Graphs and Combinatorics (2025) 41:47
https://doi.org/10.1007/s00373-025-02909-6

ORIGINAL PAPER

VC-Dimension of Hyperplanes Over Finite Fields

Ruben Ascoli[1] · Livia Betti[2] · Justin Cheigh[3] · Alex Iosevich[2] · Ryan Jeong[4] · Xuyan Liu[5] · Brian McDonald[6] · Wyatt Milgrim[7] · Steven J. Miller[3] · Francisco Romero Acosta[8] · Santiago Velazquez Iannuzzelli[9]

Received: 19 May 2024 / Accepted: 27 February 2025

Abstract

Let $\mathbb{F}_q^d$ be the d-dimensional vector space over the finite field with q elements. For a subset $E \subseteq \mathbb{F}_q^d$ and a fixed nonzero $t \in \mathbb{F}_q$, let $\mathcal{H}_t(E) = \{h_y : y \in E\}$, where $h_y : E \to \{0, 1\}$ is the indicator function of the set $\{x \in E : x \cdot y = t\}$. Two of the authors, with Maxwell Sun, showed in the case $d = 3$ that if $|E| \geq Cq^{\frac{11}{4}}$ and q is sufficiently large, then the VC-dimension of $\mathcal{H}_t(E)$ is 3. In this paper, we generalize the result to arbitrary dimension by showing that the VC-dimension of $\mathcal{H}_t(E)$ is d whenever $E \subseteq \mathbb{F}_q^d$ with $|E| \geq C_d q^{d-\frac{1}{d-1}}$.

Keywords VC-dimension · Point configurations · Learning theory

Mathematics Subject Classification 12E99 · 68Q32

Communicated by Yoshio Okamoto.

✉ Ruben Ascoli
rascoli3@gatech.edu

1 Department of Mathematics, Georgia Institute of Technology, Atlanta, GA, USA

2 Department of Mathematics, University of Rochester, Rochester, NY, USA

3 Department of Mathematics, Williams College, Williamstown, MA, USA

4 Department of Mathematics, University of Cambridge, Cambridge, England

5 Department of Mathematics, Brown University, Providence, RI, USA

6 Department of Mathematics, University of Georgia, Athens, GA, USA

7 Department of Mathematics, University of California Los Angeles, Los Angeles, CA, USA

8 Department of Mathematics, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA

9 Department of Mathematics, Northwestern University, Evanston, IL, USA

Published online: 15 March 2025

1 Introduction

Vapnik and Chervonenkis [23] introduced the VC-dimension in 1971 in the context of learning theory. For an introduction to the subject, see for example [2]. Given a domain X and a collection $\mathcal{H}$ of functions $h : X \to \{0, 1\}$, consider the learning task of trying to identify an unknown element $f \in \mathcal{H}$ by sampling finitely many points $x_1, ..., x_m \in X$ from an unknown probability distribution D, and recording the values $f(x_1), ..., f(x_m)$. One desires an algorithm which takes this input and produces a hypothesis $h \in H$ which with high probability has small error with respect to f. To make this precise, we introduce some definitions.

Definition 1.1 Given a set X, a probability distribution D, and a labeling function $f : X \to \{0, 1\}$, let h be a hypothesis; that is, $h : X \to \{0, 1\}$. Define

$$L_{D,f}(h) = \mathbb{P}_{x\sim D}[h(x) \neq f(x)],$$

where $\mathbb{P}_{x\sim D}$ means that x is being sampled according to the probability distribution D.

Definition 1.2 A hypothesis class $\mathcal{H}$ is PAC (probably approximately correct) learnable if there exists a function

$$m_{\mathcal{H}} : (0, 1)^2 \to \mathbb{N}$$

and a learning algorithm with the following property: For every $\epsilon, \delta \in (0, 1)$, for every distribution D over X, and for every labeling function $f : X \to \{0, 1\}$, if there is some hypothesis $h \in \mathcal{H}$ such that $L_{D,f}(h) = 0$, then when running the learning algorithm on $m \geq m_{\mathcal{H}}(\epsilon, \delta)$ i.i.d. examples generated by D, and labeled by f, the algorithm returns a hypothesis h such that, with probability at least $1 - \delta$ (over the choice of $(x_1, ..., x_m) \sim D^m$),

$$L_{D,f}(h) \leq \epsilon.$$

The VC-dimension characterizes PAC learnability, in light of the fundamental theorem of statistical learning; $\mathcal{H}$ is PAC learnable if and only if the VC-dimension is finite. Moreover, there are quantitative bounds for $m_{\mathcal{H}}(\epsilon, \delta)$ based on $\mathrm{VCdim}(\mathcal{H})$, with smaller VC-dimension allowing smaller effective sample sizes. In order to define the VC-dimension, we must first define shattering.

