Nearly Orthogonal Sets over Finite Fields

Dror Chawin
School of Computer Science, The Academic College of Tel Aviv-Yaffo, Israel

Ishay Haviv
School of Computer Science, The Academic College of Tel Aviv-Yaffo, Israel

─── **Abstract** ───

For a field $\mathbb{F}$ and integers d and k, a set of vectors of $\mathbb{F}^d$ is called k-nearly orthogonal if its members are non-self-orthogonal and every $k + 1$ of them include an orthogonal pair. We prove that for every prime p there exists a positive constant $\delta = \delta(p)$, such that for every field $\mathbb{F}$ of characteristic p and for all integers $k \geq 2$ and $d \geq k^{1/(p-1)}$, there exists a k-nearly orthogonal set of at least $d^{\delta \cdot k^{1/(p-1)}/\log k}$ vectors of $\mathbb{F}^d$. In particular, for the binary field we obtain a set of $d^{\Omega(k/\log k)}$ vectors, and this is tight up to the $\log k$ term in the exponent. For comparison, the best known lower bound over the reals is $d^{\Omega(\log k/\log \log k)}$ (Alon and Szegedy, Graphs and Combin., 1999). The proof combines probabilistic and spectral arguments.

1 Introduction

For a field $\mathbb{F}$ and an integer d, two vectors $x, y \in \mathbb{F}^d$ are said to be orthogonal if they satisfy $\langle x, y \rangle = 0$ with respect to the inner product $\langle x, y \rangle = \sum_{i=1}^{d} x_i y_i$, where the arithmetic operations are over $\mathbb{F}$. If the vector x satisfies $\langle x, x \rangle = 0$, then we say that x is self-orthogonal and otherwise, it is non-self-orthogonal. For an integer k, a set $\mathcal{G} \subseteq \mathbb{F}^d$ is called k-nearly orthogonal if its vectors are non-self-orthogonal and every subset of $k + 1$ members of $\mathcal{G}$ includes an orthogonal pair. Let $\alpha(d, k, \mathbb{F})$ denote the maximum possible size of a k-nearly orthogonal subset of $\mathbb{F}^d$. It can be easily seen that for $k = 1$, it holds that $\alpha(d, 1, \mathbb{F}) = d$ for every field $\mathbb{F}$ and for every integer d. Indeed, the lower bound on $\alpha(d, 1, \mathbb{F})$ follows by considering the d vectors of the standard basis of $\mathbb{F}^d$, and the upper bound holds because a set of non-self-orthogonal vectors of $\mathbb{F}^d$ that are pairwise orthogonal is linearly independent, hence its size cannot exceed d.

A simple upper bound on $\alpha(d, k, \mathbb{F})$ follows from Ramsey theory. Recall that for integers s_1 and s_2, the Ramsey number $R(s_1, s_2)$ is the smallest integer r such that every graph on r vertices has an independent set of size s_1 or a clique of size s_2. For a given k-nearly orthogonal set $\mathcal{G} \subseteq \mathbb{F}^d$, consider the graph on the vertex set $\mathcal{G}$, where two distinct vertices are adjacent if and only if their vectors are not orthogonal. Since the vectors of $\mathcal{G}$ are non-self-orthogonal and lie in $\mathbb{F}^d$, this graph has no independent set of size $d + 1$, and since every $k + 1$ members of $\mathcal{G}$ include an orthogonal pair, this graph has no clique of size $k + 1$.
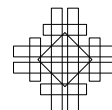
We thus obtain that

$$\alpha(d, k, \mathbb{F}) < R(d+1, k+1) \leq \binom{d+k}{k}, \tag{1}$$

where the second inequality follows by a famous upper bound on Ramsey numbers due to Erdős and Szekeres [13]. Note that for every fixed constant k, it follows from (1) that $\alpha(d, k, \mathbb{F}) \leq O(d^k)$.

The problem of determining the values of $\alpha(d, k, \mathbb{F})$ for $\mathbb{F}$ being the real field $\mathbb{R}$ was proposed by Erdős in 1988 (see [24]). For all integers d and k, it holds that

$$\alpha(d, k, \mathbb{R}) \geq k \cdot d, \tag{2}$$

as follows by considering the union of k pairwise disjoint orthogonal bases of $\mathbb{R}^d$. Erdős conjectured that this bound is tight for $k = 2$, that is, $\alpha(d, 2, \mathbb{R}) = 2 \cdot d$, and his conjecture was confirmed a few years later by Rosenfeld [27] (see also [12]). In 1992, Füredi and Stanley [16] showed that $\alpha(4, 5, \mathbb{R}) \geq 24$, which implies that the bound in (2) is not tight in general, and conjectured that there exists a constant c such that $\alpha(d, k, \mathbb{R}) \leq (k \cdot d)^c$ for all integers d and k. Their conjecture was disproved in 1999 by Alon and Szegedy [3], who applied the probabilistic method to prove that there exists a constant $\delta > 0$ such that for all integers d and $k \geq 3$, it holds that

$$\alpha(d, k, \mathbb{R}) \geq d^{\delta \cdot \log k / \log \log k},$$

where here and throughout, all logarithms are in base 2. More recently, Balla, Letzter, and Sudakov [5] proved that $\alpha(d, k, \mathbb{R}) \leq O(d^{(k+1)/3})$ for every fixed constant k, improving on the upper bound of (1) for the real field. Balla [4] further provided a bipartite analogue of the aforementioned result of [3]. Nevertheless, the currently known upper and lower bounds on $\alpha(d, k, \mathbb{R})$ for general d and k are rather far apart (see, e.g., [22, Chapter 10.2]).

