On distinct sums and distinct distances

Gábor Tardos
Rényi Institute, Budapest

Abstract The paper [10] of J. Solymosi and Cs. Tóth implicitly raised the following arithmetic problem. Consider n pairwise disjoint s element sets and form all $\binom{s}{2}n$ sums of pairs of elements of the same set. What is the minimum number of *distinct* sums one can get this way? This paper proves that the number of distinct sums is at least n^{d_s}, where $d_s = 1/c_{\lceil s/2 \rceil}$ is defined in the paper and tends to e^{-1} as s goes to infinity. Here e is the base of the natural logarithm. As an application we improve the Solymosi-Tóth bound on an old Erdős problem: we prove that n distinct points in the plane determine $\Omega(n^{\frac{4e}{5e-1}-\epsilon})$ distinct distances, where $\epsilon > 0$ is arbitrary. Our bound also finds applications in other related results in discrete geometry. Our bounds are proven through an involved calculation of entropies of several random variables.

Mathematics Subject Classifications (2000): 52C10, 11B75

1. Introduction

For an n by s matrix $A = (a_{ij})$ we define $S(A) = \{a_{ij} + a_{ik} | 1 \le i \le n, 1 \le j < k \le s\}$ the set of pairwise sums of entries from the same row. Let $f_s(n)$ be the minimum size $|S(A)|$ for a real n by s matrix with all its sn entries being pairwise distinct.

The goal of this paper is to study the asymptotic behavior of $f_s(n)$, especially for large constant values of s.

The motivation for this problem comes from the breakthrough paper of J. Solymosi and Cs. Tóth [10]. They proved an $\Omega(n^{6/7})$ bound for an old problem of P. Erdős [4], the minimum number distinct distances n points determine in the plane. This result substantially improved earlier works of L. Moser [6], F. R. K. Chung [2], F. R. K. Chung, E. Szemerédi and W. T. Trotter [3], and L. Székely [12]. See [7] for the background of this intriguing old Erdős problem and for further references.

Solymosi and Tóth implicitly use $f_3(n) = \Omega(n^{1/3})$ in their proof, and a closer look reveals that any stronger bound for $f_s(n)$, with a constant s would improve their result. Section 4 has the details, Corollary 15 states the bound we get on the number of distinct distances in the plane.

We have that $f_2(n) = 1$, $f_3(n) = \Theta(n^{1/3})$ and $f_4(n) = \Theta(n^{1/3})$ but for $f_5(n)$ and above the correct order of magnitude is unknown. The best current bounds for f_5 and f_6 are:

$$n^{4/11} \le f_5(n) \le f_6(n) = O(n^{2/5}).$$

These bounds are special cases of a construction of Imre Ruzsa and Theorem 1 below. This special case of Theorem 1 (with a worse constant factor) has a much simpler proof than the full theorem as shown in Section 5.

The best upper bound on $f_s(n)$, i.e., the best construction is due to Imre Ruzsa [9], he proves

$$f_s(n) = O(n^{\frac{1}{2}-\frac{1}{2s-2}})$$

for even s. The lower bound of the following Theorem is stated for odd values of s. It is interesting to note that both the known lower and upper bounds are identical for the functions $f_{2k-1}(n) \le f_{2k}(n)$ but for $k \ge 3$ we have no other indication for these functions being close to each other.

Theorem 1. *For an integer $k \ge 2$ we have*

$$f_{2k-1}(n) \ge n^{\frac{1}{c_k}},$$

where for $k \le 14$ we have

$$c_k = \sum_{i=0}^{k} \frac{1}{i!} + \frac{1}{(k-1)k!},$$

while for $k \geq 14$ we have

$$c_k = \sum_{i=0}^{k} \frac{1}{i!} + \frac{k^3 - 7k^2 + 20k - 40}{(k^4 - 8k^3 + 26k^2 - 46k + 40)k!}.$$

Notice, that both definitions of c_k gives the same value for c_{14}.

It is easy to see, that the limit of the values c_k as k goes to infinity is e, the base of the natural logarithm. Thus we have the following

Corollary 2. *For every $\epsilon > 0$ we have a positive integer $s = s(\epsilon)$ with*

$$f_s(n) \geq n^{1/e-\epsilon}.$$

Note, that the limit of the exponent in the Ruzsa construction is $1/2$, so the lower and upper bounds are far apart.

In Sections 2 and 3 we give the proof of Theorem 1. In Section 4 we apply it (or rather Corollary 2) to get an improvement over the Solymosi-Tóth bound on the number of distinct distances n point determine in the plane (see Corollary 15). We also give references to other related problems where Corollary 2 could be used in discrete geometry. In Section 5 we give an elementary and simple proof of the first nontrivial case of Theorem 1: we prove that $f_5(n) = \Omega(n^{4/11})$. We close the paper with concluding remarks and open problems in Section 6.

