LECTURE ON ALMOST ARITHMETIC PROGRESSIONS IN THE PRIMES

SAURABH SHRIVASTAVA

ABSTRACT. In this talk we will try to understand the paper titled 'Almost arithmetic progressions in the primes and other large sets' by Jonathan M. Fraser. This concerns the problem of finding 'almost' arithmetic progressions of arbitrary length in 'large' subsets of integers. We shall see several important results in this direction. The celebrated result due to B. Green and T. Tao states that the primes contain arbitrarily long arithmetic progressions. J M Fraser in the paper mentioned above proved a weaker version of this result. In particular, he proved that the primes contain arbitrarily long 'almost' arithmetic progressions. In other words, it means that primes get arbitrarily close to arbitrarily long arithmetic progressions. The result is of course a corollary of the Green-Tao theorem. However, the importance of the result comes from the fact that its proof is elementary and it generalises to the setting of other subsets of integers.

1. Arithmetic progressions in sets of positive upper asymptotic density

Arithmetic progression (AP) of length k: $x, x+\Delta, x+2\Delta, \dots, x+(k-1)\Delta$ for some $x, \Delta \in \mathbb{N}$

Van der Waerden (1927): In every colouring of the integers by finitely many colours, there exists arbitrarily long monochromatic APs.

Erdos and Turán conjecture (1936): Every 'large set' of integers must contain arbitrarily long APs. The term 'large' in the above requires some discussion. The most natural qualification for a set to be large is a suitable density property in the integers.

Definition 1.1 (Asymptotic density (or natural density) of sets)**.** *The upper asymptotic density* $\overline{d}$ *of a set* $A \subseteq \mathbb{N}$ *is defined by*

$$\overline{d}(A) := \limsup_{n\to\infty} \frac{\#(A \cap \{1,2,3,\dots,n\})}{n} = \limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \chi_A(i),$$

and the lower asymptotic density $\underline{d}$ *of a set* $A \subseteq \mathbb{N}$ *is defined by*

$$\underline{d}(A) := \liminf_{n\to\infty} \frac{\#(A \cap \{1,2,3,\dots,n\})}{n} = \liminf_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \chi_A(i),$$

If both $\underline{d}(A)$ *and* $\overline{d}(A)$ *are equal, the asymptotic density* $d(A)$ *is said to exist and is defined to be the common number.*

There is another useful notion of density commonly know as the logarithmic density of a set. It is given as follows.

Definition 1.2 (Logarithmic density of sets)**.** *The maximal upper logarithmic density* $\overline{\delta}$ *of a set* $A \subseteq \mathbb{N}$ *is defined by*

$$\overline{\delta}(A) := \limsup_{n\to\infty} \sup_{m\geq 0} \frac{\log(\#(A \cap [m+1, m+n]))}{\log n},$$

Date: January 24, 2020.

The lower logarithmic density $\underline{\delta}(A)$ may be defined in a similar fashion. If both $\underline{\delta}(A)$ and $\overline{\delta}(A)$ are equal, the logarithmic density $\delta(A)$ is said to exist and is defined to be the common number.

Remark 1.3. *There are example of sequences for which the logarithmic density exists and the asymptotic density does not exist.*

K. F. Roth (1953) proved the Erdos and Turán conjecture for $k = 3$. More precisely, he proved that every set $A \subseteq \mathbb{N}$ of upper asymptotic density $\overline{d}(A) > 0$ contains a 3-term AP (and hence many 3-term APs).

Remark 1.4.

- *There are quantitative versions of this question in the following sense. For evert $0 < \epsilon < 1$ and k, there exists an $N(\epsilon, k)$ such that if we take $N > N(\epsilon, k)$ and $A \subseteq \{1, 2, 3, \ldots, N\}$ with $\#A \geq \epsilon N$, then A contains an AP of length k.*
- *Roth's proof uses the Fourier analytic arguments.*

E. Szemerédi (1975) proved the Erdos and Turán conjecture. In particular, he proved that every set $A \subseteq \mathbb{N}$ of upper asymptotic density $\overline{d}(A) > 0$ contains arbitrarily long APs. His proof is based on combinatorial arguments and certain estimates on exponentials.

