Math 216A Notes, Week 10
Scribe: Oliver Thistlethwaite

Disclaimer: These notes are not nearly as polished (and quite possibly not nearly as correct) as a published paper. Please use them at your own risk.

1. The Proof of Roth’s Theorem

1.1. The Fourier Transform and Counting Progressions. Our goal here is to complete the proof we outlined last week that if \( A \subseteq \{1, \cdots, N\} \) with \( |A| \geq \delta N \), we wanted to show \( A \) has nontrivial solutions to \( x + y = 2z \).
Recall the Fourier transform:
\[
\hat{f}(r) = \sum_{k=0}^{N-1} f(k) e^{-2\pi ikr/N}.
\]
We (still) have the following useful properties.

(1) (orthogonality)
\[
\frac{1}{N} \sum_{k=0}^{N-1} e^{-2\pi ik/N} = \begin{cases} 
0 & \text{if } k \equiv 0 \pmod{N}, \\
1 & \text{otherwise}
\end{cases}
\]

(2) (Parseval’s identity)
\[
\sum_{k=0}^{N-1} |f(k)|^2 = \frac{1}{N} \sum_{k=0}^{N-1} |\hat{f}(r)|^2
\]

(3)
If \( h(x) = \sum_k f(k)g(k-x) \),
\[
\hat{h}(r) = \hat{f}(r)\overline{g(r)}
\]
and \( |\hat{f}(k)| |\hat{g}(k)| \leq \sum_x \left| \sum_y f(y)g(y-x) \right| \)

As we saw last time, we can use the Fourier transform to count the number of progressions in a subset of \( \mathbb{Z}_n \). There’s a problem, though, in that what we’re counting are progressions in \( \mathbb{Z}_n \), which aren’t necessarily progressions in \( \mathbb{Z} \). To work around this, we will restrict our attention to a class of progressions that we know remain progressions in the integers.
Let $M_A = A \cap \left[ \frac{N}{3}, \frac{2N}{3} \right]$. Note that if $x, z, y$ is a $\mathbb{Z}_N$ progression (ie $x + y = 2z$) and $x \in M_A$ and $z \in M_A$, it is also a $\mathbb{Z}$-progression. The number of such progressions is

$$= \sum_{x, z \in M_A \text{ and } y \in A} \chi(x + y - 2z = 0)$$

(by 1) $$= \sum_{x, z \in M_A \text{ and } y \in A} \frac{1}{N} \sum_k e^{-\frac{2\pi i}{N} (x + y - 2z)k}$$

$$= \sum_k \frac{1}{N} \left( \sum_{x \in M_A} e^{-\frac{2\pi i}{N} xk} \right) \left( \sum_{y \in A} e^{-\frac{2\pi i}{N} yk} \right) \left( \sum_{z \in M_A} e^{-\frac{2\pi i}{N} z(-2k)} \right)$$

Now define $1_A$ to be the function that is 0 on $A$ and 1 off $A$. We get the number of progressions is

$$\frac{1}{N} \sum_{k=0}^{N-1} \hat{1}_{M_A}(k) \hat{1}_A(k) \hat{1}_{M_A}(-2k)$$

1.2. Random-Like Sets Have Many Progressions. Note for $k = 0$ we have,

$$\hat{1}_{M_A}(0) = \sum_{X=0}^{N-1} 1_{M_A}(X) = |M_A|.$$

We can peel the $k = 0$ term off from the above, writing the total number of progressions as

$$\delta |M_A|^2 + \frac{1}{N} \sum_{k=0}^{N-1} \hat{1}_{M_A}(k) \hat{1}_A(k) \hat{1}_{M_A}(-2k).$$

Here’s one way of thinking about this expression. In $\mathbb{Z}_N$, there are $|M_A|^2$ total progressions such that both $x$ and $z$ are in $M_A$. If $A$ was a random set having density $\delta$ would expect about a $\delta$-fraction of these progressions to also have $y \in A$. This corresponds to the first term in the above expression, and is quite large (proportional to $N^2$). The remaining sum is in a sense an error term accounting for the fact that $A$ isn’t random. As long as the error is much smaller than the main term, the whole expression will be large.

