Furstenberg–Sárközy

Talk by Dima Zakharov

Notes by Sanjana Das

December 13, 2024

§1 Introduction

A few weeks ago, Ben and Mehtaab proved the following theorem.

Theorem 1.1 (Green–Sawhney)
If we have a set $A \subseteq [X]$ with $|A| > X\exp(-\log^c X)$, then A contains two integers which differ by a square — i.e., there exist $x, y \in A$ with $x - y = z^2$.

This is a classical problem in additive combinatorics or combinatorial number theory. People got interested in this kind of question in the 1970s. Furstenberg used dynamics to show that if you have a dense set A, then it contains a square difference; this gets a bound of $o(X)$. Sárközy came up with a Fourier analytic argument which gave a quantitative bound of $X/\log^c X$. Then there was a sequence of improvements, which ended up with Bloom–Maynard getting a bound of $X/(\log X)^{\omega(1)}$ (the denominator has $\log X$ to some slowly growing power, which is something like $\log\log\log X$) by refining the approach.

Dima's goal today is to explain where the drastic improvement in Theorem 1.1 comes from (there's a clear reason why you win by so much).

§2 The Fourier analytic approach

Dima will start explaining how you can prove something like this with *some* bound. We'll follow Sárközy's Fourier analytic approach, and at some point we'll get to a step where inequalities from Boolean analysis come into the picture.

The argument is by a density increment strategy. Suppose $|A| = \alpha X$ (so α is the density of A). We'll work with the characteristic function $f = \mathbf{1}_A$.

The first thing we want to do is count how many pairs A has such that their difference is a square. For this, we define another function g as the characteristic function of the set of squares at most X.

Now we can write down a formula for the number of $x, y \in A$ whose difference is a square — $x - y = z^2$ is a nice linear pattern, so we can compute this number using the Fourier transform, and we get

$$\#\{x, y \in A \mid x - y = z^2\} = \int_{\mathbb{R}/\mathbb{Z}} |\widehat{f}(\theta)|^2 \cdot \widehat{g}(\theta)\, d\theta. \tag{2.1}$$

Here we're using the Fourier transform from functions on $\mathbb{Z}$ to functions on $\mathbb{R}/\mathbb{Z}$ — if we hae a function $f\colon \mathbb{Z} \to \mathbb{C}$, then

$$\widehat{f}(\theta) = \sum_{z\in\mathbb{Z}} f(z)e(-\theta z).$$

If you've seen a proof of Roth's theorem or something like that, then (2.1) isn't shocking; you can take this definition and plug it into

$$\#\{x, y \in A \mid x - y = z^2\} = \sum_{x,y} f(x)f(y)g(x-y)$$

to get some long sum, and the only terms that don't die are the ones where the frequencies have the relevant linear relation.

§2.1 Understanding the Fourier transform of g

Now we have a nice formula for the number of square differences. We don't know much about f (the Fourier transform of our set A), but g is some *fixed* function, so we should be able to understand $\widehat{g}$ pretty well.

In fact, this is pretty well-known. By definition, we have

$$\widehat{g}(\theta) = \sum_{z \le X^{1/2}} e(-z^2\theta).$$

This is a quadratic exponential sum, and you can estimate these things pretty well using van der Korput's method. Weyl's estimate basically tells you that if θ is not a rational number of small height, then we get lots of cancellation in the sum. (One way to think about this is that if we had z instead of z^2, then this would be a geometric series, which we could sum up; and if θ is not a rational, then this would always be small.) Here's a precise statement.

Lemma 2.1

Suppose that $|\widehat{g}(\theta)| > \delta X^{1/2}$. Then there exist $p, q \le \delta^{-C}$ such that

$$\left|\theta - \frac{p}{q}\right| < \delta^{-C} X^{-1}.$$

Note that $\widehat{g}(\theta)$ is a sum of $X^{1/2}$ terms, so when we're considering whether it's large, we should compare it to $X^{1/2}$. And this lemma says that the only way it can be large is if θ is very close to a rational number.

