Perturbed identity matrices have high rank: proof and applications

Noga Alon *

Abstract

We describe a lower bound for the rank of any real matrix in which all diagonal entries are significantly larger than the absolute value of all other entries, and discuss several applications of this result to the study of problems in Geometry, Coding Theory, Extremal Finite Set Theory and Probability.

1 Introduction

Let $B = (b_{i,j})$ be an n by n real matrix. It is easy and well known that if for every i, $|b_{i,i}| > \sum_{j \neq i} |b_{i,j}|$, then B is of full rank. Indeed, assuming this is false, let $c = (c_j)$ be a nonzero column vector so that $Bc = 0$. Let $|c_r| = \max_i |c_i|$ (> 0) and consider the component number r of Bc. The absolute value of this component is

$$|\sum_j b_{r,j}c_j| \geq |b_{r,r}c_r| - \sum_{j \neq r} |b_{r,j}c_j| \geq |c_r|(|b_{r,r}| - \sum_{j \neq r} |b_{r,j}|) > 0,$$

contradicting the assumption $Bc = 0$ and proving that B indeed has full rank. In particular, this implies that if $b_{i,i} = 1$ for all i and $|b_{i,j}| \leq \frac{1}{n}$ for all distinct indices i, j, then the rank of B is n.

Suppose we relax the conditions above, and only assume that each diagonal entry is at least $1/2$ and the absolute value of each other entry is at most ϵ. In this case one can also establish a lower bound for the rank of B, as stated in the following theorem.

Theorem 1.1 *There exists an absolute positive constant c so that the following holds. Let B be an n by n real matrix with $b_{i,i} \geq 1/2$ for all i and $|b_{i,j}| \leq \epsilon$ for all $i \neq j$, where $\frac{1}{2\sqrt{n}} \leq \epsilon < 1/4$. Then the rank of B satisfies*

$$rank(B) \geq \frac{c}{\epsilon^2 \log(1/\epsilon)} \log n.$$

*Schools of Mathematics and Computer Science, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel. Email: nogaa@tau.ac.il. Research supported in part by the Israel Science Foundation, by a USA-Israel BSF grant, and by the Hermann Minkowski Minerva Center for Geometry at Tel Aviv University.

This theorem is a slight variation of a result proved in [1]. In this short survey we present the proof of the theorem, and describe several applications in various areas. Some of these applications are known, and some are new.

The rest of the paper is organized as follows. In Section 2 we present the proof of the theorem, in Sections 3, 4, 5, 6 and 7 we describe its applications in Geometry, Coding Theory, Extremal Finite Set Theory, the investigation of pseudo-random sequences, and the study of small sample spaces supporting nearly independent random variables. The final Section 8 contains some concluding remarks and open problems.

2 Perturbed identity matrices

It is convenient to first prove the following variant of Theorem 1.1.

Theorem 2.1 *There exists an absolute positive constant c so that the following holds. Let B be an n by n real matrix with $b_{i,i} = 1$ for all i and $|b_{i,j}| \leq \epsilon$ for all $i \neq j$. If the rank of B is d, and $\frac{1}{\sqrt{n}} \leq \epsilon < 1/2$, then*

$$d \geq \frac{c}{\epsilon^2 \log(1/\epsilon)} \log n.$$

This result is proved in [1]. For completeness, we reproduce the proof (omitting the final detailed computation). We need the following well known lemma proved, among other places, in [7], [1].

Lemma 2.2 *Let $A = (a_{i,j})$ be an n by n real, symmetric matrix with $a_{i,i} = 1$ for all i and $|a_{i,j}| \leq \epsilon$ for all $i \neq j$. If the rank of A is d, then*

$$d \geq \frac{n}{1 + (n-1)\epsilon^2}.$$

In particular, if $\epsilon \leq \frac{1}{\sqrt{n}}$ then $d > n/2$.