Note that if $k \neq 0$ we have

$$\hat{1}_{M_A}(0) \leq \sum_{x=0}^{N-1} |1_{M_A}(x)| = |M_A|.$$

The definition of “random-like” that we will use is that actually all of the nonzero Fourier coefficients are much smaller than the $k = 0$ one.

**Definition.** A set $A$ is $\epsilon$-uniform if $|1_A(k)| \leq \epsilon N$ for all $k \neq 0$. 
If $A$ is $\epsilon$-uniform then,

$$\frac{1}{N} \sum_{k=0}^{N-1} \hat{1}_{MA}(k) \hat{1}_A(k) \hat{1}_{MA}(-2k)$$

$$\leq \frac{\epsilon N}{N} \sum_{k=1}^{N-1} |\hat{1}_{MA}(k)||\hat{1}_{MA}(-2k)|$$

(by Cauchy-Schwarz) $\leq \epsilon \left( \sum_{k=0}^{N-1} |\hat{1}_{MA}(k)|^2 \right)^{\frac{1}{2}} \left( \sum_{k=0}^{N-1} |\hat{1}_{MA}(-2k)|^2 \right)^{\frac{1}{2}}$

(if $N$ is odd) $\leq \epsilon \left( \sum_{k=0}^{N-1} |\hat{1}_{MA}(k)|^2 \right)$

$$= \epsilon M \left( \sum_{k=0}^{N-1} |1_{MA}(k)|^2 \right)$$

(by Parseval) $= \epsilon N|M_A|$

So an $\epsilon$-uniform set has at least $\delta|M_A|^2 - \epsilon N|M_A| = |M_A|(|\delta|M_A| - \epsilon N)$ good progressions. What we’ve shown so far is that if $|A| = \delta N$, one of the following holds.

1. $|M_A| < \frac{\delta N}{4}$
2. $A$ is not $\frac{\delta^2}{8}$-uniform.
3. $A$ contains at least $\frac{\delta^3 N^2}{32}$ good progressions.

There are only $\delta N$ trivial progressions, so in the last case if $\frac{\delta^3 N^2}{32} > \delta N$, we’re done.

1.3. The Non-Uniform Case: Density Increment. We still need to handle the first two cases above. We consider the first case first. If $M_A$ contains at most $\frac{\delta N}{4}$ elements of $A$, then by pigeonhole we must have either

$$|A \cap \{1, \ldots, N/3\}| \geq \frac{3\delta N}{8} = \frac{N}{3} \left( \delta + \frac{\delta}{24} \right)$$

or

$$|A \cap \{2N/3 + 1, \ldots, N\} | \geq \frac{3\delta N}{8} = \frac{N}{3} \left( \delta + \frac{\delta}{24} \right)$$

In other words, one of the two progressions $\{1, \ldots, N/3\}$ and $\{2N/3 + 1, \ldots, N\}$ has a higher density of elements of $A$ than the original interval. A key fact here is that the equation $x + y = 2z$ is translation invariant – the problem of looking for progressions in a subset of $\{1, \ldots, N/3\}$ is exactly the same as in $\{2N/3 + 1, \ldots, N\}$. So in either case we’ve reduced down to a new case of our original Roth’s theorem where $N$ is smaller by a factor of 3, but the density has increased by a constant factor.

The more difficult part is showing that in fact the same thing happens in the second case. The difference is that instead of showing that $A$ has increased density on an interval, we’ll show it has increased density on a progression. The idea is that if we show this, we’ve again reduced down to the original problem with a higher density, since (for example), the map $x \to \frac{x-3}{4}$ shows that looking for progressions in $\{3, 7, \ldots, 4N + 3\}$ is the same as looking for progressions in $\{1, \ldots, N\}$...almost.
The problem, once again, is the difference between \( \mathbb{Z}_N \)-progressions and progressions in the integers. Division by 4 can hypothetically lead to wraparound that turns a good progression into something which isn’t an integer progression anymore. One way around this is the following definition and technical lemma.