WHy is this bound a reasonable thing to expect? For example, we can think about θ close to 0. We have $\widehat{g}(0) = X^{1/2}$. When θ is small, we have

$$\widehat{g}(\theta) = \sum_{z \le X^{1/2}} e(z^2\theta),$$

and z^2 is of order X. We only start seeing cancellation once the phase $z^2\theta$ wraps around the unit circle, so we need to make $\theta \gg 1/X$ to start seeing cancellation. Lemma 2.1 says that as long as we step away by somewhat more than this (by a margin of δ^{-C}), then we get lots of cancellation and there's no way to get a large Fourier coefficient.

We can use Lemma 2.1 to localize our integral (2.1) to X^{-1}-neighborhoods around rationals. Meanwhile, we can understand what $\widehat{g}$ is doing around rationals pretty well too.

Lemma 2.2

We have $|\widehat{g}(p/q)| \le q^{-1/2} X^{1/2}$.

The point is that near rationals, $\widehat{g}$ can be large; but as we increase the denominator q, it should start decaying eventually. And this lemma controls how fast it decays.

Proof. We have a sum $\sum_{z \le X^{1/2}} e(z^2\theta)$, and we can split it into segments of length q; so it's enough to bound the sum over one set of residues mod q, i.e.,

$$\sum_{z=0}^{q-1} e(z^2/q).$$

And this is the classic Gauss sum, whose value we know exactly — we have

$$\left|\sum_{z=0}^{q-1} e(z^2/q)\right| = q^{1/2}.$$

(The way you prove this is by squaring the left-hand side and doing a bit of computations. The moral is that you get perfect square-root cancellation for this quadratic sum — if you put *random* phases then you'd get square-root cancellation, so this is basically the best thing you could hope for.)

So on one segment of length q, we get a value of $q^{1/2}$, which is better thought of as $q^{-1/2} \cdot q$ (since q is the maximum possible value we could have gotten); and when we sum over all segments, we accumulate this saving of $q^{-1/2}$ to get $q^{-1/2}X^{1/2}$. □

§2.2 Rewriting the integral

Now we can rewrite (2.1): for each rational p/q, we consider the piece of the integral on the small arc around p/q — so we define

$$I_{p/q} = \int_{|\theta| \le X^{-1}} |\widehat{f}(p/q+\theta)^2|^2 \widehat{g}(p/q+\theta)\, d\theta$$

(we're being a bit sloppy with the range of integration — we really need something a bit wider than X^{-1} — but this is morally true). Then our integral (2.1) is basically

$$\int_{\mathbb{R}/\mathbb{Z}} |\widehat{f}(\theta)|^2 \widehat{g}(\theta)\, d\theta \approx \sum_{p,q:q<Q} I_{p/q},$$

where Q is some cutoff threshold we'll talk about later (by Lemma 2.2, we only need to consider q which are not too large).

And we'll simplify our life by replacing θ with 0 in our definition of $I_{p/q}$ — we're integrating θ over a fairly narrow interval, so $\widehat{f}$ and $\widehat{g}$ can't change too much on this interval. So we basically have

$$|I_{p/q}| \approx |\widehat{f}(p/q)|^2 \cdot q^{-1/2}X^{1/2} \cdot X^{-1}$$

(where we replace $p/q+\theta$ with p/q and use Lemma 2.2 to bound $\widehat{g}(p/q)$ by $q^{-1/2}X^{1/2}$; the X^{-1} comes from the length of the interval).

So we basically get the sum of the Fourier coefficients of our set A at rational numbers.

§2.3 A density increment

Our strategy is to do a density increment. So we first separate out the 0 term — I_0 is basically the expected count of square differences based on the density of A (i.e., the number of square differences if A were a random set of density α), so we have

$$I_0 \approx \alpha^2 X^2 \cdot X^{1/2}X^{-1} = \alpha^2 X^{3/2}.$$

The basic idea is that we know there's this large positive contribution from I_0, and if we don't have any square differences, then there has to be some significant *negative* contribution — so there should be a large character at some low-height rational, and that will give us a density increment.

