SINGULAR GAUSS SUMS, POLYA-VINOGRADOV INEQUALITY FOR $GL(2)$ AND GROWTH OF PRIMITIVE ELEMENTS

SATADAL GANGULY AND C. S. RAJAN

Abstract. We establish an analogue of the classical Polya-Vinogradov inequality for $GL(2, \mathbb{F}_p)$, where p is a prime. In the process, we compute the 'singular' Gauss sums for $GL(2, \mathbb{F}_p)$. As an application, we show that the collection of elements in $GL(2, \mathbb{Z})$ whose reduction modulo p are of maximal order in $GL(2, \mathbb{F}_p)$ and whose matrix entries are bounded by x has the expected size as soon as $x \gg p^{1/2+\varepsilon}$ for any $\varepsilon > 0$.

1. Introduction

Let χ be a non-principal Dirichlet character of modulus q. The well-known Polya-Vinogradov estimate for character sums is given by (see [Dav, Ch. 23])

$$\sum_{x \leq n < x+y} \chi(n) = O(\sqrt{q} \log q), \tag{1.1}$$

where x and $y > 0$ are any integers. Here, by the notation $f(y) = O(g(y))$ or $f(y) \ll g(y)$, f being any function and g being a positive function defined on a domain Y, we mean that there is a constant $c > 0$ such the bound $|f(y)| \leq cg(y)$ holds for all $y \in Y$. The trivial bound for such a character sum is y and one can easily obtain the bound q. Thus the Polya-Vinogradov bound indicates cancellations in a sum of the character values along an interval as soon as the length of the interval becomes somewhat larger than $\sqrt{q} \log q$.

It is natural to consider what the analogue of the Polya-Vinogradov estimate should be for groups more general than $(\mathbb{Z}/q\mathbb{Z})^*$. To start with, one may consider the following broad question:
Do cancellations occur in a sum of the type

$$\sum_{H(A) \leq x} \chi_\rho(A), \tag{1.2}$$

where ρ is a non-trivial complex representation of $G(\mathbb{Z}/q\mathbb{Z})$, G being a linear algebraic group defined over $\mathbb{Z}$, $\chi_\rho = \mathrm{Tr} \circ \rho$ its character (where Tr denotes the *trace* map), and $H : G(\mathbb{Z}) \to \mathbb{R}$ is some suitable height function that measures the 'size'

2010 *Mathematics Subject Classification.* Primary 11T24, Secondary 20C33.

of A? An affirmative answer would amount to obtaining non-trivial bound for this sum in terms of q that is uniform over x.

The proofs of the classical Polya-Vinogradov bound guides us to the groups to which we should attempt to generalize it. All the proofs utilize harmonic analysis on the abelian group $\mathbb{Z}/q\mathbb{Z}$ in one way or the other. For example, one can expand the Dirichlet character in a finite Fourier series in terms of the additive characters where the Fourier coefficients are essentially the classical Gauss sums. Then one needs to estimate a finite geometric series and use the classical bound $O(\sqrt{q})$ for Gauss sums to obtain (1.1).

The analogy of Gauss sums with L-functions and the use of abelian harmonic analysis in the method of Tate-Godement-Jacquet for proving analytic properties of the standard L-functions attached to cusp forms on $GL(n)$ suggests that a natural generalization should be to the group $GL(n, \mathbb{Z}/q\mathbb{Z})$. We restrict to the case $q = p$, an odd prime, for simplicity. Apart from the group $GL(n, \mathbb{Z}/p\mathbb{Z})$ being a natural generalization of the group $GL(1, \mathbb{Z}/p\mathbb{Z}) \simeq (\mathbb{Z}/p\mathbb{Z})^*$, the key point is that, similar to the classical case, we can utilize abelian harmonic analysis on the additive group $M(n, \mathbb{Z}/p\mathbb{Z})$, the group of $n \times n$ matrices over $\mathbb{F}_p$, to study the sum (1.2). Since the group $M(n, \mathbb{Z}/p\mathbb{Z})$ is self-dual, one new aspect that arises is the evaluation of singular Gauss sums, attached to singular matrices in $M(n, \mathbb{Z}/p\mathbb{Z})$.

