The corners problem over finite fields

Talk by Ashwin Sah
Notes by Sanjana Das
February 16, 2024

This is based on work due to Shkredov from 2004 and 2005, and notes by Ben Green.

§1 Introduction

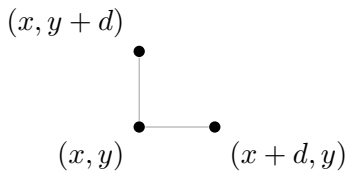
There are several problems which ask for how large a set can be if it avoids a certain pattern. For example, Roth's theorem is about avoiding 3-APs, and Szemerédi's theorem is about avoiding k-APs.

Definition 1.1. We use $r_3(N)$ to denote the size of the largest subset of $[N]$ avoiding $\{x, x+d, x+2d\}$ with $d \neq 0$ (i.e., avoiding 3-APs).

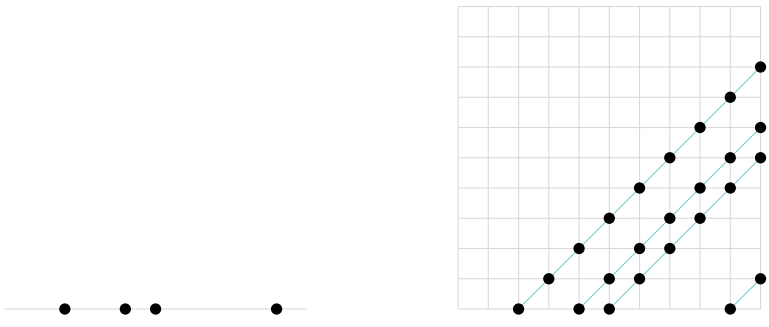
There's multiple ways to extend the notion of such a pattern. One is to replace d, $2d$, ... with some sort of polynomial. Another is to go multidimensional. Today we'll go multidimensional — once you've proven Szemerédi's theorem it implies the existence of most *linear* patterns you'd care about in one dimension, so we'll now consider a two-dimensional pattern.

Definition 1.2. We use $r(N)$ to denote the maximum size of a corner-avoiding subset $A \subseteq [N]^2$, where a corner is a set $\{(x,y), (x+d,y), (x,y+d)\}$ (with $d \neq 0$).

Note that d is allowed to be negative.



An upper bound for corners implies one for 3-APs via a simple projection argument — given any set avoiding 3-APs, we can draw them on the x-axis and take the lines from each at a 45° angle to get a corner-avoiding set of roughly N times the size.



(This set is corner-avoiding because if we had a corner and projected it down along lines of this angle, then we'd get a 3-AP.)

Today we'll consider the following theorem.

Theorem 1.3
We have $r(N) = o(N^2)$.

This was first proved by Ajtai–Szemerédi (1974) using Szemerédi's theorem. Then Solymosi (2003) proved it using the triangle removal lemma — this is a nice argument similar to how you prove Roth's theorem using the triangle removal lemma.

But the bounds on $r(N)$ that we get from these proofs are pretty poor quantitatively — they involve log iterated a large number of times. The first quantitatively 'good' bound comes from the work of Shkredov that we'll be talking about today — in 2004 he proved a bound of $r(N) \le N^2(\log\log\log N)^{-c}$ for some $c > 0$, and in 2005 he improved the bound to $N^2(\log\log N)^{-c}$.

We're going to work over finite fields, to illustrate the key points — so we'll consider $r(\mathbb{F}_2^n)$ instead of $r(N)$ (defined in the same way). We'll also see comparisons to Roth's theorem, which we'll do in $\mathbb{F}_3^n$ (we can't work over $\mathbb{F}_2$ because our pattern has a coefficient of 2, but the difference isn't important).

§2 Roth's theorem

First we'll review the proof of Roth's theorem in $\mathbb{F}_3^n$ (which states that $r_3(\mathbb{F}_3^n) = o(3^n)$). We can prove Roth's theorem using a 'structure vs. pseudorandomness' approach, as encapsulated by the following statement.

Theorem 2.1
Given a set $A \subseteq \mathbb{F}_3^n$ of size $|A| = \alpha \cdot 3^n$, at least one of the following holds:

(1) (*Pseudorandom*) A has roughly the correct amount of 3-APs given its density, i.e.,

$$\#(\text{3-APs in } A) \approx \alpha^3 N^2.$$

(2) (*Structure*) There is some hyperplane H for which

$$\frac{|A \cap H|}{|H|} \ge \alpha + \frac{\alpha^2}{4}.$$

(3) (*Exit*) n is too small relative to the density of A — specifically, $\alpha^2 < 2/3^n$.

Given this, in order to prove Roth's theorem, given a set A we iterate Theorem 2.1 roughly $O(1/\alpha)$ times — whenever we're in (2), we restrict A to that hyperplane H (which boosts its density). Then we iterate roughly $O(1/\alpha)$ times; if $1/\alpha \ll n$ then we must eventually hit (1) (we can't keep iterating (2) forever because then our density would cross 1, and we're not going to hit (3) because we defined the original density to be large enough).

