VC-DIMENSION OF SPHERICAL HYPOTHESIS CLASSES OVER $\mathbb{F}_q^d$

LIVIA BETTI

ABSTRACT. Consider a set X and a set of binary functions $\mathcal{H}$ on the points of X. A subset $C \subset X$ is shattered by $\mathcal{H}$ if the restriction of $\mathcal{H}$ to the points of C yields all $2^{|C|}$ possible functions on the points of C. The VC-Dimension of $\mathcal{H}$ is the size of the largest shattered subset. Vapnik and Chervonenkis introduced the concept of VC-dimension in 1970, which characterizes hypothesis classes of binary functions and has applications in learning theory. In this paper, I outline work I did this past summer regarding hypothesis classes of the form $\mathcal{H}_t(E) = \{h_{y,z} : y, z \in E\}$ where $h_{y,z}(x) = 1$ if $||x-y|| = ||x-z|| = t$ and 0 otherwise. We found that for $d \geq 3$, $\text{VCdim}\mathcal{H}_t(E) = \text{VCdim}\mathcal{H}_t(\mathbb{F}_q^d)$ when $|E| \geq Cq^{d-\frac{1}{d-1}}$, with a slightly stronger result for $d = 3$. Recently, the methods in this paper were applied to the hypothesis class $\mathcal{H}_t(E) = \{h_y : y \in E\}$ where $h_y(x) = 1$ if $x \cdot y = t$ and 0 otherwise. I follow the proof of these results closely, and I further expand on this by exploring the result of adding more parameters to the original hypothesis class.

CONTENTS

1. INTRODUCTION

VC-dimension is a concept in learning theory pertaining to the complexity of a hypothesis class. This concept was first introduced by Vladimir Vapnik and Alexey Chervonenkis in 1970. In particular, VC-dimension is useful in determining which hypothesis classes $\mathcal{H}$ are PAC-learnable, which we will soon define. In [8], the authors present examples of finite classes of functions $\mathcal{H}$ which are learnable and show that the class of all functions over an infinite size domain is not learnable. This raises the question of what kind of classes can be learned, and when infinite classes of functions can be learned. We now formalize the concept of VC-dimension. In general, we follow the notation of [8]. We denote our domain by X, which is the set of points we wish to label. In this paper, we will explore $X = \mathbb{F}_q^d$. Then, we denote a prediction rule as h, which is a function from our domain, X, to our labels, Y. Specifically, $Y = \{0, 1\}$, and thus h is a binary classifier. We note that VC-dimension is only defined for binary classifiers, although generalizations exist.

Definition 1.1. *We say that a set of points $C \subset X$ is shattered by $\mathcal{H}$ if the restriction of $\mathcal{H}$ to C yields all $2^{|C|}$ possible functions on the points of C.*

Date: May 5, 2023.

Based on joint work done at the SMALL REU with Ruben Ascoli, Justin Cheigh, Alex Iosevich, Ryan Jeong, Xuyan Liu, Brian McDonald, Wyatt Milgrim, Steven J. Miller, Francisco Romero Acosta, and Santiago Velazquez Iannuzzelli.

Definition 1.2. *The VC-dimension of $\mathcal{H}$ is the size of the largest shattered subset of X. Specifically, $\mathcal{H}$ has VC-dimension n if there exists a subset of size n shattered by $\mathcal{H}$, but no subset of size $n+1$ is shattered by $\mathcal{H}$. We denote the VC-dimension of $\mathcal{H}$ by $VCdim(\mathcal{H})$.*

We illustrate two examples of VC-dimension, one where $X = \mathbb{R}^2$ and one where $X = \mathbb{F}_q^d$.

Example 1.3. *Consider $X = \mathbb{R}^2$ along with the hypothesis class of axis-aligned rectangles. Therefore, $\mathcal{H} = \{h_{(a,b,c,d):a\leq b,\ c\leq d}\}$ with*

$$h_{(a,b,c,d)}(x_1, x_2) \ = \ \begin{cases} 1 \ \text{ if } \ a \leq x_1 \leq b, \ c \leq x_2 \leq d \\ 0 \ \text{ otherwise} \end{cases} .$$

Then, we see that $\mathcal{H}$ can shatter 4 points by arranging them 4 points in a diamond. However, $\mathcal{H}$ cannot shatter 5 points since for any set of points $\{y_1, y_2, y_3, y_4, y_5\}$, there is no $h \in \mathcal{H}$ such that h is 1 on the subset of the leftmost, rightmost, lowest, and highest point. Therefore, $VCdim(\mathcal{H}) = 4$.

In our next example, $X = \mathbb{F}_q^d$ and $\mathcal{H}$ is the set of spheres of radius t. This result was proved by Nathaneal Grand, Mandar Juvekar and Maxwell Sun during the 2021 Tripods REU [7].

Example 1.4. *Consider $X = \mathbb{F}_q^d$ and $\mathcal{H} = \{h_y\}$ where*

$$h_y(x) \ = \ \begin{cases} 1 \ \text{ if } \ ||x-y||= t \\ 0 \ \text{ otherwise} \end{cases} ,$$

where $||x-y||= (x_1-y_1)^2 + \cdots + (x_d-y_d)^2$. We first notice that $VCdim(\mathcal{H}) \leq d+1$. For any set of $d+2$ points in $\mathbb{F}_q^d$, if these points are in general position (no set of k points lie on a $(k-2)$-dimensional space for $k = 2, \ldots, d+1$), then $d+1$ of them determine a sphere. So, there is no function h which is 1 on all of these points. If these points are not in general position, then there is one point x that lies on a sphere determined by the remaining $d+1$ of them. However, then the restriction of $\mathcal{H}$ to this set of points does not yield the function which is 0 on the point x and 1 on the remaining points. We defer the proof that $VCdim(\mathcal{H}) \geq d+1$ to Appendix A, as we use the result of Theorem 4.3 and Lemma 4.4.

2. Learning Theory Perspective

The study of VC-dimension, and in particular, VC-dimension over finite fields, is largely motivated by concepts in computational learning theory. We consider the task of learning a classifier $c \in \mathcal{H}$, where $\mathcal{H}$ is a set of binary classifiers over a set E. Let $\mathcal{D}$ be a distribution over E that is unknown to the learner. The learner is given access to values of $c(x)$, where the input x is sampled according to the distribution $\mathcal{D}$. Based of these values, the learner provides a classifier $h \in \mathcal{H}$ which is close to the true classifier, c, with high probability. We define a loss function $L_{\mathcal{D},c} : \mathcal{H} \to [0, 1]$ by

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

where $x \sim \mathcal{D}$ indicates that x is sampled according to the distribution $\mathcal{D}$. The loss function is defined with respect to a specific distribution $\mathcal{D}$ and the true classifier c. The loss of a particular classifier $h \in \mathcal{H}$ is the probability that the function value of h at x disagrees with the true classifier evaluated at x, when x is drawn according to the distribution $\mathcal{D}$. We proceed by defining learnability:

Definition 2.1. *The hypothesis class $\mathcal{H}$ is PAC-learnable if there exists a function $m_{\mathcal{H}} : (0, 1)^2 \to \mathbb{N}$ and an algorithm $\mathcal{A}$ such that for all $\varepsilon, \delta \in (0, 1)$, for any fixed classifier $c \in \mathcal{H}$ and any distribution $\mathcal{D}$ over X, if $m \geq m_{\mathcal{H}}(\varepsilon, \delta)$ i.i.d samples are drawn according to $\mathcal{D}$, then if the algorithm $\mathcal{A}$ produces a classifier h after running on these samples, then $L_{\mathcal{D},c}(h) \leq \varepsilon$ with probability at least $1 - \delta$.*

Therefore, we consider a hypothesis class learnable if there exists an algorithm that can choose an $h \in \mathcal{H}$ so that the probability that h approximates the true classifier c ($L_{\mathcal{D},c} \leq \varepsilon$) is arbitrarily close to 1 (probability at least $1 - \delta$). The concept of VC-dimension provides information about the PAC-learnability of a set of classifiers, as we see in the fundamental theorem of machine learning:

Theorem 2.2. *The hypothesis class $\mathcal{H}$ has finite VC-dimension if and only if $\mathcal{H}$ is PAC-learnable.*

Therefore, by the fundamental theorem of machine learning, any hypothesis class with finite VC-dimension is PAC-learnable, which indicates that VC-dimension characterizes PAC-learnability.

Previous research has investigated spherical hypothesis classes over $\mathbb{F}_q^2$. In particular, Fitzpatrick, Iosevich, Wyman, and McDonald considered the hypothesis class $\mathcal{H}_t^2(E) = \{h_y : y \in E\}$ such that $h_y : E \to \{0, 1\}$ is defined by

$$h_y(x) = \begin{cases} 1 & \text{if} \quad ||x - y||= t \\ 0 & \text{otherwise} \end{cases} .$$

Here, t is a fixed number in $\mathbb{F}_q$ and the authors use $||x||= x_1^2 + x_2^2$. The hypothesis class $\mathcal{H}_t^2(E)$ is interpreted as the indicator functions of spheres of radius t in $\mathbb{F}_q^2$ with centers in E. We can also consider a graphical representation of this hypothesis class. Consider the graph $G = (V, E)$, where $V = X$ and E is the edge set connecting two points x and y if and only if $||x - y||= t$. This transforms the VC-dimension problem into a point configuration problem of finding a particular subgraph.

We note that $\mathcal{H}_t^2(\mathbb{F}_q^2)$ is a set of binary classifiers on the whole space, whereas $\mathcal{H}_t^2(E)$ is restricted to a subset of this vector space. The authors consider the following question: how large does E need to be to guarantee that $\mathcal{H}_t^2(E)$ has the same VC-dimension as $\mathcal{H}_t^2(\mathbb{F}_q^2)$? They found that if $|E|\geq Cq^{15/8}$ for a sufficiently large constant C, then the VC-dimension of $\mathcal{H}_t^2(E)$ is equal to 3 [6].

In this paper, we explore two hypothesis classes. The first hypothesis class we look at is $\mathcal{H}_t^d(E) = \{h_{u,v} : (u, v) \in E \times E\}$ where

$$h_{u,v}(x) = \begin{cases} 1 & \text{if} \quad ||u - x||= ||v - x||= t \\ 0 & \text{otherwise} \end{cases} .$$

The binary classifiers $h_{u,v}$ are interpreted as the indicator functions on the intersection of two spheres of radius t, one centered at u and one centered at v. We have the following result:

Theorem 2.3. *If $E \subset \mathbb{F}_q^d$, $d \geq 2$, and*

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

for a constant C depending only on d, then the VC-dimension of $\mathcal{H}_t^d(E)$ is equal to d.

The next hypothesis class we look at uses the dot product. We consider $\mathcal{H}_t^d(E)^* = \{h_y : y \in E\}$, where

$$h_y(x) = \begin{cases} 1 & \text{if} \quad x \cdot y = t \\ 0 & \text{otherwise} \end{cases} .$$

For this hypothesis class, we have the following theorem:

Theorem 2.4. *If $E \subset \mathbb{F}_q^d$, and $|E|\geq Cq^{d-\frac{1}{d-1}}$ for a constant depending only on d, and for q sufficiently large, the VC-dimension of $\mathcal{H}_t^d(E)^*$ is equal to d.*

We note that as previously mentioned, these results can be interpreted in terms of graph theory. We can consider a graph G with vertices in E and edges between vertices x, y if and only if $||x - y||= t$ or $x \cdot y = t$ respectively. Note that changing the binary classifiers changes the configuration of the subgraph we are looking for.

3. Preliminaries for Theorem 2.3

We now introduce preliminary definitions and notation used in proving our theorems. As described with regards to the hypothesis class introduced by Fitzpatrick, Wyman, Iosevich, and McDonald, the problem of determining the VC-dimension of this hypothesis class reduces to finding a point configuration in a corresponding graphical representation. In the following, we define a distance graph.

