The VC-dimension of Unions: Learning, Geometry and Combinatorics

Mónika Csikós* Andrey Kupavskii† Nabil H. Mustafa‡

Abstract

The VC-dimension of a set system is a way to capture its complexity, and has been a key parameter studied extensively in machine learning and geometry communities. In this paper, we make substantial progress on bounding the VC-dimension of k-fold unions and intersections of basic geometric set systems, including settling an open question in machine learning that was first studied in the 1989 foundational paper of Blumer, Ehrenfeucht, Haussler and Warmuth [4].

*Karlsruhe Institute of Technology, Karlsruhe, Germany.

†Moscow Institute of Physics and Technology and EPFL, Switzerland. Email: kupavskii@yandex.ru. The work of Andrey Kupavskii has been supported in part by the Swiss National Science Foundation Grants 200021-137574 and 200020-14453 and by the grant N 15-01-03530 of the Russian Foundation for Basic Research.

‡Université Paris-Est, LIGM, Equipe A3SI, ESIEE Paris, France. Email: mustafan@esiee.fr. The work of Nabil H. Mustafa has been supported by the grant ANR SAGA (JCJC-14-CE25-0016-01).

1 Introduction

Let $(X, \mathcal{R})$ be a set system, where X is a set of elements and $\mathcal{R}$ is a set of subsets of X. In the theory of learning, each set of $\mathcal{R}$ is also called a *concept*, and $\mathcal{R}$ a *concept class* on X. For any integer $k \geq 2$, define the *k-fold union* of $\mathcal{R}$ to be the set system

$$\mathcal{R}^{k\cup} = \{R_1 \cup \cdots \cup R_k \colon R_1, \ldots, R_k \in \mathcal{R}\}.$$

Similarly, one can define the *k-fold intersection* of $\mathcal{R}$ as the set system $\mathcal{R}^{k\cap}$ consisting of all subsets derived from the common intersection of at most k subsets of $\mathcal{R}$. Note that as the subsets $R_1, \ldots, R_k$ need not necessarily be distinct, we have $\mathcal{R} \subseteq \mathcal{R}^{k\cup}$ and $\mathcal{R} \subseteq \mathcal{R}^{k\cap}$.

VC-dimension. One of the fundamental measures for capturing the 'complexity' or 'richness' of a set system, with applications across several fields, is the *Vapnik-Chervonenkis dimension*, or in short the *VC-dimension*, of a set system. Given $(X, \mathcal{R})$, for any set $Y \subseteq X$, define the *projection* of $\mathcal{R}$ onto Y as the set system:

$$\mathcal{R}|_Y = \{Y \cap R \colon R \in \mathcal{R}\}.$$

We say that $\mathcal{R}$ *shatters* Y if $|\mathcal{R}|_Y| = 2^{|Y|}$; in other words, all subsets of Y can be derived by intersection with sets of $\mathcal{R}$. Then the VC-dimension of $\mathcal{R}$, denoted VC-dim$(\mathcal{R})$, is the size of the largest subset of X that can be shattered by $\mathcal{R}$.

Originally defined in statistics and probability, it has turned out to be a key parameter in several areas; this paper concerns three of them—learning theory, geometry and combinatorics.

Learning theory. In learning theory, the VC-dimension of a concept class measures the difficulty of learning a concept of the class. The foundational paper of Blumer, Ehrenfeucht, Haussler and Warmuth [4] states that "the essential condition for distribution-free learnability is finiteness of the Vapnik-Chervonenkis dimension". One of the theorems they prove is the following.

Theorem A (Blumer *et al.* [4])**.** *Let $(X, \mathcal{R})$ be a set system, and k be any positive integer. Then*

$$\Omega\left(\text{VC-dim}\left(\mathcal{R}\right) \cdot k\right) = \text{VC-dim}\left(\mathcal{R}^{k\cup}\right) = O\Big(\text{VC-dim}\left(\mathcal{R}\right) \cdot k \log k\Big).$$

$$\Omega\left(\text{VC-dim}\left(\mathcal{R}\right) \cdot k\right) = \text{VC-dim}\left(\mathcal{R}^{k\cap}\right) = O\Big(\text{VC-dim}\left(\mathcal{R}\right) \cdot k \log k\Big).$$

They also considered the question of whether the bound of Theorem A is tight in the most basic geometric case when $X \subseteq \mathbb{R}^d$ is a set of points and $\mathcal{R}$ is the projection of the family of all half-spaces of $\mathbb{R}^d$ onto X. Here they proved that the VC dimension of the k-fold union of half-spaces in two dimensions is exactly $2k + 1$. For general dimensions $d \geq 4$, they bound the VC-dimension of the k-fold union of half-spaces by $O(d \cdot k \log k)$, following from Theorem A together with the fact that the VC-dimension of the set system induced by half-spaces in $\mathbb{R}^d$ is $d + 1$. The same bound holds for the k-fold intersection of half-spaces in $\mathbb{R}^d$.

Eisenstat and Angluin [8] proved, by giving a probabilistic construction of an abstract set system, that the upper bound of Theorem A is asymptotically tight if VC-dim$(\mathcal{R}) \geq 5$ and that for VC-dim$(\mathcal{R}) = 1$, a lower bound of k holds and that it is tight. A few years later, Eisenstat [7] filled the gap by showing that $\text{VC-dim}\left(\mathcal{R}^{k\cup}\right) = \Omega\left(\text{VC-dim}\left(\mathcal{R}\right) \cdot k \log k\right)$ even if $\text{VC-dim}\left(\mathcal{R}\right) \geq 2$. Later Dobkin and Gunopulos [6] showed that the VC-dimension of the k-fold union of half-spaces in three dimensions is upper-bounded by $4k$.

