THE VC DIMENSION OF QUADRATIC RESIDUES IN
FINITE FIELDS

BRIAN MCDONALD, ANURAG SAHAY, AND EMMETT L. WYMAN

Abstract. We study the Vapnik–Chervonenkis (VC) dimension
of the set of quadratic residues (i.e. squares) in finite fields, \( \mathbb{F}_q \),
when considered as a subset of the additive group. We conjec-
ture that as \( q \to \infty \), the squares have the maximum possible
VC-dimension, viz. \( (1 + o(1)) \log_2 q \). We prove, using the Weil
bound for multiplicative character sums, that the VC-dimension
is \( \geq (\frac{1}{2} + o(1)) \log_2 q \). We also provide numerical evidence for
our conjectures. The results generalize to multiplicative subgroups
\( \Gamma \subseteq \mathbb{F}_q^\times \) of bounded index.

1. Introduction

Consider the setting of an Abelian group written additively. For any
subset, one defines the Vapnik–Chervonenkis dimension (VC-dimension
for conciseness) of \( S \) as follows.

Definition 1.1 (VC-dimension in groups). Let \((G, +)\) be an Abelian
group, and let \( S \subseteq G \) be an arbitrary subset. We say that \( S \) shatters
\( Y = \{y_1, \cdots, y_n\} \subseteq G \) if for any \( \emptyset \subseteq A \subseteq Y \), there is an \( x \in G \) such
that
\[
y_j \in A \iff y_j \in (S + x)
\]
for \( 1 \leq j \leq n \). Here, and throughout, we assume that \( |Y| = n \) so that
\( y_j \neq y_k \) for \( 1 \leq j < k \leq n \). Further, we define the VC-dimension of \( S \)
to be the cardinality of the largest set \( Y \) that is shattered by \( S \), and
we denote it by \( \text{VCdim}_G(S) \).

Readers familiar with the definition of VC-dimension in the language
of set systems or classifiers (see, e.g., [26, Chapter 6]) should note that
this is the usual notion of VC-dimension relative to the set system \( \mathcal{P}_S \)
of translates of \( S \),
\[
\mathcal{P}_S = \{S + x : x \in G\},
\]
where \( S + x = S + \{x\} = \{z + x : z \in S\} \), or equivalently relative to
the classifiers
\[
\mathcal{H}_S = \{1_{S+x} : x \in G\},
\]
where here and throughout, \( 1_S : G \to \{0, 1\} \) is the indicator function of the set \( S \). Further, a standard argument shows that \( \text{VCdim}_G(S) \leq \log_2 |G| \) for every \( S \), where \( |\cdot| \) denotes cardinality.

A class of ambient groups of interest to us are the vector spaces over finite fields – i.e., \( G = \mathbb{F}_q^d \) for prime power \( q \) and some dimension \( d \geq 1 \). A specific instance of this type was considered by the Fitzpatrick, Iosevich, McDonald, and Wyman [14], where they investigated a circle in a plane \( G = \mathbb{F}_q^2 \). Another example of this type was considered by Iosevich, McDonald, and Sun [18] where they considered \( G = \mathbb{F}_q^3 \), and sets defined using dot products.

Our objective in this paper is to investigate the VC-dimension of sets with multiplicative structure in \( \mathbb{F}_q \), when considered as subsets of the additive group \((\mathbb{F}_q, +)\). Our particular interest is in the set of quadratic residues (i.e. squares) \( S_q \subseteq \mathbb{F}_q \) given by

\[
S_q = \{ x^2 : x \in \mathbb{F}_q^\times \}.
\]

We remark at this point that our analysis can be easily adjusted to include the number 0 in \( S_q \) if the reader desires. For economy of notation, however, we adopt the point of view that 0 is “half” an element of \( S_q \), a decision which we shall explain later. Concretely, we mean that

\[
1_{S_q}(y) = \begin{cases} 
1 & \text{if } y = x^2, x \in \mathbb{F}_q^\times, \\
1/2 & \text{if } y = 0, \\
0 & \text{otherwise}, 
\end{cases}
\]

where \( 1_{S_q} \) is the indicator function of \( S_q \). This requires some modification to the meaning of shattering from Definition 1.1; for our purposes, it suffices to add the extra requirement that the \( x \in G \) such that \( A = Y \cap (S_q + x) \) satisfies \( x \notin Y \).

Another viewpoint on our problem is provided by using graph-theoretic terminology. In this article, graphs are directed, do not have multiple edges, and may have loops. We denote graphs also by the letter \( G \), but this should cause no confusion, as the the underlying vertex set of our graphs shall be Abelian groups. For any vertex \( v \in G \), the (out)-neighbourhood of \( v \) is given by

\[
N(v) = \{ w \in G : v \rightarrow w \text{ in } G \},
\]

and the VC-dimension of the set family of neighbourhoods is called the VC-dimension of the graph. This is encapsulated in the following definition.

**Definition 1.2** (VC-dimension in graphs). Let \( G = (V, E) \) be a directed graph. We say that \( E \) shatters \( Y = \{y_1, \ldots, y_n\} \subseteq G \) if for any
∅ ⊆ A ⊆ Y, there is an \( x \in G \) such that
\[
y_j \in A \iff x \rightarrow y_j \text{ in } G
\]
for \( 1 \leq j \leq n \). Further, we define the VC-dimension of \( G = (V, E) \) to be the cardinality of the largest set \( Y \) that is shattered by \( E \), and we denote it by \( \text{VCdim}_G'(E) \).

The astute reader will notice that the above generalizes Definition 1.1 as for a group \( G \) and a subset \( S \subseteq G \), if we set \( E \) to be the edge-set of the Cayley (di)graph \( \text{Cay}(G, S) \), then
\[
\text{VCdim}_G(S) = \text{VCdim}_G'(E).
\]
Thus, if we define \( \mathcal{P}_q = \text{Cay}(\mathbb{F}_q, S_q) \) to be the usual notion of the Paley (di)graph, we see that the problem we are interested in is exactly the question of the VC-dimension of \( \mathcal{P}_q \). Paley graphs are an important example of pseudorandom graphs (see, for example, the seminal work of Chung, Graham, and Wilson [9]).

