Square root Bound on the Least Power Non-residue using a Sylvester-Vandermonde Determinant

Michael Forbes * Neeraj Kayal † Rajat Mittal ‡ Chandan Saha §

Abstract

We give a new elementary proof of the fact that the value of the least k^{th} power non-residue in an arithmetic progression $\{bn + c\}_{n=0,1\ldots}$, over a prime field $\mathbb{F}_p$, is bounded by $7/\sqrt{5} \cdot b \cdot \sqrt{p/k} + 4b + c$. Our proof is inspired by the so called *Stepanov method*, which involves bounding the size of the solution set of a system of equations by constructing a non-zero low degree auxiliary polynomial that vanishes with high multiplicity on the solution set. The proof uses basic algebra and number theory along with a determinant identity that generalizes both the Sylvester and the Vandermonde determinant.

1 Introduction

Let $\mathbb{F}_p$ be the prime field with p elements. An element $a \in \mathbb{F}_p$ is called a k^{th} power non-residue if there is no $b \in \mathbb{F}_p$ such that $b^k = a$. Bounding the value of the least k^{th} power non-residue in a prime field $\mathbb{F}_p$, where $k \mid p-1$, is a fundamental problem in number theory and algebra. It has an important application in finding roots over finite fields. For instance, it is known from the work of Vinogradov [Vin72] (see also Proposition 7 in [Evd94]) that given a k^{th} non-residue, all the k^{th} power roots of an element $a \in \mathbb{F}_p$ i.e. all x such that $x^k = a$, can be found in $(k \cdot \log p)^{O(1)}$ time. It is a major open problem in number theory to show that the least k^{th} power non-residue is bounded by $(\log p)^{O(1)}$. Indeed, such a bound is already known under the powerful assumption of the Extended Riemann Hypothesis (ERH). It follows from the work of Ankeny [Ank52] and Bach [Bac82] that assuming ERH , the value of the least k^{th} non-residue in $\mathbb{F}_p$ is bounded by $O(\log^2 p)$, where k is a prime dividing $p-1$. However, such a strong bound is not yet shown without the assumption of any unproven conjecture. We now briefly mention the known results on ERH-free bounds on least power non-residues.

1.1 Earlier work

The Pólya-Vinogradov inequality (see Chapter 23 in [Dav00]) states that

$$\left| \sum_{m+1 \le x \le m+n} \chi(x) \right| \le \sqrt{p} \cdot \log p,$$

where χ is a *non-principal* character modulo p. Taking χ to be the quadratic character, it immediately follows that the least quadratic non-residue in $\mathbb{F}_p$ is bounded by $\sqrt{p} \log p$. In

*Massachusetts Institute of Technology
†Microsoft Research India
‡Rutgers University
§Max Planck Institute for Informatics

1919, this bound was improved by Vinogradov (see [Vin54, Vin85]), who showed that the least quadratic non-residue in $\mathbb{F}_p$ is less than $p^{\frac{1}{2\sqrt{e}}} \log^2 p$. In a subsequent work, Vinogradov [Vin27] also showed that if $k \mid p-1$ and $k > m^m$, where m is an integer greater than 8, then the least k^{th} power non-residue is less than $p^{1/m}$ for all sufficiently large values of p. Later, in 1957, Burgess [Bur57] improved upon Vinogradov's result and showed that the least quadratic non-residue is in fact bounded by $p^{\frac{1}{4\sqrt{e}}+\epsilon}$ for any small enough $\epsilon > 0$. A simple account of Burgess' theorem can be found in the work of Stepanov [Ste75] (see also [Kar68]). We note that the proofs of Vinogradov and Burgess' results involve sophisticated analytic arguments on character sums. On the other hand, using purely elementary methods Brauer [Bra32] showed that the length of the largest sequence of consecutive k^{th} power residues or non-residues is bounded by $\sqrt{2p} + 2$. Later, Hudson [Hud74] gave an elementary argument to show that the value of the smallest k^{th} power non-residue in an arithmetic progression $\{bn + c\}_{n=0,1...}$ is bounded by $2^{11/4} b^{5/2} p^{2/5} + 6b^3 p^{1/5} + 2b^2$, if p is sufficiently large. Surely, these bounds are worse than the best known bounds of Burgess and Vinogradov. Nevertheless, it is perhaps interesting to know how much elementary methods can achieve in proving non-trivial bounds for power residues and non-residues.

1.2 Our results

We give a simple proof of the following fact.

Theorem 1.1. *The value of the least k^{th} power non-residue in an arithmetic progression $\{bn + c\}_{n=0,1,...}$ over $\mathbb{F}_p$ is bounded by $7/\sqrt{5} \cdot b \cdot \sqrt{\frac{p-1}{k}} + 4b + c$.*

Notice that, for $k \geq p^{1/5}$, the bound given by Theorem 1.1 is better than the bounds shown by Hudson [Hud74] and Brauer [Bra32]. Our proof is inspired by the *polynomial method*, which was introduced by Stepanov to give elementary proofs of many of the significant special cases of Weil's theorem on rational points on curves. The reader is encouraged to refer to the book by Schmidt [Sch04] for an account of the elementary methods used in studying equations over finite fields. (For a quick introduction to some of the main results in this area refer to Tao's blog entry [Tao09].)