For $d \geq 4$, the current best upper-bound for the k-fold union and the k-fold intersection of half-spaces in $\mathbb{R}^d$ is still the one given by Theorem A almost 30 years ago, while the lower-bound has remained $\Omega\left(\text{VC-dim}(\mathcal{R}) \cdot k\right)$. We refer the reader to the PhD thesis [11] for a summary of the bounds on VC-dimensions of these basic combinatorial and geometric set systems. The resolution of the VC-dimension of k-fold unions and intersections of half-spaces is left as one of the main open problems in the thesis.

In fact, while the upper-bound of Theorem A applies to geometric set systems, we did not have, till now, a single example of a geometric set system $\mathcal{R}$ with a non-linear lower-bound—i.e., beyond $\Omega\left(\text{VC-dim}\left(\mathcal{R}\right) \cdot k\right)$—on the VC-dimension of its k-fold union or k-fold intersection.

Our first result proves a non-linear bound for a geometric set system even in the plane:

Theorem 1 (Section 3)**.** *Let k be a given positive integer. Then there exists a set P of points in $\mathbb{R}^2$ such that the set system $\mathcal{L}$ induced on P by lines in the plane satisfies*

$$\text{VC-dim}\left(\mathcal{L}^{k\cup}\right) = \Omega\left(\text{VC-dim}(\mathcal{L}) \cdot k \cdot \left(\frac{\log k}{\log\log k}\right)^{\frac{1}{3}}\right) = \Omega\left(k \cdot \left(\frac{\log k}{\log\log k}\right)^{\frac{1}{3}}\right).$$

Next we resolve the question for half-spaces in $\mathbb{R}^d$, $d \geq 4$.

Theorem 2 (Section 4)**.** *Let k be a given positive integer, and $d \geq 4$ an integer. Then there exists a set P of points in $\mathbb{R}^d$ such that the set system $\mathcal{R}$ induced on P by half-spaces satisfies*

$$\text{VC-dim}\left(\mathcal{R}^{k\cup}\right) = \Omega\Big(\text{VC-dim}(\mathcal{R}) \cdot k \log k\Big) = \Omega\Big(d \cdot k \log k\Big).$$

$$\text{VC-dim}\left(\mathcal{R}^{k\cap}\right) = \Omega\Big(\text{VC-dim}(\mathcal{R}) \cdot k \log k\Big) = \Omega\Big(d \cdot k \log k\Big).$$

Relations to Geometry. The breakthrough work of Haussler and Wezl [10] showed that the size of a key structure in discrete and computational geometry, ϵ-nets (we refer the reader to the books [5, Chapter 4], [13, Chapter 10], [16, Chapter 15] for detailed information), is directly linked with the VC-dimension of the corresponding set system:

Theorem B (Epsilon-net Theorem [10])**.** *Let $(X, \mathcal{R})$ be a set system and $\epsilon > 0$ be a given parameter. Then there exists an ϵ-net for $(X, \mathcal{R})$ of size at most*

$$O\left(\frac{\text{VC-dim}(\mathcal{R})}{\epsilon} \log \frac{1}{\epsilon}\right).$$

Thus the sizes of ϵ-nets can be upper-bounded in terms of the VC-dimension of the set system. On the other hand, Alon [1] showed a super-linear lower-bound for ϵ-nets induced by lines in the plane; in a very recent breakthrough, Balogh and Solymosi [3] improved Alon's lower bound to get the following:

Theorem C ([3])**.** *Given any $\epsilon > 0$, there exists a set P of points in the plane such that any ϵ-net for P for the set system induced on P by lines must have size at least*

$$\frac{1}{2\epsilon}\frac{\left(\log\frac{1}{\epsilon}\right)^{\frac{1}{3}}}{\log\log\frac{1}{\epsilon}}.$$

We refer the reader to the chapter [14] for further details on recent progress in the area of ϵ-nets.

2 Our Results

Our results exploit links between learning theory, geometry and combinatorics: the open problem on the VC-dimension of the k-fold union of half-spaces in $\mathbb{R}^d$ in learning theory is resolved by noting its connection to ϵ-nets in geometry (items **1.** and **2.** below), and the recent breakthrough of Balogh and Solymosi on ϵ-nets can be generalized by putting it in the framework of VC-dimension of k-fold unions (item **3.** below). More precisely:

1. The starting point of our work is the following observation that links the study of k-fold unions in learning theory literature to the one for ϵ-nets in discrete geometry. While considerable work has been done in the precise relation between VC-dimension of set systems and *upper-bounds* on the sizes of ϵ-nets, it turns out the connection goes both ways:

Theorem 3 (Section 5)**.** *Let $(X, \mathcal{R})$ be a set system and $\epsilon > 0$ be a given parameter. Then any ϵ-net for $(X, \mathcal{R})$ must have size at least*

$$\Omega\left(\frac{\text{VC-dim}(\mathcal{R})}{\epsilon}\cdot f\left(\frac{1}{2\epsilon}\right)\right),$$

where $f : \mathbb{R} \to \mathbb{R}$ is any function satisfying $\text{VC-DIM}\left(\mathcal{R}^{k\cup}\right) \geq \text{VC-dim}(\mathcal{R}) \cdot k \cdot f(k)$ *for* $k = \frac{1}{2\epsilon}$.

This can be seen as complement to Theorem B: the VC-dimension of the k-fold union gives a *lower bound* for the sizes of ϵ-nets for $\mathcal{R}$.

2. We show an optimal lower-bound on the VC-dimension of the k-fold union and the k-fold intersection of half-spaces in $\mathbb{R}^d$ follows via [12], thus settling affirmatively one of the main open question studied by Eisenstat and Angluin [8], Johnson [11], and Eisenstat [7], and matching the $O(d \cdot k \log k)$ upper bound of Theorem A.