Soon after the appearance of [3], Codenotti, Pudlák, and Resta [11] considered the question of determining the values of $\alpha(d, k, \mathbb{F})$ for finite fields $\mathbb{F}$, focusing on the case where $\mathbb{F}$ is the binary field $\mathbb{F}_2$ and $k = 2$. Note that $\alpha(d, 2, \mathbb{F}_2)$ can be formulated as the largest possible size of a family of subsets of $[d] = \{1, \ldots, d\}$, where each subset has an odd size and among every three of the subsets there are two that intersect at an even number of elements. While the original motivation of [11] to construct large 2-nearly orthogonal sets over $\mathbb{F}_2$ arrived from the rigidity approach to lower bounds in circuit complexity, this challenge enjoys additional diverse applications from the area of information theory, related to distributed storage, index coding, and hat-guessing games (see, e.g., [10, 7, 6, 20] and Section 1.3).

Codenotti et al. [11] provided an explicit construction of a 2-nearly orthogonal subset of $\mathbb{F}_2^d$ of size $4d - 8$ for every $d \geq 2$, implying that $\alpha(d, 2, \mathbb{F}_2) \geq 4d - 8$, and asked whether this bound is tight. Another explicit construction, of size $16 \cdot \lfloor d/6 \rfloor$, was provided by Blasiak, Kleinberg, and Lubetzky [10]. The question of [11] was answered negatively in [17], where it was shown by an explicit construction that $\alpha(d, 2, \mathbb{F}_2) \geq d^{1+\delta}$ for some constant $\delta > 0$ and for infinitely many integers d. Recalling that $\alpha(d, 2, \mathbb{R}) = 2 \cdot d$, this demonstrates that the value of $\alpha(d, k, \mathbb{F})$ behaves quite differently for different fields $\mathbb{F}$ already for $k = 2$.

1.1 Our Contribution

In the present paper, we prove lower bounds on the quantities $\alpha(d, k, \mathbb{F})$ for finite fields $\mathbb{F}$. For the binary field $\mathbb{F}_2$, we prove the following result.

▶ **Theorem 1.** *There exists a constant $\delta > 0$ such that for all integers $d \geq k \geq 2$, it holds that*

$$\alpha(d, k, \mathbb{F}_2) \geq d^{\delta \cdot k / \log k}.$$

The bound provided by Theorem 1 gets close to the upper bound given in (1) and is tight up to the $\log k$ term in the exponent. Note that the assumption that d is sufficiently large compared to k is essential, because otherwise the bound guaranteed by the theorem might exceed the total number of vectors in $\mathbb{F}_2^d$. It is interesting to compare Theorem 1 to the best known lower bound on $\alpha(d, k, \mathbb{F})$ where $\mathbb{F}$ is the real field $\mathbb{R}$, namely, the bound $d^{\Omega(\log k / \log \log k)}$ achieved in [3].

We proceed with an extension of Theorem 1 to general finite fields (in fact, fields with finite characteristic), stated as follows.

▶ **Theorem 2.** *For every prime p there exists a constant $\delta = \delta(p) > 0$, such that for every field $\mathbb{F}$ of characteristic p and for all integers $k \geq 2$ and $d \geq k^{1/(p-1)}$, it holds that*

$$\alpha(d, k, \mathbb{F}) \geq d^{\delta \cdot k^{1/(p-1)} / \log k}.$$

Note that the special case of Theorem 2 with $p = 2$ coincides with Theorem 1.

In fact, inspired by a recent paper of Balla [4], we provide a bipartite analogue of Theorem 2 with the same asymptotic bound. The bipartite setting concerns sets $\mathcal{G}$ of non-self-orthogonal vectors of $\mathbb{F}^d$, such that for every two subsets $G_1, G_2 \subseteq \mathcal{G}$ of size k each, some vector of G_1 is orthogonal to some vector of G_2. Notice that such a set $\mathcal{G}$ is $(k-1)$-nearly orthogonal, hence our result for the bipartite setting strengthens Theorem 2. Our bipartite analogue of Theorem 2 is motivated by questions on the dimension of orthogonal representations over finite fields of H-free graphs, where H is the complete bipartite graph $K_{k,k}$. We describe these questions and related work in Section 1.3. The applications of our results to this context are presented in the full version of the paper. Among other things, we determine up to a multiplicative constant the largest possible ratio between the clique cover number and the minimum dimension of an orthogonal representation over the binary field for graphs on n vertices.

The proofs of Theorems 1 and 2 borrow the probabilistic approach of Alon and Szegedy [3] for producing k-nearly orthogonal sets over $\mathbb{R}$, which relies on the randomized graph product technique developed by Berman and Schnitger [9] and by Feige [14]. However, in order to use this approach in the finite field setting and to establish significantly larger k-nearly orthogonal sets, we combine it with several additional ideas. For Theorem 1, we combine the approach of [3] with elementary linear-algebraic arguments. For Theorem 2 and for its bipartite analogue, we further use spectral techniques to prove pseudo-random properties of a related graph family. The analysis involves a result of Le Anh Vinh [28], which builds on a result of Alon and Krivelevich [2]. As a by-product of our proof technique, we obtain an essentially tight estimation for the number of sets of non-self-orthogonal vectors that are pairwise non-orthogonal in vector spaces over prime order fields.

1.2 Proof Overview

We offer here a high-level description of the proofs of Theorems 1 and 2. We start by presenting the approach of Alon and Szegedy [3] in their proof for the existence of large k-nearly orthogonal sets over $\mathbb{R}$, and then explain how we adapt it to the finite field setting and how we establish significantly larger sets.

