.

The question I’m interested in is this. Suppose that you have a dense subset of  $[2]^n$ . Sperner tells you that you can find

a pair  $A \subset B$  in  $\mathcal{A}$ , and that gives us a 1-dimensional combinatorial subspace. But what if we want a  $k$ -dimensional subspace?

Here's how I think one can argue. Actually, the further I go the more I realize that it's not completely easy after all. But I think it shouldn't be too bad. I think for convenience I'll use equal-slices measure. In that measure one obtains that if  $\mathcal{A}$  is dense then the proportion of pairs  $A \subset B$  that belong to  $\mathcal{A}$  is also positive. So by averaging I think one can find a set  $D$  such that the set of all  $A$  disjoint from  $D$  such that both  $A$  and  $A \cup D$  belong to  $\mathcal{A}$  is itself dense (in  $[2]^{[n] \setminus D}$ ). By induction that set contains a  $(k-1)$ -dimensional combinatorial subspace and we are done. (This argument is unchecked and I'm not certain the details work out, but if they don't then some version of it with repeated localizations certainly ought to.)

What I'd now like to do is prove that a dense set  $\mathcal{A} \subset [3]^n$  of complexity 1 of the special form discussed in the previous comment contains a combinatorial subspace with dimension that tends to infinity with  $n$ . (Eventually the plan would be to prove more than this: we would like to cover  $\mathcal{A}$  evenly with such subspaces so that a set that correlates with  $\mathcal{A}$  has to correlate with one of the subspaces. But we should walk before we run.) That counts as a separate task so I'll attempt it in the next comment.

96.  [gowers](#) Says:

[February 21, 2009 at 6:32 pm](#) | [Reply](#)

582. Dealing with complexity-1 sets.

Actually, I suddenly feel lazy, for the following reason. It seems as though the multidimensional Sperner proof generalizes rather straightforwardly, and because of that I temporarily can't face checking the details (and thereby heaping vagueness on vagueness — it is probably soon going to be a good idea to do the Sperner argument carefully).

Here's the rough thought though. In [section 2 on this wiki page](#) I showed that a set of the given form contains a combinatorial line. If the Sperner idea above works, then it should work here too: we just find a set  $D$  such that for a dense set of points  $x$  in  $[3]^{[n] \setminus D}$  we can fix  $x$ , let  $D$  be a wildcard set, and thereby obtain a combinatorial line. If so, then once again we should be in a position to say that we have a  $(k-1)$ -dimensional combinatorial subspace in the set of all such  $x$ , and hence a  $k$ -dimensional combinatorial subspace in  $\mathcal{A}$  itself.

Let me take a break now, but end with the following summary of the proof strategy. As I've already said, I would greatly welcome any feedback, so that I can get an idea of which bits look the dodgiest ...

1. As in 580 (but it was closely modelled on arguments that have been discussed already in earlier comments such as [comment 413](#), part of [comment 439](#), and [comment 441](#)) we prove that if  $\mathcal{A}$  contains no combinatorial lines then, locally at least, there is a dense complexity-1 set that is disjoint from  $\mathcal{A}$ . By averaging, we get a dense complexity-1 set inside which  $\mathcal{A}$  has a density increment. Moreover, that complexity-1 set is of the simple form where it depends just on the set-systems to which the 1-set and 2-set belong, with the 3-set free to be whatever it likes.
2. Every dense complexity-1 set of this form contains a multidimensional subspace (as optimistically sketched in 581 and earlier in this comment). Almost certainly this can be lifted up to a proof that a dense complexity-1 set of this kind can be uniformly covered by multidimensional subspaces. And then we get a density increment on a subspace.
3. Therefore, by the usual density-increment strategy, we can keep passing to subspaces until the density reaches 1 (or  $2/3$  if we feel particularly clever), by which point we must have found a combinatorial line. Done!

If there is a problem with covering a special complexity-1 set uniformly with subspaces, then what we could do next is remind ourselves exactly how the Ajtai-Szemerédi proof of the corners theorem goes, since the averaging arguments they did there are likely to be helpful here.



97.  [jozsef](#) Says:

[February 21, 2009 at 7:10 pm](#) | [Reply](#)

583. Multidimensional Sperner.

Tim, isn't it the same question that we considered in 135?

98.  [gowers](#) Says:

[February 21, 2009 at 7:39 pm](#) | [Reply](#)

584. Multidimensional Sperner

Ah — I remembered only Gil's comment 29. So the main point of what I'm saying is not so much that multidimensional Sperner is a new and exciting result — with a proof like that it couldn't be — but that if  $\mathcal{A}$  is a special set of complexity 1, then any 12-cube in  $\mathcal{A}$  (by which I mean that you fix some coordinates and some wildcard sets that take values 1 or 2) automatically extends to a combinatorial subspace. That's quite a nice way of thinking about why special complexity-1 sets are good things. So at least part of the argument sketched in 582 seems to be correct. But there's still the question of finding not just one combinatorial subspace but a uniform covering by combinatorial subspaces. I would be very surprised if that turned out to be a major problem. (In fact, I think I am beginning to see a possible strategy for a proof.)

99.  [gowers](#) Says:

[February 21, 2009 at 7:52 pm](#) | [Reply](#)

Metacomment.

This thread is going to run out soon. I have recently learned from Luca Trevisan that WordPress now allows threading — that is, indented replies to existing comments. I think the current numbering system has served us surprisingly well, and if we introduced threading then we would lose the easily viewed chronological order of comments. But perhaps there would be compensating advantages. If you want to express a view in the following poll, then by all means do, though I suspect that we may have a split vote ...

If I allowed threading, I would be able to choose how much nesting was allowed. One possibility would be to allow only one level, say, and to ask contributors to use the facility as little as possible (perhaps restricting it to short replies to existing comments that do not themselves invite further discussion). That is what I mean by option 3 below.

[polldaddy poll=1390386]

Oh. It seems that what I've just typed works only in a post and not in a comment. So I'd better do that.

100.  [Terence Tao](#) Says:

[February 21, 2009 at 9:10 pm](#) | [Reply](#)

585. Oops

My lemma in 578 might still be true, but I realise now that the proof I supplied is rubbish. Still thinking about it...

101.  [gowers](#) Says:

[February 21, 2009 at 9:17 pm](#) | [Reply](#)

586. Oops

All I can say is that it convinced me at the time ...

102.  [Jozsef](#) Says:

[February 21, 2009 at 10:43 pm](#) | [Reply](#)

587. Dimensions

I'm trying to follow the uniformity/density increment arguments and I have the same feeling like with the planar square problem. (show that every dense subset of the integer grid contains four points of the form  $(x,y), (x+d,y), (x,y+d), (x+d,y+d)$ ) One can try to prove it directly, it is very difficult but probably possible, extending methods of Ajtai-Szemerédi or Shkredov. On the other hand, transferring the problem to dimension three from 2d, makes it more accessible. So, my question is the following: Is there a natural way to consider larger cubes, where a special configuration is (more) dense and a projection of it to the smaller cube,  $3^n$ , is a combinatorial line?

103.  [Terence Tao](#) Says:

[February 22, 2009 at 8:02 am](#) | [Reply](#)

588. Dimensions

Given that the planar square theorem implies the  $k=4$  version of Szemerédi's theorem, I would imagine that Fourier or graph regularity based methods are insufficient for the task.

104.  [Gowers](#) Says:

[February 22, 2009 at 11:02 am](#) | [Reply](#)

589. Dimensions

I agree with Terry, in the sense that I don't think that deducing squares from 3D corners makes the problem fundamentally easier (just ask Michael Lacey ...), and that either problem would have to involve quadratic methods or 3-uniform hypergraphs or something of equivalent difficulty.

Nevertheless, it does seem that the 3D corners problem is the natural way to solve the 2D squares problem, so one can still ask whether it is more natural to find a low-dimensional subspace with high diameter rather than a high-dimensional subspace with low diameter. Here I think there are two possible answers.

Suppose, for simplicity, that we are thinking about a high-dimensional version of Sperner's theorem (but I think what I say generalizes easily to  $\text{DHJ}(k)$ ). That is, we would like not just one wildcard set but several. For further simplicity let's suppose that all we want is two wildcard sets. Then one proof is to use averaging arguments of the kind I was outlining in 581. This is fairly easy. Another proof is to establish  $\text{DHJ}(4)$  and observe that if you write the numbers 0 to 3 in binary then a combinatorial line in  $[4]^n$  when written out gives you a 2-dimensional combinatorial subspace in  $[2]^n$ .

The obvious reaction to this is that the second proof is *much* harder. However, it also gives a much stronger result, because the two wildcard sets have the same cardinality. Furthermore, the stronger result can be seen to be more useful: if you look at unions of slices, then the first result gives you a quadruple  $(a, a+d, a+d', a+d+d')$ , which it is easy to get from Cauchy-Schwarz, whereas the second gives you an arithmetic progression  $(a, a+d, a+2d)$ .


These simple observations raise obvious questions, which I think are close to the kinds of things Jozsef is wondering about. For example, it doesn't seem to be possible to deduce Szemerédi's theorem for progressions of length 4 from the existence of a 2D subspace with equal-sized wildcard sets, so does that mean that the true difficulty of that problem is roughly comparable to that of  $\text{DHJ}(3)$ ? In particular, does it either imply or follow from  $\text{DHJ}(3)$ ? I find this quite an interesting question (though questions often seem interesting when you haven't yet thought about them).

105.  [gowers](#) Says:

[February 22, 2009 at 2:42 pm](#) | [Reply](#)

590. Dimensions

Just realized that the question above was possibly slightly silly. The proof that you can get two wildcard sets of the same size actually gives that one consists of odd numbers and the other of even numbers. So if you look at the number of 1s in odd places plus twice the number of 1s in even places then you get an arithmetic progression of length 4. In other words, what you deduce from DHJ(4) implies Szemerédi's theorem for progressions of length 4, exactly as it should. Strictly speaking, it is just about conceivable that DHJ(3) implies the weaker statement, but I doubt it somehow.

106.  [Ryan O'Donnell](#) Says:

[February 23, 2009 at 12:29 am](#) | [Reply](#)

591. Re Oops/#578.

Hmm, I too was pretty sold on #578, but then I seemed to founder when working out the exact details. My plan was to look at the noise stability of  $f$  at some parameter  $1 - \gamma$ . If this quantity is even a smidgen greater than  $\mathbb{E}[f]^2$  then one can do a density increment.

Otherwise,  $f$  is extremely noise-sensitive at  $1 - \gamma$ . (Or, “uniform at scale  $\gamma^n$ ” as Terry called it.) This strongly feels like it should imply lots of Sperner pairs. But I really ran into difficulties when trying to prove this via the Fourier methods of #476...#567....

107.  [Ryan O'Donnell](#) Says:

[February 23, 2009 at 12:30 am](#) | [Reply](#)

592. Re Oops/#578.

More precisely, it seemed hard to simultaneously control all of the quantities  $\hat{f}_p(A) - \hat{f}_{1/2}(A)$ , even for  $A$  smallish. (I managed to do it for  $A = \emptyset$ , at least 😊.)

108.  [Randall](#) Says:

[February 23, 2009 at 1:08 am](#) | [Reply](#)

593. DHJ (2.7):

I think I managed to remove all hint of ergodic reduction in a proof of DHJ (2.6)...of course if you look at the proof (Plain TeX code will be at the end of this msg), it's because the reduction (to sets, not to stationarity) is inside. Still, it's pretty painless and the Ramsey theory used has been trimmed again, to just Ramsey's theorem this time. Also it doesn't use DHJ (2) as a lemma. I've actually called it DHJ (2.7) because it seems to be very slightly stronger than what Terry called (2.6).

More generally, I have been looking carefully at how to do a “straight translation” of the proof of DHJ (3) and am stuck. The issue arises already in the ergodic proof of the Szemerédi corners theorem. Basically, a set  $A$  of positive density in  $\mathbb{Z}^2$  is used to generate a measure preserving action of  $\mathbb{Z}^2$ , together with a positive measure set  $B$ ; recurrence properties of  $B$  imply existence of configurations for  $A$ . The issue arises because once the system is fixed, one can approximate  $B$  by a “relatively almost periodic” function  $f$ ... $f$  will have a kind of compactness property over a certain factor, meaning that given  $\epsilon > 0$  one can find an  $\epsilon$ -net (composed of functions) of finite cardinality  $M$  such that, on the fibers over the factor in question, the entire orbit of  $f$  will be  $\epsilon$ -close to some member of the net (on most fibers

for most of the orbit). The point is that this number  $M$  is completely fixed by the set  $B$ , which is fixed by the set  $A$ , but the ergodic proof doesn't give any way to bound  $M$  as a function of the density of  $A$ . For all the (usual) ergodic proof says to the contrary, this  $M$  might go to infinity for a sequence of  $A$ 's having measure bounded away from zero. This seems to be a not-very-desirable artefact of the ergodic proof, one that I don't understand very well at present.

\*\*\*\*\*

\magnification=\magstep 1

\noindent {\bf DHJ (2.7):} For all  $\delta > 0$  there exists  $n = n(\delta)$  such that if  $A \subset [3]^n$  with  $|A| > \delta 3^n$ ,

there exists  $\alpha \in \mathbb{N}$ ,  $\alpha > 0$ . Let  $m > \frac{3}{\delta}$

$\delta$  and choose by Ramsey's theorem an  $r$  such that for any 8-coloring of the 2-subsets of  $[r]$ , there is an  $m$ -subset

$B \subset [r]$  all of whose 2-subsets have the same color. Let  $\delta = \delta - \frac{\delta}{4 \cdot 3^r}$  and put

$n_1 = n$

$\delta + \frac{\delta}{4 \cdot 3^r}$ ). Finally put  $n = r + n_1$  and suppose  $A \subset [3]^n$  with  $|A| > \delta 3^n$ .

For each  $v \in [3]^r$ ,

let  $E_v = \{w \in [3]^{n_1} : vw \in A\}$ . If  $|E_v| > (\delta + \frac{\delta}{4 \cdot 3^r}) 3^{n_1}$  for some  $v$  we are done; otherwise

$|E_v| > \frac{\delta}{3}$  for every  $v \in [3]^r$ .

\medskip\noindent Some notation: for  $i \in [r]$  and  $xy \in \{10, 21, 02\}$ , let  $v_i^{xy} \in [3]^r$  be the word

$z_1 z_2 \dots z_r$ ,

where  $z_a = x$  if  $0 \leq a \leq i$  and  $z_a = y$  otherwise. Color  $\{1, j\}$ ,  $0 \leq i < j < r$ , according to whether or not the sets

$E_{v_i^{xy}} \cap E_{v_j^{xy}}$  are empty or not,  $xy \in \{10, 21, 02\}$ . (This is an 8-coloring.) By choice of  $r$  we can find

$0 \leq k_1 < k_2 < \dots < k_m < r$  such that  $\big\{ \{k_i, k_j\} : 0 \leq i < j < m \big\}$  is monochromatic for this coloring.

By pigeonhole,

for, say,  $xy = 10$ , there are  $i < j$  such that  $E_{v_{k_i}^{xy}} \cap E_{v_{k_j}^{xy}} \neq \emptyset$ , hence non-empty for all  $i, j$

by monochromaticity and similarly for  $xy = 21, 02$ . Now for  $xy = 10, 21, 02$ , pick  $u_{xy} \in E_{v_{k_1}^{xy}} \cap E_{v_{k_2}^{xy}}$  and

put  $q_{xy} = s_1 s_2 \dots s_r$ , where  $s_i = x$ ,  $0 \leq i < k_1$ ,  $s_i = 3$ ,  $k_1 \leq i < k_2$ , and  $s_i = y$ ,  $k_2 \leq i < r$ .

Finally put

$w_{xy} = q_{xy} u_{xy}$ . Then  $w_{xy}(x) = v_{k_2}^{xy} u_{xy}$  and  $w_{xy}(y) = v_{k_1}^{xy} u_{xy}$  are in  $A$ ,  $xy \in \{10, 21, 02\}$ .

Hence  $n = n(\delta)$ , contradicting  $\delta < \delta$ .

\end



109. Ryan O'Donnell Says:

[February 23, 2009 at 5:00 am](#) | [Reply](#)

594. Sperner calculations.

Re #567:  $\lambda_{p,q} = \sqrt{p/(1-p)} \sqrt{(1-q)/q}$  by the way (if I remember my calculation correctly). In particular,  $(1-\epsilon)/(1+\epsilon)$  if  $p, q = 1/2 \pm \epsilon/2$ .



110. gowers Says:

[February 23, 2009 at 10:18 am](#) | [Reply](#)

## 595. Sperner calculations

Ryan, thanks for doing that calculation! At some point soon I think I'll write up the whole thing fairly completely and put it on the wiki, which would make it easier to assess what it teaches us, if anything.

I'm also working on step 1 of 582, as I feel that I can get something rigorously proved here too. Again, anything I manage to do properly will go on the wiki. But for now I want to report on the failure of my first attempt to write it up, because it says something interesting (though in retrospect pretty obvious) about equal-slices measure.

I had previously come to the conclusion that a defect of equal-slices measure was that it didn't restrict nicely to equal-slices measure. The mini-realization I have come to is to understand what *does* happen when you restrict it. Suppose a sequence  $x$  is generated Ryan-style: that is, you randomly choose  $p+q+r=1$  and pick each coordinate independently to be 1,2,3 with probability  $p,q,r$ . And now suppose that you restrict to the set of all  $x$  that take certain values on a subset  $E \subset [n]$ . You then find yourself with a classic problem of Bayesian statistics: given the values of  $x_i$  with  $i \in E$ , what is the new probability distribution on the triples  $p+q+r=1$ ? And qualitatively the answer is clear: if  $E$  is at all large, then with immensely high probability the triple  $(p,q,r)$  is approximately proportional to the triple (number of 1s in  $E$ , number of 2s in  $E$ , number of 3s in  $E$ ).

In other words, as soon as you restrict a few coordinates, you end up with something pretty close to a weighted uniform measure (by which I mean that each coordinate is chosen independently with fixed probabilities that are not necessarily equal to  $1/3$ ).

So it seems that equal-slices measure is useful for truly global arguments, but that it collapses to more obvious measures once one starts to restrict.

What about if we restrict to other sorts of subspaces? For example, suppose we partition  $[n]$  into  $s$  wildcard sets of size  $t$ . If you are given that a sequence is constant on some large wildcard set, then by far the most likely way that that could have come about is if one of  $p,q$  and  $r$  is approximately equal to 1. So the measure is concentrated around sequences that are approximately constant.

Going back to step 1 of 582, I still think it should be possible to do it with localization. If I manage, it will go straight up on the wiki and I will report on it here.

111.  [gowers](#) Says:

[February 23, 2009 at 11:06 am](#) | [Reply](#)

Metacomment. This 500s thread has nearly run out. I've created a new post for an 800s thread, so new comments should go there unless they are quite clearly more appropriate here. But once we get to 599 that's it!

112.  [Ryan O'Donnell](#) Says:

[February 23, 2009 at 6:06 pm](#) | [Reply](#)

## 596. Re Oops/Sperner.

I think I can correct Terry's #578 and put it together with our Fourier-based calculations. The writeup is a bit long. I'm going to try to post it here in WordPress using Luca's LaTeX converter. It's my fault if this goes horribly awry; if it does, I'll just post a link to the pdf....

The following alternate argument for Sperner still feels "wrong" to me (I will say why at the end), but I can't see the "right" one yet so I'll write this one up for now.

Let **Formula does not parse** be chosen jointly from  $\{0, 1\} \times \{0, 1\}$  as follows (I use **boldface** to denote random variables): The bit **Formula does not parse** is 0 or 1 with probability  $1/2$  each. If **Formula does not parse** then

**Formula does not parse**. If **Formula does not parse** then **Formula does not parse** with probability  $1 - \epsilon$  and **Formula does not parse** with probability  $\epsilon$ . Finally, let **Formula does not parse** be chosen jointly from  $\{0, 1\}^n \times \{0, 1\}^n$  by using the single-bit distribution  $n$  times independently. Note that **Formula does not parse** always. This distribution is precisely the one Tim uses, with  $p = 1/2$ ,  $q = 1/2 + \epsilon/2$ .

Let  $f : \{0, 1\}^n \rightarrow [-1, 1]$  and let's consider **Formula does not parse**. (We are eventually interested in  $f$ 's with range  $\{0, 1\}$  and mean  $\delta$ , but let's leave it slightly more general for now.) As per Tim's calculations, **Formula does not parse** where  $\hat{f}_\epsilon$  denotes Fourier coefficients with respect to the  $(1/2 + \epsilon/2)$ -biased measure and **Formula does not parse** is defined by **Formula does not parse** (AKA **Formula does not parse**). Separate out the  $S = \emptyset$  term here and use Cauchy-Schwarz on the rest to conclude **Formula does not parse**. Let's compare **Formula does not parse** to  $\hat{f}(\emptyset)$ . Write  $\pi$  and **Formula does not parse** for the density functions of **Formula does not parse**, **Formula does not parse** respectively. Then **Formula does not parse**. By Cauchy-Schwarz, the absolute value of this is upper-bounded by **Formula does not parse** where  $\|f\|_2$  denotes **Formula does not parse**. One easily checks that **Formula does not parse** and since **Formula does not parse** and  $\pi$  are product distributions the RHS of (2) is easy to compute. One can check explicitly that **Formula does not parse** and therefore (2) is **Formula does not parse**. Naturally we will be considering **Formula does not parse**, and in this regime the quantity is bounded by, say, **Formula does not parse**. Hence we have shown **Formula does not parse** and in particular if  $f$  has range  $\{0, 1\}$  and mean  $\mu$  (AKA  $\delta$ ) then **Formula does not parse**. This is not very interesting unless **Formula does not parse**, so let's indeed assume **Formula does not parse** and then we can also use **Formula does not parse**.

We now these deductions in (1). Note that the second factor on the RHS in (1) is at most the square-root of **Formula does not parse**.

Also, using (3) for the LHS in (1) we conclude **Formula does not parse** where **Formula does not parse**.

Let's simply fix **Formula does not parse** at this point. Doing some arithmetic, it follows that if we can bound **Formula does not parse**

(AKA  $f$  is "uniform at scale **Formula does not parse**" as Terry might say) then (4) implies **Formula does not parse**. So long as **Formula does not parse** we've established existence of a Sperner pair (AKA non-degenerate combinatorial line). Since this probability is **Formula does not parse**, we're done assuming  $n \geq O(\log^2(1/\mu)/\mu)$ . (6)


Thus things come down to showing (5). Now in general, there is absolutely no reason why this should be true. The idea, though, is that if it's *not* true then we can do a density increment. More precisely, it is very easy to show (one might credit this to an old result of Linial-Mansour-Nisan) that **Formula does not parse** is precisely **Formula does not parse**, where **Formula does not parse** is a "random restriction with wildcard probability **Formula does not parse**" (and the inner  $\mathbb{E}[\cdot]$  is with respect to the uniform distribution). In other words, **Formula does not parse** is a combinatorial subspace formed by fixing each coordinate randomly with probability **Formula does not parse** and leaving it "free" with probability **Formula does not parse**. Hence if (5) fails then we have **Formula does not parse**.

In particular, since  $f$  is bounded it follows that **Formula does not parse** with probability at least  $\mu^3/128$  over the choice of **Formula does not parse**. It's also very unlikely that **Formula does not parse** will have fewer than, say, **Formula does not parse** wildcards; a large-deviation bound shows this probability is at most **Formula does not parse**. Since **Formula does not parse**, by choosing the constant in (6) suitably large we can make this large-deviation bound strictly less than  $\mu^3/128$ . Thus we conclude that there is a positive probability of choosing some **Formula does not parse** which both has at least **Formula does not parse** free coordinates and also has  $\mathbb{E}[f|_V]^2 \geq \mu^2 + \mu^3/128 \Rightarrow \mathbb{E}[f|_V] \geq \mu + \mu^2/500$ . I.e., we can achieve a density increment.


If I'm not mistaken, this kind of density increment (gaining  $\mu^2/C$  at the expense of going down to  $c\sqrt{\mu}\sqrt{n}$  coordinates, with (6) as the base case) will ultimately show that we need the initial density to be at least  $1/\log \log n$  (up to  $\log \log \log n$  factors?) in order to win. Only a couple of exponentials off the truth 😊




The incorrect quantitative aspect here isn't quite the reason I feel this argument is "wrong". Rather, I believe that no density increment should be necessary. (Actually, we probably know this is the case, by Sperner's proof of Sperner.) In other words, I believe that **Formula does not parse for any  $f$** , assuming **Formula does not parse**.

113.  *Ryan O'Donnell* Says:  
[February 23, 2009 at 6:09 pm](#) | [Reply](#)

597. Wow, indeed that didn't go so great. I should have studied the output more carefully before posting. Sorry about that. I'll post the pdf in Tim's new thread...

114.  *gowers* Says:  
[February 23, 2009 at 6:19 pm](#) | [Reply](#)

Metacomment (so as not to use precious numbers): Ryan, I've had a look at what you wrote and I'm afraid it's beyond my competence to fix — it's full of strange html-ish stuff that I don't understand.

115.  *Ryan O'Donnell* Says:  
[February 24, 2009 at 5:31 am](#) | [Reply](#)

596. Here is the corrected version. I should have read Luca's documentation better...

I think I can correct #578 and put it together with the Fourier-based calculations. The following alternate argument for Sperner still feels "wrong" to me (I will say why at the end), but I can't see the "right" one yet so I'll write this one up for now.

Let  $(\mathbf{x}_1, \mathbf{y}_1)$  be chosen jointly from  $\{0, 1\} \times \{0, 1\}$  as follows (I use **boldface** to denote random variables): The bit  $\mathbf{x}_1$  is 0 or 1 with probability  $1/2$  each. If  $\mathbf{x}_1 = 1$  then  $\mathbf{y}_1 = 1$ . If  $\mathbf{x}_1 = 0$  then  $\mathbf{y}_1 = 0$  with probability  $1 - \epsilon$  and  $\mathbf{y}_1 = 1$  with probability  $\epsilon$ . Finally, let  $(\mathbf{x}, \mathbf{y})$  be chosen jointly from  $\{0, 1\}^n \times \{0, 1\}^n$  by using the single-bit distribution  $n$  times independently. Note that  $\mathbf{x} \leq \mathbf{y}$  always. This distribution is precisely the one Tim uses, with  $p = 1/2, q = 1/2 + \epsilon/2$ .

Let  $f : \{0, 1\}^n \rightarrow [-1, 1]$  and let's consider  $\mathbf{E}[f(\mathbf{x})f(\mathbf{y})]$ . (We are eventually interested in  $f$ 's with range  $\{0, 1\}$  and mean  $\delta$ , but let's leave it slightly more general for now.) As per Tim's calculations,

$$\mathbf{E}[f(\mathbf{x})f(\mathbf{y})] = \sum_{S \subseteq [n]} \hat{f}(S) \hat{f}_\epsilon(S) (1 - \epsilon')^{|S|},$$

where  $\hat{f}_\epsilon$  denotes Fourier coefficients with respect to the  $(1/2 + \epsilon/2)$ -biased measure and  $\epsilon'$  is defined by

$1 - \epsilon' = \sqrt{(1 - \epsilon)/(1 + \epsilon)}$  (AKA  $\lambda_{1/2, 1/2 + \epsilon/2}$ ). Separate out the  $S = \emptyset$  term here and use Cauchy-Schwarz on the

$$\begin{aligned} & \text{rest to conclude } \left| \mathbf{E}[f(\mathbf{x})f(\mathbf{y})] - \hat{f}(\emptyset) \hat{f}_\epsilon(\emptyset) \right| \\ & \leq \sqrt{\sum_{|S| \geq 1} \hat{f}(S)^2 (1 - \epsilon')^{|S|}} \sqrt{\sum_{|S| \geq 1} \hat{f}_\epsilon(S)^2 (1 - \epsilon')^{|S|}}. \quad (1) \end{aligned}$$

Let's compare  $\hat{f}_\epsilon(\emptyset)$  to  $\hat{f}(\emptyset)$ . Write  $\pi$  and  $\pi_\epsilon$  for the density functions of  $\mathbf{x}, \mathbf{y}$  respectively. Then

$$\begin{aligned} \hat{f}_\epsilon(\emptyset) - \hat{f}(\emptyset) &= \mathbf{E}[f(\mathbf{y})] - \mathbf{E}[f(\mathbf{x})] \\ &= \mathbf{E} \left[ \frac{\pi_\epsilon(\mathbf{x})}{\pi(\mathbf{x})} f(\mathbf{x}) \right] - \mathbf{E}[f(\mathbf{x})] = \mathbf{E} \left[ \left( \frac{\pi_\epsilon(\mathbf{x})}{\pi(\mathbf{x})} - 1 \right) f(\mathbf{x}) \right]. \end{aligned}$$

By Cauchy-Schwarz, the absolute value of this is upper-bounded by  $\sqrt{\mathbf{E} \left[ \left( \frac{\pi_\epsilon(\mathbf{x})}{\pi(\mathbf{x})} - 1 \right)^2 \right]} \cdot \|f\|_2$ ,

where  $\|f\|_2$  denotes  $\sqrt{\mathbf{E}[f(\mathbf{x})^2]}$ . One easily checks that  $\mathbf{E} \left[ \left( \frac{\pi_\epsilon(\mathbf{x})}{\pi(\mathbf{x})} - 1 \right)^2 \right] = \mathbf{E} \left[ \frac{\pi_\epsilon(\mathbf{x})^2}{\pi(\mathbf{x})^2} \right] - 1$ , (2)

and since  $\pi_\epsilon$  and  $\pi$  are product distributions the RHS of (2) is easy to compute. One can check explicitly that

$$\mathbf{E}[\pi_\epsilon(\mathbf{x}_1)^2 / \pi(\mathbf{x}_1)^2] = 1 + \epsilon^2$$

and therefore (2) is  $(1 + \epsilon^2)^n - 1$ . Naturally we will be considering  $\epsilon \ll 1/\sqrt{n}$ , and in this regime the quantity is bounded by, say,  $4\epsilon^2 n$ . Hence we have shown  $|\hat{f}_\epsilon(\emptyset) - \hat{f}(\emptyset)| \leq 2\epsilon\sqrt{n} \cdot \|f\|_2$  and in particular if  $f$  has range  $\{0, 1\}$  and mean  $\mu$  (AKA  $\delta$ ) then  $|\hat{f}_\epsilon(\emptyset) - \mu| \leq 2\epsilon\sqrt{n} \cdot \sqrt{\mu}$ . (3)

This is not very interesting unless  $2\epsilon\sqrt{n} \cdot \sqrt{\mu} \leq \mu$ , so let's indeed assume  $\epsilon \leq \sqrt{\mu}/\sqrt{n}$  and then we can also use  $\hat{f}_\epsilon(\emptyset) \leq 2\mu$ .

We now these deductions in (1). Note that the second factor on the RHS in (1) is at most the square-root of

$$\sum_S \hat{f}_\epsilon(S)^2 = \mathbf{E}[f(\mathbf{y})^2] = \mathbf{E}[f(\mathbf{y})] = \hat{f}_\epsilon(\emptyset) \leq 2\mu \leq 4\mu.$$

Also, using (3) for the LHS in (1) we conclude  $|\mathbf{E}[f(\mathbf{x})f(\mathbf{y})] - \mu^2| \leq 2\mu^{3/2} \cdot \epsilon\sqrt{n} + 2\sqrt{\mu}\sqrt{\mathbb{S}_{1-\epsilon'}(f) - \mu^2}$ , (4)

where  $\mathbb{S}_{1-\epsilon'}(f) = \sum_S \hat{f}(S)^2(1 - \epsilon')^{|S|}$ .

Let's simply fix  $\epsilon = (1/8)\sqrt{\mu}/\sqrt{n}$  at this point. Doing some arithmetic, it follows that if we can bound  $\mathbb{S}_{1-\epsilon'}(f) - \mu^2 \leq \mu^3/64$  (?) (5)

(AKA  $f$  is “uniform at scale  $\epsilon'n$ ” as Terry might say) then (4) implies  $\mathbf{E}[f(\mathbf{x})f(\mathbf{y})] \geq \mu^2/2$ .

So long as  $\mathbf{P}[\mathbf{x} = \mathbf{y}] < \mu^2/2$  we've established existence of a Sperner pair (AKA non-degenerate combinatorial line).

Since this probability is  $(1 - \epsilon/2)^n \leq \exp(-\epsilon n/2) = \exp(-\Omega(\sqrt{\mu}\sqrt{n}))$ , we're done assuming  $n \geq O(\log^2(1/\mu)/\mu)$ . (6)

Thus things come down to showing (5). Now in general, there is absolutely no reason why this should be true. The idea, though, is that if it's *not* true then we can do a density increment. More precisely, it is very easy to show (one might credit this to an old result of Linial-Mansour-Nisan) that  $\mathbb{S}_{1-\epsilon'}(f)$  is precisely  $\mathbf{E}_{\mathbf{V}}[\mathbf{E}[f|\mathbf{V}]^2]$ , where  $\mathbf{V}$  is a “random restriction with wildcard probability  $\epsilon'$ ” (and the inner  $\mathbf{E}[\cdot]$  is with respect to the uniform distribution). In other words,  $\mathbf{V}$  is a combinatorial subspace formed by fixing each coordinate randomly with probability  $1 - \epsilon'$  and leaving it “free” with probability  $\epsilon'$ . Hence if (5) fails then we have  $\mathbf{E}[\mathbf{E}[f|\mathbf{V}]^2] \geq \mu^2 + \mu^3/64$ .

In particular, since  $f$  is bounded it follows that  $\mathbf{E}[f|\mathbf{V}]^2 \geq \mu^2 + \mu^3/128$  with probability at least  $\mu^3/128$  over the choice of  $\mathbf{V}$ . It's also very unlikely that  $\mathbf{V}$  will have fewer than, say,  $(\epsilon'/2)n$  wildcards; a large-deviation bound shows this probability is at most  $\exp(-\Omega(\epsilon'n))$ . Since  $\epsilon' \approx \epsilon = (1/8)\sqrt{\mu}/\sqrt{n}$ , by choosing the constant in (6) suitably large we can make this large-deviation bound strictly less than  $\mu^3/128$ . Thus we conclude that there is a positive probability of choosing some  $\mathbf{V} = V$  which both has at least  $(\epsilon'/2)n = \Omega(\sqrt{\mu}\sqrt{n})$  free coordinates and also has  $\mathbf{E}[f|V]^2 \geq \mu^2 + \mu^3/128 \Rightarrow \mathbf{E}[f|V] \geq \mu + \mu^2/500$ .

I.e., we can achieve a density increment.

If I'm not mistaken, this kind of density increment (gaining  $\mu^2/C$  at the expense of going down to  $c\sqrt{\mu}\sqrt{n}$  coordinates, with (6) as the base case) will ultimately show that we need the initial density to be at least  $1/\log \log n$  (up to  $\log \log \log n$  factors?) in order to win. Only a couple of exponentials off the truth 😊

The incorrect quantitative aspect here isn't quite the reason I feel this argument is “wrong”. Rather, I believe that no density increment should be necessary. (Actually, we probably know this is the case, by Sperner's proof of Sperner.) In other words, I believe that  $\mathbf{E}[f(\mathbf{x})f(\mathbf{y})] \geq \Omega(\mu^2)$  for any  $f$ , assuming  $\epsilon \ll \sqrt{\mu}/\sqrt{n}$ .

## Leave a Reply

Logged in as [Ryan O'Donnell](#). [Logout »](#)

Submit Comment

☐ Notify me of follow-up comments via email.

---

[Blog at WordPress.com.](#)

[Entries \(RSS\)](#) and [Comments \(RSS\)](#).

⌵

# A reading seminar on density Hales-Jewett

11 February, 2009 in [math.DS](#) | Tags: [polymath1](#), [density Hales-Jewett theorem](#), [Yitzhak Katznelson](#), [Hillel Furstenberg](#), [Randall McCutcheon](#) | by [Terence Tao](#)

As part of the [polymath1 project](#), I would like to set up a reading seminar on this blog for the following three papers and notes:

1. H. Furstenberg, Y. Katznelson, “[A density version of the Hales-Jewett theorem for  \$k=3\$](#) “, Graph Theory and Combinatorics (Cambridge, 1988). Discrete Math. 75 (1989), no. 1-3, 227–241.
2. R. McCutcheon, “[The conclusion of the proof of the density Hales-Jewett theorem for  \$k=3\$](#) “, unpublished.
3. H. Furstenberg, Y. Katznelson, “[A density version of the Hales-Jewett theorem](#)“, J. Anal. Math. 57 (1991), 64–119.

As I understand it, paper #1 begins the proof of DHJ(3) (the  $k=3$  version of density Hales-Jewett), but the proof is not quite complete, and the notes in #2 completes the proof using ideas from both paper #1 and paper #3. Paper #3, of course, does DHJ( $k$ ) for all  $k$ . For the purposes of the polymath1 project, though, I think it would be best if we focus exclusively on  $k=3$ .

While this seminar is of course related in content to the main discussion threads in the polymath1 project, I envision this to be a more sedate affair, in which we go slowly through various sections of various papers, asking questions of each other along the way, and presenting various bits and pieces of the proof. The papers require a certain technical background in ergodic theory in order to understand, but my hope is that if enough other people (in particular, combinatorialists) ask questions here (and “naive” or “silly” questions are *strongly* encouraged) then we should be able to make a fair amount of the arguments here accessible. I also hope that some ergodic theorists who have been intending to read these papers already, but didn’t get around to it, will join with reading the papers with me.

This is the first time I am trying something like this, and so we shall be using the carefully thought out protocol known as “making things up as we go along”. My initial plan is to start understanding the “big picture” (in particular, to outline the general strategy of proof), while also slowly going through the key stages of that proof in something resembling a linear order. But I imagine that the focus may change as the seminar progresses.

I’ll start the ball rolling with some initial impressions of paper #1 in the comments below. As with other threads in this project, I would like all comments to come with a number and title, starting with 600 and then incrementing (the numbers 1-599 being reserved by other threads in this project).

## 36 comments

[Comments feed for this article](#)

11 February, 2009 at 6:36 pm

## Terence Tao



600. Initial impressions of paper #1

Paper #1 outlines a proof of DHJ(3): given any  $\delta > 0$ , that any subset  $A$  of  $[3]^n = \{0, 1, 2\}^n$  of density at least  $\delta$  contains a combinatorial line if  $n$  is sufficiently large depending on  $\delta$ . (The paper refers to combinatorial lines as “HJ sequences”, and  $[3]^n$  as  $\Omega_n$ , but I am translating to be more compatible with the notation already used in other threads.) It proceeds by first reducing the problem to one concerning a family of measurable sets on a probability space, indexed by words, and for which there is a family of (non-commuting!) measure-preserving transformations relating these sets to each other. This reduction is performed in Proposition 3.1.

A key point in this reduction is that a certain “stationarity” property is obtained on this family of sets. To obtain this stationarity, one needs a version of the Carlson-Simpson theorem, which seems to be an infinitary strengthening of the colouring Hales-Jewett theorem. (It looks vaguely reminiscent of Hindman’s theorem; presumably there is a relation.)

With the above reductions, the game is now to show that a certain triple intersection of three sets has positive measure (Theorem B). This is not actually done in paper #1, but is covered in #2 or #3. In #1, the simpler (but still non-trivial) result is shown that all pairwise intersections of the three sets can simultaneously have positive measure (Theorem 4.1). (It would be interesting to see what the combinatorial analogue of this weaker statement is. Getting just one of the pairs to have positive measure sounds similar to the “DHJ(2.5)” I formulated in [my comment 130](#), which turns out to be equivalent to DHJ(2) and thus an easy consequence of Sperner’s theorem.)

The proof of Theorem 4.1 is established from some soft functional analysis results about IP-systems of unitary operators in Section 5.

In Section 6, there is a discussion of what is needed to establish the full strength of Theorem B, and I guess this is completed in #2.

My initial plan is to go to Section 3 and start understanding the equivalence between the combinatorial problem and the ergodic theory one. There seems to be several stages to this equivalence; a relatively easy correspondence between a combinatorial problem and an ergodic problem without stationarity, and then the reduction to the stationary case which requires the Carlson-Simpson type theorems.

11 February, 2009 at 8:01 pm

## Terence Tao



601. Proposition 3.1

Proposition 3.1 of paper #1 equates the combinatorial statement DHJ(3) with various more ergodic-flavoured versions of the same statement. There are three main statements in this equivalence, (a), (b), and (c). (There are also minor variants (a\*), (b\*), (c\*) of (a), (b), and (c), but I think we won't need to focus on them too much.)

DHJ(3)(a) is the combinatorial version of DHJ(3), i.e. for any  $\delta > 0$  and for  $n$  sufficiently large depending on  $\delta$ , every subset of  $[3]^n$  of density at least  $\delta$  contains a combinatorial line.

DHJ(3)(b) is a preliminary ergodic version of DHJ(3). Define a  $[3]^n$ -system to be a collection of subsets  $(E_w)_{w \in [3]^n}$  in some probability space  $(X, \mathcal{X}, \mu)$ , indexed by strings in  $[3]^n$ .

**DHJ(3)(b):** If  $\delta > 0$  and  $n$  is sufficiently large depending on  $\delta$ , and  $(E_w)_{w \in [3]^n}$  is a  $[3]^n$ -system with all sets  $E_x$  having measure  $\mu(E_w)$  greater than or equal to  $\delta$ , then there exists a combinatorial line  $w^0, w^1, w^2$  such that  $E_{w^0} \cap E_{w^1} \cap E_{w^2} \neq \emptyset$ .

This can be stated in a probabilistic language: if  $A$  is a *random* subset of  $[3]^n$  with the property that each element  $w$  of  $[3]^n$  belongs to  $A$  with probability at least  $\delta$ , then with positive probability,  $A$  contains a combinatorial line. (Indeed, just take  $A$  to be the set of those  $w \in [3]^n$  for which  $E_w$  holds.)

To deduce DHJ(3)(b) from DHJ(3)(a), observe from linearity of expectation that the expected density of  $A$  is at least  $\delta$ , and so  $A$  has density at least  $\delta/2$  (say) with positive probability.

To deduce DHJ(3)(a) from DHJ(3)(b), one can use random averaging arguments (as discussed for instance in [comment 460](#)). [Paper #1 uses a somewhat different derivation of this implication which I will skip here.] To recap that argument briefly: we pick a medium integer  $m$  between 1 and  $n$ , and pick a random  $m$ -dimensional subspace inside  $[3]^n$ . Pulling back the dense subset  $A$  of  $[3]^n$  to  $[3]^m$ , we get a random subset  $A'$  of  $[3]^m$  with each point lying in  $A'$  with probability at least  $\delta/2$  (say), if  $n$  is large enough depending on  $m$ . The claim then follows.

DHJ(3)(c) is DHJ(3)(b) with an additional stationarity hypothesis thrown in, and will be discussed subsequently.

[12 February, 2009 at 2:10 am](#)

[gowers](#)



602.

Dear Terry,

I think it's a very interesting idea to have an online reading seminar. Of course, it requires us to do the reading, which is quite a big investment to make, and presumably why you envisage a sedate affair. I think what would be most useful to me is basically what you encourage: rather than reading the paper very carefully and then explaining bits I understand here, I would prefer to try to use this post and its associated comments to read the paper sloppily and present my very partial understanding, or even complete misunderstanding, in the hope that others will be able to say, "Yes, you're getting there — what you need



to grasp next is this,” or “No, that’s not quite it — if that were the case then so would this be but it clearly isn’t,” and things like that. So when I get a moment, I’ll try to skim-read Randall’s paper and give some kind of initial reaction to it. I haven’t yet looked at it at all, but my eventual aim would be to obtain some kind of non-ergodic understanding of the proof. (Here I mean non-ergodic in the shallow sense of trying to explain things without getting completely into the ergodic language, rather than in the deep sense of genuinely doing without ergodic theory.)

[12 February, 2009 at 4:32 am](#)

**zugzwang trebuchet**



603.

Hi gowers,  
Why not work out the ergodic details here?

[12 February, 2009 at 8:19 am](#)

**[David Speyer](#)**



604.

I’m going to try to restate proposition 3.1 to see if I understand it. If I get it right, this will probably be redundant with comment 601 above, but if I get it wrong, it will be one of those naive questions Terry asked for.

Prop. 3.1 says that six things are equivalent, called (a), (a\*), (b), (b\*), (c) and (c\*). In each case, the unstarred version is a uniform variant on the starred version. So the starred version will say “If we have a sequence  $A_n$  where  $\limsup$  (some measure of the size of  $A_n$ )  $> 0$  then...” while the unstarred version will say “For every  $\epsilon > 0$ , there is an integer  $N$  such that, if  $n > N$  and (some measure of the size of  $A_n$ )  $> \epsilon$  then...”.

We’re going to concentrate on the unstarred versions.

(a) is what we’re trying to prove. For every  $\epsilon$ , there is an  $N$  such that any subset of  $[3]^N$  of cardinality more than  $\epsilon 3^N$  contains a combinatorial line.

(b) says that we can’t break (a) by a random construction. Suppose we had some random process, like drawing ping pong balls out of a bag. The set of balls is called  $X$ , and the probability that an individual ball  $x \in X$  is drawn will be  $\mu(x)$ . For each ball  $x$ , our process produces a particular subset of  $[3]^n$ , which I will denote  $A(x)$ . Confusingly (to me), H and K use the opposite indexing; for each  $w \in [3]^N$ , they write  $B_w$  for the set of all  $x$  such that  $w \in A(x)$ , but they don’t have a notation for  $A(x)$ .

For fixed  $\epsilon$ , if  $N$  is large enough, the following holds: is long as the probability of any particular element  $w$

of  $[3]^N$  being in  $A(x)$  is at least  $\epsilon$ , then there will be some particular combinatorial line which occurs in  $A(x)$  with nonzero probability.

Some comments:

Notice that (b)'s framework is general enough to destroy any probabilistic construction I might attempt. In particular, there is no hypothesis that the events " $w_1$  is in  $A(x)$ " and " $w_2$  is in  $A(x)$ " are independent in any way.

The only kind of probabilistic construction I could imagine which would violate the hypotheses of (a) is a process which, although they produce sets of density  $\geq \epsilon$ , have much lower probability of including certain particular elements of  $[3]^N$ . However, the problem is so symmetric that I see no reason that this would help. Presumably, the proof of (b) $\rightarrow$ (a) is to make this idea rigorous.

Finally, I described things above with a finite collection of ping pong balls, each with a positive probability. H and K phrase things in terms of a general measure space. I assume this generality will be helpful, although it seems pointless so far.

Whew, that's long! I'll save condition (c) for later, as I need to get going now.

[12 February, 2009 at 11:51 am](#)

**[Terence Tao](#)**



605.

Dear Tim: Well, I don't really have an organised plan as to where this will go, but I would like to get as many contributions (from as many backgrounds) as possible, to take full advantage of [Metcalf's law](#). So if some people read slowly and carefully, others comment loosely and freely, and yet others translate from ergodic theory to combinatorics or back, I think that will be an instructive mix. (One advantage of an internet-based seminar, as opposed to a physical one, is that one can have multiple concurrent discussions on different aspects of a paper without too much confusion.)

Incidentally, the McCutcheon notes rely heavily on paper #1, so I think it is best to start with that paper first, though it may still be valuable to skim through #2 at this preliminary stage nevertheless.

Dear David: Welcome! I'm glad you took the initiative to jump in (there is no roster or anything for this seminar :-). Please don't worry about redundancy - in fact, I think the more perspectives we have on an individual component of the argument, the better.

Regarding the implication of (b) from (a), I can give you a simpler version of it, in which we are not looking for combinatorial lines in  $[3]^n$ , but rather algebraic lines  $(x, x+r, x+2r)$  in  $(\mathbb{Z}/3\mathbb{Z})^n$ ; the analogue of DHJ(3) here is called "Roth's theorem in  $(\mathbb{Z}/3\mathbb{Z})^n$ ". Given a (deterministic) set  $A$  in  $(\mathbb{Z}/3\mathbb{Z})^n$  of density at least  $\delta$ , we can define a random set  $A(x)$  in  $(\mathbb{Z}/3\mathbb{Z})^n$  by setting  $A(x) := A + x$  where  $x$  is a random element of  $(\mathbb{Z}/3\mathbb{Z})^n$ . Note that each element  $w$  of  $(\mathbb{Z}/3\mathbb{Z})^n$  will lie in  $A(x)$  with probability at least  $\delta$ , and if  $A(x)$  contains an algebraic line with positive probability, then  $A$  certainly contains an algebraic line also. So

in this case the deduction of (a) from (b) is pretty simple.

This doesn't work in the DHJ setting because combinatorial lines are not translation invariant, but one can do a somewhat different trick, embedding  $[3]^m$  randomly in  $[3]^n$  in a combinatorial line-preserving manner, rather than by translating  $[3]^n$  randomly back to  $[3]^n$ .

[12 February, 2009 at 1:13 pm](#)

[Terence Tao](#)



606. Stationarity

In Proposition 3.1(b), the  $[3]^n$ -system  $(E_w)_{w \in [3]^n}$  (which is called  $(B_w)_{w \in \Omega_n}$  in the notation of #1, and is equivalent to the notion of a random subset  $A(x)$  of  $[3]^n$  as discussed in Speyer.604), has no structure other than having every set  $E_w$  be large. But the deepest part of Proposition 3.1 is a reduction to a special type of  $[3]^n$ -system, namely a *stationary*  $[3]^n$ -system. This is a system where the sets  $E_w \subset X$  take the form

$$E_w = U_w^{-1} E_\emptyset, \quad (1)$$

where

$$U_{w_1 \dots w_n} := U_1^{w_1} U_2^{w_2} \dots U_n^{w_n}, \quad (2)$$

$E_\emptyset \subset X$  is a set, and the  $U_i^j : X \rightarrow X$  for  $1 \leq i \leq n$  and  $j = 0, 1, 2$  are measure-preserving invertible transformations on  $X$ . [These transformations are *not* assumed to commute with each other, and this is apparently a significant technical difficulty in the rest of the argument.]

Clearly, if (1) holds, and  $E_\emptyset$  has measure at least  $\delta$ , then all the  $E_w$  have measure at least  $\delta$ . So DHJ(3)(b) certainly implies

**DHJ(3)(c):** If  $\delta > 0$  and  $n$  is sufficiently large depending on  $\delta$ , and  $(E_w)_{w \in [3]^n}$  is a stationary  $[3]^n$ -system with  $E_\emptyset$  having measure at least  $\delta$ , then there exists a combinatorial line  $(w^0, w^1, w^2)$  such that  $E_{w^0} \cap E_{w^1} \cap E_{w^2}$  has positive measure.

The non-trivial fact is that DHJ(3)(c) also implies DHJ(3)(b). I guess I'll try to discuss that in future comments.

It is tempting to try to describe stationarity in terms of a measure-preserving action on  $X$  by the free group  $F_3$  on three generators  $e^0, e^1, e^2$ , but this isn't quite right, because the  $U_i^j$  are not constant in  $i$ . But if one adds a discrete time variable  $i$ , thus replacing  $X$  with  $\mathbb{Z} \times X$ , then one can recover a measure-preserving action  $(T_g : \mathbb{Z} \times X \rightarrow \mathbb{Z} \times X)_{g \in F_3}$  of  $F_3$ , with each generator  $e^j$ ,  $j = 0, 1, 2$  of  $F_3$  mapping  $(i, x)$  to  $(i - 1, U_i^j x)$  for  $i \in \mathbb{Z}$  and  $x \in X$  (here we have to define  $U_i^j$  for non-positive  $i$  in some arbitrary fashion, e.g. setting these transformations to be the identity). With this action, and embedding  $[3]^n$  inside  $F_3$  by identifying a string  $w_1 \dots w_n$  with the group element  $e^{w_1} \dots e^{w_n}$ , one can rewrite (1), (2) as

$$\{n\} \times E_w = T_w^{-1}(\{0\} \times E_\emptyset) \quad (3)$$

and so the task is to show that there exists a combinatorial line  $(w^0, w^1, w^2)$  such that

$$T_{w^0}^{-1}(\{0\} \times E_\emptyset) \cap T_{w^1}^{-1}(\{0\} \times E_\emptyset) \cap T_{w^2}^{-1}(\{0\} \times E_\emptyset) \neq \emptyset. \quad (4)$$

At this stage I do not know whether this group action perspective is useful, although it does make the problem look a little bit like simpler ergodic recurrence theorems. For instance, Roth's theorem is equivalent to the assertion that if  $T$  acts in a measure-preserving way on a probability space  $X$ , then for every set  $E$  of positive measure, there exists an arithmetic progression  $(w^0, w^1, w^2)$  of integers such that

$$(T^{w^0})^{-1}(E) \cap (T^{w^1})^{-1}(E) \cap (T^{w^2})^{-1}(E) \neq \emptyset. \quad (5)$$

[12 February, 2009 at 1:40 pm](#)

**Kristal Cantwell**



I would be interested in this seminar. I have downloaded the three papers and will be looking at them in the next few days along with the comments here.

[12 February, 2009 at 1:56 pm](#)

**Terence Tao**



607. Stationarity cont.

Welcome, Kristal! Feel free to jump in and contribute with any questions, comments, or suggestions, even if (or especially if) they overlap with those already in this post.

One small comment, continuing 606: it appears that the stationarity gives quite strong constraints on the  $E_w$ . Most obviously, they are all forced to have the same measure, but there are other constraints too. For instance, if  $n=2$ , then  $E_{00} \cap E_{10}$  has to have the same measure as  $E_{02} \cap E_{12}$ , since the latter set can be obtained from the former set by applying the measure-preserving  $(U_2^2)^{-1}U_2^0$ . To put it another way, the random set  $A(x) \subset [3]^2$  is just as likely to contain the pair  $(0,0), (1,0)$  as it is to contain the pair  $(0,2), (1,2)$ .

More generally, stationarity seems to be saying that if one looks at a slice

$A(x) \cap ([3]^{n-m} \times \{w_{n-m+1} \dots w_n\})$  of  $A(x)$  formed by freezing the last  $m$  positions of  $A(x)$ , that the distribution of that slice (or more precisely, the identification of that slice with a subset of  $[3]^{n-m}$ ) is in fact independent of the choice of the frozen coordinates

$w_{n-m+1}, \dots, w_n$ . (This would explain the terminology “stationary”.) I do not know yet whether this property already captures the full strength of stationarity, or whether stationarity in fact says something more.

[12 February, 2009 at 2:06 pm](#)

## Terence Tao



### 608. Symmetry breaking

I wanted to repeat a remark of [Austin.54](#) here: the original formulation DHJ(3)(a) of the problem, together with its probabilistic variant DHJ(3)(b), are invariant under permutation of the indices in  $[n]$ , but the notion of stationarity is not invariant (being dependent on the ordering of the indices), and so DHJ(3)(c) has “broken” the permutation symmetry. There is nothing particularly harmful about this - the standard proofs of colouring Hales-Jewett also break symmetry, for instance - but it seems like a noteworthy point nonetheless.

[12 February, 2009 at 6:43 pm](#)

## David Speyer



609: My first dumb question: In (b) and (c), would it add any generality to allow the measure space  $X$  to depend on  $n$ ? When I think of sampling methods I might use, it seems natural to me to use different sample spaces  $X_n$  for different  $n$ . But I suspect that I could use  $\prod_{n=1}^{\infty} X_n$  as my sample space, and thus reduce to H and K's set up.

[12 February, 2009 at 8:46 pm](#)

## Randall



610: Fixed some typos I found on an airplane today and put a better version up. The most potentially confusing of these were in Lemma 1 (continued), the first display on page 4 and in the definition of  $\delta$  on page 6. There may be others so please speak up if you see anything that looks incongruous.

The notion of stationarity is more familiar for (linear) processes. If  $(X_i)$  are random variables taking values say in  $\{a, b, \dots, z\}$ , then we call  $(X_i)$  a process and we say that the process is stationary if the probability that  $X_1 X_2 X_3 = \text{“cat”}$  is the same as the probability that  $X_{17} X_{18} X_{19} = \text{“cat”}$  (for all cats and all shifts).

In the current setup, you can view the sets  $B_w$  as random variables  $X_w$  taking values in  $\{0, 1\}$ . Stationarity says: take any finite set of words. Call that set  $I$ . The probability of seeing a certain pattern of 0s and 1s in the random variables  $X_w$ ,  $w$  in  $I$ , is exactly the same as the probability of seeing the same pattern of 0s and 1s in the random variables  $X_{vw}$ ,  $w$  in  $I$ , where  $v$  is any word. Here “ $v$ ” plays the role 16 did in the cat example, the pattern of 0s and 1s plays the role of “cat”, and  $I$  plays the role of  $\{1, 2, 3\}$ .

I can certainly sympathize with the sentiment that it seems quite unclear what 3.1 can possibly do to help. It gives you some measure preserving transformations, but nothing resembling honest ergodic theory, really, because although there are measure preserving transformations around, they don't even constitute an

action (as Terry pointed out). I guess the thing to do at this early stage is just have faith....

[12 February, 2009 at 8:51 pm](#)

**[Terence Tao](#)**



611: the measure space  $X$

David, I think you're right, the actual choice of space  $X$  is more or less irrelevant. In a lot of these problems one can take the Cantor space  $\{0, 1\}^{\mathbb{Z}}$  (or the unit interval) as a universal space (note that any countably generated probability space can be lifted up to either of these two spaces).

One of the philosophies of probability theory, by the way, is to try to hide the underlying space  $X$  as much as possible; thus one talks about events  $E$  and their probabilities, random variables and their expectations and correlations, etc. without explicit mention of the probability space and probability measure, etc. It's like studying a manifold or variety through its ring of functions rather than by looking at its constituent points. It has the advantage that one can concatenate any number of random processes together (e.g. flip infinitely many coins, and then pull balls out of urns infinitely often, etc.) without having to think about what happens to the underlying probability space.

In ergodic theory, though, one often takes almost the opposite view: one starts with an abstract measure space  $X$  (with some groups acting on it) and then creates some factors  $Y$  of that space which often have a very explicit algebraic structure (e.g. they are the homogeneous space for some Lie group). One then studies the geometry and dynamics of that space  $Y$  from a very concrete point of view, thus moving from the category of abstract measure spaces and measure preserving transformations to something much more algebraic.

[13 February, 2009 at 1:53 am](#)

**[DHJ — possible proof strategies « Gowers's Weblog](#)**

[...] is his comment 460, which again is responded to by other comments. Terry has also begun an online reading seminar on the Furstenberg-Katznelson proof, using not just the original paper of Furstenberg and [...]

[13 February, 2009 at 11:56 am](#)

**Kristal Cantwell**



612. Hindman's theorem is a special case of one of their versions of the Carlson-Simpson theorem according to the paper.

[13 February, 2009 at 8:56 pm](#)

**[Terence Tao](#)**





## 613. Stationarity II

DHJ(3)(c) finds combinatorial lines in stationary  $[3]^n$ -systems  $(U_w^{-1}E)_{w \in [3]^n}$ . It turns out to be convenient to eliminate the role of  $n$ , and instead work with stationary  $[3]^\omega$ -systems  $(U_w^{-1}E)_{w \in [3]^\omega}$ , where  $[3]^\omega := \bigcup_{n=0}^{\infty} [3]^n$  is the space of all words of finite length with alphabet  $\{0,1,2\}$ . (Paper #1 uses the notation  $\Omega^f$  for  $[3]^\omega$ .) It is clear that DHJ(3)(c) implies

**DHJ(3)(c\*):** If  $(U_w^{-1}E)_{w \in [3]^\omega}$  is a stationary  $[3]^\omega$ -system, with  $E$  having positive measure, then there is a combinatorial line  $(w^0, w^1, w^2)$  such that  $U_{w^0}^{-1}E \cap U_{w^1}^{-1}E \cap U_{w^2}^{-1}E \neq \emptyset$ .

In a similar spirit, DHJ(3)(b) is easily seen to be equivalent to

**DHJ(3)(b\*):** If  $(E_w)_{w \in [3]^\omega}$  is a  $[3]^\omega$ -system, with the  $E_w$  having measure uniformly bounded away from zero, then there is a combinatorial line  $(w^0, w^1, w^2)$  such that  $E_{w^0} \cap E_{w^1} \cap E_{w^2} \neq \emptyset$ .

To finish the proof of Proposition 3.1, we need to deduce DHJ(3)(b\*) from DHJ(3)(c\*). Broadly speaking, the idea is to start with a  $[3]^\omega$ -system and repeatedly use Carlson-Simpson-type Ramsey theorems to pass to subsystems in which become increasingly stationary.

A  $[3]^\omega$ -system can be viewed as a collection of random sets  $A_n(x) \subset [3]^n$ , one for each  $n$ . As noted already in 610, stationarity implies that  $A_n(x) \times \{v\}$  has the same distribution as  $A_{n+m}(x) \cap ([3]^n \times \{v\})$  for all  $n$  and all words  $v \in [3]^m$ , where of course we identify  $[3]^{n+m}$  with  $[3]^n \times [3]^m$  in the obvious manner.

[13 February, 2009 at 10:30 pm](#)

## Terence Tao



## 614. A possible simplification

It appears to me that one can in fact get stationarity for free (and avoid invoking the Carlson-Simpson theorem here). The basic observation is that when one deduces DHJ(3)(b) from DHJ(3)(a) by using a random  $m$ -dimensional subspace inside  $[3]^n$ , the random set obtained in this fashion already is essentially stationary by construction. I'll write up the details on the wiki and report back here when I'm done.

(This does not completely eliminate the need for Carlson-Simpson, though, as such results are also used later in the paper.)

[14 February, 2009 at 9:12 am](#)

## Ryan O'Donnell



615. Let me try to recap Terry's points, passing from DHJ(3)(a) to DHJ(3)(c), in my own language.

Let  $A \subseteq \{0, 1, 2\}^n$  be a given set of density  $\delta$ .

Let  $m = m(\delta, n)$  be some moderate integer; perhaps  $n^{1/3}$  when  $\delta$  is a “constant”.

Let  $S \subseteq [n]$  be a random subset of cardinality  $m$ . Imagine choosing a “random restriction”  $x$  to the coordinates  $\bar{S} = [n] \setminus S$ . I.e.,  $x \in \{0, 1, 2\}^{\bar{S}}$  is chosen uniformly at random.

Then  $A$  induces a probability distribution on subsets of  $\{0, 1, 2\}^S$ . Call the induced subset  $A_x \subseteq \{0, 1, 2\}^S$ .

Let  $y$  denote a string in  $\{0, 1, 2\}^S$ . I will use the notation “ $(x, y)$ ” for the string in  $\{0, 1, 2\}^n$  gotten by appropriately piecing  $x$  and  $y$  together.

We claim that for \*each\* string  $y \in \{0, 1, 2\}^S$ , the probability (over choice of  $x$ ) that  $y \in A_x$  is at least  $\delta/2$ . On one hand, note that  $(x, y)$  is \*not\* a uniformly distributed string in  $\{0, 1, 2\}^n$ . For example, if  $y$  is the all-0's string, then the distribution on  $(x, y)$  is slightly biased towards those strings with more 0's.

Still if  $m$  is “small enough” then the distribution will be very close to uniform (except around the “extreme slices” of  $\{0, 1, 2\}^n$ , but  $A$  has almost all its density outside these slices anyway). And this should prove the claim with a minimum of calculations.

At this point we're at DHJ(3)(b); on to DHJ(3)(c). “Stationarity” here would seem to be that for any further restriction  $z$  to coordinates  $S' \subseteq S$ , the resulting distribution on the set  $A_{(x,z)}$  (a subset of  $\{0, 1, 2\}^{S \setminus S'}$ ) is the same.

Again, this is not true, but it should be “almost” true. And it seems like we're only going for “almost stationarity”, not “stationarity” anyway.

In fact, it seems to me that whereas “(almost) stationarity” importantly involves an \*ordering\* on the coordinates  $S$ , it seems this argument doesn't really. I.e.,  $S'$  need not be a “suffix” of  $S$  or anything.

[14 February, 2009 at 10:10 am](#)

[Terence Tao](#)



616.

Thanks Ryan, this fleshes out what I was trying to say.

I've started a running “combinatorial translation” of paper #1 on the wiki at

[http://michaelnielsen.org/polymath1/index.php?title=Furstenberg-Katznelson\\_argument](http://michaelnielsen.org/polymath1/index.php?title=Furstenberg-Katznelson_argument)

It's likely to evolve in unstable ways over time (in particular, I am likely to overhaul the notation as I understand more of the big picture) but hopefully will eventually settle down.

As you note, the random sampling construction gives additional stationarity properties beyond those used

in Paper #1. In particular, the random sets  $A_m \subset [3]^m$  are also stationary wrt permutations of the index set  $[m]$ . It's not clear whether this additional symmetry could be useful though (it might even be harmful, because later parts of the proof may need to break this symmetry and so trying to hold on to it may be counterproductive).

[14 February, 2009 at 10:35 am](#)

**[Terence Tao](#)**



617. Recap

I'm now going to try to recapitulate the previous discussion in light of the new simplification, thus moving away from the presentation in Paper #1. (I guess I am changing the rules a little bit here from a "reading seminar" to a "revisionist seminar", but I said that the rules would be made up as we went along :-)) I will try to keep the wiki page above current with the latest revision of thinking.

Let  $A^{(n)}$  be a dense line-free subset of  $[3]^n$ , where  $n$  is very large. For every small integer  $m$ , we can then form a random line-free subset  $A_m^{(n)} \subset [3]^m$  by randomly embedding  $[3]^m$  inside  $[3]^n$  (by picking  $m$  random positions for the wildcards, and then picking the frozen positions randomly), and then pulling  $A^{(n)}$  back by this map. One can think of  $A_m^{(n)}$  as a random variable taking values in the power set  $X_m := 2^{[3]^m}$ , so its probability distribution is a probability measure  $\mu_m^{(n)}$  on  $X_m$ .

By construction,  $A_m^{(n)}$  is stationary with respect to the permutation group  $S_m$  of  $[m]$ , i.e.  $(X_m, \mu_m^{(n)})$  is invariant wrt the action of  $S_m$ . Also, for any word  $w \in [3]^r$ , the *slicing map*  $U_{m+r}^w : X_{m+r} \rightarrow X_m$  defined by

$$U_{m+r}^w(A_{m+r}) := \{v \in [3]^m : vw \in A_{m+r}\}$$

is *almost stationary* in the sense that for fixed  $m, r$ , and  $w$ , the random variables  $A_m^{(n)}$  and  $U_{m+r}^w(A_{m+r}^{(n)})$  have asymptotically the same distribution in the limit  $n \rightarrow \infty$ .

Taking a subsequence limit as  $n \rightarrow \infty$  (and using the Helly selection principle, a.k.a. the vague sequential compactness of probability measures), we can then find random line-free sets  $A_m \subset [3]^m$  (or equivalently, a probability measure  $\mu_m$  on  $X_m$ ) which is stationary with respect to the permutation groups  $S_m$  acting on  $X_m$ , and also with respect to the slicing maps  $U_{m+r}^w : X_{m+r} \rightarrow X_m$ . Also, if the  $A^{(n)}$  all have density at least  $\delta$ , then so do the  $A_m$ : in particular,  $A_0$  will be non-empty with probability at least  $\delta$ , which implies that for any  $v \in [3]^m$ , that  $v \in A_m$  with probability at least  $\delta$  also.

