A Sauer–Shelah–Perles Lemma for Lattices

Zeev Dvir* Yuval Filmus† Shay Moran‡

July 13, 2018

Abstract

We extend the Sauer–Shelah–Perles lemma to an abstract setting that is formalized using the language of lattices. Our extension applies to all finite lattices with nonvanishing Möbius function, a rich class of lattices which includes all geometric lattices (or matroids) as a special case.

For example, our extension implies the following result in Algebraic Combinatorics: let $\mathcal{F}$ be a family of subspaces of $\mathbb{F}_q^n$. We say that $\mathcal{F}$ *shatters* a subspace U if for every subspace $U' \leq U$ there is $F \in \mathcal{F}$ such that $F \cap U = U'$. Then, if $|\mathcal{F}| > \begin{bmatrix} n \\ 0 \end{bmatrix}_q + \cdots + \begin{bmatrix} n \\ d \end{bmatrix}_q$ then $\mathcal{F}$ shatters some $(d+1)$-dimensional subspace (where $\begin{bmatrix} n \\ k \end{bmatrix}_q$ denotes the number of k-dimensional subspaces in $\mathbb{F}_q^n$).

1 Introduction

Vapnik–Chervonenkis dimension [31, 32], or VC dimension for short, is a combinatorial parameter of major importance in discrete and computational geometry [9, 17, 19], statistical learning theory [7, 32], and other areas [2, 12, 15, 20]. The VC dimension of a family of binary vectors $F \subseteq \{0,1\}^n$ is the largest cardinality of a set shattered by the family, that is, a set $S \subseteq \{1, \ldots, n\}$ such the projection of F into the coordinates of S is $\{0,1\}^S$. The Sauer–Shelah–Perles lemma [27, 28, 32] states that the largest cardinality of a family on n points with VC dimension d is $\binom{n}{0} + \cdots + \binom{n}{d}$, a bound achieved by the family $\{S : |S| \leq d\}$ (where we identify an n-bit string with its set of 1's and vice versa).

VC dimension (and the attendant Sauer–Shelah–Perles lemma) has been extended to various settings, such as non-binary vectors [1, 16, 18, 30], integer vectors [33], Boolean matrices with forbidden configurations [3, 4], multivalued functions [16], continuous spaces [25], graph powers [8], and ordered variants [5]. In this paper, we formulate a new generalization of VC dimension, to graded lattices, and prove a Sauer–Shelah–Perles lemma for lattices with nonvanishing Möbius function, a rich class of lattices which includes the lattice of subspaces of a finite vector space as well as all geometric lattices (flats of matroids).

VC dimension for ranked lattices One can rephrase the definition of shattering for Boolean vectors in terms of the lattice of subsets of $\{1, \ldots, n\}$ (see Section 2 for the relevant notations and definitions regarding lattices and the Möbius function). A family F of subsets of $\{1, \ldots, n\}$ shatters a set S if for all $T \subseteq S$, the family F contains a set A with $A \cap S = T$. This definition readily generalizes to lattices (indeed, to meet-semilattices): a family F of elements in a lattice L shatters an element $S \in L$ if for all $T \leq S$, the family F contains an element A such that $A \wedge S = T$. If the lattice is ranked, then it is natural to define the VC dimension of a family as the maximum rank of an element it shatters.

Given a lattice Λ, the family F_d consisting of all elements of rank at most d has VC dimension d. This family contains $\begin{bmatrix} \Lambda \\ 0 \end{bmatrix} + \cdots + \begin{bmatrix} \Lambda \\ d \end{bmatrix}$ elements, where $\begin{bmatrix} \Lambda \\ d \end{bmatrix}$ is the number of elements in Λ of rank d. The classical Sauer–Shelah–Perles lemma states that when Λ is the lattice of all subsets of $\{1, \ldots, n\}$, the family F_d

*Department of Computer Science and Department of Mathematics, Princeton University. Email: zeev.dvir@gmail.com. Research supported by NSF CAREER award DMS-1451191 and NSF grant CCF-1523816. Work on this paper was done while the author was a Von-Neumann fellow at the Institute for Advanced Study, supported by a Simons foundation fellowship.

†Computer Science Department, Technion — Israel Institute of Technology. Email: yuvalfi@cs.technion.ac.il. Taub Fellow — supported by the Taub Foundations. The research was funded by ISF grant 1337/16.