For some integers t and m, let V denote the set of all 2^t real vectors of length t whose values are in $\{-1, +1\}$, and let Q denote the set of all vectors of the form $v_1 \otimes \cdots \otimes v_m$ for vectors $v_1, \ldots, v_m$ of V, where $\otimes$ stands for the tensor product operation on vectors (for the definition and properties of this operation, see Section 2). Note that the vectors of Q are of length t^m. For some integer n, let $\mathcal{G}$ be a random set of n vectors chosen uniformly and independently from Q. The argument proceeds by estimating the probability that $\mathcal{G}$ is k-nearly orthogonal.

The probabilistic analysis in [3] uses a result of Frankl and Rödl [15], which asserts that for some $\varepsilon > 0$, every subset of V whose vectors are pairwise non-orthogonal has size at most $2^{(1-\varepsilon) \cdot t}$. A key observation, which follows by standard properties of the tensor product operation, is that every subset of Q whose vectors are pairwise non-orthogonal is contained in a set of the form

$$\{v_1 \otimes \cdots \otimes v_m \mid v_j \in A_j \text{ for all } j \in [m]\} \tag{3}$$

for sets $A_1, \ldots, A_m \subseteq V$, where the vectors of each A_j are pairwise non-orthogonal. Referring to a set of the form (3) as a box, it suffices to show that with positive probability, no box contains $k + 1$ of the random members of $\mathcal{G}$. The above result of [15] implies that the size of every box is at most $2^{m \cdot (1-\varepsilon) \cdot t}$, hence the probability that a random vector chosen from Q lies in a fixed box does not exceed $2^{m \cdot (1-\varepsilon) \cdot t}/|Q| \approx 2^{-\varepsilon \cdot m \cdot t}$. By applying the union bound, first over all choices of $k + 1$ members of $\mathcal{G}$ for a fixed box, and then over all boxes, one may obtain an upper bound on the probability that $\mathcal{G}$ includes $k + 1$ vectors that lie in a common box and thus on the probability that $\mathcal{G}$ is not k-nearly orthogonal. Note that the number of boxes does not exceed the number of m-tuples of subsets of V, hence it is at most $2^{m \cdot 2^t}$. It turns out that for every integer $k \geq 3$, it is possible to choose the values of t and m such that for $d = t^m$ and $n = d^{\Omega(\log k / \log \log k)}$, the set $\mathcal{G} \subseteq \mathbb{R}^d$ is k-nearly orthogonal with positive probability. This yields the existence of the desired k-nearly orthogonal sets over $\mathbb{R}$ and completes the description of the argument of [3].

We next describe our proof of Theorem 1, which asserts the existence of large k-nearly orthogonal sets over the binary field $\mathbb{F}_2$. Following the approach of [3], for integers t and m, we consider the set V of all non-self-orthogonal vectors in $\mathbb{F}_2^t$ and the set Q of all the vectors in $\mathbb{F}_2^{t^m}$ that can be represented as a tensor product of m vectors of V. Note that the vectors of Q are non-self-orthogonal. As before, we let $\mathcal{G}$ be a random set of n vectors chosen uniformly and independently from Q, and we estimate the probability that $\mathcal{G}$ is k-nearly orthogonal.

The first question to ask here is how large can be a subset of V whose vectors are pairwise non-orthogonal. A simple linear-algebraic argument shows that its size cannot exceed $2^{(t+1)/2}$. Indeed, consider a set of pairwise non-orthogonal vectors of V, and add a 1 entry at the end of each of them. This gives us a set of self-orthogonal vectors of $\mathbb{F}_2^{t+1}$ that are pairwise orthogonal. It follows that the subspace of $\mathbb{F}_2^{t+1}$ spanned by those vectors is contained in its orthogonal complement, and as such, its dimension cannot exceed $(t + 1)/2$. This yields the desired bound of $2^{(t+1)/2}$ on the size of the given set. In fact, this argument not only provides a bound on the size of any set of pairwise non-orthogonal vectors of V, but also provides valuable information about its structure. Namely, every such set is contained in some subspace of $\mathbb{F}_2^t$ of dimension at most $(t + 1)/2$, obtained by omitting the last coordinate of the subspace of $\mathbb{F}_2^{t+1}$ considered above (see Lemma 4).

As in the case of the real field, one may observe that every subset of Q whose vectors are pairwise non-orthogonal is contained in a box defined as in (3). Moreover, the size of every such box is bounded by $2^{m \cdot (t+1)/2}$, hence the probability that a random vector chosen from Q lies in a fixed box is bounded by $2^{m \cdot (t+1)/2}/|Q| \approx 2^{-m \cdot t/2}$. By applying the union bound,

first over all choices of $k+1$ members of $\mathcal{G}$ for a fixed box, and then over all boxes, it is possible to bound the probability that some $k+1$ vectors of $\mathcal{G}$ lie in a common box and thus to bound the probability that $\mathcal{G}$ is not k-nearly orthogonal. The crucial point that allows us to obtain significantly larger sets over $\mathbb{F}_2$ than over the reals is a stronger bound on the number of needed boxes. As explained above, every set of pairwise non-orthogonal vectors of V is contained in a subspace of $\mathbb{F}_2^t$ of dimension bounded away from t. The number of such subspaces clearly does not exceed 2^{t^2}, hence the number of boxes that the union bound should take care of is at most $2^{m \cdot t^2}$ (in comparison to the $2^{m \cdot 2^t}$ boxes over $\mathbb{R}$). This allows us to show that for every integer $k \geq 2$, it is possible to choose the values of t and m such that for $d = t^m$ and $n = d^{\Omega(k/\log k)}$, the set $\mathcal{G} \subseteq \mathbb{F}_2^d$ is k-nearly orthogonal with positive probability.