Theorem 2 (Section 4)**.** *Let k be a given positive integer, and $d \geq 4$ an integer. Then there exists a set P of points in $\mathbb{R}^d$ such that the set system $\mathcal{R}$ induced on P by half-spaces satisfies*

$$\text{VC-dim}\left(\mathcal{R}^{k\cup}\right) = \Omega\Big(\text{VC-dim}(\mathcal{R}) \cdot k\log k\Big) = \Omega\Big(d \cdot k\log k\Big).$$

$$\text{VC-dim}\left(\mathcal{R}^{k\cap}\right) = \Omega\Big(\text{VC-dim}(\mathcal{R}) \cdot k\log k\Big) = \Omega\Big(d \cdot k\log k\Big).$$

Remark 1. This statement also provides a non-probabilistic proof of Eisenstat's lower-bound [7].
Remark 2. Observe that if $\overline{\mathcal{R}} := \{\mathbb{R}^d \setminus R \mid R \in \mathcal{R}\}$, then VC-dim$(\overline{\mathcal{R}})$ = VC-dim$(\mathcal{R})$ and so

$$\mathcal{R}^{k\cap} = \overline{\overline{\mathcal{R}}^{k\cup}}.$$

holds by the De Morgan laws. Since for half-spaces $\overline{\mathcal{R}} = \mathcal{R}$, the first claim of Theorem 2 implies the second one, i.e., the same lower-bound for $\mathcal{R}^{k\cap}$, settling another question posed by Eisenstat and Angluin [8].
Remark 3. This implies that the coreset constructions in [9] require an additional $\log k$ factor in the coreset size coming from the VC-dimension of the k-fold intersection of half-spaces. See [2] for details.

3. We next turn to the set system induced on a set of points in the plane by lines. We show that the powerful technique of Balogh and Solymosi [3] can be improved and extended to prove a lower-bound on the VC-dimension of the k-fold union of this set system induced by lines in the plane:

Theorem 1 (Section 3)**.** *Let k be a given positive integer. Then there exists a set P of points in $\mathbb{R}^2$ such that the set system $\mathcal{L}$ induced on P by lines in the plane satisfies*

$$\text{VC-dim}\left(\mathcal{L}^{k\cup}\right) = \Omega\left(\text{VC-dim}(\mathcal{L}) \cdot k \cdot \left(\frac{\log k}{\log\log k}\right)^{\frac{1}{3}}\right) = \Omega\left(k \cdot \left(\frac{\log k}{\log\log k}\right)^{\frac{1}{3}}\right).$$

As an immediate corollary (using Theorem 3), we get an improvement to Theorem C:

Corollary 4. *Given any $\epsilon > 0$, there exists a set P of points in the plane such that any ϵ-net for P for the set system induced on P by lines must have size*

$$\Omega\left(\frac{1}{\epsilon}\left(\frac{\log\frac{1}{\epsilon}}{\log\log\frac{1}{\epsilon}}\right)^{\frac{1}{3}}\right).$$

3 Proof of Theorem 1.

We first present the proof outline: we start by taking an appropriate point set P in $\mathbb{R}^r$ (for a parameter r to be chosen later), with the property that P and all its sufficiently dense subsets contain many r-tuples of collinear points. We will then pick each point of P independently with probability p (a parameter to be chosen later) to get the required point-set R. The key idea is the use of the powerful 'container method' to analyze the failure probability, which in our case will be that R contains a subset of size m (a parameter to be chosen later) *without* any r collinear points. A random projection onto the plane then gives the required point set.

Our proof closely follows the one of Balogh and Solymosi, with two key differences:

(i) one can choose the construction and parameters more carefully to improve the bound by poly $\log\log$ factors. The initial set in [3] is simply the grid $[n]^r$; we will pick a more subtle construction with finer properties which allow us to get an improved bound, and

(ii) the use of containers is flexible and powerful enough to allow one to deduce the stronger statement on lower-bound for the VC dimension of the k-fold union. The value of 'm' above in [3] is set to $\frac{|R|}{2}$, following the initial work of Alon [1]—stated as the 'Alon idea' in [3]. This suffices to get a lower-bound on ϵ-nets, but not for lower-bounding the VC-dimension of the k-fold union of the set system on R induced by lines. Towards this end, we need to prove a stronger property with a lower value of m.

Now we turn to the proof of the theorem.

3.1 Initial set P: pointset with many collinear tuples

Denote by $[a, b]$ the set of all integers between a and b. In what follows, all asymptotic notation will be with respect to the parameter n; for brevity, we will denote $f(n) = g(n) \pm o(g(n))$ by $f(n) \sim g(n)$, and $f(n) \leq g(n) + o(g(n))$ by $f(n) \lesssim g(n)$.

Let $n, r \in \mathbb{N}$ and $t \in \mathbb{R}$ be parameters to be fixed later, such that $1 \ll t = o(n)$. Our initial starting point-set in $\mathbb{R}^r$ will be:

$$P := \Big\{(x_1, \ldots, x_r) \in [-n, n]^r \colon |x_i| \leq n - |x_1| \text{ for } i = 2, \ldots, r\Big\}.$$

Now consider the family $\mathcal{L}_t$ of all lines that pass through at least one point of P, and have the direction vector in the following set L_t:

$$L_t := \Big\{(x_1, \ldots, x_r) \colon \frac{n}{t} \leq x_1 \leq \frac{2n}{t},\ |x_i| \leq \frac{n}{t} \text{ for } i = 2, \ldots, r, \text{ and } x_1, \ldots, x_r \text{ are relatively prime}\Big\}.$$

Note that

$$\text{if } (x_1, \ldots, x_r) \text{ are relatively prime integers, then } (\alpha x_1, \ldots, \alpha x_r) \in \mathbb{Z}^r \text{ if and only if } \alpha \in \mathbb{Z}. \quad (1)$$

We are going to work with the set of points P and the subsets of P induced by lines in $\mathcal{L}_t$.