The main idea behind Stepanov's method is to construct a non-zero auxiliary polynomial that vanishes with high multiplicity on the solution set of a system of equations. Now, if the degree of the auxiliary polynomial is also 'small' then this can be used to upper bound the size of the solution set. We use this theme of bounding a solution set size via a low-degree auxiliary polynomial to give a new elementary proof of the square root bound on the least power non-residue/residue in any arithmetic progression. But, it turns out that the only 'not so easy' part of our proof is showing that the auxiliary polynomial thus constructed is non-zero. We resolve this difficulty by using an interesting determinant identity that generalizes the determinant of both the Sylvester and the Vandermonde matrix. Proving this determinant identity constitutes the main technical contribution of our work. We hope that this identity on a generalized Sylvester-Vandermonde matrix is of independent interest and may find applications elsewhere.

2 The Polynomial Method

In this section, we describe our approach to proving Theorem 1.1. At the heart of our argument is the following lemma.

Lemma 2.1. *A system of univariate polynomials $\mathcal{S} = \{(x + a_i)^t - \theta_i\}_{1 \le i \le r}$, where $\theta_i, a_i \in \mathbb{F}_p$ and a_i's are distinct, has at most $2t/(r-1)+3$ common roots, if $r \le 2/\sqrt{5} \cdot \sqrt{t} + 1$ and $p > 2t$.*

Before we prove this lemma, let us at first see how it implies Theorem 1.1. (To keep the presentation simple, we avoid the use of the floor/ceiling notations. The analysis can be made more precise, at the cost of making the constant 4 in Theorem 1.1 and the constant 3 in Lemma 2.1 slightly worse.)

In Lemma 2.1, take $a_i = b \cdot (i-1)$, $t = (p-1)/k$ and $\theta_i = 1$ for all $1 \le i \le r$. Set $r = 2/\sqrt{5} \cdot \sqrt{t} + 1$. If the sequence of elements $bj + c, bj + c + b, bj + c + 2b, \ldots, bj + c + b(r-1)$ are k^{th} power residues then surely, $bj + c$ is a common root of the system $\mathcal{S}$. By Lemma 2.1, there are at most $2t/(r-1)+3$ common roots of $\mathcal{S}$. In the worst case, all these common roots can possibly be consecutive elements of the arithmetic progression $\{bn + c\}_{n=0,1,\ldots}$. Therefore, the first index m for which $bm + c$ is a k^{th} power non-residue, can be at most $2t/(r-1)+3+r = 7/\sqrt{5} \cdot \sqrt{t} + 4$. The same argument can be used to prove a slightly general form of Theorem 1.1, as stated in the following corollary.

Corollary 2.2. *The length of the largest sequence of consecutive k^{th} power residues or non-residues in an arithmetic progression $\{bn + c\}_{n=0,1,\ldots}$ is bounded by $7/\sqrt{5} \cdot \sqrt{p - 1/k} + 4$.*

The rest of this section and the following section (Section 3) are devoted to the proof of Lemma 2.1.

2.1 Proof of Lemma 2.1

The strategy we employ to bound the number of common roots of $\mathcal{S}$, denoted by $\nu(\mathcal{S})$ henceforth, is inspired by what is known as the 'Stepanov method' (also called the 'polynomial method'). The idea is to show the existence of a non-zero polynomial $F(x)$ of *small* degree N (say) such that if α is a common root of $\mathcal{S}$ then α is also a root of $F(x)$ with multiplicity M (say). If this happens then we immediately know that $\nu(\mathcal{S})$ can be at most $\frac{N}{M}$. By making N as small as possible and M as large as possible, we can arrive at an upper bound for $\nu(\mathcal{S})$.

Let us see how to put this idea at work. Choose $F(x)$ to be of the form,

$$F(x) = \sum_{i=1}^{r} G_i(x)(x + a_i)^{t+M-1+s}, \tag{1}$$

where G_i's are polynomials of degree at most d (say) and M is the multiplicity parameter mentioned above. The parameters d and M will be fixed eventually in terms of t and r. Define s as,

$$\begin{aligned} s &= 0, \quad \text{if } (r-1) \mid (t+M-1) \\ &= (r-1) - ((t+M-1) \mod (r-1)), \quad \text{otherwise.} \end{aligned}$$

The role of the parameter s is to make $t + M - 1 + s$ perfectly divisible by $r - 1$, a technical requirement for the analysis in Section 3 to go through. Let us take a short digression and clarify a bit more the purpose of the parameter s.

One might wonder as to why we do not assume, for the sake of simplicity, that $s = 0$ and $t + M - 1$ is divisible by $r - 1$. At some point in our argument we need to establish linear independence of a certain linear system. If the coefficient matrix associated to the linear system is a square matrix then all we need to show is that the corresponding determinant is non-zero. It turns out, it can be shown that such a determinant is non-zero by using an identity involving *derivatives* of the determinant function. Whereas, for a non-square system it is a little more tedious.