VC-dimension can be thought of as a measure of complexity when viewed internally from a structure (e.g., the group structure or the graph structure). Thus, if a set \( S \subseteq G \) has high additive structure, one might expect a low VC-dimension. Conversely, if \( S \) has low additive structure, one should expect a high VC-dimension. There are numerous recent works in the combinatorics literature where the assumption of bounded VC-dimension delivers strong structure on the involved sets (see, for example, [2], [27], [15], [16], [21], [1]).

Since \( S_q \) is clearly multiplicatively structured, and since one expects that addition and multiplication in \( \mathbb{F}_q \) are only loosely correlated due to the sum-product phenomenon (see, for example, [7], [29], [25], and references therein), it is natural to expect that \( S_q \) will have the maximum possible VC-dimension, \( \log_2 q \). Furthermore, since \( \mathcal{P}_q \) is pseudorandom, it is similarly natural to expect that it would have maximum possible VC-dimension.

To state our expectations in this regard, we introduce two quantities \( \alpha_q \) and \( \beta_q \). The former is defined by
\[
\alpha_q = \frac{\text{VCdim}_{\mathbb{F}_q}(S_q)}{\log_2 q}.
\]
For the latter, we define first \( m_q \) to be the largest integer such that all subsets \( Y \subseteq \mathbb{F}_q \) of size \( |Y| \leq m_q \) are shattered by \( S_q \). Then,
\[
\beta_q = \frac{m_q}{\log_2 q}.
\]
The quantity \( m_q \) is called the testing dimension by some authors (for example [3] and [24]). It is straightforward to see that \( 0 \leq \beta_q \leq \alpha_q \leq 1 \).
Finally, we further define
\[ \overline{\alpha} = \limsup_{q \to \infty} \alpha_q, \quad \underline{\alpha} = \liminf_{q \to \infty} \alpha_q, \]
\[ \overline{\beta} = \limsup_{q \to \infty} \beta_q, \quad \underline{\beta} = \liminf_{q \to \infty} \beta_q. \]  \hspace{1cm} (1)

We can now formally state our aforementioned expectation.

**Conjecture 1.3.** Let \( \underline{\alpha} \) be as defined in (1). Then, \( \overline{\alpha} = 1 \). In particular, it follows that \( \alpha = \underline{\alpha} \) and further, that
\[ \text{VCdim}_{\mathbb{F}_q}(\mathcal{S}_q) = (1 + o(1)) \log q. \]

We are not able to prove this conjecture. However, we have partial progress, in the form of the following theorem.

**Theorem 1.4.** Let \( \underline{\beta} \) be as defined in (1). Then, \( \beta \geq 1/2 \). In particular, for any \( \epsilon > 0 \), we have that for all \( q \gg \epsilon \), and for every subset \( Y \subseteq \mathbb{F}_q \) with \( |Y| \leq (\frac{1}{2} - \epsilon) \log q \), the set of squares \( \mathcal{S}_q \) shatters \( Y \).

In particular, it is immediate from this theorem that \( \alpha \geq 1/2 \), and further that
\[ \text{VCdim}_{\mathbb{F}_q}(\mathcal{S}_q) \geq (\frac{1}{2} - \epsilon) \log q. \]

for all \( q \gg \epsilon \).

Our results in this regard follow from the broader context where \( \mathcal{S}_q \) (which is a multiplicative subgroup of index 2) is replaced by a more general multiplicative subgroup \( \Gamma \subseteq \mathbb{F}_q^\times \). In particular, recall that since \( \mathbb{F}_q^\times \) is cyclic, for every \( r | (q - 1) \), \( \mathbb{F}_q^\times \) has exactly one subgroup of index \( r \), namely the set of \( r \)th powers,
\[ \Gamma^{(r)} = \{ x^r : x \in \mathbb{F}_q^\times \}, \]
so that \( \Gamma^{(2)} = \mathcal{S}_q \). Further, as with squares, we adopt the point of view that 0 is a “fuzzy” member of \( \Gamma^{(r)} \) with weight \( 1/r \). In fact, we assume that for each multiplicative coset \( t\Gamma^{(r)} = \{ tx^r : x \in \mathbb{F}_q^\times \} \), 0 is a fuzzy member of \( t\Gamma^{(r)} \) with weight \( 1/r \). That is, for \( t \neq 0 \),
\[ 1_{t\Gamma^{(r)}}(y) = \begin{cases} 
1 & \text{if } y \in t\Gamma^{(r)}, \\
1/r & \text{if } y = 0, \\
0 & \text{otherwise}. 
\end{cases} \] \hspace{1cm} (2)

Analogous to the \( r = 2 \) case, we define
\[ \alpha^{(r)}_q = \frac{\text{VCdim}_{\mathbb{F}_q}(\Gamma^{(r)})}{\log q}, \quad \beta^{(r)}_q = \frac{m^{(r)}_q}{\log q}, \]
where now $m_q^{(r)}$ is the largest integer such that all subsets $Y \subseteq \mathbb{F}_q$ of size $|Y| \leq m_q^{(r)}$ are shattered by $\Gamma^{(r)}$. We further define
\[
\tilde{\alpha}^{(r)} = \limsup_{q \to \infty} \alpha_q^{(r)}, \quad \tilde{\alpha}^{(r)} = \liminf_{q \to \infty} \alpha_q^{(r)},
\]
\[
\tilde{\beta}^{(r)} = \limsup_{q \to \infty} \beta_q^{(r)}, \quad \tilde{\beta}^{(r)} = \liminf_{q \to \infty} \beta_q^{(r)}.
\]
(3)

Then, provided that $\Gamma^{(r)}$ is of bounded index (i.e., $r \ll 1$), we expect the natural generalization of Conjecture 1.3 to hold, and are able to prove the analogous partial progress, with $1/2$ replaced by $(\log r)/2$.

**Conjecture 1.5.** Let $\alpha^{(r)}$ be as defined in (3). Then, $\alpha^{(r)} = 1$. In particular, it follows that $\tilde{\alpha}^{(r)} = \alpha^{(r)}$ and further that

\[
\text{VCdim}_{\mathbb{F}_q}(\Gamma^{(r)}) = (1 + o(r)) \log_2 q.
\]

**Theorem 1.6.** Let $\beta^{(r)}$ be as defined in (3). Then, $\beta^{(r)} \geq (\log r)/2$. In particular, for any $\epsilon > 0$, we have that for all $q \gg r, \epsilon$, and for every subset $Y \subseteq \mathbb{F}_q$ with $|Y| \leq (1/2 - \epsilon) \log r$, the set of $r$th powers $\Gamma^{(r)}$ shatters $Y$.

