# Lecture 27: Roth's Theorem in $\mathbb{F}_{3}^{n}$ 

Apr. 26, 2007
Lecturer: Ryan O'Donnell
Scribe: Ryan O'Donnell

## 1 Fourier analysis over other product domains

So far in the course we have been concerned almost exclusively with functions on the set $\{-1,1\}^{n}$ (sometimes written $\mathbb{F}_{2}^{n}$ ). What about other product spaces?

As alluded to in Problem \#1 on Homeworks \#4 and \#5, we can develop some kind of "orthogonal decomposition" for functions on any product probability space $X^{n}$. This will give us some subset of our Fourier analysis, and it's useful when one doesn't have any particular structure on the set $X$. As an example, we might study social choice functions on $m$ candidates, and then the set $X=\{1,2, \ldots, m\}$ has no particular structure beyond its cardinality.

On the other hand, sometimes $X$ has some additional structure in which we're interested. A good example of this is when $X$ is an abelian group; i.e., it has an additive structure. Let's consider the simplest case (which is also essentially the most general case for finite $X$ ), namely $X=\mathbb{Z}_{m}$.

### 1.1 Fourier analysis over $\mathbb{Z}_{m}$

We will be interested in subsets of $\mathbb{Z}_{m}^{n}$, and more generally, functions $f: \mathbb{Z}_{m}^{n} \rightarrow \mathbb{R}$. In fact, it is convenient even to generalize to functions

$$
f: \mathbb{Z}_{m}^{n} \rightarrow \mathbb{C}
$$

We will also generally be interested in the counting measure on $\mathbb{Z}_{m}^{n}$; or more conveniently, the uniform probability distribution. As usual, we will often let $\boldsymbol{x}$ denote a uniformly chosen element of $\mathbb{Z}_{m}^{n}$ and write $\mathbf{E}[\cdot], \operatorname{Pr}[\cdot]$, etc. with respect to this distribution.

The set of functions $f: \mathbb{Z}_{m}^{n} \rightarrow \mathbb{C}$ forms a vector space over $\mathbb{C}$, with the natural notion of addition and scalar multiplication for functions. It is easy to see that this vector space has dimension $m^{n}$. We will again introduce an inner product:

$$
\langle f, g\rangle=\underset{x \in \mathbb{Z}_{m}^{n}}{\mathbf{E}}[f(x) \overline{g(x)}] .
$$

(We have to take the complex conjugate of $g(x)$ to make this into a complex inner product.)
We are also interested in the fact that $\mathbb{Z}_{m}^{n}$ has an additive structure (i.e., an abelian group structure). The set of complex numbers also has an abelian group structure under multiplication. Indeed, so does the following subgroup:

Definition 1.1 $S_{1}$ denotes $\{z \in \mathbb{C}:|z|=1\}$.

Definition 1.2 A character is a function $\chi: \mathbb{Z}_{m}^{n} \rightarrow S_{1}$ which is a group homomorphism; i.e., $\chi(x+y)=\chi(x) \chi(y)$.

The following fact is very easy:
Fact 1.3 If $\chi$ and $\phi$ are characters, so are $\bar{\chi}$ and $\chi \cdot \phi$. Hence so is $\chi \cdot \bar{\phi}$.
We have:
Proposition 1.4 Let $\chi$ be a character. Then either $\chi \equiv 1$ or $\mathbf{E}[\chi]=0$.
Proof: If $\chi \not \equiv 1$, pick some $y \in \mathbb{Z}_{m}^{n}$ so that $\chi(y) \neq 1$. Now when $\boldsymbol{x} \in \mathbb{Z}_{m}^{n}$ is uniformly random, so is $\boldsymbol{x}+y$. Hence

$$
\underset{\boldsymbol{x}}{\mathbf{E}}[\chi(\boldsymbol{x})]=\underset{\boldsymbol{x}}{\mathbf{E}}[\chi(\boldsymbol{x}+y)]=\underset{\boldsymbol{x}}{\mathbf{E}}[\chi(y) \chi(\boldsymbol{x})]=\chi(y) \underset{\boldsymbol{x}}{\mathbf{E}}[\chi(\boldsymbol{x})] .
$$

Since $\chi(y) \neq 1$, it must be that $\mathbf{E}[\chi]=0$.
Corollary 1.5 The set of all characters is orthonormal.