‡School of Mathematics, IAS. Email: shaymoran1@gmail.com. Research supported by the National Science Foundation under agreement No. CCF-1412958 and by the Simons Foundations.

has maximum size among all families of VC dimension d. Our main result in this paper generalizes this to ranked lattices with nonvanishing Möbius function:

Theorem 1.1. *If L is a ranked lattice of rank r in which $\mu(x,y) \neq 0$ for all $x \leq y$ then for all $0 \leq d \leq r$, any family of VC dimension d contains at most $\begin{bmatrix} L \\ 0 \end{bmatrix} + \cdots + \begin{bmatrix} L \\ d \end{bmatrix}$ elements.*

The condition on the Möbius function is somewhat mysterious, and we do not know whether it is necessary (i.e. whether any ranked lattice that satisfies the conclusion of Theorem 1.1 has a nonvanishing Möbius function). Fortunately, there is a large class of lattices for which this condition is satisfied, namely geometric lattices. These are lattices whose elements are the flats of a finite matroid. A particularly compelling example of a geometric lattice is the lattice of all subspaces of $\mathbb{F}_q^n$, where $\mathbb{F}_q$ is a finite field of order q. Another example is the Boolean lattice, which is the lattice of all subsets of $\{1,\ldots,n\}$; this is the setting of classical VC theory.

Theorem 1.1 follows from a stronger result, whose version for the Boolean lattice is due to Pajor [26] and Aharoni and Holzman (unpublished).

Theorem 1.2. *If L is a ranked lattice in which $\mu(x,y) \neq 0$ for all $x \leq y$ then every family $F \subseteq L$ shatters at least $|F|$ elements.*

On the proof The original proofs of the Sauer–Shelah–Perles lemma used induction on n. Alon [1] and Frankl [10] gave an alternative proof using combinatorial shifting, and Frankl and Pach [11], Anstee [3], Gurvits [14], Smolensky [29], and Moran and Rashtchian [24] gave other proofs using the polynomial method. Our proof employs the polynomial method, whose use of inclusion-exclusion begets the condition on the Möbius function in our main theorems. The other proof methods — induction and shifting — seem to fail even for the particular case of subspace lattices.

2 Preliminaries

Posets A *poset* is a partially ordered set. Unless mentioned otherwise, all posets we discuss are finite. We will use $\leq$ to denote the partial order. An *antichain* is a collection of elements which are pairwise incomparable. An element x is *covered* by y, denoted $x \lessdot y$, if $x < y$ and no element z satisfies $x < z < y$. We can describe a poset using its *Hasse diagram*, in which the edges correspond to the covering relation, and lower elements are smaller.

A *meet-semilattice* is a poset in which any two elements x, y have an element $z \leq x, y$ such that $w \leq z$ whenever $w \leq x, y$. The element z is denoted $x \wedge y$, and is called the *meet* of x, y. The dual operation is the *join* $x \vee y$. A poset in which any two elements have both a meet and a join is known as a lattice. The meet of all elements in a meet-semilattice is called the *minimal element*, denoted by 0.

A meet-semilattice is *ranked* if each element x is associated with a non-negative integer rank $r(x)$, subject to the following two constraints (which completely specify the rank): $r(0) = 0$, and $r(y) = r(x)+1$ if $x \lessdot y$. Not every meet-semilattice can be ranked. The rank of a meet-semilattice is the maximum rank of an element. We denote the number of elements of rank d in a poset Λ by $\begin{bmatrix} \Lambda \\ d \end{bmatrix}$, and the number of elements of rank at most d by $\begin{bmatrix} \Lambda \\ \leq d \end{bmatrix}$.

The standard example of a lattice is the *Boolean lattice* of all subsets of $\{1,\ldots,n\}$ ordered by inclusion. The meet of two elements is their intersection, and the join of two elements is their union. The rank of a subset is its cardinality.

Möbius function The Möbius function of a finite poset is a function $\mu(x,y)$ defined for any two elements $x \leq y$ in the following way: $\mu(x,x) = 1$, and for $x < y$,

$$\mu(x,y) = -\sum_{z\colon x \leq z < y} \mu(x,z).$$

For example, on the Boolean lattice the Möbius function is $\mu(x,y) = (-1)^{|y \setminus x|}$, and on the integer lattice (the integers $1,\ldots,n$ ordered by divisibility) the Möbius function is $\mu(x,y) = \mu(y/x)$, where $\mu(\cdot)$ is the number-theoretic Möbius function.

The Möbius function is important due to the two *Möbius inversion formulas*:

Lemma 2.1 (Möbius inversion). *If f, g are two real-valued function on a poset then*

$$f(x) = \sum_{y \geq x} g(y) \text{ for all } x \Longleftrightarrow g(x) = \sum_{y \geq x} \mu(x,y) f(y) \text{ for all } x.$$

and

$$f(y) = \sum_{x \leq y} g(x) \text{ for all } y \Longleftrightarrow g(y) = \sum_{x \leq y} \mu(x,y) f(x) \text{ for all } y.$$

We say that a poset has *nonvanishing Möbius function* if $\mu(x,y) \neq 0$ for all $x \leq y$ in the poset. For example, the Boolean lattice has nonvanishing Möbius function, and the integer lattice has nonvanishing Möbius function iff n is squarefree.