Proof: Let $\lambda_1, \ldots, \lambda_n$ denote the eigenvalues of A, then their sum is the trace of A, which is n, and at most d of them are nonzero. Thus, by Cauchy-Schwartz, $\sum_{i=1}^{n} \lambda_i^2 \geq d(n/d)^2 = n^2/d$. On the other hand, this sum is the trace of $A^t A$, which is precisely $\sum_{i,j} a_{i,j}^2 \leq n + n(n-1)\epsilon^2$. Hence $n + n(n-1)\epsilon^2 \geq n^2/d$, implying the desired result. $\Box$

Lemma 2.3 *Let $B = (b_{i,j})$ be an n by n matrix of rank d, and let $P(x)$ be an arbitrary polynomial of degree k. Then the rank of the n by n matrix $(P(b_{i,j}))$ is at most $\binom{k+d}{k}$. Moreover, if $P(x) = x^k$ then the rank of $(P(b_{i,j}))$ is at most $\binom{k+d-1}{k}$.*

Proof: Let $\mathbf{v_1} = (v_{1,j})_{j=1}^n, \mathbf{v}_2 = (v_{2,j})_{j=1}^n, \ldots, \mathbf{v_d} = (v_{d,j})_{j=1}^n$ be a basis of the row-space of B. Then the vectors $(v_{1,j}^{k_1} \cdot v_{2,j}^{k_2} \cdots v_{d,j}^{k_d})_{j=1}^n$, where $k_1, k_2, \ldots, k_d$ range over all non-negative integers whose sum

is at most k, span the rows of the matrix $(P(b_{i,j}))$. In case $P(x) = x^k$ it suffices to take all these vectors corresponding to $k_1, k_2, \ldots, k_d$ whose sum is precisely k. □

Proof of Theorem 2.1: We may and will assume that B is symmetric, since otherwise we simply apply the result to $(B + B^t)/2$ whose rank is at most twice the rank of B. If $\epsilon \leq 1/n^\delta$ for some fixed $\delta > 0$, the result follows by applying Lemma 2.2 to a $\lfloor \frac{1}{\epsilon^2} \rfloor$ by $\lfloor \frac{1}{\epsilon^2} \rfloor$ submatrix of B. Thus we may assume that $\epsilon \geq 1/n^\delta$ for some fixed, small $\delta > 0$. Put $k = \lfloor \frac{\log n}{2 \log(1/\epsilon)} \rfloor$, $n' = \lfloor \frac{1}{\epsilon^{2k}} \rfloor$ and note that $n' \leq n$ and that $\epsilon^k \leq \frac{1}{\sqrt{n'}}$. By Lemma 2.3 the rank of the n' by n' matrix $(b_{i,j}^k)_{i,j \leq n'}$ is at most $\binom{d+k}{k} \leq (\frac{e(k+d)}{k})^k$. On the other hand, by Lemma 2.2, the rank of this matrix is at least $n'/2$. Therefore

$$(\frac{e(k+d)}{k})^k \geq \frac{n'}{2} = \frac{1}{2} \lfloor \frac{1}{\epsilon^{2k}} \rfloor,$$

and the desired result follows by some simple manipulation, that can be found, for example, in [3]. □

Proof of Theorem 1.1: Let $C = (c_{i,j})$ be the n by n diagonal matrix defined by $c_{i,i} = 1/b_{i,i}$ for all i. Then every diagonal entry of CB is 1 and every off-diagonal entry is of absolute value at most 2ϵ. The result thus follows from Theorem 2.1. □

3 Distortion in low dimension embeddings

A well known lemma of Johnson and Lindenstraus, proved in [9] (see also [13]), asserts that for any $\epsilon > 0$, any set A of n points in an Euclidean space can be embedded in an Euclidean space of dimension $k = c(\epsilon) \log n$ with distortion at most ϵ. That is, there is a mapping $f : A \mapsto R^k$ such that for any $a, b \in A$, the distance between $f(a)$ and $f(b)$ is at least the distance between a and b, and at most that distance multiplied by $1 + \epsilon$. The proof gives that $c(\epsilon) \leq O(\frac{1}{\epsilon^2})$. Theorem 2.1 can be used to show that this is nearly tight: $c(\epsilon)$ must be at least $\Omega(\frac{1}{\epsilon^2 \log(1/\epsilon)})$, even for embedding the set of points of a simplex. This is stated in the following proposition, proved in [1].