Coming back to the main flow of the proof, let us see what is required from the polynomial $F(x)$. Denote the ℓ^{th} derivative of F with respect to x by $F^{(\ell)}$ and let $T = t + M - 1 + s$. Also, $G^{(j)}$ denotes the j^{th} derivative of G. If α is a root of the system $\mathcal{S}$ then we require $F^{(\ell)}(\alpha) = 0$ for all $0 \leq \ell \leq M-1$ since we want α to be a root of F with multiplicity M. This means,

$$F^{(\ell)}(\alpha) = \sum_{i=1}^{r} \sum_{j=0}^{\ell} c_j(T) \cdot \theta_i \cdot G_i^{(j)}(\alpha) \cdot (\alpha + a_i)^{M-1+s-(\ell-j)} = 0, \tag{2}$$

where $c_j(T) = \prod_{k=0}^{\ell-j-1} (T-k)$ is a constant and $(\alpha + a_i)^t$ is evaluated to θ_i since α is a root of $\mathcal{S}$. Suppose that the coefficients of the polynomials G_i's, in Equation 1, are variables. Also, treat the expression given in Equation 2 as a polynomial in α of degree $(d + M - 1 + s - \ell)$ with coefficients as linear forms in the variables (that are coefficients of the G_i's). By equating these coefficients to zeroes, we can ensure that $F^{(\ell)}(\alpha)$ is zero. Therefore, for any particular ℓ, Equation 2 imposes $(d + M + s - \ell)$ homogeneous linear constraints, yielding a total of $M \cdot (d+s) + \frac{M(M+1)}{2}$ homogeneous equations in $(d+1) \cdot r$ variables (as ℓ runs from 0 to $M-1$). Thus, in order that we get a nontrivial solution for the coefficients of G_i's, it is sufficient to satisfy the the following condition,

$$M \cdot (d+s) + \frac{M(M+1)}{2} < (d+1) \cdot r. \tag{3}$$

Further, we also need to ensure that this solution is such that $F(x) \neq 0$. The degree of the polynomial $F(x)$ is $N = d + t + M - 1 + s$. If the number of variables $(d+1) \cdot r$ is greater than $d + t + M + s$ then surely there is a nontrivial setting of the coefficients of G_i's that makes $F(x) = 0$. However, such a situation can be possibly averted if we also put the restriction that

$$(d+1) \cdot r \leq d + t + M + s. \tag{4}$$

Indeed, we show (in Section 3) that Condition 4 is sufficient to guarantee $F(x) \neq 0$ if the coefficients of the G_i's are not all zeroes. To summarize, Condition 3 ensures that we are able to find nontrivial G_i's by solving the homogeneous linear equations arising from Equation 2, for $0 \leq \ell \leq M-1$. Whereas, Condition 4 guarantees that the polynomial $F(x)$, defined in Equation 1, is non-zero if not all the G_i's are zeroes - the proof of this appears in Section 3.

Putting together Condition 3 and 4, and using $D = d + 1$, we get the following overall condition to satisfy.

$$M \cdot (D+s) + \frac{M(M-1)}{2} < D \cdot r \leq D + t + M + s - 1.$$

Since our objective is to minimize the quantity $\frac{N}{M} = \frac{D+t+M+s-2}{M}$, we would like to minimize $D + t + M + s - 1$, which being lower bounded by $D \cdot r$, the best we could possibly do is to choose D such that,

$$\begin{aligned} D \cdot r &= D + t + M + s - 1 \\ \Rightarrow D &= \frac{t + M + s - 1}{r - 1} \end{aligned} \tag{5}$$

This setting of D satisfies Condition 4. Now, let us see how to satisfy Condition 3. Choose $M = r/2$ and put $D = (t + M + s - 1)/(r-1)$ as in Equation 5. Using the fact that $s \leq r - 1$ and then simplifying further, Condition 3 reduces to the following quadratic inequality:

$$5r^2 - 17r - (4t - 14) < 0.$$

It is easy to check that this is satisfied if $r \leq 2/\sqrt{5} \cdot \sqrt{t} + 1$. We are almost done. Recall that the maximum size of $\nu(\mathcal{S})$, the set of common solutions of $\mathcal{S}$, is bounded by N/M, where $N = \deg(F)$. Hence,

$$|\nu(\mathcal{S})| \leq \frac{N}{M} < \frac{D + t + M + s - 1}{M} = \frac{D \cdot r}{M} \quad \text{(using Equation 5)}$$

Since $M = r/2$, $|\nu(\mathcal{S})| \leq 2D$. Once again, using the value of D it is easy to derive that

$$|\nu(\mathcal{S})| \leq \frac{2t}{r-1} + 3.$$

This proves Lemma 2.1 except the lemma:

Claim 2.3. *If $D = \frac{t+M+s-1}{r-1}$ then $F(x) = 0$ if and only if $G_i(x) = 0$, for all $1 \leq i \leq r$.*

The next section is devoted to the proof of this statement. The main ingredient of the proof is an identity involving a generalized Sylvester-Vandermonde determinant. The condition "$p > 2t$" (in Lemma 2.1) also appears in this proof.