The key properties of P and $\mathcal{L}_t$ are summarized in the following claim.

Claim 4.1. *Let P and $\mathcal{L}_t$ be as defined above. Then*

(i) $|P| \sim \frac{(2n)^r}{r}$,

(ii) $|L_t| \sim \frac{(2n)^r}{2t^r}$,

(iii) each point of P lies on $|L_t|$ lines of $\mathcal{L}_t$,

(iv) each line of $\mathcal{L}_t$ contains at most $2t$ points from P, and

(v) $\frac{(2n)^r}{2t}|L_t| \lesssim |\mathcal{L}_t| \lesssim \frac{2(2n)^r}{t}|L_t|$.

Proof.

(i)

$$|P| = (2n+1)^{r-1} + 2\sum_{i=1}^{n} \left(2\left(n-i\right)+1\right)^{r-1} \sim \frac{(2n)^r}{r}.$$

(ii) It follows from Nymann [15] that the number of k-tuples of integers from $[m]^x$ that are relatively prime is asymptotically $\frac{m^x}{\zeta(x)}$, where $\zeta(x) = \sum_{i=1}^{\infty} i^{-x}$ is the Riemann zeta function. It is easy to see that $\zeta(x) \to 1$ as $x \to \infty$. Therefore, even looking at $x_2, \ldots, x_r \in \left[-\frac{n}{t}, \frac{n}{t}\right]$, we conclude that the number of such $(r-1)$-tuples that are relatively prime is asymptotically $(\frac{2n}{t})^{r-1}$. Therefore, almost all r-tuples in $\left[\frac{n}{t}, \frac{2n}{t}\right] \times \left[-\frac{n}{t}, \frac{n}{t}\right]^{r-1}$ are relatively prime, and the number of r-tuples in L is asymptotically $\frac{(2n)^r}{2t^r}$.

(iii) Follows from the definition of $\mathcal{L}_t$.

(iv) Property (1) implies that if $p \in P$ lies on a line with direction vector $v \in L_t$ with coefficient α', then α' must be an integer. The claim follows as the first coordinate of p lies in the interval $[-n, n]$ while the absolute value of the first coordinate of the direction vector v is at least $\frac{n}{t}$.

(v) Note that every line in $\mathcal{L}_t$ passes through one of the points of the following set:

$$U_t := \left\{(x_1, \ldots, x_r)\colon |x_1| \leq \frac{2n}{t},\ x_i \in [-n, n]^r \text{ for } i = 2, \ldots, r\right\}.$$

Therefore we can upper-bound the size of $\mathcal{L}_t$ by the number of incidences with U_t, i.e., $|\mathcal{L}_t| \leq |L_t||U_t| \sim 2\frac{(2n)^r}{t} \cdot |L_t|$. On the other hand, each line from $\mathcal{L}_t$ passes through at most four points of U_t for the same reasons as the ones used to prove (iv). So each line is counted at most four times in the number of incidences with the set U_t, which gives the lower bound of $|\mathcal{L}_t| \geq \frac{|L_t||U_t|}{4}$.

□

Lemma 5. *Let $\gamma \geq \frac{8r^2}{t}$ be a given positive parameter. Then every subset of P of size $\gamma|P|$ contains*

$$\gtrsim \frac{\gamma^r n^{2r}}{t \cdot r^{2r}} \qquad \textit{collinear } r\textit{-tuples lying on lines from } \mathcal{L}_t.$$

Proof. Let S be any subset of P with $|S| = \gamma|P|$. Then Claim 4.1(iii) implies that the number of point-line incidences between S and $\mathcal{L}_t$ is $|S||L_t| = \gamma|P||L_t|$. Thus the average number of points of S lying on a line of $\mathcal{L}_t$, using Claim 4.1 (v), is

$$\frac{\gamma|P||L_t|}{|\mathcal{L}_t|} \geq \frac{\gamma|P||L_t|}{\frac{2(2n)^r}{t}|L_t|} \sim \frac{\gamma\frac{(2n)^r}{r}}{\frac{2(2n)^r}{t}} = \frac{\gamma t}{2r}.$$

Note that by the choice of γ, we have $\frac{\gamma t}{2r} \geq 4r$, and thus $\left(\frac{\gamma t}{2r} - r\right) \geq \frac{\gamma t}{4r}$. By the convexity of $\binom{x}{k}$ and Claim 4.1(v), we can conclude that the number of collinear r-tuples in S is at least

$$\binom{\frac{\gamma t}{2r}}{r}|\mathcal{L}_t| \gtrsim \frac{\left(\frac{\gamma t}{2r} - r\right)^r}{r!} \cdot \frac{(2n)^r}{2t}|L_t| \gtrsim \frac{\left(\frac{\gamma t}{4r}\right)^r}{r!} \cdot \frac{(2n)^r}{2t}|L_t| \gtrsim \frac{\left(\frac{\gamma t}{4r}\right)^r}{r!} \cdot \frac{(2n)^r}{2t} \cdot \frac{(2n)^r}{2t^r} = \frac{\gamma^r n^{2r}}{tr^r} \cdot \frac{1}{4r!} \gtrsim \frac{\gamma^r n^{2r}}{t \cdot r^{2r}},$$

where the last inequality used the bound $4r! \lesssim (r/2)^r$. □

3.2 Setting up containers

In this section we prepare notations for, and state the main technical tool—the container theorem for hypergraphs. Roughly speaking, container theorems state that all independent sets in a hypergraph are contained in a relatively small number of subsets, each of which is relatively small. Then, after taking a random subset R of vertices of the hypergraph, this type of statement helps us to bound the size of the largest independent set in R by simply applying union bound over all subsets of each of the containers.