This is all very well and good, but there is one minor technical problem which will cause difficulty in later parts of paper #1: the slicing maps  $U_{m+r}^w : X_{m+r} \rightarrow X_m$  are not invertible. We would like to upgrade them to be invertible (thus upgrading the semigroupoid of transformations  $U_{m+r}^w$  to a groupoid), thus giving us DHJ(3)(c\*) in its full strength. This can be done, but requires extending the measure spaces  $X_m$  a bit (or, in probabilistic language, adding a whole bunch of "dice" or "random number generators"). I plan to talk about this in a later comment.

[14 February, 2009 at 11:56 am](#)

## [Terence Tao](#)



### 618. Inversion of maps

There is a simple lemma that says that any finite probability-preserving surjective map can be inverted if one throws in a random number generator that draws uniformly from  $[0,1]$ . More precisely:

**Lemma** Let  $U : X \rightarrow Y$  be a surjective map from one finite probability space to another which preserves the probability measure (in the sense that  $U$  pushes the probability measure on  $X$  forward to the probability measure on  $Y$ ). Then one can lift this map to an *invertible* measure-preserving map  $\tilde{U} : X \times [0,1] \rightarrow Y \times [0,1]$ , with product measure.

I prove this lemma in the wiki notes, but let's just illustrate this with the simplest example: the non-injective map from the two-element space  $\{a,b\}$  (with probability  $1/2$  of each) to the one-element space  $\{c\}$ . One invertible lift of this map is the map  $\tilde{U} : \{a,b\} \times [0,1] \rightarrow \{c\} \times [0,1]$  which maps  $(a,t)$  to  $(c,t/2)$  and  $(b,t)$  to  $(c,(t+1)/2)$ . Many other choices are possible; this lift is not unique.

Applying this lift to all of the generators  $U_{m+1}^0, U_{m+1}^1, U_{m+1}^2 : X_{m+1} \rightarrow X_m$ , we can now find probability spaces  $\tilde{X}_m := X_m \times [0,1]$  and invertible measure-preserving maps  $\tilde{U}_{m+1}^i : \tilde{X}_{m+1} \rightarrow \tilde{X}_m$  that commute with the slicing maps  $U_{m+1}^i : X_{m+1} \rightarrow X_m$ . Concatenating these maps then give us more invertible measure-preserving maps  $\tilde{U}_{m+r}^w : \tilde{X}_{m+r} \rightarrow \tilde{X}_m$  for all words  $w$  of length  $r$ . We are now basically in the situation of DHJ(3)(c\*), which the rest of Paper #1 aims to prove. (It's slightly different because we've "graded" the underlying probability space  $X$ , replacing it with a sequence  $X_m$  of probability spaces, but I don't think this will make much of a difference.)

[Remark: the lemma here is my interpretation of Lemma 3.2 in Paper #1; the statement there is false as stated, but can be repaired by throwing in the  $[0,1]$  as is done above.]

[14 February, 2009 at 12:53 pm](#)

## [Terence Tao](#)



### 619.

A frivolous comment, for the combinatorially minded: the Lemma in 618 is in some sense a continuous version of (a special case of) Hall's marriage theorem, and serves much the same purpose (viz. conjuring up a bunch of artificial but useful perfect matchings out of thin air).

[14 February, 2009 at 2:01 pm](#)

## [Ryan O'Donnell](#)



## 620. Inversion of maps.

I'm getting slightly lost in notation and also slightly lost because I don't quite know where this is going. However it seems to me that we might not eventually need a completely general Lemma like the one in Terry.618. Some natural randomized-inverse map might work in our context; e.g., tack on a random substring from  $[3]^r$ .

[14 February, 2009 at 2:39 pm](#)

**[Terence Tao](#)**



## 621. Inversion of maps

Sorry about the notation shifting - I also don't fully know where we are going either, and this I think is reflected in the notation...

We're not inverting a map from  $[3]^{m+r}$  to  $[3]^m$ , but rather a map from  $2^{[3]^{m+r}}$  to  $2^{[3]^m}$  (which, I guess, is induced by an embedding of  $[3]^m$  in  $[3]^{m+r}$ ). Given a word  $w$  of length  $r$ , and a set  $A_{m+r} \subset [3]^{m+r}$ , we have the slice  $U_{m+r}^w(A_{m+r}) \subset [3]^m$ , defined as

$$U_{m+r}^w(A_{m+r}) := \{x \in A_m : xw \in A_{m+r}\}.$$

Stationarity tells us that if we take the random set  $A_{m+r} \subset [3]^{m+r}$  and look at a slice  $U_{m+r}^w(A_{m+r})$  of it (think of  $w$  as fixed), one gets back something with the same distribution as the random set  $A_m \subset [3]^m$ . But one cannot initially invert this procedure: starting with the random set  $A_m \subset [3]^m$ , one cannot *deterministically* lift it back to a set in  $[3]^{m+r}$  that has the same distribution as  $A_{m+r}$  (and which agrees with to  $A_m$  on the slice) - there is not enough entropy in the domain and too much in the range. But one can do this once one has a random number generator, which I am modeling by a uniform distribution on  $[0,1]$  - this gives us an infinite amount of entropy on both sides and removes the obstruction to invertibility, as per "Hilbert's hotel". (One has to take care that one only performs "reversible computations" in order to maintain invertibility, but this is a minor technical detail.)

[14 February, 2009 at 8:05 pm](#)

**Ryan O'Donnell**



622. Understood, I'm just thinking that since we got to  $[3]^m$  in the first place by making a large random restriction, perhaps we can "remember" this and "unrestrict"  $r$  coordinates later. Just guessing... I'm probably being too speculative right now. Best to see where the next parts of the arguments go.

[14 February, 2009 at 11:02 pm](#)

**[Terence Tao](#)**



623. Oh, that's an interesting idea. The artificial nature of the transformations used in the Furstenberg-Katznelson argument bothered me a bit, and this would be a much more natural way to do it. I don't quite see how to set it up so that everything is completely reversible (i.e. the relevant maps  $U_{m+r}^w$  are invertible), so that entropy is neither created nor destroyed, but I can imagine that it is possible.

I've read a little further into the paper (an interesting experiment, reading a paper collaboratively a small fragment at a time). With the above revisionist notation, we now have a sequence  $(X_m, \mu_m)$  of probability spaces and invertible measure-preserving maps  $U_{m+r}^w : X_{m+r} \rightarrow X_m$  for all words  $w$ , obeying a semigroup law. The claim is now that for every set  $E \subset X_0$  of positive measure, that there exists a combinatorial line  $w^0, w^1, w^2$  in some cube  $[3]^n$  for some  $n$  such that the triple intersection of  $(U_n^{w^0})^{-1}(E)$ ,  $(U_n^{w^1})^{-1}(E)$ , and  $(U_n^{w^2})^{-1}(E)$  has positive measure (or equivalently, that the random set  $A_n \subset [3]^n$  contains the line  $w^0, w^1, w^2$  with positive probability).

In Paper #1, a somewhat weaker result is proven, namely there exists a line  $w^0, w^1, w^2$ , such that the three *pairwise* intersections of  $(U_n^{w^0})^{-1}(E)$ ,  $(U_n^{w^1})^{-1}(E)$ , and  $(U_n^{w^2})^{-1}(E)$  have positive measure (thus the random set  $A_n$  will contain any two of the points on this line with positive probability). The notes in #2 finish the job to get all three points on the line. This weaker statement is already not obvious to me combinatorially - it's some sort of triple version of the DHJ(2.5) statement we discussed back around 130 or so. (I think it is asserting the existence of a positive  $r$  for any dense  $A \subset [3]^n$  with  $r$  wildcards such that for each  $ij=01, 12, 20$ , there exists a combinatorial line which intersects  $A$  in the  $i$  and  $j$  position.) I think I'll ask it at the 500 thread to see if anyone can come up with a combinatorial proof.

[15 February, 2009 at 4:19 pm](#)

### [Terence Tao](#)



624. Compactification of the string space

I've scanned through the rest of the paper, and it seems like one has no choice but to start understanding the statement of the Carlson-Simpson theorem (and its variants) before progressing much further, although one can at least take these statements as "black boxes". This theorem is an infinite-dimensional version of the colouring Hales-Jewett theorem, but to state it properly, we need some notation.

I'm using  $[3]^n := \{0, 1, 2\}^n$  for the strings of length  $n$ , and  $[3]^\omega := \bigcup_{n=1}^\infty [3]^n$  for the finite length strings.

We can give  $[3]^\omega$  a "3-adic" metric by defining the distance between two distinct strings  $x, y$  (possibly of different length) to be  $d(x, y) := 3^{-l}$ , where  $l$  is the first digit at which  $x$  and  $y$  differ (or terminate).

The metric space  $[3]^\omega$  is totally bounded but incomplete. One can complete it by attaching the space  $[3]^\infty$  of infinite strings. The space  $[3]^\omega \cup [3]^\infty$  is totally bounded and complete, with  $[3]^\omega$  as a dense subspace, thus  $[3]^\omega \cup [3]^\infty$  is a compactification of  $[3]^\omega$ . (Roughly, one can think of  $[3]^\omega$  as being like the terminating decimals base 3 in  $[0, 1]$ , and  $[3]^\omega \cup [3]^\infty$  as being like the entire interval  $[0, 1]$ .)



[Paper #1 uses  $\Omega_n, \Omega^f, \Omega$  in place of  $[3]^n, [3]^\omega, [3]^\infty$  is uniformly continuous if and only if it extends to a continuous map  $F : [3]^\omega \cup [3]^\infty \rightarrow X$ . One can think of a uniformly continuous  $F$  as an operation that takes a string as input and spits out a point in  $X$  as output, but is increasingly insensitive to fluctuations in the  $n^{\text{th}}$  and higher digits as  $n \rightarrow \infty$ .

Of course, not every map on  $[3]^\omega$  is continuous. (A good example of a discontinuous map: the map  $F : [3]^\omega \rightarrow \{0, 1\}$  that counts the parity of the number of 1s in the string.) But the Carlson-Simpson theorem is an assertion that every map  $F$  from  $[3]^\omega$  to a compact set (or a finite set) becomes uniformly continuous when restricted to an “infinite-dimensional combinatorial subspace”. This allows one to extend  $F$  at some key points  $\omega$  of  $[3]^\infty$ . The proof is going to proceed by taking the shift maps  $\tilde{U}_r^{w^i}$  (viewed as unitary operators) and taking a suitable limit (in the weak operator topology) as  $w^i \rightarrow \omega$  and working with the limiting operators (which turn out to be orthogonal projections). I don't know yet why this is going to be a useful thing to do.

In a subsequent comment I will state the Carlson-Simpson theorem properly, both in the finite colouring version (which is the standard formulation) and in the compact version (which is the version preferred by paper #1).

[16 February, 2009 at 1:00 pm](#)

**Ryan O'Donnell**



625. In the analysis of boolean functions there is general intuition of continuous/measurable functions vs. discontinuous/nonmeasurable functions. Roughly, the former are those functions which are “asymptotically noise stable” (in the sense of Benjamini-Kalai-Schramm: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.48.2927>) or have “analogue” functions on Gaussian space or the sphere. I wonder if this viewpoint will be helpful.

[16 February, 2009 at 2:59 pm](#)

**Terence Tao**



626. Colouring theorems I: finitary theorems

In this post and the next I want to describe a square of four Ramsey theorems: two finitary theorems, and two infinitary theorems. It is the latter which is used in Paper #1.

The first finitary theorem is the **colouring Hales-Jewett theorem**. The  $k=3$  version of this theorem asserts that if  $[3]^n$  is partitioned into  $c$  colour classes, and  $n$  is sufficiently large depending on  $c$ , then one of the colour classes contains a combinatorial line.

Iterating this theorem, we conclude that  $[3]^n$  is partitioned into  $c$  colour classes, and  $n$  is sufficiently large depending on  $c, m$ , then one of the colour classes contains an  $m$ -dimensional combinatorial subspace.

The second finitary theorem is the **Graham-Rothschild theorem**. The  $k=3$  version of this theorem asserts that if, instead of colouring the vertices of  $[3]^n$  as in Hales-Jewett, one colours the *combinatorial lines* in  $[3]^n$ , partitioning the space of all such lines into  $c$  colours, then (if  $n$  is sufficiently large depending on  $c, m$ ) one can find an  $m$ -dimensional combinatorial subspace in which all lines in that space have the same colour.

Recall that a combinatorial line can be viewed as a string of 0s, 1s, 2s, and a wildcard  $*$ , with the wildcard appearing at least once (e.g.  $01*2*$  is a combinatorial line). Identifying the wildcard with the symbol 3, we can thus view a combinatorial line as a point in  $[4]^n$  containing at least one 4. Thus one can rephrase the  $k=3$  Graham-Rothschild theorem in another way:

**Graham-Rothschild theorem.** If the vertices of  $[4]^n$  are partitioned into  $c$  colour classes, then (if  $n$  is sufficiently large depending on  $m, c$ ) there is an  $m$ -dimensional subspace (with none of the fixed coordinates equal to 3) such that all points in this subspace (with at least one coordinate equal to 3) have the same colour.

Written this way, Graham-Rothschild looks a lot like colouring Hales-Jewett, and it is in fact possible to deduce the latter from the former (the precise deduction escapes me right now, though).

The  $k=1$  version of the Graham-Rothschild theorem implies Folkman's theorem: if the positive integers are finitely coloured, then for any  $m$ , one can find  $m$  positive integers such that all finite sums of those integers (other than the trivial sum 0) have the same colour. This can be shown by considering the colouring on  $[2]^n \setminus \{0\}$  that colours a vertex by the colour given to the number of 1s on that vertex.

In the next comment I describe the infinitary strengthenings of the colouring Hales-Jewett and the Graham-Rothschild theorems, which are the Carlson-Simpson theorem and Carlson's theorem respectively.

16 February, 2009 at 3:21 pm

Terence Tao



627. Colouring theorems II. Infinitary theorems

The infinitary strengthening of the Hales-Jewett theorem is the Carlson-Simpson theorem. To define it, I need some notation. Define an *infinite-dimensional combinatorial subspace* of  $[3]^\omega$  to be an object indexed by an infinite string of 0s, 1s, 2s, and an infinite number of wildcards  $*_1, *_2, *_3, \dots$ , where each wildcard  $*_i$  appears in a non-empty consecutive block  $I_i$ , with each  $I_i$  appearing to the left of  $I_{i+1}$ . For instance, the following string describes an infinite-dimensional combinatorial subspace:

2110  $*_1$   $*_1$ 03012  $*_2$   $*_2$   $*_2$  2221111  $*_3$  201  $*_4$   $*_4$   $\dots$

This gives an embedding of  $[3]^\omega$  into  $[3]^\omega$  by mapping any  $n$ -digit string into the string formed from the above infinite string by substituting these digits into the first  $n$  wildcards  $*_1, \dots, *_n$ , and then truncating everything after  $I_n$ . For instance, in the above example, 20 maps to

21102203012000.

**Carlson-Simpson theorem** If  $[3]^\omega$  is finitely coloured, then one of the colour classes contains an infinite-dimensional combinatorial subspace.

This theorem implies the colouring Hales-Jewett theorem but is significantly stronger, in particular it does not seem to be deducible solely from the colouring Hales-Jewett theorem (much as the infinite pigeonhole principle cannot be deduced from its finite counterpart).

It implies a compact version:

**Carlson-Simpson theorem, compact version** If  $F : [3]^\omega \rightarrow X$  is a map from  $[3]^\omega$  into a compact metric space  $X$ , then there exists an infinite-dimensional combinatorial subspace  $\phi : [3]^\omega \rightarrow [3]^\omega$  such that  $F \circ \phi : [3]^\omega \rightarrow X$  is uniformly continuous (i.e. it extends continuously to the compactification  $[3]^\omega \cup [3]^\infty$ ).

It is not hard to see that the two versions are equivalent (one has to cover the compact metric space by an increasingly fine mesh of small balls, which can then be used to colour  $[3]^\omega$ ). In paper #1 it is remarked that the compact case is in fact slightly easier to prove than the finite colouring case (I suppose that the induction is somehow cleaner.)

In a similar spirit, the Graham-Rothschild theorem has the following infinitary generalisation:

**Carlson's theorem** Suppose that the combinatorial lines in  $[3]^\omega$  are finitely coloured. Then there is an infinite-dimensional combinatorial subspace on which all lines have the same colour.

Or equivalently:

**Carlson's theorem** Suppose that  $[4]^\omega$  is finitely coloured. Then there is an infinite-dimensional combinatorial subspace with no fixed coordinate equal to 3, such that any element of this space with at least one 3 has the same colour.

Again, there is a compact version:

**Carlson's theorem** Suppose that  $F : [4]^\omega \rightarrow X$  is a map into a compact metric space  $X$ . Then there is an infinite-dimensional combinatorial subspace with no fixed coordinate equal to 3, such that the restriction of  $F$  to this space is uniformly continuous on those points with at least one 3.

As I understand it, Carlson's theorem was discovered independently by Carlson and by Furstenberg-Katznelson.

The  $k=1$  case of Carlson's theorem is essentially Hindman's theorem (the infinitary generalisation of Folkman's theorem). Given that the cleanest proof of Hindman's theorem involves idempotent ultrafilters, I would imagine that this is how Carlson's theorem is also proved. Needless to say, such theorems are not easily finitisable.

[17 February, 2009 at 9:59 pm](#)

[Terence Tao](#)



## 628. Convergence to an orthogonal projection

Recall that our objective is to show that, given any system of measure-preserving transformations  $U_{m+|w|}^w : X_{m+|w|} \rightarrow X_m$  obeying a groupoid law, and any set  $E \subset X_0$  of positive measure, that there exists a combinatorial line  $w = (w^0, w^1, w^2)$  in some  $[3]^n$  such that  $(U_n^{w^i})^{-1}(E) \cap (U_n^{w^j})^{-1}(E)$  has positive measure for  $ij=01,12,20$ .

The measure of  $(U_n^{w^i})^{-1}(E) \cap (U_n^{w^j})^{-1}(E)$  can also be written as

$$\langle 1_E, U_n^{w^i} (U_n^{w^j})^{-1} 1_E \rangle_{L^2(X_0)}, \quad (1)$$

where we let measure-preserving transformations  $T : X \rightarrow X$  act on functions  $f : X \rightarrow \mathbb{R}$  in the usual manner,  $Tf := f \circ T^{-1}$ .

The key claim is

**Claim** There exists a sequence of combinatorial lines  $w_m = (w_m^0, w_m^1, w_m^2) \in [3]^{n_m}$  such that for each  $ij=01,12,20$ , the unitary operators  $U_{n_m}^{w_m^i} (U_{n_m}^{w_m^j})^{-1}$  converge in the weak operator topology to orthogonal projections  $P_{ij}$ .

Indeed, if this claim held, then (1) for the word  $w_m$  would converge as  $m \rightarrow \infty$  to

$$\langle 1_E, P_{ij} 1_E \rangle = \|P_{ij} 1_E\|_{L^2}^2.$$

On the other hand, since

$$\langle 1, U_{n_m}^{w_m^i} (U_{n_m}^{w_m^j})^{-1} 1_E \rangle = \mu_0(E),$$

we must have  $\langle 1, P_{ij} 1_E \rangle = \mu_0(E) > 0$  and hence  $\|P_{ij} 1_E\|_{L^2} > 0$ , and so (1) is positive for sufficiently large  $m$ .

The main tool in proving the claim is the Carlson-Simpson theorem.

This is a somewhat infinitary argument; at present I do not see what the combinatorial analogue of it will be.

[20 February, 2009 at 11:30 am](#)

## Terence Tao



## 629. The invertible maps $U_{m+1}^j$ and group structure

I am beginning to understand a bit better the strange collection of invertible measure-preserving transformations  $U_{m+1}^j : X_{m+1} \rightarrow X_m$  that appear in the papers 1, 2, 3, which in turn generate the

compound maps  $U_{m+|w|}^w : X_{m+|w|} \rightarrow X_m$ .

Firstly, as Ryan proposed in 622, there is a nice probabilistic interpretation of these transformations that does not require the artificial device of introducing a new source of randomness, such as the uniform probability measure on  $[0,1]$ . Roughly speaking, one can view each  $X_m$  as the space of random  $m$ -dimensional combinatorial subspaces in  $[3]^n$ , where each of the  $m$  wildcards  $*_1, \dots, *_m$  appear exactly once, plus some additional data which is a bit tricky to describe cleanly (details are [on the wiki](#)).

The transformation  $U_{m+1}^j : X_{m+1} \rightarrow X_m$  for  $j=1,2,3$  is the map that takes the random  $m+1$ -dimensional subspace to an  $m$ -dimensional slice, formed by replacing the final wildcard  $*_{m+1}$  with  $j$ .

To invert this transformation (i.e. extending an  $m$ -dimensional subspace with an  $m+1$ -dimensional one), what one would like to do is start with the random  $m$ -dimensional subspace, search the fixed coordinates randomly for an instance of  $j$ , and replace this  $j$  with a wildcard  $*_{m+1}$ . Of course, this is a random map rather than a deterministic one; so what one has to do is to “roll the dice in advance” and store the choices of fixed coordinates one would use to extend the transformation as part of the data used to define an element of  $X_m$ . In fact one needs to roll the dice once for every element of the free group  $F_3$  on three generators, representing all the future possibilities of extending and contracting the space.

The construction actually gives more than the invertible maps  $U_{m+|w|}^w : X_{m+|w|} \rightarrow X_m$ ; it also gives a permutation action on each  $X_n$ , and one can multiply all the  $X_n$  to get an honest-to-goodness measure-preserving group action  $U_w : X \rightarrow X$  of the free group  $F_3$ , rather than this funny groupoid action that Furstenberg and Katznelson use. However (as was pointed out all the way back in Austin.6), all this additional structure is actually a hindrance to the proof, because one has to let go of it in order to take advantage of the powerful Ramsey theorems available such as Carlson’s theorem.

I still feel that there should be some clean algebraic substitute for the concept of a group action that can clarify conceptually the dynamics of this system, but I do not yet know what it should be (something like a semi-groupoid action). The key thing seems to be that the structure one has generates lots of “IP-systems”, which I will discuss next.

[20 February, 2009 at 11:48 am](#)

[Terence Tao](#)



630. An IP convergence lemma

Let  $U_1, U_2, \dots \in G$  be a sequence of group elements, then we can define  $U_\alpha \in G$  for any finite set  $\alpha$  of natural numbers by the formula

$$U_{\{a_1, \dots, a_n\}} := U_{a_1} \dots U_{a_n}$$

when  $a_1 < \dots < a_n$ . The  $U_\alpha$  form an *IP-system*, which means that they have the restricted multiplicativity property

$$U_\alpha U_\beta = U_{\alpha \cup \beta}$$

whenever  $\alpha$  is to the left of  $\beta$ . (This is an example of the “semi-groupoid action” concept I am struggling to formalise in a clean algebraic fashion.)

If  $G$  has a topology on it, we say that the IP-system  $U_\alpha$  converges to a limit  $U \in G$  if every neighbourhood of  $U$  contains the  $U_\alpha$  “for sufficiently large  $\alpha$ ”, which means that the least element  $\min(\alpha)$  of  $\alpha$  is sufficiently large.

Hindman’s theorem is equivalent to the claim that any IP-system in a compact metrisable space has a convergent sub-IP-system (I will not define what a sub-IP-system is here; again, I do not feel that I have that the correct algebraic formalism for all this yet.).

The following lemma is the key in Paper #1 (together with Carlson’s theorem, which has a similar flavour to Hindman’s theorem) to proving the claim in 628:

**Lemma.** Let  $U_\alpha$  be an IP system of unitary operators on a separable Hilbert space  $H$  that converges (in the weak operator topology) to a limit  $P$ . Then  $P$  is an orthogonal projection.

[Note that the weak operator topology on the operators with norm at most 1 on  $H$  is compact metrisable.]

Proof. Since the  $U_\alpha$  have operator norm at most 1,  $P$  has norm at most 1 as well. Since the only idempotents of norm at most 1 are the orthogonal projections, it suffices to show that  $P^2 = P$ .

Let  $v \in H$ . Then  $U_\alpha^* v$  converges weakly to  $P^* v$  as  $\alpha \rightarrow \infty$ . But for fixed  $\alpha$ ,  $U_\beta^* U_\alpha^* v$  converges weakly to  $P^* U_\alpha^* v$  as  $\beta \rightarrow \infty$ ; thus

$$\lim_{\alpha \rightarrow \infty} \lim_{\beta \rightarrow \infty} U_\beta^* U_\alpha^* v = (P^*)^2 v,$$

where the limits are in the weak sense. On the other hand, using the IP property, the limit on the left is  $P^* v$ . Thus  $P^* = (P^*)^2$  and so  $P$  is idempotent.

[21 February, 2009 at 8:25 pm](#)

[Terence Tao](#)



631. An informal combinatorial translation of the ergodic proof of DHJ(2)

I’m beginning to understand (though still in a somewhat fuzzy, informal way) what the ergodic argument in Paper #1 would translate to in purely combinatorial terms. For simplicity I’ll start with DHJ(2) - the claim that any dense subset  $A$  of  $[2]^n$  contains a combinatorial line.

Let  $m$  be an integer that grows slowly with  $n$  (think of  $m$  as being like  $\lfloor \log \log n \rfloor$ , for instance), and let  $V$  be a random  $m$ -dimensional combinatorial subspace of  $[2]^n$ , where I will be deliberately vague as to what distribution  $V$  is drawn from here. One can use this random subspace  $V$  to create various real-valued random variables  $f(V)$  which depend only on what  $A$  is doing on  $V$ . For instance, one could consider the indicator random variable  $1_{E_0}(V)$ , which equals 1 when the bottom corner of  $V$  (formed by sending all  $m$  wildcards to 0) lies in  $A$ ; one could consider the random variable  $d(V)$  that measures the density of  $A$  in  $V$ ;



and so forth.

Now, the random subspace  $V$  has a lot of fixed coordinates, and so with very high probability it's going to have a lot of fixed 0s and fixed 1s (about  $n/2$  of each). Let  $V'$  be the random subspace parallel to  $V$  formed by taking one of the fixed 0s at random and flipping it to a 1. Then every random variable  $f(V)$  that is associated to  $V$  has a “shifted” version  $f(V')$ . Note that we expect  $V'$  to have almost the same distribution as  $V$ , and so  $f(V)$  and  $f(V')$  should also have about the same distribution; this is a *stationarity* property. However, this does not mean that  $f(V)$  and  $f(V')$  are necessarily close to each other.

There are two fundamental types of random variable:

\* A random variable  $f(V)$  is *periodic* if  $f(V')$  and  $f(V)$  are close, thus  $\mathbb{E}|f(V) - f(V')|$  is small.

\* A random variable  $f(V)$  is *mixing* if  $f(V')$  and  $f(V)$  are uncorrelated, thus  $\mathbb{E}f(V)f(V')$  is small.

Examples: any constant random variable is periodic. If the set  $A$  has “low influence” (which means that flipping a bit from 0 to 1 is unlikely to affect membership in  $A$ ) then every random variable  $f(V)$  is in fact periodic. At the other extreme, if  $A$  is a random set of a fixed density  $\delta$ , then any random variable  $f(V)$  of mean zero will be mixing, and only the constants will be periodic.

Now, here is the basic fact (modulo some lies):

**Fact.** Every random variable  $f(V)$  can be decomposed (uniquely) into a periodic random variable  $Pf(V)$  and a mixing random variable  $(1-P)f(V)$ . Furthermore, the periodic random variables and mixing random variables are always uncorrelated (i.e. orthogonal to each other).

This fact is a heavily disguised version of the lemma in 630, and can also be viewed as a sort of “IP mean ergodic theorem”. It is not quite true as stated; to be true, one has to redefine  $V'$  to not be formed by flipping just one fixed coordinate of  $V$ , but by flipping an unspecified but bounded number of fixed coordinates. (The number of digits to flip will be determined by a Ramsey-type theorem, such as Folkman’s theorem, Graham-Rothschild, etc. The Ramsey theorem will also tell us what distribution  $V$  is to be drawn from, and in particular how many times each wildcard will appear.).

Assume this fact. Now we can prove DHJ(2). If we let  $1_{E_0}(V)$  be the event that the bottom left corner of  $V$  lies in  $A$ , it suffices to show that

$$\mathbb{E}1_{E_0}(V)1_{E_0}(V')$$

is large.

Observe that the claim would be easy if  $1_{E_0}(V)$  was periodic, since then  $1_{E_0}(V) \approx 1_{E_0}(V')$ . The claim would also be easy if  $1_{E_0}(V)$  was mixing, since then the two random variables  $1_{E_0}(V), 1_{E_0}(V')$  would be roughly independent. In general, the Fact tells us that we are in a combination of the two situations. Splitting both  $1_{E_0}(V)$  and  $1_{E_0}(V')$  into periodic and mixing components, we get four terms, three of which are small, leaving us with

$$\mathbb{E}P1_{E_0}(V)P1_{E_0}(V') \approx \mathbb{E}P1_{E_0}(V)^2. \quad (1)$$

But observe that the mixing function  $(1 - P)1_{E_0}(V)$  is orthogonal to the periodic function 1, and thus has zero expectation. Since  $1_{E_0}(V)$  has expectation  $\delta$ , we conclude that  $P1_{E_0}(V)$  has expectation  $\delta$ , and so (by Cauchy Schwarz) the expression (1) is large, and we are done.

In my next comment I will try to indicate how DHJ(3) works in some special cases. Here there is not just one projection  $P$ , but now three separate projections  $P$ , reflecting interchange between 0 and 1, 1 and 2, or 2 and 0. Things are relatively easy when one of the projections is extreme (either everything is periodic, or only the constants are periodic), but the case in which all projections are intermediate (so some random variables are periodic and others are mixing) is quite tricky.

[21 February, 2009 at 9:27 pm](#)

[Terence Tao](#)



632. Some special cases of DHJ(3)

Now we move from  $k=2$  to  $k=3$ . Again, we consider an  $m$ -dimensional random subspace  $V$  of  $[3]^n$ . Instead of a single companion subspace  $V'$ , we now have three spaces  $V_{0 \rightarrow 1}, V_{1 \rightarrow 2}, V_{2 \rightarrow 0}$ , formed from  $V$  by flipping a fixed 0 to a 1, or a fixed 1 to a 2, or a fixed 2 to a 0. We can then define the notion of a random variable  $f(V)$  being 01-periodic or 01-mixing (resp. 12-periodic or 12-mixing, etc.) by comparing  $f(V)$  with  $f(V_{0 \rightarrow 1})$ , etc. We then get decompositions  $f(V) = P_{01}f(V) + (1 - P_{01})f(V)$  into 01-periodic and 01-mixing components, etc.

What makes life difficult here is that the three orthogonal projections  $P_{01}, P_{12}, P_{20}$  can be in general position with respect to each other (in particular, they need not commute). But things are simpler when one of the orthogonal projections is trivial.

At one extreme, consider the case when  $P_{12}$  (say) is the identity - in other words, that *every* random variable  $f(V)$  is 12-periodic. This is basically saying that the original set  $A$  is essentially insensitive to flipping a 1 to 2 or vice versa. Because of this, we can basically identify 1 and 2 together and reduce the  $k=3$  problem to the  $k=2$  problem.

A more non-trivial case is the other extreme, in which  $P_{12}$  is the expectation operator - i.e. every random variable  $f(V)$  of expectation zero is 12-periodic, or equivalently,  $f(V)$  and  $g(V_{\{1 \rightarrow 2\}})$  are independent for every  $f, g$ .

Now let  $f(V), g(V), h(V)$  be random variables. We claim that if  $f$  is 01-mixing OR if  $g$  is 20-mixing, then

$$\mathbb{E} h(V) f(V_{0 \rightarrow 1}) g(V_{0 \rightarrow 2}) \approx 0, \quad (1)$$

where  $V_{0 \rightarrow 2}$  is defined by taking the *same* digit that was flipped from 0 to 1 in the definition of  $V_{0 \rightarrow 1}$ , and flipping it to 2 instead.

To show this, it suffices (by a certain “IP van der Corput lemma”, which one should think of as a variant of Cauchy-Schwarz) to eliminate  $h$  and show that

$$\mathbb{E}f(V_{00\rightarrow 01})g(V_{00\rightarrow 02})f(V_{00\rightarrow 11})g(V_{00\rightarrow 22}) \approx 0 \quad (2)$$

where the  $V_{00\rightarrow ij}$  are formed by choosing two of the fixed 0s of  $V$  and flipping them to  $i$  and  $j$  (and it is the same two digits being flipped for all four of the  $V_{00\rightarrow ij}$ ).

The triviality of  $P_{12}$  (or more precisely, a variant of this with  $m$  replaced by  $m+1$ ) can then be used to show that the random variables  $f(V_{00\rightarrow 01})f(V_{00\rightarrow 11})$  and  $g(V_{00\rightarrow 02})g(V_{00\rightarrow 22})$ , because these random variables involve parallel  $m+1$ -dimensional subspaces which differ from each other by flipping a single 1 to a 2. Assuming this, we can simplify the expression in (2) to

$$[\mathbb{E}f(V_{00\rightarrow 01})f(V_{00\rightarrow 11})][\mathbb{E}g(V_{00\rightarrow 02})g(V_{00\rightarrow 22})].$$

But one of these factor vanishes since  $f$  is 01 mixing or  $g$  is 20 mixing, and we are done.

For general  $f, g$ , we can now use the Fact from the previous comment to obtain the “generalised von Neumann theorem”.

$$\mathbb{E}h(V)f(V_{0\rightarrow 1})g(V_{0\rightarrow 2}) \approx \mathbb{E}h(V)P_{01}f(V)P_{02}g(V).$$

Now we can prove DHJ(3). It suffices to show that

$$\mathbb{E}1_{E_0}(V)1_{E_0}(V_{0\rightarrow 1})1_{E_0}(V_{0\rightarrow 2})$$

is large. Using the generalised von Neumann theorem, this expression can be rewritten as

$$\mathbb{E}1_{E_0}(V)P_{01}1_{E_0}(V)P_{20}1_{E_0}(V).$$

But, as we saw in 631,  $P_{01}1_{E_0}$  has positive inner product with  $1_{E_0}$ . Furthermore  $P_{01}1_{E'}$  has positive inner product with  $1_{E'}$  for any subset  $E'$  of  $E_0$  of positive measure. One can check that  $P_{01}$  is monotone, and so  $P_{01}1_{E_0}$  has positive inner product with  $1_{E'}$  for any  $E'$  of positive measure; thus  $P_{01}1_{E_0}$  is positive almost surely on  $E_0$  (indeed,  $P_{01}1_{E_0}$  is the conditional expectation of  $E_0$  to a certain factor), and similarly for  $P_{20}1_{E_0}$ , and we are done.

25 February, 2009 at 11:37 am

**Terence Tao**



633. Thread moving

I have now got an informal combinatorial translation of Paper #2 on the wiki at

[http://michaelnielsen.org/polymath1/index.php?title=Furstenberg-Katznelson\\_argument](http://michaelnielsen.org/polymath1/index.php?title=Furstenberg-Katznelson_argument)

and it is now being discussed on the 800 thread at

<http://gowers.wordpress.com/2009/02/23/brief-review-of-polymath1/>

so I am now moving discussion of this paper over to that thread.

⌂

# Gowers's Weblog

Mathematics related discussions

---

« [To thread or not to thread](#)  
[DHJ\(3\): 851-899](#) »

## Brief review of polymath1

By gowers

I don't have much to say mathematically, or rather I do but now that there is a [wiki associated with polymath1](#), that seems to be the obvious place to summarize the mathematical understanding that arises in the comments on the various blog posts here and over on Terence Tao's blog (see blogroll). The reason I am writing a new post now is simply that the 500s thread is about to run out.

So let me quickly make a few comments on how things seem to be going. (At some point in the future I will do so at much greater length.) Not surprisingly, it seems that we have reached a stage that is noticeably different from how things were right at the beginning. Certain ideas that emerged then have become digested by all the participants and have turned into something like background knowledge. Meanwhile, the discussion itself has become somewhat fragmented, in the sense that various people, or smaller groups of people, are pursuing different approaches and commenting only briefly if at all on other people's approaches. In other words, at the moment the advantage of collaboration is that it is allowing us to do things in parallel, and efficiently because people are likely to be better at thinking about the aspects of the problem that particularly appeal to them.

Whether there will be a problem with lack of communication I don't know. But perhaps there are enough of us that it won't matter. At the moment I feel rather optimistic that we will end up with a new proof of DHJ(3) (but that is partly because I have a sketch that I have not subjected to appropriately stringent testing, which always makes me feel stupidly optimistic). In fact, what I'd really like to see is *several* related new proofs emerging, each drawing on different but overlapping subsets of the ideas that have emerged during the discussion. That would reflect in a nice way the difference between polymath and more usual papers written by monomath or oligomath.

Finally, a quick word on threading. The largest number of votes (at the time of writing) have gone to allowing full threading, but it is not an absolute majority: those who want unrestricted threading are outnumbered by those who have voted either for limited threading or for no threading at all. I think that points to limited threading. I've allowed the minimum non-zero amount. I can't force you to abide by any rules here, but I can at least give my recommendation, which is this. For polymath comments, I'd like to stick as closely as possible to what we've got already. So if you have a genuinely new point to make, then it should come with a new number. However, if you want to give a quick reaction to another comment, then a reply to it is appropriate. If you have a longish reply, then it should appear as a new comment, but here there is another use of threading that could be very helpful, which is to add replies such as, "I have a counterexample to this conjecture — see comment 845 below." In other words, it will allow forward referencing as well as backward referencing. Comments on this post will start at 800, and if yours is the  $n$ th reply to comment 8\*\*, then you should number it 8\*\*.n. Going back to the question of when to reply to a comment, a good rule of thumb is that you should do so only if your reply very much has the feel of a reaction and not of a full comment in itself.

Also, it isn't possible to have threading on some posts and not on others, but I'd be quite grateful if we didn't have threaded comments on any posts earlier than this one. And a quick question: does anyone know what happens to the threaded comments if you turn the threading feature off again, which is something I might find myself wanting to do?

This entry was posted on February 23, 2009 at 10:57 am and is filed under [polymath1](#). You can follow any responses to this entry through the [RSS 2.0](#) feed. You can [leave a response](#), or [trackback](#) from your own site.

## 119 Responses to “Brief review of polymath1”

1.  *Ryan O'Donnell* Says:

[February 23, 2009 at 6:12 pm](#) | [Reply](#)

800. Here is a writeup of a Fourier/density-increment argument for Sperner, implementing Terry's #578 using some of the Fourier calculations Tim and I were doing:

<http://www.cs.cmu.edu/~odonnell/wrong-sperner.pdf>

Hope it's mostly bug-free!

2.  *jozsef* Says:

[February 23, 2009 at 6:14 pm](#) | [Reply](#)

Tim - First of all I have to say that I did enjoy to read the blog and to contribute a bit. On the other hand it became less and less comfortable to me posting any comments. You wrote that

“Meanwhile, the discussion itself has become somewhat fragmented, in the sense that various people, or smaller groups of people, are pursuing different approaches and commenting only briefly if at all on other people's approaches.”

Let me tell you what is my experience. I'm reading posts using notations from ergodic theory or Fourier analysis with great interest, but I will rarely make any comment as I'm not expert on that field. Some post are still accessible to me and some others aren't. There is no way that I will understand posts referring to noise stability unless I learn something more about it, which I'm not planning to do in the near future unless it turns out to be crucial for the project. Maybe the blog is still accessible for the majority of readers but I'm finding it more and more difficult to follow. For the same reasons I feel a bit awkward about posting notes. It's tempting to work on the problems alone without sharing it with the others but this is clearly against the soul of the project.

3.  *gowers* Says:

[February 23, 2009 at 6:28 pm](#) | [Reply](#)

Metacomment. Jozsef, in the light of what you say, I invite others to give their reactions to how things are going for them. My feeling, as I said in the post, is that we are entering uncharted territory (or rather, a different uncharted territory from the initial one) and it is not clear that the same rules should still apply. What we have done so far has, in my view, been a wonderful way of doing very quickly and thoroughly the initial exploration of the problem, and it feels as though the best way of making further progress will be if we start dealing with technicalities and actually write up some statements formally. I am hoping to do that in the near future, Ryan is doing something along those lines, Terry recently put a Hilbert-spaces lemma on the wiki, etc.

The obvious hazard with this is that people may end up writing things that others do not understand, or can understand only if they are prepared to invest a lot of time and effort. So if we want to continue to operate as a collective unit, so to speak, it is extremely important for people who write formal proofs to give generous introductions explaining what they are doing, roughly how the proof goes, why they think it could be useful for DHJ(3) (or something related to DHJ(3)), and so on. This is perhaps an area where people who have been following the discussion without actually posting comments could be helpful—just by letting us know which bits you find hard to understand. For example, if you see something on the wiki that is clearly meant to be easily comprehensible but it in fact isn't, then it would be



helpful to know. (However, some of the wiki articles are over-concise simply because we haven't had time to expand them yet.)

Do others have any thoughts about where we should go from here?

4.  [gowers](#) Says:

[February 23, 2009 at 6:47 pm](#) | [Reply](#)

801. Fourier/Sperner

Ryan, re the remark at the end of the thing you wrote up, if we do indeed have an expression of the form  $\mathbb{E}_{p \leq q} \sum_{A \subset [n]} \lambda_{p,q}^{[A]} \hat{f}_p(A) \hat{f}_q(A)$  for the total weight  $\mathbb{E}_{U \subset V} f(U) f(V)$  for  $\mathbb{E}_{U \subset V} f(U) f(V)$  when everything is according to equal-slices density, then it seems to me not completely obvious that one couldn't prove some kind of positivity result for the contribution from each  $A$ , especially given your calculation that  $\lambda_{p,q}$  splits up as a product. But probably you've thought about that—I can see that it's a problem that  $\lambda_{p,q}$  is not symmetric in  $p$  and  $q$ , but can one not make it symmetric by talking instead about pairs of disjoint sets?

Or perhaps it's false that you have positivity and the disjoint-pairs formulation is what you need to get a similar result where it becomes true. I hope this vague comment makes some sense to you. I'll see if I can think of a counterexample. In fact, isn't this a counterexample: let  $n=1$ , and let  $f(0)=-1$ ,  $f(1)=1$ . Then the expectation of  $f(x)f(y)$  over combinatorial lines  $(x,y)$  is  $-1$  and the expectation of  $f$  is  $0$ .

- o  [gowers](#) Says:

[February 23, 2009 at 7:25 pm](#)

801.1 Please ignore the last paragraph of this comment — for analytic purposes we need to consider degenerate lines, and as you say the usual proof of Sperner proves positivity. Hmm — not sure that the threading adds much here.

- o  [Ryan O'Donnell](#) Says:

[February 23, 2009 at 8:05 pm](#)

801.2: Yes, it's not clear to me why  $\mathbb{E}_{p \leq q} [\hat{f}_p(A) \hat{f}_q(A)]$  needs to be nonnegative. This is a shame, because: a) we'd be extremely done if we could show this; b) as  $p, q \rightarrow 1/2$  the quantity approaches  $\hat{f}(A)^2$ , which is of course nonnegative. Unfortunately, if you make  $p, q$  extremely close to  $1/2$  so as to force it nonnegativity, you'll probably be swamped with degenerate lines.

5.  [Ryan O'Donnell](#) Says:

[February 23, 2009 at 8:20 pm](#) | [Reply](#)

Metacomment: I sympathise with Jozsef's position. Luckily, some of the directions we're working on are in areas I'm familiar with, so I can comment there. But to be honest, I know nothing about ergodic methods, am extremely shaky on what triangle-removal is, and am only kind of on top of Szemerédi's regularity lemma. So indeed it takes me quite a while just to read comments on these topics. That's fine with me though.

Thing is, even in the areas I'm familiar with it's awfully tough to keep up with a certain pair of powerhouses who post regularly to the project 😊 In a perfect world I'd post more definitions, add more intuition to my comments, update the wiki with stuff about noise sensitivity and so forth — but it already takes me many hours just to quasi-keep up with Tim and Terry and generate non-nonsensical comments. I'm fine with this too, though.

Finally, I agree that it's also gotten to the point where it's sort of impossible to explore certain directions — especially

ones with calculations — “online”. You probably won’t be surprised to learn I spent a fair bit of time working out the pdf document in #800 by myself on paper. I hope this isn’t too much against the spirit of the project, but I couldn’t find any way to do it otherwise. I would feel like I were massively spamming people if I tried to compute like this online, with all the associated wrong turns, miscalculations, and crazy mistakes I make.



6. *Ryan O'Donnell Says:*

[February 23, 2009 at 8:25 pm](#) | [Reply](#)

802. Sperner/Fourier.

By the way, I’m pretty sure one can also do Roth’s Theorem (or at least, finding “Schur triples” in  $\mathbb{F}_2^n$ ) in this way. It might sound ridiculous to say so, since Roth/Meshulam already gave a highly elegant density-increment/Fourier-analysis proof.

But the point is that:

- . it works in a non-”arithmetic”/”algebraic” way
- . it works by doing density increments that restrict just to \*combinatorial\* subspaces
- . it demonstrates that the method can sometimes work in cases where one needs to find \*three\* points in the set satisfying some relation.

I can supply more details later but right now I have to go to a meeting...



7. *Gil Says:*

[February 23, 2009 at 8:54 pm](#) | [Reply](#)

Metacomment.1

(If this .1 will automatically create a threading I will consider it a miracle.)

I think the nature of this collaboration via different people writing blog remarks is somewhat similar to the massive collaboration in ordinary math where people are writing papers partially based on earlier papers.

Here the time scale is quicker and the nature of contributions is more tentative. Half baked ideas are encouraged.

It is not possible to follow everything or even most of the things or even a large fraction of them. (Not to speak about understanding and digesting and remembering). I think it can be useful that people will not assume that other people read earlier things and from time to time repeat and summarize and overall be clear as possible. (I think comments of the form: “please clarify” can be useful.) On the other hand, like in usual collaborations people often write things mainly for themselves (e.g. the Hardy-Littlewood rules) and this is very fine.

(A question: is Ryan’s write-up a correct new proof of (weak) Sperner using density increment and Fourier? (The file is called wrong-Sperner but reading it I realize that this means the proof is right but not yet the “Fourier/Sperner proof from the book” or something like that. Please clarify.)

There are various avenues which I find very interesting even if not directly related to the most promising avenues towards DHJ proof and even some ideas I’d like to share and explore. (One of these is a non-density-decreasing Fourier strategy.)

One general avenue is various reductions (usually based on simple combinatorial reasonings) : Jozsef mentioned Moser( $k=6$ )-> DHK( $k=3$ ) and some reductions to Szemerédi ( $k=4$ ) were considered. Summarizing those and pushing for further reductions can be fruitful.

Also the equivalence of DHJ for various measures (I suppose this is also a reduction business) is an interesting aspect

worth further study. When there is group acting such equivalences are easy and of general nature. But for DHJ I am not sure how general they are.



o [Ryan O'Donnell](#) Says:

[February 23, 2009 at 11:32 pm](#)

Gil wrote: “(The file is called wrong-Sperner but reading it I realize that this means the proof is right but not yet the “Fourier/Sperner proof from the book” or something like that. Please clarify.)”

Yes — I forgot to mention that. I think/hope the writeup's correct, but I called it “wrong-Sperner” because I still feel there's something not-from-the-Book about it.



8. [Terence Tao](#) Says:

[February 23, 2009 at 10:07 pm](#) | [Reply](#)

Metacomment.2

It certainly does feel that the project is developing into a more mature phase, where it resembles the activity of a highly specialised mathematical subfield, rather than a single large collaboration, albeit at very different time scales than traditional activity. (The wiki is somewhat analogous to a “Journal of Attempts to Prove Density Hales Jewett”, and the threads are analogous to things like “4th weekly conference on Density Hales Jewett” 😊.) Also we are hitting the barrier that the number of promising avenues of research seems to exceed the number of people actively working on the project.

But I think things will focus a bit more (and become more “polymathy”) once we identify a particularly promising approach to the problem (I think we already have a partially assembled skeleton of such, and a significant fraction of the tools needed have already been sensed). This is what is going on in the 700 thread, by the way; we are focusing largely on one subproblem at a time (right now, it's getting a good Moser(3) bound for  $n=5$ ) and we seem to be using the collaborative environment quite efficiently.

My near-term plan is to digest enough of the ergodic theory proofs that I can communicate in finitary combinatorial language a formal proof of DHJ(2) that follows the ergodic approach, and a very handwavy proof of DHJ(3). (The Hilbert space lemma I'm working on is a component of the DHJ(2) analysis.) The finitisation of DHJ(2) looks doable, but is already quite messy (more “wrong” than Ryan's “wrong” proof of DHJ(2)), and it seems to me that formally finitising the ergodic proof of DHJ(3), while technically possible, is not the most desirable objective here. But there does seem to be some useful ideas that we should be able to salvage from the ergodic proof that I would like to toss out here once I understand them properly. (For instance, there seems to be an “IP van der Corput lemma” that lets one “win” when one has 01-pseudorandomness (which roughly means that if one takes a medium dimensional slice of  $A$  and then flips one of the fixed digits from 0 to 1, then the pattern of  $A$  on the shifted slice is independent of the pattern of  $A$  on the original slice). I would like to understand this lemma better. The other extreme, that of 01-insensitivity, is tractable by Shelah's trick of identifying 0 and 1 into a single letter of the alphabet, and the remaining task is to apply a suitable structure theorem to partition arbitrary sets into 01-structured and 01-pseudorandom components, analogously to how one would apply the regularity lemma to one part of the tripartite graph needed to locate triangles.)



9. [gowers](#) Says:

[February 23, 2009 at 10:34 pm](#) | [Reply](#)

803. Sperner/Fourier


Re the discussion in 801, this is a proposal for getting positivity. I haven't checked it, and it could just obviously not work.

The reason I make it is that I just can't believe the evidence in front of me: the trivial positivity in "physical space" just *must* be reflected by equally trivial positivity in "frequency space". I wonder if the problem, looked at from the random-permutation point of view, is that to count pairs of initial segments it is not completely sensible to count them with the smaller one always first — after all, the proof of positivity goes by saying that that is actually half the set of pairs of initial segments. So maybe we should look at a sum of the form  $\mathbb{E}_{A:B} f(A)f(B)$ , where  $A:B$  is a hastily invented notation for " $A \subset B$  or  $B \subset A$ ."

How could we implement this Ryan style? Well, obviously we would average over all pairs of probabilities  $(p,q)$ . And if we choose a particular pair, then the obvious thing to do would be to choose independent random variables  $(X_1, \dots, X_n)$  and  $(Y_1, \dots, Y_n)$  in such a way that the resulting set will give you a probability  $p$  of being in  $A$ , a probability  $q$  of being in  $B$ , and a certainty that  $A \subset B$ . To achieve that in a unified way, for each  $i$  we choose a random  $t \in [0, 1]$  and set  $X_i = 1$  if and only if  $t \leq p$  and  $Y_i = 1$  if and only if  $t \leq q$ .

I haven't even begun to try any calculations here, but I would find it very strange if one didn't get some kind of positivity over on the Fourier side.

Having said that, the fact that one expands  $f$  in a different way for each  $p$  does make me not quite as sure of this positivity as I'm pretending to be.

- o  Ryan O'Donnell Says:  
[February 23, 2009 at 11:43 pm](#)


#### 803.1 Positivity:

In my experience, obvious positivity on the physical side doesn't always imply obvious positivity on the Fourier side. Here is an example (although it goes in the reverse direction).

Let  $f : \{0, 1\}^n \rightarrow \mathbb{R}$  be arbitrary and consider  $\mathbb{E}[f(x)f(y)]$ , where  $x$  is uniform random and  $y$  is formed by flipping each bit of  $x$  with probability  $1/5$ . (Note that  $(x, y)$  has the same distribution as  $(y, x)$ , so there's no asymmetry here.)


Now  $x$  and  $y$  are typically quite far apart in Hamming distance, and since  $f$  can have positive and negative values, big and small, it seems far from obvious that  $\mathbb{E}[f(x)f(y)]$  should be nonnegative.

But this is obvious on the Fourier side: it's precisely  $\sum_S (3/5)^{|S|} \hat{f}(S)^2$ .

- o  gowers Says:  
[February 24, 2009 at 12:05 am](#)


#### 803.2 Positivity

That's interesting. The non-obvious assertion is that a certain matrix (where the value at  $(x,y)$  is the probability that you get  $y$  when you start from  $x$  and flip) is positive definite. And the Fourier transform diagonalizes it. And that helps to see why it's genuinely easier on the Fourier side: given a quadratic form, there is no reason to expect it to be easy to see that it is positive definite without diagonalizing it. (However, I am tempted to try with your example ...)

- o  gowers Says:  
[February 24, 2009 at 11:52 am](#)

#### 803.3 Positivity

Got it! Details below in an hour's time (after my lecture). It will be comment 812 if nobody else has commented by then.


10.  [Terence Tao](#) Says:  
[February 23, 2009 at 11:15 pm](#) | [Reply](#)

804. DHJ(2.7)


Randall: I tried to wikify your notes from 593 at

[http://michaelnielsen.org/polymath1/index.php?title=DHJ\(2.7\)](http://michaelnielsen.org/polymath1/index.php?title=DHJ(2.7))

Unfortunately due to a wordpress error, an important portion of your Tex (in particular, part of the **statement** of DHJ(2.7)) was missing (the portion between a less than sign and a greater than sign - unfortunately this was interpreted as HTML and thus eaten). So I unfortunately don't understand the proof. Could you possibly take a look at the page and try to restore the missing portion? Thanks!

- o  [Randall](#) Says:  
[February 24, 2009 at 5:04 am](#)

Sorry about that. After several missteps, I may have managed to correct it in full....

11.  [Terence Tao](#) Says:  
[February 23, 2009 at 11:22 pm](#) | [Reply](#)


805. Sperner

(Previous comment should be 804.)

Ryan's notes (which, incidentally, might also be migrated to the wiki at some point) seem to imply something close to my Lemma in 578 that had a bogus proof, namely that if  $f$  is non-uniform in the sense that  $\mathbb{E}f(x)g(y)$  or  $\mathbb{E}g(x)f(y)$  is large for some  $g$ , then  $f$  has significant Fourier concentration at low modes (i.e. the energy  $\sum_{|S| \leq \varepsilon n} |\hat{f}(S)|^2$  is large), and hence  $f$  correlates with a function of low influence (e.g.  $f$  correlates with  $T_\varepsilon f$ , where  $T_\varepsilon$  is the Fourier multiplier that multiplies the  $S$  Fourier coefficient by  $(1 - 1/\varepsilon n)^{|S|}$ ).

If this is the case, then (because polynomial combinations of low influence functions are still low influence) one can iterate this in the usual energy-increment manner (or as in the proof of the regularity lemma) to obtain a decomposition  $1_A = f_{U^\perp} + f_U$ , where  $f_U$  is uniform and  $f_{U^\perp}$  has low influence, is non-negative and has the same density as  $1_A$ . (This is an oversimplification; there are error terms, and one has to specify the scales at which uniformity or low influence occurs, but ignore these issues for now.)

If this is the case, then the Sperner count  $\mathbb{E}1_A(x)1_A(y)$  should be approximately equal to  $\mathbb{E}f_{U^\perp}(x)f_{U^\perp}(y)$ . But if  $f_{U^\perp}$  is low influence, this should be approximately  $\mathbb{E}f_{U^\perp}(x)^2$ . Meanwhile, we have  $\mathbb{E}f_{U^\perp}(x) = \delta$ , so by Cauchy-Schwartz we get at least  $\delta^2$ , as desired.

- o  [Terence Tao](#) Says:  
[February 23, 2009 at 11:24 pm](#)

805.1

p.s. Something very similar goes on in the ergodic proof of DHJ(2). What I call  $f_{U^\perp}$  here will be called  $P1_A$ , where  $P$  is a certain "weak limit" of shift operators. The key point is that  $P$  is going to be an orthogonal

projection in  $L^2$  (this is the upshot of a Hilbert space lemma that I am working on on the wiki.)

12.  *Ryan O'Donnell* Says:

[February 24, 2009 at 12:01 am](#) | [Reply](#)

806. Sperner.

Re #805, Terry, one thing I had trouble with was — roughly speaking — showing that indeed  $\mathbb{E}[f_{U^\perp}(x)f_{U^\perp}(y)]$  is close to  $\mathbb{E}[f_{U^\perp}(x)^2]$ .

It seemed logical that this would be true but I had technical difficulties actually making the switch from  $y$  to  $x$ . In some sense that's why I wrote the notes, to really convince myself that one could do something. Unfortunately, that something was to define  $f_{U^\perp}$  all the way down to just the lowest mode; i.e., to look at density rather than energy.

Perhaps we could try to think about, at a technical level, how to really pass from  $(x, y)$  to  $(x, y)$  for low-influence functions...

o  *Terence Tao* Says:

[February 24, 2009 at 1:49 am](#)

806.2 Sperner

Well, I would imagine that  $f_{U^\perp}$  would have Fourier transform concentrated in sets  $S$  of size  $\epsilon n$ , and this should mean that the average influence of  $f_{U^\perp}$ , i.e.  $\mathbb{E} |f_{U^\perp}(x) - f_{U^\perp}(y)|^2$ , where  $x$  and  $y$  differ by just one bit, should be small (something like  $O(\epsilon)$ ). This should then be able to be iterated to extend to  $x$  and  $y$  being several bits apart rather than just one bit. (In order to do this properly, one will probably need to have a whole bunch of different scales in play, not just a single scale  $\epsilon n$ . There are various pigeonhole tricks that let one find a whole range of scales at which all statistics are the same [I call this the "finite convergence principle" in my blog], so as a first approximation let's pretend that concentration at one scale  $\epsilon n$  is the same as concentration at other small scales such as  $\epsilon^2 n$ . The point is that while the  $L^2$  energy of  $f$  could be significant in the range  $\epsilon^2 n \leq |S| \leq \epsilon n$  for a single  $\epsilon$ , the pigeonhole principle prevents it from being significant for all such choices of  $\epsilon$ .)

13.  *Ryan O'Donnell* Says:

[February 24, 2009 at 12:08 am](#) | [Reply](#)

807. Fourier/density-increment.

In #802, I guess I'm getting the terminology all wrong. I should say that another problem I think the same method will work for is the problem in  $\mathbb{F}_2^n$  where you're looking for "lines" of length 3 where in the wildcard coordinates you're allowed either (0,1,1), (1,0,1), or (1,1,0).

I don't know if this problem has a name. It's not quite a Schur triple. I personally would call it the "Not-Two" problem because in every coordinate you're allowed (0,0,0), (1,1,1), (0,1,1), (1,0,1), or (1,1,0): anything where the number of 0's is "Not Two".

o  *Gil* Says:

[February 24, 2009 at 12:59 am](#)

807.1

A question and 3 remarks:



Q: I remember when we first talked about it there were obvious difficulties for why Fourier+increasing density strategy won't work. At the end what is the basic change that makes it work? Is it changing the measure?

R1: sometimes it is equally convenient (and even more) not to work with  $\mu_p$  measures and their associated Fourier transform but to consider functions on  $[0,1]^n$  and use the Walsh orthonormal basis even for every  $[0,1]$ . So given  $p$  we represent our set by a subset of the solid cube and use the "fine" Walsh transform all the time.

R2: If changing the measure can help when small Fourier coefficients do not suffice for quasirandomness, is there hope it will help for case  $k=4$  of Szemerédi?

R3: there are some problems regarding influences threshold phenomena where better understanding of the relation between equal sliced density and specific  $\mu_p$  densities are needed. Maybe some of the tools from here can help.

14.  [Gil](#) Says:

[February 24, 2009 at 1:13 am](#) | [Reply](#)

808. Another Fourier/Sperner approach

Let me mention briefly a non density-increasing approach to Sperner. You first note that not having lines with 1 wild card implies that  $\sum \hat{f}^2(S)|S| = \hat{f}^2(S)(n - |S|)$ . If you can show that the non empty Fourier coefficients are concentrated (in terms of  $|S|$ ) then I think you can conclude that the density is small.

You want to control expression of the form  $\hat{f}^2(S) \binom{n-|S|}{k}$ . Those are related to the number of elements in the set when we fixed the value of  $n-k$  variables and look over all the rest. These numbers are controlled by the density of Sperner families for subsets of  $\{1,2,\dots,k\}$ . So this looks like circular thing but maybe it allows some bootstrapping.

15.  [Ryan O'Donnell](#) Says:

[February 24, 2009 at 5:42 am](#) | [Reply](#)

809. Response to 806.2 (Sperner).

Terry wrote, "This should then be able to be iterated to extend to  $x$  and  $y$  being several bits apart rather than just one bit."

Hmm, but one source of annoyance is that  $y$  is not just  $x$  with  $\epsilon n$  or so bits flipped — it's  $x$  with  $\epsilon n$  or so 0's flipped to 1's. As you get more and more 1's in there, the intermediate strings are less and less distributed as uniformly random strings. So it's not 100% clear to me that you can keep appealing to average influence, since influence as defined is a uniform-distribution concept.

Overall I agree that this is probably more of a nuisance than a genuine problem, but it was stymieing me a bit so I thought I'd ask.

o  [Terence Tao](#) Says:

[February 24, 2009 at 6:07 am](#)

809.2 Sperner and influence


Ryan, I think the trusty old triangle inequality should deal with things nicely. (I would need a binomial model of bit-flipping rather than a Poisson model, but I'm sure this makes very little difference.)

Suppose  $f_{U^\perp}$  has average influence  $\mu$ ; thus flipping a bit from a 0 to 1 or vice versa would only affect things by  $\mu$  on the average. Conditioning, we conclude that if we flip just one bit from a 0 to 1, we expect  $f_{U^\perp}$  to change by at most  $2\mu$  on the average. Now we iterate this process  $\epsilon n$  times. As long as  $\epsilon n$  is much less than  $\sqrt{n}$  (as in your

notes), there is no significant distortion of the underlying probability distribution and we see that if  $y$  differs by  $\varepsilon n$  bits from  $x$ , then  $f_{U^\perp}(y)$  and  $f_{U^\perp}(x)$  will differ by  $O(\varepsilon n \mu)$  on the average.

This will be enough for our purposes as long as the influence  $\mu$  is a really small multiple of  $1/\varepsilon n$ . One can't quite get this by working at a single scale - one gets  $O(1/\varepsilon n)$  instead - but one can do so if one works with a whole range of scales and uses the energy increment argument. I've sketched out the details on the wiki at

[http://michaelnielsen.org/polymath1/index.php?title=Fourier-analytic\\_proof\\_of\\_Sperner](http://michaelnielsen.org/polymath1/index.php?title=Fourier-analytic_proof_of_Sperner)

16.  [Terence Tao](#) Says:  
[February 24, 2009 at 6:27 am](#) | [Reply](#)

810. DHJ(2.7)

Randall, thanks for cleaning up the file! It is a very nice proof, and I'd like to try to describe it in my own words here.

DHJ(2.7) is the strengthening of DHJ(2.6) that gives us three *parallel* combinatorial lines (i.e. they have the same wildcard set), the first of which hits the dense set  $A$  in the 0 and 1 position, the second hits it in the 1 and 2 position, and the third in the 2 and 0 position.


To describe Randall's argument, I'd first like to describe how Randall's argument gives yet another proof of DHJ(2) which is quite simple (and gives civilised bounds). It uses the density increment argument: we want to prove DHJ(2) at some density  $\delta$  and we assume that we've already proven it at any significantly bigger density.

Now let  $A$  be a subset of  $[2]^n$  of density  $\delta$  for some  $n$ . We split  $n$  into a smallish  $r$  and a biggish  $n-r$ , thus viewing  $[2]^n$  as a whole bunch of  $(n-r)$ -dimensional slices, each indexed by a word in  $[2]^r$ .

If any of the big slices has density significantly bigger than  $\delta$ , we are done; so we can assume that all the big slices have density not much larger than  $\delta$  (e.g. at most  $\delta + \delta^2/2^r$ ). Because the total density is  $\delta$ , we can subtract and conclude that all the big slices have density close to  $\delta$ .

Now we look at the  $r+1$  slices indexed by the words  $0^r, 0^{r-1}1, 0^{r-2}1^2, \dots, 01^{r-1}, 1^r$ . Each of these slices intersects the set  $A$  with density about  $\delta$ . Thus by the pigeonhole principle, if  $r$  is much bigger than  $1/\delta$ , then two of the  $A$ -slices must share a point in common, i.e. there exists  $w \in [2]^{n-r}$  and  $i, j$  such that  $0^i 1^{r-i} w, 0^j 1^{r-j} w$  both lie in  $A$ . Voila, a two-dimensional line.

The same argument gives DHJ(2.7). Now there are three different graphs on  $r+1$  vertices, the 01 graph, the 12 graph, and the 20 graph.  $i$  and  $j$  are connected on the 01 graph if the  $0^i 1^{r-i}$ -slice and  $0^j 1^{r-j}$ -slice of  $A$  have a point in common; similarly define the 12 graph and the 20 graph. Together, this forms an 8-coloured graph on  $r+1$  vertices. By Ramsey's theorem, if  $r$  is big enough there is a monochromatic subgraph of size  $\gg 1/\delta$ . But by the preceding pigeonhole argument, we see that none of the 01 graph, the 12 graph, or the 20 graph can vanish completely here, so they must instead all be complete, and we get three parallel combinatorial lines, each intersecting  $A$  in two of the three positions.

17.  [ryanworldwide](#) Says:  
[February 24, 2009 at 6:53 am](#) | [Reply](#)

811. Prelude to some more density-increment arguments.

I was hoping to give an illustration of the #578 technique for a problem in  $[3]^n$ . But it got too late for me to finish writing it, so I'll just give the "background info" I managed to write. This probably ought to go in the wiki rather than in a post, but having spent a bit of time to get Luca's converter working for me, I didn't have the energy to also convert to the third, wiki format. More on the #578 method tomorrow.

Here, then, some basics on noise and correlated product spaces; for more, see e.g. [this paper of Mossel](#).

Let  $\Omega$  be a small finite set; for example,  $\Omega = [3]$ . Let  $\mu$  be a probability distribution on  $\Omega$ . Abusing notation, write also  $\mu = \mu^{\otimes n}$  for the corresponding product distribution on  $\Omega^n$ . For  $0 \leq \rho \leq 1$ , define the *noise operator*  $T_\rho^\mu$ , which acts on functions  $\Omega^n \rightarrow \mathbb{R}$ , as follows:  $(T_\rho^\mu f)(x) = \mathbf{E}_{\mathbf{y} \sim_\rho^\mu x} [f(\mathbf{y})]$ ,

where  $\mathbf{y} \sim_\rho^\mu x$  denotes that  $\mathbf{y}$  is formed from  $x$  by doing the following, independently for each  $i \in [n]$ : with probability  $\rho$ , set  $y_i = x_i$ ; with probability  $1 - \rho$ , draw  $y_i$  from  $\mu$  (i.e., “rerandomize the coordinate”). (I use **boldface** for random variables.)

