GROWTH IN GROUPS I: SUM-PRODUCT

NICK GILL

1. A first look at growth

Throughout these notes A is a finite set in a ring R. For $n \in \mathbb{Z}^+$ Define

$$\underbrace{A+\cdots+A}_{n} = \{a_1+\cdots+a_n \mid a_1,\ldots,a_n \in A\};$$

$$\underbrace{A\cdots A}_{n} = \{a_1\cdots a_n \mid a_1,\ldots,a_n \in A\};$$

We call the former object a *sumset*, the latter a *product set*. The study of *growth* is the study of how the size of a sumset (resp. a product set) varies as n increases.

In particular suppose that a particular class of sets satisfies the following inequality:

$$|\underbrace{A+\cdots+A}_{n}| \geq C|A|^{1+\epsilon} \tag{1.1}$$

for all A in the class, where $n \in \mathbb{Z}^+$ while C and ϵ are positive real numbers. Then we say that a set in this class *grows under addition*. If no such inequality holds, then we would say that a set in this class *does not grow* in general. The key ingredient in (1.1) is ϵ; by growth we really mean *exponential growth*.

Obviously an analagous inequality can be studied for the product set. The study of the sum-product phenomenon is, in the first instance, a study of how, for a given A, the size of the sumset relates to the size of the product set.

2. Examples

Let us begin with an obvious situation where growth does not occur:

Lemma 2.1. *If $(A,+)$ is a group, then $|\underbrace{A+\cdots+A}_{n}| = |A|$ for all n. If $(A,\cdot)$ is a group, then $|\underbrace{A\cdots A}_{n}| = |A|$ for all n.*

Now to some more enlightening examples. Observe that if a group $(G,*)$ is abelian, then

$$|A*A| \leq \frac{1}{2}|A|(|A|+1).$$

This represents a situation of maximal growth; can it ever happen? Let $R = \mathbb{Z}$.

It is a pleasure to thank Harald Helfgott for introducing me to this beautiful area of mathematics; an area to which he has contributed a great deal.

E.g. 1 Suppose that A is a geometric progression:

$$A = \{a_0, a_0q, \dots, a_oq^{d-1}\}.$$

If we choose $a_0 = 1, q = 2$, then it is an easy exercise to establish that this is a situation of maximal growth under addition. However more significantly we observe the following facts:

- The set A does not grow under multiplication: $|A \cdot A| \leq 2|A|$.
- The set A grows under addition: there are constants $C, \epsilon > 0$ such that $|A + A| \geq C|A|^{1+\epsilon}$.

E.g. 2 Now suppose that A is an arithmetic progression:

$$A = \{a_0, a_0 + d, a_0 + 2d, \dots, a_0 + (k-1)d\}.$$

It turns out that this situation is a kind of opposite to the previous (as can be easily established by taking logarithms). In particular:

- The set A does not grow under addition: $|A + A| \leq 2|A|$.
- The set A grows under multiplication: there are constants $C, \epsilon > 0$ such that $|A \cdot A| \geq C|A|^{1+\epsilon}$.

It turns out that these situations are key: a set of integers which does not grow is related in a fundamental way to a progression. This is the content of *Freiman's theorem*, one statement of which is the following:

Theorem 1. [TV06, Thm. 5.12] *Let A be a set in $\mathbb{Z}$ such that*

$$|A + A| \leq 3|A| - 3.$$

Then A is contained in an arithmetic progression of length at most $2|A| + 2$.

Note that the *length* of an arithmetic progression is simply the number of terms it contains. The bound on the length of an arithmetic progression is crucial here (all sets lie inside an arithmetic progression of some length).

Clearly an analagous result can be stated which relates product sets to geometric progressions. We will not investigate Freiman theorems in these lectures; there is a huge volume of work in this area [TV06].

3. The sum-product principle

We work inside a commutative ring. We have seen that sets that do not grow under $+$ lie inside an arithmetic progression. Similarly sets that do not grow under $\cdot$ lie inside a geometric progression.

We make the following claim: a set cannot lie inside a (short) arithmetic progression and a (short) geometric progression simultaneously. This claim lies at the heart of what we can call the sum-product principle: *either a set grows under addition or a set grows under multiplication.*

In the context of real numbers, the principle is encoded in the following conjecture of Erdős and Szemerédi.

Conjecture. *Let A be a finite set of real numbers. For every $\epsilon > 0$ there exists a constant $C_\epsilon > 0$ such that*

$$\max(|A + A|, |A \cdot A|) \geq C_\epsilon |A|^{2-\epsilon}. \tag{3.1}$$

In other words (up to a constant) either the product set or the sumset is (just about) as big as the square of the size of A. Note that the statement

$$\max(|A+A|, |A \cdot A|) \geq C_\epsilon |A|^2$$

cannot be substituted for (3.1); see [TV06, Ex. 8.3.6].