3 A Generalized Sylvester-Vandermonde Determinant

Recall, from Equation 1, that $F(x)$ is defined as $F(x) = \sum_{i=1}^{r} G_i(x) \cdot (x + a_i)^T$, where $T = t + M + s - 1$. Suppose $G_i = \sum_{j=0}^{d} c_{ij} x^j$. Then,

$$F(x) = \sum_{i=1}^{r} \sum_{j=0}^{d} c_{ij} x^j (x + a_i)^T.$$

Proving Claim 2.3 essentially means proving this: if $F(x) = 0$ then $c_{ij} = 0$, for all i and j. Suppose, on the contrary, that this is false. Then, the polynomials $\{x^j(x + a_i)^T\}_{i,j}$, for $1 \leq i \leq r$ and $0 \leq j \leq d$, are $\mathbb{F}$-linearly dependent. In other words, the following matrix,

$$\mathsf{V} = \left(\begin{array}{cccccccc}
\binom{T}{T}a_1^T & \cdots & \binom{T}{T-d}a_1^{T-d} & \cdots & \cdots & \binom{T}{0}a_1^0 & & \\
& \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \\
& & \binom{T}{T}a_1^T & \cdots & \cdots & \binom{T}{d}a_1^d & \cdots & \binom{T}{0}a_1^0 \\
\hline
\binom{T}{T}a_2^T & \cdots & \binom{T}{T-d}a_2^{T-d} & \cdots & \cdots & \binom{T}{0}a_2^0 & & \\
& \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \\
& & \binom{T}{T}a_2^T & \cdots & \cdots & \binom{T}{d}a_2^d & \cdots & \binom{T}{0}a_2^0 \\
\hline
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\hline
\binom{T}{T}a_r^T & \cdots & \binom{T}{T-d}a_r^{T-d} & \cdots & \cdots & \binom{T}{0}a_r^0 & & \\
& \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \\
& & \binom{T}{T}a_r^T & \cdots & \cdots & \binom{T}{d}a_r^d & \cdots & \binom{T}{0}a_r^0
\end{array}\right)$$

must be singular. But, we show, in Lemma 3.1, that V cannot be singular if a_i's are distinct and $p > 2t$. This leads us to the necessary contradiction and hence a proof of Claim 2.3.

Remark - Notice that, V is a square matrix since the number of rows $D \cdot r$ equals the number of columns $D + T$, by the choice of $D = T/(r-1)$ in Claim 2.3. We call V a generalized Sylvester-Vandermonde matrix because when $r = 2$, it becomes the Sylvester matrix of the two polynomials $(x + a_1)^T$ and $(x + a_2)^T$, whereas when $D = 1$, it is the Vandermonde matrix (scaled appropriately).

We now prove the following identity.

Lemma 3.1 (Sylvester-Vandermonde identity)**.** *The* $\det(\mathsf{V}) = C \cdot \prod_{1 \le i < j \le r} (a_i - a_j)^{D^2}$*, where* $C = \prod_{\ell=0}^{T+d} \binom{T+d}{\ell} / \prod_{j=0}^{d} \binom{T+d}{j}^r$.

It is not hard to check that $C \neq 0$ in $\mathbb{F}_p$, if $p > 2t$ (just use the facts that $s \le r-2$ and r is an integer less than or equal to $2/\sqrt{5} \cdot \sqrt{t} + 1$).

3.1 Proof of Lemma 3.1

First, we show that $\det(\mathsf{V})$, viewed as a polynomial in $a_1, \ldots, a_r$, is divisible by $(a_1 - a_2)^{D^2}$. Then, by symmetry, $\det(\mathsf{V})$ is also divisible by $(a_i - a_j)^{D^2}$ for every pair (i, j) with $i < j$. Hence, $\det(\mathsf{V})$ is divisible by $\prod_{1 \le i < j \le r} (a_i - a_j)^{D^2}$. By looking at the matrix V, it is easy to infer that the highest degree of a_1 in $\det(\mathsf{V})$ (once again, viewed as a polynomial in $a_1, \ldots, a_r$) is at most $D \cdot T$. Since the degree of a_1 in the expression $\prod_{1 \le i < j \le r} (a_i - a_j)^{D^2}$ is $D^2 \cdot (r-1) = D \cdot T$ (as $D = T/(r-1)$), $\det(\mathsf{V})$ must be of the form $C \cdot \prod_{1 \le i < j \le r} (a_i - a_j)^{D^2}$, where C is just a function of T, d and r, but not the a_i's.

In the proof, it will be more convenient if we express matrix V in terms of polynomials. Notice that the rows of V can be identified with the coefficient vectors of the polynomials $(x+a_1)^T, x(x+a_1)^T, \ldots, x^d(x+a_1)^T, (x+a_2)^T, x(x+a_2)^T, \ldots, x^d(x+a_2)^T, \ldots$ and so on. Let us abuse notations slightly and write V as,

$$\mathsf{V} = \left(\begin{array}{c} (x+a_1)^T \\ x(x+a_1)^T \\ \vdots \\ x^d(x+a_1)^T \\ \hline (x+a_2)^T \\ \vdots \\ x^d(x+a_2)^T \\ \hline \vdots \\ \hline (x+a_r)^T \\ \vdots \\ x^d(x+a_r)^T \end{array}\right) \overset{\text{row operations}}{\longmapsto} \mathsf{V}' = \left(\begin{array}{c} (x+a_1)^T \\ (x+a_1)^{T+1} \\ \vdots \\ (x+a_1)^{T+d} \\ \hline (x+a_2)^T \\ \vdots \\ (x+a_2)^{T+d} \\ \hline \vdots \\ \hline (x+a_r)^T \\ \vdots \\ (x+a_r)^{T+d} \end{array}\right), \tag{6}$$

meaning that V is formed by the coefficient vectors of these polynomials. Let R_{ij} be the row of V standing for the coefficient vector of $x^j(x+a_i)^T$. Consider the following row operations on V.