So applying the triangle inequality to the fact that the rest of the sum cancels out the 0th term (and multiplying everything by $X^{1/2}$), we have

$$\sum_{p,q:q<Q} |\widehat{f}(p/q)|^2 q^{-1/2} \gtrsim \alpha^2 X^2. \tag{2.2}$$

At this point, we can specify what the cutoff Q should be. On the right-hand side, we have the 0th Fourier coefficient. All the other Fourier coefficients are at most the 0th, and we have a decay factor of $q^{-1/2}$. Once this decay factor is smaller than some power of α, the remaining terms on the left won't matter; so we can take $Q = \alpha^{-C}$ — past this point, we can just apply the trivial estimate $|\widehat{f}(\theta)| \leq |\widehat{f}(0)|$, and the total contribution of these remaining terms will be small.

And now we've done enough work to see Sárközy's result. If we just do the naive bound, we get that one of the Fourier coefficients on the left has to be polynomially large — we get that

$$|\widehat{f}(p/q)|^2 > \alpha^C |\widehat{f}(0)|^2$$

for some $q < \alpha^{-C}$. And now we can do a density increment — we slice everything into progressions mod q, and this large coefficient says that on one of these cosets, you get a density boost

$$\alpha \mapsto \alpha + \alpha^C,$$

And because these are progressions mod q and $q < \alpha^{-C}$, the size of the ground set changes by

$$X \mapsto \alpha^C X.$$

Iterating this gives us a bound of $\alpha = \log^{-c} X$.

§2.4 Idea for an improvement

Now we want to do better, so we want to make a stronger density increment. And doing this better is where Boolean analysis becomes relevant.

The vague idea is that we have Fourier coefficients $\widehat{f}(p/q)$, which correspond to the characters $e(z \cdot p/q)$. And if we have several such characters with coprime q, these characters should behave kind of independently from each other. This is sort of analogous to having linearly independent characters over some finite abelian group; there, you have lemmas which estimate their L^2 mass, or *level d inequalities*.

To explain this, we need to do some Boolean analysis; so we'll step away from this for a while and talk about level d inequalities.

§3 Level d inequalities

Here's the setup. (Somehow, we'll eventually see that these two settings are very similar to each other.) We have a function $f\colon \{0,1\}^n \to \{0,1\}$ on the Boolean cube. Then we can talk about its Fourier coefficients — the characters on the Boolean cube are of the form

$$\chi_S(x_1, \ldots, x_n) = \prod_{i \in S} (2x_i - 1)$$

(i.e., they come from taking some subset of coordinates and multiplying ± 1's on those coordinates), and the Fourier coefficients are the scalar products of f with the corresponding characters, i.e.,

$$\widehat{f}(S) = \mathbb{E}[\chi_S \cdot f].$$

The level d inequality gives you an upper bound on the total weight of Fourier coefficients of small size.

We'll first state and prove a level 1 inequality.

Lemma 3.1 (Chang's lemma)
Let $\mathbb{E}[f] = \alpha$. Then we have

$$\sum |\widehat{f}(\{i\})|^2 \leq C \log(1/\alpha) \cdot \alpha^2.$$

So we have the standard basis on the Boolean cube (the singleton sets), and we're computing the Fourier coefficients with respect to this basis. And Lemma 3.1 says that this sum is much better than what you'd get from the trivial bound $\sum_S |\widehat{f}(S)|^2 = \|f\|_2^2 = \alpha$.

The way to think about this sum is that we have a function on the Boolean cube, and we're taking its *degree* 1 *component* $f^{=1}$ (the degree 0 component is the density of f, and the degree 1 component is its linear bias).

Definition 3.2. We define $f^{=1} = \sum_i \chi_{\{i\}} \widehat{f}(\{i\})$ as the projection of f onto characters of weight 1.

(In other words, we're taking the Fourier expansion of f, and only looking at the characters of size 1.) And the claim is that the L^2 norm of $f^{=1}$ isn't that much larger than that of $f^{=0}$ — we have

$$\|f^{=1}\|_2^2 \leq C \log(1/\alpha) \cdot \|f^{=0}\|_2^2.$$

(The trivial bound would be $\|f\|_2^2 = \alpha$, so we're saving nearly a factor of α.)

