An Introduction to Szemeredi’s Theorem and Green-Tao’s Theorem.

Tatchai Titichetrakun

MUIC Math Seminar

cocoanutmath@gmail.com

January 24, 2018
What is additive combinatorics?

- Count / Estimate additive structures in sets e.g. Arithmetic Progressions.
- Structure of sets with some additive properties. e.g. a set $A$ with small $|A + A|$.
- Some models are easier and can be served as a role model e.g. vector space model (so-called Finite field model $\mathbb{F}_p^n$, where $p$ is fixed and $n$ is large). More tools and less technical than $\mathbb{Z}_N$.
- Related to other areas such as Harmonic Analysis, Ergodic Theory, Linear Algebra (polynomial methods), Computer Science.
Basic Notion

- If $X$ is a finite set, we write the average as
  \[ \mathbb{E}_{x \in X} f(x) = \frac{1}{|X|} \sum_{x \in X} f(x) \]

- Cauchy-Schwartz’s Inequality
  \[
  \left( \mathbb{E}_{x \in X, y \in Y} f(x)g(y) \right)^2 \leq \left( \mathbb{E}_{x \in X} f(x) \right)^2 \left( \mathbb{E}_{y \in Y} g(y) \right)^2
  \]
  \[
  = \mathbb{E}_{x \in X, x' \in X'} f(x)f(x') \mathbb{E}_{y \in Y, y' \in Y'} g(y)g(y')
  \]
  (Vertex sets doubling)
Szemeredi’s Theorem

**Theorem (Infinite version)**

Suppose \( A \subseteq \mathbb{Z} \) and \( A \) has positive upper density i.e.

\[
\limsup_{N \to \infty} \frac{|A \cap \{1, \ldots, N\}|}{N} > 0
\]

Then \( A \) contains a (non trivial) \( k \)-term arithmetic progressions.

**Theorem (Finite version)**

Given \( 0 < \alpha < 1 \). If \( A \subseteq \{1, \ldots, N\} \), \( |A| \geq \alpha N \).

Suppose \( N \geq c(k, \alpha) \)

Then \( A \) contains a (non trivial) \( k \)-term arithmetic progressions.

Note: Varnavides’s average argument: Actually \( A \) contains at least \( c(\alpha, k)N^2 \) \( k \)-arithmetic progressions.
Suppose $f : \mathbb{Z}_N \rightarrow [0, 1]$ and $\mathbb{E}_{x \in \mathbb{Z}_N} f(x) \geq \alpha$. Then

$$\mathbb{E}_{x, t \in \mathbb{Z}_N} f(x)f(x + t) \ldots f(x + (k - 1)t) \geq c(\alpha, k) - o(1)$$

All of these statements of Szemeredi’s Theorem are equivalent.
Erdos Conjecture.
Szemeredi’s Theorem

- Case $k = 3$ is proved by Roth (1953).
- Case $k > 3$:
  - Szemeredi - Combinatorics (1975)
  - Furstenberg - Ergodic Theory (1977)

Quantitative bounds: Let $r_k(N)$ denote the size of the largest subset of $\{1, \ldots, N\}$ not containing $k$-AP.

- $r_3(N) \geq N e^{c \sqrt{\log N}}$ (Berhrand 1946)
- $r_3(N) \leq c N (\log \log N)^4 \log N$ (Bloom 2016)
- $r_4(N) \leq c N \log c N (\log \log N)^2 - 2^{k+9}$. (Gowers 2001)
- $r_3(F_3^n) \leq O(2.756^n)$ (Croot-Lev-Pach, Ellenberg-Gijswijt 2016). Note: this is the same shape as the lower bound obtained by Edel (2004). This is done using an algebraic technique called Polynomial Method.
Szemeredi’s Theorem

- Case \( k = 3 \) is proved by Roth (1953).
- Case \( k > 3 \):
  - Szemeredi - Combinatorics (1975)
  - Furstenberg - Ergodic Theory (1977)
- Quantitative bounds: Let \( r_k(N) \) denote the size of the largest subset of \( \{1, \ldots, N\} \) not containing \( k \)-AP.
  - \( r_3(N) \geq N \exp(-c\sqrt{\log N}) \) (Bertrand 1946)
  - \( r_3(N) \leq c \frac{N(\log \log N)^4}{\log N} \) (Bloom 2016)
  - \( r_4(N) \leq \frac{N}{\log^c N} \) (Green-Tao 2017) Note: currently only know that \( c < 1 \).
  - \( r_k(N) \leq c \frac{N}{(\log \log N)^{2^{-2k+9}}} \) (Gowers 2001)
  - \( \frac{r_3(F_3^n)}{3^n} \leq O(2.756^n) \) (Croot-Lev-Pach, Ellenberg-Gijswijt 2016). Note: this is the same shape as the lower bound obtained by Edel (2004). This is done using an algebraic technique called Polynomial Method.
Finite Fourier Analysis

- \( f : \mathbb{Z}_N \to \mathbb{C} \), Dual Group \( \hat{G} \) is the group of character \( e_r : G \to S^1 \) where \( r \in G \). \( e_r(x) := e^{2\pi i (r \cdot x)} \)
- linear phase functions on \( \mathbb{Z}_N \): \( e_r(x) = e^{2\pi i r x / N} \),
- Fourier transform:
  \[ \hat{f} : \hat{\mathbb{Z}}_N = \mathbb{Z}_N \to \mathbb{C}, \hat{f}(r) = \mathbb{E}_x f(x)e^{r \cdot x} \]
- (Alternately) \( f : \mathbb{Z} \to \mathbb{C}, \hat{f} : S^1 \to S^1, \hat{f}(r) = \int_0^1 f(x)e^{-2\pi i r x} \, dx \)
Finite Fourier Analysis