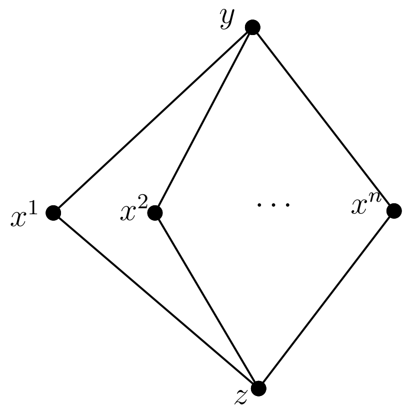

FIGURE 1. Representation of an n-prism in the distance graph $\mathcal{G}_t(E)$

Definition 3.1. *For a set $E \subset \mathbb{F}_q^d$, let $\mathcal{G}_t(E)$ be the graph with vertices elements of the set E and an edge between x and y if and only if $||x - y||= t$. We call $\mathcal{G}_t(E)$ the distance graph of E.*

The problem of shattering a set with $\mathcal{H}_t^d(E)$ is analagous to finding a specified point configuration in $\mathcal{G}_t(E)$. The desired point configuration will be discussed in detail after defining relevant terms.

Definition 3.2. *Let $S_t = \{x \in \mathbb{F}_q^d : ||x||= t\}$ denote the sphere of radius t centered at 0. Furthermore, we use $S_t(\cdot)$ to denote the indicator function of the sphere, so that $S_t(x) = 1$ if $||x||= t$ and 0 otherwise.*

Our next definition defines an important part of our desired point configuration. In order to shatter d points $\{x^1, \ldots, x^d\}$, we need the restriction of $\mathcal{H}_t^d(E)$ to this set to have the function that yields 1 on every element. Therefore, we need to find $y, z \in E$ such that $h_{y,z}(x^i) = 1$ for $1 \leq i \leq d$. So, we need to find y, z such that $||y - x^i||= ||z - x^i||= t$ for $1 \leq i \leq d$. This motivates the following definition.

Definition 3.3. *The $(n + 2)$-tuple $P = (y, z, x^1, \ldots, x^n) \in (\mathbb{F}_q^d)^{n+2}$ is an n-prism if for all $i \leq n$, $||x^i - y||= ||x^i - z||= t$. The tail of P, denoted $\mathcal{T}(P)$, is the set $\{y, z\}$. The center of P, denoted $\mathcal{C}(P)$, is the set $\{x^1, \ldots, x^n\}$.*

For a set $\{x^1, \ldots, x^n\} \subset E$ to be shattered by $\mathcal{H}_t^d(E)$, $\{x^1, \ldots, x^n\}$ must be the center of some n-prism in E.

Definition 3.4. *An n-prism $P = (y, z, x^1, \ldots, x^n)$ is non-degenerate if all of its points are distinct.*

We see that in $\mathcal{G}_t(E)$, a d-prism is a complete bipartite subgraph with vertex sets $\{x^1, \ldots, x^d\}$ and $\{y, z\}$. As previously explained, to shatter the set $\{x^1, \ldots, x^d\}$, this set has to be the center of a prism. Furthermore, for every $I \subset \{1, \ldots, d\}$, we need y^I such that $||y^I - x^i||= t$ for $i \in I$ and $||y^I - x^j||\neq t$ for $j \notin I$. Then, together with z, we see that $h_{y^I,z}$ is a binary classifier yielding 1 on x^i for $i \in I$ and 0 on x^j for $j \notin I$. This leads into our next definition.

Definition 3.5. *An element $y \in E$ is a pole of a set $A \subset E$ if*

$$y \in \bigcap_{x \in A} (S_t + x).$$

We denote the set of poles of A by Pole(A).

Therefore, in addition to finding a d-prism, we also need to find a pole of each subset of the center $\{x^1, \ldots, x^d\}$. So, for each $I \subset \{x^1, \ldots, x^d\}$, we need y^I a pole of $\{x^i : i \in I\}$. However, we have the additional condition that $||y^I - x^j||\neq t$ for $j \notin I$. Equivalently, we need y^I to not be a pole of x^j for $j \notin I$.

Definition 3.6. *For a d-prism P with center $\mathcal{C}$, a subset $A \subset C$ is P-bad, or bad in P, if*

$$\bigcap_{x \in A} (S_t + x) \subset \bigcup_{y \in \mathcal{C} \setminus A} (S_t + y),$$

or equivalently

$$\textit{Pole}(A) \subset \bigcup_{y \in \mathcal{C} \setminus \mathcal{A}} \textit{Pole}(\{y\}).$$

We say that P admits a bad set if there is some subset $A \subset \mathcal{C}$ that is P-bad.

Therefore, to find a set of size d that is shattered by $\mathcal{H}_t^d(E)$, we need to find a d-prism that admits no bad sets; the shattered set will be the center of the d-prism. In general, we will show that for E satisfying the hypotheses of Theorem 2.3, asymptotically there is a positive proportion of nondegenerate d-prisms that admit no bad sets. For the purposes of counting prisms, we also require that the center is affinely independent, which we will define below. This allows us to count the size of the set of poles of each subset of the center, i.e., this allows us to obtain an upper bound on

$$\left| \bigcap_{i \in I} (S_t + x^i) \right|.$$

Definition 3.7. *For a vector space V over a field $\mathbb{F}$, the vectors $v_1, \dots, v_k$ are affinely independent if $\lambda_1 v_1 + \cdots + \lambda_k v_k = 0$ with $\lambda_1 + \cdots + \lambda_k = 0$ implies that $\lambda_1 = \cdots = \lambda_k = 0$.*

Intuitively, affine independence is similar to linear independence without the restriction that the span of these vectors must contain the origin. We have the following lemma, regarding the unique determination of a d-dimensional sphere:

Lemma 3.8. *A set of $d+1$ affinely independent points uniquely determines a d-dimensional sphere.*

Proof. An equation for a sphere in $\mathbb{F}_q^d$ with the norm $||x|| = x_1^2 + \cdots + x_d^2$ is

$$\sum_{j=1}^{d} (x_j - y_j)^2 = r,$$

where $y = (y_1, \dots, y_d)$ is the center and r is the radius under the specified norm. Therefore,

$$2 \sum_{j=1}^{d} y_j x_j + \left(r - \sum_{j=1}^{d} y_j^2 \right) = \sum_{j=1}^{d} x_j^2$$

A set of $d+1$ affinely independent points $\{x^1, \dots, x^{d+1}\}$ leads to $d+1$ equations in the variables $y_1, \dots, y_d, r - \sum_{j=1}^d y_j^2$. Note that $r - \sum_{j=1}^d y_j^2$ is treated as a variable so that these $d+1$ equations are linear. Then,

$$\begin{pmatrix} 2x_1^1 & \cdots & 2x_d^1 & 1 \\ 2x_1^2 & \cdots & 2x_d^2 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 2x_1^{d+1} & \cdots & 2x_d^{d+1} & 1 \end{pmatrix}$$

has rank $d+1$ due to the affine independence of $\{x^1, \dots, x^{d+1}\}$. Thus, these equations have a unique solution so that the d-dimensional sphere is uniquely determined. □

We remark that d-dimensional spheres refer to spheres of the form $S_r + y$ for some $y \in \mathbb{F}_q^d$. We also refer to n-dimensional spheres, for $n \leq d$.

Definition 3.9. *An n-dimensional sphere is the intersection of a d-dimensional sphere $S_r + y$ with an n-dimensional affine subspace.*

We also have the following lemma regarding a lower-dimensional sphere determined by a set of affine points.

Lemma 3.10. *A set of $k+1$ affinely independent points, $k \leq d$, lie on the intersection of at most $d-k+1$ spheres of a fixed radius in $\mathbb{F}_q^d$.*

Proof. Consider a set of $k+1$ affinely independent points $\{x^1, \dots, x^{k+1}\}$. Then, suppose this set of points lie on a sphere of radius r centered at $y = (y_1, \dots, y_d)$. Then, each sphere has the equation

$$\sum_{j=1}^{d}(x_j^i - y_j)^2 = r$$

for $1 \leq i \leq k+1$. We determine how many solutions there are for y. In the previous lemma, we saw that a set of $d+1$ affinely independent points determine a d-dimensional sphere. So, without loss of generality, suppose $\{x^1, \dots, x^{k+1}\} \subset S_t$, for some t. Expanding and subtracting the equation for $i = 1$, we have equations of the form

$$\sum_{j=1}^{d}((x_j^i)^2 - (x_j^1)^2) - 2\sum_{j=1}^{d} y_j(x_j^i - x_j^1) = 0$$

so that as $||x^i|| = ||x^1|| = t$, then

$$2\sum_{j=1}^{d} y_j(x_j^i - x_j^1) = 0.$$

As $\{x^1, \dots, x^{k+1}\}$ are affinely independent, then $\{x^2 - x^1, \dots, x^{k+1} - x^1\}$ are a set of k linearly independent points. Therefore, the solution space has $d-k$ solutions, so that $\{x^1, \dots, x^{k+1}\}$ lie on the intersection of at most $d-k+1$ spheres (using that we already supposed $\{x^1, \dots, x^{k+1}\} \subset S_t$). $\square$

As previously mentioned, we impose the condition of affine independence on the center of a prism, as we will use properties of affinely independent sets to count prisms.

Definition 3.11. *A d-prism P is affinely nondegenerate if P is nondegenerate and its center is affinely independent.*

Therefore, we revise our earlier goal in order to show that asymptotically there is a positive proportion of affinely nondegenerate d-prisms that admit no bad sets.

4. Proof of Theorem 2.3

We start by showing that $VCdim(\mathcal{H}_t^d(E)) \leq d$. We do so by showing that $VCdim(\mathcal{H}_t^d(\mathbb{F}_q^d) \leq d$, which implies the desired inequality since $E \subset \mathbb{F}_q^d$ implies $VCdim(\mathcal{H}_t^d(E)) \leq VCdim(\mathcal{H}_t^d(\mathbb{F}_q^d))$. Therefore, consider a set of size $d+1$, namely $\{x^1, \dots, x^{d+1}\}$. We will show that this set cannot be shattered by $\mathcal{H}_t^d(\mathbb{F}_q^d)$. Suppose $\mathcal{H}_t^d(\mathbb{F}_q^d)$ can shatter $\{x^1, \dots, x^{d+1}\}$. Then, there exists y, z such that $h_{y,z}(x^i) = 1$ for $1 \leq i \leq d+1$. So, $||x^i - y|| = ||x^i - z|| = t$ for each i, $1 \leq i \leq d+1$. We have two cases.

If $\{x^1, \dots, x^{d+1}\}$ are affinely independent, then $\{x^1, \dots, x^{d+1}\}$ determine a d-dimensional sphere by Lemma 3.8. Hence, $y = z$ and thus we cannot shatter $\{x^1, \dots, x^{d+1}\}$.

If $\{x^1, \dots, x^{d+1}\}$ are affinely dependent, without loss of generality, let $\{x^1, \dots, x^{n+1}\}$ be the largest affinely independent subset of $\{x^1, \dots, x^{d+1}\}$. Then, by Lemma 3.10, these points lie on the intersection of at most $d-n+1$ spheres. However, to guarantee that $\{x^1, \dots, x^{n+1}\}$ is not a bad set, we need this set to lie on the intersection of $d+2-n$ spheres, i.e., to guarantee the existence of $y^1, \dots, y^{d+1-n}, z$ such that $h_{y^j,z}(x^i) = 1$ for $1 \leq i \leq j$ and $h_{y^j,z}(x^i) = 0$ for $j \leq i \leq d+1$. Thus, we have a contradiction.