H. Furstenberg (1977) gave a new proof of Szemerédi's theorem using ergodic theory.

W. T. Gowers (1998) gave a new proof of Szemerédi's theorem for $k = 4$. Later, in 2011 he gave yet another proof of Szemerédi's theorem.

There are few more different proofs of the Szemerédi's theorem.

2. Arithmetic progressions in the primes

First, note that the primes have the upper asymptotic density zero.

van der Corput (1939) proved that the primes contains infinitely many 3-term APs.

The existence of longer APs in the primes remained open for a long time. However, there were many interesting results in the direction. For example,

D.R. Heath-Brown (1981) proved that there are infinitely many 4-term APs consisting of three primes and a number which is either a prime or a product of two primes.

B. J. Green (2005) gave another proof of Roth's theorem for the primes. He showed that the primes contains infinitely many 3-term APs.

APs of the primes using computational tools

- A. Moran, P. Pritchard and A. Thyssen (1995) found AP of primes of length 22.
- Later, M. Frind, P. Underwood, and P. Jobling (2005) found AP of primes of length 23.
- The largest known arithmetic progression of primes contains 24 primes by J. Wroblewski (2007)

B. J. Green and T. Tao (2008) proved Szemerédi's theorem for the primes by showing that the primes contains infinitely many k-term APs for every $k \geq 1$..

The Green-Tao theorem solves the following Erdos and Turán conjecture for the primes.

Erdos and Turán conjecture (1936): If $A \subseteq \mathbb{N}$ is such that $\sum_{a\in A} \frac{1}{a} = \infty$, then A contains arbitrarily long APs.

3. Almost APs of arbitrary length in the primes

Recently, J. M. Fraser and H. Yu proved a weakened version of the Erdos and Turán conjecture. Their proof is elementary. We shall discuss this in details here.

Definition 3.1. *A set $A \subseteq \mathbb{N}$ is said to contain almost APs of arbitrarily length if for all $k \in \mathbb{N}$ and $\epsilon > 0$, there exists an AP P of length k with gap size Δ such that*

$$\sup_{p\in P} \inf_{a\in A} |p-a| \leq \epsilon\Delta.$$

Theorem 3.2. *[J. M. Fraser (2019)] If $A \subseteq \mathbb{N}$ is such that $\overline{\delta}(A) = 1$, then A contains arbitrarily long almost APs.*

This proves a weakened version of the Erdos and Turán conjecture. For, we have the following lemma.

Lemma 3.3. *If $A \subseteq \mathbb{N}$ is such that $\sum_{a\in A} \frac{1}{a} = \infty$, then $\overline{\delta}(A) = 1$.*

In particular, the primes contain arbitrarily long almost APs as we know that $\sum_{p \text{ prime}} \frac{1}{p} = \infty$.

Proof of Lemma 3.3: List the elements of A in increasing order as $a_1 < a_2 < \cdots < a_n < \ldots$. Suppose on the contrary that $\overline{\delta}(A) < 1$. Then by the definition of the upper logarithmic density, there exists $s \in (0,1)$ such that for all $m \geq 0$ and $n \in \mathbb{N}$ we have

$$\#(A \cap [m+1, m+n]) \lesssim n^s.$$

Let us consider $A_N = [2^N, 2^{N+1})$ and write

$$\begin{aligned}
\sum_{j=1}^{\infty} \frac{1}{a_j} &= \sum_{N=1}^{\infty} \sum_{j:a_j\in A_N} \frac{1}{a_j} \\
&\leq \sum_{N=1}^{\infty} \#(A_N) 2^{-N} \\
&\lesssim \sum_{N=1}^{\infty} 2^{(s-1)N} < \infty \text{ as } s < 1,
\end{aligned}$$

which is a contradiction. This completes the proof. $\square$

Proof of Theorem 3.2: The proof is by contradiction. Suppose A does not contain arbitrarily long almost APs. Then there exists $k \geq 3$ and $\epsilon > 0$ such that for any given arithmetic progression P of length k and gap size Δ, we have

$$\sup_{p\in P} \inf_{a\in A} |p-a| > \epsilon\Delta.$$

Observe that we may assume, possibly by taking a smaller ϵ, that $\frac{1}{2\epsilon}$ is an integer. We shall perform an inductive argument to get a contradiction.

