Electronic Colloquium on Computational Complexity, Report No. 135 (2014)

Sign rank, VC dimension and spectral gaps

Noga Alon* Shay Moran† Amir Yehudayoff‡

Abstract

We study the maximum possible sign rank of $N \times N$ sign matrices with a given VC dimension d. For $d = 1$, this maximum is 3. For $d = 2$, this maximum is $\tilde{\Theta}(N^{1/2})$. Similar (slightly less accurate) statements hold for $d > 2$ as well. We discuss the tightness of our methods, and describe connections to combinatorics, communication complexity and learning theory.

We also provide explicit examples of matrices with low VC dimension and high sign rank. Let A be the $N \times N$ point-hyperplane incidence matrix of a finite projective geometry with order $n \geq 3$ and dimension $d \geq 2$. The VC dimension of A is d, and we prove that its sign rank is larger than $N^{\frac{1}{2}-\frac{1}{2d}}$. The large sign rank of A demonstrates yet another difference between finite and real geometries.

To analyse the sign rank of A, we introduce a connection between sign rank and spectral gaps, which may be of independent interest. Consider the $N \times N$ adjacency matrix of a Δ regular graph with a second eigenvalue in absolute value λ and $\Delta \leq N/2$. We show that the sign rank of the signed version of this matrix is at least Δ/λ. A similar statement holds for all regular (not necessarily symmetric) sign matrices. We also describe limitations of this approach, in the spirit of the Alon-Boppana theorem.

1 Introduction

Boolean matrices (that is, matrices with $0, 1$ entries) and sign matrices (with ± 1 entries) naturally appear in many areas of research. We use them to represent set systems and

*Sackler School of Mathematics and Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv 69978, Israel and School of Mathematics, Institute for Advanced Study, Princeton, NJ 08540. nogaa@tau.ac.il. Research supported in part by a USA-Israeli BSF grant, by an ISF grant, by the Israeli I-Core program and by the Oswald Veblen Fund.

†Departments of Mathematics and Computer Science, Technion-IIT, Israel. shaymrn@cs.technion.ac.il.

‡Department of Mathematics, Technion-IIT. Email: amir.yehudayoff@gmail.com. Horev fellow – supported by the Taub foundation. Research also supported by ISF and BSF.

ISSN 1433-8092

graphs in combinatorics, concept classes in learning theory, and boolean functions in communication complexity.

We study the sign rank of sign matrices with low VC dimension, describe a general connection between the sign rank and spectral properties of the corresponding[1] boolean matrix, and provide applications to geometry, combinatorics, communication complexity, and learning theory.

The sign rank of a sign matrix S is defined as

$$\text{sign-rank}(S) = \min\{\text{rank}(M) : M_{i,j}S_{i,j} > 0 \text{ for all } i, j\},$$

where the rank is over the real numbers.

It captures the minimum dimension of a real space in which the matrix can be embedded using half spaces[2]. It is known to be related to large margin classifiers and kernel classifiers [31, 19, 20, 21, 13, 40], to be equivalent to unbounded error communication complexity [33], and appears in various other areas of research.

The VC dimension of a boolean matrix B is defined as follows. A subset C of the columns of B is called shattered if each of the $2^{|C|}$ different patterns of zeros and ones appears in some row in the restriction of B to the columns in C. The VC dimension of B is the maximum size of a shattered subset of columns.

The VC dimension captures the size of the minimum ϵ-net for the underlying set system [23, 27], it captures the PAC learnability [38] of the underlying concept class [11, 39], is related to one round communication complexity under product distributions [28], and is related to many other combinatorial properties and questions.

1.1 Sign rank of matrices with low VC dimension

The VC dimension is bounded from above by the sign rank. However, it is long known that the sign rank is not bounded from above by any function of the VC dimension. Alon, Haussler, and Welzl [7] provided examples of $N \times N$ matrices with VC dimension 2 for which the sign rank tends to infinity with N. Ben-David et al. in [9] used ideas from [6] together with estimates concerning the Zarankiewicz problem to show that many matrices with constant VC dimension (at least 4) have high sign rank.

We further investigate the problem of determining or estimating the maximum possible sign rank of $N \times N$ matrices with VC dimension d. Denote this maximum by $f(N, d)$. We are mostly interested in fixed d and N tending to infinity.

[1] There is a standard transformation of a boolean matrix B to the sign matrix $S = 2B - J$, where J is the all 1 matrix. The matrix S is called the signed version of B, and the matrix B is called the boolean version of S.

[2] That is, the columns correspond to points in $\mathbb{R}^k$ and the rows to half spaces (i.e. collections of all points $x \in \mathbb{R}^k$ so that $\langle x, v \rangle \geq \theta$ for some fixed $v \in \mathbb{R}^k$ and $\theta \in \mathbb{R}$).

We observe that there is a dichotomy between the behaviour of $f(N,d)$ when $d=1$ and when $d>1$. The value of $f(N,1)$ is 3, but for $d>1$, the value of $f(N,d)$ tends to infinity with N. We now discuss the behaviour of $f(N,d)$ in more detail, and describe our results.

We start with the case $d=1$. The following theorem and claim imply that for all $N\geq 4$,

$$f(N,1)=3.$$

The following theorem which was proved in [7] shows that for $d=1$, matrices with high sign rank do not exist. For completeness, we provide our simple and constructive proof in Section 3.

Theorem 1 ([7])**.** *If the VC dimension of a sign matrix M is one then its sign rank is at most* 3*.*

We also mention that the bound 3 is tight (see Section 3 for a proof).

Claim 2. *For $N\geq 4$, the $N\times N$ signed identity matrix (i.e. the matrix with* 1 *on the diagonal and* -1 *off the diagonal) has VC dimension one and sign rank* 3*.*

Next, we consider the case $d>1$, starting with lower bounds on $f(N,d)$. As mentioned above, two lower bounds were previously known: The authors of [7] showed that $f(N,2)\geq\Omega(\log N)$. In [9] it is shown that $f(N,d)\geq\omega(N^{1-\frac{2}{d}-\frac{1}{2^{d/2}}})$, for every fixed d, which provides a nontrivial result only for $d\geq 4$.

We prove the following stronger lower bound.

Theorem 3. *The following lower bounds on $f(N,d)$ hold:*

1. $f(N,2)\geq\Omega(N^{1/2}/\log N)$.
2. $f(N,3)\geq\Omega(N^{8/15}/\log N)$.
3. $f(N,4)\geq\Omega(N^{2/3}/\log N)$.
4. *For every fixed $d>4$,*

$$f(N,d)\geq\Omega(N^{1-(d^2+5d+2)/(d^3+2d^2+3d)}/\log N).$$

The understand part 4 better, notice that

$$\frac{d^2+5d+2}{d^3+2d^2+3d}=\frac{1}{d}+\frac{3d-1}{d^3+2d^2+3d},$$

which is close to $1/d$ for large d. The proofs are described in Section 4, where we also discuss the tightness of our arguments, and surprising connections to two other counting problems.

What about upper bounds on $f(N,d)$? It is shown in [9] that for every matrix in a certain class of $N \times N$ matrices with constant VC dimension, the sign rank is at most $O(N^{1/2})$. The proof uses the connection between sign rank and communication complexity. However, there is no general upper bound for the sign rank of matrices of VC dimension d in [9], and the authors explicitly mention they are unable to get such a result.

Here we prove the following upper bounds, using a concrete embedding of matrices with low VC dimension in real space.

Theorem 4. *For every fixed $d \geq 2$,*

$$f(N,d) \leq O(N^{1-1/d}).$$

In particular, this determines $f(N,2)$ up to a logarithmic factor:

$$\Omega(N^{1/2}/\log N) \leq f(N,2) \leq O(N^{1/2}).$$

The above results imply existence of sign matrices with high sign rank. However, their proofs use counting arguments and hence do not provide a method of certifying high sign rank for explicit matrices. In the next section we show how one can derive a lower bound for the sign rank of many explicit matrices.

1.2 Sign rank and spectral gaps

Spectral properties of boolean matrices are known to be deeply related to their combinatorial structure. Perhaps the best example is Cheeger's inequality which relates spectral gaps to combinatorial expansion [16, 3, 4, 2, 24]. Here, we describe connections between spectral properties of boolean matrices and the sign rank of their signed versions.

