Large sets of nearly orthogonal vectors

Noga Alon * Mario Szegedy †

Abstract

It is shown that there is an absolute positive constant $\delta > 0$, so that for all positive integers k and d, there are sets of at least $d^{\delta \log_2(k+2)/\log_2 \log_2(k+2)}$ nonzero vectors in R^d, in which any $k+1$ members contain an orthogonal pair. This settles a problem of Füredi and Stanley.

1 Introduction

For two positive integers d and k, let $\alpha(d,k)$ denote the maximum possible cardinality of a set of nonzero vectors in R^d such that among any $k+1$ members of the set there is an orthogonal pair. More generally, for three positive integers d and $k \geq l \geq 1$, let $\alpha(d,k,l)$ denote the maximum possible cardinality of a set P of nonzero vectors in R^d such that any subset of $k+1$ members of P contains some $l+1$ pairwise orthogonal vectors. Thus $\alpha(d,k) = \alpha(d,k,1)$. Trivially, $\alpha(d,1) = d$ and Rosenfeld [5] proved, using an interesting algebraic argument, that $\alpha(d,2) = 2d$ for every d. Füredi and Stanley [4] observed that $\alpha(2,k) = 2k$, and proved that $\alpha(4,5) \geq 24$, that for every fixed d and l the limit $lim_{k \mapsto \infty} \alpha(d,k,l)/k$ is equal to its supremum, and that for every fixed l there exists some $\delta_l > 0$ and d_0 such that this supremum is at least $(1+\delta_l)^d$ for all $d > d_0$ and at most

$$(1+o(1))\sqrt{\pi d/(2l)}((l+1)/l)^{d/2-1},$$

where the $o(1)$ term tends to zero as d tends to infinity.

They conjectured that for every $l \geq 1$ there is some $g = g(l)$ ($< \infty$) such that $\alpha(d,k,l) \leq (dk)^g$ for every d and k.

In this note we show that this conjecture is false for every admissible l by proving the following result.

*Department of Mathematics, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel and AT & T Bell Labs, NJ, USA. Email: noga@math.tau.ac.il. Research supported in part by a USA Israeli BSF grant, by a grant from the Israel Science Foundation and by the Hermann Minkowski Minerva Center for Geometry at Tel Aviv University.

†AT & T Bell Labs, Murray Hill, NJ 07974, USA. Email: ms@research.att.com

Theorem 1.1 *For every $l \geq 1$ there exists an $\epsilon = \epsilon_l > 0$ such that for every positive integer t divisible by 4 and satisfying $t > l$, and for every positive integer s, if $d = t^s$ and $k = \lfloor 2^{t+1}/(\epsilon t) \rfloor$, then*

$$\alpha(d,k,l) \geq 2^{\epsilon t s/2}.$$

Note that by the above result, and by the obvious monotonicity properties of α, for every fixed $l \geq 1$ there are some $\delta = \delta(l) > 0$ and $k_0(l)$ such that for every $k \geq k_0(l)$ and every $d \geq 2 \log k$,

$$\alpha(d,k,l) \geq d^{\delta \log(k+2)/\log\log(k+2)}, \tag{1}$$

where here and in what follows all logarithms are in base 2. Indeed, given $l \geq 1$ and $k \geq k_0(l)$, $d \geq 2 \log k$, let $\epsilon = \epsilon_l$ be as in Theorem 1.1, and let t be the largest integer divisible by 4 for which $k \geq \frac{2^{t+1}}{\epsilon t}$. Note that $t = (1+o(1)) \log k$. Next let s be the largest integer such that $d \geq t^s$. Then $s \geq \Omega(\log d/\log\log k)$ and hence, by Theorem 1.1,

$$\alpha(d,k,l) \geq \alpha(t^s, \lfloor 2^{t+1}/(\epsilon t) \rfloor, l) \geq 2^{\epsilon t s/2} \geq d^{\Omega(\epsilon \log k/\log\log k)},$$

implying (1).

Moreover, since $\alpha(d,k,l) \geq Max\{d,k\}$ for all d and $k \geq l \geq 1$ the assumptions that $d \geq 2 \log k$ and that $k \geq k_0(l)$ can be dropped (by changing δ, if necessary) and we thus conclude that (1) holds for all d and all $k \geq l \geq 1$.

This shows that the above conjecture is false for all values of l. The special case $l = 1$ implies, in particular, that there exists a *fixed* k so that for every sufficiently large dimension d there is a collection of, say, d^{1000} nonzero vectors in R^d so that among any $k+1$ of those some two are orthogonal.

The two main ingredients in our proof are a result of Frankl and Rödl [3], whose relevance to this problem is mentioned already by Füredi and Stanley in [4], and the basic idea of Feige in [2] which is based on the the technique of Berman and Schnitger [1]. It is worth noting that the gap between the upper and lower bounds for $\alpha(d,k,l)$ is still large, and the problem of determining the asymptotic behaviour of this function more precisely, as well as that of determining the precise value of the function for various small values of the parameters, remain wide open.