## Proof:

$$
\langle\chi, \chi\rangle=\mathbf{E}[\chi(\boldsymbol{x}) \overline{\chi(\boldsymbol{x})}]=\mathbf{E}[|\chi(\boldsymbol{x})|]=\mathbf{E}[1]=1
$$

On the other hand, if $\chi$ and $\phi$ are distinct characters, then the character $\chi \bar{\phi} \not \equiv 1$ (this uses the fact that $\bar{\phi}=1 / \phi$ ). Thus $\langle\chi, \phi\rangle=\mathbf{E}[\chi \bar{\phi}]=0$ as needed.

In fact, we will now exhibit $m^{n}$ distinct characters, and hence these characters form an orthonormal basis for the inner product space.
Definition 1.6 Let $\omega=\exp (2 \pi i / m)$, the mth root of unity. Given $\alpha \in \mathbb{Z}_{m}^{n}$, we define $\chi_{\alpha}: \mathbb{Z}_{m}^{n} \rightarrow$ $\mathbb{C}$ by

$$
\chi_{\alpha}(x)=\omega^{\alpha_{1} x_{1}+\cdots+\alpha_{n} x_{n}}=\omega^{\alpha \cdot x},
$$

and it's easy to see that this is a character. Also, it is well-defined in the sense that it gives the same value regardless of the integer representative of each $\alpha_{i}$ or $x_{j} \bmod m$ (since $\omega^{m}=1$ ).

The following facts are easy to see:
Fact 1.7 1. $\chi_{\alpha} \overline{\chi_{\alpha^{\prime}}}=\chi_{\alpha-\alpha^{\prime}} .2 . \alpha \neq 0 \Rightarrow \chi_{\alpha} \not \equiv 1$.
It follows that all $m^{n}$ of these characters are distinct (and the group of them, under function multiplication, is isomorphic to $\mathbb{Z}_{m}^{n}$ ). Hence indeed they form an orthonormal basis for the inner product space of functions $f: \mathbb{Z}_{m}^{n} \rightarrow \mathbb{C}$.

We now again have a Fourier expansion of any $f: \mathbb{Z}_{m}^{n} \rightarrow \mathbb{C}$,

$$
f(x)=\sum_{\alpha \in \mathbb{Z}_{m}^{n}} \hat{f}(\alpha) \chi_{\alpha}(x),
$$

where

$$
\hat{f}(\alpha)=\left\langle f, \chi_{\alpha}\right\rangle .
$$

And again, by orthonormality, Plancherel and Parseval hold:

## Fact 1.8

$$
\langle f, g\rangle=\sum_{\alpha} \hat{f}(\alpha) \overline{\hat{g}(\alpha)}
$$

and hence

$$
\underset{\boldsymbol{x}}{\mathbf{E}}\left[|f(\boldsymbol{x})|^{2}\right]=\langle f, f\rangle=\sum_{\alpha}|\hat{f}(\alpha)|^{2}
$$

We also have $\mathbf{E}[f]=\hat{f}(0)$ and $\widehat{f+g}=\hat{f}+\hat{g}$, as usual.

### 1.2 Finite fields

Finally, since we have so much group structure floating around, it's inevitable that subgroups will arise. Things are much more convenient if the only subgroups of $\mathbb{Z}_{m}$ are the trivial ones ( $\{0\}$ and $\mathbb{Z}_{m}$ ); in this case, all subgroups of $\mathbb{Z}_{m}^{n}$ will be isomorphic to $\mathbb{Z}_{m}^{n^{\prime}}$ for some $n^{\prime} \leq n$. This happens if and only if $m$ is some prime, $p$. In this case, $\mathbb{Z}_{p}$ can be thought of as a field, $\mathbb{F}_{p}$, and then we can think of $\mathbb{F}_{p}^{n}$ itself as a vector space, with vector subspaces (instead of "subgroups"). For convenience, we will restrict to this case.

