Additive Combinatorics Lecture 2

Leo Goldmakher
Scribe: Gal Gross

Jan. 17th, 2014

Recall our discussion of Erdős's multiplication table problem: how many different numbers are there in the $N \times N$ multiplication table? Erdős discovered that the answer is $o(N^2)$. That is, if $A(N)$ is the number of different integers in the $N \times N$ multiplication table, then as $N \to \infty$ we have $A(N)/N^2 \to 0$.

Theorem (Erdős)**.** The number of distinct entries appearing in the $N \times N$ multiplication table, denoted $A(N)$, is in $o(N^2)$.

We shall sketch a proof of this theorem. The overarching idea is to look at the number of distinct prime factors of integers up to N^2. Almost all such numbers have $\approx \log\log N$[*] prime factors, while almost all entries in the multiplication table have $\approx 2\log\log N$ prime factors. We introduce the notation $\omega(n) = \sum_{p|n} 1$, the number of distinct prime factors of n. Note that the behaviour of $\omega(n)$ is very hard to predict. Particularly, from time to time it simply equals 1 (whenever we hit a prime). Instead of trying to predict $\omega(n)$, we can describe its *average* behaviour.

Lemma 1.
$$\frac{1}{x}\sum_{n\le x}\omega(n) = \log\log x + O(1).$$

The proof of this lemma relies on Euler's famous result that the sum of the reciprocal of primes diverges.

Lemma (Euler)**.**
$$\sum_{\substack{p\le x\\ p\,\text{prime}}} \frac{1}{p} = \log\log x + O(1).$$

Proof of Lemma 1. In the following calculation we take p to be a prime. We calculate:

$$\begin{aligned}\frac{1}{x}\sum_{n\le x}\omega(n) &= \frac{1}{x}\sum_{n\le x}\sum_{p|n}1 = \frac{1}{x}\sum_{p\le x}\sum_{\substack{n\le x\\ p|n}}1 = \frac{1}{x}\sum_{p\le x}\sum_{d\le \frac{x}{p}}1 = \frac{1}{x}\sum_{p\le x}\left\lfloor\frac{x}{p}\right\rfloor = \frac{1}{x}\sum_{p\le x}\left(\frac{x}{p}+O(1)\right)\\ &= \frac{1}{x}\sum_{p\le x}\frac{x}{p} + O\left(\frac{1}{x}\sum_{p\le x}1\right) = \sum_{p\le x}\frac{1}{p} + O\left(\frac{\pi(x)}{x}\right) = \log\log x + O(1).\end{aligned}$$

Recall that $\pi(x)$ denotes the number of primes $\le x$, so that $\pi(x)/x \in O(1)$. □

Now since $\log\log x$ grows extremely slowly, the upshot of Lemma 1 is that for "most" $n \le x$ we have $\log\log n \approx \log\log x$ and therefore $\omega(n) \approx \log\log n$. This intuitive reasoning is made precise by a famous theorem of Hardy and Ramanujan, sometimes referred to as Hardy-Ramanujan *normal order* theorem.

Theorem (Hardy-Ramanujan, 1917)**.** Let $\varepsilon > 0$. Then for almost every n,

$$\omega(n) = \log\log n + O_\varepsilon\Big((\log\log n)^{\frac{1}{2}+\varepsilon}\Big)$$

[*] Recall that in this class $\log x$ is the natural logarithm $\ln x$.

The phrase "for almost every n" means that the number of exceptions is tiny; more precisely, the number of $n \leq x$ for which the theorem fails to hold is $o(x)$.

In 1934, the Hungarian mathematician Pál Turán (who also frequently collaborated with Erdős) gave a simpler proof of the above theorem. His statement of the theorem is also more immediately usable for our purposes.

Theorem (Turán, 1934)**.**

$$\frac{1}{x}\sum_{n\leq x}(\omega(n)-\log\log x)^2 \ll \log\log x$$

These theorem(s) characterize the error in our approximation, which is sufficiently small. To conclude we now have that for almost every $n \leq N^2$

$$\omega(n) \approx \log\log N^2 = \log(2\log N) = \log 2 + \log\log N \approx \log\log N.$$

By contrast, we expect almost every entry appearing in the multiplication table to have roughly double that number of distinct prime factors, since $\omega(ab) = \omega(a) + \omega(b) \approx 2\log\log N$. The problem with this heuristic is that $\omega(ab) = \omega(a)+\omega(b)$ iff a and b are coprimes. This motivates the following definition. Define $A^*(N) = \{(a,b) \,:\, a,b \in \mathbb{N} \text{ and } \gcd(a,b) = 1\}$. The following claim solves our problem:

Lemma 2.

$$|A^*(N)| = o(N^2) \implies |A(N)| = o(N^2).$$

Exercise 1. Prove Erdős's Theorem using the two lemmas above.

Proof of Lemma 2. First note that

$$n \in A(N) \implies n = a\cdot b \implies \frac{a}{\gcd(a,b)}\cdot\frac{b}{\gcd(a,b)} = \frac{n}{\gcd(a,b)^2}.$$

Therefore, if we denote $d_n := \gcd(a,b)^2$, then $\frac{n}{\gcd(a,b)^2} \in A^*(N/d_n)$ and so

$$|A(N)| \leq \sum_{d\leq N}\left|A^*\left(\frac{N}{d}\right)\right|.$$

Exercise 2. Carefully justify the last inequality.