Proposition 3.1 *Let $P_0, P_1, \ldots, P_n$ be a set of $n + 1$ points in R^k, and suppose that the distance between any two of them is at least 1 and at most $1+\epsilon$, where $\frac{1}{\sqrt{n}} \leq \epsilon \leq \frac{1}{10}$. Then $k \geq \frac{c'}{\epsilon^2 \log(1/\epsilon)} \log n$, where c' is an absolute positive constant.*

Proof: Put one of the points, say P_0, at the origin, and shift all other points by at most ϵ making sure that their distance from P_0 is exactly 1. By the triangle inequality the distance between any pair of the shifted points is still $1 + O(\epsilon)$. Therefore, if v_i is the k-dimensional vector representing the i-th point, then the gram matrix $C = (v_i^t \cdot v_j)$ is an n by n matrix in which all diagonal entries are 1, and all other entries are $1/2 + O(\epsilon)$. Moreover, the rank of this matrix is at most k. Therefore, the rank of $B = 2C - J$, where J is the all 1 n by n matrix, is at most $k + 1$. By Theorem 2.1 this rank is at least $\Omega(\frac{1}{\epsilon^2 \log(1/\epsilon)} \log n)$, supplying the required lower bound for the dimension k. □

4 Coding Theory

A binary code of length k is a set $C \subset \{0,1\}^k$ of binary vectors with k coordinates. The code is called ϵ-*balanced* if the Hamming distance between any two code-words is at least $\frac{1-\epsilon}{2}k$ and at most $\frac{1+\epsilon}{2}k$. For each vector $v = (v_1, v_2, \ldots, v_k) \in C$, let $x(v)$ denote the vector

$$x(v) = (\ (-1)^{v_1}, (-1)^{v_2}, \ldots, (-1)^{v_k}\ ) \in \{-1,1\}^k.$$

Note that for any two $u, v \in C$, the inner product between $x(u)$ and $x(v)$ is precisely $k - 2h(u,v)$, where $h(u,v)$ is the Hamming distance between u and v.

It follows that for $\epsilon = 0$, every two vectors $x(u), x(v)$ corresponding to distinct code-words of an ϵ-balanced code are orthogonal, and hence the number of code-words is at most k. Any Hadamard matrix of order k (if one exists) shows that this is tight, hence this is tight for all powers of 2 as well as for many other values of k divisible by 4 (see, e,g., [8] for more information about the existence of Hadamard matrices.)

For positive values of ϵ the problem of determining or estimating the largest possible cardinality of an ϵ-balanced code of length k is more complicated. Note, first, that ϵ should be at least $1/k$, since otherwise any ϵ-balanced code of length k is, in fact, 0-balanced. A simple probabilistic argument (or an obvious variant of the Gilbert Varshamov bound) shows that there are ϵ-balanced codes of length k with at least $2^{\Omega(\epsilon^2 k)}$ codewords. Theorem 2.1 provides a quick upper bound, as follows.

Proposition 4.1 *There exists an absolute positive constant a so that for all $\frac{1}{\sqrt{k}} \leq \epsilon < 1/2$ the cardinality of any ϵ-balanced code of length k is at most $2^{a\epsilon^2 \log(1/\epsilon)k}$.*

Proof: Let $C \subset \{0,1\}^k$ be an ϵ-balanced code of length k and maximum cardinality. Put $n = |C|$ and note that we may assume that $n \geq k$. Let X be the n by k matrix whose rows are the $|C|$ vectors $\frac{x(v)}{\sqrt{k}}$, $v \in C$. Let B be the n by n matrix defined by $B = (b_{u,v}) = XX^t$. Then each diagonal entry $b_{u,u}$ of B is 1, whereas each other entry of it $b_{u,v}$ for $u \neq v,\ u, v \in C$, satisfies $|b_{u,v}| = |\frac{1}{k}(k - 2h(u,v))| \leq \epsilon$. Therefore, by Theorem 2.1,

$$k \geq \mathrm{rank}(X) \geq \mathrm{rank}(B) \geq \frac{c}{\epsilon^2 \log(1/\epsilon)} \log n,$$

supplying the desired result. □

Note that the assertion of the last proposition, at least for fixed ϵ and large n, can be also deduced, in a completely different manner, from the Linear Programming technique of Delsarte and the McEliece-Rodemich-Rumsey-Welch bound (see, e.g., [12], page 559).

When ϵ is smaller than $\frac{1}{\sqrt{k}}$ we can repeat the above proof but apply Lemma 2.2 instead of Theorem 2.1, as stated in the next Proposition.