## 2 Arithmetic combinatorics

Definition 2.1 Roughly, arithmetic combinatorics is the study of arithmetic structure in subsets of $\mathbb{Z}$ or $\{1, \ldots, N\}$ or $\mathbb{Z}_{N}$, or perhaps $\mathbb{F}_{p}^{n}$. One is especially interested in subset sums and subset products -

If $A$ is a set, $A+A:=\{a+b: a, b \in A\}$, and $A \cdot A:=\{a b: a, b \in \mathbb{Z}\}$

- and also in arithmetic progressions and similar structures. One asks questions like, "If $A$ is large enough, must it contain long arithmetic progressions?", and "If $|A+A|$ is not much larger than $|A|$, must it have some special structure?"

We will only talk about "additive combinatorics", meaning we will only sum elements, not multiply them.

### 2.1 Arithmetic progressions in dense sets

We will start with the following problem of finding arithmetic progressions in dense sets.
Definition 2.2 (For any additive group.) An arithmetic progression (AP) of length $k$ is a list $x, x+$ $d, x+2 d, \ldots, x+(k-1) d$, where the elements are distinct (in particular, $d \neq 0$ ).

Theorem 2.3 (Van der Waerden 1927) Suppose the integers are colored with finitely many colors. Then there are arbitrarily long monochromatic APs.

In 1936, Erdős and Turán conjectured that the coloring was just a distraction, and that the reason this worked is that any set of positive density must have arbitrarily long APs. They asked:

Question: What is the size $r_{k}(N)$ of the largest subset of $\{1,2, \ldots, N\}$ with no length- $k$ AP? Is it $o(N)$ ?

This proved to be a huge unsolved problem for a long time. Seven years after this, Behrend gave a very pretty 3-AP-free construction, showing that the $o(N)$ couldn't be too small:

Theorem 2.4 (Behrend 1946) $r_{3}(N) \geq \Omega(N / \exp (-\sqrt{\log N})$.
This has never been improved.
Ten years after this, the first positive progress was made:
Theorem $2.5\left(\right.$ Roth 1956) $r_{3}(N) \leq O(N / \log \log N)$.
Roth's proof was by Fourier analysis.
Finally, 39 years after Erdős and Turán's conjecture, Szemerédi solved the problem:
Theorem 2.6 (Szemerédi 1975) For each $k, r_{k}(N)<o(N)$.
Szemerédi's proof involved reducing the question to graph theory, and the introduction of the famous Szemerédi Regularity Theorem, which is used pretty much nonstop in Property Testing of graphs.

As for the explicit $o(N)$ function Szemerédi proved, well, you don't want to know (let's just say that iterated Ackermann functions are involved....). Sixty-five years after Erdős and Turán, Gowers managed to get a more "sensible" bound:

Theorem 2.7 (Gowers 2001) For each $k$, there is a $c_{k}>0$ such that $r_{k}(N) \leq O\left(N /(\log \log N)^{c_{k}}\right)$. (Specifically, $c_{k}=1 / 2^{2^{k+9}}$.)

But in fact, there is still work to be done, as Erdős and Turán made the even stronger conjecture:
Conjecture 2.8 (Erdős and Turán 1936) If $A \subseteq \mathbb{Z}^{+}$with $\sum_{x \in A} 1 / x=\infty$ then $A$ has arbitrarily long APs. Nearly equivalently: For all $k, r_{k}(N) \leq O(N / \log N)$.

Since the sum of reciprocals of primes diverges, this would give arbitrarily long APs of primes. This in particular was a very famous conjecture. It was recently solved by Green and Tao - not by density arguments alone, but by using the fact that the primes are in some sense "pseudorandomly" distributed:

Theorem 2.9 (Green and Tao 2005) The primes contain arbitrarily long APs.
As a matter of fact, no one can prove the $O(N / \log N)$ conjecture even for $k=3$; the best we have there is:

Theorem 2.10 (Bourgain 1999) $r_{3}(N) \leq O(N \sqrt{\log \log N / \log N})$.
There is still a big gap here between Bourgain and Behrend, and this is considered a big open problem.