Now, the fact that $|A^*(N)| = o(N^2)$ formally tells us that $\forall\varepsilon > 0\,, \exists C_\varepsilon > 0$ s.t. $|A^*(M)| \leq \varepsilon M^2$ whenever $M > C_\varepsilon$. Fix an $\varepsilon > 0$ and the corresponding $C_\varepsilon > 0$, and let N be sufficiently large. Picking up where we left off:

$$\begin{aligned}\sum_{d\leq N}\left|A^*\left(\frac{N}{d}\right)\right| &= \sum_{d\leq \frac{N}{C_\varepsilon}}\left|A^*\left(\frac{N}{d}\right)\right| + \sum_{\frac{N}{C_\varepsilon}<d\leq N}\left|A^*\left(\frac{N}{d}\right)\right| \leq \sum_{d\leq\frac{N}{C_\varepsilon}}\varepsilon\frac{N^2}{d^2} + \sum_{\frac{N}{C_\varepsilon}<d\leq N}\frac{N^2}{d^2}\\ &= \varepsilon N^2\sum_{d\leq\frac{N}{C_\varepsilon}}\frac{1}{d^2} + N^2\sum_{\frac{N}{C_\varepsilon}<d\leq N}\frac{1}{d^2} \ll \varepsilon N^2 + N^2\sum_{\frac{N}{C_\varepsilon}<d\leq N}\frac{1}{(N/C_\varepsilon)^2}\\ &\ll \varepsilon N^2 + C_\varepsilon^2 N = \left(\varepsilon + \frac{C_\varepsilon^2}{N}\right)N^2.\end{aligned}$$

Exercise 3. Why is the proof complete?

□

* * *

* *

Let us now shift gears. Recall the following conjecture from the previous lecture:

Conjecture (Erdős-Szemerédi)**.**

$$\forall A \subseteq \mathbb{Z} \ \ \max(|A+A|, |A \cdot A|) = |A|^{2-o(1)}.$$

Incidentally, in cases where the statement holds and $|A+A| = |A|^{2-o(1)}$, we say that "A is *random* with respect to addition." The interesting idea behind this piece of jargon is that the typical behaviour of "random" sets with respect to arithmetical operations gives us a way of measuring randomness.

The strongest positive result addressing this conjecture was achieved by József Solymosi in 2009.

Theorem (Solymosi, 2009)**.**

$$\forall A, B \subseteq \mathbb{R} \quad |A \cdot B| \cdot |A+B| \cdot |B+B| \gg \frac{|A|^2 |B|^2}{\log(|A| |B|)}.$$

This result was later extended to $\mathbb{C}$, though we shall prove it only for the case where $A, B \subseteq \mathbb{R}_{>0}$. We also note that the $\log(|A| |B|)$ factor at the denomenator of the RHS is in fact $\log(|C|)$ where $C \coloneqq \min(|A|, |B|)$. Finally, note the following immediate corollary to Solymosi's theorem.

Corollary.

$$\forall A \subseteq \mathbb{R} \ \ \max(|A+A|, |A \cdot A|) \gg |A|^{\frac{4}{3}-o(1)}.$$

Before we proceed to prove Solymosi's theorem, we stop to define the following useful notation.

Definition. Let $\oplus$ be a binary operation. Then

$$r_{A \oplus B}(x) \coloneqq |\{(a,b) \, : \, a \oplus b = x, a \in A, b \in B\}|.$$

To famliarize yourself with this notation, convince yourself that

$$\sum_{x \in A \oplus B} r_{A \oplus B}(x) = |A \times B| = |A| \cdot |B|.$$

Proof of Solymosi's Theorem. We assume throughout that $A, B \subseteq \mathbb{R}_{>0}$. Solymosi's first key insight was to observe that

$$\sum_{x \in A \cdot B} r_{A \cdot B}(x)^2 = \sum_{m \in B \div A} r_{B \div A}(m)^2. \tag{1}$$

Exercise 4. Prove the equality above.

The LHS of this equality is called the "multiplicative energy" (introduced by Terence Tao?) of A and B. One way to interpret that sum intuitively is $\left|\{(a,b); (a',b') \in (A \times B)^2 \, : \, ab = a'b'\}\right|$, though it is not clear how to use this characterisation. For this proof we will trivially bound LHS from below.

Coming back to Solymosi's insight, $r_{B \div A}$ on the RHS can be interpreted geometrically. We have

$$r_{B \div A} = |\{(a,b) \in A \times B \, : \, b/a = m\}|,$$

so we are counting the number of $(A \times B)$-lattice points on the line $\mathscr{L}_m$ of slope m through the origin, $y = mx$. This geometrical insight will allow us to bound the RHS from above. Combining this with the lower bound on the LHS will yield the desired result. Let us start with the trivial lower bound on the LHS. From Cauchy-Schwarz we have

$$\left(\sum_{x \in A \cdot B} r_{A \cdot B}(x)\right)^2 \leq \left(\sum_{x \in A \cdot B} r_{A \cdot B}(x)^2\right)\left(\sum_{x \in A \cdot B} 1^2\right) \iff |A|^2 |B|^2 \leq |A \cdot B| \sum_{x \in A \cdot B} r_{A \cdot B}(x)^2.$$

Therefore, in light of equation (1), suffices it to show that

$$\sum_{m \in B \div A} r_{B \div A}(x)^2 \ll |A+A| \cdot |B+B| \cdot \log(|A| |B|).$$

This is exactly what we shall do in our next lecture. In preperation:

Exercise 5. $|A+A| \cdot |B+B| = |(A \times B) + (A \times B)|$.

□