THE CONVEX SUMSET PROBLEM

ADAM CUSHMAN

Abstract. The sum-product conjecture, posed in 1983 by Erdős and Szemerédi [ES83], posits that any sufficiently large set must have a 'relatively large' number of distinct sums or products between its elements. A similar conjecture extends this idea, positing that for convex sets there must be a 'relatively large' number of distinct sums between its elements. Both conjectures remain open and far from being solved. This report provides an entirely self-contained overview of some known results, primarily focused on the latter conjecture, and the methods used to achieve them.

1. Introduction and Motivation

For any sets A, B and binary operation $\cdot$ which acts on elements of A and B, we define

$$A \cdot B = \{a \cdot b : a \in A \ , \ b \in B\} .$$

Observe the following example which motivates the study of these problems.

Let $A = \{1, 2, 3, \cdots, n\}$ and $G = \{2, 2^2, 2^3, \cdots, 2^n\}$.

Notice that A is given by an arithmetic sequence and G by a geometric sequence. We are going to calculate $|A + A|, |AA|, |G + G|$, and $|GG|$.

Observe that

$$\begin{aligned} |A + A| &= |\{2, 3, \cdots, 2n\}| \\ &= 2n - 1 \\ &= 2|A| - 1, \end{aligned}$$

and by the same argument,

$$\begin{aligned} |GG| &= \left|\left\{2^2, 2^3, \cdots, 2^{2n}\right\}\right| \\ &= 2|G| - 1. \end{aligned}$$

We have

$$|G + G| = \left|\left\{2^i + 2^j : i, j \in \{1, \cdots, n\}\right\}\right|$$

For all i, j, $2^i + 2^j$ is a number written in base 2. By the uniqueness of binary representations (see appendix), we have that $2^i + 2^j$ is distinct for every choice of $\{i, j\}$. Therefore,

$$\begin{aligned} |G + G| &\geq |\{\{i, j\} : i, j \in \{1, \cdots, n\}\}| \\ &= \binom{n}{2} + n \\ &= \binom{|G| + 1}{2} \end{aligned}$$

Date: August 26, 2024.

Finally, I prove in the appendix that there exists $c \in \mathbb{R}^+$ such that

$$|AA| \geq \frac{c\,|A|^2}{\log(|A|)}.$$

Observing the trivial bounds $2\,|S| - 1 \leq |S \cdot S| \leq \binom{|S|}{2}$ for any set S and any commutative operation $\cdot$, it is clear that both AA and $G + G$ are, as $|A|\,, |G| \to \infty$, almost as large as they can be, and $A + A, GG$ are as small as they can be.

This is the phenomenon which motivates this problem. One questions is: "does there exist a set for which neither the sum nor the product set is large?" The sum-product conjecture states that such a set does not exist. Another question we study in this report is: "what determines if the sum set is large or the product set is large?" A similar conjecture which partially addresses this question states that the sum set is large when the set itself is convex.

The rest of the report will proceed with preliminary definitions and ideas, a precise statement of the main problems we'll focus on, proving a collection of important results, and applying these results to the conjectures we are concerned with.

2. Preliminaries

As a useful shorthand, for any natural number n, let

$$[n] = \{1, 2, \cdots, n\}.$$

For a set S, and function $f : S \to \mathbb{R}$,

$$f(S) = \{f(s) : s \in S\}.$$

The study of these problems requires the notion of orders of magnitude. For any functions $f, g : \mathbb{R} \to \mathbb{R}$, write

$$f(x) \gg g(x) \text{ as } x \to \infty$$

if

$$\exists x_0, c \in \mathbb{R}^+ \text{ s.t. } x > x_0 \implies |f(x)| \geq c\,|g(x)|\,,$$

write

$$f(x) \ll g(x) \text{ as } x \to \infty$$

if

$$\exists x_0, c \in \mathbb{R}^+ \text{ s.t. } x > x_0 \implies |f(x)| \leq c\,|g(x)|\,,$$

and write

$$f(x) \asymp g(x) \text{ as } x \to \infty$$

if

$$f(x) \ll g(x) \text{ and } f(x) \gg g(x) \text{ as } x \to \infty.$$

We write $\ll_\epsilon, \gg_\epsilon$ if the constant depends on ϵ. For example,

$$f(x) \gg_\epsilon g(x)^\epsilon$$

means

$$\forall \epsilon > 0\ ,\ f(x) \gg g(x)^\epsilon.$$

More precisely, there is some function $c : \mathbb{R}^+ \to \mathbb{R}^+$ so

$$\forall \epsilon > 0, \exists x_0 \in \mathbb{N} \text{ s.t. } x > x_0 \implies |f(x)| \geq c(\epsilon)\,|g(x)^\epsilon|\,.$$

We write $\lesssim, \gtrsim$ if, along with a constant factor, there is also a logarithmic factor. That is

$$f(x) \lesssim g(x)$$

if there is some $c \in \mathbb{R}$ such that

$$f(x) \ll (\log(x))^c g(x).$$

Throughout this report, the asymptotic parameter (in this case x) will always tend to ∞, so it will no longer be mentioned. Oftentimes the parameter will not even be in the expression. For example, if for some set A we write

$$|A+A| \gg |A|,$$

it is taken to mean that A is defined implicitly by $|A|$, and $|A|$ is the parameter which tends to ∞.

For any sets A, B and any binary operation $\cdot$ acting on elements of A and B, define the representation function $r_{A \cdot B} : A \cdot B \to \mathbb{N}$ by

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

Throughout this report the shorthand

$$\delta_{A,B}(x) = r_{A-B}(x) \ , \ \sigma_{A,B}(x) = r_{A+B}(x) \ , \ \delta_A(x) = \delta_{A,A}(x) \ , \ \sigma_A(x) = \sigma_{A,A}(x)$$

will be used.

For any sets A, B, define the Additive Energy $E(A,B)$ and Multiplicative Energy $M(A,B)$ by

$$E(A,B) = \left|\left\{(a_1,a_2,b_1,b_2) \in A^2 \times B^2 : a_1 - b_1 = a_2 - b_2\right\}\right|$$

and

$$M(A,B) = \left|\left\{(a_1,a_2,b_1,b_2) \in A^2 \times B^2 : \frac{a_1}{b_1} = \frac{a_2}{b_2}\right\}\right|.$$

Observe that

$$E(A,B) = \sum_{x \in A-B} \delta_{A,B}(x)^2$$

and

$$M(A,B) = \sum_{x \in \frac{A}{B}} r_{\frac{A}{B}}(x)^2.$$

This definition is symmetric in the sense that a 4-tuple $(a_1,a_2,b_1,b_2) \in A^2 \times B^2$ is a solution to

$$a_1 - b_1 = a_2 - b_2$$

if and only if it is a solution to

$$a_1 + b_2 = a_2 + b_1,$$

and therefore

$$E(A,B) = \sum_{x \in A-B} \delta_{A,B}(x)^2 = \sum_{x \in A+B} \sigma_{A,B}(x)^2.$$

There is a similar argument for multiplicative energy. Any 4-tuple $(a_1,a_2,b_1,b_2) \in A^2 \times B^2$ with nonzero entries is a solution to

$$\frac{a_1}{b_1} = \frac{a_2}{b_2}$$

if and only if it is a solution to

$$a_1 b_2 = a_2 b_1.$$

There are at most

$$\sum_{i=1}^{4}\binom{4}{i}=15$$

4-tuples with zero entries, so

$$M(A,B)=\sum_{x\in\frac{A}{B}}r_{\frac{A}{B}}(x)^2\asymp\sum_{x\in AB}r_{AB}(x)^2.$$

We also define higher energies

$$E_n(A,B)=\sum_{x\in A-B}\delta_{A,B}(x)^n,$$

so

$$E(A)=E_2(A),$$

and as a shorthand use

$$E_n(A)=E_n(A,A).$$

Similar definitions and shorthand are used for multiplicative energy.

We can relate the energies to the sizes of the sum and product sets by the Cauchy-Schwarz Inequality.

$$|A|\,|B|=\sum_{x\in A+B}\sigma_{A,B}(x)\leq|A+B|^{\frac{1}{2}}\,E(A,B)^{\frac{1}{2}},$$

and

$$|A|\,|B|=\sum_{x\in AB}r_{AB}(x)\leq|AB|^{\frac{1}{2}}\,M(A,B)^{\frac{1}{2}}.$$

Similar inequalities can be derived for $|A-B|$ and $\left|\frac{A}{B}\right|$.

Let $I\subset\mathbb{R}$ be an interval. We call a function $f:I\to\mathbb{R}$ convex if for any 2 points $x_1,x_2\in I$, with $x_1\neq x_2$, and any $\lambda\in(0,1)$,

$$f(\lambda x_1+(1-\lambda)x_2)<\lambda f(x_1)+(1-\lambda)f(x_2).$$

A finite set $A\subset\mathbb{R}$ is convex if there is a function $f:[1,|A|]\to\mathbb{R}$ such that

$$A=\{f(i):i\in\{1,\cdots,|A|\}\}.$$

A property of convex functions which will come up later, and is worth proving now, is

Lemma 2.1. *Let $I\subset\mathbb{R}$ be an interval. Let $f:I\to\mathbb{R}$ be convex. Let $T\subset\mathbb{R}^2$.*

Take $\ell=\{(x,f(x)):x\in I\}$ to be the graph of f. We have that

$$\forall t_1,t_2\in T\ ,\ |(\ell+t_1)\cap(\ell+t_2)|\leq 1.$$

That is that translations of the graph of a convex function intersect in at most one point.

Proof. We first show that for any x_1,x_2 with $x_1<x_2$ and $t>0$, we have

$$f(x_1+t)-f(x_1)<f(x_2+t)-f(x_2).$$

Take $\lambda_1=\frac{x_1-x_2}{x_1-x_2-t}$ to get the result

$$f(x_1+t)=f(\lambda_1x_1+(1-\lambda_1)(x_2+t))<\lambda_1f(x_1)+(1-\lambda_1)f(x_2+t).$$

Take $\lambda_2=\frac{-t}{x_1-x_2-t}$ to get the result

$$f(x_2)=f(\lambda_2x_1+(1-\lambda_2)(x_2+t))<\lambda_2f(x_1)+(1-\lambda_2)f(x_2+t).$$

Observing that $\lambda_1 = 1 - \lambda_2$ and adding the inequalities we get

$$f(x_1 + t) + f(x_2) < f(x_1) + f(x_2 + t),$$

or

$$f(x_1 + t) - f(x_1) < f(x_2 + t) - f(x_2).$$

Without loss of generality, we can consider ℓ and a single translation $t_0 = (t_1, t_2)$. A point of intersection between these curves is a solution to the equation

$$(x_1, f(x_1)) = (x_2 + t_1, f(x_2) + t_2),$$

or

$$\begin{cases} x_1 = x_2 + t_1 \\ f(x_1) = f(x_2) + t_2 \end{cases},$$

or

$$f(x_2 + t_1) = f(x_2) + t_2.$$

A second point of intersection is a solution to

$$f(x_3 + t_1) = f(x_3) + t_2$$

where $x_3 \neq x_2$.