Proving strong lower bounds on the sign rank of sign matrices turned out to be a difficult task. The authors of [6] were the first to prove that there are sign matrices with high sign rank, but they have not provided explicit examples. Later on, a breakthrough of Forster [18] showed how to prove lower bounds on the sign rank of explicit matrices, proving, specifically, that Hadamard matrices have high sign rank. Razborov and Sherstov proved that there is a function that is computed by a small depth three boolean circuit, but with high sign rank [35]. It is worth mentioning that no explicit matrix whose sign rank is significantly larger than $N^{\frac{1}{2}}$ is known.

We focus on the case of regular matrices, but a similar discussion can be carried more generally. A boolean matrix is Δ regular if every row and every column in it has exactly Δ ones, and a sign matrix is Δ regular if its boolean version is Δ regular.

An $N \times N$ real matrix M has N singular values $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_N \geq 0$. The largest singular value of M is also called its spectral norm

$$\|M\| = \sigma_1 = \max\{\|Mx\| : \|x\| \leq 1\},$$

where $\|x\|^2 = \langle x, x\rangle$ with the standard inner product. The second largest singular value of M is denoted here by

$$\sigma(M) = \sigma_2.$$

If the ratio $\sigma(M)/\|M\|$ is bounded away from one, or small, we say that M has a spectral gap.

We prove that if B has a spectral gap then the sign rank of S is high.

Theorem 5. *Let B be a Δ regular $N \times N$ boolean matrix with $\Delta \leq N/2$, and let S be its signed version. Then,*

$$\textit{sign-rank}(S) \geq \frac{\Delta}{\sigma(B)}.$$

In many cases a spectral gap for B implies that it has pseudorandom properties. This theorem is another manifestation of this phenomenon since random sign matrices have high sign rank (see [6]).

The theorem above provides a non trivial lower bound on the sign rank of S. There is a non trivial upper bound as well. The sign rank of a Δ regular sign matrix is at most $2\Delta+1$. Here is a brief explanation of this upper bound (see [6] for a more detailed proof). Every row i in S has at most 2Δ sign changes (i.e. columns j so that $S_{i,j} \neq S_{i,j+1}$). This implies that for every i, there is a real univariate polynomial G_i of degree at most 2Δ so that $G_i(j)S_{i,j} > 0$ for all $j \in [N] \subset \mathbb{R}$. To see how this corresponds to sign rank at most $2\Delta + 1$, recall that evaluating a polynomial G of degree 2Δ on a point $x \in \mathbb{R}$ corresponds to an inner product over $\mathbb{R}^{2\Delta+1}$ between the vector of coefficients of G, and the vector of powers of x.

Our proof of Theorem 5 and its limitations are discussed in detail in Section 2.

1.3 Sign rank of finite geometries

In this section, we provide explicit examples of matrices with low VC dimension and high sign rank, which also demonstrate another difference between finite and Euclidean geometries.

Let $d \geq 2$ and $n \geq 3$. Let P be the set of points in a d dimensional projective space of order n, and let H be the set of hyperplanes in the space (i.e. the set of $d-1$

dimensional subspaces). For $d = 2$, this is just a projective plane with points and lines. It is known (see, e.g., [10]) that

$$|P| = |H| = N_{n,d} := n^d + n^{d-1} + \ldots + n + 1 = \frac{n^{d+1} - 1}{n - 1}.$$

Let $A \in \{\pm 1\}^{P \times H}$ be the signed point-hyperplane incidence matrix:

$$A_{p,h} = \begin{cases} 1 & p \in h, \\ -1 & p \notin h. \end{cases}$$

Theorem 6. *The matrix A is $N \times N$ with $N = N_{n,d}$, its VC dimension is d, and its sign rank is larger than $N^{\frac{1}{2} - \frac{1}{2d}}$.*

The sign rank of A is at most $2N_{n,d-1} + 1 = O(N^{1-\frac{1}{d}})$, due to the observation in [6] mentioned above. To see this, note that every point in the projective space is incident to $N_{n,d-1}$ hyperplanes.

For example, plugging $d = 2$, we obtain that the sign rank of the point-line incidence matrix of the projective plane is at least $\Omega(N^{\frac{1}{4}})$ and at most $O(N^{\frac{1}{2}})$. We do not know if any of these bounds is tight. The bound $O(N^{\frac{1}{2}})$ is the best possible using the construction from [6] based on the number of sign changes, since in every ordering of the columns there is a row with $\Omega(N^{\frac{1}{2}})$ sign changes: The hamming distance between every two different columns is $2n = \Omega(N^{\frac{1}{2}})$. Thus, for every ordering of the columns, between every two successive columns there is a sign change in $2n$ rows. This means that the total number of sign changes over all rows is $(N - 1)2n = \Omega(N^{\frac{3}{2}})$, and in particular there is at least one row with $\Omega(N^{\frac{1}{2}})$ sign changes.

Theorem 6 is related to two previous works mentioned above. The authors of [7] have considered the special case of projective planes ($d = 2$) and showed that the sign rank tends to infinity as the order n tends to infinity, but no strong lower bound on the rate was proved. The authors of [9] provided a better existential lower bound than the one given by the theorem above (at least for large d), but the lower bound is not for an explicit matrix.

We now discuss the proof. It is well known that the VC dimension of A is d, but we provide a brief explanation. The VC dimension is at least d by considering any set of d independent points (i.e. so that no strict subset of it spans it). The VC dimension is at most d since every set of $d + 1$ points is dependent in a d dimensional space.

The lower bound on the sign rank follows immediately from Theorem 5, and the following known bound on the spectral gap of these matrices.

Lemma 7. *If B is the boolean version of A then*

$$\frac{\sigma(B)}{\Delta} = \frac{n^{\frac{d-1}{2}}(n-1)}{n^d - 1} \leq N_{n,d}^{-\frac{1}{2}+\frac{1}{2d}}.$$

The proof is so short that we include it here.

Proof. We use the following two known properties (see, e.g., [10]) of projective spaces. Both the number of distinct hyperplanes through a point and the number of distinct points on a hyperplane are $N_{n,d-1}$. The number of hyperplanes through two distinct points is $N_{n,d-2}$. The first property implies that A is $\Delta = N_{n,d-1}$ regular. These properties also imply

$$BB^T = (N_{n,d-1} - N_{n,d-2})\, I + N_{n,d-2} J = n^{d-1} I - N_{n,d-2} J.$$

Therefore, all singular values except the maximum one are $n^{\frac{d-1}{2}}$. □

1.4 Applications

1.4.1 Communication complexity

We briefly explain the notions from communication complexity we use. For formal definitions, background and more details, see the textbook [29].

For a function f and a distribution μ on its inputs, define $D_\mu(f)$ as the minimum communication complexity of a deterministic[3] protocol that correctly computes f with error $1/3$ over inputs from μ. Define

$$D^\times(f) = \max\{D_\mu(f) : \mu \text{ is a product distribution}\}.$$

Define the unbounded error communication complexity $U(f)$ of f as the minimum communication complexity of a randomized private coin[4] protocol that correctly computes f with probability strictly larger than $1/2$ on every input.

Two works of Sherstov [37, 36] showed that there are matrices with small distributional communication complexity under product distributions, but whose randomized complexity is almost as large as possible. In [37] the separation is as strong as possible but it is not for an explicit function, and the separation in [36] is not as strong but the underlying function is explicit.

[3] In the distributional setting, every randomized protocol for f can be replaced by a deterministic protocol for f without increasing the error nor the communication.

[4] In the public coin model, every boolean function has unbounded communication complexity at most two.

The matrix A with $d = 2$ and $n \geq 3$ in our example from Section 1.3 corresponds to the following communication problem: Alice gets a point $p \in P$, Bob gets a line $\ell \in L$, and they wish to decide whether $p \in \ell$ or not. Let $f : P \times L \to \{0, 1\}$ be the corresponding function and let $m = \lceil \log_2(N) \rceil$. A trivial protocol would be that Alice sends Bob using m bits the name of her point, Bob checks whether it is incident to the line, and outputs accordingly.

Theorem 6 implies the following consequences. Even if we consider protocols that use randomness and are allowed to err with probability less than but arbitrarily close to $\frac{1}{2}$, then still one can not do considerably better than the above trivial protocol. However, if the input $(p, \ell) \in P \times L$ is distributed according to a product distribution then there exists an $O(1)$ protocol that errs with probability at most $\frac{1}{3}$.