Definition 1.3 Let X be a set and $\mathcal{H}$ a collection of functions from X to $\{0, 1\}$. We say that $\mathcal{H}$ shatters a finite set $C \subset X$ if the restriction of $\mathcal{H}$ to C yields every possible function from C to $\{0, 1\}$.

Definition 1.4 Let X and $\mathcal{H}$ be as above. We say that a non-negative integer d is the VC-dimension of $\mathcal{H}$ if there exists a set $C \subset X$ of size n that is shattered by $\mathcal{H}$, and no subset of X of size $n + 1$ is shattered by $\mathcal{H}$.

For a subset $E \subseteq \mathbb{F}_q^d$, and a fixed nonzero $t \in \mathbb{F}_q^d$, consider the hypothesis class

$$\mathcal{H}_t(E) := \{h_y : y \in E\},$$

where $h_y : E \to \{0, 1\}$ is defined by $h_y(x) = 1$ if and only if $x \cdot y = t$. Our main theorem establishes the VC-dimension of this hypothesis class in arbitrary dimension $d \geq 3$, for sufficiently large sets $E \subseteq \mathbb{F}_q^d$.

Theorem 1.5 *For $d \geq 3$, if $|E| \geq C_d q^{d-\frac{1}{d-1}}$ for an appropriate constant C_d depending only on d, and q is sufficiently large, then the VC-dimension of $\mathcal{H}_t(E)$ is equal to d.*

Remark 1.6 Theorem 1.5 holds if we replace the dot product with any non-degenerate bilinear form. To check this, note that Theorem 2.1 holds for any non-degenerate bilinear form (as observed in [6]). Moreover, our counting argument based on incidence geometry of hyperplanes will go through in the exact same way.

Iosevich, McDonald, and Sun [15] studied this hypothesis class in the case $d = 3$, and showed that when $|E| \geq Cq^{\frac{11}{4}}$, the VC-dimension of $\mathcal{H}_t(E)$ is 3. Still in the case $d = 3$, the exponent was improved from 11/4 to 5/2 by Pham, Senger, Tait, and Thu-Huyen [21]. Both of these results employed a similar unfolding technique via Cauchy–Schwarz, thereby reducing the argument to the construction of a much simpler graph.

Theorem 1.5, on the other hand, relies on a modified unfolding technique using Hölder's inequality, folding the graph in a different way which is more readily generalized to higher dimensions. In the case $d = 3$, our theorem recovers the exponent 5/2 of [21].

We sketch an outline of the proof of Theorem 1.5 and how it is different from previous papers' techniques, illustrating what we mean by unfolding. The authors of [15] and [21] used the symmetry shown in Fig. 1 to demonstrate the shattering of three points. In particular, the graph on the right side, after adding three leaves, represents the shattering of three points. For $v, z, y \in E$, let $f(v, z, y)$ count the number of choices for $u, x \in E$ so that each edge in the graph on the left hand side represents a pair of points whose dot product is equal to t. Then $f(v, z, y)^2$ counts choices for $u, u', x, x' \in E$ so that each edge in the graph on the right hand side represents a pair of points whose dot product is equal to t. This observation helps to study the VC-dimension by showing the abundance of the graph on the left, which then demonstrates the abundance of the graph on the right via Cauchy–Schwarz applied to f.

In those results, much of the construction is done before the Cauchy–Schwarz unfolding, so that the only other consideration afterward is the addition of leaves, which is achieved by a straightforward pigeonhole argument. In our proof of Theorem 1.5, instead we apply Hölder's inequality at the very beginning, unfolding a single edge into a star as in Fig. 2. The difficulty is that after showing the abundance of such d-stars, it is not immediately clear whether any of them actually corresponds to a shattering of d points. This motivates our definition of so-called "bad" sets in Sect. 2, which help us enumerate stars which fail to represent a shattering of d points in this sense. With this idea, Theorem 1.5 is reduced to showing that most d-stars in $\mathbb{F}_q^d$ do not have any bad subsets of their vertex set.

