SUMS AND PRODUCTS ALONG SPARSE GRAPHS

NOGA ALON, OMER ANGEL, ITAI BENJAMINI, AND EYAL LUBETZKY

Abstract. In their seminal paper from 1983, Erdős and Szemerédi showed that any n distinct integers induce either $n^{1+\varepsilon}$ distinct sums of pairs or that many distinct products, and conjectured a lower bound of $n^{2-o(1)}$. They further proposed a generalization of this problem, in which the sums and products are taken along the edges of a given graph G on n labeled vertices. They conjectured a version of the sum-product theorem for general graphs that have at least $n^{1+\varepsilon}$ edges.

In this work, we consider sum-product theorems for sparse graphs, and show that this problem has important consequences already when G is a matching (i.e., $n/2$ disjoint edges): Any lower bound of the form $n^{1/2+\delta}$ for its sum-product over the integers implies a lower bound of $n^{1+\delta}$ for the original Erdős-Szemerédi problem.

In contrast, over the reals the minimal sum-product for the matching is $\Theta(\sqrt{n})$, hence this approach has the potential of achieving lower bounds specialized to the integers. We proceed to give lower and upper bounds for this problem in different settings. In addition, we provide tight bounds for sums along expanders.

A key element in our proofs is a reduction from the sum-product of a matching to the maximum number of translates of a set of integers into the perfect squares. This problem was originally studied by Euler, and we obtain a stronger form of Euler's result using elliptic curve analysis.

1. Introduction

1.1. **Sums and products.** Let A be a set of elements of some ring R. The *sum-set* of A, denoted by $A+A$, and the *product-set* of A, denoted by $A\times A$, are defined to be

$$A + A \stackrel{\triangle}{=} \{x + y : x, y \in A\} , \qquad A \times A \stackrel{\triangle}{=} \{x \cdot y : x, y \in A\} .$$

The *sum-product phenomenon* states that, in various settings, every set A has either a "large" sum-set or a "large" product-set. The intensive study of this area was pioneered by Erdős and Szemerédi in their celebrated paper [21] from 1983, studying sum-products over the integers. They showed that for some fixed $\varepsilon > 0$, any set $A \subset \mathbb{Z}$ has $\max\{|A + A|, |A \times A|\} \geq |A|^{1+\varepsilon}$, and conjectured that in fact ε can be taken arbitrarily close to 1.

Research of N. Alon was supported in part by a USA Israeli BSF grant, by a grant from the Israel Science Foundation, by an ERC Advanced Grant and by the Hermann Minkowski Minerva Center for Geometry at Tel Aviv University.

Research of O. Angel supported by NSERC and the University of Toronto.

The Erdős-Szemerédi sum-product problem for a set $A \subset \mathbb{Z}$ remains open, despite the considerable amount of attention it has received. The value of ε in the above statement was improved by Nathanson [30] to $\frac{1}{31}$, by Ford [23] to $\frac{1}{15}$ and by Chen [16] to $\frac{1}{5}$. In 1997, a beautiful proof of Elekes [18] yielded a version of the sums and products over the reals (where the exponent is also believed to be $2 - o(1)$) with $\varepsilon = \frac{1}{4}$, via an elegant application of the Szemerédi-Trotter Theorem. Following this approach, Solymosi [32, 33] improved the bound on ε for sums and products over $\mathbb{R}$ to $\frac{3}{11} - o(1)$ and finally to $\frac{1}{3} - o(1)$. For more details on this problem and related results, cf. [10, 12–15, 19] as well as [36] and the references therein.

Different variants of the sum-product problem were the focus of extensive research in the past decade, with numerous applications in Analysis, Combinatorics, Computer Science, Geometry, Group Theory. Most notable is the version of the sum-product theorem for finite fields [7], where one must add a restriction that $A \subset \mathbb{F}_p$ is not too large (e.g., almost all of $\mathbb{F}_p$) nor too small (e.g., a subfield of $\mathbb{F}_p$). See, e.g., [4, 5, 24] for further information on sum-product theorems over finite fields and their applications, and also [6, 35] for such theorems over more general rings.

1.2. **Sums and products along a graph.** In their aforementioned paper [21], Erdős and Szemerédi introduced the following generalization of the sum-product problem, where an underlying geometry (in the form of a graph on n labeled vertices) restricts the set of the pairs used to produce the sums and products. Formally, we let the sum-product with respect to a ring R be the following graph parameter:

Definition (Sum-product of a graph)**.** *Let $G = (V, E)$ be a graph. Given $A = (a_u)_{u \in V}$, an injective map of V into some ring R, define the sum-set $A \overset{G}{+} A$ and the product-set $A \overset{G}{\times} A$ as follows:*

$$A \overset{G}{+} A \triangleq \{a_u + a_v : uv \in E\} \ , \qquad A \overset{G}{\times} A \triangleq \{a_u \cdot a_v : uv \in E\} \ .$$

The sum-product *of G over R, denoted by $\mathtt{SP}_R(G)$, is the smallest possible value of $\max\{|A \overset{G}{+} A| \ , \ |A \overset{G}{\times} A|\}$ over all injections $A : V \to R$.*

In other words, each ordered set $A \subset R$ of cardinality $|V(G)|$ is associated with a sum-set and a product-set according to G as follows: The elements of A correspond to the vertices, and we only consider sums and products along the edges. Thus the original sum-product problem of Erdős-Szemerédi corresponds to $\mathtt{SP}_{\mathbb{Z}}(K_n)$, where K_n is the complete graph on n vertices.

In the above notation, Erdős and Szemerédi conjectured the following:

Conjecture (Erdős-Szemerédi [21]). *For all $\alpha, \varepsilon > 0$ and every sufficiently large n, if $G = (V, E)$ is a graph on n vertices satisfying $|E| \geq n^{1+\alpha}$ then*

$$\mathtt{SP}_{\mathbb{Z}}(G) \geq |E|^{1-\varepsilon} \ . \tag{1.1}$$

Erdős and Szemerédi note that the above conjecture is likely to hold also over the reals. However, for sparse graphs (graphs that contain $O(|V|)$ edges) there is a fundamental difference between the sum-product behavior over the integers and the reals. As stated in [21], Erdős had originally thought that (1.1) also holds when G is a graph on n vertices with at least cn edges for some $c > 0$. It was then shown by A. Rubin that the analogue of (1.1) for sparse graphs does not hold over $\mathbb{R}$, yet the question of whether or not it holds over $\mathbb{Z}$ remains open. See [11], where the author relates this question of Erdős to a famous conjecture of W. Rudin [31].

Clearly, for any ring R and graph $G = (V, E)$ we have that $\mathtt{SP}_R(G) \leq |E|$. We will mostly be interested in a choice of either $\mathbb{Z}$ or $\mathbb{R}$ for the ring R, and as we later state, these satisfy

$$\mathtt{SP}_{\mathbb{Z}}(G) \geq \mathtt{SP}_{\mathbb{R}}(G) \geq \sqrt{|E|} \quad \text{for any graph } G = (V, E) \ . \tag{1.2}$$

Thus, when G is a sparse graph with n edges, the order of $\mathtt{SP}_{\mathbb{Z}}(G)$ is between $\sqrt{n}$ and n. Our main focus in this paper is the case where G is a matching, i.e., a graph consisting of disjoint edges. The sum-product problem corresponding to this graph over the integers is already challenging, and as the next theorem demonstrates, it has an immediate implication for the original Erdős-Szemerédi problem:

Theorem 1. *Let M be a matching of size n. The following holds:*

$$\mathtt{SP}_{\mathbb{Z}}(M) = O\left(\mathtt{SP}_{\mathbb{Z}}(K_n)/\sqrt{n}\right) \ , \tag{1.3}$$

$$\mathtt{SP}_{\mathbb{Z}}(M) \leq n/\log(n)^{\varepsilon} \ \text{ for some } \varepsilon > 0 \ . \tag{1.4}$$

In particular, if the sum-product of M over $\mathbb{Z}$ is $\Omega(n^{1/2+\delta})$ for some $\delta > 0$, then every n-element subset $A \subset \mathbb{Z}$ satisfies $\max\{|A+A|, |A\times A|\} \geq \Omega(n^{1+\delta})$.

Note that (1.3) translates any nontrivial lower bound in the sparse setting to one for the dense setting. In particular, the best-possible lower bound of $n^{1-o(1)}$ for a matching of size n would imply that $\mathtt{SP}_{\mathbb{Z}}(K_n) \geq n^{3/2-o(1)}$, improving upon the currently best known bound of $n^{4/3-o(1)}$. Moreover, (1.4) points out a relation between the upper bounds in these two settings: Just as in the sum-product problem for the complete graph, the ε in the upper bound of $n^{1-\varepsilon}$ for the sum-product of a matching is essential.

1.3. **Sum-products and Euler's problem on translates of squares.** Our next main result reduces the problem of obtaining a lower bound on $\mathtt{SP}_{\mathbb{Z}}(M)$, the sum-product of a matching over the integers, to bounding the

maximum possible number of translates of a set of integers into the set of perfect squares, denoted by $\textsc{Squares} \triangleq \{z^2 : z \in \mathbb{Z}\}$. A special case of this problem was studied by Euler [22], and as we soon state, this problem fully captures the notion of a nontrivial lower bound on $\mathtt{SP}_{\mathbb{Z}}(M)$.