This is a very useful inequality, so Dima will give a proof. And this is for level 1; there's an analogous inequality for level d, and that generalization seems to give you much more power somehow. (Chang's lemma is widely used in additive combinatorics; the level d generalization seems to be a stronger thing, but it hasn't been used that much.)

Proof. We want to estimate $\|f^{=1}\|_2^2$. Since $f^{=1}$ is an orthogonal projection of f, this is the same thing as $\langle f^{=1}, f\rangle$. Now by Hölder, we have

$$\langle f^{=1}, f\rangle \leq \|f^{=1}\|_p \|f\|_q$$

(where we'll choose p and q later). And since f is $\{0,1\}$-valued, $\|f\|_q$ is easy to compute — it's just $\alpha^{1/q}$.

So now we need to compute $\|f^{=1}\|_p$. For this, we'll write $f^{=1} = \sum \chi_i \cdot \alpha_i$; then by definition, we have

$$\|f^{=1}\|_p^p = \mathbb{E}_x[f^{=1}(x)^p].$$

We'll assume p is an even integer; then we can expand this expression to get

$$\|f^{=1}\|_p^p = \mathbb{E}_x\left[\left(\sum \chi_i \alpha_i\right)^p\right] = \mathbb{E}_x\left[\sum_{i_1,\ldots,i_p} \alpha_{i_1} \cdots \alpha_{i_p} \chi_{i_1} \cdots \chi_{i_p}\right].$$

And now we can observe that there's lots of cancellation in this sum — $\chi_{i_1} \cdots \chi_{i_p}$ is a product of characters on the Boolean cube, so either it's trivial or it has average 0. For example, if we had two terms (i.e., $p = 2$), then we'd have

$$\mathbb{E}[\chi_i \chi_j] = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

The terms with average 0 don't matter. And to get a trivial term, we need some sort of pairing — i.e., we need $i_1 = i_2$, $i_3 = i_4$, $\dots$, $i_{p-1} = i_p$, or any other permutation of this. This ends up giving

$$\sum_{i_1,\dots,i_{p/2}} \alpha_{i_1}^2 \cdots \alpha_{i_{p/2}}^2 \cdot \frac{p!}{(p/2)! \cdot 2^{p/2}}$$

(where the factorial term is the number of ways to get the same term, coming from the number of ways to group indices into pairs — equivalently, the number of perfect matchings of the complete graph). This works out to roughly

$$p^{p/2} \left(\sum \alpha_i^2\right)^{p/2} = p^{p/2} \|f^{=1}\|_2^p.$$

(The sum of squares of Fourier coefficients of $f^{=1}$ is its L^2 norm.)

So to summarize, we've shown that

$$\|f^{=1}\|_p^p \le p^{p/2} \|f^{=1}\|_2^p.$$

Now we can plug this into our original bound $\|f^{=1}\|_2^2 \le \|f^{=1}\|_p \cdot \alpha^{1/q}$ to get

$$\|f^{=1}\|_2^2 \le \sqrt{p} \cdot \|f^{=1}\|_2 \cdot \alpha^{1/q}.$$

Thankfully, one factor of $\|f^{=1}\|_2$ cancels out, and we get

$$\|f^{=1}\|_2 \le \sqrt{p} \cdot \alpha^{1-1/p}$$

(here we're using the fact that $1/p + 1/q = 1$, because p and q are Hölder exponents).

So we've got a bound, and now we want to optimize it. The best way to optimize it turns out to be to make $\alpha^{-1/p}$ irrelevant. So we take $p = \log(1/\alpha)$ to make $\alpha^{-1/p}$ constant; then we get

$$\|f^{=1}\|_2^2 \le \log^{1/2}(1/\alpha) \cdot \alpha,$$

which is what we wanted (the expression we're trying to bound is $\|f^{=1}\|_2^2$). □

This was some clever computation, and the point is that it gives you a much better bound than the trivial one. And this bound is actually sharp — for example, it's sharp for the majority function

$$f_{\mathsf{Maj}} = \begin{cases} 1 & \text{if } \sum x_i > n/2 + k\sqrt{n} \\ 0 & \text{otherwise.} \end{cases}$$

This bound is very useful, and this computation will pop up again, closer to the end.

