We will prove the quantitative non-linear Roth theorem via a density-increment argument. Let $A \subset [N]$ be free of the nonlinear Roth configuration. Observe that if $P = a + q[N']$ is any arithmetic progression, then the rescaled set

$$A' := \{n : a + qn \in A \cap P\} \subset [N']$$

is free of the (if $q \neq 1$, different) configuration

$$x, x+y, x+qy^2 \qquad (y \neq 0). \tag{10}$$

If q is greater than the length of the interval N', then every subset of $[N']$ lacks the configuration (10), meaning that there is no leverage with which one can continue a density-increment iteration. But, if one can obtain a density-increment with q small, then there is at least some hope of continuing the argument. This is exactly what Sean and I did:

Lemma 9. *Let $q \in \mathbf{N}$ and set $M := (N/q)^{1/2}$. Suppose that $A \subset [N]$ has density α and contains no nontrivial polynomial progressions*

$$x, x+y, x+qy^2 \qquad (y \neq 0).$$

Then, either

1. *$N \gg q^2\alpha^{-O(1)}$ or*
2. *there exist $a', q', N' \in \mathbf{N}$ with $q' \ll \alpha^{-O(1)}$ and $\alpha^{O(1)}q^{-1}M \ll N' \leq M$ such*

that

$$|A \cap (a' + qq' \cdot [N'])| \geq \left(\alpha + \Omega(\alpha^{O(1)})\right) N'.$$

All of the implied constants in the above lemma are absolute. In particular, they do not depend on q; this allows one to run a density-increment argument to prove a quantitative version of the non-linear Roth theorem.

Exercise 18. Use this to prove that if $A \subset [N]$ is free of the non-linear Roth configuration, then

$$|A| \ll \frac{N}{(\log\log N)^{O(1)}}.$$

We will illustrate the proof of the density-increment lemma in the case $q = 1$ only; adapting the argument to work for general q simply consists of carrying the q throughout the same proof.

Suppose that $A \subset [N]$ has no non-linear Roth configurations. Then, $\Lambda(1_A, 1_A, 1_A) = 0$. On the other hand, $1_A = f_A + \alpha 1_{[N]}$, and so using the trilinearity of Λ, we have that $\Lambda(1_A, 1_A, 1_A)$ equals

$$\Lambda(1_A, f_A, 1_A) + \alpha\Lambda(1_A, 1_{[N]}, f_A) + \alpha^2\Lambda(f_A, 1_{[N]}, 1_{[N]}) + \alpha^3\Lambda(1_{[N]}, 1_{[N]}, 1_{[N]}).$$

Exercise 19. Prove that $\Lambda(1_{[N]}, 1_{[N]}, 1_{[N]}) \gg 1$.

As a consequence of the exercise and the fact that $\Lambda(1_A, 1_A, 1_A) = 0$, at least one of the following must hold:

1. $\left|\Lambda(f_A, 1_{[N]}, 1_{[N]})\right| \gg \alpha$,

(2) $\left|\Lambda(1_A, 1_{[N]}, f_A)\right| \gg \alpha^2$, or
(3) $|\Lambda(1_A, f_A, 1_A)| \gg \alpha^3$.

By far the most difficult case to consider (and the only one in which new ideas are needed) is the third. A density-increment of the desired form can be deduced in the second case by using ideas from the (Fourier-analytic) proof of Sárközy's theorem, or by modifying the argument used in the third case very slightly. Deducing a density-increment in the first case is even simpler.

Exercise 20. Deduce the density-increment lemma in the case $\left|\Lambda(f_A, 1_{[N]}, 1_{[N]})\right| \gg \alpha$.

So, we will assume that $|\Lambda(1_A, f_A, 1_A)| \gg \alpha^3$, and show how to deduce a density-increment.

Last time, we showed that when f, g, and h are 1-bounded, then largeness of $|\Lambda(f, g, h)|$ implies largeness of the average of Gowers box norms

$$\mathbf{E}_{|a|,|b|<\delta' M} \|h\|_{\square^3_{2a[\delta'' M], 2b[\delta'' M], 2(a+b)[\delta'' M]}(N)}. \tag{11}$$

The new innovations of our proof were the following ingredients:

(1) "Quantitative concatenation": this average of Gowers box norms can be bounded above by the U^s-norm of h for some $s \ll 1$.
(2) "Degree lowering": control of $\Lambda(f, g, h)$ by the U^s-norm implies control by the U^{s-1}-norm, after passing to an arithmetic progression with bounded common difference.