We next discuss the proof of Theorem 2 for finite fields of prime order. First, note that the argument described above does not extend in a straightforward manner to the field $\mathbb{F}_p$ for a prime $p > 2$. Specifically, it no longer holds that every set of non-self-orthogonal vectors of $\mathbb{F}_p^t$ that are pairwise non-orthogonal is contained in a subspace of dimension bounded away from t. To see this, consider the set $\{e_1, e_1 + e_2, e_1 + e_3, \ldots, e_1 + e_t\}$, where e_i stands for the ith vector of the standard basis of $\mathbb{F}_p^t$. For every prime $p > 2$, the vectors of this set are non-self-orthogonal and are pairwise non-orthogonal, and yet, they span the entire space $\mathbb{F}_p^t$. Nevertheless, we show that similarly to the binary case, the size of a set of non-self-orthogonal vectors of $\mathbb{F}_p^t$ that are pairwise non-orthogonal is bounded from above by roughly $p^{t/2}$. This result is established using a spectral argument. Namely, we consider a suitable family of graphs and derive their pseudo-random properties using the second largest eigenvalue determined in [28].

However, in order to apply the probabilistic argument with a relatively small collection of boxes, it is not sufficient to bound the size of sets of non-self-orthogonal vectors of $\mathbb{F}_p^t$ that are pairwise non-orthogonal. We further need some understanding of the structure of those sets. Specifically, one has to come up with a relatively small collection of relatively small subsets of $\mathbb{F}_p^t$, such that every set of non-self-orthogonal vectors of $\mathbb{F}_p^t$ that are pairwise non-orthogonal is contained in one of them. To do so, we show that for some integer $s = O(t^{p-1})$, there exists a function $g : \mathbb{F}_p^t \to \mathbb{F}_p^s$ satisfying that for all $v_1, v_2 \in \mathbb{F}_p^t$, the vectors v_1 and v_2 are orthogonal if and only if their images $g(v_1)$ and $g(v_2)$ are not. We use this function to produce the required collection, and it turns out that its size is bounded by the number of subspaces of $\mathbb{F}_p^s$ that are spanned by vectors of the image of g, and thus by $(p^t)^s = p^{O(t^p)}$. Then, the number of boxes on which the union bound has to be applied is $p^{O(m \cdot t^p)}$, resulting in the k-nearly orthogonal sets declared in Theorem 2.

1.3 Related Work

This paper is concerned with the problem of determining the largest possible size $\alpha(d, k, \mathbb{F})$ of a k-nearly orthogonal subset of $\mathbb{F}^d$ for a field $\mathbb{F}$ and integers d and k. In addition to the requirement that every $k+1$ members of such a set include an orthogonal pair, we require the vectors of the set to be non-self-orthogonal. Over the reals, the latter condition simply means that the vectors are nonzero, but over finite fields, it significantly affects the problem. Without this requirement, the upper bound (1) which relies on Ramsey theory does not hold, and the size of the sets may grow exponentially in d even for $k = 1$. The maximum size of a set of pairwise orthogonal vectors, which are not necessarily non-self-orthogonal, was determined precisely in the late sixties for the binary field by Berlekamp [8] and for any prime order field by Zame [29]. Further extensions to general bilinear forms were recently provided by Mohammadi and Petridis [23].

Another variant of the problem asks, for a given integer ℓ, to find large sets of non-self-orthogonal vectors of $\mathbb{F}^d$ such that every $k + 1$ of them include $\ell + 1$ pairwise orthogonal vectors. While the present paper focuses on the case $\ell = 1$, the work of Alon and Szegedy [3] does consider this general setting over the reals and shows that for every integer $\ell \geq 1$ there exists a constant $\delta = \delta(\ell) > 0$, for which there exists a set of at least $d^{\delta \cdot \log k / \log \log k}$ vectors of $\mathbb{R}^d$ satisfying the above property. The challenge of extending our results to this setting is left for future research.

The quantities $\alpha(d, k, \mathbb{F})$ can be represented in terms of orthogonal representations of graphs, a notion that was proposed in the seminal paper of Lovász [21] that introduced the celebrated ϑ-function. A d-dimensional orthogonal representation of a graph over a field $\mathbb{F}$ is an assignment of a non-self-orthogonal vector of $\mathbb{F}^d$ to each vertex, such that the vectors assigned to two distinct non-adjacent vertices are orthogonal. As mentioned earlier, one may associate with every set $\mathcal{G}$ of non-self-orthogonal vectors of $\mathbb{F}^d$, a graph on the vertex set $\mathcal{G}$ with edges connecting pairs of non-orthogonal vectors. Notice that such a $\mathcal{G}$ is k-nearly orthogonal if and only if the graph associated with it is K_{k+1}-free, i.e., contains no copy of the complete graph on $k + 1$ vertices. Therefore, $\alpha(d, k, \mathbb{F})$ is closely related to the largest possible number of vertices in a K_{k+1}-free graph that admits a d-dimensional orthogonal representation over $\mathbb{F}$ (with a minor difference between the quantities, caused by the fact that the vectors of an orthogonal representation are not necessarily distinct). The problem of determining the smallest $d = d(n)$ such that every H-free graph on n vertices has a d-dimensional orthogonal representation over the reals was studied in the literature for various graphs H with a particular attention to cycles (see, e.g., [26, 5]). Note that the bipartite setting studied in [4] corresponds to the case where H is a balanced complete bipartite graph. The applications of the results of the present paper to this context are given in its full version.