Proposition 4.2 *Suppose $\epsilon = \frac{1}{w\sqrt{k}}$, where $w > 1$. Then the cardinality of any ϵ-balanced code C of length k is smaller than $k\frac{w^2}{w^2-1}$.*

Proof: Put $n = |C|$. Applying Lemma 2.2 to the matrix B defined from the code C as in the previous proof, we conclude that

$$k \geq \mathrm{rank}(B) \geq \frac{n}{1 + (n-1)/(w^2 k)},$$

implying the desired bound. □

Thus, in particular, if $w \geq \sqrt{2}$ then $n < 2k$, and if w tends to infinity with k, then $n \leq (1+o(1))k$.

5 Cross intersecting pairs

Extremal Finite Set Theory deals with various instances of the problem of determining or estimating the maximum or minimum possible cardinality of a collection of subsets of a k-element set that satisfies some given conditions. Rank arguments are often useful in obtaining results in this area, see, e.g., [10] for several examples. It is therefore not surprising that one can apply Theorems 1.1 and 2.1 (or Lemma 2.2) in the investigation of problems of this type. Here we only describe one representative example.

Proposition 5.1 *Let c, α be positive constants satisfying $c\alpha > 1$. Let $(X_i, Y_i)_{1 \leq i \leq n}$ be a collection of n pairs of subsets of a k-element set. Suppose that $X_i \cap Y_i = \emptyset$ for all $i \in [n] = \{1, 2, \ldots, n\}$ and that for all distinct $i, j \in [n]$,*

$$| \, |X_i \cap Y_j| - c(k+1) \, | < \frac{\sqrt{k+1}}{\alpha}.$$

Then the number of pairs, n, satisfies $n < \frac{c^2\alpha^2}{c^2\alpha^2 - 1}(k+1)$.

Proof: Let X be the n by k matrix whose rows are the incidence vectors of the sets X_i, and let Y be the k by n matrix whose columns are the incidence vectors of the sets Y_j. Then the product $Z = XY$ is an n by n matrix in which each diagonal entry is zero, and each other entry deviates from $c(k+1)$ by at most $\frac{\sqrt{k+1}}{\alpha}$. Let J be the n by n matrix in which all entries are 1, and define $B = \frac{1}{c(k+1)}(c(k+1)J - Z)$. Then, each diagonal entry of B is 1, and the absolute value of each other entry in it is at most $\frac{\sqrt{k+1}}{\alpha}\frac{1}{c(k+1)} = \frac{1}{c\alpha\sqrt{k+1}}$. Note that the rank of B is at most $k+1$, as the rank of Z does not exceed k. On the other hand, by Lemma 2.2, the rank of B is larger than

$$\frac{n}{1 + n/(c^2\alpha^2(k+1))}.$$

It follows that

$$k + 1 > \frac{n}{1 + n/(c^2\alpha^2(k+1))},$$

implying that $n < \frac{c^2\alpha^2}{c^2\alpha^2 - 1}(k+1)$, as needed. □

By the above proposition, whenever $c\alpha$ is bounded away from 1, the maximum possible number of pairs is linear in k. The existence of Hadamard matrices shows that for an appropriate c this number is at least $(1 - o(1))k$ even if α is arbitrarily large, implying that the above estimate is nearly tight.

6 Pseudo-randomness

In a series of papers, Mauduit and Sárközy studied finite pseudo-random binary sequences $E_N = (e_1, \ldots, e_N) \in \{-1, 1\}^N$. In particular, they investigated in [11] a certain measure of pseudo-randomness, defined as follows.

Given $k, M \leq N$ and $D = \{d_1, \ldots, d_k\}$, where the d_i are integers with $1 \leq d_1 < \ldots < d_k \leq N - M + 1$, define

$$V(E_N, M, D) = \sum_{0 \leq n < M} \prod_{1 \leq i \leq k} e_{n+d_i} = \sum_{0 \leq n < M} \prod_{d \in D} e_{n+d}.$$

The *correlation measure* of order k of E_N is defined as

$$C_k(E_N) = \max\{|V(E_N, M, D)| \quad M \text{ and } D \text{ such that } \; M - 1 + d_k \leq N\}.$$

Improving an estimate of [6], the following is proved in [3] (among other related results).

