Blocking partial designs and block-compatible sequences

Noga Alon, Princeton, nalon@math.princeton.edu

Abstract

We prove that there is a partial design with n vertices in which each block is of size at least $\Omega(\sqrt{n})$, so that every set that intersects all blocks contains at least $\Omega(\log n)$ points of one of them. We also show that the number of sequences $n \geq x_1 \geq x_2 \geq x_m \geq 2$ so that there is a block design on n elements with blocks of sizes $x_1, x_2, \ldots, x_m$ is at least $2^{\Omega(n^{1/2} \log n)}$. This settles two problems of Erdős.

1 Results

We consider two related open problems of Erdős on block designs. Recall that a family of subsets $A_1, A_2 \ldots, A_m$ of a finite set X is a (pairwise balanced) block design if every pair of distinct elements of X is contained in exactly one of the subsets A_i. It is a partial design if every pair of distinct elements of X is contained in at most one of the subsets A_i (equivalently, if $|A_i \cap A_j| \leq 1$ for all $1 \leq i < j \leq m$.)

The first problem deals with partial designs and appears in [6], see also [4], problem number 664.

Problem 1.1. *Is it true that for every fixed positive constant $c < 1$ there is a finite constant $C = C(c)$ so that the following holds. For every m and n and for every family of subsets $\{A_1, A_2, \ldots, A_m\}$ of $[n] = \{1, 2, \ldots, n\}$ that satisfies $|A_i| > c\sqrt{n}$ for all $1 \leq i \leq m$, and $|A_i \cap A_j| \leq 1$ for all $1 \leq i < j \leq m$, there is a subset $B \subset [n]$ so that $0 < |B \cap A_i| \leq C$ for all $1 \leq i \leq m$?*

The second problem appears in [5], page 35, see also [4], problem number 732.

Problem 1.2. *Call a sequence $n \geq x_1 \geq x_2 \geq \ldots \geq x_m \geq 2$* block-compatible for *$n$ if there is a pairwise balanced block design $A_1, A_2, \ldots, A_m$ of m subsets of $[n]$ such that $|A_i| = x_i$ for $1 \leq i \leq m$. Is there an absolute constant $c > 0$ so that for all large n there are at least $e^{cn^{1/2} \log n}$ sequences that are block-compatible for n ?*

In this note we show that the answer to the first problem is "no" and the answer to the second is "yes". The proofs are short, based on appropriate modifications of the family of lines of a projective plane which form a block design with $m = n = q^2 + q + 1$ subsets of cardinality $q + 1 = (1 + o(1))\sqrt{n}$ of a set of size $n = q^2 + q + 1$. It is well known that such a plane exists for any prime power q.

2 Proofs

Throughout the proofs we assume, whenever this is needed, that the parameter n is sufficiently large. All logarithms are in base 2, unless otherwise specified. To simplify the presentation we omit all floor and ceiling signs whenever these are not crucial. We make no serious attempt to optimize the absolute constants that appear in the proofs.

The following result settles Problem 1.1

Theorem 2.1. *Let q be a (large) prime power and put $m = n = q^2 + q + 1$. Then there is a partial design consisting of m subsets $A_1, A_2, \ldots, A_m$ of an n element set P, so that $|A_i| > 0.4\sqrt{n}$ for all $1 \le i \le m$, $|A_i \cap A_j| \le 1$ for all $1 \le i < j \le m$, and for any subset B of P that has a nonempty intersection with all sets A_j, there is some $1 \le i \le m$ so that $|B \cap A_i| \ge 0.1 \log n$.*

Proof. Let P be the set of $n = q^2 + q + 1$ points of a projective plane of order q, and let $L_1, L_1, \ldots, L_m \subset P$ be the sets of points of the m lines of this plane. Thus each point lies in $q + 1$ lines and each line is of size $q + 1$. For each $1 \le i \le m$ let A_i be a random subset of L_i obtained by picking every point of L_i, randomly and independently, to lie in A_i with probability $1/2$. Note that the choices of distinct subsets A_i here are independent. By the standard estimates for Binomial distributions (c.f., e.g., [1], Appendix A) together with the union bound, with high probability (that is, with probability tending to 1 as q (or n) tend to infinity) the following conditions hold:

1. Each set A_i is of cardinality $(1/2 + o(1))\sqrt{n}$.
2. Each point lies in $(1/2 + o(1))\sqrt{n}$ of the sets A_i.
3. $|A_i \cap A_j| \le 1$ for all $1 \le i < j \le m$.