It is worth mentioning here the similar concept of d-dimensional orthogonal bi-representations of graphs over a field $\mathbb{F}$. Here, one has to assign to each vertex a pair of non-orthogonal vectors of $\mathbb{F}^d$, such that the pairs (u_1, u_2) and (v_1, v_2) assigned to distinct non-adjacent vertices satisfy $\langle u_1, v_2 \rangle = \langle v_1, u_2 \rangle = 0$. Note that d-dimensional orthogonal representations can be viewed as a special case of d-dimensional orthogonal bi-representations by replacing every vector v with the pair (v, v). This notion was proposed by Peeters [25] to provide an alternative definition for the minrank parameter of graphs, introduced by Haemers [18] in the study of the Shannon capacity. While we do not mention here the original definition of the minrank of a graph over a field $\mathbb{F}$, it turns out that the latter is precisely the smallest integer d for which the graph admits a d-dimensional orthogonal bi-representation over $\mathbb{F}$. This quantity of graphs has attracted considerable attention due to its various applications in computational complexity and in information theory. Notably, the question of determining the smallest possible minrank over a given field $\mathbb{F}$ over all H-free graphs on n vertices has been explored for various graphs H (see, e.g., [11, 1, 19, 17]). The particular setting in which $\mathbb{F}$ is the binary field and H is the triangle graph is the focus of an intensive recent line of work (see [7, 6, 20]).

1.4   Outline

The rest of the paper is organized as follows. In Section 2, we provide some useful properties of the tensor product operation on vectors. In Section 3, we prove Theorem 1. Its proof is elementary, and we present it separately from the proof of Theorem 2 for didactic reasons. The extension of Theorem 1 to general finite fields and to the bipartite setting and the applications of our results to the context of orthogonal representations of graphs can be found in the full version of the paper.

2   Preliminaries

Our proofs crucially use the tensor product operation on vectors. For a field $\mathbb{F}$ and integers t_1, t_2, the tensor product $w = u \otimes v$ of two vectors $u \in \mathbb{F}^{t_1}$ and $v \in \mathbb{F}^{t_2}$ is defined as the vector in $\mathbb{F}^{t_1 \cdot t_2}$, whose coordinates are indexed by the pairs (i_1, i_2) with $i_1 \in [t_1]$ and $i_2 \in [t_2]$ (ordered lexicographically), defined by $w_{(i_1,i_2)} = u_{i_1} \cdot v_{i_2}$. Note that the representation of w as a tensor product of two vectors of lengths t_1 and t_2 is not necessarily unique. Note further that for integers t and m and for given vectors $v_1, \ldots, v_m \in \mathbb{F}^t$, the vector $v_1 \otimes \cdots \otimes v_m$ lies in $\mathbb{F}^{t^m}$ and consists of all the t^m possible products of m values, one taken from each vector v_j with $j \in [m]$. When all the vectors $v_1, \ldots, v_m$ are equal to a single vector v, their tensor product $v_1 \otimes \cdots \otimes v_m$ can be written as $v^{\otimes m}$.

It is well known and easy to verify that for vectors $u_1, \ldots, u_m \in \mathbb{F}^t$ and $v_1, \ldots, v_m \in \mathbb{F}^t$, the two vectors $u = u_1 \otimes \cdots \otimes u_m$ and $v = v_1 \otimes \cdots \otimes v_m$ satisfy

$$\langle u, v \rangle = \prod_{j=1}^{m} \langle u_j, v_j \rangle. \tag{4}$$

It thus follows that u and v are orthogonal if and only if u_j and v_j are orthogonal for some $j \in [m]$. In particular, the vector u is self-orthogonal if and only if u_j is self-orthogonal for some $j \in [m]$.

For m sets $A_1, \ldots, A_m \subseteq \mathbb{F}^t$, we let $A_1 \otimes \cdots \otimes A_m$ denote the collection of all vectors of the form $v_1 \otimes \cdots \otimes v_m$ with $v_j \in A_j$ for each $j \in [m]$. While this notation will be convenient for us throughout this paper, let us stress that it does not coincide with the usual tensor product operation on linear subspaces. As before, when all the sets $A_1, \ldots, A_m$ are equal to a single set A, the collection $A_1 \otimes \cdots \otimes A_m$ can be written as $A^{\otimes m}$. For a set $G \subseteq \mathbb{F}^{t^m}$ whose vectors are represented as tensor products of m vectors of $\mathbb{F}^t$, the jth projection of G is the set of all vectors $v \in \mathbb{F}^t$ for which there exists a member $v_1 \otimes \cdots \otimes v_m$ of G with $v_j = v$.

We will need the following simple claim.

$\triangleright$ **Claim 3.** For a field $\mathbb{F}$ and two integers t and m, let $A_1, \ldots, A_m$ be sets of nonzero vectors of $\mathbb{F}^t$ such that the first nonzero value in each vector is 1. Then, $|A_1 \otimes \cdots \otimes A_m| = \prod_{j=1}^{m} |A_j|$.

Proof. Fix a field $\mathbb{F}$. We will prove that for all integers t_1 and t_2 and for all sets $A_1 \subseteq \mathbb{F}^{t_1}$ and $A_2 \subseteq \mathbb{F}^{t_2}$ of nonzero vectors whose first nonzero value is 1, it holds that
1. the vectors of $A_1 \otimes A_2$ are nonzero with first nonzero value 1, and
2. $|A_1 \otimes A_2| = |A_1| \cdot |A_2|$.
This, applied iteratively $m - 1$ times, completes the proof of the claim.

Let A_1 and A_2 be sets as above. For the first item, consider a vector $v_1 \otimes v_2 \in A_1 \otimes A_2$ with $v_1 \in A_1$ and $v_2 \in A_2$. Since v_1 and v_2 are nonzero, it follows that $v_1 \otimes v_2$ is nonzero as well. Additionally, the first nonzero value of $v_1 \otimes v_2$ is the product of the first nonzero values of v_1 and v_2, and is thus equal to 1.