In either case, $VCdim(\mathcal{H}_t^d(\mathbb{F}_q^d)) \leq d$ which implies $VCdim(\mathcal{H}_t^d(E)) \leq d$. The remainder of the proof is dedicated to finding a set of d points that is shattered by $\mathcal{H}_t^d(E)$. In general, to find a set of size d that is shattered by $\mathcal{H}_t^d(E)$, we wish to find a nondegenerate d-prism P in E with center $\mathcal{C} = \{x^1, \dots, x^d\}$ and tail $\mathcal{T} = \{y, z\}$ that does not admit any bad sets. Then, we see that $x^1, \dots, x^d$ is shattered by $\mathcal{H}_t^d$. The classifier $h_{y,z}$ is valued 1 on the center. Then, as there are no P-bad sets, for any $A \subset \mathcal{C}$, let

$$w \in \left(\bigcap_{x \in A}(S_t + x)\right) \setminus \left(\bigcap_{y \in \mathcal{C} \setminus \mathcal{A}}(S_t + y)\right)$$

then, $h_{w,y}$ specifies the set $A \subset \mathcal{C}$. Lastly, as this is done asymptotically, then we can choose any point v not in the point configuration determined by P and $S_t + x$ for $x \in \mathcal{C}$ to get $h_{y,v}$ which yields 0 on the center.

First, we introduce a theorem that we use to determine a lower bound on the number of nondegenerate d-prisms. In the distance graph on E, a d-prism in E is a subgraph determined by d distinct paths of length 2 between two points, which make up the tail. This is a case of Theorem 1.1 from [4].

Theorem 4.1. *Let $E \subset \mathbb{F}_q^d$, where $d \geq 2$ and $|E| > \frac{2k}{\log 2} q^{\frac{d+1}{2}}$. Suppose that $t \neq 0$. Define*

$$\Gamma_k = |\{(x^1, \dots x^{k+1}) \in E \times \cdots \times E : ||x^i - x^{i+1}|| = t, 1 \leq i \leq k\}|.$$

Then,

$$\Gamma_k = \frac{|E|^{k+1}}{q^k} + \mathcal{D}_k \quad \text{where} \quad |\mathcal{D}_k| \leq \frac{2k}{\log 2} q^{\frac{d+1}{2}} \frac{|E|^k}{q^k}.$$

In particular, for E satisfying the hypotheses of Theorem 2.3,

$$\Gamma_2 \geq \frac{|E|^3}{2q^2}.$$

As previously mentioned, this theorem comes from [4]. Note that for E satisfying

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

we see that

$$\begin{aligned} \Gamma_2 &= \frac{|E|^3}{q^2} + \mathcal{D}_2 \\ &\geq \frac{|E|^3}{q^2} - \frac{4}{\log 2} q^{\frac{d+1}{2}} \frac{|E|^2}{q^2} \\ &= \frac{|E|^3}{q^2} \left(1 - \frac{4}{\log 2} q^{\frac{d+1}{2}} \frac{1}{|E|}\right) \end{aligned}$$

Using that $|E| \leq q^d$, we have

$$\Gamma_2 \geq \frac{|E|^3}{q^2} \left(1 - \frac{4}{\log 2} q^{\frac{1-d}{2}}\right) \geq \frac{|E|^3}{2q^2}$$

for q reasonably large.

We first prove the theorem for $d = 2$. We employ techniques used in [6] to prune the set E so that each point in E is saturated with adjacent points in the distance graph of E. Then, we use a lemma from the same authors to find points y, z such that $h_{y,z}(x^1) = h_{y,z}(x^2) = 1$ for some x^1, x^2.

Proof of $d = 2$. By Theorem 3.1 in [6], if $|E| \geq 4q^{\frac{3}{2}}$, we have

$$\sum_{x \in E} |E \cap (S_t + x)| \geq \frac{|E|^2}{2q}. \tag{4.1}$$

Furthermore, $|E| > 4 \cdot 99q$ for $q \geq 99^2$, which implies that $99|E| \leq \frac{1}{4}|E|^2 q^{-1}$. Therefore,

$$\begin{aligned} \sum_{x \in E} |E \cap (S_t + x)| &\leq 99|\{x \in E : |E \cap (S_t + x) \leq 99| + \sum_{\substack{x \in E \\ |E \cap (S_t + x)| \geq 100}} |E \cap (S_t + x)| \\ &\leq \frac{1}{4}|E|^2 q^{-1} + \sum_{\substack{x \in E \\ |E \cap (S_t + x)| \geq 100}} |E \cap (S_t + x)| \end{aligned}$$

Together with Equation 4.1, this implies

$$\sum_{\substack{x\in E\\ |E\cap(S_t+x)|\geq 100}} |E\cap(S_t+x)|\geq \frac{1}{4}|E|^2 q^{-1}.$$

Then, by Cauchy-Schwarz we have

$$\begin{aligned}\frac{1}{16}|E|^4 q^{-2} &\leq \left(\sum_{\substack{x\in E\\ |E\cap(S_t+x)|\geq 100}} |E\cap(S_t+x)|\right)^2\\ &\leq |\{x\in E : |E\cap S_t(x)\geq 100\}|\left(\sum_{x\in E}|E\cap(S_t+x)|^2\right).\end{aligned}$$

We have that

$$\sum_{x\in E}|E\cap(S_t+x)|^2 = \sum_{x,y,z} E(x)E(y)E(z)S_t(x-y)S_t(x-z) \tag{4.2}$$

where $E(\cdot)$ and $S_t(\cdot)$ represent indicator functions on the respective sets. Therefore, Equation 4.2 counts the number of paths of length 2 in E. By Theorem 4.1, if $|E|> \frac{4}{\log 2}q^{\frac{3}{2}}$, then the number of paths of length 2 is $\leq 2\frac{|E|^3}{q^2}$. Therefore,

$$\frac{1}{16}|E|^4 q^{-2} \leq 2\frac{|E|^3}{q^2}|\{x\in E : |E\cap(S_t+x)|\geq 100\}|$$

so that

$$|\{x\in E : |E\cap(S_t+x)|\geq 100\}|\geq \frac{1}{32}|E|.$$

Therefore, let $E' = \{x\in E : |E\cap(S_t+x)|\geq 100\}$. Then, by Lemma 4.1 of [6], as $|E'|\geq 4q^{\frac{7}{4}}$, there exists distinct $x,y,z,w\in E'$ such that

$$||x-y||= ||y-w||= ||w-z||= ||x-z||= t.$$

Therefore, we see that the set $\{x,w\}$ is shattered. The above shows us the existence of y,z such that $h_{y,z}(x) = h_{y,z}(w) = 1$. Then, as x,w are chosen from our pruned set E', they have ≥ 100 neighbors in the distance graph of E'. Choosing u,v such that $||u-x||= t$, $||v-x||\neq t$ and $||u-w||\neq t$, $||v-w||= t$, we have $h_{u,y}(x) = 1$, $h_{u,y}(w) = 0$ and $h_{v,y}(x) = 0$, $h_{v,y}(w) = 1$. Together with $h_{y,z}$, we see that $\{x,w\}$ is shattered. □

Now, we consider general $d\geq 2$. Using Theorem 4.1, we obtain the following lower bound for the number of nondegenerate d-prisms in E.

Theorem 4.2. *Let $E\subset\mathbb{F}_q^d$, $d\geq 3$. Let $N_d(E)$ be the number of nondegenerate d-prisms in E. If $|E|> \frac{4}{\log 2}q^{\frac{d+1}{2}}$, then*

$$N_d(E)\gtrsim_d \frac{|E|^{d+2}}{q^{2d}}.$$

Here, the notation $A\gtrsim B$ indicates that $A\geq cB$ for some constant c. A subscript indicates that the constant may depend on another value, and in this case, the constant may depend on d.

Proof. Let $k_{(x,y)}$ be the number of path of length 2 from x to y in the distance graph of E. Then, we have that

$$N_d(E) = \sum_{\substack{x,y\in E\\ x\neq y}} d!\binom{k_{(x,y)}}{d} = \sum_{\substack{x,y\in E\\ x\neq y}} k_{(x,y)}(k_{(x,y)}-1)\cdots(k_{(x,y)}-d+1). \tag{4.3}$$

This is because for x, y distinct, $\binom{k_{(x,y)}}{d} = 0$ if $k_{(x,y)} < d$, i.e. if there are fewer than d paths between x and y they admit no d-prisms. And, if $k_{(x,y)} \geq d$, then there are $d!\binom{k_{(x,y)}}{d}$ is the number of tuples $(x^1, \dots, x^d)$ that can be the center, accounting for order.

For each $(x, y) \in E^2$, we define

$$k'_{(x,y)} = \max(k_{(x,y)} - d + 1, 0).$$

Then, by Equation 4.3,

$$N_d(E) \geq \sum_{\substack{x,y\in E\\ x\neq y}} (k'_{(x,y)})^d. \tag{4.4}$$

By Theorem 4.1, we have that

$$\sum_{x,y\in E} k_{(x,y)} = \Gamma_2 \geq \frac{|E|^3}{2q^2}.$$

Then, notice that $k_{(x,y)}$ where $x = y$ is twice the number of one paths with one of the endpoints $x = y$. Therefore,

$$\begin{aligned}\sum_{\substack{x,y\in E\\ x=y}} k_{(x,y)} &= 2\Gamma_1\\ &= 2\left(\frac{|E|^2}{q} + \mathcal{D}_1\right)\\ &\leq 2\left(\frac{|E|^2}{q} + \frac{2}{\log 2}q^{\frac{d+1}{2}}\frac{|E|}{q}\right)\\ &= 2\frac{|E|^2}{q}\left(1 + \frac{2}{\log 2}q^{\frac{d+1}{2}}\frac{1}{|E|}\right)\\ &\leq 3\frac{|E|^2}{q}\end{aligned}$$

using that $|E| > \frac{4}{\log 2}q^{\frac{d+1}{2}}$. Thus,

$$\sum_{\substack{x,y\in E\\ x\neq y}} k_{(x,y)} \geq \frac{|E|^3}{2q^2} - 3\frac{|E|^2}{q} \gtrsim \frac{|E|^3}{q^2}$$

as $|E| \gg q$. So, we have

$$\sum_{\substack{x,y\in E\\ x\neq y}} k'_{(x,y)} \;\geq\; \sum_{\substack{x,y\in E\\ x\neq y}} (k_{(x,y)} - d + 1) \;\gtrsim\; \frac{|E|^3}{q^2} - (d-1)|E|^2 \gtrsim_d \frac{|E|^3}{q^2} \tag{4.5}$$

as $\frac{|E|^3}{q^2} \gg (d-1)|E|^2$ allows us to bound $(d-1)|E|^2$ by a small constant (dependent on d) times $\frac{|E|^3}{q^2}$.

Recall, Hölder's inequality:

$$\left(\sum_{i=1}^{n} a_i^r b_i^s\right)^{r+s} \leq \left(\sum_{i=1}^{n} a_i^{r+s}\right)^r \left(\sum_{i=1}^{n} b_i^{r+s}\right)^s.$$

Taking $n = |E|^2$, $a_i = k'_{(x,y)}$ by arbitrarily indexing the pairs (x, y), $b_i = 1$, $r = 1$, $s = d-1$, we have

$$\left(\sum_{\substack{x,y\in E\\ x\neq y}} k'_{(x,y)}\right)^d \leq \left(\sum_{\substack{x,y\in E\\ x\neq y}} (k'_{(x,y)})^d\right)(|E|^2)^{d-1}.$$

Therefore, applying the inequality we obtained in Equation 4.5,

$$\left(\sum_{\substack{x,y\in E\\ x\neq y}} (k'_{(x,y)})^d\right) \gtrsim_d \left(\frac{|E|^3}{q^2}\right)^d \frac{1}{(|E|^2)^{d-1}} = \frac{|E|^{d+2}}{q^{2d}}.$$

Thus, by Equation 4.4,

$$N_d(E) \geq \sum_{\substack{x,y\in E\\ x\neq y}} (k'_{(x,y)})^d \geq \frac{|E|^{d+2}}{q^{2d}}.$$

□

Now that we have a lower bound on the number of d-prisms, we wish to determine asymptotically how many of these d-prisms are affinely nondegenerate. To estimate the number of points on a d-dimensional sphere, we have the following theorem, which is a special case of a theorem from [3] which was adapted from a theorem proved by Minkowski.