Claim 2.2. *With high probability there is no subset $B \subset P$ of cardinality at most $0.3\sqrt{n}\log n$ that intersects all the sets A_i.*

Proof: Let B be a fixed set of cardinality at most $0.3\sqrt{n}\log n$. Consider the set of pairs $S = \{(b, L_i) : b \in B, 1 \le i \le m, b \in L_i\}$. Clearly $|S| = |B|(q + 1)$. If $J = J(B)$ is the set of all indices i so that $|B \cap L_i| \ge 0.4 \log n$ then $|S| \ge |J| 0.4 \log n$. Thus

$$|J| \le \frac{|B|(q + 1)}{0.4 \log n} \le (3/4 + o(1))n < 0.8n.$$

Therefore there are at least $0.2n$ lines L_i that contain less that $0.4 \log n$ points of B. For each such line L_i, the probability that A_i contains no point of B is at least $2^{-0.4 \log n} \ge n^{-0.4}$. Therefore, the probability that there is no such line (that is, that B intersects all these sets A_i) is at most

$$(1 - 1/n^{0.4})^{0.2n} \le e^{-0.2n^{0.6}}.$$

Since the total number of possible sets B as above is smaller than

$$n^{0.3\sqrt{n}\log n} = 2^{0.3\sqrt{n}(\log n)^2} = o(2^{0.2n^{0.6}}) \ (\ = o(e^{0.2n^{0.6}}) \ ),$$

the union bound implies the assertion of the claim.

Returning to the proof of the theorem, fix a choice of the sets A_i that satisfy the conditions 1, 2, 3 above and the assertion of Claim 2.2. If a set B intersects all subsets A_i then it must satisfy $|B| > 0.3\sqrt{n}\log n$. Since each point is contained in $(1/2+o(1))\sqrt{n}$ of the sets A_i this implies, by averaging, that the intersection of B with some set A_i is of size at least $|B|(1/2+o(1))\sqrt{n}/n > 0.1\log n$. This completes the proof. □

The next result settles problem 1.2.

Theorem 2.3. *Let q be a large prime power and put $n = q^2+q+1$. Let $S = (x_1 \geq x_2 \geq x_3 \geq \ldots \geq x_m)$ be any sequence of integers satisfying*

$$q+1 \geq x_1 \geq x_2 \geq x_3 \ldots \geq x_n \geq 3,$$

$$m = n + \sum_{i=1}^{n}[\binom{q+1}{2} - \binom{x_i}{2}],$$

and $x_i = 2$ for all $n < i \leq m$. Then S is block-compatible for n. Therefore, the number of sequences that are block-compatible for n is at least

$$\binom{n+q-2}{q-2} = 2^{(0.5+o(1))n^{1/2}\log n}.$$

Proof. Let P and $L_1, L_2, \ldots, L_n \subset P$ be, as in the proof of Theorem 2.1, the set of points of a projective plane of order $q+1$ and the sets of points of the n lines of P. For each $1 \leq i \leq n$ let X_i be a subset of cardinality x_i of L_i. Consider the block design consisting of the n blocks X_i together with the following additional

$$\sum_{i=1}^{n}[\binom{q+1}{2} - \binom{x_i}{2}]$$

blocks of cardinality 2: for each $1 \leq i \leq n$, every pair of distinct elements of L_i which is not contained in X_i. This is clearly a block design and the ordered sequence of cardinalities of its blocks is the sequence S. This completes the first part of the proof. For the estimate note that the number of possibilities for the subsequence $x_1 \geq x_2 \geq \ldots \geq x_n \geq 3$ in the above construction is the number of ordered sequences of $q-1$ nonnegative integers whose sum is n, which is

$$\binom{n+q-2}{q-2}.$$

□

3 Concluding remarks

- It is easy to see that the estimate in Theorem 2.1 is tight up to constant factors, for every partial design in which all blocks are of sizes $\Theta(\sqrt{n})$. Indeed the following simple fact can be proved by choosing the set B randomly and by applying the union bound and the standard estimates for Binomial distribution.