One approach to proving Theorem 2.1 is using the Fourier transform.

Definition 2.2. For a finite abelian group G and a function $f\colon G \to \mathbb{C}$, we define the Fourier coefficient of f at ξ (for each ξ in the dual group of G, which we denote by $\widehat{G}$) as

$$\widehat{f}(\xi) = \mathbb{E}_{x \in G}[f(x)\overline{\xi(x)}]$$

Fact 2.3 (Fourier inversion) — For any $x \in G$, we have $f(x) = \sum_{\xi \in \widehat{G}} \widehat{f}(\xi)\xi(x)$.

In our setting, the ξ's are given by $\xi(x) = e^{2\pi i(a \cdot x)/p}$ for each $a \in \mathbb{F}_p^n$.

Definition 2.4. We say a set $A \subseteq G$ is η-Fourier uniform if $\max_{\xi \neq 0} |\widehat{\mathbf{1}_A}(\xi)| \leq \eta$.

The Fourier coefficient at $\xi = 0$ is just the density of A, and Fourier uniformity means that all other Fourier coefficients are small. (This condition basically gives a L^∞ bound on $\widehat{\mathbf{1}_A}$.)

Then the point of how we prove Roth's theorem is that if A is $\alpha^2/2$-uniform, then we can prove (1) by using its Fourier transform — we can write the 3-AP density in A in terms of the Fourier coefficients of $\mathbf{1}_A$, and the $\xi = 0$ term will give the right contribution corresponding to the density of A, while all the other terms will have small total contribution (using some L^2–L^∞ inequality).

Meanwhile, if A is *not* $\alpha^2/2$-uniform, then ξ gives a direction to consider hyperplanes in, and A will need to have substantially higher density on one of those hyperplanes.

§3 A first attempt for corners

Now we'll consider the corners problem over $\mathbb{F}_2^n$ — so we've got a set $A \subseteq \mathbb{F}_2^n \times \mathbb{F}_2^n$ of density α (and we'll let $N = 2^n$). We'd like to prove an analog of Theorem 2.1, so as a first attempt, we'd like to prove that A must fall into one of the following cases.

(1) (*Pseudorandom*) A has roughly the right number of corners (which is $\alpha^3 N^3$).

(2) (*Structure*) A has density at least $\alpha + f(\alpha)$ on a coset of $H \times H$, where $H \subseteq \mathbb{F}_2^n$ is some subspace of $\mathbb{F}_2^n$ with codimension at most $g(\alpha)$. (Think of $f(\alpha)$ and $g(\alpha)$ as some polynomials in α.)

(3) (*Exit*) n is too small.

Here (1) and (3) are essentially the exact same as their analogs in Theorem 2.1. For (2), we might not expect to be able to get a *hyperplane* H (as in Theorem 2.1) — we might have to go down in dimension by some amount greater than 1 — but as long as that amount is controlled just by α, we're fine with it.

Remark 3.1. It's fine to use the same set H in both coordinates — if we got sets H_1 and H_2 for which A had a density increment on a coset of $H_1 \times H_2$, then we could take H to be their intersection.

However, we *do* need to choose this subspace (that we're getting a density increment on) to be coordinate-aligned — i.e., we can't choose a subspace defined by $a \cdot x + b \cdot y = 0$ for arbitrary $a, b \in \mathbb{F}_2^n$ — because corners are rectilinear (i.e., they care about the axes). So we really do need to take the subspace to be a product (and not just an arbitrary subspace) to be able to iterate.

If we could get a statement like this, then we could run the same iteration argument as in Roth's theorem. Unfortunately, this dream is not going to work. If you tried to go through the proof of Theorem 2.1 but with corners, the step you'd get stuck at is trying to extract the subspace H. And there's a good reason for this — there are actually counterexamples to this statement, as we'll now see.

If we're trying to come up with examples, the first natural example is a completely random set, but that'll have the right number of corners. Here we've got a product space, so we'll try 'diagonally embedding' a 1-dimensional random set instead, and this *does* give a counterexample.

Example 3.2

Consider a random subset $B \subseteq \mathbb{F}_2^n$ sampled at rate β, and let $A = B \times B$ (so $\alpha = \beta^2$).

To count the number of corners in this set, we want to count the number of x, y, and d such that (x, y), $(x+d, y)$, and $(x, y+d)$ are all in $B \times B$, meaning that x, $x+d$, y, and $y+d$ are all in B. This occurs for about a β^4-fraction of x, y, and d (as there's four things that need to be in B), so the number of corners is roughly $\beta^4 N^3 = \alpha^2 N^3$ (as opposed to $\alpha^3 N^3$, which is what we'd expect looking just at the density). This means (1) doesn't hold. (Intuitively, the point is that some of the coordinates are repeated, so that we only have four conditions on something being in B, rather than six.)