Theorem 4.3. *For $S_t \subset \mathbb{F}_q^d$ defined above,*

$$q^{d-1} - q^{\frac{d}{2}} < |S_t| < q^{d-1} + q^{\frac{d}{2}}.$$

Therefore, the number of points on a d-dimensional sphere, $|S_t|$, is asymptotically q^{d-1}.

Next, we characterize the points on a d-sphere which lie in an affine subspace of $\mathbb{F}_q^d$.

Lemma 4.4. *Let A be an n-dimensional affine subspace of $\mathbb{F}_q^d$. Then, $|A \cap S_t| \leq 2q^{n-1}$.*

Proof. Let V be an n-dimensional linear subspace of $\mathbb{F}_q^d$. Fix $b \in A$. Then, for this fixed basepoint b, we can represent A as follows:

$$A = \{b + w : w \in V\}.$$

Let $v_1, \ldots v_n$ be a basis for V. Then, for any $a \in A$, we have

$$a = b + c_1v_1 + c_2v_2 + \cdots + c_nv_n$$

for $c_1, \ldots, c_n \in \mathbb{F}_q$ constants. Let $v_0 = b$ and $c_0 = 1$ so that $a = \sum_{i=0}^n c_iv_i$. Furthermore, for $x, y \in \mathbb{F}_q^d$, let $x \cdot y = x_1y_1 + \cdots + x_dy_d$. Then,

$$||a|| = \left|\left|\sum_{i=0}^n c_iv_i\right|\right| = \sum_{j=1}^d \left(\sum_{i=0}^n c_iv_{ij}\right)^2$$

where v_{ij} denotes the jth coordinate of v_i. Thus,

$$||a|| = \sum_{j=1}^d \left(\sum_{i=0}^n c_iv_{ij}\right)^2 = \sum_{j=1}^d \sum_{0\leq i,k\leq n} c_ic_kv_{ij}v_{kj} = \sum_{0\leq i,k\leq n}\sum_{j=1}^d c_ic_kv_{ij}v_{kj} = \sum_{0\leq i,k\leq n}\sum_{j=1}^d v_{ij}v_{kj}$$

$$= \sum_{0\leq i,k\leq n} c_ic_k(v_i\cdot v_k) = c_n||v_n|| + c_n\sum_{i=1}^{n-1} c_i(v_n\cdot v_i) + \sum_{0\leq i,k\leq n} c_ic_k(v_i\cdot v_k)$$

As $a \in A \cap S_t$, then we want $||a|| = t$. Then, fixing $c_1, \ldots, c_{n-1}$ (note $c_0 = b$ is already fixed), and letting $\alpha = ||v_n||$, $\beta = \sum_{i=1}^{n-1} c_i(v_n \cdot v_i)$ and $\gamma = \sum_{0\leq i,k\leq n} c_ic_k(v_i \cdot v_k)$, which are constant, we have the following equation for c_n:

$$c_n^2\alpha + c_n\beta + \gamma = t.$$

As this equation is quadratic in c_n, then there are at most 2 choices for c_n after fixing $c_1, \ldots, c_{n-1}$. As there are q choices for c_i, then there are q^{n-1} choices for $c_1, \ldots, c_{n-1}$. Thus, $|A \cap S_t| \leq 2q^{n-1}$. □

Corollary 4.5. *If A is an n-dimensional affine subspace of $\mathbb{F}_q^d$, then $|A \cap (S_t + y)| \leq 2q^{n-1}$ for any $y \in \mathbb{F}_q^d$.*

Proof. The proof of this fact follows from $|A \cap (S_t + y)| = |y + (A - y) \cap S_t| = |(A - y) \cap S_t|$, where $A - y$ is again an n-dimensional affine subspace. $\square$

Using these lemmas, we prove the following lemma, which shows that given restrictions on the size of E, a positive proportion of d-prisms in E are affinely nondegenerate.

Lemma 4.6. *Let $N'_d(E)$ be the number of affinely nondegenerate prisms in $E \subset \mathbb{F}_q^d$, $d \geq 3$, and assume that $d = 3$ or $|E| \geq C_d q^{d - \frac{1}{d-1}}$. Then,*

$$\frac{N_d(E) - N'_d(E)}{N_d(E)} \leq C'_d$$

where C'_d is a constant with respect to q (but not d), with $C'_d < 1$.

Proof. First, consider $d = 3$. Then, if a nondegenerate d-prism had an affinely dependent center $\mathcal{C}$, this would imply that $\mathcal{C}$ lies in a 1-dimensional affine subspace A. By Lemma 4.4, for a pole y of $\mathcal{C}$, $|A \cap (S_t + y)| \leq 2$, which is a contradiction as a nondegenerate prism has 3 points in its center. Therefore, for $d = 3$, all nondegenerate d-prisms are affinely nondegenerate.

Next, consider $d > 3$. To determine the number of affinely nondegenerate prisms in E, we bound the number of nondegenerate prisms with affinely dependent centers. Consider a d-prism P with tail $\{y, z\}$. We will count the number of choices for the center. As the center is affinely dependent, then the center consists of a set of $d - 1$ points $(x^1, \ldots, x^{d-1})$ and a point x^d that lies on the affine subspace A_0 generated by $\{x^1, \ldots, x^{d-1}\}$. As each x^i must lie on a path of length 2 from y to z, then there are $k_{(y,z)}$ choices for x^i and therefore, $k_{(y,z)}^{d-1}$ choices for the set $(x^1, \ldots, x^{d-1})$.

Then, notice that A_0 has dimension

$$\text{rank} \begin{pmatrix} x^2 - x^1 \\ x^3 - x^1 \\ \vdots \\ x^{d-1} - x^1 \end{pmatrix} \leq d - 2$$

Therefore, as x^d lies on the subspace A_0 and must be distance t from y, then $x^d \in A_0 \cap (S_t + y)$. By Lemma 4.4, $|A_0 \cap (S_t + y)| \leq 2q^{d-3}$ so that there are $\leq 2q^{d-3}$ choices for x^d.

Thus, we see that there are $\leq 2q^{d-3} k_{(y,z)}^{d-1}$ choices for the affinely dependent center $\{x^1, \ldots, x^d\}$. Up to reordering, there are $\leq d!\, 2q^{d-3} k_{(y,z)}^{d-1}$ prisms with tail $\{y, z\}$ and affinely dependent center. So, using that $k_{(y,z)} \leq 2q^{d-2}$ (which applies Lemma 4.4 as the intersection of 2 spheres lies on a $d-1$-dimensional affine space),

$$N_d(E) - N'_d(E) \leq d!\, 2q^{d-3} \sum_{y,z \in E} (k_{(y,z)})^{d-1} \leq d!\, 2q^{d-3} (2q^{d-2})^{d-2} \sum_{y,z \in E} k_{(y,z)}.$$

Then, from Theorem 4.1, $\sum_{y,z \in E} k_{(y,z)} = \Gamma_2 \lesssim_d \frac{|E|^3}{q^2}$. So,

$$N_d(E) - N'_d(E) \leq d!\, 2q^{d-3} (2q^{d-2})^{d-2} \sum_{y,z \in E} k_{(y,z)} \lesssim_d q^{d^2 - 3d + 1} \frac{|E|^3}{q^2} = q^{d^2 - 3d - 1} |E|^3.$$

Lastly, from Theorem 4.2, $N_d(E) \gtrsim_d \frac{|E|^{d+2}}{q^{2d}}$, so that

$$\frac{N_d(E) - N'_d(E)}{N_d(E)} \leq C''_d \frac{q^{d^2 - 3d - 1} |E|^3}{\frac{|E|^{d+2}}{q^{2d}}} = C''_d \frac{q^{d^2 - d - 1}}{|E|^{d-1}}.$$

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

$$\frac{N_d(E) - N'_d(E)}{N_d(E)} \leq C''_d \frac{q^{d^2 - d - 1}}{|E|^{d-1}} \leq \frac{C''_d}{C_d} = C'_d < 1$$

for C_d a sufficiently large constant depending only on d. $\square$

Corollary 4.7. *For E satisfying the hypotheses of Theorem 4.2, we have*
$$N_d'(E) \gtrsim_d \frac{|E|^{d+2}}{q^{2d}}.$$

Now that we have counted the number of affinely nondegenerate prisms, we look at asymptotically how many of these affinely nondegenerate prisms do not admit a bad set. The following lemma provides an upper bound on the number of poles for an affinely independent set.

Lemma 4.8. *Suppose that the set of distinct points $\{a_i\}_{i=1}^k$ are affinely independent. Then,*
$$\left|\bigcap_{i=1}^{k}(S_t + a_i)\right| \leq 2q^{d-k}.$$

Proof. We have that $\{a_i\}_{i=1}^k$ are affinely independent if and only if $\{a_j' : a_j' = a_j - a_1\}_{j=2}^k$ are linearly independent. For notational convenience, let $a_1' = \mathbf{0}$ and represent $a_j' = (\alpha_{j,1}, \ldots, \alpha_{j,d})$. Then, we have
$$\left|\bigcap_{i=1}^{k}(S_t + a_i)\right| = \left|\bigcap_{j=1}^{k}(S_t + a_j')\right|,$$
where
$$\bigcap_{j=1}^{k}(S_t + a_j') = \{x \in \mathbb{F}_q^d : ||x||= t \text{ and } ||x - a_j'||= t \text{ for } j \geq 2\}.$$
Then, any $x \in \bigcap_{j=1}^k(S_t + a_j')$ satisfies the following equations:
$$\begin{aligned}
(x_1 - \alpha_{2,1})^2 + (x_2 - \alpha_{2,2})^2 + \cdots + (x_d - \alpha_{2,d})^2 &= t\\
(x_1 - \alpha_{3,1})^2 + (x_2 - \alpha_{3,2})^2 + \cdots + (x_d - \alpha_{3,d})^2 &= t\\
&\vdots\\
(x_1 - \alpha_{k,1})^2 + (x_2 - \alpha_{k,2})^2 + \cdots + (x_d - \alpha_{k,d})^2 &= t
\end{aligned}$$
Expanding these equations and using $||x||= t$, we have
$$\begin{aligned}
2x_1\alpha_{2,1} + 2x_2\alpha_{2,2} + \cdots + 2x_d\alpha_{2,d} &= ||a_2'||\\
2x_1\alpha_{3,1} + 2x_2\alpha_{3,2} + \cdots + 2x_d\alpha_{3,d} &= ||a_3'||\\
&\vdots\\
2x_1\alpha_{k,1} + 2x_2\alpha_{k,2} + \cdots + 2x_d\alpha_{k,d} &= ||a_k'||
\end{aligned}$$

As a_j' are linearly independent for $2 \leq j \leq k$, then the matrix of a_j' has full rank, i.e. has dimension $k - 1$. Therefore, the solution space, A, corresponding to this system of equations is an affine subspace of dimension $d - (k - 1)$. As $x \in \cap_{i=1}^k(S_t + a_i)$ requires that $||x||= t$, then we see $|A \cap S_t|\leq 2q^{d-k}$. Thus,
$$\left|\bigcap_{i=1}^{k}(S_t + a_i)\right| = \left|\bigcap_{j=1}^{k}(S_t + a_j')\right| \leq 2q^{d-k}.$$
□

As previously mentioned, to find a d-prism that admits no bad sets, we will determine an asymptotic upper bound on the number of d-prisms a given set B is bad in. To this end, given a lower bound on the size of Pole(B), we wish to find an affinely independent subset of Pole(B). In our final Lemma, Lemma 4.10, this will allow us to place a restriction other center points $a_i \notin B$, for $i = 1$ to $d - k$.