2. The proof—Reduction to a linear program

Let us fix the positive integers s, n and an n by s real matrix A. Our proof does not use in full generality the assumption that all entries of A are distinct. It is enough to make the slightly weaker assumption that no two rows of A have *two* common entries. Our goal is to prove a lower bound on $|S(A)|$.

Let $I = \{1, 2, 3, \ldots, s\}$ be the set of column indices. For subsets $U, V \subseteq I$ and for an s-tuple $R = (a_1, \ldots, a_s)$ we define the *UV pattern* $p_{U,V}(R)$ of R to be a sequence of real numbers consisting of the *differences* $a_i - a_j$ for $i, j \in U$ and for $i, j \in V$ and the *sums* $a_i + a_j$ for $i \in U$ and $j \in V$. We define

$$H(U, V) = H(p_{UV}(R)),$$

where H denotes the entropy and R is a uniformly distributed random row of A. All entropies and all logarithms in this paper are binary.

The next lemma stating linear constraints on the entropies $H(U, V)$ is crucial for the proof.

Lemma 3. *Let $U, U', V, V' \subseteq I$. We have*
(a) $H(U, V) = H(V, U)$;
(b) $H(U, V) \leq H(U', V')$ if $U \subseteq U'$ and $V \subseteq V'$;
(c) $H(U, V) = 0$ if $U = \emptyset$ and $|V| = 1$;
(d) $H(U, V) \leq \log|S(A)|$ if $U \neq V$ and $|U| = |V| = 1$;
(e) $H(U, V) = \log n$ if $U \cap V \neq \emptyset$ and $|U \cup V| > 1$;
(f) $H(U \cup U', V \cup V') + H(U \cap U', V \cap V') \leq H(U, V) + H(U', V')$ if $(U \cap U') \cup (V \cap V') \neq \emptyset$.

Proof: We use the well known properties of the entropy to prove this lemma.

A. Range. For a random variable F that has k possible values $H(F) \leq \log k$ with equality if F is distributed uniformly.

Part (c) follows since $p_{U,V}(R)$ is constant in that case.

Part (d) also follows since $p_{U,V}(R)$ consists of a single value from $S(A)$ in that case.

Part (e) of the lemma also follows from the above property. The pattern $p_{U,V}(R)$ contains $2a_i$ for the index $i \in U \cap V$ and thus it determines a_i and with it all other values a_j with $j \in U \cup V$. As two entries uniquely determine the row of the matrix A we have that $p_{U,V}(R)$ is different for all the n rows of A, thus it is uniformly distributed among n possible values.

B. Monotonicity. If the value of a random variable F uniquely determines the value of another random variable G then we have $H(F) \geq H(G)$.

Part (b) of the lemma follows as the pattern $p_{U',V'}(R)$ contains all entries of the pattern $p_{U,V}(R)$.

Part (a) of the lemma also follows as the patterns $p_{U,V}(R)$ and $p_{V,U}(R)$ contain the same entries, so they mutually determine each other.

C. Submodularity. Suppose that the value of *either one* of the random variables F_1 and F_2 determines the value of the random variable G_1 and the values of the random variables F_1 and F_2 *together* determine the value of the random variable G_2. In this case we have $H(G_1) + H(G_2) \leq H(F_1) + H(F_2)$.

For part (f) of the lemma we use the submodularity of entropy as stated above. Clearly the pattern $p_{U\cap U',V\cap V'}(R)$ is determined by either one of $p_{U,V}(R)$ and $p_{U',V'}(R)$. We need to show an entry $a_i \pm a_j$ in $p_{U\cup U',V\cup V'}(R)$ is determined by the two patterns $p_{U,V}(R)$ and $p_{U',V'}(R)$. Indeed, the term $a_i \pm a_j$ in the former pattern can be expressed as a sum or difference of the terms $a_i \pm a_k$ and $a_j \pm a_k$ in the latter patterns if $k \in (U \cap U') \cup (V \cap V')$. ■

Lemma 3 contains linear constraints on the entropies $H(U,V)$ and $\log|S(A)|$, thus solving them as a linear program provides a bound on $|S(A)|$. This is indeed the route we will take. The rest of the proof of the lower bound of $|S(A)|$ uses solely Lemma 3. We remark here that the linear program defined by Lemma 3 has a unique optimal solution for all values of s except for $s = 27$ or $s = 28$ where the optimal solutions are the convex combinations of two extremal optimal solutions.

