SUMS versus PRODUCTS in Number Theory, Algebra and Erdős Geometry

György Elekes

July 19, 2002

Abstract

In this paper we review some structure theorems which characterize certain "nearly optimal" configurations, mostly in Erdős–type Geometry problems and also in Algebra. Some of the latter will even generalize Freiman's theorem on sumsets to composition sets of linear functions. As far as we know, these are the first Freiman–type results for non–Abelian groups.

One common feature of the problems we study is that they can usually be re–formulated in terms of $n \times n$ or $n \times n \times n$ Cartesian products in the two or three dimensional Euclidean plane/space and certain straight lines, curves or surfaces which pass through many points of the Cartesian product (at least cn for curves and cn^2 for surfaces, respectively, for a fixed $c > 0$ and large n). That is why we shall also review certain bounds on incidences like those of Szemerédi–Trotter and Pach–Sharir.

On the influence of Paul Erdős

Writing an article for a volume dedicated to the memory of Erdős and reviewing results which are closely related to his beautiful conjectures, problems and theorems, certainly requires no further explanation of his influence on the subject. However, the impact of his personality on the author himself may be worth describing in detail. Actually, I would like to mention two of my most remarkable experiences with him.

In 1971 I was a student graduating from Eötvös University, fortunate enough to attend a Colloquium Lecture of Erdős on, among others, infinite combinatorics. One of the problems really caught my phantasy. That year E.P. came to Budapest several times and I — previously unknown to him — went to the Institute and asked several questions on the topic. And Prof. Erdős, one of the most famous mathematicians of the world for decades, answered the long series of sometimes boring questions of a (that time quite illiterate) beginner with incredible patience, explaining details in his usual, clear and elegant ways. Last but not least, a few months later he paid \$25 for the problem *he* helped me solve.

The other — for me very instructive — story which also sheds some light on his way of mathematical thinking, happened in the late eighties, at a Geometry workshop in Visegrád, Hungary. An informal group of participants were discussing, among others, the maximal number of squares or equilateral triangles which can be found in a set of n points of the plane. We soon came to the conclusion that (portions of) a square or triangular lattice, respectively, must give the optimal configurations. When I suggested considering regular pentagons, Erdős immediately grasped the question and we started working on it that night. After some (unsuccessful) attempts of finding upper bounds, he turned by 180 degrees: "OK; we'd better define pentagonal lattices" he said. I answered what (almost) everyone else would: "But Uncle Paul, you know, there are no pentagonal lattices". "Right" he replied; "that is why we should define one". The reader can find the end of the story in Example 3.16 and Theorem 3.17 below.

Introduction

Everywhere in mathematics, whenever we determine the maximum or minimum of a quantity, it is interesting to describe those configurations for which this extremum is attained. Sometimes even the stability of the extremal structures is interesting i.e., if we are not far from the best possible value, can the structure change very much or must it remain close to the optimal one? Such questions, of course, tend to be harder than those of the first type.

A good example is the characterization of small sumsets. If we have an n element set of real numbers then there must exist at least $2n-1$ distinct two–term sums; moreover, it is not difficult to show that $2n-1$ is only attained for arithmetic progressions. However, it was the main result of a 100-page paper of Freiman that if we have at most, say, $1000n$ distinct two–term sums then the set is contained in a (not too large) "generalized" arithmetic progression (see Theorem 1.3 below).

In this paper we review some structure theorems which characterize such "nearly optimal" configurations, mostly in Erdős–type Geometry problems and also in Algebra. Some of them, e.g., Theorems 3.4 and 3.8, will even generalize Freiman's aforementioned result to linear functions. As far as we know, these are the first Freiman–type theorems for non–Abelian groups.

One common feature of the problems we study is that they can usually be re–formulated in terms of $n \times n$ or $n \times n \times n$ Cartesian products in the two or three dimensional Euclidean plane/space and certain straight lines, curves or surfaces which pass through many points of the Cartesian product (at least cn for curves and cn^2 for surfaces, respectively, for a fixed $c > 0$ and large n). That is why we shall heavily rely upon certain bounds on incidences (to be described in Section 1.2).

The structure of this paper is like a tree in an orchard: it has roots, a trunk, and several branches with fruits. For a tourist (i.e., reader), probably the latter are most interesting so we start listing them first. Beyond the global geometric flavor all over, one branch (Section 3.1) grows fruits (i.e., results) with some algebraic spices: structure theorems on affine transforms of the real line and their compositions can be found there, including the already mentioned generalizations of Freiman's Theorem. The next branch (Section 3.2) leans against the former one and uses it to characterize those point sets which contain many subsets similar or homothetic to each other. Sets which determine few distinct directions or distances are the topic of Sections 3.3 and 3.4, respectively. Finally, certain "circle grids" are described in Section 3.5.

However, all these fruits and branches are supported by the trunk of the tree; i.e., the results which characterize Cartesian products and curves or surfaces with many points. Our structure theorems typically state that such configurations must contain many curves or surfaces from a family described by as few as *one* parameter — though, say, straight lines form a two-parameter family, hyperbolas $y = a + b/(x - c)$ a three-parameter one, etc. An example (which, in fact, lies in the core of the whole survey) states that *if each of $c_1 n$ straight lines contains $c_2 n$ or more points of an $n \times n$ Cartesian product then at least $c_3 n$ of them must be parallel or concurrent* (i.e., pass through a common point, see Theorem 2.13).

The trunk and a young shoot — Section 2.1 on the "hybrid problem", on which no further branches and fruits (i.e., applications) have been found so far — grow directly from the roots formed by Sections 1.1 on sumsets and 1.2 on the theory of incidences.

A common feature of the results shown in the "middle" part is that most proofs use the paradigm "if you have many functions but not enough, compose them to get more". Sometimes these compositions will be all distinct and then the incidence bounds work well. However, sometimes they may coincide and in such cases a detailed analysis of the possible coincidences becomes necessary, after which certain "commutator pairs" will be the main tools we use.

Last but not least, all the results presented in this survey suggest a general philosophy: in a wide variety of problems it is impossible to mix up arithmetic and geometric progressions in order to construct nearly optimal configurations. In other words, addition and multiplication cannot tolerate each other. If a sumset is small then the product set is large (Theorems 2.1 and 2.3); if the composition set of some linear functions is small then they are graphs of *either* parallel *or* concurrent lines (Theorem 3.4); if two collinear point sets determine few directions then they must be close to *either* an arithmetic *or* a geometric progression (a special case of Theorem 3.29); etc.

I would like to express my gratitude to Vera Sós whose comments have increased the possible values of this survey exponentially.

Chapter 1

The roots

This first part presents the two groups of theorems (or, rather, two theories) from which all further results developed.

Sumsets are the topic of Section 1.1. The theorem of Freiman and its "statistical" variants are mentioned here, together with a common generalization. These will be used later on in Sections 3.1 through 3.5.

Section 1.2 studies bounds on the number of incidences between planar point sets and families of straight lines or curves. Also, some classical applications e.g., Beck's "Two extremities" Theorem 1.19 and the "weak" Dirac–Motzkin problem (Theorem 1.20) will be mentioned there.

1.1 Small sumsets

For two (usually finite) subsets X, Y of the real or complex numbers we put $X + Y \stackrel{\text{def}}{=} \{x + y\ ;\ x \in X,\ y \in Y\}$ and call it a *sumset.* In his celebrated works [Fre66, Fre73], Freiman studied the structure of "small" sumsets, i.e., those for which $|X + X| \leq C|X|$ for some fixed positive C.

How can one make $X + X$ this small? One example is the case of an arithmetic progression when $|X+X| = 2|X|-1$. Also, some "generalized" arithmetic progressions will do the trick:

Definition 1.1 Let d and $n_1, \ldots, n_d$ be positive integers and $\Delta_1, \Delta_2, \ldots, \Delta_d$ arbitrary real or complex numbers. A set G is a *generalized arithmetic progression* ("arithmetic GP" for short) of dimension d and size $n = n_1 \cdot n_2 \cdot \ldots \cdot n_d$ if

$$G = \left\{ \sum_{i=1}^{d} k_i \cdot \Delta_i\ ;\ 0 \leq k_i < n_i \text{ for } i = 1 \ldots d \right\},$$

and these elements are all distinct.

In what follows, $\mathcal{G}^{d,n}$ will denote the class of arithmetic GP's of dimension *not exceeding* d and size *at most* n.

Remark 1.2 An arithmetic GP does, indeed, have few distinct sums; it is not difficult to show that $|G+G| \leq 2^d|G|$.

The following result says that there are essentially no other examples of small sumsets.

Theorem 1.3 *(Freiman [Fre66, Fre73], Ruzsa [Ruz92, Ruz94], Bilu [Bil99]) If* $|X|, |Y| \geq n$ *and* $|X+Y| \leq Cn$ *then* $X \cup Y$ *is contained in an arithmetic GP* $G \in \mathcal{G}^{d^*, C^*n}$*, where* $d^* = d^*(C)$ *and* $C^* = C^*(C)$ *do not depend on* n*.*

Remark 1.4 Theorem 1.3 remains valid in any torsion–free Abelian group.

A "statistical" version of this result was found by Balog and Szemerédi in [BS94]. Let X be a subset of the reals or the complex numbers and E a set of unordered pairs of X (i.e., the edge set of an undirected graph $H(X,E)$ on vertex set X). Put

$$X +_E X \stackrel{\text{def}}{=} \{x' + x'' \ ; \ (x', x'') \in E\}.$$

Also, $X +_E Y$ can be defined similarly for bipartite graphs $H(X,Y,E)$ on vertex sets X, Y.

Theorem 1.5 *(Balog–Szemerédi [BS94]) If* $|E| \geq \alpha|X|^2$ *and* $|X +_E X| \leq C|X|$ *then some* $\alpha^*|X|$ *elements of* X *are contained in an arithmetic GP* $G \in \mathcal{G}^{d^*, C^*|X|}$*, where* $d^* = d^*(C,\alpha)$*,* $C^* = C^*(C,\alpha)$ *and* $\alpha^* = \alpha^*(C,\alpha)$ *do not depend on* $|X|$*.*

Their proof is based upon Szemerédi's famous "Regularity Lemma" [Sze78]. Theorem 1.5 was used for settling a conjecture of Erdős on three-term arithmetic progressions (see Theorem 1.7 below). First we mention a simple consequence of Theorem 1.5 [BS94].

Corollary 1.6 *If a set of* n *real numbers contains* cn^2 *three-term arithmetic progressions then it contains some* c^*n *elements of an arithmetic GP* $G \in \mathcal{G}^{d^*, C^*n}$*.*

Theorem 1.7 *(Erdős Conjecture, Balog–Szemerédi [BS94]) If a set* $\mathcal{A}$ *of* n *real numbers contains* cn^2 *three-term arithmetic progressions then it also contains a* k*–term arithmetic progression, provided that* $n > n_0(c,k)$*.*

Outline of the proof: Such a set, by the previous Corollary, contains many elements of an arithmetic GP of dimension $d \leq d^*$. The largest of these dimensions, say n_1, must be at least $\sqrt[d]{c^*n}$, which can be arbitrarily large if n is large. Moreover, one of the "fibers" in this direction must intersect $\mathcal{A}$ in at least c^*n_1 elements. By a famous result of Szemerédi [Sze75], a k–term arithmetic progression can be found. ∎

Balog–Szemerédi's Theorem 1.5 says nothing about the number of edges in E spanned by $G \cap X$. Such a slightly stronger extension was proven by Laczkovich and Ruzsa [LR96].

Theorem 1.8 *(Laczkovich–Ruzsa [LR96]) In Theorem 1.5 above, $G \cap X$ can be required to span at least $\beta|X|^2$ edges of E, for some $\beta = \beta(C, \alpha)$.*

Their result was applied to deduce some consequences concerning planar point sets which contain many similar copies of a given pattern, see Theorem 3.19 below.

None of the above two statistical results can assert the *total coverability* of X by structures like arithmetic GP's, since their assumption allows lots of elements of X not related to any others (not occurring in any sums) at all.

Under the slightly stronger "uniform statistical" assumption that each element occurs in "many" pairs (which is still weaker than that of Theorem 1.3) X can already be covered with a bounded number of arithmetic GP's.

Theorem 1.9 *(Elekes–Ruzsa [ER03b]) Let $X \subset \mathbb{C}$ be finite, $\alpha > 0$ fixed and $H(X, E)$ a graph with all degrees at least $\alpha|X|$. Let, moreover, an arbitrary positive ε be given. Now if $|X +_E X| \le C|X|$ then X can be decomposed into disjoint parts $X_1, X_2, \dots, X_k$ with the following properties.*

(1) each X_i is contained in an arithmetic GP $G_i \in \mathcal{G}^{d^*, C^*|X|}$ (which may not be disjoint);

(2) each X_i spans at least $\gamma|X|^2$ edges of E (which implies that $k \le 1/\gamma$);

(3) there are at most $\varepsilon|X|^2$ "leftover" edges outside the X_i, i.e.,

$$\sum_{1 \le i < j \le k} |E \cap (X_i \times X_j)| \le \varepsilon |X|^2,$$

where $d^* = d^*(\alpha, C, \varepsilon)$ and $C^* = C^*(\alpha, C, \varepsilon)$ do not depend on $|X|$; nor does $\gamma = \gamma(\alpha, C)$ depend on ε or on $|X|$.

This result does, indeed, generalize all the previous Theorems on sumsets. Of those, the statistical versions are direct applications while to prove Theorem 1.3, one can just take $\varepsilon = \gamma(1/2, C)/2$; the resulting decomposition cannot have two or more parts X_i.

1.2 Incidences

In this section we review certain facts concerning point-line and/or point-curve incidences in the plane.

We must emphasize that all the forthcoming results are specific to the Euclidean plane $\mathbb{R}^2$ (or at most the complex plane $\mathbb{C}^2$) i.e., they do not hold in finite planes. Historically, the first combinatorial distinction between Euclidean and finite projective planes was a theorem of Gallai (see [EP95]) who — solving an old problem of Sylvester — proved that if n points are not all collinear then there is a straight line which contains *exactly two* of them.

Beyond straight lines we shall also consider certain families of curves, as well. They will be defined in full generality in Definition 1.23. At the moment for our purposes it suffices to recall that a plane curve is called algebraic if it is defined as the zero set $\{(x,y) \in \mathbb{C}^2 \ ; \ F(x,y)=0\}$ of a bivariate polynomial $F \in \mathbb{C}[x,y]$. Also, the degree of such a curve is the (total) degree of its defining polynomial F.

1.2.1 The general problem of nearly optimal incidence structures

Problem 1.10 *What is the structure of those configurations of* n *planar points and* g *curves of degree at most* r *(i.e., they can be described by polynomial equations of degree not exceeding* r*) in which the number of incidences is maximal or asymptotically maximal?*

It is not surprising that nothing is known — except for some trivial cases — in this generality. Even the case of straight lines is unsolved.

In Section 2.2 we shall focus on the characterization of those configurations which consist of $n \times n$ Cartesian products $A \times B$ with $|A| = |B| = n$, $A, B \subset \mathbb{R}$ or $\mathbb{C}$, and curves which contain a positive proportion of the maximal possible number of elements of the point set (i.e., at least cn points for a fixed $c > 0$).

Definition 1.11 Given a positive integer k, a curve γ is *k–rich* in a point set $\mathcal{P} \subset \mathbb{R}^2$ if $|\gamma \cap \mathcal{P}| \geq k$.