Theorem 6.1 ([3], Theorem 1.2) *There is an absolute constant $c > 0$ for which the following holds. For any positive integers ℓ and N with $\ell \leq N/3$, we have*

$$\max\{C_2(E_N), C_4(E_N), \ldots, C_{2\ell}(E_N)\} \geq c\sqrt{\ell N}$$

for all $E_N \in \{-1, 1\}^N$.

The proof is a simple consequence of Theorem 2.1. Here is a sketch. Fix a sequence $E_N = (e_1, e_2, \ldots, e_N)$ for which the above maximum is as small as possible, and denote it by T. For every subset A of at most ℓ distinct members of $\{1, 2, \ldots, 2N/3\}$, consider the $\{-1, 1\}$-vector $x(A)$ of length $N/3$ whose i-th coordinate, for $1 \leq i \leq N/3$ is the product $\prod_{a \in A} e_{i+a}$. The set of all vectors $x(A)$ is a set of $\sum_{j=0}^{\ell} \binom{2N/3}{j}$ vectors. The inner product of any two distinct vectors in this set is, in absolute value, at most T. Therefore, the gram matrix of the vectors, divided by $N/3$, has 1 in each diagonal entry, and an element of absolute value at most $3T/N$ in each other entry. It follows, by Theorem 2.1, that its rank is at least

$$\Omega(\frac{N^2}{T^2 \log(N/T)} \log[\sum_{j=0}^{\ell} \binom{2N/3}{j}]).$$

however, this rank is at most $2N/3$, implying that $2N/3$ is at least as large as the last expression. This implies the assertion of the theorem by some simple calculation which is omitted. For more details see [3], where it is also shown that this estimate is sharp up to a logarithmic factor.

7 Derandomization

7.1 Nearly independent random variables

Let $X = \{X_1, X_2, \ldots, X_n\}$ be a set of random variables over a sample space S of size m, and suppose each variable takes values in $\{-1, 1\}$. For every subset $Y \subset [n]$, let X_Y denote the random variable

$X_Y = \Pi_{i \in Y} X_i$. The family X is called *ϵ-biased* if for every nonempty Y,

$$|Prob[X_Y = 1] - Prob[X_Y = -1]| \leq \epsilon.$$

Note that it is more common to consider random variables attaining values in $\{0, 1\}$, and look at their linear combinations over Z_2, but the above definition is equivalent.

It is known (see [2]) that if S is a *uniform* sample space of size m supporting an ϵ-biased set X as above, where $\epsilon \geq 2^{-n/2}$, then $m \geq \Omega(\frac{n}{\epsilon^2 \log(1/\epsilon)})$. Here we show that the same lower bound applies even without the assumption that S is uniform.

Theorem 7.1 *Let $X = \{X_1, X_2, \ldots, X_n\}$ be an ϵ-biased set of n random variables over a sample space $S = \{s_1, s_2, \ldots, s_m\}$ of size m. If $\epsilon \geq 2^{-n/2}$ then $m \geq \Omega(\frac{n}{\epsilon^2 \log(1/\epsilon)})$. If $\epsilon < 2^{-n/2}$ then $m \geq \Omega(2^n)$.*

Proof: Suppose $S = \{s_1, s_2, \ldots, s_m\}$, where $Prob(s_i) = p_i$. Define a 2^n by m matrix $U = (U_{Y,s_j})$ whose rows are indexed by the family of all subsets Y of $[n]$, and whose columns are indexed by the points of S as follows: $U_{Y,s_j} = X_Y(s_j)\sqrt{p_j}$.

Put $A = UU^T$ and note that for every two subsets Y_1, Y_2 of $[n]$,

$$A_{Y_1,Y_2} = Prob[X_{Y_1 \oplus Y_2} = 1] - Prob[X_{Y_1 \oplus Y_2} = -1].$$

Therefore, all diagonal entries of A are 1, whereas all off-diagonal entries are, in absolute value, at most ϵ. By Theorem 2.1, if $\epsilon \geq 2^{-n/2}$ then

$$m \geq rank(A) \geq \Omega(\frac{\log(2^n)}{\epsilon^2 \log(1/\epsilon)}),$$

completing the proof for $\epsilon \geq 2^{-n/2}$. The result for $\epsilon < 2^{-n/2}$ follows from the case $\epsilon = 2^{-n/2}$. □

Remark: A similar proof implies that the size m of any (not necessarily uniform) sample space that supports a family of n random variables in which every set of k is ϵ-biased, where $\epsilon \geq [\binom{n}{\lfloor k/2 \rfloor}]^{-1/2}$, satisfies

$$m \geq \Omega(\frac{k \log(n/k)}{\epsilon^2 \log(1/\epsilon)}).$$

As is the case with Theorem 7.1, this is tight, up to the $\log(1/\epsilon)$ term.