4. Incidence theorems and sum-product

It turns out that the sum-product principle has a very interesting connection to a certain type of incidence theorem. A good starting point for this material is [TV06, Chapter 8]; we will give only the briefest treatment of these connections.

Let P be a finite set of points L and L a finite set of lines in a plane $\mathbb{F}^2$ where F is a field. Define

$$I(P,L) = |\{(p,l) \in P \times L \mid p \in l\}|.$$

In other words $I(P,L)$ is the set of incidences between P and L.

We are interested in upper bounds on the quantity $I(P,L)$. We can use the Cauchy-Schwarz inequality to quickly prove that

$$I(P,L) \leq \min(|P|^{\frac{1}{2}}|L| + |P|, |L|^{\frac{1}{2}}|P| + |L|).$$

Now restrict to the situation where $\mathbb{F} = \mathbb{R}$, the real numbers. The Szemerédi-Trotter theorem states that

Theorem 2. *Let P be a finite set of points and L a finite set of lines. Then*

$$I(P,L) \leq 4|P|^{\frac{2}{3}}|L|^{\frac{2}{3}} + 4|P| + |L|.$$

Some notes about this result:

- The proof uses facts about the *crossing number* of a graph in $\mathbb{R}^2$.
- The result can be generalized to the situation where L is a finite set of curves, where a curve is a continuous injective embedding of the compact interval $[0,1]$ into $\mathbb{R}^2$.
- Although, as we shall see, a Szemerédi-Trotter theorem exists for finite fields, it is proved in a very different way. The machinery of crossing numbers does not exist in this context. In particular this means that in the finite field context results about sets of curves (as opposed to sets of lines) are not known. Such a result would be of great interest.

We now use the Szemerédi-Trotter theorem to state Elekes' proof of the Erdos-Szemerédi conjecture for $\epsilon \geq \frac{3}{4}$.

Theorem 3. *Let A be a finite set of real numbers. There exists a constant C such that*

$$\max(|A+A|, |A \cdot A|) \geq C|A|^{\frac{5}{4}}).$$

Proof. This proof comes from [TV06, §8.3]. Let $P = \{(a,b) \mid a \in A+A, b \in A \cdot A\}$; P is a subset of the plane and

$$|P| = |A+A| \cdot |A \cdot A|.$$

Consider the set L of lines of the form $\{(x,y) \mid y = a(x-b)\}$ where $a, b \in A$. Clearly $|L| = |A|^2$; moreover, each such line constains at least $|A|$ points in P, namely the points

$(b+c, ac)$ with $c \in P$. Thus $I(P,L) \geq |A|^3$. Applying the Szemerédi-Trotter theorem we obtain, for some positive number C,

$$|A|^3 \leq C\left(|A+A||A\cdot A|^{\frac{2}{3}}(|A|^2)^{\frac{2}{3}} + |A+A||A\cdot A| + |A|^2\right),$$

and we are done. □

Solymosi has improved Elekes' argument to hold for $\epsilon \geq \frac{8}{11}$.

5. Growth under commuting actions

This section is lifted wholesale from [Hel, §3]. It describes Helfgott's extension of the idea of sum-product from the field setting to the setting of *groups acting on groups*.

The story starts with an investigation into the sum-product phenomenon over finite fields. The starting point for this investigation is the well-known theorem of Bourgain, Katz and Tao [BKT04], which was extended by Glibichuk and Konyagin [GK07] so that the final statement is as follows:

Theorem 4. *Let A be a finite set in $\mathbb{Z}/p\mathbb{Z}$ with $|A| < p^{1-\epsilon}$. Then there exists $\delta > 0$ and $C > 0$ depending only on ϵ such that*

$$\max(|A+A|, |A\cdot A|) \geq C|A|^{1+\delta}. \tag{5.1}$$

Let us make some remarks about this result before proceeding with Helfgott's generalization.

- The finite fields of order a non-prime are not covered by this theorem. In fact the theorem is not true in this setting: consider A a subfield of $\mathbb{F}_{p^a}$. Since any such theorem would have to include a list of counter-examples, investigation has mainly focussed on the much cleaner setting of $\mathbb{Z}/p\mathbb{Z}$.
- Note that constants δ and C **do not** depend on the prime p. This is the key point: the growth constants in the finite field setting do not depend on the characteristic of the field.
- The bound $|A| < p^{1-\epsilon}$ is necessary. Consider a set A consisting of all elements of the field except one. Then
$$\max(|A+A|, |A\cdot A|) = |A|+1$$
which violates (5.1) as $|A| \to \infty$. We interpret this clause in English as *"the set has room to grow"*. Thus the theorem could be stated as follows: *If A is a subset of $\mathbb{Z}/p\mathbb{Z}$ with room to grow, then A does indeed grow.*

5.1. Statement and consequences.

The generalization we wish to study is a consequence of the *pivoting argument* that appears in [GK07, Lem. 3.2 – Cor. 3.5]. The new statement is as follows.