It is clear that Theorem 1.4 is a corollary of this one. We now focus on this theorem, and provide an overview of its proof. Broadly speaking, we proceed by showing that $\Gamma^{(r)}$ shatters $Y = \{y_1, \ldots, y_n\}$ provided a certain expression involving the indicator functions of the multiplicative cosets $t\Gamma^{(r)}$ evaluated at $y_j - x$ is positive. Then, we Fourier-expand these indicator functions using the (multiplicative) characters on the quotient group $\mathbb{F}_q^*/\Gamma^{(r)}$. Since $\mathbb{F}_q^*$ is cyclic, so is $\mathbb{F}_q^*/\Gamma^{(r)}$, and hence everything can be written in terms of its generator $\chi_r$. Upon doing so, the task reduces to finding nontrivial bounds for sums of the form

\[
\sum_{x \in \mathbb{F}_q} \chi_r(f(x)),
\]

where $f(x)$ is a polynomial of the form
\[
f(x) = \prod_{j=1}^n (y_j - x)^{k_j},
\]
for some $0 \leq k_j < r$. At this point, we use the Weil bound for multiplicative character sums to upper bound these sums point-wise, which delivers the result. This method is uniform in the choice of $Y$ provided $|Y| = n$ is small enough in terms of $q$, which is what enables us to control the quantity $\tilde{\beta}^{(r)}$. We remark that an averaging argument that is elementary (relative to the Weil bound, at least) can be used to prove
the weaker result that $\alpha \geq 1/2$. It seems that breaking the 1/2-barrier here would require new insights.

After we proved our theorem, we discovered that the key idea of applying bounds on complete character sums to show pseudorandomness of the quadratic residues is very, very well known; it was already known to Davenport in 1939 [12], though, of course, he did not have access to the Weil bound via the Riemann hypothesis then, and had to resort to zero-free regions for the associated $L$-functions instead. The reader may peruse [8], [17], [6], [5], [20], [23], [10], [11], [13] for a selection of such applications similar to ours, many of which appear to be independent of each other. See also, [22] for a recent example in the literature which uses similar ideas.

In Section 4, we present numerical evidence towards Conjecture 1.3. We restrict our attention there to quadratic residues in prime fields, i.e., $q$ prime and $r = 2$. Our programs were written in the Julia language [4] and are available with their outputs on Dropbox[1] and GitHub[2]. A crucial observation of independent interest that underlies our code is the following:

**Proposition 1.7.** The property of being shattered by $S_q$ is translation-dilation invariant. In other words, for any $Y \subseteq \mathbb{F}_q$, $a \in \mathbb{F}_q^\times$ and $b \in \mathbb{F}_q$,

$$Y \text{ is shattered by } S_q \iff aY + b \text{ is shattered by } S_q,$$

where $aY + b = \{ay + b : y \in Y\}$.

It is clear that this will save time when computing the VC dimension of $S_q$.

We considered three numerical experiments:

1. (Section 4.2) A direct computation of the VC-dimension of $S_q$ for prime values of $q$ going up to 300.
2. (Section 4.3) A computation of the size of the largest shattered arithmetic progression in $\mathbb{F}_q$ for prime values of $q$ going up to 200000.
3. (Section 4.4) A set of plots of the likelihoods that a random set whose size is a given proportion of the theoretical maximum VC-dimension, $\log_2 q$, is shattered for prime values

$$58 \approx 2^{5/0.85} \leq q \leq 2^{12/0.7} \approx 144716.$$

While we are not able to compute experiment (1) out to large primes, we do verify that the VC-dimension of $S_q$ is $\lfloor \log_2 q \rfloor - 1$ for about 57%
of the primes $5 \leq q \leq 300$, while the remainder have the maximum theoretical VC-dimension $\lfloor \log_2 q \rfloor$ (see Figure 1). In fact, by cutting the search short upon reaching a shattered set of size $\lfloor \log_2 q \rfloor - 1$, we are able to verify the VC-dimension of $S_q$ is at least $\lfloor \log_2 q \rfloor - 1$ for primes up to 512. Perhaps $\text{VCdim} S_q < \lfloor \log_2 q \rfloor - 1$ for some $q$, but it is not possible to tell at the small scales at which our experiment runs. If this happens, it likely happens for $q$ just larger than a power of 2, such as 257 or 521. In fact, the search for a shattered set of the desired size takes quite a bit longer for such primes.

For experiment (2), the length of the longest shattered arithmetic progression, in general, varies between $\frac{1}{2} \log_2 q$ and $\log_2 q$, but certainly seems to concentrate around $\frac{3}{4} \log_2 q$, as indicated in Figure 2.

The third experiment seems to indicate that, probabilistically speaking, most subsets of size $\lfloor 0.7 \log_2 q \rfloor$ are shattered by $S_q$, while most subsets of size $\lfloor 0.85 \log_2 q \rfloor$ are not shattered. It is not clear from the data if this remains true for larger primes. (See Figure 3.)

Finally, in Section 5, we discuss some interesting open problems related to our paper.

**Notation.** For $A,B,$ and $C$ quantities that depend on various parameters with $C \geq 0$ always, we use the Vinogradov notation $A - B \ll C$ interchangeably with the Bachmann-Landau notation $A = B + O(C)$ to mean that there exists a constant $K > 0$ such that $|A - B| \leq KC$. We
allow the constant $K$ to depend on some parameters, in which case such dependence is denoted by subscripts such as $A \ll_{\epsilon} C$ or $A = O_{\epsilon}(C)$.

Acknowledgements. BM is partially supported by Alex Iosevich’s NSF grant DMS-2154232. AS received summer support from the same grant in 2022. ELW is partially supported by NSF grant DMS-2204397, and by the AMS-Simons travel grants.

All three authors would like to thank Alex Iosevich for encouragement and useful discussions. AS would like to thank Will Sawin and Igor Shparlinski for useful discussions on character sums. Finally, AS would also like to thank Joe Neeman for useful discussions on VC-dimensions in graphs.