7.2 Nearly minwise independent permutations

A family $\mathcal{F}$ of permutations of $[n] = \{1, 2, \ldots, n\}$ is an *ε-approximate k-restricted min-wise independent family* (or an *(ε, k)-min-wise independent family*, for short) if for every nonempty subset X of at most k elements of $[n]$, and for any $x \in X$, the probability that in a random element π of $\mathcal{F}$, x is the minimum element of $\pi(X)$, deviates from $1/|X|$ by at most $\varepsilon/|X|$. This notion can be defined for

the uniform case, when the elements of $\mathcal{F}$ are picked according to a uniform distribution, or for the more general, biased case, in which the elements of $\mathcal{F}$ are chosen according to a given distribution D.

The notion of (ε, k)-min-wise independent families was introduced by Broder et al. [5], motivated by applications in data mining. It is shown in [5] that there are such families of size at most $O\left(\frac{k^2}{\varepsilon^2}\log\left(\frac{n}{k}\right)\right)$ and that each such family must be of size at least $\Omega\left(k^2(1-\sqrt{8\varepsilon})\right)$ in the uniform case, and at least $\Omega\left(\min\left\{k2^{k/2}\log\left(\frac{n}{k}\right), \frac{\log(1/\varepsilon)(\log n - \log\log(1/\varepsilon))}{\varepsilon^{1/3}}\right\}\right)$ in the biased case.

The lower estimates are improved in [4], where the following two results are proved. Note that both supply lower bounds for the biased case, that improve even the known bounds for the uniform case.

Theorem 7.2 *For any $1/3 > \varepsilon > 0$ and $k \geq 3$, and all sufficiently large n, the following holds. Let $\mathcal{F} \subset S_n$ be an (ε, k)-min-wise independent family of permutations of $[n]$, with respect to a distribution D on $\mathcal{F}$. Then*

$$|\mathcal{F}| \geq \Omega\left(\frac{k}{\varepsilon^2\log(1/\varepsilon)}\log n\right).$$

Theorem 7.3 *For any $1/3 > \varepsilon > 0$ and $k \geq 3$, and all sufficiently large n, the following holds. Let $\mathcal{F} \subset S_n$ be an (ε, k)-min-wise independent family of permutations of $[n]$, with respect to a distribution D on $\mathcal{F}$. Then*

$$|\mathcal{F}| \geq \Omega\left(\frac{k^2}{\varepsilon\log(1/\varepsilon)}\log n\right).$$

The proofs are based on Theorem 1.1, together with some additional linear-algebra arguments. Here is the proof of the first result.

Proof of Theorem 7.2: Let $\mathcal{F}$ be an (ε, k)-min-wise independent family of permutations of $[n]$, with respect to the distribution D, where $\varepsilon > 0$, $k \geq 3$ and n is large. Put $s = k/3$, $L = n/s$ and partition $[n]$ into L pairwise disjoint sets $X_0, X_1, \ldots, X_{L-1}$, each of size s, where $X_0 = \{1, 2, \ldots, s\}$. Put $\mathcal{F} = \{\pi_1, \pi_2, \ldots, \pi_d\}$, $m = L - 1$, and define, for each $h \in [s]$, an m by d matrix $U^{(h)} = (u_{ij}^{(h)})$ as follows:

$$u_{ij}^{(h)} = \begin{cases} \sqrt{\mathrm{Prob}_D(\pi_j)} & \text{if } \min(\pi_j(X_0 \cup X_i)) = \pi_j(h) \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Define $V^{(h)} = (v_{ij}^{(h)}) = U^{(h)}(U^{(h)})^T$ and observe that $v_{ii}^{(h)}$ is precisely the probability that h is the minimum element of $X_0 \cup X_i$ (according to the distribution D on $\mathcal{F}$), whereas for $i \neq j$, $v_{ij}^{(h)}$ is the probability that h is the minimum element of $X_0 \cup X_i \cup X_j$ according to the same distribution. By the assumption on $\mathcal{F}$ and D, each $v_{ii}^{(h)}$ deviates from $\frac{1}{2s}$ by at most $\frac{\varepsilon}{2s}$, and each $v_{ij}^{(h)}$ for $i \neq j$ deviates from $\frac{1}{3s}$ by at most $\frac{\varepsilon}{3s}$. In addition, by the definition of the matrices $U^{(h)}$, for any distinct $h, g \in [s]$, $U^{(h)}(U^{(g)})^T = 0$.