Using this notion, we can say that the main results of Section 2.2 describe the structure of families of straight lines, hyperbolae, graphs of polynomials or of algebraic curves of degree r, which are cn–rich in $n \times n$ Cartesian products. (In case of straight lines, for every Cartesian product there are n vertical and n horizontal lines with n points each; that is why we shall only be interested in "non–trivial" lines which are neither vertical nor horizontal.)

Instead of asking for structures, only determining the maximum number of incidences (or, perhaps, just its order magnitude) is a much better understood problem. Such bounds are the topic of the next section.

1.2.2 Bounds on the number of incidences

The following estimate was conjectured by Erdős and proven by Szemerédi–Trotter [ST83] for real points and lines. Recently Csaba Tóth has extended it to $\mathbb{C}^2$, i.e., to complex point-line configurations.

Theorem 1.12 *(Szemerédi–Trotter [ST83], Cs. Tóth [Tót03]) The maximum number* $I(n,m)$ *of incidences between* n *points and* m *straight lines of the real* *or complex plane satisfies*

$$I(n,m) = O(n^{2/3}m^{2/3} + n + m).$$

Remark 1.13 a. One might expect that this result also generalizes to set-systems where any two sets share at most one point in common. However, as a finite projective plane of order q shows, $n = q^2 + q + 1$ points and $m = q^2 + q + 1$ lines (subsets) can produce $I = (q^2 + q + 1)(q + 1) \approx q^3$ incidences which exceeds the order of magnitude of $n^{2/3}m^{2/3} + n + m \approx q^{8/3}$.

b. Thus the above Theorem also depends on the *topology* of the underlying planes. Indeed, all the known proofs of the real version rely heavily upon the fact that a straight line cuts $\mathbb{R}^2$ into two parts. However, the topology of $\mathbb{C}^2$ is much different: the four (Euclidean) dimensional space is not cut into two by a two (Euclidean) dimensional affine subspace. That is why it is not at all obvious to extend any of the proofs of the real version to the complex case.

c. Even the incidence structures of the real and the complex plane are different. E.g., Gallai's Theorem (mentioned at the beginning of Section 1.2) does not hold in $\mathbb{C}^2$: nine appropriate points of an elliptic cubic have the property that a straight line through any two of them will always pass through a third one, as well.

A simple and elegant proof of the original real Szemerédi–Trotter result was found by L. Székely [Szé97]. Cs. Tóth's complex proof follows the "old" way.

This bound is asymptotically best possible. Before presenting an example, we mention that the result also has another version which is sometimes — actually quite frequently — more convenient to use. Moreover, it is even equivalent to the original statement (but we do not need this fact).

Corollary 1.14 *(Szemerédi–Trotter [ST83] and Cs. Tóth [Tót03]) There is an absolute constant (denoted by C_{SzTr} in the sequel) such that, given a set of N points in the real or complex plane, the number of k–rich lines is at most*

$$C_{SzTr} \cdot \max\left\{\frac{N^2}{k^3}, \frac{N}{k}\right\}. \blacksquare$$

Remark 1.15 Of the two terms in the "max{ }", the first one dominates if $k \leq \sqrt{N}$ and the second one otherwise.

Also this result is asymptotically best possible. For $k \geq \sqrt{N/2}$, just use N/k lines with k points on each of them. Otherwise the following example works. (The original construction of Erdős used some Number Theory to show that a $\sqrt{N} \times \sqrt{N}$ lattice and its "richest" lines would also do the trick.)

1.16 Example Let $\mathcal{P} = \{1, 2, \dots, k\} \times \{1, 2, \dots, \lfloor N/k \rfloor\}$ and consider the lines $y = mx + b$, where $m = 1, 2, \dots, \lfloor N/(2k^2) \rfloor$ and $b = 1, 2, \dots, \lfloor N/(2k) \rfloor$.

There are $\lfloor N/(2k^2)\rfloor \cdot \lfloor N/(2k)\rfloor \geq (1/16)N^2/k^3$ of them (since $\lfloor z\rfloor \geq z/2$ if $z \geq 1$). Moreover, substituting $x = 1, 2, \ldots, k$ into any of these linear equations, we get one of the values $1, 2, \ldots, \lfloor N/k\rfloor$ by

$$y = mx + b \leq \frac{N}{2k^2} \cdot k + \frac{N}{2k} = \frac{N}{k}.$$

Therefore, all the lines contain at least k points of $\mathcal{P}$.

The following observation asserts that "medium rich" lines cannot contain very many pairs. More precisely, let $\mathcal{P} = \{P_1, \ldots, P_N\}$ be a set of N points in the plane and $b \leq \sqrt{N}$ an arbitrary number. Call a straight line L "medium rich" (with respect to b) if $b \leq |L \cap \mathcal{P}| \leq \sqrt{N}$. We shall bound the number of pairs P_iP_j located on these lines.

Corollary 1.17 *The number of such pairs is at most CN^2/b, where C is an absolute constant.*

Proof: Let $l_1, l_2, \ldots, l_t$ be the lines which contain at least b and at most $\sqrt{N}$ points of $\mathcal{P}$ and put $k_i = |l_i \cap \mathcal{P}| \leq \sqrt{N}$ for $i = 1 \ldots t$. Then, ("$\leq$" in the middle is implied by Corollary 1.14),

$$\begin{aligned}
\#\text{ of pairs } &\leq \sum_{i=1}^{t} \binom{k_i}{2} < \sum_{i=1}^{t} k_i^2 = \sum_{j=0}^{\log\lceil\sqrt{N}/b\rceil} \sum_{b\cdot 2^j \leq k_i < b \cdot 2^{j+1}} k_i^2 \leq \\
&\leq \sum_{j=0}^{\log\lceil\sqrt{N}/b\rceil} \left(b \cdot 2^{j+1}\right)^2 \cdot C_{SzTr} \cdot \frac{N^2}{\left(b \cdot 2^j\right)^3} = \\
&= 4 \cdot C_{SzTr} \cdot \frac{N^2}{b} \sum_{j=0}^{\log\lceil\sqrt{N}/b\rceil} 2^{-j} < 8 \cdot C_{SzTr} \cdot \frac{N^2}{b} \blacksquare
\end{aligned}$$

A similar argument — using the "second case" of Corollary 1.14 and the Remark after it — proves the following.

Corollary 1.18 *There is an absolute constant c with the following properties. Let $\mathcal{P} = \{P_1, \ldots, P_N\}$ be a set of N points in the plane and $B \geq \sqrt{N}$ an arbitrary number. Then the number of pairs P_iP_j located on those straight lines which contain at least $\sqrt{N}$ and at most B points of $\mathcal{P}$ cannot exceed cNB.* $\blacksquare$

Another interesting consequence is a result of Beck. Originally, it asserted that any set of n points has at least one of the following extreme properties: "many" points are collinear or they determine "many" distinct straight lines.

Here we show a "statistical" version of Beck's Theorem. The notion of "statistical" results originates from Balog and Szemerédi [BS94]. They used it for statements in which not all pairs of n objects are considered, just a positive proportion of them, i.e., some cn^2 pairs. These can be represented e.g., by a "dense" graph on the set of objects as its vertex set.

Theorem 1.19 ("Two Extremities" Theorem) *(Beck [Bec83]) Let $\mathcal{P}$ be a set of n points in the plane and $H(\mathcal{P}, E)$ a graph with vertex set $\mathcal{P}$ and a set E of at least cn^2 edges. Draw a straight line through each pair which is connected by an edge of E. Then at least one of the following two must hold (with a constant $c' = c'(c)$, independent of n):*

(i) some $c'n^2$ pairs determine all distinct lines;

(ii) or some $c'n^2$ pairs determine the very same line (and then, of course, at least $c''n$ points must be collinear, e.g., for $c'' = \sqrt{2c'}$).

Proof: (unpublished) With some modifications, we follow Beck's original proof [Bec83]:

Denote by $L_1, L_2, \dots, L_m$ the straight lines determined by pairs from E. Put $p_1, p_2, \dots, p_m$ and $e_1, e_2, \dots, e_m$ for the number of points (of $\mathcal{P}$) and of graph edges, respectively, which are on the L_i. (Of course, $e_i \le \binom{p_i}{2}$, i.e., $\sqrt{2e_i} \le p_i$ for all $i \le m$.)
Then we have

$$\sum_{i=1}^{m} e_i = |E| \ge cn^2.$$

For a sufficiently large constant C_0 and another, small c_0, decompose this sum as follows:

$$\sum e_i = \sum_{1 \le e_i < C_0} e_i + \sum_{C_0 \le e_i < n} e_i + \sum_{n \le e_i < c_0 n^2} e_i + \sum_{c_0 n^2 \le e_i} e_i.$$

These four sums will be denoted by $\sum_1, \sum_2, \sum_3, \sum_4$, respectively.

Here $\sum_2$ and $\sum_3$ can be made less than $cn^2/4$ each (which is just one quarter of all pairs) for suitable constants C_0 and c_0, by Corollaries 1.17 and 1.18, respectively. Now the Theorem comes easily since then either $\sum_4$ is non-zero whence the existence of at least one line with many pairs — i.e., case (ii) holds — or, otherwise, $\sum_1 \ge cn^2/2$ — i.e., case (i) holds for a suitable c'. ∎

This also implies an affirmative answer to the "weak" Dirac–Motzkin problem.

Theorem 1.20 *(Beck [Bec83]) Given n non-collinear points in the plane, it is always possible to select one (say P_0) such that the pairs P_0P_i ($i \ne 0$) determine at least cn distinct straight lines for a fixed $c > 0$, independent of n.*

The still unsolved "strong" Dirac–Motzkin problem asks the same with $c = 1/2$ and n large enough.

Finally we mention that not only the number of pairs but also that of the collinear triples can be estimated, as well. Here we only show a result for Cartesian products. (The proof follows that of Corollary 1.17.)

Corollary 1.21 *There is an absolute constant C such that, given any $n \times n$ Cartesian product $X \times Y$ in the plane, at most $Cn^4 \log n$ of its triples can be collinear.* ∎

1.2.3 Pseudolines and bounded-degree curves

A family of simple continuous curves (which do not intersect themselves) is a *pseudoline system* if any two has at most one common point.

Theorem 1.22 *(Clarkson et al. [CEG+90]) The Szemerédi–Trotter Theorem (Theorem 1.12) and Corollary 1.14 hold for pseudoline systems in the* real *plane* $\mathbb{R}^2$.

Unfortunately, nothing is known for *complex* pseudolines.

In what follows Γ will denote a family of simple (i.e., not self-intersecting) continuous real curves. We say that Γ is a family of r degrees of freedom if the curves are "almost uniquely" determined by r of their points. More precisely this means the following.

Definition 1.23 Γ is a family of r degrees of freedom if there is a fixed positive integer s such that

(i) for any r points of the plane, at most s members of Γ pass through all of them;

(ii) any two members of Γ have at most s points in common.

This notion originates from Pach and Sharir [PS90, PS98]. They found the following generalization of Corollary 1.14.

Theorem 1.24 *(Pach–Sharir [PS90, PS98]) Let $\mathcal{P} \subset \mathbb{R}^2$ be a set of N points, Γ a family of curves of r degrees of freedom, and $c > 0$. Then the number of k–rich members of Γ cannot exceed*

$$\begin{cases} C \cdot \frac{N^r}{k^{2r-1}} & \text{if } k \leq c\sqrt{N}; \\ \\ C \cdot \frac{N}{k} & \text{if } k > c\sqrt{N}, \end{cases}$$

where $C = C(c, r, s)$ only depends on c, r and the parameter s in Definition 1.23, but not on N or k.

This bound, unlike Corollary 1.14, is NOT known to give the correct order of magnitude for r at least 3 and $k \leq \sqrt[3]{N}$. The best lower bound found so far for the case $k \leq \sqrt[r]{N}$ is $N^r/k^{r(r+1)/2}$.

1.25 Example Let $\mathcal{P} \stackrel{\text{def}}{=} \{1, 2, \ldots, k\} \times \{1, 2, \ldots, \lfloor N/k \rfloor\}$ and Γ the set of curves described by the equations

$$y = a_{r-1}x^{r-1} + a_{r-2}x^{r-2} + \ldots a_1 x + a_0, \qquad \text{where } a_i \in \left\{1, 2, \ldots, \left\lfloor \frac{N}{rk^{i+1}} \right\rfloor \right\}$$

which really form a family of r degrees of freedom and consists of at least $cN^r/k^{r(r+1)/2}$ curves, all k–rich in $\mathcal{P}$.

Problem 1.26 *Does this example give the best order of magnitude?*

It is worth noting that Theorem 1.24 implies that the number of those curves of a family of r degrees of freedom which are all cn–rich in an $n \times n$ Cartesian product is *linear*, independently of r.

Corollary 1.27 *Let $X, Y \subset \mathbb{R}$, $|X| = |Y| = n$ and $\mathcal{P} = X \times Y \subset \mathbb{R}^2$. Moreover, let Γ be a family of curves of r degrees of freedom. Then the number of those $\gamma \in \Gamma$ which are cn–rich in $\mathcal{P}$ is $O(n)$, where the implicit constant depends on c, r and the parameter s in Definition 1.23, but not on n.*

Finally we mention that recently Endre Szabó has proven this statement for bounded–degree complex algebraic curves, as well (see Corollary 2. in [Sza01]) — though the Pach–Sharir Theorem is not known in this generality.

Theorem 1.28 *(E. Szabó [Sza01]) Corollary 1.27 holds in $\mathbb{C}^2$ for algebraic curves of degree at most r.*

Chapter 2

A shoot and the trunk.

This middle part presents the core of our survey. Section 2.1 contains results from Additive Number Theory which typically state that if a set of numbers is close to an arithmetic progression then it must be far from a geometric one, i.e., sums and products cannot really be mixed up with each other. We call this section a "shoot" (with no branches as yet) since no further applications have been found so far. The other Section 2.2 (the "trunk"), in the contrary, forms the basis of all results given in the third part. It presents structural results for certain nearly extremal configurations concerning straight lines, algebraic curves and surfaces.

Throughout both sections, the idea of the proofs will be similar. First of all, we use the Theory of Incidences everywhere. Moreover, another common feature is that, from the assumptions, we typically have some — but not enough — lines or curves with many points of a Cartesian product. In order to find more, we usually compose pairs of those functions which describe the given curves (or, perhaps, a function with the inverse of another one). These compositions will sometimes be all distinct — like those in the proof of the "hybrid" bound — while sometimes they can coincide. In such cases an elaborate study of these coincidences becomes necessary — like in all the "one-parameter-structure" results, where, once this detailed analysis is done, certain "commutator pairs" will work as our basic tools.

2.1 The "hybrid" problem of Erdős and Szemerédi.

During this section $\mathcal{A}$ will denote a finite subset of the *non-zero* real or complex numbers, and n the number of its elements.
As usual, $\mathcal{A}+\mathcal{A}$ and $\mathcal{A}\cdot\mathcal{A}$ stand for the set of all pairwise sums $\{a+a' \,;\, a, a' \in \mathcal{A}\}$ and products $\{a \cdot a' \,;\, a, a' \in \mathcal{A}\}$, respectively.
The following "hybrid" problem (i.e., one which mixes sums and products) was posed by Erdős and Szemerédi (see [ES83].)

For a given n, how small can one make $|\mathcal{A}+\mathcal{A}|$ and $|\mathcal{A}\cdot\mathcal{A}|$ simultaneously?