Definition (Square translates)**.** *Let F_k denote the maximum number of translates of a set A that are contained within the set of perfect squares, taken over every k-element subset $A \subset \mathbb{Z}$:*

$$F_k \triangleq \max_{A \subset \mathbb{Z}\ ,\ |A|=k} \#\{x : A + x \subset \textsc{Squares}\}\ . \tag{1.5}$$

Further let $F_k(n)$ denote this maximum with the added constraint that $|a| \le n$ for all $a \in A$:

$$F_k(n) \triangleq \max_{\substack{A \subset \{-n,\ldots,n\} \\ |A|=k}} \#\{x \in \{-n,\ldots,n\}\ :\ A + x \subset \textsc{Squares}\}\ . \tag{1.6}$$

Recall that $\sqrt{|E|}$ is a lower bound on $\mathtt{SP}_{\mathbb{R}}(G)$ for any graph $G = (V, E)$. The following theorem shows that, while this bound is tight for a matching M over $\mathbb{R}$, there is an equivalence between a nontrivial lower bound for $\mathtt{SP}_{\mathbb{Z}}(M)$ and a uniform upper bound on F_k for some integer k.

Theorem 2. *Let M be a matching of size n. The following holds:*

1. *We have $\mathtt{SP}_{\mathbb{R}}(M) = \lceil \sqrt{n}\ \rceil$.*
2. *If $F_k = \infty$ for any k, then $\mathtt{SP}_{\mathbb{Z}}(M) = \lceil \sqrt{n}\ \rceil$ for all n.*
3. *Conversely, if $F_k < \infty$ for some k, then $\mathtt{SP}_{\mathbb{Z}}(M) = \Omega\big(n^{k/(2k-1)}\big)$. Furthermore, for any $t = t(n)$ and any $A \subset \{-t,\ldots,t\}$ we have*

$$\max\{|A \overset{M}{+} A|\ ,\ |A \overset{M}{\times} A|\} = \Omega\big(n^{k/(2k-1)}[F_k(4t^2)]^{-1/(2k-1)}\big)\ .$$

As a corollary of the above theorem, we obtain a nontrivial lower bound of $n^{2/3}$ in case the elements of A are all polynomial in n.

Corollary 3. *Let M be a matching of size n, and A be a mapping of its vertices to distinct integers, such that $|a_v| \le n^{O(1)}$ for all $v \in M$. Then $\max\{|A \overset{M}{+} A|\ ,\ |A \overset{M}{\times} A|\} \ge n^{2/3-o(1)}$.*

In fact, the statement of Corollary 3 holds as long as $|a| \le n^{c \log\log n}$ for all $a \in A$ and some constant $c > 0$.

Euler [22, Chapter 2.XIV, Article 223] studied translates of sets of three integers into the set of perfect squares, corresponding to the parameter F_3. He provided examples where nontrivial translates exist, and showed how to find such translates in general if they are known to exist. In Section 4 we extend Euler's results, and use elliptic curves to construct sets of three integers for which there are infinitely many such translates ($F_3 = \infty$).

The parameter F_4, together with the results of Theorem 2, enables us to deduce another lower bound on $\mathtt{SP}_{\mathbb{Z}}(M)$, assuming a major conjecture in arithmetic geometry — the Bombieri-Lang conjecture for rational points on varieties of general type.

Corollary 4. *Assume the Bombieri-Lang conjecture, and let M denote a matching of size n. Then $\mathtt{SP}_{\mathbb{Z}}(M) = \Omega(n^{4/7})$.*

Notice that, combining the above lower bound (assuming the Bombieri-Lang conjecture) with Theorem 1 yields a lower bound of $\varepsilon = \frac{1}{14}$ for the sum-product of the complete graph. While this does not improve the best known exponent for $\mathtt{SP}_{\mathbb{Z}}(K_n)$, as we later state, it does improve all known sum-product bounds for graphs with slightly smaller degrees (e.g., of average degree $n^{1-\delta}$ for certain small $\delta > 0$).

See, for instance, [8, 17] for other implications of the Bombieri-Lang conjecture on problems involving the perfect squares.

1.4. **Sums along expander graphs.** Up till now, we considered sums and products along graphs, where each of the sum-set and product-set could be small (yet they could not both be small at the same time): For instance, the sum-set along a matching can consist of a single element. The final part of this paper investigates the smallest possible size of the sum-set along other underlying geometries. Note that this problem is trivial for dense graphs, as the maximal degree of a graph G is clearly a lower bound on $|A \overset{G}{+} A|$.

As we later explain, a straightforward extension of one of our basic arguments for the sum-product of a matching gives that, for instance, if G is a vertex-transitive graph on n vertices with odd-girth ℓ then

$$|A \overset{G}{+} A| \geq n^{1/\ell} \qquad \text{for any } A \subset \mathbb{R} \,. \tag{1.7}$$

In particular, when G is a disjoint union of triangles, $|A \overset{G}{+} A| \geq n^{1/3}$, and we later show that this bound is tight. It is natural to assume that the sum-set along G should be forced to be larger if G had, in some sense, stronger interactions between its vertices, and specifically, if the graph is an *expander* (defined below). Surprisingly, our results show that the sum-set along an n-vertex expander can be of size only $O(\log n)$, and this is best possible.

The *conductance* of a graph $G = (V, E)$, denoted by $\Phi(G)$, is defined as

$$\Phi(G) \triangleq \min_{\substack{S \subset V \\ \mathrm{vol}(S) \neq 0}} \frac{e(S, \overline{S})}{\min\{\mathrm{vol}(S)\,,\, \mathrm{vol}(\overline{S})\}} \,, \tag{1.8}$$

where $\overline{A}$, $\mathrm{vol}(A)$ and $e(A, B)$ denote the complement of A, its volume (the sum of its degrees) and the number of edges between A and B respectively.

For a real $\delta > 0$ and a graph G without isolated vertices, we say that G is a δ-(edge)-expander if $\Phi(G) > \delta$. For further information on these objects and their numerous applications, cf., e.g., [26].

The next theorem characterizes the smallest possible cardinality of the sum-set of $A \subset \mathbb{Z}$ along an expander.

Theorem 5. *For any $0 < \delta < \frac{1}{2}$ there exist constants $C, c > 0$ such that:*

1. If G is a δ-expander on n vertices then

$$|A \overset{G}{+} A| \geq c \log n \qquad \text{for any } A \subset \mathbb{Z} \text{ , } |A| = n \text{ .}$$

2. There exists a regular δ-expander G on n vertices such that

$$|A \overset{G}{+} A| \leq C \log n \qquad \text{for } A = \{1, 2, \ldots, n\} \text{ .}$$

1.5. **Organization.** The rest of this paper is organized as follows. Section 2 contains the proof of Theorem 1, which provides upper bounds on $\mathtt{SP}_{\mathbb{Z}}(M)$ and relates it to $\mathtt{SP}_{\mathbb{Z}}(K_n)$. In Section 3 we prove Theorem 2, which gives lower bounds for $\mathtt{SP}_{\mathbb{Z}}(M)$ in terms of the parameters F_k (translates of k integers into the squares). Section 4 focuses on this problem of translates of a set into the squares: We first analyze F_3, and extend Euler's result using elliptic curves. We then discuss F_4 and its implication on the sum-product of the matchings. In Section 5 we study sum-sets along other geometries, and prove Theorem 5, which establishes tight bounds for sums along expanders. The final section, Section 6, contains concluding remarks and open problems.

2. Upper bounds for the sum-product of a matching

In this section, we prove Theorem 1, which provides upper bounds for the sum-product of a matching over the integers. Throughout this section, let M denote a matching consisting of n disjoint edges. We begin with a simple lemma.

Lemma 2.1. *Suppose the edges of a graph $G = (V, E)$ are properly coloured with k colours, and let Δ be the maximal degree in G. Then G contains a matching of at least $|E|/(4\Delta)$ edges involving at most $k/(2\Delta)$ colours.*

Proof. Repeatedly select all edges of the most used colour and delete all edges adjacent to them, until at least $|E|/(4\Delta)$ edges have been selected. Up to that point at most $|E|/2$ edges are deleted, so at each step at least $|E|/(2k)$ edges are selected. Thus the number of steps is at most $2k/(4\Delta)$. ∎

2.1. **A sub-linear upper bound: proof of** (1.4)**.** The desired upper bound given in inequality (1.4) is equivalent to the following statement: There is a fixed $\varepsilon > 0$ so that, for every sufficiently large n, there exists an ordered set A of $2n$ distinct integers satisfying

$$|A \overset{M}{+} A| \leq \frac{n}{(\log n)^{\varepsilon}} \quad \text{and} \quad |A \overset{M}{\times} A| \leq \frac{n}{(\log n)^{\varepsilon}} \ .$$

We need the following result of Erdős [20].

Lemma 2.2. *There is a fixed $\varepsilon > 0$ such that for every sufficiently large N, the number of integers which are the product of two integers, each no greater than N, is smaller than $N^2/[128(\log N)^{2\varepsilon}]$.*

Let $N = 16n$ be a large integer, let I be the interval of all integers in

$$\Big[N - \frac{N}{32(\log N)^{\varepsilon}}, N + \frac{N}{32(\log N)^{\varepsilon}}\Big) \ ,$$

and let $G_0 = (V, E)$ be the graph on the set of vertices $\{1, 2, \ldots, N\}$ in which i and j are connected iff $i + j \in I$. Note that every vertex of G_0 has degree at least $d/2$ and at most d, where $d = \frac{N}{16(\log N)^{\varepsilon}}$, and in particular $|E| \geq Nd/4$. Assign each edge of G_0 a colour according to the product of its endpoints, and note that G_0 is now properly coloured with at most $k = \frac{N^2}{128(\log N)^{2\varepsilon}}$ colours, due to Lemma 2.2.