Unfortunately, container statements are rather technical. Let us first introduce some notation. Let $\mathcal{H}$ be a r-uniform hypergraph on vertex set $V(\mathcal{H})$ and let d be the average degree of a vertex of $\mathcal{H}$. The degree $d(S)$ of a subset $S \subset V(\mathcal{H})$ is simply the number of edges of $\mathcal{H}$ containing S. For every $j \in [r]$, let us denote by Δ_j the maximum degree of a j-element subset, that is,

$$\Delta_j := \max \{d(S)\colon S \subset V(\mathcal{H}), |S| = j\}.$$

For any $S \subseteq V(\mathcal{H})$, let $\mathcal{H}[S]$ denote the hypergraph induced by S, and by $e(\mathcal{H}[S])$ its number of hyperedges. For any $\tau \in (0, 1)$, define

$$\Delta(\mathcal{H}, \tau) := \sum_{j=2}^{r} \frac{\Delta_j}{d \cdot \tau^{j-1} \cdot 2^{\binom{j-1}{2}+1-\binom{r}{2}}}.$$

The role of $\Delta(\mathcal{H}, \tau)$, roughly speaking, is to measure the "goodness" of $\mathcal{H}$: smaller values of $\Delta(\mathcal{H}, \tau)$ imply better bounds. Note that, from this perspective, we are interested in lower-bounding d and upper-bounding Δ_j.

We will use the following variant of a container statement:

Theorem D ([18]). *Let $\mathcal{H}$ be a r-uniform hypergraph on the vertex set $\{1, \ldots, N\}$. Let $0 < \varepsilon, \tau < 1/2$. Suppose that $\tau < 1/(200 \cdot r \cdot r!^2)$ and $\Delta(\mathcal{H}, \tau) \leq \varepsilon/(12r!)$. Then there exists $c = c(r) \leq 1000 \cdot r \cdot r!^3$ and a collection of vertex subsets $\mathcal{C}$ such that*

1) every independent set in $\mathcal{H}$ is a subset of some $A \in \mathcal{C}$,

2) for every $A \in \mathcal{C}$, $e(\mathcal{H}[A]) \leq \varepsilon \cdot e(\mathcal{H})$, and

3) $\log |\mathcal{C}| \leq cN\tau \cdot \log\left(\frac{1}{\varepsilon}\right) \cdot \log\left(\frac{1}{\tau}\right)$.

We will work with the r-uniform hypergraph $\mathcal{H}$, where $V(\mathcal{H}) := P$, and edges are all r-tuples of collinear points on lines from $\mathcal{L}_t$. Note that $\mathcal{L}_t$ depends on a parameter t, which we set to be $t = \frac{r^5}{2}$.

Next we calculate the parameters of $\mathcal{H}$ that are needed to apply Theorem D. Applying Lemma 5 with $\gamma = 1$ (note here that $\frac{8r^2}{t} = \frac{16}{r^3} < 1$), the total number of collinear r-tuples in P is lower-bounded by $\frac{n^{2r}}{tr^{2r}}$, and so the average degree d of $\mathcal{H}$ can be lower-bounded by

$$d \geq \frac{\frac{n^{2r}}{tr^{2r}} \cdot r}{|P|} = \frac{\frac{2n^{2r}}{r^5 r^{2r}} \cdot r}{\frac{(2n)^r}{r}} = \frac{n^r}{r^3 r^{2r} 2^{r-1}} \geq \frac{n^r}{r^{3r+2}}.$$

On the other hand, since each line contains at most $2t = r^5$ points, we get that

$$\Delta_j \leq \binom{r^5 - j}{r - j} \leq r^{5(r-j)} \quad \text{for } j = 2, \ldots, r.$$

Set $\tau = n^{-1-\frac{1}{2r}}$ and r be such that $r \leq \frac{1}{20}\frac{\log n}{\log\log n}$. Then note the following inequalities:

$$n^r \tau^{j-1} = \frac{n^r}{n^{j-1+\frac{j}{2r}-\frac{1}{2r}}} = n^{r-j}\cdot n^{1+\frac{1}{2r}-\frac{j}{2r}} > n^{r-j}\cdot n^{\frac{1}{2}} \quad \text{for all } j = 1,\ldots,r.$$

$$r^{r-j} \leq 2^{r^2-j^2} \leq 2^{2r(r-j)} = n^{o(r-j)} \quad \text{for all } j = 2,\ldots,r-1.$$

$$r^r \leq (\log n)^{\frac{\log n}{20\log\log n}} = n^{\frac{1}{20}}.$$

Therefore, we may conclude that for this choice of τ and r, we have

$$\begin{aligned}\Delta(\mathcal{H},\tau) &\leq \sum_{j=2}^{r} \frac{\Delta_j}{d\cdot\tau^{j-1}\cdot 2^{\binom{j-1}{2}+1-\binom{r}{2}}} \leq \sum_{j=2}^{r}\frac{r^{5(r-j)+3r+2}}{n^r\cdot\tau^{j-1}\cdot 2^{j^2-r^2}} \leq \sum_{j=2}^{r}\frac{r^{5(r-j)+3r+2}\cdot 2^{r^2-j^2}}{n^r\cdot\tau^{j-1}}\\ &\leq \sum_{j=2}^{r}\frac{n^{o(r-j)+\frac{3}{20}+o(1)}}{n^{r-j}\cdot n^{\frac{1}{2}}} \ll n^{-\frac{1}{3}}.\end{aligned}$$