Now we'll state a generalization of this inequality, which is also good to know — a similar inequality, but for degree d rather than degree 1. (We basically just replace 1 by d in the correct places.)

Definition 3.3. We define $f^{=d} = \sum_{|S|=d} \chi_S \widehat{f}(S)$.

Lemma 3.4 (Level d inequality)

For $d \le \log(1/\alpha)$, we have

$$\|f^{=d}\|_2^2 \le \frac{C \log(1/\alpha)^d}{d} \cdot \|f^{=0}\|_2^2.$$

This is a generalization of the level 1 inequality which is very useful. Dima won't tell us the proof, which is more complicated.

§4 The connection

To see the connection between the two settings, let's first try to do some wishful setting. Let's say we have our sum (2.2) from before. For simplicity, we'll only consider the case where q is prime, so we have

$$\sum_{q \le Q \text{ prime}} |\widehat{f}(p/q)|^2 q^{-1/2} \gtrsim \alpha^2 X^2. \tag{4.1}$$

(Thinking about only the case where q is prime is reasonable — there are a lot of primes, so they should constitute a decent chunk of the sum.)

And the idea is that $\widehat{f}(p/q)$ in this setting is kind of analogous to $\widehat{f}(\{i\})$ in the Boolean setting. Why? Here we're taking an integral of f against $e(p/q \cdot x)$, while in the Boolean setting we're taking an integral against $e(e_i \cdot x)$ (where $e_i \in \{0,1\}^n$ is the ith basis vector). And the only thing we used in the proof of the level 1 inequality was that these basis vectors e_i are linearly independent — so when we expanded out the pth power, we got cancellation unless the coefficient of each summed to 0. And the ratios p/q are also kind of like that — if we have a sum

$$\frac{p_1}{q_1} \pm \frac{p_2}{q_2} \pm \cdots,$$

then it's usually nonzero. This tells you that you can hope to get cancellation and make things work.

If we *could* get a level 1 inequality, what would it say? We'd like to say that

$$\sum_{q \le Q \text{ prime}} |\widehat{f}(p/q)|^2 \le C \log(1/\alpha) \cdot \|f^{=0}\|_2^2 \tag{4.2}$$

(since the left-hand side is what we think of as $\|f^{=1}\|_2^2$).

§4.1 A better density increment

If we can prove (4.2) then we win — we get an upper bound of $C \log(1/\alpha) \cdot \alpha^2 X^2$ for the unweighted sum, while in (4.1) we have a weighted sum with an extra $q^{-1/2}$ factor. So we can consider some dyadic range which dominates the sum (4.1); then we get

$$\sum_{q \sim q_0} q_0^{-1/2} |\widehat{f}(p/q)|^2 \gtrsim \frac{1}{\log(1/\alpha)} \cdot \alpha^2 X^2$$

(where $q \sim q_0$ means that q is in the dyadic interval around q_0, and the $\log(1/\alpha)$ term is because we did dyadic pigeonhole).

On the other hand, we have an upper bound on this sum from (4.2) (the level 1 inequality), so we get

$$\sum_{q \sim q_0} q_0^{-1/2} |\widehat{f}(p/q)|^2 \le q_0^{-1/2} \cdot C \log(1/\alpha) \cdot \alpha^2 X^2.$$

And the only way this can happen is if q_0 is very small, so we get $q_0 < \log(1/\alpha)^4$. This is the reason why we get a better bound this way — the level 1 (or level d) inequality says we're within a log factor of the best possible bound $\alpha^2 X^2$, and this kind of forces q_0 to be very small (compared to the trivial upper bound $q_0 < Q = \alpha^{-C}$).