In other words, denoting by

$$m(\mathcal{A}) \stackrel{\text{def}}{=} \max\{|\mathcal{A}+\mathcal{A}|\ ,\ |\mathcal{A}\cdot\mathcal{A}|\},$$

a lower estimate should be found for

$$g(n) \stackrel{\text{def}}{=} \min_{|\mathcal{A}|=n} m(\mathcal{A}).$$

Remark The philosophy behind the question is that *either* of $|\mathcal{A}+\mathcal{A}|$ or $|\mathcal{A}\cdot\mathcal{A}|$ is easy to minimize — just take an arithmetic or geometric (i.e., exponential) progression for $\mathcal{A}$. However, in both of these examples, the other set becomes very large.

In their aforementioned paper, Erdős and Szemerédi managed to prove the existence of a small but positive constant ε such that $g(n) \geq n^{1+\varepsilon}$ for all n.

Later on, Nathanson and K. Ford found the lower bounds $n^{32/31}$ and $n^{16/15}$, respectively [Nat97, For98].

The following result [Ele97b] improves the exponent to 5/4.

Theorem 2.1 *There is a positive absolute constant c such that, for every n–element set $\mathcal{A}$,*

$$c \cdot n^{5/2} \leq |\mathcal{A}+\mathcal{A}| \cdot |\mathcal{A}\cdot\mathcal{A}|,$$

whence $c' \cdot n^{5/4} \leq \max\{|\mathcal{A}+\mathcal{A}|\ ,\ |\mathcal{A}\cdot\mathcal{A}|\}$.

Proof: Denote the elements of $\mathcal{A}$ by $a_1, a_2, \ldots, a_n$, and define the following n^2 functions.

$$f_{j,k}(x) \stackrel{\text{def}}{=} a_j(x - a_k) \qquad \text{for } 1 \leq j,k \leq n.$$

Lemma 2.2 *For every $j,k \leq n$, the function $f_{j,k}$ maps at least n elements of $\mathcal{A}+\mathcal{A}$ to some elements of $\mathcal{A}\cdot\mathcal{A}$.*

(Indeed, the image of $a_k + a_i$ is $a_j \cdot a_i \in \mathcal{A}\cdot\mathcal{A}$, for every $a_i \in \mathcal{A}$.)

From a geometric point of view, the above Lemma asserts that the graph of each of the functions $f_{j,k}$ contains n or more points of $\mathcal{P} \stackrel{\text{def}}{=} (\mathcal{A}+\mathcal{A}) \times (\mathcal{A}\cdot\mathcal{A})$. Put $N = |\mathcal{P}| = |\mathcal{A}+\mathcal{A}| \cdot |\mathcal{A}\cdot\mathcal{A}|$. Then, by applying Corollary 1.14 to $\mathcal{P}$ and the f_{jk} (with $k = n$ there), we get

$$n^2 \leq C_{SzTr} \cdot \frac{N^2}{n^3},$$

i.e., $N \geq C^{-1/2} n^{5/2}$. ∎

Another related problem was also posed in [ES83] by Erdős and Szemerédi. They asked whether $|\mathcal{A}+\mathcal{A}| \leq C|\mathcal{A}|$ implies $|\mathcal{A}\cdot\mathcal{A}| \geq c|\mathcal{A}|^{2-\varepsilon}$, for all $\varepsilon > 0$ and $c = c(C,\varepsilon)$. The affirmative answer was found in [ER03a].

Theorem 2.3 *(Elekes–Ruzsa [ER03a]) If $|\mathcal{A}| = n$ and $|\mathcal{A} + \mathcal{A}| \leq Cn$ then $|\mathcal{A} \cdot \mathcal{A}| \geq cn^2/\log n$.*

In this bound one cannot completely get rid of the log factor in the denominator, as shown by the example $\mathcal{A} = \{1, 2, \ldots, n\}$, for which $|\mathcal{A} \cdot \mathcal{A}| \leq n^2/\log^\alpha n$ for a positive exponent α.

Instead of Theorem 2.3, we prove the following more general statement (which, for $|\mathcal{A} + \mathcal{A}| < |\mathcal{A}|^{7/6}$, is also better than the bound found in Theorem 2.1).

Theorem 2.4 *(Elekes–Ruzsa [ER03a]) If $|\mathcal{A}| = n$ then $\big(|\mathcal{A}+\mathcal{A}|+n\big)^4 \cdot |\mathcal{A} \cdot \mathcal{A}| \geq cn^6/\log n$.*

Proof: This one will be a consequence of Corollary 1.21. Put $s = |A + A|$ and $p = |A \cdot A|$. Now $A \times A$ can be covered by p hyperbolae of the form $xy = \lambda$ ($\lambda \in A \cdot A$). An average hyperbola contains n^2/p points thus n^4/p^2 pairs of points of $A \times A$. By the convexity of $f(x) = \binom{x}{2}$, the number of pairs $\langle (a_i, a_j), (a_k, a_l) \rangle$ for which $a_i a_j = a_k a_l$ is at least cn^4/p, for a suitable absolute constant $c > 0$. From these, we can form $n^2 \cdot cn^4/p = cn^6/p$ collinear triples of the type (a_u, a_v), $(a_i + a_u, a_k + a_v)$, $(a_l + a_u, a_j + a_v)$ in $((A+A) \cup A) \times ((A+A) \cup A)$ and each such triple is counted at most three times. Applying Corollary 1.21 yields $cn^6/p \leq C(s+n)^4 \log n$, whence the required inequality. ∎

It is not difficult to demonstrate a similar bound on $\mathcal{A}/\mathcal{A}$, but one can do even better.

Theorem 2.5 *(Elekes–Ruzsa [ER03a]) If $|\mathcal{A}| = n$ and $|\mathcal{A} + \mathcal{A}| \leq Cn$ then $|\mathcal{A}/\mathcal{A}| \geq cn^2$.*

The proof is based upon the following lemma (cf. Theorem 3.28 below).

Lemma 2.6 *Let $\mathcal{P} = X \times X$ be an $N \times N$ Cartesian product in the plane and $H(\mathcal{P}, E)$ a graph on it with $|E| \geq cN^4$. Then the pairs involved in E determine at least $c'N^2$ distinct directions, for some $c' = c'(c)$.*

Proof: Put $n = N^2$ and use Theorem 1.19. The second alternative ($c''n = c''N^2$ collinear points) is impossible here if N is large enough so we get $c'n^2 = c'N^4$ distinct straight lines of which at most N^2 can be parallel (since each contains at least one point of $\mathcal{P}$).

Proof of Theorem 2.5: Apply this Lemma to $\mathcal{P} = \big((\mathcal{A}+\mathcal{A}) \cup \mathcal{A}\big) \times \big((\mathcal{A}+\mathcal{A}) \cup \mathcal{A}\big)$, $N = (C+1)n$, and the n^4 pairs (a_u, a_v), $(a_u + a_i, a_v + a_j)$ which all determine segments of slopes $a_j/a_i \in \mathcal{A}/\mathcal{A}$. ∎

One may expect that in Theorems 2.3 or 2.5 the role of addition and multiplication/division must be symmetric; however, no bound better than $n^{3/2}$ implied by Theorem 2.1 is known for the converse situation with "+" and "·" swapped.

2.1.1 Convexity and Sumsets.

The results of this section are from [ENR99]. It will turn out that products had no special role in Theorem 2.1; only the concavity of the log function was important in that statement. The methods we use are still the same: define sufficiently many distinct compositions of the functions in question and use incidence bounds.

Theorem 2.7 *Let $\mathcal{A} \subset \mathbb{R}$ be a finite set, $|\mathcal{A}| = n$, and let f be a strictly convex (or concave) function, defined on an interval containing $\mathcal{A}$. Write*

$$f(\mathcal{A}) = \{f(a) : a \in \mathcal{A}\}.$$

Then we have

$$|A \pm A| \cdot |f(A) \pm f(A)| \geq cn^{5/2},$$

where we are free to write $+$ or $-$ in place of the $\pm$ signs and $c > 0$ is an absolute constant.

(Theorem 2.1 is the special case when $f(x) = \log x$.)

Proof: Without loss of generality assume that f is strictly monotonic (otherwise, at the cost of a constant factor, we cut $\mathcal{A}$ into two parts). Consider the graph of the functions $f_{ij}(x) = f(x \mp a_i) \pm f(a_j)$ for $a_i, a_j \in \mathcal{A}$. They are all distinct and they even form a pseudoline system, since any two translates of a strictly convex or concave graph can intersect each other in at most one point. Moreover, they are all n–rich in $\mathcal{P} = (\mathcal{A} \pm \mathcal{A}) \times (f(\mathcal{A}) \pm f(\mathcal{A}))$, since each f_{ij} passes through the points $(a_i \pm a_k, f(a_j) \pm f(a_k))$, for all $a_k \in \mathcal{A}$. Applying Theorem 1.22 (the pseudoline version of Corollary 1.14) to $\mathcal{P}$ and the graphs of the f_{ij} yields the required inequality. ∎

Put $1/\mathcal{A} = \{1/a \ ; \ a \in \mathcal{A}\}$. In [ES83], Erdős and Szemerédi also asked whether at most one of $\mathcal{A} + \mathcal{A}$, $1/\mathcal{A} + 1/\mathcal{A}$ and $\mathcal{A} + 1/\mathcal{A}$ can be small. It turns out that already one of the first two must be large.

Corollary 2.8 $|\mathcal{A} + \mathcal{A}| \cdot |1/\mathcal{A} + 1/\mathcal{A}| \geq cn^{5/2}$.

Proof: Use Theorem 2.7 for $f(x) = 1/x$. ∎

Moreover, the third quantity can never be small.

Theorem 2.9 $|\mathcal{A} + 1/\mathcal{A}| \geq cn^{5/4}$.

The proof [ENR99] uses a generalization of Theorem 2.7.

Problem 2.10 *Do the above Corollary and Theorem hold for subsets of $\mathbb{C}$?*

Let $\mathcal{B} = \{b_1 < b_2 < \ldots < b_n\}$ with the property that $b_i - b_{i-1} < b_{i+1} - b_i$ holds for each $1 < i < n$. Answering a problem of Erdős, it was shown by Hegyvári [Heg86] that $|B - B| \geq cn \log n / \log \log n$. Theorem 2.7 yields a better bound.

Corollary 2.11 $|B - B| \geq cn^{3/2}$.

Proof: By assumption there exists a function f, strictly convex on an interval $I \supset \{1, 2, \ldots, n\}$, whose graph contains all the points (i, b_i) for $1 \leq i \leq n$. Let $\mathcal{A} = \{1, 2, \ldots, n\}$; then $\mathcal{B} = f(\mathcal{A})$. Using Theorem 2.7 and the fact that $|\mathcal{A} - \mathcal{A}| = n - 1 < n$, we have

$$|\mathcal{B} - \mathcal{B}| = |f(\mathcal{A}) - f(\mathcal{A})| \geq \frac{cn^{5/2}}{|\mathcal{A} - \mathcal{A}|} > cn^{3/2}. \blacksquare$$

Problem 2.12 *Can the exponent 3/2 be improved to* $2 - \varepsilon$*? Can the exponents 5/2 in all the previous results be substituted by* $3 - \varepsilon$*?*

2.2 Finding one–parametric families.

In this section we present some structure theorems which form the basis of the further applications shown in the last part. They all derive qualitative consequences from quantitative (numerical) assumptions. A typical result states that if a large set of lines, curves, or surfaces contains many points of a Cartesian product, then many of them must be contained in a family described by as few as one parameter. The methods — beyond incidence bounds — include commutator pairs and their younger brother: the paradigm "compose your functions to get more".

2.2.1 2+2=3 versions of the Linear Theorem

In what follows, we denote by $\mathcal{L}$ the set of non-constant real or complex linear functions $x \mapsto ax + b$ $(a \neq 0)$.

Four equivalent versions of the main result of this section (the "Linear Theorems" 2.13–2.16) will be presented here in three different settings:

- cN–rich lines of $N \times N$ Cartesian products;
- small image sets $\Phi(H) \stackrel{\text{def}}{=} \{\phi(h)\ ;\ \phi \in \Phi, h \in H\}$ for $H \subset \mathbb{C}$ and $\Phi \subset \mathcal{L}$;
- small composition sets $\Phi \circ \Psi \stackrel{\text{def}}{=} \{\phi \circ \psi\ ;\ \phi \in \Phi, \psi \in \Psi\}$ for $\Phi, \Psi \subset \mathcal{L}$.

As for the latter, we mention both symmetric and asymmetric composition sets; that is why there will be four versions altogether.

All these results state that certain sets of straight lines — that form a family of two degrees of freedom in general — must contain "many" concurrent or parallel lines which, of course, form a family of just one degree of freedom [Ele97a].

Theorem 2.13 (Linear Theorem, Cartesian product version) *For every* $c_1, c_2 > 0$ *there exists a* $c_3 = c_3(c_1, c_2)$ *with the following property.*
If each of $c_1 N$ *straight lines contains* $c_2 N$ *or more points of an* $N \times N$ *Cartesian*

product then at least c_3N of them must be parallel or concurrent (i.e., pass through a common point).

We do not give a direct proof; rather, we shall prove one of the three forthcoming equivalent statements.

For arbitrary subsets $H \subset \mathbb{C}$ and $\Phi \subset \mathcal{L}$ put $\Phi(H) \stackrel{\text{def}}{=} \{\phi(h) \ ; \ \phi \in \Phi, h \in H\}$ and call it an *image set.* Similarly, for $E \subset \Phi \times H$, let the *statistical image set* $\Phi_E(H)$ be defined by $\{\phi(h) \ ; \ (\phi, h) \in E\}$.

Theorem 2.14 (Linear Theorem, image set version) *Let $H \subset \mathbb{C}$, $\Phi \subset \mathcal{L}$ and $E \subset \Phi \times H$ with $N \le |\Phi|, |H|, |\Phi_E(H)| \le CN$ for an $|E| \ge cN^2$. Then there exists a $\Phi' \subset \Phi$ which consists of either parallel or concurrent lines and $|E \cap (\Phi' \times H)| \ge c'N^2$.*

Theorem 2.14 obviously implies the previous one while to prove the other implication it is sufficient to delete from Φ those ϕ which occur in less then $cN/(2C)$ pairs of E and then substitute CN for N.

Now we define composition sets of families of linear functions. They are relatives of the sumsets defined in Section 1.1. An even closer relation will be the topic of Section 3.1 below.

The set $\mathcal{L}$ of non-constant real or complex linear functions, as defined above, forms a group with operation "$\circ$", i.e., composition $\phi \circ \psi : x \mapsto \phi(\psi(x))$. For $\Phi, \Psi \subset \mathcal{L}$, we put $\Phi \circ \Psi = \{\phi \circ \psi \ ; \ \phi \in \Phi, \psi \in \Psi\}$ and call it a *composition set.* Similarly, for $E \subset \Phi \times \Psi$, we define

$$\Phi \circ_E \Psi \stackrel{\text{def}}{=} \{\phi \circ \psi \ ; \ (\phi, \psi) \in E\},$$

and call it a *statistical composition set.*

Theorem 2.15 (Linear Theorem for composition sets) *For all $c, C > 0$ there exists a $c^* = c^*(c, C) > 0$ with the following property.*
Let $\Phi, \Psi \subset \mathcal{L}$ and $E \subset \Phi \times \Psi$ with $|\Phi|, |\Psi| \le N$ and $|E| \ge cN^2$. Assume, moreover, that

$$|\Phi \circ_E \Psi| \le CN.$$

Then there are $\Phi^* \subset \Phi$ and $\Psi^* \subset \Psi$ for which $|(\Phi^* \times \Psi^*) \cap E| \ge c^* N^2$ and

- *(i) either both Φ^* and Ψ^* consist of functions whose graphs are all parallel (but the directions may be different for Φ^* and Ψ^*);*
- *(ii) or both Φ^* and Ψ^* consist of functions whose graphs all pass through a common point (which may be different for those in Φ^* and in Ψ^*).*

Since $(\mathcal{L}, \circ)$ is non–Abelian, the size and structure of $\Phi \circ \Psi$ and of $\Psi \circ \Phi$ can be very different, e.g., for $\Phi = \{x \mapsto 2^t x + 1 \ : \ t = 1 \ldots N\}$ and $\Psi = \{x \mapsto 2^t x \ : \ t = 1 \ldots N\}$, where $|\Phi \circ \Psi| = 2N - 1$ while $|\Psi \circ \Phi| = N^2$. That is why we shall also consider *symmetric* composition sets $(\Phi\circ_E\Psi)\cup(\Psi\circ_E\Phi)$, as opposed to *asymmetric* ones like $\Phi\circ_E\Psi$. The following statement involves Ψ^{-1} instead of just Ψ, for reasons which will become clear later in Section 2.2.2.