By Lemma 2.1, there is a matching in G_0 consisting of at least $|E|/(4d) \geq N/16$ edges which are coloured by at most $k/(2d) = \frac{N}{16(\log N)^{\varepsilon}}$ distinct colours. Thus we have found $N/16$ disjoint pairs of integers with at most $\frac{N}{16(\log N)^{\varepsilon}}$ distinct sums and as many products. ■

2.2. **From matchings to dense graphs: proof of** (1.3)**.** We prove a stronger statement than the one given in Theorem 1, and bound $\mathtt{SP}_{\mathbb{Z}}(M)$ in terms of the sum-product of any sufficiently dense graph (rather than the complete graph). This is formalized by the following theorem.

Theorem 2.3. *Let $G = (V, E)$ be a graph on N vertices with maximum degree at most $D \geq 10(\log N)^2$ and average degree at least d, such that N is large enough and $d \geq 5\sqrt{D}$. Suppose that $S \leq \frac{Nd}{16\sqrt{D}}$ and that $A = \{a_v : v \in V\}$ are distinct integers satisfying*

$$|A \overset{G}{+} A| \leq S \quad \text{and} \quad |A \overset{G}{\times} A| \leq S \ .$$

Then there is a matching M of $n = \frac{Nd}{32D}$ edges in G, so that

$$|A \overset{M}{+} A| \leq \frac{2S}{\sqrt{D}} \quad \text{and} \quad |A \overset{M}{\times} A| \leq \frac{2S}{\sqrt{D}} \ .$$

Proof. Fix $p = \frac{1}{\sqrt{D}}$, and let R be a random subset of $A \overset{G}{+} A$ obtained by picking every element $s \in A \overset{G}{+} A$, randomly and independently, with probability p. Let H be the spanning subgraph of G consisting of all edges uv so that $a_u + a_v \in R$.

By standard large deviation estimates for binomial distributions, with high probability the total size of R is smaller than $2Sp = 2S/\sqrt{D}$, and the maximum degree in H is smaller than $2Dp = 2\sqrt{D}$. (Note that the degree of each vertex in H is indeed a binomial random variable, as each edge of G incident with the vertex remains in H randomly and independently with probability p.) Moreover, we claim that the number of edges of H is at least $\frac{Nd}{4\sqrt{D}}$ with probability at least $1/5$, hence with positive probability H satisfies all of these conditions. To see this last claim, let $m_i \le N/2$ be the number of edges in G with sum i. Then

$$\operatorname{Var}|E(H)| = p(1-p)\sum_i m_i^2 < p\sum_i \frac{N}{2} m_i = \frac{N|E(G)|}{2\sqrt{D}}.$$

By Chebyshev's inequality, $\mathbb{P}(|E(H)| < p|E(G)|/2) \le \frac{4\sqrt{D}}{d} \le 4/5$, implying the claim.

Fix a choice of H for which the above conditions hold. Assign to each edge $e = uv$ of H, a colour given by the numbers associated to its endpoints: $a_u \cdot a_v$. Note that this is a proper colouring of H with at most S colours.

Applying Lemma 2.1 to H, yields a matching in H consisting of at least $\frac{Nd}{32D}$ edges, with at most $\frac{S}{4\sqrt{D}}$ colours. This matching gives $\frac{Nd}{32D}$ pairs of integers with at most $\frac{2S}{\sqrt{D}}$ sums and $\frac{S}{4\sqrt{D}}$ products, as required. ∎

Remark. The assumption $D \ge 10(\log N)^2$ can be easily relaxed, as it is not essential that all degrees in H will be at most $2\sqrt{D}$, it suffices to ensure that no set of $\frac{Nd}{16D}$ vertices captures more than $\frac{Nd}{8\sqrt{D}}$ edges. It is also not difficult to prove a version of the above theorem starting with the assumption that $|A \overset{G}{+} A| \le S$ and $|A \overset{G}{\times} A| \le T$, where S and T are not necessarily equal. Similarly, the requirement $d \ge 5\sqrt{D}$ can be relaxed (if one accepts larger sum and product sets) by splitting the edges with a given sum into subsets for the construction of H.

An immediate application of the last theorem is the following.

Corollary 2.4. *If G is a D-regular graph on N vertices with $D \ge 10(\log N)^2$ and there exists a set of N distinct integers A so that $|A \overset{G}{+} A| \le S$ and $|A \overset{G}{\times} A| \le S$, then there is a matching M of size $n = \frac{N}{32}$ and a set B of $2n$ distinct integers so that $|B \overset{M}{+} B| \le \frac{2S}{\sqrt{D}}$ and $|B \overset{M}{\times} B| \le \frac{2S}{\sqrt{D}}$.*

In particular, for G being a complete graph this implies that if there is a set A of N distinct integers so that $|A+A| \leq S$ and $|A \times A| \leq S$, then there is a matching M of size $\Omega(N)$ and a set B of $2|M|$ distinct integers so that

$$|B \overset{M}{+} B| \leq O(S/\sqrt{N}) \quad \text{and} \quad |B \overset{M}{\times} B| \leq O(S/\sqrt{N}) \ .$$

This proves (1.3), and completes the proof of Theorem 1. ■

3. Lower bounds for matchings and translates into squares

In this section, we prove Theorem 2, which relates lower bounds for the sum-product of the matching over the integers to the square-translates problem defined in the introduction.

Proof of Theorem 2. We first elaborate on inequality (1.2), which stated that any graph $G = (V, E)$ satisfies $\mathtt{SP}_{\mathbb{Z}}(G) \geq \mathtt{SP}_{\mathbb{R}}(G) \geq \sqrt{|E|}$. This follows immediately from the next simple observation:

Observation 3.1. *Let $\mathbb{F}$ be a field and $G = (V, E)$. Then any injection $A : V \to \mathbb{F}$ satisfies $|A \overset{G}{+} A| \cdot |A \overset{G}{\times} A| \geq |E|$.*

Indeed, since any quadratic polynomial over $\mathbb{F}$ has at most 2 roots, any two elements $\{x, y\} \in \mathbb{F}$ are uniquely determined by their sum $s = x+y$ and their product $p = xy$. In particular, when the characteristic of $\mathbb{F}$ is other than 2,

$$\{x, y\} = \left\{ \frac{s \pm \sqrt{s^2 - 4p}}{2} \right\} \ . \tag{3.1}$$

The above bound is tight (up to rounding) whenever it is possible to take a square-root of elements in $\mathbb{F}$ (in fact, a slightly weaker condition already suffices). To demonstrates this over $\mathbb{R}$ and any $n = m^2$ for $m \geq 1$, let X denote a set of m reals chosen uniformly from the interval $[5, 6]$. Clearly, every pair $s, p \in X$ satisfies

$$s^2 - 4p \geq 25 - 24 > 0 \ ,$$

and furthermore, with probability 1 there exist $2n$ distinct solutions to the m^2 equations of the form (3.1), as s, p range over all possible values in X. This shows that

$$\mathtt{SP}_{\mathbb{R}}(M) = \lceil \sqrt{n} \rceil \ ,$$

even with the added constraint $|A \overset{G}{+} A| = |A \overset{G}{\times} A|$. Next, we wish to relate $\mathtt{SP}_{\mathbb{Z}}(M)$ to the parameters F_k, defined in (1.5).

To prove Item (2) of the theorem, assume that indeed $F_k = \infty$ for all k. We need to show that if M is a matching comprising $n = m^2$ edges, then

there exists a set A of $2n$ distinct integers such that both $|A \overset{M}{+} A| = m$ and $|A \overset{M}{+} A| = m$.

Set $K = 2m^3$. By the assumption on $\{F_k\}$ we have $F_{m+1} > K$. In particular, there exist two sets of distinct integers, $X = \{x_1, \ldots, x_K\}$ and $Y = \{y_0, y_1, \ldots, y_m\}$, such that

$$x + y \in \textsc{Squares} \quad \text{for all } x \in X \text{ and } y \in Y.$$

By translating X, Y in opposite directions (recall that $F_{m+1} \geq K + 1$) we may assume that $0 = y_0 \in Y$, and so $X \subset \textsc{Squares}$. We may also assume $4 \mid x_i$ for all i, (otherwise, multiply X and Y by 4), and set $\tilde{X} = \{\sqrt{x} : x \in X\} \subset \mathbb{Z}$.

Let $P = \{-\frac{1}{4}y_1, \ldots, -\frac{1}{4}y_m\}$. We claim that there exists a subset $S \subset \tilde{X}$ of size m, such that all the solutions to (3.1) with $s \in S$, $p \in P$ are distinct.

To see this, first notice that if $s \in \tilde{X}$ and $p \neq p' \in P$, then $s^2 - 4p$ and $s^2 - 4p'$ are two distinct squares by our assumption on X and Y. It thus follows that there are $2m$ distinct solutions to (3.1) for this s and all $p \in P$. Let A_s denote this set of $2m$ integer solutions.

Consider the graph H on the vertex set $\tilde{X}$, where two distinct vertices $s, s' \in \tilde{X}$ are adjacent if and only if they share a common solution to (3.1), that is, if $A_s \cap A_{s'}$ is nonempty.