2. Preliminaries from Fourier analysis and algebraic geometry

We review some basic facts about Fourier analysis on finite Abelian groups which may be found in any relevant textbooks, such as [28] or [30]. For a finite Abelian group $G$ written multiplicatively, and a subgroup $H \subseteq G$, the characters in $\widehat{G/H}$ satisfy the orthogonality
relation for $x, t \in G$

$$\frac{1}{|G/H|} \sum_{\chi \in \hat{G}/H} \chi(x)\overline{\chi(t)} = \begin{cases} 1 & \text{if } x \in tH, \\ 0 & \text{otherwise.} \end{cases}$$

where for any finite Abelian group $G$, $\hat{G}$ denotes the dual group of its characters $\chi : G \to S^1$, and where here we are abusing notation and writing $\chi(x)$ for $\chi(xH)$. Specializing to $G = \mathbb{F}_q^\times$ and $H = \Gamma(r)$, this reads

$$1_{\Gamma(r)}(x) = \frac{1}{r} \sum_{\chi \in \mathbb{F}_q^\times/\Gamma(r)} \chi(x)\overline{\chi(t)},$$

for $x, t \neq 0$. We now extend the domain of $\chi$ to 0, by defining $\chi(0) = 0$ for nontrivial characters and $\chi_0(0) = 1$ for the trivial character $\chi_0$. If $q$ is a prime and $r = 2$, this is consistent with the definition of the Legendre symbol, which is partially the motivation for our convention. Further, with the fuzzy notion of set membership encoded in (2), the above relation then holds for all $x \in \mathbb{F}_q, t \in \mathbb{F}_q^\times$. Hopefully, this justifies our non-standard choices in this regard.

As remarked in the introduction, since $\mathbb{F}_q^\times$ is cyclic, $\mathbb{F}_q^\times/\Gamma(r) = \langle \chi_r \rangle$ where $\chi_r$ is a (not necessarily unique) character that has order $r = |\mathbb{F}_q^\times/\Gamma(r)|$ and hence generates the dual group. Thus, one can write the above as

$$1_{\Gamma(r)}(x) = \frac{1}{r} \left( 1 + \sum_{k=1}^{r-1} \chi_r^k(x)\overline{\chi_r^k(t)} \right),$$

where we have separated the trivial character.

We will use (as a black box) a deep result from algebraic geometry, namely the Weil bound for character sums.

Lemma 2.1 (Weil). For $r \geq 2$, suppose that $f \in \mathbb{F}_q[x]$ has $n$ distinct roots and that $f$ is not an $r$th power. Then, we have

$$\left| \sum_{x \in \mathbb{F}_q} \chi_r(f(x)) \right| \leq (n - 1)\sqrt{q}.$$
Lemma 2.2. Let \( r \geq 2, n \geq 1, Y = \{y_1, \cdots, y_n\} \subseteq \mathbb{F}_q \) and \( t = (t_1, \cdots, t_n) \in (\mathbb{F}_q^\times)^n \). Then,

\[
\Pr_{x \in \mathbb{F}_q}[y_j - x \in t_j \Gamma^{(r)}, 1 \leq j \leq n] = \frac{1}{r^n} + O\left(\frac{n}{q^{1/2}}\right),
\]

where the implicit constant can be chosen to be 1 and hence is uniform in all parameters, and \( x \) is sampled from the uniform distribution.

Note that for a fixed \( j, y_j, t_j \),

\[
\Pr_{x \in \mathbb{F}_q}[y_j - x \in t_j \Gamma^{(r)}] = \frac{1}{r},
\]

since \( y_j - x \) is also uniform in \( \mathbb{F}_q \), and a coset of \( \Gamma^{(r)} \) has measure \( 1/r \). Thus, the above lemma can be viewed as saying that for a fixed choice of \( n \), the values of \( x - y_j \) all equidistribute in the cosets of \( \Gamma^{(r)} \) independently of each other as \( q \to \infty \). This is what we mean by the pseudorandomness of bounded index subgroups.

Proof of Lemma 2.2. For conciseness, let \( E \subseteq \mathbb{F}_q \) be the event in the lemma (namely, that \( y_j - x \in t_j \Gamma^{(r)} \) for every \( 1 \leq j \leq n \)). Further, for \( 1 \leq j \leq n \), define \( E_j \subseteq \mathbb{F}_q \) to be the event that \( y_j - x \in t_j \Gamma^{(r)} \). Then, clearly \( E = \bigcap_j E_j \), and hence

\[
1_E(x) = \prod_{j=1}^n 1_{E_j}(x) = \prod_{j=1}^n 1_{t_j \Gamma^{(r)}}(y_j - x).
\]

We now use \([4]\) to Fourier-expand each indicator function in the right-most expression, obtaining

\[
\frac{1}{r^n} \prod_{j=1}^n \left(1 + \sum_{k=1}^{r-1} \chi_r^k(y_j - x) \overline{\chi_r^k(t_j)}\right).
\]

Expanding out the product, we find that a convenient index for the resulting summand is \( k = (k_1, \cdots, k_n) \), where \( k_j \) is the choice of \( k \) for a given \( j \) taken from the inner sum. Thus,

\[
1_E(x) = \frac{1}{r^n} \left(1 + \sum_{\mathbf{k} \in [0, r-1]^{n}} \prod_{k \neq 0}^n \chi_r^{k_j}(y_j - x) \overline{\chi_r^{k_j}(t_j)}\right),
\]

where \( 0 \leq \mathbf{k} \leq r - 1 \) means that \( 0 \leq k_j \leq r - 1 \) for each \( 1 \leq j \leq n \), and \( \mathbf{k} \neq \mathbf{0} \) means that \( k_j \) is not identically 0.