$$\mathsf{R}_{ij} \mapsto \sum_{k=0}^{j} \binom{j}{k} \cdot a_i^k \cdot \mathsf{R}_{ij-k}, \qquad \text{for } 1 \le i \le r \text{ and } 0 \le j \le d.$$

Equivalently, after the row operations, the coefficient vector of $x^j(x+a_i)^T$ gets replaced by that of the polynomial,

$$\sum_{k=0}^{j} \binom{j}{k} \cdot a_i^k \cdot x^{j-k}(x+a_i)^T = (x+a_i)^{T+j}.$$

This leaves us with a transformed matrix V', as shown in Equation 6, such that $\det(\mathsf{V}) = \det(\mathsf{V}')$. To show that $(a_1 - a_2)^{D^2}$ divides $\det(\mathsf{V}')$, view $\det(\mathsf{V}') = f(a_1)$ as a polynomial in a_1. Let $f^{(\ell)}(a_1)$ denote the ℓ^{th} order derivative of $f(a_1)$ with respect to a_1. It is sufficient if we are able to show that $a_1 = a_2$ is a root of $f^{(\ell)}(a_1)$, for all $0 \le \ell < D^2$.

Claim 3.2. *Let* $f(a_1) = \det(\mathsf{V}')$ *and* $f^{(\ell)}(a_1) = \frac{\partial^{(\ell)}}{\partial a_1^{(\ell)}} f(a_1)$. *Then* $a_1 = a_2$ *is a root of* $f^{(\ell)}(a_1)$, *for all* $0 \leq \ell < D^2$.

Proof. To prove this claim, we need the following identity involving derivatives of a determinant. Let $A = (a_{i,j})$ be an $n \times n$ matrix whose entries are real functions of x. Then,

$$\frac{d^\ell}{dx^\ell} \det(A) = \sum_{\ell_1+\ell_2+\cdots+\ell_n=\ell} \binom{\ell}{\ell_1, \ell_2, ..., \ell_n} \det \begin{pmatrix} \frac{d^{\ell_1}}{dx^{\ell_1}} a_{1,1} & \frac{d^{\ell_1}}{dx^{\ell_1}} a_{1,2} & \cdots & \frac{d^{\ell_1}}{dx^{\ell_1}} a_{1,n} \\ \vdots & \vdots & & \vdots \\ \frac{d^{\ell_n}}{dx^{\ell_n}} a_{n,1} & \frac{d^{\ell_n}}{dx^{\ell_n}} a_{n,2} & \cdots & \frac{d^{\ell_n}}{dx^{\ell_n}} a_{n,n} \end{pmatrix},$$

where $\binom{\ell}{\ell_1,\ell_2,...,\ell_r}$ is the multinomial coefficient. Now imagine applying this identity to $f^{(\ell)}(a_1)$, the ℓ^{th} order derivative of $\det(\mathsf{V}')$, where the entries of V' are viewed as functions of a_1. It is clear from Equation 6 that except for the first D rows of V', the rest are independent of the variable a_1. Denote by R_{ij}, the row of V' generated by the coefficients of $(x + a_i)^{T+j}$. We write $\frac{d^\ell}{da_1^\ell}\mathsf{R}_{ij}$ to mean the row formed by applying the operator $\frac{d^\ell}{da_1^\ell}$ to every entry of R_{ij}. Therefore, we have the following identity. (For economy of space, we switch to the transpose notation.)

$$f^{(\ell)}(a_1) = \sum_{\ell_1+\ell_2+\cdots+\ell_D=\ell} \binom{\ell}{\ell_1, \ell_2, ..., \ell_D} \det \left([\frac{d^{\ell_1}}{da_1^{\ell_1}}\mathsf{R}_{10}, \ldots, \frac{d^{\ell_D}}{da_1^{\ell_D}}\mathsf{R}_{1d}, \mathsf{R}_{20}, \ldots, \mathsf{R}_{2d}, \ldots, \mathsf{R}_{r0}, \ldots, \mathsf{R}_{rd}]^T \right).$$

Notice one nice property of the polynomial representation of V': In the above equation, $\frac{d^{\ell_{j+1}}}{da_1^{\ell_{j+1}}}\mathsf{R}_{1j}$ is exactly the row formed by the coefficients of $\frac{d^{\ell_{j+1}}}{da_1^{\ell_{j+1}}}(x + a_1)^{T+j} = c_j \cdot (x + a_1)^{T+j-\ell_{j+1}}$, where c_j is a constant (depending only on T, j and ℓ_{j+1}). Now let us see how large ℓ needs to be so that $(a_1 - a_2)$ does not divide $f^{(\ell)}(a_1)$.