Matroids and geometric lattices A *matroid* over a finite set U is a finite non-empty collection of subsets of U called *indepedent sets*, satisfying the following two axioms: if a set is independent, then so are all its subsets; and if A, B are independent and $|A| > |B|$, then there exists an element $x \in A \setminus B$ such that $B \cup \{x\}$ is also independent.

The *rank* of a subset $S \subseteq U$ is the maximum cardinality of a subset of S which is independent. The rank of a matroid is the rank of U. A *flat* is a subset of U whose supersets all have higher rank.

Given a matroid, we can construct a poset whose elements are all flats of the matroid, ordered by inclusion. This poset forms a ranked lattice, and a lattice formed in this way is called a *geometric lattice*. The rank of an element in the lattice is the rank of the corresponding flat in the matroid. Weisner's theorem implies that geometric lattices have nonvanishing Möbius functions:

Theorem 2.2 (Weisner). *The Möbius function of a geometric lattice satisfies $(-1)^{r(y)-r(x)}\mu(x,y) > 0$ for all $x \leq y$.*

For a proof, see [13, Corollary 16.3].

The collection of all subsets of $\{1, \ldots, n\}$ forms a matroid of rank n whose flats are all subsets of $\{1, \ldots, n\}$. The corresponding geometric lattice is the Boolean lattice described above. A more interesting example of a matroid is the collection of all subsets of $\mathbb{F}_q^n$ which are linearly independent, which forms a matroid of rank n whose flats are all subspaces of $\mathbb{F}_q^n$. The corresponding geometric lattice is called the *subspace lattice* of $\mathbb{F}_q^n$.

3 VC theory for lattices

3.1 Definitions

In order to develop VC theory on lattices, we need to define two concepts: shattering and VC dimension. We start with the more basic concept, shattering:

Definition 3.1 (Shattering). Let Λ be a meet-semilattice. A set $F \subseteq \Lambda$ *shatters* an element $y \in \Lambda$ if for all $x \leq y$ there exists an element $z \in F$ such that $z \wedge y = x$.

We denote the set of all elements shattered by F by $\mathrm{Str}(F)$.

We comment that the definition can be extended further to general posets: in this case, the condition $z \wedge y = x$ should be understood as follows: $z \wedge y$ exists, and equals x.

A basic property of shattering is that it is hereditary:

Lemma 3.2. *Let Λ be a meet-semilattice. If a set $F \subseteq \Lambda$ shatters an element $z \in \Lambda$ and $y \leq z$, then F also shatters y. In other words, $\mathrm{Str}(F)$ is downwards-closed.*

Proof. Let $x \leq y$. Since F shatters z and $x \leq z$, there exists an element $w \in F$ satisfying $w \wedge z = x$. Since $y \leq z$, the same element satisfies $w \wedge y = w \wedge (y \wedge z) = (w \wedge z) \wedge y = x \wedge y = x$. □

Having defined shattering, the definition of VC dimension is obvious:

Definition 3.3. Let Λ be a ranked meet-semilattice. The *VC dimension* of a non-empty set $F \subseteq \Lambda$, denoted $\mathrm{VC}(F)$, is the maximum rank of an element shattered by F.

These definitions specialize to the classical ones in the case of the Boolean lattice.

3.2 Proof of the main result

Our main result is Theorem 1.2, from which all other results follow as corollaries.

Theorem 3.4 (Restatement of Theorem 1.2)**.** *If Λ is a meet-semilattice with nonvanishing Möbius function then for all $F \subseteq \Lambda$,*

$$|F| \le |\operatorname{Str}(F)|.$$

Let $\mathbb{F}$ be an arbitrary field of characteristic zero. We will prove Theorem 1.2 by giving a spanning set of size $|\operatorname{Str}(F)|$ for the F-dimensional vector space $\mathbb{F}[F]$ of $\mathbb{F}$-valued functions on F. Theorem 3.4 is then implied since the cardinality of any spanning is at least the dimension. The spanning set we will construct will consist of functions of the form given by the next definition.

Definition 3.5. For $x \in \Lambda$, the function $\chi_x \colon \Lambda \to \mathbb{F}$ is given by

$$\chi_x(y) = \mathbf{1}_{\{y \ge x\}},$$

that is, $\chi_x(y) = 1$ if $y \ge x$, and otherwise $\chi_x(y) = 0$.