Let U be the ms by d matrix defined by $U^T = [(U^{(1)})^T, (U^{(2)})^T, \ldots, (U^{(s)})^T]$. Then $V = UU^T$ is a block-diagonal matrix whose blocks are the matrices $V^{(h)}$, implying that its rank is the sum of ranks of the matrices $V^{(h)}$.

The crucial claim now is that the rank of each matrix $V^{(h)}$ is at least $\Omega(\frac{1}{\varepsilon^2 \log(1/\varepsilon)} \log m)$. Indeed, if we subtract from $V^{(h)}$ the rank-one matrix in which every entry is exactly $\frac{1}{3s}$, and multiply the result by $6s$, we get a matrix in which each diagonal entry is at least $\frac{1}{2}$, and each off-diagonal entry is in absolute value at most 2ε. As the above subtraction and multiplication can change the rank by at most 1, the assertion of the claim follows from Theorem 1.1. Combining this with the fact that for all large n ($n > k^2$ will suffice here), $\log m > 0.5 \log n$, and the fact that $|\mathcal{F}| = d \geq \text{rank}(V)$, the assertion of the theorem follows. $\Box$

The proof of Theorem 7.3 is similar, with an extra combinatorial argument. The idea is to replace the family of sets $\{X_1, X_2, \ldots, X_{L-1}\}$ in the proof above by a larger family of s-subsets of $[n] - X_0$, so that the intersection of every two of them is at most ϵs. The full details appear in [4].

8 Concluding remarks

The proof of Theorems 1.1 and 2.1 can be easily modified to supply a more general result, as follows.

Theorem 8.1 *Let $B = (b_{i,j})$ be an n by n real, symmetric matrix of rank d, and let $P(z)$ be an arbitrary polynomial of degree k. Then the following inequality holds*

$$\binom{d+k}{k} \geq \frac{[\sum_{i=1}^n P(b_{i,i})]^2}{\sum_{i,j=1}^n P^2(b_{i,j})}.$$

Indeed, this follows by noticing that the proof of Lemma 2.2 implies the known fact that the rank of any real, symmetric matrix is at least the ratio between the square of its trace, and the trace of its square, and by applying this fact, together with the assertion of Lemma 2.3 to the matrix $P(b_{i,j})$. As mentioned in Section 2, here too, the symmetry assumption is not very crucial, as any matrix can be made symmetric by averaging it with its transposed, a process that does not change the rank by more than a factor of 2, maintains the trace, and does not increase the trace of the square.

The main open problem concerning the assertion of Theorems 1.1 and 2.1 is whether it is possible to remove the $\log(1/\epsilon)$ term in their statement when n is sufficiently large as a function of ϵ. If possible, this would be tight up to a constant factor, as shown by many of the applications described throughout the paper, where the gap between the upper and lower bounds is $\Theta(\log(1/\epsilon))$. Note that when $\epsilon = \frac{1}{\sqrt{n}}$, the $\log(1/\epsilon)$-term cannot be omitted.

In most of the proofs throughout the paper, and in particular, in the proof of Theorem 2.1, we made no attempt to optimize the absolute constants involved. In some cases these constants may be of interest, and it is thus worthwhile to note that the estimates can be improved by replacing the

polynomial $P(z) = z^k$ used in the proof of Theorem 2.1 by an appropriate Chebyshev Polynomial. Indeed, the proof suggests that the best choice of a polynomial P of degree k for which we consider the matrix $P(b_{i,j})$, is the polynomial of degree P for which the maximum value of $|P(z)|$ over $z \in [-\epsilon, \epsilon]$ is minimum, among all polynomials P satisfying $P(1) = 1$. It is known (see [14]) that the optimal polynomial P for this problem can be obtained as follows.