Lemma 4.9. *Suppose that B is a bad set with $|Pole(B)|> 2q^{a-1}$. For every $y, z \in Pole(B)$, there exists a subset $J \subset Pole(B)$ such that $J \cup \{y, z\}$ are affinely independent and $|J|= a$.*

Proof. Fix $b \in B$. Then, we construct a sequence of sets $J_0 \subset J_1 \subset J_2 \subset \cdots \subset J_a$ with $|J_i|= i$ and $J_i \cup \{y, z\}$ affinely independent. Let $J_0 = \emptyset$ and $J_1 = b$.

Now, suppose we have chosen $J_0, J_1, \ldots, J_i$. Let A be the $(i+1)$-dimensional subspace generated by $J_i \cup \{y, z\}$. Then, by Lemma 4.4 and its Corollary, $|A \cap \text{Pole}(B)| \leq |A \cap (S_t + b)| \leq 2q^i$. Therefore, as $|\text{Pole}(B)| > 2q^{a-1}$, there exists a point $p \in \text{Pole}(B)$ such that $p \notin A$. Let $J_{i+1} = J_i \cup p$. We see that $|J_{i+1}|= i + 1$ and $J_{i+1} \cup \{y, z\}$ is affinely independent. Our desired set J is J_a. □

We have a final lemma, that allows us to complete the proof of our main result:

Lemma 4.10. *Fix some set B with $|B|= k$. Then, B is bad in at most $C_d q^{d^2-kd-d+k-1}$ affinely nondegenerate prisms.*

Proof. Suppose B is bad in an affinely nondegenerate prism $P = (y, z, x^1, \ldots, x^d)$, with $\mathcal{C}(P) = \{x^1, \ldots, x^d\}$, $\mathcal{T}(P) = \{y, z\}$. Let $M_B(E)$ be the number of other affinely nondegenerate prisms Q in which B is bad. We will show that $M_B(E) \leq C_d q^{d^2-kd-d+k-1}$.

First, we determine an upper bound for the size of $\text{Pole}(B)$. As B is a bad set, then

$$\text{Pole}(B) \subset \bigcup_{a\in\mathcal{C}(P)\setminus B} (S_t + a) = \bigcup_{a\in\mathcal{C}(P)\setminus B} (S_t + a) \cap \left(\bigcap_{b\in B}(S_t + b)\right). \tag{4.6}$$

Therefore,

$$|\text{Pole}(B)| \leq \sum_{a\in\mathcal{C}(P)\setminus B} \left|(S_t + a) \cap \left(\bigcap_{b\in B}(S_t + b)\right)\right| \leq 2(d-k)q^{d-k-1}$$

by Lemma 4.8, as $(S_t + a) \cap \left(\bigcap_{b\in B}(S_t + b)\right)$ is the intersection of $k + 1$ spheres.

First, note that if $|\text{Pole}(B)|= 2$, then there is only one choice of tail so that $\{y, z\} \in \text{Pole}(B)$. As $|(S_t + y) \cap (S_t + z)| \leq 2q^{d-2}$, then we have $(2q^{d-2})^{d-k} = 2^{d-k}q^{(d-2)(d-k)}$. As $k \leq d - 1$, then $q^{(d-2)(d-k)} = q^{d^2-kd-2(d-k)} \leq q^{d^2-kd-d+k-1}$ so that B is bad in at most $C_d q^{d^2-kd-d+k-1}$ affinely nondegenerate prisms.

We proceed under the assumption that $|\text{Pole}(B)| > 2$. Let ℓ be the smallest positive integer such that $|\text{Pole}(B)| \leq 2q^\ell$. Notice that $\ell \leq d - k - 1$. Then, we have fewer than $(2q^\ell)^2 = 4q^{2\ell}$ choices of tail for Q. Fix a choice $\{y, z\}$ for the tail of Q. We proceed by counting the number of choices for the center of Q.

By our choice of ℓ, $|\text{Pole}(B)| > 2q^{\ell-1}$. Therefore, by Lemma 4.9, there exists a $J \subset \text{Pole}(B)$ with $|J|= \ell$ and $J \cup \{y, z\}$ are affinely independent. Choose such a set J.

Next, define a function $\phi : E \setminus B \to \mathcal{P}(J)$, where the range is the power set of J, by $\phi(x) = J \cap \text{Pole}(x)$. Consider $A = (a_1, a_2, \ldots, a_{d-k}) \in (E \setminus B)^{d-k}$, a $(d - k)$-tuple with a_i distinct. Let $T_A = (\phi(a_1), \phi(a_2), \ldots, \phi(a_{d-k})) \in (\mathcal{P}(J))^{d-k}$.

Suppose $\mathcal{C}(Q) = B \cup A$. If B is Q-bad, then we have

$$\bigcup_{i=1}^{d-k} \phi(a_i) = \bigcup_{i=1}^{d-k}(J \cap \text{Pole}(a_i)) = J \cap \left(\bigcup_{i=1}^{d-k} \text{Pole}(a_i)\right) = J \tag{4.7}$$

as $J \subset \text{Pole}(B)$ and $\bigcup_{i=1}^{d-k} \text{Pole}(a_i) \supset \text{Pole}(B)$ by the hypothesis that B is Q-bad.

We have that the center points other than B must satisfy Equation 4.7. Notice there are finitely many ways to choose sets $\phi(a_i)$ such that $\bigcup_{i=1}^{d-k} \phi(a_i) = J$. Therefore, consider $Y_i = \phi(a_i)$ for the above fixed A_i. We denote $Y_i = \{y_{j1}, \ldots, y_{j(n_i)}\}$ where $n_i = |Y_i|$. We see that

$$a_i \in (S_t + y) \cup (S_t + z) \cup \bigcap_{j=1}^{|Y_i|}(S_t + y_{ji}).$$

As $Y_i \cup \{y, z\} \subset J \cup \{y, z\}$ which is affinely independent, then, by Lemma 4.8, we have that

$$\left|(S_t + y) \cup (S_t + z) \cup \bigcap_{j=1}^{|Y_i|}(S_t + y_{ji})\right| \leq 2q^{d-2-|Y_i|}.$$

By Equation 4.7, $\sum_{i=1}^{d-k}|Y_i|\geq \ell$. Therefore,

$$\begin{aligned}
M_B(E) \leq 4q^{2\ell} \sum_{\substack{(Y_1,\dots,Y_{d-k})\\ \cup Y_i = J}} \prod_{i=1}^{d-k} 2q^{d-2-|Y_i|} &\leq 4q^{2\ell} \sum_{\substack{(Y_1,\dots,Y_{d-k})\\ \cup Y_i = J}} (2q^{d-2})^{d-k} \prod_{i=1}^{d-k} q^{-|Y_i|} \\
&= 4q^{2\ell} \sum_{\substack{(Y_1,\dots,Y_{d-k})\\ \cup Y_i = J}} (2q^{d-2})^{d-k} q^{-\sum_{i=1}^{d-k}|Y_i|} \\
&\leq 4q^{2\ell} \sum_{\substack{(Y_1,\dots,Y_{d-k})\\ \cup Y_i = J}} (2q^{d-2})^{d-k} q^{-\ell} \\
&= C_d q^{d^2-kd-2d+2k+\ell}
\end{aligned}$$

for some constant C_d that accounts for the number of choices for Y_i such that $\cup Y_i = J$ as well as the additional constants 4 and 2^{d-k}. We have that $q^{d^2-kd-2d+2k+\ell}$ is maximized when ℓ is maximized. As $\ell \leq d-k-1$, then

$$C_d q^{d^2-kd-2d+2k+\ell} \leq C_d q^{d^2-kd-2d+2k+d-k-1} = C_d q^{d^2-kd-d+k-1}.$$

Therefore,

$$M_B(E) \lesssim_d q^{d^2-kd-d+k-1}.$$

$\square$

Now, we use the previous lemma to prove Theorem 2.3.

Proof. As the number of bad sets of size k (i.e., the number of choices for B) is $\leq |E|^k$, then if $M_k(E)$ is the number of affinely nondegenerate prisms that do not admit a bad set of size k, then

$$M_k(E) \lesssim_d |E|^k q^{d^2-kd-d+k-1}.$$

Then, let $M(E)$ be the number of affinely nondegenerate prisms that admit no bad sets, so that $M(E) = \sum_{i=1}^{d-1} M_k(E)$. Then,

$$\begin{aligned}
M(E) = \sum_{i=1}^{d-1} M_k(E) &\leq C_d \sum_{i=1}^{d-1} |E|^k q^{d^2-kd-d+k-1} \\
&\leq C_d(d-1)|E|^{d-1} q^{d^2-(d-1)d-d+(d-1)-1} \\
&= C_d(d-1)|E|^{d-1} q^{d-2} \\
&< C_d d|E|^{d-1} q^{d-2}.
\end{aligned}$$

Recall $N_d'(E)$ is the total number of affinely nondegenerate prisms. If $d=3$ or $|E|\geq C_d q^{d-\frac{1}{d-1}}$, by Lemma 4.7, $N_d'(E) \geq C\frac{|E|^{d+2}}{q^{2d}}$. In this case, $N_d'(E) > M(E)$ whenever

$$C\frac{|E|^{d+2}}{q^{2d}} > C_d d|E|^{d-1} q^{d-2}.$$

The above equation holds whenever

$$|E|\geq C_d' q^{d-\frac{2}{3}}$$

for C_d' some constant which depends on d. When $d=3$, this is the strongest bound on $|E|$. Otherwise, if $d>3$, we required that $|E|\geq C_d q^{d-\frac{1}{d-1}}$. Therefore, E satisfying the hypothesis of Theorem 2.3, contains an affinely nondegenerate prisms that admits no bad sets, which is what we wanted to show. $\square$

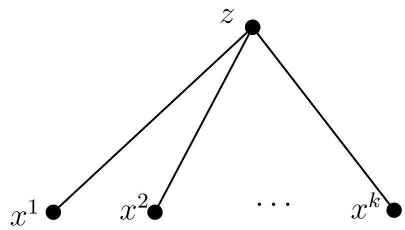

FIGURE 2. k-star in the dot-product graph $\mathcal{G}_t(E)$

5. PRELIMINARIES FOR THEOREM 2.4

We now move onto proving Theorem 2.4. The proof of this Theorem was based on work done by the authors of [1]. Much of the proof of this theorem was finalized by the seventh author of [1], Brian McDonald.

Recall the hypothesis class we are now interested in is $\mathcal{H}_t^d(E)^* = \{h_y : y \in E\}$, where

$$h_y(x) = \begin{cases} 1 & \text{if} \quad x \cdot y = t \\ 0 & \text{otherwise} \end{cases}.$$

We proceed by defining a dot-product graph, which is the analog of the distance graph in Section 3.

Definition 5.1. *The dot-product graph $\mathcal{G}_t(E)$ is a graph with vertices as points in E and edges $x \sim y$ if and only if $x \cdot y = t$.*

Then, we use the following notation to denote the degree of a vertex in the dot-product graph.

Definition 5.2. *The function $\psi : E \to \mathbb{N}$ counts the number of neighbors of $x \in E$ in the dot-product graph $\mathcal{G}_t(E)$. We have*

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

Next, we define a k-star, which becomes our new configuration of interest.

Definition 5.3. *A $(k+1)$-tuple $(y, x^1, \ldots, x^k) \subset (\mathbb{F}_q^d)^{k+1}$ is a k-star if $y \cdot x^i = t$ for $1 \leq i \leq k$. A k-star $(y, x^1, \ldots, x^k)$ is nondegenerate if each x^i is distinct. We call $\{x^1, \ldots, x^k\}$ the leaf set.*

To shatter d points, we are interested in a d-star, as we require the existence of a y such that $h_y(x^i) = 1$ for all x^i in the shattered set $\{x^1, \ldots, x^d\}$. For such a y, $y \cdot x^i = t$ for all i. Next, we define the analog of a bad set for these classifiers.