Adding these equations, we see that 2 points of intersection can occur only if

$$f(x_2 + t_1) - f(x_2) = f(x_3 + t_1) - f(x_3)$$

which yields a contradiction.

□

A function $f : I \to \mathbb{R}$ is concave iff $-f$ is convex. An important symmetry to notice is that concave sets A, defined by a concave function f satisfy

$$|A + A| = |(-A) + (-A)|.$$

That is, any results throughout the rest of this report concerning the size of convex sets also hold for concave sets.

The final tool we'll introduce is dyadic partitioning. Let $S \subset \mathbb{R}$ be a finite set, $f : S \to \mathbb{R}$ be a function, and M be the maximum value of $f(x)$ for $x \in S$. We partition a sum of $f(x)$ in the following way

$$\sum_{x \in S} f(x) = \sum_{j \leq \log_2(M)} \sum_{\substack{x \in S \\ 2^{j-1} \leq f(x) < 2^j}} f(x).$$

Out of the $\log_2(M)$ partitions of the sum, one of them must be the largest, so

$$\sum_{x \in S} f(x) \leq \log_2(M) \sum_{\substack{x \in S \\ 2^{k-1} \leq f(x) < 2^k}} f(x)$$

for some $k \leq \log_2(M)$.

We use this notation introduce the technique. Throughout this report, we write $\Delta = 2^{k-1}$, and

$$D = \{x \in S : f(x) \asymp \Delta\}$$

so that

$$\sum_{x \in S} f(x) \ll \log(M) \sum_{x \in D} f(x) \asymp \log(M) \Delta |D|.$$

This technique is useful in this problem because a factor of $\log(M)$ will be negligible in most circumstances. This allows us to sum only over x for which $f(x)$ is a particular order and use this to find an upper bound.

3. Statement of Problems

We'll now move on to precise statements of the mentioned problems, and statements of the most modern results.

The idea that there does not exist a set with a small sum and product set is stated precisely as

Conjecture 3.1 (Sum-Product Conjecture)**.** *For every finite set $A \subset \mathbb{R}$,*

$$\max\left(|A+A|, |A \cdot A|\right) \gtrsim |A|^2$$

The idea that convexity opposes additive structure is stated precisely as

Conjecture 3.2 (Convex Sumset Conjecture)**.** *For finite and convex set A,*

$$|A+A| \gtrsim |A|^2$$

Both of these conjectures are sharp in the sense that the power of log is not negligible. In the appendix I prove that

$$|\log([n]) + \log([n])| = |[n] \cdot [n]| = o(n^2).$$

To date, the best results for both of these conjectures are proven in [RS21].

Theorem 3.3. *For finite sets $A \subset \mathbb{R}$,*

$$\max\left(|A+A| \ , \ |A \cdot A|\right) \gtrsim |A|^{\frac{4}{3}+\frac{2}{1167}}$$

and

Theorem 3.4. *For finite and convex sets $A \subset \mathbb{R}$,*

$$|A+A| \gtrsim |A|^{\frac{30}{19}}$$

These results will not be covered in this report. In this report, the strongest results proven are

Theorem 3.5. *For finite sets $A \subset \mathbb{R}$,*

$$\max\left(|A+A|, |AA|\right) \gtrsim |A|^{\frac{4}{3}}.$$

Theorem 3.6. *For finite and convex sets $A \subset \mathbb{R}$,*

$$|A+A| \gtrsim |A|^{\frac{20}{13}}.$$

and

Theorem 3.7. *For finite and convex sets $A \subset \mathbb{R}$,*

$$|A-A| \gtrsim |A|^{\frac{8}{5}}.$$

Theorem 3.5 is the weakest amongst these results, although it is still near the best known result. Theorem 3.7 gives the best known exponent on $|A|$, and Theorem 3.6 is obtained by a slight variation on the same methods.

4. Graphs and the Crossing Number Inequality

A profoundly useful theorem in the study of sum-product conjectures is the Szemeredi-Trotter theorem, a statement about systems of points and lines.

The easiest way to prove this theorem is through the use of the crossing number inequality, a statement about how planar a given graph can be. The purpose of this section is to give a proof of the crossing number inequality. Due to the brevity of this section, the definitions and proofs given will not be as rigorous as they should be.

The first paper to employ the crossing number inequality to proving the Szemeredi-Trotter theorem is [SZE97]. This argument was largely taken from [Tao07].

There are pictures included to help illustrate some ideas.

A graph G is a pair $G = (V, E)$ where each $e \in E$ is of the form $e \subset V$ with $|e| = 2$. We call the set V the vertices, and the set E the edges. A drawing is a representation of a graph with vertices as points in the plane and edges as curves between their respective vertices. For example:

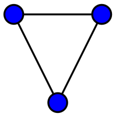

Figure 1. A drawing of a graph with $V = \{A, B, C\}$ and $E = \{\{A, B\}, \{B, C\}, \{A, C\}\}$

A graph is connected if for any 2 vertices, there exists a sequence of edges which join them. Note that connectedness is a property of a graph and not of the drawing of a graph.

There are infinitely many ways to draw any given graph. A crossing in a drawing of a graph is an intersection between 2 curves which represent edges. The crossing number of a graph is the minimum number of crossings a drawing of the graph can have. Denote this by $\mathrm{cr}(G)$. A graph G is called planar if its crossing number is 0.

A precise statement of the Crossing Number Inequality is

Theorem 4.1. *Let $G = (V, E)$ be a connected graph. If $|E| \geq 4|V|$ then*

$$\mathrm{cr}(G) \gg \frac{|E|^3}{|V|^2}.$$

For a drawing of a planar graph, we call any region of the plane which is bounded by edges a face. We also call the unbounded region of the plane a face. Here is an example of a drawing of a planar graph with labelled faces:

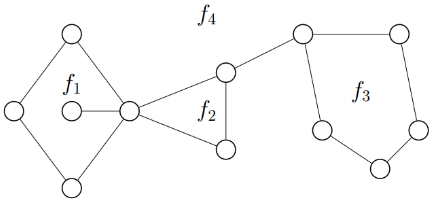

Figure 2. Drawing of Planar Graph with Labeled Faces f_i

Observe that any non-planar graph $G = (V, E)$ can be turned into a planar graph by removing at most $\operatorname{cr}(G)$ edges from E. Therefore, a bound on the number of edges a graph can have and remain planar yields a bound on the crossing number for any graph. A famous theorem relating the vertices and edges of planar graphs is

Theorem 4.2 (Euler's Formula for Planar Graphs)**.** *Let $G = (V, E)$ be a connected planar graph, with $|V| \geq 1$, and consider some drawing with 0 crossings. Let F be the set of all faces of this drawing.*

$$|V| - |E| + |F| = 2.$$

Proof. We prove this by induction on the number of edges. For the base case, $|E| = 0$, G consists of a single vertex and a single face, so

$$|V| - |E| + |F| = 2.$$

Now suppose that this equality holds for all graphs with no more than $e - 1$ edges, and consider a graph with e edges.

If G contains no cycles, there is only one face, so we may remove a vertex and a corresponding edge, which results in a graph with $e - 1$ edges satisfying Euler's formula. Because we removed one vertex and one edge, the original graph also satisfies Euler's formula.

The next case is G containing at least one cycle. If this is the case, we may remove an edge from the cycle, thereby decreasing the amount of faces by one. The remaining graph satisfies Euler's formula, and therefore the original graph does too.

By induction on the number of edges, Euler's formula holds for all connected planar graphs. □

The dependence on $|F|$ in Euler's formula can be removed by using its obvious dependence on $|E|$.

We call an edge incident to a face if the edge is one of the bounding edges which define the face. Define $\chi : F \times E \to \{0, 1\}$ as the incidence function, so $\chi(f, e) = 1$ if f and e are incident, and $\chi(f, e) = 0$ otherwise. The total number of face edge incidences is

$$I(F, E) = \sum_{f \in F} \sum_{e \in E} \chi(f, e).$$

Because the crossing number inequality is a statement about order of magnitude, we may assume $|E| \geq 3$, so that every face is incident to at least 3 edges. It follows that

$$I \geq \sum_{f \in F} 3 = 3\,|F|\,.$$

Every edge is incident to at most 2 faces, so it follows that

$$I \leq \sum_{e \in E} 2 = 2\,|E|\,.$$

Therefore

$$3\,|F| \leq 2\,|E|$$

or

$$|F| \leq \frac{2}{3}\,|E|\,.$$

Applying this to Euler's formula,

$$|V| - |E| + \frac{2}{3}|E| \geq 2$$

or

$$|E| \leq 3|V| - 6$$

when $|E| \geq 3$. Now suppose that $G = (V, E)$ is non-planar and connected. As mentioned before, G may be turned planar by removing at most $\operatorname{cr}(G)$ edges. Therefore, for any graph G with $|E| \geq 3$,

$$|E| - \operatorname{cr}(G) \leq 3|V| - 6$$

or

$$\operatorname{cr}(G) > |E| - 3|V|.$$

To further improve this inequality, we apply the probabilistic method to the deletion of vertices of G.

Let each $v \in V$ be removed with a probability $1 - p$, $p \in (0, 1)$. Let the remaining set of vertices be V'.

An edge is removed whenever either of the corresponding vertices are removed. Let the remaining set of edges be E'. Let the remaining graph be $G' = (V', E')$. We have that if $|E'| \geq 3$,

$$\operatorname{cr}(G') \geq |E'| - 3|V'|,$$

and so

$$\mathbb{E}(\operatorname{cr}(G')) \geq \mathbb{E}(|E'| - 3|V'|),$$

or, by the linearity of the expected value,

$$\mathbb{E}(\operatorname{cr}(G')) \geq \mathbb{E}(|E'|) - 3\mathbb{E}(|V'|).$$

Each $v \in V$ is removed with probability $1 - p$, so

$$\mathbb{E}(|V'|) = p|V|.$$

Each edge remains only when both corresponding vertices remain. Each vertex remains independently with a probability p, so

$$\mathbb{E}(|E'|) = p^2|E|.$$

We can bound $\mathbb{E}(\operatorname{cr}(G'))$ by considering a drawing of G with the minimum number of crossings. Each crossing remains and only when both corresponding edges remain, each of which occurs independently with probability p^2.