Combining the first point with iterations of the second point implies that one has control of $\Lambda(f, g, h)$ by a U^1-norm localized to an arithmetic progression with small common difference. Precisely, we obtain the following result.

Theorem 12. *Let $f, g, h : \mathbf{Z} \to \mathbf{C}$ be 1-bounded functions supported in the interval $[N]$ and $\delta > 0$. Set $M := \sqrt{N}$ and suppose that*

$$|\Lambda(f, g, h)| \geq \delta.$$

Then, there exist $q', N' \in \mathbf{N}$ with $q' \ll \delta^{-O(1)}$ and $\delta^{O(1)} q^{-1} M \ll N' \leq M$ such that

$$\frac{1}{N} \sum_{x \in \mathbf{Z}} \left|\mathbf{E}_{y \in [N']} g(x + q'y)\right| \gg \delta^{O(1)},$$

provided $N \gg \delta^{-O(1)}$.

From this statement, one can mimic the end of our proof of the density-increment lemma used to prove Sárközy's theorem to obtain a density-increment.

Exercise 21. Prove the density-increment lemma in the $q = 1$ case assuming the above theorem.

8. Quantitative concatenation

Concatenation results are used to bound averages of Gowers box norms in terms of Gowers uniformity norms, which we understand much better. The first such results were proven by Tao and Ziegler in their work on countig polynomial progressions in the primes, but these were purely qualitative in nature, and thus not suitable for proving quantitative bounds in the polynomial Szemerédi theorem. The first quantitative concatenation result was proven by Sean and I in our work on the nonlinear Roth theorem. We showed that the average of Gowers box norms (11) is bounded above by (some power of) $\|h\|_{U^5_{[\delta'\delta''N]}(N)}$. The main tool we used was repeated applications of the Cauchy–Schwarz inequality, along with a bit of Fourier analysis and an inverse theorem for certain Gowers box norms of degree 2. Recently, Kuca introduced a new proof of quantitative concatenation that does not require the degree 2 box norm inverse theorem and also can be applied in more general multidimensional situations.

All proofs of concatenation require an elaborate sequence of applications of the Cauchy–Schwarz inequality. To illustrate the main idea, we will sketch a proof of the most basic case.

The first observation noted in all quantitative concatenation arguments is that if $a, b \in [M]$ are sufficiently "independent" in the sense that $\gcd(a, b)$ is small, then the terms $an-bm$ as we range over $n, m \in [M']$ should relatively evenly cover the integers divisible by $\gcd(a, b)$ in $\{-MM', \dots, MM'\}$. We make this precise below, where we write $a[M'] - b[M']$ to mean the multiset of elements of the form $an - bm$ with $n, m \in [M']$.

Lemma 10. *Let $a, b \in \mathbf{Z}$ be nonzero, $\varepsilon > 0$, $N, M, M' > 0$ with $M \le M'$, and $f : \mathbf{Z} \to \mathbf{C}$ be 1-bounded and supported on $[N]$. Assume that $\gcd(a, b) \le \varepsilon^{-1}$ and $\varepsilon M \le |a|, |b| \le 2M$. Suppose that*

$$\|f\|^2_{\square^1_{a[M']-b[M']}(N)} \ge \delta$$

and that $N \ge \frac{2}{\delta^2\varepsilon^2}$. Then,

$$\|f\|_{U^2_{[MM']}(N)} \gg (\delta\varepsilon)^2.$$

The proof of this lemma is purely Fourier-analytic, and so in the interest of time we omit it.

It turns out that the condition on the greatest common divisor of a and b in the above lemma is generically true for pairs (a, b) with $|a|, |b| < N$. In fact, it is true even with fixed shifts, as can be seen with some elementary number theory.