Here are some simple facts about the noise operator:

Fact 1:  $\mathbf{E}_{\mathbf{x} \sim \mu} [T_\rho^\mu f(\mathbf{x})] = \mathbf{E}_{\mathbf{x} \sim \mu} [f(\mathbf{x})]$ .

Fact 2: For  $p \geq 1$  we have  $\|T_\rho^\mu f\|_{\mu, p} \leq \|f\|_{\mu, p}$ , where  $\|f\|_{\mu, p}$  denotes  $\mathbf{E}_{\mathbf{x} \sim \mu} [|f(\mathbf{x})|^p]^{1/p}$ .

Fact 3: The  $T_\rho^\mu$ 's form a semigroup, in the sense that  $T_{\rho_1}^\mu T_{\rho_2}^\mu = T_{\rho_1 \rho_2}^\mu$ .

Fact 4:  $\mathbf{E}_{\mathbf{x} \sim \mu} [f(\mathbf{x}) \cdot T_\rho^\mu f(\mathbf{x})] = \mathbf{E}_{\mathbf{x} \sim \mu} [(T_{\sqrt{\rho}}^\mu f(\mathbf{x}))^2]$ . We use the notation  $\mathbb{S}_\rho^\mu(f)$  for this quantity.

Fact 5:  $\mathbb{S}_{1-\gamma}^\mu(f) = \mathbf{E}_\mu [(T_{\sqrt{1-\gamma}}^\mu f)^2] = \mathbf{E}_\mathbf{V} [\mathbf{E}_\mu [f|\mathbf{V}]^2]$ , where: a)  $\mathbf{V}$  is a randomly chosen combinatorial subspace, with each coordinate being fixed (from  $\mu$ ) with probability  $1 - \gamma$  and left “free” with probability  $\gamma$ , independently; b)  $f|_\mathbf{V}$  denotes the restriction of  $f$  to this subspace, and  $\mathbf{E}_\mu [f|\mathbf{V}]$  denotes the mean of this restricted function, under  $\mu$  (the product distribution on the free coordinates of  $\mathbf{V}$ ).

18.  [gowers](#) Says:

[February 24, 2009 at 1:32 pm](#) | [Reply](#)

## 812. Positivity

This concerns Ryan's example of a quantity that is easily seen to be non-negative on the Fourier side, and not so easily seen to be non-negative in physical space. Except that I now have an easy argument in physical space. I don't know how relevant this is, but there is some potential motivation for it, which is that perhaps the simple idea behind the proof could help in Ryan's quest for a “non-wrong” proof of Sperner.

The problem, laid out in 803.1, was this. Let  $f$  be a function defined on  $\{0, 1\}^n$ . Now choose two random points  $x, y$  in  $\{0, 1\}^n$  as follows:  $x$  is uniform, and  $y$  is obtained from  $x$  by changing each bit of  $x$  independently with probability  $1/5$ . Now show that  $\mathbf{E} f(x)f(y)$  is non-negative for all real functions  $f$ .


There are various ways of describing a proof of this. Here is the one I first thought of, and after that I'll give a slicker version. I'm going to do a different calculation. Again I'll pick a random  $x$ , but this time I'll create *two* new vectors  $y$  and  $z$  out of  $x$ , which will be independent of each other (given  $x$ ) and obtained from  $x$  by flipping coordinates with probability  $\lambda$ . And I'll choose  $\lambda$  so that  $2\lambda(1 - \lambda) = 1/5$ , which guarantees that the probability that the  $i$ th coordinates of  $y$  and  $z$  are equal is  $1/5$ . Then I'll work out  $\mathbf{E} f(y)f(z)$ . It's obviously positive because it is equal to  $\mathbf{E}_x (\mathbf{E} [f(y)f(z)|x])^2$ . But also, the joint distribution of  $y$  and  $z$  is equal to the joint distribution of  $x$  and  $y$ , since  $y$  is uniform and the digits of  $z$  are obtained from  $y$  by flipping independently with probability  $1/5$ . (Proof: the events  $y_i = z_i$  are independent and have probability  $4/5$ .)

Now for the second way of seeing it. This time I'll generate  $y$  from  $x$  in two stages. First I'll flip each digit with probability  $\lambda$ . And then I'll do that again. Now each time you flip with probability  $\lambda$  you are multiplying  $f$  by the symmetric matrix  $A$  defined by  $A_\lambda(x, y) = \binom{n}{|x-y|} \lambda^{|x-y|} (1 - \lambda)^{n-|x-y|}$  (which is the same as convolving  $f$  by the function  $w(z) = \binom{n}{|z|} \lambda^{|z|} (1 - \lambda)^{n-|z|}$ .) So we end up with  $\langle f, A_\lambda^2 f \rangle$ , which, since  $A_\lambda$  is symmetric, is  $\langle A_\lambda f, A_\lambda f \rangle$ .


Note that this proof fails if you flip with probability  $p$  greater than  $1/2$ , because then you can't solve the equation

$2\lambda(1 - \lambda) = p$ . But that's all to the good, because the result is false when  $p > 1/2$ . When  $p = 1/2$  it is "only just true" since flipping with probability  $1/2$  gives you the uniform distribution for  $y$ , so if  $f$  averages zero then  $\mathbb{E}f(x)f(y)$  is zero too.


I'm sure there's yet another way of looking at it (and in fact these thoughts led to the above proof) which is to think of the random flipping operation as the exponential of an operation where you flip with a different probability. (To define that, I would flip  $N$  times, each time with probability  $c/N$ , and take limits.) I would then expect some general nonsense about exponentials to give the positivity. In the end I have taken a square root instead.

o  ryanworldwide Says:  
[February 24, 2009 at 4:06 pm](#)

812.1. Thanks Tim! I believe this proof is in fact Fact 4 from #811 (with its  $\rho = 3/5$ ).

o  ryanworldwide Says:  
[February 24, 2009 at 4:27 pm](#)

812.2. I was pleased with myself for figuring out the analogous trick in our equal-slices setting... until I realized it was exactly what you said way back in #572, first paragraph. Drat.

19.  Sune Kristian Jakobsen Says:  
[February 24, 2009 at 1:50 pm](#) | [Reply](#)

813. An optimistic conjecture.

The hyper-optimistic conjecture says that  $c_n\mu \leq \bar{c}_n\mu$ . Here I would like to suggest an "optimistic conjecture": There exist a number  $C$  such that  $c_n\mu \leq C\bar{c}_n\mu$  for all  $n$ .


The hyper-optimistic conjecture implies the optimistic conjecture and the optimistic conjecture implies DHJ.

Let  $A_\epsilon$  be the set of elements  $x$  in  $A$ , such that  $A$  has measure at least  $\epsilon$  in the slice  $x$  belong to.

We know that if  $A$  is a line-free set, the sum of the measure of  $A$  in the slices  $\Gamma_{a+r,b,c}, \Gamma_{a,b+r,c}, \Gamma_{a,b,c+r}$  is at most 2. In particular  $A$  cannot contain more than  $2/3$  of each of three slices in an equilateral triangle. Thus, the slices where  $A$  has a density greater than  $2/3$  forms a triangle-free subset of  $\Delta_n := \{(a, b, c) \in \mathbb{Z}_+^3 : a + b + c = n\}$ . So, if the optimistic conjecture is false, we know that for biggest line-free set  $A$ , the density of  $A_{2/3+\epsilon}$  in  $A$  go to 0 as  $n \rightarrow \infty$ .

I think the following would imply DHJ:

For every  $\epsilon$  there exist a  $C$  such that for every set  $A$  with  $A = A_\epsilon$  (no element in  $A$  is "epsilon-lonesome" in its slice) there exist a equilateral triangle-free set  $D \subset \Delta_n$  with more than  $\mu(A)/C$  elements.


20.  ryanworldwide Says:  
[February 24, 2009 at 4:11 pm](#) | [Reply](#)

814. Re Sperner & Influences & 809.2.


Hi Terry, not to be ridiculously nitpicky, but could I ask one more question?

It seems to me that:

- $f_{U^\perp}$  might end up having range  $[0, 1]$  rather than  $\{0, 1\}$ ;
- average influence is defined in terms of the \*squared\* change in  $f_{U^\perp}$ ;
- we don't have a triangle inequality for  $\ell_2^2$ . (It wouldn't be a problem if  $f_{U^\perp}$  had range  $\{0, 1\}$  since  $\ell_2^2$  is a metric on  $\{0, 1\}$ .)


- 
[ryanworldwide](#) Says:  
[February 24, 2009 at 4:33 pm](#)

814.1. Maybe one gets around it by playing with the scales & using triangle inequality for  $\ell_2$ ; I'm thinking about it...

- 
[ryanworldwide](#) Says:  
[February 24, 2009 at 5:10 pm](#)

814.2. Oops.

That was a dumb question, as it turns out. Assuming  $f_{U^\perp}$  is bounded, squared-differences are bounded by 4 times absolute-value-differences, and then use the triangle inequality. Got it.

21.  [ryanworldwide](#) Says:  
[February 24, 2009 at 5:56 pm](#) | [Reply](#)

815. Re Sperner.

I think I can answer my own question from #814. I'll say what Terry was saying in #809.2, being a little bit imprecise. Let  $T \ll \sqrt{n}$  be an integer. Define random variables  $x_0, \dots, x_T$  as follows:  $x_0$  is a uniformly random string;  $x_t$  is formed from  $x_{t-1}$  by picking a random coordinate and, if that coordinate is 0 in  $x_{t-1}$ , changing it to a 1. Now the distribution on  $(x_0, x_T)$  is pretty much like the distribution on  $(x, y)$  from before, with  $\epsilon = T/n$ . Probably, as Terry says, one should ultimately wrap a Poisson choice of  $T$  around this entire process.

Anyway, for fixed  $T$ , by telescoping we have

$$\mathbf{E}[f(x_0)g(x_T)] = \mathbf{E}[f(x_0)g(x_0)] + \sum_{t=1}^T \mathbf{E}[f(x_0)(g(x_t) - g(x_{t-1}))]. \quad (1)$$

To bound the “error term” here, use Cauchy-Schwarz on each summand. For a given  $1 \leq t \leq T$  we have

$$|\mathbf{E}[f(x_0)(g(x_t) - g(x_{t-1}))]| \leq \|f\|_2 \sqrt{\mathbf{E}[(g(x_t) - g(x_{t-1}))^2]}. \quad (2)$$


Let  $Mg$  be the function defined by  $Mg(y) = \mathbf{E}_{y'}[(g(y') - g(y))^2]$ , where  $y'$  is formed from  $y$  by taking one step in the Markov chain we used in defining the  $x_t$ 's. Hence the expression inside the square-root in (2) is  $\mathbf{E}[Mg(x_{t-1})]$ . We would like to say that this expectation is close to  $\mathbf{E}[Mg(x_0)]$  because the distributions on  $x_0$  and  $x_{t-1}$  are very similar. We can use the argument in my “wrong-Sperner” notes for this. Using  $t \leq T \ll \sqrt{n}$ , I think that'll give something like  $|\mathbf{E}[Mg(x_{t-1})] - \mathbf{E}[Mg(x_0)]| \leq O(t/\sqrt{n})\|Mg(x_0)\|_2 \leq O(t/\sqrt{n})\sqrt{\mathbf{E}[Mg(x_0)]}$ , where on the right I've now assumed that  $g$  is bounded, hence  $Mg$  is bounded, hence  $(Mg)^2 \leq Mg$  pointwise.

But  $\mathbf{E}[Mg(x_0)]$  is precisely (well, maybe up to a factor of 2 or something) the “energy” or “average influence” of  $g$ ; write it as  $\mathcal{E}(g)$ . So we get  $\mathbf{E}[(g(x_t) - g(x_{t-1}))^2] = \mathcal{E}(g) \pm O(t/\sqrt{n})\sqrt{\mathcal{E}(g)}$ .

Let's assume now that in fact  $T \leq \sqrt{\mathcal{E}(g)}\sqrt{n}$ . Then the error in the above is essentially negligible. Now plugging this into (2) we get that each of the  $T$  error terms in (1) is at most  $O(1)\|f\|_2\sqrt{\mathcal{E}(g)}$ . So overall, we conclude

$$\mathbf{E}[f(x_0)g(x_T)] = \mathbf{E}[f(x_0)g(x_0)] \pm O(T \cdot \|f\|_2 \cdot \sqrt{\mathcal{E}(g)}). \quad (3)$$

Just to repeat, the assumptions needed to deduce (3) were that  $g$  is bounded and that  $T \leq \sqrt{\mathcal{E}(g)}\sqrt{n}$ .

- 
[Terence Tao](#) Says:  
[February 24, 2009 at 8:10 pm](#)

815.1 Sperner

Ryan, thanks for fleshing in the details and sorting out the  $\ell^1$  vs  $\ell^2$  issue. I just wanted to point out that we may have located the “DHJ(0,2)” problem that Boris brought up a while back - namely, DHJ(0,2) should be

DHJ(2) for “low influence” sets. I think we now have a satisfactory understanding of the DHJ(2) problem from an obstructions-to-uniformity perspective, namely that arbitrary sets can be viewed as the superposition of a low influence set (or more precisely, a function  $f_{U^\perp}$ ) plus a DHJ(2)-uniform error.

DHJ(1,3) may now need to be tweaked to generalise to functions  $f$  which are polynomial combinations of 01-low influence, 12-low influence, and 20-low influence functions, where 01-low influence means that  $f(x)$  and  $f(y)$  are close whenever  $y$  is formed from  $x$  by flipping a random 0 bit to a 1 or vice versa, etc. (With our current definition of DHJ(1,3), we are considering products of indicator functions which have zero 01-influence, zero 12-influence, and zero 20-influence respectively.)

22.  [gowers](#) Says:

[February 24, 2009 at 6:47 pm](#) | [Reply](#)

816. Obstructions.


I hope I'm going to have time to do some serious averaging arguments this evening. They will be private calculations (though made public as soon as they work), but let me give an outline of what I am hoping to make rigorous. The broad aim is to prove that a set with no combinatorial lines has a significant local correlation with a set of complexity 1. This is a statement that has been sort of sketched in various comments already (by Terry and by me and possibly by others too), but I now think that writing it out in a serious way will be a very useful thing to do and should get us substantially closer to a density-increment proof of DHJ(3). Here is a rough description of the steps.

1. A rather general argument to prove that whenever we feel like restricting to a combinatorial subspace, we can always assume that it has density at most  $\delta - \eta$ , where  $\eta$  is an arbitrary function of  $\delta$ . This kind of argument is standard, and has already been mentioned by Terry: basically if you can find a subspace with increased density then you happily pass to that subspace and you've already completed your iteration. If you can't do that, then the proportion of subspaces (according to any reasonable distribution of your convenience) with substantially smaller density is tiny. The minor (I hope) technical challenge is to get a version of this principle that is sufficiently general that it can be used easily whenever it is needed.
2. Representing combinatorial lines as  $(U, V, W)$ , we know that for an average  $W$  (each element chosen with probability  $1/3$  — I'm going for uniform measure here) the density of  $(U, V)$  such that  $(U, V, W)$  is in  $\mathcal{A}$  is  $\delta$ .
3. Also, by 1, for almost all  $W$  we find that the set of points  $(U, V, W \cup W')$  where  $|W'| \leq m$  (for some  $m$  with  $1 \ll m \ll n$ ) has density almost  $\delta$ .
4. Combining 2 and 3, we obtain a  $W$  such that the density of  $(U, V)$  with  $(U, V, W)$  in  $\mathcal{A}$  is at least  $\delta/2$ , say, and the density of  $(U, V, W \cup W') \in \mathcal{A}$  is almost as large as  $\delta$  (at least).
5. But the points of the latter kind have to avoid those  $U$  and  $V$  for which the point  $(U, V, W) \in \mathcal{A}$ , which is a dense complexity-1 obstruction in the “unbalanced” set of  $(U, V, W')$  that have union  $[n] \setminus W$  and have  $W'$  of size at most  $m$ .
6. Randomly restricting, we can get a similar statement, but this time for a “balanced” set—that is, one where  $U', V', W'$  have comparable sizes.
7. From that it is straightforward to get a (local) density increment on a special set of complexity 1 (as defined [here](#)).

Now I think I've basically shown that a special set of complexity 1 contains large combinatorial subspaces, though I need to check this. But what we actually need is a bit stronger than that: we need to cover sets of complexity 1 uniformly by large combinatorial subspaces, or perhaps do something else of a similar nature. (This is where the Ajtai-Szemerédi proof could come in very handy.) But if everything from 1 to 7 works out—I'm not sure how realistic a hope that is but even a failure would be instructive—then we'll be left with a much smaller-looking problem to solve.




I'll report back when I either get something working or see where some unexpected difficulty lies.

23.  [ryanworldwide](#) Says:  
[February 24, 2009 at 8:27 pm](#) | [Reply](#)

817. Sperner.

Just to clarify, the thing I wrote in #800 says that if you don't mind restricting to combinatorial subspaces (which we usually don't, unless we're really trying to get outstanding quantitatives), then the decomposition  $f = f_{U^\perp} + f_U$  we seek can be achieved trivially: you just take  $f_{U^\perp} = \mathbb{E}[f]$ .

24.  [Terence Tao](#) Says:  
[February 25, 2009 at 2:35 am](#) | [Reply](#)

818. Ergodic proof of DHJ(3)

I managed to digest Randall's lecture notes on the completion of the Furstenberg-Katznelson proof of DHJ(3) (the focus of the 600 thread) to the point where I now have an informal combinatorial translation of the argument at

[http://michaelnielsen.org/polymath1/index.php?title=Furstenberg-Katznelson\\_argument](http://michaelnielsen.org/polymath1/index.php?title=Furstenberg-Katznelson_argument)

that avoids any reference to infinitary concepts, at the expense of rigour and precision. Interestingly, the argument is morally based on a reduction to something resembling DHJ(1,3), but more complicated to state. We are trying to get a lower bound for

$$\mathbb{E} f(\ell(0)) f(\ell(1)) f(\ell(2)) (1)$$

where  $f$  is non-negative, bounded, and has positive density, and  $\ell$  ranges over all lines with “few” wildcards (and I want to be vague about what “few” means). The first reduction is to eliminate “uniform” or “mixing” components from the second two factors and reduce to

$$\mathbb{E} f(\ell(0)) f_{01}(\ell(1)) f_{20}(\ell(2)) (2)$$

where  $f_{01}, f_{20}$  are certain “structured” components of  $f$ , analogous to  $f_{U^\perp}$  from the Sperner theory. They have positive correlation with  $f$ , and in fact are positive just about everywhere that  $f$  is positive.

What other properties do  $f_{01}, f_{20}$  have? In a perfect world, they would be “complexity 1” sets, and in particular one would expect  $f_{01}$  to be describable as some simple combination of 01-low influence sets and 12-low influence sets (and similarly  $f_{20}$  should be some simple combination of 20-low influence sets and 12-low influence sets). Here,  $ij$ -low influence means that the function does not change much if an  $i$  is flipped to a  $j$  or vice versa.

Unfortunately, it seems (at least from the ergodic approach) that this is not easily attainable. Instead,  $f_{01}$  obeys a more complicated (and weaker) property, which I call “01-almost periodicity relative to 12-low influence”, with  $f_{20}$  obeying a similar property. Very roughly speaking, this is a “relative” version of 01-low influence: flipping digits from 0 to 1 makes  $f_{01}$  change, but the way in which it changes is controlled entirely by functions that have low 12-influence. (This is related to the notion of “uniform almost periodicity” which comes up in my paper on the quantitative ergodic theory proof of Szemerédi's theorem.)

It is relatively painless (using Cauchy-Schwarz and energy-increment methods) to pass from (1) to (2). To deal with (2) we need some periodicity properties of  $f_{01}, f_{20}$  on small “monochromatic” spaces (the existence of which is ultimately guaranteed by Graham-Rothschild) which effectively let us replace  $f_{01}(\ell(1))$  with  $f_{01}(\ell(0))$  and  $f_{20}(\ell(2))$  with  $f_{20}(\ell(0))$  on a large family of lines  $\ell$  (and more importantly, a large 12-low influence family of lines). From this fact, and the previously mentioned fact that  $f_{01}, f_{20}$  are large on  $f$ , we can get DHJ(3).

The argument as described on the wiki is far from rigorous at present, but I am hopeful that it can be translated into a rigorous finitary proof (though it is not going to be pleasant - I would have to deploy a lot of machinery from my quantitative ergodic theory paper). Perhaps a better approach would be to try to export some of the ideas here to the Fourier-type approaches where there is a better chance of a shorter and more quantitatively effective argument.



o [Randall](#) Says:

[February 25, 2009 at 5:06 pm](#)

818.1

I haven't read this closely enough to have even an initial impression, however, much of it looks (somewhat) familiar.

First, I notice you removed your discussion of stationarity... instead (tell me if I misread), and in multiple settings, you seem to do a Cesaro average, over lines, rather than employing something like Graham-Rothschild to get near-convergence along lines restricted to a subspace. Most striking of these are instances of using the so-called IP van der Corput lemma. Looking at its proof, this does indeed look kosher, but it's rather surprising to me all the same; assuming I'm understanding this at least somewhat correctly, did you give any thought to whether the ergodic proof itself could be tidily rewritten to accommodate this averaging method?

Modulo the above, the main part of the argument I still don't (in principle) understand is how you bound  $h$ , the number of functions used to approximate the almost periodic component of  $f$ , independently of  $n$ . (This is the part of the argument I got stuck on in my own thoughts.) I see now that you solved this issue in your quantitative ergodic proof of Szemerédi, which I printed last night as well, though I haven't read deeply enough yet to see how. Am I to assume that something similar happens here, or is the answer different in the two cases?



o [Terence Tao](#) Says:

[February 25, 2009 at 8:50 pm](#)

818.2

Yes, the argument is distilled from your notes, though as you see I messed around with the notation quite a bit.


The stationarity is sort of automatic if you work with random subspaces of a big space  $[3]^n$ , and I am implicitly using it all over the place when rewriting one sort of average by another.

I am indeed hoping that Cesaro averaging may be simpler to implement than IP averaging, and may save me from having to use Graham-Rothschild repeatedly. There are a lot of things hidden in the sketch that may cause this to bubble up. For instance, I am implicitly using the fact that certain shift operators  $U_\alpha$  converge (in some Cesaro or IP sense) to an orthogonal projection, and this may require a certain amount of Graham-Rothschild type trickery (I started writing some separate notes on a finitary Hilbert space IP Khintchine recurrence theorem which will be relevant here.)

I admit I'm a bit sketchy on how to deal with  $h$  not blowing up. A key observation here is that of statistical sampling: if one wants to understand an average  $\mathbb{E}_{h \in H} c_h$  of bounded quantities over a very large set  $H$ , one can get quite a good approximation to this expression by picking a relatively small number  $h_1, \dots, h_m$  of representatives of  $H$  at random and looking at the local average  $\mathbb{E}_{j=1, \dots, m} c_{h_j}$  instead. (This fact substitutes for the fact used in the Furstenberg approach that Volterra integral operators are compact and hence approximable by finite rank operators; or more precisely, the Furstenberg approach needs the relative version of this over some factor  $Y$ .) I haven't worked out completely how this trick will mesh with the IP-systems involved, but I'm hoping that I can throw enough Ramsey theorems at the problem to make it work out.

Perhaps one thing that helps out in the finitary setting that is not immediately available in the ergodic setting is

that there are more symmetries available; in particular, the non-commutativity of the IP systems that makes the ergodic setup so hard seems to be less of an issue in the finitary world (the operations of flipping a random 0 to a 1 and flipping a random 0 to a 2 essentially commute since there are so many 0s to choose from). There is a price to pay for this, which is that certain Ramsey theorems may break the symmetry and so one may have to choose to forego either the Ramsey theorem or the symmetry. This could potentially cause a problem in my sketch; as I said, I have not worked out the details (given the progress on the Fourier side of things, I had the vague hope that maybe just the concepts in the sketch, most notably the concept of almost periodicity relative to a low influence factor, could be useful to assist the other main approach to the problem, as I am not particularly looking forward to rewriting my quantitative ergodic theory paper again.)

25.  ryanworldwide Says:

[February 25, 2009 at 4:53 am](#) | [Reply](#)

819. #816 and #818 look quite exciting; I plan to try to digest them soon. Meanwhile, here is a Moser-esque problem I invented for the express purpose of being solvable. I hope it might give a few tricks we can use (but it might not be of major help due to the PS of #528).

Let's define a combinatorial *bridge* in  $[3]^n$  to be a triple of points  $(x, y, z) \in \{-1, 0, 1\}^n$  formed by taking a string with zero or more wildcards and filling in the wildcards with either  $(0, -1, 0)$  or  $(0, 1, 0)$ . If there are zero wildcards we call the bridge degenerate. I think I can show, using the ideas from #800, that if  $f : \{-1, 0, 1\}^n \rightarrow \{0, 1\}$  has mean  $\delta > 0$  and  $n$  is sufficiently large as function of  $\delta$ , then there is a nondegenerate bridge  $(x, y, z)$  with  $f(x) = f(y) = f(z) = 1$ .

Roughly, we first use a density-increment argument to reduce to the case when  $f$  is extremely noise sensitive; i.e.,  $\|T_{1-\gamma}f\|_2^2$  is only a teeny bit bigger than  $\delta^2$ . Here  $\gamma$  is something very small to be chosen later. Next, we pick a suitable distribution on combinatorial bridges  $(x, y, z)$ ; basically, choose a random one where the wildcard probability is  $\epsilon \ll \sqrt{\delta/n}$ . Now the key is that under this distribution, there is "imperfect correlation" between the random variable  $(x, y)$  and the random variable  $z$  — and similarly, between  $x$  and  $(y, z)$ . Here I use the term in the sense of the Mossel paper in #811. Because of this (see Mossel's Lemma 6.2),  $\mathbf{E}[f(x)f(y)f(z)]$  is practically the same as  $\mathbf{E}[T_{1-\gamma}f(x) \cdot T_{1-\gamma}f(y) \cdot T_{1-\gamma}f(z)]$ , when  $\gamma \leq \delta^{O(1)}\epsilon$ . But this is extremely close to  $\delta^3$  because  $\mathbf{E}[T_{1-\gamma}f(x)] = \mathbf{E}[T_{1-\gamma}f(y)] = \mathbf{E}[T_{1-\gamma}f(z)] = \delta$  and because the error can be controlled with Hoeffding's inequality in terms of  $\|T_{1-\gamma}f - \delta\|_2^2$ , which is teeny by the density-increment reduction.

More details here: <http://www.cs.cmu.edu/~odonnell/bridges.pdf>. I can try to port this to the wiki soon.



o Terence Tao Says:

[February 25, 2009 at 8:38 pm](#)

Ryan, I'm not sure what you mean by "fill the wildcards by either  $(0, -1, 0)$  or  $(0, +1, 0)$ ". Wouldn't this always make  $x$  and  $z$  equal in a bridge? Perhaps an example would clarify what you mean here.



o ryanworldwide Says:

[February 25, 2009 at 10:09 pm](#)

819.2: Er, whoops. You're right. Even easier than I thought 😊 To make this problem more interesting, I think it will work if the triples allowed on wildcards are, say:  $(-, 0, -)$ ,  $(-, 0, 0)$ ,  $(+, 0, 0)$ ,  $(+, 0, +)$ .

But: a) this is not as nice-looking, and b) I think it'll actually take slightly more work.

So, um, never mind for now.


26.  gowers Says:

[February 25, 2009 at 6:44 pm](#) | [Reply](#)

820. Density increment.

I've now finished a wiki write-up that is supposed to establish (and does unless I've made a mistake, which is possible) that if  $\mathcal{A}$  is a line-free set of uniform density  $\delta$  then you can pass to a combinatorial subspace of dimension  $m$ , as long as  $m = o(n^{1/2})$ , and find a special subset of complexity 1 in that subspace of density at least  $c\delta^2$ , such that the equal-slices density of  $\mathcal{A}$  inside that special subset is at least  $\delta + c\delta^2$ .

To be less precise about it, I think I've shown that if  $\mathcal{A}$  contains no combinatorial line, then you get a density increase on a nice set. I start with uniform density and switch to equal-slices density, but that is deliberate, and explained in the write-up, which, by the way, is [here](#). So my attention is now going to turn to trying to copy either the Ajtai-Szemerédi proof, or Shkredov's proof, of the corners theorem. I feel optimistic that this can be done, given that special sets of complexity 1 can be shown to contain combinatorial subspaces—though that on its own is not enough.

 [ryanworldwide](#) Says:

[February 26, 2009 at 1:02 am](#)

820.1. Hi Tim — nice. I read through the wiki proof and agree that it should be correct. It is quite interesting how the passage between uniform and equal-slices seems necessary. On one hand I worry a bit that the different measures might get out of control; on the other hand, the optimistic way to look at it is that we may eventually get so proficient at passing between measures that it'll be a very useful tool.

 [gowers](#) Says:

[February 26, 2009 at 10:14 am](#)

820.2

One observation that makes me feel slightly less worried about passing from one measure to another is that you can start with equal-slices measure instead. Then as a first step you pass to uniform measure on a subspace.

Then you carry out the argument as given (passing to a subspace of that subspace). And now you've ended up with the same measure that you started with.

27.  [ryanworldwide](#) Says:

[February 25, 2009 at 10:11 pm](#) | [Reply](#)

821. Here is (I believe) a simple proof of Terry's most basic structure theorem. The setting is as follows:  $\Omega$  is a finite set and  $\mu$  is a probability distribution on  $\Omega$ . We also write  $\mu = \mu^{\otimes n}$  for the product distribution on  $\Omega^n$ . We work in the space of functions  $f : \Omega^n \rightarrow \mathbb{R}$  with inner product  $\langle f, g \rangle = \mathbf{E}_{x \sim \mu}[f(x)g(x)]$ . Henceforth, all probabilities are wrt  $\mu$  unless specified.

**Theorem:** Let  $f : \Omega^n \rightarrow \mathbb{R}$  have  $\text{Var}[f] \leq 1$ . Let  $0 < \epsilon < 1$ . Let  $\frac{1}{n} \leq \sigma_1 \leq \tau_1 \leq \sigma_2 \leq \tau_2 \leq \dots \leq \sigma_k \leq \tau_k \leq \epsilon$  be a sequence of “scales” (or “times”) satisfying the following conditions: (a)  $\sigma_i/\tau_{i-1} \geq 3 \ln(1/\epsilon)/\epsilon$  for all  $i$ ; (b)  $k \geq 1/\epsilon^2$ .


Then there exists  $i$  such that  $f$  can be written as  $f = f_{\text{stab}} + f_{\text{sens}}$ , where:

1.  $\mathbf{E}[f_{\text{stab}}] = \mathbf{E}[f]$ .
2.  $\|f_{\text{stab}}\|_p \leq \|f\|_p$  for all  $1 \leq p \leq \infty$ ; hence  $f_{\text{stab}}$  has range  $[0, 1]$  if  $f$  does (and then  $f_{\text{sens}}$  has range  $[-1, 1]$ ).
3.  $0 \leq \langle f_{\text{stab}}, f_{\text{sens}} \rangle \leq 3\epsilon$ .
4.  $\mathcal{E}(f_{\text{stab}}) \leq 1/(\tau_i n)$ , where  $\mathcal{E}(g)$  denotes the “average influence of  $g$ ”: i.e.,  $\mathbf{E}[(g(x) - g(x'))^2]$ , where  $x \sim \mu$  and  $x'$  is

formed by rerandomizing a random coordinate of  $x$ .

5.  $\mathcal{S}_{1-\sigma_i}(f_{\text{sens}}) \leq 3\epsilon^2$ , where  $\mathcal{S}_{1-\sigma}(g)$  denotes the “noise stability of  $g$  at  $1 - \sigma$ ”: i.e.,  $\mathbf{E}[g(x)g(x')]$ , where  $x \sim \mu$  and  $x'$  is formed by rerandomizing each coordinate of  $x$  with probability  $\sigma$ .

Proof in next post.

28.  ryanworldwide Says:

[February 25, 2009 at 10:12 pm](#) | [Reply](#)

822. *Proof:* For simplicity I present the proof when  $\Omega = \{0, 1\}$  and  $\mu$  is the uniform distribution; to get the full version, replace the Fourier transform with the “Hoeffding AKA Efron-Stein orthogonal decomposition”.

For an interval  $J = [a, b]$  of natural numbers, let  $\mathcal{W}_J(f)$  denote  $\sum_{a \leq |S| \leq b} \hat{f}(S)^2$ . The condition  $\text{Var}[f] \leq 1$  is equivalent to  $\mathcal{W}_{[1, n]} \leq 1$ , and we will use this frequently. Now consider the intervals of the form  $J_i = [\epsilon/\tau_i, 2 \ln(1/\epsilon)/\sigma_i]$ . By hypothesis (a) these intervals are disjoint, and by hypothesis (b) there are at least  $1/\epsilon^2$  of them. By Pigeonhole, we can fix a particular  $i$  such that  $\mathcal{W}_{J_i} \leq \epsilon^2$ . Henceforth write  $\sigma = \sigma_i$ ,  $\tau = \tau_i$ .

Now set  $f_{\text{stab}} = T_{1-\tau}f$  and  $f_{\text{sens}} = f - f_{\text{stab}}$ . Conditions 1 and 2 are satisfied. Condition 4 holds because  $n\mathcal{E}(f_{\text{stab}}) = \sum_S |S| \hat{f}_{\text{stab}}(S)^2 = \sum_S |S| (1-\tau)^{2|S|} \hat{f}(S)^2$

and this is at most  $1/\tau$  because  $|S|(1-\tau)^{2|S|} \leq 1/\tau$  for all  $S$ . For Condition 3,  $\langle f_{\text{stab}}, f - f_{\text{stab}} \rangle = \sum_S a(|S|) \hat{f}(S)^2$ , where  $a(s) = (1-\tau)^s - (1-\tau)^{2s}$ .

We have  $a(0) = 0$  so we can drop that term. We have  $a(s) \leq s\tau$  always so we get a contribution of at most  $\epsilon$  from the terms with  $|S| \leq \epsilon/\tau$ . We also have  $a(s) \leq (1-\tau)^s \leq \exp(-\tau s)$  always so we get a contribution of at most  $\epsilon^2$  from the terms with  $|S| \geq 2 \ln(1/\epsilon)/\tau$ . And the remaining terms have  $|S| \in [\epsilon/\tau, \ln(1/\epsilon)/\tau] \subset J_i$ , so we get a contribution of at most  $\epsilon^2$  from them.

The proof of Condition 5 is similar. We have  $\mathcal{S}_{1-\sigma}(f_{\text{sens}}) = \sum_S (1-\sigma)^{|S|} \hat{f}_{\text{sens}}(S)^2 = \sum_S b(|S|) \hat{f}(S)^2$ ,

where  $b(s) = (1-\sigma)^s (1 - (1-\tau)^s)^2$ .

Again,  $b(0) = 0$ . We have  $b(s) \leq (\tau s)^2$  always so we get a contribution of at most  $\epsilon^2$  from the terms with  $|S| \leq \epsilon/\tau$ . We also have  $b(s) \leq (1-\sigma)^s$  always so we get a contribution of at most  $\epsilon^2$  from the terms with  $|S| \geq 2 \ln(1/\epsilon)/\sigma$ . And the remaining terms have  $|S| \in J_i$  so we get a contribution of at most  $\epsilon^2$  from them.  $\square$


o  Terence Tao Says:

[February 25, 2009 at 10:48 pm](#)

822.1 This looks about right to me. The game is always to first use pigeonhole to find a nice big gap in the energy spectrum, and then use that gap to split  $f$  into opposing pieces. (Sometimes one also needs a (small) third term to deal with the small amount of energy stuck inside the gap.)


One thing that works in your favour here is that the averaging operators  $T_{1-\tau}$  are positivity preserving, so if  $f$  is positive then so is  $f_{\text{stab}}$  (and related to this is a useful comparison principle: if  $f$  is pointwise bigger than  $g$ , then  $f_{\text{stab}}$  is pointwise bigger than  $g_{\text{stab}}$ .) Things get more tricky when one doesn't have this positivity preserving property, because it is absolutely essential later on that  $f_{\text{stab}}$  be non-negative. In my paper with Ben we had to introduce the machinery of partitions and conditional expectations to deal with this. One side effect of this is that it can force the scales to be exponentially separated (e.g.  $1/\sigma_i \gg \exp(1/\sigma_{i+1})$ ) rather than just separated by a large constant depending only on  $\epsilon$ ). This leads to the type of tower-exponential bounds which are familiar from the regularity lemma. Once we get to the fancier structure theorems, we may start seeing these sorts of bounds emerge, but I agree that in the simple case here we don't have to deal with all that because of the handy  $T_{1-\tau}$  operators.

Combining this with your notes that control the Sperner count by the noise stability, it looks like we have a pretty solid Fourier-analytic proof of DHJ(2), which was one of our early objectives (suggested first, I believe, by Gil).

◦  ryanworldwide Says:

[February 26, 2009 at 1:00 am](#)

822.2: Yes, I think with slightly more work one can get an “Adequate-Sperner” Fourier proof; I think it will require density  $\Omega(1/\sqrt{\log n})$ . This is still not quite right, but is closer.

29.  Ryan O'Donnell Says:

[February 26, 2009 at 6:54 am](#) | [Reply](#)

823. Structure Theorem.

A small note: Unless I'm mistaken, the structure theorem around Lemma 1 in Terry's wiki notes on Furstenberg-Katznelson — which decomposes functions on  $[3]^n$  into a 12-stable and a 12-sensitive part — can be proved in the same way as in #821.

Specifically, employing Tim's  $(U, V, W)$  notation, I think you just need to look at the expectation over  $U$  of the Fourier weight of  $f_U$  on the various intervals  $J_i$ . Here  $f_U$  denotes the restricted function, with 0's fixed into  $U$ , on binary inputs (1 and 2). Everything seems to go through the same, using the fact that restricting functions with  $\$U\$$  commutes with flipping 1's and 2's, and that  $\mathbb{E}_U[\text{Var}[f_U]] \leq \text{Var}[f]$ .

◦  Ryan O'Donnell Says:

[February 26, 2009 at 6:56 pm](#)

823.1. The weird  $\mathbb{V}_{\mathcal{D}}\setminus$  at the end here is meant to be “Var”.

*Fixed now — Tim.*

30.  gowers Says:

[February 26, 2009 at 11:48 am](#) | [Reply](#)

824. DHJ(3) general strategy

I have a hunch, after looking at [Van Vu's quantitative version of Ajtai-Szemerédi](#), that it may be possible to strengthen the argument on the wiki to give the whole thing. If this were not a polymath project I would of course go away and try to make the idea work, keeping quiet if it didn't and giving full details if it did. But it is polymath, so let me say in advance roughly what I hope will happen.

In the corners problem, it is fairly easy to prove that a corner-free set has a density increment on a Cartesian product of two dense sets. The problem is what to do then. Ajtai and Szemerédi's strategy is to use a more sophisticated averaging argument, making heavy use of Szemerédi's theorem, to obtain a density increment on a highly structured Cartesian product: it is a product of two arithmetic progressions with the same common difference.

Initially, it looks very discouraging that they have to use Szemerédi's theorem, but I now believe that that may just be an artificial byproduct of the fact that they are working in the grid and want to get another grid at the next stage of the induction. But in the Hales-Jewett world, you don't actually need long arithmetic progressions anywhere, and I think the right analogue will in fact turn out to be multidimensional subspaces of  $[2]^n$ . My actual idea is slightly more precise than this, but hard to explain. Basically, I want to revisit the argument on the wiki, but aim straight for a density increase on a subspace, modelling the argument as closely as I can on the Ajtai-Szemerédi argument for the corners



problem.

If that works, then it will probably give a tower-type bound. A follow-up project would then be to try to do a more powerful Shkredov-type argument to get a bounded number of exponentials. But I'm getting way ahead of myself by even discussing this.

Let me also slip in a metacomment. I find it slightly inconvenient that, now that we have modest threading, I can't just scroll to the bottom of the comments and see what has been added. As Jason Dyer pointed out, there is a trade-off here. On the one hand, it makes it slightly harder if you are trying to keep up with the discussion in real time, but on the other it probably means that the comments are better organized if you want to come back to them later. In the end, the second consideration probably trumps the first, because it's a long-term convenience at the cost of a mild short-term inconvenience. But I still think that we should be quite careful about when and how much we use the threading.

31.  [gowers](#) Says:

[February 26, 2009 at 5:32 pm](#) | [Reply](#)

825. Ajtai-Szemerédi approach.

No time to justify this, or at least not for a few hours (and possibly not before tomorrow morning) but I am now very optimistic indeed that the Ajtai-Szemerédi approach to corners can be modified to give DHJ(3). As usual, this surge of optimism may be followed by a realization that there are some serious technical obstacles, but I think that the existing write-up of the complexity-1 correlation contains methods for dealing with the technicalities that will arise. I plan to write up (i) a sketch proof of Ajtai-Szemerédi on the wiki, (ii) an outline of an argument here on the blog (which will include the AS-to-DHJ(3) dictionary I will be using), and (iii) a detailed sketch of an argument on the wiki. Those are in decreasing order of probability that I can actually finish them without getting stuck.

32.  [Terence Tao](#) Says:

[February 27, 2009 at 12:45 am](#) | [Reply](#)

826. Structure theorem

I now think there is a way to do the DHJ(2) structure theory without explicit mention of the Fourier transform (at the cost of worse quantitative bounds), which may be important when trying to extend the arguments to the DHJ(3) structure theory sketched in my wiki notes.

The point is to replace the self-adjoint operators  $T_{1-\tau}$  that Ryan uses by a “one-sided” counterpart. Given a function  $f : [2]^n \rightarrow \mathbb{R}$ , and a parameter  $\lambda > 0$ , let  $S_\lambda f : [2]^n \rightarrow \mathbb{R}$  be defined by  $S_\lambda f(x) := \mathbb{E}f(y)$ , where  $y$  is formed from  $x$  by allowing each coordinate of  $x$ , with probability  $\lambda/n$ , of being overwritten by a 1. (This is in contrast to Ryan's rerandomisation, which would cause  $x$  to be overwritten by a random 0/1 bit rather than by 1.)

Informally, one can think of  $y$  as taking about  $\lambda/2$  of the 0-bits of  $x$  (the exact number basically obeys a Poisson distribution) and flipping them to 1s.

Observe that  $(x,y)$  always forms a combinatorial line. Thus lower bounds on  $\mathbb{E}f(x)S_\lambda f(x)$  imply lots of combinatorial lines in the support of  $f$ .

The key observation is that the  $S_\lambda$  are basically a semigroup (for  $\lambda$  much smaller than  $n$ ):  $S_\lambda S_{\lambda'} \approx S_{\lambda+\lambda'}$ . In particular, if  $\lambda \ll \lambda'$ , we have the absorption property  $S_\lambda S_{\lambda'} \approx S_{\lambda'}$ . Because of this, one can show that the  $S_\lambda$  converge (in the strong operator topology) to an orthogonal projection  $P$ ; to see this finitarily, one would inspect the energies  $\|S_\lambda f\|_2^2$ , which are basically monotone decreasing in  $\lambda$ , and locate a long gap in the energy spectrum in which these energies plateau in  $\lambda$ . I think this gives us the same type of structure/randomness decomposition as before. Note that for large  $\lambda$ ,  $S_\lambda f$  is approximately  $S_1$ -invariant, which is basically the same thing as having low influence.



o [Terence Tao](#) Says:

[February 27, 2009 at 12:46 am](#)

correction:  $(x,y)$  form a combinatorial line most of the time; there is a degenerate case when  $x=y$  (analogous to the case of a degenerate arithmetic progression). But for large  $\lambda$ , this case is very rare and can be ignored.



33. [gowers](#) Says:

[February 27, 2009 at 1:12 am](#) | [Reply](#)

827. Ajtai-Szemerédi

Just to say that the easy step of my three-step plan is now done: an informal write-up of the Ajtai-Szemerédi proof of the corners theorem can now be found [on this wiki page](#). I have to go to bed now so parts (ii) and (iii) will have to wait till tomorrow.



34. [gowers](#) Says:

[February 27, 2009 at 1:44 pm](#) | [Reply](#)

828. Progress report

Part (ii) is, perhaps predictably, giving me a bit more of a headache. But it's an interesting headache in that I am stuck on something that feels as though it shouldn't be too hard. I don't know how much time I'll have to think about it, so let me say what the problem is.

Call a subset  $\mathcal{B} \subset [3]^n$  *simple* if whether or not  $x$  belongs to  $\mathcal{B}$  depends only on the 1-set of  $x$ . In set language, there is a set-system  $\mathcal{U}$  such that  $\mathcal{B} = \{(U, V, W) : U \in \mathcal{U}\}$ . The question I'm struggling with is this. Let  $\mathcal{A}$  be a dense subset of  $[3]^n$  that correlates with a dense simple set  $\mathcal{B}$ . Does that imply that there is a combinatorial subspace with dimension tending to infinity on which  $\mathcal{A}$  has a density increment?

I cannot believe that this is a truly hard problem, but I haven't yet managed to solve it. I think a solution would be very useful. Oh, and I really don't mind what measure is used in any solution ...



o [Ryan O'Donnell](#) Says:

[February 27, 2009 at 4:26 pm](#)

828.1: What if  $\mathcal{A} = \mathcal{B}$  = strings with an even number of 1's?



o [gowers](#) Says:

[February 27, 2009 at 4:50 pm](#)

828.2: Then take all sequences  $x$  such that  $x_1 = x_2, x_3 = x_4$ , etc., and you have a combinatorial subspace of dimension  $n/2$  where the density has gone all the way up to 1. Were you imagining that wildcard sets had to have size 1? That is certainly not intended to be a requirement in my question.



o [Terence Tao](#) Says:

[February 27, 2009 at 5:29 pm](#)

828.3

I can answer your question as "yes", but in a totally unhelpful way: DHJ(3) implies that *any* dense set has a density increment on a large combinatorial subspace.

Of course, the whole point is to find a DHJ(3)-free proof of this fact... I don't currently have time to think about it right now, but I might look at it later.



- o [Terence Tao](#) Says:  
[February 27, 2009 at 5:33 pm](#)

828.4

Hmm, actually thinking about it a little bit there may be an energy-increment type proof.

Firstly, one should be able to show that a 1-set contains lots of combinatorial subspaces where it is denser, by concatenating the 2 and 3-sets together and applying DHJ(2).

Next, one should be able to apply some sort of structure theorem to decompose an arbitrary set (or function) into a 1-set plus something orthogonal to 1-sets. Actually one needs to mess around with scales and get a decomposition into a 1-set  $f_{U^\perp}$  at some large scale plus something  $f_U$  orthogonal to 1-sets at finer scales (in particular, it should have mean zero at finer scales). (Let me be vague about what "scale" means, but every time you pass to a large combinatorial subspace, you're supposed to descend one scale.)

Now take the guy  $f_{U^\perp}$  which is a 1-set at a large scale and find lots of combinatorial subspaces in it at the finer scale where it is denser. The guy  $f_U$  which is orthogonal to 1-sets at finer scale should not disrupt this density increment.

Have to run, unfortunately... may flesh this out later



- o [Ryan O'Donnell](#) Says:  
[February 27, 2009 at 5:52 pm](#)

828.5. Re 828.2:

Ah, no, I guess I was misunderstanding what a combinatorial subspace is. I always think of this as, "fixing some of the coordinates and leaving the rest of them 'free'." It had not occurred to me that one is also allowed to insist on things like " $x_i = x_j$ ", although I now see that there is no reason not to.



- 35. [gowers](#) Says:  
[February 27, 2009 at 6:37 pm](#) | [Reply](#)

829. Ajtai-Szemerédi

Let me try to explain the motivation for my simple-sets question. As I said earlier, I am trying to model a proof of DHJ(3) on the Ajtai-Szemerédi corners proof (an account of which is given [here on the wiki](#)). I have the following rough dictionary in mind.

1. A vertical line corresponds to a set of sequences  $(U, V, W)$  where you fix the 1-set  $U$ .
2. A horizontal line corresponds to a set of sequences  $(U, V, W)$  where you fix the 2-set  $V$ .
3. A line of slope -1, or diagonal, corresponds to a set of sequences  $(U, V, W)$  where you fix the 3-set  $W$ .
4. An arithmetic progression in  $[n]$  with length tending to infinity corresponds to a combinatorial subspace in  $[2]^n$  with dimension tending to infinity.
5. A subgrid with width tending to infinity corresponds to a combinatorial subspace of  $[3]^n$  with dimension tending to

infinity.

6. A corner corresponds to a combinatorial line.

That's about it. So now let me say where the first problem arises when one tries to translate the AS proof to the DHJ(3) context. Obviously step 1 is fine: with any sensible model of a random combinatorial subspace you can assume that the density of  $A$  in that subspace is almost always at least  $\delta - \eta$ , where  $\delta$  is the density of  $A$ .

Step 2 says that if you have a positive density of vertical lines inside which  $A$  has density noticeably less than  $\delta$ , then you also have a positive density of vertical lines inside which  $A$  has density noticeably *greater* than  $\delta$ . If we let  $B$  be the set of  $x$ -coordinates associated with these vertical lines, then we can use Szemerédi's theorem to get an arithmetic progression  $P$  of these "over-full" vertical lines, with small common difference. Then we can partition the horizontal lines into arithmetic progressions  $Q_i$ , each of which is a translate of  $P$ , and by averaging we find a subgrid  $P \times Q_i$  inside which  $A$  has density noticeably larger than  $\delta$ , which gives us a density increment and we're done. (Of course, then one has to go on and say what happens if we don't have a positive density of sparse vertical lines — this is, as I say, just the first step of the argument.)

Let me gloss over the question of which measure I want to use: I feel that that is just a technicality and the main point is to get the conceptual argument working. If you accept item 1 in the dictionary, then the obvious first step is to define  $U$  to be underfull if the density of pairs  $(V, W)$  (out of all pairs such that  $(U, V, W)$  is a point in  $[3]^n$ ) is noticeably less than  $\delta$ . Then if there is a positive density of underfull  $U$ , then we have a positive density of overfull  $U$  as well. Now I don't have a problem finding a nice big combinatorial subspace consisting entirely of overfull  $U$ s (which is what item 4 suggests I should be looking for) but I don't then have a nice analogue of the fact that a Cartesian product of two APs with the same common difference is a grid. Or rather, maybe I do, but I'm not sure how to get at it. The analogue should be something like that if I have such a combinatorial subspace (by which I mean a system of sets  $U_0, U_1, \dots, U_k$  such that the elements of the system are all sets of the form  $\bigcup_{i \in E} U_i$  with  $0$  belonging to  $E$ ), then on average the combinatorial subspaces that fix  $U_0$  and treat  $U_1, \dots, U_k$  as wildcards should be slightly too dense. But for a fixed combinatorial subspace this doesn't seem to have to be true, so I need somehow to build a combinatorial subspace such that it is true.

I can feel this explanation getting less clear as I proceed, but perhaps it's at least clear globally what I want to do: the "simple set" is the set of all  $(U, V, W)$  such that  $U$  is overfull. So trivially  $A$  intersects it too densely. If we can get a density increment on a subgrid, then we're done, which means that Step 3 is complete and we may assume that virtually no  $U$  are underfull. (Given the difficulties associated with this step, it's not altogether clear that this statement is strong enough, but it would be encouraging. I might assume it and see what happens if I try to press on with the later steps.)



36. Ryan O'Donnell Says:

[February 27, 2009 at 7:50 pm](#) | [Reply](#)

830. Combinatorial subspaces; 828.4 vs. 828.2.

It seems to me that an argument along the lines in 828.4 at some point has to find combinatorial subspaces which include these  $x_i = x_j$  constraints, as in Tim's 828.2. If the combinatorial subspaces are just of the "fixing-some-coordinates" type then I'm not sure (yet) how to get around 828.1.

But will arguments as in 828.4 actually produce such subspaces?



37. gowers Says:

[February 27, 2009 at 8:14 pm](#) | [Reply](#)

831. Finding combinatorial subspaces

Ryan, that's an interesting point. The kind of arguments I had in mind are different though: they are inductive ones where you build a subspace up dimension by dimension, using Sperner's theorem (or similar results) at each stage. A prototype for such an argument is the proof I sketched on the wiki of the [basic multidimensional Sperner theorem](#). This argument leads naturally to larger wildcard sets.

38.  [Randall](#) Says:

[February 27, 2009 at 10:28 pm](#) | [Reply](#)

832. 1-sets

I was trying to think about Tim's question from the ergodic perspective; if I am translating correctly the outlook is not encouraging. Perhaps though I am overgeneralizing.

A 1-set is just a set, membership in which is impervious to switchings of 0s with 2s, right? Okay, so, if I understand Terry's translations from the ergodic world and back again, the functions that are measurable with respect to 1-sets would be in some sense analogous to the  $\tau^{-1}$  rigid factor...if that's right, then in order to think about the question ergodically, one should think whether it would help at all to know that  $1_A$  had non-zero projection onto this factor. And I don't see how it would. In particular, it isn't the case that things orthogonal to this factor don't contribute to the averages you're interested in, so I don't see how decomposing would help. Yes, it is easy to see from DHJ (2) what the positive contribution would be restricted to this rigid factor, but you lose positivity of the relative AP components; it would not seem an easy task to show that they have non-negative contribution without positivity, especially given that the argument as it now stands uses positivity very heavily (it throws a lot away because it can).

On the other hand, a non-zero projection onto this factor is sufficient to get (what Terry would call, I think) both 01 over 02 and 12 over 02 almost periodic components, and in a special way, so it might in theory be easier to get a density increment for these guys in the combinatorial world, but at least from an ergodic perspective, I can't see any strategy that might simplify the proof from what it already is. Short, that is, of a detailed analysis of what is going on in terms of finite-dimensional modules instead of "almost periodicity", but that is potentially a step in the direction of harder, not easier.

Of course, all I say depends on this analogy between 1-sets and the 02-rigid factor being what I suggest it is, and I don't really see that part well at all (I'm mostly just guessing because I don't understand the picture very well from the combinatorial side). (Also of course someone may be posting a proof simultaneously with my posting my doubts.)

o  [Terence Tao](#) Says:

[February 27, 2009 at 10:55 pm](#)

832.1

Yeah, a 1-set is what I would call a 02-low influence (in fact, a 02-zero influence) set, and it would correspond to a  $\tau^{-1}$  invariant set in the ergodic theory language. Tim's approach actually deals with intersections of 1-sets and 2-sets (which we have been calling "complexity 1 sets"), which would be a combination of a 02-low influence and 01-low influence set, or a set in the factor generated by both the  $\tau^{-1}$ -invariants and the  $\sigma^{-1}$ -invariants. (One may also want to throw in the 12-invariant sets for good measure.) But never mind this distinction for now.

Perhaps the place where Tim's approach and the ergodic approach differ, though, is that Tim can always stop the argument as soon as a density increment is found on a large subspace, whereas this trick is not available in the ergodic world. So I don't think Tim needs the strong (and presumably false) claim that "functions orthogonal to complexity 1 sets have negligible impact on the combinatorial line count". He only needs the weaker claim that "if a set has no combinatorial lines, then it must have non-trivial correlation with a complexity 1 set", combined with the (as yet not fully established) claim that "a set with non-trivial correlation with a complexity 1 set has a

density increment on a large combinatorial subspace”.

Hmm. Maybe one thing to put on my “to do” list is to find an ergodic translation of Tim’s argument on the wiki that line-free sets correlate with complexity 1 sets. (If you haven’t already noticed, I’m a big fan of translating back and forth between these two languages.)



o [Randall](#) Says:

[February 27, 2009 at 11:06 pm](#)

832.2

Actually I just got on here again to retract everything I said (I should think more carefully before posting). I now think, from the ergodic perspective, you trivially get a “measure increment” if  $1_A$  correlates projects non-trivially onto the 02-rigid factor, for the reason that on fibers, the measure is not constant. Just pick a bunch of fibers on which the measure is too big and then a subspace on which you have intersection of those fibers (DJH (2)).



o [Terence Tao](#) Says:

[February 27, 2009 at 11:39 pm](#)

832.2

Hmm, I tried my hand at translating Tim’s argument to ergodic theory, and it came out weird... in order to describe it, I have to leave the measure space  $X$  and work in a certain extension of it, vaguely analogous to the Host-Kra parallelopiped spaces  $X^{[k]}$  (or the Furstenberg self-joinings used in Tim Austin’s proof of convergence for multiple commuting shifts).

It’s something like this: consider the space of all quadruples

$(x, \sigma(\alpha)\rho^{-1}(\alpha)x, \sigma(\alpha\beta)\rho^{-1}(\alpha\beta)x, \sigma(\alpha)\tau(\beta)\rho^{-1}(\alpha\beta)x)$  in  $X^4$ , where  $\alpha$  is a word to the left of  $\beta$  (or maybe to the right; I may have the signs messed up). There should somehow be a natural measure  $\tilde{\mu}$  associated to these quadruples, which I currently do not know how to define properly; each of the four projections of this measure to  $X$  should be the original measure  $\mu$ .

The last three points of this quadruple form a combinatorial line (I may have messed up the order of operations a bit). So if  $A$  is line-free, the function  $1_A(x_4)$  on  $X^4$  will have strong negative correlation with  $1_A(x_2)1_A(x_3)$ . But the function  $1_A(x_2)$  has a “02-invariance property” in the sense that it is insensitive to changes to  $\beta$ , while the function  $1_A(x_3)$  has a certain “01-invariance property” in the sense that it is insensitive to interchanges between  $\alpha$  and  $\beta$ . So  $A$  is correlating with a combination of a 02-invariant set and a 01-invariant set, but the correlation only takes place in the extension  $(X^4, \tilde{\mu})$  of  $(X, \mu)$  and the invariance seems to be a bit fancier than just  $\tau\rho^{-1}$ -invariance or  $\sigma\rho^{-1}$ -invariance.



39. [Ryan O'Donnell](#) Says:

[February 27, 2009 at 10:36 pm](#) | [Reply](#)

832. Following up on 830, perhaps Randall or Terry would be able to answer:

Is there an analogue in the ergodic proof of restrictions to these subspaces with more than one “type” of wildcard? (In other words, subspaces with these “ $x_i = x_j$ ” constraints?) In my limited understanding of the ergodic proofs, the operators used therein seemed to be more aligned with just the simpler combinatorial subspaces.



o [Terence Tao](#) Says:



[February 27, 2009 at 10:59 pm](#)

Ryan, I'm not sure I understand your question. A  $k$ -dimensional combinatorial subspace is described by a word with  $k$  wildcards, but each wildcard can be used more than once. For instance,  $xyyzz21$  would describe the three-dimensional subspace of  $[3]^8$  in which  $x_1 = x_2$ ,  $x_3 = x_4$ ,  $x_5 = x_6$ ,  $x_7 = 2$ ,  $x_8 = 1$ . A combinatorial line is the case when  $k=1$ , so we have one wild card  $x$ , which can be used many times over.

o  [Jason Dyer](#) Says:

[February 27, 2009 at 11:06 pm](#)

Are you meaning like how Moser's has a wildcard that goes 1-2-3 while the other goes 3-2-1? So, generalizing, there would be six types of wildcards? (123, 132, 213, 312, 231, 321 — obviously certain combinations would be equivalent to Moser)

o  [Jason Dyer](#) Says:

[February 27, 2009 at 11:26 pm](#)

To clarify further, any pair of distinct wildcards from those six (123, 132, 213, 312, 231, 321) will fix one element and permute the other two, so they are all equivalent to Moser. However, we could take all the way up to six different wildcards, resulting in lines I believe have never been named. For example, taking  $x = 123$ ,  $y = 321$  and  $z = 213$ , then a line formed with  $11xyz$  would be  $11132$ ,  $11221$ ,  $11313$ .

o  [Jason Dyer](#) Says:

[February 27, 2009 at 11:29 pm](#)

Argh. There are pairs that don't fix any of the elements, (123-231 for example) which would also be something different than Moser's.

o  [Ryan O'Donnell](#) Says:

[February 28, 2009 at 9:34 pm](#)

832.x. Re Terry's reply: Right, what I'm saying is, in my limited experience, I hadn't noticed any ergodic-inspired finitary arguments that involved combinatorial subspaces with more than one type of wildcard.

40.  [gowers](#) Says:

[February 28, 2009 at 1:40 am](#) | [Reply](#)

833. Fourier and Sperner

On a slightly different topic, I still haven't given up hope of a "right" Fourier proof of Sperner's theorem. On the wiki I have now written up [an argument that almost gets there](#). The good news is that it gives a Fourier-type decomposition with the property that a suitable equal-slice-weighted sum of  $f(A)f(B)$  over all pairs  $A \subset B$  transforms into an expression where the contribution from every Fourier mode (when this claim is suitably interpreted) is positive. The bad news is that so far the only way I have of proving this positivity, which boils down to proving that a certain explicitly defined integral kernel is positive definite, is using an inverse Fourier argument, so I can't quite claim that the proof is a Fourier proof of Sperner. But now that I know that this kernel is positive definite, I think it will be just a matter of time before one of us comes up with a direct proof, and then there will be a nice clean Fourier proof of Sperner (or at least, I think it would deserve to be called that).

41.  [Terence Tao](#) Says:

[February 28, 2009 at 5:07 am](#) | [Reply](#)

#### 834. Density increment

I think I can now prove Tim's claim in 828, namely that a set that correlates with a 1-set has a density increment on a large dimensional subspace.

First observe, by the greedy algorithm and DHJ(2), that if  $n$  is large enough depending on  $m$  and  $\varepsilon$ , then any dense subset  $A$  of  $[2]^n$  can be partitioned into  $m$ -dimensional subspaces, plus a residual set of density at most  $\varepsilon$ . Indeed one just keeps using DHJ(2) to locate and then delete  $m$ -dimensional subspaces from the set. Similarly, the complement of  $A$  can also be partitioned into  $m$ -dimensional spaces plus a residual set of density at most  $\varepsilon$ .

Now, one should be able to pull this statement to  $[3]^n$  and show that if  $A$  is a dense 1-set of  $[3]^n$ , then one can partition both  $A$  and its complement into  $m$ -dimensional subspaces, plus a residual of small density. I'm not sure what "density" should mean here; equal-slices may be a better bet than uniform at this point.

Now, if  $B$  is a dense set which correlates with a dense 1-set  $A$ , then by picking a moderately large  $m$ , picking  $\varepsilon$  tiny compared to  $1/m$ , and then  $n$  really large, we should be able to use the above partition and the pigeonhole principle to get a density increment of  $B$  on one of these  $m$ -dimensional subspaces.

42.  [gowers](#) Says:

[February 28, 2009 at 9:35 am](#) | [Reply](#)

#### 835. Density increment

Terry, that's similar to things I was trying, and I'm wondering whether it will run into similar difficulties. The point where I get stuck is where you say, "Now, one should be able to pull this statement to  $[3]^n$ ." How might one do this? Well, given any combinatorial subspace in  $[2]^n$  you can easily define an associated combinatorial subspace in  $[3]^n$  by simply allowing the wildcards to take value 2 as well. (For convenience I am taking  $[2] = \{0, 1\}$  and  $[3] = \{0, 1, 2\}$ .) And the result will be a collection of disjoint subspaces. But they won't anything like partition  $[3]^n$  because so far all the fixed coordinates are 0 or 1. We can correct that by allowing the fixed coordinates that used to be 0 to be either 0 or 2, but I think that doesn't get us out of the woods.

What one really needs to answer is the following question: given a sequence  $x$  in  $[3]^n$ , to what combinatorial subspace does it belong?

Here's an example. Suppose  $n = 5$  and you are presented with  $x = 110022$ . And suppose that one of the combinatorial subspaces you took in  $[2]^n$  was  $V = aabcbcb$ . Then you'd be forced to put  $x$  into  $V$ , since its 1-set is contained in  $V$ . But if you now allow the wildcards to take the value 2 as well, you can't get  $x$ .

o  [gowers](#) Says:

[February 28, 2009 at 11:37 am](#)


835.1 Maybe one could use a variant of the argument where having chosen the first wildcard set you then inductively cover evenly everything else. Sorry — that's a bit vague but I have no time to clarify it just yet. In the above example, having chosen  $aa$  you might e.g. then partition into things like  $bcbcb$ ,  $bbccb$ ,  $bccbb$ ,  $cbbbb$ , etc. Except that even that wouldn't be right because once you'd chosen your second wildcard set you'd play the same game with the third, fourth, etc. Sorry, that's probably still completely unclear.

o  [Jason Dyer](#) Says:


[February 28, 2009 at 4:23 pm](#)

I haven't had time to test this yet, but would it help to triplicate the original set into a 9-uniform hypergraph (with subgraphs A, B, and C) such that subgraph A uses 0 or 2, B uses 0 or 1, and C uses 1 or 2?


Then somehow before the last step everything would need to be recombined into the original graph.

- 
[gowers](#) Says:  
[February 28, 2009 at 5:04 pm](#)

835.3 Jason, I'm not sure I follow what you are saying here.


- 
[Randall](#) Says:  
[February 28, 2009 at 5:24 pm](#)

Hopefully this issue will be resolved in a simpler way soon anyway, but could the “measure increment” observation I made in my last post be fashioned into a proof? In the regular F-K conversion they only cared about combinatorial lines, but presumably one could arrange that for any set  $I$  of words, if the intersection over  $I$  of  $T_w^{-1}C$  is non-empty, then you see a subspace-homomorphic image of  $I$  in the original set  $A$  you were dealing with. You could do this for 2 sets simultaneously, a 1-set  $A$  and a correlated set  $B$ . I am guessing that  $A$  would wind up associated with a set in your measure space that would be measurable with respect to the  $\tau_\rho^{-1}$  rigid factor, and that  $B$  would wind up associated with a set that had non-zero projection onto that factor, and you would then wrap up the proof as indicated above. Of course you'd have to convert this to a combinatorial proof, which presumably Terry is good at (energy incrementation?) But again, one hopes it's not that hard, or that something already suggested can be made to work....


- 
[Terence Tao](#) Says:  
[February 28, 2009 at 7:30 pm](#)

Huh, so the pullback of a combinatorial subspace of  $[2]^n$  to  $[3]^n$  is not a combinatorial subspace. That's a bit weird. I agree that my previous comment doesn't quite work as stated then.

However, it does seem that the pullback of a combinatorial subspace of  $[2]^n$  is an “average” of combinatorial subspaces of  $[3]^n$  (or what Gil would call a “fractional cover”). Basically, there is a random way to convert a 2-dimensional subspace e.g. 01aa01bbb10ccc to a 3-dimensional subspace by randomly converting 0s to 0s or 2s, and a wildcard such as a to a, 0, or 2. One has to carefully select the probability measures here (in particular, to make choices between equal slices and uniform) to make everything work, but perhaps it can be pulled off. (One may also want to ensure that each wildcard appears a lot of times so that when one pulls back, one (with very high probability) doesn't accidentally erase all copies of any given wildcard.) As Gil observed, if you can fractionally cover most of a set  $A$  and its complement by large combinatorial subspaces, and  $B$  has a large correlation with  $A$ , then  $B$  has a density increment on one of these subspaces.

- 
[Terence Tao](#) Says:  
[February 28, 2009 at 7:42 pm](#)

No, that's rubbish, I take it back; the pullback of, say, aa, is the set  $\{00, 11, 12, 21, 22\}$ , and I can't cover that space by lines. Strange.