Corollary 8. *The unbounded error communication complexity of f is*[5]

$$U(f) \geq \frac{m}{4} - O(1).$$

The distributional communication complexity of f under product distributions is

$$D^{\times}(f) \leq O(1).$$

These two seemingly contradicting facts are a corollary of the high sign rank and the low VC dimension of A, using two known results. The upper bound on $D^{\times}(f)$ follows from the fact that $\text{VCdim}(A) = 2$, and the work of Kremer et al. [28] which used the PAC learning algorithm to construct an efficient (one round) communication protocol for f under product distributions. The lower bound on $U(f)$ follows from that $\text{sign-rank}(A) \geq \Omega(N^{1/4})$, and the result of Paturi and Simon [33] which showed that unbounded error communication complexity is equivalent to the logarithm of the sign rank. See [37] for more details.

1.4.2 Learning theory

Learning theory started with Valiant's seminal paper [38], in which PAC learning was introduced. Vapnik and Chervonenkis [39] and Blumer et al. [11] proved that PAC learnability is exactly captured by VC dimension. Specifically, a concept class of constant VC dimension, like A, can be PAC learnt using $O(1)$ many examples.

Large margin classifiers concern finding an efficient embedding of the concept class in real space, and using the geometry of Euclidean space to perform the learning (see e.g. [13, 40, 31] and references within). One example is Klivans and Servedio's algorithm for learning DNF formulas [26].

[5]By taking larger values of d, the constant $\frac{1}{4}$ may be increased to $\frac{1}{2} - \frac{1}{2d}$.

The example above shows that although A can be PAC learnt with a constant number of examples, if we try to learn A via embedding it in a real space then the dimension must be extremely high (and the margin small by the Johnson-Lindenstrauss lemma [25], see [9]).

1.4.3 Geometry

Differences and similarities between finite geometries and real geometry are well known. An example of a related problem is finding the minimum dimension of Euclidean space in which we can embed a given finite plane (i.e. a collection of points and lines satisfying certain axioms). By embed we mean that there are two one-to-one maps e_P, e_L so that $e_P(p) \in e_L(\ell)$ iff $p \in \ell$ for all $p \in P, \ell \in L$. The Sylvester-Gallai theorem shows, for example, that Fano's plane can not be embedded in any finite dimensional real space if points are mapped to points and lines to lines.

How about a less restrictive meaning of embedding? One option is to allow embedding using half spaces, that is, an embedding in which points are mapped to points but lines are mapped to half spaces. Such embedding is always possible if the dimension is high enough: Every plane with point set P and line set L can be embedded in $\mathbb{R}^P$ by choosing $e_P(p)$ as the p'th unit vector, and $e_L(\ell)$ as the half space with positive projection on the vector with 1 on points in ℓ and -1 on points outside ℓ. The minimum dimension for which such an embedding exists is captured by the sign rank of the underlying incidence matrix (up to a ± 1).

Corollary 9. *A finite projective plane of order $n \geq 3$ can not be embedded in $\mathbb{R}^k$ using half spaces, unless $k > N^{1/4} - 1$ with $N = n^2 + n + 1$.*

Roughly speaking, the corollary says that there are no efficient ways to embed finite planes in real space using half spaces.

1.4.4 Counting graphs

Here we describe an application of our method for proving Theorem 4 to counting graphs with a given forbidden substructure.

Let $G = (V, E)$ be a graph (not necessarily bipartite). The universal graph $U(d)$ is defined as the bipartite graph with two color classes A and $B = 2^A$ where $|A| = d$, and the edges are defined as $\{a, b\}$ iff $a \in b$. The graph G is called $U(d)$-free if for all two disjoint sets of vertices $A, B \subset V$ so that $|A| = d$ and $|B| = 2^d$, the bipartite graph consisting of all edges of G between A and B is not isomorphic to $U(d)$.

In Theorem 24 of [1] which improves Theorem 2 there, it is proved that for $d \geq 2$, the number of $U(d+1)$-free graphs on N vertices is at most

$$2^{O(N^{2-1/d}(\log N)^{d+2})}.$$

The proof in [1] is quite involved, consisting of several technical and complicated steps. Our methods give a different, quick proof of an improved estimate, replacing the $(\log N)^{d+2}$ term by a single $\log N$ term.

Theorem 10. *For every fixed $d \geq 1$, the number of $U(d+1)$-free graphs on N vertices is at most $2^{O(N^{2-1/d}\log N)}$.*

The proof of the theorem is given in Section 4.1.3.

2 Sign rank and spectral gaps

The lower bound on the sign rank uses Forster's argument [18], who showed how to relate sign rank to spectral norm. He proved that if S is an $N \times N$ sign matrix then

$$\text{sign-rank}(S) \geq \frac{N}{\|S\|}.$$

We would like to apply Forster's theorem to the matrix S in our explicit examples. The spectral norm of S, however, is too large to be useful: If S is $\Delta \leq N/3$ regular and x is the all 1 vector then $Sx = (2\Delta - N)x$ and so $\|S\| \geq N/3$. Applying Forster's theorem to S yields that its sign rank is $\Omega(1)$, which is not informative.

Our solution is based on the observation that Forster's argument actually proves a stronger statement. His proof works as long as the entries of the matrix are not too close to zero, as was already noticed in [19]. We therefore use a variant of the spectral norm of a sign matrix S which we call star norm and denote by[6]

$$\|S\|^* = \min\{\|M\| : M_{i,j}S_{i,j} \geq 1 \text{ for all } i,j\}.$$

Three comments seem in place. (i) We do not think of the star norm as a norm. (ii) It is always at most the spectral norm, $\|S\|^* \leq \|S\|$. (iii) Every M in the above minimum satisfies $\text{sign-rank}(M) = \text{sign-rank}(S)$.

Theorem 11 ([19])**.** *Let S be an $N \times N$ sign matrix. Then,*

$$\textit{sign-rank}(S) \geq \frac{N}{\|S\|^*}.$$

[6]The minimizer belongs to a closed subset of the bounded set $\{M : \|M\| \leq \|S\|\}$.

For completeness, in Section 2.2 we provide a short proof of this theorem (which uses the main lemma from [18] as a black box).

To get any improvement using this theorem, we must have $\|S\|^* \ll \|S\|$. It is not a priori obvious that there is a matrix S for which this holds. The following lemma shows that spectral gaps yield such examples.

Theorem 12. *Let S be a Δ regular $N \times N$ sign matrix with $\Delta \leq N/2$, and B its boolean version. Then,*

$$\|S\|^* \leq \frac{N \cdot \sigma(B)}{\Delta}.$$

In other words, every regular sign matrix whose boolean version has a spectral gap has a small star norm. Theorem 11 and Theorem 12 immediately imply Theorem 5. In Section 1.3, we provided concrete examples of matrices with a spectral gap, that have applications in communication complexity, learning theory and geometry.

Proof of Theorem 12. Define the matrix

$$M = \frac{N}{\Delta} B - J.$$

Observe that since $N \geq 2\Delta$ it follows that $M_{i,j} S_{i,j} \geq 1$ for all i, j. So,

$$\|S\|^* \leq \|M\|.$$

Since B is regular, the all 1 vector y is a right singular vector of B with singular value Δ. Specifically, $My = 0$. For every x, write $x = x_1 + x_2$ where x_1 is the projection of x on y and x_2 is orthogonal to y. Thus,

$$\langle Mx, Mx \rangle = \langle Mx_2, Mx_2 \rangle = \frac{N^2}{\Delta^2} \langle Bx_2, Bx_2 \rangle.$$

Note that $\|B\| \leq \Delta$ (and hence $\|B\| = \Delta$). Indeed, since B is regular, there are Δ permutation matrices $B^{(1)}, \ldots, B^{(\Delta)}$ so that B is their sum. The spectral norm of each $B^{(i)}$ is one. The desired bound follows by the triangle inequality.

Finally, since x_2 is orthogonal to y,

$$\|Bx_2\| \leq \sigma(B) \cdot \|x_2\| \leq \sigma(B) \cdot \|x\|.$$

So,

$$\|M\| \leq \frac{N \cdot \sigma(B)}{\Delta}.$$

□

2.1 Limitations

It is interesting to understand whether the approach above can give a better lower bound on sign rank. There are two parts to the argument: Forster's argument, and the upper bound on $\|S\|^*$. We can try to separately improve each of the two parts.