  Proposition 3.1. *For any two positive constants $c_1 < c_2$ there are two positive constants $C_1(c) < C_2(c)$ so that the following holds. For every m and n and for every family of subsets $\{A_1, A_2, \ldots, A_m\}$ of $[n] = \{1, 2, \ldots, n\}$ that satisfies $c_1\sqrt{n} \le |A_i| \le c_2\sqrt{n}$ for all $1 \le i \le m$, and $|A_i \cap A_j| \le 1$ for all $1 \le i < j \le m$, there is a subset $B \subset [n]$ so that $C_1 \log n < |B \cap A_i| \le C_2 \log n$ for all $1 \le i \le m$.*

  In fact the conclusion holds, of course, even without any assumption on the sizes of the intersections of pairs of blocks.

- Problem 1.1 for block designs (and not for partial designs) was also asked by Erdős in [5]. This remains open although we suspect that the answer here is negative as well. We suggest the following conjecture which, if true, would establish this negative answer.

  Conjecture 3.2. *Let q be a (large) prime power, put $n = q^2 + q + 1$, let P be the set of n points of a projective plane of order q and let $L_1, L_2, \ldots, L_n$ be the sets of points of its lines. Let R be a random subset of P obtained by picking each point of P randomly and independently to lie in R with probability $1/2$. Then with high probability the smallest cardinality of a subset B of R that intersects all the subsets $L_1 \cap R, L_2 \cap R, \ldots, L_n \cap R$ satisfies $|B|/q > f(q)$ for some function $f(q)$ tending to infinity as q tends to infinity. In fact, this may even be true with $f(q) = \Omega(\log q)$.*

  This conjecture remains open, although related results have been proved in [2], [3] using the container method. The parameters in these papers are very different and it seems that a proof here would require additional ideas.

- As mentioned by Erdős in [5], the $e^{cn^{1/2}\log n}$ lower bound for the number of block-compatible sequences for n is tight up to the absolute constant c. For completeness we include a brief proof of the upper bound. It is clear that if $n \ge x_1 \ge \ldots \ge x_m \ge 2$ is block-compatible for n, then $m \le \binom{n}{2} < n^2/2$. Therefore, the number of choices of all x_i which are, say, at most $2\sqrt{n}$ is smaller than
$$\binom{n^2/2 + 2\sqrt{n}}{2\sqrt{n}} < 2^{4\sqrt{n}\log n}.$$
  Next, observe that for each block-compatible sequence x_i for n, the number of indices i with $x_i > 2\sqrt{n}$ is smaller than $\sqrt{n}$. Indeed, otherwise the size of the union of a set of blocks

$A_i, i \in I$ with $|A_i| = x_i > 2\sqrt{n}$ and $|I| = \sqrt{n}$ in a block design realizing the sequence is at least

$$\sum_{i \in I} |A_i| - \sum_{i,j \in I, i<j} |A_i \cap A_j| > \sqrt{n} \cdot 2\sqrt{n} - \binom{\sqrt{n}}{2} > n,$$

which is impossible. Thus there are less than $\sqrt{n}$ such large x_i, and the number of choices for those is at most

$$\binom{n+\sqrt{n}}{\sqrt{n}} < 2^{\sqrt{n}\log n}.$$

This gives the required $2^{O(\sqrt{n}\log n)}$ upper bound for the total number of block-compatible sequences for n.

- Call a sequence $n \geq x_1 \geq x_2 \geq \ldots \geq x_m \geq 2$ *line-compatible for* n if there is a set P of n points in the Euclidean plane R^2 so that for the family $L_1, L_2, \ldots, L_m$ of all lines in R^2 determined by the points of P, $|L_i \cap P| = x_i$ for $1 \leq i \leq m$. Note that every line-compatible sequence for n is also block-compatible for n, but the converse is not true. Erdős conjectured in [5] (see also [4], problem 733) that the number of sequences which are line compatible for n is only $2^{O(n^{1/2})}$. This upper bound was proved by Szemerédi and Trotter in [9]. Note that in view of Theorem 2.3 this is much smaller than the number of block-compatible sequences for n.

Indeed, there are far more block designs on n points than designs that can be described by the lines determined by a set of points in the plane. This is demonstrated by the following result.