- \( f: \mathbb{Z}_N \to \mathbb{C} \), Dual Group \( \hat{G} \) is the group of character \( e_r: G \to S^1 \)
  where \( r \in G \). \( e_r(x) := e^{(r \cdot x)} := e^{2\pi i(r \cdot x)} \)
- linear phase functions on \( \mathbb{Z}_N \): \( e_r(x) = e^{2\pi irx/N} \)
- Fourier transform:
  \[
  \hat{f}: \mathbb{Z}_N = \mathbb{Z}_N \to \mathbb{C}, \hat{f}(r) = \mathbb{E}_x f(x) e^{(r \cdot x)}
  \]
- (Alternately) \( f: \mathbb{Z} \to \mathbb{C} \), \( \hat{f}: S^1 \to S^1 \)
  \( \hat{f}(r) = \int_0^1 f(x) e^{-2\pi irx} \, dx \)
- Orthogonal relation \( \langle e_m, e_n \rangle = \begin{cases} 0 & \text{if } m \neq n \\ 1 & \text{if } m = n \end{cases} \)
- Normalization \( \|f\|_p^p = \mathbb{E}_x |f(x)|^p \), \( \|\hat{f}\|_p^p = \sum_r |\hat{f}(r)|^p \)
  \[
  \langle f, g \rangle = \mathbb{E}_x f(x) g(x), \quad \langle \hat{f}, \hat{g} \rangle = \sum_r \hat{f}(r) \hat{g}(r)
  \]
- convolution \( f * g(x) = \mathbb{E}_y f(y) g(x - y) \)
  \( \hat{f} \hat{g} = \hat{f \ast g} \)
- Plancherel: \( \langle f, g \rangle = \langle \hat{f}, \hat{g} \rangle \).
Gowers Uniformity norms ($U^k$—norms)

- It can be defined in terms of cubic structure $\{0, 1\}^k$ or iterated additive differences.
- Example $k = 2, 3$.

$$
\|f\|_{U^2}^4 = \mathbb{E}_{x, y, x', y'} f(x + x')f(x + y')f(y + x')f(y + y')
= \mathbb{E}_{x, h_1, h_2} f(x)f(x + h_1)f(x + h_2)f(x + h_1 + h_2)
$$
Gowers Uniformity norms ($U^k$—norms)

- It can be defined in terms of cubic structure \(\{0, 1\}^k\) or iterated additive differences.
- Example \(k = 2, 3\).

\[
\|f\|_{U^2}^4 = \mathbb{E}_{x,y,x',y'} f(x + x')f(x + y')f(y + x')f(y + y') = \mathbb{E}_{x,h_1,h_2} f(x)f(x + h_1)f(x + h_2)f(x + h_1 + h_2)
\]

\[
\|f\|_{U^3}^8 = \mathbb{E}_{x,y,z,x',y'z'} f(x + y + z) \\
\times f(x' + y + z)f(x + y' + z)f(x + y + z') \\
\times f(x' + y' + z)f(x' + y + z')f(x' + y' + z) \\
\times f(x' + y' + z') \\
= \mathbb{E}_{x,h_1,h_2,h_3} f(x)f(x + h_1)f(x + h_2)f(x + h_3) \\
\times f(x + h_1 + h_2)f(x + h_1 + h_3)f(x + h_2 + h_3)f(x + h_1 + h_2 + h_3)
\]

- This norms corresponds to generalized inner product of \(2^k\) functions.
- Additive differences can define linear function, quadratic functions etc.
Uniformity norms

- Inverse theorem: Find a set of structure (functions) $\mathcal{F}_d$ such that if $\|f\|_{U_d} > \eta$ then $\langle f, g \rangle \geq c(\eta)$ for some $g \in \mathcal{F}_d$. 
Uniformity norms

- Inverse theorem: Find a set of structure (functions) $\mathcal{F}_d$ such that if $\|f\|_{U^d} > \eta$ then $\langle f, g \rangle \geq c(\eta)$ for some $g \in \mathcal{F}_d$.
- $\|f\|_{U^d}$ is small means $f$ does not correlate with any structure in $\mathcal{F}_d$.
- $\|f\|_{U^d} \leq \|f\|_{U^{d+1}}$ i.e. $\mathcal{F}_d \subset \mathcal{F}_{d+1}$.
- $\|f\|_{U^2} = \|\hat{f}\|_4$. 

Proof:

\[ \|f\|_4^2 \leq \|\hat{f}\|_2^2 \leq \|\hat{f}\|_\infty \|f\|_2 \leq \|\hat{f}\|_\infty \|f\|_2. \]
Uniformity norms

- Inverse theorem: Find a set of structure (functions) $\mathcal{F}_d$ such that if $\|f\|_{U^d} > \eta$ then $\langle f, g \rangle \geq c(\eta)$ for some $g \in \mathcal{F}_d$.
- $\|f\|_{U^d}$ is small means $f$ does not correlate with any structure in $\mathcal{F}_d$.
- $\|f\|_{U^d} \leq \|f\|_{U^{d+1}}$ i.e. $\mathcal{F}_d \subset \mathcal{F}_{d+1}$.
- $\|f\|_{U^2} = \|\hat{f}\|_4$. **Proof:**
  \[
  \|f\|_{U^2}^4 = \mathbb{E}_{x+y=z+w} f(x)f(y)f(z)f(w)
  = \langle f \ast f, f \ast f \rangle
  = \langle \hat{f} \ast \hat{f}, \hat{f} \ast \hat{f} \rangle
  = \sum_r |\hat{f}(r)|^4 = \|\hat{f}\|_{L^4}^4
  \]

**Corollary (U^2 inverse theorem)**

Suppose $f$ is 1-bounded and $\|f\|_{U^2} \geq c$ then there is a linear phase function $e_r$ such that $\hat{f}(r) = \langle f, e_r \rangle \geq c^2$

**Proof:** $c^2 \leq \|\hat{f}\|_4^2 \leq \|\hat{f}\|_\infty \|\hat{f}\|_2 = \|\hat{f}\|_\infty \|f\|_2 \leq \|\hat{f}\|_\infty$
Observation: If $|\widehat{1}_A - \alpha| \leq \alpha^2/2$ then $A$ has roughly expected numbers of 3-term arithmetic progressions.

$$\sum_{x \in A, y \in A, z \in A} \frac{1}{N} \sum_{r=0}^{N} e^{-\frac{2\pi i}{N}(x+y-2z)r} = \frac{1}{N} \sum_{r} \widehat{1}_A(r)^2 \widehat{1}_A(-2r)$$

$$= \alpha^3 N^2 + \frac{1}{N} \sum_{r \neq 0} \widehat{1}_A(r)^2 \widehat{1}_A(-2r)$$

Function $f$ is $\alpha-$uniform if $\sup_{r \neq 0} |\hat{f}(r)| \leq \alpha$.

$$f(x) = \sum_r \hat{f}(r) e^{rx} = \sum_{r: \hat{f}(r) \text{ large}} + \sum_{r: \hat{f}(r) \text{ medium}} + \sum_{r: \hat{f}(r) \text{ small}}$$

Spectral graph theory: discrepancy is measured by the second largest eigenvalue of the adjacency matrix (cheeger-type discrepancy bound).