**Definition.** A progression in \( \mathbb{Z}_N \) of length \( L \) and difference \( d \) is called **nonoverlapping** if \( dL < N \).

Note that a nonoverlapping \( \mathbb{Z}_N \) progression is a union of at most 2 \( \mathbb{Z} \) progressions.

**Lemma.** (technical) If \( B' \) is a nonoverlapping \( \mathbb{Z}_N \) progression on which \( A \) has density at least \( \delta + \epsilon' \) then there exists a \( \mathbb{Z} \) progression \( p \) with \( |p| \geq 1/2 \epsilon' |B'| \) such that \( A \) has density at least \( \delta + 1/2 \epsilon' \) on \( p \).

The point here is that this lets us just look for non-overlapping progressions in \( \mathbb{Z}_N \) where \( A \) has increased density.

**Proof.** Write \( B' = p_1 \sqcup p_2 \) where \( |p_1| \leq |p_2| \) and \( p_1 \) and \( p_2 \) are \( \mathbb{Z} \) progressions. If \( |p_1| \leq 1/2 \epsilon' |B'| \) then

\[
|A \cap p_2| \geq (\delta + 1/2 \epsilon') |A| \\
\geq (\delta + 1/2 \epsilon') |p_2|.
\]

If \( |p_1|, |p_2| \geq 1/2 \epsilon' |B'| \) then \( A \) has density \( \delta + \epsilon' \) on one of them. \( \square \)

**Lemma.** (big density increment) If \( A \) is not \( \epsilon \)-regular ie \( |1_A(r)| \geq \epsilon N \) for some \( r \neq 0 \), there exists a nonoverlapping \( \mathbb{Z}_N \)-progression \( B' \) of length at least \( \sqrt{N}/25 \) such that \( |A \cap B'| \geq (\delta + \frac{1}{4} \epsilon') |B'| \).

**Proof.** Let

\[
f_A(x) = 1_A(x) - \delta = \begin{cases} 
1 - \delta & \text{if } x \in A \\
-\delta & \text{if } x \notin A
\end{cases}
\]

then \( \hat{f}_A(k) = \hat{1}_A(k) \) for all \( k \neq 0 \) and \( \hat{f}_A(0) = 0 \).

The key idea is that if \( B' = B + X \) (a translated \( B \)) then \( |A \cap (B + X)| \geq (\delta + \frac{1}{4} \epsilon') |B| \) is true if and only if

\[
\sum_y f_A(y) 1_B(y - x) \geq \frac{1}{4} \epsilon |B|.
\]

By our bound (3) from earlier, we can make this happen by making \( 1_B \) share a large Fourier coefficient with \( f \).

**Claim.** There exists a \( d \) such that \( 0 < d \leq \sqrt{N} \) and \( |rd| \leq \sqrt{N} \) (here \( |.| \) denotes the distance from 0 mod \( N \)).

**Proof.** (of claim)
Consider the map \( x \rightarrow (x, xd \pmod{N}) \) from \( \mathbb{Z}_N \) to \([1, N] \times [1, N] \). What this lemma is saying is that if we divide the above box into \( N \) smaller boxes, at least one point in \( \mathbb{Z}_N \) gets mapped to either the lower left or upper left. But if this doesn’t happen, there must be two points \( x_1 \) and \( x_2 \) mapped to in the same box. Then either \( x_1 - x_2 \) or \( x_2 - x_1 \) gets mapped to the one of the boxes by construction (our map is linear). \( \square \)

Take the \( d \) from the claim and let \( B \) be the progression of length \( \lfloor \sqrt{N/8\pi} \rfloor \) given by \( \{\cdots , -2d, -d, 0, d, 2d, \cdots \} \) (with endpoints chosen so the progression is nearly symmetric around 0). By construction this is nonoverlapping.