But on the other hand, A is 'uniform' with respect to subspaces of the form $H \times H$ (i.e., it has roughly the density you'd expect on any coset of such a space), because B is. So (2) doesn't hold either.

This shows that a direct analog of Theorem 2.1 isn't going to work here — there's an issue that we have sets A which don't look pseudorandom in the sense that they don't have the right number of corners, but also are not structured in the sense of (2). (In this example, we actually have *more* corners than we'd expect, not fewer; but it's unclear how we could leverage this.)

§4 The main idea

We'll now get to the key idea of Shkredov, which is very beautiful. The idea is that we still want to get a density increment on *some* kind of set, as in (2), but we can't guarantee one on sets of the form $H \times H$ (where H is a subspace), so we're going to relax what kinds of sets we're looking for.

Question 4.1. What class of sets can we get a density increment on (meaning that if A doesn't have the right number of corners and has reasonable density, then A has a density increment on a set from this class)?

Example 3.2 shows that we can't just take sets which are algebraically structured — we have to allow some sort of looseness.

The first thing that jumps out is that the counterexample in Example 3.2 is caused by a diagonal embedding, where A is a product; so as a first attempt, what if we try to get a density increment onto a product set — i.e., a set of the form $E_1 \times E_2$ (for large but otherwise arbitrary E_1 and E_2)?

It turns out that a statement like this will be true. But then the problem is that E_1 and E_2 are arbitrary sets, so how do you iterate? (With Roth's theorem, iteration was nice because we passed down to hyperplanes, where everything continued to work in the same way; but here when we pass down to $E_1 \times E_2$ we'll have these weird host sets that we can't do much with.)

It might be able to iterate for a few more steps, by still doing Fourier analysis with respect to the original space. But the problem is that as you iterate multiple times, the host sets E_1 and E_2 become sparse, and this Fourier analysis is going to lose factors based on their densities; and this means we lose. So to fix this, we have to somehow 'relativize' the argument with respect to $E_1 \times E_2$ (rather than doing the argument with Fourier analysis in the original set).

But we can't do this with arbitrary sets E_1 and E_2. So we're going to try to require E_1 and E_2 to be *very Fourier uniform* — this will allow us to do the argument relative to $E_1 \times E_2$. (In some sense, such sets end up being just as good as subspaces for our purposes, but they really have to be very uniform — specifically, their uniformity needs to be really good with respect to their densities.)

§4.1 The trichotomy

Here's the right trichotomy that we'll use (as an analog to Theorem 2.1), stated informally.

Theorem 4.2 (Informal trichotomy for corners)
For any set $A \subseteq \mathbb{F}_2^n \times \mathbb{F}_2^n$ of density α, one of the following holds (letting $N = 2^n$):

(1) A has roughly $\alpha^3 N^3$ corners.

(2) A has density at least $\alpha + (\alpha/2)^{100}$ on a translate of some $E_1 \times E_2 \subseteq H \times H$ where $H \subseteq \mathbb{F}_2^n$ is a subspace of low codimension, and E_1 and E_2 are subsets of H of size at least $\mathsf{poly}(\alpha)\,|H|$ which are very Fourier uniform with respect to H.

(3) n is too small.

We're not yet going to quantify what 'very' means in (2), but if E_1 and E_2 have densities β_1 and β_2 in H, then we'll need them to be η-Fourier uniform where η is good with respect to α as well as β_1 and β_2, such that a set with this uniformity are in some sense as good a host set as the original subspace itself.

Remark 4.3. By a 'translate' of $E_1 \times E_2$, we mean a set obtained by taking a fixed vector and adding it to each vector in $E_1 \times E_2$. (The only reason we write this is because if we wanted to avoid translating, we'd need to consider cosets of $H \times H$, which would be affine subspaces rather than actual subspaces, and then we'd have to define Fourier coefficients in affine subspaces.)

§4.2 A proof outline

Ben Green breaks the proof down into the following three steps.

(1) (*Generalized von Neumann*) We show that if A is 'close' to the constant function α in some appropriate norm, then it has roughly the right number of corners. This doesn't have a direct analog in the proof of Roth's theorem we described. But you can also handle the pseudorandom case of Roth's theorem (where we assume A is Fourier uniform and show it has the right 3-AP count) by using the Gowers U^2 norm and Cauchy–Schwarz — to show that if A is close to constant in U^2 norm then it has the right 3-AP count — and then relating the U^2 norm to Fourier coefficients. And then you can think of this step as replacing the U^2 norm with a different norm — it'll be a 'rectangle norm' that's sort of counting C_4's in some sense (we'll define this norm later).

(2) (*Density increment on a product set*) Now we consider the case where A is 'far' from the constant function α in norm; then we show that it has a density increment on some product set $F_1 \times F_2$ (where F_1 and F_2 are reasonably large, but we don't impose any other conditions on them yet).