Any improvement over Forster's argument would be very interesting, but as mentioned there is no significant improvement over it even without the restriction induced by VC dimension, so we do not discuss it further.

To improve the second part, we would like to find examples with the biggest spectral gap possible. The Alon-Boppana theorem [32] optimally describes limitations on spectral gaps. The second eigenvalue σ of a Δ regular graph is not too small,

$$\sigma \geq 2\sqrt{\Delta - 1} - o(1),$$

where the $o(1)$ term vanishes when N tends to infinity (a similar statement holds when the diameter is large [32]). Specifically, the best lower bound on sign rank this approach can yield is roughly $\sqrt{\Delta}/2$, at least when $\Delta \leq N^{o(1)}$.

But what about general lower bounds on $\|S\|^*$? It is well known that any $N \times N$ sign matrix S satisfies $\|S\| \geq \sqrt{N}$. We prove a generalization of this statement.

Lemma 13. *Let S be an $N \times N$ sign matrix. For $i \in [N]$, let γ_i be the minimum between the number of 1's and the number of -1's in the i'th row. Let $\gamma = \gamma(S) = \max\{\gamma_i : i \in [N]\}$. Then,*

$$\|S\|^* \geq \frac{N - \gamma}{\sqrt{\gamma} + 1}.$$

This lemma provides limitations on the bound from Theorem 12. Indeed, $\gamma(S) \leq \frac{N}{2}$ and $\frac{N-\gamma}{\sqrt{\gamma}+1}$ is a monotone decreasing function of γ, which implies $\|S\|^* \geq \Omega(\sqrt{N})$. Interestingly, Lemma 13 and Theorem 12 provide a quantitively weaker but a more general statement than the Alon-Boppana theorem: If B is a Δ regular $N \times N$ boolean matrix with $\Delta \leq N/2$, then

$$\frac{N \cdot \sigma(B)}{\Delta} \geq \frac{N - \Delta}{\sqrt{\Delta} + 1} \quad \Rightarrow \quad \sigma(B) \geq \left(1 - \frac{\Delta}{N}\right)\left(\sqrt{\Delta} - 1\right).$$

This bound is off by roughly a factor of two when the diameter of the graph is large. When the diameter is small, like in the case of the projective plane which we discuss in more detail below, this bound is actually almost tight: The second largest singular value of the boolean point-line incidence matrix of a projective plane of order n is $\sqrt{n}$ while this matrix is $n + 1$ regular (c.f., e.g., [5]).

It is perhaps worth noting that in fact here there is a simple argument that gives a slightly stronger result for boolean regular matrices. The sum of squares of the singular values of B is the trace of B^tB, which is $N\Delta$. As the spectral norm is Δ, the sum of squares of the other singular values is $N\Delta - \Delta^2 = \Delta(N-\Delta)$, implying that

$$\sigma(B) \geq \sqrt{\frac{\Delta(N-\Delta)}{N-1}},$$

which is (slightly) larger than the bound above.

Proof of Lemma 13. Let M be a matrix so that $\|M\| = \|S\|^*$ and $M_{i,j}S_{i,j} \geq 1$ for all i,j. Assume without loss of generality[7] that γ_i is the number of -1's in the i'th row of S. If $\gamma = 0$, then S has only positive entries which implies $\|M\| \geq N$ as claimed. So, we may assume $\gamma \geq 1$. Let t be the largest real so that

$$t^2 = \frac{(N-\gamma-t)^2}{\gamma}. \tag{1}$$

That is, if $\gamma = 1$ then $t = \frac{N-\gamma}{2}$ and if $\gamma > 1$ then

$$t = \frac{-(N-\gamma) + \sqrt{(N-\gamma)^2 + (\gamma-1)(N-\gamma)^2}}{\gamma - 1}.$$

In both cases,

$$t = \frac{N-\gamma}{\sqrt{\gamma}+1}.$$

We shall prove that

$$\|M\| \geq t.$$

There are two cases to consider. One is that for all $i \in [N]$ we have $\sum_j M_{i,j} \geq t$. In this case, if x is the all 1 vector then

$$\|M\| \geq \frac{\|Mx\|}{\|x\|} \geq t.$$

The second case is that there is $i \in [N]$ so that $\sum_j M_{i,j} < t$. Assume without loss of

[7] Multiplying a row by -1 does not affect $\|S\|^*$.

generality that $i = 1$. Denote by C the subset of the columns j so that $M_{1,j} < 0$. Thus,

$$\begin{aligned}\sum_{j \in C} |M_{1,j}| &> \sum_{j \notin C} M_{1,j} - t \\ &\geq |[N] \setminus C| - t && (|M_{i,j}| \geq 1 \text{ for all } i, j) \\ &\geq N - \gamma - t. && (|C| \leq \gamma)\end{aligned}$$

Convexity of $x \mapsto x^2$ implies that

$$\left(\sum_{j \in C} |M_{1,j}| \right)^2 \leq |C| \sum_{j \in C} M_{1,j}^2,$$

so by (1)

$$\sum_j M_{1,j}^2 \geq \frac{(N - \gamma - t)^2}{\gamma} = t^2.$$

In this case, if x is the vector with 1 in the first entry and 0 in all other entries then

$$\|(M)^T x\| = \sqrt{\sum_j M_{1,j}^2} \geq t = t\|x\|.$$

Since $\|(M)^T\| = \|M\|$, it follows that $\|M\| \geq t$.

□

2.2 A proof of Forster's theorem

Forster's argument is based on the following lemma, which he proved.

Lemma 14 ([18])**.** *Let $X \subset \mathbb{R}^k$ be a finite set in general position, i.e., every k vectors in it are linearly independent. Then, there exists an invertible matrix B so that*

$$\sum_{x \in X} \frac{1}{\|Bx\|^2} Bx \otimes Bx = \frac{|X|}{k} I,$$

where I is the identity matrix, and $Bx \otimes Bx$ is the rank one matrix with (i, j) entry $(Bx)_i (Bx)_j$.

The lemma shows that every X in general position can be linearly mapped to BX that is, in some sense, equidistributed. In a nutshell, the proof of the lemma is by finding $B_1, B_2, \ldots$ so that each B_i makes $B_{i-1}X$ closer to being equidistributed, and finally using that the underlying object is compact, so that this process reaches its goal.

Proof of Theorem 11. Let M be a matrix so that $\|M\| = \|S\|^*$ and $M_{i,j}S_{i,j} \geq 1$ for all i, j. Clearly, sign-rank(S) = sign-rank(M). Let X, Y be two subsets of size N of unit vectors in $\mathbb{R}^k$ with k = sign-rank(M) so that $\langle x, y\rangle M_{x,y} > 0$ for all x, y. Lemma 14 says that we can assume

$$\sum_{x \in X} x \otimes x = \frac{N}{k} I; \tag{2}$$

If necessary replace X by BX and Y by $(B^T)^{-1}Y$, and then normalize (the assumption required in the lemma that X is in general position may be obtained by a slight perturbation of its vectors).

The proof continues by bounding $D = \sum_{x\in X, y \in Y} M_{x,y}\langle x, y\rangle$ in two different ways. First, bound D from above: Observe that for every two vectors u, v, Cauchy-Schwartz inequality implies

$$\langle Mu, v\rangle \leq \|Mu\|\|v\| \leq \|M\|\|u\|\|v\|. \tag{3}$$

Thus,

$$\begin{aligned} D &= \sum_{i=1}^{k} \sum_{x\in X} \sum_{y \in Y} M_{x,y} x_i y_i \\ &\leq \sum_{i=1}^{k} \|M\| \sqrt{\sum_{x \in X} x_i^2} \sqrt{\sum_{y\in Y} y_i^2} && \text{((3))} \\ &\leq \|M\| \sqrt{\sum_{i=1}^{k} \sum_{x\in X} x_i^2} \sqrt{\sum_{i=1}^{k}\sum_{y \in Y} y_i^2} = \|M\|N. && \text{(Cauchy-Schwartz)} \end{aligned}$$

Second, bound D from below: Since $|M_{x,y}| \geq 1$ and $|\langle x, y\rangle| \leq 1$ for all x, y, using (2),

$$D = \sum_{x\in X}\sum_{y \in Y} M_{x,y}\langle x, y\rangle \geq \sum_{x \in X}\sum_{y\in Y} (\langle x, y\rangle)^2 = \sum_{y\in Y}\sum_{x \in X} \langle y, (x \otimes x) y\rangle = \frac{N}{k}\sum_{y \in Y}\langle y, y\rangle = \frac{N^2}{k}.$$

□

3 VC dimension one

Our goal in this section to show that sign matrices with VC dimension one have sign rank at most 3, and that 3 is tight. Before reading this section, it may be a nice exercise to prove that the sign rank of the $N \times N$ signed identity matrix is exactly three (for

$N \geq 4$).