Next, note that any $a \in \mathbb{Z}$ (in fact even in $\mathbb{R}$) and $p \in P$ can correspond to at most one possible value of s such that a is a solution of (3.1) with this pair (s, p) (namely, the only possible value for s is $a + \frac{p}{a}$, where here we used the fact that $p \neq 0$ for all $p \in P$). It then follows that the degree of any $s \in \tilde{X}$ in H is at most

$$|A_s||P| - 1 < 2m^2 \ .$$

In other words, H is graph on K vertices with maximal degree less than $2m^2$, and thus has an independent set (an induced subgraph containing no edges) of size at least $K/(2m^2) = m$. Furthermore, such an independent set can easily be found via the Greedy algorithm (sequentially processing $\tilde{X}$ and adding vertices that are not incident to the current induced subgraph).

Combined with Observation 3.1, this implies that $\mathtt{SP}_{\mathbb{Z}}(M) = m$, proving Item (2).

It remains to prove Item (3). Let $A = \{a_u : u \in V\}$ be a set of $|V|$ distinct integers, let $S = A \overset{M}{+} A$ and $P = A \overset{M}{\times} A$ denote the sum-set and product-set of A along M resp., and set $m \triangleq \max\{|S|, |P|\}$. Define the following $m \times m$ binary matrix B, indexed by the elements of S and P (if either S or P has

less than m elements, B may have all-zero rows or columns respectively):

$$B_{s,p} = \begin{cases} 1 & s = a_u + a_v \text{ and } p = a_u a_v \text{ for some } e = (u,v) \in E \ , \\ 0 & \text{otherwise} \ . \end{cases} \tag{3.2}$$

By definition there are two distinct integer solutions to (3.1) for any (s,p) such that $B_{s,p} = 1$. In particular,

$$s^2 - 4p \in \textsc{Squares} \ \text{ for any } (s,p) \text{ such that } B_{s,p} = 1 \ . \tag{3.3}$$

Let $t = t(n)$, and consider $F_k(4t^2)$, defined in (1.6) as the maximum number of translates that k integers $\{a_1, \ldots, a_k\} \subset \{-(2t)^2, \ldots, (2t)^2\}$ can have into the set of perfect squares. It then follows from (3.2) and (3.3) that, if $|a_u| \leq t$ for all $u \in V$, then there are at most $r \stackrel{\triangle}{=} F_2(4t^2)$ translates of any set $\{s_1^2, \ldots, s_k^2\}$ with $s_1, \ldots, s_k \in S$ into the squares. Similarly, there are at most r translates of any set $\{-4p_1, \ldots, -4p_k\}$ with $p_1, \ldots, p_k \in P$ into the squares. It follows that B does not contain a $k \times (r+1)$ minor consisting of all 1's. Equivalently, B represents a bipartite graph G with color classes of size m each, which has $e(M) = n$ edges and does not contain a copy of the subgraph $K_{k,r+1}$.

The case $k = 2$ is somewhat simpler and has interesting consequences, and so we deal with it first. In what follows we need a special case of a well known result of Kövári, Sós and Turán. For completeness, we reproduce its (simple) proof. Let $N(u)$ and $d(u)$ denote the neighborhood of a vertex u and its degree resp., and further let $N(u,v)$ and $d(u,v)$ denote the common neighborhood of two vertices u, v and its size (the co-degree) respectively. According to these notations, a standard calculation shows that the total of all co-degrees in G is

$$\begin{aligned} D \stackrel{\triangle}{=} \sum_{u \in V(G)} \binom{d(u)}{2} &= \frac{1}{2} \sum_u (d(u))^2 - e(G) \\ &\geq \frac{\left(\sum_u d(u)\right)^2}{4m} - e(G) = \frac{e(G)^2}{m} - e(G) \ , \end{aligned}$$

where the inequality was due to Cauchy-Schwartz (recalling that G is a graph on $2m$ vertices). Dividing by $\binom{2m}{2}$ and using the fact that $e(G) \leq \binom{2m}{2}$, we obtain that the average co-degree in G is at least

$$\frac{D}{\binom{2m}{2}} \geq \frac{e(G)^2}{2m^3} - 1 \ .$$

On the other hand, as G contains no $K_{2,r+1}$, this quantity is necessarily at most r, and so

$$r \geq \frac{e(G)^2}{2m^3} - 1 = \frac{n^2}{2m^3} - 1 \ .$$

Rearranging, we have

$$m \geq \Big(\frac{n^2}{2(r+1)}\Big)^{1/3} ,$$

which by definition of m, r gives that for all $A \subset \{-t, \ldots, t\}$

$$\max\{|A \overset{M}{+} A| \, , \, |A \overset{M}{\times} A|\} = \Omega\big(n^{2/3}[F_2(4t^2)]^{-1/3}\big) \, , \tag{3.4}$$

thus proving Item (3) of the theorem for $k = 2$.

Note that at this point we can infer Corollary 3. Indeed, letting $a, b \in \mathbb{Z}$, any $x \in \mathbb{Z}$ that translates $\{a, b\}$ into the squares satisfies

$$a + x = y_1^2 \;\text{ and }\; b + x = y_2^2 \;\text{ for some } y_1, y_2 \in \mathbb{Z} \, ,$$

and so

$$a - b = y_1^2 - y_2^2 = (y_1 - y_2)(y_1 + y_2) \, .$$

It follows that the number of such translates corresponds to the number of divisors of $a-b$. As it is well known that the number of divisors of an integer N is at most $\exp\big[O(\frac{\log N}{\log\log N})\big] \leq N^{o(1)}$, we deduce that

$$F_2(N) \leq N^{o(1)} \, .$$

Combining this with (3.4) immediately implies the required lower bound $\max\{|A \overset{M}{+} A| \, , \, |A \overset{M}{\times} A|\} \geq n^{2/3-o(1)}$ whenever every $a \in A$ has $|a| \leq n^{O(1)}$.

To generalize the lower bound to any fixed k, we apply the general theorem of Kövári, Sós and Turán [28] on the density of binary matrices without certain sub-matrices consisting only of 1 entries. We use the following version of this theorem (see, e.g., [27, Chapter 2.2], and also [29]):

Theorem 3.2 (Kövári-Sós-Turán)**.** *Let $k \leq r$ be two integers, and let G be a bipartite graph with m vertices in each of its parts. If G does not contain $K_{k,r}$ as a subgraph, then*

$$e(G) \leq (r-1)^{1/k}(m-k+1)m^{1-1/k} + (k-1)m \, .$$

As noted above, with $r = F_k(4t^2)$, for any k rows of B there can be at most r columns forming a sub-matrix consisting only of 1 entries, and vice versa. Equivalently, the bipartite graph G does not contain $K_{k,r+1}$ as a subgraph. Recalling that $e(G) = n$, we obtain that

$$n \leq r^{1/k} m^{(2k-1)/k} + (k-1)m \, .$$

Either $m > \frac{n}{2(k-1)}$, in which case we are done, or else this yields

$$m \geq (n/2)^{k/(2k-1)} \, r^{-1/(2k-1)} \, ,$$

that is,

$$\max\{|A \overset{M}{+} A| \, , \, |A \overset{M}{\times} A|\} = \Omega\big(n^{k/(2k-1)}[F_k(t^2)]^{-1/(2k-1)}\big) \, .$$

This concludes the proof of Theorem 2. ∎

4. Translates of a set into the squares

4.1. **Euler's problem: translates of three integers into the squares.** In this section, we study the parameter F_3: We are interested in integer solutions to the following set of 3 equations in 4 variables (X, Y_1, Y_2, Y_3):

$$Y_i^2 = X + a_i \qquad (i = 1, 2, 3) \ , \tag{4.1}$$

where the a_i's are distinct integers. By clearing denominators, it is equivalent to consider rational solutions rather than integer ones.

Euler [22, Chapter 2.XIV] studied this question in the following form:

> "To find a number, x, which, added to each of the given numbers, a,b,c, produces a square"

After demonstrating that this is impossible in some families of parameters, he concludes (in the following $m = b - a$ and $n = c - a$):

> "...it is not easy to choose such numbers for m and n as will render the solution possible. The only means of finding such values for m and n is to imagine them, or to determine them by the following method."

Euler's method is to start with a given solution (assuming one is available), and look for others, using transformations of certain quartics into squares. He considers the integers $\{0, 2, 6\}$, so $\frac{1}{4}$ is a solution, and proceeds to find the solution $\left(\frac{191}{60}\right)^2$. Euler further claims that this method can be used recursively to find other solutions. Following his line of arguments gives the following recursion relation: If x^2 is a solution for the integers $\{0, 2, 6\}$, that is, $y^2 = x^2 + 2$ and $z^2 = 6 + x^2$ for some $y, z \in \mathbb{Q}$, then

$$x' = \frac{(x^4 - 12)(x + y)}{2xyz\sqrt{2 + 2x(x + y)}} = \frac{x^4 - 12}{2xyz}$$

also provides such a solution, since in that case it is easy to verify that

$$(x')^2 + 2 = \frac{(x^4 + 4x^2 + 12)^2}{(2xyz)^2} \ , \qquad (x')^2 + 6 = \frac{(x^4 + 12x^2 + 12)^2}{(2xyz)^2} \ .$$

Plugging in $x = \frac{191}{60}$ gives the additional solution $(x')^2$ for $x' = \frac{1175343361}{1154457480}$ (the next element obtained via this recursion has 38-digit numerator and denominator). Euler does not discuss when this method may guarantee an aperiodic series of translations (though this may be shown using similar elementary methods).

In what follows, we present a general framework for obtaining sets of 3 integers with infinitely many translates into the squares, using elliptic curves. This approach further provides a quantitative lower bound on the number

of translates, in terms of the height of the elements of the original set (i.e., a lower bound on $F_3(n)$). We begin by showing that indeed $F_3 = \infty$.