For the second item, it suffices to show that for all vectors $u_1, v_1 \in A_1$ and $u_2, v_2 \in A_2$, if $u_1 \otimes u_2 = v_1 \otimes v_2$ then $u_1 = v_1$ and $u_2 = v_2$. So suppose that $u_1 \otimes u_2 = v_1 \otimes v_2$. For contradiction, suppose further that $u_1 \neq v_1$, and consider the blocks of length t_2 in $u_1 \otimes u_2$ and $v_1 \otimes v_2$ that correspond to the first entry in which u_1 and v_1 differ. These blocks must be distinct, because u_2 and v_2 are nonzero vectors whose first nonzero value is 1, which implies that one is not a multiple of the other by a field element different from 1. This clearly contradicts the assumption $u_1 \otimes u_2 = v_1 \otimes v_2$. Now, given that $u_1 = v_1$, consider the blocks of length t_2 in $u_1 \otimes u_2$ and $v_1 \otimes v_2$ that correspond to the first nonzero value of u_1 (and v_1). Since those blocks are equal to u_2 and v_2 respectively, the assumption $u_1 \otimes u_2 = v_1 \otimes v_2$ implies that $u_2 = v_2$, and we are done. $\triangleleft$

3   Nearly Orthogonal Sets over the Binary Field

In this section, we prove the existence of large k-nearly orthogonal sets over the binary field and confirm Theorem 1. We start by proving that for every integer t, there exists a relatively small collection of relatively small subsets of $\mathbb{F}_2^t$, such that every set of non-self-orthogonal vectors of $\mathbb{F}_2^t$ that are pairwise non-orthogonal is contained in one of them.

▶ **Lemma 4.** *Let t be an integer. There exists a collection $\mathcal{C}$ of subsets of $\mathbb{F}_2^t$ such that*
1. $|\mathcal{C}| \leq 2^{t^2}$,
2. *for every $C \in \mathcal{C}$, it holds that $|C| \leq 2^{(t+1)/2}$, and*
3. *for every set $A \subseteq \mathbb{F}_2^t$ with $\langle v_1, v_2 \rangle \neq 0$ for all $v_1, v_2 \in A$, there exists a set $C \in \mathcal{C}$ such that $A \subseteq C$.*

Proof. Fix an integer t, and define $\mathcal{C}$ as the collection of all subspaces of $\mathbb{F}_2^t$ of dimension $\lfloor \frac{t+1}{2} \rfloor$. We prove that $\mathcal{C}$ satisfies the three properties required by the lemma. Firstly, the number of subspaces of $\mathbb{F}_2^t$ clearly does not exceed the number of possible bases of such subspaces, hence it holds that $|\mathcal{C}| \leq 2^{t^2}$, as required for Item 1 of the lemma. Secondly, every subspace of $\mathbb{F}_2^t$ of dimension $\lfloor \frac{t+1}{2} \rfloor$ includes $2^{\lfloor (t+1)/2 \rfloor}$ vectors, implying that $\mathcal{C}$ satisfies Item 2 of the lemma.

Finally, to prove that $\mathcal{C}$ satisfies Item 3 of the lemma, let A be a subset of $\mathbb{F}_2^t$ with $\langle v_1, v_2 \rangle \neq 0$ for all $v_1, v_2 \in A$. Consider the subset $A' = \{(v, 1) \mid v \in A\}$ of $\mathbb{F}_2^{t+1}$ obtained by adding a 1 entry at the end of each vector of A. Observe that our assumption on A implies that for every two (not necessarily distinct) vectors $u_1 = (v_1, 1)$ and $u_2 = (v_2, 1)$ of A' with $v_1, v_2 \in A$, it holds that $\langle u_1, u_2 \rangle = \langle v_1, v_2 \rangle + 1 = 0$. Let $W = \mathrm{span}(A')$ denote the subspace of $\mathbb{F}_2^{t+1}$ spanned by the vectors of A'. Since every two vectors of A' are orthogonal, it follows that every two vectors of W are orthogonal. Letting $W^\perp$ stand for the orthogonal complement of W, this implies that $W \subseteq W^\perp$, and thus $\dim(W) \leq \dim(W^\perp)$. By $\dim(W) + \dim(W^\perp) = t + 1$, we derive that $\dim(W) \leq \lfloor \frac{t+1}{2} \rfloor$. Now, let $\widetilde{W}$ denote the projection of W on the first t coordinates (omitting the last one). Observe that $\widetilde{W}$ forms a subspace of $\mathbb{F}_2^t$ of dimension at most $\lfloor \frac{t+1}{2} \rfloor$ and that $\widetilde{W} = \mathrm{span}(A)$, so in particular, $A \subseteq \widetilde{W}$. By the definition of $\mathcal{C}$, this yields that there exists a set $C \in \mathcal{C}$ such that $A \subseteq C$. This completes the proof.   ◀

We are ready to prove Theorem 1.

Proof of Theorem 1. For an integer t, let V denote the set of all non-self-orthogonal vectors of $\mathbb{F}_2^t$, that is,

$$V = \{v \in \mathbb{F}_2^t \mid \langle v, v \rangle = 1\},$$

and notice that $|V| = 2^{t-1}$. For an integer m, let Q denote the set of all vectors obtained by applying the tensor product operation on m vectors of V, that is, $Q = V^{\otimes m} \subseteq \mathbb{F}_2^{t^m}$. Since the vectors of V are nonzero, we can apply Claim 3 to obtain that

$$|Q| = |V|^m = 2^{m \cdot (t-1)}. \tag{5}$$

Note that the vectors of Q are non-self-orthogonal, because every vector $v = v_1 \otimes \cdots \otimes v_m \in Q$ satisfies $\langle v, v \rangle = \prod_{j=1}^m \langle v_j, v_j \rangle = 1$, where the last equality holds because the vectors $v_1, \ldots, v_m$ are members of V.