**Claim.** \( |\widehat{1}_B(r)| \geq \frac{1}{2} |B| \).

**Proof.** (of claim)

\[
|\widehat{1}_B(r)| - |B| = \left| \sum_x 1_B(x)[e^{-2\pi i x r/N} - 1] \right| \\
\leq \sum_{|\ell| \leq \frac{1}{2}|B|} \left[ e^{-2\pi i \ell d r/N} - 1 \right]
\]

Then \( rd \leq \sqrt{N} \) and \( |\ell| \leq \lfloor \sqrt{N/16\pi} \rfloor \). Each term is less than or equal to \( |e^{-2\pi i \ell r/N} - 1| \leq \frac{1}{2} \). \( \square \)

Applying property 3 we have,

\[
(\epsilon N) \cdot \frac{1}{2} |B| \leq |\widehat{f}_A(r)||\widehat{1}_B(r)| \\
\leq \sum_x |\sum_y f_A(y)1_B(y-x) |
\]

Thus by the pigeonhole principle, there exists an \( x \) such that \( \frac{e|B|}{2} \leq \sum_y f_A(y)1_B(y-x) \). \( \square \)

### 1.4. The Endgame, and Further Remarks.

In case 2, \( A \) had a \( \mathbb{Z}_N \)-progression of length at least \( \sqrt{N/25} \) where \( A \) has density \( \delta + \frac{\delta^2}{62} \). This implies there exists a \( \mathbb{Z} \)-progression of length \( \frac{\delta^2}{1000} \sqrt{N} \) where \( A \) has density \( \delta + \frac{\delta^2}{64} \).

If we iterate \( \frac{64}{25} \) times then either we eventually get a uniform set (in which case we’re done), or the density becomes \( 2\delta \). If we iterate \( \frac{64}{25} \) more times then again either we are done or the
density reaches $4\delta$. If we iterate $\frac{64}{\delta}(1 + \frac{1}{2} + \cdots + \frac{1}{3\pi})$ times then either we’re done or the density is at least $2^\ell \delta$. But then if we iterate more than $128\delta^{-1}$ times, the density becomes at least 1 which is impossible. Therefore we must have encountered a uniform set before then.

We need $N$ large enough so after $128\delta^{-1}$ iterations $|B'| \geq \frac{32}{\delta}$ (to make sure that when we finally find our uniform set it’s large enough that we actually have a non-trivial progression on it). After some calculations we see $N \geq \exp(\exp(c\delta^{-1}))$ works.

It is still an open question what the best bound on $N$ is here. The best known bound is due to (Sanders, 2010) which states if $A \leq \{1, \cdots, N\}$ with $|A| \geq N(\log N)^{1+o(1)}$ then $A$ has a 3 term arithmetic progression. One related conjecture is

**Conjecture.** (Erdos-Turan) If $\sum_{x \in A} \frac{1}{x} = \infty$ then $A$ has arbitrarily long arithmetic progressions.

Note that Sanders’ bound is tantalizingly close here – if we could replace $1 + o(1)$ by any number bigger than 1, the conjecture would be proven. A special case of this conjecture is the Green-Tao theorem that the primes contain arbitrarily lone arithmetic progressions.

2. **Lower Bounds on Roth’s Theorem**

We can rewrite the version of Roth’s theorem we proved as stating that if $A \subseteq [1, N]$ and $|A| > \delta N$ where $\delta > \frac{C}{\log(\log N)}$ then $A$ contains a 3 term progression.

We now switch our attention to lower bounds, that is to say showing the existence of a large subset of $[1, N]$ without a 3 term arithmetic progression. We want $\delta \to 0$ as slowly as possible.