## 3 Roth's Theorem in $\mathbb{F}_{3}^{n}$

Heuristic: Suppose one has an additive combinatorics problems over $\{1, \ldots, N\}$. Then for any fixed $p$, there is an analogous problem over $\mathbb{F}_{p}^{n}$, where $N=p^{n}$. (One may have to take $p=2,3$, or 5 to avoid some silly problems.) Further, in this finite field setting, the problem is much cleaner and easier.

We will now give an example by proving Roth's Theorem in $\mathbb{F}_{3}^{n}$ (the result was first observed by Meshulam in 1995):

Theorem 3.1 $r_{3}\left(\mathbb{F}_{3}^{n}\right) \leq O\left(3^{n} / n\right)$. (I.e., $\leq O(N / \log N)$.) I.e., if $A \subseteq \mathbb{F}_{3}^{n}$ has $|A| \geq O(N / \log N)$, then $A$ contains some $x, x+d, x+2 d$, with $d \neq 0$.
(Note that if one works really hard at transferring this to the setting of $\{1, \ldots, N\}$, one gets Bourgain's theorem, not the Erdős-Turán conjecture.)

The reason we don't take $\mathbb{F}_{2}^{n}$ here is that $x+2 d=x$ always in this case, so there are no (nontrivial) 3-APs. In $\mathbb{F}_{3}^{n}$ though, note that a 3-AP has the property that in each of the 3 coordinate one sees either all the same value in $\mathbb{Z}_{3}$ or all different values in $\mathbb{Z}_{3}$. Basically, we're interested in the (card) game of Set:

Remark $3.2 r_{3}\left(\mathbb{F}_{3}^{n}\right)$ is the most number of cards you can have face-up in the game Set (with $n$ features rather than 4) without having any Set available.

Incidentally, the best lower bound known comes from coding theorists, who also study this problem. They call 3-AP-free sets in $\mathbb{F}_{3}^{n}$ "caps", and their best examples are by taking specific constant-size examples and doing a recursive construction:

Theorem 3.3 (Edel 2004) There is a constant $c=2.74 \ldots$ such that $r_{3}\left(\mathbb{F}_{3}^{n}\right) \geq c^{n}$.
Major open problem: Prove $r_{3}\left(\mathbb{F}_{3}^{n}\right)<o\left(3^{n} / n\right)$ or $r_{3}\left(\mathbb{F}_{3}^{n}\right) \geq(3-\epsilon)^{n}$ for every $\epsilon>0$.

### 3.1 The proof

Let $A \subseteq \mathbb{F}_{3}^{n}$ be given, with density $\mu=|A| / 3^{n}$. Let $f: \mathbb{F}_{3}^{n} \rightarrow\{0,1\}$ be the indicator function of $A$, so $f(0)=\mu$.

How can we find a $3-\mathrm{AP} x, x+d, x+2 d$ in $A$ ? Write $y=x+d$, so this is $x, y, x+2(y-x)=$ $-(x+y)$ (and we need $x \neq y$ ). Here is the idea:

Idea: Pick $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{F}_{3}^{n}$ independently and uniformly and hope that $\boldsymbol{x}, \boldsymbol{y},-(\boldsymbol{x}+\boldsymbol{y}) \in$; i.e., $f(\boldsymbol{x}) f(\boldsymbol{y}) f(-(\boldsymbol{x}+\boldsymbol{y}))=1$. (One needs to also hope that $\boldsymbol{x} \neq \boldsymbol{y}$ so that the 3 elements are distinct.)

Remark 3.4 If the three events were independent (e.g., perhaps if A were "randomly distributed") then the success probability would be basically $\mu^{3}$. (Since the probability $\boldsymbol{x}=\boldsymbol{y}$ is $3^{-n} \ll \mu^{3}$, we can just ignore the $\boldsymbol{x}=\boldsymbol{y}$ problem.) Since this is positive, there would exist a 3-AP.