Theorem 5. *Let G be a group and Υ an abelian group of automorphisms of G. Let $Y \subset \Upsilon$ be a non-empty set such that*

$$\text{if } y(g) = g \text{ for } y \in Y^{-1}Y,\ g \in G, \text{ then either } y = e \text{ or } g = e. \tag{5.2}$$

Then, for any non-empty $A \subset G$ and any $Y_0 \subset \Upsilon$, $A_0 \subset G$, either

$$|A \cdot Y(a_1)| \geq |A| \cdot |Y| \tag{5.3}$$

or

$$|\{y_2(a)\cdot y(y_2(a_0))\cdot y(a_2^{-1}\cdot a_1)\cdot y(y_1(a_0^{-1}))\cdot y_1(a^{-1}) : a\in A, y\in Y\}| \geq |A|\cdot|Y|. \tag{5.4}$$

for some $a_0 \in A_0$, $a_1, a_2 \in A$, $y_1, y_2 \in Y$, *or*

$$|\{y_2(a)\cdot y_0(y(a_2^{-1}a_1))\cdot y_1(a^{-1}) : a\in A, y\in Y\}| \geq |A|\cdot|Y| \tag{5.5}$$

for some $y_0 \in Y_0$, $a_1, a_2 \in A$, $y_1, y_2 \in Y$, *or*

$$|\{y_2(a)\cdot y(a_2^{-1}a_1)\cdot y_1(a^{-1}) : a\in A, y\in Y\}| > \frac{|A||Y||\mathscr{O}|}{|A||Y|+|\mathscr{O}|} \geq \frac{1}{2}\min(|A||Y|, |\mathscr{O}|), \tag{5.6}$$

where $a_1, a_2 \in A$, $y_1, y_2 \in Y$, *and* $\mathscr{O}$ *is the union of the orbits of the elements of* A *under the operations* $a \mapsto a_0 \cdot a$ *(for all* $a_0 \in A_0$*) and* $a \mapsto y_0(a)$ *(for all* $y_0 \in Y_0$*).*

It should be easy to see that the inequalities (5.3)–(5.5) must all be equalities; we phrase them as inequalities simply because we are interested in lower bounds on growth.

Note too that condition (5.2) translates into group theory terms as follows: *all the elements in* $Y^{-1}Y$ *are fixed-point free automorphisms of* G.

If we take $A_0 = A$ and $Y_0 = Y \cup Y^{-1}$, Thm. 5 acquires a particularly simple form:

Corollary 5.1. *For any group* G *and any abelian group* Υ *of automorphisms of* G. *Then, for any* $A \subset G$ *and any* $Y \subset \Upsilon$ *satisfying (5.2),*

$$|(Y_2(A))_6| > \frac{1}{2}\min(|A||Y|, |R|),$$

where $R = \langle\langle Y\rangle(\langle A\rangle)\rangle$ *is the set of all products of elements of the form* $y(a)$ *with* $a \in \langle A\rangle$ *and* $y \in \langle Y\rangle$.

Proof of Corollary 5.1. Set $A_0 = A \cup A^{-1}$, $Y_0 = Y \cup Y^{-1}$ and apply Thm. 5. It remains only to prove that the union $\mathscr{O}$ of the orbits of the elements of A under the action of $x \mapsto a \cdot x$ ($a \in A$) and $x \mapsto y(x)$ ($y \in Y$) is equal to the set R described in the statement. It is clear that $\mathscr{O} \subset R$.

To prove $R \subset \mathscr{O}$, we proceed by induction: let $R(n)$ be the set of all products of at most n elements of the form $y(a)$, $a \in A \cup A^{-1}$, $y \in \langle Y\rangle$. Assume $R(n) \subset \mathscr{O}$. (This is certainly true for $n = 0$, since the identity element $e = a \cdot a^{-1}$ is in $\mathscr{O}$.) We wish to prove $R(n+1) \subset \mathscr{O}$. Any $g \in R(n+1)$ can be written in the form $y(a) \cdot h$, where $y \in \langle Y\rangle$ and $a \in A \cup A^{-1}$. Now $y(a) \cdot h = y(a \cdot y^{-1}(h))$. Because $h \in R(n)$, and because y is a homomorphism, $y^{-1}(h)$ is also in $R(n)$. Since $R(n) \subset \mathscr{O}$, $y^{-1}(h)$ must be in $\mathscr{O}$. Then $y(a \cdot y^{-1}(h))$ must also be in $\mathscr{O}$. Thus every element of $R(n+1)$ is in $\mathscr{O}$. □

Examples. Before we prove Thm. 5, let us see two of its consequences.