Our first step is to use averaging to decrease the exponential number of variables to less than s^2 of them. For integers $i, j \geq 0$, $1 \leq i + j \leq s$ we define

$$H_{i,j} = 1 - \frac{1}{\log n \binom{n}{i}\binom{n-i}{j}} \sum_{U,V} H(U,V),$$

where the summation extends over all $\binom{n}{i}\binom{n-i}{j}$ pairs of disjoint subsets U and V of I with $|U| = i$ and $|V| = j$. (We consider the values $H_{i,j}$ to form a matrix H with some entries of this matrix missing. We will only use the values $H_{i,j}$ satisfying $0 \leq i, j \leq k$ and $1 \leq i + j \leq 2k - 1$ where $k = \lceil s/2 \rceil$.)

Lemma 4. *For i, j nonnegative integers with $1 \leq i + j \leq s$ we have:*
(a) (symmetry) $H_{i,j} = H_{j,i}$;
(b) (monotonicity) $H_{i,j} \geq H_{i+1,j}$ if $i + j \leq s - 1$;
(c) $H_{0,1} = 1$;
(d) $H_{1,1} \geq 1 - \log|S(A)|/\log n$;
(e) (convexity) $H_{i-1,j} + H_{i+1,j} \geq 2H_{i,j}$ if $i \geq 1$ and $2 \leq i + j \leq s - 1$;
(f) $H_{i,j} \geq H_{i+1,j} + H_{i,j+1}$ if $i + j \leq s - 1$.

We could also state the non-negativity of these variables, but we will not use it.

Proof: Parts (a), (b), (c), and (d) of this lemma follows from the corresponding parts of Lemma 3 by simple averaging.

Part (e) follows from part (f) of Lemma 3, here the averaging is over the four-tuples of sets $U, V, U', V' \subseteq I$ satisfying $|U| = |U'| = i$, $|U \cap U'| = i - 1$, $V = V'$, $|V| = j$ and $(U \cup U') \cap V = \emptyset$.

Finally for part (f) of this lemma consider two disjoint subsets U and V' of I (not both the empty set) and an index $k \in I \setminus (U \cup V')$. Applying Lemma 3(f) for U, $U' = U \cup \{k\}$, $V = V' \cup \{k\}$, and V' one gets

$$H(U,V') + H(U',V) \leq H(U,V) + H(U',V').$$

Here Lemma 3(e) applies and yields $H(U',V) = \log n$, thus we have

$$H(U,V') + \log n \leq H(U,V) + H(U',V').$$

Part (f) of the lemma follows from averaging over all pairs of disjoint subsets $U, V' \subseteq I$ with $|U| = i$ and $|V'| = j$ and for all possible indices k. ■

We remark that the linear program defined by Lemma 4 is already tractable by standard linear programming methods for small values of s but as we will see, considering the cases $s \leq 28$ only can be misleading.

3. Solving the linear program

In this rather technical section we combine the inequalities in Lemma 4 to prove an upper bound on $H_{1,1}$ and thus a lower bound on the size of $S(A)$.

The optimal solution of the linear program in Lemma 4 is the same for an odd number s and for the next even number (and is unique unless s is either 27 or 28). Since our goal is simply to prove a lower bound on $|S(A)|$ we assume $s = 2k - 1$ for some integer $k \geq 3$.

Lemma 5. $H_{k-2,k-1} \leq \frac{3}{k+1} H_{0,k-1}$

Proof: By Lemma 4(e) the columns of the matrix H are convex, therefore we have

$$\frac{H_{0,k-1} - H_{k-2,k-1}}{k-2} \geq H_{k-3,k-1} - H_{k-2,k-1} \geq \frac{H_{k-2,k-1} - H_{k,k-1}}{2}.$$

Using parts (f), (b) and (a) of Lemma 4 we get

$$H_{k-3,k-1} - H_{k-2,k-1} \geq H_{k-3,k} \geq H_{k-1,k} = H_{k,k-1}.$$

Combining the last two displayed inequalities we get

$$\frac{H_{0,k-1} - H_{k-2,k-1}}{k-2} \geq \frac{H_{k-2,k-1} - H_{k,k-1}}{2} \geq \frac{H_{k-2,k-1} - (H_{k-3,k-1} - H_{k-2,k-1})}{2}$$

$$\geq \frac{H_{k-2,k-1} - \frac{H_{0,k-1} - H_{k-2,k-1}}{k-2}}{2},$$

yielding the claimed statement by rearrangement. ■

Lemma 6. *Suppose we have $H_{j-1,j} \leq \alpha H_{0,j}$ for some $3 \leq j < k$ and $\alpha > 0$. If $(j-3)\alpha \leq 2$ then we also have $H_{j-2,j-1} \leq \beta H_{0,j-1}$ for $\beta = (2+\alpha)/(j+\alpha) > 0$ and $(j-4)\beta \leq 2$ is also satisfied.*