- 
[Jason Dyer](#) Says:  
[February 28, 2009 at 7:44 pm](#)

835.5 I was meaning based on “allowing the fixed coordinates that used to be 0 to be either 0 or 2” you have two

more subspaces where:

coordinates that used to be a 0 are either 1 or 2, and coordinates that used to be a 0 are a 1 and  
coordinates that used to be a 0 are either 0 or 1, and coordinates that used to be a 1 are a 2.

However, I have not been able to get anything useful out of this.

43.  [gowers](#) Says:

[February 28, 2009 at 5:24 pm](#) | [Reply](#)

836 Density increment

I'm going to return to full polymath mode and think online rather than offline, concentrating on the problem of getting a density increment on a subspace if you already have one on a simple set. (Recall that I am defining a simple set to be a set  $\mathcal{A}$  of sequences  $x$  of the form  $\{(U, V, W) : U \in \mathcal{U}\}$ .) Since I'll be talking about  $[2]^n$  quite a bit, it will be natural to take  $[3] = \{0, 1, 2\}$  and  $[2] = \{0, 1\}$ . Set theorists would no doubt approve.

I am aiming for an inductive construction. That is, I want to fix some coordinates and choose a wildcard set in such a way that I am not dead. To say that more precisely, the assumption we are given to start with is that for every  $U \in \mathcal{U}$  the density of points  $(U, V, W) \in \mathcal{A}$  is at least  $\delta + \eta$ . I would now like to find a disjoint quadruple  $(U_1, V_1, W_1, X_1)$  that does *not* partition  $[n]$ , or even come close to doing so, and I would like the following to be the case. Let  $Z_1$  be the complement of  $U_1 \cup V_1 \cup W_1 \cup X_1$ . Then there are many  $U \subset Z_1$  such that the density of  $(V, W)$  with  $(U_1 \cup X_1 \cup U, V_1 \cup V, W_1 \cup W) \in \mathcal{A}$  is at least  $\delta + \eta$  (or perhaps very slightly less), and the same is true, with the same  $U$ , if we move  $X_1$  over to  $V_1 \cup V$  or  $W_1 \cup W$ .

I've got to go in a moment, but it occurs to me that one might be able to get away with less. I've tried to choose a 1-dimensional subspace  $S_1$  such that for many  $U$  the density is good for *every* assignment of the wildcards. But perhaps it's enough just to get the *average* density good when we set the wildcards, and perhaps that's easier to prove.

Incidentally, here's a problem that ought to count as a first exercise. To make sense of it, we need to go for a functional version. So let's suppose that  $f : [3]^n \rightarrow [0, 1]$  is a function of mean  $\delta$  and suppose that there is a dense set system  $\mathcal{U}$  such that for every  $U \in \mathcal{U}$  the average of  $f(U, V, W)$  over all  $(V, W)$  that make a point with  $U$  is at least  $\delta + \eta$ . Can we find a combinatorial line on which  $f$  averages at least  $\delta + \eta'$  (where  $\eta'$  is  $99\eta/100$ , say)?

Of course, we can do it by quoting DHJ(3) but is there an elementary proof?

That's all I've got time for for now.

o  [gowers](#) Says:

[February 28, 2009 at 5:48 pm](#)

836.1 Just to be clear, to make sense of that last problem, one should either use equal-slices measure or ask for  $\mathcal{U}$  to be dense in the probability measure where each element of a set is chosen independently with probability  $1/3$ .

44.  [Terence Tao](#) Says:

[February 28, 2009 at 7:51 pm](#) | [Reply](#)

837. Density increment

Randall, I think the problem we're seeing in the combinatorial world (that the pullback of a  $[2]^n$ -combinatorial subspace is not a  $[3]^n$ -combinatorial subspace) is reflected in the ergodic world that the  $\tau\rho^{-1}$  IP-system does not commute with

the  $\sigma\rho^{-1}$ -IP-system, and so DHJ(2) for the  $\tau\rho^{-1}$ -invariant factor does not seem to quite give us what we want (applying  $\sigma\rho^{-1}(w)$  to a bunch of  $\tau\rho^{-1}$ -fibers does not give another bunch of  $\tau\rho^{-1}$ -fibers).



○ [Terence Tao](#) Says:

[February 28, 2009 at 7:52 pm](#)

Ahh, but there's some sort of "asymptotic commutativity" if one separates the 01 interchange and 02 interchange operations sufficiently. Let me think about this...



○ [Terence Tao](#) Says:

[February 28, 2009 at 8:10 pm](#)

837.2

OK, I think I can make Randall's argument work. In order to exploit asymptotic commutativity, I first need to weaken the notion of a 1-set to a "local 1-set", which is a 1-set on large slices of  $[3]^n$ ; more precisely, A is a local 1-set if there exists a small set  $I \subset [n]$  of "bad" coordinates such that whenever one fixes those coordinates (thus reducing the n-dimensional cube to an  $n - |I|$ -dimensional cube), the slice of A is a 1-set. (Equivalently, A is insensitive to 0-2 interchanges on coordinates outside of I, but could be extremely sensitive to such changes within I). Every global 1-set is a local 1-set, but not conversely.

Suppose a set B has a density increment with a global 1-set A of non-trivial size. Then by a greedy algorithm, we can find a local 1-set A' of non-trivial size on which B has a "near-maximal density increment" with B in the sense that any other local 1-set A'' (with perhaps slightly more bad coordinates than A', and not too much smaller than A') does not have a significantly higher B-density than A' did. (There will be a lot of parameter juggling to sort out to quantify everything here; I will ignore this issue here.)

OK, let's look at the local 1-set A', which has a small number of bad coordinates I'. Let m be a medium sized number. Pick m random small wildcard sets  $I_1, \dots, I_m$  (which with high probability will be disjoint from each other, and from I').

If we then pick a word w at random from  $[3]^n$ , then by DHJ(2.5) and the local 1-set nature of A', with positive probability, the combinatorial space  $V(w)$  of  $2^m$  elements formed by taking w and overwriting each of  $I_1, \dots, I_m$  with 0s, 1s, or 2s, will lie in A'. Call the set of w that do this *good*.

The key point is that the set A'' of good w is itself a local 1-set outside of the bad coordinates  $I'' := I' \cup I_1 \cup \dots \cup I_m$ . And so by hypothesis, B enjoys essentially the same density increment on A'' that it did on A'. But on the other hand, A'' is the union of (parallel) m-dimensional combinatorial subspaces, and so we get a density increment on one of these spaces.



○ [Terence Tao](#) Says:

[February 28, 2009 at 8:18 pm](#)

small correction:  $V(m)$  should of course have  $3^m$  elements, not  $2^m$ . (DHJ(2.5) places  $2^m$  of the elements of  $V(m)$  inside A', and the local 1-set nature of A' then automatically extends this membership of A' to the rest of  $V(m)$ .)

45.



[Ryan O'Donnell](#) Says:

[February 28, 2009 at 10:48 pm](#) | [Reply](#)

838. Structure Theorems; specifically, question for Terry re #826:

Terry, could you clarify on your statement therein, “the energies  $\|S_\lambda f\|_2^2 \dots$  are basically monotone decreasing in  $\lambda$ ”?

Suppose  $f : \{0, 1\}^n \rightarrow \mathbb{R}$  is of the form  $f(x) = h(\sum_{i=1}^n x_i)$ , where  $h : \mathbb{R} \rightarrow \mathbb{R}$  is a nonnegative increasing function. Doesn't it seem as though  $\|S_\lambda f\|_2^2$  will be *increasing* in  $\lambda$ ?



o Ryan O'Donnell Says:

[February 28, 2009 at 10:49 pm](#)

838.1 The nonparsing formula there is  $f : \{0, 1\}^n \rightarrow \mathbb{R}$ . *Now fixed* — Tim



o Terence Tao Says:

[February 28, 2009 at 11:03 pm](#)

838.2

Ryan, I guess one should specify that  $f$  should be bounded within  $[-1, 1]$ , and “basically” means “up to  $o(1)$  errors, and with  $\lambda$  small compared with  $\sqrt{n}$ ”.

In your example  $h$  should also be bounded between  $[-1, 1]$ . In that case, flipping  $\lambda$  0's to 1's will only have an effect of increasing  $f$  by  $O(\lambda/n) = o(1)$  or so on average.

The reason for the monotone decrease is because of the absorption formula  $S_{\lambda'} S_\lambda f \approx S_{\lambda+\lambda'} f$ . Since  $S_{\lambda'}$  is basically a contraction for  $\lambda' = o(\sqrt{n})$  (though it does cease to contract for large  $\lambda'$ , of course), we see that  $\|S_{\lambda+\lambda'} f\|_{L^2} \leq \|S_\lambda f\|_{L^2} + o(1)$ .

I guess the “local” picture (small  $\lambda$ ) is looking a bit different from the “global” picture. Locally, the operation of flipping 0s to 1s is a measure-preserving operation; globally, of course, it isn't. (Tragedy of the commons!)



46. Ryan O'Donnell Says:

[February 28, 2009 at 11:04 pm](#) | [Reply](#)

839. Fourier proof of Sperner.

I semi-checked to myself that one can prove Sperner by increment-free, purely Fourier arguments. One uses:

- . the structure theorem (#821).
- . Mossel's “Lemma 6.2” to handle the  $\langle f_{\text{sens}}, f_{\text{stab}} \rangle$  and  $\langle f_{\text{sens}}, f_{\text{sens}} \rangle$  parts.
- . the triangle inequality argument (#815) to handle the  $\langle f_{\text{stab}}, f_{\text{stab}} \rangle$  parts.

However, this argument is still quite unsatisfactory to me. For one, it requires  $\delta \gg 1/\sqrt{\log \log n}$ . For another, it requires selecting the parameter called  $\epsilon$  in #800 (equivalently,  $\lambda/n$  in Terry's #826) *after* using the Structure Theorem. And most “wrongly”, this parameter must be  $\Theta_\delta(1)/n$ .

In particular, the following theorem is true (I'm 99% sure, at least; proof by “pushing shadows around with Kruskal-Katona”) — however, I don't think we have *any* Fourier-based proof at all:

Theorem: Let  $f : \{0, 1\}^n \rightarrow \{0, 1\}$  have density  $\delta > 0$ . Fix  $p = 1/2 - \epsilon$ ,  $q = 1/2 + \epsilon$ , where  $\epsilon$  is, say,  $1/n^{.75}$ . (Or set  $p = 1/2$  if you like.) Then  $\mathbb{E}[f(X_{p,q})f(Y_{p,q})] > 0$ , where I'm using the notation from [Tim's wiki entry](#).



47. Terence Tao Says:

[February 28, 2009 at 11:06 pm](#) | [Reply](#)

Metacomment.



One side effect of the threaded system is that we hit 100 comments long before the number assigned to the comments increases by 100; we're at 838 now but the thread is already longer than most of the other threads. Perhaps we may wish to renew this thread well before 899 (e.g. at 850)?

48.  [Ryan O'Donnell](#) Says:

[February 28, 2009 at 11:23 pm](#) | [Reply](#)

840. Sperner.

This is very tiny comment. I just wanted to point out that one can prove the “correct” density-Sperner Theorem under the uniform distribution, in the same way Tim proves it under the equal-slices distribution.

Assume  $\mathbb{E}[f] = \delta$  under the uniform distribution. Pick a random chain  $C$  from  $(0, 0, \dots, 0)$  up to  $(1, 1, \dots, 1)$  and then choose  $u, v \sim \text{Binomial}(n, 1/2)$ , independently. Define  $x = C(u)$ , the  $u$ th string in the chain, and  $y = C(v)$ . Then  $\mathbb{E}[f(x)f(y)] = \mathbb{E}_C[\mathbb{E}_u[f(C(u))]^2] \geq \mathbb{E}_{C,u}[f(C(u))]^2 = \delta^2$ , as  $x = C(u)$  is uniformly distributed.

So we have a distribution on pairs of strings  $(x, y)$  such that  $x$  and  $y$  both have the uniform distribution, and such that they form a nondegenerate “line” with probability at least  $\delta^2 - \Theta(1/\sqrt{n})$ .

PS: If someone wants to express  $\mathbb{E}[f(x)f(y)]$  here in terms of Fourier coefficients, I'd be very happy to see it.

o  [gowers](#) Says:

[March 1, 2009 at 11:56 am](#)

840.1 Ryan, I'm not sure how tiny that comment is, and must think about it. Maybe it will turn out that putting the binomial distribution on maximal chains is always more convenient than putting equal-slices measure on the cube. And clearly this same trick works for  $[3]^n$ . This feels to me like a potentially useful further technique to add to our armoury, and when I get the chance I think I'll add something about it to the wiki.

I agree that the Fourier calculation looks more or less compulsory. And something that's rather nice about this measure is that it combines the uniform distribution on  $[2]^n$  with a very natural non-uniform distribution on combinatorial lines, so it provides a potential answer to a question we were trying to answer way back in the Varnavides thread. Indeed, another exercise we should do is prove a Varnavides-type version of DHJ(3) for this measure on the combinatorial lines (given DHJ(3) itself).

Oh no wait a moment — it's not obvious how to generalize to DHJ(3) because we don't have an analogue of maximal chains. This is something else to think about.

49.  [Gil Kalai](#) Says:

[March 1, 2009 at 7:00 am](#) | [Reply](#)

841. A problem where some of the techniques developed here are potentially useful.

Let  $f$  be a monotone Boolean function. Let  $\psi_k(f) = \int I_k^p(f)$  be the integral of the influence of the  $k$ -th variable w.r.t. the probability measure  $\mu_p$ . ( $\psi_k(f)$  is called the “Shapley value” of  $k$ .) Let  $\epsilon$  be a fixed small real number and let  $T$  be the difference between  $p$ - $q$  where  $p$  is the value for which the probability that  $f$  is 1 is  $1 - \epsilon$  and  $q$  is the value for which the probability that  $f$  is 1 is  $\epsilon$ . ( $T$  is called the length of the “threshold interval for  $f$ .”)

It is known that if all  $\psi_k(f) < \delta$  then  $T$  is small as well. The known bound is  $T < 1/C \log \log(1/\delta)$ . The proof relies on connecting the influences for different  $\mu_p$ s. It looks that one  $\log$  could be eliminated and that a more careful understanding of the relation between influences and other properties of  $f$  for different values of  $p$  may be useful.

50.  [gowers](#) Says:

[March 1, 2009 at 12:00 pm](#) | [Reply](#)

Metacomments.

1. Terry I agree that with threading we need shorter threads, if you'll excuse the dual use of the word "thread". Let's indeed go for 850.

2. The other thing is that I've been meaning to say for ages that I loved your notion that the wiki could be thought of as an online journal of density Hales-Jewett studies, and the discussions as weekly conferences.

51.  [gowers](#) Says:

[March 1, 2009 at 12:28 pm](#) | [Reply](#)

842. General remark

I have a very non-mathematically busy day today, so I've got time for just one brief post. First, let me say that I'm quite excited by Terry's 837.2. I haven't fully digested it yet but I have partially digested it enough to see that it is very definitely of the sort of flavour I was hoping for. It's frustrating not to be able to spend a few hours today playing around with the argument. And then tomorrow and Tuesday I have two heavy teaching days.

So instead, let me throw out a small thought/question. Recall that in the dictionary I gave in 829, long arithmetic progressions go to combinatorial subspaces in  $[2]^n$ . Therefore, if the general plan of modelling an argument on Ajtai-Szemerédi works, it will have to give a new proof of the corners theorem that is similar to their proof but avoids the use of Szemerédi's theorem. As I've already explained, that sounds discouraging at first, but I think it may in fact not be. But I've subsequently realized that these thoughts lead inexorably to a suggestion that Jozsef made, way back in [comment 2](#) (!), that one should look at a generalization of the corners problem in which you don't have long APs in the two sets of which you take the Cartesian product. Maybe the time is ripe for having a go at that problem, and in particular seeing whether the Ajtai-Szemerédi approach can be pushed through for it.

52.  [Ryan O'Donnell](#) Says:

[March 1, 2009 at 5:25 pm](#) | [Reply](#)

843. Measures:

One reason I said the comment in #840 is tiny is that I'm finally starting to catch up the fact (well known to the rest of you, I think) that if you don't mind density increment arguments (and we certainly don't!) then the underlying probability measure is quite unimportant. Or rather, you can pass freely between any "reasonable" measures, using the arguments sketched [here](#) on the wiki.

In fact, I think "equal-slices" is a good "base measure" to always return to. It has the nice feature that if you have equal-slice density  $\delta$ , by density-increment arguments you can assume that you have density very nearly  $\delta$  under *almost all* product measures.

53.  [Ryan O'Donnell](#) Says:

[March 1, 2009 at 5:29 pm](#) | [Reply](#)

844. Density increment / the problem from #828:

What if  $\mathcal{A} = \mathcal{B}$  is a random simple set in  $[3]^n$  of density  $\delta$ , meaning that its  $\mathcal{U}$  is a random set of density  $\delta$  under the  $1/3$ -biased distribution?

How should we try to find a combinatorial subspace to increment on?

(Perhaps we shouldn't; perhaps instead we should exploit the fact that this  $\mathcal{A}$  is extremely 23-insensitive.)



o [Terence Tao](#) Says:

[March 1, 2009 at 5:58 pm](#)

Ryan, that's right; being 20-insensitive, we can invoke DHJ(2.5) and get large subspaces in  $\mathcal{A} = \mathcal{B}$ , which of course is a healthy density increment.



o [Ryan O'Donnell](#) Says:

[March 1, 2009 at 7:36 pm](#)

844.2. Ah, sorry if you said this already in an earlier comment, Terry; I'm still catching up.

54. *Anonymous* Says:

[March 1, 2009 at 7:06 pm](#) | [Reply](#)

845. Density increment in cubes

Tim, Re:842. In case you are considering corners in cubes, let me mention a few things here. Reading some recent blogs, I thought that you are close to prove a density increment result which would imply the corner result. I write the statement first and then I will argue that it proves a Moser type result.

Given an  $IP_n$  set (or an  $n$ -dimensional Hilbert cube if you wish) with  $2^n$  elements. Consider the elements as vertices of a graph. The graph is  $c$ -dense, i.e. the number of edges is  $c2^{2n}$ . The graph contains a huge hole, there is an  $\alpha 2^n$  empty subgraph. Prove that then it contains a  $d$ -dimensional subcube on the vertices spanning at least  $(c + \delta)2^{2d}$  edges.  $\delta > 0$  might depend on  $c$  and  $\alpha$  but not on  $n$ , and  $d$  grows with  $n$  (arbitrary slowly). Now I will try to post the next part as a new thread.



55. [jozsef](#) Says:

[March 1, 2009 at 7:11 pm](#) | [Reply](#)

845.1

oops, I didn't give my name for the previous note



56. [jozsef](#) Says:

[March 1, 2009 at 7:26 pm](#) | [Reply](#)

846. A Moser type theorem

845.2 ... and apparently I don't know how to generate a thread.


Still going backwards, let me state the theorem which would follow from 845: Every dense subset of  $[3]^n$  contains three elements, formed by taking a string with one or more wildcards  $*^1$  and  $*^2$  in it and replacing the first type wildcards by 1,2 and 3, and the second type wildcards by 4,3, and 2, respectively. For example 123412, 223312, and 323212 is such triple.



57. [jozsef](#) Says:

[March 1, 2009 at 7:29 pm](#) | [Reply](#)

846.1. In the previous statement we were in  $[4]^n$ , not in  $[3]^n$ .

58.  *josef* Says:  
[March 1, 2009 at 7:56 pm](#) | [Reply](#)


846 contd.

With Ron Graham we proved a colouring version of 846. For any  $\log n$  colouring of  $[4]^n$  there is a monochromatic triple like in 846. The paper is available at

[http://www.math.ucsd.edu/~sbutler/ron/pre\\_cube.pdf](http://www.math.ucsd.edu/~sbutler/ron/pre_cube.pdf)


The reasonable bound for the colouring version shows that this problem might be easier to work with than with DHJ  $k=3$ . In the next post I will try to prove that 845 implies 846, that is, an “arithmetic density increment” in a graph with a huge hole implies a Moser type statement.

(Sorry about the number of posts, but I don't want to write long ones as even the shorter ones are full with typos)

59.  *josef* Says:  
[March 1, 2009 at 9:29 pm](#) | [Reply](#)

847. (part one)

Here we show that a density increment argument implies the Moser-type theorem in 846. It isn't new that a density increment would give us what we are looking for, the new feature is that it might be easier to prove density increment in a graph with a huge empty subgraph. For practical reasons we switch to a bipartite settings. Consider a bipartite graph  $G(A,B)$ , between two copies of the  $n$ -dimensional cube,  $A = B = \{0, 1\}^n$ .  $G(A,B)$  has  $2^{n+1}$  vertices and, say,  $c2^{2n}$  edges. There is a natural one to one mapping between the set of edges and a pointset of  $[4]^n$ . The two vertices of an edge are 0-1 sequences. From the two sequences we get one as follows. The  $i$ -th position of the new sequence (0,1,2, or 3) is given by  $x + 2^y$  where  $x$  is the number (0 or 1) in the  $i$ -th position of the vertex in  $A$  and  $y$  is the number in the  $i$ -th position of the vertex in  $B$ . We say that two edges span the same cube if substituting all 1-s and 2-s by a wildcard,  $*$ , we get the same sequence. Observe that if  $(v1,v2)$  and  $(w1,w2)$ , two edges of  $G(A,B)$ , span the same cube and  $(v1,w2)$  is in  $G(A,B)$ , then we are done, there are three points formed by taking a string with one or more wildcards  $*^1$  and  $*^2$  in it and replacing the first type wildcards by 1,2 and 3, and the second type wildcards by 4,3, and 2, respectively. (For example  $*^123*^212$  gives 123412, 223312, and 323212.) Equivalently, it will give a corner on the Cartesian product of an  $n$ -dimensional Hilbert cube. I will continue, but I would like to see if the typing is correct so far or not.

60.  *josef* Says:  
[March 1, 2009 at 10:41 pm](#) | [Reply](#)

847. (part two)

We said that two edges span the same cube if substituting all 1-s and 2-s by a wildcard we get the same 0,3,\* sequence. Similarly we can say that every edge,  $e$ , spans a particular subcube; substitute all 1-s and 2-s with wildcards to get a 0,3,\* sequence. Any edge, which differs from  $e$  in the  $*$  positions only, is in the subcube spanned by  $e$ . Our moderate goal here is to show that if there are no three edges of the form  $(v1,v2)$ ,  $(w1,w2)$ , and  $(v1,w2)$  in  $G(A,B)$  that  $(v1,v2)$  and  $(w1,w2)$  span the same subcube - in which case we were done - then there is a subcube spanned by many edges containing many other edges (see in post 845). The vertices of the spanning edges provide us the large empty subgraph. We will follow a simple algorithm. At first consider the set of edges spanning the whole cube. Those are edges without 0 or 3. If there are at least  $\alpha n$  such edges, then stop. If there are less than  $\alpha n$  such edges, then select the densest  $n-1$  dimensional subcube. (Such subcubes are spanned by edges having exactly one 0 and no 3-s and by edges with one 3 and no 0-s.) If the densest subcube is spanned by at least  $\alpha(n-1)$  edges then stop, otherwise select the densest  $n-2$  dimensional subcube. If we repeat the algorithm and  $\alpha \ll c$ , then it should terminate in a  $c$ -dense  $d$ -dimensional subcube spanned by at least  $\alpha 2^d$  edges. By the initial assumption there is no edge connecting the end-vertices of the spanning edges. It gives us the huge empty bipartite graph,  $K_{\alpha n/2, \alpha n/2}$ . I didn't do the actual calculations, but it seems

correct to me.

61.  [gowers](#) Says:

[March 2, 2009 at 12:18 am](#) | [Reply](#)

848. Density increment

Terry, I've tried to understand your argument properly but there are a few points where I've had difficulty working out what you mean. I think they all stem from one main point, which comes in the last paragraph where you say "by hypothesis". I'm not sure I know what the hypothesis is. You established that it was impossible to get a substantial density *increase* on a local 1-set, but here you seem to need that you can't have a substantial density *decrease*. Have I misunderstood something? I hope so. (I'm actually writing this offline in my car so by the time I get to post it perhaps this matter will have been cleared up.)

62.  [gowers](#) Says:

[March 2, 2009 at 12:19 am](#) | [Reply](#)

849. Different measures (also written earlier but offline)

I want to think some more about Ryan's observation that the permutations trick can be used with the uniform measure. In particular, I want to understand from a non-permutations point of view what the resulting measure is on pairs  $A \subset B$ .

So let's fix two sets  $A \subset B$  of sizes  $r$  and  $s$ , respectively. The probability that a random permutation gives both  $A$  and  $B$  as initial segments is  $\binom{n}{s}^{-1} \binom{s}{r}^{-1}$  and the probability that we choose those two initial segments (if we binomially choose a random  $r$  and a random  $s$  without conditioning on their order) is  $2^{-2n} \binom{n}{r} \binom{n}{s}$ . So the probability that we choose the sets  $A$  and  $B$  is  $2^{-2n} \binom{n}{r} \binom{s}{r}^{-1}$ , which equals  $2^{-2n} \frac{n!(s-r)!}{s!(n-r)!}$ , which equals  $2^{-2n} \binom{n}{s} \binom{n-r}{s}^{-1}$ .

Therefore, the probability of choosing  $B$  conditional on having chosen  $A$  is  $2^{-n} \binom{n}{s} \binom{n-r}{s}^{-1}$ . If  $A$  and  $B$  were independent, this probability would of course be  $2^{-n}$ , so the extra weight is ... hmm, I'm not really getting anywhere here.

63.  [gowers](#) Says:

[March 2, 2009 at 12:20 am](#) | [Reply](#)

850. Different measures (again written earlier)

How about the Fourier expansion? Suppose we choose  $U$  and  $V$  according to this distribution. That is, we choose a random permutation, and then choose initial segments  $r$  and  $s$  independently with binomial probabilities on  $r$  and  $s$ . What is the expected value of  $w_A(U)w_B(V)$ ?

To make the calculation friendlier, I'm going to work out something else that I hope will turn out to be the same. It's more like Ryan's  $p$ - $q$  approach. To pick a random point I'll pick a number  $m$ , binomially distributed with parameters  $n$  and  $1/2$ , and I'll then set  $p = m/n$  and choose a point with the measure  $\mu_p$ . Damn, that doesn't work, because the probability that I end up choosing the empty set is strictly greater than  $2^{-n}$  (because there's a chance I'll pick it if  $p$  takes a typical value round  $1/2$ , and also a probability  $2^{-n}$  of taking  $p = 0$ ). OK, it's believable that any average of binomial distributions that ends up with mean  $n/2$  is guaranteed to be less concentrated than the binomial distribution itself, so probably this idea was doomed.

OK I don't at the moment see a conceptual argument for calculating the Fourier expression. Perhaps it's a question of a

brute-force calculation that one hopes will cancel a lot. But it would be disappointing if the expectation of  $w_A(U)w_B(V)$  above were not zero when  $A \neq B$ .

64.  [gowers](#) Says:

[March 2, 2009 at 12:22 am](#) | [Reply](#)

Metacomment: I'm about to start a new thread, which will start at 851. So perhaps further comments on this thread could be limited to a few small replies.

## Leave a Reply

Logged in as [Ryan O'Donnell](#). [Logout »](#)

☐ Notify me of follow-up comments via email.

---

[Blog at WordPress.com.](#)

[Entries \(RSS\)](#) and [Comments \(RSS\)](#).

☺



# Gowers's Weblog

Mathematics related discussions

« [Brief review of polymath1](#)  
[Problem solved \(probably\)](#) »


## DHJ(3): 851-899

By gowers

Once again there is not a huge amount to say in this post. Since the last post there have been a few additions to the [polymath1 wiki](#) that may be of some use. In particular, there is now a collection of fairly complete write-ups of related results (see the section entitled “Complete proofs or detailed sketches of potentially useful results”) to which I hope we will add soon. Also on the wiki is [an account of the Ajtai-Szemerédi proof of the corners theorem](#), which seems to have some chance of serving as a better model for a proof of DHJ(3) than the proof via the triangle-removal lemma. Meanwhile, progress has been made in understanding and to some extent combinatorializing the ergodic-theoretic proof of DHJ(3), ideas from which have fed into the discussion. As with the last post, this one is mainly to stop the number of comments getting too large. We're now down to 50 comments per post (except that it was 51 for the last one and will be 49 for this), since with the new threading we seem to be averaging at least one reply per comment.

This entry was posted on March 2, 2009 at 12:37 am and is filed under [polymath1](#). You can follow any responses to this entry through the [RSS 2.0](#) feed. You can [leave a response](#), or [trackback](#) from your own site.

### 113 Responses to “DHJ(3): 851-899”

1.  [Terence Tao](#) Says:  
[March 2, 2009 at 2:03 am](#) | [Reply](#)


851. Density increment

I'll start off (perhaps a bit awkwardly) by responding to Tim's query from 848, which was asking about how a hypothesis that a density increment is near-maximal implies that there are also no significant density *decrements*. This is basically just coming from subtraction.

Let  $A$  be a set, and suppose that  $B$  is a structured set (specifically, a local 1-set of non-trivial size) on which one has a near-maximal density  $\delta_B := |A \cap B|/|B|$ . Then any structured subset  $B'$  of  $B$  of non-trivial size also has to have density close to  $\delta_B$ . It can't have a much bigger density, since otherwise we could pass from  $B$  to  $B'$  and get a density increment; and it can't have a much smaller density, since otherwise we could pass from  $B$  to  $B \setminus B'$  and also get a density increment.

There are three parameters that will need to be juggled here to make it all work: the density of  $A$  in  $B$ , the density of  $B$  in  $[3]^n$ , and the number of bad coordinates in  $B$ . The last parameter should not cause a difficulty; as  $n$  is so large, we can afford to turn quite a lot of coordinates bad and still retain a lot of usable structure. So it should boil down to making

sure that the greedy algorithm that locates  $B$  terminates in a bounded amount of time without sending the density of  $B$  to zero. The way I think of it is like this: if  $B$  contains a small subset  $B'$  on which  $A$  has too low of a density, then one can remove  $B'$  from  $B$  to improve the density without significantly shrinking the size of  $B$ . I don't think this process can be iterated indefinitely, as the density will eventually shoot up to 1.

- 
[Terence Tao](#) Says:  
[March 2, 2009 at 2:06 am](#)


851.1

In ergodic theory land, what we are doing here is stating the obvious fact that if a function  $f$  has mean  $\delta$  and is non-constant, then there is a set  $B$  of non-trivial measure on which  $f$  is uniformly bounded below by  $\delta + c$  for some  $c > 0$ . Here,  $f$  is the “conditional expectation of  $1_A$  to the sigma-algebra generated by local 1-sets”. The problem in the finitary world is that the local 1-sets don't actually form a sigma-algebra and so  $f$  only exists in some “virtual” sense rather than a tangible one, and so one has to do various greedy algorithm contortions to simulate any ergodic theory computation involving  $f$ .

- 
[gowers](#) Says:  
[March 2, 2009 at 9:04 am](#)


851.2 Terry, what worries me in your explanation above is the “of non-trivial size”. When you apply the result, you apply it to the set  $A''$  of all good words. Now it appears to me that the density of this set goes to zero as  $m$  goes to infinity (and we need  $m$  to go to infinity in order to carry on with the density-increment argument), and therefore I don't see how ...

Ah, I think I do see what you are saying now, and I have in fact used this kind of argument myself sometimes. The point (which you yourself have already said) is that if you remove only a small set then you don't greatly change the size of  $B$ . But this can work only if your structured sets are closed under subtraction. So for instance if you had a maximal density on a subspace, you couldn't conclude that you had near maximal density on small subspaces of that subspace. But if you remove a local 1-set from a local 1-set you end up with a local 1-set so it's OK.

- 
 2. [Ryan O'Donnell](#) Says:  
[March 2, 2009 at 4:08 am](#) | [Reply](#)

852. Different measures.

Tim, re your #849, I also found that binomial probability distribution on  $A \subseteq B$ 's surprisingly ... “awkward” to analyze. I suppose one can just really grind things out and see if, say, the Fourier expansion turns out nicely after all.

- 
 3. [Terence Tao](#) Says:  
[March 2, 2009 at 6:17 am](#) | [Reply](#)

Metacomment.

As I had some idle moments, I decided to spin out a timeline of the story so far for polymath1:

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

It's remarkably nonlinear, but that's how maths research goes, I guess.

Also: this post needs a “polymath1” tag.

4.  [Terence Tao](#) Says:

[March 2, 2009 at 8:05 am](#) | [Reply](#)

#### 853. Density increment

I just realised that the “density increment” trick underlying Randall’s “dense fibres” argument that I translated into a combinatorial argument is not really a density increment, but rather a “mass” increment argument, essentially the same as that used to prove the classical [Hahn decomposition](#) in measure theory (which, by coincidence, I [taught in my class](#) a few weeks ago).

Let’s say  $A \subset [3]^n$  has density  $\delta$ , and let  $f = 1_A - \delta$  be the balanced function. Let  $\mathcal{B}$  be a family of “structured sets” (such as local 1-sets); this family is secretly supposed to be behaving like a sigma-algebra, but let us try to avoid using that fact. Suppose we have a large set  $B$  in  $\mathcal{B}$  on which  $A$  has a density increment, and more specifically we have  $\mathbb{E}_{x \in [3]^n} f(x) 1_B(x) \geq c$  for some  $c > 0$ ; note that this hypothesis forces  $B$  to be somewhat dense. We now want to restrict  $B$  to the fibres of  $\mathcal{B}$  on which  $f$  is positive. One way to do this is to try to maximise the “mass”  $\mathbb{E}_{x \in [3]^n} (f(x) - c/2) 1_{B'}(x)$  as  $B'$  ranges over elements of  $\mathcal{B}$ ; note that we are not dividing by the density of  $B'$ , so this measures a mass increment rather than a density increment. Setting  $B' = B$  we know that the mass can be positive. If we find a  $B'$  that maximises the mass, we know that any non-trivial structured subset  $B''$  of  $B'$  will have an  $A$ -density significantly larger than  $\delta$  (e.g. larger than  $\delta + c/4$ ), since otherwise we could remove this “negative mass” portion  $B''$  from  $B'$  and increase the mass of  $B'$  substantially. (Not coincidentally, this is *exactly* how one proves the Hahn decomposition theorem.)

5.  [gowers](#) Says:

[March 2, 2009 at 10:09 am](#) | [Reply](#)

#### 854. Ajtai-Szemerédi

First a general remark to add further motivation to the Ajtai-Szemerédi approach. Let’s briefly think about why it has never been extended to prove a result for three-dimensional corners. If you try to do it, you soon see that what you want in place of the applications of Szemerédi’s theorem is applications of the full 2-dimensional Szemerédi theorem, and it seems that the original Ajtai-Szemerédi argument is not powerful enough to do that (not that I’ve thought hard about why).

But if we managed to push through to a proof of DHJ(3), it seems not outrageous to hope that we might be able to get a multidimensional version, and that that might be what was needed for DHJ(4), and so on. That’s all a bit pie-in-the-sky at the moment, but that’s all it’s meant to be.

Assuming that the Randall/Terry argument is fine (as I do, but for my own benefit I hope to write it up on the wiki soon), where have we actually got to? Here is my assessment, which gets a bit vague in the details. What it definitely gives us is what seems (and that’s a pretty confident “seems”) to be the appropriate analogue of step 2 of the Ajtai-Szemerédi argument [as it is presented on the wiki](#).

That tells us that for almost every  $U$  we have pretty much the expected density of  $(V, W)$  such that  $(U, V, W) \in \mathcal{A}$ . Let us call such a  $U$  good. The obvious analogue of Step 3 is then this: we can find  $W$  with a positive density of pairs  $(U, V)$  such that  $U$  is good and  $(U, V, W) \in \mathcal{A}$ .

An unthinking analogue of the fourth step would simply be that amongst these pairs  $(U, V)$  can be found a combinatorial subspace of dimension tending to infinity. Given the experience with Step 2, I think it pretty unlikely that this will be good enough, but let’s try to understand why not.

Fifth step: in the Ajtai-Szemerédi argument they have a long AP of points on the diagonal, which gives rise to a long AP  $P$  of vertical lines. They then partition the horizontal lines into long APs with the same common difference as  $P$ , using

the fact that an AP of vertical lines intersects an AP of horizontal lines with the same common difference in a grid.

What could the analogue of this be in the subspace world? Well, an AP of vertical lines translates to taking a combinatorial subspace of  $U$ s and then extending each  $U$  in all possible ways to a point  $(U, V, W)$  in  $[3]^n$ . And when would a structure like that intersect a similar one for  $V$ s in a combinatorial subspace? If we denote by  $\langle Z_0, Z_1, \dots, Z_m \rangle$  the combinatorial subspace that consists of all finite unions of the  $Z_i$  that include  $Z_0$ , and we write  $U(Z_0, \dots, Z_m)$  for the set of all  $(U, V, W)$  with  $U \in \langle Z_0, Z_1, \dots, Z_m \rangle$ , then the answer is that  $U(Z_0, Y_1, \dots, Y_m)$  intersects  $V(Z_0, Z_1, \dots, Z_m)$  in a combinatorial subspace if  $Y_0$  and  $Z_0$  are disjoint, and  $Y_i = Z_i$  for  $i = 1, 2, \dots, m$ .

This is almost saying is that the combinatorial subspaces are translates of each other, but there are two additional complications. First, the sets that determine the translates (that is,  $Y_0$  and  $Z_0$ ) have to be disjoint. Secondly, we cannot partition  $[2]^n$  with translates of a given subspace because all points in all translates will be constant on the wildcard sets, whereas a general point definitely isn't.

So there's a lot of sorting out to do if an analogue of Step 5 is to be made to work. But what's encouraging is that these difficulties are rather similar in flavour to the difficulties that we've already encountered and dealt with in Step 2.

6.  [gowers](#) Says:

[March 2, 2009 at 10:16 am](#) | [Reply](#)

Metacomment. A quick remark about teaching. Terry comments that he has just taught the Hahn decomposition theorem, and implies (I think, reading between the lines) that this has helped him to have certain thoughts more quickly and easily than he would otherwise have had them. So that prompts me to say that I am lecturing probability this term, and noticed when I was thinking about Fourier analysis and equal-slices measure how it made a definite difference to be completely 100% on top of things like expectations of products of independent random variables, conditional probability, etc. etc. The odd thing in that case was that the concepts were quite basic, but nevertheless there was a gain: what was gained was some kind of clarity and fluency that replaced the normal process of muddling through first and tidying up afterwards.

o  [Terence Tao](#) Says:

[March 2, 2009 at 9:08 pm](#)

Well, when I taught Hahn decomposition a few weeks ago, I thought to myself "Ah, another density/mass increment argument! These arguments are everywhere!". But other than reinforcing my (already quite strong) belief that energy/density/mass increment arguments are an essential tool in analysis, I don't think it directly helped me except in the *a posteriori* sense, when my attempts to explain Randall's argument more coherently begin to ring a bell and I consciously went back to look at the proof of Hahn decomposition. But perhaps there was indeed some unconscious "priming" due to the fact that I have been spending the last few weeks revisiting the foundations of soft analysis to teach to the UCLA freshman graduate students.

(Certainly, in order to convert ergodic theory arguments to finitary ones, one does have to think really hard about deconstructing extremely basic notions in measure theory and functional analysis - for instance, deconstructing the construction of conditional expectation, or the Riesz representation theorem, or the approximation of measurable sets by open sets, or the fact that every bounded monotone sequence converges, are essential first exercises before one can do anything even remotely nontrivial, e.g. finitising the mean ergodic theorem.)

7.  [gowers](#) Says:

[March 2, 2009 at 6:33 pm](#) | [Reply](#)

855. Density increment.

I'm trying to write up the Randall/Terry argument on the wiki, but it's very slow going as I keep getting stuck. And


each time I get stuck it takes me ages to decide whether I've uncovered a genuine difficulty or simply not got the point. And I'm in that position yet again.

Here's my problem this time. Suppose I've passed to a local 1-set  $A'$  where  $B$  has maximal density. So we've got some coordinates belonging to an index set  $I$ , and membership of  $A'$  depends just on what happens inside  $I$  and what the 1-set of a sequence is outside  $I$ . (I'm using "1-set" in two senses here.) Let's call this an  $(I, 1)$ -set.

Now I'm very happy with the idea that if we pass to a sub- $(I, 1)$ -set, then we can assume that  $B$  has near-maximal density there too: if it hasn't, then we just subtract off the subset. But that's not what is going on in the argument here. Here we define a  $(I \cup I_1 \cup \dots \cup I_m)$ -set, and it's much less obvious to me that we can subtract it off, since if we do, then we are not left with a  $(I, 1)$ -set.

Could we use the fact that we have a  $(I_1 \cup \dots \cup I_m)$ -set after we've done the subtraction? I don't see how, because the amount of measure we've subtracted is tiny (since the probability that a word  $w$  is good goes to zero as  $m$  goes to infinity) and the size of the set of bad coordinates has gone up a lot.

I suppose what I'm really saying is that I'd be very grateful to see this argument written out in full detail. (Another problem, but I think I can solve this one, is finding a suitable measure on the set of  $I_1, \dots, I_m$  to make it true that words are good with positive probability.)


- o  *Ryan O'Donnell* Says:  
[March 2, 2009 at 6:48 pm](#)

855.2. Hmm, perhaps our #855's are coming to a similar difficulty, to be discussed further.

I just wanted to say in this short reply that in your parenthesis at the end seems to match the parenthesis in the second paragraph of what I wrote below. I don't quite know how to solve this issue, but I do like to remember that the Gunderson-Rodl-Sidorenko paper Jozsef pointed us to shows that whenever you have a subset of  $\{0, 1\}^m$  of positive density, it must completely contain an  $\Omega(\log \log n)$ -dimensional subspace. (This is a quantification of the "multidimensional Sperner" argument on the wiki, I think.)

- o  *Randall* Says:  
[March 2, 2009 at 8:48 pm](#)

Terry will answer this far more fully at some point but in the meantime I will say this much: once  $m$  is fixed, the positive probability that  $V(w)$  lies in  $A'$  is bounded below by something like  $\delta^{2^m}$  and the size of the  $\alpha_i$  can be safely bounded above by some finite amount....

- 8.  *Ryan O'Donnell* Says:  
[March 2, 2009 at 6:42 pm](#) | [Reply](#)

856. I wonder if I could ask for some clarification on the Randall/Terry argument; specifically, on #837.2.

In particular, I'm trying to follow the last paragraph. (Actually, I don't quite understand the appeal to DHJ(2.5) in the previous paragraph either, but I think I understand my own way around that.)

Could we clarify precisely what the definition of  $A''$  should be? If it's defined exactly as written in #837.2, then I wouldn't so much say it's the union of parallel  $m$ -dimensional subspaces so much as I would say it's union of  $(|I_1| + \dots + |I_m|)$ -dimensional subcubes. Which would be fine, but this doesn't seem to square with the idea that  $A'' \subseteq A'$  as discussed in #851.

On the other hand, perhaps we shouldn't take  $A''$  to be *all* good  $w$ , but just those  $w$  such that the part in  $I_1 \cup \dots \cup I_m$

is actually *in* the large-dimensional subspace  $V(m)$ . But my worry here is that in general (if the “wildcard sets” are large) that this  $A''$  will constitute only a negligible ( $o_n(1)$ ) fraction of  $A'$ , so it should be impossible to conclude that  $B$  has a density increment on  $A''$ .



o Ryan O'Donnell Says:

[March 2, 2009 at 6:43 pm](#)

856.2. Oops, sorry for the simultaneous post with Tim's #855; please consider what I wrote #856. [Changed -- Tim]



9. [gowers](#) Says:

[March 2, 2009 at 7:43 pm](#) | [Reply](#)

857. Varnavides version of multidimensional Sperner

Ryan, it was amusing to see just how many similarities there were between our near-simultaneous comments. While we're waiting for a response, here's roughly how I expected to fill in the details with the appeal to DHJ(2.5). The obvious worry, which you have expressed before in a different context, is that you have to ensure that the wildcard sets can take plenty of different sizes: it's not enough to go for “easy” combinatorial subspaces where they all have size 1, since  $\mathcal{U}$  could be the set of all sets of even size.

The thought I had was that one could define the measure in a way that more or less guaranteed that the proof worked. You would first find an  $M$  such that every subset  $A$  of density  $\delta/2$  in  $[3]^M$  contained a subspace of dimension  $m$  where you got an element of  $A$  whenever you fixed the wildcards to be 1 or 2. Then you would choose your random wildcard sets  $I_1, \dots, I_m$  as follows. First choose a random “easy” combinatorial subspace of dimension  $M$ , with ground set  $F$ , and then pick, uniformly at random,  $m$  disjoint non-empty subsets of  $F$ . I'm not 100% sure that works but it feels as though a double-counting argument should now do the trick.



10. [gowers](#) Says:

[March 2, 2009 at 8:35 pm](#) | [Reply](#)

858. Density increment

I'm going to try to approach this question in small stages, proving little lemmas without being sure they are useful — though I will try to explain why I think they should be.

Suppose that one tries to build up a subspace inductively on which  $\mathcal{A}$  has a density increment. The problem one faces is this. We can assume that for every  $U \in \mathcal{U}$  there are too many  $(V, W)$  such that  $(U, V, W) \in \mathcal{A}$ , and we can even find a combinatorial subspace of such  $U$ , but we do not have a good grip on what happens when we convert that into a combinatorial subspace in  $[3]^n$ .

Let's just look at combinatorial lines. I'd like to find a subset  $E \subset [n]$  and a combinatorial line inside  $[3]^E$  such that for all its three points we get too many ways of completing them with 2s and 3s to form an element of  $\mathcal{A}$ . The problem is that while I may be able to find several  $U \subset E$  with too many completions, at some point I'm going to need  $U_1 \subset U_2$  such that not only do they both have too many  $(V, W)$ s, but if I fill  $U_2 \setminus U_1$  with 2s or 3s then there are *still* too many  $(V, W)$ s.

Now there might be some hope of that if I could say that for almost every  $U \in \mathcal{U}$  it was the case that for almost all small sets  $G$ , if I fill  $G$  with 2s or 3s, then there are still too many  $(V, W)$ s that complete to a combinatorial line. Why? Because then I could use the fact that there is a positive density of lines (all this depends on the appropriate measures being used) to find a combinatorial line such that  $U_2 \setminus U_1$  behaved itself.



So can we say that about  $G$ ? Well, suppose that for a positive proportion of the  $U \in \mathcal{U}$  there is a positive proportion of small sets  $G$  that have too few extensions when you fill them with 2s. Then there must be a positive proportion of small sets  $G$  such that you get too many extensions when you fill them with 2s. And then you can pick a random  $G$  and get a dense set of  $U$ s such that you get too many extensions. So if we drop down to  $[n] \setminus G$  then we have a strengthening of our original hypothesis in that combinatorial subspace.

So now suppose we've reached the point where we can't play this game any more. Then my hope is that you get that for almost every  $U \in \mathcal{U}$  (and therefore all once you chuck out the bad ones) it is the case that for almost all  $G$  you can fill them with 2s or 3s and still get roughly the right number of extensions. But with positive probability  $U \cup G$  also belongs to  $\mathcal{U}$  (by one of our Varnavides-versions of Sperner) and then perhaps we have something.

That's written quickly and carelessly, so I'm not sure it's right. But if it is, then maybe one can run more or less the same argument with  $m$ -dimensional subspaces instead.



11. [Terence Tao](#) Says:

[March 2, 2009 at 8:44 pm](#) | [Reply](#)

858. Dense fibres argument

Responding to 855, 856, etc.

Firstly, I forgot to define  $A''$  in 837.2, as pointed out by Ryan.  $A''$  is the union of the  $V(w)$  as  $w$  ranges over all good words. It has a smallish density that depends on the density of the original 1-set  $A$ , the constants in DHJ(2.5), the number  $m$ , and the size of the wildcard sets  $I_1, \dots, I_m$ , but - crucially - the density does not depend on the size of the pre-existing bad coordinate set  $I'$ . And this is important, because the size of  $I'$  is going to explode (but, importantly, will always be of size  $O(1)$  as far as  $n$  is concerned).

As noted by Tim, if  $A'$  is an  $(I', 1)$  set, then  $A' \setminus A''$  is not going to be an  $(I', 1)$  set - it's going to be merely an  $(I'', 1)$ -set, where  $I'' = I' \cup I_1 \cup \dots \cup I_m$ . I like to think of the coordinates  $I_1, \dots, I_k$  as being "consumed" or "used up" by the procedure of passing from  $A'$  to  $A' \setminus A''$ , in that the operation of interchanging 1 to 2 on these coordinates has been "spent". But the remaining  $n - |I''| \approx n$  interchanging operations remain "unspent", and that's good enough for our purposes. (This is the "IP" philosophy from ergodic theory; it's OK if each operation can only be performed once, so long as one has a huge number of essentially identical copies of that operation to spend.)


I don't think we can just fix  $I'$  in advance and maximise some mass over all  $(I', 1)$  sets; one has to do some greedy algorithm of the following sort:

1. We start with a 1-set  $A$  on which the function  $f := 1_A - \delta$  has mass  $\mathbb{E}f 1_A \geq c$ , and a desired dimension  $m$  that one wants to get a  $c/4$ -density increment on.
2. DHJ(2.5) should tell us that any 1-set of density at least  $c/100$  should contain a lot of  $m$ -dimensional combinatorial subspaces in which the wildcard sets  $I_1, \dots, I_m$  have size at most  $r$  for some  $r = O_{m,c}(1)$ . Indeed, if one chooses a random such subspace (using an appropriate measure), the probability that the entire subspace lies in the set should be at least  $\varepsilon$  for some  $\varepsilon \gg_{m,c} 1$ .
3. One should be able to localise DHJ(2.5) to  $(I, 1)$ -sets rather than 1-sets, so long as  $|I|$  is much smaller than  $n$ , by the first moment method.
4. Initialise  $I' = \emptyset$  and  $A' = A$ .
5.  $A'$  is a  $(I', 1)$  set with  $\mathbb{E}(f - c/2) 1_{A'} \geq c/2$ . In particular,  $A'$  has density at least  $c/2$ .
6. Suppose that there exists disjoint  $I_1, \dots, I_m$  in  $[n] \setminus I'$  of size at most  $r$ , and a  $(I'', 1)$ -set  $A''$  of density at least  $\varepsilon$  in  $A'$ , where  $I'' := I' \cup I_1 \cup \dots \cup I_m$  such that  $B$  has density at most  $\delta + c/4$  on  $A''$ . Then replace  $I', A'$  by  $I'', A''$ ; this


increases  $\mathbb{E}(f - c/2)1_{A'}$  by at least  $c\varepsilon/4$ . Now return to Step 5.

7. We can only loop  $O_{c,\varepsilon}(1)$  times, and so the net size of  $I'$  at the end of the day is  $O_{c,\varepsilon,m,r}(1)$ , and we end up with an  $(I', 1)$  set  $A'$  of density at least  $c/2$  such that  $B$  has density at least  $\delta + c/4$  on every subset  $A''$  of the form discussed in Step 6. Now invoke DHJ(2.5) to find an  $A''$  of this form which is the union of  $m$ -dimensional subspaces, and we obtain the desired density increment.

[Amazing how much messier the finitary argument is from the ergodic theory argument, where one simply takes the set where the conditional expectation of  $1_B$  to the sigma-algebra of  $\mathcal{O}_2$ -invariant sets is bigger than  $\delta + c/2$  to be  $A'$ !]

- o  [Terence Tao](#) Says:  
[March 2, 2009 at 8:51 pm](#)

A small correction: the  $\varepsilon$  in Step 6 should actually be  $3^{-mr}\varepsilon$ . (The set of good  $w$  has density  $\varepsilon$ , but the set  $A'$ , formed from the union of the  $V(w)$ , could have a slightly smaller density of  $3^{-mr}\varepsilon$ . But this is no big deal - as long as it doesn't depend on  $n$  or the size of the wildcard set  $I'$ , one is OK.)

- 12.  [Terence Tao](#) Says:  
[March 2, 2009 at 8:58 pm](#) | [Reply](#)

## 860. Multidimensional Sperner


I believe one can get the right sort of multidimensional Sperner just by iterating the right sort of DHJ(2).

If  $A$  is a subset of  $[2]^n$  of density  $\delta$ , then there should exist an  $r = O_\delta(1)$  such that if one picks a random  $r'$  between 1 and  $r$ , then picks a random wildcard set  $I$  of size  $r'$ , then picks a random combinatorial line with that wildcard set, then with probability at least  $\varepsilon = \varepsilon(\delta, r) > 0$ , the line will lie in  $A$ . (In fact one can probably take  $r$  to be about  $1/\delta$  and  $\varepsilon$  to be about  $\delta^2$ , using the usual chain proof of Sperner. If one uses equal-slices measure then things may be particularly favourable.)


In particular, if we pick  $r'$  and  $I$  randomly, then the set  $A' \subset [2]^{n \setminus I}$ , defined as the set of all words  $w \in [2]^{n \setminus I}$  whose associated combinatorial line with wildcard set  $I$  lies in  $A$ , will have density at least  $\varepsilon/2$  with positive probability. In particular there exists  $r', I$  for which this statement is true.

If we iterate this  $m$  times (replacing  $\delta$  by  $\varepsilon/2$  at each stage) we conclude that there exist disjoint wildcard sets  $I_1, \dots, I_m$  of size  $O_{m,\delta}(1)$  such that the proportion of combinatorial  $m$ -spaces with these wildcards that completely lie in  $A$  is at least  $\varepsilon_m$  for some  $\varepsilon_m = \varepsilon_m(\delta) > 0$ .

This is the multidimensional DHJ(2). The standard derivation of DHJ(2.5) from DHJ(2) should then give the multidimensional DHJ(2.5) needed for my argument.


- o  [Terence Tao](#) Says:  
[March 2, 2009 at 9:00 pm](#)

The argument here, of course, is identical to the usual proof of the Szemerédi cube lemma (that dense subsets of  $[n]$  contain high-dimensional cubes), by first iterating the fact given that a subset  $A$  of  $[n]$  of density  $\delta$  one can find a positive  $h$  such that  $A' := (A + h) \cap A$  has density  $\gg \delta^2$ .

- o  [josef](#) Says:  
[March 2, 2009 at 10:39 pm](#)

Terry, I'm not sure that I understand what are you saying here. Would it lead to a Varnavides-type result? Here

are the numbers that I see; If a set of subsets of  $[n]$  has at least  $c_d 2^n / n^{1/2^d}$  elements, then it contains a  $d$ -dimensional subspace. This bound is close to be sharp. Using this we see that a  $c$ -dense subset of  $2^{[n]}$  contains at least  $c n^{1-1/2^d} 2^n$   $d$ -dimensional subspaces, which is not much. Your calculation suggests much more.

- 
[Terence Tao](#) Says:  
[March 2, 2009 at 11:12 pm](#)


Dear Jozsef,

One has to count the subspaces in an appropriate equal-slices fashion to get the right Varnavides-type result. (This was discussed quite a while back, I think in the 1-199 thread, but no harm in reviving it here.)

Let's start with  $m=1$ : if  $A$  has density  $c$  in  $2^{[n]}$ , then a random maximal chain  $\emptyset = S_0 \subset S_1 \subset \dots \subset S_n = [n]$  in  $2^{[n]}$  will intersect  $A$  in about  $c\sqrt{n}$  positions in the middle  $O(\sqrt{n})$  of the chain. [The calculation is a little simpler here if one uses equal-slices density for  $A$  instead of uniform density, but never mind that.] The pigeonhole principle then tells us that if we randomly choose an  $r$  between 1 and  $O(1/c)$ , and randomly pick  $i = n/2 + O(\sqrt{n})$ , then  $S_i, S_{i+r}$  will both land in  $A$  with probability about  $c^2$ .

To put it another way, if we pick  $r$  randomly between 1 and  $O(1/c)$ , and then randomly select from all the  $2^{n-r} \binom{n}{r}$  combinatorial lines with  $r$  wildcards, then the probability that this line will lie in  $A$  is  $\gg c^2$  (there may be some logarithmic losses of  $\log \frac{1}{c}$  due to the Chernoff inequality etc.) Note that this is **not** the same as choosing a combinatorial line uniformly at random among all  $3^n - 2^n$  such lines, or even amongst all such lines with  $O(1/c)$  wildcards.

This Varnavides-type theorem gives a weighted count of combinatorial lines in the set; a line with  $r$  wildcards has a weight proportional to  $1/2^{n-r} \binom{n}{r}$ . This is the one which I think one should iterate, and not the theorem based on raw (unweighted) cardinality of lines.

- 
[jozsef](#) Says:  
[March 2, 2009 at 11:32 pm](#)

Thank you Terry!

- 
[gowers](#) Says:  
[March 3, 2009 at 12:06 am](#) | [Reply](#)

861. Density increment.

I'm going to have yet another go at expressing the argument in terms that I understand. Or rather, I'm going to try to prove the result, using what Terry has written as a huge hint as to what kind of argument I can expect to work. However, I do not plan to give anything like full details in this comment: I just want to convince myself of the argument.

The set-up once again: I have a 1-set  $\mathcal{A}$  (I'd better call it  $\mathcal{A}$ , following Terry, to avoid confusion, though earlier I've been calling it  $\mathcal{B}$  and the dense set that correlates with it  $\mathcal{A}$ ) defined as the set of all  $(U, V, W)$  such that  $U \in \mathcal{U}$ ; I have a set  $\mathcal{B}$  of density  $\delta$  such that the density of  $\mathcal{B}$  inside  $\mathcal{A}$  is at least  $\delta + \eta$ ; I want a density increase for  $\mathcal{B}$  on a combinatorial subspace with dimension tending to infinity.

Step 1. For every  $U$  define  $d(U)$  to be the probability that if you randomly partition  $[n] \setminus U$  into two sets  $V, W$  then the point  $(U, V, W)$  belongs to  $\mathcal{B}$ . Then we are assuming that the average of  $d(U)$  over all  $U \in \mathcal{U}$  is  $\delta + \eta$ . (One of the details I'm going to be very non-explicit about is what measures I am using.) By an easy averaging argument we may pass to a dense subset of  $\mathcal{U}$  such that  $d(U) \geq \delta + \eta/2$  for every  $U$  in the dense subset. Let's replace  $\mathcal{U}$  by this subset

and  $\eta$  by  $\eta/2$  and still call them  $\mathcal{U}$  and  $\eta$ .

Step 1.5. The obvious thing to do next is pass to an  $m$ -dimensional combinatorial subspace (in the  $[2]^n$  sense) that lives inside  $\mathcal{U}$ . Let us write it as  $Z_0 \cup FU(Z_1, \dots, Z_m)$ , which stands for the set of all sets  $Z_0 \cup Z$ , where  $Z$  is a union of some of the  $Z_i$ . (The  $F$  stands, unnecessarily, for “finite”.) But if we do that then we find ourselves wondering why we bothered, since if we treat  $Z_0$  as a fixed set of 1s and the  $Z_i$  as wildcard sets and try to fix the remaining coordinates to get a combinatorial subspace of  $[3]^n$  with a density increase, then we simply don't succeed. It could be that as soon as we fill the  $Z_i$  with constant strings of 2s or 3s, we end up with sequences that give rise to hardly any extensions that belong to  $\mathcal{B}$ . (Just to explain slightly more, our aim is to find a combinatorial subspace such that the average density of possible extensions to a point in  $\mathcal{B}$  is too high, so that then we can turn things round and say that a random extension comes from too many points in the combinatorial subspace, which gives us our density increment.)

Step 2. At this point, we say to ourselves, “Yes, but if that disaster happened all the time, then surely it would have to be compensated for by exploitable density increases elsewhere.” So now we are trying to produce an averaging argument. At this stage, let's forget that  $m$  is supposed to tend to infinity, and treat it as a very large fixed constant. (This is a standard move — if we can always do it for sufficiently large  $n$ , then we simply work out what  $n$  we needed in terms of  $m$ , invert that function, and we've got our function of  $n$  that tends to infinity. Terry has been using this trick a lot — in fact, easy though it is, it should go on the Tricky.)

Given that we are trying to produce an averaging argument, we had better look not just for *one* combinatorial subspace, but for a dense set of combinatorial subspaces. So now let's assume we've got a Varnavides-type result for multidimensional Sperner that tells us that with positive probability (depending on the density of  $\mathcal{U}$  and on  $m$  but not on  $n$ ) if we choose our  $Z_0, Z_1, \dots, Z_m$  randomly (according to some carefully chosen probability distribution) then  $Z_0 \cup FU(Z_1, \dots, Z_m) \subset \mathcal{U}$ . Here I'm thinking of  $Z_0$  as being a fairly typical  $U$  and  $Z_1, \dots, Z_m$  as being very small sets.

Now let's suppose that however we do that we find that, to our annoyance, there is some way of assigning 1s, 2s and 3s to the sets  $Z_1, \dots, Z_m$  and filling  $Z_0$  with 1s, such that the density of ways of filling the rest of  $[n]$  with 2s and 3s to get an element of  $\mathcal{B}$  is a bit less than  $\delta + \eta$  (by some rather small amount that depends on  $m$ ). So each  $Z_0$  is associated with a positive density (depending on  $m$ ) of “bad” sequence fragments, formed by taking some small sets  $Z_1, \dots, Z_m$  and assigning 1s, 2s and 3s to them. Turning this round we can find a many fragments with a lot of “bad”  $Z_0$ s.

Now if I generate a sequence by picking  $Z_0, Z_1, \dots, Z_m$  randomly, filling  $Z_0$  with 1s the rest with 1s, 2s and 3s and the rest of  $[n]$  with 2s and 3s, then I get almost exactly the same distribution as if I had just chosen a random sequence (or at least, I believe that one can easily ensure this). And I think we can use that to argue that if many fragments give rise to several “bad”  $Z_0$ s (by which I mean that the density of extensions by 2s and 3s that give rise to elements of  $\mathcal{B}$  is too low) then there must be many fragments that give rise to several “good”  $Z_0$ s. But then by averaging we can fix one of those fragments and pass to a  $(|Z_1| + \dots + |Z_m|)$ -codimensional subspace inside which we have a 1-set consisting of sets  $U$  for which  $d(U)$  has gone up very very slightly (by an amount that depends on  $m$ ).

Step 3. And now we just iterate. If the density goes up by  $c = c(m, \delta, \eta)$  each time, and if our sets  $Z_i$  have size bounded above by  $r$ , then we need something like  $mr/c$  to be distinctly less than  $n$ . If that is the case, then the iteration will stop at some point where we can find a combinatorial subspace  $Z_0 \cup FU(Z_1, \dots, Z_m)$  inside  $\mathcal{U}$  with no possible bad assignments of 1s, 2s and 3s to the  $Z_i$  with  $i \geq 1$ . And that means that for every element of that combinatorial subspace we get too many 2,3-extensions that belong to  $\mathcal{B}$ . Turning that round, we pick 2s and 3s randomly outside  $Z_0 \cup \dots \cup Z_m$  and get on average a density increase in the subspace.

Terry, I'd be interested to know whether you think that's exactly the same as your argument, or roughly the same, or the same apart from an error that I've introduced, or correct and slightly different, or wrong for a reason that's hard to correct but once you manage to correct it you get precisely your argument, or what.



o [Terence Tao](#) Says:


[March 3, 2009 at 12:35 am](#)

861.1

Tim, I think this is essentially the same argument, but I like how you use an additional pigeonhole principle to freeze (and then discard) all the “bad” coordinates, thus avoiding the need to distinguish between global and local 1-sets. It’s not as if there was anything I was planning to do with those bad coordinates anyway, so one may as well pigeonhole them away.

(Incidentally, if one does some Ramsey-theoretic preprocessing of the set, it may turn out that one has enough “stationarity” that one does not need to pass to subspaces at all... but this is overkill and would in any event produce worse quantitative bounds.)


I guess the bounds will in fact end up being reasonably civilised at this stage; the bad set  $I$  of coordinates is only increasing linearly with the number of iterations, as you point out, not exponentially or anything.

o  [Terence Tao](#) Says:  
[March 3, 2009 at 12:41 am](#)

Some minor remarks:

- I don’t think the averaging argument in Step 1 is actually needed; one can work with the global density of  $B$  in  $A$  rather than the local densities  $d(U)$  of  $B$  associated to  $U$ . (If it later turns out that there are lots of  $U$ s in  $A$  with too low of a  $B$ -density, then this will only help us when the time comes to make our density/mass increment.)

- In order not to have to keep balancing the density of  $B$  in  $A$ , against the density of  $A$ , it is probably better to stare at the mass  $\mathbb{E}(1_A - \delta - \eta/2)1_B$  rather than the density, as this (a) seems to go up by a non-trivial amount at every stage of the iteration, and (b) it automatically provides a lower bound on the density of  $A$ , which is needed to stop the amount removed from  $A$  at each stage of the iteration from shrinking to zero.

14.  [gowers](#) Says:  
[March 3, 2009 at 1:28 am](#) | [Reply](#)

862. Density increment.

I want to throw out another thought, which is that there is something we would be silly not to try to do, though it may turn out not to be feasible. Actually, it may be one of those cases where Randall and Terry will be able to give an informed guess whether it has any chance of working by looking at things from an ergodic perspective. It looks as though we now know that correlation with a 1-set implies a density increase, but what about correlation with a 12-set?

What is a 12-set? It is what on the wiki I called a special set of complexity 1: you take set systems  $\mathcal{U}$  and  $\mathcal{V}$  and define  $\mathcal{A}$  to be the set of all  $x$  with 1-set in  $\mathcal{U}$  and 2-set in  $\mathcal{V}$ . Equivalently, it is the set of  $(U, V, W)$  with  $U \in \mathcal{U}$  and  $V \in \mathcal{V}$ . (Thus, a 1-set is a 12-set where  $\mathcal{V} = [3]^n$ .)

If the argument on the wiki that says that line-free sets correlate with 12-sets is correct, and if correlation with a 12-set implies correlation with a subspace, then the whole problem is solved, so the motivation for this question is rather obvious.

A few preliminary thoughts. First, I think it is easy to prove (by imitating the DHJ(2.5) argument) that a dense 12-set  $\mathcal{A}$  contains multidimensional subspaces. You just choose some random copy of  $[2]^M$  inside  $[3]^n$  (where  $M$  is small enough for a random point in this random copy to be approximately uniformly distributed). In that copy,  $\mathcal{A}$  will on average be dense, and therefore contain a multidimensional subspace in the  $[2]^n$  sense. And then, since  $\mathcal{A}$  is a 12-set, we can allow the wildcards to be 3 and we still live inside  $\mathcal{A}$ . (I think this argument has probably appeared several times


already in our discussions — I may even be one of the people to have given it.)

That proof should give a Varnavides-type statement too. So what would the next stage be?

Let's choose a random combinatorial subspace as follows. First we choose random large sets  $U_0$  and  $V_0$ . Then we choose random small wildcard sets  $Z_1, \dots, Z_m$ . And then we ... er ... hope that the density of  $\mathcal{B}$  inside the resulting subspace (where we assign values 1 to  $U_0$ , 2 to  $V_0$ , anything we like to the  $Z_i$ , and 3 to the rest) is at least  $\delta + \eta/2$ .