(a) Let $G = \mathbb{F}_p$ (as an additive group), $\Upsilon = \mathbb{F}_p^*$ (acting on G by multiplication), $A_0 = \{1\}$, $G_0 = e$. Then condition (5.2) is easily seen to be satisfied: it just says that, in a field, if $y \cdot g = g$, then either $y = 1$ or $g = 0$. (The same is true in any ring without zero divisors.) Thus we may apply Thm. 5, and we obtain that, for any $A \subset \mathbb{F}_p$ and any $Y \subset \mathbb{F}_p^*$,

$$|Y\cdot A + Y\cdot A - Y\cdot A - Y\cdot A + Y^2 - Y^2| > \frac{1}{2}\min(|A||Y|, p). \tag{5.7}$$

(This is the result of Glibichuk and Konyagin's mentioned before; see [GK07, §3].) We may set $Y = A$, and then a few applications of the Plünnecke-Ruzsa estimates ([TV06], Cor. 6.29) suffice to derive from (5.7) the conclusion that

$$|A \cdot A + A \cdot A| \geq |A| \cdot (\frac{1}{2}\min(|A|, p/|A|))^{1/6}$$

for every subset A of $\mathbb{F}_p^*$. Now the Katz-Tao lemma [TV06, Lem. 2.53] implies that for every $A \subset \mathbb{F}_p^*$ with $|A| < p^{1-\delta}$, $\delta > 0$, we have either $|A + A| > |A|^{1+\epsilon}$ or $|A \cdot A| > |A|^{1+\epsilon}$, where $\epsilon > 0$ depends only on $\delta > 0$, and we have the sum-product theorem (Thm. 4).

(b) Let G be the group of upper-triangular matrices in $\mathrm{SL}_n(K)$ with 1's on the diagonal. Let Υ be the group of diagonal matrices, acting on G by *conjugation* (not multiplication). Let $Y \subset \Upsilon$ be a set of matrices such that the map $g \mapsto g_{ii}g_{jj}^{-1}$ (i.e., a *root of* $\mathrm{SL}_n(K)$ *relative to* Υ, in the parlance of groups of Lie type) is injective on Y for all $1 \leq i, j \leq n$ distinct.

Then (5.2) is satisfied, and so, by Corollary 5.1,

$$|(Y_2(A))_6| \geq \frac{1}{2}\min(|A||Y|, |R|),$$

where $R = \langle\langle Y\rangle(\langle A\rangle)\rangle$.

5.2. **Proof.** We have now to prove the "generalized sum-product theorem" (Thm. 5). Before we do this, let us make some philosophical comments about *pivoting*. This technique is now understood to occupy centre stage in the study of growth in groups. Typically the situation is as follows:

We have a number of groups $G_1, \dots, G_d$ which somehow interact with each other (they are subgroups of each other, or automorphism groups, or what have you). We take sets $A_1, \dots, A_d$ inside these groups, and we want to prove that *growth occurs*. In this context a *pivot* is an element ξ of the group G_1 (say) such that some function

$$\phi_\xi : A_1 \times \cdots \times A_d \to H$$

is injective. The group H may be equal to G_i for some i depending on the context.

We will need to choose the function ϕ_ξ very carefully. First of all we need ϕ_ξ to be defined using a bounded number of operations from the groups $G_1, \dots, G_d$, applied to the element ξ. Thus, in the proof below, we have $A \subset G$, $Y \subset \Gamma$, and we define

$$\phi_\xi : A \times Y \to G,\ (g, y) \mapsto g \cdot y(\xi).$$

After this, we need ϕ_ξ to have properties that we can exploit under different suppositions. Let S be the set of *non-pivots* in the group G_1.

(a) If A_1 contains a pivot ξ, then the injectivity of ϕ_ξ implies that

$$|\phi_\xi(A_1, \dots, A_d)| \geq |A_1||A_2| \cdots |A_d|.$$

In other words we have *growth*.

(b) At the other extreme, if $\langle A_1\rangle$ does not contain a pivot, i.e. $\langle A_1\rangle \cap S = \emptyset$, then we use the definition of ϕ_ξ to deduce information about S, and in turn about $A_1, \dots, A_d$. This information should be enough to prove *growth*.

(c) Finally we have the possibility that there is a pivot in $\langle A_1 \rangle$, but not in A_1. In this case there is some element $s \in S$ such that one of the group operations moves s out of S. In other words $g_i(s) \notin S$ for some $g_i \in G_i$. In the example below, there is an element $s \in S$ such that either $gs \notin S$ or $y(s) \notin S$ for some $g \in A, y \in Y$.

Now we set $\xi = g_i(s)$, and observe again that $|\phi_\xi(A_1, \dots, A_d)| \geq |A_1||A_2| \cdots |A_d|$. We need to deal somehow with the fact that s is not in our set A. This is typically the difficult case; we outline one approach in the situation below, but different situations will require different techniques.