Definition 5.4. *Let $S = (y, x^1, \ldots, x^d)$ be a d-star with leaf set $L = \{x^1, \ldots, x^d\}$. Suppose $A \subset L$. We define*

$$\mathcal{Q}(A) = \{x \in E : x \cdot x^i = t \; \forall x^i \in A\}.$$

We say that a subset $A \subset L$ is bad in S if

$$\mathcal{Q}(A) \subset \bigcup_{y \in L \setminus A} \mathcal{Q}(\{y\}).$$

In general, we may refer to a hyperplane which we denote H_z:

Definition 5.5. *A hyperplane H_z is the hyperplane given by the equation $x \cdot z = t$, i.e., $H_z = \{x \in \mathbb{F}_q^d : x \cdot z = t\}$. Note that if z is in the leaf set of some k-star, then $H_z = \mathcal{Q}(\{z\})$.*

Therefore, we wish to find a d-star in E that admits no bad sets.

6. Proof of Theorem 2.4

We first show that $VCdim(\mathcal{H}_t^d(E)^\star) \leq d$ by showing $VCdim(\mathcal{H}_t^d(\mathbb{F}_q^d)^\star) \leq d$. Suppose $\mathcal{H}_t^d(\mathbb{F}_q^d)^\star$ can shatter $d+1$ points $\{x^1, \ldots, x^{d+1}\}$. We have two cases:

If $\{x^1, \ldots, x^{d+1}\}$ are linearly independent, then we have $d+1$ equations $x^i \cdot y = t$ with solution space dimension 0. Therefore, we have no y such that $h_y(x^i) = 1$ for all $1 \leq i \leq d+1$ so that $\{x^1, \ldots, x^{d+1}\}$ is not shattered.

If $\{x^1, \ldots, x^{d+1}\}$ are linearly dependent, then $x^i \in \text{Span}(\{x^1, \ldots, x^{d+1}\} \setminus \{x^i\})$. This implies that any y such that $y \cdot x^j = t$ for all $j \neq i$ also has $y \cdot x^i = t$. Therefore, $\{x^i\}$ is a bad set and thus $\{x^1, \ldots, x^{d+1}\}$ is not shattered.

So, we see that $VCdim(\mathcal{H}_t^d(E)^\star) \leq VCdim(\mathcal{H}_t^d(\mathbb{F}_q^d)^\star) \leq d$. The remainder of the proof is dedicated to finding a set of size d that is shattered by $\mathcal{H}_t^d(E)^\star$.

We first note the following result, which comes Theorem 2.1 in [2], which was adapted from [5].

Lemma 6.1. *We have* $|\{(x,y) \in E^2 : x \cdot y = t\}| = \frac{|E|^2}{q} + O\left(q^{\frac{d-1}{2}}|E|\right)$.

Using this lemma, we prove the following lower bound on the number of non-degenerate k-stars.

Lemma 6.2. *Let $N_k(E)$ be the number of non-degenerate k-stars in E, i.e.,*

$$N_k(E) = |\{(y, x^1, \ldots, x^k) \in E^{k+1}\} : x^i \textit{ distinct}, y \cdot x^i = t \ \forall i\}|.$$

If $|E| \geq C_k q^{\frac{d+1}{2}}$ for a constant C_k dependent on k, then

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

Proof. We have that

$$N_k(E) = \sum_{x \in E} k! \binom{\psi(x)}{k} = \sum_{\substack{x \in E \\ \psi(x) \geq k}} \psi(x)(\psi(x)-1) \cdot \cdots \cdot (\psi(x)-k+1).$$

Let $\phi(x) = \max(\psi(x) - k + 1, 0)$. Then,

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

We proceed by determining a lower bound for $\sum_{x \in E} \phi(x)^k$. We apply Lemma 6.1,

$$\begin{aligned}
\sum_{x \in E} \phi(x) &\geq \sum_{x \in E} \psi(x) - k + 1 = \sum_{x \in E} \psi(x) - \sum_{x \in E} (k-1) \\
&\geq \frac{|E|^2}{q} + O\left(q^{\frac{d-1}{2}}|E|\right) - |E|(k-1) \\
&\geq \frac{|E|^2}{q}\left(1 - \frac{(k-1)q}{|E|}\right) \geq 2^{-\frac{1}{k}} \frac{|E|^2}{q}
\end{aligned}$$

as $|E| \gg q$. Then, by applying Hölder's inequality with $n = |E|$, $a_i = \phi(x)$, $b_i = 1$, $r = 1$, $s = k-1$, we have

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

Therefore, using the lower bound on $\sum_{x \in E} \phi(x)$, we have

$$\frac{|E|^{2k}}{2q^k} \leq |E|^{k-1} N_k(E)$$

so that

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

$\square$

Now that we have a lower bound on the number of non-degenerate k-stars, we determine a lower bound on the number of d-stars with a linearly independent leaf set.

Lemma 6.3. *Let $\mathcal{N}_d(E)$ be the number of d-stars with linearly independent leaf set. Then, if $|E|\geq C_d q^{d-\frac{1}{d-1}}$ for q sufficiently large, we have*

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

Proof. To obtain this lower bound, we will determine an upper bound on the number of d-stars with linearly dependent leaf set $\{x^1, \ldots, x^d\}$. A linearly independent leaf set consists of a tuple of points $(x^1, \ldots, x^{d-1})$ and a point $x^i \in \text{Span}(\{x^1, \ldots, x^{d-1}\})$, up to reordering.

For a given $y \in E$, there are $\psi(y)$ points $x \in E$ such that $x \cdot y = t$. Then, there are $\psi(y)^{d-1}$ choices for $(x^1, \ldots, x^{d-1})$. And, we must have

$$x^d \in \text{Span}(\{x^1, \ldots, x^{d-1}\}) \cap \{x \in E : x \cdot y = t\}.$$

Notice that $\text{Span}(\{x^1, \ldots, x^{d-1}\})$ and $\{x \in E : x \cdot y = t\}$ are distinct hyperplanes, as $\mathbf{0} \in \text{Span}(\{x^1, \ldots, x^{d-1}\})$ but $\mathbf{0} \notin \{x \in E : x \cdot y = t\}$. Furthermore, these hyperplanes have nonempty intersection as they both contain x^d. Therefore, we see that as x^d must lie on a $(d-2)$-dimensional subspace, so that there are at most q^{d-2} choices for x^d. Then, accounting for reordering, we have that the number of stars with linearly dependent leaf set is

$$d!\, q^{d-2} \sum_{y \in E} \psi(y)^{d-1} \leq d!\, q^{d-2} q^{(d-2)(d-1)} \sum_{y \in E} \psi(y)$$

where we use that $\psi(y) \leq q^{d-1}$ as for a given y, the equation $x_1 y_1 + \cdots + x_d y_d = t$ has q^{d-1} solutions for a fixed t. Then,

$$\begin{aligned} d!\, q^{d-2} q^{(d-2)(d-1)} \sum_{y \in E} \psi(y) &= d!\, q^{d-2} q^{(d-2)(d-1)} \left(\frac{|E|^2}{q} + O\left(q^{\frac{d-1}{2}} |E|\right)\right) \\ &\lesssim d!\, q^{d(d-2)} \frac{|E|^2}{q}. \end{aligned}$$

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

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

Therefore, using the result of Lemma 6.2,

$$\mathcal{N}_d(E) \geq \frac{|E|^{d+1}}{2q^d} - \frac{|E|^{d+1}}{6q^d} = \frac{|E|^{d+1}}{3q^d}.$$

$\square$

In our following lemma, for a given bad set $B = \{b_1, \ldots, b_k\}$, we find a maximal linearly independent subset of $\mathcal{Q}(B)$.

Lemma 6.4. *Suppose $B = \{b_1, \ldots, b_k\}$ is bad in some star $S = (y, x^1, \ldots, x^d)$, with $|\mathcal{Q}(B)|> q^{r-1}$. Then, for each $y \in \mathcal{Q}(B)$, there exists a $J \subset \mathcal{Q}(B)$ with $|J|= r$ such that $\{y\} \cup J$ is linearly independent.*

Proof. Fix a $b \in B$. Then, $x \in \mathcal{Q}(B)$ implies that $x \cdot b = t$. Therefore,

$$\mathcal{Q}(B) \subset H_b,$$

where H_b is the hyperplane given by $x \cdot b = t$. Let J be a maximal subset of $\mathcal{Q}(B)$ such that $\{y\} \cup J$ is linearly independent. Then, any $z \in \mathcal{Q}(B)$ must have $\{y, z\} \cup J$ linearly dependent since we chose J to be maximal. Thus, $z \in \text{Span}(\{y\} \cup J)$ so that

$$\mathcal{Q}(B) \subset \text{Span}(\{y\} \cup J). \tag{6.1}$$

Therefore, by Equations 6 and 6.1,

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

As H_b is an affine subspace and $\text{Span}(\{y\} \cap J)$ a linear subspace, then $H_b \cap \text{Span}(\{y\} \cap J)$ is an affine subspace of dimension $a \leq |J|$ and thus has q^a elements. Therefore,

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

which implies that $|J| > r-1$, i.e., $|J| \geq r$. □

In our final lemma, we bound above the number of d-stars with linearly independent leaf set that admit a bad set of size k.

Lemma 6.5. *If $B = \{b_1, \ldots, b_k\} \subset E$ with $1 \leq k \leq d-1$ is bad in at least one d-star in E, then B is bad in at most $C'_d q^{d^2-kd-d+k}$ d-stars in E for some C'_d only depending on d.*

Proof. Suppose that B is bad in a d-star $\mathcal{S}$. We will bound the number of ways B can be extended to d-stars.

Consider an arbitrary d-star $\mathcal{S}' = (y, b_1, \ldots, b_k, x_{k+1}, \ldots, x_d)$ with leaf set L in which B is bad. We note that we can consider a d-star of this form as any d-star in which B is bad is of the form $\mathcal{S}'$ up to reordering of the leaf set.

First, note that if $|\mathcal{Q}(B)| = 1$, then there is one choice for y such that $y \cdot x = t$ for all x in the leaf set of a prism. Then, as the choices for the remaining points in the leaf set lie on the hyperplane determined by y, then there are $(q^{d-1})^{(d-k)} = q^{d^2-kd-d+k}$ choices for the remaining points in the leaf set, so that B is bad in at most $q^{d^2-kd-d+k}$ d-stars in E.

As $|\mathcal{Q}(B)| \geq 1$, then let ℓ be the smallest positive integer such that $|\mathcal{Q}(B)| \leq q^{\ell}$. Then, there are $\leq q^{\ell}$ choices for the point y in $\mathcal{S}'$. Fix such a $y \in \mathcal{Q}(B)$.

By choice of ℓ, $|\mathcal{Q}(B)| > q^{\ell-1}$. By Lemma 6.4, there exists a $J \subset \mathcal{Q}(B)$ such that $\{y\} \cup J$ is linearly independent and $|J| = \ell$. We define $\phi : E \setminus B \to \mathcal{P}(J)$, the power set of J, by

$$\phi(x) = J \cap \mathcal{Q}(\{x\}).$$

If $\{x_{k+1}, \ldots, x_d\} = L \setminus B$, since B is bad in $\mathcal{S}'$, we must have

$$Q(B) \subset \bigcup_{i=k+1}^{d} Q(\{x_i\})$$

so that

$$J = \bigcup_{i=k+1}^{d} J \cap Q(\{x_i\}) = \bigcup_{i=k+1}^{d} \phi(x_i).$$

Consider any $Z \subset J$. For all $x \in E$ with $\phi(x) = Z$, as $\phi(x) = Z \subset \mathcal{Q}(\{x\})$, then $x \cdot z = t$ for all $z \in Z$. Furthermore, for x to be in the leaf set of $\mathcal{S}'$, we require $x \cdot y = t$. Therefore, we have a set of $|Z|+1$ equations which are linear in the coordinates of x. As $\{y\} \cup J$ is linearly independent, $\{y\} \cup Z$ is linearly independent as $Z \subset J$. This implies that the $|Z|+1$ equations which are linear in the coordinates of x have a solution space with dimension $d - (|Z|+1) = d - 1 - |Z|$. So, there are $\leq q^{d-1-|Z|}$ choices for x in the leaf set of $\mathcal{S}'$ with $\phi(x) = Z$.