This is already enough to get a Behrend-type bound — for the new density increment, now we have a large Fourier coefficient $|\widehat{f}(p/q)| > \log^{-C}(1/\alpha) \cdot |\widehat{f}(0)|$, where $q < \log^C(1/\alpha)$. So when we do our density increment, we get

$$\alpha \mapsto \alpha(1 + \log^{-C}(1/\alpha)),$$
$$X \mapsto X \log^{-C}(1/\alpha).$$

And this gives Behrend-type bounds.

§5 A level d inequality for the arithmetic setting

This is the reason why if we had a level 1 inequality as in (4.2) (or more generally, a level d inequality) for $\widehat{f}(p/q)$, we'd be happy; now Dima will explain what we have to do to actually get it.

The idea is that we have a function on $[X]$, and we want to model this set $[X]$ by a large cyclic group — for a set of primes $\mathcal{Q}$, we'll use the cyclic group

$$G = \prod_{q \in \mathcal{Q}} (\mathbb{Z}/q\mathbb{Z}).$$

The idea is that we want to relate the characters at p/q with basis vectors in some product space (like $\{i\}$ in the Boolean setting), and this is the natural product space to consider. So we want to model $[X]$ by this large product space; note that $|G| \gg X$ for the set of primes $\mathcal{Q}$ we'll care about, but this is still a model in some sense.

And if we have a function f on $[X]$, we want to model it by a function F on G with the same low-degree Fourier coefficients, in some sense — we somehow want $\widehat{f}(p/q)$ to be related to $\widehat{F}(\{q\} \cdot p)$ (this notation is a bit questionable, but the point is that G is a product space, so it has Fourier coefficients in the basis directions corresponding to each component).

And we want a level d inequality for f to correspond to a level d inequality for F.

At the moment, there are two issues.

(1) What is F?

(2) The group $G = \prod_q (\mathbb{Z}/q\mathbb{Z})$ is not the Boolean cube — in particular, the components in this product are quite large (i.e., $|\mathbb{Z}/q\mathbb{Z}| \gg 1$). So it's not immediately clear what the analog of the level 1 (or level d) inequality should be for functions on this space.

But both of these issues can be resolved; that's what we'll talk about now.

§5.1 Issue 2 — a global level d inequality

We'll start with the second issue. We'll describe this in a general setting: you have a collection of finite sets $\Omega_1, \dots, \Omega_n$ (in our setting, Ω_i is the ith cyclic group $\mathbb{Z}/q\mathbb{Z}$ in our product), and we write $\Omega_{[n]} = \Omega_1 \times \cdots \times \Omega_n$.

If we have a function $f\colon \Omega_{[n]} \to \{0,1\}$, we'd like to talk about the level d component of f. In the Boolean setting, the level 1 component of f was the sum of characters of weight 1. Here the Ω_i are arbitrary finite sets, so there are no characters. But we can still define the level 1 component of f as a linear combination of functions that depend only on one coordinate — i.e.,

$$f^{=1}(x) = \sum_{i=1}^{n} g_i(x_i)$$

where each $g_i\colon \Omega_i \to \mathbb{R}$ is a function on the ith space in our product with $\mathbb{E}[g_i] = 0$. (We say $\mathbb{E}[g_i] = 0$ because we can subtract out the mean, which we'd consider the level 0 part.) You can also define the level d part of f in a similar way.

We want a bound on $\|f^{=1}\|_2^2$ — we have a function f on our product space with mean $\mathbb{E}[f] = \alpha$, and we're trying to find an upper bound on its level 1 part.

> **Question 5.1.** Can we say that $\|f^{=1}\|_2^2 \le C \log(1/\alpha) \cdot \alpha^2$?

If we could show this, then we'd be happy. It's important that the constant C should not depend on the sizes of the sets Ω_i — in our application, we're taking primes q up to α^{-C}, so the sizes of these sets are in principle comparable to our density.

Unfortunately, it turns out that in this generality, you have no chance — this statement is false. There's actually a very simple counterexample.

Example 5.2

Suppose that f only depends on one coordinate, e.g., $f(x) = \mathbf{1}_{x_1=c}$ for some $c \in \Omega_1$. Also suppose that $|\Omega_1| = \alpha^{-1}$, so that $\mathbb{E}[f] = \alpha$.