Suppose $(a_1 - a_2) \nmid f^{(\ell)}(a_1)$. Then there exist a term L in the above summation that is not divisible by $(a_1 - a_2)$. Let that term be identified by some tuple $(\ell_1, \ell_2, \ldots, \ell_D)$. Observe that this term,

$$\begin{aligned} L &= \det \left([\frac{d^{\ell_1}}{da_1^{\ell_1}}\mathsf{R}_{10}, \ldots, \frac{d^{\ell_D}}{da_1^{\ell_D}}\mathsf{R}_{1d}, \mathsf{R}_{20}, \ldots, \mathsf{R}_{2d}, \ldots, \mathsf{R}_{r0}, \ldots, \mathsf{R}_{rd}]^T \right) \\ &= \det \left([c_0(x + a_1)^{T-\ell_1}, \ldots, c_d(x + a_1)^{T+d-\ell_D}, (x + a_2)^T, \ldots, (x + a_2)^{T+d}, \ldots]^T \right). \end{aligned}$$

If any of the exponents $\{T - \ell_1, T + 1 - \ell_2, \ldots, T + d - \ell_D\}$ is greater or equal to T then $(a_1 - a_2) \mid L$; since otherwise some row $c_j(x + a_1)^{T+j-\ell_{j+1}}$ becomes equal to some other row $(x+a_2)^{T+k}$ (up to a multiple of c_j), when a_1 is replaced by a_2. Also, if any two of the exponents $\{T-\ell_1, T+1-\ell_2, \ldots, T+d-\ell_D\}$ are the same then $L = 0$. This leaves us with only one option - the set $\{T-\ell_1, T+1-\ell_2, \ldots, T+d-\ell_D\}$ is 'dominated' by the set $\{T-1, T-2, \ldots, T-D\}$. (We say a set S_1 is *dominated* by another set S_2 if for every element $e_1 \in S_1$ there is a unique element $e_2 \in S_2$ such that $e_1 \leq e_2$.) Therefore,

$$\begin{aligned} \sum_{j=0}^{d} T + j - \ell_{j+1} &\leq \sum_{k=1}^{D} T - k \\ \Rightarrow \ell &\geq DT + \frac{d(d+1)}{2} - DT + \frac{D(D+1)}{2} \\ &= D^2 \quad \text{(Taking } d = D - 1\text{)} \end{aligned}$$

□

It follows from Claim 3.2 and the discussion before that $\det(\mathsf{V})$ is of the form $C \cdot \prod_{1 \le i < j \le r} (a_i - a_j)^{D^2}$, where C is a constant depending only on T, d and r. What remains to be done, in order to complete the proof of Lemma 3.1, is to show that $C = \prod_{\ell=0}^{T+d} \binom{T+d}{\ell} / \prod_{j=0}^{d} \binom{T+d}{j}^r$. The proof of this is included in Appendix A.

4 Discussion

Although, the square root bound of Corollary 2.2 is not the best known bound for this problem, it may be worthwhile exploring the 'polynomial method' further to see if the bound can be strengthened, or if nontrivial bounds of some other related problems can be derived through it. Towards this, we have the following three questions in mind.

Question 4.1. *(Strengthening Lemma 2.1)* Is it possible to give a better bound on the number of common solutions of the system $\mathcal{S}$ (defined in Lemma 2.1), perhaps by considering an auxiliary polynomial of the form,

$$F(x) = \sum_{\sigma \in S_\ell} G_\sigma(x) \cdot \prod_{i \in \sigma} (x + a_i)^{T+M-1},$$

where S_ℓ is the set of all ℓ-tuples with ℓ distinct elements chosen from $\{1, \ldots r\}$? ($|S_\ell| = \binom{r}{\ell}$).

Note, in our case, the auxiliary polynomial $F(x)$ is defined with $\ell = 1$.

Question 4.2. *(Simultaneous quadratic character)* Is it possible to use our approach to show that for any pair of distinct elements $a, b \in \mathbb{F}_p$, the value of the largest possible m for which $\chi((a+i) \cdot (b+i)) = 1$, for all $0 \le i \le m$, is $O(\sqrt{p} \log p)$? Here $\chi(a)$ denotes the quadratic character of a.

Question 4.3. *(Least primitive element in $\mathbb{F}_p$)* Is it possible to show that the value of the least primitive element in $\mathbb{F}_p$ is $O(\sqrt{p} \cdot \mathsf{poly}(\log p))$ using our approach?

Acknowledgement

This research was done when the authors F, M and S were interning at Microsoft Research India. The authors are thankful to MSR India for providing an excellent environment for research.

References

[Ank52] Nesmith C. Ankeny. The Least Quadratic Nonresidue. *Annals of Mathematics*, 55:65–72, 1952.

[Bac82] Eric Bach. Fast Algorithms under the Extended Riemann Hypothesis: A Concrete Estimate. In *STOC*, pages 290–295, 1982.

[Bra32] A. Brauer. Ueber die Verteilung der Potenzreste. *Math. Z.*, 35:39–50, 1932.

[Bur57] D. A. Burgess. The distribution of quadratic residues and non-residues. *Mathematika*, 4(8):106–112, 1957.

[Dav00] Harold Davenport. *Multiplicative Number Theory*. Springer-Verlag, New York, 3rd edition, 2000.