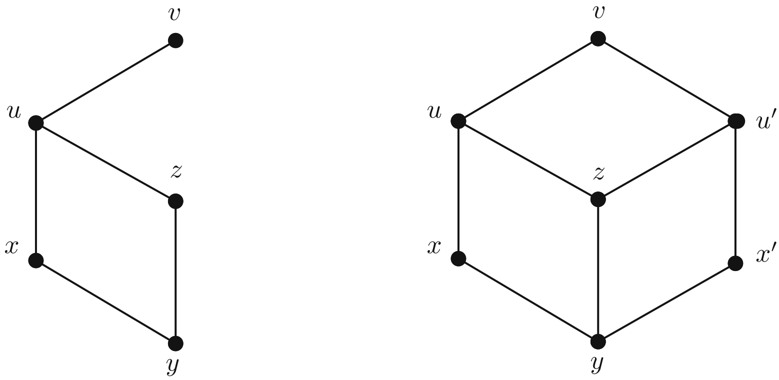

Fig. 1 Cauchy–Schwarz unfolding technique used in [15] and [21]

1.1 Related Work

Similar results have also been obtained in the context of distance problems over finite fields. In this setting, the relevant hypothesis class is $\mathcal{H}_t^{dist}$, defined as follows. For $x \in \mathbb{F}_q^d$, let

$$||x|| = x_1^2 + \cdots + x_d^2.$$

For a subset $E \subseteq \mathbb{F}_q^d$, and a fixed nonzero $t \in \mathbb{F}_q$, let

$$\mathcal{H}_t^{dist}(E) := \{f_y : y \in E\},$$

where $f_y : E \to \{0, 1\}$ is defined by $f_y(x) = 1$ if and only if $||x - y|| = t$. Fitzpatrick, Iosevich, McDonald, and Wyman [10] showed in the case $d = 2$ that if $|E| \geq Cq^{\frac{15}{8}}$, q sufficiently large, then $\text{VCdim}(\mathcal{H}_t^{dist}(E)) = 3$. The exponent $\frac{15}{8}$ was recently improved to $\frac{13}{7}$ by Thang Pham [20], refining the method of [10]. In the case when $E = \mathbb{F}_q^2$ this is trivial, and one may see by induction that in general

$$\text{VCdim}(\mathcal{H}_t^{dist}(\mathbb{F}_q^d)) = d + 1.$$

In the dot product setting, on the other hand, we have

$$\text{VCdim}(\mathcal{H}_t(\mathbb{F}_q^d)) = d.$$

This disparity comes from the fact that a sphere in $\mathbb{F}_q^d$ is determined by $d + 1$ points in general position, whereas a hyperplane in $\mathbb{F}_q^d$ is determined by d points in general position.

In dimensions $d \geq 3$, it is still an open problem whether one can find a threshold $\alpha \in (0, d)$ so that whenever $E \geq C_d q^{\alpha}$ for some constant C_d independent of q,

$$\text{VCdim}(\mathcal{H}_t^{dist}(E)) = d + 1.$$

The strongest partial result in arbitrary dimension is a corollary of the main theorem from a previous result by the authors of this paper [1]. In that paper, we considered a related hypothesis class with two parameters. Let

$$\mathcal{H}_t^*(E) := \{h_{u,v} : u, v \in E\},$$

where $h_{u,v}(x) = 1$ if and only if $||x - u|| = ||x - v|| = t$. In [1], the authors showed that whenever

$$|E| \geq \begin{cases} Cq^{\frac{7}{4}} & d = 2 \\ Cq^{\frac{7}{3}} & d = 3 \\ Cq^{d-\frac{1}{d-1}} & d \geq 4 \end{cases}$$

and q is sufficiently large, the VC-dimension of $\mathcal{H}_t^*(E)$ is equal to d. It follows that with the same restriction on the size of $E \subseteq \mathbb{F}_q^d$, the VC-dimension of $\mathcal{H}_t^{dist}(E)$ is either d or $d + 1$ [1, Section 5].

In contrast to the situation for distances, in this paper we are able to find the VC-dimension exactly, for sufficiently large sets $E \subseteq \mathbb{F}_q^d$. The techniques used are related to those from [1], but new ideas were needed to overcome the difficulty that the property $||x - y|| = t$ is translation invariant, whereas the property $x \cdot y = t$ is not.

The results discussed above can be expressed in terms of graph embeddings $\phi : G \hookrightarrow \mathcal{G}_t(E)$ for appropriate graphs G, where $\mathcal{G}_t(E)$ is the distance (resp. dot product) graph, i.e., the vertices are points in E, with an edge $x \sim y$ whenever $||x - y|| = t$ (resp. $x \cdot y = t$). For relevant results on graph embeddings in the distance and dot product graphs, see for example [3, 6, 10, 12–15, 21].