Theorem 2.16 (Linear Theorem for symmetric composition sets) *Let $\Phi, \Psi \subset \mathcal{L}$ and $E \subset \Phi \times \Psi$ be as above and assume that*

$$|(\Phi\circ_E\Psi^{-1})\cup(\Psi^{-1}\circ_E\Phi)| \leq CN.$$

Then the conclusion of Theorem 2.15 holds.

Of course, the foregoing Theorem 2.15 is stronger than this Theorem 2.16. (We stated them separately because it is the "weaker" symmetric version whose proof will be outlined below.) Moreover, they can even be shown to be equivalent both to each other and to Theorem 2.14.

Claim 2.17 *Theorem 2.16 implies Theorem 2.14 and the latter implies Theorem 2.15.*

Outline of the proof: (see [Ele97a] for more details). In order to show the first implication, we define a three–partite graph on vertex sets H, Φ and $\Phi_E(H)$ by using the edges in E between H and Φ and connecting $\phi \in \Phi$ to $\phi(x) \in \Phi_E(H)$ if $(\phi, x) \in E$. Then we count the two-paths $\langle x_1, \phi, \phi(x_2)\rangle$ (NOT just $\phi(x_1)$!!) and find at least $c'N^3$ of them. Therefore, there are $c''N^4$ four–cycles $\langle x, \phi_1, y, \phi_2, x\rangle$ in this graph. We conclude that $c'''N^2$ pairs $\langle \phi_1, \phi_2\rangle$ have at least $c''''N$ common neighbors both in H and in $\Phi_E(H)$. Denote by E^* the graph formed by these pairs and observe that, for $(\phi_1, \phi_2) \in E^*$, the functions $\phi_2^{-1}\circ\phi_1$ and $\phi_1\circ\phi_2^{-1}$ have graphs which are $c''''N$–rich in $H \times H$ and $\Phi_E(H) \times \Phi_E(H)$, respectively. According to Corollary 1.14 (or 1.27), the number of such straight lines is $O(N)$. Therefore, $|(\phi_1\circ_{E^*}\phi_2^{-1})\cup(\phi_2^{-1}\circ_{E^*}\phi_1)| \leq C'N$; thus we can use Theorem 2.16 to find the required sub–structure Φ'.
As for the second implication, first pick a "sufficiently general" $u \in \mathbb{C}$, i.e., one for which the values $\psi(u)$ $(\psi \in \Psi)$ are all distinct. Put $H = \Psi(\{u\})$ and define the edge set $E' = \{(\phi, \psi(u))\ ;\ (\phi, \psi) \in E\}$ on vertex set $\Phi \cup H$. Observe that $|\Phi_{E'}(H)| \leq |\Phi\circ_E\Psi| \leq CN$ and use Theorem 2.14 to find many concurrent or parallel ϕ. Repeating for $\Psi^{-1}\circ_E(\Phi')^{-1}$ (at the cost of some technical details, see [Ele97a]) finishes the proof. ∎

2.2.2 Proof of the Linear Theorems

Of the four equivalent versions, it is the "symmetric composition set" form, i.e., Theorem 2.16, whose proof we are going to present here.

Since we are studying the non-Abelian group $\mathcal{L}$, it is quite natural to define some notions that can be considered as relatives of the usual commutators.

Definition 2.18 For $\phi, \psi \in \mathcal{L}$, the pair $(\phi\circ\psi^{-1}, \psi^{-1}\circ\phi)$ is called the *commutator pair* defined by ϕ and ψ. (This name originates from M. Simonovits.)

Also, for $\Phi, \Psi \subset \mathcal{L}$ and $E \subset \Phi \times \Psi$, the *commutator graph* $\hat{G}_E(\hat{V}, \hat{E})$ defined by E, has edge set $\hat{E}$ which consists of the corresponding commutator pairs, i.e.,

$$\hat{V} = (\Phi \circ_E \Psi^{-1}) \cup (\Psi^{-1} \circ_E \Phi); \text{ and}$$
$$\hat{E} = \{(\phi \circ \psi^{-1}, \psi^{-1} \circ \phi) \ ; \ (\phi, \psi) \in E\}.$$

Of course, the two terms of a commutator pair are identical (and the corresponding edge in $\hat{G}$ is a self–loop) iff ϕ and ψ commute.

Remark 2.19 The commutator graph has the interesting property that, in any of its connected components, the vertices represent functions with all identical leading coefficients (i.e., equal slopes). This is because the two endpoints of any edge (the two members of a commutator pair) obviously have this property.

Different pairs may define identical commutator pairs. The following observation gives a sufficient condition for the commutator graph to have all distinct edges.

Lemma 2.20 *If, for $\phi, \psi \in E$, the intersection points of the straight lines ϕ and ψ all exist and are distinct, then no two edges of the commutator graph coincide.*

Proof: If (u, v) is the point of intersection of ϕ and ψ then u is the unique fixed point of $\psi^{-1} \circ \phi$ and v is the unique fixed point of $\phi \circ \psi^{-1}$. ■

Now we prove the "symmetric" Linear Theorem.
Proof of Theorem 2.16: We shall use the *dual* of Beck's Theorem 1.19 where the role of lines and points are exchanged, concluding that many intersection points are distinct or many coincide. First we apply this to the straight lines $\Phi \cup \Psi$ (actually functions which we identify with their graphs) and edge set E. If many pairs $(\phi, \psi) \in E$ have identical points of intersection then we are done; even the common points of Φ^* and Ψ^* coincide. Otherwise there are many distinct points of intersection, whence, by Lemma 2.20, the commutator graph $\hat{G}_E(\hat{V}, \hat{E})$ has many distinct edges. It is not difficult to show that a "dense" graph (where the number of edges is quadratic in the number of vertices) contains a connected component which still has a quadratic number of edges. These commutator edges, on the one hand, connect vertices with all equal slopes by Remark 2.19; while on the other hand, they are generated by many edges from E. There again, we find a connected component with a quadratic number of such original edges; their end-vertices in Φ and Ψ will do as Φ^* and Ψ^*, respectively (see [Ele97a] for more details). ■

2.2.3 Hyperbolas and the projective groups of $\mathbb{R}$ and $\mathbb{C}$.

Let $\mathcal{P}$ denote the group of non-degenerate projective mappings of $\mathbb{C}$, i.e., the set of non-constant linear fractions $x \mapsto \frac{ax+b}{cx+d}$ (where $ad - bc \neq 0$). Note that linear functions are also contained in $\mathcal{P}$; that is why we shall consider straight lines as special hyperbolas in the sequel. The Linear Theorem 2.13 can be generalized to hyperbolas as follows [EK01].

Theorem 2.21 (Hyperbola Theorem, Cartesian product version) *For every $c > 0$ there is a $c_0 = c_0(c)$ with the following property.*
Let $X \times Y \subset \mathbb{C}^2$ be an $N \times N$ Cartesian product and $\mathcal{H}$ a set of cN hyperbolas of equations $y = \gamma(x)$ (for some $\gamma \in \mathcal{P}$). If each $h \in \mathcal{H}$ contains at least cN points of $X \times Y$ then there exists a $\mathcal{H}_0 \subset \mathcal{H}$ with $|\mathcal{H}_0| \geq c_0 N$, and linear fractions f, $g \in \mathcal{P}$ for which

$$\begin{aligned} &\mathcal{H}_0 \subset \{x \mapsto f\big(g(x) + t\big) \; ; \; t \in \mathbb{C}\}; \qquad \textit{or} \\ &\mathcal{H}_0 \subset \{x \mapsto f\big(g(x) \cdot t\big) \; ; \; t \in \mathbb{C}\}. \end{aligned} \tag{2.1}$$

This result generalizes the Linear Theorem 2.13 since parallel and concurrent lines are graphs of functions of the first and second type in (2.1), respectively. Moreover, — just like its linear counterpart — it follows from its forthcoming composition set version.

Remark 2.22 As for *real* linear fractions, it is possible that we must use *complex* coefficients in the foregoing expansion. E.g., the graphs of the functions $h_m(x) = (x + \tan(m))/(1 - x\tan(m))$ $(1 \leq m \leq N/2)$ are $N/2$–rich in $\{\tan(1), \ldots, \tan(N)\} \times \{\tan(1), \ldots, \tan(N)\}$. However, they can only be written as

$$h_m(x) = \alpha\left(\frac{x+i}{x-i} \cdot \frac{\tan(m)+i}{\tan(m)-i}\right)$$

where $\alpha(\zeta) = i(\zeta - 1)/(\zeta + 1)$. This — not really aesthetic — situation can be avoided by also allowing a third form: for *real* fractions f, g,

$$\mathcal{H}_0 \subset \left\{x \mapsto f\left(\frac{g(x)+t}{1 - g(x)\cdot t}\right) \; ; \; t \in \mathbb{R}\right\},$$

in (2.1), which turns out to be a special case of the (complex) second version there.

In what follows we consider $\mathcal{P}$ with the composition as group operation and state a result on composition sets.

Theorem 2.23 (Projective Theorem, function version) *For every $C > 0$ there is a $c_0 = c_0(C)$ with the following property.*
Let $\Phi, \Psi \subset \mathcal{P}$ with $N \leq |\Phi|, |\Psi| \leq CN$ and $E \subset \Phi \times \Psi$ with $|E| \geq N^2$. If $|\Phi \circ_E \Psi| \leq CN$, then there exist $\Phi_0 \subset \Phi$, $\Psi_0 \subset \Psi$ with $|E \cap (\Phi_0 \times \Psi_0)| \geq c_0 N^2$, such that either both are of the first type or both are of the second type in (2.1), perhaps with different pairs of functions (f, g) for Φ_0 and Ψ_0.

The proof [EK01] is similar to that of the Linear Theorems in the previous section. Actually, an equivalent symmetric version can be proven using commutator pairs while in the proof of equivalence, in place of Corollary 1.14, Corollary 1.27 is used in the real case and Theorem 1.28 for complex linear fractions.

This result can also be reformulated in terms of subgroups and cosets as follows [EK01].

Theorem 2.24 (Projective Theorem, subgroup version) *If Φ and Ψ are as above then there exist $\phi \in \Phi$, $\psi \in \Psi$, and an Abelian subgroup $S \subset \mathcal{P}$ such that*

$$|\Phi \cap \phi\, S| \geq c_0 N;$$
$$|\Psi \cap S\, \psi| \geq c_0 N.$$

It would be interesting to know whether something similar to Theorem 2.24 holds for other groups beyond $\mathcal{P}$. One cannot expect the original statement without any modifications, as shown by the following example of Endre Szabó.

2.25 Example Let $N = m^4$ and Φ consist of the following matrices:

$$M(x, y, z) = \begin{pmatrix} 1 & x & y \\ 0 & 1 & z \\ 0 & 0 & 1 \end{pmatrix},$$

where $x, z = 1, 2, \ldots, m$ and $y = 1, 2, \ldots, m^2$. Then $|\Phi| = m^4 = N$, moreover

$$M(x, y, z) \cdot M(a, b, c) = \begin{pmatrix} 1 & x + a & b + xc + y \\ 0 & 1 & z + c \\ 0 & 0 & 1 \end{pmatrix};$$

thus we only get $|\Phi \circ \Phi| \leq CN$ distinct products. (The coefficient $C = 12$ will do since $1 \leq x + a \leq 2m$, $1 \leq z + c \leq 2m$ and $1 \leq b + xc + y \leq 3m^2$.)
On the other hand, $M(x, y, z)$ and $M(a, b, c)$ only commute if $xc = az$, i.e., the vectors (x, z) and (a, c) are parallel. We conclude that any coset of an Abelian subgroup only contains $m^3 = N^{3/4} = o(N)$ elements of Φ.

It may well be true that e.g., even for higher dimensional matrix groups, Theorem 2.24 holds with S a nilpotent subgroup.

2.2.4 Polynomial and rational curves.

As usual, we denote by $\mathbb{R}[t]$ and $\mathbb{C}[t]$ the ring of polynomials with real or complex coefficients, respectively. Similarly, $\mathbb{R}(t)$ and $\mathbb{C}(t)$ stands for rational functions (quotients of polynomials) over $\mathbb{R}$ and $\mathbb{C}$. For short, the graph of a polynomial or rational function will be called a polynomial curve or a rational curve, respectively.

The following result [ER00] generalizes Theorems 2.13 and 2.21 to rich polynomial and rational curves. The proof reduces the general statement to those two special cases, using some Commutative Algebra.

Theorem 2.26 *(Elekes–Rónyai [ER00]) For every $c > 0$ and positive integer d there is a $c_0 = c_0(c, d)$ with the following property.*
Let $X \times Y \subset \mathbb{C}^2$ be an $N \times N$ Cartesian product and $\mathcal{F} \subset \mathbb{C}(t)$ a set of cN rational functions of degree at most d. If the graph of each $f_i \in \mathcal{F}$ contains at least cN points of $X \times Y$ then there exists an $\mathcal{F}_0 \subset \mathcal{F}$ with $|\mathcal{F}_0| \geq c_0 N$, and rational functions $f,\ g \in \mathbb{C}(t)$ for which

$$\begin{aligned} &\mathcal{F}_0 \subset \{x \mapsto f\big(g(x) + s\big)\ ;\ s \in \mathbb{C}\}; \quad \textit{or} \\ &\mathcal{F}_0 \subset \{x \mapsto f\big(g(x) \cdot s\big)\ ;\ s \in \mathbb{C}\}. \end{aligned}$$

Moreover, if $\mathcal{F} \subset \mathbb{C}[t]$ or $\mathcal{F} \subset \mathbb{R}[t]$ consists of polynomials, then also f and g can be required to be polynomials in $\mathbb{C}[t]$ or $\mathbb{R}[t]$, respectively.
Finally, if $\mathcal{F} \subset \mathbb{R}(t)$ consists of real rational functions and we insist on real $f, g \in \mathbb{R}(t)$, then a third type (similar to that in Remark 2.22) must also be allowed.

2.2.5 Polynomial and rational surfaces.

Now we turn our attention to rational and polynomial surfaces in three dimensions (i.e., graphs of functions from one of $\mathbb{C}(x, y)$, $\mathbb{C}[x, y]$, $\mathbb{R}(x, y)$ or $\mathbb{R}[x, y]$). In the plane, the graph of any function can be cN–rich in a suitably chosen $N \times N$ Cartesian product. (That is why we were only interested in large families of such curves.) This is not the case for $N \times N \times N$ Cartesian products and cN^2–rich surfaces in the three-space. Of course, any plane $z = ux + vy + w$ will contain $N^2/2$ points of $\{1/u, 2/u, \dots, N/u\} \times \{1/v, 2/v, \dots, N/v\} \times \{1+w, 2+w, \dots, N+w\}$. However, already for very simple surfaces like rational or polynomial ones, this is impossible in general. It turns out that surfaces which are rich like those mentioned above, must have a rather special form (which, again, has a certain "either sums or products" flavor). The problem originates from [Ele98b].