[Evd94] Sergei Evdokimov. Factorization of polynominals over finite fields in subexponential time under GRH. In *ANTS*, pages 209–219, 1994.

[GKP89] Ronald L. Graham, Donald E. Knuth, and Oren Patashnik. *Concrete Mathematics: A Foundation For Computer Science*. Addison-Wesley, 1st edition, 1989.

[Hud74] Richard H. Hudson. Power Residues and Nonresidues in Arithmetic Progressions. *Transactions of the American Mathematical Society*, 194:277–289, 1974.

[Kar68] A. A. Karatsuba. Character sums and primitive roots in finite fields. *Soviet Math-Dokl*, 9(3):755–757, 1968.

[Sch04] Wolfgang M. Schmidt. *Equations over Finite Fields: An Elementary Approach*. Kendrick Press, Inc., 2nd edition, 2004.

[Ste75] S. A. Stepanov. Contructive methods in the theory of equations over finite fields. *Proc. Steklov Inst. Math*, 132:271–281, 1975.

[Tao09] Terence Tao. The least quadratic nonresidue, and the square root barrier. `http://terrytao.wordpress.com/2009/08/18/`, 2009.

[Vin27] J. M. Vinogradov. On the Bound of the Least Non-Residue of nth Powers. *Transactions of the American Mathematical Society*, 29(1):218–226, 1927.

[Vin54] I.M. Vinogradov. *Elements of Number Theory*. Dover Publication, 1954.

[Vin72] I.M. Vinogradov. *Basic Number Theory*. Moscow, 1972.

[Vin85] I.M. Vinogradov. *Selected works*. Springer, 1985. Translated from Russian.

A The constant in Lemma 3.1

Since the monomial $\prod_{i=1}^{r} a_i^{D^2(r-i)}$ has coefficient 1 in the product $\prod_{1\le i<j\le r}(a_i-a_j)^{D^2}$ (viewed as a polynomial in the a_i's), the constant C in Lemma 3.1 is the same as the coefficient of the monomial $\prod_{i=1}^{r} a_i^{D^2(r-i)}$ in $\det(\mathsf{V})$.

Claim A.1. *The coefficient of the monomial* $\prod_{i=1}^{r} a_i^{D^2(r-i)}$ *in* $\det(\mathsf{V})$ *is* $\prod_{\ell=0}^{T+d}\binom{T+d}{\ell}/\prod_{j=0}^{d}\binom{T+d}{j}^{r}$.

Proof. By definition, $\det(\mathsf{V}) = \sum_{\sigma\in S_m}\operatorname{sign}(\sigma)\cdot\prod_{\ell\in[m]} v_{\ell,\sigma(\ell)}$, where $m = Dr$, S_m is the symmetric group of degree m, and $v_{i,j}$ is the $(i,j)^{th}$ entry of V. Note that, every product $\prod_{\ell\in[m]} v_{\ell,\sigma(\ell)}$ is a monomial in the a_i's with an attached coefficient. We need to find out, which all permutations σ give rise to the monomial $\prod_{i=1}^{r} a_i^{D^2(r-i)}$.

Let both R_i and C_i denote the set $\{D(i-1)+1,\ldots,D(i-1)+D\}$, so that $[Dr] = R_1\cup\cdots\cup R_r = C_1\cup\cdots\cup C_r$. We think of the R_i's as partitioning the rows and the C_i's as partitioning the columns of V. For instance, the rows with indices in R_i contain only those terms involving the variable a_i. A crucial observation here is the following. The monomial $\prod_{i=1}^{r} a_i^{D^2(r-i)}$ is generated by exactly those permutations σ that induce bijections between $R_i \leftrightarrow C_i$, for all $1\le i\le r$. This gives us a strategy to find the coefficient of $\prod_{i=1}^{r} a_i^{D^2(r-i)}$.

Define the matrix M_i as the $D\times D$ submatrix of V which is induced by the rows R_i and the columns C_i. Since each term in $\det(M_i)$ has the same degree in a_i, which is $D^2(r-i)$, the coefficient of $\prod_i a_i^{D^2(r-i)}$ can be obtained from the product $\prod_{1\le i\le r}\det(M_i)$.

Notice that $\det(M_i) = a_i^{D^2(r-i)} \cdot \det(H_i)$, where H_i is the following matrix formed by the binomial coefficients of the terms in M_i.

$$H_i = \begin{pmatrix} \binom{T}{T-D(i-1)} & \binom{T}{T-D(i-1)-1} & \cdots & \binom{T}{T-D(i-1)-d} \\ \binom{T}{T-D(i-1)+1} & \ddots & \ddots & \binom{T}{T-D(i-1)-(d-1)} \\ \vdots & \ddots & \ddots & \vdots \\ \binom{T}{T-D(i-1)+d} & \binom{T}{T-D(i-1)+(d-1)} & \cdots & \binom{T}{T-D(i-1)} \end{pmatrix}.$$