This is not hard; intuitively, our norm is basically a C_4 count and $F_1 \times F_2$ is a box, so we can do basically the same thing as Chung–Graham–Wilson (where they show several notions of pseudorandomness in a graph are equivalent; here the two relevant ones are the graph having C_4 count roughly what you'd expect from its number of edges (which intuitively corresponds to A being close to α in norm), and the graph having roughly the same density between any two large sets (which corresponds to A having roughly the same density on all large product sets $F_1 \times F_2$)).

(3) (*Uniformizing the product set*) This is the key step of the argument. From (2), we're starting off with some arbitrary product set $F_1 \times F_2$ in some host subspace W. And then we want to use this to find a small codimension subspace W' (of the original space) and a coset of W' on which we can basically replace $F_1 \times F_2$ with a product set $E_1' \times E_2'$ where E_1' and E_2' are Fourier uniform, and A still has a density increment on this smaller host $E_1' \times E_2'$.

(Then when we iterate, you can think of W' as the new host subspace we're in, and $E_1' \times E_2'$ as the new product set that A lives in. We'll maybe have worse densities than before, so we'll need to uniformize a *lot* at this stage to be able to loop back; but this is the basic sketch.)

§5 The proof

Next, we'll write an outline of the concrete statements that you prove, and then we'll discuss some points about these statements — we'll especially focus on (3) (the uniformization process), because that's the thing that's particularly novel in the proof; but there's also an additional novelty in (1) (the generalized von Neumann step), because you need to make sure that the argument relativizes correctly with respect to the pseudorandom host sets.

§5.1 Some definitions

First, we'll define what it means to be 'sufficiently' uniform.

Definition 5.1. Given a density $\alpha \in (0,1)$, we define Struct_α as the collection of all translates of sets of the form $E_1 \times E_2$ such that:

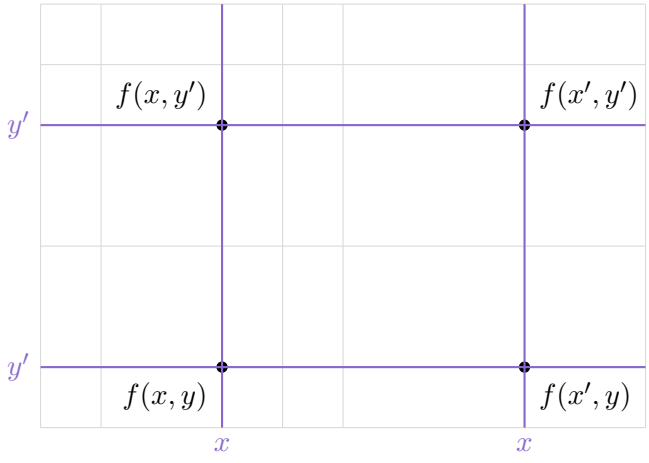
- There is some subspace $H \subseteq \mathbb{F}_2^n$ such that E_1 and E_2 are both subsets of H, with sizes $|E_i| = \beta_i |H|$.
- E_1 and E_2 are both $(\alpha\beta_1\beta_2/2)^{100}$-Fourier uniform with respect to H.

This quantifies the notion of uniformity that we want in step (3) (or in case (2) of Theorem 4.2) — Struct_α is essentially the collection of 'structured sets' that we're trying to get a density increment on. (The reason for naming it this way is that Ben Green's notes consider several problems along these lines, and for each problem he defines the correct kind of 'structured set' that you want to pass to — for example, in Roth's theorem, the structured sets are just hyperplanes of codimension 1.)

Next, we'll define the relevant norm.

Definition 5.2. For a function $f: S \to \mathbb{R}$ where $S = E_1 \times E_2$ is a product set, we define its rectangle norm, denoted $\|f\|_S$, by

$$\|f\|_S^4 = \mathbb{E}_{x,x' \in E_1} \mathbb{E}_{y,y' \in E_2} f(x,y) f(x',y) f(x,y') f(x',y').$$



The reason we referred to this norm as a C_4 count is that we can imagine a graph with vertex set E_1 on one side and E_2 on the other, where we draw an edge for each $(e_1, e_2) \in A$ (where $f = \mathbf{1}_A$); then in this norm we're taking two vertices on each side and checking whether they form a C_4.

Finally, we'll set up a bit of convenient notation (since we'll be considering the density of A on various sets).

Notation 5.3. We use $\delta_S(A)$ to denote $|A \cap S| / |S|$.

§5.2 A quantitative version of the outline

We'll now give a quantitative version of the outline from Subsection 4.2. First, step (1) — that if the indicator of A is very close to constant (i.e., to the relative density of A), then we have many corners — corresponds to the following statement.

Proposition 5.4

Let $S = E_1 \times E_2$ be in Struct_α with $\delta_S(A) \geq \alpha$. Then if $\|\mathbf{1}_A - \delta_S(A)\|_S^4 \leq (\alpha/2)^{100}$, then A has at least $(\alpha\beta_1\beta_2/2)^{100} |A|^3$ corners.

Remark 5.5. This is different than the way we've written step (1) earlier — we're not claiming that A has the correct number of corners, just a lower bound (which is much smaller than the correct number). But you should actually be able to get that A has the correct count of corners; this is just simpler.