Classical results - every odd numbers is a sum of 3 primes. Counting prime 3-AP (Vinogradov, Chowla, Van der Corput 1930s).
Density Increment Method ($L^\infty$-method)

Suppose we are working in $\mathbb{F}_p^n$. If $|\widehat{1_A}(t)| \geq \alpha^2/2$. By Fourier expansion and averaging over cosets of $\langle t \rangle^\perp$, we have:

Example (\textit{i}th-step of iteration)

Set $A_i \subseteq \mathbb{F}_3^n$ with density $\alpha_i$ on $H_i$. Then either $A_i$ is $\alpha_i^2/2$ uniform or there is a subspace $H_{i+1} \leq H_i$ of codimension 1 such that for some $x \in H_i$ we have $A_i \cap (H_{i+1} + x)$ has density $\geq \alpha_i + \alpha_i^2/4$. 
Density Increment Method ($L_\infty$-method)

Suppose we are working in $\mathbb{F}_p^n$. If $|\widehat{1_A}(t)| \geq \alpha^2/2$. By Fourier expansion and averaging over cosets of $\langle t \rangle^\perp$, we have:

Example ($i^{th}$—step of iteration)

Set $A_i \subseteq \mathbb{F}_3^n$ with density $\alpha_i$ on $H_i$. Then either $A_i$ is $\alpha_i^2/2$ uniform or there is a subspace $H_{i+1} \leq H_i$ of codimension 1 such that for some $x \in H_i$ we have $A_i \cap (H_{i+1} + x)$ has density $\geq \alpha_i + \alpha_i^2/4$. 

- We have to find a good notion of structure (with complexity $\omega$ associated with the structure) and a uniformity norm of functions. Suppose $S$ is a structure and $A$ has density $\delta_S(A)$ on $S$ then one the following holds.
  - (generalized von Neumann Theorem) If the function $1_A - \delta(A)$ is small in the uniformity norm then the set $A$ should contains the required configuration.
  - If the function has large uniformity norm then you can find a structure $S'$ with $\omega(S') \leq \omega(S) + 1$ such that the density of $A$ on $S'$ is increased by a fixed amount.
We want to use the idea of the iterations in vector space but in general the group $\mathbb{Z}_N$ has no subspace.

In general, for $\Gamma = \{r_1, \ldots, r_d\} \subseteq G^*$, we define the Bohr set

$$B(\Gamma, \varepsilon) = \{x \in G : \|\frac{r_j \cdot x}{N}\|_{\mathbb{R}/\mathbb{Z}} \leq \varepsilon, 1 \leq j \leq d\}$$

This is an approximation of the subspace $\Gamma^\perp$. 

Bohr sets $B(K, \rho) \subseteq \mathbb{Z}_N$ contain AP of size $\geq \rho N^{1/K}$ centered at 0. In fact, it contains a proper multidimensional arithmetic progression. 

Bourgainization: Smooth Bohr sets $B + B' \approx B$. 

From finite fields to integers
We want to use the idea of the iterations in vector space but in general the group $\mathbb{Z}_N$ has no subspace.

In general, for $\Gamma = \{r_1, \ldots, r_d\} \subseteq G^*$, we define the Bohr set

$$B(\Gamma, \varepsilon) = \{x \in G : \| \frac{r_j \cdot x}{N} \|_{\mathbb{R}/\mathbb{Z}} \leq \varepsilon, 1 \leq j \leq d\}$$

This is an approximation of the subspace $\Gamma^\perp$.

For Roth’s theorem we use the following Bohr set

$$B(r, \varepsilon) = \{x \in \mathbb{Z}_N : |1 - e^{rx}| \leq \|rx/N\|_{\mathbb{R}/\mathbb{Z}} \leq \varepsilon\}$$

We can decompose $\{1, \ldots N\}$ into long sub arithmetic progressions $P_i$ of length $N^{1/K}$ such that $x$ on each $P_i$ that $e(r \cdot x)$ are roughly in the same direction.

Bohr sets $B(K, \rho) \subseteq \mathbb{Z}_N$ contains AP of size $\geq \rho N^{1/K}$ centered at 0. In fact, it contains a proper multidimensional arithmetic progression.

Bourgainization: Smooth Bohr sets $B + B' \approx B$. 

Example (Fourier uniformity is not enough for 4-AP.)

There exists $\varepsilon > 0$ such that for every $\delta$

$$A = \{ x \in \mathbb{F}_p^n : x \cdot x = 0 \}$$

is $\delta$– Fourier uniform but has more than $\alpha^3 > \alpha^4 + \varepsilon$ proportions of 4-AP.
4-term arithmetic progressions

Example (Fourier uniformity is not enough for 4-AP.)

There exists $\varepsilon > 0$ such that for every $\delta$

$$A = \{ x \in \mathbb{F}_p^n : x \cdot x = 0 \}$$

is $\delta-$ Fourier uniform but has more than $\alpha^3 > \alpha^4 + \varepsilon$ proportions of 4-AP.

- 3-AP has linear discrepancy: $x - 2(x + d) + (x + 2d) = 0$
- 4-AP has quadratic discrepancy

$$x^2 - 3(x + d)^2 + 3(x + 2d)^2 - (x + 3d)^2 = 0.$$ 

This is the only obstruction when using Fourier transform to count 4-AP.