Theorem 4.1. *The parameter F_3 is unbounded. Moreover, there exist distinct integers $\{a_1, a_2, a_3\}$ with infinitely many translates into the set of perfect rational squares.*

Proof. We may assume that there is at least one solution to (4.1), and without loss of generality $X = 0$ is a solution, so we have $\sqrt{a_i} \in \mathbb{Q}$ for all i and can instead consider the equations

$$Y_1^2 - Y_i^2 = a_1 - a_i \qquad (i = 2, 3) \ .$$

For some $t, u \in \mathbb{Q}$ to be later specified, let

$$Y_1 = \sqrt{a_1} + u \ , \quad Y_2 = \sqrt{a_2} + tu \ . \tag{4.2}$$

It follows that

$$Y_1^2 - Y_2^2 = a_1 - a_2 + \big((1-t^2)u + 2(\sqrt{a_1} - t\sqrt{a_2})\big)\, u \ ,$$

thus if $t^2 \neq 1$ then $u = \frac{2(\sqrt{a_1} - t\sqrt{a_2})}{t^2-1}$ is the unique non-zero rational such that

$$Y_1^2 - Y_2^2 = a_1 - a_2$$

(Note that, for $t = \pm 1$, only $u = 0$ satisfies this equality, since $a_1 \neq a_2$). Using this substitution, we find

$$Y_3^2 = Y_1^2 - (a_1 - a_3) = (\sqrt{a_1} + u)^2 - (a_1 - a_3) = \frac{Q(t)}{(t^2-1)^2} \ ,$$

where $Q(t)$ is a monic quartic with known coefficients (derived from a_1, a_2, a_3). Therefore, to solve (4.1) we need $Q(t)$ to be a rational square.

Let $G(t)$ be a quadratic and $H(t)$ linear so that $Q(t) = G^2(t) + H(t)$. If $Q(t)$ is a rational square, let

$$T_0 \triangleq G(t) + \sqrt{Q(t)} \ , \qquad\qquad S_0 \triangleq tT_0 \ . \tag{4.3}$$

We have

$$\begin{aligned} 0 = Q(t) - G^2(t) - H(t) &= T_0(T_0 - 2G(t)) - H(t) \\ &= T_0^2 - 2T_0 G(S_0/T_0) - H(S_0/T_0) \ , \end{aligned}$$

which upon multiplying by T_0 becomes a polynomial relation between S_0, T_0. Next, let S, T be affine changes of S_0, T_0 as follows:

$$\begin{aligned} T &= \tfrac{1}{2}T_0 + \frac{3a_1a_2 - 2a_1a_3 - 2a_2a_3 + a_3^2}{3a_3^2} \ , \\ S &= \tfrac{1}{2}S_0 + \frac{\sqrt{a_1a_2}}{2a_3}T_0 - \frac{(a_1a_2)^{3/2} - \sqrt{a_1^3a_2}a_3 - \sqrt{a_1a_2^3}a_3 - \sqrt{a_1a_2}a_3^2}{a_3^3} \ . \end{aligned}$$

Integer assignment	Sum	Product
283815 •——• 17974425	$4 \cdot 4564560$	5101411431375
597975 •——• 8531145	$2 \cdot 4564560$	5101411431375
1954575 •——• 2609985	4564560	5101411431375
−1711710 •——• 19969950	$4 \cdot 4564560$	−34182763114500
−2852850 •——• 11981970	$2 \cdot 4564560$	−34182763114500
−3993990 •——• 8558550	4564560	−34182763114500
−6607744 •——• 24865984	$4 \cdot 4564560$	−164308056580096
−9042176 •——• 18171296	$2 \cdot 4564560$	−164308056580096
−10737584 •——• 15302144	4564560	−164308056580096

TABLE 1. Optimal sum-product mapping for a matching of size 9 over the integers: 3 sums and 3 products, found using the elliptic curve $S^2 = T^3 - 63T + 162$.

This brings the polynomial relation between S_0, T_0 to an elliptic curve in standard form:

$$S^2 = T^3 + \alpha T + \beta \ ,$$

where

$$\alpha = \frac{-\sum_i a_i^2 + \sum_{i<j} a_i a_j}{3a_3^2} \ , \ \beta = \frac{2\sum_i a_i^3 - 3\sum_{i \neq j} a_i^2 a_j + 12 a_1 a_2 a_3}{27 a_3^3} \ . \quad (4.4)$$

We next show that $F_3 = \infty$. Consider the choice $a_1 = \frac{4}{9}$, $a_2 = \frac{16}{9}$, $a_3 = \frac{1}{9}$. By (4.4), this produces the elliptic curve

$$S^2 = T^3 - 63T + 162 \ ,$$

which has positive rank (namely, rank 1, computed via SAGE). This gives rise to infinitely many rational points (T, S), and using (4.3), we can recover the value of $t = S/T$ from each of them. Recalling that u is uniquely determined by t, we now return to (4.2) and obtain the rational points Y_1, Y_2, Y_3 from each pair (T, S), as required. ■

Table 1 demonstrates how the above analysis provides an optimal family of 9 pairs of distinct integers, inducing only 3 sums and 3 products:

$$\mathtt{SP}_{\mathbb{Z}}(M) = 3 \quad \text{when } M \text{ is the matching on 9 edges} \ ,$$

where the lower bound follows from (1.2).

Remark. In general, the above analysis leads, for any integer n, to an explicit construction giving a family of $3n$ pairs of distinct integers, inducing 3 distinct sums and n distinct products.

Remark. An alternative way for proving Theorem 4.1 is to consider the curve $y^2 = (x+a_1)(x+a_2)(x+a_3)$. This curve contains the rational points of the curve defined in (4.4), and one may obtain infinitely many of them by starting from one of the points and repeatedly doubling it.

4.2. **A quantitative lower bound.** Recall that $F_2(n) = \exp\left[\Theta\left(\frac{\log n}{\log\log n}\right)\right]$, which implies that

$$F_3(n) \le F_2(n) \le n^{o(1)} \ .$$

The next theorem provides a lower bound on $F_3(n)$:

Theorem 4.2. *The function F_3 satisfies $F_3(n) = \Omega\left((\log n)^{5/7}\right)$.*

Proof. We need the following well-known facts concerning elliptic curves.

The Mordell-Weil Theorem states that, for any elliptic curve $E(\mathbb{Q})$, the group of rational points on the curve is a finitely generated abelian group: $E(\mathbb{Q}) \cong E_{\text{torsion}} \bigoplus \mathbb{Z}^r$, where E_{torsion} is the *torsion group* (points of finite order) and r is the *rank* of the curve. Mazur's Theorem characterizes the torsion group of any $E(\mathbb{Q})$ as one of 15 given (small) groups.

The *logarithmic height* of a rational $x = \frac{p}{q}$, denoted by $h(x)$, is defined as

$$h(p/q) \stackrel{\triangle}{=} \max\{\log|p| \ , \ \log|q|\} \ .$$

For a point $P = (x,y) \in E(\mathbb{Q})$ we let $h(P) \stackrel{\triangle}{=} h(x)$. Further define the *canonical height* of $P \in E(\mathbb{Q})$ to be

$$\hat{h}(P) \stackrel{\triangle}{=} \frac{1}{2}\lim_{m\to\infty}\frac{h(mP)}{m^2} \ .$$

The following theorem states some well known properties of the canonical height:

Theorem 4.3 (canonical height)**.** *The following holds:*

1. *The canonical height $\hat{h}$ is quadratic and satisfies the parallelogram law:*

$$\hat{h}(mP) = m^2 \cdot \hat{h}(P) \quad \textit{and} \quad \hat{h}(P+Q) + \hat{h}(P-Q) = 2(\hat{h}(P) + \hat{h}(Q)) \ .$$

2. *The canonical height $\hat{h}$ is roughly logarithmic:*

$$\left|\hat{h}(P) - h(P)\right| < K \quad \textit{for some } K = K(E) \ .$$

By the above facts, we can now deduce the following corollary for the number of rational points on $E(\mathbb{Q})$ with a given bound on their numerators and denominators:

Corollary 4.4. *In an elliptic curve $E(\mathbb{Q})$ of rank r, the number of points with $\hat{h}(P) < M$ has order $M^{r/2}$. Consequently this is also the number of points with numerator and denominator bounded by $O(\exp(M))$.*

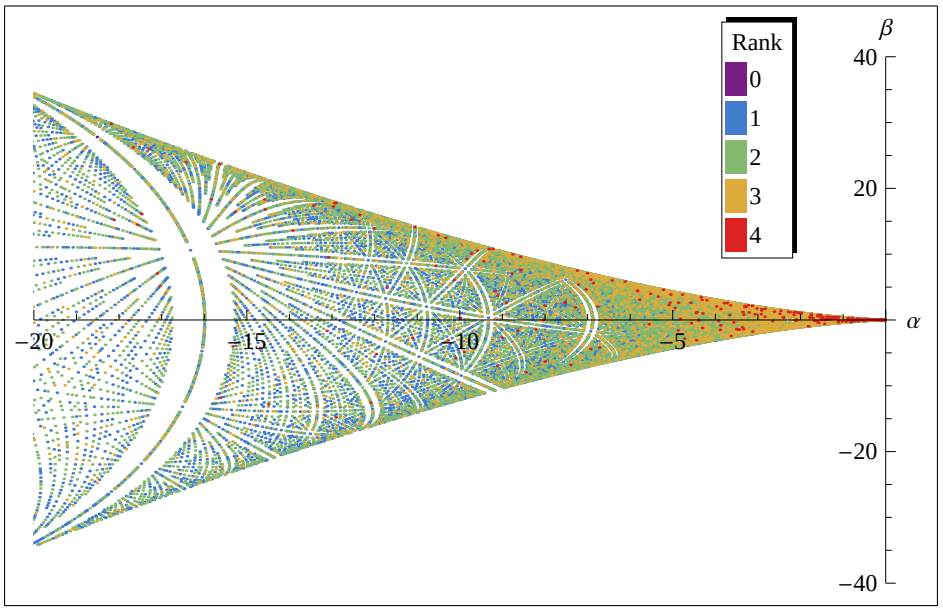

FIGURE 1. Ranks of all curves corresponding to triplets $\{a_1, a_2, a_3\} \subset \{0, 1, \ldots, 100\}$, as given by (4.4).