We apply Theorem D with our choice of τ and r, and with $\epsilon = n^{-\frac{1}{4}}$. It is easy to see that the conditions of the theorem are satisfied. Indeed,

$$\tau \leq n^{-1} \ll n^{-\frac{1}{6}+o(1)} \ll r^{-2r} \ll 1/(200\cdot r\cdot r!^2),$$

as well as that

$$\Delta(\mathcal{H},\tau) \ll n^{-\frac{1}{3}} = n^{-\frac{1}{4}-\frac{1}{12}} \ll \varepsilon r^{-r} \ll \frac{\varepsilon}{12r!}.$$

Note that the constant $c(r)$ from the theorem satisfies $c(r) \ll r^{3r}$. The conclusion of the theorem gives a collection of subsets $\mathcal{C} \subset 2^P$, such that for every $A \in \mathcal{C}$ we have

$$e(\mathcal{H}[A]) \leq n^{-\frac{1}{4}} e(\mathcal{H}) \ll n^{2r-\frac{1}{4}}, \tag{2}$$

where the second inequality is due to the fact that $e(\mathcal{H}) \ll n^{2r}$, and such that

$$\log \mathcal{C} \ll r^{3r} n^r n^{-1-\frac{1}{2r}} \log^2 n \ll r^{3r} n^{r-1-\frac{1}{3r}}. \tag{3}$$

Here the last inequality is due to the fact that $\log^2 n = n^{\frac{2\log\log n}{\log n}} \leq n^{\frac{1}{10r}}$.

Using Lemma 5 with $\gamma = r^{-2}$ (which is a valid choice of γ, since $\gamma t = r^3 \gg r^2$), we conclude that in any subset of size at least $\gamma|P|$ there are at least $n^{2r}/r^{4r+o(r)} \geq n^{2r-\frac{1}{5}+o(1)}$ induced hyperedges. Therefore, any subset $A \in \mathcal{C}$ has size at most $\gamma|P|$ by inequality (2).

3.3 Constructing R

In this section we use the properties of the hypergraph $\mathcal{H}$, developed in Section 3.2, to obtain a point set for which the VC-dimension of the k-fold union of the set system induced by lines is bounded from below.

We would like to use the point set P with lines from $\mathcal{L}_t$ directly, but the problem is that it contains many collinear $(r+1)$-tuples. From Claim 4.1 (iv), the number of collinear $(r+1)$-tuples in the set P with lines from $\mathcal{L}_t$ is at most

$$\binom{2t}{r+1}|\mathcal{L}_t| \leq \frac{(2t)^{r+1}}{(r+1)!} \cdot \frac{2(2n)^r}{t} \cdot \frac{(2n)^r}{2t^r} = \frac{(2t)^{r+1}(2n)^{2r}}{(r+1)!t^{r+1}} = \frac{2^{r+1}2^{2r}}{(r+1)!}n^{2r} \ll \left(\frac{24}{r}\right)^r n^{2r}.$$

Let R' be a random subset constructed by picking each point of P with probability p, a parameter we will set shortly. We want the number of points in R' to be much bigger than the number of collinear $(r+1)$-tuples present in R'. In other words, we set p such that

$$\mathbb{E}\left[|R'|\right] = p|P| \sim p\frac{(2n)^r}{r} \gg p^{r+1}\left(\frac{24}{r}\right)^r n^{2r}. \tag{4}$$

Using Markov's inequality, it is easy to see that this inequality is satisfied, with high probability, for $p = \frac{r}{20n}$.

Set $m = \frac{p|P|}{r^{3/2}}$. Now we argue that, with high probability, R' has the property that any subset of R' of size m is *not* an independent set; or stated another way, any subset of R' of size m contains r collinear points. Using containers and the fact that $\gamma = r^{-2}$, we conclude via the union bound that the probability that there exists an independent set of size m in $\mathcal{H}[R']$ is at most

$$|\mathcal{C}| \cdot \binom{\gamma|P|}{m} \cdot p^m \leq 2^{r^{3r}n^{r-1-\frac{1}{3r}}} \left(\frac{\gamma e p|P|}{m}\right)^m = 2^{r^{3r}n^{r-1-\frac{1}{3r}}} \left(\frac{e}{r^{\frac{1}{2}}}\right)^m \leq 2^{r^{3r}n^{r-1-\frac{1}{3r}}-m}.$$

Therefore, w.h.p., R' does not contain any independent sets of size m, provided $r^{3r}n^{r-1-\frac{1}{3r}} - m \to -\infty$. As $|P| \sim \frac{(2n)^r}{r}$, we have $m \sim \frac{(2n)^r}{r} \cdot \frac{r}{20n} \cdot \frac{1}{r^{3/2}} = \frac{(2n)^{r-1}}{10r^{3/2}}$ and thus

$$\frac{m}{r^{3r}n^{r-1-\frac{1}{3r}}} \sim \frac{\frac{(2n)^{r-1}}{10r^{3/2}}}{r^{3r}n^{r-1-\frac{1}{3r}}} = \frac{2^{r-1}}{10r^{\frac{3}{2}}} \cdot \frac{n^{\frac{1}{3r}}}{r^{3r}} \gg \frac{n^{\frac{1}{3r}}}{r^{3r}}, \quad \text{for } r \text{ a large-enough constant.}$$

We want the last expression to tend to infinity, which holds if $n \geq r^{10r^2}$. This, in turn, holds for

$$r = \sqrt{\frac{\log n}{6\log\log n}}. \tag{5}$$

Note that this is the only place in the proof where we actually need r to be of order $\log^{0.5-o(1)} n$, and not $\log^{1-o(1)} n$, and the reason for it is the factor r^{3r} in (3).