For an integer n, let $\mathcal{Z} = (z_1, \ldots, z_n)$ be a random sequence of n vectors chosen uniformly and independently from Q (repetitions allowed). The vectors of $\mathcal{Z}$ are clearly non-self-orthogonal, because the vectors of Q are. We will show that for a given integer k and for

an appropriate choice of the integers t, m, and n, it holds with positive probability that for every set $I \subseteq [n]$ of size $|I| = k+1$ the vectors of $\{z_i \mid i \in I\}$ include an orthogonal pair. We start with some preparations.

By Lemma 4, there exists a collection $\mathcal{C}$ of subsets of $\mathbb{F}_2^t$ such that

1. $|\mathcal{C}| \leq 2^{t^2}$,
2. for every $C \in \mathcal{C}$, it holds that $|C| \leq 2^{(t+1)/2}$, and
3. for every set $A \subseteq \mathbb{F}_2^t$ with $\langle v_1, v_2 \rangle \neq 0$ for all $v_1, v_2 \in A$, there exists a set $C \in \mathcal{C}$ such that $A \subseteq C$.

Consider the collection

$$\mathcal{B} = \{C^{(1)} \otimes C^{(2)} \otimes \cdots \otimes C^{(m)} \mid C^{(j)} \in \mathcal{C} \text{ for all } j \in [m]\}.$$

It follows from Item 1 above that

$$|\mathcal{B}| \leq |\mathcal{C}|^m \leq 2^{m \cdot t^2}. \tag{6}$$

It further follows from Item 2 that for every set $B = C^{(1)} \otimes C^{(2)} \otimes \cdots \otimes C^{(m)} \in \mathcal{B}$, it holds that

$$|B| \leq \prod_{j=1}^m |C^{(j)}| \leq 2^{m \cdot (t+1)/2}. \tag{7}$$

We next claim that for every set $G \subseteq Q$ of pairwise non-orthogonal vectors, there exists a set $B \in \mathcal{B}$ such that $G \subseteq B$. To see this, consider such a set $G \subseteq Q$, and for each $j \in [m]$, let $A^{(j)} \subseteq \mathbb{F}_2^t$ denote the jth projection of G (see Section 2). Using the property of tensor product given in (4), the fact that every two (not necessarily distinct) vectors of G are not orthogonal implies that for each $j \in [m]$, every two (not necessarily distinct) vectors of $A^{(j)}$ are not orthogonal, hence by Item 3, there exists a set $C^{(j)} \in \mathcal{C}$ such that $A^{(j)} \subseteq C^{(j)}$. This implies that $G \subseteq B$ for the set $B = C^{(1)} \otimes C^{(2)} \otimes \cdots \otimes C^{(m)} \in \mathcal{B}$, as desired.

Now, for a given integer k, consider the event $\mathcal{E}$ that there exists a set $I \subseteq [n]$ of size $|I| = k+1$ for which the vectors of $\{z_i \mid i \in I\}$ are pairwise non-orthogonal. As shown above, such vectors lie in some set of the collection $\mathcal{B}$. For every fixed $B \in \mathcal{B}$, we apply the union bound to obtain that the probability that there exists a set $I \subseteq [n]$ of size $|I| = k+1$ such that $\{z_i \mid i \in I\} \subseteq B$ is at most

$$\binom{n}{k+1} \cdot \left(\frac{|B|}{|Q|}\right)^{k+1} \leq \left(\frac{n \cdot |B|}{|Q|}\right)^{k+1} \leq \left(\frac{n \cdot 2^{m \cdot (t+1)/2}}{2^{m \cdot (t-1)}}\right)^{k+1} = \left(\frac{n}{2^{m \cdot (t-3)/2}}\right)^{k+1},$$

where for the second inequality we have used (5) and (7). We apply again the union bound, this time over all the sets of $\mathcal{B}$, and use (6) to obtain that

$$\Pr[\mathcal{E}] \leq |\mathcal{B}| \cdot \left(\frac{n}{2^{m \cdot (t-3)/2}}\right)^{k+1} \leq 2^{m \cdot t^2} \cdot \left(\frac{n}{2^{m \cdot (t-3)/2}}\right)^{k+1}. \tag{8}$$

We finally set the parameters of the construction, ensuring that the event $\mathcal{E}$ occurs with probability smaller than 1. Let $d \geq k$ be two integers. Note that constant values of k can be handled, using $\alpha(d, k, \mathbb{F}_2) \geq d$, by an appropriate choice of the constant δ from the assertion of the theorem. For a sufficiently large k, set $t = \lfloor k/8 \rfloor$, and let m be the largest integer such that $d \geq t^m$. The assumption $d \geq k$ implies that $m \geq 1$. Set $n = \lfloor 2^{m \cdot t/4} \rfloor$. We obtain that

$$\Pr[\mathcal{E}] \leq 2^{m \cdot t^2} \cdot \left(\frac{1}{2^{m \cdot (t/4 - 3/2)}}\right)^{k+1} \leq 2^{m \cdot t^2} \cdot \left(\frac{1}{2^{m \cdot t/8}}\right)^{k+1} < 1,$$

where the first inequality holds by combining (8) with our choice of n, the second by the assumption that k and t are sufficiently large (specifically, $t/4 - 3/2 \geq t/8$ for $t \geq 12$), and the third by our choice of t. This implies that there exists a choice for the n vectors of the sequence $\mathcal{Z}$ for which the event $\mathcal{E}$ does not occur, that is, no $k+1$ of them are pairwise non-orthogonal. For this choice, let $\mathcal{G} = \{z_i \mid i \in [n]\} \subseteq \mathbb{F}_2^{t^m}$ denote the set that consists of the vectors of $\mathcal{Z}$. Since the event $\mathcal{E}$ does not occur, no vector appears in $\mathcal{Z}$ more than k times, hence $|\mathcal{G}| \geq n/k$. We thus obtain, using the monotonicity of $\alpha(d, k, \mathbb{F}_2)$ with respect to d, that

$$\alpha(d, k, \mathbb{F}_2) \geq \alpha(t^m, k, \mathbb{F}_2) \geq n/k \geq 2^{\Omega(m \cdot t)} \geq 2^{\Omega((\log d) \cdot t/\log t)} \geq d^{\Omega(t/\log t)} \geq d^{\Omega(k/\log k)}.$$