2 The proof

Proof of Theorem 1.1. The *tensor product* $x = v_1 * v_2 * \ldots * v_s$ of s vectors

$$v_i = (v_{i,1}, v_{i,2}, \ldots, v_{i,t}), \ (1 \leq i \leq s)$$

in R^t is a vector in R^{t^s} whose coordinates, indexed by the ordered s-tuples

$$(i_1, i_2, \ldots, i_s) : 1 \leq i_j \leq t$$

are defined by

$$x_{(i_1,i_2,\ldots,i_s)} = v_{1,i_1} v_{2,i_2} \cdots v_{s,i_s}.$$

It is easy and well known that the inner product $x \cdot y$ of the vector x above with $y = u_1 * u_2 * \ldots * u_s$ ($u_i \in R^t$) is simply the product $\prod_{i=1}^{s}(v_i \cdot u_i)$ of all the inner products $v_i \cdot u_i$ (computed in R^t). Therefore, x and y are orthogonal if and only if there is some index i for which v_i and u_i are orthogonal. If F is a set of vectors in R^{t^s} consisting of vectors each of which is a tensor product of s vectors in R^t, the *j^{th}-projection of F* is the set of all vectors v in R^t such that there is some member $v_1 * v_2 * \ldots * v_s$ of F with $v_j = v$.

Let l be a positive integer. For an integer t, let Q_t denote the set of all 2^t real vectors of length t whose coordinates are $+1$ and -1. Frankl and Rödl [3] proved that there exists an $\epsilon = \epsilon_l > 0$ such that for every $t > l$ which is divisible by 4, any subset of Q_t of cardinality at least $2^{(1-\epsilon)t}$ contains $l+1$ pairwise orthogonal vectors. Define a subset F of 2^{ts} nonzero vectors in R^d, where $d = t^s$, as follows

$$F = \{v_1 * v_2 \ldots * v_s : v_i \in Q_t\}.$$

If $G_1, G_2, \ldots, G_s$ are subsets of Q_t, and each G_i does not contain $l+1$ pairwise orthogonal vectors, then the set

$$B = \{v_1 * v_2 * \ldots * v_s : v_i \in G_i\}$$

is called a *dangerous box*. Note that trivially, each dangerous box contains at most $2^{(1-\epsilon)ts}$ vectors, since each G_i in the definition above is of size at most $2^{(1-\epsilon)t}$. Note also that the number of dangerous boxes is clearly less than $2^{2^t s}$ (since there are less than 2^{2^t} possible choices for each G_i). A crucial observation is that any subset S of F that contains no $l+1$ pairwise orthogonal vectors is contained in a dangerous box. Indeed, simply define G_i to be the i^{th}-projection of S. Then S lies in the box determined by the sets G_i, and no G_i can contain $l+1$ pairwise orthogonal vectors (since otherwise the corresponding members of S are pairwise orthogonal as well, contradicting the assumption).

Let P be a random set of vectors obtained by choosing, randomly, independently (and with repetitions) $n = \lceil 2^{\epsilon st/2} \rceil$ members of F. To complete the proof we show that with positive probability every subset of more than $k = \lfloor 2^{t+1}/(\epsilon t) \rfloor$ members of P contains $l+1$ pairwise orthogonal vectors.

For each dangerous box B, let E_B be the event that P contains more than k members of B. By the observation above, if none of the events E_B occurs, then P contains no subset of cardinality $k+1$ without $l+1$ pairwise orthogonal members, as needed. It thus remains to estimate the probability of each event E_B. For a fixed box B,

$$Prob[E_B] \leq \binom{n}{k+1} (\frac{|B|}{|F|})^{k+1} \leq 2^{-\epsilon ts(k+1)/2}.$$

Since there are less than $2^{2^t s}$ dangerous boxes, the probability that at least one event E_B occurs is smaller than

$$2^{2^t s} 2^{-\epsilon ts(k+1)/2} \leq 1.$$

Therefore, with positive probability every subset of cardinality $k+1$ of P contains $l+1$ pairwise orthogonal members. In particular, such a P exists, showing that for $d = t^s$ and $k = \lfloor 2^{t+1}/(\epsilon t) \rfloor$,

$$\alpha(d, k, d) \geq |P| = n \geq 2^{\epsilon st/2},$$

and completing the proof. $\Box$

References

[1] P. Berman and G. Schnitger, *On the complexity of approximating the independent set problem*, Information and Computation 96 (1992), 77-94.

[2] U. Feige, *Randomized graph products, chromatic numbers, and the Lovász θ-function*, Proc. of the 27^{th} ACM STOC, ACM Press (1995), 635-640.

[3] P. Frankl and V. Rödl, *Forbidden intersections*, Trans. AMS 300 (1987), 259-286.

[4] Z. Füredi and R. Stanley, *Sets of vectors with many nearly orthogonal pairs (Research Problem)*, Graphs and Combinatorics 8 (1992), 391-394.

[5] M. Rosenfeld, *Almost orthogonal lines in E^d*, DIMACS Series in Discrete Math. 4 (1991), 489-492.