The difficulty in extending the techniques of this paper to the distance setting is that we would need to solve the same graph embedding problem in a lower dimensional space, and generally these problems are easier in higher dimensions. In particular, the analog of our methods would again only show that $\text{VCdim}(\mathcal{H}_t^{dist}(E)) \geq d$ when E is large enough, leaving open whether $\text{VCdim}(\mathcal{H}_t^{dist}(E)) = d + 1$.

2 Proof of Main Theorem

Consider a large subset $E \subseteq \mathbb{F}_q^d$, and a fixed nonzero $t \in \mathbb{F}_q$. We will use Theorem 2.1 from [6], which counts pairs $(x, y) \in E^2$ with $x \cdot y = t$.

Theorem 2.1 ([6]) *For non-negative functions $f, g : \mathbb{F}_q^d \to \mathbb{R}$,*

$$\sum_{x \cdot y = t} f(x)g(y) = q^{-1}||f||_{L^1}||g||_{L^1} + R(t),$$

where

$$|R(t)| \leq ||f||_{L^2}||g||_{L^2}q^{\frac{d-1}{2}}.$$

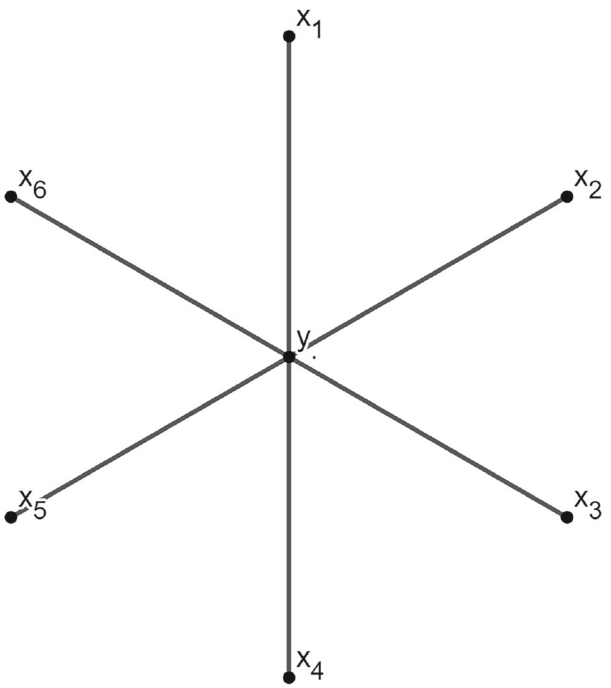

Fig. 2 A 6-star realized as a subgraph of the dot product graph $\mathcal{G}_t(E)$

In particular, when f, g are both chosen to be the indicator function of E, we see that

$$|\{(x, y) \in E^2 : x \cdot y = t\}| = \frac{|E|^2}{q} + O\left(q^{\frac{d-1}{2}}|E|\right),$$

and the error term is much smaller than the main term when $|E| = \omega\left(q^{\frac{d+1}{2}}\right)$. We use this fact, along with Hölder's inequality, to count the number of k-stars in the dot-product graph on E.

Definition 2.2 A (k+1)-tuple $(y, x_1, ..., x_k)$ of points in $\mathbb{F}_q^d$ is a k-star if $y \cdot x_i = t$ for each $i = 1, ..., k$. If all the x_i are distinct, we say $(y, x_1, ..., x_k)$ is a non-degenerate k-star.

Definition 2.3 Let $\mathcal{G}_t(E)$ be the dot product t graph on E, i.e., the graph with vertex set E and an edge $x \sim y$ whenever $x \cdot y = t$.

Lemma 2.4 *Let*

$$N_k(E) := \left|\left\{(y, x_1, ..., x_k) \in E^{k+1} : x_i \text{ distinct}, \ y \cdot x_i = t \ \forall i\right\}\right|$$

be the number of non-degenerate k-stars in $\mathcal{G}_t(E)$. If $|E| \geq C_k q^{\frac{d+1}{2}}$ for an appropriate constant C_k depending only on k, then

$$N_k(E) \geq \frac{|E|^{k+1}}{2q^k}.$$

Proof For $x \in E$, let

$$\psi(x) = \sum_{\substack{y \in E \\ x \cdot y = t}} 1$$

be the number of neighbors of x in $\mathcal{G}_t(E)$. Then

$$N_k(E) := \sum_{x \in E} \psi(x)(\psi(x) - 1) \cdots (\psi(x) - k + 1)$$

$$\geq \sum_{x \in E} \phi(x)^k,$$

where $\phi(x) = \max(\psi(x) - k + 1, 0)$. By Hölder's inequality,

$$\left(\sum_{x \in E} \phi(x)\right)^k \leq \left(\sum_{x \in E} \phi(x)^k\right)\left(\sum_{x \in E} 1\right)^{k-1} \leq |E|^{k-1} N_k(E).$$