To see this, take a basis $P_1, \ldots, P_r$ for the abelian group $E(\mathbb{Q})$. It is now straightforward to verify that the requisite points are (up to the slight effect of the torsion group) the points $\sum_{i=1}^{r} a_i P_i$ where $a_i < \sqrt{M}$.

Let a_1, a_2, a_3 be distinct rational points, and let r denote the rank of the elliptic curve defined in (4.4). By the above discussion, there are $M^{r/2}$ rational solutions with denominators bounded by $O(\exp(M))$. Clearing denominators results in $M^{r/2}$ integer solutions to the system corresponding to $a_1', a_2', a_3' \in \mathbb{Z}$, where all absolute values are at most $N = O(\exp(M^{1+r/2}))$. Equivalently, $M = \Omega\left((\log N)^{2/(r+2)}\right)$, giving the following estimate on $F_3(n)$:

$$F_3(n) \geq \Omega((\log n)^{r/(r+2)}) \ .$$

In particular, one can verify that a choice of $a_1 = 3$, $a_2 = 34$, $a_3 = 89$ (obtained by a computer search using SAGE) for the a_i's produces a curve of rank 5, implying the desired result. ∎

Remark. The above curve of rank 5 was found by a computer search (see Figure 1). Not every elliptic curve can be represented by (4.4), and it is even unknown if there are elliptic curves of arbitrarily large rank.

It appears that the curves obtained from (4.4) have rank 0 only when a_1, a_2, a_3 are of the form $xy, xy + x^2, xy + y^2$ for some x, y. The rank distribution for curves obtained from (4.4) appears to concentrate on rank 1 as the height of the a_i's tends to infinity (e.g., random samples of curves

from the ranges $\{L, \ldots, 2L\}$ with $L \in \{10^3, 10^4, 10^5\}$ gave rank 1 in about 0.33, 0.48 and 0.70 fraction of the samples resp.). This is in contrast with all elliptic curves, where it is believed that the rank is 0 and 1 with density 1/2 each.

Further note that the constants implicit in Corollary 4.4 depend on the curve. A uniform (in the curve) bound on the number of points of height at most m in an elliptic curve may lead to an improved upper bound.

4.3. **Translates of four or more integers and curves of higher genus.** We next turn our attention to the parameters F_k for $k > 3$. Recalling the general framework of the problem, we are interested in integer solutions to the set of k equations in the $k+1$ variables $(X, Y_1, \ldots, Y_k)$:

$$Y_i^2 = X + a_i \qquad (i = 1, \ldots, k) \ , \tag{4.5}$$

where the a_i's are assumed to be distinct coefficients.

Let $\mathcal{S} \subset \mathbb{C}^{k+1}$ denote the set of all complex solutions. By adding a point at infinity, $\mathcal{S}$ can be compactified into a one (complex) dimensional manifold.

Lemma 4.5. *The genus of $\mathcal{S}$ is $1 + (k-3)2^{k-2}$.*

Proof. Using the relation to the Euler characteristic $\chi(\mathcal{S}) = 2 - 2g(\mathcal{S})$, it suffices to show that

$$\chi(\mathcal{S}) = (3-k)2^{k-1} \ .$$

This is achieved by means of the Riemann-Hurwitz formula: For a map $\pi : \mathcal{S} \to \mathcal{S}'$ which is $N \to 1$ except at some ramification points of $\mathcal{S}'$ we have

$$\chi(\mathcal{S}) = N\chi(\mathcal{S}') + \sum e_p - c_p$$

where the sum is over the ramification points and e_p is the ramification index at p and c_p the cycle index.

We apply this to our manifold $\mathcal{S}$ and the Riemann sphere $\mathcal{S}'$, and with the map $\pi(X, Y_1, \ldots, Y_k) = X$. Any X has 2^k pre-images with coordinates given by the square roots of $X + a_i$. The ramification points are $X = -a_i$ for each i and $X = \infty$. At $X = a_i$ we have only 2^{k-1} pre-images (here we use that all a_i's are distinct) and so

$$e_{-a_i} - c_{-a_i} = 2^k - 2^{k-1} = 2^{k-1} \ .$$

The singularity at $X = \infty$ is of the same type (2^{k-1} coinciding points of ramification index 2, and thus $e_\infty - c_\infty = 2^{k-1}$). Combining these we find

$$\chi(\mathcal{S}) = 2^k \cdot 2 - k2^{k-1} - 2^{k-1} = (3-k)2^{k-1},$$

as required. ∎

Note that for $k = 3$, the genus of $\mathcal{S}$ is 1 and thus $\mathcal{S}$ is an elliptic curve, as we have already seen in the above analysis of this case. For $k > 3$ the genus is larger, and the understanding of rational points on $\mathcal{S}$ is relatively scant.

It is well known that rational points in curves of high genus are uncommon: Indeed, Falting's Theorem states that the number of rational points on any curve of genus $g > 1$ is finite. The following result of Caporaso, Harris and Mazur further states that this quantity is uniformly bounded from above, if we accept a major conjecture in Arithmetic Geometry.

Theorem 4.6 (Caporaso-Harris-Mazur [9])**.** *Assume the Bombieri-Lang conjecture. Then for any $g > 1$ there is some constant $B(g)$ such that the number of rational points on any curve of genus g is at most $B(g)$.*

Combining this with Lemma 4.5 implies that, if we accept the Bombieri-Lang conjecture, then for any $k \geq 4$ there is some constant B depending only on k such that the number of solutions to (4.5) is at most B. In other words, if the Bombieri-Lang conjecture holds, then $F_k \leq B(1 + (k-3)2^{k-2}) < \infty$ for any $k \geq 4$. Together with Theorem 2 (Part (3)), this proves Corollary 4.

Remark. The curve determined by (4.5) is not completely general, and so it may be possible to get bounds on the number of solutions without assuming the Bombieri-Lang conjecture. One approach is to multiply the equations and note that $P(X) := \prod(X + a_i) = (\prod Y_i)^2$ is a square, where $P(X)$ is some polynomial of degree k with distinct integer roots. This determines a curve of lower genus, but still of genus greater than 1 for $k > 4$.

4.4. **Consequences for the original Erdős-Szemerédi problem.** We have shown that, assuming a plausible conjecture in Number Theory, for every matching M of size n and every set B of $2n$ distinct integers,

$$\max\{|B \overset{M}{+} B|, |B \overset{M}{\times} B|\} \geq \Omega(n^{4/7}).$$

Together with Theorem 2.3, this implies that if G is any graph on n vertices and at least n^2/k edges, and A is any set of n distinct integers, then

$$\max\{|A \overset{G}{+} A|, |A \overset{G}{\times} A|\} \geq \Omega\Big(\min\Big\{\frac{n^{15/14}}{k^{4/7}}\ ,\ \frac{n^{3/2}}{k}\Big\}\Big). \tag{4.6}$$

Indeed, if $S = \max\{|A \overset{G}{+} A|\ ,\ |A \overset{G}{\times} A|\}$ then, by Theorem 2.3 with $D = n$ and $d = n/k$, either $S = \Omega(n^{3/2}/k)$, or there is a matching M of size $\Omega(n/k)$ and a set of of $2|M|$ distinct integers B so that

$$\Omega\left((n/k)^{4/7}\right) \leq \max\{|B \overset{M}{+} B|\ ,\ |B \overset{M}{\times} B|\} \leq O\Big(\frac{S}{\sqrt{n}}\Big),$$

supplying the desired lower bound for S.

It is worth noting that without assuming any unproven conjectures, one can prove that for every G as above,

$$\max\{|A \overset{G}{+} A| \ , \ |A \overset{G}{\times} A|\} \geq \Omega\left(\frac{n^{10/9-o(1)}}{k^{19/9-o(1)}}\right) . \tag{4.7}$$

This can be done as follows. Suppose

$$\max\{|A \overset{G}{+} A| \ , \ |A \overset{G}{\times} A|\} = cn \ ,$$

where G has n vertices and at least n^2/k edges. By the proof of Gowers [25] of the Balog-Szemerédi Theorem [3] (see also [34]), there are two subsets $A', B' \subset A$ so that

$$\begin{aligned} &|A'| = |B'| \geq \Omega(n/k) \qquad \text{and} \\ &|A' + B'| \leq O(c^3 k^5 n) \ , \ |A' \times B'| \leq O(c^3 k^5 n) \ . \end{aligned}$$

However, if we plug in the best current lower bound for the sum-product of the complete graph, due to Solymosi [33], we have that

$$\max\{|A' + B'| \ , \ |A' \times B'|\} \geq \Omega((n/k)^{4/3-o(1)}) \ ,$$

implying that $c^3 \geq \Omega(n^{1/3-o(1)}/k^{19/9+o(1)})$, which gives the desired estimate

$$cn \geq \Omega\left(\frac{n^{10/9-o(1)}}{k^{19/9+o(1)}}\right) .$$

Note that for $k > n^{1/19}$ this does not give any nontrivial bound.