Proof of Thm. 5. Take $\xi \in G, y_1, y_2 \in Y, y_1 \neq y_2$ and define two functions:

$$\phi_\xi : A \times Y \to G,\ (g, y) \mapsto (g \cdot y(\xi))$$
$$\delta_{y_1,y_2} : G \to G,\ g \mapsto y_2(g) \cdot (y_1(g))^{-1}$$

First of all we prove that the last two functions are well-defined; at the same time we demonstrate that they are injective.

$$\begin{aligned} & \delta_{y_1,y_2}(g_1) = \delta_{y_1,y_2}(g_2) \\ \iff & y_2(g_1) \cdot (y_1(g_1))^{-1} = y_2(g_2) \cdot (y_1(g_2))^{-1} \\ \iff & y_2(g_2^{-1} g_1) = y_2(g_2^{-1} g_2) \\ \iff & y_1^{-1} y_2(g_2^{-1} g_1) = g_2^{-1} g_2 \\ \iff & y_1 = y_2 \text{ or } g_2^{-1} g_1 \in Z \end{aligned}$$

Since we have prescribed that $y_1 \neq y_2$, δ_{y_1,y_2} we conclude that δ_{y_1,y_2} is well-defined and injective. Now observe that

$$\begin{aligned} & \phi_\xi(g_1, y_1) = \phi_\xi(g_2, y_2) \\ \iff & g_1 y_1(\xi) = g_2 y_2(\xi) \\ \iff & g_2^{-1} g_1 = y_2(\xi)(y_1(\xi))^{-1} \\ \iff & g_2^{-1} g_1 = y_2(\xi)(y_1(\xi))^{-1} \\ \iff & \xi \in \delta_{y_1,y_2}^{-1}(g_2^{-1} g_1) \end{aligned}$$

Thus ϕ_ξ is injective provided $\xi \notin \delta_{y_1,y_2}^{-1}(\{a_2^{-1} \cdot a_1\})$ for all $a_1, a_2 \in A$ and all distinct $y_1, y_2 \in Y$. If ϕ_ξ is injective then we refer to ξ as a *pivot*.

We face two cases, depending on whether or not the set

$$S = \bigcup_{\substack{a_1,a_2 \in A \\ y_1,y_2 \in Y \\ y_1 \neq y_2}} \delta_{y_1,y_2}^{-1}(a_2^{-1} \cdot a_1) \tag{5.8}$$

contains the orbit $\mathscr{O}$. The set $\mathscr{O}$ contains all "easily constructible" elements; if $\mathscr{O}$ is not contained in S, we can construct an element not in S, i.e., a valid pivot.

Case 1: $\mathscr{O} \not\subset S$. (Read: there is a pivot.)

The set $\mathscr{O}$ is the union of orbits of the elements of A under certain actions. Hence, if $\mathscr{O} \not\subset S$, we have that either $A \not\subset S$ or there is an element s of S that is taken out of S by one of the actions: that is, either $a_0 \cdot s \notin S$ for some $a_0 \in A_0$ or $y_0(s) \notin S$ for some

$y_0 \in Y_0$. Call these three cases (a), (b) and (c). In case (a), we let ξ be any element of A not in S; in case (b), we let $\xi = a_0 \cdot s$; finally, in case (c), we let $\xi = y_0(s)$.

Now we are almost done. We have a map

$$\phi_\xi : (g, y) \mapsto g \cdot y(\xi) \tag{5.9}$$

from $A \times Y \to G$. Because $\xi \notin S$, the map is injective. The map has been constructed in a finite number of steps from the elements of A and Y, since ξ was defined that way.

Let us work out the meaning and implications of this last statement case by case.

Case 1(a): $A \not\subset S$; ξ an element of A not in S. Since ϕ_ξ is injective,

$$|A \cdot Y(\xi)| \geq |A| \cdot |Y|.$$

We have proven (5.3).

Case 1(b): $\xi = a_0 \cdot s$. Since ϕ_ξ is injective,

$$|A \cdot Y(\xi)| \geq |A| \cdot |Y|. \tag{5.10}$$

Now we must do a little work: ξ is defined in terms of s, and the definition of s involves the map δ_{y_1,y_2}^{-1}, which we must now somehow remove. Because δ_{y_1,y_2} is injective, (5.10) implies

$$|\delta_{y_1,y_2}(A \cdot Y(\xi))| \geq |A| \cdot |Y|. \tag{5.11}$$

Now, for any $a \in A$, $y \in Y$,

$$\begin{aligned} \delta_{y_1,y_2}(a \cdot y(\xi)) &= y_2(a \cdot y(\xi)) \cdot (y_1(a \cdot y(\xi)))^{-1} \\ &= y_2(a) \cdot y_2(y(\xi)) \cdot (y_1(y(\xi)))^{-1} \cdot (y_1(a))^{-1} \\ &= y_2(a) \cdot y(y_2(\xi)(y_1(\xi))^{-1}) \cdot (y_1(a))^{-1}. \end{aligned} \tag{5.12}$$

(It is here that the fact that Υ is abelian is finally used.) Recall that the definition of δ_{y_1,y_2} is $\delta_{y_1,y_2}(\xi) = y_2(\xi)(y_1(\xi))^{-1}$.