If that always fails to be the case, then we argue as follows. With some small but positive probability  $p$ , the random combinatorial subspace is a subset of  $\mathcal{A}$ . That means that for some random assignment of 1s, 2s and 3s to the  $Z_i$  we get too few pairs  $(U_0, V_0)$  belonging to  $\mathcal{B}$  out of the ones that belong to  $\mathcal{A}$ . (By the way, I should confess that I'm losing the thread here and don't really know whether what I'm writing is correct. But I'll just plough on.) So that should be balanced by some other random assignment that gives too many pairs. And then we can get a subspace with a small density increase on a 12-set.

I must go to bed in the knowledge that that argument could be anywhere along the spectrum that goes from complete nonsense at one end to a proof of DHJ(3) at the other. (For now I'll leave unspecified what measure I am placing on that spectrum.) Perhaps someone will be able to tell me by the time I get up tomorrow.

15.  [gowers](#) Says:  
[March 3, 2009 at 2:57 pm](#) | [Reply](#)

#### 863. Density increment

Here's a slightly more precise, but still far from checked, version of what I said in 862.

1. This is a way that one could choose a random point  $(U, V, W) \in [3]^n$ . You choose a random disjoint pair  $(U_0, V_0)$  and some random small wildcard sets  $Z_1, \dots, Z_m$ . You assign values to the wildcards, fill  $U_0$  with 1s,  $V_0$  with 2s and put 3s everywhere else. That gives you your point, and it's more or less uniformly distributed.
2. Now let's condition on  $Z_1, \dots, Z_m$  and the assignment of their values. (I'll call this a random sequence fragment.) If the probability that  $(U, V, W) \in \mathcal{B}$  given the random sequence fragment is ever more than  $\delta + \eta$  (plus a tiny tiny amount) then we've got a density increase on a 12-set in a subspace, and can iterate.
3. Therefore, this conditional probability cannot be more than  $\delta + \eta$  or we're done.
4. But there is a positive probability that  $U_0 \cup FU(Z_1, \dots, Z_m) \subset \mathcal{U}$  and  $V_0 \cup FU(Z_1, \dots, Z_m) \subset \mathcal{V}$ . Moreover, the density of  $\mathcal{B}$  inside the subspace  $\langle U_0, V_0; Z_1, \dots, Z_m \rangle$  (I hope it's easy to guess what that means) is at most  $\delta + \eta/2$  or we're done. So, turning things round, there is in fact a positive probability that the conditional probability in 3 is somewhat less than  $\delta + \eta$ .
5. Therefore, by 1 (I think) there is a positive probability that it is somewhat more than  $\delta + \eta$  and we're done.

Again, that sort of feels as though it could be correct, but it also feels as though it could collapse when I try to write it out properly.

o  [gowers](#) Says:  
[March 4, 2009 at 2:13 am](#)

863.2 I see now that Step 2 is not very clear because I am not mentioning  $\mathcal{U}$  and  $\mathcal{V}$ .

If in Step 1 we discard the point and try again unless  $U \in \mathcal{U}$  and  $V \in \mathcal{V}$  then we end up with a roughly uniformly distributed point in  $\mathcal{A}$ . (That is, the restriction of the distribution to  $\mathcal{A}$  is roughly uniform.)



Then in Step 2 I want to fix  $Z_1, \dots, Z_m$ , and the probability I'm talking about is the probability of being in  $\mathcal{B}$  given that you are in  $\mathcal{A}$ . It is this that cannot go above  $\delta + \eta$  unless it is possible to get a density increase and an iteration.

If nobody sees an obvious flaw in this approach by tomorrow morning (my time) then I suppose I'll try to work out some details properly. I'm slightly suspicious of Steps 4 and 5.

One tiny remark for clarification: this is not supposed to be the Ajtai-Szemerédi approach: it's a different argument that would combine the density increase on a 12-set (as given on the wiki) with a Randall/Terry style argument to get from that to a density increase on a subspace. My dream for the project as a whole is that we'll end up with lots of different proofs ...


16.  [Randall](#) Says:

[March 3, 2009 at 8:45 pm](#) | [Reply](#)

864. Speculative reply to 862.

From the ergodic perspective, it looks like correlation with a dense 12-set implying a mass increment is still on the order of double recurrence. However, the picture I seem to be getting is that maybe it's worth pursuing anyway. Here is the idea. Say we assume some fancy double recurrence result (FDR) that is not, on the face of things, as general as DHJ(3). Like for example, what are called IP Szemerédi or IP Roth on the Wiki (comments 2 and 469, respectively). And say we can use this to show that correlation with a dense 12-set implies a density increment on a subspace. Then, assuming that really was all that was needed in the first place, we will have reduced DHJ to FDR, which might not be close to finishing, but might at least be progress.

The reason I think this might be feasible is, in the ergodic world the projection  $P$  onto the asymptotically 02 invariant sets commutes with the projection  $Q$  onto the asymptotically 01 invariant sets. So,  $PQ$  ought to project to  $L^2$  of some larger sigma-algebra. Correlating with something measurable with respect to this sigma-algebra gets you big fibers, you just need a way to bring these fibers back to themselves along a subspace. That's where FDR comes in. I think you can probably work in a product space with genuine corners now. The idea is you want to show that  $A \cap T_{\{01\}}A \cap T_{\{02\}}A$  is non-empty (I suppress  $\alpha$  in the notation, which tells you which coordinates to flip...also you want only to flip things that are zero; strictly speaking the notation I am using probably doesn't mean anything taken literally, it's only meant to be suggestive.) Well, on the horizontal coordinate, 1s and 2s are interchangeable, so  $T_{\{01\}} = T_{\{02\}}$  (more or less). On the vertical coordinate, 0s and 1s are interchangeable, so  $T_{\{01\}} = \text{Id}$  (more or less). So basically, you have  $T_{\{01\}}$  acting on each coordinate, and you need the putative corners type result (FDR) to bring sets in the Cartesian product back to themselves.

o  [Terence Tao](#) Says:  
[March 4, 2009 at 12:28 am](#)

Dear Randall, I think in fact that double recurrence for 12-sets collapses to single recurrence by the argument sketched out in 862. It's easiest to explain for the corners problem: to find corners  $(x,y), (x+r,y), (x,y+r)$  in a Cartesian product  $A \times B$ , it suffices to just find the latter two points  $(x+r,y), (x,y+r)$  of the corner inside this product, as this will automatically place the third guy inside the corner as well. So the corners theorem in this case follows trivially from the pigeonhole principle; and for similar reasons, DHJ(3) for 12-sets follows from DHJ(2).

In ergodic language: if  $f = f_1 f_2$ , where  $f_1$  is  $S$ -invariant and  $f_2$  is  $T$ -invariant, then the double recurrence integral  $\int f S^n f T^n f$  collapses to the single recurrence integral  $\int S^n (f_1^2 f_2) T^n (f_2^2 f_1)$ . In the DHJ world, if  $f = f_1 f_2$ ,  $f_1$  is  $\rho\sigma^{-1}$ -invariant, and  $f_2$  is  $\tau\sigma^{-1}$ -invariant, then

$$\int \sigma(f) \rho(f) \tau(f) \approx \int \rho(f_1^2 f_2) \tau(f_1 f_2^2)$$

at least for words  $w$  that are sufficiently “large”.



o [Randall](#) Says:

[March 4, 2009 at 3:34 am](#)

864.2 Clarification.

The set  $A$  I am trying to bring back is not a 12-set. It is a set that is measurable with respect to the sigma-algebra of 12-sets. (Notice that, unlike i-sets, ij-sets do not form an algebra.)

In general, the aim of my last post was to outline a method for obtaining a density increment on a subspace for a set that correlates with a 12-set. If I understand Tim correctly, that may be the last piece. What I hope is possible is to fill in that gap with a double recurrence theorem that is, on the face of things, weaker (and hopefully easier to prove directly) than DHJ (3).



o [Randall](#) Says:

[March 4, 2009 at 3:36 am](#)

864.3 Obviously I meant “measurable with respect to the sigma algebra generated by the 12-sets”; although what I actually typed is an amusing oxymoron.



17. [Ryan O'Donnell](#) Says:

[March 4, 2009 at 1:41 am](#) | [Reply](#)

865. Structure Theorem/FK-McCutcheon approach.

I know we have been making progress lately on Tim's Ajtai-Szemerédi approach, but I've been thinking also about Terry's approach, started in #818, of finitising Randall's proof of the Furstenberg-Katznelson argument. Call me craven, but there's something nice about also working an angle that “in principle” is known to succeed. Perhaps we might even end up with a glorious amalgam of the two approaches.

The crux of Terry's [wiki version](#) seems to be a proposed structure theorem which splits any function  $f$  into a “01-almost periodic relative to 12-low influence” part and a “01-uniform relative to 12-low influence” part. It's slightly tricky for me to keep even the definitions in mind, and I had a moment of despair contemplating actually executing such a structure theorem.

Hence I fantasise that some kind of simple structure theorem (a la #821 and #826) might also do the job. Let me throw this fantasy out there and see if it can survive for a bit.

Put a graph on  $[3]^n$  where vertices are connected by an edge if they have Hamming distance 1. Define an operator  $L^{01}$  on functions  $f : [3]^n \rightarrow \mathbb{R}$  by  $L^{01}f(x) = f(x) - \mathbb{E}[f(x')]$ , where  $x'$  is formed from  $x$  by choosing a random coordinate and flipping it between 0 and 1 (if it is not 2). Similarly define the operator  $L^{12}$ . For  $t \geq 0$ , define also the operator  $H_t^{01} = e^{-tL^{01}}$ , and similarly  $H_t^{12}$ . Note for intuition that if we had defined  $H_t^{01}$  for functions on  $[2]^n$ , we would have gotten the operator  $T_{e^{-t/n}}$  (as used in #822). These operators form a semigroup.

So my hope is to have the structure theorem be

$$f = H_t^{12} H_{t'}^{01} f + (f - H_t^{12} H_{t'}^{01} f)$$

for some carefully chosen small quantities  $t, t'$ . Let me give an extraordinarily non-rigorous argument for why this might work out.

First, is  $H_t^{12} H_{t'}^{01} f$  “01-almost periodic relative to 12-low influence”? Well, in fact perhaps it's even 01-low influence.

The idea here is that if  $t$  and  $t'$  are both very small, then hopefully  $H_t^{12}$  and  $H_{t'}^{01}$  approximately commute. Then  $H_t^{12} H_{t'}^{01} f \approx H_{t'}^{01} H_t^{12} f$ , and the latter function has 01-low influence because it has an  $H_{t'}^{01}$  out front.

And what about  $g := (f - H_t^{12} H_{t'}^{01} f)$ ? Is it “01-uniform relative to 12-low influence”? At the grossest level, I take this to mean, is it true that  $\langle H_s^{01} g, h \rangle$  is small for any 12-low influence  $h$ ? (Note that  $H_s^{01}$  should be self-adjoint so this should be the same as asking if  $\langle g, H_s^{01} h \rangle$  is small for 12-low influence  $h$ .) By definition, this correlation is

$$\langle H_s^{01} f, h \rangle - \langle H_s^{01} H_t^{12} H_{t'}^{01} f, h \rangle.$$

Again, in the second term here we hope to commute the  $H^{12}$  operator to the front, use self-adjointness to put it on the other side, and get

$$\langle H_s^{01} f, h \rangle - \langle H_{s+t'}^{01} f, H_t^{12} h \rangle.$$

Now I again fantasise that perhaps by picking  $s$  and/or  $t$  carefully (at a “plateau in the energy spectrum”) we can say that  $H_s^{01} f \approx H_{s+t'}^{01} f$ . Then we’d get

$$\langle H_s^{01} f, h \rangle - \langle H_s^{01} f, H_t^{12} h \rangle$$

which equals

$$\langle H_s^{01} f, h - H_t^{12} h \rangle.$$

But  $h$  is 12-low influence so  $h \approx H_t^{12} h$  and hence this is indeed small.

—

I’m normally much soberer than this post would suggest.



o [Terence Tao](#) Says:

[March 4, 2009 at 4:44 am](#)

Dear Ryan,

I’m hopeful on this long-term strategy too, but from experience with finitising the ergodic proof of Szemerédi’s theorem, I’m pretty sure that the proof is going to be yuckier than the Ajtai-Szemerédi-flavoured approach.

01-uniformity of a function  $f$  relative to 12-low influence is a bit stronger than what you’re saying. It’s the assertion that

$$\mathbb{E} g(\ell(0)) f(\ell(1)) h(\ell(1))$$

is small for all bounded  $g$  and all 12-low influence  $h$ , where  $\ell$  ranges over lines. The point here is that  $h$  depends not only on the fixed coordinates of  $h$ , but also on the location of the wildcards. The statement you’re saying is roughly equivalent to saying that (1) is small in the case when  $h$  doesn’t depend on the location of the wildcards.


Perhaps I can motivate things a bit better by considering the analogous notion when counting three-term progressions. Here, the analogue of “12-low influence” is simply “constant”, and a function  $f : \mathbb{Z}/N\mathbb{Z} \rightarrow [-1, 1]$  is “uniform” (the analogue of “01-uniform relative to 12-low influence”) if one has

$$\mathbb{E}_{n,r} g(n) f(n+r) h(r)$$

small for all bounded  $g$  and all bounded  $h$  which are independent of  $n$ .

Unfortunately, the structure theorems that show up in ergodic theory, once one leaves the “1-step” world and

moves to the “higher-step” or “relative” world, are a bit messy, especially in the finitary world. In principle the machinery from my paper will be relevant here, but I’m reluctant to unpack it all now; I’m hoping that the Ajtai-Szemerédi thread may lead to some new insights or simplifications first. (Also, Tim Austin may soon be coming out with some new work that could also assist with this; more on this once it is more firm.)

- o  [Terence Tao](#) Says:  
[March 4, 2009 at 4:46 am](#)

I forgot to say that the key point in the length three progressions example was that while  $h$  did not depend on the  $n$  parameter (which roughly corresponds to the “fixed positions” of the line), it still depended on the  $r$  parameter (which correspond to the “wildcards” of the line). Because of this, I don’t see an easy structure theorem here... the cheapest way is to take Fourier transforms and extract out the large Fourier modes, but this can’t really be done just by heat operators.

- o  [Randall](#) Says:  
[March 4, 2009 at 5:50 am](#)

865.3


Ryan, it might help to consider the functions

$$f(U, V, W) = e(\sum_V -y_i + \sum_W y_i),$$

$$g(U, V, W) = e(\sum_V y_i),$$

$$h(U, V, W) = e(\sum_W -y_i).$$

Then  $f(w(0))g(w(1))h(w(2)) = 1$  for all variable words  $w$ , so your structure theory had better not be counting  $g$  as “01-uniform relative to 12-low-influence” or some similar category. On the other hand,  $g$  looks like it may be orthogonal to 12-low-influence.

- 18.  [Terence Tao](#) Says:  
[March 4, 2009 at 5:11 am](#) | [Reply](#)

866. 12-sets

Randall raises a good point about 12-sets not being closed under union, though they are still closed under intersection. To borrow a notation I floated a while back, one should make a distinction between “basic 12-sets” (the intersection of a 1-set and a 2-set, which to continue the topological analogy might be thought of as “sub-basic 12-sets”), and “general 12-sets”, which are unions (a bounded number of) of basic 12-sets. In the iterative process that is carving out the 12-set  $A'$  on which to locate a density increment, we are removing basic 12-sets from  $A'$  at each iteration, and so the complexity of the 12-set  $A'$  (i.e. the number of basic 12-sets needed to union together to make  $A'$ ) increases with the iteration.

This is going to cause a problem. Some sort of regularity lemma might be needed here (the ergodic analogue of this would be the Lebesgue differentiation theorem: every general 12-set is very dense inside a basic 12-set). But this might not be enough.

The other alternative is to take advantage of the freedom in the finitary world to restrict to smaller subspaces in one’s hunt for a density increment; this is a trump card that the ergodic world doesn’t really have with current technology. There is a chance that by continually restricting, one may only have to work with basic 12-sets and not general 12-sets. I’ll think about it...

 [gowers](#) Says:

[March 4, 2009 at 11:07 am](#)

866.1 This connects in an amusing way with a terminological difficulty I had, and that you picked up, on the wiki. A basic 12-set is like a rank-1 tensor, and at one point I tried to use the word “rank” instead of “complexity” for that reason. But I was inconsistent about it.

If we went on to DHJ(4) things would get more complicated still. For example, there's an important distinction between a 12-set (an intersection of a 1-set and a 2-set) and a  $\{1,2\}$ -set (a set  $A$  such that whether or not  $(U,V,W,X)$  belongs to  $A$  depends only on the pair  $(U,V)$ ). Then a  $\{1,2\}$ -function of rank at most  $k$  would be a sum of at most  $k$  basic  $\{1,2\}$ -functions, etc. etc.

 [Randall](#) Says:

[March 4, 2009 at 7:39 pm](#)

866.2

Terry, having read Tim's very exciting posts more carefully, I see what you meant now. Yes, I raised the point of the 12-sets not forming an algebra, but apparently I was too caught up in the ergodic paradigm to ignore it, as I should have. In the ergodic setting one develops a habit of thinking not about how one function correlates with another, but rather how a function projects to a subspace. Hence my counterproductive impulse to immediately go searching for an algebra on which to project.

19.  [gowers](#) Says:

[March 4, 2009 at 11:57 am](#) | [Reply](#)

867. Density increment

I'm going to give a sort of commentary on 863 (which I reproduce here) as a first step towards either making it more precise or discovering a problem with it.

*1. This is a way that one could choose a random point  $(U, V, W) \in [3]^n$ . You choose a random disjoint pair  $(U_0, V_0)$  and some random small wildcard sets  $Z_1, \dots, Z_m$ . You assign values to the wildcards, fill  $U_0$  with 1s,  $V_0$  with 2s and put 3s everywhere else. That gives you your point, and it's more or less uniformly distributed.*

I don't foresee any problems with this step. It would use the uniform measure on  $[3]^n$  and exploit the fact that slices near  $(n/3, n/3, n/3)$  all have roughly the same size. The size of the wildcard sets would be some large constant that depended on the density of  $\mathcal{B}$ .

*2. Now let's condition on  $Z_1, \dots, Z_m$  and the assignment of their values. (I'll call this a random sequence fragment.) If the probability that  $(U, V, W) \in \mathcal{B}$  given the random sequence fragment is ever more than  $\delta + \eta$  (plus a tiny tiny amount) then we've got a density increase on a 12-set in a subspace, and can iterate.*

Here is a more precise formulation of what I mean, taking into account comment 863.1. Let me establish my terminology. A *sequence fragment* is a set  $Z \subset [n]$  and a function  $\sigma$  from  $Z$  to  $\{1, 2, 3\}$ . An *extension* of a sequence fragment  $(Z, \sigma)$  is a sequence  $x \in [3]^n$  such that the restriction of  $x$  to  $Z$  is  $\sigma$ . If  $\mathcal{A}$  is any subset of  $[3]^n$ , I shall write  $\mathcal{A}(Z, \sigma)$  for the set of all extensions of  $(Z, \sigma)$  that belong to  $\mathcal{A}$ , which can be naturally identified with a subset of  $[3]^{[n] \setminus Z}$ .

The precise statement of Step 2 is now this. If we can ever find a sequence fragment  $(Z, \sigma)$  such that  $|\mathcal{B}(Z, \sigma)|/|\mathcal{A}(Z, \sigma)|$  is a bit bigger than  $\delta + \eta$ , then we can restrict the coordinates in  $Z$  to  $\sigma$  and iterate. Or at least, we can do that as long as  $\mathcal{A}(Z, \sigma)$  isn't too small.

One thing that makes me slightly anxious is the fact that the densities of the sets  $\mathcal{A}(Z, \sigma)$  may vary quite widely. But it occurs to me here that we may be able to do some kind of regularization by passing to a sequence fragment that maximizes this density and observing that the relative density of  $\mathcal{B}(Z, \sigma)$  must also be preserved or we would have found a density increment somewhere along the line. I'm pretty convinced this can be done if it's needed, so I think for now I won't worry about this problem too much (which makes me much more confident about the double counting arguments later on).

I have to go so will continue this later.

20.  [gowers](#) Says:

[March 4, 2009 at 1:51 pm](#) | [Reply](#)

868. Density increment.

4. But there is a positive probability that  $U_0 \cup FU(Z_1, \dots, Z_m) \subset \mathcal{U}$  and  $V_0 \cup FU(Z_1, \dots, Z_m) \subset \mathcal{V}$ . Moreover, the density of  $\mathcal{B}$  inside the subspace  $\langle U_0, V_0; Z_1, \dots, Z_m \rangle$  (I hope it's easy to guess what that means) is at most  $\delta + \eta/2$  or we're done. So, turning things round, there is in fact a positive probability that the conditional probability in 3 is somewhat less than  $\delta + \eta$ .

This is a bit easier to say in the new language. If  $Z_1, \dots, Z_m$  are disjoint sets, let's write  $\sum_i \eta_i Z_i$  for the sequence fragment that takes the value  $\eta_i$  on  $Z_i$  (so  $Z = Z_1 \cup \dots \cup Z_m$ ). And if  $x \in [3]^{[n] \setminus Z}$  then let's write  $x + \sum_i \eta_i Z_i$  for the extension of this fragment that takes value  $x_i$  at each  $i \in [n] \setminus Z$ .

For any given  $Z_1, \dots, Z_m$  let  $\mathcal{A}(Z_1, \dots, Z_m)$  be the intersection of all the sets  $\mathcal{A}(\sum_i \eta_i Z_i)$ . That is,  $\mathcal{A}(Z_1, \dots, Z_m)$  consists of all  $x \in [3]^{[n] \setminus Z}$  such that  $x + \sum_i \eta_i Z_i \in \mathcal{A}$  for every  $\eta_1, \dots, \eta_m \in [3]$ . It is easy to check (and I've done so — I promise) that  $\mathcal{A}(Z_1, \dots, Z_m)$  is a 12-set. In fact, I think I'll need to give a quick proof of that later on.

At this point we use the multidimensional Sperner type result to say that if we choose  $(Z_1, \dots, Z_m)$  at random then the expected density of  $\mathcal{A}(Z_1, \dots, Z_m)$  is positive (meaning bounded below by a positive constant that depends only on  $\delta, \eta, m$  and the sizes of the sets  $Z_i$ ). This is saying that if  $\mathcal{A}$  is a dense 12-set, then a positive density of all possible  $m$ -dimensional combinatorial spaces (with very small wildcard sets) is contained in  $\mathcal{A}$ . This bit is not thoroughly checked, but I don't think there's much doubt that it can be with a bit of effort.

Now let's suppose that  $x \in \mathcal{A}(Z_1, \dots, Z_m)$ . Let  $U_0$  be the 1-set of  $x$  and let  $V_0$  be the 2-set of  $x$ . Then we must have that the union of  $U_0$  and any subcollection of the  $Z_i$  belongs to  $\mathcal{U}$ , and similarly for  $V_0$  and  $\mathcal{V}$ . And the converse is true too, since  $\mathcal{A}$  is a 12-set. Let  $\mathcal{U}(Z_1, \dots, Z_m)$  be the set of all  $U_0$  such that  $U_0 \cup FU(Z_1, \dots, Z_m) \subset \mathcal{U}$  (to use my notation above), and similarly for  $\mathcal{V}(Z_1, \dots, Z_m)$ . Then  $\mathcal{A}(Z_1, \dots, Z_m)$  is (naturally identified with) the set of all  $x \in [3]^{[n] \setminus Z}$  with 1-set in  $\mathcal{U}(Z_1, \dots, Z_m)$  and 2-set in  $\mathcal{V}(Z_1, \dots, Z_m)$ . So it really is a 12-set in  $[3]^{[n] \setminus Z}$ .

Now let's fix a choice of  $Z_1, \dots, Z_m$  such that  $\mathcal{A}(Z_1, \dots, Z_m)$  has positive density. And let's write  $\mathcal{A}'$  for  $\mathcal{A}(Z_1, \dots, Z_m)$ , and similarly for  $\mathcal{U}'$  and  $\mathcal{V}'$ .

For every  $x \in \mathcal{A}'$ , the subspace of all points  $x + \sum \eta_i Z_i$  contains at most  $(\delta + \eta/2)3^m$  points in  $\mathcal{B}$  or we have our density increase on a subspace. Therefore, if we choose a random  $(\eta_1, \dots, \eta_m)$ , the proportion of  $x \in \mathcal{A}'$  such that  $x + \sum \eta_i Z_i \in \mathcal{B}$  is on average at most  $\delta + \eta/2$ . So choose  $(\eta_1, \dots, \eta_m)$  such that this proportion is at most  $\delta + \eta/2$ .

Now let's suppose that  $\mathcal{B}$  has near maximal density  $\delta + \eta$  in  $\mathcal{A}(\sum \eta_i Z_i)$ , which I will denote by  $(\mathcal{U}, \mathcal{V})(\sum \eta_i Z_i)$ . This set naturally partitions into four parts, according to whether the 1-set/2-set of  $x$  belongs/does not belong to  $\mathcal{U}'/\mathcal{V}'$ . The density of  $\mathcal{B}$  in the  $\mathcal{U}'$ — $\mathcal{V}'$  part is too small, so elsewhere it must be too big. That gives us a density increase for  $\mathcal{B}$  on a 12-set and we can iterate.

All that can go wrong is if  $\mathcal{B}$  does not have near maximal density in  $\mathcal{A}(\sum \eta_i Z_i)$ . But that would have to happen with positive probability, which means that somewhere we would get a density increase.



Metacomment.

At the moment, as is clear, I haven't quite got to the point of plunging in and writing things up formally, but the formality is increasing and I'm still not getting any sense that it's on the point of collapse. This raises questions about how we should proceed if we get to the writing-up stage. The wiki has been great for that, but I think the inconvenience of it would start to bite quite hard if one were actually writing things up in complete detail. (Unfortunately, Luca hasn't yet written a LaTeX2wiki converter ...)

What I'd ideally like is to start writing things in LaTeX but in such a way that others can edit it. I'm not sure if that is technically feasible, however. Another possibility would be to write a skeleton version on the wiki, with statements of the main lemmas and things like that, and then work on the proofs in LaTeX. Or I could just go ahead and do what Ryan did and post a link to a pdf, which anybody could comment on, and which I'd happily send to anyone if they wanted to add to it or make changes (though we would then have to know at all times who was in charge of the latest version). Or we could do something like that but split it into a number of subfiles, one for each section.



o *Jon Says:*

[March 5, 2009 at 3:00 pm](#)

870. Metacomment

Perhaps using the wiki for LaTeX may not be that cumbersome, the wiki has the great advantage of being a common repository already, plus it has the tools to compare changes between versions.

The idea would be to use the wiki simply as a storing device for the latest common LaTeX version, not a place to edit or view the paper. Each time one user A would like to read and/or change something in the LaTeX (say one section at a time as you suggest), A would simply copy-paste the latest wiki version of that section into a blank LaTeX template file on A's computer and continue editing and monitoring the pdf locally. When done A would simply: (1) copy-paste that new LaTeX section onto the current wiki version; (2) check whether some changes to that section by some other user B have been made during A's local editing.

If not, then no problem: A's version is now the common current one and appears as such on the wiki. On the other hand if A sees that a new version by B had appeared in between, then A would quickly edit the wiki of that section and add as first line "merging in progress" (a signal preventing disciplined other users to make further changes). Then A and B would need to discuss their respective version in the wiki discussion page or a blog thread, reach a satisfactory common one, add it to the wiki, and finally remove the "merging in progress" tag.

Since only a few people are working on the project this situation should be fairly rare and localized, so that independent parts of the common overall LaTeX file would progress quickly. Archived versions of the whole pdf file, say one per day, might be storable on a separate blog thread for example.




21. *Gil Kalai Says:*

[March 4, 2009 at 4:12 pm](#) | [Reply](#)

869. Here is a variant on Sperner that I wonder what is the situation for it, and if our discussion is relevant.


Supposet that for every  $k$  dimensional discrete cube you have a specific "forbidden" pair of maximal distance elements. E.g. for  $k=5$  you can exclude the pair  $\{(10100), (010111)\}$ . How large can be a subset of  $\{0,1\}^n$  so that for every  $k$  whenever you fix  $n-k$  coordinates you do not have the forbidden pair for the remaining  $k$  coordinates. (The case that for every  $k$  you forbid  $\{(000\dots), (11\dots, 1)\}$  is Sperner.

we can try even to forbid pairs which depend on the identity of the  $n-k$  coordinates or on the content of the  $n-k$  coordinates (but not both as a random large density subset shows).

22.  [gowers](#) Says:  
[March 5, 2009 at 8:28 pm](#) | [Reply](#)

870. Progress report.

No time to do anything more today. This is to say that I've written a [skeleton proof on the wiki](#) and will try to flesh it out in the near future.

23.  [Ryan O'Donnell](#) Says:  
[March 6, 2009 at 9:39 pm](#) | [Reply](#)

871. Passing between measures.


I noted that Tim's skeleton proof on the wiki at one point sketches why, if you have density  $\delta$  in the equal-slices measure, you can pass to a largish combinatorial subspace on which you have density  $\geq \delta - o(1)$  in the uniform measure.

I wrote up this proof carefully [on the wiki](#). I tried to make it crisp; hopefully that didn't introduce major mistakes.


Nothing unexpected happened; the result is that you can pass to a subspace of dimension at least  $\epsilon n$  while losing only  $O(\epsilon\sqrt{n}) + \exp(-\Omega(\epsilon n))$  additively in the density.

If someone wants to amuse themselves, they can try to evaluate this quantity that arises:

$(1/k^2) \int_{p_1+\dots+p_k=1} \sqrt{\sum_{i=1}^k 1/p_i - k^2}$ . It's not hard to check that this is  $O(\sqrt{k})$  (I think), but is it in fact  $O(1)$ ?


o  [Ryan O'Donnell](#) Says:  
[March 6, 2009 at 9:42 pm](#)

871.1. In the amusement, it should be  $1/k$  out front, not  $1/k^2$ .

o  [Gil Kalai](#) Says:  
[March 6, 2009 at 10:50 pm](#)

871.2 2 questions: How general is the passing between measure phenomenon goes. When there is a group action say for the cup set problem, whenever you prove a bound for any measure the same bound apply to the uniform measure by averaging. What is the situation here? It looks ok for moving from equal slice to uniform (when you pass the a larger subspace) but how general measures you can start with.

I am probably trying to get away with not reading some detailed proofs and missed some crucial postings, but let me still ask: on the very conceptual and general level, what is the main ingredient that allows fourier proofs to work after all the initial examples of various sets with irregular number of lines and no large fourier coefficients?

24.  [gowers](#) Says:  
[March 6, 2009 at 11:07 pm](#) | [Reply](#)

872. Fourier etc.

Gil, I'm not quite certain to which Fourier proofs you are referring. But if you mean how can a density-increment strategy have a chance of working if it is not the case that the wrong number of lines implies a large Fourier coefficient, then the answer is (i) that one can keep localizing to subspaces and (ii) that one can directly prove correlation with a 12-set (on a suitable subspace) rather than via some kind of expansion. If you are referring to Fourier approaches to

Sperner, the answer (depending on which approach you are talking about — both are written up on the wiki) is either that some non-Fourier ingredients are included, or that equal-slices measure is used so that some of the troublesome examples no longer work.

Basically though, if I had to choose a one-word answer to your question, it would be “localization”.



o [Gil](#) Says:

[March 18, 2009 at 10:23 pm](#)

872.1 (just trying to catch up) I suspect this “localization” means that at the end the emerging argument has a strong (probably central) ingredient of Szemerédi-like-regularity-lemma; so we have a distinction between Roth-like-density-increment and Szemerédi-like-density-increment.



25. [gowers](#) Says:

[March 6, 2009 at 11:35 pm](#) | [Reply](#)

873. Probable collapse.

I was beginning to think that the argument [sketched out on the wiki](#) had the ring of truth about it. But as so often happens the effort of writing it up has thrown up a problem that seems serious. I’ve only just realized it, so I’m not sure that it can’t be got round by some sort of trick, or more elaborate argument, but at the moment it feels to me as though a new idea is still needed.

Here, in brief, is the difficulty, which I noticed when trying to write up Step 5 in full detail. The basic idea, which I got from Terry’s 837.2, was to have *two* iterations going on, one in which one tries to get a density increase on a subspace, and the other in which one tries to get a density increase on a 12-set. The problem, which escaped me before, is that the density increase on the 12-set depends on the density of that 12-set. (It comes from the Varnavides density of multidimensional subspaces that you can find in the 12-set.) So there is no guarantee that this inner iteration will terminate. And unfortunately, at least as the argument goes at the moment, the density of the 12-set may drop quite a bit when one does the inner iteration.

I’ll probably think a bit more about whether something can be done about this, but I think progress is more likely to come either from returning to the Ajtai-Szemerédi template or from a new approach to getting from increased density on a 12-set to increased density on a subspace.

Incidentally, one might wonder whether Terry’s argument has the same problem, but it doesn’t, and the reason it doesn’t is instructive. He makes use of the principle that if you have a function  $f$  defined on a set  $X$  and can find a small subset  $Y$  where  $f$  is less dense than average, then you can remove that subset from  $X$ . If  $Y$  is small, then you will get only a small density increase on the complement of  $Y$ , but you will have removed only a small set from  $X$  so things are OK. Now contrast that with a situation where you have a function  $f$  defined on a set  $X_1 \times X_2$  and it is less dense on a subset  $Y_1 \times Y_2$ . By averaging, we get a density increase on one of  $X_1 \times (X_2 \setminus Y_2)$ ,  $(X_1 \setminus Y_1) \times Y_2$  or  $(X_1 \setminus Y_1) \times (X_2 \setminus Y_2)$ . However, if  $Y_1$  and  $Y_2$  both have density  $\epsilon$  and the density increase is in the last of these sets, then we get a density increase that’s proportional to  $\epsilon^2$  but have to drop to a subset of size around  $(1 - 2\epsilon)$  times the size of the original set. And this is too expensive if you do it over and over again. Basically the difference is that  $(1 - 1/n)^n$  converges to a positive limit while  $(1 - 1/n)^{n^2}$  converges to zero.




o [gowers](#) Says:

[March 7, 2009 at 12:08 am](#)


873.1 Actually, the situation is slightly worse: I now see that the problem that I discussed in the last paragraph above is precisely the problem that Terry was talking about in 866 and Randall in 864.2.

A question that occurred to me just now was whether one might be able to deal with low-complexity members of the sigma-algebra of 12-sets. But I now see that that still creates problems, since then we don't get a multidimensional Sperner-type theorem that just depends on the density of the set (or rather, finding such a result doesn't seem any easier than DHJ(3) itself).


26.  [Terence Tao](#) Says:  
[March 7, 2009 at 1:32 am](#) | [Reply](#)

874.

Unfortunately I won't have much time to devote to this project for the next few days, but one possible way around the problem may be to factor the density increment into two pieces. Suppose that  $f$  has a density increment on  $X_1 \times X_2$ , then roughly speaking this means that  $f1_{X_1}$  has a density increment on  $X_2$ . One could try to hold  $X_1$  fixed and start passing to subspaces to increment the density of  $f1_{X_1}$  until it reaches some saturation point, and then  $f$  would have some sort of "relative density increment" on  $X_1$ , and then one could pass to subspaces again to clean up  $X_1$ . This is terribly vague and there are a large number of issues, including the fact that 1-sets and 2-sets are not independent (so the analogy with Cartesian products is slightly misleading), though perhaps some preliminary regularity lemma type arguments might deal with that problem.

- o  [gowers](#) Says:  
[March 7, 2009 at 11:18 am](#)

874.1 Terry, I don't think I understand your suggestion even vaguely. For instance, if in the  $\mathbb{Z}_n^2$  world we choose random sets  $V, X_1, X_2 \subset \mathbb{Z}_n$  and take all points  $(x, y) \in X_1 \times X_2$  such that  $x + y \in V$ , then we have a density increase on  $X_1 \times X_2$ , but it doesn't seem to be possible to do much by looking at each variable separately.

27.  [gowers](#) Says:  
[March 7, 2009 at 11:05 am](#) | [Reply](#)

875.

Now that this has happened, I want to suggest another question, which I had been slightly suppressing. It was always a bit of a worry that we were trying to get from a density increase on a 12-set to a density increase on a subspace when we did not know an analogous argument in  $[n]^2$ . I had persuaded myself that this might be OK because we were passing to combinatorial subspaces instead of long APs, but in retrospect that was not a very convincing argument. So my question is this. Suppose we allow ourselves to use Szemerédi's theorem as a black box and we have a subset  $A \subset [n]^2$  of density  $\delta$  that correlates with a dense Cartesian product  $X \times Y$ . Can we get a density increase on a grid (defined to be a product  $P \times Q$ , where  $P$  and  $Q$  are APs with the same common difference and length tending to infinity)?

If we use arguments that would also prove a functional version, then it's fairly easy to prove that what we are really trying to do here is prove that the characteristic function of  $X \times Y$  can be approximated by a positive linear combination of characteristic functions of grids. Roughly, the argument goes like this. If you can't approximate it (in  $L_1$ ) then the Hahn-Banach theorem gives you a bounded function  $f$  that has average at most zero on all grids but average 1 on  $X \times Y$ . Taking  $g$  to be  $\delta + \eta f$  for some small  $\eta$  we get a function with a density increase on  $X \times Y$  but no density increases on grids. Conversely, if the approximation is possible, then an easy averaging argument gets you from a density increase on  $X \times Y$  to a density increase in one of the grids.

In one dimension, the analogous question is easy to answer. If you have a dense subset  $X \subset [n]$  then you can apply Szemerédi's theorem to find an arithmetic progression  $P \subset X$ . If you remove  $P$  from  $X$  you still have a dense set, so can repeat. Keep going until you're left with just a small proportion of  $P$  and you have then approximated  $P$  by a union of arithmetic progressions.

The problem in two dimensions is that while it is not too hard to find a single grid inside  $X \times Y$  (by averaging find  $s$  such that  $X \cap (Y + s)$  is dense and apply Szemerédi's theorem), when you remove this grid, you no longer have a Cartesian product. And perhaps it is also worth mentioning that the trivial argument does not work: if you approximately partition both  $X$  and  $Y$  into APs and take their Cartesian products, you won't get grids, because the common differences in the two directions will not necessarily be the same.

It is for this reason that I think our best hope is to go back to Ajtai-Szemerédi, because they avoided this problem.

28.  [gowers](#) Says:

[March 7, 2009 at 5:24 pm](#) | [Reply](#)

876. Ajtai-Szemerédi

My natural instinct is to go off and have a hard think about how to adapt the Ajtai-Szemerédi proof, but I want to try to do things the polymath way, which may well mean a few comments that are vague, or go nowhere, or where I can't quite explain properly what I'm trying to do. However, here goes.

What we appear to be able to do is choose  $W$  such that there are many pairs  $(U, V)$  with  $(U, V, W) \in \mathcal{A}$ , and with the additional property that for almost all of the  $U$  involved in those pairs we have almost the expected density of  $(V', W')$  such that  $(U, V', W') \in \mathcal{A}$ . Having fixed such a  $W$ , we can also find a combinatorial subspace in  $[2]^n$  consisting entirely of  $U$  with both properties.

What we would like to do at that point is say that there are many combinatorial subspaces in  $[3]^n$  that project down to this combinatorial subspace in  $[2]^n$ .

Temporarily forgetting that last paragraph, let's rephrase the Ajtai-Szemerédi argument as follows. By the usual averaging argument, we may assume that almost all grids have almost the right density. But if we choose a random grid, then with positive probability we find that it has a positive proportion of horizontal lines that are forbidden, because all the vertical lines point to points in a dense diagonal, and several horizontal lines point to points in the same dense diagonal. In other words, with positive probability, we get a density increment on a 1-set rather than a 12-set.

I have to go for a little while, but I'm going to think along these lines for a bit.

29.  [Randall](#) Says:

[March 7, 2009 at 6:28 pm](#) | [Reply](#)

877. More speculation

I had a look on the wiki at the Ajtai-Szem proof page (I am somewhat embarrassed to admit I had never seen this proof before), and I have a few initial thoughts that may or may not lead in a positive direction.

First, I don't see how multidimensional DHJ (2) can be the analog for Szemerédi's theorem, given the disparity in depth between the results. In particular, it seems to me that whatever the analog for Szem is going to be, it has got to be a multiple recurrence theorem. I can't think of what that would be in the DHJ setup, though. Now, as it happens I have long been pushing for a reduction of the problem to something else anyway, so thinking along these lines, here is a general proof strategy.

Step 1. Reduce DHJ (3) to IP Roth using the "density increment on a subspace for a set correlated with a 12-set" strategy I tried to outline in a speculative fashion in 864. (This step is indeed speculative; I have no idea whether or how it can work.)

Step 2. Reduce IP Roth to "IP\* Szemerédi" via the Ajtai-Szemerédi argument. (I have not thought about this at all. It might be really easy, though, if the analogs are the right ones, given how easy the Ajtai-Szem argument is.)

Step 3. Come up with a combinatorial proof of IP\* Szemerédi. (No real idea how this might go....)

IP\* Szemerédi: Let  $\delta > 0, k \in \mathbb{N}$ . There is an  $n$  such that if  $A \subset [n]^{[n]}$  with  $|A| > \delta n^n$  then  $A$  contains a configuration of the form  $\{f + i \cdot 1_\alpha : 0 \leq i < k\}$  for some  $\alpha \subset [n]$  and some  $f : [n] \rightarrow [n]$ .



o [gowers](#) Says:

[March 7, 2009 at 6:46 pm](#)

877.1 Randall, that first point of yours is one I've mentioned a couple of times, but I haven't really properly justified my view that it shouldn't be a problem. So I'm going to think about it now. I may well end up deciding that it is a problem after all.



o [Randall](#) Says:

[March 8, 2009 at 2:44 am](#)

877.2

Well there is certainly no reason not to use subspaces in that way; in fact it's part of the Furstenberg-Katznelson argument (cf. first three lines of p. 7 of my notes). Something different seems to be going on in the Ajtai-Szemerédi argument, though, where Szemerédi's theorem seems to be doing virtually all of the work.



30. [gowers](#) Says:

[March 7, 2009 at 6:42 pm](#) | [Reply](#)

878. Ajtai-Szemerédi

Hmm, what I said above is actually *not* the Ajtai-Szemerédi argument because they don't choose a random grid like that. Instead, they choose a single set of vertical lines and a random grid that runs across in the other direction.

So let's think more like that. I'll start by ignoring the requirement that the vertical lines have to have roughly the right density. Moving over to the  $[3]^n$  world, I find my "dense diagonal" — that is, a  $W$  such that many  $(U, V, W)$  belong to  $\mathcal{A}$ . Now I want to look for a combinatorial subspace with a good property of some kind. What should that property be?

Let  $\mathcal{U}$  be the set of all  $U$  such that  $(U, [n] \setminus (U \cup W), W) \in \mathcal{A}$  and let  $\mathcal{V}$  be the set of all  $V$  such that  $([n] \setminus (V \cup W), V, W) \in \mathcal{A}$ . Then a good property would be if all the points in the combinatorial subspace had their 1-sets in  $\mathcal{U}$ . Then we would know that no point with its 2-set in  $\mathcal{V}$  could belong to  $\mathcal{A}$ , or something along those lines, which would give us correlation with a 2-set. That would give us a density increase as long as we also knew that the density of  $\mathcal{A}$  in the combinatorial subspace was almost maximal.

I'm finding this boringly hard to do on screen. I'll allow myself a little bit of offline time to try to clarify what I'm saying.



31. [gowers](#) Says:

[March 7, 2009 at 9:44 pm](#) | [Reply](#)

879. Correlation with 1-sets

Nothing conclusive to report, so instead I want to revisit the question of showing that correlation with a 1-set implies a density increase on a subspace. In particular, I want to get a feel for whether the double iteration is necessary.

It occurs to me that a Hahn-Banach argument ought to prove that the question is equivalent to showing that the



characteristic function of a 1-set can be approximated in  $L_1$  by a positive linear combination of characteristic functions of subspaces. At some point I'll check that, and maybe even put it on the wiki, but for now I'll assume it. So how should we write an arbitrary dense 1-set as a positive combination of subspaces?

I think I see a way. No time to be fully detailed, but I'm aiming for something that's equivalent to what we did above. The first step would be to choose small wildcard sets  $(Z_1, Z_2, \dots, Z_m)$  such that a positive proportion (depending on  $m$ ) of the combinatorial subspaces with those wildcard sets are subsets of the given 1-set.

All these combinatorial subspaces are disjoint, so we can safely remove them. Let  $Z = Z_1 \cup \dots \cup Z_m$  and partition the rest of  $[3]^n$  according to how points restrict to  $Z$ .

For every sequence  $z \in [3]^Z$ , let  $\mathcal{U}(z)$  be the set of all sequences  $y \in [3]^{[n] \setminus Z}$  such that  $(y, z)$  belongs to the 1-set  $\mathcal{U}$ . Also, let  $\mathcal{U}'$  be the set of all  $y \in [3]^{[n] \setminus Z}$  such that  $(y, z) \in \mathcal{U}$  for every  $z$  that is constant on all the  $Z_i$ . The precise partition we shall take is this. For each  $z$  that is *not* constant on all the  $Z_i$  we take the set  $\mathcal{U}(z)$ . And for each  $z$  that *is* constant on all  $Z_i$  we take the set  $\mathcal{U}(z) \setminus \mathcal{U}'(z)$ .

This has given us a 1-set inside each combinatorial subspace obtained by fixing the coordinates in  $Z$ . And the average density of those 1-sets is down by a small factor (depending on  $m$ ) from the density of the original 1-set. So we can iterate the procedure.

I think that's a nice clean way of presenting the Randall/Terry result about 1-sets. Of course, it's still using something similar to the double iteration.

32.  *jozsef* Says:

[March 7, 2009 at 11:55 pm](#) | [Reply](#)

880. Shelah's v.s. DHJ

It seems to be a serious difficulty to follow the double iteration in a density incremental argument. It might be helpful to check what would be a Shelah-like density proof look like for  $k=3$ . The first step is actually the same; prove that for any 2-colouring of  $\mathbb{Z}^n$  there are (many) "flip-flop" subspaces. A  $d$ -dimensional subspace of  $\mathbb{Z}^n$  can be represented by  $d$  classes of wildcards, the elements from same class always have the same characters. Two elements in the subspace are neighbours if they differ only in one wildcard class where one is 2 and the other is 1. The subspace is flip-flop if any pair of neighbours have the same colour. The second step would be to show that there is a monochromatic line in  $\mathbb{Z}^d$ . In our case it would mean to show that there is a flip-flop  $d$ -subspace with at least  $c_d 3^d$  elements where we allow  $c_d$  to go to 0 slowly. There are two advantages; first that we allow  $c_d$  to go to 0, second that the number of flip-flop subspaces is independent of the original density as it follows from the two colouring of  $\mathbb{Z}^n$ .

This looks quite promising to me, but let me check first what did I write here...

The first part of Shelah's proof shows that

o  *gowers* Says:

[March 8, 2009 at 12:54 am](#)

880.1 Jozsef, can you explain where the 2-colourings come in?

o  *jozsef* Says:

[March 8, 2009 at 1:19 am](#)

The points are red or blue depending if they are in our dense set or not. I will write more details soon.

33.  *jozsef* Says:

[March 8, 2009 at 1:39 am](#) | [Reply](#)

880.2 For DHJ, we say that a subspace is flip-flop if there are no neighbours that one is in our set and the other isn't. (One might think that we can't gain anything from this if most of the pairs are not from the set, however we will never make any statistics on the number of neighbour pairs inside or outside of our set in a subspace.) To prove that there are many flip-flop subspaces we can follow the original colouring proof; Colour every element of our dense subset by red and the points in the complement by blue. I will try to find a link to the proof or I will write it up myself. Then, the Varnavides type argument gives many flip-flop subspaces. The number of d-dimensional flip-flop subspaces is independent of the density of our pointset, but it certainly depends on the dimension d. The second observation is that every flip-flop subspace is sparse or there is a line. I think I should write up this part.

 [József Says:](#)

[March 8, 2009 at 1:58 am](#)

880.2 For the proof of the existence of flip-flop subspaces I have find two books on Google, "Ramsey Theory" by Ronald L. Graham, Bruce L. Rothschild, Joel H. Spencer, and Jukna's "Extremal Combinatorics". There is a nice paper A. Nilli, "Shelah's proof of the Hales–Jewett theorem", Mathematics of Ramsey theory (Algorithms Combin.), 5, Springer (1990) pp. 150–151, but I was unable to find it online. I think that the original name was "flip top" for a colouring of a subspace where top neighbours received the same colour, but the top isn't special, the bottom pair would work as well, so I've changed it to flip-flop as it's more appropriate (and funny)

 [József Says:](#)

[March 8, 2009 at 4:54 am](#)

All proofs I know for the existence of flip-flop subspaces are recursive. (see the references above) For a d-dimensional flip-flop subspace one needs the recursion  $r_{i+1} = r_i^2 3^{r_i}$  with  $r_1 = 3$ . n should be at least  $r_1 + r_2 + \dots + r_d$  to guarantee a d-dimensional flip-flop subspace in  $\mathfrak{Z}[n]$ . This was also the type of proof I knew for the Sperner subspace theorem, but checking Tim's write up in the Wiki, I realized that his proof is somewhat different.

 [József Says:](#)

[March 8, 2009 at 5:46 am](#)

Well, Tim's proof in the Wiki isn't significantly different from the "traditional" proof but it's elegantly written. Note that the recursive proofs give very uneven subspaces; the sizes of wildecards are increasing recursively as well. It isn't a problem when one considers HJ where every point has its colour, however this property makes it difficult the use of such subspaces for density problems.

 [József Says:](#)

[March 8, 2009 at 7:41 am](#)

After the second reading of Tim's proof I see now that one can choose p, q, and  $\delta$  that one can get "balanced" subspaces.

34.  [gowers Says:](#)

[March 8, 2009 at 1:51 am](#) | [Reply](#)

881. Correlation with 12-sets


I've got to go to bed, but an idea has occurred to me. Maybe I'll see by the morning that it's nonsense. But a 12-set is

just an intersection of a 1-set with a 2-set. So maybe one can use 879 to partition (almost all of) the 1-set into subspaces, and then use 879 again to partition the intersection of the 2-set with each of those subspaces into further subspaces, thereby ending up with a partition of the 12-set into subspaces.

For Cartesian products in grids it would work like this. Given  $X \times Y$ , you first partition  $X \times [n]$  into grids, which is easy. And then inside each of those grids you partition the intersection of that grid with  $[n] \times Y$  into further grids.

If that second argument is correct then (i) I don't know why I didn't spot it before and (ii) it suggests that the first one has a good chance of being correct. And if the first one is correct, it seems to do DHJ(3).

Off to bed while this still feels good ...

- 
[Terence Tao](#) Says:  
[March 8, 2009 at 3:51 am](#)


881.2

This looks like it would work in the  $[n]^2$  world, and give an answer to your 875 (and would also formalise my 874, for that matter). The one thing to bear in mind is that Szemerédi allows one to take the spacing of the long arithmetic progressions in  $X$  or in  $Y$  to be of size  $O(1)$  rather than  $O(n)$ , by working locally. (Meanwhile, the length of the progressions is something like  $\log \log \log \log \log \log \log n$ .) That way, you don't lose too much when taking the GCD of two different spacings.

Of course, in the Hales-Jewett world, we don't have GCD, but the trick of rendering a few coordinates "bad" and working with local 1-sets, etc. rather than global ones may help. (We may eventually have to also break out the Ramsey theory to make the local statistics match the global statistics; this is related to Furstenberg-Katznelson's "strong stationarity" which, after talking to Tim Austin a bit, I suspect we may have to exploit to finish off this problem.)

- 
[gowers](#) Says:  
[March 8, 2009 at 8:27 am](#)

881.3 Terry, I think I don't need to worry about GCDs of spacings. In the corners world I just partition a 1-set into grids, and then the restriction of the 2-set to each grid is still a 2-set, so I partition its restriction to each grid into further grids. What's more, this can be seen as a sort of dualized version of what Ajtai and Szemerédi do themselves. I've woken up still feeling very good about this, and plan to get wikifying straight away.

35.  [gowers](#) Says:  
[March 8, 2009 at 11:10 am](#) | [Reply](#)

882. Progress report

I am in the middle of wikifying the latest DHJ(3) attempt. This time I would actually be prepared to put money on the argument working (unlike last time, when there were too many slightly complicated bits that I felt I didn't fully understand). So far, I've written up a [new proof of the corners theorem](#) to serve as a template for the new DHJ(3) argument. The new proof of the corners theorem is not totally new: it is more like a reorganization of the ideas that go into the Ajtai-Szemerédi argument. Nevertheless, it simplifies things in a way that is crucial for the DHJ(3) proof.

- 
[Randall](#) Says:  
[March 8, 2009 at 4:38 pm](#)

882.1

I have looked at your new proof of corners and it really does make less mysterious what Szemerédi is doing. When I initially read the original proof of Ajtai/Szemerédi yesterday, it struck me that Szemerédi's theorem was being used not once but twice... what was confusing was that it was used once on the diagonal, then again on one of the coordinates. The use on the diagonal gave an impression that DHJ (2) would go proxy for it in the DHJ (3) case. This struck me (everyone else too, I gather) as odd, given the disparity in depth, etc.... Something had to explain the fact that one wasn't doing something about "compactness relative to the diagonal" or some such, and Szemerédi was the only culprit on offer. Your proof uses Szemerédi's theorem twice also, once on each of the coordinates; indeed, it now appears that what Szemerédi's theorem is actually doing in that proof is going proxy for a relative compactness over the diagonal notion. And, in the DHJ (3) case, what you have in mind to fill in here is the idea of partitioning a dense 1 set into dense subspaces, if I understand correctly. (So it's that which corresponds to the use of Szemerédi after all, not DHJ (2)). And this isn't all that surprising anymore, given that the proof of that seemed to involve (at least at the very superficial level I understand it at) a look at the decomposition over the diagonal.

Aesthetically, all of this seems dead on, so I will not take your bet (and will indeed be quite depressed if something else is amiss).

36.  *josef* Says:

[March 8, 2009 at 8:45 pm](#) | [Reply](#)

883. Strong Sperner

It is very possible that by now Tim is just polishing the write up of a combinatorial DHJ, but still let me go back to the unevenness of multidimensional Sperner or flip-flop subspaces which one can get by recursive arguments. It would be better to have a control on the arithmetic structure of such subspaces. In his Wiki article Tim describes a strong version of multidimensional Sperner. Unfortunately the argument there uses DHJ what we don't want to use. On the other hand we might get a similar result by using multidimensional Szemerédi. Given a dense subset of  $2^{[n]}$  denoted by  $A$ . Take a random permutation of  $[n]$ . An element of  $A$  is "d-nice" if it consists of  $d$  intervals, each has length  $n/2d \pm \sqrt{n}$ , and each interval begins at position  $id$  for some  $0 \leq i \leq n/d$ . (Suppose that  $d$  divides  $n$ ) Any interval like this can be represented as a point in a  $d$ -dimensional  $[\sqrt{n}]^d$  cube. If it's dense then multidimensional Szemerédi gives us a strong Sperner.

o  *gowers* Says:

[March 8, 2009 at 9:06 pm](#)

Metacomment: I wouldn't say I'd got to the polishing stage exactly, but if you want to see what's going on it's [here](#). At the moment, I simply don't see anywhere where it can go wrong, but I've had that feeling about wrong proofs in the past, so I won't feel entirely happy until I've got a bit further.

However, the main point of this comment is to say that I strongly support your idea of looking at flip-flop subspaces. What I would really like is for the polymath collaboration to produce a polyproof. If the Ajtai-Szemerédi approach works, then that's just the start: I'd like a triangle-removal approach, a Shkredov approach, and an ergodic approach, and if a Shelah-influenced approach is potentially feasible too then I'm very interested.

37.  *Ryan O'Donnell* Says:

[March 8, 2009 at 9:05 pm](#) | [Reply](#)

884. Wikification.

Here is a short writeup of the [multidimensional Sperner stuff](#) roughly following Terry.860, as used in Tim.879. I will wikify it soon. In fact, this was more or less already on the wiki, in Tim's [latest additions](#). Here are the parameters:

Let  $A \subseteq [2]^n$  have density  $\delta$  and let  $Y_1, \dots, Y_d$  be arbitrary disjoint subsets of  $[n]$  of cardinality  $r \geq (1/\delta)^{O(2^d)}$ .

Choose a random nondegenerate  $d$ -dimensional subspace as follows. For each  $i$ , choose a random nondegenerate combinatorial line in  $[2]^{Y_i}$ , uniformly from the  $3^r - 1$  possibilities. (Actually, you can choose the line from virtually any reasonable distribution.) Form the final subspace by taking the Cartesian product of these lines, and then filling in the coordinates outside the  $Y_i$ 's uniformly at random.

Then this entire subspace is contained in  $A$  with probability at least

$\exp(-\exp(O(1) \ln(1/\delta) 2^d))$ .

The  $O(1)$  can more or less be 2.



◦ *jozsef* Says:

[March 8, 2009 at 9:23 pm](#)

884.1. It is a nice write up! There is a gap between the upper and lower bounds in Gunderson-Rodl-Sidorenko. As I remember, the density is between  $n^{-1/2^d}$  and  $n^{-d/2^d}$ . Ryan, do you think that you can close the gap?



◦ *Ryan O'Donnell* Says:

[March 8, 2009 at 9:38 pm](#)

884.2. Done; it's [here](#). I'd change Tim's writeup to point to it, but he seems to be editing it currently 😊



◦ *gowers* Says:

[March 8, 2009 at 10:04 pm](#)

884.3 I've added a link. (I've kept what I wrote too because I quite like having different styles of explanation on the wiki, even of the same result.)



◦ *Ryan O'Donnell* Says:

[March 8, 2009 at 10:06 pm](#)

884.4. Hi Jozsef — not sure about closing the gap... Actually, as written it doesn't quite even match [GDR]: it requires density approximately  $n^{-1/2^{d+1}}$  rather than their  $n^{-1/2^d}$ .



38. *gowers* Says:

[March 9, 2009 at 12:54 am](#) | [Reply](#)

885 DHJ(3)

I'm now pretty confident that the modified-Ajtai-Szemerédi-based approach to DHJ(3) is in the bag. I have a complete informal write-up [on this page of the wiki](#), though some of the ingredients (such as getting a density increment on a 12-set in a subspace) are on other pages. A certain amount of work will be needed to get it into an acceptable form for a journal article, but not, I hope, as much as all that.

If everyone else shares my belief that it works, then I'd be more interested in pressing on and doing DHJ(k), or at least thinking about its feasibility, than in rushing to write it up with all the numbers put in. Also, it seems to me that the statement about line-free sets correlating locally with 12-sets should be provable using localization rather than equal-slices measure, and that would bring it into line with the rest of the proof. So that's something else I think we should try to do before writing anything up properly.

In my next comment I'm going to speculate a little about DHJ(4).



o Ryan O'Donnell Says:

[March 9, 2009 at 1:02 am](#)

885.2. Hi Tim. I agree it's looking pretty solid. Thanks for all the wikification! I plan to check it over tonight.



o [gowers](#) Says:

[March 9, 2009 at 1:36 am](#)

885.3 That's great — I've got to go to bed pretty soon, but I'll look forward to checking in the morning to see whether it still looks solid to you by then. Thanks for your wikification too!



39. [gowers](#) Says:

[March 9, 2009 at 1:32 am](#) | [Reply](#)

886 DHJ(4)

How might the argument generalize to DHJ(4)? Probably a good way to think about this is to try to deduce the 3D corners result from the full 2D Szemerédi theorem. This wouldn't be much use as a proof of the corners theorem because nobody knows a proof of the full 2D Szemerédi theorem that does not also give 3D corners. (Probably one could falsify that last sentence in silly ways, but I think it's basically true.) However, there is reason to hope that (i) the multidimensional DHJ(3) theorem can be obtained from the one-dimensional theorem by some kind of trickery and (ii) it can be used as an ingredient for proving DHJ(4) in the way that multidimensional Sperner was used for proving DHJ(3). Eventually, of course, I want to come back to (i) and (ii) but for now I'll stick to the easier world of 3D corners.

For 2D corners, the first step is to find a dense diagonal. The nice thing about a dense diagonal is that it gives rise to lots of forbidden points: indeed, if  $(x,y)$  and  $(x',y)$  both belong to the same diagonal, with  $x < x'$ , then  $(x,y)$  is not allowed in the set.

Even better, the set of all forbidden points has a nice Cartesian-product structure. (In fact, it's the points of a Cartesian product that lie below the diagonal, but that will contain a large Cartesian product.)

The analogue of a diagonal for 3D corners is a plane of constant  $x + y + z$ . How can a dense diagonal forbid other points? Answer: if you have a suitably aligned equilateral triangle  $(x + d, y, z), (x, y + d, z), (x, y, z + d)$  then you forbid the point  $(x, y, z)$ . Now the 2D corners theorem tells us that a dense diagonal plane contains many such equilateral triangles, so we end up forbidding a good lot of points. What is rather less clear is what kind of structure that set of points has. In fact, it's so unclear that I think I'd better stop this comment because I do not immediately have anything useful to say about it.

Actually, perhaps I do. What would be very nice would be to get a density increase on a dense (12,23,13)-set. By that I mean a dense set  $B$  of the form  $\{(x, y, z) : (x, y) \in R, (y, z) \in S, (x, z) \in T\}$ . That would be nice because it is the natural analogue of a 12-set (natural, that is, for anyone who has thought about hypergraph regularity and that kind of thing).

If the world is a friendly place, it will turn out that the set of points that form the bottom vertex of a 3D corner with the other three vertices in the dense diagonal plane is a dense (12,23,13)-set. Is it? Yes of course it is! It consists of all points  $(x,y,z)$  such that three conditions hold. The first is that if you go in the  $z$  direction until you hit the diagonal plane, you hit it at a point in  $A$ . But that condition depends on  $(x,y)$  only. The other two conditions depend on  $(y,z)$  and  $(x,z)$ .



OK, this is looking good. So now we've got our dense  $(12,23,13)$ -set that's disjoint from  $A$ . By averaging we find one with which  $A$  correlates. So now all we have to do is partition a dense  $(12)$ -set (where this does not mean the same as a  $12$ -set, but rather it means a set that depends just on  $(x,y)$ ) into large 3D grids. And that we can easily do using 2D-Szemerédi! If the  $(12)$ -set is all  $(x,y,z)$  such that  $(x,y) \in B$ , then by 2D Szemerédi we can partition almost all of  $B$  into large grids with fairly small width. For each such grid  $G$  we can then partition  $G \times [n]$  into large 3D grids, and we're done. The rest of the argument is almost exactly as before.

Obviously, this technique is going to work to show that Szemerédi in  $d$  dimensions implies corners in  $d+1$  dimensions.

So it looks very promising for DHJ( $k$ ). The first target, it seems to me, is to get a multidimensional version of DHJ(3). Somehow the whole thing feels as though it is not going to be too hard ...



o [gowers](#) Says:

[March 9, 2009 at 2:04 am](#)

886.1 No time to write it now, but I see how to deduce multidimensional DHJ(3) from DHJ(3). And indeed it is not hard.



o [Randall](#) Says:

[March 9, 2009 at 6:17 am](#)

886.2 General  $k$

I had a look at the FK proof for general  $k$  and found some very interesting parallels between their proof and what Tim has been doing in the past 48 hours. (And what he proposes above.) Quite striking, really, right down to the trick of cutting things up in one dimension first, then the other. (I must have forgotten this trick, as I didn't consider using it for  $k=3$ .) At any rate, the general outline suggested for  $k=4$  (and beyond) looks terribly sound.



40. [Ryan O'Donnell](#) Says:

[March 9, 2009 at 5:42 am](#) | [Reply](#)

887. Wikification.

I started going through the proof Tim has sketched from the beginning, trying to fill in a few small details.

I thought briefly about removing the use of equal-slices density in the [first part of the argument](#), wherein it is shown that line-free sets correlate with  $12$ -sets. It wasn't immediately clear to me how to do this. Therefore I decided to leave it alone, and work out the "technicality" of passing from relative density under equal-slices to relative density under uniform, discussed in the last paragraph of [the proof sketch](#).

Specifically, this requires the details in the "more details" section of the [currently abortive original density-increment plan](#).

Therefore, I worked for a bit to clean these up. As usual, no surprises; everything is fine. Indeed, one can do it passing to subspaces of  $\Omega(n)$  dimension. I added the last 1% to Tim's sketch and put it in the [passing between measures](#) wiki article. The only minor innovation is noting that you can write equal-slices *exactly* as a mixture of uniform-distributions-on-subspaces.



41. [Ryan O'Donnell](#) Says:

[March 9, 2009 at 7:21 am](#) | [Reply](#)

888. More wikification: equal-slices.

I added the following observation (which I assume was clear to most everyone already) to the wiki entry on [equal-slices measure](#), which helped me understand the “Hang on” part of the proof that [line-free sets correlate with 12-sets](#).

Another equivalent way to draw from the equal-slices distribution is as follows. Start with a string of  $n$  “dots” •. Next, place a “bar” | randomly in one of the  $n + 1$  “slots” between (and to the left and right of) the dots. Next, place a second bar randomly in one of the  $n + 2$  slots formed by the string of  $n$  dots and one bar. (At this point we have determined the “slice”.) Next, fill in all the dots to the left of the leftmost bar with 1’s; fill in all the dots between the two bars with 3’s (not 2’s!); and, fill in all the dots to the right of the rightmost bar with 2’s. Delete the bars. Finally, randomly permute the resulting string of 1’s, 2’s, and 3’s.


With this viewpoint, it may be easier to understand the joint distribution of the 1-set and the 2-set of a string drawn from equal-slices. Specifically, it is one that is useful for proving density-Sperner’s theorem.

**Fact:** Let  $z$  be a string drawn from the equal-slices distribution, in the manner described above. Let  $x \in [2]^n$  be the string that would have been formed had we filled in all the dots to the left of the *first* bar with 1’s and all the dots to its right with 2’s. Similarly, let  $y \in [2]^n$  be the string that would have been formed had we filled in all the dots to the left of the *second* bar with 1’s and all the dots to its right with 2’s. Then the following should be easy to verify:

- (i)  $x$  and  $y$  are both distributed according to the equal-slices distribution on  $[2]^n$  (but not independently);
- (ii)  $x, y, z$  form a combinatorial line in  $[3]^n$ ; in particular,  $x$  and  $y$  are “comparable” in  $[2]^n$ , i.e., either  $x \leq y$  or  $x \geq y$ ;
- (iii)  $\Pr[\text{line is degenerate}] = \Pr[x = y] = 2/(n + 2)$ .

From these facts we can derive the density version of Sperner’s Theorem:

**Theorem:** Suppose  $A \subseteq [2]^n$  has equal-slices density  $\delta$ . Then according to the above distribution on  $(x, y) \in [2]^n \times [2]^n$ , we get a nondegenerate combinatorial line in  $A$  with probability at least  $\delta^2 - \frac{2}{n+2}$ .