The expected value of the number of crossings remaining in the drawing is $p^4 \operatorname{cr}(G)$. There is no guarantee that this drawing is optimal to minimize the crossings of G', but we may conclude that

$$\mathbb{E}(\operatorname{cr}(G')) \leq p^4 \operatorname{cr}(G),$$

and therefore that

$$p^4 \operatorname{cr}(G) \geq \mathbb{E}(\operatorname{cr}(G)) \geq p^2|E| - 3p|V|$$

for any $p \in (0, 1)$.

Because we are only concerned with an order of magnitude bound, we assume $|E| \geq 4|V|$, and take $p = \frac{4|V|}{|E|}$. This yields the result

$$\operatorname{cr}(G) \geq \frac{|E|}{\left(\frac{4|V|}{|E|}\right)^2} - \frac{3|V|}{\left(\frac{4|V|}{|E|}\right)^3} = \frac{1}{16}\left(\frac{|E|^3}{|V|^2} - \frac{3|E|^3}{4|V|^2}\right) \gg \frac{|E|^3}{|V|^2}.$$

5. The Szemeredi-Trotter Theorem

With this, we can employ the argument of [SZE97] to prove the Szemeredi-Trotter Theorem. A precise statement of the Szemeredi-Trotter Theorem, first proven in [ST83] is

Theorem 5.1 (Szemeredi-Trotter Theorem)**.** *Let $P \subset \mathbb{R}^2$ be a finite set of points. Let $\mathcal{L}$ be a finite set of curves in $\mathbb{R}^2$.*

Let $\chi : P \times \mathcal{L} \to \{0,1\}$ be the incidence function between a point and a line, so

$$\chi(p,l) = \begin{cases} 1 & \textit{if } p \in l \\ 0 & \textit{otherwise} \end{cases}$$

If any two $l \in \mathcal{L}$ intersect in at most one point, then the total number of point-curve incidences,

$$I(P,\mathcal{L}) = \sum_{(p,l)\in P\times\mathcal{L}} \chi(p,l)$$

satisfies

$$I(P,\mathcal{L}) \ll |P|^{\frac{2}{3}} |\mathcal{L}|^{\frac{2}{3}} + |P| + |\mathcal{L}|.$$

Proof. This is proven by turning a system of curves and points into a graph. We first omit all points and curves which contribute to one or fewer incidences.

For each remaining curve, if there are n incidences along it, we partition it into $n-1$ curves, each with 2 incidences. These become the edges of the drawing of some graph. Let the set of edges be E.

If $I_0(P,\mathcal{L})$ is the remaining number of incidences,

$$|E| \geq I_0(P,\mathcal{L}) - |\mathcal{L}|.$$

Let the vertices of the graph, V, be the remaining points in the system. Obviously, $|V| \leq |P|$.

Because any two curves intersect in at most one point, $\operatorname{cr}(G)$ is at most $|\mathcal{L}|^2$.

Supposing that $|E| \geq 4|V|$, we have that

$$|\mathcal{L}|^2 \geq \operatorname{cr}(G) \gg \frac{(I_0(P,\mathcal{L}) - |\mathcal{L}|)^3}{|P|^2}$$

or

$$I_0(P,\mathcal{L}) \ll |\mathcal{L}|^{\frac{2}{3}} |P|^{\frac{2}{3}} + |\mathcal{L}|.$$

If $|V| \geq \frac{1}{4}|E|$, then $|V| \gg |E|$, or

$$|P| \gg I_0(P,\mathcal{L}) - |\mathcal{L}| \implies I_0(P,\mathcal{L}) \ll |P| + |\mathcal{L}|.$$

Finally, the remaining incidences not yet counted is

$$I(P,\mathcal{L}) - I_0(P,\mathcal{L}) \ll |\mathcal{L}| + |P|,$$

so

$$I(P,\mathcal{L}) \ll |P|^{\frac{2}{3}} |\mathcal{L}|^{\frac{2}{3}} + |P| + |\mathcal{L}| .$$

□

The Szemeredi Trotter theorem has a handful of direct applications to the sum product conjecture and the convex sumset conjecture. The main theorem which leads to these results is proven in [ENR00].

Theorem 5.2. *Let $A \subset \mathbb{R}$ be finite, with $|A| = n$.*
Label the elements of A so that $a_1 < a_2 < \cdots < a_n$.
Let $f : [a_1, a_n] \to \mathbb{R}$ be convex. Let $S = \{(a, f(a)) : a \in A\}$ and $T \subset \mathbb{R}^2$ be finite.
We have

$$|S+T| \gg \max\left(|S|^{\frac{3}{2}} |T|^{\frac{1}{2}}, |S|\,|T|\right).$$

Proof. Let

$$L_t = \{(x, f(x)) + t : x \in [a_1, a_n], t \in T\},$$

and let

$$\mathcal{L} = \{L_t : t \in T\}.$$

For every $x \in A$, $(x, f(x)) + t \in S + T$. Therefore, there are $|A| = |S|$ incidences between L_t and the point set $S + T$, for all $t \in T$. It follows that there are $|S|\,|T|$ total incidences. The set $\mathcal{L}$ consists of translations of the graph of a convex function, so the Szemeredi-Trotter theorem is satisfied. Thus

$$|S|\,|T| \ll |S+T|^{\frac{2}{3}} |T|^{\frac{2}{3}} + |S+T| + |T|.$$

Trivially,

$$|S+T| \geq |T|,$$

so

$$|S|\,|T| \ll \max\left(|S+T|^{\frac{2}{3}} |T|^{\frac{2}{3}}, |S+T|\right),$$

or

$$|S+T| \gg \max\left(|S|^{\frac{3}{2}} |T|^{\frac{1}{2}}, |S|\,|T|\right).$$

□

The following corollaries are also proven in [ENR00].

Corollary 5.3. *For convex and finite sets $A \subset \mathbb{R}$, and finite sets $B \subset \mathbb{R}$,*

$$|A+B| \gg |A|\,|B|^{\frac{1}{2}}.$$

Proof. Let $A \subset \mathbb{R}$ be finite. Let $n = |A|$. Let $f : [1, n] \to \mathbb{R}$ be the convex function for which

$$A = \{f(i) : i \in [n]\}.$$

Take

$$S = \{(i, f(i)) : i \in [n]\}$$

and

$$T = [n] \times B.$$

We have

$$S + T \subset ([n] + [n]) \times (A + B),$$

so

$$|S+T| \leq |[n]+[n]| \, |A+B| \ll |A| \, |A+B| \, .$$

Apply Theorem 5.2 to get

$$|A| \, |A+B| \gg |S+T| \gg \max \left(|A|^{\frac{3}{2}} \left(|A| \, |B| \right)^{\frac{1}{2}} , |A| \left(|A| \, |B| \right) \right)$$

or

$$|A| \, |A+B| \gg |A|^2 \, |B|^{\frac{1}{2}} \implies |A+B| \gg |A| \, |B|^{\frac{1}{2}} \, .$$

□

This will be used to achieve stronger results later, but in particular, this gives the result

Corollary 5.4. *For finite and convex sets $A \subset \mathbb{R}$,*

$$|A+A| \gg |A|^{\frac{3}{2}} \, .$$

We can also use Theorem 5.2 to get a result on the Sum-Product conjecture.

Corollary 5.5. *For finite sets $A \subset \mathbb{R}^+$ of positive real numbers,*

$$\max \left(|A+A| \, , |A \cdot A| \right) \gg |A|^{\frac{5}{4}} \, .$$

Proof. Let $A \subset \mathbb{R}$ be finite.

Label the elements of A so that $a_1 < \cdots < a_n$. Let $f : [a_1, a_n] \to \mathbb{R}$ be concave or convex. Take

$$S = \{ (a, f(a)) : a \in A \} \, ,$$

and

$$T = A \times f(A).$$

Observe that

$$S + T \subset (A+A) \times \left(f(A) + f(A) \right) ,$$

so that

$$|S+T| \ll |A+A| \, |f(A)+f(A)| \, .$$

Apply Theorem 5.2 to get

$$|A+A| \, |f(A)+f(A)| \gg |S+T| \gg \max \left(|A|^{\frac{3}{2}} \left(|A|^2 \right)^{\frac{1}{2}} , |A| \, |A|^2 \right)$$

or

$$\max \left(|A+A| \, , |f(A)+f(A)| \right) \gg |A|^{\frac{5}{4}}$$

for any convex or concave function $f : [a_1, a_n] \to \mathbb{R}$.

If $A \subset \mathbb{R}^+$, we may take $f(x) = \log (x)$, immediately yielding the desired result. □

The Szemeredi-Trotter theorem also provides more general tools which can be used alongside other results to sharpen the above bounds.

Theorem 5.6. *Let $A \subset \mathbb{R}$ be convex, then for every finite set $B \subset \mathbb{R}$ we have*

$$|\{ x \in A - B : \delta_{A,B}(x) \geq \tau \}| \ll \frac{|A| \, |B|^2}{\tau^3} .$$

Proof. Let $|A| = n$. Let $f : [1, n] \to \mathbb{R}$ be the convex function defining the set A. Take

$$\ell_{a,b} = \{(x, f(x)) + (a, b) : x \in [1, n]\}$$

and

$$\mathcal{L} = \{\ell_{a,b} : a \in [n], b \in -B\}.$$

Take

$$P = ([n] + [n]) \times (A - B),$$

and let $P_\tau \subset P$ be the largest subset of P for which every point has at least τ lines passing through it.

Observe that

$$|A| \, |\{x \in A - B : \delta_{A,B}(x) \geq \tau\}| \asymp |\{(p_1, p_2) \in P : \delta_{A,B}(p_2) \geq \tau\}| \tag{1}$$

Additionally,

$$\forall n_0 \in [n] + [n] \ , \ \forall x \in [n] \ , \ \exists a \in [n] \text{ s.t. } a + x = n_0,$$

so if $(p_1, p_2) \in P$ satisfies the RHS of 1, the number of solutions to

$$\ell_{a,b}(x) = (p_1, p_2)$$

for some a, b, x is τ.

That is, every point in the RHS of 1 has at least τ lines passing through it, so

$$|A| \, |\{x \in A - B : \delta_{A,B}(x) \geq \tau\}| \ll |P_\tau| \, .$$

It is clear that

$$I(P_\tau, \mathcal{L}) \geq \tau \, |P_\tau| \, ,$$

so by the Szemeredi-Trotter theorem

$$\tau \, |P_\tau| \ll |P_\tau|^{\frac{2}{3}} \, |\mathcal{L}|^{\frac{2}{3}} + |P_\tau| + |\mathcal{L}| \, .$$

Because the Theorem is trivial for $\tau \ll 1$ and $\tau > \min(|A|, |B|)$, we may assume $1 \ll \tau \leq \min(|A|, |B|)$.

