Oberwolfach Report: Combinatorics, January 2014

The Green-Tao theorem and a relative Szemerédi theorem

Yufei Zhao

(joint work with David Conlon and Jacob Fox)

The celebrated Green-Tao theorem [4] states that the primes contain arbitrarily long arithmetic progressions. The main technical result in their work is a relative Szemerédi theorem. Szemerédi's theorem [5] states that any subset of the integers with positive upper density contains arbitrarily long arithmetic progressions. A relative Szemerédi theorem is a result where the ground set is no longer $\mathbb{Z}$ but some sparse pseudorandom subset (or more generally some measure).

Green and Tao proved a relative Szemerédi theorem where the ground set satisfies certain pseudorandomness conditions known as the linear forms condition and the correlation condition. They then constructed a majorizing measure to the primes, using ideas from the work of Goldston and Yıldırım [3], so that this majorizing measure satisfies the desired pseudorandomness conditions.

In our work [1, 8], we significantly simplify the proof of the Green-Tao theorem. We show that a relative Szemerédi theorem holds under much weaker pseudorandomness hypotheses. Namely, we only need a weaker version of Green-Tao's linear forms condition, and we completely get rid of the need for a correlation condition.

Here is a weighted formulation of Szemerédi's theorem (it is equivalent to the usual statement about subsets of $\mathbb{Z}$ with positive density).

Theorem 1 (Szemerédi's theorem, weighted version)**.** *Let $k \geq 3$ and $0 < \delta \leq 1$ be fixed. Let $f\colon \mathbb{Z}_N \to [0,1]$ be a function satisfying $\mathbb{E}[f] \geq \delta$. Then*

$$\mathbb{E}_{x,d\in\mathbb{Z}_N}[f(x)f(x+d)f(x+2d)\cdots f(x+(k-1)d)] \geq c(k,\delta) - o_{k,\delta}(1) \tag{1}$$

for some constant $c(k,\delta) > 0$ which does not depend on f or N.

For a relative Szemerédi theorem, we have a pseudorandom majorizing measure $\nu\colon \mathbb{Z}_N \to [0,\infty)$, and $f\colon \mathbb{Z}_N \to [0,\infty)$ is assumed to satisfy $0 \leq f \leq \nu$ pointwise. Here the function ν is normalized so that $\mathbb{E}[\nu] = 1 + o(1)$. For instance, one can think of ν as $\frac{N}{|S|}1_S$ for some pseudorandom subset $S \subseteq \mathbb{Z}_N$, and f as $1_A\nu$ with some $A \subseteq S$.

Definition 2 (Linear forms condition)**.** A nonnegative function $\nu = \nu^{(N)} : \mathbb{Z}_N \to \mathbb{R}_{\geq 0}$ is said to obey the *k-linear forms condition* if one has

$$\mathbb{E}_{x_1^{(0)},x_1^{(1)},\dots,x_k^{(0)},x_k^{(1)}\in\mathbb{Z}_N}\Big[\prod_{j=1}^{k}\prod_{\omega\in\{0,1\}^{[k]\setminus\{j\}}}\nu\Big(\sum_{i=1}^{k}(i-j)x_i^{(\omega_i)}\Big)^{n_{j,\omega}}\Big] = 1 + o(1) \tag{2}$$

for any choice of exponents $n_{j,\omega} \in \{0,1\}$.

Example 3. For $k = 3$, condition (2) says that

$$\begin{aligned}\mathbb{E}_{x,x',y,y',z,z'\in\mathbb{Z}_N}[&\nu(y+2z)\nu(y'+2z)\nu(y+2z')\nu(y'+2z')\cdot\\ &\cdot\nu(-x+z)\nu(-x'+z)\nu(-x+z')\nu(-x'+z')\cdot\\ &\cdot\nu(-2x-y)\nu(-2x'-y)\nu(-2x-y')\nu(-2x'-y')] = 1 + o(1)\end{aligned} \tag{3}$$

and similar conditions hold if one or more of the twelve ν factors in the expectation are erased.