On the other hand, the estimate (4.6) (which depends on the validity of the Bombieri-Lang conjecture) gives a nontrivial bound for all $k < n^{1/8}$. Furthermore, our bound improves upon (4.7) already for $k > n^{5/194}$.

5. Sums along expanders

In this section, we study sum-sets along various underlying geometries, and obtain tight bounds for the case of expander graphs (Theorem 5). Note that we no longer consider the product-set along the graph.

5.1. **Lower bound for sums along general graphs.**

We begin with a straightforward lower bound for the size of the sum-set along a given graph:

Observation 5.1. *Let $\mathbb{F}$ be a field of characteristic* $\mathrm{char}(\mathbb{F}) \neq 2$. *Then for any graph G and an injective map A from its vertices to $\mathbb{F}$ we have*

$$|A \overset{G}{+} A| \geq [2k \cdot e_k(G)]^{1/k} \quad \textit{for any odd integer } k \geq 3 \ ,$$

where $e_k(G)$ denotes the number of cycles of length k in G.

To prove this, let A be a mapping from the vertices of G to $\mathbb{F}$, and consider a cycle

$$u = v_0, v_1, \ldots, v_k = u \ ,$$

where $v_0 v_k \in E$ and $v_i v_{i+1} \in E$ for $i = 0, \ldots, k-1$. We then have that

$$2A_u = \sum_{i=1}^{k} (-1)^{i-1} (A_{v_{i-1}} + A_{v_i}) \ ,$$

and so A_u is uniquely determined by the sums on the edges of the cycle. Therefore, for any directed cycle as above (fixing the starting point $u = v_0$ and the orientation), we must have a different sequence of k sums along the edges (otherwise, for two cycles starting at $u \neq u'$ we would get $A_u = A_{u'}$, while for the two orientations of the same cycle we would get $A_{v_1} = A_{v_{k-1}}$). Altogether, any undirected cycle gives rise to $2k$ distinct sequences of sums (accounting for both orientations). This implies the desired lower bound.

Note that we cannot infer a bound on $|A \overset{G}{+} A|$ in terms of $e_k(G)$ when $\mathrm{char}(\mathbb{F}) = 2$ or k is even. To see this, suppose one cycle has labels $a_1, \ldots, a_k$. Disjoint cycles may then be labeled $a_i + (-1)^i x$ with an arbitrary x to get the same k sums. It is thus possible to choose at least $|\mathbb{F}|/k^2$ values of x (and in fact even more, with a careful choice of the values $a_1, \ldots, a_k$), so that the labels are all distinct, whereas $|A \overset{G}{+} A| = k$.

The lower bound of Observation 5.1 can be asymptotically tight: To demonstrate this for $k = 3$, we consider the graph G comprising $\binom{m}{3}$ disjoint triangles, and construct for it an injection A such that

$$|A \overset{G}{+} A| = m \ .$$

Let S be a set of m distinct sums, all even, to be specified later. We assign each of the triangles a different triplet of these sums. This determines the integer values at each of the vertices. To conclude the construction, we need S to yield distinct integer values at different vertices; this is achieved, for instance, by choosing

$$S = \{2^i : i = 1, \ldots, m\} \ ,$$

since $2^{i_1} + 2^{j_1} - 2^{k_1} \neq 2^{i_2} + 2^{j_2} - 2^{k_2}$ for any two sets $\{i_1, j_1, k_1\} \neq \{i_2, j_2, k_2\}$ (alternatively, one can obtain A which is injective **whp**, by choosing the elements of S independently and uniformly over some large ground set).

It is not difficult to extend this example and show that Observation 5.1 gives the optimal order of $|A \overset{G}{+} A|$ whenever G is a disjoint union of cycles of odd length k.

5.2. **Tight bounds for sums along expanders.** We now prove Theorem 5, showing that the sum-set along an n-vertex expander can be of size $O(\log n)$, and this is best possible.

Recall the definition of an expander given in the introduction. In the special case where the graph $G = (V, E)$ is d-regular, we call G a δ-expander if for every set X of at most $|V|/2$ vertices, the number of edges from X to its complement is at least $\delta d|X|$.

An (n, d, λ)-graph is a connected d-regular graph on n vertices, in which the absolute value of each nontrivial eigenvalue is at most λ. This notion was introduced by the first author in the 1980's, motivated by the observation that such graphs in which λ is much smaller than d exhibit strong pseudo-random properties. In particular, it is easy to show (see, e.g., [2]) that

$$\text{every } (n, d, \lambda)\text{-graph is a } \delta\text{-expander for } \ \delta = \frac{d-\lambda}{2d} \ . \tag{5.1}$$

Theorem 5.2. *For every fixed $\delta < \frac{1}{2}$ there is a constant $c = c(\delta)$ so that the following holds: For any sufficiently large n there is a δ-expander G on n vertices such that $|A \overset{G}{+} A| \le c \log n$ for $A = \{1, 2, \ldots, n\}$.*

Proof. For an abelian group Λ and a subset $T \subset \Lambda$, the *Cayley sum graph* $G = G(\Lambda, T)$ of Λ with respect to T is the graph whose set of vertices is Λ, in which yz is an edge for each $y, z \in \Lambda$ satisfying $y + z \in T$. Clearly, this is a $|T|$-regular graph.

Let D be the adjacency matrix of G. It is well known (cf., e.g., [1]) that its eigenvalues can be expressed in terms of T and the characters of Λ. Indeed, for every character χ of Λ and every $y \in \Lambda$,

$$(D\chi)(y) = \sum_{s \in T} \chi(s - y) = \Big(\sum_{s \in T} \chi(s) \Big) \overline{\chi(y)} \ .$$

Applying D again, it follows that

$$D^2 \chi = \Big| \sum_{s \in T} \chi(s) \Big|^2 \chi \ .$$

Therefore, the eigenvalues of the symmetric matrix D^2 are precisely the expressions $|\sum_{s \in T} \chi(s)|^2$, where the characters are the corresponding eigenvectors, and as the characters are orthogonal, these are all eigenvalues. It then follows that each nontrivial eigenvalue of the graph $G = G(\Lambda, T)$ is, in absolute value, $|\sum_{s \in T} \chi(s)|$ for some nontrivial character χ of Λ (it is not difficult to determine the signs as well, but these are not needed here).

In particular, for the additive group $\mathbb{Z}_n$ and for $T \subset \mathbb{Z}_n$, every nontrivial eigenvalue of the Cayley graph of $\mathbb{Z}_n$ with respect to T is, in absolute value, $|\sum_{s \in T} \omega^s|$, where ω is a nontrivial n-th root of unity. The following lemma is proved in [1] by a simple probabilistic argument.

Lemma 5.3 ([1]). *For every integer $d \leq n^{2/3}$ there exists a subset $T \subset \mathbb{Z}_n$ of cardinality d so that for every nontrivial n-th root of unity ω*

$$\Big|\sum_{s\in T}\omega^s\Big| \leq 3\sqrt{d}\sqrt{\log(10n)}.$$

To complete the proof of Theorem 5.2, assume n is sufficiently large as a function of δ. Let T satisfy the assertion of the lemma, with $|T| = c' \log n$, where c' is chosen to ensure that

$$\frac{c'\log n - 3\sqrt{c'\log n}\sqrt{\log(10n)}}{2c'\log n} \geq \delta.$$

Since $\delta < 1/2$ it is obvious that $c' = c'(\delta) > 0$ can be chosen to satisfy this inequality, and thus the graph $G = G(\mathbb{Z}_n, T)$ has the desired expansion properties, by (5.1). Every sum of endpoints of an edge in it is equal, modulo n, to a member of T, and as we are considering addition over the integers, this means that for $A = \{1, 2, \ldots, n\}$,

$$|A \overset{G}{+} A| \leq 2|T| = 2c'\log n \ ,$$

as required. ■

It remains to show that the logarithmic estimate is tight. This is done by considering the diameter of graphs G for which $A \overset{G}{+} A = T$ is a relatively small set.

Lemma 5.4. *Let $G = (V, E)$ be a graph on n vertices and $A = \{a_v : v \in V\}$ be a set of distinct integers. If $|A \overset{G}{+} A| = s$ and the diameter of G is r, then*

$$2^s\binom{r+s}{s} \geq n/2 \ . \tag{5.2}$$

In particular, if $r \leq L\log n$ for some $L > 1$ then $s \geq \Omega(\log_L n)$.

Proof. Put $T = A \overset{G}{+} A$. Fix a vertex $u \in V$. If

$$u = v_0 v_1 v_2 \ldots v_l = w$$

is a path of length l in G starting at u, then there are (not necessarily distinct) elements $g_1, g_2, \ldots, g_l \in T$ so that

$$a_w = g_l - g_{l-1} + g_{l-2} - \cdots + (-1)^{l-1}g_1 + (-1)^l a_u \ .$$

It follows that for every $a_w \in A$, either the difference $a_w - a_u$ or the sum $a_w + a_u$ can be expressed as the inner product of an integral vector x of length s with ℓ_1-norm at most r with the vector $(g : \ g \in T)$. The number of choices for the vector x is at most $2^s\binom{r+s}{s}$, implying (5.2).