We have

$$\tau \, |P_\tau| \gg |P_\tau|$$

so

$$|P_\tau| \ll \max\left(\frac{|A| \, |B|}{\tau}, \frac{|A|^2 \, |B|^2}{\tau^3}\right),$$

but

$$\frac{|A| \, |B|}{\tau^2} \gg 1,$$

and therefore it follows that

$$|P_\tau| \ll \frac{|A|^2 \, |B|^2}{\tau^3}.$$

Substituting,

$$|\{x \in A - B : \delta_{A,B}(x) \geq \tau\}| \ll \frac{|A| \, |B|^2}{\tau^3}.$$

□

An immediate corollary of this is

Corollary 5.7. *Let $A \subset \mathbb{R}$ be a convex and finite set. Let $B \subset \mathbb{R}$ be finite. Order elements $s_i \in A - B$ such that*

$$\delta_{A,B}(s_1) \geq \delta_{A,B}(s_2) \geq \cdots \geq \delta_{A,B}(s_{|A-B|}).$$

For every $1 \leq r \leq |A-B|$ we have

$$\delta_{A,B}(s_r) \ll \frac{|A|^{\frac{1}{3}} |B|^{\frac{2}{3}}}{r^{\frac{1}{3}}}.$$

Proof.

$$r = |\{x \in A - B : \delta_{A,B}(x) \geq \delta_{A,B}(s_r)\}| \ll \frac{|A| |B|^2}{\delta_{A,B}(s_r)^3} \implies \delta_A(s_r) \ll \frac{|A|^{\frac{1}{3}} |B|^{\frac{2}{3}}}{r^{\frac{1}{3}}}.$$

□

Later in the next section we'll use these results to find bounds on the additive energies between certain sets, but first we'll introduce some simpler results which use additive and multiplicative energy to show why such bounds are useful.

6. Additive and Multiplicative Energy Estimates

Recall that

$$|A+A| \geq \frac{|A|^4}{E(A)}$$

and

$$|AA| \geq \frac{|A|^4}{M(A)}.$$

Observe that finding an upper bound on $E(A)$ or $M(A)$ in terms of $|A+A|$, $|AA|$, and $|A|$ yields a sum product theorem. In this section we showcase a result which employs this idea. We also give other, more complicated, estimates on additive energies involving a convex set which will be used later.

The aforementioned result is Theorem 3.5, found in [Sol09]. It gives a stronger result on the sum-product conjecture than the Szemeredi-Trotter theorem.

Theorem. *Let $A \subset \mathbb{R}^+$ be finite.*

$$\max(|A+A|, |AA|) \gtrsim |A|^{\frac{4}{3}}.$$

Proof. We begin with a construction. Consider the set A^2, along with the smallest set of lines through the origin which cover A^2.

The claim is that each line represents an element of $\frac{A}{A}$. This is easy to see, two pairs $(a_1, a_2), (b_1, b_2) \in A^2$ give the same representation as a quotient if and only if

$$\frac{a_2}{a_1} = \frac{b_2}{b_1}.$$

Observe that this is the slope of the line through the origin and the points $(a_1, a_2), (b_1, b_2)$. This shows that the number of lines in the construction is $\left|\frac{A}{A}\right|$, the slope of each line is an element in $\frac{A}{A}$, and the number of points on a line of slope m is the number of representations of m as a quotient, $r_{\frac{A}{A}}(m)$.

We'll now prove 2 facts about the set of vector sums of 2 points lying on consecutive lines.

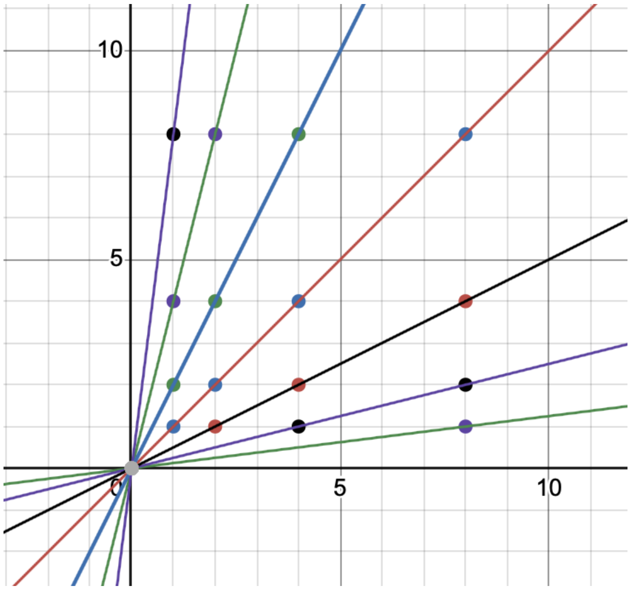

Figure 3. Example with $A = \{1, 2, 4, 8\}$.

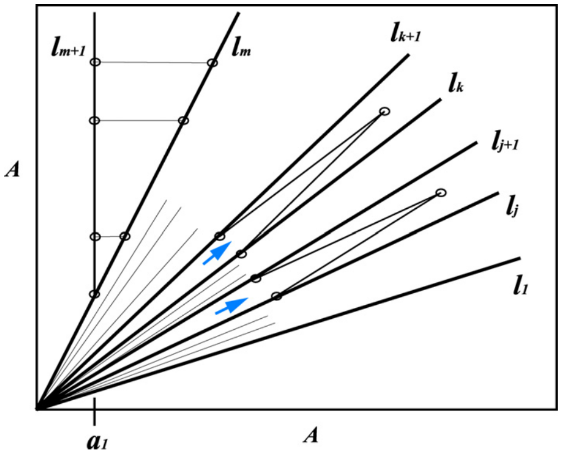

Figure 4. Vector sums of 2 points lying on consecutive lines. Figure taken from [Sol09], where the argument was first given.

Firstly, that the set of vector sums of points along either line is disjoint for each choice of consecutive lines. In figure 4, this is represented by the blue arrow between lines l_j, l_{j+1} and l_k, l_{k+1}. To show this, consider 2 consecutive lines and the set of all vector sums between a point on each line. If our points are (a_1, a_2) and (b_1, b_2), with

$$\frac{a_2}{a_1} > \frac{b_2}{b_1},$$

then the slope of their sum is

$$\frac{a_2 + b_2}{a_1 + b_1}$$

which satisfies

$$\frac{b_2}{b_1} < \frac{a_2 + b_2}{a_1 + b_1} < \frac{a_2}{a_1}.$$

That is, the vector sum must “lie between” the two lines which the original vectors are on. A consequence of this is that, for any pairs of consecutive lines, the vector sums of all points along the lines are disjoint.

The remaining claim is that for any choice of points on consecutive lines, the vector sum is distinct. To show this, consider solutions to

$$\lambda_1 v + \lambda_2 w = \lambda_3 v + \lambda_4 w.$$

We have a solution if and only if

$$(\lambda_1 - \lambda_3)\, v + (\lambda_2 - \lambda_4)\, w = 0,$$

where $\lambda_1 \neq \lambda_3$ or $\lambda_2 \neq \lambda_4$. This exists only if v and w are linearly dependent, which is untrue if $\mathrm{Span}\,(v), \mathrm{Span}\,(w)$ are distinct lines in $\mathbb{R}^2$.

We are able to use these facts to prove Theorem 3.5. Begin by applying dyadic partitioning on $M(A)$ to get

$$M(A) = \sum_{x \in \frac{A}{A}} r_{\frac{A}{A}}(x)^2 \ll \log\left(\left|\frac{A}{A}\right|\right) \tau^2 \left|S\right|$$

for some τ, where $S = \left\{x \in \frac{A}{A} : r_{\frac{A}{A}}(x) \asymp \tau\right\}$.

Consider a reduced system of points and lines, consisting only of the $|S|$ many lines which have $\asymp \tau$ many points on them. Consider the set of all vector sums between points over all consecutive lines. Because all pairs of lines give disjoint sets of sums, each with $\asymp \tau^2$ many distinct sums, there are $\tau^2 |S|$ many vector sums. The set of all vector sums between points of A^2 is a subset of $(A+A)^2$, so

$$\tau^2 \left|S\right| \leq \left|A + A\right|^2.$$

Therefore

$$\frac{\left|A\right|^4}{\left|AA\right|} \leq M(A) \ll \log\left(\left|\frac{A}{A}\right|\right) \left|A + A\right|^2 \ll \log\left(\left|A\right|\right) \left|A + A\right|^2$$

so

$$\max\left(\left|A + A\right|, \left|AA\right|\right) \gtrsim \left|A\right|^{\frac{4}{3}}.$$

□

Many arguments involving energy estimates are not as straightforward as finding an upper bound. We are often interested in quantites such as $E(A, A+A), E(A-A), E_3(A)$ etc.

In the remaining part of this section, we’ll prove two theorems concerning these quantities which we will then apply to prove more advanced results in the next section. Both of these results come as corollaries of the Szemeredi-Trotter Theorem.

Theorem 6.1. *For finite and convex sets $A \subset \mathbb{R}$,*

$$E_3(A) \lesssim \left|A\right|^3.$$

Proof. Recall that upon ordering a_i such that $\delta_A(a_1) \geq \delta_A(a_2) \geq \cdots \geq \delta_A(a_{|A-A|})$, we have that

$$\delta_A(a_r) \ll \frac{|A|}{r^{\frac{1}{3}}}.$$

With this,

$$\begin{aligned} E_3(A) &= \sum_{x\in A-A} \delta_A(x)^3 \\ &\ll |A|^3 \sum_{r=1}^{|A-A|} \frac{1}{r} \\ &\asymp |A|^3 \int_1^{|A-A|} \frac{1}{r}\,\mathrm{d}r \\ &\asymp |A|^3 \log\left(|A-A|\right) \\ &\lesssim |A|^3 \end{aligned}$$

□

Theorem 6.2. *For finite and convex sets $A\subset\mathbb{R}$, and finite sets $B\subset\mathbb{R}$*

$$E(A,B) \ll |A|\,|B|^{\frac{3}{2}}\,.$$

Proof. Denote the elements of $A-B$ by s_i where $\delta_{A,B}(s_1)\geq\cdots\geq\delta_{A,B}(s_{|A-B|})$

Let $P=\left\{x\in A-B : \delta_{A,B}(x)\geq|B|^{\frac{1}{2}}\right\}$, and let $P^*=(A-B)\setminus P$.

$$\begin{aligned} \sum_{x\in P}\delta_{A,B}(x)^2 &= \sum_{i=1}^{|P|}\delta_{A,B}(s_r)^2 \\ &\ll |A|^{\frac{2}{3}}\,|B|^{\frac{4}{3}}\sum_{i=1}^{|P|}\frac{1}{r^{\frac{2}{3}}} \\ &\asymp |A|^{\frac{2}{3}}\,|B|^{\frac{4}{3}}\,|P|^{\frac{1}{3}} \\ &\ll |A|^{\frac{2}{3}}\,|B|^{\frac{4}{3}}\left(\frac{|A|\,|B|^2}{|B|^{\frac{1}{2}}}\right)^{\frac{1}{3}} \\ &= |A|\,|B|^{\frac{3}{2}}\,, \end{aligned}$$

and

$$\sum_{x\in P^*}\delta_{A,B}(x)^2 < |B|^{\frac{1}{2}}\sum_{x\in P^*}\delta_{A,B}(x) = |A|\,|B|^{\frac{3}{2}}\,.$$

Therefore,

$$E(A,B)=\sum_{x\in P}\delta_{A,B}(x)^2+\sum_{x\in P^*}\delta_{A,B}(x)^2 \ll |A|\,|B|^{\frac{3}{2}}\,.$$

□

7. Application of Energy Estimates

This section will employ results from the previous section to prove Theorems 3.7, 3.6. This argument can be found in [SS11b].