- Quadratic exponentials are still too special to form complete $U^3$-obstruction. We will also need quadratic functions defined locally on Bohr sets.
- We have too many quadratic exponential functions to form orthonormal basis.
Szemeredi’s Theorem: Multidimensional case

- Higher dimensions $\mathbb{Z}^d$: Arithmetic progressions replaced by Affine copies of any finite configurations $F = \{v_0, \ldots, v_k\} \subseteq \mathbb{Z}^d$ i.e. $x + tF$. 
Simplices in $\mathbb{Z}^d$ can be parametrized in $(d + 1)$-partite hypergraph, by pairwise linearly independent linear forms.

Theorem (Triangle Removal Lemma; Ruzsa-Szemeredi 1976)

If $G$ is a graph of $n$ vertices contains fewer than $\varepsilon n^3$ triangles then it is possible to delete $o_{\varepsilon \to 0}(n^2)$ edges from $G$ to make it triangle free.

Equivalently, If $G$ more than $\delta n^2$ edge-disjoint triangles for some $0 < \delta < 1$. Then it in fact contains $c(\delta)n^3$ triangles.
Theorem (Triangle removal lemma; Functional version)

Let \((X, \mu_X), (Y, \mu_Y), (Z, \mu_Z)\) be probability spaces. Suppose 
\(f_1 : X \times Y \to [0, 1], f_2 : Y \times Z \to [0, 1], f_3 : X \times Z \to [0, 1]\) are measurable functions. Let \(\varepsilon > 0\). Suppose 
\[
\Lambda_3(f_1, f_2, f_3) := \int_X \int_Y \int_Z f_1(x, y)f_2(y, z)f_3(x, z)d\mu_X d\mu_Y d\mu_Z \leq \varepsilon
\]

Then we can find measurable functions \(\tilde{f}_1 : X \times Y \to [0, 1], \tilde{f}_2 : Y \times Z \to [0, 1], \tilde{f}_3 : X \times Z \to [0, 1]\) such that \(\|f_i - \tilde{f}_i\|_1 = o_{\varepsilon \to 0}(1)\) for \(i = 1, 2, 3\) such that \(\tilde{f}_1 \tilde{f}_2 \tilde{f}_3\) vanishes entirely, in particular \(\Lambda_3(\tilde{f}_1, \tilde{f}_2, \tilde{f}_3) = 0\).

High complexity expression \(\Lambda_3(f, g, h)\) can be manipulated purely in lower complexity operations.
A σ—algebra (or Factor) on a finite set \( X \) can be considered as a finite partition into atoms of a set \( X \).

Complexity of a sigma-algebra is the number of its generator.

If \( \mathcal{B}' \supset \mathcal{B} \) then we say that \( \mathcal{B}' \) is finer or is a refinement of \( \mathcal{B} \). Note that \( \mathcal{B} \cap \mathcal{B}' \) is a common refinement of both \( \mathcal{B} \) and \( \mathcal{B}' \).

Define the function \( \mathbb{E}(f \mid \mathcal{B})(x) = \frac{1}{|\mathcal{B}(x)|} \sum_{x \in \mathcal{B}(x)} f(x) \)

\[
\begin{array}{cccc}
1 & 2 & 3 & 4 \\
5 & 6 & 7 & 8 \\
9 & 10 & 11 & 12 \\
13 & 14 & 15 & 16 \\
\end{array}
\begin{array}{cc}
14/4 & 22/4 \\
46/4 & 54/4 \\
\end{array}
\begin{array}{c}
136/16 \\
\mathcal{B}_{\text{trivial}} \\
\end{array}
A function is $\mathcal{B}$–measurable if it is constant on each atom of $\mathcal{B}$. For $\mathcal{B}' \supset \mathcal{B}$, the space of $\mathcal{B}'$–measurable functions form a subspace of $L^2(\mathcal{B})$. $\mathbb{E}(f|\mathcal{B}')$ is the orthogonal projection of $f$ into $L^2(\mathcal{B}')$.

$\mathbb{E}(f|\mathcal{B})$ is the information on $f$ probed by $\mathcal{B}$.

$\|f - \mathbb{E}(f|\mathcal{B})\|_{U^k}$ small means there is not much information (structure, recognizable by $\| \cdot \|_{U^k}$) left after probed by $\mathcal{B}$.

$\|\mathbb{E}(f|\mathcal{B})\|_{L^2(\mu)}$ is called energy. It is bounded if $f$ is bounded.

**Example** $\mathbb{E}(1_A|\mathcal{B}_{\text{trivial}})(x) = |A|/|X|$.

If $A$ is dense random set with density $\alpha$ then $\|1_A - \alpha\|_{U^k} = \|1_A - \mathbb{E}(1_A|\mathcal{B}_{\text{trivial}})\|_{U^k}$ is small.
Energy increment

- $U^k$- norm uniform functions are orthogonal to lower order sets/functions.

$$\|f(x, y, z)\|_{U^3}^8 = \mathbb{E}_{x, y, z, x', y', z'} f(x, y, z) f(x', y, z) f(x, y', z) f(x, y, z') f(x', y', z) f(x, y', z') f(x', y', z')$$

$$\|g\|_{U^3} > \eta \text{ implies } g \text{ correlates with structure,}$$

$$\mathbb{E}_{x, y, z} g(x, y, z) F(x, y) G(y, z) H(x, z) > c\eta^8,$$
Energy increment

- $U^k$- norm uniform functions are orthogonal to lower order sets/functions.

$$\|f(x, y, z)\|_{U^3}^8 = \mathbb{E}_{x,y,z,x',y'z'} f(x, y, z) f(x', y, z) f(x, y', z) f(x, y, z')$$
$$\times f(x', y', z) f(x', y, z') f(x, y', z') f(x', y', z')$$

$\|g\|_{U^3} > \eta$ implies $g$ correlates with structure,

$$\mathbb{E}_{x,y,z} g(x, y, z) F(x, y) G(y, z) H(x, z) > c\eta^8,$$

By Fubini’s Theorem, we can think of $F, G, H$ as sets.

- Cooperate the set $F(x, y) G(y, z) H(x, z)$ to define a sigma-algebra $B'$.