Observe that (3) is the density of $K_{2,2,2}$ in some tripartite graph constructed from ν.

Here is our main result. (This version is from [8]. In [1] we prove the same result without the additional conclusion that $c(k,\delta)$ can be made the same as in Theorem 1.)

Theorem 4 (Relative Szemerédi theorem)**.** *Let $k \geq 3$ and $0 < \delta \leq 1$ be fixed. Let $\nu\colon \mathbb{Z}_N \to [0,\infty)$ satisfy the k-linear forms condition. Assume that N is sufficiently large and relatively prime to $(k-1)!$. Let $f\colon \mathbb{Z}_N \to [0,\infty)$ satisfy $0 \leq f(x) \leq \nu(x)$ for all $x \in \mathbb{Z}_N$ and $\mathbb{E}[f] \geq \delta$. Then*

$$\mathbb{E}[f(x)f(x+d)f(x+2d)\cdots f(x+(k-1)d)|x,d \in \mathbb{Z}_N] \geq c(k,\delta) - o_{k,\delta}(1), \tag{4}$$

where $c(k,\delta)$ is the same constant which appears in Theorem 1. The rate at which the $o_{k,\delta}(1)$ term goes to zero depends not only on k and δ but also the rate of convergence in the k-linear forms condition for ν.

Here is an overview of the proof. Starting with $0 \leq f \leq \nu$, by a *dense model theorem*, one can approximate f by some $\tilde{f}\colon \mathbb{Z}_N \to [0,1]$. Note that $\tilde{f}$ is bounded by 1, so it is a dense object. The dense model $\tilde{f}$ is a good approximation of f in terms of a certain cut norm, which we won't define here. A counting lemma (which is the main new technical advance in our work) implies that the k-AP count in f must be similar to the k-AP count in $\tilde{f}$. By Theorem 1, the k-AP count in $\tilde{f}$ can be bounded from below by $c(k,\delta) - o(1)$, and thus the same must be true for the k-AP count in f due to the counting lemma.

This is the same strategy taken in the original Green-Tao proof. However, instead of approximating f by $\tilde{f}$ in terms of a cut norm, they rely on a notion based on the Gowers uniformity norm. By using the cut norm, we obtain a cheaper dense model theorem, at the cost of a trickier counting lemma. This simplifies the proof significantly.

The strategy described above was the approach taken in [8]. A different approach was taken in our earlier paper [1] where we transfer the hypergraph removal lemma to obtain a sparse relative hypergraph removal lemma, from which we deduce the relative Szemerédi theorem (this is similar in spirit to Tao's proof [6, 7] that the Gaussian primes contain arbitrary constellations).

We are currently preparing an exposition [2] of the complete proof of the Green-Tao theorem (assuming Szemerédi's theorem), incorporating all the simplifications that have taken place since the original proof.

References

[1] D. Conlon, J. Fox, and Y. Zhao, *A relative Szemerédi theorem.*, arXiv:1305.5440.

[2] D. Conlon, J. Fox, and Y. Zhao, *The Green-Tao theorem: an exposition*, in preparation.

[3] D. A. Goldston and C. Y. Yıldırım, *Higher correlations of divisor sums related to primes. I. Triple correlations*, Integers **3** (2003), A5, 66.

[4] B. Green and T. Tao, *The primes contain arbitrarily long arithmetic progressions*, Ann. of Math. **167** (2008), no. 2, 481–547.

[5] E. Szemerédi, *On sets of integers containing no k elements in arithmetic progression*, Acta Arith. **27** (1975), 199–245.

[6] T. Tao, *The Gaussian primes contain arbitrarily shaped constellations*, J. Anal. Math. **99** (2006), 109–176.

[7] T. Tao, *A variant of the hypergraph removal lemma*, J. Combin. Theory Ser. A **113** (2006), no. 7, 1257–1280.

[8] Y. Zhao, *An arithmetic transference proof of a relative Szemerdi theorem*, Math. Proc. Cambridge Philos. Soc., to appear.