Throughout this argument, in the same convention as [SS11b], we'll use the notation

$$A_x = A\cap(A+x)=\{a\in A : a-x\in A\}=\{a\in A : \exists a_0\in A \text{ s.t. } a-a_0=x\}\,.$$

Observe that

$$|A_x| = \delta_A(x).$$

We require the following lemmata, proven in [SS11a] and [SS11b].

Lemma 7.1. *For every set $A \subset \mathbb{R}$ we have*

$$\sum_x E\left(A, A_x\right) = E_3(A).$$

Proof. For a set S, let the function S be the indicator function, so

$$S(x) = \begin{cases} 1 \text{ if } x \in S \\ 0 \text{ otherwise} \end{cases}$$

We begin by observing that

$$A_x(s)A_x(s+t) = A(s)A(s+x)A(s+t)A(s+t+x) = A_t(s)A_t(s+x),$$

and therefore that

$$\delta_{A_x}(t) = \sum_s A_x(s)A_x(s+t) = \sum_s A_t(s)A_t(s+x) = \delta_{A_t}(x).$$

With this, we have

$$\begin{aligned}\sum_x E(A, A_x) &= \sum_s \sum_x \delta_A(s)\delta_{A_x}(s) \\ &= \sum_s \delta_A(s) \left(\sum_x \delta_{A_x}(s) \right) \\ &= \sum_s \delta_A(s) \left(\sum_x \delta_{A_s}(x) \right) \\ &= \sum_s \delta_A(s) \left|A_s\right|^2 \\ &= E_3(A)\end{aligned}$$

□

Lemma 7.2. *For every set $A \subset \mathbb{R}$, and $P \subset A - A$, if η is the number for which*

$$\sum_{x \in P} |A_x| = \eta \left|A\right|^2,$$

then

$$\sum_{x \in P} |A \pm A_x| \geq \frac{\eta^2 \left|A\right|^6}{E_3(A)}.$$

Proof.

$$\begin{aligned}|A| \left|A_x\right| = \sum_{s \in A + A_x} \sigma_{A,A_x}(s) = \sum_{s \in A - A} \delta_{A,A_x}(s) \\ \leq E\left(A, A_x\right)^{\frac{1}{2}} \left|A \pm A_x\right|^{\frac{1}{2}}\end{aligned}$$

so

$$\begin{aligned}\eta\,|A|^3 &= \sum_{x\in P}|A|\,|A_x| \\ &\leq \sum_{x\in P}E(A,A_x)^{\frac{1}{2}}\,|A\pm A_x|^{\frac{1}{2}} \\ &\leq \left(\sum_{x\in P}E(A,A_x)\right)^{\frac{1}{2}}\left(\sum_{x\in P}|A\pm A_x|\right)^{\frac{1}{2}}\end{aligned}$$

and therefore

$$\sum_{x\in P}|A\pm A_x| \geq \frac{\eta^2\,|A|^6}{E_3(A)}.$$

□

We're now able to prove the theorems. Recall the following theorems

Theorem. *For finite and convex sets $A\subset\mathbb{R}$,*

$$|A+A| \gtrsim |A|^{\frac{20}{13}}.$$

and

Theorem. *For finite and convex sets $A\subset\mathbb{R}$,*

$$|A-A| \gtrsim |A|^{\frac{8}{5}}.$$

Proof of Theorems 3.7, 3.6. Denote the difference and sum sets by $D=|A-A|$ and $S=|A+A|$. Observe that

$$A-A_x\subset D\cap D_x$$

and

$$A+A_x\subset S\cap S_x.$$

We use this observation along with Lemma 7.2 to find bounds on the additive energies between D and S and A.

Consider the popular sets of differences P,P' defined by

$$P=\left\{x\in A-A:\delta_A(x)\geq\frac{|A|^2}{2\,|A-A|}\right\}$$

and

$$P'=\left\{x\in A-A:\delta_A(x)\geq\frac{|A|^2}{2\,|A+A|}\right\}.$$

Let P^* and $P^{'*}$ be their respective compliments. We have

$$\sum_{x\in P^*}|A_x| < \frac{|A|^2}{2\,|A-A|}\cdot|P^*|\leq\frac{|A|^2}{2},$$

so

$$\sum_{x\in P}|A_x| \gg |A|^2. \tag{2}$$

We also have

$$\sum_{x\in P'^{*}} |A_x|^2 < \frac{|A|^2}{2\,|A+A|} \cdot |A|^2 \,,$$

so

$$\sum_{x\in P'} |A_x|^2 \gg \frac{|A|^4}{|A+A|}.$$

We can directly apply (2) to Lemma 7.2. We have

$$\eta\,|A|^2 = \sum_{x\in P} |A_x| \gg |A|^2 \implies \eta \gg 1$$

so

$$\sum_{x\in P} |A \pm A_x| \gg \frac{|A|^6}{E_3(A)} \gtrsim |A|^3 \,.$$

Recall that $A - A_x \subset D \cap D_x$, or $|A - A_x| \leq |D_x| = \delta_D(x)$. It follows that

$$\sum_{x\in P} \delta_D(x) \gtrsim |A|^3 \,.$$

By the definition of P it follows that

$$E(A,D) \geq \sum_{x\in P} \delta_A(x)\delta_D(x) \geq \frac{|A|^2}{2\,|A-A|} \sum_{x\in P} \delta_D(x) \gtrsim \frac{|A|^5}{|A-A|}.$$

By applying Theorem 6.2,

$$\frac{|A|^5}{|A-A|} \lesssim |A|\,|A-A|^{\frac{3}{2}} \implies |A-A| \gtrsim |A|^{\frac{8}{5}} \,.$$

Recall that we have

$$\frac{|A|^4}{|A+A|} \ll \sum_{x\in P'} |A_x|^2 \,.$$

Additionally, by applying Theorem 5.6 and Corollary 5.7 we have that

$$\begin{aligned}
\sum_{x\in P' : |A_x| \gg \frac{|A+A|}{|A|}} |A_x|^2 &\ll \sum_{r=1}^{\frac{|A|^6}{|A+A|^3}} \left(\frac{|A|}{r^{\frac{1}{3}}}\right)^2 \\
&\asymp |A|^2 \int_1^{\frac{|A|^6}{|A+A|^3}} \frac{1}{r^{\frac{2}{3}}}\ \mathrm{d}r \\
&\asymp |A|^2 \cdot \frac{|A|^2}{|A+A|} \\
&= \frac{|A|^4}{|A+A|}
\end{aligned}$$

and therefore that

$$\frac{|A|^4}{|A+A|} \ll \sum_{x\in P' : |A_x| \ll \frac{|A+A|}{|A|}} |A_x|^2 \,.$$

We apply a dyadic partitioning to get

$$\frac{|A|^4}{|A+A|} \ll \sum_{x\in P' : |A_x| \ll \frac{|A+A|}{|A|}} |A_x|^2$$
$$\lesssim \Delta \sum_{x\in D} |A_x|$$

where

$$D = \left\{ x \in P' : \frac{|A+A|}{|A|} \gg |A_x| \asymp \Delta \right\}.$$

This gives

$$\sum_{x\in D} |A_x| \gtrsim \frac{|A|^5}{|A+A|^2}.$$

Applying Lemma 7.2, we get

$$\sum_{x\in D} |A+A_x| \gtrsim \frac{|A|^9}{|A+A|^4}.$$

Recalling that $A+A_x \subset S_x$, we have

$$\sum_{x\in D} \delta_S(x) \geq \sum_{x\in D} |A+A_x| \gtrsim \frac{|A|^9}{|A+A|^4},$$

which, by the definition of P' gives

$$\frac{|A|^{11}}{|A+A|^5} \lesssim \sum_{x\in D} \delta_S(x)\delta_A(x) \leq E(A,S) \ll |A|\,|A+A|^{\frac{3}{2}},$$

or

$$|A+A| \gtrsim |A|^{\frac{20}{13}}.$$

□

Acknowledgements. This report is based on work supported by NSF grant DMS-2051032. I'd like to express my gratitude to the Department of Mathematics at Indiana University Bloomington, as well as thanks to my advisor Shukun Wu, who greatly assisted me in my studies.

References

[ENR00] György Elekes, Melvyn B Nathanson, and Imre Z Ruzsa, *Convexity and sumsets*, Journal of Number Theory **83** (2000), no. 2, 194–201.

[ES83] P. Erdős and E. Szemerédi, *On sums and products of integers*, pp. 213–218, Birkhäuser Basel, Basel, 1983.

[For08] Kevin Ford, *The distribution of integers with a divisor in a given interval*, Annals of Mathematics **168** (2008), 367–433.

[RS21] Misha Rudnev and Sophie Stevens, *An update on the sum-product problem*, 2021.

[Sol09] József Solymosi, *Bounding multiplicative energy by the sumset*, Advances in Mathematics **222** (2009), 402–408.

[SS11a] T. Schoen and I. D. Shkredov, *On a question of cochrane and pinner concerning multiplicative subgroups*, 2011.

[SS11b] Tomasz Schoen and Ilya D. Shkredov, *On sumsets of convex sets*, 2011.

[ST83] E. Szemerédi and W. T. Trotter, *Extremal problems in discrete geometry*, Combinatorica **3** (1983), no. 3, 381–392.

[SZE97] LASZLO A. SZEKELY, *Crossing numbers and hard erdős problems in discrete geometry*, Combinatorics, Probability and Computing **6** (1997), no. 3, 353–358.

[Tao07] Terence Tao, *The crossing number inequality*, 2007.