Alternatively, if $\mu>.667$, we would always have success, by the union bound.
Obviously, the idea looks very much like the BLR linearity test from Lecture 2, and indeed we will use Fourier analysis to analyze the success probability:
Proposition 3.5 The probability that $x, y,-(x+y)$ are all in $A$ is $\sum_{\alpha} \hat{f}(\alpha)^{3}$.

## Proof:

$$
\begin{aligned}
\underset{\boldsymbol{x}, \boldsymbol{y}}{\mathbf{E}}[f(\boldsymbol{x}) f(\boldsymbol{y}) f(-(\boldsymbol{x}+\boldsymbol{y}))] & =\sum_{\alpha, \beta, \gamma} \hat{f}(\alpha) \hat{f}(\beta) \hat{f}(\gamma) \underset{\boldsymbol{x}, \boldsymbol{y}}{\mathbf{E}}\left[\chi_{\alpha}(\boldsymbol{x}) \chi_{\beta}(\boldsymbol{y}) \chi_{\gamma}(-(\boldsymbol{x}+\boldsymbol{y}))\right] \\
& =\sum_{\alpha, \beta, \gamma} \hat{f}(\alpha) \hat{f}(\beta) \hat{f}(\gamma) \underset{\boldsymbol{x}, \boldsymbol{y}}{\mathbf{E}}\left[\chi_{\alpha}(\boldsymbol{x}) \chi_{\beta}(\boldsymbol{y}) \chi_{-\gamma}(\boldsymbol{x}) \chi_{-\gamma}(\boldsymbol{y})\right] \\
& =\sum_{\alpha, \beta, \gamma} \hat{f}(\alpha) \hat{f}(\beta) \hat{f}(\gamma) \underset{\boldsymbol{x}, \boldsymbol{y}}{\mathbf{E}}\left[\chi_{\alpha-\gamma}(\boldsymbol{x}) \chi_{\beta-\gamma}(\boldsymbol{y})\right] \\
& =\sum_{\alpha, \beta, \gamma} \hat{f}(\alpha) \hat{f}(\beta) \hat{f}(\gamma) \underset{\boldsymbol{x}}{\mathbf{E}}\left[\chi_{\alpha-\gamma}(\boldsymbol{x})\right] \underset{\boldsymbol{y}}{\mathbf{E}}\left[\chi_{\beta-\gamma}(\boldsymbol{y})\right] \\
& =\sum_{\gamma} \hat{f}(\gamma)^{3},
\end{aligned}
$$

where in the last step we used that $\alpha-\gamma$ and $\beta-\gamma$ must be 0 or else the expectation is 0 .
Corollary 3.6 Suppose that $|\hat{f}(\alpha)|<\mu^{2} / 2$ for all $\alpha \neq 0$. Then A contains a 3 -AP. (Assuming $\mu \geq 2 /\left(3^{n}\right)^{1 / 3}=2 / N^{1 / 3}$.)

Proof: We have

$$
\begin{aligned}
\operatorname{Pr}_{\boldsymbol{x}, \boldsymbol{y}}[\boldsymbol{x}, \boldsymbol{y},-(\boldsymbol{x}+\boldsymbol{y}) \in A] & =\sum_{\alpha} \hat{f}(\alpha)^{3} \\
& =\mu^{3}+\sum_{\alpha \neq 0} \hat{f}(\alpha)^{3} \quad \text { (the latter is a real number) } \\
& \geq \mu^{3}-\sum_{\alpha \neq 0}|\hat{f}(\alpha)|^{3} \\
& \geq \mu^{3}-\left(\mu^{2} / 2\right) \sum_{\alpha \neq 0}|\hat{f}(\alpha)|^{2} \\
& \geq \mu^{3}-\left(\mu^{2} / 2\right) \sum_{\alpha}|\hat{f}(\alpha)|^{2} \\
& =\mu^{3}-\left(\mu^{2} / 2\right) \mathbf{E}\left[|f|^{2}\right] \quad \text { (Parseval) } \\
& =\mu^{3}-\left(\mu^{2} / 2\right) \mu \\
& =\mu^{3} / 2 .
\end{aligned}
$$

So the probability is positive. Further, the probability that $\boldsymbol{x}=\boldsymbol{y}$ is $3^{-n}=1 / N$, negligible compared to $\mu^{3} / 2$. Hence there is a positive probability of finding a genuine 3 -AP.