For a set $G \subseteq \Lambda$,

$$X(G) = \{\chi_x : x \in G\}.$$

In the case of the Boolean lattice, we can think of the elements of the lattice as encoded by sets $S \subseteq \{1, \dots, n\}$ as well as by Boolean variables $x_1, \dots, x_n$. The reader can verify that

$$\chi_S = \prod_{i \in S} x_i.$$

Definition 3.5 extends this idea to general posets.

We will show that $\mathbb{F}[F]$ is spanned by $X(\operatorname{Str}(F))$. The first step is showing that $X(\Lambda)$ is a basis for $\mathbb{F}[\Lambda]$, which for the Boolean lattice just states that every function on $\{0,1\}^n$ can be expressed uniquely as a multilinear polynomial:

Lemma 3.6. *The set $X(\Lambda)$ is a basis for $\mathbb{F}[\Lambda]$.*

Proof. Since $|X(\Lambda)| = |\Lambda| = \dim \mathbb{F}[\Lambda]$, it suffices to show that $X(\Lambda)$ is linearly independent. Consider any linear dependency of the form $\ell := \sum_x c_x \chi_x = 0$. We will show that $c_x = 0$ for all $x \in \Lambda$, and so $X(\Lambda)$ is linearly independent.

Arrange the elements of Λ in an order $x_1, \dots, x_N$ such that $x_i < x_j$ implies $i < j$. We prove that $c_{x_i} = 0$ by induction on i. Suppose that $c_{x_j} = 0$ for all $j < i$. Then, in particular $c_{x_j} = 0$ for all $x_j < x_i$ and therefore

$$0 = \ell(x_i) = \sum_j c_{x_j} \chi_{x_j}(x_i) = \sum_{j \,:\, x_j \le x_i} c_{x_j} = c_{x_i}. \qquad \square$$

The crucial step of the proof of Theorem 3.4 is an application of (generalized) inclusion-exclusion, which shows that if F does not shatter z then $\chi_z|_F$ can be expressed as a linear combination of $\chi_w|_F$ for $w < z$. In the case of the Boolean lattice, the argument is as follows. Suppose that F does not shatter S, say $A \cap S \neq T$ for all $A \in F$. Then all elements of F satisfy

$$\prod_{i \in T} x_i \prod_{j \in S \setminus T} (1 - x_j) = 0,$$

which implies that over F,

$$\prod_{i \in S} x_i = \sum_{R \subsetneq S \setminus T} (-1)^{|S \setminus (T \cup R)|+1} \prod_{j \in T \cup R} x_j.$$

The argument for general posets is very similar, and uses Möbius inversion:

Lemma 3.7. *Suppose that $z \notin \operatorname{Str}(F)$. There exist coefficients γ_y such that for all $p \in F$,*

$$\chi_z(p) = \sum_{y < z} \gamma_y \chi_y(p).$$

Proof. For an element $p \in F$, define the following two functions:

$$f_p(x) = \mathbf{1}_{\{x \le p \wedge z\}}, \qquad g_p(y) = \mathbf{1}_{\{y = p \wedge z\}}.$$

Clearly $f_p(x) = \sum_{y \ge x} g_p(y)$, and so Lemma 2.1 shows that $g_p(x) = \sum_{y \ge x} \mu(x,y) f_p(y)$. Since $f_p(y) = 0$ unless $y \le z$, we can restrict the sum to the range $x \le y \le z$. When $y \le z$, the condition $y \le p \wedge z$ is equivalent to the condition $y \le p$, and so we conclude that

$$g_p(x) = \sum_{x \le y \le z} \mu(x,y) f_p(y) = \sum_{x \le y \le z} \mu(x,y) \chi_y(p).$$

Since $z \notin \mathrm{Str}(F)$, there exists an element $x \le z$ such that $p \wedge z \ne x$ for all $p \in F$. In other words, $g_p(x) = 0$ for all $p \in F$. Therefore all $p \in F$ satisfy

$$\chi_z(p) = -\sum_{x \le y < z} \frac{\mu(x,y)}{\mu(x,z)} \chi_y(p),$$

using the nonvanishing of the Möbius function. □

We can now complete the proof, employing exactly the same argument used for the Boolean lattice.

Proof of Theorem 3.4. Lemma 3.6 shows that $X(\Lambda)$ is a basis for $\mathbb{F}[\Lambda]$, and so the functions χ_x, restricted to the domain F, span $\mathbb{F}[F]$. We will show that every function in $\mathbb{F}[F]$ can be expressed as a linear combination of functions in $X(\mathrm{Str}(F))$.