Appendix

Theorem (Uniqueness of Binary Representations)**.** *For any $x \in \mathbb{N}_0$, there is a unique sequence of numbers $(a_i)_{i=0}^{I}$, with $I \leq \log_2(x)$ and $a_i \in \{0,1\}$ such that*

$$x = \sum_{i \leq \log_2(x)} a_i 2^i$$

Proof. Let $x \in \mathbb{N}_0$ and suppose that there are two binary representations for x, that is, there are $r, s \in \mathbb{N}_0$, and sequences $(a_i)_{i=0}^{r-1}$ and $(b_i)_{i=0}^{s-1}$ such that

$$x = 2^r + a_{r-1}2^{r-1} + \cdots + a_0 = 2^s + b_{s-1}2^{s-1} + \cdots + b_0.$$

We first prove that both binary representations must be of the same degree, or have the same highest power of 2. Without loss of generality, suppose $s > r$. It follows that

$$x = 2^r + a_{r-1}2^{r-1} + \cdots + a_0 \leq \sum_{i \leq r} 2^i = 2^{r+1} - 1 < 2^s \leq 2^s + b_{s-1}2^{s-1} + \cdots + b_0,$$

a contradiction. With the fact that any representations must be of the same degree, we have that

$$0 = (a_{r-1} - b_{r-1})\, 2^{r-1} + \cdots + (a_0 - b_0)$$

is of degree 0, and therefore that for all i, $a_i = b_i$, or that the binary representation is unique. □

Past this point in the appendix, we will freely use shorthand which was discussed in the preliminaries section. We also introduce the notation

$$f(x) = o(g(x))$$

if

$$\lim_{x \to \infty} \frac{f(x)}{g(x)} = 0.$$

We'll introduce the following theorems together due to their relatedness.

Theorem.

$$|[n] \cdot [n]| \gg \frac{n^2}{\log(n)}.$$

Theorem.

$$|[n] \cdot [n]| = o(n^2).$$

This is known as the "multiplication table theorem" because the quantity

$$|[n] \cdot [n]|$$

is the number of distinct numbers in an $n \times n$ multiplication table:

×	1	2	3	4	5	6	7	8	9	10
1	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
2	2	4	6	8	10	**12**	**14**	**16**	**18**	**20**
3	3	6	9	12	**15**	18	**21**	**24**	**27**	**30**
4	4	8	12	16	20	24	**28**	**32**	**36**	**40**
5	5	10	15	20	**25**	30	**35**	40	**45**	**50**
6	6	12	18	24	30	36	**42**	**48**	**54**	**60**
7	7	14	21	28	35	42	**49**	**56**	**63**	**70**
8	8	16	24	32	40	48	56	**64**	**72**	**80**
9	9	18	27	36	45	54	63	72	**81**	**90**
10	10	20	30	40	50	60	70	80	90	**100**

FIGURE 5. 10 × 10 multiplication table with distinct numbers highlighted in red.

These theorems are just bounds on the asymptotic behavior of $|[n]\cdot[n]|$. In [For08], it is proven that the exact order of this quantity is

$$|[n]\cdot[n]| \asymp \frac{n^2}{\log(n)^{\delta}(\log\log(n))^{\frac{3}{2}}},$$

where

$$\delta = 1 - \frac{1+\log\log(2)}{2}.$$

The exact asymptotic behavior of this quantity is not known. To prove the first, we require a lemma and a prerequisite theorem.

Lemma (Abel's Summation Formula). *Let $(a_n)_{n=1}^{\infty}$ be a sequence of real numbers. Let $A : \mathbb{R}\to\mathbb{R}$ be defined by*

$$A(t) = \sum_{n\leq t} a_n.$$

For $x\in\mathbb{R}$ and any differentiable function $\phi : [1, y] \to \mathbb{R}$,

$$\sum_{n\leq x} a_n\phi(n) = A(x)\phi(x) - \int_1^x A(t)\phi'(t)\ \mathrm{d}t$$

Proof.

$$\begin{aligned}
\sum_{n\leq x} a_n\phi(n) &= a_1\phi(1) + \cdots + a_{\lfloor x\rfloor}\phi(\lfloor x\rfloor) \\
&= A(1)\phi(1) + (A(2)-A(1))\,\phi(2) + \cdots + (A(\lfloor x\rfloor) - A(\lfloor x\rfloor - 1))\,\phi(\lfloor x\rfloor) \\
&= (\phi(1)-\phi(2))\,A(1) + \cdots + (\phi(\lfloor x\rfloor - 1) - \phi(\lfloor x\rfloor))\,A(\lfloor x\rfloor - 1) + \phi(\lfloor x\rfloor)A(\lfloor x\rfloor) \\
&= \phi(\lfloor x\rfloor)A(\lfloor x\rfloor) - \sum_{i=2}^{\lfloor x\rfloor}(\phi(i)-\phi(i-1))\,A(i-1) \\
&= \phi(\lfloor x\rfloor)\,A(\lfloor x\rfloor) - \sum_{i=2}^{\lfloor x\rfloor}\left(\int_{i-1}^{i}\phi'(t)\ \mathrm{d}t\right)A(i-1)
\end{aligned}$$

Observe that, for $i \in \mathbb{Z}$ and $t \in [i-1, i)$,

$$A(t) = A(\lfloor t \rfloor) = A(i-1),$$

so

$$\begin{aligned}\sum_{n \leq x} a_n \phi(n) &= \phi\left(\lfloor x \rfloor\right) A(\lfloor x \rfloor) - \sum_{i=2}^{\lfloor x \rfloor} \left(\int_{i-1}^{i} \phi'(t)\ \mathrm{d}t \right) A(i-1) \\ &= \phi(\lfloor x \rfloor) A(\lfloor x \rfloor) - \sum_{i=2}^{\lfloor x \rfloor} \int_{i-1}^{i} A(t)\phi'(t)\ \mathrm{d}t \\ &= \phi(\lfloor x \rfloor) A(\lfloor x \rfloor) - \int_{1}^{\lfloor x \rfloor} A(t)\phi'(t)\ \mathrm{d}t\end{aligned}$$

and

$$\begin{aligned}\int_{1}^{\lfloor x \rfloor} A(t)\phi'(t)\ \mathrm{d}t &= \int_{1}^{x} A(t)\phi'(t)\ \mathrm{d}t - \int_{\lfloor x \rfloor}^{x} A(t)\phi'(t)\ \mathrm{d}t \\ &= \int_{1}^{x} A(t)\phi'(t)\ \mathrm{d}t - A(\lfloor x \rfloor)\left(\phi(x) - \phi(\lfloor x \rfloor)\right) \\ &= \int_{1}^{x} A(t)\phi'(t)\ \mathrm{d}t - A(x)\phi(x) + A(\lfloor x \rfloor)\phi(\lfloor x \rfloor).\end{aligned}$$

Substituting this we get

$$\begin{aligned}\sum_{n \leq x} a_n \phi(n) &= \phi(\lfloor x \rfloor) A(\lfloor x \rfloor) - \int_{1}^{\lfloor x \rfloor} A(t)\phi'(t)\ \mathrm{d}t \\ &= A(x)\phi(x) - \int_{1}^{x} A(t)\phi'(t)\ \mathrm{d}t\end{aligned}$$

□

The following prerequisite theorem gives the order of magnitude of 3 functions which will prove useful.

Theorem 7.3. *Define the functions $\theta, \psi, \pi : [1, \infty) \to \mathbb{R}$ by*

$$\theta(x) = \sum_{\substack{p \leq x \\ p\ prime}} \log\left(p\right),$$

$$\psi(x) = \sum_{\substack{p^\alpha \leq x \\ p\ prime \\ \alpha \in \mathbb{N}}} \log\left(p\right),$$

$$\pi(x) = \sum_{\substack{p \leq x \\ p\ prime}} 1.$$

We have

$$\psi(x) \asymp \theta(x) \asymp x$$

and

$$\pi(x) \asymp \frac{x}{\log(x)}.$$

Proof. The functions θ and ψ are clearly related by

$$\psi(t) = \prod_{\substack{p \leq N^{\frac{1}{\alpha}} \\ p \text{ prime} \\ \alpha \in \mathbb{N}}} \log(p) = \sum_{\alpha \in \mathbb{N}} \theta\left(t^{\frac{1}{\alpha}}\right) = \theta(t) + \sum_{\alpha \geq 2} \theta\left(t^{\frac{1}{\alpha}}\right).$$

Note that the sum over α has only finitely many terms. The sum terminates when

$$2 \geq t^{\frac{1}{\alpha}} \implies \alpha \leq \log_2(t).$$

A trivial upper bound on $\theta(t)$ is

$$\theta(t) = \sum_{\substack{p \leq t \\ p \text{ prime}}} \log(p) \leq t \log(t),$$

so

$$\begin{aligned} \psi(t) &= \theta(t) + \sum_{2 \leq \alpha \leq \log_2(t)} \theta(t^{\frac{1}{\alpha}}) \\ &\leq \theta(t) + \log_2(t)\, \theta(t^{\frac{1}{2}}) \\ &\leq \theta(t) + \frac{t^{\frac{1}{2}} \log(t)^2}{\log(2)} \end{aligned}$$

or

$$\psi(t) \ll \max\left(\theta(t), t^{\frac{1}{2}} \log(t)^2\right).$$

By showing $\theta(t) \ll t$, we will have shown $\psi(t) \ll \theta(t) \ll t$. Because $\theta(t) \leq \psi(t)$, we will have shown that $\psi(t) \asymp \theta(t) \ll t$.

We have

$$\theta(t) = \sum_{\substack{p \leq t \\ p \text{ prime}}} \log(p) = \log\left(\prod_{\substack{p \leq t \\ p \text{ prime}}} p\right),$$

so it is sufficient to show that

$$\prod_{\substack{p \leq t \\ p \text{ prime}}} p \ll e^t.$$

It is also sufficient to prove it for $t \in \mathbb{N}$ because $\theta(t) = \theta(\lfloor t \rfloor)$.

For some natural number t, and a prime p,

$$t + 1 < p \leq 2t + 1 \implies p \mid \binom{2t+1}{t} = \frac{(2t+1)!}{t!(t+1)!}.$$

Therefore, for any t,

$$\prod_{\substack{t+1 < p \leq 2t+1 \\ p \text{ prime}}} p \mid \binom{2t+1}{t} \implies \prod_{\substack{t+1 < p \leq 2t+1 \\ p \text{ prime}}} p \leq \binom{2t+1}{t},$$

which gives us

$$2\prod_{\substack{t+1<p\leq 2t+1\\ p\text{ prime}}} p \leq 2\binom{2t+1}{t} \leq (1+1)^{2t+1} \implies \prod_{\substack{t+1<p\leq 2t+1\\ p\text{ prime}}} p \leq 4^t.$$

The rest follows by induction on t. Because we are proving a statement about order of magnitude, the base case is trivial. Now suppose that for some $t \in \mathbb{N}$,

$$\prod_{\substack{p\leq m\\ p\text{ prime}}} p \ll e^t.$$

If t is odd, the induction follows trivially. If t is even, let $t = 2m$, so

$$\prod_{\substack{p\leq 2m\\ p\text{ prime}}} p \ll e^{2m}.$$

We have

$$\begin{aligned}\prod_{\substack{p\leq 2m+1\\ p\text{ prime}}} p = \prod_{\substack{p\leq m+1\\ p\text{ prime}}} p &\prod_{\substack{m+1<p\leq 2m+1\\ p\text{ prime}}} p\\ &\ll e^{m+1}4^m\\ &\ll e^{m+1}e^m = e^{2m+1}.\end{aligned}$$

Thus, $\psi(t) \asymp \theta(t) \ll t$. To prove $\psi(t) \asymp \theta(t) \asymp t$, it suffices to show that $\psi(t) \gg t$.