Let us start by recalling a geometric interpretation of sign rank. Let M by an $R \times C$ sign matrix. A d-dimensional embedding of M using half spaces consists of two maps e_R, e_C so that for every row $r \in [R]$ and column $c \in [C]$, we have that $e_R(r) \in \mathbb{R}^d$, $e_C(c)$ is a half space in $\mathbb{R}^d$, and $M_{r,c} = 1$ iff $e_R(r) \in e_C(c)$. The important property for us is that if M has a d-dimensional embedding using half spaces then its sign rank is at most $d+1$. The $+1$ comes from the fact that the hyperplanes defining the half spaces do not necessarily pass through the origin.

Our goal in this section is to embed M with VC dimension one in the plane using half spaces. The embedding is constructive and uses the following known claim (see, e.g., [17]).

Claim 15. *Let M be an $R \times C$ sign matrix with VC dimension one so that no row appears twice in it, and every column c is shattered (i.e. the two values ± 1 appear in it). Then, there is a column $c_0 \in [C]$ and a row $r_0 \in [R]$ so that $M_{r_0,c_0} \neq M_{r,c_0}$ for all $r \neq r_0$ in $[R]$.*

Proof. For every column c, denote by $ones_c$ the number of rows $r \in [R]$ so that $M_{r,c} = 1$, and let $m_c = \min\{ones_c, R - ones_c\}$. Assume without loss of generality that $m_1 \leq m_c$ for all c, and that $m_1 = ones_1$. Since all columns are shattered, $m_1 \geq 1$. To prove the claim, it suffices to show that $m_1 \leq 1$.

Assume towards a contradiction that $m_1 \geq 2$. For $b \in \{1, -1\}$, denote by $M^{(b)}$ the submatrix of M consisting of all rows r so that $M_{r,1} = b$. The matrix $M^{(1)}$ has at least two rows. Since all rows are different, there is a column $c \neq 1$ so that two rows in $M^{(1)}$ differ in c. Specifically, column c is shattered in $M^{(1)}$. Since $\mathrm{VCdim}(M) = 1$, it follows that c is not shattered in $M^{(-1)}$, which means that the value in column c is the same for all rows of the matrix $M^{(-1)}$. Therefore, $m_c < m_1$, which is a contradiction. □

The embedding we construct has an extra structure which allows the induction to go through: The rows are mapped to points on the unit circle (i.e. set of points $x \in \mathbb{R}^2$ so that $\|x\| = 1$).

Lemma 16. *Let M be an $R \times C$ sign matrix of VC dimension one so that no row appears twice in it. Then, M can be embedded in $\mathbb{R}^2$ using half spaces, where each row is mapped to a point on the unit circle.*

The lemma immediately implies Threorem 1 due to the connection to sign rank discussed above.

Proof. The proof follows by induction on C. If $C = 1$, the claim trivially holds.

The inductive step: If there is a column that is not shattered, then we can remove it, apply induction, and then add a half space that either contains or does not contain all points, as necessary.

So, we can assume all columns are shattered. By Claim 15, we can assume without loss of generality that $M_{1,1} = 1$ but $M_{r,1} = -1$ for all $r \neq 1$.

Denote by r_0 the row of M so that $M_{r_0,c} = M_{1,c}$ for all $c \neq 1$, if such a row exists. Let M' be the matrix obtained from M by deleting the first column, and row r_0 if it exists, so that no row in M' appears twice. By induction, there is an appropriate embedding of M' in $\mathbb{R}^2$.

The following is illustrated in Figure 1. Let $x \in \mathbb{R}^2$ be the point on the unit circle to which the first row in M' was mapped to (this row corresponds to the first row of M as well). The half spaces in the embedding of M' are defined by lines, which mark the borders of the half spaces. The unit circle intersects these lines in finitely many points. Let y, z be the two closest points to x among all these intersection points. Let y' be the point on the circle in the middle between x, y, and let z' be the point on the circle in the middle between x, z. Add to the configuration one more half space which is defined by the line passing through y', z'. If in addition row r_0 exists, then map r_0 to the point x_0 on the circle which is right in the middle between y, y'.

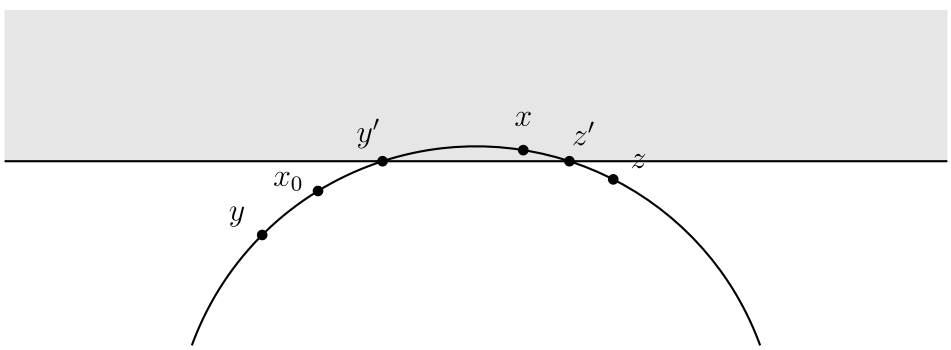

Figure 1: An example of a neighbourhood of x. All other points in embedding of M' are to left of y and right of z on the circle. The half space defined by the line through y', z' is coloured light gray.

This is the construction. Its correctness follows by induction, by the choice of the last added half space which separates x from all other points, and since if x_0 exists it belongs to the same cell as x in the embedding of M'.

□

We conclude the section by showing that the bound 3 above can not be improved.

Proof of Claim 2. One may deduce the claim from Forster's argument, but we provide a more elementary argument. It suffices to consider the case $N = 4$. Consider an arrangement of four half planes in $\mathbb{R}^2$. These four half planes partition $\mathbb{R}^2$ to eight cones with different sign signatures, as illustrated in Figure 2. Let M be the 8×4 sign matrix whose rows are these sign signatures. The rows of M form a distance preserving cycle (i.e. the distance along cycle is hamming distance) of length eight in the discrete cube of dimension four[8]. Finally, the signed identity matrix is not a submatrix of M. To see this, note that the four rows of the signed identity matrix have pairwise hamming distance two, but there are no such four points (not even three points) on this cycle of length eight.

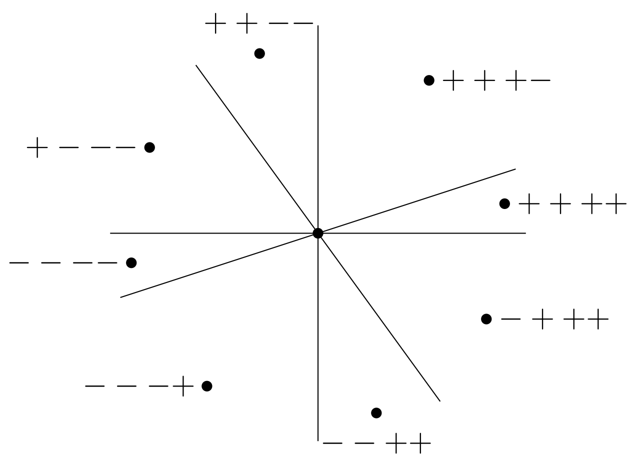

Figure 2: Four lines defining four half planes, and the corresponding eight sign signatures.

□

4 Sign rank and VC dimension

In this section we study the maximum possible sign rank of $N \times N$ matrices with VC dimension d, presenting the proofs of Theorems 4 and 3. We also show that the arguments supply a new, short proof and an improved estimate for a problem in asymptotic enumeration of graphs studied in [1].