Now delete one point in R' from each collinear $(r+1)$-tuple of R', obtaining the set $R \subset R'$. Clearly, R contains no collinear $(r+1)$-tuples, and with high probability, it has size $(1-o(1))\,p|P|$ (by inequality (4)), and no independent sets of size m. Project R on the plane so that no new collinearities appear. This implies that, with r set by (5), there exists a set R whose projection in the plane satisfies these two conditions:

- $\mathcal{H}[R]$ does not contain independent sets of size at least m, and
- R has size $(1-o(1))p|P|$.

Given the integer k, set the value of n so that $k = \frac{2|R|}{r} \sim \frac{2p|P|}{r} \sim n^{r-1+o(1)}$. Then

$$\log r \sim \frac{1}{2} \log\log n \sim \frac{1}{2} \log\log k - \frac{1}{2} \log r,$$

which implies that $\log\log n \sim \frac{2}{3} \log\log k$ and that

$$r^3 = \frac{r \log n}{6 \log\log n} = \frac{\log k}{6 \log\log n} \sim \frac{\log k}{4 \log\log k}.$$

We have that

$$|R| = \frac{kr}{2} \sim \frac{k}{2} \left(\frac{\log k}{4 \log\log k} \right)^{\frac{1}{3}}. \tag{6}$$

3.4 Lower-bound on VC-dimension of the k-fold union induced by lines

It remains to show a lower-bound on the VC-dimension of the k-fold union of the set system induced by lines on R. In other words, given any $S \subseteq R$, we want to show that there exist k lines in the plane such that *i)* S lies in the union of these k lines, and *ii)* no point of $R \setminus S$ lies on any of these lines. This would imply that R is shattered by the k-fold union of lines in the plane, and thus the bound in inequality (6) gives a lower-bound on the size of a shattered set, which then gives the required lower-bound on the VC-dimension of k-fold union of the set system induced by lines in the plane.

Take any subset $S \subset R$. We construct the required set $\mathcal{A}$ of k lines iteratively. Initially set $\mathcal{A} = \emptyset$. If $|S| \geq m$, then as shown above, there is a collinear r-tuple in S. This means that there is a line l_1 that intersects S in r points and does not intersect $R \setminus S$—here we need the fact that there are no $(r+1)$-tuples in R! Add l_1 to $\mathcal{A}$, remove the covered points from S and iterate as long as the remaining set still has size at least m. This can continue for at most $\frac{|R|}{r} = \frac{k}{2}$ steps, after which it must be that the size of the remaining set S is less than m. So far, $|\mathcal{A}| \leq \frac{k}{2}$.

When $|S| < m \sim \frac{|R|}{r^{3/2}} \ll k$, then we add one general position line per point of S to $\mathcal{A}$, which gives an additional $o(k)$ lines. We have thus added at most k lines in total to $\mathcal{A}$, and by construction, these lines cover all the points of S and no point of $R \setminus S$. This holds for any $S \subseteq R$, and thus R is shattered by the set system induced by the k-fold union of lines in the plane. This completes the proof.

4 Proof of Theorem 2.

Proof. The proof will need the following lemma from [12]:

Lemma 6. *Let $n, d \geq 2$ be integers. Then there exists a set $\mathcal{B}$ of $\lfloor \frac{d}{2} \rfloor (n+3)2^{n-2}$ axis-parallel boxes in $\mathbb{R}^d$ such that for any subset $\mathcal{S} \subseteq \mathcal{B}$, one can find a 2^{n-1}-element set Q of points in $\mathbb{R}^d$ with the property that*

(i) $Q \cap B \neq \emptyset$ for any $B \in \mathcal{B} \setminus \mathcal{S}$, and

(ii) $Q \cap B = \emptyset$ for any $B \in \mathcal{S}$.

By a standard lifting transform (e.g., see [17, 12]), given a set $\mathcal{B}$ of boxes in $\mathbb{R}^d$, there exists a function $\pi : \mathcal{B} \rightarrow \mathbb{R}^{2d}$ mapping boxes in $\mathcal{B}$ to points in $\mathbb{R}^{2d}$ such that for any point $p \in \mathbb{R}^d$, there exists a corresponding half-space in $\mathbb{R}^{2d}$, denoted by H_p, such that a box $B \in \mathcal{B}$ contains p if and only if H_p contains the point $f(B) \in \mathbb{R}^{2d}$. For any $\mathcal{B}' \subseteq \mathcal{B}$, set $\pi(\mathcal{B}') = \{f(B) \colon B \in \mathcal{B}'\}$.

Apply Lemma 6 with $n = \lfloor \log k \rfloor + 1$ in $\mathbb{R}^{\lfloor d/2 \rfloor}$ to get a set $\mathcal{B}$ of boxes in $\mathbb{R}^{\lfloor d/2 \rfloor}$. Then $P = \pi(\mathcal{B})$ will be the required point-set in $\mathbb{R}^d$; i.e., we claim that P is shattered by the set system induced by the k-fold union of half-spaces in $\mathbb{R}^d$. To see that, let P' be any subset of P. Set $\mathcal{S} = \pi^{-1}(P \setminus P')$. By Lemma 6, there exists a set Q of $2^{n-1} = 2^{\lfloor \log k \rfloor} \leq k$ points in $\mathbb{R}^{\lfloor d/2 \rfloor}$ such that each box in $\mathcal{S}$ contains no point of Q, and each box in $\mathcal{B} \setminus \mathcal{S}$ contains at least one point of Q. We show that the set $\mathcal{H}(P') = \{H_p \colon p \in Q\}$ of half-spaces separates P' from P. By the property of the lifting map $\pi(\cdot)$,

$$\text{each box in } \mathcal{S} \text{ contains no point of } Q \implies \text{each point in } P \setminus P' \text{ is contained in no half-space of } \mathcal{H}(P'),$$

$$\text{each box in } \mathcal{B} \setminus \mathcal{S} \text{ contains a point of } Q \implies \text{each point in } P' \text{ is contained in some half-space in } \mathcal{H}(P').$$