Theorem 2.27 *(Elekes–Rónyai [ER00]) For every $c > 0$ there is an $n_0 = n_0(c)$ with the following property.*
Assume that $F \in \mathbb{C}(x, y)$ has the property that for an $N > n_0$ there exists an $N \times N \times N$ Cartesian product $X \times Y \times Z \subset \mathbb{C}^3$ in which the graph of F is cN^2–rich. Then there exist rational functions $f, g, h \in \mathbb{C}(t)$ such that

$$\begin{aligned} &F(x, y) = f\big(g(x) + h(y)\big); \quad \textit{or} \\ &F(x, y) = f\big(g(x) \cdot h(y)\big). \end{aligned}$$

Also, the special cases concerning polynomials and real rational functions mentioned in Theorem 2.26 remain valid.

Remark 2.28 Such functions are really sufficiently rich in appropriate Cartesian products $X \times Y \times Z$. E.g., if F can be expressed in terms of a sum or product then we can make $g(X)$ and $h(Y)$ an arithmetic or geometric progression, respectively. Even generalized arithmetic or geometric progressions will do. It can also be shown using Theorem 1.5 that each of $g(X)$, $h(Y)$ and $f^{-1}(Z)$ must contain $c^* N$ elements of such a progression of size $c^{**} N$.

The following is a simple analytic consequence which may be useful in certain applications (e.g., Theorem 3.37).

Corollary 2.29 *(Elekes–Rónyai [ER00]) If c and F are as above then*

(1) the quantity

$$q_1(x,y) \stackrel{\text{def}}{=} \frac{\partial F}{\partial x} / \frac{\partial F}{\partial y}$$

is the product/quotient of two functions of one variable each; indeed, using Theorem 2.27, in the first case $q_1(x,y) = g'(x)/h'(y)$; while the second case can be reduced to this one since $f(g(x)\cdot h(y)) = f(\exp(\log g(x)+\log h(y))) = \phi(\gamma(x)+\chi(y))$;

(2) also,

$$q_2(x,y) \stackrel{\text{def}}{=} \frac{\partial^2(\,\log|\,q_1(x,y)|\,)}{\partial x \partial y} = 0$$

identically (by (1)), wherever q_2 is well-defined.

2.2.6 Algebraic surfaces.

Problems of Combinatorial Geometry can often be transformed into polynomial (i.e., algebraic) equations. Several questions on distances or incidences can be reduced to studying whether the zero set

$$S_F \stackrel{\text{def}}{=} \{(x,y,z) \in \mathbb{R}^3 \ ; \ F(x,y,z) = 0\}$$

of a polynomial $F \in \mathbb{R}[x,y,z]$ of three variables can contain many points of an $N \times N \times N$ Cartesian product of the three–space. In the previous section we could see surfaces of equation $z = f(x,y)$ i.e., $F(x,y,z) = z - f(x,y) = 0$; now we show some structure theorems on general trivariate polynomials.

A purely algebraic result of E. Szabó [Sza01] is the following. (Note that it only requires less than N^2 points of a Cartesian product on the surface.)

Theorem 2.30 *(E. Szabó [Sza01]) If the surface S_F contains $cN^{1{,}95}$ points of an $N \times N \times N$ Cartesian product then F is "locally the pull–back of a one dimensional algebraic group operation". (We do not define these notions here.)*

In some sense, this is a generalization of Theorem 2.26. (Here — beyond the additive and multiplicative groups of $\mathbb{C}$ — there also occurs the Abelian group $\mathbb{C}/\mathbb{Z}^2$ related to elliptic cubics.)

Assume that the algebraic surface $S = S_F$ is not "vertical cylindric", i.e., $F \in \mathbb{C}[x,y,z]$ does depend on z. Then its tangent planes are non-vertical at almost every point $P \in S_F$. (Here "almost every" is not meant in measure theoretic sense; rather, it stands for "on a dense open subset relative to S".) In a sufficiently small neighborhood of such a $P \in S_F$, the surface is the graph of an analytic function $\phi_P(x, y)$.

What we really need is a consequence of Theorem 2.30 [ES03]. It, again, is a relative of Theorem 2.26 for general algebraic curves.

Theorem 2.31 *(Elekes–E. Szabó [ES03]) For any positive integer r there exists an $n_0 = n_0(r)$ with the following property.*
Let $F \in \mathbb{C}[x, y, z]$ of degree r depend on each of x, y and z. Assume that the algebraic surface $S = S_F$ contains at least $N^{1,95}$ points of an $N \times N \times N$ Cartesian product, for an $N > n_0$. Then, for almost every point $P \in S$, there exist analytic functions f, g, h in one variable such that the function ϕ_P defined above (i.e., which describes S in a neighborhood of P) can be expressed as

$$\phi_P(x, y) = f(g(x) + h(y)).$$

One might wonder why the product form disappeared. However, as mentioned in Corollary 2.29 part (1), they can be expressed in terms of sums if we also allow analytic functions like exp and log. (Actually, we must allow even more complicated functions in this theorem, like the Weierstrass $\wp$ function.)

Also, Corollary 2.29 holds for algebraic surfaces, as well.

Corollary 2.32 *(Elekes–E. Szabó [ES03]) If S is as in Theorem 2.31 then*

$$q_1(x, y) \stackrel{\text{def}}{=} \frac{\partial \phi_P}{\partial x} \Big/ \frac{\partial \phi_P}{\partial y} \qquad \textit{and}$$

$$q_2(x, y) \stackrel{\text{def}}{=} \frac{\partial^2 (\log |q_1(x, y)|)}{\partial x \partial y}$$

satisfy the statement of Corollary 2.29; especially, $q_2 = 0$ for almost all $P \in S$.

Chapter 3

Branches and some fruits

3.1 Generalizations of Freiman's Theorem

The main result of this section, on the one hand, generalizes Freiman's result Theorem 1.3, while, on the other hand, it characterizes small composition sets. We have seen (in Theorem 2.15) that such composition sets must contain many functions with parallel or concurrent graphs. Now we show that they must even be contained in the union of a bounded number of such parallel or concurrent bunches.

We start with two examples of small composition sets. Recall that $\mathcal{L}$ denotes the set of non-constant real or complex linear functions $x \mapsto ax + b$ $(a \neq 0)$.

3.1 Example Let $G \subset \mathbb{C}$ be an arithmetic GP, $\Phi, \Psi \subset \mathcal{L}$ and C a positive integer. We say that the pair (Φ, Ψ) is an arithmetic GP-type structure based upon G with C slopes if there are non-zero complex numbers $s_1, s_2, \ldots, s_C$ such that

$$\Phi^{-1} \cup \Psi \;\; = \;\{x \mapsto s_i x + g \; ; \;\; 1 \le i \le C \;\; \text{and } g \in G\}.$$

Such structures really determine small composition sets since it is known from Remark 1.2 that arithmetic GP's have small sumsets.

Remark 3.2 At first glance it might seem awkward that Φ^{-1} occurs together with Ψ. However, a closer look shows that "$\Phi \circ \Psi$ is small" is equivalent to "$\Psi^{-1} \circ \Phi^{-1}$ is small"; thus Φ^{-1} and Ψ must play equal roles.

Geometric GP's can be defined similarly to arithmetic ones (Definition 1.1); instead of sums of linear combinations of the differences Δ_i we consider products $\pm \prod q_i^{k_i}$ of powers of positive quotients q_i. (Note that for convenience we allow both positive and negative signs here.) Also, we shall use $\mathcal{G}^{d,n}$ for geometric, as well as arithmetic, GP's — hopefully this will cause no ambiguity.

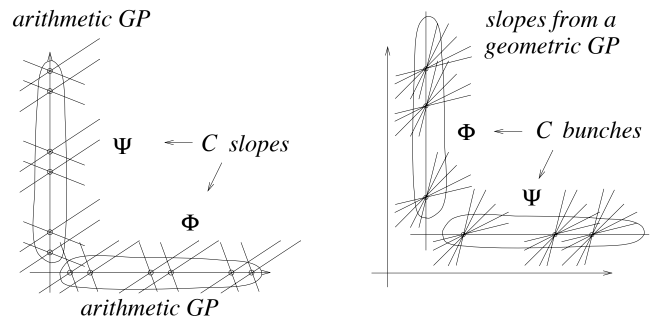

Figure 3.1: Φ (not Φ^{-1}) and Ψ in an **(a)** arithmetic **(b)** geometric GP-type structure.

3.3 Example Let $G \subset \mathbb{C}$ be a geometric GP, $\Phi, \Psi \subset \mathcal{L}$ and C a positive integer. We say that the pair (Φ, Ψ) is a geometric GP-type structure based upon G with C bunches if there are complex numbers $u_1, u_2, \dots, u_C$ and v such that

$$\Phi^{-1} \cup \Psi \;\; = \; \{x \mapsto g(x - u_i) + v \; ; \;\; 1 \leq i \leq C \;\; \text{and } g \in G\}.$$

These structures, again, determine small composition sets by the "geometric progression" version of Remark 1.2 for product sets. Moreover, no other, essentially different example exists beyond these. This is formulated in the following statement which is the main result of this section.

Theorem 3.4 *(Elekes [Ele98a]) For every $C > 0$ there are $C^* = C^*(C) > 0$, $C^{**} = C^{**}(C) > 0$ and $d^* = d^*(C) > 0$ with the following property. If $\Phi, \Psi \subset \mathcal{L}$ with $|\Phi|, |\Psi| \geq N$ and*

$$|\Phi \circ \Psi| \leq CN$$

then (Φ, Ψ) is contained in an arithmetic or in a geometric GP-type structure with $\leq C^*$ slopes or bunches, respectively, based upon an arithmetic or geometric $G \in \mathcal{G}^{d^*, C^{**}N}$.

Outline of the proof: First use Theorem 2.15 to find a regular subset Φ^* (we shall not need Ψ^* here). This is contained in a left coset of one of the two types of Abelian subgroups of $\mathcal{L}$: functions of slope 1 or those which pass through a common point (u, u) on the line $y = x$. Therefore, Ψ, too, must be contained in a bounded number of such *right* cosets (otherwise the composition set would be too large). This implies that the original Φ is contained in a bounded number of left cosets. Then we use Freiman's Theorem 1.3 or its multiplicative counterpart to find suitable arithmetic or geometric GP's, respectively. Finally, we collect these in one — larger — GP (see [Ele98a]). ∎

Remark 3.5 As mentioned before, this theorem is a common generalization of Theorem 1.3 and of its product version, as well. They are equivalent to the special cases when $\Phi = \Psi = \{x + a_i \ ; \ a_i \in X\}$ or they are of type $a_i x$ $(a_i \in X)$, respectively. (Of course, this gives no new proof for Freiman's Theorem since it was used in the foregoing argument.)

Also an "uniform" statistical version can be deduced from Theorem 1.9.

Theorem 3.6 *Let $\alpha > 0$ be fixed, $\Phi, \Psi \subset \mathcal{L}$ as in Theorem 3.4, and $H(\Phi, \Psi, E)$ a bipartite graph with all degrees at least αN. If*

$$|\Phi \circ_E \Psi| \leq CN$$

then (Φ, Ψ) is contained in a bounded number of arithmetic and geometric GP-type structures.

We also show a related result on the structure of small image sets.

Definition 3.7 Let G be an arithmetic GP and C a positive integer. We say that $\Phi \subset \mathcal{L}$ and $H \subset \mathbb{C}$ is an arithmetic GP-type structure based upon G with C slopes if there are non-zero real or complex numbers $s_1, s_2, \ldots, s_C$ such that

$$\begin{aligned} H &= G; \\ \Phi &= \{x \mapsto s_i(x+g) \ ; \ \ 1 \leq i \leq C \ , \ g \in G\}; \quad \text{and so} \\ \Phi(H) &= \bigcup_{i=1}^{C} s_i(G+G). \end{aligned}$$

Similarly, if G is a geometric GP then $\Phi \subset \mathcal{L}$ and $H \subset \mathbb{C}$ is a geometric GP-type structure based upon G with C bunches if there are real or complex numbers u and $v_1, v_2, \ldots, v_C$ such that

$$\begin{aligned} H &= (G \cup \{0\}) + u; \\ \Phi &= \{x \mapsto g(x-u) + v_i \ ; \ \ 1 \leq i \leq C \ , \ g \in G\}; \quad \text{and so} \\ \Phi(H) &= \bigcup_{i=1}^{C} \big((G \cdot G \cup \{0\}) + v_i\big). \end{aligned}$$

Theorem 3.8 *For every $C > 0$ there are $C^* = C^*(C) > 0$ and $d^* = d^*(C) > 0$ with the following property. If $\Phi \subset \mathcal{L}$ and $H \subset \mathbb{C}$ with $|\Phi|, |H| \geq N$ and*

$$|\Phi(H)| \leq CN$$

then (Φ, H) is contained in an arithmetic or in a geometric GP-type structure with $\leq C^*$ slopes or with $\leq C^*$ bunches, respectively, based upon an arithmetic or geometric $\mathcal{G}^{d^*, C^* N}$.

This result, again, generalizes the two versions of Freiman's Theorem 1.3.

It would be interesting to know whether also an "uniform" statistical version, similar to Theorem 3.6, can be stated. In terms of Cartesian products, this question is equivalent to the following.

Problem 3.9 *Is it true that if each of cN straight lines are cN–rich in an $N \times N$ Cartesian product then the lines are contained in the union of $C = C(c)$ arithmetic and geometric GP-type structures?*

An affirmative answer would be implied by the following conjecture of József Solymosi. We say that some straight lines are in general position if no two are parallel and no three pass through a common point.

Conjecture 3.10 *(J. Solymosi, unpublished) Among the lines which are cN–rich in an $N \times N$ Cartesian product, at most $C = C(c)$ can be in general position.*

3.2 Similar and homothetic subsets.

How many subsets of a set of n points in the Euclidean plane (or in higher dimensions) can be similar/homothetic to each other, or to a prescribed pattern? The study of this problem was initiated in [EE94].

Definition 3.11 Given a set (or "pattern") $\mathcal{P} = \{P_1, P_2, \ldots, P_t\}$ of $t \geq 2$ points and another set $\mathcal{A} = \{A_1, A_2, \ldots, A_n\}$ in the r dimensional space $\mathbb{R}^r$, we denote the number of those subsets of $\mathcal{A}$ which are similar to $\mathcal{P}$ by

$$s^{(r)}(\mathcal{P}, \mathcal{A}) \stackrel{\text{def}}{=} \#\{\mathcal{A}' \subset \mathcal{A}\ ;\ \mathcal{A}' \sim \mathcal{P}\}$$

where "$\sim$" means "similar", i.e. a magnified/shrunk and possibly rotated image. Similarly, the number of subsets homothetic to $\mathcal{P}$ is

$$h^{(r)}(\mathcal{P}, \mathcal{A}) \stackrel{\text{def}}{=} \#\{\mathcal{A}' \subset \mathcal{A}\ ;\ \mathcal{A}' \approx \mathcal{P}\}$$

where "$\approx$" means "homothetic", i.e. a magnified/shrunk image in the same position (rotations are not allowed).

We shall be interested in the order of magnitude of these two quantities for large values of $n = |\mathcal{A}|$. The following planar bound is obvious.

Proposition 3.12 $h^{(2)}(\mathcal{P}, \mathcal{A}) \leq s^{(2)}(\mathcal{P}, \mathcal{A}) \leq 2n(n-1)$.