- $\|f - \mathbb{E}(f|B)\|_{U^3} > \eta \implies \exists$ sigma algebras $B'$ (complexity increased at most 1) such that

$$\|f - \mathbb{E}(f|B')\|_{L^2}^2 \geq \|f - \mathbb{E}(f|B)\|_{L^2}^2 + c\eta^{16}$$
\[ \| f - \mathbb{E}(f|\mathcal{B}) \|_{U^3} > \eta \implies \exists \ \text{sigma algebras } \mathcal{B}' \ (\text{complexity increased at most } 1) \text{ such that} \]

\[ \| f - \mathbb{E}(f|\mathcal{B}') \|^2_{L^2} \geq \| f - \mathbb{E}(f|\mathcal{B}) \|^2_{L^2} + c\eta^{16} \]

**Proof:**

\[ \| \mathbb{E}(f|\mathcal{B}') \|^2_{L^2} - \| \mathbb{E}(f|\mathcal{B}) \|^2_{L^2} = \| \mathbb{E}(f|\mathcal{B}') - \mathbb{E}(f|\mathcal{B}) \|^2_{L^2} \]

\[ \geq |\langle \mathbb{E}(f|\mathcal{B}') - \mathbb{E}(f|\mathcal{B}), F(x, y)G(y, z)H(x, z) \rangle|^2 \]

\[ = |\langle f - \mathbb{E}(f|\mathcal{B}), F(x, y)G(y, z)H(x, z) \rangle|^2 \]

\[ > c\eta^{16} \]
After finishing the iteration of energy increment, with $g = f = E(f|B)$, one has (This theorem holds for any $U^k$. with some changes in parameters.)

**Theorem (Koopman-von Neumann for $U^2$)**

Let $\varepsilon > 0$ and $f : G \to [0, 1]$ is then there is a linear factor $B$ with $\dim B \leq \varepsilon^{-4}$ such that

$$f = E(f|B) + g, \|g\|_{U^2} \leq \varepsilon$$

**Example:** with $k = 2$, $G = \mathbb{F}_2^n$, we will need linear factors $B$:

- Let $r_1, \ldots, r_k \in G$. Atoms of $B$ is defined to be cosets of $\langle r_1, \ldots, r_k \rangle^\perp$
- If $r_1, \ldots, r_k$ are linearly independent then we say that $\dim(B) = k$.
  Numbers of atoms is $2^k$. Each atom has codim $= k$. 
**Theorem**

Let \( f \to [0, 1], \varepsilon > 0 \) Let \( F = F_\varepsilon \) be a growth function, there is a factors \( B \) with \( \dim(B) \ll O_{F,\varepsilon}(1) \) such that

\[
f = f_{str} + f_{psd} + f_{sml}
\]

- \( f_{str} = \mathbb{E}(f|B) \)
- \( \|f_{psd}\|_{U^k} < F(\dim B)^{-1} \)
- \( \|f_{err}\|_{L^2} < \varepsilon \)
- \( f_{str} + f_{sml} \in [0, 1] \)

We will need \( F \) to be exponential and we will obtain the Tower-type bound.

\( M(\varepsilon) = \) number of atoms.

- Gowers (1997) : \( M(\varepsilon) = \text{Tower}(\Theta(\varepsilon^{-c})) \).
- Fox, Miklos, Lovasz (2017) : \( M(\varepsilon) = T(\Theta(\varepsilon^{-2})) \)
\( \varepsilon_0 = 1, B_0 = B_{\text{trivial}}. \)

If we apply Koopman-von Neumann to \( B_i \) we obtain a refinement \( B_{i+1} \) such that \( \dim B_{i+1} \leq \dim B_i + \varepsilon_i^{-4} \). So

\[
\dim(B_i) \leq \varepsilon_0^{-4} + \cdots + \varepsilon_{i-1}^{-4}
\]

Let \( \varepsilon_i = F(\varepsilon_0^{-4} + \cdots + \varepsilon_{i-1}^{-4})^{-1}. \)

So

\[
\| f - \mathbb{E}(f|B_{i+1}) \|_{U^2} \leq \varepsilon_i = F(\varepsilon_0^{-4} + \cdots + \varepsilon_{i-1}^{-4})^{-1} \leq F(\dim(B_i))^{-1}
\]

By Pigeonhole Principle \( \exists i \leq \varepsilon^{-2}, \| \mathbb{E}(f|B_{i+1}) \|_{L^2}^2 - \| \mathbb{E}(f|B_i) \|_{L^2}^2 \leq \varepsilon^2. \)

We have the decomposition \( f = f_{\text{str}} + f_{\text{psd}} + f_{\text{sml}} \) with

\[
f = \mathbb{E}(f|B_i) + [f - \mathbb{E}(f|B_{i+1})] + [(\mathbb{E}(f|B_{i+1})) - \mathbb{E}(f|B_i)]
\]
Generalized von Neumann Theorem

- We wish to show the expression that counts the number of 3-term arithmetic progressions by the $U^2$-norm i.e.

$$\Lambda(f, g, h) = \mathbb{E}_{x, y} f(2x - y)g(x)h(y) \leq \min\{\|f\|_{U^2}, \|g\|_{U^2}, \|h\|_{U^2}\}$$

Apply Cauchy-Schwartz’s inequality in $y$ and use that $|g| \leq 1$ we have

$$|\Lambda|^2 \leq \mathbb{E}_{x, y, y'} f(2x - y)f(x - 2y')h(y)h(y')$$

Now Apply the cauchy schwartz’s inequality one more time in $x$-variables and use that $|h| \leq 1$ we obtain

$$|\Lambda|^4 \leq \mathbb{E}_{x, x', y, y'} f(2x - y)f(2x' - y)f(2x - y')f(2x' - y') = \|f\|_{U^2}$$

- **Complexity of linear systems (Gowers-Wolf):** True-complexity is the least $s$ such that all $\psi^s$ are linearly independent.

$$(True\ \text{complexity}) \leq (\text{Cauchy} - \text{Schwartz's complexity})$$
Decomposing $\Lambda(f, g, h)$

Decompose

$$\Lambda((f_{str} + f_{sml} + f_{psd}), (g_{str} + g_{sml} + g_{psd}), (h_{str} + h_{sml} + h_{psd}))$$