Recall that \( \chi_r \) is a multiplicative character, and hence that \( \chi_r^k(t) = \chi_r(t^k) \), and further that \( \chi_r(x) \chi_r(x') = \chi_r(xx') \). In particular, if we
define
\[ b(k, t) = \chi_r \left( \prod_{j=1}^{n} t_j^{k_j} \right), \]
and
\[ f_k(x) = \prod_{j=1}^{n} (y_j - x)^{k_j}, \]
then the above can be written as
\[ 1_E(x) = \frac{1}{r^n} \left( 1 + \sum_{0 \leq k \leq r - 1 \atop k \neq 0} b(k, t) \chi_r(f_k(x)) \right). \]

Now, we find by linearity of expectation that
\[ \Pr_{x \in \mathbb{F}_q} [E] = \mathbb{E}_x[1_E(x)] = \frac{1}{r^n} + \frac{1}{r^n} \sum_{0 \leq k \leq r - 1 \atop k \neq 0} b(k, t) \mathbb{E}_x[\chi_r(f_k(x))]. \]

Here the expectation is over \( x \in \mathbb{F}_q \). Now, \(|b(k, t)| \leq 1\). Thus, by the triangle inequality,
\[ \left( \Pr_{x \in \mathbb{F}_q} [E] - \frac{1}{r^n} \right) \leq \frac{1}{r^n} \sum_{0 \leq k \leq r - 1 \atop k \neq 0} \left| \mathbb{E}_x[\chi_r(f_k(x))] \right| \]
\[ \leq \max_{0 \leq k \leq r - 1 \atop k \neq 0} \left| \mathbb{E}_x[\chi_r(f_k(x))] \right|. \]

It now suffices to show that uniformly for \( 0 \leq k \leq r - 1, k \neq 0 \), we have the bound
\[ \mathbb{E}_x[\chi_r(f_k(x))] \leq \frac{n}{q^{1/2}}. \]
This follows by multiplying the Weil bound (Lemma 2.1) by \( 1/q \) on both sides. To see this, note that \( f_k \in \mathbb{F}_q[x] \) is clearly not an \( r \)th power, since \( k_j < r \) for each \( j \), and further note that \( f_k \) has at most \( n \) distinct roots, as its roots are contained in \( Y = \{y_1, \cdots, y_n\} \). This completes the proof. \( \square \)

3. Proof of Theorem 1.6

Let \( \epsilon > 0 \), and
\[ n \leq \left( \frac{1}{2} - \epsilon \right) \log_r q. \tag{5} \]
We will first show that for \( q \) large enough in terms of \( r \) and \( \epsilon \), this will imply that
\[ \Pr_{x \in \mathbb{F}_q} [y_j - x \in t_j \Gamma^{(r)}, 1 \leq j \leq n] > 0. \tag{6} \]
uniformly in \( Y = \{ y_1, \ldots, y_n \} \) and \( t \in (\mathbb{F}_q^\times)^n \). By Lemma \[2.2\] we see that the probability above is at least
\[
\frac{1}{r^n} - \frac{n}{q^{1/2}} \geq \frac{1}{q^{1/2}} - \frac{\log_q q}{q^{1/2}}.
\]
Since \( r \ll 1 \), \( \log_q q = o(q^{2\epsilon}) \) as \( q \to \infty \). Thus,
\[
\Pr_{x \in \mathbb{F}_q} \left[ y_j - x \in t_j \Gamma^{(r)}, 1 \leq j \leq n \right] \gg q^{-1/2+\epsilon} > 0,
\]
as claimed.

To prove the theorem, it clearly suffices to show that for any set of size \( n \) satisfying \[5\], \( Y \) is shattered by \( \Gamma^{(r)} \). Let \( Y = \{ y_1, \ldots, y_n \} \). To prove that \( Y \) is shattered by \( \Gamma^{(r)} \), we need to find, for every \( \emptyset \subseteq A \subseteq Y \), an element \( x \in \mathbb{F}_q \) with the property that
\[
A = Y \cap (\Gamma^{(r)} + x).
\]

We do this as follows. Fix \( t \in \mathbb{F}_q^\times \setminus \Gamma^{(r)} \). Now, define \( t = (t_1, \ldots, t_n) \) by
\[
t_j = \begin{cases} 1 & \text{if } y_j \in A, \\ t & \text{if } y_j \notin A. \end{cases}
\]

Applying \[6\] to this choice of \( t \), we find that, since this probability is non-zero, there must be an \( x \in \mathbb{F}_q \) with the property that \( y_j - x \in t_j \Gamma^{(r)} \) for each \( 1 \leq j \leq n \). In particular, since \( t \notin \Gamma^{(r)} \), this means that \( y_j - x \notin \Gamma^{(r)} \) when \( t_j = t \) and \( y_i - x \in \Gamma^{(r)} \) when \( t_j = 1 \). Thus, \( y_j \in (\Gamma^{(r)} + x) \) if and only if \( t_j = 1 \), which by construction, happens if and only if \( y_j \in A \). Since this is true for any choice of \( A \), we have shown that \( Y \) is shattered by \( \Gamma^{(r)} \), thus obtaining Theorem \[1.6\].

4. Numerical evidence for Conjecture \[1.3\]

First, we prove Proposition \[1.7\]. Some care must be taken to resolve issues surrounding whether \( 0 \in \Gamma^{(r)} \) – for example, by recalling the modified definition of shattering proposed in the introduction.

**Proof of Proposition \[1.7\]** The invariance under translations is trivial, since the set of translates of \( S_q \) is obviously invariant under translations. A similar argument shows that if \( a \in S_q \), then \( aY \) is shattered by \( S_q \) if and only if \( Y \) is shattered by \( S_q \).

It remains to show that if \( a \in \mathbb{F}_q^\times \setminus S_q \), the \( aY \) is shattered by \( S_q \) if and only if \( Y \) is shattered by \( S_q \).

We suppose \( Y = \{ y_1, \ldots, y_n \} \) and let \( t \in \{0, 1\}^n \) be a bit-string. We say that \( Y \) realizes \( t \) if there is a translate \( x \in \mathbb{F}_q \) so that
\[
y_j \in (S_q + x) \iff t_j = 1.
\]
We see that an equivalent way to saying that $Y$ is shattered by $S_q$ is that $Y$ realizes every $t \in \{0, 1\}^n$. Let $t' = (1, 1, \ldots, 1) - t$. Then we see that

$$Y \text{ realizes } t \iff aY \text{ realizes } t',$$

since if $x \notin Y$ is the translation such that $t = 1_{Y \cap (S_q + x)}$, then $ax \notin aY$ is a translation for which $t' = 1_{aY \cap (S_q + ax)}$. If $Y$ is shattered by $S_q$, then every choice of $t \in \{0, 1\}^n$ is realized by $Y$. But $t'$ runs over all of $\{0, 1\}^n$ in that case, thus establishing that $aY$ is shattered by $S_q$, as desired.

We now discuss the three algorithms described in the introduction in order. Preceding this, however, we discuss our code that detects if a given subset is shattered $S_q$ that is common to all algorithms.