Next we'll quantify step (2), which states that if A is far from its density in norm, then we can get a big product set where we have a density increment. This is just a statement about the norm; it doesn't have anything to do with the host sets.

Proposition 5.6

Suppose that $S = E_1 \times E_2$ (for any E_1 and E_2) and that $A \subseteq S$ is such that $\delta_S(A) = \alpha$ and $\|\mathbf{1}_A - \alpha\|_S^4 \geq \eta$. Then there are subsets $F_1 \subseteq E_1$ and $F_2 \subseteq E_2$ of relatively big sizes — i.e., with $|F_i| \geq 2^{-100}\eta |E_i|$ for each $i \in \{1, 2\}$ — such that

$$\delta_{F_1 \times F_2}(A) \geq \alpha + 2^{-100}\eta^2.$$

Finally, the key proposition is the one for (3) (the uniformization step).

Proposition 5.7

Let $\alpha, \tau, \sigma \in (0, 1)$, and let $W \subseteq \mathbb{F}_2^n$ be a subspace of size $|W| \geq \exp(16\sigma^{-2}\delta^{-1}\tau^{-1})$. Let $S' = F_1 \times F_2 \subseteq W \times W$ be such that $\delta_W(A) = \alpha$ and $\delta_{S'}(A) = \alpha + \tau$. Let $\delta_i = |F_i| / |W|$ and $\delta = \delta_1\delta_2$.

Then there is a subspace $W' \subseteq W$ and shifts $t_1, t_2 \in W$ such that if we define $E_1' = (F_i - t_i) \cap W'$ and $S'' = E_1' \times E_2'$, then the following statements all hold:

(1) $\dim W - \dim W' \leq 8\sigma^{-2}\delta^{-1}\tau^{-1}$.

(2) $|S''| \geq \delta\tau |W'|^2 / 2$.

(3) E_1' and E_2' are 2σ-Fourier uniform subsets of W'.

(4) $\delta_{S''}(A - (t_1, t_2)) \geq \alpha + \tau/8$.

This is kind of a mouthful. But in words, we think of α as the original density, τ as the increment size we got on an arbitrary product set (from step (2)), and σ as our target uniformity. (When we apply the

proposition, σ will be really small.)

Here, W is the subspace we're starting with (it's essentially the same thing as the subspace H corresponding to $E_1 \times E_2$ in the definition of Struct_α); we don't want W to be too small, but this is more of an exit condition. (In particular, in comparison to Theorem 4.2, W corresponds to $\mathbb{F}_2^n$ and the size condition on W corresponds to the exit condition (3).) And we've got a product set S', which we can think of as the output of Proposition 5.6 — so we've got a density increment of τ on S'. And we use δ_i to denote the density of F_i in W, and δ to denote the density of their product S' in $W \times W$.

And in the conclusion, W' is essentially the same thing as H from Theorem 4.2; it's the subspace hosting our new pseudorandom hosts E_1' and E_2', and S'' is this new pseudorandom host (i.e., when we iterate, it's going to become the S in Proposition 5.4). The second condition says that S'' is still pretty big, the third condition (that E_1' and E_2' are uniform) is the key point that allows us to iterate, and the fourth condition says that we didn't sacrifice too much of our density increment — we started with a density increment of τ onto an arbitrary product, and we've gotten a density increment comparable to τ onto this really *uniform* product S''.

Remark 5.8. As a technical remark, it's important that Proposition 5.6 makes no reference to σ (i.e., the uniformity of E_1 and E_2) — this is because when we apply Proposition 5.7, we'll need σ to be very small relative to δ and τ (i.e., we'll need to be able to uniformize in a meaningful set — if δ were really small relative to σ instead, then you'd trivially have the level of uniformity coming from any sparse set being uniform).

Given these three statements, here's how we finish the proof — we essentially run the three statements in order, take the output of the third and plug it back into the first, and iterate.

Proof of Theorem 1.3. Suppose we start with a corner-avoiding set $A \subseteq \mathbb{F}_2^n \times \mathbb{F}_2^n$ of density α (and let $E_1 = E_2 = \mathbb{F}_2^n$ and $S = E_1 \times E_2$). Then plugging this into Proposition 5.4, we get $\delta_S(A) \geq (\alpha/2)^{100}$, and then Proposition 5.6 gives that A has a density increment of at least $\tau = (\alpha^2/8)^{100}$ on some reasonably large product set $F_1 \times F_2$. Then Proposition 5.7 (with $W = \mathbb{F}_2^n$) gives new sets $A' \subseteq E_1' \times E_2' \subseteq W' \times W'$ (where A' is obtained by shifting A and intersecting it with the new product set $E_1' \times E_2'$) where A' has density at least $\alpha + \tau/8 \geq \alpha + (\alpha^2/16)^{100}$ on $E_1' \times E_2'$, E_1' and E_2' are sufficiently uniform, and W' has small codimension in W (so we haven't lost too much in the dimension). (We take σ to be small enough that we can then apply Proposition 5.4 to these new sets.)