- Terms with $f_{psd}$ is controlled by $\|f_{psd}\|_{U^2} = O(1/F(M))$.
- Consider

$$\Lambda((f_{srt} + f_{sml}), (g_{srt} + g_{sml}), (h_{srt} + h_{sml}))$$

$$\geq \frac{|V|^3}{|G|^3} \frac{1}{|V|^3} \sum_{x_i \in a_i + V} (f_{str}(x_1)g_{str}(x_2)h_{str}(x_3) - 3\varepsilon^{10})$$

$$\geq 2^{-3\text{dim}(B)}\varepsilon^{10}, \text{ if the sum of } f_{str}, g_{str}, h_{str} \text{ is bigger than error term}$$

- Choose $F(d) = 2^{3d} \varepsilon^{-100}$

Note: Calculation of $\Lambda(f_{str}, g_{str}, h_{str})$ is called the counting lemma.
Density Increment Method

- Roth’s Theorem (Roth 1953) Szemeredi’s Theorem (Gowers, 1998, 2001)
- Sarkozy’s Theorem: $n, n + r^2$.
- Shrkedov’s corner theorem in $\mathbb{Z}^2$ (2005)

More general patterns - Energy increment method ($L^2$—method)
- Ergodic Theory: Polynomial pattern, Multidimension, Primes, etc.
Furstenberg’s Corresponding Principle: Szemeredi’s Theorem follows from the following estimate for measure preserving system \((X, B, \mu, T : X \to X)\) i.e. \(\mu(T^{-1}(X)) = \mu(x)\).

\[
\lim \inf_{n \to \infty} \frac{1}{N} \sum_{n=1}^{N} f \cdot f(Tx) \cdot \ldots \cdot f(T^{k-1}x) > 0
\]

i.e. if \(\mu(A) > 0\) then

\[
\mu(A \cap T^{-n}(A) \cap \ldots \cap T^{-kn}(A)) > 0
\]

This is called multiple recurrence or non-conventional recurrence.

Furstenburg showed in 1976 that the liminf is positive. It was eventually shown in 2005 (Host-Kra) that the limit of the average actually exists in \(L^2\). This leads to structural theorem of measure preserving system.
Factors

- Ergodicity is a notion in the hierarchy of Mixing which is very useful. For example, every measuring preserving system can be decomposed into ergodic component.
- Idea of Factor is the decomposition $L^2(X, B, \mu) = H \oplus H^\perp$ into structural part and mixing part. $f = \mathbb{E}(f|B) + (f - \mathbb{E}(f|B))$.
- In case $k = 3$, we have decomposition into Kronecker Factor and weak-mixing part. Kronecker’s factor corresponds rotation on the unit circle (which can be thought as Fourier transforms or Eigenfunctions of $T$).
- The structural part of Higher $U^k$- norm corresponds to Nilsystems.
Theorem (Arithmetic Progression in Primes: Green-Tao 2004)

Let \( f : \mathbb{Z}_N \rightarrow \mathbb{R}_{\geq 0}, \) \( f(x) \leq \nu(x) \) for some pseudorandom measure \( \nu \) with \( \mathbb{E}_{x \in \mathbb{Z}_N} f(x) \geq \alpha \) then

\[
\mathbb{E}_{x, y \in \mathbb{Z}_N} f(x)f(x + y) \ldots f(x + (k - 1)y) \geq c(\alpha, k) - o_M(1).
\]
Theorem (Arithmetic Progression in Primes: Green-Tao 2004)

Let \( f : \mathbb{Z}_N \to \mathbb{R}_{\geq 0}, f(x) \leq \nu(x) \) for some pseudorandom measure \( \nu \) with \( \mathbb{E}_{x \in \mathbb{Z}_N} f(x) \geq \alpha \) then

\[
\mathbb{E}_{x,y \in \mathbb{Z}_N} f(x)f(x+y)...f(x+(k-1)y) \geq c(\alpha, k) - o_M(1).
\]

- **Hardy-Littlewood Conjecture**: Apart from local obstruction - Primes should behave randomly (no conspiracy among primes). W-trick is a way to get rid of local obstruction for small primes (large primes have less effect).
- **Almost primes** (here-primes with only large prime factors) is much easier to understand than the primes and primes has positive upper density on almost primes.
Theorem (Arithmetic Progression in Primes: Green-Tao 2004)

Let \( f : \mathbb{Z}_N \to \mathbb{R}_{\geq 0}, f(x) \leq \nu(x) \) for some pseudorandom measure \( \nu \) with
\[
E_{x \in \mathbb{Z}_N} f(x) \geq \alpha \text{ then }
\]
\[
E_{x,y \in \mathbb{Z}_N} f(x)f(x+y)\ldots f(x+(k-1)y) \geq c(\alpha, k) - o_M(1).
\]

- Hardy-Littlewood Conjecture: Apart from local obstruction - Primes should behave randomly (no conspiracy among primes). W-trick is a way to get rid of local obstruction for small primes (large primes have less effect).
- Almost primes (here-primes with only large prime factors) is much easier to understand than the primes and primes has positive upper density on almost primes.
- "Green-Tao measure \( \nu \)" normalized support of almost primes (numbers with only large prime factors).
- It can be viewed as a smooth cut off of \( \Lambda(n) = \sum_{d \leq n} \mu(d) \log(n/d) : \)
\[
\Lambda_R = \frac{1}{\log R} \left( \sum_{d \mid n, d \leq R} \mu(d) \log(R/d) \right)^2, R = N^{1/10}
\]
Pseudorandom conditions

Assume \( L_1, \ldots, L_k \) are pairwise linearly independent linear forms.

**Theorem (Linear forms conditions)**

\[ \mathbb{E}_{x \in \mathbb{Z}_N^d} \nu(L_1(x))\nu(L_2(x)) \ldots \nu(L_k(x)) = 1 + o(1) \]

- Equivalently, the system of linear equations
  \[ L_1(x) = y_1, L_2(x) = y_2, \ldots, L_l(x) = y_k \]
  Has expected probability that \( y_1, \ldots, y_k \) are all almost primes.