Therefore, we count the number of d-stars with B a bad set by summing over the number of ways to partition the set J into $(Z_1, \ldots, Z_{d-k})$ so that $\cup Z_i = J$. We have the number of d-stars with B a bad set is bounded above by

$$\begin{aligned} q^{\ell} \sum_{\substack{(Z_1,\ldots,Z_{d-k}) \\ \cup Z_i = J}} \prod_{i=1}^{d-k} q^{d-1-|Z_i|} &= q^{(d-1)(d-k)+\ell} \sum_{\substack{(Z_1,\ldots,Z_{d-k}) \\ \cup Z_i = J}} \prod_{i=1}^{d-k} q^{-|Z_i|} \\ &= q^{d^2-kd-d+k+\ell} \sum_{\substack{(Z_1,\ldots,Z_{d-k}) \\ \cup Z_i = J}} q^{-\sum_{i=1}^{d-k}|Z_i|}. \end{aligned}$$

As $\cup Z_i = J$, then $\sum_{i=1}^{d-k}|Z_i|\geq |J|= \ell$. So,

$$q^{d^2-kd-d+k+\ell} \sum_{\substack{(Z_1,\dots,Z_{d-k})\\ \cup Z_i=J}} q^{-\sum_{i=1}^{d-k}|Z_i|} \leq q^{d^2-kd-d+k+\ell} \sum_{\substack{(Z_1,\dots,Z_{d-k})\\ \cup Z_i=J}} q^{-\ell} \leq C''_d q^{d^2-kd-d+k},$$

where C''_d accounts for the number of ways to partition the set J into $(Z_1, \dots, Z_{d-k})$ so that $\cup Z_i = J$. Then, if $C'_d \geq d!\, C''_d$ to account for the number of ways to order the leaf set to include the bad set B, then we see that B is bad in at most $C'_d q^{d^2-kd-d+k}$ d-stars in E. □

We have the following Corollary, which follows directly from our lemma above.

Corollary 6.6. *Let $M_k(E)$ be the number of d-stars that have a bad set of size k. Then, $M_k(E) \leq C'_d|E|^k q^{d^2-kd-d+k}$.*

Proof. From Lemma 6.5, a given bad set $B = \{b_1, \dots, b_k\}$ of size k is bad in at most $C'_d q^{d^2-kd-d+k}$ d-stars in E. As there are at most $|E|^k$ ways to choose B a bad set of size k in E, then the desired inequality follows. □

Using these results, we are able to proof Theorem 2.4.

Proof. Let $M(E)$ be the number of d-stars that admit a bad set of any size. Then,

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

As $|E|> q^{d-1}$, then $|E|^k q^{d^2-kd-d+k} \leq |E|^{d-1}q^{d-1}$, so that

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

Then, note that if $|E|\geq C_d q^{d-\frac{1}{2}}$, then

$$M(E) \leq C'_d(d-1)|E|^{d-1}q^{d-1} < \frac{|E|^{d+1}}{3q^d} = \mathcal{N}_d(E).$$

which ensures the existence of a d-star with linearly independent leaf set that admits no bad sets. As Lemma 6.3 requires $|E|\geq C_d q^{d-\frac{1}{d-1}}$, we adopt this tighter restriction. □

7. Extension to intersection of m spheres

We now define a general hypothesis class that represents the intersection of m-spheres in $E \subset \mathbb{F}_q^d$. Let $m \geq 2$. We define the hypothesis class $\mathcal{H}_t^d(E)^m = \{h_{u_1,\dots,u_m} : (u_1, \dots, u_m) \in E^m\}$ where

$$h_{u_1,\dots,u_m}(x) = \begin{cases} 1 \text{ if } ||u_i - x||= t \text{ for } i = 1, \dots, m \\ 0 \text{ otherwise} \end{cases}.$$

Then, as a direct consequence of Theorem 2.3, we have the following theorem regarding the VC-dimension of $\mathcal{H}_t^d(E)^m$:

Theorem 7.1. *If $E \subset \mathbb{F}_q^d$, $d \geq m$, and*

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

for a constant C depending only on d, then the VC-dimension of $\mathcal{H}_t^d(E)^m$ is $d - m + 2$.

This theorem is a generalization of Theorem 2.3 and states the same result when $m = 2$. Therefore, we proceed to prove this when $m \geq 3$.

Proof. As in the proof of Theorem 2.3, we start by showing that $VCdim(\mathcal{H}_t^d(E)) \leq d - m + 2$. We will show that a set of size $d - m + 3$, $\{x^1, \ldots, x^{d-m+3}\}$, cannot be shattered by $\mathcal{H}_t^d(\mathbb{F}_q^d)$. Suppose $\mathcal{H}_t^d(\mathbb{F}_q^d)$ can shatter $\{x^1, \ldots, x^{d-m+3}\}$. Then, there exists $u_1, \ldots, u_m$ such that $h_{u_1,\ldots,u_m}(x^i) = 1$ for $1 \leq i \leq d - m + 2$. So, $||x^i - u_j|| = t$ for each $i,\ 1 \leq i \leq d + 1$ and $j,\ 1 \leq j \leq m$. We have two cases.

If $\{x^1, \ldots, x^{d-m+3}\}$ is affinely independent, then by Lemma 3.10, this set lies on the intersection of at most $m - 1$ spheres, which contradicts the existence of $u_1, \ldots u_m$ above.

If $\{x^1, \ldots, x^{d-m+3}\}$ are affinely dependent, without loss of generality, let $\{x^1, \ldots, x^{n+1}\}$ be the largest affinely independent set. Then, by Lemma 3.10, these points lie on the intersection of at most $d - n + 1$ spheres. However, to guarantee that $\{x^1, \ldots, x^{n+1}\}$ is not a bad set, we need this set to lie on the intersection of $d - n + m$ spheres. Thus, we have a contradiction. In either case, $VCdim(\mathcal{H}_t^d(\mathbb{F}_q^d)) \leq d$ which implies $VCdim(\mathcal{H}_t^d(E)) \leq d$. The remainder of the proof is dedicated to finding a set of d points that is shattered by $\mathcal{H}_t(E)$.

The remainder of the proof of this Theorem follows almost immediately from the proof of Theorem 2.3. In the proof of Theorem 2.3, we proved the existence of an affinely nondegenerate d-prism that admits no bad sets in E (asymptotically). Let $P = (y, z, x^1, \ldots, x^d)$ be such a prism.

Consider the $(d-1)$-prism defined by $P' = (y, z, x^1, \ldots, x^{d-m+2})$. Then, as P admits no bad sets, then there exists a $w_1, \ldots, w_{m-2} \in E$ such that $\{w_1, \ldots, w_{m-2}\} \subset \text{Pole}(\{x^1, \ldots, x^{d-m+2}\})$ where $y, z, w_1, \ldots, w_{m-2}$ are distinct. Note that we obtain such w_i by choosing

$$w_i \in \text{Pole}(\{\{x^1, \ldots, x^{d-i}\}) \setminus \text{Pole}(\{x^{d-i+1}, \ldots, x^d\}).$$

This method of choosing $w_1, \ldots, w_{m-2}$ ensures that such w_i are distinct. Then, we see that $h_{y,z,w_1,\ldots,w_{m-2}}$ is valued 1 on $\{x^1, \ldots, x^{d-m+2}\}$.

To show the existence of functions $h \in \mathcal{H}_t^d(E)^m$ that yield 1 on a subset A of $\{x^1, \ldots, x^{d-m+2}\}$ and 0 on the remaining points, it suffices to find a v such that $v \in \text{Pole}(A)$ but $v \notin \text{Pole}(A')$ where $A' = \{x^1, \ldots, x^{d-m+2}\} \setminus A$. Then, together with $y, z, w_2, \ldots, w_{m-2}$, we will have $h_{y,z,v,w_2,\ldots,w_{m-2}}(x) = 1$ for $x \in \{x^1, \ldots, x^{d-m+2}\}$ if and only if $x \in A$.

As $A \subset \{x^1, \ldots, x^{d-m+2}\} \subset \{x^1, \ldots, x^d\}$, then since P is an affinely non-degenerate d-prisms that admits no bad sets, then A is not bad in P. Thus, there exists a u such that $u \in \text{Pole}(A) \setminus \text{Pole}(\{x^1, \ldots, x^d\} \setminus A)$. Take $v = u$. Then, $v \in \text{Pole}(A)$ but $v \notin \text{Pole}(A')$ as $\text{Pole}(A') \subset \text{Pole}(\{x^1, \ldots, x^d\} \setminus A)$.

This completes the proof. □

Therefore, we see that with the same restrictions on the size of the subset E, we can obtain the desired configuration which has a set of size $d - m + 2$ which is shattered by $\mathcal{H}_t^d(E)^m$. It is natural to ask whether the exponent correlated with the size of E can be improved. Using the methodology presented in this paper, it is unlikely that this exponent can be improved. There are a couple ways that these methods can be applied beyond the classifiers for the intersection of two spheres.

1. The first approach would be to look for a new point configuration $(w_1, \ldots, w_m, x^1, \ldots, x^{d-m+2})$ entirely. In the proof of the above theorem, we used the count of the number of nondegenerate d-prisms from Lemma 6.2. However, our structure of interest is a sub-structure of a d-prism. Therefore, counting d-prisms will lead to a poorer bound on the count for our point configuration of interest, as we are operating under constraints imposed by extra points in the center of the d-prism. However, we were not able to determine an intuitive method of counting such point configurations, nor were we able to find helpful bounds in the literature. Note that the new point configuration of interest is no longer composed of paths of length two between two points.

2. A second approach would be to count the number of $(d-m+2)$-prisms. For a given $(d-m+2)$-prism with center $\{x^1, \ldots, x^{d-m+2}\}$, this would provide a configuration with two points y, z such that $(S_t + y) \cap (S_t + z) \supset \{x^1, \ldots, x^{d-m+2}\}$. Therefore, it remains to find points $w_3, \ldots, w_m$ such that $\{x^1, \ldots, x^{d-m+2}\} \subset S_t + w_i$ for $1 \leq i \leq m$. It is difficult to employ this restriction after counting the number of $(d - m + 2)$-prisms as in Lemma 6.2 since the existence of $w_3, \ldots, w_m$ depends on $|E \cap \text{Pole}(\{x^1, \ldots, x^{d-m+2}\})|$, and we have not guaranteed the existence of such poles in E.

8. Conclusion and Future Work

In this paper, we explore an interesting VC-dimension problem for two different hypothesis classes over $\mathbb{F}_q^d$, which transformed the problem of shattering a set of points into a problem of finding a specific point configuration within a subset of $\mathbb{F}_q^d$. Namely, our results imply that for $E \subset \mathbb{F}_q^d$ of a certain size restriction, the VC-dimension of $\mathcal{H}_t^d(E)$ and $\mathcal{H}_t^d(E)^\star$ match that of $\mathcal{H}_t^d(\mathbb{F}_q^d)$ and $\mathcal{H}_t^d(\mathbb{F}_q^d)^\star$ respectively. We see that for q large, i.e., analyzing these results asymptotically, $|E| \ll q^d = |\mathbb{F}_q^d|$, as the exponent on q in the size of E is smaller than d. So, consider q large, our results express that smaller subsets of vector space retain the complexity of the whole vector space.