And then we iterate — we apply Proposition 5.4, then Proposition 5.6, then Proposition 5.7 again, until we hit density 1. At each step we're not losing too much in the dimension, so we end up getting a bound of $r(\mathbb{F}_2^n) \lesssim N^2/(\log\log N)^c$ for some small c. □

Remark 5.9. Where does the second log come from (in comparison to Roth's theorem, where this proof only ends up with one log)? It's there because we have some loss in the uniformization — the β_i's sort of explode when we iterate. If we begin with β_1 and β_2 (representing the densities of our original pseudorandom hosts E_1 and E_2 in the original subspace W), then Proposition 5.6 gives a product set with $|F_i| \geq (\alpha/4)^{100} |E_i| \geq (\alpha/4)^{100} \cdot \beta_i |W|$, which means $\delta \geq (\alpha/4)^{200}\beta_1\beta_2$, while $\tau \geq (\alpha^2/8)^{100}$. And $\delta\tau/2$ corresponds to $\beta_1'\beta_2'$ because of the second condition of Proposition 5.7 (where β_1' and β_2' are the values of β_1 and β_2 we'll have on the next iteration), so we'll have $\beta_1'\beta_2' \approx (\alpha^4/128)^{100}\beta_1\beta_2$. And when we apply Proposition 5.7, we need σ to be small enough that we can iterate Proposition 5.4, so we're taking $2\sigma \approx (\alpha\beta_1'\beta_2'/2)^{100}$.

And we need at most $(16/\alpha^2)^{100}$ iterations to hit density 1, so in the end we'll have $\beta_1\beta_2 \approx \alpha^{\alpha^{-200}}$ (where we've dropped several constants), which means we're losing roughly $\alpha^{-\alpha^{-200}}$ in the dimension in the end. This means we need $\alpha^{-\alpha^{-200}}$ to be small compared to n (to ensure we don't hit the exit condition before density 1), so α^{-1} should be a small power of $\log n \approx \log\log N$.

Remark 5.10. For integers, this strategy ends up giving $r(N) \lesssim N/(\log\log\log N)^c$ (for small $c > 0$). Shkredov later improved the bound for $\mathbb{F}_2^n$ to have one log and the bound for integers to have two log's.

Remark 5.11. Finally, we'll remark that the example from Example 3.2 is compatible with this outline — Proposition 5.6 is going to identify B, and the uniformization step (Proposition 5.7) is going to chop it up and find some uniform component of B.

§5.3 Proof sketch of Proposition 5.4

First, for Proposition 5.6, the point is that the rectangle norm is really counting C_4's, and then we can use a slightly modified Chung–Graham–Wilson type argument. The really interesting pieces are Propositions 5.4 and 5.7; we'll first say a couple of words about Proposition 5.4, and then we'll focus on Proposition 5.7.

We can think of Proposition 5.4 as some type of Cauchy–Schwarz argument, in analogy to the Gowers Cauchy–Schwarz argument relating the U^{k-2} norm to k-APs (where you do iterated Cauchy–Schwarz, duplicating variables one at a time). But the important thing is that we need things to be relativized — we need to be taking expectations relative to the quasirandom host sets E_i, rather than H itself. (Otherwise we'd be losing factors corresponding to the densities of the host sets, and this would be bad.)

This means you have to be careful — when we use Cauchy–Schwarz to duplicate a variable x (as with the Gowers argument for k-APs) we usually get x and $x + h$, and we need to be working in a subspace in order to say that if x and $x + h$ are uniform then so is h.

So in order to relativize correctly, we include terms like $\mathbf{1}_{E_1}(x)$ and $\mathbf{1}_{E_1}(x + h)$ and $\mathbf{1}_{E_2}(y)$ and $\mathbf{1}_{E_2}(y + h)$ and so on, alongside everything else, when we use Cauchy–Schwarz (and we're still choosing x, y, and h from the host subspace). And in the end, we use the quasirandomness of E_1 and E_2 to show that the associated terms end up averaging out and being roughly the correct constant (you can do this using Fourier analysis or some pseudorandomness arguments involving Cayley graphs and codegrees).

§5.4 Proof of Proposition 5.7 — the uniformization step

Now we'll talk about how to perform the uniformization step. The way we wrote Proposition 5.7, we start with A, and we want to cut up our original space $W \times W$ into smaller pieces and eventually find one piece where A gets a density increment and the piece is 'regular' (i.e., E_1' and E_2' are uniform). But the way we'll *actually* show this is by getting a sort of 'regular decomposition' — in analogy to Szemerédi's regularity lemma, you can imagine we're taking something like a weakly regular decomposition of $F_1 \times F_2$ on $W \times W$ (meaning that we're chopping up $W \times W$, and $F_1 \times F_2$ is supposed to be regular (i.e., uniform) on most of the pieces). This decomposition will have some error pieces of two types — either they're *small* (these pieces are referred to as *expired*) or *non-uniform*. We'll show that the contribution of these pieces is small, and we'll choose parameters so that they're so small that even if we kill them, our set A still has a reasonable density increment.