In other words, the union of the half-spaces in $\mathcal{H}(P')$ contains precisely the set P'. As this is true for any $P' \subseteq P$, the k-fold union of half-spaces in $\mathbb{R}^d$ shatters P. Finally, we have

$$|P| = |\mathcal{B}| = \lfloor \frac{d}{2} \rfloor (\lfloor \log k \rfloor + 3) 2^{\lfloor \log k \rfloor - 2} = \Omega(d \cdot k \log k),$$

as desired. □

5 Proof of Theorem 3.

Proof. Set $k = \frac{1}{2\epsilon}$, and let $d = \text{VC-dim}(\mathcal{R})$. As $\text{VC-DIM}(\mathcal{R}^{k\cup}) \geq \frac{d}{2\epsilon} \cdot f(\frac{1}{2\epsilon})$, we can assume that $|X| = \frac{d}{2\epsilon} \cdot f(\frac{1}{2\epsilon})$ and that it is shattered by $\mathcal{R}^{\frac{1}{2\epsilon}\cup}$. We will show that if N is an ϵ-net for $\mathcal{R}$, then

$$|N| \geq \frac{|X|}{2} = \frac{d}{4\epsilon} \cdot f\left(\frac{1}{2\epsilon}\right).$$

Suppose that $N < \frac{|X|}{2}$. Since X is shattered by $\mathcal{R}^{\frac{1}{2\epsilon}\cup}$, there exists a set $S' \in \mathcal{R}^{\frac{1}{2\epsilon}\cup}$ containing precisely the elements in $X \setminus N$. In other words, we can find $\frac{1}{2\epsilon}$ sets $S_1, \ldots, S_{\frac{1}{2\epsilon}} \in \mathcal{R}$ such that

$$\left(S_1 \cup \cdots \cup S_{\frac{1}{2\epsilon}}\right) \cap X = X \backslash N.$$

Each set S_i contains no point of N and by the pigeonhole principle, one of them must have size at least

$$\frac{|X \setminus N|}{\frac{1}{2\epsilon}} \geq \frac{|X|/2}{\frac{1}{2\epsilon}} = \epsilon|X|,$$

and which is not hit by N. This contradicts the fact that N was an ϵ-net. □

References

[1] Noga Alon. A non-linear lower bound for planar epsilon-nets. *Discrete & Computational Geometry*, 47(2):235–244, 2012.

[2] Olivier Bachem, Mario Lucic, and Andreas Krause. Scalable and distributed clustering via lightweight coresets. *CoRR*, abs/1702.08248, 2017.

[3] J. Balogh and J. Solymosi. On the number of points in general position in the plane. *ArXiv e-prints*, April 2017.

[4] A. Blumer, A. Ehrenfeucht, D. Haussler, and M. K. Warmuth. Learnability and the Vapnik-Chervonenkis dimension. *J. ACM*, 36(4):929–965, October 1989.

[5] B. Chazelle. *The Discrepancy Method: Randomness and Complexity.* Cambridge University Press, Cambridge, New York, 2000.

[6] D. P. Dobkin and D. Gunopulos. Concept learning with geometric hypotheses. In *Proceedings of the Eighth Annual Conference on Computational Learning Theory*, COLT '95, pages 329–336, New York, NY, USA, 1995. ACM.

[7] D. Eisenstat. k-fold unions of low-dimensional concept classes. *Information Processing Letters*, 109(23-24):1232–1234, 2009.

[8] D. Eisenstat and D. Angluin. The VC dimension of k-fold union. *Information Processing Letters*, 101(5):181 – 184, 2007.

[9] Dan Feldman and Michael Langberg. A unified framework for approximating and clustering data. In *Proceedings of the 43rd ACM Symposium on Theory of Computing, STOC 2011, San Jose, CA, USA, 6-8 June 2011*, pages 569–578, 2011.

[10] D. Haussler and E. Welzl. ε-nets and simplex range queries. *Discrete & Computational Geometry*, 2:127–151, 1987.

[11] H. Johnson. *Definable families of finite Vapnik Chernonenkis dimension.* PhD thesis, University of Maryland, 2008. `http://drum.lib.umd.edu/handle/1903/8174`.

[12] A. Kupavskii, N. H. Mustafa, and J. Pach. New lower bounds for epsilon-nets. In *32nd International Symposium on Computational Geometry (SoCG)*, pages 54:1–54:16, 2016.

[13] Jirí Matousek. *Lectures in Discrete Geometry.* Springer-Verlag, New York, NY, 2002.

[14] N. H. Mustafa and K. Varadarajan. Epsilon-approximations and epsilon-nets. In J. E. Goodman, J. O'Rourke, and C. D. Tóth, editors, *Handbook of Discrete and Computational Geometry.* CRC Press LLC, 2017, to appear.

[15] J.E. Nymann. On the probability that k positive integers are relatively prime. *Journal of Number Theory*, 4:469–473, 1972.

[16] J. Pach and P. K. Agarwal. *Combinatorial Geometry.* John Wiley & Sons, New York, NY, 1995.

[17] J. Pach and G. Tardos. Tight lower bounds for the size of epsilon-nets. In *Symposium on Computational Geometry*, pages 458–463, 2011.

[18] D. Saxton and A. Thomason. Hypergraph containers. *Inventiones mathematicae*, pages 1–68, 2015.