Observe that for some number $N \in \mathbb{N}$, the prime factorization of $N!$ is of the form

$$N! = \prod_{\substack{p\leq N\\ p\text{ prime}}} p^{\alpha(N,p)},$$

where

$$\alpha(N,p) = \sum_{i\in\mathbb{N}} \left\lfloor \frac{N}{p^i} \right\rfloor = \sum_{i\leq \log_p(N)} \left\lfloor \frac{N}{p^i} \right\rfloor = \sum_{i\leq \log_2(N)} \left\lfloor \frac{N}{p^i} \right\rfloor.$$

Therefore,

$$\binom{2n}{n} = \frac{(2n)!}{(n!)^2} = \prod_{\substack{p\leq 2n\\ p\text{ prime}}} p^{\sum_{i\in\mathbb{N}}\left(\left\lfloor \frac{2n}{p^i}\right\rfloor - 2\left\lfloor \frac{n}{p^i}\right\rfloor\right)}.$$

In general for some $x \in \mathbb{R}, k \in \mathbb{N}$, if $k \leq x < k+1$, then

$$2k \leq 2x < 2k+2$$

so

$$2\lfloor x \rfloor \leq \lfloor 2x \rfloor \leq 2\lfloor x \rfloor + 1.$$

It follows that

$$\left\lfloor \frac{2n}{p^i} \right\rfloor - 2\left\lfloor \frac{n}{p^i} \right\rfloor \leq 1$$

so

$$\sum_{i\in\mathbb{N}} \left(\left\lfloor \frac{2n}{p^i} \right\rfloor - 2\left\lfloor \frac{n}{p^i} \right\rfloor \right) \leq \sum_{i\leq \log_p(2n)} 1 = \left\lfloor \frac{\log{(2n)}}{\log{(p)}} \right\rfloor,$$

and therefore

$$\log \binom{2n}{n} \leq \sum_{\substack{p \leq 2n \\ p \text{ prime}}} \left\lfloor \frac{\log(2n)}{\log(p)} \right\rfloor \log(p) = \psi(2n).$$

Observing that

$$\log \binom{2n}{n} = \log \left(\frac{n+1}{1} \cdot \frac{n+2}{2} \cdot \dots \cdot \frac{n+n}{n} \right) \geq \log(2^n) \gg n,$$

it follows that

$$\psi(t) \gg t.$$

We have shown

$$\psi(t) \asymp \theta(t) \asymp t.$$

By applying Abel's summation formula,

$$\begin{aligned} \pi(x) &= \sum_{\substack{p \leq x \\ p \text{ prime}}} 1 \\ &= \sum_{\substack{p \leq x \\ p \text{ prime}}} \log(p) \cdot \frac{1}{\log(p)} \\ &= \theta(x) \cdot \frac{1}{\log(x)} + \int_2^x \frac{\theta(t)}{t \log^2(t)} \, \mathrm{d}t \\ &\asymp \frac{x}{\log(x)} + \int_2^x \frac{1}{\log^2(t)} \, \mathrm{d}t \end{aligned}$$

and

$$\begin{aligned} \frac{x}{\log(x)} &\asymp \int_2^x \left(\frac{t}{\log(t)} \right)' \mathrm{d}t \\ &= \int_2^x \frac{1}{\log(t)} \, \mathrm{d}t - \int_2^x \frac{1}{\log^2(t)} \, \mathrm{d}t \\ &\gg \int_2^x \frac{1}{\log^2(t)} \, \mathrm{d}t \end{aligned}$$

so

$$\pi(x) \asymp \frac{x}{\log(x)}.$$

□

We are now able to prove the first theorem.

Theorem.

$$|[n] \cdot [n]| \gg \frac{n^2}{\log(n)}.$$

Proof. Let p_i be the i-th prime number and k be chosen such that p_k is the largest prime $p_k \leq n$. Consider the set of numbers

$$P = \{p_i \cdot m : i \leq k \ , \ m \leq p_i\}.$$

We have $P \subset [n] \cdot [n]$ and $p_i \cdot m$ is distinct for every choice of i, m.

It follows that, by applying Abel's summation formula,

$$\begin{aligned}|[n]\cdot[n]| &\geq |P| \\ &= \sum_{\substack{p\leq n\\ p \text{ prime}}} p \\ &= \pi(n)\cdot n - \int_2^n \pi(t)\ \mathrm{d}t \\ &\gg \pi(n)\cdot n \\ &\gg \frac{n^2}{\log(n)}\end{aligned}$$

□

We require 2 additional lemmata to prove the second theorem.

Lemma. *Let X be a real random variable with variance σ^2. For any $t \in \mathbb{R}^+$,*

$$\mathbb{P}\left(|X - \mathbb{E}(X)| \geq t\right) \leq \frac{\sigma^2}{t^2}$$

Proof. We prove this for the case of discrete X. The proof for continuous X follows similarly.

Let N be the number of possible values of X.

For any interval $I \subset \mathbb{R}$, let $1_I : \mathbb{R} \to \{1, 0\}$ be defined by

$$1_I(x) = 1 \text{ if } x \in I\ ,\ 1_I(x) = 0 \text{ if } x \notin I.$$

We have

$$\mathbb{P}(X \geq t) = \frac{\sum_x 1_{[t,\infty)}(x)}{N} \leq \frac{\sum_x \frac{x}{t}}{N} = \frac{\sum_x \frac{x}{N}}{t} = \frac{\mathbb{E}(X)}{t}$$

so

$$\mathbb{P}\left(|X - \mathbb{E}(x)| \geq t\right) = \mathbb{P}\left(|X - \mathbb{E}(X)|^2 \geq t^2\right) \leq \frac{\sigma^2}{t^2}.$$

□

Lemma.

$$\sum_{\substack{p\leq n\\ p \text{ prime}}} \frac{1}{p} = \log\log(n) + O(1).$$

Proof. Recall that for some number $N \in \mathbb{N}$, the prime factorization of $N!$ is of the form

$$N! = \prod_{\substack{p\leq N\\ p \text{ prime}}} p^{\alpha(N,p)},$$

where

$$\alpha(N,p) = \sum_{i\in\mathbb{N}} \left\lfloor \frac{N}{p^i} \right\rfloor = \sum_{i\leq \log_p(N)} \left\lfloor \frac{N}{p^i} \right\rfloor = \sum_{i\leq \log_2(N)} \left\lfloor \frac{N}{p^i} \right\rfloor.$$

It follows that

$$\begin{aligned}\log(N!) &= \sum_{\substack{p\leq N\\ p\text{ prime}}} \alpha(N,p)\log(p)\\ &= \sum_{\substack{p\leq N\\ p\text{ prime}\\ i\leq\log_p(N)}} \left\lfloor\frac{N}{p^i}\right\rfloor \log(p)\\ &= \sum_{\substack{p\leq N\\ p\text{ prime}\\ i\leq\log_p(N)}} \left(\frac{N}{p^i}-\delta(p)\right)\log(p)\\ &= N\sum_{\substack{p\leq N\\ p\text{ prime}\\ i\leq\log_p(N)}} \frac{\log(p)}{p^i} - \sum_{\substack{p\leq N\\ p\text{ prime}\\ i\leq\log_p(N)}} \delta(p)\log(p).\end{aligned}$$

We have that

$$i\leq\log_p(N) \iff p^i\leq N,$$

so

$$\sum_{\substack{p\leq N\\ p\text{ prime}\\ i\leq\log_p(N)}} \log(p)=\psi(N),$$

and therefore

$$\sum_{\substack{p\leq N\\ p\text{ prime}\\ i\leq\log_2(N)}} \frac{\log(p)}{p^i}\leq\frac{\log(N!)}{N}+\frac{\psi(N)}{N}.$$

We also have, via a Riemann sum,

$$\begin{aligned}\log(N!) &= \sum_{i=1}^N \log(i)\\ &= \int_2^N \log(i)\ \mathrm{d}i+O(1)\\ &= [i\log(i)-i]|_{i=1}^N+O(1)\\ &= N\log(N)-N+O(1).\end{aligned}$$

This leads to

$$\sum_{\substack{p\leq N\\ p\text{ prime}\\ i\leq\log_2(N)}} \frac{\log(p)}{p^i}=\frac{N\log(N)-N+O(1)}{N}+\frac{O(N)}{N}=\log(N)+O(1).$$

Moreover, we have that

$$\begin{aligned}\sum_{\substack{p\leq N\\ p\text{ prime}\\ 2\leq i\leq\log_2(N)}}\frac{\log(p)}{p^i} &\leq \sum_{x,i\geq 2}\frac{\log(x)}{x^i}\\ &= \sum_{x\geq 2}\sum_{i\geq 2}\frac{\log(x)}{x^i}\\ &= \sum_{x\geq 2}\frac{1}{x^2}\left(\frac{\log(x)}{1-\frac{1}{x}}\right)\\ &= \sum_{x\geq 2}\frac{\log(x)}{x^2-x}\end{aligned}$$

which, by the limit comparison test, converges if

$$\sum_{x\geq 2}\frac{\log(x)}{x^2}$$

converges.

Apply L'Hopital's rule to see that

$$\forall\epsilon>0\ ,\ \log(x)=o(x^{\epsilon}),$$

and therefore that

$$\begin{aligned}\sum_{x\geq 2}\frac{\log(x)}{x^2} &= \sum_{x\geq 2}\frac{o(1)}{x^{2-\epsilon}}\\ &\ll \sum_{x\geq 2}\frac{1}{x^{2-\epsilon}}\\ &\asymp \int_2^{\infty}\frac{1}{x^{2-\epsilon}}\ \mathrm{d}x\end{aligned}$$

which converges for $\epsilon<1$. It follows that

$$\sum_{\substack{p\leq N\\ p\text{ prime}}}\frac{\log(p)}{p}=\log(N)+O(1).$$

Applying Abel's summation formula,

$$\begin{aligned}
\sum_{\substack{p\leq N\\ p\text{ prime}}} \frac{1}{p} &= \sum_{\substack{p\leq N\\ p\text{ prime}}} \frac{\log(p)}{p}\cdot\frac{1}{\log(p)}\\
&= \left(\sum_{\substack{p\leq N\\ p\text{ prime}}} \frac{\log(p)}{p}\right)\cdot\frac{1}{\log(N)} + O(1) + \int_2^N \frac{\sum_{\substack{p\leq t\\ p\text{ prime}}} \frac{\log(p)}{p}}{t\log^2(t)}\,\mathrm{d}t\\
&= \frac{\log(N)+O(1)}{\log(N)} + O(1) + \int_2^N \frac{\log(t)+O(1)}{t\log^2(t)}\,\mathrm{d}t\\
&= O(1) + \int_2^N \frac{1}{t\log(t)}\,\mathrm{d}t + \int_2^N \frac{O(1)}{t\log^2(t)}\,\mathrm{d}t\\
&= \log\log(N) + O(1)
\end{aligned}$$

□

We are now able to prove

Theorem.

$$|[n]\cdot[n]| = o(n^2).$$

Proof. Define the prime power counting function $\Omega:\mathbb{N}\to\mathbb{N}$ by

$$\Omega(x) = \sum_{\substack{p^\alpha | x\\ p\text{ prime}\\ \alpha\in\mathbb{N}}} 1.$$

Let $N\in\mathbb{N}$. We prove this theorem by examining the distribution of $\Omega(X)$ for an uniformly distributed random variable X of natural numbers $x\leq N$. Let $Y=\Omega(X)$.