Proof: We consider the following four inequalities:

$$H_{0,j} + H_{1,j-1} \leq H_{0,j-1},$$

by Lemma 4(f);

$$\frac{H_{j-2,j-1} - H_{j,j-1}}{2} \leq \frac{H_{0,j-1} - H_{j-2,j-1}}{j-2},$$

by the convexity of column $j-1$ (Lemma 4(e));

$$\frac{H_{j-2,j-1} - H_{j,j-1}}{2} \leq \frac{H_{1,j-1} - H_{j-2,j-1}}{j-3},$$

for $j > 3$ by the convexity of the same column; finally

$$H_{j,j-1} \leq \alpha H_{0,j},$$

by assumption and symmetry (Lemma 4(a)). We sum these inequalities with the non-negative coefficients α, $2-(j-3)\alpha$, $(j-3)\alpha$, and 1, respectively, and rearrange to get the inequality $H_{j-2,j-1} \leq \beta H_{0,j-1}$ as claimed in the lemma. Notice that for $j = 3$ the third inequality is not valid but we use it with zero coefficient. Simple calculation yields the claimed bound on β. ■

We need a closed form for the continued fraction in the next lemma. Note that as a consequence of the lemma, the corresponding infinite continued fraction evaluates to e, the base of the natural logarithm.

Lemma 7. *For an integer $k \geq 1$ and real $x < k^2/(k-1)$ we have*

$$3 - \cfrac{1}{4 - \cfrac{2}{5 - \cfrac{3}{6 - \cdots \; - \cfrac{k-1}{k+1-\cfrac{k-1}{k+2-x}}}}} = \sum_{i=0}^{k} \frac{1}{i!} + \frac{k-x}{(k^2-(k-1)x)k!}.$$

For $k = 1, 2, 3, \ldots$ the left hand side of the equation in the lemma is understood to be

$$3 - x,$$

$$3 - \frac{1}{4-x},$$

$$3 - \cfrac{1}{4 - \cfrac{2}{5-x}},$$

etc.

Proof: The proof is by induction on k. The $k = 1$ case is trivial. For $k > 1$ we use the inductive hypothesis for $k' = k-1$ and $x' = (k-1)/(k+2-x) < (k-1)^2/(k-2)$, and a simple calculation yields the lemma. ■

Instead of Theorem 1 we prove the somewhat stronger statements of Theorems 8 and 10.

Theorem 8. *Let $2 \leq k \leq 14$, $n \geq 1$, and let A be an n by $(2k-1)$ real matrix with no two distinct rows sharing more than a single entry. Then we have*

$$|S(A)| \geq n^{1/c_k},$$

with $c_k = \sum_{i=0}^{k} \frac{1}{i!} + \frac{1}{(k-1)k!}$.

Proof: Previously in this section we assumed $k \geq 3$ so the $k = 2$ ($s = 3$) case must be dealt with separately. One can either solve the linear program of Lemma 4 (there are only four distinct relevant variables in this case) or use direct reasoning as in the beginning of Section 5. This $s = 3$ case was already discussed in [10].

For $k \geq 3$ we prove a bound $H_{j-1,j} \leq \alpha_j H_{0,j}$ by reverse induction on $j = k-1, \ldots, 2$. We use Lemma 5 to get $\alpha_{k-1} = 3/(k+1)$ as the bases of our induction. We use Lemma 6 for the inductive step to get $\alpha_{j-1} = (2+\alpha_j)/(j+\alpha_j)$. Notice that the $(j-3)\alpha_j \leq 2$ condition is satisfied at $j = k-1$ because of the $k \leq 14$ assumption and this condition is preserved by Lemma 6. Rewriting the recursion to $1 - \alpha_{j-1} = (j-2)/((j+1) - (1-\alpha_j))$ and writing $1 - \alpha_{k-1} = (k-2)/(k+1)$ we get the following continued fraction expansion for α_2:

$$1 - \alpha_2 = \cfrac{1}{4 - \cfrac{2}{5 - \cfrac{3}{6 - \cdots \; - \cfrac{k-2}{k+1}}}}.$$