Consider any function $f \in \mathbb{F}[F]$. Since $X(\Lambda)$ spans $\mathbb{F}[f]$, there exist coefficients c_x such that $f = \sum_x c_x \chi_x$. Define the potential function

$$\Phi(\vec{c}) = \sum_{\substack{x \notin \mathrm{Str}(F): \\ c_x \ne 0}} N^{r(x)},$$

where $N = |\Lambda| + 1$, and choose a representation which minimizes $\Phi(\vec{c})$. If $\Phi(\vec{c}) > 0$ then choose $z \notin \mathrm{Str}(F)$ satisfying $c_z \ne 0$ of maximal rank. Lemma 3.7 shows that

$$f = \sum_{x \ne z} c_x \chi_x + \sum_{y < z} \gamma_y c_z \chi_y.$$

The corresponding coefficient vector $\vec{d}$ satisfies $\Phi(\vec{d}) < \Phi(\vec{c})$, contradicting the choice of $\vec{c}$. We conclude that $\Phi(\vec{c}) = 0$, and so f is a linear combination of functions in $X(\mathrm{Str}(F))$.

Concluding, we have shown that $X(\mathrm{Str}(F))$ spans $\mathbb{F}[F]$. Hence $|\mathrm{Str}(F)| = |X(\mathrm{Str}(F))| \ge \dim \mathbb{F}[F] = |F|$. □

It is natural to ask whether the condition of nonvanishing of the Möbius function is necessary, a question which remains open. Figure 1 portrays two examples of lattices in which the Möbius function does vanish, and the result of Theorem 3.4 indeed fails to hold:

1. The lattice $\{0,1,2\}$ ordered by the usual order (equivalently, the lattice $\{1,p,p^2\}$ ordered by divisibility). The Möbius function vanishes: $\mu(0,2) = 0$. The set $F = \{1,2\}$ shatters only one element: $\mathrm{Str}(F) = \{0\}$.

2. The lattice $\{[i,j] : 1 \le i \le j \le 3\} \cup \{\emptyset\}$ of intervals ordered by inclusion (where i,j are integers). The Möbius function vanishes: $\mu(\emptyset, \{1,2,3\}) = 0$. The set $F = \{x : 2 \in x\}$ of size 4 shatters only 3 elements: $\mathrm{Str}(F) = \{\emptyset, \{1\}, \{3\}\}$.

(a) Path lattice (b) Interval lattice

Figure 1: Lattices with vanishing Möbius function

3.3 Some corollaries

Theorem 3.4 immediately implies Theorem 1.1.

Corollary 3.8 (Restatement of Theorem 1.1)**.** *If Λ is a ranked meet-semilattice with nonvanishing Möbius function then for all $F \subseteq \Lambda$,*

$$|F| \leq \begin{bmatrix} \Lambda \\ \leq \mathrm{VC}(F) \end{bmatrix}.$$

Furthermore, for every $d \leq r(\Lambda)$ the inequality is tight for some $F \subseteq \Lambda$ of VC dimension d.

Proof. Suppose that $\mathrm{VC}(F) = d$. If $|F| > \begin{bmatrix} \Lambda \\ \leq d \end{bmatrix}$ then, according to Theorem 3.4, also $|\mathrm{Str}(F)| > \begin{bmatrix} \Lambda \\ \leq d \end{bmatrix}$. However, this implies that $\mathrm{Str}(F)$ must contain a set of rank larger than d, contradicting the assumption $\mathrm{VC}(F) = d$. This proves the inequality.

To show that the inequality is tight for all $d \leq r(\Lambda)$, consider the set $F_d = \{x : r(x) \leq d\}$. This is a set containing $\begin{bmatrix} \Lambda \\ \leq d \end{bmatrix}$ elements which shatters all elements of rank d but no element of rank $d + 1$, and so satisfies $\mathrm{VC}(F_d) = d$. □

We can generalize Corollary 3.8 to arbitrary antichains to obtain a further corollary.

Corollary 3.9. *Let Λ be a ranked meet-semilattice with nonvanishing Möbius function and let $A \subseteq \Lambda$ be a maximal antichain. Then, if $F \subseteq \Lambda$ does not shatter any element of A then*

$$|F| \leq |F_A|, \text{ where } F_A = \{x \in \Lambda : x < y \text{ for some } y \in A\}.$$

Furthermore, F_A does not shatter any element of A.

Corollary 3.8 is the special case of Corollary 3.9 in which A consists of all elements of rank $\mathrm{VC}(A)+1$.

Proof of Corollary 3.9. Let us start by showing that $|F| \leq |F_A|$. If $|F| > |F_A|$ then, according to Theorem 3.4, also $|\mathrm{Str}(F)| > |F_A|$. Therefore F shatters some element x such that $x \not< y$ for all $y \in A$. Since A is a maximal antichain, either $x \in A$ or $x \geq y$ for some $y \in A$. In both cases F shatters some set in A (in the second case, according to Lemma 3.2).