So the first step — which is the big one — is to get the decomposition. The second step is to forget the 'bad' pieces. And the third step is to use averaging (or Markov) to show that we still get a density increment on one of the good (i.e., uniform) pieces.

§5.4.1 A proof sketch

First, we'll discuss the high-level ideas behind how we get the regular decomposition. We're going to do this by an iterative algorithm, where we iteratively take partitions (very much like the proof of the Szemerédi regularity lemma) and use an energy argument.

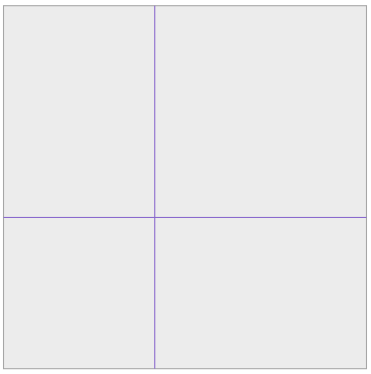
Let $\dim W = n$ (we're chopping up the space $W \times W$). At step 0, we'll just have the whole set $W \times W$ (as one part). And we're going to repeatedly partition so that at every further step t, we'll have a partition $W \times W = \bigcup_{i \in I_t} C^{(i)}$, where $\mathcal{I}_t$ is some index set that grows with time, and each $C^{(i)}$ is a cell in our partition of the form

$$c^{(i)} = (W^{(i)} + t_1^{(i)}) \times (W^{(2)} + t_2^{(i)})$$

for some subspace $W^{(i)} \subseteq W$ with $\dim W^{(i)} \geq n - t$.

Imagine we start out with $W \times W$ as a big square; in the first step, we just have a single piece, and we have $F_1 \times F_2$ on this big square. If $F_1 \times F_2$ is uniform on this big square, then we're done (and if it's very low-density, then it's also uniform, so we're also done). So we can imagine that it's not uniform.

This means either F_1 is non-uniform in W, or F_2 is. So we can find some vector along which to cut (i.e., we split W into two subspaces along a certain vector). Now we have four cells (since we cut W into two in both directions).

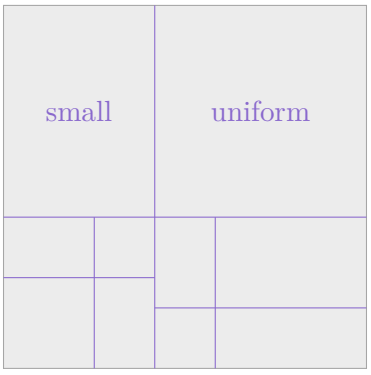


$W \times W$

Intuitively, because we were non-uniform in one of these directions (i.e., either F_1 or F_2 was non-uniform, and we took a cut showing this), the cut in that direction should give us an energy increment, and the cut in the other direction won't change that (as in general, cutting can only increase energy). So taking this partition will give us an energy increment — if we failed to be σ-uniform, then we'll get an energy increment of roughly σ^2.

Then we look at the pieces we have left after this cut. On some of these pieces $F_1 \times F_2$ will be small; we ignore those pieces. On some it'll be uniform; we're done with those pieces. So we only need to deal with the pieces on which $F_1 \times F_2$ is non-uniform.

And for each of these pieces, we can find some other direction on which one of F_1 and F_2 is non-uniform, and then split them up along that direction. (We can use different directions for the different pieces; in the end, we'll just be passing to a single cell.)



$W \times W$

At time t, each cell has only been subdivided t times, so its corresponding subspaces have codimension at most t.

And as long as there's 'many' non-uniform cells (e.g., at least a $\tau\delta/16$-fraction of $W \times W$ — the reason for this parameter is that it's roughly the amount of badness we're able to throw away without losing the density increment), we can get an energy increment of $\Omega(\sigma^2\tau\delta)$ (where δ is the original density of $F_1 \times F_2$). This means the process should continue for at most roughly $\sigma^{-2}\tau^{-1}\delta^{-1}$ steps (or else the energy would cross 1), so that's the codimension of each cell we end up with.

And once we've got this decomposition, we can throw away the bad pieces (i.e., the ones where $F_1 \times F_2$ is non-uniform or small). These pieces have very little contribution to the density increment (they certainly aren't contributing to the density increment if $F_1 \times F_2$ is very small on them, and the non-uniform cells occupy a small fraction of $W \times W$ relative to the density increment). So A still has reasonable density increment on average over all the 'good' pieces (i.e., one where $F_1 \times F_2$ is uniform and not small), which means we can find some good piece where it has such a density increment.

§5.4.2 The energy argument

Finally, we'll describe in more quantitative detail how the energy argument works. Suppose we've got a partition of $W \times W$ into cells $\bigcup_{i\in\mathcal{I}_t} C^{(i)}$, where each cell $C^{(i)}$ is of the form

$$C^{(i)} = (W^{(i)} + t_1^{(i)}) \times (W^{(i)} + t_2^{(i)})$$

for some subspace $W^{(i)} \subseteq W$ of codimension at most t.