The Chebyshev polynomials of the first kind $T_k(z)$ can be defined by $T_0(Z) = 1$, $T_1(z) = z$ and $T_{k+1}(z) = 2T_k(z) - T_{k-1}(z)$ for all $k \geq 1$. Equivalently, $T_k(z) = \cosh(k \cosh^{-1}(z))$, where $\cosh(z) = \frac{e^z+e^{-z}}{2}$. It is known that if $[a, b]$ is a real interval where $b > a > 0$, then among all polynomials t of degree k that satisfy $t(0) = 1$, the one for which the maximum of the absolute value in $[a, b]$ is minimal, is the polynomial

$$t_k(z) = \frac{T_k(\frac{a+b-2z}{b-a})}{T_k(\frac{a+b}{b-a})}.$$

For this polynomial,

$$\max_{z\in[a,b]} |t_k(z)| = t_k(a) = \frac{1}{T_k(\frac{a+b}{b-a})}.$$

It follows that for our purpose, the best polynomial of degree k is obtained by taking $a = 1 - \epsilon$, $b = 1 + \epsilon$ and $P_k(z) = t_k(1 - z)$ for t_k as above. Therefore, $P_k(1) = 1$, and the maximum value of $|P_k(z)|$ in $[-\epsilon, \epsilon]$ is $T_k(1/\epsilon)^{-1}$. Since $\cosh^{-1}(z) = \ln(z + \sqrt{z^2 - 1})$ and $T_k(z) = \cosh(k \cosh^{-1}(z))$, it is not difficult to check that for small ϵ, $T_k(1/\epsilon)^{-1}$ is roughly $\frac{\epsilon^k}{2^{k-1}}$ for all $k \geq 1$. I follows that by using this polynomial instead of the polynomial z^k in the proof of Theorem 2.1, if ϵ is small and k is large, one can roughly replace ϵ by $\epsilon/2$ in the conclusion of the theorem, improving its estimate by roughly a factor of 4. This does not shed any light on the problem of deciding whether or not the $\log(1/\epsilon)$-term in the statement of the theorem can be removed for sufficiently large n.

References

[1] N. Alon, Problems and results in extremal combinatorics, I, Discrete Math. 273 (2003), 31-53.

[2] N. Alon, O. Goldreich, J. Hastad and R. Peralta, Simple constructions of almost k-wise independent random variables, Proc. 31^{st} IEEE FOCS, St. Louis, Missouri, IEEE (1990), 544-553. Also: Random Structures and Algorithms 3 (1992), 289-304.

[3] N. Alon, Y. Kohayakawa, C. Mauduit, C. G. Moreira and V. Rödl, Measures of pseudorandomness for finite sequences: minimal values, Combinatorics, Probability and Computing 15 (2006), 1-29.

[4] N. Alon, T. Itoh, and T. Nagatani, On (ε, k)-min-wise independent permutations, submitted.

[5] A. Broder, M. Charikar, A. Frieze, and M. Mitzenmacher, Min-wise independent permutations, JCSS 60 (2000), 630-659. A preliminary version appeared in *Proc. of the 30th Annual ACM Symposium on Theory of Computing*, 327–336, 1998.

[6] J. Cassaigne, C. Mauduit and A. Sárközy, On finite pseudorandom binary sequences, VII, The measures of pseudorandomness, Acta Arith. 103 (2002) 97–118.

[7] B. Codenotti, P. Pudlák and G. Resta, Some structural properties of low-rank matrices related to computational complexity, Theoret. Comput. Sci. 235 (2000), 89–107.

[8] M. Hall,*Combinatorial Theory*, Second Edition, Wiley, 1986.

[9] W. B. Johnson and J. Lindenstrauss, Extensions of Lipschitz mappings into a Hilbert space, Contemp. Math. 26, AMS, Providence, RI (1984), 189-206.

[10] S. Jukna, *Extremal Combinatorics*, Springer-Verlag, Berlin, 2001.

[11] C. Mauduit and A. Sárközy, On finite pseudorandom binary sequences, I, Measure of pseudorandomness, the Legendre symbol, Acta Arith. 82 (1997), 365–377.

[12] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*, North-Holland Publishing Co., Amsterdam, 1977.

[13] J. Matoušek, *Lectures on Discrete Geometry*, Springer, 2002.

[14] T. J. Rivin, *The Chebyshev Polynomials*, John Wiley, New York, 1990.