Pseudorandom versus Structure: What we"ve shown here is a kind of "pseudorandom versus linear" test. One makes the following definition:

Definition 3.7 We say that $A$ (or more generally any $f: \mathbb{F}_{p}^{n} \rightarrow[0,1]$ ) is $\eta$-uniform or $\eta$-pseudorandom if $|\hat{f}(\alpha)| \leq \eta$ for all $\alpha \neq 0$.

What do we do if $A$ is not $\mu^{2} / 2$-uniform? Let $\eta:=\mu^{2} / 2$, and suppose $|\hat{f}(\beta)| \geq \eta$, where $\beta \neq 0$. What this means is that $A$ is slightly positively correlated with one of the three "hyperplanes" $\beta \cdot x=0, \beta \cdot x=1$, or $\beta \cdot x=2$ :

Proposition 3.8 Suppose $f: \mathbb{F}_{p}^{n} \rightarrow \mathbb{R}$ has $\mathbf{E}[f]=\mu$ and $|\hat{f}(\beta)| \geq \eta$, where $\beta \neq 0$. Then there exists some $c \in \mathbb{F}_{p}$ such that $\mathbf{E}_{\boldsymbol{x}}[f(\boldsymbol{x}) \mid \beta \cdot \boldsymbol{x}=c] \geq \mu+\eta / 2$.

Proof: Let $g=f-\mu$, so $\mathbf{E}[g]=0$ and $|\hat{g}(\beta)| \geq \eta$. We can write

$$
|\hat{g}(\beta)|=\left|\underset{\boldsymbol{x}}{\mathbf{E}}\left[g(\boldsymbol{x}) \omega^{\beta \cdot \boldsymbol{x}}\right]\right|
$$

$$
=\left|\frac{1}{p} \underset{\boldsymbol{x}}{\mathbf{E}}\left[g(\boldsymbol{x}) \cdot \omega^{0} \mid \beta \cdot \boldsymbol{x}=0\right]+\frac{1}{p} \underset{\boldsymbol{x}}{\mathbf{E}}\left[g(\boldsymbol{x}) \cdot \omega^{1} \mid \beta \cdot \boldsymbol{x}=1\right]+\cdots+\frac{1}{p} \underset{\boldsymbol{x}}{\mathbf{E}}\left[g(\boldsymbol{x}) \cdot \omega^{p-1} \mid \beta \cdot \boldsymbol{x}=p-1\right]\right|
$$

$$
\leq \frac{1}{p}|\underset{\boldsymbol{x}}{\mathbf{E}}[g(\boldsymbol{x}) \mid \beta \cdot \boldsymbol{x}=0]|+\frac{1}{p}|\underset{\boldsymbol{x}}{\mathbf{E}}[g(\boldsymbol{x}) \mid \beta \cdot \boldsymbol{x}=1]|+\cdots+\frac{1}{p}|\underset{\boldsymbol{x}}{\mathbf{E}}[g(\boldsymbol{x}) \mid \beta \cdot \boldsymbol{x}=p-1]|
$$

$$
=: \frac{1}{p}\left(\left|\delta_{0}\right|+\left|\delta_{1}\right|+\cdots+\left|\delta_{p-1}\right|\right) .
$$

Since this is at least $\eta$, by averaging we could conclude that there exists some $c \in \mathbb{F}_{p}$ such that $\left|\delta_{c}\right| \geq \eta$. This is not quite what we're looking for, though, since we want to ensure we get a positive $\delta_{c}$. So simply further observe that the average of all the $\delta_{i}$ 's is clearly $\mathbf{E}[g]$, which is 0 . Hence from

$$
\eta \leq \frac{1}{p}\left(\left|\delta_{0}\right|+\left|\delta_{1}\right|+\cdots+\left|\delta_{p-1}\right|\right)
$$