- Example \( \|\nu - 1\|_{U_k} = o(1) \).
Correlations

\[(c,d) \in \mathcal{A} \times \mathcal{A}\]

\[(a,b) \in \mathcal{A} \times \mathcal{A}\]

\[\nu \otimes \nu(x, y) = \nu(x)\nu(y)\] cannot be (pseudo)random.
correlations

\[(c, d) \in A \times A\]

\[(a, b) \in A \times A\]

\[\nu \otimes \nu(x, y) = \nu(x)\nu(y)\] cannot be (pseudo)random.

Twin Primes conjecture:

\[\sum \Lambda(n)\Lambda(n + 2)\]

we have system of linear forms \(x, x + 2\).
Transference Principle

Write \( \| f \| := \| \hat{f} \|_\infty \).  

Lemma (Properties of dual norms of \( \| \hat{f} \|_\infty \).

- **Algebra norm** : \( \| \phi_1 \cdot \phi_2 \|^* \leq \| \phi_1 \|^* \| \phi_2 \|^* \)
- **\( L^\infty \) – compatibility** \( \| \phi \|_\infty \leq \| \phi \|^* \)
- **Real-compatibility** \( \| \text{Re}(\phi) \|^* \leq \| \phi \|^* \)
- **Duality** \( \forall f \in \mathbb{C}^N, \exists \phi \in \mathbb{C}^N, \| \phi \|^* = 1 \) such that \( \| f \| = \text{Re}(\langle f, \phi \rangle) \).

Theorem (Transference Principle (Gowers, RTTV))

Suppose \( \| \hat{\nu} - \hat{1} \|_\infty \leq \theta \) then for any \( 0 \leq f \leq \nu \) we can find \( 0 \leq g \leq 1 \) such that

\[
\| \hat{f} - \hat{g} \|_\infty \ll \log(1/\theta)^{-3/2}
\]
Lemma (Supporting Hyperplane Theorem)

Let $C$ be a convex subset of $\mathbb{C}^N$ and $x \notin \text{int}(C)$. Then there exists $\phi \in \mathbb{C}^N \setminus \{0\}$ such that for all $y \in \overline{C}$ we have

$$\text{Re}\langle y, \phi \rangle \leq \text{Re}\langle x, \phi \rangle$$

Corollary (minimax)

Let $A$ and $B$ be non-empty compact convex subsets of $\mathbb{C}^N$ and at least one of $A$ or $B$ is a convex hull of finitely many points. Then there exist $a_0 \in A$, $b_0 \in B$ such that for any $a \in A$, $b \in B$ we have

$$\text{Re}\langle a, b_0 \rangle \leq \text{Re}\langle a_0, b \rangle$$
Proof: Write By contra-positive assume \( \| f - g \| > \varepsilon \) we will show that \( \| \nu - 1 \| > \exp(-C\varepsilon^{-2/3}) \). Assume now that \( \| \nu - 1 \| < 1 \).

- By (iv) for each \( 0 \leq g \leq 1 \) there exists \( \phi_g, \| \phi_g \|^* = 1 \) such that \( \text{Re} \langle f - g, \phi_g \rangle > \varepsilon \).
- Consider \( A = \{ g - f : 0 \leq g \leq 1 \}, B = \{ \phi : \| \phi \|^* \leq 1 \} \) then A and B are compact, convex, nonempty. Also A is a convex hull of the finite set \( \{1_S - f : S \subseteq [N]\} \).
- By minimax theorem, we can find \( 0 \leq g \leq 1 \) and \( \| \phi_0 \|^* \leq 1 \) such that
  \[
  \text{Re} \langle f - g, \phi_0 \rangle \geq \text{Re} \langle f - g_0, \phi_g \rangle > \varepsilon
  \]
- Let \( \psi := \text{Re}(\phi_0), \psi_+ = \min\{\psi, 0\}, g = 1_{\psi \geq 0} \) then
  \[
  \langle \nu, \psi_+ \rangle \geq \langle f, \psi_+ \rangle \geq \langle f, \psi \rangle = \text{Re} \langle f, \phi_0 \rangle > \text{Re} \langle g, \phi_0 \rangle + \varepsilon = \langle 1_{[N]}, \psi_+ \rangle + \varepsilon
  \]
- This implies
  \[
  \langle \nu - 1, \psi_+ \rangle > \varepsilon
  \]
Also by (ii) since \( \| \psi \|_\infty \leq \| \phi_0 \|_\infty \leq \| \phi_0 \|^* \leq 1 \). By Weierstrass Polynomial Approximation theorem there is a Polynomial \( P \) of degree \( \leq C \varepsilon^{-2/3} \), height \( \leq \exp(C \varepsilon^{-3/2}) \) such that

\[
\| P \circ \psi - \psi^+ \| < \varepsilon/4
\]

We have

\[
\langle \nu - 1, P \circ \psi \rangle = \langle \nu - 1, \psi^+ \rangle + \langle \nu - 1, P \circ \psi - \psi^+ \rangle \\
\geq \varepsilon - \| \nu - 1 \|_1 \| P \circ \psi - \psi \|_\infty \geq \frac{1}{2} \varepsilon
\]

i.e. \( \| \nu - 1 \| > \exp(-C \varepsilon^{-2/3}) \).

This implies \( \| \nu - 1 \| \cdot \| P \circ \psi \|^* \geq \varepsilon/2 \)

By (iii), \( \| \psi \|^* \leq \| \psi_0 \|^* \leq 1 \). By (i) , \( \| P \circ \psi \|^* \leq \exp(\varepsilon^{-2/3}) \).

So \( \| \nu - 1 \| > \exp(-C \varepsilon^{-2/3}) \).
Soft and Hard Obstructions

- **Soft Obstruction (Green-Tao’s Theorem):**
  - Dual function $\mathcal{D}f$ of $f$ is the function that $\langle f, \mathcal{D}f \rangle = \|f\|_{U^k}^2$
  - We may not know explicitly the exact structure of $\mathcal{D}f$.