Lemma 11. *Fix integers $|c|, |c'| \le M$. The number of pairs of integers (a, b) with $|a|, |b| \le M$ such that $\gcd(a + c, b + c') > \varepsilon^{-1}$ is $\ll \varepsilon M^2$.*

Proof. The the greatest common divisor of $a+c$ and $b+c'$ is certainly at most $|a+c| \leq |a|+|c| \leq 2M$, we have

$$\begin{aligned}\sum_{\substack{|a|,|b|\leq M \\ \gcd(a+c,b+c')>\varepsilon^{-1}}} 1 &\leq \sum_{\varepsilon^{-1}<d\leq 2M}\left(\sum_{|m|\leq 2M} 1_{d|m}\right)^2 \\ &\leq \sum_{\varepsilon^{-1}<d\leq 2M}\left(\frac{4M}{d}+1\right)^2 \\ &\ll M^2\sum_{d>\varepsilon^{-1}}\frac{1}{d^2} \ll \varepsilon M^2,\end{aligned}$$

as desired. □

The main idea underlying all quantitative concatenation arguments is that the situation arising in Lemma 10 can be guaranteed by using the Cauchy–Schwarz inequality and a change of variables to mix the different arithmetic progressions appearing in averages of Gowers box norms. Indeed, suppose that

$$\mathbf{E}_{|a|<M}\|f\|^2_{\square^1_{a[M']}(N)} \geq \delta.$$

Then, by making a change of variables in the definition of the Gowers box norm, we have,

$$\mathbf{E}_{|a|<M}\frac{1}{N}\sum_{x\in\mathbf{Z}}\mathbf{E}_{h,h'\in[M']}f(x)\overline{f}(x+a(h'-h)) \geq \delta.$$

Applying the Cauchy–Schwarz inequality to double a and (h,h') then yields

$$\mathbf{E}_{|a|,|b|<M}\frac{1}{N}\sum_{x\in\mathbf{Z}}\mathbf{E}_{h,h',k,k'\in[M']}f(x+b(k'-k))\overline{f}(x+a(h'-h)) \geq \delta^2.$$

Making the change of variables $x\mapsto x-b(k-k')$ reveals that the left-hand side above equals

$$\mathbf{E}_{|a|,|b|<M}\|f\|^2_{\square^1_{a[M']-b[M']}(N)}.$$

Now, the contribution to the average coming from pairs (a,b) that have large greatest common divisor too large or for which a is too small to apply Lemma 10 is negligible, and so it follows that

$$\|f\|_{U^2_{[MM']}(N)} \gg \delta^{O(1)}.$$

9. Degree-lowering

Thus, supposing that $|\Lambda(f,g,h)| \geq \delta$, with an appropriate choice of $\delta',\delta'' \ll \delta^{O(1)}$, one can obtain that that $\|h\|_{U^s_{[\gamma N]}(N)} \gg \delta^{O(1)}$ for some $s \ll 1$ and $\gamma \gg \delta^{O(1)}$. This implication actually implies that if $H:\mathbf{Z}\to\mathbf{C}$ is the *dual function*

$$H(x) := \mathbf{E}_{y\in[M]}f(x-y^2)g(x+y-y^2),$$

then, in fact, $\|H\|_{U^s_{[\gamma N]}(N)} \gg \delta^{O(1)}$. Indeed, we have by a change of variables that

$$\delta \leq \left| \frac{1}{N} \sum_{x \in \mathbf{Z}} \mathbf{E}_{y \in [M]} f(x-y^2) g(x+y-y^2) h(x) \right|$$

so that, by an application of the Cauchy–Schwarz inequality and another change of variables,

$$\begin{aligned} \delta^2 &\leq \frac{1}{N} \sum_{x \in \mathbf{Z}} \mathbf{E}_{y,z \in [M]} f(x-y^2) \overline{f}(x-z^2) g(x+y-y^2) \overline{g}(x+z-z^2) \\ &= \frac{1}{N} \sum_{x \in \mathbf{Z}} \mathbf{E}_{z \in [M]} \overline{f}(x-z^2) \overline{g}(x+z-z^2) H(x) \\ &= \frac{1}{N} \sum_{x \in \mathbf{Z}} \mathbf{E}_{z \in [M]} \overline{f}(x) \overline{g}(x+z) H(x+z^2), \end{aligned}$$

which equals $\Lambda(\overline{f}, \overline{g}, H)$, all arguments of which are still 1-bounded functions (supported on intervals of length $\asymp N$). This maneuver is now known as *stashing*.