We recall one of the motivations for this paper, which was the paper by Fitzpatrick, Iosevich, Wyman, and McDonald [6] that investigated $\mathcal{H}_t^d(E)' = \{h_y : y \in E\}$, where

$$h_y(x) = \begin{cases} 1 & \text{if } ||y - x|| = t \\ 0 & \text{otherwise} \end{cases}.$$

In particular, they focused on the case $d = 2$ and found for $|E| \geq Cq^{15/8}$, the $VCdim(\mathcal{H}_t^2(E)) = VCdim(\mathcal{H}_t^2(\mathbb{F}_q^2)) = 3$. However, the analogous theorem for a general dimension d remains an open problem. As in [1], we remark that the methodology in this paper cannot be used to determine a bound on the size of E so that $\mathcal{H}_t^d(E)'$ shatters $d+1$ points. Finding an affinely non-degenerate $(d+1)$-prism that admits no bad sets would imply that we are able to shatter $d+1$ points with the classifiers $\mathcal{H}_t^d(E)$, which contradicts $VCdim(\mathcal{H}_t^d(E)) \leq d$. If instead we looked at the distance analog of a $(d+1)$-star, which is of the form $(y, x^1, \ldots, x^{d+1})$ with $||y - x^i|| = t$ for $1 \leq i \leq x^{d+1}$, our proof breaks down when bounding the number of affinely nondegenerate prisms that do not admit a bad set of size d. This is because by Lemma 4.8,

$$\left| \bigcap_{i=1}^{d} (S_t + x^i) \right| \leq 2.$$

Therefore, if we wanted an analogous result of Lemma 4.10, we require a positive integer ℓ such that $|\text{Pole}(B)| > 2q^{\ell-1}$, which does not exist for the set $B = \{x^1, \ldots, x^d\}$. So, asking this question for the hypothesis class $\mathcal{H}_t^d(E)'$ introduces several difficulties.

In general, future directions would include further exploring the VC-dimension of $\mathcal{H}_t^d(E)'$, the spherical classifiers. Furthermore, as mentioned earlier, other techniques can be developed to explore the lower bound on $|E|$ so that $\mathcal{H}_t^d(E)^m$ has VC-dimension $d - m + 2$ as to try to improve the exponent on the size of such subsets E.

9. Acknowledgements

I would first like to thank my advisor, Professor Iosevich, for his support and dedication to this research that made this thesis possible. I would also like to thank my peers, Ruben Ascoli, Justin Cheigh, Ryan Jeong, Xuyan Liu, Brian McDonald, Wyatt Milgrim, Steven J. Miller, Francisco Romano Acosta, and Santiago Velazquez Iannuzelli, who I collaborated with on the ideas summarized in this paper. Lastly, I would like to thank my other committee members, Professor Gonek and Professor Wyman for the endless academic support during my undergraduate career.

References

[1] R. Ascoli, L. Betti, J. Cheigh, A. Iosevich, R. Jeong, X. Liu, B. McDonald, W. Milgrim, S. J. Miller, F. R. Acosta, and S. V. Iannuzzelli. VC-dimension and distance chains in $\mathbb{F}_q^d$, 2022.

[2] R. Ascoli, L. Betti, J. Cheigh, A. Iosevich, R. Jeong, X. Liu, B. McDonald, W. Milgrim, S. J. Miller, F. R. Acosta, and S. V. Iannuzzelli. VC-dimension of hyperplanes over finite fields. Forthcoming, 2023.

[3] M. Bennett, D. Hart, A. Iosevich, J. Pakianathan, and M. Rudnev. Group actions and geometric combinatorics in $\mathbb{F}_q^d$. *Forum Mathematicum*, 29, 2013.

[4] M. Bennett, J. Chapman, D. Covert, D. Hart, A. Iosevich, and J. Pakianathan. Long paths in the distance graph over large subsets of vector spaces over finite fields. 2014. URL https://arxiv.org/pdf/1406.0107.pdf.

[5] D. Covert, D. Hart, A. Iosevich, D. Koh, and M. Rudnev. Generalized incidence theorems, homogeneous forms, and sum-product estimates in finite fields, 2008.

[6] D. Fitzpatrick, A. Iosevich, B. McDonald, and E. Wyman. The VC-dimension and point configurations in $\mathbb{F}_q^2$. 2021. URL https://arxiv.org/abs/2108.13231.

[7] N. Grand, M. Juvekar, and M. Sun. VC dimension and configurations in $\mathbb{F}_q^d$. https://people.math.rochester.edu/faculty/iosevich/TripodsPresentationsErdos.pdf, 2021.

[8] S. Shalev-Shwartz and S. Ben-David. *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press, 2014.

APPENDIX A. VC-DIMENSION OF HYPOTHESIS CLASS OF SPHERES

Again, the following method was introduced by Nathanael Grand, Mandar Juvekar, and Maxwell Sun in the 2021 Tripods REU [7]. We show that $VCdim(\mathcal{H}) = d + 1$. Without loss of generality, suppose that $t = 1$. We work with $t = 1$ so that our algebraic manipulations involve looking at the unit sphere centered at 0, S_1, and considering standard basis vectors e_j with the jth component 1 and the other components 0. Notice that such an argument can be generalized to any t by working with a sphere of radius t, S_t, and noting that $|S_t|= O(q^{d-1}) = |S_1|$, and replacing the standard basis vectors with a scaling by t.

As $|S_1|= O(q^{d-1})$ by Theorem 4.3, choose $a \in S_1$, $a \neq e_j$. Let $C = \{e_1, \dots, e_d\}$ and consider the set $T_a = C \cup \{a\}$. We will show there exists a choice of a so that for all subsets $C_0 \subset C$, there exists a y with $h_y(x) = 1$ for $x \in C_0$ and $h_y(x) = 0$ for $x \in T_a \setminus C_0$, and for all subsets $C_0 \cup \{a\}$, there exists a y with $h_y(x) = 1$ for $x \in C_0 \cup \{a\}$ and $h_y(x) = 0$ for $x \in C \setminus C_0$. We do so by fixing a y for each condition and showing that the set of a that do not satisfying the conditions of the provided classifier has size of order q^{d-2}. As $|S_1|= O(q^{d-1})$, this shows that asymptotically, there are a positive number of choices for our final point a for which T_a is shattered.

First, we find a y such that $h_y(x) = 0$ for all $x \in T_a$. By taking $y = 3e_1$, then $||y - e_j||\neq 1$ for all $1 \leq j \leq d$ and

$$||a - y||= ||a - 3e_1||= (a_1 - 3)^2 + \sum_{j=2}^{d} a_j^2 = 10 - 6a_1$$

so that $10 - 6a_1 = 1$ if and only if $6a_1 = 9$. We notice that this is a hyperplane (an affine subspace of dimension $d - 1$), and thus as a lies on S_1, then by Lemma 4.4, there are at most $O(q^{d-2})$ choices for a that satisfy the above equation. As $|S_1|= O(q^{d-2})$, choose a such that a does not lie on this hyperplane.

Next, we find y such that $h_y(x) = 1$ for $x \in C_0$ and $h_y(x) = 0$ for $x \in T_a \setminus C_0$. Consider subsets $C_0 \subset C$ such that $a \notin C_0$. Without loss of generality, suppose $C_0 = \{e_1, \dots, e_i\}$ for $1 \leq i \leq d$. Let y be such that $y_j = 2/i$ for $1 \leq j \leq i$ and $y_j = 0$ for $i + 1 \leq j \leq d$. Then, for $e_j \in C_0$,

$$||e_j - y||= (i - 1)\left(\frac{2}{i}\right)^2 + \left(1 - \frac{2}{i}\right)^2 = 1,$$

and for $e_j \notin C_0$,

$$||e_j - y||= i\left(\frac{2}{i}\right)^2 + 1 \neq 1.$$

We want to choose a such that $||a - y||\neq 1$. As previously mentioned, we consider the set of a such that $||a - y||= 1$ and show that the size of this set is small compared to $|S_1|$. We have

$$||a - y||= \sum_{j=1}^{i}\left(a_j - \frac{2}{i}\right)^2 + \sum_{j=1}^{d} a_j^2 = 1 + \frac{4}{i}\left(1 - \sum_{j=1}^{i} a_j\right)$$

which is $= 1$ if and only if $1 - \sum_{j=1}^{i} a_j = 0$. Again, this is a hyperplane (an affine subspace of dimension $d - 1$). So, by Lemma 4.4, there are $O(q^{d-2})$ such choices of a. As we did above, choose a so that a is not on this hyperplane.

Lastly, we find y such that $h_y(x) = 1$ for $x \in C_0 \cup \{a\}$ and $h_y(x) = 0$ for $x \in C \setminus C_0$. Once again, without loss of generality, suppose $C_0 = \{e_1, \dots, e_i\}$. If $i = 0$, take $y = 2a$. Then, $||2a - e_j||= 1$ if and only if $\sum_{k\neq j} 4a_k^2 + (2a_j - 1)^2 = 1$ which happens if and only if $a_j = 1$. Again, this forms a hyperplane so that $||2a - e_j||= 1$ holds for $O(q^{d-2})$ choices of a. Choose a so that a does not lie on this hyperplane.

Now, we proceed having chosen a so that a lies on neither hyperplane specified by the above conditions. This was possible as there is asymptotically a positive number of choices for a that does not lie on these hyperplanes.

If $i = d$, then take y to be the origin, which satisfies $h_y(x) = 1$ for $x \in C_0 \cup \{a\}$ as all of these points lie on the unit sphere. Otherwise, take y with the following:

$$y_j = \begin{cases} \frac{2a_{i+1}^2}{\left(1-\sum_{j=1}^i a_j\right)^2 + ia_{i+1}^2} & \text{if } 1 \le j \le i \\ \frac{2a_{i+1}\left(1-\sum_{j=1}^i a_j\right)}{\left(1-\sum_{j=1}^i a_j\right)^2 + ia_{i+1}^2} & \text{if } j = i+1 \\ 0 & \text{otherwise} \end{cases} .$$

Then, for $1 \le j \le i$, we have

$$\begin{aligned} ||y - e_j|| =& (i-1)\left(\frac{2a_{i+1}^2}{\left(1-\sum_{j=1}^i a_j\right)^2 + ia_{i+1}^2}\right)^2 \\ &+ \left(\frac{2a_{i+1}^2}{\left(1-\sum_{j=1}^i a_j\right)^2 + ia_{i+1}^2} - 1\right)^2 + \left(\frac{2a_{i+1}\left(1-\sum_{j=1}^i a_j\right)}{\left(1-\sum_{j=1}^i a_j\right)^2 + ia_{i+1}^2}\right)^2 \end{aligned}$$

which is $= 1$. Furthermore,

$$||y - e_{i+1}|| = i\left(\frac{2a_{i+1}^2}{\left(1-\sum_{j=1}^i a_j\right)^2 + ia_{i+1}^2}\right)^2 + \left(\frac{2a_{i+1}\left(1-\sum_{j=1}^i a_j\right)}{\left(1-\sum_{j=1}^i a_j\right)^2 + ia_{i+1}^2} - 1\right)^2 \neq 1,$$

and for $i + 2 \le j \le d$,

$$||y - e_j|| = i\left(\frac{2a_{i+1}^2}{\left(1-\sum_{j=1}^i a_j\right)^2 + ia_{i+1}^2}\right)^2 + \left(\frac{2a_{i+1}\left(1-\sum_{j=1}^i a_j\right)}{\left(1-\sum_{j=1}^i a_j\right)^2 + ia_{i+1}^2}\right)^2 + 1 \neq 1.$$

Lastly, we have

$$\begin{aligned} ||y - a|| =& \sum_{j=1}^{i}\left(\frac{2a_{i+1}^2}{\left(1-\sum_{j=1}^i a_j\right)^2 + ia_{i+1}^2} - a_j\right)^2 \\ &+ \left(\frac{2a_{i+1}\left(1-\sum_{j=1}^i a_j\right)}{\left(1-\sum_{j=1}^i a_j\right)^2 + ia_{i+1}^2} - a_{i+1}\right)^2 + \sum_{j=i+2}^{d} a_j^2 \\ =& 1 \end{aligned}$$

Therefore, there exists a y with $h_y(x) = 1$ for $x \in C_0$ and $h_y(x) = 0$ for $x \in T_a \setminus C_0$, and for all subsets $C_0 \cup \{a\}$, there exists a y with $h_y(x) = 1$ for $x \in C_0 \cup \{a\}$ and $h_y(x) = 0$ for $x \in C \setminus C_0$.