2.1. **Attempt 1: The First Moment.** Take a random $A$ where for each $x \in [1, N]$ we have $P(x \in A) = \delta$ (assuming $\delta N \to \infty$, $A$ almost surely has size approximately $\delta N$). Note that

$\{1, \ldots, N\}$ contains approximately $\frac{(N)}{2} \approx \frac{N^2}{4}$ nontrivial progressions. ($(\frac{N}{2})$ choices for $x$ and $y$ and a $\frac{1}{2}$ chance that $\frac{x+y}{2} \in \mathbb{Z}$.) Each progression fails with probability $\delta^3$ so $P(\text{fail}) \leq \delta^3 \frac{N^2}{4}$. By Markov’s inequality, we’re done if this probability to be strictly less than 1, which will happen if $\delta < N^{-\frac{1}{2}}$.

2.2. **Attempt 2: Deletion.** Suppose the number of progressions in $A$ is not too large. We can form a new set $A'$ from $A$ by deleting one $X$ from each progression. So $|A'| = |A| - \{\text{num of deleted progressions}\}$, and by construction $A'$ is progression free.

By linearity of expectation, we have $E(|A'|) = \delta N - \delta^3 \frac{N^2}{4}$. If we take $\delta \approx CN^{-\frac{1}{2}}$ for some small $C$ then $E(|A'|) > C'N^{\frac{1}{2}}$, so there must be a set this large without progressions.

2.3. **Attempt 3: The Local Lemma.** We would like to apply the local lemma, taking the bad events to be progressions that could be in $A$. Say $P_1 \sim P_2$ if $|P_1| \cap |P_2| \neq 0$. The maximum degree of the dependency graph is approximately $CN$ for some $C$. To use the local lemma, we need to have $c\delta^3(1 + CN) < 1$, so it seems like we could get away with taking $\delta = C'N^{-\frac{1}{2}}$.

However, this method doesn’t actually work! The problem is that while the probability given by the local lemma is nonzero for any fixed $N$, it is also very small. So small, in fact, that it’s less than the probability we have approximately $\delta N$ elements in $A$. So we can’t use the
Local Lemma directly to guarantee a set which is both large and progression-free. (Note that a similar argument would “prove” there are large sets without 1 term progressions).

2.4. Attempt 4: Go Greedy. Let us construct a progression-free set as follows. Take 0 ∈ A and 1 ∈ A, and for x ≥ 2, x ∈ A if and only if x does not form a progression with two numbers already in A. So A = 0, 1, 3, 4, 9, 10, 12, · · · . If we look at A in base 3 we obtain A = 0, 1, 10, 11, 100, 101, 110, · · · . So we see

\[ A = \{ x \mid x \text{ has no 2 in base 3}\}. \]

The density of A is \( N^{\log_3 2 - 1} \approx N^{-0.37} \).

2.5. Behrend’s Idea: Structured Sets. Behrend in 1946 had the idea to start in \( \{0, \cdots, m-1\}^d \) with \( m \) and \( d \) to be determined later. A 3-term arithmetic progression here means \( x + y - 2z = (0, \cdots, 0) \). A quick observation is if \( x, y, z \) are in an arithmetic progression then they are colinear. Hence if S is any subset of a sphere then S has no 3-term arithmetic progression. There are \( m^d \) points on each sphere \( \{x_1^2, \cdots, x_d^2 = C\} \) for some \( 0 \leq C \leq d(m-1)^2 \). Using the pigeonhole principle, we see there is some sphere with \( \frac{m^{d-2}}{d} \) lattice points.

Let \( S \) be those points. We can define a projection

\[ f(x_1, \cdots x_d) = X_1 + (2m)x_2 + (2m)^2x_3 + \cdots + (2m)^{2d}x_d. \]

This is one-to-one on \( S \), so \( |f(S)| = |S| > \frac{m^{d-2}}{d} \).