Here is the precise "degree-lowering" statement that we prove:

Proposition 3. *Fix $M = \sqrt{N}$. Let $f, g : \mathbf{Z} \to \mathbf{C}$ be 1-bounded functions supported on the interval $[N]$, and define the dual function H as above. Suppose that $s \in \mathbf{N}$ with $s \geq 3$. If*

$$\|H\|_{U^s_{[\delta' N]}(N)} \geq \delta,$$

then either $N \ll_s (\delta\delta')^{-O_s(1)}$, or

$$\|H\|_{U^{s-1}_{[\delta' N]}(N)} \gg_s \delta^{O_s(1)}.$$

Thus, if $\delta' \ll \delta^{O(1)}$ and $N \gg \delta^{-O(1)}$ and one knows that $\|H\|_{U^s_{[\delta' N]}(N)} \geq \delta$ for some $s \ll 1$, then one can repeatedly apply the above proposition to deduce that $\|H\|_{U^2_{[\delta' N]}(N)} \gg \delta^{O(1)}$. One can then apply the U^2-inverse theorem to obtain that there exists $\xi \in \mathbf{T}$ for which

$$\left| \frac{1}{N} \sum_{x \in \mathbf{Z}} H(x) e(\xi x) \right| \gg \delta^{O(1)}$$

which, after inserting in the definition of the dual function and making a change of variables, yields

$$\left| \frac{1}{N} \sum_{x \in \mathbf{Z}} \mathbf{E}_{y \in [M]} f(x) g(x+y) e(\xi(x+y^2)) \right| \gg \delta^{O(1)}.$$

The next ingredient we need is the following Fourier-analytic statement, whose proof is also an exercise.

Lemma 12 (Major arc lemma)**.** *Let $\xi \in \mathbf{T}$. If for some 1-bounded functions $f, g : \mathbf{Z} \to \mathbf{C}$ supported on $[N]$ we have that*

$$\left| \frac{1}{N} \sum_{x \in \mathbf{Z}} \mathbf{E}_{y \in [M]} f(x) g(x+y) e(\xi(x+y^2)) \right| \geq \delta,$$

then there exists $q \ll \delta^{-O(1)}$ such that $\|q\xi\| \ll \frac{\delta^{-O(1)}}{N}$.

Exercise 22. Prove the major arc lemma.

Applying the major arc lemma, we obtain that there exists $q \ll \delta^{-O(1)}$ such that $\|q\xi\| \ll \frac{\delta^{-O(1)}}{N}$, i.e., ξ is very close to a rational $\frac{a}{q}$ with small denominator. Applying the triangle inequality and partitioning the average over $y \in [M]$ into an average of averages over arithmetic progressions with common difference q then makes the $e(\xi y^2)$ term roughly constant, and allows us to deduce that

$$\frac{1}{N} \sum_{x \in \mathbf{Z}} \left| \mathbf{E}_{y \in [N']} g(x+qy) \right| \gg \delta^{O(1)}$$

for some $N' \gg \delta^{O(1)} M$, as desired.

We will finish by illustrating how to prove Proposition 3 in the case that $s = 3$. For the remainder of the lecture, we will drop the Fejér kernel weights and the averaging ranges in the definition of the Gowers norms for ease of exposition.

It's an immediate consequence of the definition that $\mathbf{E}_a \|\Delta_a H\|_{U^2}^4 = \|H\|_{U^3}^8$. Thus, by the assumption $\|H\|_{U^3}^8 \geq \delta$, it follows from the pigeonhole principle that for many choices of the differencing parameter a, we have

$$\|\Delta_a H\|_{U^2}^4 \gg \delta.$$

Applying the U^2-inverse theorem, we deduce the existence of a function $\phi : \mathbf{Z} \to \mathbf{T}$ such that for many (i.e., $\gg \delta^{O(1)} N$) choices of a, we have

$$\left| \frac{1}{N} \sum_{x \in \mathbf{Z}} \Delta_a H(x) e(\phi(a)x) \right| \gg \delta, \tag{12}$$

so that by positivity,

$$\frac{1}{N} \sum_{a \in \mathbf{Z}} \left| \frac{1}{N} \sum_{x \in \mathbf{Z}} \Delta_a H(x) e(\phi(h)x) \right|^2 \gg \delta^2.$$

I claim that if we could replace the function ϕ by a constant $\beta \in \mathbf{T}$ not depending on a, we would be done. Indeed, this follows by one application of the Cauchy–Schwarz inequality: If

$$\frac{1}{N} \sum_{a \in \mathbf{Z}} \left| \frac{1}{N} \sum_{x \in \mathbf{Z}} \Delta_a H(x) e(\beta x) \right|^2 \gg \delta^{O(1)},$$

then expanding the square, we have

$$\frac{1}{N^3}\sum_{x\in\mathbf{Z}}\sum_{a,a'\in\mathbf{Z}} H(x)\overline{H(x+a)H(x+a')}H(x+a+a')e\left(-\beta a'\right) \gg \delta^{O(1)},$$

so that, after Cauchy–Schwarzing to double the a variable and eliminate $e(-\beta a')$, we deduce that

$$\|H\|_{U^2}^4 \gg \delta^{O(1)}.$$

It remains to show that such a β exists.