We further have

$$2H_{1,1} - H_{0,1} \leq H_{2,1} = H_{1,2} \leq \alpha_2 H_{0,2} \leq \alpha_2 (H_{0,1} - H_{1,1}),$$

by Lemma 4(e), Lemma 4(a), the statement above, and Lemma 4(f), respectively. By rearrangement, and using $H_{0,1} = 1$ (Lemma 4(c)) we get the bound $H_{1,1} \leq (1+\alpha_2)/(2+\alpha_2)$. By Lemma 4(d) we get

$$\log n / \log |S| \leq \frac{1}{1 - H_{1,1}} \leq 2 + \alpha_2 = 3 - \cfrac{1}{4 - \cfrac{2}{5 - \cfrac{3}{6 - \cdots \; - \cfrac{k-2}{k+1}}}}.$$

Lemma 7 provides a closed form for the continued fraction of the above statement and thus proves the theorem. ■

In order to use Lemma 6 recursively as in the proof of Theorem 8 we need a base case. Lemma 5 cannot be used for this if $k > 14$ as the inequality required for Lemma 6 to apply ($(j-3)\alpha \leq 2$) does not hold for $j = k-1$ and $\alpha = 3/(k+1)$. The next lemma provides the base $j = k-2$ case for large k.

Lemma 9. *If $k \geq 14$ we have $H_{k-3,k-2} \leq \frac{2k+3}{k^2-k+4} H_{0,k-2}$.*

Proof: We combine six inequalities to get the desired bound. We use

$$H_{k-2,k-1} \leq \frac{3}{k+1} H_{0,k-1}$$

and

$$H_{k-2,k-1} \leq \frac{3}{k} H_{1,k-1}.$$

The former inequality is provided by Lemma 5, while the latter inequality can be proven the same way. We also use

$$H_{0,k-1} + H_{1,k-2} \leq H_{0,k-2}$$

and

$$H_{1,k-1} + H_{2,k-2} \leq H_{1,k-2}$$

provided by Lemma 4(f). Finally we also use

$$\frac{H_{k-3,k-2} - H_{k-2,k-1}}{2} \leq \frac{H_{1,k-2} - H_{k-3,k-2}}{k-4}$$

and

$$\frac{H_{k-3,k-2} - H_{k-2,k-1}}{2} \leq \frac{H_{2,k-2} - H_{k-3,k-2}}{k-5},$$

both coming from symmetry (Lemma 4(a)) and the convexity of column $k-2$ (Lemma 4(e)). We sum the above six inequalities with coefficients $(k+1)(2k+3)$, $k(k-14)$, $3(2k+3)$, $3(k-14)$, $3(k-4)(k+17)$, and $3(k-5)(k-14)$, in this order, and rearrange to obtain the bound of the Lemma. Note, that all these coefficients are non-negative if $k \geq 14$. ∎

Theorem 10. *Let $k \geq 14$, $n \geq 1$, and let A be an n by $(2k-1)$ real matrix with no two distinct rows sharing more than a single entry. Then we have*

$$|S(A)| \geq n^{1/c_k},$$

with $c_k = \sum_{i=0}^{k} \frac{1}{i!} + \frac{k^3-7k^2+20k-40}{(k^4-8k^3+26k^2-46k+40)k!}$.

Proof: We copy the proof Theorem 8. We prove a bound $H_{j-1,j} \leq \alpha_j H_{0,j}$ by reverse induction on $j = k-2, \ldots, 2$. We use Lemma 9 for the base case $j = k-2$ and $\alpha_{k-2} = (2k+3)/(k^2-k+4)$. Lemma 6 gives $\alpha_{j-1} = (2+\alpha_j)/(j+\alpha_j))$ since $(j-3)\alpha_j \leq 2$ for $j = k-2$ and so it is true for all values of j considered. As in the proof of Theorem 8 we get a continued fraction expansion of α_2, in this case it is:

$$1 - \alpha_2 = \cfrac{1}{4 - \cfrac{2}{5 - \cfrac{3}{6 - \cdots \; k-2 - \cfrac{k-4}{k-1 - \cfrac{k^2-3k+1}{k^2-k+4}}}}}.$$

Just as in the proof of Theorem 8 we get

$$\log n / \log |S| \leq 2 + \alpha_2.$$

Thus we have

$$\log n / \log |S| \leq 3 - \cfrac{1}{4 - \cfrac{2}{5 - \cfrac{3}{6 - \cdots \; k-2 - \cfrac{k-4}{k-1 - \cfrac{k^2-3k+1}{k^2-k+4}}}}}.$$

Lemma 7 provides a closed form for the continued fraction of the above statement and thus proves the Theorem. ∎

4. Distinct distances in the plane

As mentioned in the introduction, the problem considered in this paper is a byproduct of the paper [10] by J. Solymosi and Cs. Tóth. One of their lemmas can be stated in our notation as follows.