42.  Ryan O'Donnell Says:  
[March 9, 2009 at 8:30 am](#) | [Reply](#)

889. Last wikification of the night.

Okay, using the above mentality I was able to rewrite in my own words Tim’s proof that line-free sets correlate with 12-sets. I added these words [to the wiki](#), modulo the passing from uniform density to equal-slices density (which is still in that article and also partially [here](#)). It’s pretty late at night for me, so I hope I got it right.

43.  gowers Says:  
[March 9, 2009 at 10:12 am](#) | [Reply](#)

890. Multidimensional DHJ(3)

My main teaching days are Mondays and Tuesdays this term, and today and tomorrow are the last two such days of term. So I’ll be fairly busy, but I hope I’ll still have a bit of time for blogging and wikification. Here I want, as a pre-wikification exercise, to sketch a proof that DHJ(k) implies multidimensional DHJ(k). I’ve woken up with the feeling that DHJ(k) is going to go through almost as easily as DHJ(3). If that is the case, it will be unexpected for two reasons. First, it will give a proof of Szemerédi’s theorem that has a strong claim to be the simplest known. (The only rival I can think of is a particularly clean approach via infinitary hypergraphs, due to Elek and Szegedy, but I may be wrong.) Secondly, it would be the first proof of Szemerédi’s theorem for which “the general case is the case  $k=3$ ”. By that I mean that in all other proofs you have to go at least as far as  $k=4$  before it’s obvious how to generalize, and in some you have to go to  $k=5$ . (Perhaps a true understanding of the problem would require a proof that generalizes straightforwardly from the  $k=2$  case ...)

Back to multidimensional DHJ(k). Here's what I think works. Let  $\mathcal{A}$  be a density- $\delta$  subset of  $[k]^n$  and let  $M$  be large enough so that every subset of  $[k]^M$  of density at least  $\theta$  contains a combinatorial line. Now split  $[k]^n$  up into  $[k]^M \times [k]^{n-M}$ . For a proportion at least  $\delta/2$  of the points  $y$  in  $[k]^{n-M}$  the set of  $x \in [k]^M$  such that  $(x, y) \in \mathcal{A}$  has density at least  $\delta/2$ . Therefore, by DHJ(k) (with  $\theta = \delta/2$ ) we have a combinatorial line. Since there are fewer than  $(k+1)^M$  to choose from, by the pigeonhole principle we can find a combinatorial line  $L \in [k]^M$  and a set  $\mathcal{A}_1$  of density  $\delta/2(k+1)^M$  in  $[k]^{n-M}$  such that  $(x, y) \in \mathcal{A}$  whenever  $x \in L$  and  $y \in \mathcal{A}_1$ . And now by induction we can find an  $(m-1)$ -dimensional subspace in  $\mathcal{A}_1$  and we're done.

This gives a truly horrible bound, and should mean that if DHJ(k) goes through as I expect (and Randall also expects, I'm glad to see from 886.2), the bound that comes out at the end will probably be of Ackermann type, so it will be comparable to the bounds that come out of the hypergraph approach. (A small challenge that I know some people out there would enjoy is to try to see how this approach to Szemerédi fits in with the general philosophy that all the different proofs are at some deep level manifestations of closely related ideas. There are distinct echoes of hypergraphs in this proof, and yet it is far easier than hypergraph regularity and counting — what is going on? Possibly that we are “cheating” by continually passing to subspaces, but why can't we do that with hypergraphs? Or can we? Perhaps there's a way of passing to subgraphs without throwing away too many degenerate simplices. Hmm ... I quite like that but no time to pursue it just at the moment.)



o [gowers](#) Says:

[March 9, 2009 at 10:28 am](#)

Multidimensional DHJ(3) is [now wikified](#).



o [jozsef](#) Says:

[March 9, 2009 at 5:09 pm](#)

I did write up the strong multidimensional Sperner proof which uses multidim. Szemerédi on the wiki, however it might be totally uninteresting if DHJ is easier than hypergraph removal.



44. [Ryan O'Donnell](#) Says:

[March 9, 2009 at 4:17 pm](#) | [Reply](#)

891. Corrected/expanded a bit the “passing between measures” article on the wiki.



45. [Ryan O'Donnell](#) Says:

[March 9, 2009 at 5:00 pm](#) | [Reply](#)

892. Wikification.

I finished the 1% fleshing out required in the article proving that [line-free sets correlate with 12-sets](#), including all the passing back and forth between uniform and equal-slices measures. I think the only bit remaining undone here is instantiating all the parameters.



46. [jozsef](#) Says:

[March 9, 2009 at 5:27 pm](#) | [Reply](#)

wiki


Meta comment: Something happened with the wiki. It seems that it has been hacked. Be careful with the links there.

- o  [josef](#) Says:  
[March 9, 2009 at 5:42 pm](#)


Just seems to be spammers changing the page. I reversed it.

- o  [josef](#) Says:  
[March 9, 2009 at 5:46 pm](#)

It wasn't enough, the page is wrong again. I'm not sure what to do. I changed it back again, but I don't think it will stay like this for long.

- o  [Terence Tao](#) Says:  
[March 9, 2009 at 5:58 pm](#)


I blocked the offending IP. If that fails, the next step would be to protect the main page by limiting edits to signed in users, I guess.

47.  [Terence Tao](#) Says:  
[March 9, 2009 at 7:06 pm](#) | [Reply](#)


893.

I've been busy, so I haven't been able to stop by much recently, but things look pretty good at this point. I agree that the DHJ(3) Ajtai-Szemerédi sketch looks pretty solid. (An amusing side note: when I had Ajtai-Szemerédi described to me, I thought that they were already doing what we were doing now, i.e. getting correlation with an unstructured Cartesian product and then partitioning that product into grids. So I was a little confused when Tim was insisting that what we were doing was not quite Ajtai-Szemerédi... but now I see the subtle difference between the two approaches.)

It looks like Tim Austin has come up with an alternate proof that is also based on correlation with 12-sets, etc. but is based on triangle-removal type strategies rather than density-increment strategies. It also requires a preliminary use of Graham-Rothschild to regularise a large number of statistics so as to make them stable under freezing of coordinates, and so is likely to give poorer bounds. But it is closer in spirit to the original intent of Polymath1. I believe Tim will come on here himself to report on this soon (he's working in an ergodic theory setting), and I will focus on trying to finitise it. (It's likely to be cleaner than the finitisation of Furstenberg-Katznelson, because one does not have to deal with relative almost periodicity.)

- o  [gowers](#) Says:  
[March 9, 2009 at 10:39 pm](#)

893.1 Re your amusing aside, I have had a very similar experience, which can sort of be deduced from my initial blog comment on the Ajtai-Szemerédi proof. (It can be found at the end of [the article on the wiki](#).) At that stage I only half remembered their proof, and likewise assumed that they must have *started* with the dense diagonal, got a global Cartesian product disjoint from  $A$ , and deduced a density increment on a grid.

48.  [Tim Austin](#) Says:  
[March 9, 2009 at 9:24 pm](#) | [Reply](#)

894. I seem to have come to this at a handy moment, following Terry's post

Having been following progress here (albeit only in fits and starts) and talking with Terry about the ideas that have

come out, it struck me late last week that in the infinitary world of stochastic processes that Furstenberg and Katznelson move to, the approach raised here for obtaining obstructions to uniformity that are built from ij-sets can actually be coupled to a lot of machinery that's already known from other things to give a new infinitary proof of their multiple recurrence result, without anything else being required. In particular, it uses an infinitary analog of 'energy increment' to improve the structure of a stochastic process, and then an appeal to an 'infinitary hypergraph removal lemma' originally motivated by some work of Terry on infinite random hypergraphs, both which I recently used to play a similar game around the multidimensional Szemerédi Theorem (arXiv 0808.2267, in case it's of interest).

In fact, it turned out that this could be written up completely in just a couple of days by judiciously cutting, pasting and re-notating writeups of other things, so this is now done and on the arXiv: once it becomes publicly visible it'll be at 0903.1633. I feel I should possibly offer my assurances that I wouldn't have rushed from a moment of realization to completing a preprint if it really hadn't been so very quick and mechanical from that point on, without requiring any input of new ideas from me.

For what it's worth, I've thought only briefly about finitizing this approach and Terry has already said most of what I could say. As it stands it will require a preliminary heavy appeal to Graham-Rothschild (Carlson-Simpson, in the infinitary world) and then proceeding in close analogy with hypergraph removal strategies. So it is rather removed from the density-increment approach that I think is now mainly being pursued here, and would look set to give much worse bounds unless some other new idea can remove the reliance on Graham-Rothschild.



o [gowers](#) Says:

[March 9, 2009 at 10:29 pm](#)

894.1 Tim, this is great news and very much in the spirit of polymath leading to multiple proofs and all-round improved understanding



49. [Randall](#) Says:

[March 9, 2009 at 9:31 pm](#) | [Reply](#)

895. Book proof of Szemerédi? Or "This is a theorem Harry Furstenberg stole from Szemerédi...we're stealin' it back."

Regarding what Tim (Gowers) said about an easy proof of Szemerédi materializing, as well as what Terry said about avoiding relative almost periodicity (which seems to be exactly what makes this proof easy as well), it seems natural whether one of us should think about writing up carefully a "book proof of Szemerédi's theorem". For starters, it seems to me that this might entail proving József's comment no. 2 from a multi-dim Sperner type of hypothesis, then pushing an induction to k-dimensional corners ala Tim's 886. I am hoping it could be easier in the details than DHJk; that should be clear by the end of the first step, though.



o [Terence Tao](#) Says:

[March 9, 2009 at 10:34 pm](#)

It may in fact be that, paradoxically, the book proof of Szemerédi may ultimately pass through DHJ. For instance, observe that the original Ajtai-Szemerédi proof of corners had to pass through Szemerédi, whereas by lifting up from  $[n]$  to cubes, we can substitute (multi-dimensional) DHJ(2) in place of Szemerédi.

The hypergraph regularity/removal proof of (multidimensional) Szemerédi is not too bad, actually, despite its reputation. The ergodic version of it, which Tim Austin wrote up in the arXiv a few months back, is perhaps a touch simpler than the Furstenberg-Katznelson proof based on repeated extensions by relatively almost periodic functions, being instead based on extending the entire system up to a more "pleasant" system enjoying a number of useful relative independence properties.



o [gowers](#) Says:

[March 9, 2009 at 10:46 pm](#)

I've been thinking about this too, and it may not be as paradoxical as all that — just a bit unexpected at first. For instance, it is generally accepted that the van der Waerden theorem is “really” the Hales-Jewett theorem (at least if you prove it combinatorially), and that starting with a subset of  $[n]$  distracts from what is actually going on. And something like that seems to have been the case here too: for a while the fact that Ajtai and Szemerédi used Szemerédi's theorem was a distraction, in that it made it seem as though their approach reached a dead end at the corners theorem when in fact the structure they should have been using was a cube. So Jozsef's comment 2 was spot on.

Randall, there may be room for disagreement over your slogan, but I can't help liking it ...



50. [gowers](#) Says:

[March 9, 2009 at 10:50 pm](#) | [Reply](#)

Metacomment: I'm doing my best not to write newly numbered comments here, since this thread is about to run out (at 899) and then we'll reach an important milestone — comment number 1000. I feel that we deserve a decent-length post before that one. (My main activity at the moment is over at the wiki. I'm currently working on generalizing the local correlation with 12-sets from  $[3]^n$  to  $[k]^n$ .)



51. [Terence Tao](#) Says:

[March 9, 2009 at 11:09 pm](#) | [Reply](#)

896. Austin's proof

I will probably focus the (limited) time I have available for this thread on trying to explicate Austin's proof in finitary language. I know you guys don't actually have access to it yet, but let me try to informally describe some of the details. Take all this *cum grano solis*; I have not yet fully digested the argument and some of the details may be slightly or perhaps even massively incorrect.

Let me take  $[3] = \{0,1,2\}$  (rather than  $\{1,2,3\}$ ) for sake of arbitrarily fixing the conventions (we've been a bit inconsistent on this). Let's define a trilinear form  $\Lambda(f, g, h)$  on functions  $f, g, h : [3]^n \rightarrow \mathbb{R}$  by the formula

$$\Lambda(f, g, h) = \mathbb{E}_{\ell} f(\ell(0))g(\ell(1))h(\ell(2))$$

where  $\ell : [3] \rightarrow [3]^n$  varies along combinatorial lines with respect to some measure that I will intentionally leave vague. DHJ(3) is equivalent to the following “triangle-removalish” type statement:

**DHJ(3)'**: Let  $f, g, h : [3]^n \rightarrow [0, 1]$  be such that  $\Lambda(f, g, h) = o(1)$ . Then  $\mathbb{E}_{x \in [3]^n} f(x)g(x)h(x) = o(1)$ .

Roughly speaking, Austin's strategy is to “regularise” the situation so that 12-sets, 01-sets are “relatively independent” over a common algebra of 1-sets, and similarly for the 12-sets and 02-sets, etc., with the 0-sets, 1-sets, 2-sets themselves being relatively independent over  $\emptyset$ -sets (which, for us, I think means “unions of large subspaces”, and which can be ignored by passing to a large subspace). I don't understand this part well yet, but it is analogous to the Szemerédi regularity lemma. There is also a preliminary reduction to “strong stationarity” which means, roughly, that the statistics of various 01-sets, etc. (e.g. the density of an overlap between a relevant 01-set and a relevant 02-set) doesn't change if we freeze a bounded number of coordinates. This reduction is obtained via Graham-Rothschild and is going to be hideously expensive as regards quantitative bounds, but never mind that for now.

1. Once one has this regularisation, this makes DHJ(3)' easy when  $f$  is an (indicator of a) (basic) 12-set,  $g$  is a basic 02-set, and  $h$  is a basic 01-set, much as it is easy to find corners connecting three sets  $A, B, C$  in  $[n]^2$  when those sets



are Cartesian products in the right fashion. Indeed things seem to collapse to DHJ(2.5) in this case. (This is analogous to the triangle removal lemma for three bipartite graphs when each of the three graphs is a complete bipartite graphs between one cell in each vertex set.)

2. Next, this implies DHJ(3)' when  $f, g, h$  are non-basic 12-sets, 02-sets, and 01-sets, i.e. finite unions of basic sets (or more precisely finite linear combinations of indicators of basic sets), but where the lower bound depends on the complexity of the partition into basic sets. (This is analogous to the triangle removal lemma for unions of complete bipartite graphs between cells.)

3. Next, this implies DHJ(3)' when  $f$  (and similarly  $g, h$ ) are "Borel 12-sets" (to continue the topological analogy much as Borel sets can be approximated by open sets), which means that given any  $\varepsilon > 0$ , one can approximate  $f$  to within  $\varepsilon$  by a non-basic 12-set of bounded complexity. This is because pigeonhole ensures that there are a lot of non-basic 12-sets which are 99% occupied by  $f$ , and one will be able to get this case from applying Case 2 to these "rich" sets, and using some relative independence properties. (This is analogous to triangle removal for unions of 99%-complete bipartite graphs between cells.)

4. Next, we obtain DHJ(3)' when  $f$  (and similarly  $g, h$ ) is the sum of a Borel 12-set and something which is highly orthogonal to all basic 12-sets, including the very small basic 12-sets coming from the cells of an approximation to the Borel 12-set. This is basically because the guy which is highly orthogonal to all basic 12-sets is so uniform as to have essentially no contribution to  $\Lambda(f, g, h)$ . (This is analogous to triangle removal for a triplet of bipartite graphs which have been regularised.)

5. Finally, we have a regularity lemma that tells us that arbitrary  $f, g, h$  decompose in this fashion (possibly after localising to a large subspace). This is a "soft" energy increment argument, analogous to that in the regularity lemma. One has to keep freezing coordinates while performing this increment argument, so it is important that one has the strong stationarity property first before one sets up the regularity argument.

Maybe I'll try to write a more coherent version of the above on the wiki at some point.



o [gowers](#) Says:

[March 9, 2009 at 11:44 pm](#)

Another wordpress bug for your collection: it doesn't like `\varepsilon` (so I changed yours to `\epsilon`).



52. [gowers](#) Says:

[March 9, 2009 at 11:58 pm](#) | [Reply](#)

897. Progress report

I've reached the first point where it isn't almost trivial to generalize the argument we have for DHJ(3) to an argument for DHJ( $k$ ). I think it's going to be a sort of middling difficulty — that is, you can't do it in five minutes but you know you'll get there in at most a few hours. (Actually, I think it may be easier than that, but I haven't yet tried.) Over on the wiki, I have written up [part of an argument](#) that generalizes to  $[k]^n$  the fact that a dense line-free set (in equal-slices measure) correlates with a dense 12-set.

The problem in the  $[k]^n$  case is to prove that the lower-complexity set you get is dense. The proof for  $[3]^n$  uses a quantitative Sperner theorem (namely that you get a positive proportion of all possible combinatorial lines if you start with a dense set — all with respect to equal-slices measure). The proof for  $[k]^n$  needs a similar statement about combinatorial lines in  $[k-1]^n$ . It should be obtainable from DHJ( $k-1$ ) by a suitable averaging argument, but it's not one I can do in my head.



o [gowers](#) Says:

[March 10, 2009 at 1:31 am](#)

897.1 I've got a vague outline of an argument, but have to go to bed. Here's how it goes. We're given a dense (in equal-slices measure) subset  $\mathcal{B} \subset [k-1]^n$ . I claim that it contains a dense set of combinatorial lines (where that means that if you randomly permute  $[n]$  and then randomly divide it up into  $k$  subintervals, filling the first  $k-1$  with 1 to  $k-1$  and treating the final one as a wildcard set, and then with positive probability that line lies in your set). I think that follows by an averaging argument where you choose a large subspace randomly by first picking the sizes of the wildcard sets and then choosing the sets themselves, and then you apply DHJ( $k-1$ ) and average up. Finally, I think that if you pick an equal-slices random point  $x$  in  $[k]^n$  and consider all the points in  $[k-1]^n$  you get by replacing the  $j$ s of  $x$  by  $1$ s,  $2$ s,  $\dots$ ,  $(k-1)$ s, then the distribution on the resulting combinatorial lines is not radically different from the distribution described above. In particular, I think the probability that all those points lie in the dense subset of  $[k-1]^n$  is positive.



o [gowers](#) Says:

[March 10, 2009 at 1:37 am](#)

897.2 Just realized that “not radically different from” might be better changed to “identical to”.



53. [gowers](#) Says:

[March 10, 2009 at 12:25 am](#) | [Reply](#)

Metacomment: I'm going to think offline for a bit about the problem in 897. Before I clock off for the night I'll say whether I think I've sorted it out, in case anyone else wants to think about it. (It feels as though it might be Ryan territory.) And I've now put up a new post in case the comment count reaches 899. *[Am clocking off now -- see 897.1 above.]*

## Leave a Reply

Logged in as [Ryan O'Donnell](#). [Logout »](#)

☐ Notify me of follow-up comments via email.

[Blog at WordPress.com.](#)

[Entries \(RSS\)](#) and [Comments \(RSS\)](#).

⌂

# Gowers's Weblog

Mathematics related discussions

---

« [DHJ\(3\): 851-899](#)  
[Polymath1 and open collaborative mathematics](#) »

## Problem solved (probably)

By gowers

Without anyone being particularly aware of it, a race has been taking place. Which would happen first: the comment count reaching 1000, or the discovery of a new proof of the density Hales-Jewett theorem for lines of length 3? Very satisfyingly, it appears that DHJ(3) has won. If this were a conventional way of producing mathematics, then it would be premature to make such an announcement — one would wait until the proof was completely written up with every single *i* dotted and every *t* crossed — but this is blog maths and we're free to make up conventions as we go along. So I hereby state that I am basically sure that the problem is solved (though not in the way originally envisaged).

Why do I feel so confident that what we have now is right, especially given that another attempt that seemed quite convincing ended up collapsing? Partly because it's got what you want from a correct proof: not just some calculations that magically manage not to go wrong, but higher-level explanations backed up by fairly easy calculations, a new understanding of other situations where closely analogous arguments definitely work, and so on. And it seems that all the participants share the feeling that the argument is “robust” in the right way. And another pretty persuasive piece of evidence is that Tim Austin has used some of the ideas to produce a new and simpler proof of the recurrence result of Furstenberg and Katznelson from which they deduced DHJ. [His preprint is available](#) on the arXiv.

Better still, it looks very much as though the argument here will generalize straightforwardly to give the full density Hales-Jewett theorem. We are actively working on this and I expect it to be done within a week or so. (Work in progress can be found [on the polymath1 wiki](#).) Better even than that, it seems that the resulting proof will be the simplest known proof of Szemerédi's theorem. (There is one other proof, via hypergraphs, that could be another candidate for that, but it's slightly less elementary.)

I have lots of thoughts about the project as a whole, but I want to save those for a different and less mathematical post. This one is intended to be the continuation of the discussion of DHJ(3), and now DHJ(k), into the 1000s. More precisely, it is for comments 1000-1049.


---

### Possibly related posts: (automatically generated)

- [DHJ — the triangle-removal approach](#)

This entry was posted on March 10, 2009 at 12:21 am and is filed under [polymath1](#). You can follow any responses to this entry through the [RSS 2.0](#) feed. You can [leave a response](#), or [trackback](#) from your own site.


### 126 Responses to “Problem solved (probably)”

1.  [Jason Dyer](#) Says:  
[March 10, 2009 at 1:32 am](#) | [Reply](#)

Metacomment.


Appropriate timing of note:

Cantwell.949: A Moser set for  $n=5$  has 124 points

2.  [Terence Tao](#) Says:  
[March 10, 2009 at 3:37 am](#) | [Reply](#)

1000. DHJ( $k$ )

A small comment: from experience with hypergraph regularity, I would imagine that DHJ( $k$ ) will not follow exactly from DHJ( $k-1$ ), but from something more like DHJ( $k-1, k$ ), and so the induction hypothesis may need to have an additional parameter than just  $k$ . (For instance, the simplex removal lemma for 3-uniform hypergraphs requires, at base, something resembling the  $K_4$ -removal lemma for graphs rather than the triangle removal lemma for graphs. While the  $K_4$ -removal and triangle removal lemmas have nearly identical proofs, I don't think one can easily deduce the former from the latter if the latter is a black box.)

-  [gowers](#) Says:  
[March 10, 2009 at 10:37 am](#)

1000.1 An  $\epsilon^2$  comment: you mean of course DHJ( $k-2, k$ ).

I think the jury is still out on whether we need to prove the full DHJ( $k-2, k$ ) before doing DHJ( $k$ ) or whether passing to subspaces the whole time makes it unnecessary. I'm hoping for the latter, but am ready to switch if we have to.

An  $\epsilon^3$  comment: you are now our official [powers-of-ten champion](#) ...

3. *Anonymous* Says:  
[March 10, 2009 at 7:20 am](#) | [Reply](#)

1001. Line and equal-slice distributions in  $[k]^n$ .


It seems to me that the right generalization of the “[another useful equivalent definition](#)” of equal-slices on  $[k]^n$  is as follows:

Pick  $z \in [k]^n$  from the equal-slices distribution. Form  $x^i \in [k-1]^n$  by changing all the  $k$ 's in  $z$  into  $i$ 's, for  $i = 1 \dots k-1$ . Finally, let  $y^1, \dots, y^{k-1}$  be a random permutation of the  $x^i$ 's.

Certainly the set  $\{y^1, \dots, y^{k-1}, z\}$  is a combinatorial line, which is degenerate only if  $z$  had no  $k$ 's.

I *believe* that each  $y^i$  is individually distributed as equal-slices on  $[k-1]^n$ , but I can't quite seem to prove it to myself at this late hour. Anyone want to prove or disprove?

(If it's true, there must be some slick way to see it, like “break up a necklace of  $n$  beads into  $k$  pieces, then pretend you did it in one the  $k!$  different orders, then...” etc.)

-  [Ryan O'Donnell](#) Says:  
[March 10, 2009 at 7:21 am](#)

1001.1. That was me writing; forgot to log in.



o [gowers](#) Says:

[March 10, 2009 at 9:12 am](#)

1001.2 I'm pretty sure it's false. In fact, isn't it obviously false, since the expected number of  $i$ 's in  $y^i$  is twice the expected number of any  $j$  for  $j \neq i$ ?

However, I think what can be proved is that the distribution on each individual  $y^i$  is not importantly different from the equal-slices distribution, in that the collection of dense sets is the same in both cases. This is analogous to what happens in the corners problem, where the number of corners containing a point is different from point to point (even if we start off with a triangular grid) but this doesn't matter because the discrepancy between the uniform measure of a set and the point-in-random-corners measure of a set depends only on the measures and not on  $n$ .



o [gowers](#) Says:

[March 10, 2009 at 10:41 am](#)

1001.3 However, I think I may be misunderstanding what you have written, since I don't quite see why you bother to permute the  $x^i$ 's (given that they were already distributed in a way that is permutation-invariant — or at least that's how it appears to me from what you say).



o [Ryan O'Donnell](#) Says:

[March 10, 2009 at 4:41 pm](#)

1001.4. Oops, yes, I think what I meant to say was “Pick a permutation  $\pi$  on  $[k-1]$  and then form  $y^i = \pi(x^i)$ , where by  $\pi(x^i)$  we mean permute the  $j$ -sets of  $x^i$  according to the permutation  $\pi$ .”

I will write a new comment on this topic later today.



4. [Randall](#) Says:

[March 10, 2009 at 7:49 am](#) | [Reply](#)

1002. 895 again.

Okay, maybe the book proof of Szemerédi goes through DHJ, but I hope not... I was trying before to get someone to bite on trying something potentially simpler than DHJ (3) that would still prove the corners theorem via the Ajtai-Szem method. (And in particular didn't use Szemerédi's theorem as a lemma.) Of course, being the token ergodic theorist, you can probably guess that I would like someone to bite so I don't have to try to check myself, because I am woefully slow and error prone when it comes to such things. So let me try again... I'll just think out loud and if none of this makes any sense, everyone can feel free to ignore.

What if I take as my basic object pairs  $(\alpha, \beta)$  of  $[n]$ . A diagonal is a set  $D_\gamma = \{(\alpha, \beta) : \alpha + \beta = \gamma\}$ , where  $+$  is multi-set union and  $\gamma$  is some fixed multi-set (of multiplicity no greater than 2 in each element). A corner is a triple of the form  $\{(\alpha, \beta), (\alpha + \eta, \beta), (\alpha, \beta + \eta)\}$ , where  $\eta$  can be a set plus an anti-set (an antiset is like a set where every element has a multiplicity of -1).

Okay, let  $A$  be a set of pairs of density at least  $\delta$ , i.e.  $|A| \geq \delta 4^n$ . Now, for some rather small, or at least not-so-big  $m$ , most of the whole space is covered by diagonals whose “multiplicity 2-or-zero” (for the defining multiset  $\gamma$ ) set of indices has size no bigger than  $m$ .

Diagonals of this sort have at least  $2^{n-m}$  members. Take one of these diagonals  $D_\gamma$  on which  $A$  is dense, say of relative density at least  $\delta/2$ . If  $A$  is corner-free, then any pair of the form  $(\alpha, \beta)$ , where  $(\gamma - \beta, \beta) \in A$  and  $(\alpha, \gamma - \alpha) \in A$  is forbidden. That should be something like  $4^{n-m-1}\delta^2$  forbidden pairs, and these pairs lie in a Cartesian product. This should mean that  $A$  has a density increment on some other Cartesian product, call it  $X \times Y$ .

The next step is presumably to use the multidimensional Sperner's theorem to carve almost all of  $B = X \times Y$  up into subgrids of some large but fixed size  $r$ . (Here a subgrid is a collection  $(\zeta \cup FU\{\beta_1, \dots, \beta_r\}) \times (\alpha \cup FU\{\beta_1, \dots, \beta_r\})$ .) I wasn't really anticipating how hard this would be but maybe you can use the iteration method we use to carve 1-spaces up. Now that I am thinking about this finally, maybe one can do it all in one go. Namely, you can always find  $\beta_1, \dots, \beta_r$  with combined size at most  $M$  such that the probability that  $(\zeta \cup FU\{\beta_1, \dots, \beta_r\}) \subset X$  and  $(\alpha \cup FU\{\beta_1, \dots, \beta_r\}) \subset Y$  is above some constant  $c$ . So for those pairs  $(\zeta, \alpha)$  where you get the containment you were shooting for, remove those grids from  $B$  and keep them as part of your partition; partition the other pairs into  $4^T$  classes according to the value taken on the subgrid associated with that pair. Each of these classes is the intersection of  $B$  with a subgrid and we iterate...just like in the proof of DHJ. And, I guess that's it. Assuming all of the above is correct, you get an increment on a subgrid.

It's almost 3 here and too late to check so I will apologize in advance if something is amiss.



o [gowers](#) Says:

[March 10, 2009 at 9:18 am](#)

1002.1 Randall, is your problem the same as Jozsef's in comment 2? In one of the very early comments I tried to translate that one into set-theoretic language and ended up with something pretty similar. This is mainly to say that I agree that it's worth seeing whether Szemerédi can be done without going the whole way to DHJ. For now I'm going to concentrate my efforts on DHJ(k) but will follow what you have to say with interest. I suppose there are two issues: can DHJ be avoided, and is the resulting proof substantially simpler?



o [Randall](#) Says:

[March 10, 2009 at 3:39 pm](#)

1002.2 Yes, it was intended to be the same (not sure whether Jozsef intended to allow inverted corners, but I am here). Oh, yes...I see your comment 7. That looks to be exactly the interpretation I am employing.

Well, perhaps Terry will wake up and say whether the alleged proof I gave is really a proof or can be made into a proof. (And whether it's simpler than what we have already.) Since he composed and typed the breakthrough post (837?) in eighteen minutes, perhaps he'll be able to do this despite being very busy with other things.

On a more general philosophical note, Tim, to we ergodic theorists (who have trouble with subscripts appearing on the numbers capital N appearing in proofs), the dynamical proof of van der Waerden (not that of Furstenberg and Weiss but the easy one, due to Blaszczuk, Plewik and Turek) is much more pleasant than any proof I have seen of HJ. (Actually, I have long wanted to see a proof, even an unpleasant one, of HJ that made efficacious use of non-metrizable topological dynamics.)




o [Randall](#) Says:

[March 10, 2009 at 6:58 pm](#)

1002.3 Obviously I meant "efficacious use of \*metrizable\* topological dynamics". There exist proofs using non-metrizable topological dynamics (e.g. Furstenberg and Katznelson's proof). It may not be possible, or it may require some non-commutative stuff like in the FK proof of DHJ. There is a Bergelson-Leibman proof that tried to use recurrence for continuous maps of compact metric spaces, but it was realized after the paper appeared that neither continuity of the transformations nor completeness of the space was actually ever used.




- 
[Randall](#) Says:  
[March 11, 2009 at 7:34 pm](#)


1002.4 No induction....

I think the idea I had to prove Szemerédi via Ajtai-Szemerédi type argument using József's formulation from comment 2 is dead in the water at  $k=3.5$ . The problem I don't see a way around is that there don't seem to be enough corners at  $k=3$  to forbid enough points at  $k=4$ . Obviously not a problem for DHJ since the number of lines at  $k=3$  is  $4^k$  which is equal to the number of points at  $k=4$ .

Oh well...if the above is correct (nobody checked?) it may be the simplest proof of the corner's theorem, at least.

- 
[Randall](#) Says:  
[March 25, 2009 at 8:57 pm](#)


Hmm...I finally had a look at the above myself, and it seems to be complete garbage for (at least) two reasons.... Perhaps now I am convinced that the DHJ proof is the right one for Szemerédi after all. (Maybe.)

- 
[gowers](#) Says:  
[March 10, 2009 at 10:49 am](#) | [Reply](#)

Metacomment: because of more spam problems on the wiki, you now have to be registered on it and logged in if you want to edit the main page.


- [Problem solved \(probably\) « OU Math Club](#) Says:  
[March 10, 2009 at 4:41 pm](#) | [Reply](#)

[...] experiment in doing math research on his blog. It sounds like it was a resounding success! Timothy Gowers announced on his blog that he is pretty convinced that the crowd on his blog has come up with a new proof of the density [...]


- 
[Terence Tao](#) Says:  
[March 10, 2009 at 7:10 pm](#) | [Reply](#)

Metacomment.

This might be a good time to set up a “meta” thread to deal with issues not directly related to finishing off the proof of DHJ(3)/DHJ(k) [1000-1049] or of computing DHJ-type numbers in small dimension [900-999]. For instance, this seems like a good time to face issues concerning writing up the results, how to perform proper attribution, etc. (At a more frivolous level, there was a suggestion on the wiki to come up with a logo for that wiki; I think that might be a fun topic for the meta thread. More generally, discussion of the strengths and weaknesses of the wiki, or the threaded format, etc. would be apropos at this time, I think.)

- 
[gowers](#) Says:  
[March 10, 2009 at 7:45 pm](#)

OK what I'll do is get on and write a post that I've been intending to write, in which I discuss general meta-issues that have arisen out of how things have gone. I think there's a lot to talk about, so comments on that should be a natural place for discussing these issues amongst others.

8.  [Terence Tao](#) Says:  
[March 10, 2009 at 7:43 pm](#) | [Reply](#)

1003. Austin's proof

I've managed to digest a key ingredient in Austin's proof (Lemma 6.1 of <http://arxiv.org/abs/0903.1633>) which may be of interest here. Roughly speaking, it says that 1-sets and 2-sets (say) are always "locally independent" of each other. A more precise statement is the following. Let  $f : [3]^n \rightarrow [-1, 1]$  be 02-insensitive (e.g. it is the indicator function of a 1-set), and let  $g : [3]^n \rightarrow [-1, 1]$  be 01-insensitive (e.g. it is the indicator of a 2-set). Suppose also that  $\mathbb{E}_{x \in [3]^n} f(x)g(x)$  is large. Then  $f$  has large average on large-dimensional spaces, or (equivalently) correlates with a 012-low influence function (a function which is almost invariant under any interchanges between 0, 1, and 2).

Proof. Let  $x$  be a random string  $[3]^n$ , let  $A \subset [n]$  be a random medium-sized set of indices, and let  $\pi_{A,02 \rightarrow 0}(x) = \pi_{A,2 \rightarrow 0}(x)$  be the string in  $[3]^n$  formed by replacing all occurrences of 0 or 2 in  $A$  with 0. If  $A$  is not too large, then  $\pi_{A,02 \rightarrow 0}(x)$  is still roughly uniformly distributed in  $[3]^n$ , and so

$\mathbb{E}_{x \in [3]^n} f(\pi_{A,02 \rightarrow 0}(x))g(\pi_{A,02 \rightarrow 0}(x))$  is large.

But as  $f$  is 02-insensitive,  $f(\pi_{A,02 \rightarrow 0}(x)) = f(x)$ . As  $g$  is 01-insensitive,  $g(\pi_{A,02 \rightarrow 0}(x)) = g(\pi_{A,02 \rightarrow 1}(x)) = g(\pi_{A,012 \rightarrow 1}(x))$ , where  $\pi_{A,012 \rightarrow 1}(x)$  is the string formed by sending every digit of  $x$  in  $A$  to 1. Thus


$\mathbb{E}_{x \in [3]^n} f(x)g(\pi_{A,012 \rightarrow 1}(x))$  is large.

Freezing  $A$ , we thus conclude that  $f$  has large mean on many  $|A|$ -dimensional subspaces; alternatively, we see that  $f$  correlates with the 012-low influence function  $G(x) := \mathbb{E}_A g(\pi_{A,012 \rightarrow 1}(x))$  (if we choose  $A$  in a suitably Poisson manner).


A corollary of this is that if one has a 1-set  $A$  and a 2-set  $B$  which are "strongly stationary" in the sense that their average local density on medium-sized subspaces is close to their global density, then  $A$  and  $B$  are roughly independent, thus the density of the 12-set  $A \cap B$  is roughly the density of  $A$  multiplied by the density of  $B$ .

As one might imagine, this fact is extremely useful for doing any sort of "counting lemma" involving 12-sets.

To be continued...

- o  [Terence Tao](#) Says:  
[March 10, 2009 at 7:44 pm](#)

A small correction: where I said that  $f$  and  $g$  could be indicators of 1-sets and 2-sets respectively, it might be better to think of them as the *balanced functions* of a 1-set and 2-set, i.e. the indicator minus the density.

- o  [Terence Tao](#) Says:  
[March 10, 2009 at 7:55 pm](#)

1003.2

Incidentally, this framework also gives a short proof of "lack of lines implies correlation with a local 12-set". With the notation as before, observe that  $x, \pi_{A,01 \rightarrow 1}(x), \pi_{A,02 \rightarrow 2}(x)$  will form a combinatorial line with high probability if  $A$  is not too small. Thus if  $\mathcal{A}$  is a set with no lines and density  $\delta$  then

$\mathbb{E}(1_{\mathcal{A}} - \delta)(x)1_{\mathcal{A}}(\pi_{A,01 \rightarrow 1}(x))1_{\mathcal{A}}(\pi_{A,02 \rightarrow 2}(x))$  is large.

But if we freeze  $A$  and the coordinates outside of  $A$  (thus passing to an  $|A|$ -dimensional subspace), then  $1_A(\pi_{A,01 \rightarrow 1}(x))$ ,  $1_A(\pi_{A,02 \rightarrow 2}(x))$  are indicators of a 2-set and a 1-set respectively, and the claim follows.



o [gowers](#) Says:

[March 10, 2009 at 11:50 pm](#)

1003.3 Terry, I'm trying to digest 1003.2. It seems to me that the displayed line is true only if there are many  $x$  such that  $\pi_{A,01 \rightarrow 1}(x)$  and  $\pi_{A,02 \rightarrow 2}(x)$  both belong to  $A$ . In the previous proof this was done with the help of Sperner and that seemed necessary. I'm not sure where the Sperner step is appearing in what you've written.



9. [Ryan O'Donnell](#) Says:

[March 10, 2009 at 9:54 pm](#) | [Reply](#)

1004. Random lines/equal slices for DHJ(k+1).

I think what I wrote in 1001 was correct. Let me restate it here for clarity:

Let  $z \in [k+1]^n$  be drawn from equal-slices. Form the string  $v^j \in [k]^n$  by changing all  $(k+1)$ 's in  $z$  to  $j$ 's, for  $j = 1 \dots k$ . Finally, pick a random permutation  $\pi$  on  $[k]$  and define strings  $x^i = v^{\pi(i)}$ , for  $i = 1 \dots k$ .

Then each string  $x^i$  is distributed as equal-slices on  $[k]^n$ . Further, since  $(v^1, \dots, v^k, z)$  is a combinatorial line ("in order"), we also have that  $\{x^1, \dots, x^k, z\}$  is a combinatorial line (possibly "out of order").

The proof is [here](#); I'll wikify it soon.



o [Ryan O'Donnell](#) Says:

[March 10, 2009 at 10:12 pm](#)

1004.1. Wikified [here](#). Hopefully this will help with the [generalization of line-free sets correlating with 12-sets](#).



o [gowers](#) Says:

[March 10, 2009 at 10:56 pm](#)

1004.2 Ryan, I'm still struggling with the statement of what you are proving. Suppose  $k=2$ . Then I pick  $z \in [3]^n$  from equal slices and form the strings  $u$  and  $v$  by turning the 3s of  $z$  into 1s and 2s, respectively. Then I randomly decide whether to interchange  $u$  to  $v$ . Now let's look at the event that the number of 1s in  $u$  is between  $0.49n$  and  $0.51n$ . If I choose  $u$  according to equal-slices that is  $.02$ . But if I choose it by starting with  $z$ , then either I start with about that number of 1s and change the 3s to 2s, or I start with about that number of 2s and change the 3s to 1s. The probability seems to me to be smaller because  $z$  is being asked to have rather a lot of 1s or rather a lot of 2s. Since you've got a wikified proof, what I'm asking here is for an explanation of where I'm going wrong.



o [Ryan O'Donnell](#) Says:

[March 10, 2009 at 11:53 pm](#)

1004.3. Hmm, I'm not quite sure what I can say to help.

In your example, when you pick  $u$  from equal-slices on  $[2]^n$  "in the normal fashion", then as you say the event  $A$  = "number of 1's in  $u$  is in  $[.49n, .51n]$ " occurs with probability exactly  $.02$ .

When you pick it in the funny way, through,  $z$ , what is the probability? As you say, half the time  $u$  is formed by

changing  $z$ 's 3's to 2's. In this case,  $A$  occurs iff  $z$  had between  $.49n$  and  $.51n$  1's. The other half of the time,  $u$  is formed by changing  $z$ 's 3's to 1's. In this case,  $A$  occurs iff  $z$  had between  $.49n$  and  $.51n$  1's-and-3's-together, iff  $z$  had between  $.49n$  and  $.51n$  2's.


Clearly the probability  $z$  has between  $.49n$  and  $.51n$  1's is the same as the probability  $z$  has between  $.49n$  and  $.51n$  2's. So we are reduced to asking why this probability is indeed  $.02$ .

Well, I'm not sure what to say, except that it is. It comes from the fact that  $\int_{.49}^{.51} 2(1-p)dp = .02$ , where  $2(1-p)$  is the density of the first component in a uniform draw from the 3-simplex. Note that this integral is a bit of a "coincidence", because it is **not** true that, e.g.,  $\int_0^{.49} 2(1-p)dp = .49$ .

The reason this all works is a bit easier to understand in the case of  $k = 2$ ; see [this explanation](#) in the older wiki article on equal-slices.


 [gowers](#) Says:  
[March 11, 2009 at 12:16 am](#)

1004.4 Thanks — I think I'm starting to get it now.

10.  [Terence Tao](#) Says:  
[March 10, 2009 at 9:55 pm](#) | [Reply](#)


1005. Quantitative bounds?

Is our understanding of the density-increment proof of DHJ(3) good enough that we can venture a probable quantitative bound for  $n$  in terms of  $1/\delta$ ? Naively I am expecting tower-exponential behaviour (with the height of the tower being either polynomial or exponential in  $1/\delta$ ).

 [gowers](#) Says:  
[March 10, 2009 at 11:08 pm](#)


1005.1 I'm pretty sure it's a tower. The rough reason is that when we drop to a subspace we obtain that subspace by applying multidimensional Sperner, which I think restricts its dimension to the log of the previous dimension. But I think the density increase we can get that way is good, so I think the bound should be a tower of height polynomial in  $1/\delta$  and therefore not worse than what you get out of the Ajtai-Szemerédi proof of the corners theorem.

Also, if we've got the stomach for it, we could think about trying to Shkredovize the argument and beat Shelah's bound for the colouring DHJ(3). That would be a serious undertaking though, as it would require us to understand the global obstructions rather than just the local ones.

11.  [Ryan O'Donnell](#) Says:  
[March 10, 2009 at 10:27 pm](#) | [Reply](#)

1006. Equal-slices.

I wonder: can we mostly work in the equal-slices measure for the whole proof (of DHJ(3), say)? What are the advantages of the uniform distribution?

 [gowers](#) Says:  
[March 10, 2009 at 11:12 pm](#)

1006.1 My take on this is that equal-slices measure is very good when you want to mix random points and random lines, but the uniform measure is better when you want to restrict to subspaces. Roughly, the reason for the latter is that if you start with equal-slices and restrict a few coordinates, then the distribution on the resulting subspace becomes much more like a uniform one, or a suitably weighted uniform one. For example, if you're told that a point's restriction to the first  $m$  coordinates has roughly equal numbers of 1s, 2s and 3s, then with very high probability it belongs to a slice with roughly equal numbers of 1s, 2s and 3s, so the distribution in that  $(n-m)$ -dimensional subspace is roughly uniform.

It's possible that we could argue that we can restrict to subspaces and get an equal-slices density increase, but I think we would still have to go via uniform, so in the end I don't see a compelling argument for doing that.



o Ryan O'Donnell Says:

[March 11, 2009 at 12:08 am](#)

1006.2. That's a good point. I guess I was idly hoping that the fact that equal-slices is [exactly equal to a \(fairly natural\) mixture of uniform distributions on subspaces](#) might allow us a clever way to stay entirely in the equal-slices world.



12. [gowers](#) Says:

[March 11, 2009 at 12:55 am](#) | [Reply](#)

1007. DHJ(k)

I think I've now proved (though with details a bit hazy at a couple of points) the analogue for  $k$  of the fact that line-free sets in  $[3]^n$  correlate locally with 12-sets. The [proposed argument](#) can probably be tidied up in a few places. I haven't yet thought about whether any of the rest of DHJ(k) is going to present problems, but it feels as though we're closing in on it pretty rapidly. That's my lot for today though.



13. [Terence Tao](#) Says:

[March 11, 2009 at 1:20 am](#) | [Reply](#)

1008. Re: 1003.3

Yes, you're right, one also needs Sperner. Without it, what one gets is that if  $\mathcal{A}$  is uniformly distributed with respect to 12-sets, then

$$\mathbb{E} 1_{\mathcal{A}} 1_{\mathcal{A}}(\pi_{A,01 \rightarrow 1}(x)) 1_{\mathcal{A}}(\pi_{A,02 \rightarrow 2}(x)) \approx \mathbb{E} \delta 1_{\mathcal{A}}(\pi_{A,01 \rightarrow 1}(x)) 1_{\mathcal{A}}(\pi_{A,02 \rightarrow 2}(x))$$

and then one has to use DHJ(2) to say that the RHS is large (which one can do, after letting  $A$  vary in a sufficiently Poisson-like manner).

I would imagine that the same thing works in higher  $k$ , perhaps this is already implicit in 1007.

14. [Top Posts « WordPress.com](#) Says:

[March 11, 2009 at 1:49 am](#) | [Reply](#)

[...] Problem solved (probably) Without anyone being particularly aware of it, a race has been taking place. Which would happen first: the comment count [...] [...]



15. Ryan O'Donnell Says:

[March 11, 2009 at 5:38 am](#) | [Reply](#)

1009. Equal-slices distribution.

I think the following is an easier explanation of what I wrote.

Claim: Pick  $z \in [k]^n$  according to equal-slices. Then pick  $J \in [k-1]$  uniformly and form  $x$  by changing all the  $k$ 's in  $z$  to  $J$ 's. Then  $x$  is distributed according to equal-slices on  $[k-1]^n$ .

Note: it's helpful to think of  $k$ 's as  $*$ 's.

Proof sketch via coupling: Draw a circle of  $n$  dots. Place bar(1) uniformly at random in the  $n+1$  slots. Then place bar(2) uniformly at random in the resulting  $n+2$  slots. Etc., placing  $k$  bars. Pick  $J \in [k-1]$  uniformly. Look at the arc of dots following bar( $k$ ) clockwise, up until the next bar. Fill in these dots with  $k$ 's. Look also at the arc of dots preceding bar( $k$ ) counterclockwise, up until the next bar. Fill in these dots with  $J$ 's. Next, for each of the remaining  $k-2$  arcs of dots, fill it all with the same digit — using up the remaining digits  $[k-1] \setminus \{J\}$  in some arbitrary way. Finally, delete the bars, cut the circle to make it a segment, and then randomly permute the whole sequence of digits. We claim the resulting sequence is distributed as equal-slices on  $[3]^n$ , and therefore we can take it as  $z$ .

Now forming  $x$  from  $z$  is the same as doing the above procedure but using  $J$ 's where we were using  $k$ 's before. But really, happens under this new procedure? You place  $k-1$  bars as if you were doing equal-slices on  $[k-1]^n$ ; then you throw in a phantom extra bar but only use it insofar as deciding to call the arc it lands in the  $J$ -arc; then you define the other  $k-2$  arcs according to the digits in  $[k-1] \setminus \{J\}$ , etc. Now granted, the arc chosen to be the  $J$ -arc is not uniformly distributed among arcs — it's biased towards the longer arcs. But who cares, since  $J$  is uniformly random on  $[k-1]$ ?

16.  Ryan O'Donnell Says:

[March 11, 2009 at 6:01 am](#) | [Reply](#)

1010. “Varnavides”-DHJ( $k$ ). (Sorry for the scare quotes, but it strikes me as funny somehow to name this concept after a person.)

Tim, I can't quite follow what you wrote at the end of your proof sketch that line-free sets in  $[k]^n$  correlate with lower-complexity sets. I was wondering if you could clarify. In particular, as you say it seems we need to show that DHJ( $k$ ) actually implies Varnavides-DHJ( $k$ ) under some appropriate measures.

As you suggest, it seems that equal-slices is what to hope for here (it works pretty well for  $k=2$  at least!). In other words, we may hope that if  $A \subseteq [k]^n$  has equal-slices measure  $\geq \delta$  then if you choose an “equal-slices-combinatorial-line” (i.e., draw from equal-slices on  $([k] \cup \{*\})^n$ ) then there is some  $c(\delta) > 0$  chance that all  $k$  points are in  $A$ .

Looking at that statement now I worry slightly that it is a bit optimistic. Well, let's hope it's true, and let me ask, Tim, about [what you wrote](#).

I guess I don't quite understand what's going on with the parameter  $M$ . On one hand, I assume it's to be “small” (although still perhaps  $\omega_{n \rightarrow \infty}(1)$ ), because otherwise bounding the number of lines in  $[k]^M$  by  $(k+1)^M$  looks worrisome.

On the other hand, this makes me worry a bit about bullet-point 3: if  $\nu$  is supported on  $M$ -dimensional subspaces then somehow I don't expect that a set of lines  $\mathcal{L}$  being large under  $\nu$  should mean it's large under a more “global” measure on lines such as equal-slices.

Quite likely I'm mistaking your argument or missing something — would you be able to clarify? Thanks!

o  [gowers](#) Says:



[March 11, 2009 at 1:16 pm](#)

1010.1 Ryan, the argument is so sloppily written (little more than a declaration that my gut feeling is that it works) that it is good that you apply a little pressure of this kind. I think I have answers to your specific queries, but that's not the same as a proof. But maybe it will tip you over to believing that it works.

First of all, my parameter  $M$  is not supposed to tend to infinity. As I said in the write-up, it's meant to be the smallest integer such that we have DHJ( $k$ ) with density  $\zeta$ , where  $\zeta$  depends only on the equal-slices density  $\theta$  of the set inside which we're trying to find lots of combinatorial lines.

So what about your second worry? The point here is that the  $M$ -dimensional subspaces I'm averaging over are absolutely *not* local. Instead, they're supposed to be far more tailored to equal-slices. What I'm not doing is fixing almost all coordinates and taking  $M$  small wildcard sets, or something like that. Instead, I'm randomly choosing the sizes of the  $M$  wildcard sets (which I allow to add up to 1, so there are in fact no fixed coordinates at all) and then uniformly at random choosing a line from the resulting subspace. That's why I'm hoping that the resulting measure on lines is sufficiently global in character.

17.  *jozsef* Says:

[March 11, 2009 at 8:00 am](#) | [Reply](#)

#### 1011. A Shelah-type argument

I'm quite optimistic that a Shelah type argument might work for the general DHJ( $k$ ) problem. It's late evening and I had a hard day, so I'm not sure how far can I go with the argument tonight. Let me give some basic definitions first. In  $k^{[n]}$  two points are neighbours if they have same coordinates everywhere but in one position where one is  $k-2$  and the other is  $k-1$ . For a subset  $S \subset k^{[n]}$  the space  $k^{[n]}$  is fliptop if a point  $p$  is in  $S$  iff any neighbour of  $p$  is also in  $S$ . (I will wikify it if the argument turns out to be useful) Our goal is to find a large fliptop subspace which is dense. By induction there will be a  $k-1$  combinatorial line which gives a full combinatorial line because of the fliptop property of the subspace. I will follow Shelah's strategy with some extra conditions to make it (hopefully) work for the density version. Suppose that  $S \subset k^{[n]}$  is  $c$ -dense and we already know the subspace version of DHJ for  $k-1$ . Partition  $[n]$  into  $m$  consecutive intervals of lengths  $l_1, l_2, \dots, l_m$ .

We will repeat two steps to get a subspace eventually. The first one should keep the density high and the second will provide the fliptop property.

Step one: For any point in  $k^{[l_m]}$  there is a certain number of points in  $S$  with that tail. Select the points from  $k^{[l_m]}$  which are tails of at least  $(c - \epsilon_1)k^{n-l_m}$  elements of  $S$ . The set of such points is denoted by  $H_1$ . Applying the induction hypothesis we know that  $k^{[l_m]}$  will contain a  $d_1$ -dimensional subspace where every element avoiding (running) coordinate  $k-1$  are from  $H_1$  (so it has many extensions that are in  $S$ ).

Step two: Consider the elements of the  $d_1$ -dimensional subspace where every coordinate is  $k-1$  or  $k-2$ . We say that two such points are equivalent if they are the tails of the same subset of  $S$ . There are no more than  $k^{n-l_m}$  such equivalence classes. Therefore if  $d_1 > k^{n-l_m}$  then we have a Sperner pair, two points where the set of  $k-1$  coordinates of a point contains the other point's set of  $k-1$  coordinates. It defines a combinatorial line in the subspace and we will work with this line in the following steps.

We will repeat steps one and two in  $l_{m-1}$ . I will continue, but I would like to read over first what I wrote.

o  *jozsef* Says:

[March 11, 2009 at 8:56 am](#)

#### 1011. contd.

Let us denote the  $k$  points of the combinatorial line selected previously by  $P_0^1, P_1^1, \dots, P_{k-1}^1$  (indexed by the running coordinates.)

Step one: select set  $H_2$  from  $k^{[l_{m-1}]}$ .

Select a point if with any  $P_i$  it is the tail of at least  $(c - \epsilon_2)k^{n-l_m-l_{m-1}}$  elements of  $S$  where  $0 \leq i \leq k-2$ .

There are two cases; either there are many such points or there is a tail of at least  $(c + \delta)k^{n-l_m-l_{m-1}}$  elements of  $S$ . (More careful calculations are needed, but let me sketch the argument first.) If there are enough points, the there is a  $d_2$ -dimensional subspace in  $k^{[l_{m-1}]}$  where every point without coordinate  $k-1$  is from  $H_2$ .

Step two: Consider the elements of the  $d_2$ -dimensional subspace where every coordinate is  $k-1$  or  $k-2$ . We say that two such points,  $a$  and  $b$ , are equivalent if for any element  $t \in k^{[n-l_m-l_{m-1}]}$ ,  $t, a, P_i$  is in  $S$  iff  $t, b, P_i$  is in  $S$ .

There are no more than  $k^{n-l_m-l_{m-1}} + 1$  such equivalence classes. If  $d_2 > k^{n-l_m-l_{m-1}} + 1$  then there is a Sperner pair, two points where the set of  $k-1$  coordinates contains the other point's set of  $k-1$  coordinates. It defines the second combinatorial line in the subspace and we will work with the two lines in the following steps. Let us denote the  $k$  points of the new combinatorial line by  $P_0^2, P_1^2, \dots, P_{k-1}^2$



o [jozsef](#) Says:

[March 11, 2009 at 6:18 pm](#)

There are a few typos in the argument. I should fix them eventually, but let me mention her one which might be quite misleading. "There are no more than  $k^{n-l_m}$  such equivalence classes." actually, there are no more than  $2^{k^{n-l_m}}$  such equivalence classes. Similarly, in the second iteration There are no more than  $2^{k^{n-l_m-l_{m-1}}+1}$  such equivalence classes.



18. [jozsef](#) Says:

[March 11, 2009 at 9:00 am](#) | [Reply](#)

I'll continue tomorrow morning, but I hope that what I wrote makes sense and maybe someone will tell me if this plan is feasible or not.



o [gowers](#) Says:

[March 11, 2009 at 9:20 am](#)

1012.1 I've just read through it and nothing jumps out at me as wrong, or obviously not going to work. I suppose my one worry is a metaworry, which is that many people must have tried to produce a density version of Shelah's argument, and there doesn't seem to be any sign of an obstacle in the above sketch. On the other hand, a balancing metareassurance is that you have a record of finding dazzlingly simple arguments.

I haven't yet thought about the part where you write "more careful calculations are needed".

Perhaps another reason to be hopeful is that one way of explaining what Shelah did is to say that he changed an inductive argument from deducing  $HJ(k)$  from  $HJ(2)$  with the repeated help of  $HJ(k-1)$  to deducing  $HJ(k)$  from  $HJ(k-1)$  with the repeated help of  $HJ(2)$ , and there is no obvious reason to think that that cannot be done for density as well.

Anyhow, if it works ... amazing!



19. [gowers](#) Says:

[March 11, 2009 at 12:54 pm](#) | [Reply](#)

1013. Shelah.

I took the car to be serviced today, which resulted in a long walk to work, during which I pondered Jozsef's argument. I can't see a problem, but neither have I thought it through fully. So I want to try to assess its feasibility by seeing if I can sketch it. (Of course, Jozsef has already sketched it, but there's no harm in having more than one sketch, and if I

manage to sketch it without looking any further at what he wrote, then it might make us feel better in the way that people feel better about computer-assisted proofs if more than one person does the computer part and different programs are used.)

What is remarkable about Jozsef's proposal if it works is that it is so similar to what Shelah did: it would make it very mysterious that the proof had not been discovered earlier (which would add another layer of mystery to the already remarkable fact that Shelah's proof was itself not discovered earlier).

The idea is this. Suppose that we are trying to prove DHJ( $k$ ) with density  $\delta$ . We begin by choosing some parameters. First, we choose  $m$  so large that DHJ( $k-1$ ) is true with density  $\delta/2$ . That is, we assume as an inductive hypothesis that every subset of  $[k-1]^m$  of density at least  $\delta/2$  contains a combinatorial line.

Next, we choose a very rapidly increasing sequence of integers (though in the context of some proofs of this kind the rate of increase is not too bad — this proof should result in a primitive recursive bound for DHJ( $k$ ), one level beyond a tower, just like Shelah's)  $r_1 \ll r_2 \ll \dots \ll r_m$ . (Jozsef called them  $l_1 \ll \dots \ll l_m$  but on my screen an  $l$  is just a vertical line so I've changed the notation.)

We shall repeatedly apply the following averaging argument: if  $a_1, \dots, a_M$  are integers between 0 and 1, and the average of the  $a_i$  is  $\delta$ , then at least  $\epsilon N$  of the  $a_i$  are at least  $\delta - \epsilon$ . The proof is trivial (by contradiction).

Now let  $n = r_1 + \dots + r_m$  and think of  $[k]^n$  as  $[k]^{r_1} \times \dots \times [k]^{r_m}$ . Let  $\mathcal{A}$  be a subset of  $[k]^n$  of density  $\delta$ . For each  $y \in [k]^{r_m}$ , let  $\mathcal{A}_y = \{x \in [k]^{n-r_m} : (x, y) \in \mathcal{A}\}$ . Then the average density of  $\mathcal{A}_y$  over  $y \in [k]^{r_m}$  is  $\delta$ , so the density of  $y$  such that  $\mathcal{A}_y$  has density at least  $3\delta/4$  is at least  $\delta/4$ .

Let  $\mathcal{B}_m$  be the set of all  $y \in [k]^{r_m}$  such that  $\mathcal{A}_y$  has density at least  $3\delta/4$ . By the pigeonhole principle, we can find a subset  $\mathcal{C}_m$  of  $\mathcal{B}_m$  of density at least  $\gamma = \delta/4k^{n-r_m}$  such that  $\mathcal{A}_y$  is the same for every  $y \in \mathcal{C}_m$ .

Let us define a "binary subspace" to be a set where you fix some of the coordinates and allow the remaining ones to take the values  $k-1$  and  $k$ . If  $t$  is small enough, and if you choose a random binary subspace of  $[k]^{r_m}$  by randomly choosing  $t$  coordinates to be the variable ones and randomly fixing the others, and if you then choose a random point in one of these random binary subspaces, the result will be very close to the uniform distribution on  $[k]^{r_m}$ . Therefore, by an easy averaging argument we can find a binary subspace inside which the density of  $\mathcal{C}_m$  is at least  $\gamma/2$ . (Note that for this to work we need  $r_m$  to be large enough in terms of  $\gamma$  and  $t$ .) And then by Sperner's theorem we can find two points in the binary subspace that form a combinatorial line (in the  $[2]^n$  sense but with  $k-1$  and  $k$  replacing 1 and 2, or 0 and 1 if you prefer).

We can extend this binary combinatorial line to a full combinatorial line in  $[k]^{m_r}$ , by allowing the wildcards to take the values  $1, 2, \dots, k-2$  as well. The result is a ...

OK, this is the first point at which I've got stuck, though I think Jozsef may have dealt with the difficulty I am facing. Time to end this comment and start a new one.

20.  [gowers](#) Says:

[March 11, 2009 at 1:06 pm](#) | [Reply](#)

1014. Shelah

The problem with what I was doing just above is this. I've just defined a line  $L = (y_1, \dots, y_k)$  in  $[k]^{m_r}$ , and I know that the sets  $\mathcal{A}_{y_{k-1}}$  and  $\mathcal{A}_{y_k}$  are the same. However, I don't know anything about  $\mathcal{A}_{y_i}$  for  $i \leq k-2$ , and it would be a serious problem if e.g. they turned out to be empty.

However, I've ignored Jozsef's step 1, which is perhaps his main new idea. He begins by using DHJ( $k-1$ ) to pass to a subspace where  $\mathcal{A}_y$  has density close to  $\delta$  for *all* sequences  $y$ , provided only that they do not have any (variable)

coordinate equal to  $k$ . (This can be done by an averaging argument similar to the one I used above to get dense binary subspaces — this time one wants dense  $(k-1)$ -ary subspaces.)

But now I've reached the point where I don't understand Jozsef's argument. How do I know that it is not the case that  $\mathcal{A}_y$  is empty whenever one of the variable coordinates of  $y$  is equal to  $k$ ? I don't see it at the moment, so I think I'd better wait for Jozsef to wake up.

◦  *jozsef* Says:

[March 11, 2009 at 5:47 pm](#)

1014.1 Dear Tim, I don't think that we need that  $\mathcal{A}_y$  is dense if one of the variable coordinates is equal to  $k$ . This is actually a crucial observation, that's why we need the flip-top property. At the last step, when only  $r_1$  remains, one word of  $r_1$  spans a dense subset of the  $\{1, 2, \dots, k-1\}^{m-1}$  part of the subspace. By induction there is a line, and by the flip-top property there is an extension to  $k$ . Note that the flip-top property was defined by considering running coordinates  $k$  as well, however it wasn't sensitive for density in any ways.

◦  *gowers* Says:

[March 11, 2009 at 7:15 pm](#)

1014.2 I'm still mystified. In Step 1, let's choose a subspace such that for every sequence  $y$  that avoids running coordinate  $k$ , there is a set of density almost  $\delta$  of  $x$  such that  $(x, y) \in \mathcal{A}$ . Now we go to Step 2. Nothing we've done so far stops  $\mathcal{A}_y$  being empty for every  $y$  that *does* involve  $k$ . So we choose our Sperner pair and have two points  $y$  and  $y'$  (where  $y'$  is obtained from  $y$  by changing some of the  $(k-1)$ s to  $k$ s). How can that help? Indeed, it seems that  $\mathcal{A}_y$  could be empty for every point in the combinatorial line you select.

◦  *jozsef* Says:

[March 11, 2009 at 7:59 pm](#)

Tim - I have to go to the University now, where I will rethink your question again, but I think that the answer is that it is not possible that you switch a running coordinate to  $k$  from  $k-1$  and the density drops because the two points are neighbours. I will get back to you soon.

◦  *gowers* Says:

[March 11, 2009 at 8:02 pm](#)

Jozsef, that gives me time to be more precise about my question. See comment 1016 below.

21. [Polymath1: Probable success « Combinatorics and more](#) Says:

[March 11, 2009 at 7:21 pm](#) | [Reply](#)

[...] northern Israel. Coming back I saw two new posts on Tim Gowers's blog entitled "Problem solved probably" and "polymath1 and open collaborative mathematics." It appears [...]

22.  *Gil Kalai* Says:

[March 11, 2009 at 7:39 pm](#) | [Reply](#)

1015. A bit off-topic

"If this were a conventional way of producing mathematics, then it would be premature to make such an announcement — one would wait until the proof was completely written up with every single *i* dotted and every *t* crossed — "

hmmm, maybe we should regard it premature also in this new mode. What do the rules say?

“but this is blog maths and we’re free to make up conventions as we go along.

So I hereby state that I am basically sure that the problem is solved”

Congratulations! I must admit that even if a full  $k=3$  proof will take 200 more comments and several more weeks the success of the project exceeded my expectations.

“(though not in the way originally envisaged).”

Actually, this is a lovely aspect and a sign of success of the open collaboration idea, isn’t it?



o [gowers](#) Says:

[March 11, 2009 at 7:49 pm](#)

1015.1 Gil, I feel slightly awkward about it, and won’t feel entirely happy until it’s written up in the conventional way. But if one regards a write-up as having a tree structure, then it’s all there on the wiki down to the small sticks, and only a few leaves are missing. Somehow there don’t seem to be any points where one says “I don’t quite see it but I’m sure it works.” It’s more like “I do see it but it’s slightly tedious to write out in full.” Despite that, I wouldn’t have written what I wrote if it had not been for the fact that other participants seem to share my confidence. But for now I will keep the “(probably)”.



23. [gowers](#) Says:

[March 11, 2009 at 8:12 pm](#) | [Reply](#)

1016 Shelah

Jozsef, I’m still trying to follow your argument in 1011. Let me try to be completely precise about where it seems to me to go wrong. Of course, it could be me who is going wrong.

Consider the following example. I take a random subset  $S$  of  $[k]^{r_1 + \dots + r_m}$  and then take the set  $\mathcal{A}$  of all sequences in  $[k]^{r_1 + \dots + r_m}$  such that the final part lies in  $S$ . This set has density  $1/2$ . Now I apply your step 1, and it happens that the  $d_1$ -dimensional subspace I choose has the following property: every point with all running coordinates less than  $k$  belong to  $S$ , and all other points belong to the complement of  $S$ . (With high probability such a subspace exists, and no steps have been taken to avoid it.)

Then the pigeonhole argument gives you that all the points with running coordinates less than  $k$  can be joined to anything you like in  $[k]^{r_1 + \dots + r_{m-1}}$ . However, if you take any line given by the application of Sperner, then both its top two points must involve a  $k$  and give the empty set. Therefore, the whole line gives the empty set (meaning that if your  $m$ th part belongs to the line then you don’t belong to  $\mathcal{A}$ ). So at the second stage of the iteration I don’t see how you can do Step 1 (because of the “with any  $P_i$ ” requirement).



o [jozsef](#) Says:

[March 11, 2009 at 8:52 pm](#)

Tim, what you wrote is perfectly right with the exception that in step one I never consider running coordinates larger than  $k-1$ . (with my old notation “larger than  $k-2$ ”) So, in your example flip top means that the neighbours given by the last  $(m-1)$   $k, k-1$  pair are simultaneously not in  $S$ . I’m still not sure that the argument works, but I don’t see any serious problem yet.



o [jozsef](#) Says:

[March 11, 2009 at 10:40 pm](#)

I have office hours now, so I will begin to write up the argument and if I won't have too many students to show up then I will post the sketch in an hour or so.

o  [gowers](#) Says:

[March 11, 2009 at 10:49 pm](#)

Just to try to be more explicit still, let's imagine that  $k = 3$  and  $r_m = 6$ . Suppose that all sequences of six 0s and 1s are the tail of every preceding sequence, and that the two points 001122 and 001222 are the tails of precisely the same set of sequences, so you choose as your line the line  $001 * 22$ . It seems to me that there is no reason for any point in that line to be the tail of any previous sequence.