Claim. \( f(S) \) has no non-trivial arithmetic progressions

Proof. Note that \( f(x) + f(y) - 2f(z) = 0 \) means

\[ (x_1 + y_1 - 2z_1) + (2m)(x_2 + y_2 - 2z_2) + \cdots + (2m)^{d-1}(x_d + y_d - 2z_d) = 0 \]

Since \( x, y, z \) was not originally a progression (\( S \) has no non-trivial progressions), there must be some largest \( j \) with \( x_j + y_j - 2z_j \neq 0 \). Then

\[ |f(x) + f(y) - 2f(z)| \geq (2m)^{j-1} \cdot 1 - (2m)^{j-2} \cdot (2m - 2) \cdots \]

\[ \geq (2m)^{j-1} - (2m - 2)((2m)^{j-2} + (2m)^{j-3} + \cdots \]

\[ = (2m)^{j-1} - (2m - 2) \cdot \frac{(2m)^{j-2}}{1 - \frac{1}{2m}} \]

\[ = (2m)^{j-1} - (2m - 2) \cdot \frac{(2m)^{j-1}}{2m - 1} > 0. \]

\[ \square \]

We also note that \( f(S) \subseteq [0, (m-1)(1 + (2m) + \cdots + (2m)^{d-1})] \subseteq [0, (2m)^d] \).

This shows for any \( m \) and \( d \) there exists an \( A \subseteq [0, (2m)^d] \) of size greater than \( \frac{m^{d-2}}{2} \) without any 3 term arithmetic progressions. To maximize \( |A| \) take \( d = \sqrt{\log N} \) and \( m = \lfloor \frac{1}{2} N^{\frac{3}{2}} \rfloor \).

Then \( |A| \geq N e^{-C \sqrt{\log N}} \) for some \( C \). In particular, \( |A| > N^{1-\epsilon} \) for any positive \( \epsilon \) and large enough \( N \).
This bound (which actually came before Roth’s Theorem) suggests that in a way the complicated iteration proof of Roth’s theorem may be necessary, since many more natural hypothetical proofs would lead to a bound that was ”too good to be true” (smaller than Behrend’s lower bound). Note also the contrast with Ramsey’s theorem: For many graph-Ramsey problems, random constructions (or nearly random constructions) are the best known bound, but for this problem the best known sets are very structured.

3. Miscellaneous Notes

3.1. The Cap Set Problem. Instead of working over \( \mathbb{Z} \) or \( \mathbb{Z}_N \), we can also ask these questions over other groups. For example, we could ask what’s become known as the ”Cap Set” problem: What is the largest \( A \subseteq \mathbb{Z}_N^3 \) without a solution to \( x + y = 2z \)?

Note that we can rewrite this equation as \( x + y + z = 0 \) (mod 3). Three points satisfy this equation exactly when they form a line in \( \mathbb{Z}_3^3 \). Let \( a_n \) be the size of the largest \( n \)-dimensional set without a solution. Meshulam showed that \( a_N = O(3^{N/3}) \) (a bound which appears on the homework).

To obtain a quick lower bound, observe \( \{0, 1\}^n \) works so \( a_n \geq 2^n \). A more general fact is if \( A \subseteq \mathbb{Z}_{n_1}^3 \) and \( B \subseteq \mathbb{Z}_{n_2}^3 \), both without 3 term arithmetic progressions then \( A \times B \subseteq \mathbb{Z}_{n_1+n_2}^3 \) also has no 3 term arithmetic progression. Thus \( 3^{n_1+n_2} \geq a_{n_1+n_2} \geq a_{n_1} \times a_{n_2} \).

Fekete’s lemma implies that \( \lim_{n \to \infty} a_{\frac{1}{n}}^\frac{1}{n} \) exists and is equal to \( \sup_n a_{\frac{1}{n}}^\frac{1}{n} \). Edel(2004) showed \( a_{\frac{1}{n}}^\frac{1}{n} > 2.217 \) for large \( n \). An major open problem is whether \( \lim_{n \to \infty} a_{\frac{1}{n}}^\frac{1}{n} = 3 \).

3.2. Longer Progressions. A question we may ask is whether we can make the proof of Roth’s theorem work for 4 term (or longer) progressions. This turns out to be quite hard. A key fact obtained from (and used heavily in) the proof of Roth’s theorem is the number of 3 term progressions in \( A \) is equal to \( \sum k \widehat{1}_A(k)^2 \widehat{1}_A(-2k) \). So in a sense the 3 term progressions are controlled by the behavior of the Fourier transform of \( A \). This turns out not to be true for longer progressions!