4.1 The upper bound

In this subsection we prove Theorem 4. The proof is short, but requires several ingredients. The first one has been mentioned already, and appears in [6]. For a sign matrix

[8] The graph with vertex set $\{\pm 1\}^4$ where every two vectors of hamming distance one are connected by an edge.

S, let $SC(S)$ denote the maximum number of sign changes (SC) along a column of S. Define $SC^*(S) = \min SC(M)$ where the minimum is taken over all matrices M obtained from S by a permutation of the rows.

Lemma 17 ([6])**.** *For any sign matrix S, sign-rank$(S) \leq SC^*(S) + 1$.*

Of course we can replace here rows by columns, but for our purpose the above version will do.

The second result we need is a theorem of Welzl [42] (see also [14]). As observed, for example, in [30], plugging in its proof a result of Haussler [22] improves it by a logarithmic factor, yielding the result we describe next. For a function g mapping positive integers to positive integers, we say that a sign matrix S satisfies a primal shatter function g if for any integer t and any set I of m columns of S, the number of distinct projections of the rows of S on I is at most $g(t)$. The result of Welzl (after its optimization following [22]) can be stated as follows[9].

Lemma 18 ([42], see also [14, 30])**.** *Let S be a sign matrix with N rows that satisfies the primal shatter function $g(t) = ct^d$ for some constants $c \geq 0$ and $d > 1$. Then $SC^*(S) \leq O(N^{1-1/d})$.*

Proof of Theorem 4. Let S be an $N \times N$ sign matrix of VC dimension $d > 1$. By Sauer's lemma [34], it satisfies the primal shatter function $g(t) = t^d$. Hence, by Lemma 18, $SC^*(S) \leq O(N^{1-1/d})$. Therefore, by Lemma 17, sign-rank$(S) \leq O(N^{1-1/d})$. □

4.1.1 On the tightness of the argument

The proof of Theorem 4 works, with essentially no change, for a larger class of sign matrices than the ones with VC dimension d. Indeed, the proof shows that the sign rank of any $N \times N$ matrix with primal shatter function at most ct^d for some fixed c and $d > 1$ is at most $O(N^{1-1/d})$. In this statement the estimate is sharp for all integers d, up to a logarithmic factor. This follows from the construction in [8], which supplies $N \times N$ boolean matrices so that the number of 1 entries in them is at least $\Omega(N^{2-1/d})$, and they contain no d by $D = (d-1)! + 1$ submatrices of 1's. These matrices satisfy the primal shatter function $g(t) = D\binom{t}{d} + \sum_{i=0}^{d-1}\binom{t}{i}$ (with room to spare). Indeed, if we have more than that many distinct projections on a set of t columns, we can omit all projections of weight at most $d-1$. Each additional projection contains 1's in at least one set of size d, and the same d-set can not be covered more than D times. Plugging this matrix in the counting argument that gives a lower bound for the sign rank using

[9] The statement in [42] and the subsequent papers is formulated in terms of somewhat different notions, but it is not difficult to check that it is equivalent to the statement below.

Lemma 21 proven below supplies an $\Omega(N^{1-1/d}/\log N)$ lower bound for the sign rank of many $N \times N$ matrices with primal shatter function $O(t^d)$.

We have seen in Lemma 17 that sign rank is at most of order SC^*. Moreover, for a fixed r, many of the $N \times N$ sign matrices with sign rank at most r also have SC^* at most r: Indeed, a simple counting argument shows that the number of $N \times N$ sign matrices M with $SC(M) < r$ is

$$\left(2 \cdot \sum_{i=0}^{r-1} \binom{N-1}{i}\right)^N = 2^{\Omega(rN \log N)},$$

so, the set of $N \times N$ sign matrices with $SC^*(M) < r$ is a subset of size $2^{\Omega(rN \log N)}$ of all $N \times N$ sign matrices with sign rank at most r. How many $N \times N$ matrices of sign rank at most r are there? by Lemma 21 proved in the next section, this number is at most $2^{O(rN \log N)}$. So, the set of matrices with $SC^* < r$ is a rather large subset of the set of matrices with sign rank at most r.

It is reasonable, therefore, to wonder whether an inequality in the other direction holds. Namely, whether all matrices of sign rank r have SC^* order of r. We now describe an example which shows that this is far from being true, and also demonstrates the tightness of Lemma 18. Namely, for every constant $d > 1$, there are $N \times N$ matrices S, which satisfy the primal shatter function $g(t) = ct^d$ for a constant c, and on the other hand $SC^*(S) \geq \Omega(N^{1-1/d})$. Consider the grid of points $P = [n]^d$ as a subset of $\mathbb{R}^d$. Denote by $e_1, \ldots, e_d$ the standard unit vectors in $\mathbb{R}^d$. For $i \in [n-1]$ and $j \in [d]$, define the hyperplane $h_{i,j} = \{x : \langle x, e_j \rangle > i + (1/2)\}$. Denote by H the set of these $d(n-1)$ axis parallel hyperplanes. Let S be the $P \times H$ sign matrix defined by P and H. That is, $S_{p,h} = 1$ iff $p \in h$. First, the matrix S satisfies the primal shatter function ct^d, since every family of t hyperplanes partition $\mathbb{R}^d$ to at most ct^d cells. Second, we show that

$$SC^*(S) \geq \frac{n^d - 1}{d(n-1)} \geq \frac{|P|^{1-1/d}}{d}.$$

Indeed, fix some order on the rows of S, that is, order the points $P = \{p_1, \ldots, p_N\}$ with $N = |P|$. The key point is that one of the hyperplanes $h_0 \in H$ is so that the number of $i \in [N-1]$ for which $S_{p_i,h_0} \neq S_{p_{i+1},h_0}$ is at least $(n^d-1)/(d(n-1))$: For each i there is at least one hyperplane h that separates p_i and p_{i+1}, that is, for which $S_{p_i,h} \neq S_{p_{i+1},h}$. The number of such pairs of points is $n^d - 1$, and the number of hyperplanes is just $d(n-1)$.

4.1.2 The number of matrices with a given VC dimension

The proof of Theorem 4 also supplies an upper bound for the number of $N \times N$ matrices with VC dimension d, and in fact with primal shatter function $O(t^d)$. Indeed, in each such matrix one can permute the rows and get a matrix in which the number of sign changes in each column is $O(N^{1-1/d})$. The number of ways to choose the permutation is $N!$, and then the number of ways to choose each column is at most $2^{O(N^{1-1/d}\log N)}$. This gives that the total number of such matrices is at most $2^{O(N^{2-1/d}\log N)}$. By the discussion above, this is tight up to the logarithm in the exponent for $d=2$, and for counting matrices with primal shatter function $O(t^d)$ it is tight up to this logarithm for any integer $d>1$, by the construction using the matrices of [8]. For VC dimension 1, it is not difficult to show that the correct number is $2^{\Theta(N\log N)}$.

4.1.3 An application: counting graphs

Proof of Theorem 10. The key observation is that whenever we split the vertices of a $U(d+1)$-free graph into two disjoint sets of equal size, the bipartite graph between them defines a matrix of VC dimension at most d. Hence, the number of such bipartite graphs is at most

$$T(N,d)=2^{O(N^{2-1/d}\log N)}.$$

By a known lemma of Shearer [15], this implies that the total number of $U(d+1)$-free graphs on N vertices is less than $T(N,d)^2=2^{O(N^{2-1/d}\log N)}$. For completeness, we include the simple details. The lemma we use is the following.

Lemma 19 ([15])**.** *Let $\mathcal{F}$ be a family of vectors in $S_1\times S_2\cdots\times S_n$. Let $\mathcal{G}=\{G_1,\dots,G_m\}$ be a collection of subsets of $[n]$, and suppose that each element $i\in[n]$ belongs to at least k members of $\mathcal{G}$. For each $1\le i\le m$, let $\mathcal{F}_i$ be the set of all projections of the members of $\mathcal{F}$ on the coordinates in G_i. Then*

$$|\mathcal{F}|^k\le\prod_{i=1}^m|\mathcal{F}_i|.$$

In our application, $n=\binom{N}{2}$ and $S_1=\ldots=S_n=\{0,1\}$. The vectors represent graphs on N vertices, each vector being the characteristic vector of a graph on N labeled vertices. The set $[n]$ corresponds to the set of all $\binom{N}{2}$ potential edges. The family $\mathcal{F}$ represents all $U(d+1)$-free graphs. The collection $\mathcal{G}$ is the set of all complete bipartite graphs with $N/2$ vertices in each color class. Each edge $i\in[n]$ belongs to at least (in fact a bit more than) half of them, i.e., $k\ge m/2$. Hence,

$$|\mathcal{F}|\le\left(\prod_{i=1}^m|\mathcal{F}_i|\right)^{2/m}\le\left((T(N,d))^m\right)^{2/m},$$

as desired. □

4.2 The lower bound

In this subsection we prove Theorem 3. Our approach follows the one of [6], which is based on known bounds for the number of sign patterns of real polynomials. A similar approach has been subsequently used in [9] to derive lower bounds for $f(N, d)$ for $d \geq 4$, but here we do it in a slightly more sophisticated way and get better bounds.