For, let I be a finite interval with $|I| > 0$. We decompose I into $\frac{k}{2\epsilon}$ equal subintervals of length $\frac{2\epsilon|I|}{k}$. Let us label these intervals (from left to right) as $1, 2, 3, \ldots, \frac{k}{2\epsilon}$. Note that each label corresponds to a subinterval of length $\frac{2\epsilon|I|}{k}$. We form congruence classes modulo $\frac{1}{2\epsilon}$ on this set of labels $1, 2, 3, \ldots, \frac{k}{2\epsilon}$. Note that the centres of the intervals with labels in the same congruence class form an arithmetic progression of length k and gap size $\frac{|I|}{k}$. In view of our assumption, we observe that at least one interval from each congruence class does not intersect with the set A, otherwise, we would get an almost AP of length k in A. This would imply that $A \cap I$ is contained in the union of $\frac{k-1}{2\epsilon}$ subintervals of length $\frac{2\epsilon|I|}{k}$. We apply this argument inductively on the subintervals which intersect with the set A. Let $J_0 = [m+1, m+n]$ where $m \geq 0$ and $n \in \mathbb{N}$.

Preform the previous step with $I = J_0$. We get that $A \cap J_0$ is contained in the union of $\frac{k-1}{2\epsilon}$ subintervals of length $\frac{2\epsilon|J_0|}{k} = (n-1)\frac{2\epsilon}{k}$. We the same procedure to each one the subintervals which intersect with A. In this step the length of subintervals, say J_i, would be $(n-1)(\frac{2\epsilon}{k})^2$ and also we notice that we would get that $A \cap J_i$ is contained in the union of $(\frac{k-1}{2\epsilon})$ such subintervals. Subsequently, we get that $A \cap J_0$ is contained in the union of $(\frac{k-1}{2\epsilon})^2$ subintervals of length $(n-1)(\frac{2\epsilon}{k})^2$. We continue this procedure and after performing it N times we would got that $A \cap J_0$ is contained in the union of $(\frac{k-1}{2\epsilon})^N$ subintervals of length $(n-1)(\frac{2\epsilon}{k})^N$.

Fix N to be the smallest positive integer so that

$$(n-1)\left(\frac{2\epsilon}{k}\right)^N < 1.$$

This means that N satisfies the estimate $N \leq \frac{\log(n-1)}{\log(\frac{k}{2\epsilon})} + 1$ Since we are working with subsets of integers. This choice of N ensures that at the Nth step each subinterval contains at most one point of the set A. Therefore, we have that

$$\begin{aligned}
\#(A \cap [m+1, m+n]) &\leq \left(\frac{k-1}{2\epsilon}\right)^N \\
&\leq \left(\frac{k-1}{2\epsilon}\right)^{\frac{\log(n-1)}{\log(\frac{k}{2\epsilon})}+1} \\
&= \left(\frac{k-1}{2\epsilon}\right)\left(\frac{k-1}{2\epsilon}\right)^{\frac{\log(n-1)}{\log(\frac{k}{2\epsilon})}} \\
&\leq \left(\frac{k-1}{2\epsilon}\right) n^{\frac{\log(\frac{k-1}{2\epsilon})}{\log(\frac{k}{2\epsilon})}}.
\end{aligned}$$

Since $\frac{\log(\frac{k-1}{2\epsilon})}{\log(\frac{k}{2\epsilon})} < 1$, the estimate above yields that the upper logarithmic density $\overline{\delta}(A) < 1$, which is a contradiction. This completes the proof. $\square$

Next, we give the proof of the fact that sum of reciprocals of primes diverges.