A separate question: is the aim to construct in each  $[k]^{r_i}$  a line  $L_i$  in such a way that the product of the lines is flip-top (that is, insensitive to changing 1s to 2s when  $k=3$ ), and  $\mathcal{A}$  is dense in that product? My difficulty is that I don't see the relevance of a statement about  $\{0, 1\}^{r_m}$  when the line  $L_m$  might have 2 as one of its fixed coordinates.

o  [jozsef](#) Says:

[March 11, 2009 at 11:42 pm](#)

I see now. One should be more careful with the selection of the flip-top lines. I still think that it is doable, but it seems that we have to consider more subspaces. I don't have much time, but I will think about it.

o  [gowers](#) Says:

[March 12, 2009 at 1:36 am](#)

I should stress here that I definitely think this is worth attempting. Even if it turns out not to work 100% straightforwardly, the following question exists and is clearly an interesting one. You have a set  $\mathcal{A} \subset [k]^n$  of density  $\delta$ . How large a subspace  $S$  can you find such that  $\mathcal{A}$  has density at least  $\delta/2$  in the subspace, and membership of  $\mathcal{A}$  in the subspace does not change if you flip a  $k$  to a  $k-1$ ? I do not see any philosophical reason for that being significantly harder than Sperner plus induction, though such a reason might emerge I suppose.

24.  [Ryan O'Donnell](#) Says:

[March 11, 2009 at 10:43 pm](#) | [Reply](#)

1017. Varnavides-DHJ(k).

In 1010.1, Tim wrote “*What I'm not doing is fixing almost all coordinates and taking  $M$  small wildcard sets, or something like that. Instead, I'm randomly choosing the sizes of the  $M$  wildcard sets (which I allow to add up to 1, so there are in fact no fixed coordinates at all) and then uniformly at random choosing a line from the resulting subspace.*”


Tim, I think I get it now; as usual I was misinterpreting the word “subspace” and was thinking you were doing what you said you weren't doing 😊 I agree now that the sketch looks relatively solid. Let me see if I can fill in some “leaves”, as you say.

o  [gowers](#) Says:

[March 11, 2009 at 10:50 pm](#)

Great — I'm off for a couple of hours and won't be able to do much more today, if anything.



25.  Ryan O'Donnell Says:

[March 11, 2009 at 11:40 pm](#) | [Reply](#)

1018. Is Equal-Slices of Equal-Slices equal to Equal-Slices?

Suppose we want to try to prove Varnavides-DHJ(3); as in #1010, let's try to show that if  $A \subseteq [3]^n$  has equal-slices density  $\delta > 0$ , and we choose a random combinatorial line by drawing its template from equal-slices on  $([3] \cup \{*\})^n$ , then with probability at least  $\epsilon(\delta) > 0$  the whole line is in  $A$  (assuming  $n$  is sufficiently large).

Tim's strategy for this involves, among other things, picking a combinatorial subspace of dimension  $M$  with *no* free coordinates, according to some distribution  $\nu$ ; then drawing a point from that subspace (thought of as  $[k]^M$ ) from some other distribution,  $\sigma$ .

It would be particularly helpful if the overall distribution is equal-slices on  $[3]^n$ . Also helpful would be if  $\lambda \nu$  is a relatively "natural" distribution; also  $\sigma$  should have full support. (It's also not essential that  $\nu$  generates subspaces with *exactly*  $M$  dimensions, but never mind that for now.)

Here is perhaps the simplest possible proposal: Let  $\nu$  be equal-slices on  $[M]^n$ , and let  $\sigma$  be equal-slices on  $[k]^M$ .

The question is: might combining these two actually give equal-slices on  $[3]^n$  *precisely*?! I will submit some evidence in a reply to this post.

o  Ryan O'Donnell Says:

[March 11, 2009 at 11:53 pm](#)

1018.1. Evidence.

Well, my first piece of "evidence" is that in these simple probability matters, sometimes things just turn out nicely 😊

Here is slightly more rigorous evidence... (I would try to simply give a proof, but I haven't the time to work on it this evening, so I thought I'd quickly throw it out there.)

For my poor brain, let me fix  $M = 100$  and fix  $k = 3$ ,  $[k] = \{R, G, B\}$  (red, green, blue).

We can think of the "outer" equal slices as choosing  $p_1 + \dots + p_{100} = 1$  uniformly, then drawing from the resulting product distribution on  $\{*_1, \dots, *_100\}^n$ . We can think of the "inner" equal slices as choosing  $q_R + q_G + q_B = 1$  uniformly, then changing each  $*_i$  to  $R$  with probability  $q_R$ ,  $G$  with probability  $q_G$ , and  $B$  with probability  $q_B$ .

If you think about it a minute, you see that the final is string is drawn from a *randomly chosen product distribution*  $s_R + s_G + s_B = 1$  on  $\{R, G, B\}^n$ . That's good! If we could just show that that  $(s_R, s_G, s_B)$  is indeed distributed as "uniform on the 3-simplex" we'd be done.


So how *is* it chosen? Well, we think of  $(q_R, q_G, q_B)$  as being chosen first. Then, think of  $p_1, \dots, p_{100}$  as being independent  $\text{Exponential}(1)$ 's. (Of course, this does not give 100 numbers adding to 1, but we use these numbers as proportions, rather than probabilities.)

Then  $(s_R, s_G, s_B)$  gets generated by putting each  $p_i$  into the "R" bucket with probability  $q_R$ , into the "G" bucket with probability  $q_G$ , and into the "B" bucket with probability  $q_B$ . Finally, you add up all the Exponentials in each bucket, and  $(s_R, s_G, s_B)$  is proportional to these totals.


So what is the distribution on the three bucket totals? It shouldn't be too hard to really work it out, but let me

write heuristically here. Once  $(q_R, q_G, q_B)$  is fixed, we expect about  $100q_R$  of the independent Exponentials to go into the R bucket, and similarly for G and B. (Even moreso do we expect this if  $100 = M$  is huge.) And thinking of  $100q_R$  as still huge, we're adding up a huge number of i.i.d. Exponentials, so the CLT will imply that the sum will almost surely be very close to  $100q_R$ . Similarly, the sum in the G bucket will almost surely be very close to  $100q_G$ , and again for B.

But in this case, the proportions will be really close to  $q_R/q_G/q_B$ . I.e., almost surely it seems  $(s_R, s_G, s_B)$  will be very nearly  $(q_R, q_G, q_B)$ . But  $(q_R, q_G, q_B)$  is distributed uniformly on the 3-simplex, as desired!

- o  Ryan O'Donnell Says:  
[March 12, 2009 at 6:00 am](#)

1018.2. I don't think it's exactly correct, but I still hope some variant is...

26.  Ryan O'Donnell Says:  
[March 12, 2009 at 6:29 am](#) | [Reply](#)


1019. ENDS of ENDS = ENDS.

I think I've got it now. Define a distribution on  $k$ -partitions of  $n$  called  $k$ -Equal Non-Degenerate Slices ( $k$ -ENDS): To get it, take  $n$  dots in a line, consider the  $n - 1$  gaps between them, and place  $k - 1$  bars in these gaps, uniformly from among the  $\binom{n-1}{k-1}$  many possibilities (i.e., "without replacement"). Then the intra-bar blocks of dots correspond to  $k$  positive integers  $m_1, \dots, m_k$  which add up to  $n$ . Associating block  $m_i$  to character  $i \in [k]$ , we get a distribution on "Non-Degenerate" slices (each character appears at least once). We can extend this to a distribution on  $[k]^n$  as usual, just by randomly permuting the string.


This is actually a kind of convenient variant on equal-slices, because when you're using equal-slices to pick combinatorial lines, it's a bit annoying to have to worry about the unlikely case of getting  $\emptyset$  wildcards.

Anyway, I think it's now easy to see that if you do  $M$ -ENDS, and then you do  $k$ -ENDS on that (as described in #1018), then you just get  $k$ -ENDS, exactly. Because all that's really happening is that you're picking  $M - 1$  bar-spots uniformly, and then you're picking  $k - 1$  of those  $M - 1$  bars to be the "final bars", uniformly. So that's just the same as picking  $k - 1$  bars uniformly to start with.

I think. Once again, it's late.

27.  Ryan O'Donnell Says:  
[March 12, 2009 at 6:30 am](#) | [Reply](#)

1019.1. "Formula does not parse" above is " $(n-1 \text{ choose } k-1)$ ". [Fixed]

28.  Ryan O'Donnell Says:  
[March 12, 2009 at 7:13 am](#) | [Reply](#)

1019.2. Some googling reveals [this paper](#) (among others) which discusses in its Section 2 the proper nomenclature for this stuff, as well as related probability distributions.

$k$ -ENDS  $\rightarrow$  "random composition into  $k$  parts"; the consequent distribution on strings is "random ordered partition into  $k$  parts".

dots  $\rightarrow$  "balls"

bars  $\rightarrow$  "walls"

29.  [gowers](#) Says:

[March 12, 2009 at 10:21 am](#) | [Reply](#)

1020 DHJ(k)

Ryan, that all looks like good news — it would be very nice if we could get the technicalities to run smoothly so that they don't obscure the main ideas, and your comments above give me hope that it will be possible.

I've got 25 minutes to spare so I'm going to try to explain why I think that DHJ(k) is basically done too. I'm assuming for this that the proof that a line-free set correlates locally with a set of complexity  $k-2$  is going to work out now, which seems likely, as it's been sketched down to a pretty fine level of detail and you've scrutinized the parts that looked most likely to go wrong if anything was going to.

In order to explain how I think the rest of the proof will go, what I need to do is explain the difference between the proof for DHJ(k) and [the proof of DHJ\(3\) outlined on the wiki here](#).

Analogue of Step 1. We need to replace 1-sets by what I call  $\{1, 2, \dots, k-2\}$ -sets. These are sets  $\mathcal{B}$  such that  $x \in \mathcal{B}$  if and only if  $(U_1(x), \dots, U_{k-2}(x)) \in \mathcal{Z}$  for some suitable collection  $\mathcal{Z}$  of disjoint  $(k-2)$ -tuples of sets and  $U_i(x)$  stands for the set of places where  $x$  takes the value  $i$ . To save writing, I'll call these *simple* sets. So the next obvious aim is to prove that a simple set can be approximately partitioned into subspaces. Once that's done, I think it is fair to say that the rest of the proof is exactly as in the  $k=3$  case.

The one other thing that needs to be changed in Step 1 is that the appeal to the multidimensional Sperner theorem has to be replaced by an appeal to multidimensional DHJ( $k-1$ ), [a proof of which is sketched on the wiki here](#). Otherwise, the changes are trivial: 3 becomes  $k$ , and  $\mathcal{U} \otimes [2]^n$  becomes our simple set  $\mathcal{B}$ . (I admit that I haven't been through Step 1 line by line checking what I'm saying — that would use up more than my 25 minutes.)

I've looked through Step 2 and I think all the changes needed are of the trivial variety (changing 3s to  $k$ s and that kind of thing). It might be nice at some point to write an abstract version of this argument, since what we're using seems to be very general — that we can partition  $[k]^n$  into subspaces very conveniently, that the structure of good sets is preserved when we do so, that dense good sets contain lots of subspaces — that kind of thing.

Step 3. And now we do the same trick as we did in DHJ(3) except that we have to apply Step 2  $k-1$  times instead of twice. The end result is that a set of complexity  $k-2$  can be almost partitioned into subspaces, and that's what we're looking for.

30. *Anonymous* Says:

[March 12, 2009 at 4:46 pm](#) | [Reply](#)

1021. Varnavides.

Hi Tim, I think the  $k$ -ary proof that line-free sets correlate with simple sets is now done to about 100% satisfaction in my mind. I'll try to wikify some details later, but to illustrate the simplicity with which it's working out, here is a complete proof of Varnavides:


Assume we have DHJ(k).

Technicality: I want to say that a combinatorial line is (also) “degenerate” if the fixed coordinates do not use each character from  $[k]$  at least once. Deducing DHJ(k) with this very slightly stronger notion of degeneracy is trivial.


Proof of Varnavides: Assume  $A \subseteq [k]^n$  has density  $\delta > 0$  under the “uniform  $k$ -composition” distribution (i.e., equal-slices conditioned on the slice having at least one of each character). Let  $M = M(\delta/2)$  denote the least integer such that DHJ(k) guarantees the existence of a nondegenerate line in any subset of  $[k]^M$  with uniform  $k$ -composition density at least  $\delta/2$ .

Think of choosing  $x \in [k]^n$  as first drawing a random  $M$ -dimensional subspace  $V$  from the uniform  $M$ -composition distribution, and then drawing  $x$  randomly from  $V \cong [k]^M$  according to the uniform  $k$ -composition distribution. Since  $x \in A$  with probability at least  $\delta$ , we know that with probability at least  $\delta/2$  over the choice of  $V$ , the  $k$ -uniform composition density of  $A|_V$  is at least  $\delta/2$ . Call such  $V$ 's "good". By DHJ(k) we know that for any good  $V$ , there is a (nondegenerate) line inside  $A|_V$ .

Now suppose we choose a random line  $L$  in  $[k]^n$  according to the uniform  $(k+1)$ -composition distribution. We claim that it is entirely within  $A$  with probability at least  $\delta/[2M^{k/2}(k+1)^M]$ . To see this, again think of  $L$  as being drawn by first choosing  $V$ , and then drawing a random line in  $V$  according to uniform  $(k+1)$ -composition. With probability at least  $\delta/2$   $V$  is good, in which case there is at least one nondegenerate line in  $A|_V$ . The probability  $L$  is picked to be it is at least  $1/[M^{k/2}(k+1)^M]$ , since the least probability under uniform  $(k+1)$ -composition of  $M$  is at least this quantity (er, or something like that).

 Ryan O'Donnell Says:  
[March 12, 2009 at 4:51 pm](#)

1021.1. That was me again, not logged in. I wanted to add that we can complete the whole "line-free implies correlation with simple set" proof exclusively in this uniform  $k$ -composition measure rather easily, I think. The only thing to really check is that if you choose a  $1 - \epsilon$  fraction of the coordinates  $S$  randomly, do uniform- $k$ -composition on  $S$ , and then independently do uniform- $k$ -composition OR uniform- $(k-1)$ -composition on  $\bar{S}$ , then the overall probability distribution has total variation distance at most  $O(\epsilon\sqrt{n})$  from the usual global uniform  $k$ -composition distribution. Which is, I'm pretty sure, true.

31.  gowers Says:  
[March 12, 2009 at 10:58 pm](#) | [Reply](#)

1022. Shelah

While I wait for reaction about how we should go ahead with writing things up (see [this comment](#)), I want to continue thinking a bit about whether the proof we have could be Shelah-ized. Terry said in a comment some time ago that he is a big fan of translating things back and forth between combinatorial and ergodic languages. I (in common with just about every mathematician) am a big fan of "completing the square," by which I mean completing puzzles of the form "A is to B as C is to ??" For example, "corner is to combinatorial line as large grid is to ??" the answer to which turned out, not completely expectedly, to be " $[2]^k$  for some large  $k$ ". And that was just one part of a bigger such puzzle: "the Ajtai-Szemerédi proof of the corners theorem is to the triviality that a dense subset of  $[n]$  contains two distinct points as ?? is to the proof of Sperner's theorem."

So here I would like to think about the puzzle, "The colour-focusing proof of the Hales-Jewett theorem is to Shelah's proof of the Hales-Jewett theorem as the proof we now have of DHJ(k) is to ??"

The reason I am drawn to this is that I once lectured both proofs of the Hales-Jewett theorem and came to a very clear understanding of the relationship between them, which I'll have to reconstruct, but I remember it as being natural enough that it will be easy to reconstruct. When I have a spare moment, I'll write something about this on the wiki, but for now let me just say that Shelah's basic idea can be interpreted as follows: whereas the old proof of HJ(k) used HJ(k-1) to reduce the problem to the almost trivial HJ(2), Shelah's new proof used HJ(2) to reduce the problem to HJ(k-1). As I say, I'll justify that at some point, but here I just want to take that thought and have a stab at applying it to DHJ(k) instead.

Of course, I'm not sure that it's possible to do so, but just having the goal in mind does, I think, suggest some questions. For instance, here's a place in the proof where DHJ(k-1) comes into the proof of DHJ(k). One first passes to a subspace, which we'll notate as though it is the whole space, such that the density of  $\mathcal{A}$  is almost  $\delta$  and the density of  $\mathcal{A}$  in  $[k-1]^n$  is still positive. Writing  $\mathcal{B}$  for  $\mathcal{A} \cap [k-1]^n$ , we use  $\mathcal{B}$  to construct a dense set  $\mathcal{C}$  of

complexity  $k - 2$  that is disjoint from  $\mathcal{A}$ . It doesn't matter too much what all this means: the main point is that once we have a set of complexity  $k - 2$  we are able to hit it with  $\text{DHJ}(k-1)$ .

So a natural question, if we are trying to Shelah-ize, is whether we can construct some set of complexity 1 (which means it's hittable with  $\text{DHJ}(2)$ , aka Sperner) and thereby, somehow, create a flip-top subspace inside which  $\mathcal{A}$  is dense. Recall that a flip-top subspace is one inside which if you change any coordinate with value  $k-1$  to a  $k$ , then you don't change whether the point  $x$  belongs to the set  $\mathcal{A}$ . So if your set is dense in such a subspace (or rather, dense in the  $[k-1]^n$  part of the subspace) then you can hit it with  $\text{DHJ}(k-1)$  and finish off the proof immediately. Because  $\text{DHJ}(k-1)$  is used just as the inductive step and not as some part of an iterative procedure for getting to the inductive step, a proof of this kind should be far better quantitatively.

In my next comment I shall think about the following question: are there any complexity-1 sets that can be derived from a set  $\mathcal{A}$ ? For now, I don't care whether we can actually use them.

32.  [gowers](#) Says:

[March 12, 2009 at 11:40 pm](#) | [Reply](#)

1023. Shelah.

Continuing on from 1022, a completely simple-minded idea would be to pass to a subspace such that  $\mathcal{A}$  is dense in  $[2]^n$  and still had density almost  $\delta$ . (I'm just going to assume that that's fine.) If we write  $\mathcal{B}$  for  $\mathcal{A} \cap [2]^n$ , then what set  $\mathcal{C}$  can we construct from  $\mathcal{B}$ ?

In the other argument, we looked at all sequences  $x$  such that, for every  $j$ , if you change all  $k$ s to  $j$ s then you get a sequence in  $\mathcal{B}$ . Here, it is clear, we are going to have to do a much more radical collapsing of the sequence. One possibility, which may be nonsense, would be to convert all non-1 coordinates into 2s. In other words, we let  $\mathcal{C}$  be the set of all  $x$  such that the 1-set of  $x$  is equal to the 1-set of some sequence in  $\mathcal{B}$ .

In the other argument, we then argued that  $\mathcal{A}$  is disjoint from  $\mathcal{C}$  if it does not contain a combinatorial line. What happens here? Well, bearing in mind that (using the Shelah proof as a guide) we are not trying to go straight for a combinatorial line, but rather for a flip-top subspace, we should perhaps assume as our hypothesis that  $\mathcal{A}$  does not contain such a subspace.

Turning that round, we would want to prove that if  $\mathcal{A}$  intersects  $\mathcal{C}$  then it contains a flip-top subspace. (This doesn't sound true, but let's see what's wrong with it first.)

For  $\mathcal{A}$  to intersect  $\mathcal{C}$  we need a sequence  $x \in \mathcal{A}$  such that if you convert all its non-1s into 2s, then you get a sequence of 1s and 2s that also belongs to  $\mathcal{A}$ . Why should that be interesting? I don't know.

But suppose instead that we managed to find quite a large subspace  $S$  of such points. Then if  $x$  belonged to  $\mathcal{A} \cap S$  and had a variable coordinate equal to 2, we would ... no, I don't see where this goes.


Part of the problem is that I seem to be going for the flip-top subspace to be contained in  $\mathcal{A}$ , whereas what I should be aiming for is merely that  $\mathcal{A}$  is *dense* in the flip-top subspace (which actually seems to be a flip-bottom subspace).

I'm going to stop this comment here — I hope I've demonstrated the flavour of the argument that I think might conceivably exist.

o  [gowers](#) Says:

[March 12, 2009 at 11:41 pm](#)

1023.1 I should have added the remark that the set  $\mathcal{C}$  has complexity 1.

33.  [Terence Tao](#) Says:  
[March 12, 2009 at 11:45 pm](#) | [Reply](#)


1024. Austin's proof

I've written down an informal (and sketchy) combinatorial translation of the  $k=3$  case of Austin's argument at

[http://michaelnielsen.org/polymath1/index.php?title=Austin%27s\\_proof](http://michaelnielsen.org/polymath1/index.php?title=Austin%27s_proof)

It's actually very much in the style of the proof of the triangle-removal lemma, especially if one takes an analytic perspective rather than a combinatorial one and talks about decomposing functions rather than regularising graphs. In particular, it's best to avoid thinking in the classical language of vertices and take a more "point-less" approach in the spirit of probability theory or ergodic theory (or operator algebras, for that matter), thus for instance one should think about averages of products of functions rather than counting triangles etc. One reason for shifting one's thinking this way is that it allows one to think about strong stationarity in a civilised way, whereas this concept becomes really difficult to work with if one insists on holding on to concrete vertices and edges. (In classical combinatorial language, being strongly stationary is like saying that a certain graph is a "statistically equivalent" to the subgraph formed by restricting of its vertex set from a big space such as  $[3]^n$  to a randomly chosen subspace; but defining what "statistically equivalent" means precisely is sort of a pain unless one decides to give up on vertices altogether.)

It's also clear that this proof will end up being messier than the pure density-increment proof (though cleaner than a finitisation of the Furstenberg-Katznelson proof), and give worse bounds, in large part because one has to apply Graham-Rothschild (with parameters given in turn from a Szemerédi-type regularity lemma!) to get the initial strong stationarity properties. It looks possible to use density increment arguments as a substitute for Graham-Rothschild, but this is still going to have quite lousy bounds.

- o  [Terence Tao](#) Says:  
[March 12, 2009 at 11:55 pm](#)

1023.1 Information-theoretic approach

Hmm, it occurs to me now that the information-theoretic approach to the regularity lemma that I laid out in

<http://front.math.ucdavis.edu/math.CO/0504472>

(where I "regularise" random variables using Shannon entropy) is actually well suited to this problem, and may lead to a relatively slick (albeit alien-looking) proof. What I may do at some point is begin by showing how by playing around with conditional Shannon entropies of various random variables, one can establish the triangle removal lemma, and I believe that this proof may extend relatively painlessly to the case under consideration. More on this later...

34.  [Ryan O'Donnell](#) Says:  
[March 12, 2009 at 11:47 pm](#) | [Reply](#)

1025. Always the uniform-ordered-partition distribution.

Following up on my comment #1006, I feel somewhat sure that we can work *entirely* within the uniform-ordered-partition distribution (i.e., "equal-nondegenerate-slices"). I'm going to try to write it down now to be sure, but if it's correct, it should mean that "DHJ(k)  $\Rightarrow$  Varnavides", "DHJ(k)  $\Rightarrow$  multidimensional-DHJ(k)", and hopefully also "line-free sets correlate with simple sets" can all be proved with simple arguments.

More precisely, they will be the *same* simple arguments we have now, but with no annoyances due to all the tedious



passing back and forth between various distributions via subspaces.



o [gowers](#) Says:

[March 12, 2009 at 11:54 pm](#)

1025.1 Wow — I very much hope you're right about this. I suppose the bit that would be hardest (but not obviously impossible) would be “simple sets can be almost partitioned into subspaces”. In the uniform measure you are then done, but in other measures you can't instantly argue that a set of density  $\alpha$  inside the simple set must have density almost  $\alpha$  on one of the subspaces.

But that could well be just another little task to add to your collection. Now the idea would be to prove that the measure on a simple set can be written as a positive linear combination of the measures on subspaces. If that too can be done cleanly in the UOPD/ENS then we are in fantastic shape.



35. [gowers](#) Says:

[March 13, 2009 at 12:26 am](#) | [Reply](#)

1026. Does triangle removal imply DHJ(3)?

A wild thought this one, especially as it's close to an idea I had several years ago and couldn't get to work. But with the benefit of a bad memory (something that is very helpful in mathematics, in my view) I now can't see what is wrong with it.

The idea is sort of ridiculous. Let's pass to a subspace in which, notating it as  $[3]^n$ ,  $\mathcal{A}$  has average density at least  $\alpha$  in the slices up to  $m$ , where  $m$  is much smaller than  $n$ . Sorry — by that I mean slices where the 1-set and 2-set are of size at most  $m$ .

The hope now is that the disjointness graph becomes dense (because two sets of size  $m$  are almost always disjoint), so we can just apply the usual triangle-removal lemma. My one reason for not completely dismissing this idea is that when I tried it all those years ago I did not have the concept of equal-slices measure to play with.

Unfortunately, it doesn't become dense. Only the UV-part of it is dense (because it is just the 1-set and the 2-set that I am assuming to be small).

Let me ask a slightly different question, which probably has a fairly easy answer, and that answer will probably kill off this little idea. Given  $n$  points in  $[3]^N$  (where  $N$  can be as big as you like), how many combinatorial lines can they contain?



o [jozsef](#) Says:

[March 13, 2009 at 1:07 am](#)

Well, if you don't care about the dimension, then the number of lines might be  $cn^2$ . Just take 0-1 sequences as  $i$  zeros followed by  $n-i$  ones (for every  $i$ ) and 0-2 sequences where the zeros are followed by two-s. For any pair of such 0-1 sequences there is a third 0-2, giving a line. So this number itself won't show that your plan won't work.



o [gowers](#) Says:

[March 13, 2009 at 1:17 am](#)

I don't understand — large should mean  $cn^3$  surely? Or perhaps  $Cn^2$  but with the three bipartite graphs dense.



o *jozsef* Says:

[March 13, 2009 at 2:10 am](#)

Probably I don't understand the question. Isn't it so that two points define the line uniquely. In any case, my example won't work.



o *jozsef* Says:

[March 13, 2009 at 3:29 am](#)

As it happened before, I didn't read your post carefully. It is an extremely busy period for me; heavy teaching, committees, etc. So, I'm just writing those "half baked remarks" as Boris would say. I will try to refrain from that in the future.



36. *Ryan O'Donnell* Says:

[March 13, 2009 at 2:24 am](#) | [Reply](#)

1027. A single distribution?

In preparation for trying to justify what I wrote in 1025, I have written [this document](#) explaining the distribution. I'll wikify it if and only if this scheme works out and proves helpful 😊



o *Ryan O'Donnell* Says:

[March 13, 2009 at 7:22 am](#)

1027.1 I wrote some more [here](#) but I've gotten too sleepy to finish tonight. I'll try to finish tomorrow and explain it properly in a post, but I have to go out of town tomorrow too...

37. [Klas Markström](#) Says:

[March 13, 2009 at 9:30 am](#) | [Reply](#)

1028.

I will make my first visible visit on this side of the project by asking one of the stupid questions discussed in the metathread.

Regarding the question from 1026. As far as I understand any pair of points specify a unique line. if that is the case  $n$  points can specify at most  $Cn^2$  lines.

Now if we have a set of points and a set of lines where every pair of points specify a unique line we have the beginnings of a projective plane, but a projective plane has the same number of points and lines.

So is there another known nice family of combinatorial objects which instead maximises the gap between the number of points and lines?



38. *gowers* Says:

[March 13, 2009 at 9:59 am](#) | [Reply](#)

1029

Klas, your "stupid" question has exposed the stupidity (without inverted commas) of mine — it's not what I meant to ask. Let me have another go.

The way we prove Sperner using equal-slices measure can be thought of like this. We partition the measure on  $[2]^n$  into

measures where the set of combinatorial lines is dense. Each measure is obtained by taking a random permutation of  $[n]$  and taking all intervals in that permuted set, and then any two points with non-zero measure define a combinatorial line.

So a vague version of what I am trying to ask is whether we can somehow find dense structures inside  $[3]^n$  and play a similar game, this time exploiting the corners theorem. However, I don't have a sensible precise version of this question, which perhaps suggests that the answer is no.

39.  [gowers](#) Says:

[March 13, 2009 at 12:47 pm](#) | [Reply](#)

1030. Shelah

Another idea. Instead of struggling to preserve the density of  $\mathcal{A}$  by constructing just one flip-top subspace, perhaps we can partition the whole of (or almost the whole of)  $[k]^n$  into flip-top subspaces. Then in at least one of them  $\mathcal{A}$  would have density almost  $\delta$  and we would be done by induction.

Actually, that's not quite right unless we're rather careful, because in order to apply induction we need  $\mathcal{A}$  to be dense in the  $[k-1]^m$  part of the flip-top subspace. Let me not worry about this for the time being.

Instead, I'd like to think about whether some kind of argument like the one we've used for partitioning 1-sets into subspaces could do the job here. I can't see it in advance so I'll just try to plunge in.

Let  $m$  be a medium-sized integer — much smaller than  $n$ , but much larger than 1 — and write  $[k]^n$  as  $[k]^{n-m} \times [k]^m$ . By the pigeonhole principle we can find a subset  $\mathcal{A}_1$  of  $[k]^{n-m}$  of density  $k^{-m}$  such that for every  $x \in \mathcal{A}_1$  the set of  $y \in [k]^m$  such that  $(x, y) \in \mathcal{A}$  is the same. Let that set be  $\mathcal{B}_1$ , and find a big subspace  $S_1$  in  $[k]^m$  that is flip-top with respect to  $\mathcal{B}_1$ . Then all the subspaces  $\{x\} \times S_1$  such that  $x \in \mathcal{B}_1$  are flip-top. (They could be disjoint from  $\mathcal{A}$ , but we don't care about this.) Let us include them in our attempted partition and remove them from  $[k]^n$ . Note that we have removed a positive proportion (depending on  $m$ ) of  $[k]^n$  as a result of this.

We now partition  $[k]^n$  according to the part that belongs to  $[k]^m$ . The result is a number of subspaces of the form  $S_y = \{(x, y) : x \in [k]^{n-m}\}$ . If  $y \notin S_1$ , then we have not thrown away any of  $S_y$  and we just repeat the first step. However, if  $y \in S_1$ , then things are more complicated, because we have thrown away some of  $S_y$  and are therefore not allowed to use that part in any new cell that we want to add to the attempted partition.

At this point I think it helps to imagine that we've somehow managed to deal with this problem for several iterations. What position would we expect to be in then? Well, we'd have restricted to some subspace (of codimension  $rm$ , where  $r$  is the number of iterations completed), and we would want to find a big chunk (which would be a union of flip-top subspaces) to remove. Hmm, it doesn't seem to work, since by this stage I'm needing to find a big orderly chunk of subspaces inside a set over which I have pretty well no control.

40.  [Ryan O'Donnell](#) Says:

[March 13, 2009 at 3:58 pm](#) | [Reply](#)

1031. New distribution.

[This writeup](#) is practically the same, but I think it sets things up the best. It's a bit long, but that was mainly just to convince *myself* what I was saying is correct.

The gist is this:

The basic objects are length- $m$  lists (ordered) of nonempty sets, where the sets' union is  $[n]$ . Such lists are in obvious correspondence with the set of "nondegenerate" points in  $[m]^n$  (where a point is degenerate if it does not include each character at least once).

The basic operations are: 1.  $\Pi$  = randomly permute the list. 2.  $C$  = do a random “coagulation”: pick  $i \in [m-1]$  uniformly, and merge the  $m$ th set into the  $i$ th set (shrinking the list length to  $m - 1$ ).

The basic distribution is: Start with the length- $n$  list  $\{\{1\}, \{2\}, \dots, \{n\}\}$ . If we first apply a  $\Pi$  operation, and then apply  $n - k$  consecutive  $C$  operations, we get a length- $k$  list, which corresponds to a point in  $[k]^n$ . This is our basic distribution on strings. It is (one can show) the “equal-nondegenerate-slices” distribution.

Finally, note that if we only did  $n - (k + 1)$  of the  $C$  operations, we’d get a string  $\lambda \in [k + 1]^n$ , which we think of as  $([k] \cup \{\star\})^n$ ; i.e., a combinatorial line (template). If we were to further apply the  $C$  operation to this, we’d get one of the  $k$  points on this line, uniformly at random. And, note that  $C\lambda$  is a string in  $[k]^n$  distributed according to our basic distribution.

---

I’m quite sure that at the very least, Varnavides and Multidimensional DHJ(k) become “trivial” now. The proof of line-free sets in  $[k]^n$  correlating with simple sets was already pretty simple assuming these two tools, but I’m hoping it also gets a paragraph proof now.

Of course, I further hope that this helps simplify the “other half” of the proof, namely the density-increment argument. I’d like to think more about it except I’m going to visit family for a few days and will likely not get the chance.



o [gowers](#) Says:

[March 14, 2009 at 9:41 am](#)

Ryan, the link to your write-up isn’t working. I’m looking forward to seeing it ...



41. [gowers](#) Says:

[March 13, 2009 at 4:08 pm](#) | [Reply](#)

1032. Shelah/wikification

I’ve now put on the wiki a very slight modification of the usual proof of the (colouring) Hales-Jewett theorem, together with Shelah’s proof. The purpose of the exercise is to show just how similar the two proofs really are: one uses HJ(k-1) to get you down to HJ(2), while the other one uses HJ(2), in pretty well exactly the same way, to get you down to HJ(k-1). If I’d been feeling silly enough, I would have interpolated between the two cases and written a general argument for how to use HJ(s) to get you down to HJ(k-s+1), which can quite clearly be done. (Indeed, it’s so clear that it can be done that maybe what’s already there can be thought of as doing it by giving enough examples to make clear what the general case is.)

I bothered to do this partly because I find it historically interesting that it took such a long time for Shelah’s proof to be discovered, and I think that presenting the two proofs side by side in that way makes it really quite mysterious. But my main motive is to try to get people interested in the process of Shelahification of the proof we (probably) have of DHJ(k). If the induction can be turned upside down so easily in the colouring version, I will need a pretty convincing argument to persuade me that it cannot be done for the density version. And if it can be done, it would be really good to do it because the bounds would drop from Ackermann to primitive recursive.

The proofs are on [the page about the colouring Hales-Jewett theorem](#).



o [jozsef](#) Says:


[March 13, 2009 at 4:34 pm](#)

Isn’t it so that the iterations give you a  $k$ -( $k-1$ ) flip-top space, then a  $(k-1)$ -( $k-2$ ) flip-top space in the smaller cube


so on? Then if you check any line, it is fliptop for any consecutive pairs, so it is monochromatic. I was wondering if a statistical variant would work or not; Take an almost fliptop subspace (many neighbours have the same colour) and a smaller almost fliptop subspace so on. At the end if at least one line is pairwise fliptop then we are done.

- o  [gowers](#) Says:  
[March 13, 2009 at 7:43 pm](#)

Jozsef, I'm not sure if this is what you're asking, but on the wiki I presented Shelah's argument in terms of flipbottom subspaces instead of fliptop ones.

- o  [jozsef](#) Says:  
[March 13, 2009 at 8:18 pm](#)


Tim, I just changed the notation there. I should have followed yours. While I think that I understand Shelah's proof completely, I'm not sure that I see exactly the limits of the technique. For example, it seems to me that we don't use that the points of every neighboring pair have the same colour. For example if 11111... is blue and 21111... is red in the first fliptop subspace, it won't make any trouble later as the line defined by the two elements is outside of the next subspace. In general, every point with coordinate 1 has exactly one "semi-neighbour" which is important in the iteration, where the 1-s are replaced by 2-s. But I don't see an easy way to consider the important pairs only.

- o  [gowers](#) Says:  
[March 13, 2009 at 8:29 pm](#)

Yes, I had a similar thought soon after writing up the wiki page. I modified the usual proof so as to ensure, unnecessarily, that it reduced HJ(k) to HJ(2). But the usual colour-focusing proof doesn't need that full strength. I therefore suspect that it is possible to modify the Shelah proof so that instead of producing a fliptop subspace you produce a subspace with some weaker property. But I haven't tried to work out any details, or even to convince myself that it's definitely possible.

42. [Gower's Polymath Experiment - Problem probably solved « Healthy Algorithms](#) Says:  
[March 13, 2009 at 6:31 pm](#) | [Reply](#)

[...] combinatorial proof of the density Hales-Jewitt theorem (DHJ). Big congratulations to them, because the problem is solved, probably. Summarizing why he spent his time on this particular problem, Terry Tao wrote: I guess DHJ is [...]

43.  [gowers](#) Says:  
[March 13, 2009 at 8:15 pm](#) | [Reply](#)

1033. Shelah

I'm still just throwing out guesses here (and seeing whether I get anywhere with them in real time — no luck so far) and here's another one.

In the argument we have now, we could present it as saying that we try to find a combinatorial line, and if we can't then we get correlation with a set of complexity  $k-2$ . If we are going to try to Shelah-ize, then we're very much hoping to find a fliptop subspace (or perhaps, as Jozsef suggests above, something slightly weaker but still sufficient) where  $\mathcal{A}$  is still dense. So perhaps we should plunge in and try to construct such a subspace, but instead of being completely determined to succeed, we should hope that *either* we will succeed *or* we will get correlation with a set of complexity 1.

The first step of what I was trying to do in 1013, following Jozsef in 1011, was as follows. Write  $[k]^n$  as  $[k]^r \times [k]^s$ , where  $s$  is much bigger than  $r$ . For each  $y \in [k]^s$ , define  $\mathcal{A}_y$  to be  $\{x \in [k]^r : (x, y) \in \mathcal{A}\}$ . We would like to find  $y_1$  and  $y_2$  that differ only in a few places where coordinates that were  $k-1$  in  $y_1$  are  $k$  in  $y_2$ , such that  $\mathcal{A}_{y_1} = \mathcal{A}_{y_2}$ . So far so good, but we would also like  $\mathcal{A}$  to have density almost as big as the starting density when restricted to  $[k]^r \times L$ , where  $L$  is the unique combinatorial line that contains both  $y_1$  and  $y_2$ .

Now if we *can't* do this, then we potentially get an interesting set inside which  $\mathcal{A}$  has reduced density. It's formed as follows: for each pair  $(y_1, y_2)$  as above, form the line  $L$ , and take the union  $X$  of all those lines.

Ignoring for now the fact that the union is not an arithmetic operation, so we can't actually conclude without extra work that if  $\mathcal{A}$  has reduced density on every  $[k]^r \times L$  that it has reduced density on  $[k] \times X$ , let's see whether  $X$  has any good structure, such as low complexity.

A sequence  $z \in [k]^s$  belongs to  $X$  if and only if we can find some  $j$  and some subset  $Z$  of the  $j$ -set of  $x$  such that when we overwrite  $Z$  with  $(k-1)$ s or  $k$ s, we get two points  $y_1$  and  $y_2$  with  $\mathcal{A}_{y_1} = \mathcal{A}_{y_2}$ . Hmm, that doesn't seem to have low complexity. Back to the drawing board.



o [jozsef](#) Says:

[March 13, 2009 at 8:47 pm](#)

Just a technical remark; when you cut  $[k]^n$  into two parts the you can play with the densities. The density of  $\mathcal{A}$  is  $c$ , say. If any  $\mathcal{A}_x$  or  $\mathcal{A}_y$  had density larger than  $c + \epsilon$  then we can go there and repeat. That means that the density of  $x$  that  $\mathcal{A}_x$  is at least  $c - \epsilon_2$  is almost 1, I think.



44. [gowers](#) Says:

[March 13, 2009 at 8:48 pm](#) | [Reply](#)

1034. Shelah

Time for a very vague thought indeed. Maybe the mistake in our thoughts so far about a density version of Shelah's proof is that we're still using colourings too much. Is there some "density equivalent" of a flip-top subspace?

To get a handle on this question, let us try to focus on the property that's good about flip-top subspaces: that if you ever find a line with all the points up to  $k-1$  of the same colour, then the whole line must be monochromatic. The obvious density version of that would be that if the first  $k-1$  points belong to  $\mathcal{A}$  then so does the  $k$ th.

One immediate small observation is that for this to be the case, all we need is that if you change some  $(k-1)$ s to  $k$ s, you don't go *out* of  $\mathcal{A}$ . We don't mind if this operation takes you *in* to  $\mathcal{A}$ . But is there perhaps some much more general circumstance under which we can deduce that  $\mathcal{A}$  contains a line from the fact that  $\mathcal{A} \cap [k-1]^n$  does? Or perhaps from the fact that  $\mathcal{A} \cap [k-1]^n$  contains *lots* of combinatorial lines? In the latter case, what is the set of points that  $\mathcal{A}$  will be forced to avoid?

I have to go now, but at first glance it still doesn't look as though there are any super-low-complexity sets floating about.



45. [Terence Tao](#) Says:

[March 14, 2009 at 4:47 am](#) | [Reply](#)

1035. Austin's proof w/o Ramsey theory or density increment

I've written up a more or less complete proof of DHJ(3) using a (rather abstract) triangle-removal approach, based on Austin's proof, at



[http://michaelnielsen.org/polymath1/index.php?title=Austin%27s\\_proof\\_II](http://michaelnielsen.org/polymath1/index.php?title=Austin%27s_proof_II)

I found that stationarity is not used as much as I had previously thought, and can in fact be obtained by a simple energy increment argument and the pigeonhole principle rather than more high-powered Ramsey-theoretic tools such as Graham-Rothschild.

The argument is a bit tedious though. It follows the approach to triangle removal used for instance in [my reworking of the hypergraph removal lemma](#), in particular, heavily using lots of conditional expectation computations, rather than the traditional language of cleaning out “bad” cells and then counting both “non-trivial” and “trivial” triangles in the surviving cells. But there are some distracting complications due to the multiplicity of scales, and the presence of some additional random choices (basically, one has to choose some random index sets  $I$  to flip from 0 to 2, etc). One has to make the whole argument “relative” or “conditioned” to these random choices, which makes the argument look stranger than it normally does.

The bounds seem to be tower-exponential in nature - not so surprising, being based on triangle removal. Amusingly, if one used uniform measure rather than equal-slices measure, one would be forced to make a double tower-exponential, because of the fact that each new scale would have to essentially be the square root of the previous. Hooray for equal slices measure!

I'm pretty sure the density-increment argument is going to end up being shorter and simpler than this one, but I'm putting the triangle removal proof here for completeness.

I'll be travelling overseas shortly, and so I'll probably be contributing very little for the next week or two.

46. [Polymath « Maxwell's Demon](#) Says:

[March 14, 2009 at 12:20 pm](#) | [Reply](#)

[...] even though it touches on my central topics of maths and communication. Now with its preliminary success feels like a good time to do so. [...]

47.  [gowers](#) Says:

[March 14, 2009 at 12:57 pm](#) | [Reply](#)

1036. Graham-Rothschild

A quick pair of closely related questions: is the density version of Graham-Rothschild known, and do we now have the tools to prove it (whether or not it is known)?

For the benefit of anyone who can't be bothered to look it up on the wiki, the Graham-Rothschild theorem is like the Hales-Jewett theorem except that the basic objects are combinatorial subspaces of some fixed dimension rather than points. For instance, a special case is that if you colour the lines in  $[3]^n$  with finitely many colours, then there is a combinatorial subspace of dimension  $m$  such that all the lines in that subspace have the same colour — provided  $n$  is big enough in terms of  $m$  and the number of colours.

It may be that the question is easy. For instance, suppose we think of a combinatorial line in  $[k]^n$  as a point in  $[k+1]^n$  with coordinates that belong to the set  $\{1, 2, \dots, k, *\}$ . If we then find in our dense set  $\mathcal{A}$  of such lines an  $m$ -dimensional subspace, what does that translate into? It gives us some wildcard sets  $Z_1, \dots, Z_m$ , each of which can be converted into an element of this alphabet, and it also gives us some fixed coordinates. The problem is that the fixed coordinates can be equal to  $*$ , so we do not end up with all the lines in some combinatorial subspace.

At first this looks problematic: given a dense subset of  $\{1, 2, \dots, k, *\}^n$ , we can't hope to find a subspace inside it with no fixed coordinates equal to  $*$ , since  $\mathcal{A}$  might consist of all points with at least one coordinate equal to  $*$ . However, that's not a problem, because our *entire space* consists of points with at least one  $*$  (since they are

combinatorial lines).

So here's a DHJ(3) variant. Suppose you have a dense subset  $\mathcal{A}$  of  $[3]^n$ . Must there be an  $m$ -dimensional subspace  $S$  such that all the fixed coordinates are equal to 1 or 2, and every point in  $S$  that includes at least one 3 is an element of  $\mathcal{A}$ ?



o [gowers](#) Says:

[March 14, 2009 at 1:28 pm](#)

Just realized that the formulation of that last question was a bit careless, since the usual example of a set with roughly equal numbers of all three coordinates is a counterexample to this when  $m=2$ . So it's not instantly obvious what a density version of Graham-Rothschild would even say, but I hope that with the help of equal-slices measure one might be able to formulate a decent statement.



o [Randall](#) Says:

[March 14, 2009 at 10:16 pm](#)

Generally, these Ramsey theorems are either "projective" (Schur, Hindman, Ramsey), "affine" (van der Waerden, HJ) or "projective/affine" (Graham-Rothschild, Carlson, CST). Nothing with a projective aspect will typically have a "naive" density version, though I don't know exactly what you are shooting for here. (Presumably you don't mean something that the set of words having an odd number of 3s would be a counterexample to.)



o [jozsef](#) Says:

[March 14, 2009 at 10:58 pm](#)

Hi Randall, It is an interesting question if these problems have density versions or not. Let's consider Schur's thm. Ben Green proved that if a subset of  $[n]$ ,  $S$ , has only a few,  $o(n^2)$ , solutions to  $x+y=z$  in  $S$  then one can remove  $o(n)$  elements from  $S$  to destroy all solutions. This might be viewed as a density version of Schur's theorem. On the other hand I don't see a simple way to deduce Schur's theorem from this density version. I tried it as it might give a bound on the triangle removal lemma, but without success.



48. [Ryan O'Donnell](#) Says:

[March 14, 2009 at 5:05 pm](#) | [Reply](#)

1037. Uniform ordered partition measure.

[Here](#) is the corrected link from 1031; in other words, add ".pdf" to the broken link. I think this measure will work well with the Graham-Rothschild problem described above; again, I'll try to write more in a few days when I get back home.



o [Ryan O'Donnell](#) Says:

[March 14, 2009 at 5:06 pm](#)

1037.1 Drat, I keep botching it! It's [here](#).



49. [jozsef](#) Says:

[March 14, 2009 at 7:17 pm](#) | [Reply](#)

1038. Subspace implies line

A simple argument shows that DHJ( $k$ ) implies  $d$ -dimensional subspaces in  $[k]^n$  for any dense subset (with large enough

n). I remember seeing a wiki article about this, but I was unable to find it this morning. It is also true that a subspace theorem implies DHJ. That looks fairly trivial but let me explain it; I want to show that the following statement is equivalent to DHJ(k): “There is a constant,  $c < 1$  that for every  $d$  there is an  $n$  that any  $c$ -dense subset of  $[k]^n$  contains a  $d$ -dimensional subspace.”

I would like to show the direction that the statement above implies DHJ(k). The other direction is already wikified. As before, write  $[k]^n$  as  $[k]^r \times [k]^s$ , where  $s$  is much bigger than  $r$ . For each  $y \in [k]^s$ , define  $\mathcal{A}_y$  to be  $\{x \in [k]^r : (x, y) \in \mathcal{A}\}$ .

Let  $Y$  denote the set of  $y \in [k]^s$ , that  $\mathcal{A}_y$  is empty. Suppose that  $\mathcal{A}$  is large, line-free, and its density is  $\delta = \Delta - \epsilon$  where  $\Delta$  is the limit of density of line-free sets and  $\epsilon < (1 - c)\Delta$ . We can also suppose that no  $\mathcal{A}_y$  has density much larger than  $\Delta$  as that would guarantee a combinatorial line. But then the density of  $Y$  is at most  $1 - c$ , so there is a  $c$ -dense set  $Z \subseteq [k]^s - Y$  such that any element is a tail of some elements of  $\mathcal{A}$ . For every  $y \in Z$  choose an  $x \in [k]^r : (x, y) \in \mathcal{A}$ . This  $x$  will be the colour of  $y$ . It gives a  $[k]^r$  colouring on  $Z$ . By the initial condition  $Z$  contains arbitrary large subspaces, so by HJ(k) we get a line in  $\mathcal{A}$ .



◦ *jozsef* Says:

[March 15, 2009 at 5:06 pm](#)

I've find Tim's post about DHJ(k) implies subspace DHJ(k).

[http://michaelnielsen.org/polymath1/index.php?title=DHJ\(k\) implies multidimensional DHJ\(k\)](http://michaelnielsen.org/polymath1/index.php?title=DHJ(k)_implies_multidimensional_DHJ(k))

The right title of the post above would be “Weak subspace DHJ(k) implies DHJ(k)”.



50. *gowers* Says:

[March 14, 2009 at 11:33 pm](#) | [Reply](#)

1039 Graham-Rothschild

I've got a new attempt at a density version of Graham-Rothschild that might have a chance of being true. I realized that equal-slices measure wouldn't rescue the previous version, since the set of lines with wildcard set of size approximately  $\sqrt{n}$  is a counterexample.

As I write, I realize that my new attempt fails too. Basically, if you've got a 2D subspace then you must have lines with wildcard sets of sizes  $x$ ,  $y$  and  $x+y$ , so a density version looks as though it would have to imply a density version of Schur's theorem, which is of course false. (I realize that I am sort of repeating what people have already said in their replies to 1036.)

Maybe one could go down the route Jozsef implicitly suggests and try proving that if  $\mathcal{A}$  is a set that contains few  $m$ -dimensional subspaces, then one can remove a small number of elements from  $\mathcal{A}$  and end up with no  $m$ -dimensional subspaces. I'm not sure that I feel like trying this though ...

51. *Klas Markström* Says:

[March 15, 2009 at 10:52 am](#) | [Reply](#)

1040.

Since the density of a subset of  $[k]^n$  is quite insensitive to the removal of an element, at least for large  $n$ , a set with positive density will contain many lines. How many lines can be guaranteed, as a function of the density and  $n$ ?

From the equal-slices proof of Sperner it is not hard to get a lower bound for DHJ(2) so I'm curious as to what can be said for DHJ(3).



52. *jozsef* Says:

[March 15, 2009 at 6:33 pm](#) | [Reply](#)

## 1041. Line-free sets correlate locally with complexity-1 sets

I would like to go back to one important part of the DHJ(3) proof, to analyze it from a slightly different angle. Let us consider our set  $\mathcal{A}$ , a dense subset of  $[3]^n$ , as a subset of  $(\mathbb{Z}/3\mathbb{Z})^n$ . Build a graph on  $\mathcal{A}$  as follows; The vertex set is  $\mathcal{A}$  and two vertices,  $a$  and  $b$ , are connected iff for  $c: a+b+c=0$   $c$  is also in  $\mathcal{A}$ . If  $\mathcal{A}$  was dense then there are many edges in any typical subspace. Now DHJ is equivalent to the statement that there are two connected elements of  $\mathcal{A}$  with the same set of 3-s and that one's set of 1-s contains the 1-s of the other. This model leads us to a density Sperner problem.

It just turned out that I have to go somewhere right now. I will come back in a few (4+) hours.



o Ryan O'Donnell Says:

[March 15, 2009 at 11:38 pm](#)

## 1042. The distribution.

Hmm. In fact, it may be even better to view strings in  $[k]^n$  being generated in the time-opposite way from the one I described. Specifically, equal-slices and its nondegenerate variant are the same as the following Polya-Eggenberger urn process:

[Non-degenerate version:] Start with one of the  $k!$  permutations of the string  $123 \dots k$ . Repeat the following  $n - k$  times: pick a random character in  $[k]$  with probability proportional to the number of times it appears already in the string. Now insert that character into a random place in the string.

[Equal-slices version, I think:] Same, except: a) start with the empty string; b) for the purposes of proportionality, pretend there is a phantom copy of each of the  $k$  characters.

This yet-another viewpoint on the equal-slices distribution helps with making “subspace arguments” (which the uniform distribution was good for): the point is, if you do a  $k$ -color Polya urn process for  $M + n$  steps and  $n \gg M$ , then the final distribution hardly depends at all on what happened in the first  $M$  steps.

Will write more when I get back in two days.



53. Gil Kalai Says:

[March 15, 2009 at 11:57 pm](#) | [Reply](#)

## 1042. Density increasing and an analog problem

(This is a little off topic) Let me mention a problem which I thought of as analogous to Roth/cap set where the gaps between lower and upper bounds are similarly shocking and the current density increasing arguments cannot help; (It is related to old work of mine with Kahn and Linial and also to some more recent work with Shelah that we did not publish.)

you have a subset  $A$  of  $\{0, 1\}^n$  of density  $c$  and you would like to find a combinatorial subcube  $F$  of dimension  $0.9n$  so that the **projection** of  $A$  to  $F$  is of large density say  $0.99$ . In other words, we want to find a set of  $0.9n$  coordinates so that for a fraction  $0.99$  of the vectors supported on this set we can find a continuation on the other coordinates that is in  $A$ . (We usually talk here about restriction to a subcube/subspace and not about projections. But traditionally sections and projections are not unrelated.)

By a density increasing argument doing it one coordinate at a time it was known from the late 80s that this can be achieved if  $c$  is  $n^{-\alpha}$  for  $\alpha = 1/5$  or so. A conjecture by Benny Chor asserts that  $c = 0.999^n$  is good enough!

I think it is a little analogous to Roth (or cap set)

a few points:

- 1) The proof is also by a slow density increasing argument (you reduce the dimension by one every time) and there are examples the such an argument cannot be improved.
- 2) There are some conjectures (by Friedgut and others) which may explain why we can get the density down to  $n^{-\beta}$  for every  $\beta$  maybe even  $\beta = \gamma \log n$  but no plans beyond it.
- 3) There are alarming examples by Ajtai and Linial of Boolean functions described by certain random depth 3 circuits (that Ryan already mentioned) that may (or a varian of) give a counter example. It is complicated to check it.

I admit that the analogy with density increasing argument for Roth-type problems is not very strong: this problem is about projection to subcubes and there it is about restrictions to subspaces or similar creatures; But there may be some connecion.

In particular I would try subsets described by low depth small size circuits (with operations over  $\{0,1,2\}$ ) as candidates for counter examples for the most ambitious conjectures regarding Roth and cap sets.

(On the positive side maybe more sophisticated density increasing arguments of the kind we talk about here can be used in this problem.)

54.  *Kristal Cantwell* Says:

[March 16, 2009 at 3:12 am](#) | [Reply](#)

1043. Graham-Rothschild

From 1036: "A quick pair of closely related questions: is the density version of Graham-Rothschild known, and do we now have the tools to prove it (whether or not it is known)?

For the benefit of anyone who can't be bothered to look it up on the wiki, the Graham-Rothschild theorem is like the Hales-Jewett theorem except that the basic objects are combinatorial subspaces of some fixed dimension rather than points. For instance, a special case is that if you colour the lines in with finitely many colours, then there is a combinatorial subspace of dimension  $m$  such that all the lines in that subspace have the same colour — provided  $n$  is big enough in terms of  $m$  and the number of colours. "

I doubt there is density version of the Graham-Rothschild theorem. If one fixes a coordinate and deletes all lines that has a constant coordinate at that point then that will only lower the number of lines by a constant factor but it will prevent the formation of any two dimensional space with all its lines monochromatic as in that case the intersection of the moving coordinates of all of the lines is the null set.

55.  *Kristal Cantwell* Says:

[March 16, 2009 at 3:19 am](#) | [Reply](#)

1044. Density theorems and Ramsey theorems

Density theorems are stronger than Ramsey theorems. Any density theorem can be converted to a Ramsey theorem as follows. One sets the density less than  $1/r$  and gets a large enough configuration so that the if the density is greater than  $1/r$  there will be the desired result than for any  $r$  coloring there will be one color with density  $1/r$  or more and we will have a monochromatic configuration in that color.

56.  *Kristal Cantwell* Says:

[March 16, 2009 at 5:49 am](#) | [Reply](#)

1045. Density Schur Theorem

I have just remembered the counterexample for a density version of Schur's theorem. One just takes the odd numbers, they have density  $1/2$  but since the sum of two odd numbers is even there are no triples  $a, b, c$  in the set such that  $a+b=c$ .



o [gowers](#) Says:

[March 16, 2009 at 7:54 am](#)

The fact that the density version of Schur's theorem is false implies that density Graham-Rothschild is false — see 1038.



57. [gowers](#) Says:

[March 16, 2009 at 11:36 am](#) | [Reply](#)

1046. Bad news and good news.

Bad news: our proof of DHJ(3) is wrong!

Good news: it isn't too hard to fix.

But it's still quite amusing that nobody noticed that [Substep 2.2](#) of the write-up on the wiki of the crucial lemma that a dense 1-set can be almost completely partitioned into subspaces was nonsense as presented. I've left the old version there as a cautionary tale. Fortunately, the reason for the mistake was that I had translated the corresponding part of Terry's argument in a sloppy way — the argument itself was sound. But it did give me a bit of a scare ...



o [Randall](#) Says:

[March 16, 2009 at 6:28 pm](#)

1046.1 Good news and polymath....

Well, perhaps it will be somewhat reassuring that the changes you have made seem to be only adding more detail, and that, from the ergodic perspective at least, it was natural to omit said detail; loosely translated (and unless I misunderstand), what you seem not to be checking more carefully is that when you remove the fibers (over some factor) that recur under a subspace, what you have left is still measurable with respect to that subspace. So perhaps the good news is actually that polymath is like math with a net. (In which case the bad news is the danger than everyone involved will internalize that piece of good news to the point where it becomes a problem.)



o [Randall](#) Says:

[March 16, 2009 at 6:32 pm](#)

I of course meant (a) “now” to be checking more carefully, and (b) measurable with respect to that “factor”.



58. [gowers](#) Says:

[March 16, 2009 at 1:09 pm](#) | [Reply](#)

1047. Wikification

I have now wikified an [abstract version of the iterative argument](#) that was wrong before. From the abstract point of view the property that I was forgetting about was the all-important sigma-algebra property. This version should be fine now and is intended to be sufficiently general to deal with DHJ(k) as well. The argument could be sharpened up a bit towards the end — I ran out of energy.



59. [gowers](#) Says:



[March 16, 2009 at 1:17 pm](#) | [Reply](#)

Metacomment. I've created a thread for comments 1050-1099, so this one should draw to a close.

60. [Michael Nielsen » The Polymath project: scope of participation](#) Says:

[March 20, 2009 at 9:29 pm](#) | [Reply](#)

[...] week, Gowers announced that the problem was “probably solved”. In fact, if the results hold up, the project [...]

61. [Concluding notes on the polymath project — and a challenge « What Is Research?](#) Says:

[March 24, 2009 at 5:27 pm](#) | [Reply](#)

[...] project, as of February 20, 2009. The project, which began around February 2, 2009, has now been declared successful by Gowers. While the original aim was to find a new proof of a special case of an already proved theorem, the [...]

62. [An Open Discussion and Polls: Around Roth's Theorem « Combinatorics and more](#) Says:

[March 25, 2009 at 9:13 am](#) | [Reply](#)

[...] (as a little spin off to the polymath1 project, (look especially here and here)) if you have any thoughts about where the truth is for these problems, or about how to [...]

63. [A gentle introduction to the Polymath project « The Number Warrior](#) Says:

[March 26, 2009 at 1:41 am](#) | [Reply](#)

[...] to an online collaborative approach, then kicked things off with a blog post. Six weeks later, the main problem he proposed was declared (essentially) solved. However, the project still continues apace, especially at threads at Terry Tao's [...]

64. [Un blog concentra a cientos de matemáticos en la demostración conjunta de un teorema « Francis \(th\)E mule Science's News](#) Says:

[April 28, 2009 at 9:27 pm](#) | [Reply](#)

[...] (probably),” Nature Physics, 5: 237, April 2009 . Hay una entrada en el blog de Tim con el mismo título. Más allá de la demostración la experiencia muestra a los jóvenes que se inician en la [...]

## Leave a Reply

Logged in as [Ryan O'Donnell](#). [Logout »](#)

☐ Notify me of follow-up comments via email.

---

[Blog at WordPress.com.](#)

[Entries \(RSS\)](#) and [Comments \(RSS\)](#).

⌂

# Gowers's Weblog

Mathematics related discussions

« [Polymath1 and open collaborative mathematics](#)  
[PCM errata II](#) »

## DHJ(3) and related results: 1050-1099

By gowers

I'm not sure how many more comment threads we will have, but we are running out of the 1000-1049 thread, so it's time for a new one. The main news to report since the last post is that progress is being made on writing up the proof of DHJ(3) and DHJ(k). At the moment it is more like preparatory sketches, but they are pretty detailed and can, as usual, be found [on the wiki](#). It looks as though some of the more technical parts will end up very streamlined thanks to work of Ryan O'Donnell: the final proof of DHJ(k) should be surprisingly short (though I hope that we will write it up with plenty of accompanying explanation so that it is not too compressed and hard to understand).

This entry was posted on March 16, 2009 at 1:16 pm and is filed under [polymath1](#). You can follow any responses to this entry through the [RSS 2.0](#) feed. You can [leave a response](#), or [trackback](#) from your own site.

### 88 Responses to "DHJ(3) and related results: 1050-1099"

1.  [gowers](#) Says:

[March 16, 2009 at 5:38 pm](#) | [Reply](#)

1050. Shelah

I still haven't given up on the idea of Shelahifying the proof of DHJ(k), but a weakness in my attempts so far is that I have no good story to tell about corners. So in this comment I am going to try to guess how a proof of the corners theorem might go if the induction worked the other way round. In other words, I want an argument that solves the higher-dimensional corners problem with the one-dimensional Szemerédi theorem as its main tool (rather than Szemerédi in one dimension down). It feels rather unlikely that this is possible, but we aren't going to know if we don't try.

I think the difference should show up only once we get to three dimensions. I've been struggling to think of some way that complexity-1 sets might come into the picture (which in this context means Cartesian products  $X \times Y \times Z \subset [n]^3$ ), and the best I've come up with so far is this. The conventional proof takes a "dense diagonal," meaning a plane of constant  $x + y + z$  that contains many points of the set  $A$ , and exploits the fact that this set contains many aligned equilateral triangles to obtain a structured set that  $A$  is forced to avoid. What if we try to do things the other way round? We start with the fact that  $A \subset [n]^3$  is dense and try to deduce from that that almost all diagonal planes are sparse. The hope is that the argument works unless  $A$  correlates with a set of complexity 1.

Let's be more precise about this. If  $A$  is a 3D-corner-free subset of  $[n]^3$  and  $(x, y, z) \in A$  then for any  $d$  we cannot

have all three points in the aligned equilateral triangle  $(x + d, y, z), (x, y + d, z), (x, y, z + d)$ . Thus, each point in  $\mathcal{A}$  rules out a large number of triangles in a large number of diagonal planes. What we would like to be able to deduce from that is that the density of  $\mathcal{A}$  in those planes is necessarily small.

This suggests the following question, which I have not had time to think about properly. For what collections  $\mathcal{C}$  of corners (in  $[n]^2$  now) is it the case that every dense subset of  $[n]^2$  contains a corner in  $\mathcal{C}$ ?

A more specific question is this. A set of complexity 1 in  $[n]^2$  is a set of the form

$K = \{(x, y) : x \in U, y \in V, x + y \in W\}$ . Let us say that  $\mathcal{C}$  *correlates with*  $K$  if there are more corners with vertices in  $K$  than there should be. Suppose that  $\mathcal{C}$  does not correlate with any dense set of complexity 1. Does it follow that every dense set has a corner in  $\mathcal{C}$ ? And can it be proved surprisingly easily?

If the answer to both questions were to be yes then we might be in business, but as I said at the beginning this feels a bit over-optimistic.

Another point to make is that the above question could need a little adjustment before it becomes sensible. For instance, it might need to talk about more local forms of correlation.

◦  *jozsef* Says:

[March 16, 2009 at 6:50 pm](#)

Tim, could you please clarify your last question? “ $\mathcal{C}$  correlates with if there are more corners with vertices in  $K$  than there should be.” (Probably you meant correlates with  $K$ .) I can’t see any condition which guarantees that  $\mathcal{C}$  is not even empty. Probably I missed something.

◦  *gowers* Says:

[March 16, 2009 at 7:29 pm](#)

OK. First, I’ve added the missing  $K$  — thanks for pointing that out. The idea I had in mind was to assume that  $\mathcal{C}$  had density  $\alpha$  in the set of all possible corners, that  $\beta$  was the density of corners with vertices in  $K$ , and that the density of such corners that belong to  $\mathcal{C}$  was significantly different from  $\alpha\beta$ .

◦  *jozsef* Says:

[March 16, 2009 at 8:15 pm](#)

I see. This sounds to me like another sparse hypergraph regularity problem. The complexity 1 set is the shadow of  $\mathcal{A}$  and we check if  $\mathcal{A}$  is quasirandom on its shadow or not.

The complexity 1 set contains many corners (it is a large Cartesian product with small sumset) so we want to conclude that  $\mathcal{A}$  contains corners or it is “very irregular”. Isn’t it your modified Ajtai-Szemerédi proof? I have to read your post again, as I lost track of how this relates to a Shelah type argument.


◦  *jozsef* Says:

[March 17, 2009 at 6:25 pm](#)


Regarding to the first question; There is a very sparse set of corners that every dense subset of  $[n]^2$  contains one from that set. Using DHJ(4) (I know that we don’t want to use it but anyways) one can show that there is a set of  $n^{2.3219\dots}$  corners that every dense set contains a corner from that set. To construct such set, consider  $[n]$  as a  $m = \log n$  dimensional cube and there is a 1-1 correspondence between the points of  $[n]^2$  and  $[4]^m$ . Take the triangles defined by combinatorial lines. By DHJ(4) every dense subset contains a corner (even a square) from this set of corners. Maybe one can find an even sparser set of corners but I don’t see how. I also can’t see a reasonable lower bound for the size of such set of corners. It seems to be an interesting problem.

- o  [Gil Kalai](#) Says:  
[March 17, 2009 at 9:29 pm](#)

Shalahifying DHJ(k) is a noble cause indeed; What about DHJ(3)? does the proof gives inverse ackerman type function for DHJ(3) as well? Is the goal of density  $1/\log \log \log \dots \log n$  just for DHJ(3) out of reach with the new proof? (for a bounded number of logs? for two logs?)

- o  [Gil](#) Says:  
[March 18, 2009 at 7:42 am](#)


Another quick question: what are the best bound known for multidimensional Sperner? (The wiki argument with DHJ( $2^k$ ) gives terrible bounds). I recall we discussed it but didnt find it. (Another question: is there an easy way to do search on the union of all threads?)