To get the desired lower bound for $N_k(E)$, it suffices to bound $\sum_{x \in E} \phi(x)$ from below. We obtain such a lower bound as a result of Theorem 2.1:

$$\sum_{x \in E} \phi(x) \geq \sum_{x \in E} (\psi(x) - k + 1) = \sum_{x \in E} \sum_{\substack{y \in E \\ x \cdot y = t}} 1 - (k-1)|E|$$

$$= \frac{|E|^2}{q} + O\left(q^{\frac{d-1}{2}}|E|\right) - (k-1)|E| \geq 2^{-\frac{1}{k}} \frac{|E|^2}{q},$$

assuming $|E| \geq C_k q^{\frac{d+1}{2}}$ for an appropriate constant C_k depending only on k. This yields

$$N_k(E) \geq \frac{|E|^{k+1}}{2q^k}.$$

□

Having obtained a lower bound for the number of k-stars in $\mathcal{G}_t(E)$, we are particularly interested in the case $k = d$, and particularly those stars $(y, x_1, ..., x_d)$ with the property that $\{x_1, ..., x_d\} \subseteq \mathbb{F}_q^d$ is a linearly independent set of vectors. Therefore, we would like to find an upper bound for the number of d-stars $(y, x_1, ..., x_d)$ formed from linearly dependent sets $\{x_1, ..., x_d\}$.

Lemma 2.5 *Let $\mathcal{N}_d(E)$ be the number of d-stars $(y, x_1, ..., x_d)$ in $\mathcal{G}_t(E)$ such that $\{x_1, ..., x_d\}$ is a linearly independent set. If*

$$|E| \geq C_d q^{d-\frac{1}{d-1}},$$

for q sufficiently large, then

$$\mathcal{N}_d(E) \geq \frac{|E|^{d+1}}{3q^d}.$$

Proof In a star $(y, x_1, ..., x_d)$, if $\{x_1, ..., x_d\}$ is linearly dependent, we may assume without loss of generality that

$$x_d \in \mathrm{Span}(x_1, ..., x_{d-1}).$$

For a given $y \in E$, there are $\psi(y)$ points $x \in E$ such that $x \cdot y = t$. Therefore, there are at most $\psi(y)^{d-1}$ choices for the first $d-1$ points $x_1, ..., x_{d-1}$. Once $y, x_1, ..., x_{d-1}$ are fixed, we see that the point x_d lies on the hyperplane $\{x \in E : x \cdot y = t\}$ as well as the hyperplane $\mathrm{Span}(x_1, ..., x_{d-1})$. These are not the same hyperplane, as only one of them contains the origin since $t \neq 0$. Moreover, their intersection is nonempty since it contains x_d, and so we conclude that x_d must be chosen from a $(d-2)$-dimensional subspace, which must have q^{d-2} points. Putting this together, we find that the number of stars $(y, x_1, ..., x_d)$ in $\mathcal{G}_t(E)$ with the set $\{x_1, ..., x_d\}$ being linearly dependent is bounded by

$$\begin{aligned} dq^{d-2} \sum_{y \in E} \psi(y)^{d-1} &\leq dq^{d-2} q^{(d-1)(d-2)} \sum_{y \in E} \psi(y) \\ &\lesssim dq^{d(d-2)} \frac{|E|^2}{q}, \end{aligned}$$

since $\phi(y) \leq q^{d-1}$ for any y. The factor of d comes from the fact that we chose $x_d \in \mathrm{Span}(x_1, ..., x_{d-1})$. The last line follows from Theorem 2.1. We find that

$$dq^{d(d-2)} \frac{|E|^2}{q} < \frac{|E|^{d+1}}{6q^d}$$

as long as $|E| \geq C_d q^{d-\frac{1}{d-1}}$ for an appropriate constant C_d. Finally, Lemma 2.4 finishes the proof of the statement. □

Definition 2.6 For a d-star $\mathcal{S} = (y, x_1, ..., x_d)$, we call $L = \{x_1, ..., x_d\}$ the leaf set. We say a subset $A = \{x_{n_1}, ..., x_{n_k}\}$ of the leaf set is bad with respect to $\mathcal{S}$ if for every $z \in E$ satisfying $z \cdot x_{n_i} = t$ for all $i = 1, ..., k$, there is some $x \in L \setminus A$ with $z \cdot x = t$ as well.