Because we are in case 1(b), there are $a_0 \in A_0$, $s \in S$ such that $\xi = a_0 \cdot s$. By the definition (5.8) of S, there are $y_1, y_2 \in Y$ distinct and $a_1, a_2 \in A$ such that $\delta_{y_1,y_2}(s) = a_2^{-1} \cdot a_1$. Then

$$\begin{aligned} y(y_2(\xi) \cdot y_1(\xi)^{-1}) &= y(y_2(a_0) \cdot y_2(s) \cdot (y_1(s))^{-1} \cdot (y_1(a_0))^{-1}) \\ &= y(y_2(a_0)) \cdot y(y_2(s)(y_1(s))^{-1}) \cdot y((y_1(a_0))^{-1}) \\ &= y(y_2(a_0)) \cdot y(\delta_{y_1,y_2}(s)) \cdot y((y_1(a_0))^{-1}) \\ &= y(y_2(a_0)) \cdot y(a_2^{-1} \cdot a_1) \cdot y((y_1(a_0))^{-1}). \end{aligned} \tag{5.13}$$

Thus

$$\delta_{y_1,y_2}(a \cdot y(\xi)) = y_2(a) \cdot y(y_2(a_0)) \cdot y(a_2^{-1} \cdot a_1) \cdot y((y_1(a_0))^{-1}) \cdot (y_1(a))^{-1}.$$

We conclude that

$$|\{y_2(a) \cdot y(y_2(a_0)) \cdot y(a_2^{-1} \cdot a_1) \cdot y(y_1(a_0^{-1})) \cdot y_1(a^{-1}) : a \in A, y \in Y\}| \geq |A| \cdot |Y|.$$

That is, the conclusion (5.4) is true.

Case 1(c): $\xi = y_0(s)$. We start as in case 1(b): (5.11) and (5.12) still hold. By the definition (5.8) of S, there are $y_1, y_2 \in Y$ distinct and $a_1, a_2 \in A$ such that $\delta_{y_1,y_2}(s) =$

$a_2^{-1} \cdot a_1$. Now, because we are in case 1(c) and not in case 1(b), we have $\xi = y_0(s)$ instead of $\xi = a_0 \cdot s$. We replace (5.13) by the following calculation:

$$\begin{aligned} y(y_2(\xi) \cdot y_1(\xi)^{-1}) &= y(y_2(y_0(s)) \cdot (y_1(y_0(s)))^{-1}) = y(y_2(y_0(s))) \cdot y(y_1(y_0(s^{-1}))) \\ &= y_0(y(y_2(s))) \cdot y_0(y(y_1(s^{-1}))) = y_0(y(y_2(s) \cdot y_1(s^{-1}))) \\ &= y_0(y(y_2(s) \cdot (y_1(s))^{-1})) = y_0(y(\delta_{y_1,y_2}(s))) = y_0(y(a_2^{-1}a_1)). \end{aligned}$$

(It is here that the fact that Υ is abelian is used for the second time.) Thus

$$\delta_{y_1,y_2}(a \cdot y(\xi)) = y_2(a) \cdot y_0(y(a_2^{-1}a_1)) \cdot (y_1(a))^{-1}.$$

We conclude that

$$|\{y_2(a) \cdot y_0(y(a_2^{-1}a_1)) \cdot y_1(a^{-1}) : a \in A, y \in Y\}| \geq |A| \cdot |Y|.$$

In other words, (5.5) holds.

Case 2: $\mathscr{O} \subset S$. (Read: there is no pivot.)

Then S must be rather large. From the definition (5.8), it becomes clear that either Y or A must be rather large. It is then no surprise that some crude techniques appropriate for large sets shall be sufficient for our task.

Since δ_{y_1,y_2} is injective for $y_1 \neq y_2$, the sets

$$R_\xi = \{(a_1, a_2, y_1, y_2) \in A \times A \times Y \times Y : y_1 \neq y_2,\ a_1 \cdot y_1(\xi) = a_2 \cdot y_2(\xi)\}$$

are disjoint as ξ ranges in G. Choose $\xi_0 \in S$ such that $|R_{\xi_0}|$ is minimal. Then

$$|R_{\xi_0}| \leq \frac{|A|^2|Y|(|Y|-1)}{|S|} < \frac{|A|^2|Y|^2}{|S|} \leq \frac{|A|^2|Y|^2}{|\mathscr{O}|}$$

and so

$$|\{(a_1, a_2, y_1, y_2) \in A \times A \times Y \times Y : a_1 \cdot y_1(\xi_0) = a_2 \cdot y_2(\xi_0)\}| < \frac{|A|^2|Y|^2}{|\mathscr{O}|} + |A| \cdot |Y|.$$

Hence

$$|A \cdot Y(\xi_0)| > \frac{|A|^2|Y|^2}{\frac{|A|^2|Y|^2}{|\mathscr{O}|} + |A| \cdot |Y|} = \frac{|A||Y||\mathscr{O}|}{|A||Y| + |\mathscr{O}|}. \tag{5.14}$$