This completes the proof. ◀

References

1 Noga Alon, H. Tracy Hall, Christian Knauer, Rom Pinchasi, and Raphael Yuster. On graphs and algebraic graphs that do not contain cycles of length 4. *J. Graph Theory*, 68(2):91–102, 2011.

2 Noga Alon and Michael Krivelevich. Constructive bounds for a Ramsey-type problem. *Graphs and Combinatorics*, 13(3):217–225, 1997.

3 Noga Alon and Mario Szegedy. Large sets of nearly orthogonal vectors. *Graphs and Combinatorics*, 15(1):1–4, 1999.

4 Igor Balla. Orthonormal representations, vector chromatic number, and extension complexity. *arXiv*, abs/2310.17482, 2023.

5 Igor Balla, Shoham Letzter, and Benny Sudakov. Orthonormal representations of H-free graphs. *Discret. Comput. Geom.*, 64(3):654–670, 2020.

6 Alexander Barg, Moshe Schwartz, and Lev Yohananov. Storage codes on triangle-free graphs with asymptotically unit rate. *arXiv*, abs/2212.12117, 2022.

7 Alexander Barg and Gilles Zémor. High-rate storage codes on triangle-free graphs. *IEEE Trans. Inform. Theory*, 68(12):7787–7797, 2022.

8 Elwyn Ralph Berlekamp. On subsets with intersections of even cardinality. *Canad. Math. Bull.*, 12(4):471–474, 1969.

9 Piotr Berman and Georg Schnitger. On the complexity of approximating the independent set problem. *Inf. Comput.*, 96(1):77–94, 1992. Preliminary version in STACS'89.

10 Anna Blasiak, Robert Kleinberg, and Eyal Lubetzky. Broadcasting with side information: Bounding and approximating the broadcast rate. *IEEE Trans. Inform. Theory*, 59(9):5811–5823, 2013.

11 Bruno Codenotti, Pavel Pudlák, and Giovanni Resta. Some structural properties of low-rank matrices related to computational complexity. *Theor. Comput. Sci.*, 235(1):89–107, 2000.

12 Louis Deaett. The minimum semidefinite rank of a triangle-free graph. *Linear Algebra Appl.*, 434(8):1945–1955, 2011.

13 Paul Erdős and George Szekeres. A combinatorial problem in geometry. *Compositio Mathematica*, 2:463–470, 1935.

14 Uriel Feige. Randomized graph products, chromatic numbers, and the Lovász ϑ-function. *Combinatorica*, 17(1):79–90, 1997. Preliminary version in STOC'95.

15 Peter Frankl and Vojtěch Rödl. Forbidden intersections. *Trans. Amer. Math. Soc.*, 300(1):259–286, 1987.

16 Zoltán Füredi and Richard P. Stanley. Sets of vectors with many orthogonal pairs. *Graphs and Combinatorics*, 8(4):391–394, 1992.

17 Alexander Golovnev and Ishay Haviv. The (generalized) orthogonality dimension of (generalized) Kneser graphs: Bounds and applications. *Theory of Computing*, 18(22):1–22, 2022. Preliminary version in CCC'21.

18 Willem H. Haemers. An upper bound for the Shannon capacity of a graph. In László Lovász and Vera T. Sós, editors, *Algebraic Methods in Graph Theory*, volume 25/I of *Colloquia Mathematica Societatis János Bolyai*, pages 267–272. Bolyai Society and North-Holland, 1978.

19 Ishay Haviv. On minrank and forbidden subgraphs. *ACM Trans. Comput. Theory (TOCT)*, 11(4):20, 2019. Preliminary version in RANDOM'18.

20 Hexiang Huang and Qing Xiang. Construction of storage codes of rate approaching one on triangle-free graphs. *Des. Codes Cryptogr.*, 91(12):3901–3913, 2023.

21 László Lovász. On the Shannon capacity of a graph. *IEEE Trans. Inform. Theory*, 25(1):1–7, 1979.

22 László Lovász. *Graphs and Geometry*, volume 65. Colloquium Publications, 2019.

23 Ali Mohammadi and Giorgis Petridis. Almost orthogonal subsets of vector spaces over finite fields. *Eur. J. Comb.*, 103:103515, 2022.

24 Jaroslav Nešetřil and Moshe Rosenfeld. Embedding graphs in Euclidean spaces, an exploration guided by Paul Erdős. *Geombinatorics*, 6:143–155, 1997.

25 René Peeters. Orthogonal representations over finite fields and the chromatic number of graphs. *Combinatorica*, 16(3):417–431, 1996.

26 Pavel Pudlák. Cycles of nonzero elements in low rank matrices. *Combinatorica*, 22(2):321–334, 2002.

27 Moshe Rosenfeld. Almost orthogonal lines in E^d. *DIMACS Series in Discrete Math.*, 4:489–492, 1991.

28 Le Anh Vinh. On the number of orthogonal systems in vector spaces over finite fields. *Electr. J. Comb.*, 15(32):1–4, 2008.

29 Alan Zame. Orthogonal sets of vectors over Z_m. *J. Comb. Theory*, 9(2):136–143, 1970.