Although we can use the estimate in [6] for the number of sign matrices with a given sign rank, we prefer to describe the argument by directly applying a result of Warren [41], described next.

Let $P = (P_1, P_2, \ldots, P_m)$ be a list of m real polynomials, each in ℓ variables. Define the semi-variety

$$V = V(P) = \{x \in \mathbb{R}^\ell : P_i(x) \neq 0 \text{ for all } 1 \leq i \leq m\}.$$

For $x \in V$, the sign pattern of P at x is the vector

$$(sign(P_1(x)), sign(P_2(x)), \ldots, sign(P_m(x))) \in \{-1, 1\}^m.$$

Let $s(P)$ be the total number of sign patterns of P as x ranges over all of V. This number is bounded from above by the number of connected components of V.

Theorem 20 ([41]). *Let $P = (P_1, P_2, \ldots, P_m)$ be a list of real polynomials, each in ℓ variables and of degree at most k. If $m \geq \ell$ then the number of connected components of $V(P)$ (and hence also $s(P)$) is at most $(4ekm/\ell)^\ell$.*

An $N \times N$ matrix M is of rank at most r iff it can be written as a product $M = M_1 \cdot M_2$ of an $N \times r$ matrix M_1 by an $r \times N$ matrix M_2. Therefore, each entry of M is a quadratic polynomial in the $2Nr$ variables describing the entries of M_1 and M_2. We thus deduce the following from Warren's Theorem stated above.

Lemma 21. *Let $r \leq N/2$. Then, the number of $N \times N$ sign matrices of sign rank at most r does not exceed $(O(N/r))^{2Nr} \leq 2^{O(rN \log N)}$.*

For a fixed r, this bound for the logarithm of the above quantity is tight up to a constant factor: As argued in Subsection 4.1.1, there are at least some $2^{\Omega(rN \log N)}$ matrices of sign rank r.

In order to derive the statement of Theorem 3 from the last lemma it suffices to show that the number of $N \times N$ sign matrices of VC dimension d is sufficiently large. We proceed to do so. It is more convenient to discuss boolean matrices in what follows (instead of their signed versions).

Proof of Theorem 3. There are 4 parts as follows.

1. The case $d = 2$: Consider the $N \times N$ incidence matrix A of the projective plane with N points and N lines, considered in the previous sections. The number of 1 entries in A is $(1 + o(1))N^{3/2}$, and it does not contain $J_{2\times 2}$ (the 2×2 all 1 matrix) as a submatrix, since there is only one line passing through any two given points. Therefore, any matrix obtained from it by replacing ones by zeros has VC dimension at most 2, since every matrix of VC dimension 3 must contain $J_{2\times 2}$ as a submatrix. This gives us $2^{(1+o(1))N^{3/2}}$ distinct $N \times N$ sign matrices of VC dimension at most 2. Lemma 21 therefore establishes the assertion of Theorem 3, part 1.

2. The case $d = 3$: Call a 5×4 binary matrix heavy if its rows are the all 1 row and the 4 rows with Hamming weight 3. Call a 5×4 boolean matrix heavy-dominating if there is a heavy matrix which is smaller or equal to it in every entry.

We claim that there is a boolean $N \times N$ matrix B so that the number of 1 entries in it is at least $\Omega(N^{23/15})$, and it does not contain any heavy-dominating 5×4 submatrix. Given such a matrix B, any matrix obtained from B by replacing some of the ones by zeros have VC dimension at most 3. This implies part 2 of Theorem 3, using Lemma 21 as before.

The existence of B is proved by a probabilistic argument. Let C be a random binary matrix in which each entry, randomly and independently, is 1 with probability $p = \frac{1}{2N^{7/15}}$. Let X be the random variable counting the number of 1 entries of C minus twice the number of 5×4 heavy-dominant submatrices C contains. By linearity of expectation,

$$\mathbb{E}(X) \geq N^2 p - 2N^{4+5} p^{1\cdot 4 + 4\cdot 3} = \Omega(N^{23/15}).$$

Fix a matrix C for which the value of X is at least its expectation. Replace at most two 1 entries by 0 in each heavy-dominant 5×4 submatrix in C to get the required matrix B.

3. The case $d = 4$: The basic idea is as before, but here there is an explicit construction that beats the probabilistic one. Indeed, Brown [12] constructed an $N \times N$ boolean matrix B so that the number of 1 entries in B is at least $\Omega(N^{5/3})$ and it does not contain $J_{3\times 3}$ as a submatrix (see also [8] for another construction). No set of 5 rows in every matrix obtained from this one by replacing 1's by 0's can be shattered, implying the desired result as before.

4. The case $d > 4$: The proof here is similar to the one in part 2. We prove by a probabilistic argument that there is an $N \times N$ binary matrix B so that the number of 1 entries in it is at least

$$\Omega(N^{2-(d^2+5d+2)/(d^3+2d^2+3d)})$$

and it contains no heavy-dominant submatrix. Here, heavy-dominant means a $1 + (d + 1) + \binom{d+1}{2}$ by $d + 1$ matrix that is bigger or equal in each entry than the matrix whose rows are all the distinct vectors of length $d+1$ and Hamming weight at least $d-1$. Any matrix obtained by replacing 1's by 0's in B cannot have VC dimension exceeding d. The result follows, again, from Lemma 21.

We start as before with a random matrix C in which each entry, randomly and independently, is chosen to be 1 with probability

$$p = \frac{1}{2} \cdot N^{\frac{2-1-(d+1)-\binom{d+1}{2}-(d+1)}{1\cdot(d+1)+(d+1)\cdot d+\binom{d+1}{2}\cdot(d-1)-1}} = \frac{1}{2N^{(d^2+5d+2)/(d^3+2d^2+3d)}}.$$

Let X be the random variable counting the number of 1 entries of C minus three times the number of heavy-dominant submatrices C contains. As before, $\mathbb{E}(X) \geq \Omega(N^2 p)$, and by deleting some of the 1's in C we get B. $\square$

5 Concluding remarks and open problems

We have given explicit examples of $N \times N$ sign matrices with small VC dimension and large sign rank. However, we have not been able to prove that any of them has sign rank exceeding $N^{1/2}$. Indeed this seems to be the limit of Forster's approach, even if we do not bound the VC dimension. Forster's theorem shows that the sign rank of any $N \times N$ Hadamard matrix is at least $N^{1/2}$. It is easy to see that there are Hadamard matrices of sign rank significantly smaller than linear in N. Indeed, the sign rank of the 4×4 signed identity matrix is 3, and hence the sign rank of its k'th tensor power, which is an $N \times N$ Hadamard matrix with $N = 4^k$, is at most $3^k = N^{\log 3/\log 4}$. It may well be, however, that some Hadamard matrices have sign rank linear in N, as do random sign matrices, and it will be very interesting to show that this is the case for some such matrices.

It will also be interesting to decide what is the correct behavior of the sign rank of the incidence graph of the points and lines of a projective plane with N points. We have seen that it is at least $\Omega(N^{1/4})$ and at most $O(N^{1/2})$.

Using our spectral technique we can give many additional explicit examples of matrices with high sign rank, including ones for which the matrices not only have VC dimension 2, but are more restricted than that (for example, no 3 columns have more than 6 distinct projections). Here is a brief description. An (N, Δ, λ)-graph is a Δ regular graph on N vertices so that the absolute value of every eigenvalue of the graph besides the top one is at most λ. There are several known constructions of (N, Δ, λ)-graphs for which $\lambda \leq O(\sqrt{\Delta})$, that do not contain short cycles. Any such graph with $\Delta \geq N^{\Omega(1)}$ provides an example with sign rank at least $N^{\Omega(1)}$, and if there is no cycle of length at most 6 then in the sign matrix we have at most 6 distinct projections on any set of 3 columns.