As before, we must somehow remove δ_{y_1,y_2}^{-1} from ξ_0. By the injectivity of δ_{y_1,y_2}, (5.14) implies

$$|\delta_{y_1,y_2}(A \cdot Y(\xi_0))| > \frac{|A||Y||\mathscr{O}|}{|A||Y| + |\mathscr{O}|}.$$

Equation (5.12) is still valid. Since $\xi \in S$, we know that $\delta_{y_1,y_2}(\xi_0) = a_2^{-1}a_1$ for some $a_1, a_2 \in A$, $y_1, y_2 \in Y$ distinct. Thus, for $a \in A$, $y \in Y$,

$$\begin{aligned} \delta_{y_1,y_2}(a \cdot y(\xi_0)) &= y_2(a \cdot y(\xi_0)) \cdot (y_1(a \cdot y(\xi_0)))^{-1} \\ &= y_2(a) \cdot y_2(y(\xi_0)) \cdot (y_1(y(\xi_0)))^{-1} \cdot (y_1(a))^{-1} \\ &= y_2(a) \cdot y(y_2(\xi_0) \cdot (y_1(\xi_0))^{-1}) \cdot (y_1(a))^{-1} \\ &= y_2(a) \cdot y(\delta_{y_1,y_2}(\xi_0)) \cdot (y_1(a))^{-1} \\ &= y_2(a) \cdot y(a_2^{-1}a_1) \cdot (y_1(a))^{-1}. \end{aligned}$$

(It is here that the fact that Υ is abelian is used for the third and last time.) Hence

$$|\{y_2(a)\cdot y(a_2^{-1}a_1)\cdot y_1(a^{-1}) : a\in A, y\in Y\}| > \frac{|A||Y||\mathscr{O}|}{|A||Y|+|\mathscr{O}|}.$$

The inequality $\frac{ab}{a+b} \geq \frac{1}{2}\min(a,b)$ is easy and true for all positive a, b. Hence we have proven (5.6). □

6. Further work

Let me finish by noting some possible avenues of further research into the sum-product phenomenon.

6.1. **Orbit structure.** The condition (5.2) in Thm. 5 is quite strong. Clearly similar results will hold under somewhat weaker conditions; for instance, a stronger version of Thm. 5 appears in [GH10]. There is no reason to think, though, that this stronger version is definitive; there's plenty of scope for investigation here!

6.2. **Incidence theorems over finite fields.** A Szemerédi-Trotter theorem exists over finite fields [BKT04]:

Theorem 6. *Let P and L be points and lines in the projective plane over $\mathbb{F}_p$. Suppose that $|P|, |L| \leq N = p^\alpha$ for some $0<\alpha<2$. Then we have*

$$\{(p,l)\in P\times L : p\in l\} \leq CN^{\frac{3}{2}\epsilon}$$

for some constants C, ϵ depending only on α.

The key point here is the presence of ϵ; the bound $CN^{\frac{3}{2}}$ can be derived easily from Cauchy-Schwarz (as described earlier). Some work has been done giving numerical values for ϵ given particular α (most notably $\alpha=\frac{1}{2}$) but there is plenty still to be done.

There is also plenty to be done in developing incidence theorems for curves; at this stage only specific instances are known. For instance it is shown in [Bou05] that the function

$$f : \mathbb{F}_p\times\mathbb{F}_p \to \mathbb{F}_p,\ (x,y)\mapsto x(x+y)$$

satisfies $|f(A,B)| > p^\beta$ for some $\beta>\alpha$ (and β depends only on α) whenever $A, B\subset\mathbb{F}_p$ and $|A|\sim|B|\sim p^\alpha, 0<\alpha<1$. (Here $\sim$ means that two quantities are the *same up to a constant.*)

6.3. **Sum-product from incidence theorem.** Bourgain, Katz and Tao prove the Szemerédi-Trotter theorem over a finite field (Thm. 6) as a corollary to a sum-product theorem over finite fields [BKT04]. It is possible that a more natural line of argument would be the reverse of this.

The idea is to "construct" a finite field from a projective plane, rather than the other (more usual) way.

The usual approach to working in projective planes involves moving **from algebra to geometry**. Start with an arithmetic system (e.g. a field) and coordinatize the points of a projective plane using elements from this system, and an extra symbol ∞. The coordinatization works by assigning ordered pairs (a,b) from your arithmetic system to an affine plane within the projective plane; points on the extra line are then assigned

symbols (a) with a from the arithmetic system, with one point being labelled (∞). We want this coordinate system to have some basic properties; in particular, linear equations in these coordinates should correspond to lines in the projective plane. This is exactly what happens in the standard coordinatizations of a desarguesian projective plane $PG_2(q)$ by a field, $\mathbb{F}_q$, with q elements.