Remark 2.7 Our definition of a bad set is designed for testing whether the set $\{x_1, ..., x_d\}$ is shattered by $\mathcal{H}_t(E)$. In particular, it follows immediately from definitions that $\{x_1, ..., x_d\} \subseteq E$ is shattered if and only if there is some $y \in E$ so that

$\mathcal{S} = (y, x_1, ..., x_d)$ is a d-star in $\mathcal{G}_t(E)$, and $\{x_1, ..., x_d\}$ admits no bad subset of size $k = 1, ..., d - 1$.

With this in mind, our strategy is to show that a generic d-star in $\mathcal{G}_t(E)$ with a linearly independent leaf set admits no bad sets. To see this, we bound the number of d-stars corresponding to a given bad set.

Definition 2.8 Given a set $B = \{b_1, ..., b_k\}$ which is bad in some d-star $\mathcal{S} = (y, x_1, ..., x_d)$ with linearly independent leaf set $L = \{x_1, ..., x_d\}$, let

$$\mathcal{Q}(B) := \{x \in E : x \cdot b_i = t \ \forall i = 1, ..., k\}.$$

If $\mathcal{Q}(B)$ is small, this restricts the number of choices for the point y in a star $\mathcal{S} = (y, x_1, ..., x_d)$ containing B. If $\mathcal{Q}(B)$ is large, on the other hand, we will see that this restricts the number of choices for the leaf set. The following lemma will allow us to separate into cases based on the size of $\mathcal{Q}(B)$.

Lemma 2.9 *Suppose that $B = \{b_1, ..., b_k\}$ is bad in some star $\mathcal{S} = (y, x_1, ..., x_d)$, and that*

$$|\mathcal{Q}(B)| > q^{r-1}.$$

Then for any $y \in \mathcal{Q}(B)$, there is a subset $J \subseteq \mathcal{Q}(B)$ of size r, not containing y, so that $\{y\} \cup J$ is linearly independent.

Proof Fix $b \in B$, so that every point $x \in \mathcal{Q}(B)$ lies on the hyperplane H_b defined by $x \cdot b = t$. Suppose that J is the largest subset of $\mathcal{Q}(B)$, with the desired property that $\{y\} \cup J$ is linearly independent and J does not contain y. For any

$$z \in \mathcal{Q}(B) \backslash \text{Span}(\{y\} \cup J),$$

we see that $\{y, z\} \cup J$ is linearly independent. Since we assumed that J is maximal, this means that

$$\mathcal{Q}(B) \backslash \text{Span}(\{y\} \cup J) = \emptyset.$$

Therefore,

$$\mathcal{Q}(B) = \mathcal{Q}(B) \cap \text{Span}(\{y\} \cup J) \subseteq H_b \cap \text{Span}(\{y\} \cup J).$$

Also note that H_b does not contain $\text{Span}(\{y\} \cup J)$ since the former does not contain 0, while the latter does. Thus, their intersection is an affine subspace of dimension at most $|J|$, having at most $q^{|J|}$ points. Therefore,

$$q^{r-1} < |\mathcal{Q}(B)| \leq |H_b \cap \text{Span}(\{y\} \cup J)| \leq q^{|J|},$$

so $|J| \geq r$. □

Lemma 2.10 *For $E \subseteq \mathbb{F}_q^d$, the number of d-stars in E with linearly independent leaf set containing a bad set of size k is at most*

$$C'_d|E|^k q^{d^2-kd-d+k},$$

for an appropriate constant C'_d.