2.  [gowers](#) Says:  
[March 17, 2009 at 8:19 pm](#) | [Reply](#)

1051. Shelah

I'm going to have a quick go at the question I asked in 1050 by following the modified Ajtai-Szemerédi proof and seeing what assumptions I need to make about  $\mathcal{C}$  to get the argument to work.

Actually, I've got to go — will do this later.


3.  [gowers](#) Says:  
[March 18, 2009 at 12:07 am](#) | [Reply](#)

1052. Shelah

OK here goes. The first step was to find a dense diagonal. Do we still want to do that? Well, the point of a dense diagonal was that it gave rise to a large Cartesian product full of forbidden points. But if we've only got a certain subset of corners, then we cannot make such a strong conclusion.

So what *can* we conclude? To investigate this, we could see what happens if the set of pairs on some diagonal that belong to corners in our system  $\mathcal{C}$  is random. If there are  $\alpha n^2$  of these pairs, then they are ruling out a ... random set of size  $\alpha n^2$ , which isn't very helpful.

Actually, what I'm doing is nonsense because the contrapositive of what I'm trying to show is that a dense “diagonal plane” forces a complexity-1 bias in the set A, which is clearly untrue. So I still don't have any feel for how, or whether, the proof can be given a Shelah version.

- o  [Randall](#) Says:  
[March 18, 2009 at 12:13 am](#)

1052.1

This may make absolutely no sense, but what if instead of looking at corners in  $\mathbb{Z}^3$  you look at four terms progressions in  $\mathbb{Z}$ ...only you can still look at your dense plane in 3 space, only now your dense set is a set of a special type. Namely, what we really seek is a four term progression in a dense set  $E$  of  $\mathbb{Z}$ , but we actually look at the set  $E' = \{(x, y, z) \in \mathbb{Z}^3 : x + 2y + 3z \in E\}$ , which is dense in  $\mathbb{Z}^3$ . It seems that even though you are only forbidding a sparse set with your dense plane, that sparse set may project back to a dense set in  $\mathbb{Z}$  (though I

have no idea why that would be, if so, so this is mere speculation).

4.  [Randall](#) Says:

[March 18, 2009 at 12:09 am](#) | [Reply](#)

1052. Szemerédi's proof of Roth's theorem

I just looked for the first time at the page about Szemerédi's combinatorial proof of Roth's theorem on the wiki. It seems to me that this proof actual resembles the polymath proof of DHJ3 as much as does the Ajtai/Szemerédi proof, if not more. After all it uses this "cube lemma" which seems to be more comparable to DHJ(2) than is Szemerédi's theorem. Indeed, isn't my post 1002, which was intended to be a dumbed-down version of our DHJ3 proof that would be just strong enough to tackle corners, essentially the same proof? (Assuming of course that 1002 is even correct....)

o  [gowers](#) Says:

[March 18, 2009 at 12:29 am](#)

I had a similar thought at some stage (though less precise): the fact that he uses combinatorial cubes is certainly suggestive. No time just now to think about your question about 1002, but we should certainly try to clarify this.

5.  [Ryan O'Donnell](#) Says:

[March 18, 2009 at 1:41 am](#) | [Reply](#)

1053. Simplifying the DHJ(3) proof.

I'll be back on my normal schedule soon, but let me just say I started thinking about the part of the proof wherein it is shown that if  $A$  has a density increment on a 12-set, then we can find a subspace (with  $\omega_n(1)$  coordinates) on which it also has a density increment.

Let's simplify by imagining that  $A$  has a density increment on a 1-set. By the argument already sketched (based on multidim Sperner), one can show that a 1-set contains not just a combinatorial line but a combinatorial subspace of dimension  $\omega_n(1)$ .

I think then one can use the new equal-slices perspective to easily show the Varnavides statement, that a 1-set contains a randomly chosen  $\omega_n(1)$ -dimensional subspace with constant probability. Perhaps this can substitute for the argument which almost partitions 1-sets into combinatorial subspaces...

o  [gowers](#) Says:

[March 18, 2009 at 8:10 am](#)

1053.1 I think that doesn't work, for two reasons. First, I don't see why a 1-set containing lots of combinatorial subspaces gives you a density increase on one of those subspaces. Why couldn't they all be disjoint from  $A$ ?

More seriously, and this was an important factor in guiding me to the partitioning-into-subspaces perspective, one can use the Hahn-Banach theorem to prove that IF it follows from the fact that  $A$  correlates with  $X$  that  $A$  correlates with a set of type  $T$ , then one MUST be able to partition  $X$  into sets of type  $T$ . (I've oversimplified here, and not explained where partitioning almost the whole set comes in — perhaps I'll wikify this argument some time soon.)

o  [gowers](#) Says:

[March 18, 2009 at 8:52 am](#)



1053.2 I didn't quite say what I meant there. The Hahn-Banach theorem tells you that you're basically forced to find at the very least some kind of fractional partition — that is, a covering by subspaces where almost every element is covered almost the same number of times. But I don't see a probabilistic way of proving that, because different elements can be in very different numbers of subspaces. (I gave up after failing to find a direct probabilistic proof that a Cartesian product in  $[n]^2$  can be evenly covered by grids.)



o [Gil](#) Says:

[March 18, 2009 at 8:29 pm](#)

Is this Hahn-Banach argument essentially LP duality? A general statement will be useful and may ring various combinatorial bells...



o [Ryan O'Donnell](#) Says:

[March 20, 2009 at 2:59 am](#)

Gil, I think think so, except that when Tim says "Hahn-Banach" here I'm not sure he actually means "using the Hahn-Banach theorem". I interpret it somehow as short for "the argument wherein you say that if one can partition  $X$  into sets of type  $T$ , then it must be that  $A$  being dense in  $X$  means  $A$  is dense in some set of type  $T$ ".

Or perhaps he does mean, "use the Hahn-Banach Theorem" (i.e., LP duality), I don't know.



o [gowers](#) Says:

[March 20, 2009 at 9:03 am](#)

My fault for writing in too much haste — sorry, I've been rather busy with other things recently (and am just about to go abroad for a few days, so I'll be going even quieter). I was actually referring to the Hahn-Banach separation theorem, which says that if a point doesn't belong to a convex body then you can find a linear functional that's large on the point and small everywhere on the convex body. (I'm oversimplifying slightly, but not too much.) In this case, if the characteristic function  $f$  of  $X$  can't be approximated by a suitable positive combination of characteristic functions of subspaces, then you have a convex set to which  $f$  does not belong. The functional given by the Hahn-Banach theorem can be used to produce a bounded function that correlates with  $X$  but not with any subspace. I'm trying to wikify the details but have kept messing it up. I'm going to work them out on paper and then I'll get back to the wikification.



6. [Ryan O'Donnell](#) Says:

[March 18, 2009 at 3:08 pm](#) | [Reply](#)

1054. Simplifying

Hi Tim, perhaps you're right, but here is an alternative that I think works.

By [known arguments](#) I think we can Varnavides-ise and multidimensional-ise any DHJ-type statement fairly easily, in the equal-slices measure. (The multidimensionalising is in the wiki for uniform measure I think, but I'm pretty sure it's easy to do under equal-slices.) In particular, we can even Varnavides+multidim-ise any such statement. Hence, e.g., starting with Sperner, it's not too hard to get an equal-slices VM-DHJ(2). [ $M$  = multidim,  $V$  = Varnavides.]

That's good, because that's a key component in pushing up from  $k = 2$  to  $k = 3$ . In particular, with VM-DHJ(2), one can get DHJ(3) for 1-sets ("Step 1" in the [overall sketch](#).) Were we calling this DHJ(1,3) or something? Let's say we were.

I checked (I think) that again, we can easily V-ise and M-ise. (One only needs to check that restrictions of 1-sets to

subspaces are still 1-sets.) So we can easily get VM-DHJ(1,3); i.e., a 1-set in  $[3]^n$  with density  $\delta$  has the property that if you pick a random equal-slices  $m$ -dimensional subspace, it's entirely within the set with probability  $c(\delta, m) > 0$ .

Let me now wave my hands and pretend that we can somehow get DHJ(2,3) as well, where I take "DHJ(2,3)" to mean the claim that a dense 12-set in  $[3]^n$  contains a line. I'm not quite sure how to do this currently, but the jump from 1-sets to 12-sets in the current proof outline is easy, so let's pretend it's easy here too. 😊 [More honestly, I think we should try to do it by using the argument that appears below.]

Again, we can V-ise and M-ise to get VM-DHJ(2,3).

Now for the main point of this post: how can we use this to conclude DHJ(3)? If I'm not mistaken (I'm probably mistaken) I think it's actually easy:

As usual, we note that if  $A \subseteq [3]^n$  has (equal-slices) density  $\delta$ , yet fails to contain a line, then it must [completely avoid a certain 12-set](#) we like to call  $U \cap V$ . This set  $U \cap V$  has equal-slices density  $\eta = \eta(\delta)$ .


Let  $S$  denote a randomly chosen (equal-slices)  $m$ -dimensional subspace of  $[3]^n$ . Note that if we draw  $x \in S$  at random (equal-slices), then  $x$  is overall distributed as equal-slices. Hence:

$$\delta = \Pr_{S,x}[x \in A] = \Pr[x \in A \mid S \subseteq U \cap V] \Pr[S \subseteq U \cap V] + \Pr[x \in A \mid S \not\subseteq U \cap V] \Pr[S \not\subseteq U \cap V].$$


Since  $A$  avoids  $U \cap V$  we have  $\Pr[x \in A \mid S \subseteq U \cap V] = 0$ . But also, by VM-DHJ(2,3), we know that  $\Pr[S \not\subseteq U \cap V] \leq 1 - c(\eta, m)$  for some tiny but positive  $c(\eta, m)$ . Hence the above equation implies

$$\delta / (1 - c(\eta, m)) \leq \Pr[x \in A \mid S \not\subseteq U \cap V].$$

Hence there exists a particular subspace  $S_0$  (not in  $U \cap V$ ) such  $\Pr_{x \in S_0}[x \in A] \geq \delta / (1 - c(\eta, m))$ , and we have the desired density increment on a (multidimensional) subspace.


-  Ryan O'Donnell Says:  
[March 18, 2009 at 3:29 pm](#)

A small point: Actually, DHJ(1,3) follows almost trivially just from Sperner (as I'm sure we noted back in the 1-200 thread). So you can get to VM-DHJ(1,3) rather quickly. About DHJ(2,3), I'm still thinking...

-  Ryan O'Donnell Says:  
[March 18, 2009 at 3:51 pm](#)

1054.2. DHJ(3) for 12-sets.

Oh, it's [already on the wiki](#). Hmm. Are we done?

-  Ryan O'Donnell Says:  
[March 18, 2009 at 4:02 pm](#)


1054.3. Okay, so just to recap. The following statements all refer to equal-nondegenerate-slices measure:

(a) DHJ(3) for 12-sets is an easy consequence of Sperner (see 1054.2).


(b) I claim that it is not too hard to Varnavides-ise, Multidimensional-ise, and indeed Varna+Multidim-ise DHJ(3) for 12-sets (using known arguments of Tim and Terry I think that are already on the wiki, but which can be streamlined a bit if you work in equal-nondegenerate-slices)

(c) Then, assuming VM-DHJ(3) for 12-sets, use the argument following “Now for the main point of this post:” in 1054.


If someone is able to agree with (c), I will endeavour to wikify (b) a bit more clearly.

- 
 Ryan O'Donnell Says:  
[March 18, 2009 at 5:04 pm](#)


1054.4. I'm slightly nervous about the dog-chasing-its-own-tail aspect of the parameters here; you know, getting too small an increment on too small a subspace. Let me think about why it's okay in the original proof sketch.

- 
 gowers Says:  
[March 18, 2009 at 5:25 pm](#)


1054.5 I see better what you're saying now. One detail troubles me, which is that the density increase you get on the subspace depends on the dimension of the subspace, which definitely isn't the case in the other argument, but I haven't yet decided whether that's a serious problem or just an interesting detail.

- 
 gowers Says:  
[March 18, 2009 at 5:39 pm](#)


1054.6 The more I think about it, the more I think that it is a problem. E.g. if on passing to an  $m$ -dimensional subspace you get a density increase of  $2^{-m}$  it seems to me that there's no way of getting the density up to 1.

- 
 Ryan O'Donnell Says:  
[March 18, 2009 at 6:50 pm](#)

Right, it seems to be a problem. But I kind of like the flavour of this angle; I wonder if there's a way to make it work... Probably won't get a chance to think about it today though.

- 
 gowers Says:  
[March 18, 2009 at 6:53 pm](#)

It would be great if you could simplify the argument. I'm busy wikifying an argument that places some limits on any proof that can possibly work — I hope it will narrow down and therefore speed up the search.

- 7. 
 Ryan O'Donnell Says:  
[March 18, 2009 at 7:47 pm](#) | [Reply](#)

1055. Yes, Tim, I'm starting to see what you say, that this “inner increment/partition argument” is necessary.

Without thinking about it too carefully, let me ask the following question: in the [general partitioning argument](#), it seems to me that the “Subspace Richness” property tends to follow in general just from the “Line-Having” property, by the standard Multidimensional-ising and Varnavides-ising (perhaps one needs some of the other properties here; for example, I've only bothered to convince myself for 1-sets and 12-sets).

If so, I guess the overall argument could go (or rather, *is* going):

DHJ(2) (Sperner)  $\Rightarrow$  12-sets in  $[3]^n$  have Lines  $\Rightarrow$  12-sets in  $[3]^n$  are Subspace Rich (VM-ise)  $\Rightarrow$  12-sets in  $[3]^n$  can be partitioned  $\Rightarrow$  DHJ(3) (because Line-free sets correlate with 12-sets)

Now presumably the argument continues:  $\text{DHJ}(3) \Rightarrow$  “lower complexity” (??) sets in  $[4]^n$  have Lines  $\Rightarrow$  lower complexity sets in  $[4]^n$  are Subspace Rich  $\Rightarrow$  lower complexity sets in  $[4]^n$  are partitionable  $\Rightarrow \text{DH}(4) \dots$

Sorry, I guess this is already more or less on the wiki already and I'm just thinking out loud / catching up. Is this right?



o [Randall](#) Says:

[March 19, 2009 at 5:18 pm](#)

I confess I have trouble following these outlines (partly because I just think rather slowly) but I think both proofs of the part in question that we have, Tim's via almost-partitioning a 1-set into subspaces and Terry's via dense fibers (really, these are essentially the same proof) proceed via...well...partitioning a 1-set into subspaces. (This is of course a tautology in the case of Tim's argument, but it's true of Terry's as well because the greedy algorithm he uses may in general throw away almost the entire 1-set in question before it locates a suitable set of “maximal density” or drives the density on the remaining part to 1.) More generally, I tend to think Tim is right than any argument that is going to work at all has got to be powerful enough to partition a 1-set into subspaces, so you might as well do exactly that because it's much easier to follow the proof in that case anyway. (I at least find Tim's argument easier to follow than Terry's, despite the fact that Terry's is closer in spirit to the ergodic heuristic...and also despite the fact that they are essentially the same argument, of course.)



8. [Ryan O'Donnell](#) Says:

[March 19, 2009 at 11:25 pm](#) | [Reply](#)

1056. Wikifying.

Just wanted to say I've started trying to build a [wiki page](#) devoted to giving the all-details-filled-in version of the density-increment proof of  $\text{DHJ}(3)$ . I'll report back when I've fleshed it out some.



o [Ryan O'Donnell](#) Says:

[March 19, 2009 at 11:26 pm](#)

Reader, you too should feel free to flesh it out, of course!



o [gowers](#) Says:

[March 20, 2009 at 5:20 pm](#)

Ryan, one aspect of the outline you've got on the wiki is different, though perhaps not truly different, from how it is at the moment. You deduce that an ij-insensitive set is subspace rich from the fact that it has subspaces, whereas what we have at the moment deduces it from the fact that a dense subset of  $[2]^n$  is subspace rich. In both cases the proof is that we can express the measure as a positive linear combination (plus an error if necessary, but I think you've managed to avoid that) of suitable measures on subspaces. The way you do it, you need to add the extra remark that the restriction of an ij-insensitive set to a subspace is still an ij-insensitive set.

I haven't checked, but I think it may be that there is a sort of commutative diagram that allows one to see that the two proofs are exactly the same, in the sense that there's a one-to-one correspondence between the sets they consider. If that is the case, then I find it slightly more natural to prove the subspace richness in  $[2]^n$  and then lift it up to ij-insensitive sets in  $[3]^n$ .



o [Ryan O'Donnell](#) Says:

[March 20, 2009 at 11:03 pm](#)

1056.3. Tim: I agree with everything you wrote 100%, and in fact at some point I independently decided that it

made more sense to commute those two steps: i.e., first show that dense sets in  $[2]^n$  are subspace rich, and then pull up to ij-insensitive sets in  $[3]^n$ .

Besides being arguably more natural, doing it this way has a small benefit: If we really really care, we can use the Gunderson-Rodl-Sidorenko argument to get that dense sets in  $[2]^n$  have large subspaces, and this paper has way better quantitative parameters than the lines->subspaces argument we use in general.

(On this point, I should note that the alternative-to-GRS lines->subspaces deduction I put on the wiki a while ago was extremely wrong, so I deleted it. At some point we should see if we can fix it; if GRS can get civilised parameters, hopefully we can also get civilised parameters for lines-in- $[3]^n$  -> subspaces-in- $[3]^n$ .)



9. *Ryan O'Donnell* Says:

[March 20, 2009 at 11:05 pm](#) | [Reply](#)

1057. Wikifying.

Just wanted to report that I've gotten about 50% of the way through making the aforementioned [stub article](#) for the fleshed-out proof. It's currently an orphan page on the wiki; when/if I get to 90%+ I'll put a link to it on the main page.



10. *Gil* Says:

[March 24, 2009 at 10:22 am](#) | [Reply](#)

While following, and cheering the emerging proof on the wiki, and resting a little from the serious efforts that took place here over the last 6 weeks or so (and I hope there will be a follow up math discussion (not just meta discussion) for a couple of weeks even after the success,) can I offer to amuse ourselves with the silliest spin-off problem that was offered in these discussions: (naturally, by me): What is the size of the largest family of subsets of  $\{1, 2, \dots, n\}$  without two sets  $A$  and  $B$  so that  $A \setminus B$  is twice as large as  $B \setminus A$ . Any slice will do and we can (as pointed out by NK) add to the middle slice also silices of size  $n/4$ ,  $n/8$  etc. Can we do better?



11. *Kristal Cantwell* Says:

[March 24, 2009 at 5:58 pm](#) | [Reply](#)

1058. Hyper-optimistic conjecture

Could the Hyper-optimistic conjecture be extended beyond 3, and if so if it is true would it imply DHJ(n) for higher values of n. For example if it is true for 4 would we have DHJ(4)?



12. *Gil* Says:

[March 24, 2009 at 11:05 pm](#) | [Reply](#)

1059. hyper-optimism

Great question. I looked at the wiki and did not find it stated beyond 3. I also do not see off hand what it will give.



13. *Kristal Cantwell* Says:

[March 25, 2009 at 8:00 pm](#) | [Reply](#)

1060. Corners theorem

Could the corner's theorem be extended to higher dimensions? Has it been already? I think that such a result might be useful in trying to get DHJ(n) from the hyper-optimistic conjecture for values of n greater than three.

Even without the corner's lemma if we could get the grid  $a+b+c+d=n$  of quadruples  $(a,b,c,d)$  with density  $\delta$  contains  $(a+r,b,c,d), (a,b+r,c,d), (a,b,c+r,d), (a,b,c,d+r)$  that and its extension into  $n$  might be useful in a possible proof.



o [gowers](#) Says:

[March 27, 2009 at 7:27 pm](#)

The corners theorem is known in all dimensions, either by ergodic theory or as a consequence of hypergraph regularity. (It also follows from DHJ but obviously that's no help if you want to use it to prove DHJ.) The hyperoptimistic conjecture can be formulated for any  $k$ , and I do not know of any counterexamples. The result you talk about with the grid is equivalent to the 3D corners theorem, if I interpret correctly what you are saying.

I recently tried to come up with a lower bound for DHJ(3) that would beat the Behrend bound applied to slices, but had absolutely no luck. But I'd be amazed if the hyperoptimistic conjecture turned out to be true ...



14. [gowers](#) Says:

[March 27, 2009 at 7:29 pm](#) | [Reply](#)

Metacomment.

What are people's views about writing up? Ryan wanted to wait, and I know others feel that there isn't a big hurry. I myself would very much like to see a complete write-up, and would like to be closely involved in it, but I don't mind waiting a bit more because I'm pretty busy over the next couple of weeks. I thought I'd ask though. For example, if I were to produce a first attempt at a plan of a paper, would that be welcomed, or would the other potential writers rather I didn't?

o [Ryan O'Donnell](#) Says:

[March 28, 2009 at 12:20 am](#)

Hi Tim. Yes, I did suggest waiting a bit, but it seems like we all collectively have done so, perhaps unintentionally. I too have sorely needed to catch up on some other projects, which won't be wrapped up till April 2. I still hope to finish the wiki outline I started.

In any case, I would be fine with the idea of writing beginning in a week or two or three, whenever we get freer. I can contribute some writing, or Tim if you want to, please do so. I'd be happy with a model wherein anybody who felt they had time to write could just notify the others that they're doing so. Tim, if you or someone manages to put out a full draft at some point, why not?



o [jozsef](#) Says:

[March 30, 2009 at 3:43 pm](#)


It would be great if Tim could write up the proof. I'm trying to write a toy version of the proof for corners on the Cartesian product  $A \times A$ , where  $A$  is a Hilbert cube. A few weeks ago my first try didn't go well as I followed the original Ajtai-Szemerédi proof. I knew it quite well and didn't pay attention to Tim's modified proof. Now I see that actually the key of the DHJ3 proof is hiding there. So, I advise the interested reader to compare Tim's version to the original, even if one knows the Ajtai-Szemerédi proof very well. The two wiki entries are:

"Ajtai-Szemerédi's proof of the corners theorem"

[http://michaelnielsen.org/polymath1/index.php?title=Ajtai-Szemer%C3%A9di%27s\\_proof\\_of\\_the\\_corners\\_theorem](http://michaelnielsen.org/polymath1/index.php?title=Ajtai-Szemer%C3%A9di%27s_proof_of_the_corners_theorem)


and "A Modification of the Ajtai-Szemerédi argument"

[http://michaelnielsen.org/polymath1/index.php?title=Modification\\_of\\_the\\_Ajtai-Szemer%C3%A9di\\_argument](http://michaelnielsen.org/polymath1/index.php?title=Modification_of_the_Ajtai-Szemer%C3%A9di_argument)

- o  [Terence Tao](#) Says:  
[April 3, 2009 at 2:07 am](#)

I plan to revisit the writeups I have of the triangle removal proof and the Furstenberg-Katznelson proof, and clean them up - right now I get the sense that they're not really readable by anyone other than myself. But it does seem like the optimal strategy is to move slowly, in case some simplifications crop up that can be taken advantage of.

For instance, I found it very useful in those two proofs to play around with substitution maps such as  $\rho_{1 \rightarrow 2}^I : [3]^n \rightarrow [3]^n$  (defined as the map that takes a string  $x$  in  $[3]^n$  and changes all the 1s in positions in  $I$  to 2s) in order to build lines (e.g.  $(x, \rho_{1 \rightarrow 2}^I(x), \rho_{1 \rightarrow 3}^I(x))$  is a line whenever  $x$  has at least one 1 in  $I$ ). This might help with some of the arguments in the density-increment argument, though I haven't checked it.

15.  [Kristal Cantwell](#) Says:  
[March 29, 2009 at 12:38 am](#) | [Reply](#)


#### 1061. Hyper-optimistic conjecture

I think I can get the hyper-optimistic conjecture implies DHJ(k).

I have to assume I have DHJ(k) for density is equivalent for DHJ(k)


for equal slices density as in  $k=3$ . Then the corners theorem makes the

density of slices a without corners arbitrarily low then the hyper-optimistic conjecture transfers the small size to the equal slices density and the logical equivalence makes the density arbitrarily low. So If I start with density epsilon I can find a large enough  $n$  such that the corner free space has density much less than epsilon and that means by the hyper-optimistic conjecture the equal slices density is much less than epsilon and hence by logical equivalence we have the regular density must be low and we are done.

16.  [Gil Kalai](#) Says:  
[March 30, 2009 at 12:41 pm](#) | [Reply](#)

#### 1062. Finer slice decompositions.

There are "finer decompositions" (compared to slices) of  $\{0, 1, 2\}^n$  that can also plays a role. For example, we can consider vectors with a '0', b '1' so that the sum of indices of coordinates with value '0' is  $x$ , and the sum of indices of coordinates with value 1 is  $y$ . If we denote these vectors by  $\Gamma_{a,b;x,y}$  we get a finer slice decomposition which can be of interest.

- o  [Gil](#) Says:  
[March 30, 2009 at 8:16 pm](#)

hmm, when we consider union of such finer slices  $\Gamma_{a,b,c;x,y,z}$  (or perhaps a different forms of slicing) there are several issues that seems relevant. (But they do not suggest that union of such fine slices will give better constructions.)

One thing is this: Consider a subset of  $\{1, 2, \dots, m\} \times \{1, 2, \dots, k\}$  (where  $km=n$ ) with out 3 terms A P. Can you get examples with higher densities than for  $\{1, 2, \dots, n\}$ ? (lets assume  $k=m$  for simplicity) (Probably this is known to experts.)

Another thing is that there are various extensions of Sperner lemma which show that under more general conditions you get the same bound. One is the 2-part Sperner lemma of Katona. Another is Kleitman's solution of Littlewood-Offord problem. So if we think its fruitful to regard DHJ(3) as extension of DHJ(2). Is there an



analog for Kleitman's theorem?

17.  *Kristal Cantwell* Says:

[March 31, 2009 at 6:22 am](#) | [Reply](#)

1063. Hyper-optimistic conjecture

Apparently this is not true for the generalization for 4 see post 1185 in the thread

<http://terrytao.wordpress.com/2009/03/14/dhj3-1100-1199-density-hales-jewett-type-numbers/>

Congratulations to Klas Markström for finding the counterexample for this.

18.  *Gil Kalai* Says:

[March 31, 2009 at 10:11 am](#) | [Reply](#)

1064. more on the analogy with Sperner

We want to find a set avoiding a configuration  $C$  inside a large set  $X$ . If we can show that for some subset  $Y$  of  $X$  the maximum density of a  $C$ -avoiding subset is at most  $c|Y|$  then usually this is good. When there is a group acting like in the cap set problem we can show the same bound for subsets of  $X$  that avoid  $C$ . But even if there is no group action this can still be useful.

For Sperner the basic observation is that if we take  $Y$  to be a chain then a line-free subset of  $Y$  contains at most 1 element. So the density of line-free subsets of chain is very small and this can be pushed to various proofs.

So the question is (and we discussed it more or less in these threads, i just repeat it): is there a good analog for “chains” for DHJ(3). A small configuration (and I do not know if small is in the order of 1 or of  $2^n$  up to  $\text{poly}(n)$ ) on which we can show (or at least conjecture) that the density of line-avoiding subset must be small. (Thinking about combinatorial subspaces is sort of “cheating” and in any case not the analog for the case of Sperner.)

Simple questions to state are: Is there any subset  $Y$  of  $\{0,1,2\}^n$  on which we can show (more easily) that the maximum density of line-free subsets is small. Is there a subset  $Y$  where this maximum density is expected to be smaller than that for the entire  $\{0,1,2\}^n$ ?

19.  *Terence Tao* Says:

[April 3, 2009 at 2:02 am](#) | [Reply](#)

Metacomment.

I gave a talk at UCLA on the three combinatorial proofs (or proof sketches) we have, namely the density increment-based proof, the triangle-removal proof (based on Austin's ergodic proof), and the finitary Furstenberg-Katznelson proof, with some notes at

<http://terrytao.wordpress.com/2009/04/02/polymath1-and-three-new-proofs-of-the-density-hales-jewett-theorem/>

20.  *gowers* Says:

[April 4, 2009 at 5:29 pm](#) | [Reply](#)

1065. Writing up.

Here's a possible structure for a write-up of the density-increment argument. I'm planning to make a start on this, but

I'll try to make it as modular as possible so that if people want to then it may be possible, without a major rewrite of the whole thing, (i) to make local changes and (ii) to switch the order about.

```
\author{Polymath}
```

```
\address{The world} [Not, of course, intended to be taken seriously.]
```

```
\begin{document}
```

```
\title{A new proof of the density Hales-Jewett theorem}
```

```
\maketitle
```

```
\section{Introduction}
```

Basic definitions and statement of the theorem.

History of and motivation for the problem.

Discussion of related results, including the corners theorem.

```
\section{A sketch proof of the corners theorem}
```

A fairly detailed sketch of the modified Ajtai-Szemerédi argument.

```
\section{Different measures on  $[k]^n$ .}
```

Definition of equal-slices measure and the Polya urn measure.

Motivation for considering these other measures.

Proofs that dense sets in any of those two measures or the uniform measure can be restricted to subspaces where they retain their density in any other of the measures.

```
\section{Three important lemmas}
```

```
\subsection{The multidimensional Sperner theorem.}
```

```
\subsection{A dense line-free set correlates locally with a 12-set.}
```

```
\subsection{A 23-insensitive set can be almost entirely partitioned into fairly large combinatorial subspaces.}
```

```
\section{A proof of the theorem for  $k=3$ .}
```

```
\section{A proof of the general theorem}
```

```
\subsection{Uniform DHJ( $k-1$ ) implies that an equal-slices-dense subset of  $[k-1]^n$  contains many combinatorial subspaces.}
```

```
\subsection{Putting everything together.}
```

```
\end{document}
```

21.  József Says:

[April 5, 2009 at 6:55 am](#) | [Reply](#)

1066. Dear Tim, I still don't have enough time to work out the details of the proof. Next week is the last week of classes here, so I will make some progress after that, I hope. Until then let me ask a question about the second outline of a density-increment argument. I don't understand the argument after Substep 2.2. In each step you keep a  $3^m$  "portion" of  $Z$  and partition the remaining  $B$  into  $\sim 3^M$  classes. Elements which partitioned into different classes will never be considered together in a subspace again. What I can't see is that why will be the intersection of the 1-set and the 2 set large in a subspace? Isn't it possible that only a small subset of the 1-set is element of a subspace  $S_i$ ? Most likely I overlooked something. I will check it tomorrow again.

22.  *jozsef* Says:

[April 5, 2009 at 5:00 pm](#) | [Reply](#)

1066. (contd.) I still can't see how to finish the sequence of iterations. The question is that what happens when Substep 2.3 ends. Given this rapidly growing family of partitions. If a partition is sparse then we don't do anything with it, if dense AND large enough, then we find a subspace in it. What happens at the end when we have many small and dense partition classes? I'm stucked. I'll get my morning coffee now...

o  *gowers* Says:

[April 5, 2009 at 5:57 pm](#)

I don't know if this fully answers your question, but when we have a dense set, we don't just find a subspace in it — we find a whole family of subspaces whose union is dense. Therefore, the number of steps in the iteration is bounded. But I'm answering this question without looking at the proof — I'll give another answer in a moment.

o  *gowers* Says:

[April 5, 2009 at 6:02 pm](#)

I've had another look, and I still don't quite understand what your difficulty is, so I'll have to hope that my first reply answers it, or at least prompts you to explain what it is in terms that I understand better.

o  *jozsef* Says:

[April 5, 2009 at 6:28 pm](#)

During the iterations if a  $B_i$  is sparse, then you do nothing with it, if it is dense then you choose  $Z_1, \dots, Z_m$ . The union of them,  $Z$ , has size at most  $M$ . It works fine until  $B_i$  has size at least  $M$ . What do you do with smaller  $B_i$ -s? I don't see that this error term - the number of elements remaining in small  $B_i - s$  is small.

o  *jozsef* Says:

[April 5, 2009 at 6:42 pm](#)

Maybe I didn't follow the right algorithm; Do we have a good bound on the number of iterations? I think that now I see; For any element, there is a very high probability that after a few iterations this element is in a selected subspace. Right?

o  *gowers* Says:

[April 5, 2009 at 6:48 pm](#)

I still don't understand, but largely for "local" reasons. For instance,  $B_i$  is a subset of a subspace  $S_i$ . Since the number of iterations is bounded, we can ensure that the dimension of  $S_i$  is much larger than  $M$  all the time. When you talk about the cardinality of  $B_i$ , do you actually mean the dimension of  $S_i$ , and when you say "at least" do you actually mean "less than"?

Also, it's not clear to me whether your problem arises when we almost partition a 1-set, or only when we look at the intersection of a 1-set with a 2-set. Assuming it's the first, then here is the rough idea of the argument.

We begin by choosing  $M$  coordinates and partitioning  $[3]^n$  into  $3^M$  subspaces according to those coordinates. We then choose  $Z_1, \dots, Z_m$  inside the  $M$  chosen coordinates such that for many  $x$  the subspace  $x + \langle Z_1, \dots, Z_m \rangle$  is a subset of  $\mathcal{B}$ . Then we remove all subspaces of the form  $x + \langle Z_1, \dots, Z_m \rangle$  from  $\mathcal{B}$ . This removes a positive fraction (depending on  $m$  and the error  $\eta$  you want to end up with) of the elements of  $\mathcal{B}$ . Also, one can check that after removing these elements, the restriction of the remaining set to any of the  $3^M$  subspaces  $S_i$  is still a 1-set. Therefore, we can iterate, except that we do nothing if we have a set of density less than  $\eta/2$  in a given subspace. Let us call such sets *discarded*. The total measure of the discarded sets cannot be more than  $\eta/2$ , since they are contained in disjoint subspaces inside which they have relative density at most  $\eta/2$ .

After a bounded number  $R$  of iterations (depending on  $m$  and  $\eta$ ) we have either discarded or put into subspaces all but at most  $\eta/2$  of  $\mathcal{B}$ , and there is always space to do this provided that  $RM$  is substantially less than  $n$ . So provided  $n$  is large enough, the argument works.

I've written that in the hope that you either agree with it or can point to the bit where I seem to be going wrong. (Even if it is wrong, I have theoretical reasons for believing that some argument like this has to work, at least if Terry's argument about correlations with 1-sets implying correlations with subspaces was correct, which it seemed to be.)

◦  *jozsef* Says:

[April 5, 2009 at 7:04 pm](#)

Tim, first of all I agree that the original argument works even for an incomplete partitioning. I don't think that the argument is incomplete, I'm just slow following it. In this particular case, I think that I didn't see the good bound on the number of iterations, I thought that the fast growing number of partition classes might be a problem.

23.  *jozsef* Says:

[April 5, 2009 at 7:35 pm](#) | [Reply](#)

So, what you proved is a strong partition in the following sense; For every  $c > 0, \epsilon > 0$  and  $m$ , there is a number  $K$ , that any  $c$ -dense 1-set has a large,  $1 - \epsilon$ -dense subset which is the disjoint union of (several) translates of  $K$   $m$ -dimensional subspaces.

24. *Ryan O'Donnell* Says:

[April 6, 2009 at 10:23 pm](#) | [Reply](#)

1068. Hi Tim, re the outline in 1065:

Looks good. One question: It looks like you've divided the proof into three main lemmas: multidim-Sperner (more generally, multidim-DHJ( $k-1$ )), line-free set correlating with intersections of  $ij$ -insensitive sets, and  $ij$ -insensitive sets being partitionable.

It seems to me that the Varnavides-version of multidim-Sperner (more generally, multidim-DHJ( $k-1$ )) may as well be considered the basic lemma. Where will this go?

Putting it into `\subsection{The multidimensional Sperner theorem}` makes sense to me, although then the actual `\section{A proof of the theorem for  $k=3$ }` might be quite short. On the other hand, if it goes into the proof section itself, then the multidim-Sperner theorem subsection will be awfully short (might as well just quote Gunderson-Rodl-Sidorenko).

The latter seems less modular to me, so I guess what I'm ultimately suggesting is that `\subsection{The`

multidimensional Sperner theorem} be more like \subsection{The Varnavides multidimensional Sperner theorem}.

Except that I strongly vote for using a more generic descriptor than “Varnavides”. There’s got to be a catchy word that indicates to the reader that not only do dense sets contain lines, a *random* line is in there with positive probability.

—

Also, I’m still of two minds as to whether “Equal-Slices” should be treated as the main distribution, with Polya as a slight variant, or vice versa.

25. *Ryan O'Donnell* Says:

[April 6, 2009 at 10:34 pm](#) | [Reply](#)

1069. Notation and terminology.

Pre-writing, it might be helpful to fix our notation and terminology. I know it’s not crucial, but I think getting just the write names and letters can really help a reader. We may also try to keep things switchable via appropriate use of macros and the xspace package in LaTeX (see the first two comments to the post here: <http://terrytao.wordpress.com/2007/06/08/advice-on-mathematical-careers-and-mathematical-writing/>), although I must confess I sometimes find it tough to rigorously stick to using macros.

Anyway, here are some questions we might debate:

- . are sets called  $A$  or  $\mathcal{A}$ ?
- . what should “Equal-Slices” measure actually be called (e.g., do probabilists already call this something else?)
- . what “Polya Urn” measure should actually be called? (same question)
- . what letter to use for each of them? what letter to use for the uniform distribution?
- . how to denote drawing a random *line* or a random *subspace* from them?
- .  $[k] = \{0, 1, \dots, k-1\}$  or  $\{1, 2, \dots, k\}$ ?
- . unify terminology of  $ij$ -set, (special) set of complexity  $t$ ,  $ij$ -insensitive set, etc.
- . questions from the previous post: what short words/phrases can we use to indicate that not only do dense subsets of  $[3]^n$  contain lines, they actually contain large-dimensional subspaces, and in fact a random large-dimensional subspace is in with positive probability?



26. *Terence Tao* Says:

[April 9, 2009 at 9:09 pm](#) | [Reply](#)

Hi,

I transferred Tim’s outline (and Ryan’s questions) to the wiki at

[http://michaelnielsen.org/polymath1/index.php?title=Outline\\_of\\_first\\_paper](http://michaelnielsen.org/polymath1/index.php?title=Outline_of_first_paper)

I figure that the wiki format might be better for the topmost level of organising the paper, and then we can switch to LaTeX once all the top-level stuff is finalised.

Some thoughts on notation:

I guess, on balance, that  $[k] = \{1, \dots, k\}$  looks slightly nicer than  $[k] = \{0, \dots, k-1\}$ ; the 0-based notation is slightly more “logical”, but we don’t seem to derive any substantial benefit from it. So I’m weakly in favour of  $\{1, \dots, k\}$ .

We can borrow from geometry and use the notation  $\text{Gr}([k]^n, d)$  to denote the  $d$ -dimensional subspaces of  $[k]^n$  (a “combinatorial Grassmanian”). I don’t know what to call the measures on this space though. Does every measure  $\mu$  on  $[k]^n$  canonically define a measure on  $\text{Gr}([k]^n, d)$ ? It seems to me that one needs some additional parameters to specify

such a measure.

“A” versus “ $\mathcal{A}$ ” - I would prefer A, as I want to think of a subset of  $[k]^n$  as a set of points, rather than a collection of sets (which is what the  $\mathcal{A}$  notation suggests to me). The one problem with using A is that we are also likely to be using A for subsets of  $[n]^2$  in the corners theorem, and if we are going to discuss the reduction of corners to DHJ then there might be a very slight notational clash there. But perhaps this is actually a good thing, since we want to use the corners argument to motivate the DHJ one...

As for the last question, what about “Hales-Jewett property” for containing lines, “subspace Hales-Jewett property” for containing subspaces, and “subspace Hales-Jewett-Varnavides property” for containing subspaces with positive probability? (thus, e.g. “dense ij-insensitive sets obey the subspace Hales-Jewett-Varnavides property”)?



27. Ryan O'Donnell Says:

[April 14, 2009 at 2:26 pm](#) | [Reply](#)

I have collected up a small amount of free time and was planning to put some time into the writing-up. I agree that it will be nice to fix the overall outline before starting. Here are some more comments:

. I propose continuing this discussion not here but on the [discussion page](#) of the wiki page Terry made for the [paper outline](#). I will copy this post to that page too.

. Since I will likely never finish it, I put a link to [my old outline of the proof](#) on the wiki's home page. I kind of like that outline too 😊 although of course it's quite similar to Tim's.

. Changed “author” from “Polymath” to “A. Polymath” (as Terry had suggested a while ago). Reasons: a) Automatic indexing systems probably like authors to have both a “first name” and a “last name”. This will help the author list to get sorted under “P”. b) The next projects could be authored by “B. Polymath”, “C. Polymath”, “ZZ. Polymath”, etc. This would also help get all projects sorted under “P”, yet differentiate the author groups. c) Subtle allusion to A. Nilli. 😊

. Address: perhaps could change to the URL of the wiki. [Perhaps we should get a stabler one?] I suppose we'll need a corresponding address for the paper-journal version though...

. I agree with Terry's votes for  $[k] = \{1, \dots, k\}$  and  $\mathcal{A}$  rather than  $\mathcal{A}$ .

. I'll admit it, I always find the terminology Grassmannian “scary”. If it were me I would just try to refer to “d-dimensional subspaces”. Not sure we need a name for the set 😊

. Naming the distributions: I think the phrase “equal slices” is unbelievably catchy, and [this](#) is cute. On the other hand, it's hard to deny the officialness of “random [ordered 3-partition](#)”. I mean, such terms are in Chapter 1 of Stanley 😊 Either becomes a bit of a mouthful when you get to its sister distribution: “equal nondegenerate slices” or “ordered weak 3-partition” (“weak” = “parts of size 0 allowed”, apparently). So I don't know...

. For the letter, who knows? I'm kind of against  $\mu$  as it sounds like the uniform distribution to me.  $\nu$  is a nice letter for measures. Or maybe  $\lambda$  (for “line”) or  $\pi$  (for “partition”)...

. For the last question, I know Varnavides's name gets attached to this notion, but I must say it is the opposite of evocative for a nonexpert like me. It actually took me a good few hundred blog posts before I understood that “Varnavides = contains random lines w. positive prob.” Guided by Tim's spirit of trying to make the paper as easy-to-read as possible, how about...

“DHJ(3) = Nonnegligible sets in  $[3]^n$  contain a line.” Then:

“Nonnegligible sets in  $[3]^n$  contain a high-dimensional subspace.”

“Nonnegligible sets in  $[3]^n$  contain a nonnegligible *fraction* of lines.”

“Nonnegligible sets in  $[3]^n$  contain a nonnegligible fraction of high-dimensional subspaces.”

Wordy, perhaps, but I think it gets the idea across straight away. Also, using the same word “nonnegligible” subtly gets across the key fact that we’re actually talking about the *same* distribution..

Or does this just sound too weird to an arithmetic combinatorialist’s ears?

28.  [gowers](#) Says:

[April 14, 2009 at 2:46 pm](#) | [Reply](#)

Ryan, I agree with almost all of what you say. My main quibble would be that I prefer “sets of positive density” to “nonnegligible sets”. Of course, that really means “density bounded away from 0 as  $n$  tends to infinity” but that kind of shorthand is very standard. So I’d end up with something more like

“Sets of positive density in  $[3]^n$  contain a positive proportion of all lines/subspaces.”

I’ve never liked Varnavides getting his name attached to a very standard averaging argument.

My vote would be for “equal-slices” with a comment about how it relates to known terminology.

I’m slightly more in favour of  $\mathcal{A}$  than you and Terry, but I see arguments on both sides and so am perfectly willing to be outvoted on that one.

29.  [Terence Tao](#) Says:

[April 14, 2009 at 6:46 pm](#) | [Reply](#)

I modified the outline slightly to reflect the preceding discussion. I also added a sample abstract, but am not seriously committed to it, feel free to modify.

I also added some queries, for instance: should we have a section explaining what Polymath is, the history of the project, and/or the names of contributors?

I also added a discussion section where we can talk about quantitative bounds, relationships to other proofs of DHJ, and other sundry topics.

30.  [Ryan O'Donnell](#) Says:

[April 14, 2009 at 9:49 pm](#) | [Reply](#)

I committed to giving a talk about this at Microsoft Research in 3 weeks, so I am thinking now more about the structure. A few thoughts that occurred to me after this morning’s work:

. It wasn’t too productive to worry much about the name of the distribution, since I invariably referred to it by its letter anyway. I was writing  $\nu_k^n$  for the distribution on  $[k]^n$ .

. It was more convenient to think of the equal nondegenerate-slices distribution as  $\nu$  and of equal-slices as the “variant”. This was because the nondegenerate version seems to show up more in the *statements* of results, whereas the pure equal-slices seems to show up more in the grungy technical details of measure arguments.

. It wasn’t too important to name the set of  $d$ -dimensional subspaces either, because the point of the distribution  $\nu$  is that lines are identified with strings in  $[4]^n$ , and more generally,  $d$ -dimensional subspaces of  $[k]^n$  are identified with strings in  $[k + d]^n$ .

. It may be convenient (I was doing this at one point) to write things like  $\nu_4^n(A) \geq \delta$  to mean, “if



one chooses a random line in  $[3]^n$ , it is entirely contained within  $A$  with probability at least  $\delta^c$ .

. I agree that “dense” is better than “nonnegligible”, after all. E.g., on my roughest sketches of slides, I was writing things like:

“DHJ(3) = if  $A \subseteq [3]^n$  with  $\nu_3(A) \geq \delta$ , then  $\nu_4(A) > 0$ ; i.e., if  $A$  has constant  $\nu_3$ -density then it has positive  $\nu_4$ -density.”

“DHJ(3) implies if  $A \subseteq [3]^n$  with  $\nu_3(A) \geq \delta$ , then  $\nu_4(A) \geq \delta' = \delta'(\delta)$ .” [the Varnavides version]

“DHJ(3) implies if  $A \subseteq [3]^n$  with  $\nu_3(A) \geq \delta$ , then  $\nu_{3+d}(A) \geq \delta' = \delta'(\delta, d)$ .” [the Varnavides version for  $d$ -dimensional subspaces, etc.]

To be continued...

31.  Ryan O'Donnell Says:

[April 14, 2009 at 9:58 pm](#) | [Reply](#)

Here is another structural idea that I was using in my slides draft — which might also be worth considering for the paper version:

First, a section on probability distributions and the various technical lemmas.

Second, a section showing how to deduce the contains-a-subspace, contains-a-constant-density-of-lines, and contains-a-constant-density-of-subspaces versions of DHJ from the simple, contains-a-line version.

Third, sections on the argument for ij-insensitive sets, partitioning these into subspaces, density increment, etc.

The idea of this structure is that it kind of isolates the more “essential” ideas involved in pulling up DHJ(3) from DHJ(2) from some of the more generic stuff. I.e., a sophisticated reader of the paper might skim through the “second” section described above and say, “yeah, yeah, okay, I get the idea from the sketches, this all seems like straightforward tricks and things”, and feel like “nothing has happened yet” after the second section.

This I think would be good; this feeling of “nothing has happened yet” is what I’d hope for. Then, when they settled in to read the third (and succeeding?) sections they’d feel like they were properly starting the argument — which is the feeling I’d actually like to convey.

32.  Ryan O'Donnell Says:

[April 15, 2009 at 1:39 am](#) | [Reply](#)

As a tiny point: It occurred to me that sequential ordering of initials — A., B., C., Polymath — will never work; it was hard enough to get comments numbered consistently. How would various groups known not to step on each other’s initials? Heck, for all I know the other, small- $n$  thread may already have taken A. Polymath. Perhaps the initials could be chosen in a problem-centric way, then; I would of course propose “D.H.J. Polymath”.

33.  Ryan O'Donnell Says:

[April 15, 2009 at 1:40 am](#) | [Reply](#)

Also: Terry, the abstract you wrote looks good to me.

34.  [gowers](#) Says:

[April 15, 2009 at 1:43 am](#) | [Reply](#)

Ryan, I definitely like your “nothing has happened yet” approach. At best, we could get the technical stuff down to some easily comprehensible black boxes that would make the later proofs very streamlined, and I think that ideal may be attainable.

35.  *Ryan O'Donnell* Says:

[April 15, 2009 at 4:05 pm](#) | [Reply](#)

I hope so too. By the way, it appears as though the blog is down.

36.  *Ryan O'Donnell* Says:

[April 15, 2009 at 4:06 pm](#) | [Reply](#)

I hope so too. By the way, it appears as though the blog is down. Or more precisely, [Michael Nielsen's whole site is down](#).

37.  *Ryan O'Donnell* Says:

[April 15, 2009 at 8:23 pm](#) | [Reply](#)

Seems like the blog just went back up.

Also, I understand now a small point that Tim clearly understood a while ago; namely, it is not necessary to show an abundance of  $d$ -dimensional subspaces to execute the partitioning step. Rather, that step only uses existence of a  $d$ -dimensional subspace.

Although one may as well prove abundance at some point, for kicks.

38.  *Ryan O'Donnell* Says:

[April 16, 2009 at 2:58 pm](#) | [Reply](#)

I've been thinking about the partitioning argument. Here are a few points:

- . Seems to me the partitioning is pretty much the center part of the argument. I mean, once everything is set up, the part about a line-free set correlating with an intersection of  $ij$ -insensitive sets is, well, no more than 2 or 3 sentences. So partitioning might deserve a section of itself, or even a billing as “the main lemma”.

- . I don't currently see a natural way to do it under something other than the uniform product distribution. That would be a bit of a shame if true, since it would be nice if we didn't have to do \*too\* much back and forth and back and forth and back with the measures. For example, I was hoping we wouldn't have to dust [this](#) argument off, but perhaps we will. I want to think about it.

- . In [Terry's UCLA talk description of it](#), he describes the partitioning step as “quotient down to  $[2]^n$ , then use the greedy algorithm to pick out large-dimensional subspaces”. Picking out large-dim subspaces in  $[2]^n$  is particularly easy if you're willing to appeal to [GRS]. Can the actual argument for the partitioning really be made this easy? (I know Terry was short on space & time and so of course could not write out a full argument.)

- . Does anybody actually know the [GRS] argument cold? (Jozsef?) From some brief glances, it looks like they're also doing a “partitioning” style argument...

o  *jozsef* Says:

[April 16, 2009 at 4:15 pm](#)

Hi Ryan, I think we should partition a 1-set,  $\mathcal{A} \times [2]^n$ , instead of a set system  $\mathcal{A} \subset [2]^n$  like in the [GRS]. Probably that's what you meant, I just want to emphasize that this partition is quite different (and as I see it is much stronger) from [GRS]. I would put [GRS] as a reference there, and would write the probabilistic argument for the almost perfect tiling. If you wish you can leave this part to me, I wanted to write more about it anyways.

39.  *Kristal Cantwell Says:*

[April 17, 2009 at 10:36 pm](#) | [Reply](#)

1070. Outline of a proof of DHJ(3)

I have an outline of a proof of DHJ(3) and an outline of its extension to DHJ(n) I am going to post them here to see if anyone sees any gaps or errors. If it works I think it would be simpler then the current proof.

Outline of proof of DHJ(3)

We get an 12 insensitive set.

We get a space of density epsilon

Actually this could be a much smaller constant function of epsilon inside the 12 insensitive set so it inherits the 12 insensitivity.

We get a space in which the only coordinates

That change are 2 and 3 with density delta much lower than epsilon.

We do this by getting a subspace of the original space which projects down to this space if the preimage that projects down to the space has density epsilon for each point then we are done otherwise.

We remove the point that has the low density get

A slight increase in the density of the remaining spaces

Associated with the preimages in the then take one of the preimages that has the slightly higher density and repeat the process this will result in a density increment if we cannot repeat the process often so it terminates

and we get a set of preimages with the desired density from which we can get the desired space.

Then we use DHJ(2) to get a combinatorial line of length two in the space.

We use the fact the space is 12 insensitive to extend the line to length 3 and we are done.

40.  *Kristal Cantwell Says:*

[April 17, 2009 at 10:39 pm](#) | [Reply](#)

1071 Outline of proof of DHJ(n)

We get an 12 insensitive set.

We get a space of density epsilon

Actually this could be a much smaller constant function of epsilon inside the 12 insensitive set so it inherits the 12 insensitivity.

We get a space in which the only coordinates

That change are two through  $n$  with density  $\delta$  much lower than  $\epsilon$ .

We do this by getting a subspace of the original space which projects down to this space if the preimage that projects down to the space has density  $\epsilon$  for each point then we are done otherwise.

We remove the point that has the low density get

A slight increase in the density of the remaining spaces associated with the preimages in the then take one of the preimages that has the slightly higher density and repeat the process this will result in a density increment if we cannot repeat the process often so it terminates and from the preimages at the terminal step we get the desired space.

We use DHJ( $n-1$ ) to get a Combinatorial line of length  $n-1$  in the space

We use the fact the space is 12 insensitive to extend the line to length  $n$  and we are done.

41.  [gowers](#) Says:

[April 17, 2009 at 11:31 pm](#) | [Reply](#)

1072. I'm afraid I'd need a lot more detail before I even began to understand that. Just to give a few examples, what do you mean by "We get" at the beginning? What does the second line mean?

I'm potentially interested though, because you seem to be going for a Shelah-type approach, which I was attempting to do in some earlier comments. Do you think you could expand on what you have written? Perhaps if it would be on the long side you could do it on the wiki — perhaps just for  $k=3$  to start with.

42.  [Kristal Cantwell](#) Says:

[April 19, 2009 at 2:09 am](#) | [Reply](#)

1073. I can't get the 12 insensitive sets as easily as I thought. The idea won't work without them. I doubt I can fix this.

43.  [Ryan O'Donnell](#) Says:

[April 19, 2009 at 3:50 am](#) | [Reply](#)

1074. Outline.

I've thought for a while and I can't really see how to carry out the partitioning argument under the equal-slices distribution. For every technical difficulty that I worked around, a new one would pop up, and I was never able to simultaneously handle all of them. (The main problem: what "progress measure" to keep track of, akin to the "total uniform-distribution mass".) I still hold the intuition that there should be a slick way to do it, but I'm stuck. Perhaps one could turn the whole argument into a "density-increment on an  $ij$ -insensitive set" argument rather than a "density-increment on a subspace" argument...?

In any case, I therefore reluctantly propose that in the writeup, the uniform distribution be treated as the "main" distribution, with equal-slices as a secondary distribution one needs to slip in and out of to facilitate the argument. I say reluctantly because I feel that the equal-slices distribution on lines is so natural that it may even be that there is some superb (polynomial!?) relationship between equal-slices point-density in  $[3]^n$  and equal-slices line-density in  $[3]^n$ . It would be nice to spend some time thinking about global obstructions.

Anyway, I guess that means the outline of the measure-changing tricks is:

1. Start with a set of uniform density  $\delta$ .
2. Argue that one can pass to a subspace where both the  $[3]^n$ -equal-slices and  $[2]^n$ -equal-slices density are at least  $\delta - o(1)$ . (Or else, one can get a subspace where the  $[3]^n$ -equal-slices density increments — in which case, one can further pass to a subspace where the uniform density increments.)
3. Either find a line, or get an equal-slices density increment inside an intersection of 13-insensitive and 23-insensitive sets.
4. Use [this argument](#) to convert this into a uniform-distribution density increment inside an intersection of 13-insensitive and 23-insensitive sets. (One has to note that that argument preserves the property of being an intersection of 13-insensitive and 23-insensitive sets.)
5. Run the [partitioning argument](#) under the uniform distribution. Note that that argument doesn't really need "subspace richness" per se, because it implicitly proves it anyway. (It's okay, I think, to have  $c$  to include a  $3^{-M}$  factor.) I.e., all it needs is existence of  $d$ -dimensional subspaces for  $[2]^n$ -uniform dense sets.
6. Thus get a uniform density-increment on a subspace.



44. Ryan O'Donnell Says:

[April 22, 2009 at 4:08 am](#) | [Reply](#)

1075. Hmm. Once more my mind changes. It's really essential to be in equal-slices mode when doing the line-free-sets-correlate-with-ij-insensitive stuff (so that a random line is in with nonnegligible probability); and even more is this so in the generalization to  $[k]^n$ .

Further, I think I see a non-horrendous way to carry out the partitioning argument under equal-slices. So, I propose going back to "equal-slices is the main measure".

Re partitioning: the argument is written for the uniform distribution, but it seems like it really only needs a *product* distribution to work. (The key need is the ability to break strings up into pieces and have the global distribution be the product of the two marginal distributions.)

Luckily, equal-slices is a mixture of product distributions. (There'll be a bit of a technicality here probably, where one will want to throw out the product distributions  $(p_1, p_2, p_3)$  which have a component extremely close to 0 or 1.) So if you can partition an  $ij$ -insensitive set into  $d$ -dimensional subspaces up to tiny error under any product distribution, then you get a "fractional cover" of it by (subspace, product distribution) pairs.

So you should be able to get a density increment for  $A$  on some  $d$ -dimensional subspace under some product distribution. Which you can convert back to a density increase under equal-slices by an easy passing-between-measures argument.

seem plausible?



45. Ryan O'Donnell Says:

[April 22, 2009 at 8:55 pm](#) | [Reply](#)

1076. Simplifying the first step in the proof.

Recall that in the first step of the proof, one does a little song and dance to arrange things so that  $A$  has density  $\geq \delta - o(1)$  under  $\nu_3$  (equal-nondegenerate-slices on 3 characters) and simultaneously density at least some  $f(\delta) > 0$  under  $\nu_2$ . (We currently get  $f(\delta) = \delta - o(1)$ , but it would be okay to get, say,  $f(\delta) = \Theta(\delta^{3/2})$  as we'll do here.) It requires some [hacking around with total variation distance between product distributions](#)..

Here is a sketch of an arguably more straightforward method.

In this sketch I will: a) ignore the distinction between equal-slices and equal-nondegenerate-slices when it's convenient (this only changes statements by an additive  $o_n(1)$ ); b) ignore the tiny technicality that when one passes to a subspace  $[3]^T$  where  $T$  is a random set of coordinates, one has to take care of the unlikely event that  $|T|$  is outrageously smaller than expected.

Define a *random restriction*  $(S, \ell)$  as follows: choose  $z \sim \nu_4^n$  and let  $S$  be the coordinates where  $z$  has 4's; also choose  $\ell \in [3]$  randomly. The meaning of "restricting according to  $(S, \ell)$ " is that we think of forming a string in  $[3]^n$  by fixing the  $S$  coordinates to be all  $\ell$ 's and leaving the remaining coordinates  $\bar{S} = [n] \setminus S$  "free". (Note that  $z$  only plays a role insofar as it determines the cardinality and location of  $S$ .) A basic observation about equal-(nondegenerate)-slices is that if we draw a random restriction  $(S, \ell)$  and then draw the free coordinates from  $\nu_3^{\bar{S}}$ , we precisely get the equal-(n.d.)-slices distribution  $\nu_3^n$ .

Hence  $\mathbf{E}_{(S, \ell)}[\nu_3^{\bar{S}}(A_{(S, \ell)})] = \delta$ , and by the usual argument we can in fact ensure that  $\nu_3^{\bar{S}}(A_{(S, \ell)}) \geq \delta - \delta^{10}$ , say, for "almost all" restrictions  $(S, \ell)$ . (If the variance of  $\nu_3^{\bar{S}}(A_{(S, \ell)})$  is tiny then this holds; otherwise, we must have a density increment for some  $(S, \ell)$ .)

Consider now drawing a random restriction  $(S, \ell)$  and filling in the free coordinates according to  $\nu_{\{i, j\}}^{\bar{S}}$ , where  $\{i, j\} = [3] \setminus \{\ell\}$ . This gives some nonstandard distribution  $\lambda$  on  $[3]^n$ .

Claim:  $\lambda(A) \geq \Omega(\delta^{3/2})$ .

Once we show this claim the argument concludes as follows: For an  $\Omega(\delta^{3/2})$  fraction of  $(S, \ell)$  we have  $\nu_{\{i, j\}}(A_{(S, \ell)}) \geq \Omega(\delta^{3/2})$ . Hence there exists some  $(S, \ell)$  for which both  $\nu_3(A_{(S, \ell)}) \geq \delta - \delta^{10}$  and  $\nu_{\{i, j\}}(A_{(S, \ell)}) \geq \Omega(\delta^{3/2})$ , and our mission is accomplished. (Recall that the latter fact, by Sperner, gives us an  $i\ell$ -insensitive  $\cap j\ell$ -insensitive set of measure  $\Omega(\delta^3)$  which  $A$  completely avoids.)


Proof of Claim: Given a small  $\gamma$  we will show that  $\lambda(A) \geq \gamma\delta - 2\gamma^3$ . This completes the proof by taking  $\gamma = \Omega(\delta^{1/2})$ .

To argue this, let  $\beta_k$  denote the distribution on the cardinality of the set of  $k$ 's in a draw from  $\nu_k^n$ . The distribution  $\lambda$  draws  $s \sim \beta_4$ , picks a random set  $S$  of cardinality  $s$ , fills it with a random 1, 2, or 3, then fills the remaining coordinates with equal-slices on the other two characters. If we had instead drawn  $s \sim \beta_3$ , the resulting distribution would have exactly been  $\nu_3$ . Hence if we have  $d\beta_4/d\beta_3 \geq \gamma$ , we could conclude  $d\lambda/d\nu_3^n \geq \gamma$  and hence  $\lambda(A) \geq \gamma\nu_3^n(A) = \gamma\delta$ . Unfortunately, this is not quite true; in fact,  $\beta_4(s)$  can be arbitrarily smaller than  $\beta_3(s)$ . However this occurs only for  $s$  which are very close to  $n$ .

Indeed, one can think of  $s \sim \beta_k$  as placing  $k - 1$  bars independently at random within a row of  $n$  dots, and taking the length of the first dot-segment thus formed (here I am conflating equal-slices and equal-n.d.-slices). Hence

$$\begin{aligned} \Pr_{\beta_4}[s] &= \Pr[\text{min of bar1, bar2, bar3} = s] \\ &\geq \Pr[\text{min of bar1, bar2, bar3} = s \mid \text{min of bar1, bar2} = s] \\ &\quad \cdot \Pr[\text{min of bar1, bar2} = s] \\ &= (1 - s/n) \Pr_{\beta_3}[s]. \end{aligned}$$

So we have  $d\lambda/d\nu_3^n \geq \gamma$  for all  $s$  with  $s/n \leq 1 - \gamma$ . But we can simply "lose" (additively) the probability contribution from  $s/n \geq 1 - \gamma$  since this occurs with probability only  $\gamma^{k-1}$  under  $\beta_k$ . A little calculation then gives the claimed  $\lambda(A) \geq \gamma\delta - 2\gamma^3$ . (Some tiny details skipped.)

46.  Ryan O'Donnell Says:  
[April 24, 2009 at 4:16 pm](#) | [Reply](#)

1077. Simplifying the partitioning argument.

It seems to me slightly simpler if we do the partitioning argument via the older notion of “ij-insensitivity on a subset of coordinates”. (“Local ij-insensitivity”, I think Terry called it?)


**Definition:** Let  $a, b \in [k]$  be distinct characters, and  $I \subseteq [n]$ . We say that  $A \subseteq [k]^n$  is “ $ab$ -insensitive on  $I$ ” if the following condition holds: for all  $x \in A$ , if  $x'$  is formed by changing  $x_i$  from  $a$  to  $b$  or  $b$  to  $a$  for some  $i \in I$ , then  $x' \in A$  also. If  $I = [n]$  we simply say that  $A$  is “ $ab$ -insensitive”.

True, this introduces an extra definition. But it has the upside of making the argument a little simpler, I think. The reason is that one doesn't have to carry around pieces of the set  $B$  living in different  $[k]^n$  universes, and one doesn't have to collect up pieces of the eventual error set. After removing a subspace, one can simply say the new set is ij-insensitive except on some additional  $m$  coordinates.

As an example justification, [here is the partitioning argument under any product distribution](#), fitting on a page.

In fact, having written that, it's now (finally) clear to me how to directly do the partitioning under equal-slices (if we want to). In the product distribution case, all we *really* used was that we could find a  $d$ -dimensional subspace in  $B$  localised to our favourite  $m$  coordinates. We can *almost* do this under equal-slices except that we don't get to pick  $m$  coordinates arbitrarily; they have to be picked randomly. Still, that's not a big problem; so long as we've only “used up”  $T_m$  insensitive coordinates, there are still  $n - T_m$  insensitive ones remaining, so a random  $m$ -subset will fall inside the insensitive coordinates with probability  $1 - o_n(1)$ .


(The benefit we get from using “local ij-sensitivity” is that we can take the equal-slices distribution of  $B$  over *all* coordinates to be the progress measure; the probability measure doesn't keep shifting on us.)

-  Ryan O'Donnell Says:  
[April 24, 2009 at 4:18 pm](#)

Typo: “ $V$ ” towards the end of the above should be “ $T_m$ ”.

-  Ryan O'Donnell Says:  
[April 24, 2009 at 4:19 pm](#)

What the heck? “ $V$ ” should be “ $T_m$ ”. But every time I wrap “ $T_m$ ” in latex-dollar-signs, it changes to a  $V$  in the published version. How strange!

47.  Ryan O'Donnell Says:  
[April 24, 2009 at 8:08 pm](#) | [Reply](#)

It seems a couple of the wiki pages on generalizations to DHJ( $k$ ) have been deleted. Well, not exactly deleted, but blanked out, by mystery IP addresses.


See [here](#) and [here](#).

I've tried to revert them properly, but can we use captchas on edits or block unregistered users from editing?

- [Klas Markström](#) Says:  
[April 25, 2009 at 8:15 am](#)



Earlier I have reverted several changes where someone replaced the proper text with a lot of non-functioning URLs.

48.  *Ryan O'Donnell* Says:  
[May 2, 2009 at 2:12 am](#) | [Reply](#)

I have started creating TeX files for the writeup. We can still discuss the outline on the [page Terry created](#).

But to get the ball rolling, here is a [new page for the TeX](#). The idea was proposed by someone in an earlier thread. On this page are wiki-links to 6 pages, named “dhj.tex”, “dhj.sty”, “intro.tex”, etc.

When you go to the wiki page, e.g., [measures.tex](#), don't try to read it as a wiki page. Just click “edit”, copy the source for the page, paste it into your favourite editor, and save the document as measures.tex.

Once you've done this for all 6 files, you should be able to run “latex” and “bibtex” to produce a document.

If you want to change a file, just go ahead in your TeX editor, and when you're done, copy/paste it back into the appropriate wiki-page source.

It's extremely bare-bones right now; I didn't even put in the title etc. yet. Also, more files/sections need to be created.

The only substantive thing done so far is writing some of the “measures” section. My goal for this section is just to get down, in the shortest amount of space, the lemmas that “doing uniform on most coordinates and equal-slices on an  $\epsilon$ -fraction is  $O(\epsilon \sqrt{n})$ -close to uniform”, “doing equal-slices on most coordinates and then either uniform or equal-slice on an  $\epsilon$ -fraction is  $O(k \epsilon \sqrt{n})$ -close to uniform”.

I got most of the way on this, but didn't finish.

## Leave a Reply

Logged in as [Ryan O'Donnell](#). [Logout »](#)

☐ Notify me of follow-up comments via email.

---

[Blog at WordPress.com.](#)

[Entries \(RSS\)](#) and [Comments \(RSS\)](#).

⌵