Suppose that $A \subset \mathbf{Z}$ is the set of shifts a for which (12) holds and that $|A| \gg \delta^{O(1)}N$. Then, we have

$$\mathbf{E}_{a\in A}\left|\frac{1}{N}\sum_{x\in\mathbf{Z}}\Delta_a H(x)e(\phi(h)x)\right|^2 \gg \delta^2.$$

Inserting in the definition of the dual function H and expanding the square reveals that the left-hand side above equals

$$\frac{1}{N^2}\sum_{x,z\in\mathbf{Z}}\mathbf{E}_{a\in A}\mathbf{E}_{\substack{y,y'\in[M]\\ w,w'\in[M]}}\Big(f(x-y^2)g(x+y-y^2)\overline{f}(x-y'^2+a)\overline{g}(x+y'-y'^2+a)$$
$$\overline{f}(z-w^2)\overline{g}(z+w-w^2)f(z-w'^2+a)g(z+w'-w'^2+a)e(\phi(a)[x-z])\Big)$$

Applying the Cauchy–Schwarz inequality to double a but leaving all other variables fixed, we obtain (using that f and g are 1-bounded) that the modulus squared of the above is bounded by

$$\frac{1}{N^2}\sum_{x,z\in\mathbf{Z}}\mathbf{E}_{a,a'\in A}\mathbf{E}_{\substack{y,y'\in[M]\\ w,w'\in[M]}}\Big(\overline{f}(x-y'^2+a)\overline{g}(x+y'-y'^2+a)f(x-y'^2+a')g(x+y'-y'^2+a')$$
$$f(z-w'^2+a)g(z+w'-w'^2+a)\overline{f}(z-w'^2+a')\overline{g}(z+w'-w'^2+a')$$
$$e([\phi(a)-\phi(a')][x-z])\Big),$$

which, since the averages over y and w are superfluous, swapping the order of summation, making the change of variables $x\mapsto x-a$ and $z\mapsto z-a$, and relabeling y' by y equals

$$\mathbf{E}_{a,a'\in A}\left|\frac{1}{N}\sum_{x\in\mathbf{Z}}\mathbf{E}_{y\in[M]}\Delta_{a'-a}f(x-y^2)\Delta_{a'-a}g(x+y-y^2)e([\phi(a)-\phi(a')]x)\right|^2$$

Thus, applying the pigeonhole principle and making the change of variables $x\mapsto x+y^2$, we deduce that

$$\left|\frac{1}{N}\sum_{x\in\mathbf{Z}}\mathbf{E}_{y\in[M]}\Delta_{a'-a}f(x)\Delta_{a'-a}g(x+y)e([\phi(a)-\phi(a')][x+y^2])\right| \gg \delta^{O(1)}$$

for many pairs $(a,a')\in A^2$.

Now, we are in a position to apply the major arc lemma again. This tells us that for all such $(a,a')\in A^2$, there exists $q_{a,a'} \ll \delta^{-O(1)}$ for which

$\|q_{a,a'}(\phi(a)-\phi(a'))\| \ll \frac{\delta^{-O(1)}}{N}$. That is, $\phi(a)-\phi(a')$ is extremely close to a rational $r \in \mathbf{T}$ with denominator $\ll \delta^{-O(1)}$. There are very few such rationals (precisely, $\ll \delta^{-O(1)}$ many). Thus, by the pigeonhole principle again, there exists a $\beta_0 \in \mathbf{T}$ for which $\phi(a)-\phi(a') \approx \beta_0$ for a positive proportion of $(a,a') \in A^2$. Using a final application of the pigeonhole principle to fix a', we obtain that $\phi(a) \approx \beta_0 + \phi(a')$ for $\gg \delta^{O(1)}N$ many $a \in \mathbf{Z}$ for which (12) holds. This implies that

$$\frac{1}{N}\sum_{a\in\mathbf{Z}}\left|\frac{1}{N}\sum_{x\in\mathbf{Z}}\Delta_a H(x)e(\beta x)\right|^2 \gg \delta^{O(1)}$$

with $\beta = \beta_0 + \phi(a')$, and thus deduce that the U^2-norm of H is large.