We first calculate $\mathbb{E}(Y)$.

$$\begin{aligned}
\mathbb{E}(Y) &= \sum_{x\leq N} \frac{1}{N}\cdot \Omega(x) \\
&= \frac{1}{N}\sum_{x\leq N} \sum_{\substack{p^\alpha | x \\ p \text{ prime} \\ \alpha\in\mathbb{N}}} 1 \\
&= \frac{1}{N}\sum_{\substack{p^\alpha \leq N \\ p \text{ prime} \\ \alpha\in\mathbb{N}}} \left\lfloor \frac{N}{p^\alpha} \right\rfloor \\
&= \frac{1}{N}\sum_{\substack{p^\alpha \leq N \\ p \text{ prime} \\ \alpha\in\mathbb{N}}} \left(\frac{N}{p^\alpha} - \delta_{p^\alpha} \right) \\
&= \sum_{\substack{p^\alpha \leq N \\ p \text{ prime} \\ \alpha\in\mathbb{N}}} \frac{1}{p^\alpha} - \frac{1}{N}\sum_{\substack{p^\alpha \leq N \\ p \text{ prime} \\ \alpha\in\mathbb{N}}} \delta_{p^\alpha} \\
&= \log\log(N) + O(1) - \frac{O(1)\psi(N)}{N} \\
&= \log\log(N) + O(1).
\end{aligned}$$

Now we calculate $V(Y)$, we have that

$$V(Y) = \mathbb{E}(Y^2) - \mathbb{E}(Y)^2 = \mathbb{E}(Y^2) - \left(\log\log(N) + O(1)\right)^2,$$

so it suffices to calculate

$$\mathbb{E}(Y^2).$$

$$\begin{aligned}
\mathbb{E}\left(Y^2\right) &= \frac{1}{N}\sum_{x\leq N}(\sum_{\substack{p^\alpha | x \\ p \text{ prime} \\ \alpha\in\mathbb{N}}} 1)^2 \\
&= \frac{1}{N}\sum_{x\leq N} \sum_{\substack{(p^\alpha, q^\beta) \\ p,q \text{ prime} \\ \alpha,\beta\in\mathbb{N} \\ p^\alpha, q^\beta | x}} 1 \\
&= \frac{1}{N}\sum_{x\leq N} \sum_{\substack{(p^\alpha, q^\beta) \\ p\neq q \\ p,q \text{ prime} \\ \alpha,\beta\in\mathbb{N} \\ p^\alpha, q^\beta | x}} 1 + \frac{1}{N}\sum_{x\leq N} \sum_{\substack{(p^\alpha, p^\beta) \\ p \text{ prime} \\ \alpha,\beta\in\mathbb{N} \\ p^\alpha, p^\beta | x}} 1
\end{aligned}$$

$$
\begin{aligned}
\mathbb{E}\left(Y^2\right) &= \frac{1}{N}\sum_{x\leq N}\sum_{\substack{(p^\alpha,q^\beta)\\ p\neq q\\ p,q \text{ prime}\\ \alpha,\beta\in\mathbb{N}\\ p^\alpha,q^\beta|x}} 1 + \frac{1}{N}\sum_{x\leq N}\sum_{\substack{(p^\alpha,p^\beta)\\ p \text{ prime}\\ \alpha,\beta\in\mathbb{N}\\ p^\alpha,p^\beta|x}} 1 \\
&= \frac{1}{N}\sum_{\substack{\left(p^\alpha,q^\beta\right)\\ p,q \text{ prime}\\ p\neq q\\ p^\alpha q^\beta\leq N\\ \alpha,\beta\in\mathbb{N}}} \left\lfloor\frac{N}{p^\alpha q^\beta}\right\rfloor + \frac{1}{N}\sum_{\substack{(p^\alpha,p^\beta)\\ p^{\max(\alpha,\beta)}\leq N\\ p \text{ prime}\\ \alpha,\beta\in\mathbb{N}}} \left\lfloor\frac{N}{p^{\max(\alpha,\beta)}}\right\rfloor \\
&\leq \sum_{\substack{\left(p^\alpha,q^\beta\right)\\ p,q \text{ prime}\\ p\neq q\\ p^\alpha q^\beta\leq N\\ \alpha,\beta\in\mathbb{N}}} \frac{1}{p^\alpha q^\beta} - \frac{1}{N}\sum_{\substack{\left(p^\alpha,q^\beta\right)\\ p,q \text{ prime}\\ p\neq q\\ p^\alpha q^\beta\leq N\\ \alpha,\beta\in\mathbb{N}}} \delta_{p^\alpha q^\beta} + 2\sum_{\substack{p^\alpha\leq N\\ \beta\leq\alpha\\ p \text{ prime}\\ \alpha,\beta\in\mathbb{N}}} \frac{1}{p^\alpha} - \frac{2}{N}\sum_{\substack{p^\alpha\leq N\\ \beta\leq\alpha\\ p \text{ prime}\\ \alpha,\beta\in\mathbb{N}}} \delta_{p^\alpha} \\
&\leq \left(\sum_{\substack{p^\alpha\leq N\\ p \text{ prime}\\ \alpha\in\mathbb{N}}} \frac{1}{p^\alpha}\right)^2 + 2\sum_{\substack{p^\alpha\leq N\\ p \text{ prime}\\ \alpha\in\mathbb{N}}} \frac{\alpha}{p^\alpha}
\end{aligned}
$$

We have that

$$
\begin{aligned}
2\sum_{\substack{p^\alpha\leq N\\ \alpha\geq 2\\ p \text{ prime}}} \frac{\alpha}{p^\alpha} &\leq 2\sum_{\substack{x\geq 2\\ \alpha\geq 2}} \frac{\alpha}{x^\alpha} \\
&= 2\sum_{x\geq 2}\left(x\sum_{\alpha\geq 2}\frac{\alpha}{x^{\alpha+1}}\right) \\
&= 2\sum_{x\geq 2}\left(x\cdot\frac{\mathrm{d}}{\mathrm{d}x}\left(\sum_{\alpha\geq 2} -\frac{1}{x^\alpha}\right)\right) \\
&= -2\sum_{x\geq 2}\left(x\cdot\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\frac{1}{x^2}}{1-\frac{1}{x}}\right)\right) \\
&= 2\sum_{x\geq 2}\frac{x(2x-1)}{\left(x^2-x\right)^2}
\end{aligned}
$$

which, via limit comparison test with

$$\frac{1}{x^2},$$

converges.

It follows that

$$2 \sum_{\substack{p^\alpha \leq N \\ p \text{ prime} \\ \alpha \in \mathbb{N}}} \frac{\alpha}{p^\alpha} = 2 \sum_{\substack{p \leq N \\ p \text{ prime}}} \frac{1}{p} + O(1) = 2 \log\log(N) + O(1),$$

and therefore that

$$\begin{aligned} \mathbb{E}\left(Y^2\right) &\leq \left(\log\log(N) + O(1)\right)^2 + 2\log\log(N) + O(1) \\ &= \left(\log\log(N) + O(1)\right)^2 . \end{aligned}$$

This gives that

$$\begin{aligned} V(Y) &= \mathbb{E}\left(Y^2\right) - \mathbb{E}(Y)^2 \\ &\leq \left(\log\log(N) + O(1)\right)^2 - \left(\log\log(N) + O(1)\right)^2 \\ &= O\left(\log\log(N)\right). \end{aligned}$$

Applying the lemma, we see that for any $\delta > 0$,

$$\mathbb{P}\left(|Y - \log\log(N)| \geq \log\log(N)^\delta\right) \leq \frac{O(\log\log(N))}{\log\log(N)^{2\delta}},$$

this gives, for any $\epsilon > 0$

$$\mathbb{P}\left(|Y - \log\log(N)| \geq \log\log(N)^{\frac{1}{2}+\epsilon}\right) \leq \frac{O(\log\log(N))}{\log\log(N)^{1+\epsilon}} = o(1).$$

We have proven the following statement: For any $\epsilon > 0$ and $N \in \mathbb{N}$, all but $o(N)$ of the numbers $x \leq N$ satisfy

$$\log\log(N) - \log\log(N)^{\frac{1}{2}+\epsilon} \leq \Omega(x) \leq \log\log(N) + \log\log(N)^{\frac{1}{2}+\epsilon} .$$

For products ab, we clearly have that

$$\Omega(ab) = \Omega(a) + \Omega(b),$$

and so all but $o(N^2)$ of the products ab with $a, b \leq N$ satisfy

$$2\log\log(N) - 2\log\log(N)^{\frac{1}{2}+\epsilon} \leq \Omega(ab) \leq 2\log\log(N) + 2\log\log(N)^{\frac{1}{2}+\epsilon} .$$

On the contrary, all but $o(N^2)$ of the numbers $x \leq N^2$ satisfy

$$\log\log\left(N^2\right) - \log\log\left(N^2\right)^{\frac{1}{2}+\epsilon} \leq \Omega(x) \leq \log\log\left(N^2\right) + \log\log\left(N^2\right)^{\frac{1}{2}+\epsilon}$$

and, for $\epsilon < \frac{1}{2}$,

$$\lim_{N\to\infty} \frac{\log\log(N^2) + \log\log(N^2)^{\frac{1}{2}+\epsilon}}{2\log\log(N) - 2\log\log(N)^{\frac{1}{2}+\epsilon}} = \frac{1}{2}.$$

That is, the majority of products ab are only a small portion of the total numbers $x \leq N^2$. Precisely,

$$|[n]\cdot[n]| - o(N^2) = o(N^2) \implies |[n]\cdot[n]| = o(N^2).$$

□

College of Arts and Sciences, Indiana University, Bloomington, IN 47405 USA
Email address: acushma@iu.edu