we can deduce

$$
\eta \leq \frac{1}{p}\left(\left|\delta_{0}\right|+\delta_{0}+\left|\delta_{1}\right|+\delta_{1}+\cdots+\left|\delta_{p-1}\right|+\delta_{p-1}\right),
$$

whence there exists $c \in \mathbb{F}_{p}$ such that $\left|\delta_{c}\right|+\delta_{c} \geq \eta$, whence $\delta_{c} \geq \eta / 2$.
Thus $\mathbf{E}[g \mid \beta \cdot \boldsymbol{x}=c] \geq \eta / 2$, which implies $\mathbf{E}[f \mid \beta \cdot \boldsymbol{x}=c] \geq \mu+\eta / 2$.
We now know that either the random strategy is guaranteed to find a $3-\mathrm{AP}$ in $A$, or there must exists some hyperplane $\beta \cdot x=c$ on which the density of $A$ is at least $\mu+\mu^{2} / 4$.

Finally, a trick: This hyperplane is isomorphic to $\mathbb{F}_{3}^{n-1}$. Specifically, there is some invertible affine transformation $T$ which maps the hyperplane (affine subspace of codimension 1) to $\mathbb{F}_{3}^{n-1}$.

This $T$ will take $A$ to some $A^{\prime} \subseteq \mathbb{F}_{3}^{n-1}$ of density at least $\mu+\mu^{2} / 4$. Now suppose we find a 3 -AP $(x, y,-(x+y))$ in $A^{\prime}$. Then $\left(T^{-1} x, T^{-1} y, T^{-1}(-(x+y))\right.$ are in $A$, and we claim they form a 3 -AP. To see this, note that $T^{-1}$ is just a linear transformation plus a "constant term". Adding the same constant to each element in a 3-AP still gives a $3-\mathrm{AP}$, so we may assume that $T^{-1}$ is simply linear. But then $T^{-1}(-(x+y))=-\left(T^{-1} x+T^{-1} y\right)$.

We are now essentially home-free. We can repeat the argument on $A^{\prime}$, and then on $A^{\prime \prime}$ if necessary, and then on $A^{\prime \prime \prime}$ if necessary, ... At each step, we either find a 3-AP, or the density increases by $\mu_{i+1} \geq \mu_{i}+\left(\mu_{i}\right)^{2} / 4$. Now of course, the density can never go above 1 . But this easily implies that we can iterate at most $8 / \mu_{0}$ times:

$$
\mu_{0} \stackrel{4 / \mu_{0} \text { times }}{\longmapsto} 2 \mu_{0} \stackrel{2 / \mu_{0} \text { times }}{\longmapsto} 4 \mu_{0} \stackrel{1 / \mu_{0} \text { times }}{\longmapsto} 8 \mu_{0} \stackrel{1 / 2 \mu_{0} \text { times }}{\longmapsto} 16 \mu_{0} \stackrel{1 / 4 \mu_{0} \text { times }}{\longmapsto} 32 \mu_{0} \stackrel{1 / 8 \mu_{0} \text { times }}{\longmapsto} 64 \mu_{0} \stackrel{1 / 16 \mu_{0} \text { times }}{\longmapsto} \cdots
$$

Every time we iterate, we lose a dimension, and hence a factor of 3 on the size of the ambient space. There is just one catch: Corollary 3.6 only works provided that $\mu_{i} \geq 2 / N^{1 / 3}$. If the ambient space shrinks too quickly, we could get caught in a position where $\mu_{i}$ is no longer greater than $2 / N^{1 / 3}$. Now we iterate at most $8 / \mu_{0}$ times, so the ambient space always has size at least $3^{-8 / \mu_{0}} \cdot N$. And the density is always at least the initial density, $\mu_{0}$. Hence we will be guaranteed to find a 3 -AP so long as

$$
\mu_{0} \geq 2 /\left(\frac{N}{3^{8 / \mu_{0}}}\right)^{1 / 3} \geq C^{1 / \mu_{0}} / N^{1 / 3}
$$

And indeed this holds if $\mu_{0} \geq O(1 / \log N) \geq O(1 / n)$.