Next, let us show that F_A does not shatter any element of A. Suppose that F_A shatters some element $a \in A$. Then some $x \in F_A$ satisfies $x \wedge a = a$, that is, $x \geq a$. Since $x \in F_A$, we know that $x < y$ for some $y \in A$. Put together, this implies that $a < y$, contradicting the fact that A is an antichain. □

A final corollary is a *dichotomy theorem*, a direct consequence of the Sauer–Shelah–Perles lemma which is the source of many of its applications. Before describing our generalized dichotomy theorem, let us briefly describe the classical one. Let $F \subseteq \{0, 1\}^X$, where X is infinite. For every finite $I \subseteq X$, we can consider the projection of F to the coordinates of I, denoted $F|_I$. The *growth function* of F is

$$\Pi_F(n) = \max_{\substack{I \subseteq X \\ |I|=n}} \big|F|_I\big|.$$

The Sauer–Shelah–Perles lemma immediately implies the following polynomial versus exponential dichotomy for the growth function:

- Either $\mathrm{VC}(F) = \infty$, in which case $\Pi_F(n) = 2^n$;
- or $\mathrm{VC}(F) = d < \infty$, in which case $\Pi_F(n) \leq n^d$.

For example, it implies that there is no F for which $\pi_F(n) = \Theta(2^{\log^2 n})$.

We can extend this result to vector spaces (we leave extensions to more general domains to the reader). Let $\mathbb{F}_q$ be a finite field, let X be an infinite set, let $\mathcal{V}$ denote the linear space of all functions $v \colon X \to \mathbb{F}_q$ with a finite support (i.e. $v(x) = 0$ for all but finitely many $x \in X$), and let $\mathcal{L}$ denote the (infinite) lattice of all finite dimensional subspaces of $\mathcal{V}$. Let $F \subseteq \mathcal{L}$ be a family. For every $I \in \mathcal{L}$, we can consider the projection $F|_I = \{V \cap I : V \in F\}$. The growth function of F is defined as in the classical case, with dimension replacing cardinality:

$$\Pi_F(n) = \max_{\substack{I \in \mathcal{L} \\ \dim(I) = n}} \big|F|_I\big|.$$

Corollary 3.8 immediately implies a dichotomy as in the classical case. In order to understand the resulting orders of growth, we need to be able to estimate $\left[\begin{smallmatrix}\Lambda \\ d\end{smallmatrix}\right]$ for subspace lattices Λ.

Lemma 3.10. *Let $\Lambda = \mathbb{F}_q^n$. For all $d \leq n$,*

$$q^{d(n-d)} \leq \begin{bmatrix}\Lambda \\ \leq d\end{bmatrix} \leq 2n^d q^{dn}.$$

In particular, $|\Lambda| \geq q^{(n^2-1)/4}$.

Proof. The number of elements of Λ of rank d is the q-binomial coefficient $\left[\begin{smallmatrix}n \\ d\end{smallmatrix}\right]_q$. There are many formulas for $\left[\begin{smallmatrix}n \\ d\end{smallmatrix}\right]_q$. The one we use is

$$\begin{bmatrix}n \\ d\end{bmatrix}_q = \sum_{\substack{A \subseteq \{1,\ldots,n\} \\ |A| = d}} q^{\sum_{i \in A} i - d(d+1)/2}.$$

Calculation shows that the summand with highest exponent, corresponding to $A = \{n-d+1, \ldots, n\}$, has exponent $d(n-d)$. Therefore

$$q^{d(n-d)} \leq \begin{bmatrix}n \\ d\end{bmatrix}_q \leq \binom{n}{d} q^{d(n-d)} \leq n^d q^{dn}.$$

This implies that

$$\begin{bmatrix}\Lambda \\ \leq d\end{bmatrix} \leq \sum_{e=0}^{d} n^e q^{en} \leq n^d q^{dn} \sum_{e=0}^{d} (nq^n)^{-e}.$$

We can assume that $n \geq 1$, and so $nq^n \geq 2$, implying that $\sum_{e=0}^{\infty} (nq^n)^{-e} \leq 2$. This proves the main inequalities. The lower bound on $|\Lambda|$ follows from taking $m = \lfloor n/2 \rfloor$. ◻

Combining the lemma with Corollary 3.8 specialized to the subspace lattice, we immediately obtain the following dichotomy theorem:

Theorem 3.11. *For every family $F \subseteq \mathcal{L}$, exactly one of the following holds:*

- *Either* $\mathrm{VC}(F) = \infty$, *in which case* $\Pi_F(n) \geq q^{(n^2-1)/4}$;
- *or* $\mathrm{VC}(F) = d < \infty$, *in which case* $\Pi_F(n) \leq 2n^d q^{dn}$.