4.1. Detecting if a subset is shattered. Common to all code is a way to determine if a subset $Y \subset \mathbb{F}_q$ of size $n$ is shattered by the shifts of the squares $S_q$. This is implemented in such a way to replace loops over the elements of $\mathbb{F}_q$ with vectorized operations. To implement this, we define a $q \times n$ matrix $A$ whose entries are 1’s and 0’s such that

$$A(x, i) = \begin{cases} 
1 & \text{if } y_i - x \in S_q \\
0 & \text{otherwise}
\end{cases}$$

Note, the $i$th column of $A$ reads as the indicator function of the reflected set of squares $-S_q$ after translating by $y_i$, and the $x$th row of $A$ reads as the restriction of the shifted set of squares $x + S_q$ to the set $Y = \{y_1, \ldots, y_n\}$. Hence, if we are to initialize such a matrix in Julia, we would first initialize the reflected set of squares modulo $q$, and then set the columns of $A$ as follows.

```julia
using CircularArrays

nSq = CircularArray(zeroes(Int64, q))
for x in 0:(q-1)/2
    nSq[-x^2] = 1
end

A = Array{Int64}(undef, q, n)
for i in 1:n
    A[:,i] = nSq[Y[i]:Y[i]+q-1]
```
end

The set $Y$ is shattered by shifts of $S_q$ if and only if every length-$n$ string of 1's and 0's is realized as a row in $A$. To detect this, we read a row as if it were a number written in binary, except we read the digits in reverse order so that the first element is the 1’s digit, the second the 2’s digit, the third the 4’s digit, and so on (the reversed order has some practical benefit for the algorithm in Section 4.3). We can carry this operation out for all rows with a single matrix multiplication:

$$A = \begin{bmatrix}
1 \\
2 \\
4 \\
\vdots \\
2^{n-1}
\end{bmatrix}.$$

The set $Y$ is shattered then if every number from 0 to $2^n - 1$ appears as an entry in the resulting vector. We can get Julia to return “true” if $Y$ is shattered and “false” if $Y$ is not shattered by writing:

```plaintext
restrictions = CircularArray(falses(2^n))
for k in A * [2^i for i in 0:n-1]
    restrictions[k] = true
end
return all(restrictions)
```

4.2. Computing the VC-dimension. This experiment computes the VC-dimension of $S_q$ for primes $q$ in a user-defined range. Broadly speaking, this is accomplished by performing an exhaustive search through the subsets of $\mathbb{F}_q$ for the largest shattered by translates of $S_q$. The search is slightly smarter than brute-force, and is described below.

The code finds the VC-dimension of $S_q$ for primes $q$ which are 5 or greater. One quickly sees that the VC-dimension of $S_q$ for such $q$ is at least $2^2 - \{0, 1\}$ is always shattered. Due to Proposition 1.7, it is enough to search over only those subsets which contain 0 and 1, as one can always find an appropriate affine transformation.

The set $\{0, 1\}$ will serve as the root of the tree of subsets over which we will search. A set $Y$ on the tree will have parent $Y \setminus \max Y$ after identifying $\mathbb{F}_q$ with the integers $\{0, 1, \ldots, q - 1\}$. Note, a set $Y$ is the ancestor to exactly $q - 1 - \max Y$ generations. To search for the largest shattered set, we start with the following iterative scheme which we will refine momentarily:
(1) Determine if \( Y \) is shattered. If so, record the size of \( Y \) if it is the largest so far. If \( Y \) is shattered and has a child, go to step (2). Otherwise, go to step (3).

(2) Take the first child \( Y \cup \{\max Y + 1\} \) of \( Y \) and assign it the name \( Y \). Go to step (1).

(3) If \( Y \) has a next sibling, namely \( Y \) with \( \max Y \) removed and \( \max Y + 1 \) included, then assign it the name \( Y \) and go to step (1). Otherwise, go to step (4).

(4) If \( Y \) is not the root \( \{0, 1\} \), take its parent \( Y \setminus \max Y \) and assign it the name \( Y \) and go to step (3). If \( Y \) is the root, exit.

This iterative scheme certainly gets the job done, but it can be improved in two ways at relatively little expense.

First, we need not search through the descendants of \( Y \) if there are none whose size exceeds the size of the largest shattered set found so far. This is an easy check, since the largest descendant of \( Y \) has size \(|Y| + q - 1 - \max Y\), the size of \( Y \) plus the number of generations following it.

The second refinement uses what we call the *shattering index* of \( Y \) in the code. This is determined by first tallying the number of times each restriction is realized, taking the minimum of these numbers, and then taking the base-2 logarithm of that. We can get Julia to return the shattering index of \( Y \) by setting up the matrix \( A \) as before and writing:

```julia
restrictions = CircularArray(zeros(2^n))
for k in A * [2^i for i in 0:n-1]
    restrictions[k] += 1
end

min = minimum(restrictions)

if min > 0
    return floor(Int64, log2(min))
else
    return -1
end
```

The \(-1\) value here is used as a flag to indicate that our set is not shattered. If \( Y \) has shattering index \( k \), then any superset of size \(|Y| + k + 1\) necessarily fails to be shattered. Hence, we should only search through the descendants of sets \( Y \) for which \(|Y| + k\) exceeds the size of the largest shattered set found so far.
4.3. Size of the largest shattered arithmetic sequence. In this experiment, we compute the largest value \( n \leq \log_2 q \) such that \( Y = \{0, 1, \cdots, n-1\} \subseteq \mathbb{F}_q \) is shattered by \( \mathcal{S}_q \) for primes \( q \) in a user-defined range.

The motivation for this experiment comes from a naive guess that perhaps a good candidate for a large shattered subset of \( \mathbb{F}_q \) for \( q \) prime is a long arithmetic progression. At this point, we see that from Proposition 1.7 it follows that if an arithmetic progression of length \( n \) is shattered, then so is the set \( \{0, 1, \cdots, n-1\} \).