From geometry to algebra. Alternatively one may start with a projective plane of order n. One can then define an arithmetic system on a set N of size n using properties of the plane as follows: One takes a line ℓ and labels all but one point (the *point at infinity*) with elements from N. One then defines a sum operation and a multiplication operation by way of various incidence configurations within the plane. In this way one effectively generates a coordinatization for the plane (and an arithmetic system) using intrinsic properties of the plane. All this is explained in [Sti05]; in particular if one's plane is Desarguesian, the above process yields an arithmetic system that is actually a field.

Let us define an intrinsic notion of growth in the projective plane. Suppose we are given a set of points $\mathbf{P} = \mathbf{P}_0(\mathbf{P})$. Now define

- $\mathbf{L}_i(\mathbf{P}), i = 0, 1, 2, \dots$ to be the set of lines incident with at least two points of $\mathbf{P}_{i-1}(\mathbf{P})$ (say $\mathbf{L}_i(\mathbf{P})$ is the set of points *defined by* $\mathbf{P}_{i-1}(\mathbf{P})$);
- $\mathbf{P}_i(\mathbf{P}), i = 1, 2, \dots$ to be the set of points incident with at least two lines of $\mathbf{L}_{i-1}(\mathbf{P})$ (say $\mathbf{P}_i(\mathbf{P})$ is the set of points *defined by* $\mathbf{L}_{i-1}(\mathbf{P})$).

Now it is easy enough to prove that, except in certain circumstances, intrinsic growth does occur:

Proposition 6.1. *Let* $\mathbf{P}$ *be a set of points in a projective plane. Then one of the following statements hold.*

(a) $|\mathbf{P}_3(\mathbf{P})| \geq \frac{1}{4}|\mathbf{P}|^2$.
(b) *more than* $\frac{1}{2}|\mathbf{P}|$ *points of* $\mathbf{P}$ *lie on a line.*
(c) $\mathbf{P}_2 = \mathbf{P}_1$ *or* $\mathbf{P}_2 = \mathbf{P}_1 \cup \{\wp\}$ *for some point* $\wp$ *in the plane.*

It would be nice if we could somehow show that the growth described in (a) actually occured earlier. So, for instance, rather than a bound on $|\mathbf{P}_3(\mathbf{P})|$, we could give a bound on $|\mathbf{L}_0(\mathbf{P})|$.

The idea of *intrinsic growth* result can be connected to a Szemerédi-Trotter result as follows:

Lemma 6.2. *Let* $\mathbf{P}$ *be a set of points in a projective plane such that any set of lines,* $\mathbf{L}$ *with* $|\mathbf{L}| = |\mathbf{P}| = N$*, will satisfy* $I(\mathbf{P}, \mathbf{L}) \leq N^{\frac{3}{2}-3\epsilon}$ *where* $0 < \epsilon < \frac{1}{4}$*. Provided* $N^\epsilon > 2$*, we conclude that* $|\mathbf{L}_0(\mathbf{P})| \geq N^{1+\epsilon}$.

In other words a Szemerédi-Trotter theorem (like Thm. 6) implies an intrinsic growth result. We know already that Szemerédi-Trotter theorems are connected to sum-product theorems, so it would seem reasonable to hope that sum-product theorems may be provable using the notion of intrinsic growth.

A fuller description of this possible approach (along with proofs of Prop. 6.1 and Lem. 6.2) is available in an unpublished preprint [GHR08].

References

[BKT04] J. Bourgain, N. Katz, and T. Tao, *A sum-product estimate in finite fields, and applications*, Geom. Funct. Anal. **14** (2004), no. 1, 27–57.

[Bou05] J. Bourgain, *More on the sum-product phenomenon in prime fields and its applications*, Int. J. Number Theory **1** (2005), no. 1, 1–32.

[GH10] Nick Gill and Harald Helfgott, *Growth in solvable subgroups of $GL_r(\mathbb{Z}/p\mathbb{Z})$*, Preprint available at `http://arxiv.org/abs/1008.5264`, 20010.

[GHR08] Nick Gill, Harald Helfgott, and Misha Rudnev, *A geometric approach to arithmetic combinatorics*, Draft preprint, 2008.

[GK07] A. A. Glibichuk and S. V. Konyagin, *Additive properties of product sets in fields of prime order*, Additive combinatorics, CRM Proc. Lecture Notes, vol. 43, Amer. Math. Soc., Providence, RI, 2007, pp. 279–286.

[Hel] H.A. Helfgott, *Growth and generation in $SL_3(\mathbb{Z}/p\mathbb{Z})$*, J. Eur. Math. Soc. (JEMS), To appear.

[Sti05] John Stillwell, *The four pillars of geometry*, Undergraduate Texts in Mathematics, Springer, New York, 2005.

[TV06] Terence Tao and Van Vu, *Additive combinatorics*, Cambridge Studies in Advanced Mathematics, vol. 105, Cambridge University Press, Cambridge, 2006.