Proof We fix a linearly independent set $B = \{x_1, ..., x_k\} \subseteq E$, $1 \leq k \leq d-1$, and count the ways to extend this to a d-star $(y, x_1, ..., x_d)$ for which B is a bad set and $\{x_1, ..., x_d\}$ is linearly independent. Note that permuting the elements of $\{x_1, ..., x_d\}$ does not change any of this data, so up to a constant depending only on d, this is the only case we need to consider. We assume B is bad in at least one d-star, $\mathcal{S}_0 = (y^0, x_1, ..., x_k, x_{k+1}^0, ..., x_d^0)$, with $\{x_1, ..., x_k, x_{k+1}^0, ..., x_d^0\}$ linearly independent, since otherwise the count is zero. To extend to a different d-star $\mathcal{S} = (y, x_1, ..., x_k, x_{k+1}, ..., x_d)$, y must be chosen from the set $\mathcal{Q}(B)$. Let ℓ be the smallest positive integer satisfying

$$|\mathcal{Q}(B)| \leq q^\ell,$$

so that there are at most q^ℓ choices for $y \in \mathcal{Q}(B)$. Given such a choice, we count the number of ways to extend the leaf set to obtain a valid star $\mathcal{S}$. Since

$$|\mathcal{Q}(B)| > q^{\ell-1},$$

Lemma 2.9 tells us that there exists a subset $J \subseteq \mathcal{Q}(B)$ with ℓ points, not containing y, such that $\{y\} \cup J$ is linearly independent. For $x \in E \setminus B$, let

$$\Phi_x = J \cap \{z \in E : x \cdot z = t\}.$$

Suppose that the leaf set of $\mathcal{S}$ is $L = A \cup B$, so that

$$A = \{x_{k+1}, ..., x_d\}.$$

If B is bad in $\mathcal{S}$, then

$$\bigcup_{i=k+1}^{d} \Phi_{x_i} = J.$$

Given some set $Z \subseteq J$, for any $x \in E$ satisfying $\Phi_x = Z$, we see that x lies on the hyperplane $H_z := \{x : x \cdot z = t\}$ for each $z \in Z$. Since we already fixed the point y in the star $\mathcal{S} = (y, x_1, ..., x_d)$, x also lies on H_y. Since $\{y\} \cup Z$ is linearly independent, this means there are at most $q^{d-1-|Z|}$ choices for $x \in E$ satisfying $\Phi_x = Z$. Therefore, summing over all possible collections of $d-k$ subsets of J whose union is J, we find that the number of stars $\mathcal{S}$ containing B in the leaf set is at most

$$q^\ell \sum_{\substack{(Z_1,...,Z_{d-k}) \\ \bigcup Z_i = J}} \prod_{i=1}^{d-k} q^{d-1-|Z_i|} = q^{(d-1)(d-k)+\ell} \sum_{\substack{(Z_1,...,Z_{d-k}) \\ \bigcup Z_i = J}} \prod_{i=1}^{d-k} q^{-|Z_i|}$$

$$= q^{d^2-kd-d+k+\ell} \sum_{\substack{(Z_1,\ldots,Z_{d-k}) \\ \bigcup Z_i = J}} q^{-\sum_{i=1}^{d-k}|Z_i|}$$

$$\leq q^{d^2-kd-d+k+\ell} \sum_{\substack{(Z_1,\ldots,Z_{d-k}) \\ \bigcup Z_i = J}} q^{-\ell}$$

$$\leq C'_d q^{d^2-kd-d+k}.$$

Here in the third line we used that $\sum_{i=1}^{d-k}|Z_i| \geq |J| = \ell$. In the fourth line, C'_d is the number of ways to write $J = \bigcup_{i=1}^{d-k} Z_i$; note that C'_d depends only on d, since $|J| = \ell < d$. □

We are now ready to prove Theorem 1.5.

Proof of Theorem 1.5 For each $k = 1, \ldots, d-1$, let $M_k(E)$ denote the number of d-stars in $\mathcal{G}_t(E)$ admitting a bad set of size k, and let $M(E)$ denote the total number of d-stars admitting a bad set of any size. If we can show that $M(E) < \frac{|E|^{d+1}}{3q^d}$, then it follows from Lemma 2.5 that there exists some d-star in $\mathcal{G}_t(E)$ which admits no bad set, and hence the VC-dimension of $\mathcal{H}_t(E)$ is equal to d. Using Lemma 2.10, we see that

$$M(E) \leq \sum_{k=1}^{d-1} M_k(E) \leq C'_d \sum_{k=1}^{d-1} |E|^k q^{d^2-kd-d+k} \leq (d-1)C'_d |E|^{d-1} q^{d-1}.$$

The last step follows from the assumption that $|E| \geq q^{d-1}$, meaning that the summand is largest when k is largest.