4 Concluding remarks

The idea of VC dimension has found many applications in mathematics and computer science, and has been studied extensively. We believe that our generalization also merits study on its own right. Theorem 3.4 shows that the Sauer–Shelah–Perles lemma extends to our generalized setting. What other properties extend? One place to initiate such a study is sets with small VC dimension. An explicit description of all sets of VC dimension 1 exists in the classical case: they correspond to forests [6, 21]. In particular, every set of VC dimension 1 can be extended to a set of size $n + 1$ without increasing the VC dimension. The same does not hold in the case of vector spaces, as the following example shows:

Lemma 4.1. *Consider a subspace lattice $\Lambda = \mathbb{F}_q^n$, where $n \geq 2$. Let U be a subspace of $\mathbb{F}_q^n$ of dimension $n-1$. The set $F = \{0, \mathbb{F}_q^n\} \cup \{\langle x \rangle : x \notin U\}$ is an inclusion-maximal set of VC dimension 1, and it contains $q^{n-1}+2$ subspaces; in comparison, $\left[{\Lambda \atop \leq 1}\right] = 1 + \frac{q^n-1}{q-1}$, which is larger by a factor of roughly $\frac{q}{q-1}$.*

Proof. Let us start by noting that $|U| = q^{n-1}$ and so $|\mathbb{F}_q^n \setminus U|$ contains $q^n - q^{n-1} = (q-1)q^{n-1}$ vectors, all of them non-zero. Since every one-dimensional vector space contains $q-1$ non-zero vectors, we see that there are q^{n-1} different one-dimensional vectors spaces of the form $\langle x \rangle$ for $x \notin U$. This implies that $|F| = q^{n-1}+2$.

Next, let us show that $\mathrm{VC}(F) = 1$. Since F clearly shatters $\langle x \rangle$ for any $x \notin U$, it suffices to show that F does not shatter any two-dimensional subspace. Indeed, if $\dim Q = 2$, then Q must intersect U at some non-zero vector x. By construction, if $V \cap Q \ni x$ for some $V \in F$ then $V = \mathbb{F}_q^n$. It follows that $V \cap Q \neq \langle x \rangle$, and so F does not shatter Q.

Finally, let us show that $\mathrm{VC}(F \cup \{V\}) \geq 2$ for any $V \notin F$. Notice first that V must contain some non-zero vector $u \in U$. If $\dim V = 1$ then this is clear, and otherwise it follows from the fact that any subspace of dimension at least 2 intersects U non-trivially. It is easy to see that $\mathbb{F}_q^n \setminus U$ spans all of $\mathbb{F}_q^n$, and so there must be some vector $w \notin U$ which is missing from V.

We claim that $F \cup \{V\}$ shatters $\langle u, w \rangle$, and so $\mathrm{VC}(F \cup \{V\}) \geq 2$. First, $0 \wedge \langle u, w \rangle = 0$ and $\mathbb{F}_q^n \wedge \langle u, w \rangle = \langle u, w \rangle$. Every other subspace of $\langle u, w \rangle$ is a one-dimensional subspace $\langle x \rangle$. If $x \notin U$ then $\langle x \rangle \in F$, and otherwise $\langle x \rangle = \langle u \rangle$, in which case $V \wedge \langle u, w \rangle = \langle u \rangle$. This shows that $F \cup \{V\}$ indeed shatters $\langle u, w \rangle$. □

Can we describe all inclusion-maximal sets of VC dimension 1? What can we say about shattering-extremal systems of small VC dimension [21–23]? We leave these questions for future work.

References

[1] Noga Alon. On the density of sets of vectors. *Discrete Math.*, 46(2):199–202, 1983.

[2] Noga Alon, Shay Moran, and Amir Yehudayoff. Sign rank versus Vapnik–Chervonenkis dimension. *Sbornik: Mathematics*, 208(12):1724, 2017.

[3] R. P. Anstee. General forbidden configuration theorems. *J. Combin. Theory Ser. A*, 40(1):108–124, 1985.

[4] R. P. Anstee and Balin Fleming. Two refinements of the bound of Sauer, Perles and Shelah, and of Vapnik and Chervonenkis. *Discrete Math.*, 310(23):3318–3323, 2010.

[5] R. P. Anstee, Lajos Rónyai, and Attila Sali. Shattering news. *Graphs Combin.*, 18(1):59–73, 2002.

[6] Shai Ben-David. 2 notes on classes with Vapnik–Chervonenkis dimension 1. *ArXiv e-prints*, July 2015.