Therefore, the coefficient of $\prod_{i=1}^{r} a_i^{D^2(r-i)}$ in $\det(\mathsf{V})$ is exactly $\prod_{1\le i\le r} \det(H_i)$. For $1 < i < r$, each of the binomial coefficients in the matrix H_i is non-degenerate, whereas for $i \in \{1, r\}$ it is easy to see that $\det(H_i) = 1$ as H_i is a triangular matrix with units along the diagonal. Suppose $1 < i < r$. After an appropriate row transformation, H_i gets transformed to,

$$H_i' = \begin{pmatrix} \binom{T+d}{T-D(i-1)+d} & \binom{T+d}{T-D(i-1)+(d-1)} & \cdots & \binom{T+d}{T-D(i-1)} \\ \binom{T+d-1}{T-D(i-1)+d} & \ddots & \ddots & \binom{T+d-1}{T-D(i-1)} \\ \vdots & \ddots & \ddots & \vdots \\ \binom{T}{T-D(i-1)+d} & \binom{T}{T-D(i-1)+(d-1)} & \cdots & \binom{T}{T-D(i-1)} \end{pmatrix},$$

so that $\det(H_i) = \det(H_i')$. Now we apply the following lemma, which we prove shortly, to find an expression for $\det(H_i')$.

Lemma A.2. *For $n, m, \ell \in \mathbb{N}$, satisfying $\ell + m \le n$,*

$$\begin{vmatrix} \binom{n+m}{\ell+m} & \binom{n+m}{\ell+m-1} & \cdots & \binom{n+m}{\ell} \\ \binom{n+m-1}{\ell+m} & \binom{n+m-1}{\ell+m-1} & \cdots & \binom{n+m-1}{\ell} \\ \vdots & \vdots & \ddots & \vdots \\ \binom{n}{\ell+m} & \binom{n}{\ell+m-1} & \cdots & \binom{n}{\ell} \end{vmatrix} = \frac{\prod_{j=0}^{m} \binom{n+m}{\ell+j}}{\prod_{j=0}^{m} \binom{n+m}{j}}$$

Taking $n \to T$, $m \to d$ and $\ell \to T - D(i-1)$, in the above lemma, we get

$$\det(H_i) = \prod_{j=0}^{d} \frac{\binom{T+d}{T-D(i-1)+j}}{\binom{T+d}{j}} = \prod_{j=0}^{d} \frac{\binom{T+d}{D(i-1)+j}}{\binom{T+d}{j}}. \tag{7}$$

Note that the condition $\ell + m \le n$, in Lemma A.2, is satisfied after the substitution since $1 < i < r$. Also, the above formula 7 evaluates to 1 for $i \in \{1, r\}$. Hence, the formula holds for all $1 \le i \le r$. Therefore, the coefficient of $\prod_{i=1}^{r} a_i^{D^2(r-i)}$ is,

$$\prod_{i=1}^{r} \det(H_i) = \prod_{i=1}^{r} \prod_{j=0}^{d} \frac{\binom{T+d}{D(i-1)+j}}{\binom{T+d}{j}} = \frac{\prod_{\ell=0}^{T+d} \binom{T+d}{\ell}}{\prod_{j=0}^{d} \binom{T+d}{j}^r},$$

as claimed. □

It remains to prove Lemma A.2.

Proof of Lemma A.2. Index the rows from the bottom - the 0^{th} row R_0 is the bottommost. Consider the row operation $\mathsf{R}_m \to (-1)^m \cdot \mathsf{R}_m + \sum_{k=0}^{m-1} (-1)^k \binom{m}{k} \frac{\binom{n+m}{\ell}}{\binom{n+k}{\ell}} \cdot \mathsf{R}_k$. Then the value of the topmost row in the i^{th} column (from the right, starting from zero) is

$$\sum_{k=0}^{m} \binom{n+k}{\ell+i} (-1)^k \binom{m}{k} \frac{\binom{n+m}{\ell}}{\binom{n+k}{\ell}} = \frac{\binom{n+m}{\ell}}{\binom{\ell+i}{\ell}} \sum_{k=0}^{m} (-1)^k \binom{m}{k} \binom{n-\ell+k}{i}$$

The claim is - the quantity inside the summation is zero for $i \in \{0, 1, \cdots, m-1\}$. This is because of the following identity involving binomial coefficients (see page-169 in the book [GKP89]).

$$\sum_{k=0}^{m} (-1)^k \binom{m}{k} \binom{s+k}{i} = (-1)^m \cdot \binom{s}{i-m},$$

which is zero as $\binom{s}{i-m} = 0$ when $i < m$ (by definition). For $i = m$, the above equation evaluates to $(-1)^m$. This means, after the transformation the value of the leftmost entry of the top row is $(-1)^m \cdot \binom{n+m}{\ell} / \binom{\ell+m}{\ell}$, whereas all the remaining entries of the row are zeroes. Recall that, in the row transformation we have multiplied the first row R_m by $(-1)^m$, and so remultiplying by $(-1)^m$ the top-left entry becomes simply $\binom{n+m}{\ell} / \binom{\ell+m}{\ell}$. Using this argument inductively on the minors, the determinant evaluates to,

$$\prod_{j=0}^{m} \frac{\binom{n+j}{\ell}}{\binom{\ell+j}{\ell}} = \prod_{j=0}^{m} \frac{\binom{n+m}{\ell+j}}{\binom{n+m}{j}}.$$

The last equality is simple to verify. □