Indeed, $P_1 \in \mathcal{P}$ is to be mapped to one of the n points of $\mathcal{A}$ and P_2 to one of the remaining $n - 1$. Once these two images are given, there remain at most two possibilities.

As we shall see, the order of magnitude of $s^{(2)}(\mathcal{P}, \mathcal{A})$ can really be quadratic for various (fixed) patterns and can be close to quadratic anyway. As for the order of magnitude of $h^{(2)}(\mathcal{P}, \mathcal{A})$, it cannot exceed $n^{3/2}$ if $\mathcal{P}$ is a proper planar (i.e., non–collinear) pattern and this, again, can be attained.

Moreover, we also consider "larger" homothetic and similar subsets, as well. This means that, together with $n = |\mathcal{A}|$, also $t = |\mathcal{P}|$ may go to infinity. E.g., we may ask for the maximal number of similar/homothetic $\sqrt{n}$–subsets or $n/100$–subsets of an n element set. In such cases, we shall use the notation

$$H^{(r)}(t,n) \stackrel{\text{def}}{=} \max\{h^{(r)}(\mathcal{P},\mathcal{A}) \ ; \ |\mathcal{P}|=t,\ |\mathcal{A}|=n\}$$

and consider H as a function of two variables. The only dimension, where the order of magnitude of H is known for general (i.e., not fixed) t and proper r–dimensional patterns $\mathcal{P}$, is $r=1$.

Theorem 3.13 *There is a positive constant C (independent of t and n) for which*

$$H^{(1)}(t,n) \le C\frac{n^2}{t}.$$

Moreover, if $n \ge 2t$ then

$$H^{(1)}(t,n) \ge \frac{1}{16}\frac{n^2}{t}.$$

Proof: [Ele99c] Assume that $|\mathcal{P}|=t,\ |\mathcal{A}|=n$ and $H=H^{(1)}(t,n)=h^{(1)}(\mathcal{P},\mathcal{A})$. Denote by $\mathcal{P}_1,\dots,\mathcal{P}_H$ the subsets of $\mathcal{A}$ which are homothetic to $\mathcal{P}$ and by f_i (for $i=1,\dots,H$) those linear functions $f_i(x)=m_ix+b_i$ which map $\mathcal{P}$ to $\mathcal{P}_i$, i.e., for which $f_i(\mathcal{P})=\mathcal{P}_i$. Then the graph of each such f_i contains t points of $\mathcal{P}\times\mathcal{A}$, i.e., those of the form $(p,f_i(p))$ for $p\in\mathcal{P}$. By the Szemerédi–Trotter Theorem (actually by its Corollary 1.14) the number of such graphs (straight lines) satisfies

$$H \le C_{SzTr}\frac{|\mathcal{P}\times\mathcal{A}|^2}{t^3} = C\frac{(tn)^2}{t^3} = C\frac{n^2}{t},$$

as required.

The lower bound comes easily by showing that an arithmetic progression of n terms contains $\lfloor\frac{n}{2t}\rfloor\lfloor\frac{n}{2}\rfloor$ (or more) homothetic copies of an arithmetic progression of t terms. ∎

Remark 3.14 The foregoing result gives the exact order of magnitude (apart from the multiplicative constant) for $2t\le n$. Even for smaller values of n, two arithmetic progressions show that

$$H^{(1)}(t,n) \ge \begin{cases} n-t+1 & \text{if } n<2t-1 \\ t+1 & \text{if } n=2t-1. \end{cases}$$

Perhaps this formula also gives the exact value, i.e., equality may hold here in place of "$\ge$".
Moreover, for higher dimensional cases, the same upper bound as in Theorem 3.13 can be demonstrated by projecting the space to a sufficiently general straight line (such a way that the images of no two points of $\mathcal{A}$ coincide). However, for *full dimensional patterns*, this bound is not known to be best possible and, probably, it is never sharp — as it has already been mentioned for fixed size $\mathcal{P}$.

For similar subsets in the plane, we have the following.

Theorem 3.15 *(Cs. Tóth [Tót03])* $cn^2/t \le S^{(2)}(t,n) \le Cn^2/t$.

Proof: The first inequality comes from $H^{(1)}(t,n) \le S^{(2)}(t,n)$ and Theorem 3.13 while the second one can be shown by repeating the previous argument using complex linear functions. ∎

Here again, nothing is known in higher dimensions, unless t is fixed.

3.2.1 Fixed patterns.

For $\mathcal{P}$ fixed and n an arbitrary positive integer, we write

$$\begin{aligned} S^{(r)}(\mathcal{P},n) &\stackrel{\text{def}}{=} \max_{|\mathcal{A}|=n} s^{(r)}(\mathcal{P},\mathcal{A}) \qquad \text{and} \\ H^{(r)}(\mathcal{P},n) &\stackrel{\text{def}}{=} \max_{|\mathcal{A}|=n} h^{(r)}(\mathcal{P},\mathcal{A}). \end{aligned}$$

By Proposition 3.12, $S^{(2)}(\mathcal{P},n) = O(n^2)$. The simplest non-collinear example, when this order of magnitude is attained, is a square as $\mathcal{P}$ and a $\sqrt{n}\times\sqrt{n}$ square lattice. There also are this many equilateral triangles in a triangle lattice. The first non-trivial question asks about $\mathcal{P}$ a regular pentagon. It might be natural to think — since no pentagonal lattice exists — that a quadratic order of magnitude is impossible here; but this is not the case [EE94].

3.16 Example Let ε be the principal fifth root of unity and define a "pentagonal pseudo-lattice" by

$$\mathcal{G}_\infty \stackrel{\text{def}}{=} \{a_3\varepsilon^3 + a_2\varepsilon^2 + a_1\varepsilon + a_0 \ ; \ a_i \in \mathbb{Z}\}.$$

Then n suitable points of $\mathcal{G}_\infty$ contain cn^2 regular pentagons. Indeed, if we consider $\mathcal{G}_m := \{a_3\varepsilon^3 + a_2\varepsilon^2 + a_1\varepsilon + a_0 \ ; \ \forall |a_i| \le m\} \subset \mathcal{G}_\infty$, then any pair of points of $\mathcal{G}_m$ will determine a regular pentagon (the first point as its center and the other one as a vertex) whose vertices are in $\mathcal{G}_{17m}$. Thus, for $n = |\mathcal{G}_{17m}|$, the order of magnitude of $S^{(2)}(\mathcal{P},n)$ is at least $|\mathcal{G}_m|^2 \ge cn^2$.

Similar constructions give the following result.

Theorem 3.17 *(Elekes–Erdős [EE94])*

(a) If $\mathcal{P} \subset \mathbb{R}^2$ can be represented in the complex plane $\mathbb{C}$ as a set of algebraic complex numbers then $S^{(2)}(\mathcal{P},n) \ge c_{\mathcal{P}} n^2$, where $c_{\mathcal{P}}$ depends on $\mathcal{P}$ but not on n.

(b) If s denotes the transcendence degree of $\mathcal{P} \subset \mathbb{C}$ then

$$S^{(2)}(\mathcal{P},n) \ge c \cdot n^{\,2 - \frac{b}{\sqrt[s+1]{\log n}}},$$

where b and c depend on $\mathcal{P}$ but not on n. Thus, for any ε, $S^{(2)}(\mathcal{P},n) \ge n^{2-\varepsilon}$.

There remains one more question: is it true that if a pattern cannot be represented as algebraic complex numbers then $S^{(2)}(\mathcal{P},n)=o(n^2)$? The answer is in the negative, since all triangles can occur a quadratic number of times.

Theorem 3.18 *(Elekes–Erdős [EE94]) For any triangle $\mathcal{T}$ and $n\equiv 1 \mod 3$,*

$$S^{(2)}(\mathcal{T},n)\geq \frac{(n-1)(n+8)}{18}.$$

It was observed by Laczkovich and Ruzsa [LR96] that the order of magnitude depends on the algebraicity of the *cross ratios* — instead of the algebraicity of the *points* themselves.

Theorem 3.19 *(Laczkovich–Ruzsa [LR96]) If the cross ratio of any four points of $\mathcal{P}\subset\mathbb{C}$ is algebraic then $S^{(2)}(\mathcal{P},n)\geq c_{\mathcal{P}}n^2$. Otherwise $S^{(2)}(\mathcal{P},n)\leq B(n)$, where $B(n)$ is a bound independent from $\mathcal{P}$, satisfying $B(n)/n^2\to 0$.*

The structure of the asymptotically extremal configurations is also described for certain patterns in [AEFM02]. An interesting phenomenon appears here. For $\mathcal{E}$ an equilateral triangle and a set $\mathcal{A}$ of n points, if $s^{(2)}(\mathcal{E},\mathcal{A})\geq(1/6+\varepsilon)n^2$, then $\mathcal{A}$ must contain a $k\times k$ portion of a triangular lattice (provided that $n>n_0(k,\varepsilon)$). However, for a smaller positive constant c, it is not difficult to find a set $\mathcal{A}$ of n points located on as few as three straight lines with $s^{(2)}(\mathcal{E},\mathcal{A})\geq cn^2$. This means that, in this case, the asymptotically extremal structure is not stable.

The problem of estimating $s^{(r)}(\mathcal{P},\mathcal{A})$ and $S^{(r)}(\mathcal{P},n)$ in the higher dimensional space $\mathbb{R}^r$ is still wide open. The simplest question is the following.

Problem 3.20 *(Elekes–Erdős [EE94]) Denote the regular r–dimensional simplex by Δ^r. Determine the order of magnitude of $S^{(3)}(\Delta^2,n)$ and $S^{(3)}(\Delta^3,n)$.*

The only estimates known that time were

$$\begin{aligned} cn^2 &\leq S^{(2)}(\Delta^2,n)\leq S^{(3)}(\Delta^2,n)=o(n^3) \quad \text{and} \\ cn^{4/3} &\leq S^{(3)}(\Delta^3,n)=o(n^3). \end{aligned}$$

Here, for equilateral triangles (in the first row), the $o(n^3)$ bound comes from the well-known fact that, among the angles determined by n points of $\mathbb{R}^3$, only $o(n^3)$ can be equal (see [CCEG79]). As for tetrahedra, the lower bound is best demonstrated by a parallelepiped lattice (which even contains at least that many homothetic tetrahedra) while the upper bound comes from the triangle bound and from the fact that $S^{(3)}(\Delta^3,n)\leq\frac{1}{2}S^{(3)}(\Delta^2,n)$.
Recently Akatsu–Tamaki–Tokuyama [ATT98] have improved the upper bounds to $cn^{2.2}$. Ábrego and Fernández-Merchant [AFM02] proved

$$S^{(4)}(\Delta^2,n)\leq S^{(5)}(\Delta^2,n)\leq cn^{3-1/9}.$$

They also mention that, for $r\geq 6$, the order of magnitude of $S^{(r)}(\Delta^2,n)$ is n^3, as shown by a generalization of Lenz' construction (the original can be found in [Erd75]).

Finally, for a cube $\square^3$, a result of Sárközi [Sár61] implies $S^{(3)}(\square^3, n) \geq cn^{5/3}$. The right order of magnitude is still unknown.

As for the number of *homothetic* copies of fixed patterns in the general r–dimensional Euclidean space, the order of magnitude is given by the following result.

Theorem 3.21 *(a) $H^{(r)}(\mathcal{P}, n) = O(n^{1+1/r})$ if $\mathcal{P}$ is not contained in any $r-1$ dimensional subspace;*

(b) This order of magnitude is attained for any pattern where the coordinates of the points are all algebraic; moreover, $H^{(r)}(\mathcal{P}, n) \geq cn^{1+1/r-\varepsilon}$ anyway;

(c) There are patterns for which $H^{(r)}(\mathcal{P}, n) = o(n^{1+1/r})$.

Here (a) and (b) can be found in [EE94] while (c) in [LR96].

3.2.2 The structure of large homothetic subsets.

The one dimensional form of the problem we study in this section is the following.

Question 3.22 *What is the structure of those pairs $\mathcal{P}$, $\mathcal{A}$ for which $h^{(1)}(\mathcal{P}, \mathcal{A})$ is* asymptotically best possible, *i.e.,* $|\mathcal{P}| = t$, $|\mathcal{A}| = n$ *and* $h^{(1)}(\mathcal{P}, \mathcal{A}) \geq c_0 n^2/t$, *for a fixed* $c_0 > 0$*?*

Historically, the first result in this direction was that of Balog-Szemerédi [BS94] who settled the special case of $\mathcal{P} = \{0, 1, 2\}$ as a lemma to prove a conjecture of Erdős on three-term arithmetic progressions. Other fixed patterns have been mentioned in the previous section.

As for the case $t \to \infty$, no structure theorem is known in general. One possible reason for this is that the (asymptotically) extremal configurations for the Szemerédi–Trotter Theorem and its Corollary 1.14 have not been characterized so far. (Perhaps this characterization may not be easy.)

However, in the "upper extreme" case when t is a positive proportion of n, it is possible to describe the asymptotically optimal configurations, i.e., those, for which $H^{(1)}(c_1 n, n) \geq c_2 n$. Of course, one can only describe that portion of $\mathcal{A}$, whose points are covered by at least one homothetic copy of $\mathcal{P}$. (The rest can be arbitrary.) That is why the following result involves

$$\mathcal{A}_{\mathcal{P}} \stackrel{\text{def}}{=} \bigcup \{\mathcal{P}' \subset \mathcal{A}\ ;\ \mathcal{P}' \approx \mathcal{P}\}.$$

The interesting feature of the statement is that *all* elements of $\mathcal{A}_{\mathcal{P}}$ can be accounted for — even those which are contained in as few as *one* copy $\mathcal{P}'$.

For arithmetic and geometric GPs, denote by "$\oplus$" the natural operation: "$+$" and "$\cdot$", respectively. Using this notation, the following structure theorem holds [Ele99c].

Theorem 3.23 *If $\mathcal{P}$, $\mathcal{A} \subset \mathbb{R}$ with $|\mathcal{A}| = n$, $|\mathcal{P}| = c_1 n$ and $h^{(1)}(\mathcal{P}, \mathcal{A}) \geq c_2 n$ then there exists an arithmetic or geometric GP $G \in \mathcal{G}^{d^*, C^* n}$ in $\mathbb{R}$ for which $\mathcal{P}$ is contained in a shifted copy of G while $\mathcal{A}_\mathcal{P}$ is contained in C^* homothetic copies of $G \oplus G$.*

The proof is an easy application of Theorem 3.8, using the functions f_i defined in the proof of Theorem 3.13. This argument (using complex f_i) also gives the following.

Theorem 3.24 *If $\mathcal{P}$, $\mathcal{A} \subset \mathbb{R}^2$ with $|\mathcal{A}| = n$, $|\mathcal{P}| = c_1 n$ and $s^{(2)}(\mathcal{P}, \mathcal{A}) \geq c_2 n$ then there exists an arithmetic or geometric GP $G \in \mathcal{G}^{d^*, C^* n}$ in $\mathbb{C}$ for which $\mathcal{P}$ is contained in a shifted copy of G while $\mathcal{A}_\mathcal{P}$ is contained in C^* similar copies of $G \oplus G$.*

Here again, no structure result is known about similar copies in higher dimensions.

Also, for the homothetic Theorem 3.23, it is not obvious, how to generalize it to higher dimensions and how to define GP's there.