The conclusion in case $r \leq L\log n$ for some $L > 1$ now follows by a simple manipulation, since $2^s\binom{r+s}{s} \leq \left(\frac{2e(r+s)}{s}\right)^s$. ■

Corollary 5.5. *Let $G = (V, E)$ be a δ-expander on n vertices, and let $A = \{a_v : v \in V\}$ be a set of distinct integers. Then $|A \overset{G}{+} A| \geq \Omega(\frac{1}{\log(1/\delta)} \log n)$.*

Proof. It is well-known (and easy) that if G is a δ-expander on n vertices, its diameter satisfies

$$\operatorname{diam}(G) \leq 2\Big\lceil \frac{|E|}{\log(1+\delta)} \Big\rceil = O\Big(\frac{\log n}{\delta}\Big) \,. \tag{5.3}$$

Indeed, by definition (1.8) and the assumption $\Phi(G) \geq \delta$,

$$e(S, \overline{S}) \geq \delta \operatorname{vol}(S) \;\text{ for any } S \subset V \text{ with } 0 \neq \operatorname{vol}(S) \leq \operatorname{vol}(\overline{S}) \,,$$

and it is straightforward to infer from this that

$$\operatorname{vol}\left(S \cup N(S)\right) \geq (1+\delta) \operatorname{vol}(S) \quad \text{ for any } S \text{ with } 0 \neq \operatorname{vol}(S) \leq \operatorname{vol}(\overline{S}) \,,$$

implying (5.3). The desired result now follows by Lemma 5.4. ∎

6. Concluding remarks and open problems

We have introduced the study of sums and products along the edges of sparse graphs, showing that it is related to the well studied investigation of the problem for dense graphs, as well as to classical problems and results in Number Theory.

There are many possible variants of the problems considered here. In particular, the results in the previous section suggest the study of the minimum possible value of $|A \overset{G}{+} A|$ for a given graph G, and its relation to the structural properties of the graph. Lemma 5.4 provides a lower bound for this quantity for graphs with a small diameter, and Observation 5.1 supplies another lower bound in terms of the number of short odd cycles in the graph (provided the characteristic of the field is not 2).

In view of the relation to the original conjecture of Erdős and Szemerédi for the dense case, an interesting open problem is whether the minimum possible value of $\mathtt{SP}_{\mathbb{Z}}(M)$ for a matching M of size n is $n^{1-o(1)}$. We believe that this is the case, but a proof will certainly require some additional ideas.

Finally, given its ramifications on the sum-product exponents of sparse graphs, it would be interesting to establish whether the parameter F_k (the maximum number of translates of k integers into the perfect squares) is indeed finite for some k.

Conjecture. *There is a finite k so that there are* no *sets $X, Y \subset \mathbb{Z}$ of sizes $|X| = |Y| = k$ that satisfy*

$$x + y \in \textsc{Squares} \text{ for all } x \in X \text{ and } y \in Y \,. \tag{6.1}$$

Note that, by our results, for all k there exist sets $X, Y \subset \mathbb{Z}$ of sizes $|X| = 3$, $|Y| = k$ that do satisfy (6.1).

Acknowledgments

We wish to thank Henry Cohn, Noam Elkies, Jordan Ellenberg, Andrew Granville, Patrick Ingram, Ram Murty, Joseph Silverman, József Solymosi, Endre Szemerédi and Terence Tao for useful discussions. This work was initiated while the second and third authors were visiting the Theory Group of Microsoft Research.

References

[1] N. Alon, *Large sets in finite fields are sumsets*, J. Number Theory **126** (2007), no. 1, 110–118.

[2] N. Alon and V. D. Milman, λ_1, *isoperimetric inequalities for graphs, and superconcentrators*, J. Combin. Theory Ser. B **38** (1985), no. 1, 73–88.

[3] A. Balog and E. Szemerédi, *A statistical theorem of set addition*, Combinatorica **14** (1994), no. 3, 263–268.

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

[5] J. Bourgain, *Some arithmetical applications of the sum-product theorems in finite fields*, Geometric aspects of functional analysis, Lecture Notes in Math., vol. 1910, Springer, Berlin, 2007.

[6] J. Bourgain, *Sum-product theorems and exponential sum bounds in residue classes for general modulus*, C. R. Math. Acad. Sci. Paris **344** (2007), no. 6, 349–352 (English, with English and French summaries).

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

[8] E. Bombieri, A. Granville, and J. Pintz, *Squares in arithmetic progressions*, Duke Math. J. **66** (1992), no. 3, 369–385.

[9] L. Caporaso, J. Harris, and B. Mazur, *Uniformity of rational points*, J. Amer. Math. Soc. **10** (1997), no. 1, 1–35.

[10] M.-C. Chang, *Factorization in generalized arithmetic progressions and applications to the Erdős-Szemerédi sum-product problems*, Geom. Funct. Anal. **13** (2003), no. 4, 720–736.

[11] M.-C. Chang, *On problems of Erdős and Rudin*, J. Funct. Anal. **207** (2004), no. 2, 444–460.

[12] M.-C. Chang, *The Erdős-Szemerédi problem on sum set and product set*, Ann. of Math. (2) **157** (2003), no. 3, 939–957.

[13] M.-C. Chang, *Some problems in combinatorial number theory*, Integers **8** (2008), no. 2, A1, 11.

[14] M.-C. Chang, *Some problems related to sum-product theorems*, Additive combinatorics, CRM Proc. Lecture Notes, vol. 43, Amer. Math. Soc., Providence, RI, 2007.

[15] M.-C. Chang and J. Solymosi, *Sum-product theorems and incidence geometry*, J. Eur. Math. Soc. (JEMS) **9** (2007), no. 3, 545–560.

[16] Y.-G. Chen, *On sums and products of integers*, Proc. Amer. Math. Soc. **127** (1999), no. 7, 1927–1933.

[17] J. Cilleruelo and A. Granville, *Lattice points on circles, squares in arithmetic progressions and sumsets of squares*, Additive combinatorics, CRM Proc. Lecture Notes, vol. 43, Amer. Math. Soc., Providence, RI, 2007.

[18] G. Elekes, *On the number of sums and products*, Acta Arith. **81** (1997), no. 4, 365–367.
[19] G. Elekes and I. Z. Ruzsa, *Few sums, many products*, Studia Sci. Math. Hungar. **40** (2003), no. 3, 301–308.
[20] P. Erdős, *Some remarks on number theory*, Riveon Lematematika **9** (1955), 45–48 (Hebrew, with English summary).
[21] P. Erdős and E. Szemerédi, *On sums and products of integers*, Studies in pure mathematics, Birkhäuser, Basel, 1983, pp. 213–218.
[22] L. Euler, *Elements of algebra*, Springer-Verlag, New York, 1984. Translated from the German by John Hewlett; Reprint of the 1840 edition; With an introduction by C. Truesdell. (Opera omnia E.388, ser. I, vol. 1, pp. 456–460).
[23] K. Ford, *Sums and products from a finite set of real numbers*, Ramanujan J. **2** (1998), no. 1-2, 59–66.
[24] M. Z. Garaev, *An explicit sum-product estimate in* $\mathbb{F}_p$, Int. Math. Res. Not. IMRN **11** (2007), Art. ID rnm035, 11 pp.
[25] W. T. Gowers, *A new proof of Szemerédi's theorem for arithmetic progressions of length four*, Geom. Funct. Anal. **8** (1998), no. 3, 529–551.
[26] S. Hoory, N. Linial, and A. Wigderson, *Expander graphs and their applications*, Bull. Amer. Math. Soc. (N.S.) **43** (2006), no. 4, 439–561 (electronic).
[27] S. Jukna, *Extremal combinatorics*, Texts in Theoretical Computer Science. An EATCS Series, Springer-Verlag, Berlin, 2001. With applications in computer science.
[28] T. Kövári, V. T. Sós, and P. Turán, *On a problem of K. Zarankiewicz*, Colloquium Math. **3** (1954), 50–57.
[29] J. Matoušek, *Lectures on discrete geometry*, Graduate Texts in Mathematics, vol. 212, Springer-Verlag, New York, 2002.
[30] M. B. Nathanson, *On sums and products of integers*, Proc. Amer. Math. Soc. **125** (1997), no. 1, 9–16.
[31] W. Rudin, *Trigonometric series with gaps*, J. Math. Mech. **9** (1960), 203–227.
[32] J. Solymosi, *On the number of sums and products*, Bull. London Math. Soc. **37** (2005), no. 4, 491–494.
[33] J. Solymosi, *An upper bound on the multiplicative energy*. preprint.
[34] B. Sudakov, E. Szemerédi, and V. H. Vu, *On a question of Erdős and Moser*, Duke Math. J. **129** (2005), no. 1, 129–155.
[35] T. Tao, *The sum-product phenomenon in arbitrary rings* (2008). preprint.
[36] T. Tao and V. Vu, *Additive combinatorics*, Cambridge Studies in Advanced Mathematics, vol. 105, Cambridge University Press, Cambridge, 2006.

Noga Alon
Sackler School of Mathematics and Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv, 69978, Israel, and
Microsoft-Israel R&D Center, Herzeliya, 46725, Israel.
E-mail address: nogaa@tau.ac.il

Omer Angel
Department of Mathematics, University of British Columbia, Vancouver, BC V6T-1Z2, Canada.
E-mail address: angel@math.ubc.ca

Itai Benjamini
Weizmann Institute, Rehovot, 76100, Israel.
E-mail address: itai.benjamini@weizmann.ac.il

Eyal Lubetzky
Microsoft Research, One Microsoft Way, Redmond, WA 98052-6399, USA.
E-mail address: eyal@microsoft.com