Therefore, $M(E) < \frac{|E|^{d+1}}{3q^d}$ whenever

$$|E| \geq C_d q^{d-\frac{1}{2}}.$$

We already needed the stronger restriction $|E| \geq q^{d-\frac{1}{d-1}}$ to apply Lemma 2.5, and this completes the proof. □

Funding This work was supported by the National Science Foundation (NSF) Grant DMS1947438 and Williams College. Alex Iosevich was also partially supported by NSF Grant HDR TRIPODS - 1934962 and NSF Grant DMS2154232. Xuyan Liu was also partially supported by the University of Michigan.

Data availability No data was used for the research described in this article.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

1. Ascoli, R., Betti, L., Cheigh, J., Iosevich, A., Jeong, R., Liu, X., McDonald, B., Milgrim, W., Miller, S.J., Romero Acosta, F., Velazquez Iannuzzeli, S.: VC-dimension and distance chains in $\mathbb{F}_q^d$. Korean J. Math. **32**(1), 43–57 (2024)
2. Ben-David, S., Shalev-Shwartz, S.: Understanding Machine Learning: From Theory to Algorithms. Cambridge University Press (2014)
3. Bennett, M., Chapman, J., Covert, D., Hart, D., Iosevich, A., Pakianathan, J.: Long paths in the distance graph over large subsets of vector spaces over finite fields, J. Korean Math. Soc. **53**, (2016)
4. Brass, P., Moser, W., Pach, J.: Research Problems in Discrete Geometry. Springer Science and Business Media (2006)
5. Bruner, A., Sharir, M.: Distinct distances between a collinear set and an arbitrary set of points. Discret. Math. **341**(1), 261–265 (2018)
6. Covert, D., Iosevich, A., Koh, D., Rudnev, M.: Generalized incidence theorems, homogeneous forms and sum-product estimates in finite fields. Eur. J. Combin. **31**(1), 306319 (2010)
7. Elekes, G.: A note on the number of distinct distances. Period. Math. Hung. **38**(3), 173–177 (1999)
8. Elekes, G., Rónyai, L.: A combinatorial problem on polynomials and rational functions. J. Combin. Theory Ser. A **89**(1), 1–20 (2000)
9. Erdős, P.: On sets of distances of n points. Am. Math. Monthly **53**, 248–250 (1946)
10. Fitzpatrick, D., Iosevich, A., McDonald, B., Wyman, E.: The VC-dimension and point configurations in $\mathbb{F}_q^2$. Discrete Comput. Geom. (2023)
11. Guth, L., Katz, N.: On the Erdős distinct distances problem in the plane. Ann. Math. **181**(1), 155–190 (2015)
12. Iosevich, A., Parshall, H.: Embedding distance graphs in finite field vector spaces. J. Korean Math. Soc. **56**(6), 1515–1528 (2019)
13. Iosevich, A., Rudnev, M.: Erdős distance problem in vector spaces over finite fields. Trans. Am. Math. Soc. **359**(12), 6127–6142 (2007)
14. Iosevich, A., Jardine, G., McDonald, B.: Cycles of arbitrary length in distance graphs on $\mathbb{F}_q^d$. Proc. Steklov Inst. Math. **314**(1), 27–43 (2021)
15. Iosevich, A., McDonald, B., Sun, M.: Dot products in $\mathbb{F}_q^3$ and the Vapnik–Chervonenkis dimension. Discrete Math. **346**(1), 9 (2023)
16. Maga, P.: Full dimensional sets without given patterns. Real Anal. Exchange **36**(1), 79–90 (2010)
17. McDonald, A., McDonald, B., Passant, J., Sahay, A.: Distinct distances from points on a circle to a generic set. Integers **21**, 13 (2021)
18. McDonald, B., Sahay, A., Wyman, W.: The VC-dimension of quadratic residues in finite fields. Discrete Math. **348**, 114192 (2025)
19. Pach, J., de Zeeuw, F.: Distinct distances on algebraic curves in the plane. Comb. Probab. Comput. **26**(1), 99–117 (2017)
20. Pham, T.: Parallelograms and the VC-dimension of the distance sets. Discrete Appl. Math. **349**, 195–200 (2024)
21. Pham, T., Senger, S., Tait, M., Thu-Huyen, N.: VC-dimension and pseudo-random graphs. Discrete Appl. Math. **365**, 231–246 (2025)
22. Sharir, M., Zahl, J.: Cutting algebraic curves into pseudo-segments and applications. J. Combin. Theory Ser. A **150**, 1–35 (2017)
23. Ya, A., Chervonenkis Vapnik, V.N.: Distinct distances between a collinear set and an arbitrary set of points. Theory Prob. Appl. **16**, 264 (1971)