Of course, arithmetic GPs consisting of numbers extend naturally to arithmetic GPs of vectors of $\mathbb{R}^r$ — we just let the differences Δ_i be such vectors. In this generality, an arithmetic GP has two parameters related to dimension: we shall call r the *vector dimension* while d the *progression dimension*. It will still hold that $|\mathcal{G} + \mathcal{G}| \leq 2^d |\mathcal{G}|$ and it also remains true that $\mathcal{G} + \mathcal{G}$ contains $|\mathcal{G}|$ homothetic copies of $\mathcal{G}$.

Geometric GPs do not really extend to $\mathbb{R}^r$. (There is no natural multiplication there.) The best we can do in order to have proper r-dimensional geometric GP-type structures is to place a copy of a one dimensional G on each of some concurrent straight lines which span $\mathbb{R}^r$. It will turn out that no other geometric GP-based configurations can produce many large homothetic subsets. (This phenomenon can be considered as an "implicit evidence" of the intuitive fact that geometric GPs are inherently one-dimensional objects.)

Definition 3.25 Let C be a positive integer. We call $\mathcal{J}$ a *geometric GP–type bunch*, of center $\overline{b} \in \mathbb{R}^r$, located on C straight lines and based upon a geometric $G \subset \mathbb{R}$, if there exist vectors $\overline{v}_1, \ldots, \overline{v}_C \in \mathbb{R}^r$ such that

$$\mathcal{J} = \{\overline{b}\} \cup \bigcup_{i=1}^{C} \{g\overline{v}_i + \overline{b}\,;\ g \in G\}.$$

Also, with the same center $\overline{b}$ and vectors $\overline{v}_i$ but with $G \cdot G$ in place of G above, we shall denote the resulting structure by $\mathcal{J} \cdot \mathcal{J}$. (Thus, in this "product", the lines do not change — just the geometric GPs on them extend.)

Note that $\mathcal{J} \cdot \mathcal{J}$ contains $|\mathcal{J}|$ homothetic copies of $\mathcal{J}$. Of course, $|\mathcal{J} \cdot \mathcal{J}| \leq 2^d |\mathcal{J}|$ still holds.

Theorem 3.26 *(Elekes [Ele99c]) If $\mathcal{P}, \mathcal{A} \subset \mathbb{R}^r$ with $|\mathcal{A}| \le n$, $|\mathcal{P}| \ge c_1 n$, and $h^{(r)}(\mathcal{P}, \mathcal{A}) \ge c_2 n$, then*

(i) either a shifted copy of an arithmetic $G \in \mathcal{G}^{d^*, C^* n}$ contains $\mathcal{P}$, while $\mathcal{A}_{\mathcal{P}}$ is contained in C^* homothetic copies of $\mathcal{G} + \mathcal{G}$;

(ii) or a geometric GP–type bunch $\mathcal{J}$ on C^* lines, based upon a geometric GP $G \in \mathcal{G}^{d^*, C^* n}$, contains $\mathcal{P}$, while $\mathcal{A}_{\mathcal{P}}$ is contained in C^* homothetic copies of $\mathcal{J} \cdot \mathcal{J}$,

where d^* and C^* only depend on c and C but neither on n nor on r.

(Note again the asymmetry between several dimensional arithmetic versus one dimensional geometric GPs; cf. Figure 3.3 below.)

It would be natural to try a proof by induction and everything goes well from $r \ge 2$ upwards. However, no inductive proof is known from $r = 1$ to $r = 2$. Rather, a result on plane sets with few directions (see Theorem 3.29) can be shown using Theorem 3.23 as in [Ele99c] — or directly from Theorem 3.8 — and that result can be generalized to higher dimensions by induction; finally, the high dimensional version will imply Theorem 3.26 (see Corollary 3.35).

3.3 Sets which determine few directions

Definition 3.27 For a finite point set $\mathcal{A} \subset \mathbb{R}^r$, we write

$$D(\mathcal{A}) \stackrel{\text{def}}{=} \#\{\text{directions of segments } \overline{A_1 A_2} \mid A_1, A_2 \in \mathcal{A},\ A_1 \ne A_2\}.$$

We do not distinguish segments $\overline{A_1 A_2}$ and $\overline{A_2 A_1}$; thus two segments have equal directions iff they are parallel.

The study of sets which determine few distinct directions was initiated by Scott [Sco70]. His following conjecture was settled in the affirmative by Ungar [Ung82].

Theorem 3.28 *(Scott's Conjecture, Ungar [Ung82]) For any non–collinear planar point set, $D(\mathcal{A}) \ge |\mathcal{A}| - 1$.*

Sets for which equality holds are called *critical* by Jamison [Jam84] and those with one more directions, i.e., $D(\mathcal{A}) = |\mathcal{A}|$, are *near–critical*. He gives an overview of the known critical and near–critical configurations of the Euclidean plane. Among others, he characterizes those such configurations $\mathcal{A}$ which lie on the union of two or three straight lines. His two basic structures are:

(a) copies of an arithmetic progression on each of two or three parallel lines (with the starting points fitted appropriately) — Jamison calls them "bi-columnar" and "tricolumnar" arrays, respectively; and

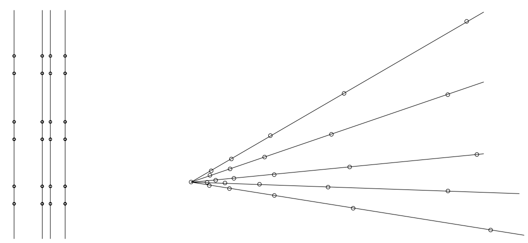

Figure 3.2: Generalized Jamison configurations

(b) copies of a geometric progression on each of the four half–lines of the two coordinate axes, plus the origin — an "exponential cross".

His results can be extended to point sets which determine more than n directions but not more than Cn, provided that a good proportion of the set is collinear. Beyond the aforementioned "Jamison configurations", we define "generalized Jamison configuration" to consist either of C copies of an arithmetic GP, one on each of C parallel straight lines, or a geometric GP–type bunch (see Definition 3.25 and Figure 3.2). Moreover, we write

$$D(\mathcal{A}_1, \mathcal{A}_2) \stackrel{\text{def}}{=} \#\{\text{directions of segments } \overline{A_1 A_2} \mid A_i \in \mathcal{A}_i\}.$$

Theorem 3.29 *Let $C > 1$ be fixed; $\mathcal{A}_1, \mathcal{A}_2 \subset \mathbb{R}^2$ with $n \le |\mathcal{A}_1|, |\mathcal{A}_2| \le Cn$ and l a straight line which contains $\mathcal{A}_1$ but no point of $\mathcal{A}_2$.*
If $D(\mathcal{A}_1, \mathcal{A}_2) \le Cn$ then $\mathcal{A}_1 \cup \mathcal{A}_2$ is contained in a "generalized Jamison configuration".

Proof: We can reduce this statement to Theorem 3.8 by applying a polarity

$$(a, b, c) \quad \leftrightarrow \quad cx + by + az = 0$$

of the projective plane, where the point with projective coordinates (a, b, c) will correspond to the line on the right and vice versa. (This mapping is known to be incidence preserving.) The lines which correspond to the original points of $\mathcal{A}_2$ are to be considered as graphs of linear functions (see [Ele99c]). ∎

Even an uniform statistical version holds. For $E \subset \mathcal{A}_1 \times \mathcal{A}_2$, write

$$D_E(\mathcal{A}_1, \mathcal{A}_2) \stackrel{\text{def}}{=} \#\{\text{directions of segments } \overline{A_1 A_2} \mid (A_1, A_2) \in E\}.$$

Theorem 3.30 *Let $\alpha > 0$ be fixed, $\mathcal{A}_1, \mathcal{A}_2$ as in Theorem 3.29 and $H(\mathcal{A}_1, \mathcal{A}_2, E)$ a bipartite graph with all degrees at least αN. If $D_E(\mathcal{A}_1, \mathcal{A}_2) \le Cn$ then $\mathcal{A}_1 \cup \mathcal{A}_2$ is contained in a bounded number of generalized Jamison configurations.*

There are several examples of point sets which determine at most Cn directions: equidistant points on a circle (or, as their affine image, on an ellipse); appropriate points on a hyperbola or parabola; moreover square lattices will also do the trick. These configurations all arise from conics, except for the last one which contains many collinear points (and so do generalized Jamison configurations). It is natural to believe the following.

Conjecture 3.31 *(Elekes [Ele99b]) For every $C > 0$ there is an $n_0 = n_0(C)$ with the following property.*
If $\mathcal{A} \subset \mathbb{R}^2$ with $|\mathcal{A}| \geq n_0$ and $D(\mathcal{A}) \leq C|\mathcal{A}|$ then $\mathcal{A}$ contains six points of a (possibly degenerate) conic.
As usual, a pair of lines is considered a degenerate conic.

It is very likely that Conjecture 3.31 holds for any number in place of six, for $|\mathcal{A}|$ large enough. It was pointed out by M. Simonovits, that one cannot expect $c^*|\mathcal{A}|$ conconic points in general (as shown by a square lattice.) However, some $|\mathcal{A}|^\alpha$ such points may exist, for a suitable $\alpha = \alpha(C) > 0$. Perhaps even $c^*|\mathcal{A}|$ can be found, provided that $\mathcal{A}$ is the vertex set of a convex polygon.

A weaker version could be the following.

Conjecture 3.32 *If a non-collinear set $\mathcal{A}$ of n points is located on an irreducible algebraic curve of degree r, and $D(\mathcal{A}) \leq Cn$, then the curve must be a conic, provided that $n > n_0(r, C)$.*

In order to support Conjectures 3.31 and 3.32 we mention that, of all polynomial curves $y = p(x)$ (where $p \in \mathbb{R}[x]$), only parabolas can accommodate n non-collinear points with at most Cn directions, provided that n is large enough as compared to the degree of p.

Theorem 3.33 *(Elekes [Ele99b]) Let $r \geq 3$ be an integer and $C > 1$. Then there exists an $n_0 = n_0(r, C)$ with the property that no polynomial curve of degree r (defined by an equation $y = p(x)$) can contain a set of $n > n_0$ points $\mathcal{A} = \{A_1, \ldots, A_n\}$ with $D(\mathcal{A}) \leq Cn$.*

Proof: Consider

$$F(x, y) \stackrel{\text{def}}{=} \frac{p(x) - p(y)}{x - y},$$

the slope of the segment which connects two generic points of the curve. After appropriate simplifications, F becomes a polynomial in two variables.
Let the given points have coordinates $A_i = \big(s_i, p(s_i)\big)$ for $i = 1, \ldots, n$; then F only takes at most Cn distinct values while x and y, independently of each other, range over $\{s_1, \ldots, s_n\}$. Therefore, by Theorem 2.27, F can be written in one of the forms

$$\begin{array}{ll} (1) & F(x, y) = f(g(x) + h(y)); \text{ or} \\ (2) & F(x, y) = f(g(x) \cdot h(y)), \end{array}$$

for suitable polynomials $f, g, h \in \mathbb{R}[z]$.

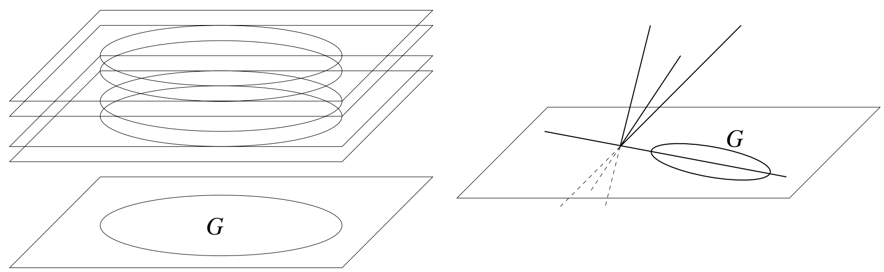

Figure 3.3: Generalized Jamison configurations in three dimensions

Now if $p(x) = a_r x^r + a_{r-1}x^{r-1} + \ldots + a_1 x + a_0$, then the highest (total) degree terms of F are $a_r(x^{r-1} + x^{r-2}y + \ldots + y^{r-1})$. Note that, for $r \geq 3$, there are three or more monomials here and they involve at least one which contains both x and y.

These leading terms must come from the highest power in $f(z)$, with the sum or product of the leading terms of g and h substituted into z, in cases (1) or (2), respectively.

We show that this is impossible.

(1) in case of $F(x, y) = f(g(x) + h(y))$, the coefficients of the highest degree terms on the right hand side come from a binomial expansion and, thus, they cannot be equal — unless $\deg f = 1$. However, in the latter case, no term can involve both x and y simultaneously, a contradiction;

(2) in case of $F(x, y) = f(g(x) \cdot h(y))$, there is just one highest degree term on the right hand side; again a contradiction. ∎

Finally we mention high dimensional sets which determine few directions. In what follows, we consider $\mathbb{R}^r$ to be a special subspace of $\mathbb{R}^k$, for $r < k$:

$$\mathbb{R}^r = \{(x_1, x_2, \ldots, x_k) \ ; \ x_{r+1} = x_{r+2} = \ldots = x_k = 0\} \subset \mathbb{R}^k.$$

Theorem 3.34 *(Elekes [Ele99c]) Let $\mathcal{A}_1, \mathcal{A}_2 \subset \mathbb{R}^k$ with $N \leq |\mathcal{A}_1|, |\mathcal{A}_2| \leq CN$. Assume, moreover, that $\mathcal{A}_1 \subset \mathbb{R}^r$ (for some $r < k$) while $\mathcal{A}_2 \cap \mathbb{R}^r = \emptyset$.*
If $D(\mathcal{A}_1, \mathcal{A}_2) \leq CN$ then

(i) either $\mathcal{A}_1$ is contained in a shifted copy of an r–dimensional arithmetic $G \in \mathcal{G}^{d^*, C^*N}$, where $G \subset \mathbb{R}^r$, while $\mathcal{A}_2$ can be covered with at most C^* shifted copies of G;

(ii) or $\mathcal{A}_1 \cup \mathcal{A}_2$ is contained in a geometric GP–type bunch $\mathcal{J}$, based upon a geometric $G \in \mathcal{G}^{d^*, C^*N}$ where $G \subset \mathbb{R}$ and $\mathcal{J}$ is located on C^* straight lines with center in $\mathbb{R}^r$.

(See Figure 3.3 for the three dimensional cases.)
Outline of the proof: The case $r = 1$, $k = 2$ is Theorem 3.29. Proceed by induction on r and k, of which we only give the details for the case $r = 2$, $k = 3$.

Assume that $\mathcal{A}_1$ is contained in the x–y plane and project all points on a generic vertical plane. Here "generic" means that non-collinear triples are not mapped to collinear points unless the three points lie in a horizontal plane; non-parallel lines are not mapped to parallel lines unless they are horizontal; and three non-concurrent lines are not mapped to concurrent lines unless one of the lines lies in a horizontal plane and the other two intersect at a point of this plane.
Now we use Theorem 3.29 for the projected image and find horizontal or concurrent lines (with common point on the image of the x–y plane) which contain the projected points. In the first case we have suitable horizontal planes and, using Theorem 1.3 and Remark 1.4, also the arithmetic GP can be found. Otherwise, for many general vertical planes (at least two suffice), the projected image can be covered by concurrent lines, hence the intersections of the pre–image planes provide the required bunch which contains the original points (using the multiplicative version of Theorem 1.3). ∎

Corollary 3.35 *Theorem 3.26 holds.*

Proof: Apply in backward direction a high dimensional version of the polarity used in the proof of Theorem 3.29. ∎

3.4 Plane sets with few distinct distances

Throughout this section $\mathbf{e}$ and $\mathbf{f}$ will denote two straight lines of the Euclidean plane and $\mathcal{A} \subset \mathbf{e}$, $\mathcal{B} \subset \mathbf{f}$ two collinear subsets located on these lines. For the set of distances between their points write

$$\mathcal{D}(\mathcal{A},\mathcal{B}) \stackrel{\text{def}}{=} \{\, \overline{AB} \ ; \ A \in \mathcal{A}, \ B \in \mathcal{B}\, \}.$$

These n^2 distances will not be all distinct in general; those which occur several times are, of course, counted just once in $\mathcal{D}(\mathcal{A},\mathcal{B})$.