Lemma 11. [10, Lemma 5] *Let $A = (a_{ij})$ be an n by 3 real matrix with all its entries pairwise distinct. Assume that $a_{i1} < a_{i2} < a_{i3}$ for $i = 1, \ldots, n$, and assume also that $a_{i3} < a_{i+1,1}$ for all but at most $t-1$ indices $i = 1, \ldots, n-1$. Then*

$$|S(A)| = \Omega(N/t^{2/3}).$$

We generalize this lemma as follows. We include the simple proof along the same lines.

Lemma 12. *Let s, and $t \leq n$ be positive integers and let $A = (a_{ij})$ be an n by s real matrix with all the ns entries pairwise distinct. Assume that $\max_j a_{ij} < \min_j a_{i+1,j}$ holds for all but at most $t-1$ of the indices $i = 1, \ldots, n-1$. Then*

$$|S(A)| \geq \left\lfloor \frac{n}{2t} \right\rfloor \cdot f_s(t).$$

Proof: We find $\lfloor \frac{n}{2t} \rfloor$ pairwise disjoint real intervals, each containing all entries of t rows of A. This can be done from left to right on the real line using a greedy strategy. Let A_I be the submatrix of A consisting of the t rows fully contained in the interval I. Clearly $|S(A_I)| \geq f_s(t)$ and $S(A_I) \subseteq S(A)$. Furthermore $S(A_I)$ and $S(A_{I'})$ are disjoint if I and I' are disjoint, yielding the lemma. ■

With the help of this lemma we get:

Theorem 13. *Let us be given n points in the plane such that from any one of the points there are at most t distinct distances to the other points. Then*

$$\frac{t^5}{f_s(t)} \geq cn^4,$$

where $c = c(s)$ is a positive constant depending on s.

Proof sketch: Solymosi and Tóth in [10] prove that n points in the plane determine $\Omega(n^{6/7})$ distinct distances. Their proof can be considered the $s = 3$ special case of our proof. The beautiful proof is based on the method of L. Székely, uses the crossing number theorem of M. Ajtai, V. Chvátal, M. Newborn, and E. Szemerédi [1] and F. T. Leighton [5], and the point-line incidence theorem of E. Szemerédi and W. T. Trotter [13]. As most of the proof goes through without a change we only sketch the differences and refer the reader to the original proof for details.

We consider the same incidences between points and circles as in [10]. We partition the points incident to a circle into s-tuples (rather than triplets as in [10]). We construct a topological graph by connecting at most a single pair of points along the circle from every s-tuple, a pair with a bisector not *rich*, i.e., not going through more than $\epsilon n^2/t^2$ points. If no such pair exists we call the s-tuple *bad*. We deduce (as in [10]) from the crossing number theorem that most of the s-tuples are bad. Now we contrast the upper bound of the Szemerédi-Trotter theorem on incidences between points and rich lines, to the lower bound obtained from Lemma 12 (in place of Lemma 11 in [10]). All calculations of the paper go through with this modification and they yield Theorem 13. ■

The next corollary is straightforward consequence of Theorem 13, while Corollary 15 below is a consequence of Corollaries 2 and 14.

Corollary 14. *If for some positive integer s and positive real α we have $f_s(n) = \Omega(n^\alpha)$ then the following holds. Any set P of n points in the plane has an element from which the number of distinct distances to the other points of P is*

$$\Omega(n^{\frac{4}{5-\alpha}}).$$

Corollary 15. *For any $\epsilon > 0$ we have that any set P of n points in the plane has an element from which the number of distinct distances to the other points of P is*

$$\Omega(n^{\frac{4e}{5e-1}-\epsilon}).$$

Subsequent to the paper [10], it turned out that the same proof technique can be used to prove generalizations of the result in [10]. The bound on the number of appearances of the k most frequent distances among n points in the plane in [11] and the bound on the number of equilateral triangles n points in the

plane determine in [8] both imply Corollary 15, and both heavily rely on Corollary 2, the main result of this paper.

5. An elementary proof

In this section we consider f_s for small values of s. Trivially $f_1(n) = 0$, $f_2(n) = 1$.
We claim that $f_3(n) \approx (6n)^{1/3}$, more precisely,

$$f_3(n) = \min\left\{m \middle| \binom{m}{3} \geq n\right\}.$$

Indeed, each row of an n by 3 matrix A of all distinct entries determines three distinct sums in $S(A)$ and these three sums in turn determine the entries and thus the row. Thus $\binom{f_3(n)}{3} \geq n$ must hold. To see the claimed equality we construct for an arbitrary integer $m \geq 3$ a matrix A_S of $n = \binom{m}{3}$ rows and three columns from a rationally independent set S of m reals. Each row of A_S is of the form

$$(\frac{x+y-z}{2}, \frac{x-y+z}{2}, \frac{-x+y+z}{2}),$$

where $\{x, y, z\}$ is a different three element subset of S for each row. Notice that all entries of A_S are different and $S(A_S) = S$.