- **Hard Obstruction. (Linear equations in Primes)**
  - To get asymptotics for
    \[
    \sum_{x \in \mathbb{Z}_N^d} \Lambda(L_1(x)) \cdots \Lambda(L_k(x))
    \]
    we need a control of exponential sum $\sum_{p<N} e(rp)$ or sum of nilsequences $\sum_{p<N} F(g(p)\Gamma)$.
**Inverse Gowers Uniformity norms theorem**

Motivation: If $G$ is $(k - 2)$–step nilpotent manifold, then the $k^{th}$ term is the sequence $x\Gamma, gx\Gamma, g^2x\Gamma, \ldots$ is determined by the first $(k - 1)^{th}$ terms (in a way that related to Geometric Sequence or Hall-Petresco Group).

**Theorem**

For each $k \geq 2, \delta \in (0, 1]$ there are constants $c(k, \delta) > 0, d(k, \delta)$ such that the following holds: If $f : \{1, \ldots, N\} \rightarrow \mathbb{C}, |f| \leq 1, \|f\|_{U^k} > \delta$ then there is a $k$-step nilsequence $\Psi(n)$ of dimension at most $d(k, \delta)$ such that

$$\left| \frac{1}{N} \sum_{n=1}^{N} f(n)\Psi(n) \right| \geq c(k, \delta).$$

- In case $U^3$ — norm we can take
  - 2-step nilsequence (Heisenberg Group).
  - Quadratic bracket phase function. $\Psi(n) = e(\phi(n))$ where $\phi(n) = \sum_{r,s \leq c_1(\delta)} \beta_{rs}\{\theta_r n\}\{\theta_s n\} + \gamma_r\{\theta_r n\}$.
  - Local quadratic average $\mathbb{E}_{y \in G} \mathbb{E}_{x \in B+y} f(x)e(q_y(x))$.
Open Questions

- What are pseudorandom conditions that are needed?
- What is "quadratic phase function" on \( \{1, \ldots, N\} \) exactly? Where do these objects come from?

\[
\| f \|_{U^2}^4 = \int_0^1 \left| \frac{1}{N} \sum_{n=1}^N f(n) e^{(2\pi i n \theta)} \right|^4 d\theta
\]

Can we find a formula of higher \( U^k \) norms in terms of nilsequences?
- \( c(\delta) \) in the \( U^3 \) norm inverse theorem is polynomial is equivalent Polynomial Freiman-Ruzsa Conjecture.
  - Sander(2012): \( c(\delta) = \delta^{O(\log^c(\delta^{-1}))} \)

Polynomial Freiman-Ruzsa Conjecture

If \( A \subseteq \mathbb{F}_2^n, |A + A| \leq K|A| \) then \( A \) can be covered by \( O(K^{O(1)}) \) translates of a subspace of \( \mathbb{F}_2^n \) that has cardinality at most \( |A| \).
Recent Developments

- Colon-Fox-Zhao Densification Trick.
  - Use the cut-norm which is weaker than the Gowers Uniformity norm.
  - Weaker pseudorandom condition is needed, in particular, counting 2-blow up of a simplex and its subgraphs.

- Helfgott and De Roton’s $L^2$–transference principle.
  - Replace the condition on Fourier estimates of $\nu$ by the boundedness of $L^2$–norm of $\nu$.

- Croot-Sisask Almost Periodicity
  - If $A \subseteq G$ has small $A + A$ then there is a dense set $X$ called almost period on $A + A$ such that for all $x \in X$, $1_A \ast 1_A(\cdot)$ and $1_A \ast 1_A(\cdot + x)$ is small in $L^p$, $p \geq 2$.

- Approximate Algebraic Structure.
- Numbers of solutions to translation-invariant equation.
- Higher order Fourier analysis on multiplicative functions.
- Function Fields Model
Recent Developments

This is first conjectured by Bergelson, Host, Kra. It is proved for $k=3$ by Green, $k=4$ by Green and Tao. It is shown by Ruzsa that the statement is false for $k \geq 5$

**Theorem (Green’s Roth Theorem with Popular Differences)**

For all $\epsilon > 0$, there is $N(\epsilon)$ such that for any abelian group $G$ with $|G| \geq N(\epsilon)$, Suppose $A \subseteq G$, $|A| = \alpha|G|$,

Then $\exists d \in G, d \neq 0$ such that the density of 3-AP with common difference $d$ in $A$ is at least $\alpha^3 - \epsilon$.

- (Fox-Pham-Zhao, 2017+) $N(\epsilon) = \text{Tower}(\Omega(\log(1/\epsilon))).$


Suggested reading

- S. PRENDIVILLE, Four Variants of the Fourier-Analytic Transference Principle.

Books:
- Ergodic Theory with a view towards Number Theory. Einsiedler, Manfred. Ward, Thomas
- Additive Combinatorics. Tao T, Vu V.
In 2013, Endre Szemerédi and his wife Anna visited Roth in Inverness, Scotland.

It is common knowledge that Klaus Roth is a mathematical genius, but who knows about his dancing skills? Who knows that he could do moves ranging from cha-cha to the jive?

In 2013 my husband Endre Szemerédi and I visited Klaus Roth. He was living in a nursing home in Inverness; he had moved there after his wife Melek’s death from cancer. We were very sorry to see him alone in a tiny private room unable to get out of bed. However, the subject of dance in general—and dancing with his wife of so many years in particular—made him lively and alert.
A: Do you think that dance and mathematics have anything in common?

K: Dance and mathematics? No, for me no. [...] 

A: How did you feel while you were dancing. Can you express it? 

K: Very frustrated because it is very difficult.

(Anna, Endre and Klaus laughing.)

A: Was it more difficult than mathematics?

K: I could explain and understand mathematics by using words. Whereas dancing is a completely different skill. Even if it is shown to me, I feel that there is no way of explaining it.
A: Which memories are dearer to you? Memories of when you solved math problems or when you danced?

K: My recollections of dancing with Melek on the whole give me more pleasure than my recollections of mathematics. It doesn’t make too much sense but that is the way it is.

However, I feel terribly lucky that I was exposed to geniuses.

E: I am certain that people felt terribly lucky that they were exposed to you.

Later on I learned from William Chen (Macquarie University, Sidney) who was a student of Professor Klaus Roth, that the dance coaches of Melek and Klaus were Alan and Hazel Fletcher, World Latin Dance Champions from 1977–1981. Professor Roth dedicated his paper “On Irregularities of Distribution. III” to Alan Fletcher.
Thank you for your attention.