Erdős posed in [Erd46] the problem of determining the minimal number of distinct distances which can occur among n points of the plane. (This question is still open, see the survey of Székely [Szé02] for more details). The study of the number of distances between two collinear point sets was initiated by Purdy (see the end of Problem 2, page 3 in the excellent collection by Moser and Pach [MP95]). Two examples of such configurations with few distances are the following.

3.36 Example (1) Let $\mathbf{e}$ and $\mathbf{f}$ be parallel while $\mathcal{A}$ and $\mathcal{B}$ two copies of an n–term arithmetic progression on $\mathbf{e}$ and $\mathbf{f}$, respectively. Then $|\mathcal{D}(\mathcal{A},\mathcal{B})| = n-1$;

(2) Let $\mathbf{e}$ and $\mathbf{f}$ be orthogonal e.g., the two axes of a coordinate system while

$$\mathcal{A} = \{(\sqrt{i},0) \ ; \ i = 1\ldots n\}$$
$$\mathcal{B} = \{(0,\sqrt{j}) \ ; \ j = 1\ldots n\}.$$

Then $|\mathcal{D}(\mathcal{A},\mathcal{B})| = 2n-1$, since $\overline{A_iB_j}^2 = i+j \in \{2,3,\ldots,2n\}$ for $A_i = (\sqrt{i},0)$ and $B_j = (0,\sqrt{j})$.

The following result was also conjectured by Purdy and solved in [ER00].

Theorem 3.37 *(Elekes–Rónyai [ER00]) For every $C > 1$ there is a bound $n_0 = n_0(C)$ such that if*

$$|\mathcal{D}(\mathcal{A}, \mathcal{B})| \leq Cn,$$

then $\mathbf{e}$ *and* $\mathbf{f}$ *are parallel or orthogonal, provided that $n > n_0$.*

Proof: Let α be the angle of $\mathbf{e}$ and $\mathbf{f}$, $\lambda = -\cos\alpha$ and consider the polynomial $P(x,y) = x^2 + 2\lambda xy + y^2$. Its graph, by assumption, is n^2–rich in an $n \times n \times Cn$ Cartesian product. We show that this implies $\lambda = 0$ or ± 1, provided that $n > n_0$.

Using Theorem 2.27 and its Corollary 2.29, the quantity

$$q_1(x,y) \stackrel{\text{def}}{=} \frac{\partial F}{\partial x} / \frac{\partial F}{\partial y} = \frac{x + \lambda y}{\lambda x + y}$$

must satisfy

$$q_2(x,y) \stackrel{\text{def}}{=} \frac{\partial^2(\,\log|\,q_1(x,y)|\,)}{\partial x \partial y} = 0,$$

wherever well-defined. Here we have

$$q_2 = \frac{\lambda}{(\lambda x + y)^2} - \frac{\lambda}{(x + \lambda y)^2} = \frac{\lambda(\lambda^2 - 1)(y^2 - x^2)}{(\lambda x + y)^2 (x + \lambda y)^2},$$

which can only vanish if $\lambda = 0$ or ± 1. ∎

Remark 3.38 Even the structure of $\mathcal{A}$ and $\mathcal{B}$ can be described here using Theorem 1.3: they must come from an arithmetic GP or the square root thereof. Also, an uniform statistical version could be stated for $\mathcal{D}_E(\mathcal{A}, \mathcal{B})$, had we defined it in this generality.

It was asked by P. Brass and J. Matousek whether a “gap theorem” holds i.e., if all other angles give essentially more distances. This was answered in the affirmative in [Ele99a].

Theorem 3.39 *There is a positive absolute constant c for which, if* $\mathbf{e}$ *and* $\mathbf{f}$ *are neither parallel nor orthogonal, then*

$$|\mathcal{D}(\mathcal{A}, \mathcal{B})| \geq cn^{5/4}.$$

3.5 Circle grids

The topic of this section is, again, related to point sets which determine few distances.

In what follows we denote by $K(P, r)$ the circle centered at $P \in \mathbb{R}^2$ and of radius r. For three points P_1, P_2, P_3 of the Euclidean plane and any subset

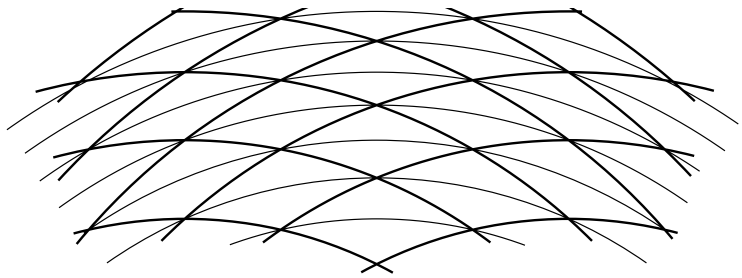

Figure 3.4: Portion of a circle grid with collinear centers.

$\mathcal{R} \subset \mathbb{R}^+$ of the positive reals, the set of points which are covered by all three families $\{K(P_i, r)\ ;\ r \in \mathcal{R}\}$ will be denoted by

$$T(P_1, P_2, P_3, \mathcal{R}) \overset{\text{def}}{=} \{X \in \mathbb{R}^2\ ;\ \forall i = 1, 2, 3\ \exists r_i \in \mathcal{R}\ \text{ such that } X \in K(P_i, r_i)\}$$

and we call them "triple points" for short. Moreover, write

$$t(P_1, P_2, P_3, n) \overset{\text{def}}{=} \max_{|\mathcal{R}|=n} \big|T(P_1, P_2, P_3, \mathcal{R})\big|.$$

We study the following question: *is $t(P_1, P_2, P_3, n) \geq cn^2$ possible for three non-collinear points P_i, a fixed $c > 0$ and infinitely many values of n?*

This problem — apart from the non-collinearity condition — originates from Erdős–Lovász–Vesztergombi [ELV89]. As for collinear triples, there exist "circle grids" like in Figure 3.4 with this large order of magnitude of triple points [Ele95]. The construction is simple: let $P_1 = (-1, 0)$, $P_2 = (0, 0)$, $P_3 = (1, 0)$ and $r_j = \sqrt{j}$.
However, the behavior of collinear and non-collinear P_i is rather different. Actually, t has a strict jump here [ES03].

Theorem 3.40 *(Elekes–E. Szabó [ES03]) If the P_i are non-collinear then, for $n > n_0$,*

$$t(P_1, P_2, P_3, n) \leq n^{1{,}95}.$$

The proof depends on the algebraic equation satisfied by the three distances between a point $X \in \mathbb{R}^2$ and the P_i. The surface described by this (trivariate) polynomial equation must obey Theorem 2.31 and its Corollary 2.32. It can be shown (using some elementary calculus) that this is only possible if the P_i are collinear.

It would be interesting to characterize those configurations which consist of three families of n circles each and the number of triple points is quadratic as a function of n. Even the following is unsolved.

Conjecture 3.41 *(L. Székely, unpublished) If $F^{(1)}$, $F^{(2)}$, $F^{(3)}$ are three families of n unit circles each with the property that, for $i = 1, 2, 3$, the circles of $F^{(i)}$ pass through a common point $X^{(i)}$, then the number of triple points is $o(n^2)$.*

Bibliography

[AEFM02] Bernardo Ábrego, György Elekes, and Silvia Fernández-Merchant. Structural results for planar sets with many similar subsets. *Combinatorica*, page to appear, 2002.

[AFM02] Bernardo Ábrego and Silvia Fernández-Merchant. On the maximum number of equilateral triangles I–II. *Discrete and Computational Geometry*, 23:129–135, 2000 and part II., to appear in 2002.

[ATT98] Tatsuya Akutsu, Hishao Tamaki, and Takeshi Tokuyama. Distribution of distances and triangles ... *Discrete and Computational Geometry*, 20:307–331, 1998.

[Bec83] József Beck. On the lattice property of the plane and some problems of Dirac, Motzkin and Erdős. *Combinatorica*, 3 (3-4):281–297, 1983.

[Bil99] Yurii Bilu. Structure of sets with small sumset. *Asterisque, SMF*, 258:77–108, 1999.

[BS94] Antal Balog and Endre Szemerédi. A statistical theorem of set addition. *Combinatorica*, 14:263–268, 1994.

[CCEG79] John H. Conway, Hallard T. Croft, Paul Erdős, and Michael J.T. Guy. On the distribution of angles .... *Journal of London Math. Soc.*, second series, 19:137–143, 1979.

[CEG+90] Kenneth Clarkson, Herbert Edelsbrunner, Leo Guibas, Micha Sharir, and Emo Welzl. Combinatorial complexity bounds for arrangements of curves and surfaces. *Discrete and Computational Geometry*, 5:99–106, 1990.

[EAH83] Paul Erdős, László Alpár, and Gábor Halász, editors. *Studies in Pure Mathematics; To the memory of Paul Turán*. Akadémiai Kiadó — Birkhauser Verlag, Budapest — Basel–Boston, Mass., 1983.

[EE94] György Elekes and Paul Erdős. Similar configurations and pseudo grids. In *Intuitive geometry. Proceedings of the 3rd international conference held in Szeged, Hungary, from 2 to 7 September, 1991.*

Amsterdam: North-Holland, Coll. Math. Soc. János Bolyai, 63, pages 85–104, 1994.

[EK01] György Elekes and Zoltán Király. On combinatorics of projective mappings. *Journal of Algebraic Combinatorics*, 14:183–197, 2001.

[Ele95] György Elekes. Circle grids and bipartite graphs of distances. *Combinatorica*, 15:167–174, 1995.

[Ele97a] György Elekes. On linear combinatorics I. *Combinatorica*, 17 (4):447–458, 1997.

[Ele97b] György Elekes. On the number of sums and products. *Acta Arithmetica*, LXXXI.4:365–367, 1997.

[Ele98a] György Elekes. On linear combinatorics II. *Combinatorica*, 18 (1):13–25, 1998.

[Ele98b] György Elekes. A problem on polynomials. *Discrete and Computational Geometry*, 19:383–389, 1998.

[Ele99a] György Elekes. A note on the number of distinct distances. *Periodica Math. Hung.*, 38(3):173–177, 1999.

[Ele99b] György Elekes. On linear combinatorics III. *Combinatorica*, 19 (1):43–53, 1999.

[Ele99c] György Elekes. On the structure of large homothetic subsets. *DIMACS series in Discrete Mathematics*, 49:101–111, 1999.

[ELV89] Paul Erdős, László Lovász, and Katalin Vesztergombi. On graphs of large distances. *Discrete and Computational Geometry*, 4:541–549, 1989.

[ENR99] György Elekes, Melvyn B. Nathanson, and Imre Z. Ruzsa. Convexity and sumsets. *Journal of Number Theory*, 83:194–201, 1999.

[EP95] Paul Erdős and George Purdy. *Extremal problems in combinatorial geometry.*, pages 809–874. Elsevier (North-Holland), Amsterdam, 1995.

[ER00] György Elekes and Lajos Rónyai. A combinatorial problem on polynomials and rational functions. *Journal of Combinatorial Theory, series A*, 89:1–20, 2000.

[ER03a] György Elekes and Imre Z. Ruzsa. The number of sums and products. page in preparation, 2003.

[ER03b] György Elekes and Imre Z. Ruzsa. Sets with few sums along a graph. page in preparation, 2003.

[Erd46] Paul Erdős. On sets of distances of n points. *Am. Math. Monthly*, 53:248–250, 1946.

[Erd75] Paul Erdős. On some problems of elementary and combinatorial geometry. *Annali di Matematica Pura Applicata Serie Quarta*, 103:99–108, 1975.

[ES83] Paul Erdős and Endre Szemerédi. *On sums and products of integers*, pages 213–218. In Erdős et al. [EAH83], 1983.

[ES03] György Elekes and Endre Szabó. Triple points of circle grids. *Combinatorica*, to appear in 2003.

[For98] Kevin Ford. Sums and products from a finite set of real numbers. *Ramanujan J.*, 2(1–2):59–66, 1998.

[Fre66] Gregory A Freiman. *Foundations of a Structural Theory of Set Addition (in Russian).* Kazan Gos. Ped. Inst., Kazan, 1966.

[Fre73] Gregory A Freiman. *Foundations of a Structural Theory of Set Addition, Translation of Mathematical Monographs vol. 37.* Amer. Math. Soc., Providence, R.I., USA, 1973.

[GN96] Ron Graham and Jaroslav Nešetřil, editors. *The Mathematics of Paul Erdős.* Springer, 1996.

[Heg86] Norbert Hegyvári. On consecutive sums in sequences. *Acta Math. Acad. Sci. Hungar.*, 48:193–200, 1986.

[Jam84] Robert E. Jamison. Planar configurations which determine few slopes. *Geometriae Dedicata*, 16:17–34, 1984.

[LR96] Miklós Laczkovich and Imre Z Ruzsa. *The Number of Homothetic Subsets.* In Graham and Nešetřil [GN96], 1996.

[MP95] William Moser and János Pach. *Research Problems in Combinatorial Geometry*, Mimeographed notes, McGill University, Montreal, 1995.

[Nat97] Melvyn B. Nathanson. On sums and products of integers. *Proc. Am. Math. Soc.*, 125 (1-2):9–16, 1997.

[PS90] János Pach and Micha Sharir. Repeated angles in the plane and related problems. *Journal of Combinatorial Theory, series A*, 59:12–22, 1990.

[PS98] János Pach and Micha Sharir. On the number of incidences between points and curves. *Combinatorics, Probability and Computing*, 7:121–127, 1998.

[Ruz92] Imre Z Ruzsa. Arithmetical progressions and the number of sums. *Periodica Math. Hung.*, 25:105–111, 1992.

[Ruz94] Imre Z Ruzsa. Generalized arithmetic progressions and sum sets. *Acta Math. Sci. Hung.*, 65:379–388, 1994.

[Sár61] András Sárközi. Lattice cubes in 3-space. *Mat. Lapok*, 12:232–245, 1961.

[Sco70] Paul R. Scott. On the sets of directions determined by n points. *Am. Math. Monthly*, 77:502–505, 1970.

[ST83] Endre Szemerédi and W. T. Trotter Jr. Extremal problems in Discrete Geometry. *Combinatorica*, 3 (3–4):381–392, 1983.

[Sza01] Endre Szabó. How to find groups? *in preparation*, 2001.

[Sze75] Endre Szemerédi. On sets of integers containing no k elements in arithmetic progression. *Acta Arithmetica*, 27:199–245, 1975.

[Sze78] Endre Szemerédi. Regular partitions of graphs. *Colloques Internationaux C.N.R.S.: Problèmes Combinatoires et Théorie des Graphes*, No 260:399–401, 1978.

[Szé97] László A Székely. Crossing numbers and hard Erdős problems in discrete geometry. *Combinatorics, Probability and Computing*, 6,No.3:353–358, 1997.

[Szé02] László A Székely. Erdős on unit distances and the Szemerédi–Trotter theorem. In *this volume*, page ??, 2002.

[Tót03] Csaba Tóth. The Szemerédi–Trotter theorem in the complex plane. *Combinatorica*, (submitted), 2003.

[Ung82] Peter Ungar. $2n$ non-collinear points determine at least $2n$ directions. *Journal of Combinatorial Theory*, 33:343–347, 1982.