We care about how much of each cell our host product set $F_1 \times F_2$ occupies (which we refer to as its relative density in the cell), as well as how uniform it is. So we'll consider $C^{(i)} \cap (F_1 \times F_2)$, which we can write as $D_1^{(i)} \times D_2^{(i)}$. We'll define

$$\delta_1^{(i)} = \frac{|D_1^{(i)}|}{|W^{(i)}|} \text{ and } \delta_2^{(i)} = \frac{|D_2^{(i)}|}{|W^{(i)}|},$$

and $\delta^{(i)} = \delta_1^{(i)}\delta_2^{(i)}$ — so $\delta^{(i)}$ is just the density of $F_1 \times F_2$ in our cell $C^{(i)}$. And the overall density of $F_1 \times F_2$ in $W \times W$ is δ, so when we average $\delta^{(i)}$ over all cells $C^{(i)}$, we should get δ.

We say $C^{(i)}$ is *expired* if $\delta^{(i)} < \delta\tau/2$ — the point is that then $F_1 \times F_2$ occupies less than a $\delta\tau/2$-fraction of all the expired cells, so throwing those cells away isn't going to affect our density increment by too much (specifically, A had a density increment of τ on $F_1 \times F_2$, which had size a δ-fraction of W; so throwing away these cells means we lose at most a $\tau/2$-fraction of $F_1 \times F_2$, and by an averaging argument we still need to have at least a density increment of $\tau/2$ on the rest).

We say $C^{(i)}$ is *uniform* if $D_1^{(i)} - t_1^{(i)}$ and $D_2^{(i)} - t_2^{(i)}$ are both σ-uniform as subsets of $W^{(i)}$. (In words, $F_1 \times F_2$ on our cell $C^{(i)}$ is uniform — the shifting is just so that we've got subsets in a real subspace and not an affine one.) And finally, we say $C^{(i)}$ is *non-uniform* if it's neither expired nor uniform. (This is a quantitative version of what we discussed earlier.)

If the total fraction of $W \times W$ occupied by non-uniform cells is less than $\tau\delta/4$, then we'll stop; the point is that when this happens, the fraction of $F_1 \times F_2$ in non-uniform cells is small, and the fraction of $F_1 \times F_2$ in expired cells is also small by the definition of an expired cell; so we can throw these cells away, and an averaging argument says that we still have a density increment on the remaining cells.

And otherwise, we iterate — we keep on subdividing all the non-uniform cells. To bound how long this process can run for, we use an energy argument — given a decomposition with cells $C^{(i)}$ for i ranging over some index set $\mathcal{I}$, we define its *energy* as

$$\frac{1}{2}\sum_{i\in\mathcal{I}} \frac{|C^{(i)}|}{|W|^2}\left((\delta_1^{(i)})^2 + (\delta_2^{(i)})^2\right).$$

Claim 5.12 — If we haven't yet stopped (so the fraction of $W \times W$ occupied by non-uniform cells is at least $\delta\tau/4$), then we can subdivide each non-uniform cell into 4 cells such that in total, we get an energy increment of $\Omega(\sigma^2 r\delta)$.

Proof. First, let's consider a single non-uniform cell $C^{(i)}$, and let's consider $D_1^{(i)} \times D_2^{(i)} \subseteq C^{(i)}$ (this is the restriction of $F_1 \times F_2$ to $C^{(i)}$). Then the fact that the cell is non-uniform means that either $D_1^{(i)} - t_1^{(i)}$ or $D_2^{(i)} - t_2^{(i)}$ has some large Fourier coefficient with respect to $W^{(i)}$; we'll assume without loss of generality that this is true for D_1.

Then this large Fourier coefficient means there is a vector $a \neq 0$ (in the dual of $W^{(i)}$) such that we can decompose

$$D_1^{(i)} = D_{10}^{(i)} \cup D_{11}^{(i)} = \{x \in D_1^{(i)} \mid a \cdot x = 0\} \cup \{x \in D_1^{(i)} \mid a \cdot x = 1\}$$

where the two pieces have densities at least $\delta_1^{(i)} + \sigma$ and at most $\delta_1^{(i)} - \sigma$ (in the analogous pieces that $W^{(i)}$ gets split into), essentially by the definition of what it means to have a large Fourier coefficient.

This changes the contribution to the energy from $d_1^{(i)}$ to at least $d_1^{(i)} + \sigma^2$ (for just this term).

And then we split up D_2 along the same vector a (which means we get 4 cells); this at least doesn't *decrease* its contribution to the energy.

And finally, this means we're increasing the energy contributed by each non-uniform cell by at least σ^2 (before we weight by the size of the cell); so if non-uniform cells occupy at least a $\delta\tau/4$-fraction of the entire space, then we get a total energy increment of at least $\sigma^2\delta\tau/4$. □