To find the largest \( n \) for which \( Y = \{0, 1, \cdots, n-1\} \) is shattered by \( \mathcal{S}_q \), the code proceeds by as described in Section 4.1 with \( n = \lfloor \log_2 q \rfloor \) (i.e., with the theoretical maximum value it could be). The main time-saving innovation here is that it suffices to do this computation only for this value of \( n \). This is achieved by the observation that if \( \{0, 1, \cdots, n-1\} \) is not shattered, then the restrictions matrix computed when checking already contains the information about whether \( \{0, 1, \cdots, n-2\} \) is shattered or not. Let \( R \) be this restriction matrix, which has dimensions \( 2^n \times 1 \). Let \( R_1 \) and \( R_2 \) be the \( 2^{n-1} \times 1 \) matrices such that

\[
R = \begin{bmatrix} R_1 \\ R_2 \end{bmatrix}.
\]

Now, let \( R' \) be the \( 2^{n-1} \times 1 \) matrix obtained by applying the component-wise logical-OR operation to \( R_1 \) and \( R_2 \). Then \( R' \) is the restriction matrix that is obtained when determining if \( \{0, 1, \cdots, n-2\} \) is shattered.

This is best illustrated by an example. Suppose \( n = 6 \). Then, the 3rd element of \( R \) represents whether the bit-string 110000 can be achieved as the indicator function of \( \{0, \cdots, 5\} \cap (\mathcal{S}_q + x) \) for some \( x \) as \( 1 + 2 = 3 \). Similarly, the 19th element of \( R \) represents whether the bit-string 110001 can be achieved as \( 1 + 2 + 16 = 19 \).

Thus, if we take the logical-OR of these two Boolean values, we get a bit which represents whether 11000 can be achieved as the indicator function for \( \{0, \cdots, 4\} \cap (\mathcal{S}_q + x) \) for some \( x \), as we are essentially neglecting to care what happens to 5. In particular, this will be stored in the 3rd position in \( R' \), which lines up with the fact that \( 1 + 2 = 3 \).

We see that this process is recursive, and thus, once we have computed the restriction matrix for \( n = \lfloor \log_2 q \rfloor \), we can easily find the actual largest value of \( n \) that works using a simple loop, given below:

```python
while !all(restrictions[0:2^n-1])
    restrictions[0:2^(n-1)-1]
    = restrictions[0:2^(m-1) - 1] .|
    restrictions[2^(n-1) : 2^n - 1]
```
4.4. Probabilities that a set is shattered. This experiment determines, up to statistical error, the likelihood that a random subset of $\mathbb{F}_q$ of size $n$ is shattered by $\mathcal{S}_q$. In particular, for each $n$ in a user-defined range, we draw 1000 subsets of size $n$ from $\mathbb{F}_q$ (pseudo-)randomly and uniformly and determine the proportion of them which are shattered. We then record the datum

$$\left( \frac{n}{\log_2 q}, \frac{\# \text{ shattered}}{1000} \right).$$

We generate a plot of probabilities by fixing $n$ and ranging over primes $q$ in $[2^{n/0.85}, 2^{n/0.7}]$ (see Figure 3).

Plotting the recorded data on the axes reveals a smooth-looking profile of shattering probabilities which seems to be monotone decreasing from nearly 1 to nearly 0 on the vertical axis as we range from 0.7 to 0.85 on the horizontal axis. There also seems to be an interface that simultaneously scales thinner and shift rightwards as $n$ increases. It would be interesting to investigate (1) the position of the interface as $n \to \infty$, (2) the scale of the interface as $n \to \infty$, and (3) the structure of the interface after an appropriate rescaling. We anticipate that this would require a combination of techniques from both algebraic geometry and probability. It would follow that nearly every subset in a relatively large range of sizes is shattered by $\mathcal{S}_q$, rather than just those close to $\frac{1}{2} \log_2 q$.

5. Directions for future research

The following are a list of open problems related to our results. We would be interested in a solution for any of these.

5.1. Sets which are not shattered. For any $\delta > 0$, is it always the case that for $q$ large enough, there is a set $Y = \{y_1, \ldots, y_n\}$ with $n = |Y| \geq (\frac{1}{2} + \delta) \log_2 q$ such that $\mathcal{S}_q$ does not shatter? It seems plausible that this is true, which would imply, together with Theorem 1.4 that $\beta = 1/2$. It is unclear whether the right approach to such an endeavour would be algebraic geometric or probabilistic.

The obvious analogue for $\Gamma^{(r)}$ is also interesting, and in particular, it seems unclear what the right value for $\beta^{(r)}$ must be. A concrete question of interest to us is where $\beta^{(r)} \to 0$ as $r \to \infty$; we remark that
Figure 3. Along the vertical axis is probability. Along the horizontal axis is the proportion $n/\log_2(q)$ and ranges in $[0.7, 0.85]$. From left to right starting at the top, the plots take $n$ to be at 5, ..., 12, respectively, and the series in each plot is generated as $q$ ranges such that $0.7 \leq n/\log_2 q \leq 0.85$. Primes $q$ were selected randomly so that the expected number of values of $q$ per plot is about 100.
our lower bound in Theorem 1.6 has this property, but it is not clear if the true value of $\beta^{(r)}$ satisfies this as well.

5.2. The probabilistic interface. For a fixed $q$ and $n$, let $p_{q,n}$ denote the probability that a random $Y \subseteq \mathbb{F}_q$, $|Y| = n$ selected with uniform probability is shattered by $S_q$. The experiments suggest there may be an interface over which $p_{q,n}$ goes from essentially 1 to essentially 0. To put it precisely, let

$$\gamma = \sup \left\{ t \in [0, 1] : \liminf_{q \to \infty} \inf_{n \leq t \log_2 q} p_{q,n} = 1 \right\},$$

$$\overline{\gamma} = \inf \left\{ t \in [0, 1] : \limsup_{q \to \infty} \sup_{n \geq t \log_2 q} p_{q,n} = 0 \right\}.$$

It is difficult to ascertain the behavior of $\gamma$ and $\overline{\gamma}$; we can only say for certain that

$$\beta \leq \gamma \leq \overline{\gamma} \leq \alpha.$$

It seems reasonable to conjecture that $p_{q,n}$ is monotone-decreasing with $n$, from which we would be able to conclude

$$\overline{\beta} \leq \gamma \leq \overline{\gamma} \leq \alpha.$$

It also seems likely from the experiment in Section 4.4 that the scale of the interface tends to 0 as $q \to \infty$, that is, $\gamma = \overline{\gamma}$. It would be interesting to understand the scale, position, and structure of this interface as $q \to \infty$.