It is easy to see, that $f_3(n) \leq f_4(n) \leq 2f_3(n)$, thus f_4 has the same order of magnitude as f_3 (namely $n^{1/3}$). Indeed add an extra column to the matrix A_S constructed above making each row add up to zero. Notice that the modified matrix A'_S has still all different entries and furthermore $S(A'_S) = S \cup (-S)$. More precisely, we claim that

$$\min\left\{m \middle| \binom{m}{3} \geq 4n\right\} \leq f_4(n) \leq 2\min\left\{m \middle| 2\binom{m}{3} \geq n\right\}.$$

Indeed, each three element subset of a row of an n by 4 matrix A of all different entries determine three different sums in $S(A)$ and a moment notice verifies that all $4n$ triplets obtained this way are distinct proving the lower bound on $f_4(n)$. For the upper bound consider an m element rationally independent set S and the $\binom{m}{3}$ by 4 matrix A'_S constructed above. Notice that by using all rows of A'_S and $-A'_S$ we get a $2\binom{m}{3}$ by 4 matrix $A''(S)$ of all different entries with $S(A''_S) = S(A'_S) = S \cup (-S)$. Notice that the lower and upper bounds differ by at most 3 (and this can be further reduced by simple observations).

The observations above on f_3 appeared already in the paper of Solymosi and Tóth [10] and were used to prove a lower bound on the number of distinct distances n points determine in the plane. It was clear from their proof, that an improved lower bound for f_s even if for a higher constant value of s would improve their bound on the number of distances. It was independently found by Gyula Károlyi and Tibor Szabó that f_4 has the same order of magnitude as f_3 and thus it cannot be used in this manner.

Next we consider $f_5(n)$ for which the actual order of magnitude is not known, but Theorem 1 gives $f_5(n) \geq n^{4/11}$. Here we give an elementary proof of this result (with a constant multiplicative factor in the bound). This proof can serve as a motivation for the general result, the techniques of the proof of Theorem 1 were introduced to generalize this elementary proof below. Note also that the lower bound for $f_s(n)$ in Theorem 1 is very close to the lower bound presented here: for arbitrary $s \geq 7$ the improvement in the exponent of n is only in the third digit after the decimal point.

Theorem 16. $f_5(n) = \Omega(n^{4/11})$.

Proof: Let A be an n by 5 real matrix of all distinct entries with $S = S(A)$ having minimal size $|S| = m = f_5(n)$. We call a real value x *heavy* if it can be expressed as $x = u - v$ with $u, v \in S$ in at least $m^{1/4}$ different ways and x is *light* otherwise. We call a row of A *heavy* it has two distinct entries b and c with $b - c$ being heavy, otherwise the row is called *light*.

Our goal is to bound the number n of rows of A in terms of m and we do this separately for heavy and light rows.

Clearly there are only m^2 ways to form a difference $x = u - v$ from values $u, v \in S$, so there are at most $m^2/m^{1/4} = m^{7/4}$ heavy numbers. Given any value x it can be expressed as the difference between two distinct entries of the same row of A in only m different ways, as the sum of these two entries is a value in S and the sum and the difference together determines the entries (we use here that all entries of A are distinct). The last two statements together bound the number of heavy rows in A by $m^{7/4} \cdot m = m^{11/4}$.

Now consider a light row $(a_1, a_2, a_3, a_4, a_5)$ of A. We identify this row with revealing limited information, and use this to bound the number of light rows. We first reveal $u = a_1 + a_2$ and $u' = a_1 + a_3$. Both numbers are from the m-element set S so we have m^2 choices for these values. With this we also identified $u - u' = a_2 - a_3$ which must be a light number. Thus writing $u - u' = (a_2 + a_4) - (a_3 + a_4) = (a_2 + a_5) - (a_3 + a_5)$ are two of the less than $m^{1/4}$ ways $u - u'$ can be expressed as the difference of two values in S. So we have less than $(m^{1/4})^2 = m^{1/2}$ choices when revealing the values $v = a_2 + a_4$, $v' = a_2 + a_5$. Just as above, the value $v - v' = a_4 - a_5$ is now revealed, and it must be light number, thus $v - v' = (a_1 + a_4) - (a_1 + a_5)$ is one of the less than $m^{1/4}$ ways $v - v'$ can be expressed as a difference of values in S. So for revealing $w = a_1 + a_4$ we have less than $m^{1/4}$ choices. At this point $a_1 = (u + w - v)/2$ is also revealed, so the row itself is identified (as all entries of A are distinct). This limits the number of light rows by $m^2 \cdot m^{1/2} \cdot m^{1/4} = m^{11/4}$.