[7] Anselm Blumer, Andrzej Ehrenfeucht, David Haussler, and Manfred K. Warmuth. Learnability and the Vapnik–Chervonenkis dimension. *J. Assoc. Comput. Mach.*, 36(4):929–965, 1989.

[8] Nicolò Cesa-Bianchi and David Haussler. A graph-theoretic generalization of the Sauer-Shelah lemma. *Discrete Appl. Math.*, 86(1):27–35, 1998.

[9] Bernard Chazelle and Emo Welzl. Quasi-optimal range searching in spaces of finite VC-dimension. *Discrete Comput. Geom.*, 4(5):467–489, 1989.

[10] Peter Frankl. On the trace of finite sets. *J. Combin. Theory Ser. A*, 34(1):41–45, 1983.

[11] Peter Frankl and János Pach. On the number of sets in a null t-design. *European J. Combin.*, 4(1):21–23, 1983.

[12] Zoltán Füredi and János Pach. Traces of finite sets: extremal problems and geometric applications. In *Extremal problems for finite sets (Visegrád, 1991)*, volume 3 of *Bolyai Soc. Math. Stud.*, pages 251–282. János Bolyai Math. Soc., Budapest, 1994.

[13] Chris Godsil. An introduction to the Moebius function. *ArXiv e-prints*, March 2018.

[14] Leonid Gurvits. Linear algebraic proofs of VC-dimension based inequalities. In *Computational Learning Theory*, pages 238–250. Springer, 1997.

[15] Hamed Hatami and Yingjie Qian. Teaching dimension, VC dimension, and critical sets in Latin squares. *J. Comb.*, 9(1):9–20, 2018.

[16] David Haussler and Philip M. Long. A generalization of Sauer's lemma. *J. Combin. Theory Ser. A*, 71(2):219–240, 1995.

[17] David Haussler and Emo Welzl. ϵ-nets and simplex range queries. *Discrete Comput. Geom.*, 2(2):127–151, 1987.

[18] M. G. Karpovsky and V. D. Milman. Coordinate density of sets of vectors. *Discrete Math.*, 24(2):177–184, 1978.

[19] János Komlós, János Pach, and Gerhard Woeginger. Almost tight bounds for ϵ-nets. *Discrete Comput. Geom.*, 7(2):163–173, 1992.

[20] Ilan Kremer, Noam Nisan, and Dana Ron. On randomized one-round communication complexity. *Comput. Complexity*, 8(1):21–49, 1999.

[21] Tamás Mészáros and Lajos Rónyai. Shattering-extremal set systems of small VC dimension. *ISRN Comb.*, 2013.

[22] Tamás Mészáros and Lajos Rónyai. Shattering-extremal set systems of VC dimension at most 2. *Electron. J. Combin.*, 21(4):Paper 4.30, 17, 2014.

[23] Shay Moran. Shattering extremal sets. Master's thesis, Saarland University, Saarbrücken, Germany, 2012.

[24] Shay Moran and Cyrus Rashtchian. Shattered sets and the hilbert function. In *MFCS*, volume 58 of *LIPIcs*, pages 70:1–70:14. Schloss Dagstuhl — Leibniz-Zentrum fuer Informatik, 2016.

[25] B. K. Natarajan. On learning sets and functions. *Mach. Learn.*, 4(1):67–97, Oct 1989.

[26] Alain Pajor. *Sous-espaces l_1^n des espaces de Banach*, volume 16 of *Travaux en Cours [Works in Progress]*. Hermann, Paris, 1985. With an introduction by Gilles Pisier.

[27] Norbert Sauer. On the density of families of sets. *J. Combinatorial Theory Ser. A*, 13:145–147, 1972.

[28] Saharon Shelah. A combinatorial problem; stability and order for models and theories in infinitary languages. *Pacific J. Math.*, 41:247–261, 1972.

[29] Roman Smolensky. Well-known bound for the vc-dimension made easy. *Computational Complexity*, 6(4):299–300, 1997.

[30] J. Michael Steele. Existence of submatrices with all possible columns. *J. Combin. Theory Ser. A*, 24(1):84–88, 1978.

[31] V. N. Vapnik and A. Ja. Červonenkis. The uniform convergence of frequencies of the appearance of events to their probabilities. *Dokl. Akad. Nauk SSSR*, 181:781–783, 1968.

[32] V. N. Vapnik and A. Ya. Chervonenkis. The uniform convergence of frequencies of the appearance of events to their probabilities. *Theor. Probability Appl.*, 16:264280, 1971.

[33] Roman Vershynin. Integer cells in convex sets. *Adv. Math.*, 197(1):248–273, 2005.