Proof of $\sum\limits_{p\ prime} \frac{1}{p} = \infty$: The proof is due to Erdos. Let us enumerate the primes in increasing order as $p_1, p_2, p_3, \ldots$. Suppose on the contrary that

$$\sum_{p\ prime} \frac{1}{p} < \infty.$$

Choose $L \in \mathbb{N}$ large enough so that $\sum\limits_{j=L+1}^{\infty} \frac{1}{p_j} < \frac{1}{2}$. For a positive integer x, write $N(x)$ to be the number of positive integers $n \leq x$ which are not divisible by any primes strictly larger than p_L. If $n \leq x$ is such an integer then we can express it as $n = b^2 c$ where c is square free and b is an integer. This can be done in the following way. Note that only prime factor of n are from first L primes $p_1, p_2, p_3, \ldots, p_L$. The number c is the product of a primes in a subset of $\{p_1, p_2, p_3, \ldots, p_L\}$ with each prime appears at most once.

With this the number of choices of c is the number of subsets of the first L primes, which in nothing but 2^L. Further, $b^2 \leq x$ and hence the number of choices of b is at most $\sqrt{x}$. These observations give us that

$$N(x) \leq 2^L \sqrt{x}.$$

Next, note that the number of positive integers $n < x$ which are divisible by a prime p is at most $\frac{x}{p}$. With our notation the number of positive integers $n < x$ which are divisible by any prime other than the first L primes is at most

$$x - N(x) \leq \frac{x}{p_{L+1}} + \frac{x}{p_{L+2}} + \cdots < \frac{x}{2}.$$

This implies that $\frac{x}{2} < N(x) \leq 2^L \sqrt{x}$. Note that the estimate above cannot hold for all positive integers x. For example, take $x = 2^{4L+2}$, then we would get that

$$2^{4L+1} < 2^L 2^{2L+1} = 2^{3L+1}.$$

This leads to a contradiction to the hypothesis that sum of reciprocal of primes is finite. $\square$

References

[1] P. Erdös and Über die Reihe, $\sum \frac{1}{p}$, Mathematica, Zutphen, B 7, (1938), 1–2.

[2] P. Erdös and P. Turán, *On some sequences of integers,* J. London Math. Soc. 11 (1936), 261–264.

[3] J. M. Fraser, *Almost arithmetic progressions in the primes and other large sets,* to appear in Amer. Math. Monthly, available at arXiv:1809.01409v2.

[4] J. M. Fraser and H. Yu, *Arithmetic patches, weak tangents, and dimension,* Bull. London Math. Soc. 50 (2018),85–95.

[5] J. M. Fraser, K. Saito, and H. Yu, *Dimensions of sets which uniformly avoid arithmetic progressions,* Int. Math. Res. Not. (IMRN) to appear.

[6] H. Furstenberg, *Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions,* J. Analyse Math. 31(1977), 204–256.

[7] W. T. Gowers, *A new proof of Szemerédi's theorem for arithmetic progressions of length four,* GAFA 8 (1998), 529–551.

[8] W. T. Gowers, *A new proof of Szemerédi's theorem,* GAFA 11 (2001), 465–588.

[9] B. J. Green, *Roth's theorem in the primes,* Ann. of Math. 161 (2005), 1609–1636.

[10] B. J. Green and T. tao, *The primes contain arbitrarily long arithmetic progressions,* Ann. of Math. 167 (2008), 481–547.

[11] D. R. Heath-Brown, *Three primes and an almost-prime in arithmetic progression,* J. London Math. Soc. 23 (1981), 396–414.

[12] I. Laba and M. Lacey, *On sets of integers not containing long arithmetic progressions,* unpublished; Available at http://www.arxiv.org/pdf/math.CO/0108155.

[13] K. F. Roth, *On certain sets of integers,* J. London Math. Soc. 28 (1953), 104–109.

[14] E. Szemerédi, *On sets of integers containing no four elements in arithmetic progression,* Acta Math. Acad. Sci. Hungar. 20 (1969), 89–104.

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

Department of Mathematics, Indian Institute Science Education and Research Bhopal, Bhopal-462066, India

Email address: {saurabhk}@iiserb.ac.in