Thus $n < 2m^{11/4}$ and so $f_5(n) = m > (n/2)^{4/11}$. ∎

6. Concluding remarks and open problems

As we have already mentioned the orders of magnitude of the functions $f_s(n)$ are unknown for $s \geq 5$. Improving either the lower or the upper bounds is a challenge.

One could hope to improve the lower bounds by the methods of this paper, i.e., considering a uniformly distributed random row of a matrix with all distinct entries, several linear functions on this row, and using inequalities on the (joint) entropies of these functions. There is room for improvement. In this paper we restricted our attention to specific collections of pairwise sums and differences only. One can try use different collections, like $H(a+b, c+d)$ or even linear combinations of a different type, like $H(a+b+c+d)$. Here a, b, c, and d represent distinct entries of the random row. Submodularity gives a lot of inequalities, like

$$H(a+b, b+c, c+d) + H(a+b+c+d) \leq H(a+b, c+d) + H(b+c, d+a),$$

but the author was unable to use these additional inequalities to obtain better bounds on f_s. One could also try to use information inequalities that are not consequences of the basic inequalities of Shannon type (monotonicity and submodularity). Such inequalities were published by Zhen Zhang and Raymond W. Yeung [14] but they seem to be too complicated to easily lend themselves to applications.

The original motivation for the problem considered in this paper is its applicability to the Erdős problem of finding the minimum number of distinct distances n points determine in the plane as formulated in Corollary 15. The Ruzsa construction shows that one cannot use this theorem directly to prove an $\Omega(n^{8/9})$ bound on the number of distinct distances. One may try to modify the lower bound proof on the number of distinct distances by letting the parameter s grow with n. Unfortunately a simple modification of the Ruzsa construction is still in the way of proving $\Omega(n^{8/9+\epsilon})$ in this way. Indeed, one can show that for every $\epsilon > 0$ there exists $\delta > 0$ with

$$f_{n^\delta}(n) = O(n^{1/2+\epsilon}).$$

Acknowledgments

The author is indebted to János Pach for enlightening discussions. Fruitful conversations with Nati Lineal, Imre Ruzsa, Gábor Simonyi, and Csaba Tóth are also acknowledged.

References

[1] M. Ajtai, V. Chvátal, M. Newborn, and E. Szemerédi, Crossing free graphs, *Ann. Discrete Math.* **12** (1982), 9–12.

[2] F. R. K. Chung, The number of different distances determined by n points in the plane, *J. Combin. Theory Ser. A* **36** (1984), 342–354.

[3] F. R. K. Chung, E. Szemerédi, and W. T. Trotter, The number of different distances determined by a set of points in the Euclidean plane, *Discrete and Computational Geometry* **7** (1992), 1–11.
[4] P. Erdős, On the set of distances of n points, *Amer. Math. Monthly* **53** (1946), 248–250.
[5] F. T. Leighton, Complexity Issues in VLSI, MIT Press, Cambridge, MA, 1983.
[6] L. Moser, On the different distances determined by n points, *Amer. Math. Monthly* **59** (1952), 85–91.
[7] J. Pach and P. K. Agarwal, Combinatorial Geometry, Wiley, New York, 1995.
[8] J. Pach, G. Tardos, On equilateral triangles determined by a finite point set in the plane, manuscript, (2001).
[9] I. Ruzsa, personal communication.
[10] J. Solymosi, Cs. Tóth, Distinct distances in the plane, *Discrete and Computational Geometry* **25** (2001), 629–634.
[11] J. Solymosi, Cs. Tóth, G, Tardos, The k most frequent distances in the plane, submitted to *Discrete and Computational Geometry*.
[12] L. Székely, Crossing numbers and hard Erdős problems in discrete geometry, *Combin. Probab. Comput.* **6** (1997), 353–358.
[13] E. Szemerédi and W. T. Trotter, extremal problems in discrete geometry, *Combinatorica* **3** (1983), 381–392.
[14] Z. Zhang, R. W. Yeung, On characterization of entropy function via information inequalities, *IEEE Transactions on Information Theory* **44** (4) (1998), 1440-1452.