5.3. VC-dimension of random graphs. Since $\mathcal{P}_q$ is a pseudorandom graph, it is a deterministic approximation to an Erdős-Rényi random graph $\text{ER}(n,p)$ with $n = q$ and $p = 1/2$. This raises the question of what the VC-dimension of an Erdős-Rényi random graph behaves like. This problem has been studied in the sparse setting as $n \to \infty$ with $p = p(n) = o(1)$ by Anthony, Brightwell, and Cooper [3]. However, there appear to be no results in the literature in the dense regime where $p$ is constant as $n \to \infty$.

An algebraic variant of the above which is perhaps of more relevance is the problem of determining the VC-dimension of a random Cayley (di)graph, $\text{RC}(G,p)$. Here, for a fixed group $G$, we pick a set $S$ where independently for each $x \in G$, we have $x \in S$ with probability $0 \leq p \leq 1$. Then, the graph we consider is $\text{Cay}(G,S)$.

One could also use other random graph models instead of $\text{ER}(n,p)$ and $\text{RC}(G,p)$ – we would be interested in any such results, as it may shed light on whether Conjecture 1.3 is correct.
References

[1] Noga Alon, Eldar Fischer, and Ilan Newman. Efficient testing of bipartite graphs for forbidden induced subgraphs. *SIAM J. Comput.*, 37(3):959–976, 2007.

[2] Noga Alon, Jacob Fox, and Yufei Zhao. Efficient arithmetic regularity and removal lemmas for induced bipartite patterns. *Discrete Anal.*, pages Paper No. 3, 14, 2019.

[3] Martin Anthony, Graham Brightwell, and Colin Cooper. The Vapnik-Chervonenkis dimension of a random graph. volume 138, pages 43–56. 1995. 14th British Combinatorial Conference (Keele, 1993).

[4] Jeff Bezanson, Alan Edelman, Stefan Karpinski, and Viral B. Shah. Julia: A fresh approach to numerical computing. *SIAM Review*, 59(1):65–98, 2017.

[5] Andreas Blass, Geoffrey Exoo, and Frank Harary. Paley graphs satisfy all first-order adjacency axioms. *J. Graph Theory*, 5(4):435–439, 1981.

[6] Béla Bollobás and Andrew Thomason. Graphs which contain all small graphs. *European J. Combin.*, 2(1):13–15, 1981.

[7] J. Bourgain, N. Katz, and T. Tao. A sum-product estimate in finite fields, and applications. *Geom. Funct. Anal.*, 14(1):27–57, 2004.

[8] F. R. K. Chung and R. L. Graham. Quasi-random subsets of $\mathbb{Z}_n$. *J. Combin. Theory Ser. A*, 61(1):64–86, 1992.

[9] F. R. K. Chung, R. L. Graham, and R. M. Wilson. Quasi-random graphs. *Combinatorica*, 9(4):345–362, 1989.

[10] Fan Chung, Ronald Graham, and Tom Leighton. Guessing secrets. *Electron. J. Combin.*, 8(1):Research Paper 13, 25, 2001.

[11] Ivan Bjerre Damgård. On the randomness of Legendre and Jacobi sequences. In *Advances in cryptology—CRYPTO ’88 (Santa Barbara, CA, 1988)*, volume 403 of *Lecture Notes in Comput. Sci.* pages 163–172. Springer, Berlin, 1990.

[12] H. Davenport. On character sums in finite fields. *Acta Math.*, 71:99–121, 1939.

[13] H. Davenport and P. Erdős. The distribution of quadratic and higher residues. *Publ. Math. Debrecen*, 2:252–265, 1952.

[14] D. Fitzpatrick, A. Iosevich, B. McDonald, and E. Wyman. The VC-dimension and point configurations in $\mathbb{F}_q^2$. *preprint, arXiv:2108.13231*, 2021.

[15] Jacob Fox, János Pach, and Andrew Suk. Erdős-Hajnal conjecture for graphs with bounded VC-dimension. *Discrete Comput. Geom.*, 61(4):809–829, 2019.

[16] Jacob Fox, János Pach, and Andrew Suk. Bounded VC-dimension implies the Schur-Erdős conjecture. *Combinatorica*, 41(6):803–813, 2021.

[17] R. L. Graham and J. H. Spencer. A constructive solution to a tournament problem. *Canad. Math. Bull.*, 14:45–48, 1971.

[18] A. Iosevich, B. McDonald, and M. Sun. Dot products in $\mathbb{F}_3$ and the Vapnik-Chervonenkis dimension. *preprint, arXiv:2203.03046*, 2022.

[19] Henryk Iwaniec and Emmanuel Kowalski. *Analytic number theory*, volume 53 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2004.

[20] Gareth A. Jones. Infinite Paley graphs. *Art Discrete Appl. Math.*, 4(2):Paper No. 2.05, 10, 2021.
[21] László Lovász and Balázs Szegedy. Regularity partitions and the topology of graphons. In An irregular mind, volume 21 of Bolyai Soc. Math. Stud., pages 415–446. János Bolyai Math. Soc., Budapest, 2010.

[22] Bernard Mans, Min Sha, Igor E. Shparlinski, and Daniel Sutantyo. Functional graphs of families of quadratic polynomials. preprint, arXiv:2208.01885, 2022.

[23] René Peralta. On the distribution of quadratic residues and nonresidues modulo a prime number. Math. Comp., 58(197):433–440, 1992.

[24] Kathleen Romanik and Carl Smith. Testing geometric objects. Comput. Geom., 4(3):157–176, 1994.

[25] M. Rudnev, G. Shakan, and I. D. Shkredov. Stronger sum-product inequalities for small sets. Proc. Amer. Math. Soc., 148(4):1467–1479, 2020.

[26] Shai Shalev-Shwartz and Shai Ben-David. Understanding machine learning: From theory to algorithms. Cambridge university press, 2014.

[27] Olof Sisask. Convolutions of sets with bounded VC-dimension are uniformly continuous. Discrete Anal., pages Paper No. 1, 25, 2021.

[28] Elias M. Stein and Rami Shakarchi. Fourier analysis: An introduction, volume 1 of Princeton Lectures in Analysis. Princeton University Press, Princeton, NJ, 2003.

[29] Terence Tao. The sum-product phenomenon in arbitrary rings. Contrib. Discrete Math., 4(2):59–82, 2009.

[30] Audrey Terras. Fourier analysis on finite groups and applications, volume 43 of London Mathematical Society Student Texts. Cambridge University Press, Cambridge, 1999.