Then by design, we have $f^{=1} \approx f$ (since the level 1 component is the part of f which only depends on one coordinate, but f itself only depends on one coordinate; the $\approx$ is because we subtract out the 0th component), so $\|f^{=1}\|_2^2 \approx \|f\|_2^2$. This means we only get the trivial bound α, rather than α^2.

So things look kind of green. But it turns out you *can* actually prove such an inequality if you know something more about the function. This counterexample is a function that depends on just one coordinate. But in our setting, we're okay with functions like this — if our function only depends on the residue mod q, then we directly get a huge density increment mod q. And it turns out that if we forbid functions like this, then an analog of this inequality *is* true.

Lemma 5.3 (Global level 1 inequality)

If $f : \Omega_{[n]} \to \{0, 1\}$ is global, then

$$\|f^{=1}\|_2^2 \leq C \log(1/\alpha) \cdot \|f^{=0}\|_2^2.$$

Intuitively, *global* means that the function doesn't depend on just a few coordinates. More precisely, if we take our product space and restrict some number of coordinates, then the density of the function shouldn't increase significantly.

Definition 5.4. We say f is global if for all $S \subseteq [n]$ and $y_S \in \Omega_S$, we have

$$\mathbb{E}[f(x_S \to y_S)] \leq K^{|S|}\mathbb{E}[f].$$

What this notation means is that we choose a set of coordinates S, and a vector y_S representing our entries in these coordinates; then we get a function on the smaller product space $\Omega_{[n]\setminus S}$ where we take some input and plug in y_S for the coordinates on S. And globalness says that this restriction isn't supposed to increase the average by too much. For example, if $K = 10$, this says that if I fix one coordinate then the density doesn't increase by a factor of more than 10; if I fix two then it doesn't increase by a factor of more than 100; and so on.

We can assume the function in our application has this property — if it doesn't, then we can just restrict to the corresponding arithmetic progression, and the density of our set will increase a *lot* there. So this is a fine thing to assume; and under this assumption we do get a level 1 inequality.

A level d version also exists; it says that

$$\|f^{=d}\|_2^2 \leq \frac{C \log(1/\alpha)^d}{d} \cdot \|f^{=0}\|_2^2$$

for $d < \log(1/\alpha)$.

Dima won't say too much about how to prove this, but a similar strategy works. Here's the proof idea for $d = 1$. First, you prove that if f is global, then $f^{=1}$ is also global. (*A priori* these are two not very related

functions, but it turns out that they remember enough about each other that they're global at the same time.) Then the Hölder argument works as before — you raise things to the power of p and get some linear combination, and globalness kind of makes sure that everything works out nicely. (But the first step is a key deivation that takes a while — it's a few pages, but it's clever.)

So this fixes one problem — now we do have a nice inequality for our product group.

§5.2 Issue 1 — constructing the model

The final step is to relate functions on $[X]$ to functions on this product space G. First, we have a map $\psi\colon [X] \to G$ where $\psi(x)$ is the tuple of reductions of x mod q — i.e.,

$$\psi(x) = (x \bmod q)_{q\in\mathcal{Q}}.$$

So we have this huge box G, and $[X]$ corresponds to some kind of mysterious subset inside this box. We need to define a function F on the whole group which extends a function $f\colon [X] \to \{0,1\}$. It turns out you don't need to be smart, and you can just extend by 0's — we define

$$F(y) = \begin{cases} f(x) & \text{if } \psi(x) = y \\ 0 & \text{otherwise.} \end{cases}$$

So this is our model function.

We want to apply the level d inequality to F. But if we try to do this, we run into trouble. The problem is that F is extremely sparse — it lives on a really tiny subset of our product space. Explicitly, if we let $Y = \psi([X]) \subseteq G$, then we have

$$\mathbb{E}[F] = \mu(Y) = \frac{X}{|G|},$$

and in the setting we care about, this is extremely small (G is a product of cyclic groups of size up to α^{-C}, and there are roughly α^{-C} primes in this product, so $|G| = \exp(\alpha^{-C})$; this is extremely large compared to X). So we can't really apply the level d inequality to F, because its density is so small that the result would be meaningless.