**Nasty Example:** Fix \( \delta \) small and let \( A = \{ a \in \mathbb{Z}_N \mid |a|^2 > \frac{N\delta}{2} \} \). Here \( |x| = \min\{ x \pmod{n}, -x \pmod{n} \} \).

We have the following facts.

- \( |A| = \delta N + o(N) \) since \( (-\frac{N\delta}{2}, \frac{N\delta}{2}) \) has approximately \( \frac{N\delta}{2} \) quadratic residues (each with two choices for \( a \)).
- \( A \) is uniform in the sense of Roth’s theorem. For any \( r \neq 0 \) \( \widehat{1}_A(r) = O(\sqrt{N\log N}) \).

The idea of the proof of the second of these is that the Fourier transform looks like a Gauss sum.

\[
\sum_x \exp \left( -\frac{2\pi i}{N} (ax^2 + bx) \right),
\]

and such sums are small in magnitude.

By our proof of Roth’s theorem, we know \( A \) has approximately \( \delta^3 N^2 \) 3 term arithmetic progressions, and we might hope \( A \) has approximately \( \delta^3 N^2 \) 4 term arithmetic progressions. But this is not the case.
The key relation here is that for any \( a \) and \( d \) we have

\[
(a)^2 - 3(a + d)^2 + 3(a + 2d)^2 - (a + 3d)^2 = 0.
\]

This implies that if \( a^2 \), \( (a + d)^2 \), and \( (a + 2d)^2 \) are all close to 0, then \( (a + 3d)^2 \) is also fairly close to 0. Let \( A' = \{ a : |a|^2 < \frac{N\delta}{14} \} \). We observe:

- \( A' \) is uniform and has density \( \frac{N\delta}{7} \).
- \( A' \) has roughly \( (\frac{N}{7})^3 N^2 \) three term arithmetic progressions.
- If \( a, a + d, a + 2d \in A' \) then \( a + 3d \in A \).

This means \( A \) has at least \( (\frac{N}{7})^3 N^2 \) four term arithmetic progressions which is much greater than \( \delta^4 N^2 \) for small enough \( \delta \).

**Gowers:** The right idea for random-like is instead to average over hypercubes. He defined the following norms.

\[
\|f\|_{U^2} = \left( \frac{1}{N^3} \sum_{x,a_1,a_2} f(x) f(x + a_1) \overline{f(x + a_2)} f(x + a_1 + a_2) \right)^\frac{1}{4}
\]

\[
\|f\|_{U^3} = \left( \frac{1}{N^4} \sum_{x,a_1,a_2,a_3} f(x) f(x + a_1) \overline{f(x + a_2)} \overline{f(x + a_3)} f(x + a_1 + a_2) f(x + a_1 + a_3) \right)^\frac{1}{8}
\]

\[
\|f\|_{U^d} = \left( \frac{1}{N^{d-1}} \sum_{x,a_1,a_2,...,a_d} f(x) f(x + a_1) \overline{f(x + a_2)} \overline{f(x + a_3)} \cdots \overline{f(x + a_d)} f(x + a_1 + a_2 + \cdots + a_d) \right)^\frac{1}{2^d}
\]

and similarly for larger \( d \). It turns out \( \|1_A\|_{U^{d-1}} \) controls arithmetic progressions of length \( d \), meaning that if the number of progressions is vastly different than that of a random set than the norm is different too. He used this to give a density-increment proof of Roth’s theorem for longer progressions.

As a sidenote, it turns out \( \|f\|_{U^2} = \sum_k |\hat{f}(k)|^4 \). So in the 3-term progression case, this ends up reducing back down to the case where \( \|f\|_{U^2} \) is small if and only if the largest Fourier coefficient of \( f \) is small.