Proposition 3.3. *1. The number of hypergraphs on n labelled vertices whose edges form a block design is $2^{\Theta(n^2\log n)}$.*

2. The number of hypergraphs whose vertices are n labelled points in R^2 and whose edges are the sets of points contained in the lines determined by the points is only $2^{\Theta(n\log n)}$.

Proof. The lower bound in the first part follows from the known lower bound for the number of Steiner triple systems on n points, which as proved in [7] is $2^{(1+o(1)n^2\log n/6}$. To prove the upper bound let $x_1, x_2, \ldots x_m$ be the sizes of the blocks. Then $\sum_i \binom{x_i}{2} = \binom{n}{2}$ and thus $\sum x_i \leq n^2$. The number of choices of the number of blocks m and their sizes x_i is at most $2^{O(\sqrt{n}\log n)}$. Given those, the number of ways to choose subsets of cardinalities $x_1, x_2, \ldots, x_m$ in $[n]$ is at most

$$\binom{n}{x_1}\binom{n}{x_2}\cdots\binom{n}{x_m} < n^{\sum_i x_i} \leq n^{n^2} = 2^{n^2\log n}.$$

The lower bound in the second part is proved by considering all possible labelings of the hypergraph consisting of a matching of $\lfloor n/3 \rfloor$ pairwise disjoint edges of size 3 and additional $\binom{n}{2} - \lfloor n/3 \rfloor$ edges of size 2. It is easy to realize this hypergraph by points in the plane and the lines they determine by placing 3 points on each of $\lfloor n/3 \rfloor$ parallel lines where no other line contains more than two points.

We prove the upper bound in the second part by using the result of [8] on the number of zero patterns of a sequence of polynomials.

If $f = (f_1, \ldots, f_a)$ is a sequence of polynomials in b variables over a field K, then the *zero pattern* of f evaluated at the point $u \in K^b$ is the set

$$Z_f(u) = \{i \in [a] : f_i(u) = 0\}.$$

Let Z_f denote the total number of distinct zero patterns that appear as u ranges over K^b. We need the following result of Rónyai, Babai, and Ganapathy [8].

Theorem 3.4 ([8])**.** *Let $f = (f_1, \ldots, f_a)$ be a sequence of polynomials in b variables over a field K, and let d_i denote the degree of f_i. Then*

$$Z_f \le \binom{b + \sum_{i=1}^{a} d_i}{b}.$$

In our case the field is R, each point can be described by 2 real variables, so $b = 2n$, and for each set of 3 points there is a degree 3 polynomial that vanishes iff they lie on a line. The zero pattern of these $\binom{n}{3}$ real polynomials of degree 3 in $2n$ variables determines all the lines, and therefore, by Theorem 3.4, the number of possible hypergraphs here is at most

$$\binom{2n + (n^3/6)3}{2n} = 2^{(4+o(1))n \log n}.$$

This completes the proof. □

References

[1] Noga Alon and Joel H. Spencer, *The Probabilistic Method, Fourth Edition*, Wiley, 2016, xiv+375 pp.

[2] József Balogh and Wojciech Samotij, An efficient container lemma, Discrete Anal. 2020, Paper No. 17, 56 pp.

[3] József Balogh and József Solymosi, On the number of points in general position in the plane, Discrete Anal. (2018), Paper No. 16, 20 pp.

[4] Thomas Bloom, Erdős' Open Problems, https://www.erdosproblems.com/all

[5] Paul Erdős, On the combinatorial problems which I would most like to see solved, Combinatorica (1981), 25–42.

[6] Paul Erdős, Some unsolved problems, in: Combinatorics, geometry and probability (Cambridge, 1993) (1997), 1-10.

[7] Peter Keevash, Counting Steiner triple systems, European Congress of Mathematics, 459–481, Eur. Math. Soc., Zürich, 2018.

[8] Lajos Rónyai, László Babai and Murali K. Ganapathy, On the number of zero-patterns of a sequence of polynomials, J. Amer. Math. Soc. 14 (2001), 717–735.

[9] Endre Szemerédi and William T. Trotter, Extremal problems in discrete geometry, Combinatorica 3 (1983), no. 3-4, 381–392.