On the other hand, F has reasonable density *relative* to Y (its relative density is α). So what we want to prove is a level d inequality relative to a uniform subset — basically, it turns out that as long as Y is kind of nicely distributed (with respect to low-degree characters), everything still works out nicely. We'll formulate this statement for the Boolean cube (in order to only have one modification at a time).

Lemma 5.5 (Relative level 1 inequality)

Suppose that $Y \subseteq \{0,1\}^n$ has small low-degree Fourier coefficients, and let $f\colon Y \to \{0,1\}$ be such that $\mathbb{E}_Y[f] = \alpha$. Then

$$\|f^{=1}\|_2^2 \le C\log(1/\alpha)\cdot\|f^{=0}\|_2^2.$$

(Here the 2-norms treat f as a function on the entire Boolean cube.)

We won't be precise with what 'small' and 'low-degree' mean; but once you have this and combine it with the global level d inequality, you get some statement which is sufficient for our application.

Proof sketch. The idea is that you run the same proof as for the ordinary level 1 inequality and see what happens; and you'll need to use Y in one place.

As before, we want to compute

$$\|f^{=1}\|_2^2 = \langle f^{=1}, f\rangle.$$

In the original proof, we applied Hölder here. But now we first use the observation that this is the same as $\langle f^{=1}\mathbf{1}_Y, f\rangle$ and *then* apply Hölder; this gives

$$\|f^{=1}\|_2^2 = \langle f^{=1}\mathbf{1}_Y, f\rangle \le \|f^{=1}\mathbf{1}_Y\|_p \|f\|_q.$$

When we try estimating the first guy, if we write $f^{=1} = \sum \alpha_i \chi_i$, then we have

$$\|f^{=1}\mathbf{1}_Y\|_p^p = \mathbb{E}\left[\left(\sum \alpha_i \chi_i(x)\right)^p \mathbf{1}_Y(x)\right].$$

When we expand this, we get

$$\mathbb{E}\left[\sum_{i_1,\dots,i_p} \alpha_{i_1}\cdots\alpha_{i_p}\chi_{i_1}\cdots\chi_{i_p}\mathbf{1}_Y\right].$$

There are two cases, depending on whether the product of characters vanishes or not (i.e., whether $\chi_{i_1}\cdots\chi_{i_p}$ is nontrivial or 1). If it multiplies to some nontrivial character χ_S, then we get a term of $\mathbb{E}[\chi_S\mathbf{1}_Y]$. And this is tiny because of the uniformity assumption on Y; so nontrivial products don't contribute much to the sum.

Meanwhile, the terms with trivial characters give roughly

$$\sum_{i_1,\dots,i_{p/2}} \alpha_{i_1}^2\cdots\alpha_{i_{p/2}}^2 \cdot p^{p/2}\cdot \mathbb{E}[\mathbf{1}_Y]$$

(this is the same thing as before, but with the extra term of $\mathbb{E}[\mathbf{1}_Y]$). So we end up with roughly

$$\|f^{=1}\|_p^p \le \|f^{=1}\|_2^p \cdot p^{p/2}\cdot \mu(Y).$$

When we plug this into the Hölder expression, we're taking the pth root of this, so we get

$$\|f^{=1}\|_2^2 \le \sqrt{p}\cdot \|f^{=1}\|_2 \cdot \mu(Y)^{1/p}\cdot \alpha^{1/q}\mu(Y)^{1/q}$$

(where the term $\alpha^{1/q}\mu(Y)^{1/q}$ comes from $\|f\|_q$). And this makes us happy — we get

$$\|f^{=1}\|_2 \le \sqrt{p}\cdot\alpha^{1-1/p}\cdot\mu(Y) = \sqrt{p}\cdot\alpha^{-1/p}\cdot\|f^{=0}\|_2,$$

as before. The key observation was that the $\mu(Y)$ in the L^p norm kind of saved us — so it was crucial to throw in the $\mathbf{1}_Y$ before using Hölder, because that gave a much better bound. □