We have shown that the maximum sign rank $f(N, d)$ of an $N \times N$ matrix with VC dimension $d > 1$ is at most $O(N^{1-1/d})$, and that this is tight up to a logarithmic factor for $d = 2$, and close to being tight for large d. It seems plausible to conjecture that $f(N, d) = \tilde{\Theta}(N^{1-1/d})$ for all $d > 1$.

Finally we note that most of the analysis in this paper can be extended to deal with $M \times N$ matrices, where M and N are not necessarily equal, and we restricted the attention here for square matrices mainly in order to simplify the presentation.

Acknowledgements

We wish to thank Rom Pinchasi, Amir Shpilka, and Avi Wigderson for helpful discussions and comments.

References

[1] N. Alon, J. Balogh, B. Bollobás and R. Morris. The structure of almost all graphs in a hereditary property. J. Comb. Theory, Ser. B 101, pages 85–110, 2011.

[2] N. Alon. Eigenvalues and expanders. Combinatorica 6, pages 83–96, 1986.

[3] N. Alon and V. D. Milman. Eigenvalues, expanders and superconcentrators. In *25th Annual Symp. on Foundations of Computer Science,* pages 320–322, 1984

[4] N. Alon and V. D. Milman. λ_1, isoperimetric inequalities for graphs, and superconcentrators. *J. Comb. Theory, Ser. B*, 38, pages 73–88, 1985.

[5] N. Alon. Eigenvalues, geometric expanders, sorting in rounds and Ramsey Theory. Combinatorica 6, pages 207–219, 1986.

[6] N. Alon, P. Frankl, and V. Rödl. Geometrical realization of set systems and probabilistic communication complexity. In *26th Annual Symposium on Foundations of Computer Science,* pages 277–280, 1985.

[7] N. Alon, D. Haussler, and E. Welzl. Partitioning and geometric embedding of range spaces of finite Vapnik-Chervonenkis dimension. In *Symposium on Computational Geometry*, pages 331–340, 1987.

[8] N. Alon, L. Rónyai and T. Szabó. Norm-graphs: variations and applications. J. Combinatorial Theory, Ser. B 76, pages 280–290, 1999.

[9] S. Ben-David, N. Eiron, and H.-U. Simon. Limitations of learning via embeddings in Euclidean half spaces. Journal of Machine Learning Research, 3, pages 441–461, 2002.

[10] A. Beutelspacher and U. Rosenbaum. *Projektive Geometrie. Von den Grundlagen bis zu den Anwendungen.* Braunschweig: Vieweg, 2nd revised and expanded ed. edition, 2004.

[11] A. Blumer, A. Ehrenfeucht, D. Haussler, and M. K. Warmuth. Classifying learnable geometric concepts with the vapnik-chervonenkis dimension (extended abstract). In *18th Annual ACM Symposium on Theory of Computing*, pages 273–282, 1986.

[12] W. G. Brown. On graphs that do not contain a Thomsen graph. Canad. Math. Bull. 9, pages 281-289, 1966.

[13] C. J. C. Burges. A tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Discov.*, 2(2), pages 121–167, 1998.

[14] B. Chazelle and E. Welzl. Quasi-optimal range searching in spaces of finite VC-dimension. Discrete Comput. Geom. 4, no. 5, 467–489, 1989.

[15] F. R. K. Chung, P. Frankl, R. L. Graham and J. B. Shearer. Some intersection theorems for ordered sets and graphs. J. Combinatorial Theory, Ser. A 43, pages 23–37, 1986.

[16] J. Dodziuk. Difference equations, isoperimetric inequality and transience of certain random walks. Trans. Am. Math. Soc., 284, pages 787–794, 1984.

[17] T. Doliwa, H.-U. Simon, and S. Zilles. Recursive teaching dimension, learning complexity, and maximum classes. In *Algorithmic Learning Theory, 21st International Conference*, volume 6331 of *Lecture Notes in Computer Science*, pages 209–223, 2010.

[18] J. Forster. A linear lower bound on the unbounded error probabilistic communication complexity. In *16th Annual IEEE Conference on Computational Complexity*, pages 100–106, 2001.

[19] J. Forster, M. Krause, S. V. Lokam, R. Mubarakzjanov, N. Schmitt, and H.-U. Simon. Relations between communication complexity, linear arrangements, and computational complexity. In *Foundations of Software Technology and Theoretical Computer Science,* volume 2245 of *Lecture Notes in Computer Science*, pages 171–182, 2001.

[20] J. Forster, N. Schmitt, and H.-U. Simon. Estimating the optimal margins of embeddings in euclidean half spaces. In *14th Annual Conference on Computational Learning Theory,* volume 2111 of *Lecture Notes in Computer Science*, pages 402–415, 2001.

[21] J. Forster and H.-U. Simon. On the smallest possible dimension and the largest possible margin of linear arrangements representing given concept classes. Theor. Comput. Sci., 350(1), pages 40–48, 2006.

[22] D. Haussler,. Sphere packing numbers for subsets of the Boolean n-cube with bounded Vapnik-Chervonenkis dimension. J. Combin. Theory Ser. A 69, no. 2, pages 217–232, 1995.

[23] D. Haussler and E. Welzl. Epsilon-nets and Simplex Range Queries In *2nd Annual Symposium on Computational Geometry,* pages 61–71, 1986.

[24] S. Hoory, N. Linial, and A. Widgerson. Expander graphs and their applications. *Bull. Am. Math. Soc., New Ser.*, 43(4), pages 439–561, 2006.

[25] W. Johnson and J. Lindenstrauss. Extensions of Lipschitz mapping into Hilbert space. Contemporary Mathematics, 26, pages 189–206, 1984.

[26] A. R. Klivans and R. A. Servedio. Learning DNF in time $2^{\tilde{O}(n^{1/3})}$. J. Comput. Syst. Sci., 68(2), pages 303–318, 2004.

[27] J. Komlós and J. Pach and G. J. Woeginger. Almost Tight Bounds for epsilon-Nets Discrete & Computational Geometry, 7, pages 163–173, 1992.

[28] I. Kremer, N. Nisan, and D. Ron. On randomized one-round communication complexity. In *27th Annual ACM Symposium on Theory of Computing,* pages 596–605, 1995.

[29] E. Kushilevitz and N. Nisan. *Communication complexity.* Cambridge University Press, 1997.

[30] J. Matoušek, E. Welzl and L. Wernisch. Discrepancy and approximations for bounded VC-dimension. Combinatorica 13, no. 4, pages 455–466, 1993.

[31] N. Linial and A. Shraibman. Learning complexity vs. communication complexity. In *23rd Annual IEEE Conference on Computational Complexity,* pages 53–63, 2008.

[32] A. Nilli. On the second eigenvalue of a graph. Discrete Math., 91(2), pages 207–210, 1991.

[33] R. Paturi and J. Simon. Probabilistic communication complexity. J. Comput. Syst. Sci., 33(1), pages 106–123, 1986.

[34] N. Sauer. On the density of families of sets. J. Combinatorial Theory, Ser. A 13, pages 145–147, 1972.

[35] A. A. Razborov and A. A. Sherstov. The sign-rank of ACˆo. In *49th Annual IEEE Symposium on Foundations of Computer Science,* pages 57–66, 2008.

[36] A. A. Sherstov. Halfspace matrices. In *22nd Annual IEEE Conference on Computational Complexity,* pages 83–95, 2007.

[37] A. A. Sherstov. Communication complexity under product and nonproduct distributions. Computational Complexity, 19(1), pages 135–150, 2010.

[38] L. G. Valiant. A theory of the learnable. Communications of the ACM, 27, pages 1134–1142, 1984.

[39] V. Vapnik and A. Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. Theory of Probability and its Applications, 16(2), pages 264–280, 1971.

[40] V. Vapnik. *Statistical learning theory.* Wiley, 1998.

[41] H. E. Warren. Lower Bounds for approximation by nonlinear manifolds. Trans. Amer. Math. Soc. 133, pages 167-178, 1968.

[42] E. Welzl. Partition trees for triangle counting and other range searching problems. In *4th Annual Symposium on Computational Geometry,* pages 23–33, 1988.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de