1.1. **A GL(2) Polya-Vinogradov bound.** For a matrix $A \in M(n, \mathbb{Z})$, we denote by $\bar{A}$ its reduction modulo p; i.e., the image of A under the reduction map $M(n, \mathbb{Z}) \to M(n, \mathbb{F}_p)$. We extend χ_ρ to a function on $M(n, \mathbb{F}_p)$ by defining it to be zero on matrices whose determinant vanish modulo p and consider χ_ρ as a function on $M(n, \mathbb{Z})$ by the reduction modulo p map. We now make a definition which is a natural choice for the height function:

Definition 1.1. For $A \in M(n, \mathbb{Z})$, we define

$$h(A) := Max\{|a_{ij}|\},$$

a_{ij} being the (i, j)-th entry of A.

Our first main theorem is the following $GL(2)$-analogue of the classical Polya-Vinogradov inequality.

Theorem 1.2. *Let ρ be a non-trivial irreducible complex representation of the group $GL(2, \mathbb{F}_p)$. Let $d(\rho)$ be the dimension of ρ. Then, for any $x \geq 1$, we have the estimate*

$$\sum_{A \in M(n,\mathbb{Z}): h(A) \leq x} \chi_\rho(A) \ll d(\rho) p^2 (\log p)^4, \tag{1.3}$$

where the implied constant is absolute and one can take it to be 16 *if $p \geq 11$.*

We now state a more general version of the theorem from which Theorem 1.2 follows easily. First we define the notion of a matrix interval.

Definition 1.3. By a *matrix interval* over the integers we shall mean a set $\mathbf{I}$ of the form $\mathbf{I} = \prod_{1\leq i,j\leq n} I_{ij}$, where each I_{ij} is an interval in $\mathbb{Z}$; i.e., I is the set of $n \times n$ integer matrices $((a_{ij}))$ such that for every fixed pair (i, j), the (i, j)-th entry a_{ij} varies over the component interval I_{ij} in $\mathbb{Z}$.

A simple example of a matrix interval to keep in mind is to take some fixed matrix A_0 and define $\mathbf{I} = \mathbf{I}(A_0, x)$ to be the collection of matrices A such that $h(A - A_0) \leq x$.

With the above definition, our theorem is:

Theorem 1.4. *Suppose $\mathbf{I} = \prod_{1\leq i,j\leq 2} I_{ij}$ is a matrix interval over the integers such that the length of each component interval satisfies the bound $|I_{ij}| \leq cp$, where $c > 0$ is a constant. Then, under the same assumptions on a representation ρ as above, we have the bound*

$$\sum_{A\in\mathbf{I}} \chi_\rho(A) \ll d(\rho)p^2(\log p)^4, \tag{1.4}$$

where the implied constant is absolute and can be taken to be $\left(\frac{c+3}{2}\right)^4$ if $p \geq 11$.

Remarks

1. Recall (see Remark 1.9) that if χ is a non-trivial character of a finite group G then

$$\sum_{g\in G} \chi(g) = 0.$$

It follows, therefore, that if $\mathbf{I} = \prod_{1\leq i,j\leq 2} I_{ij}$ is a matrix interval having component intervals I_{ij} of the type $I_{ij} = [0, r(p - 1)]$, where $r \geq 1$ is a fixed integer, then

$$\sum_{A\in\mathbf{I}} \chi_\rho(A) = 0.$$

Thus, in the situation of Theorem 1.2, we may assume that $x < p$ and apply Theorem 1.4 with $c = 1$ to obtain Theorem 1.2.

2. The trivial estimate for the sums in Equations (1.3) and (1.4) is $d(\rho)p^4$, which shows that we obtain a 'saving' of p^2 compared to the trivial estimate.

3. The dimension $d(\rho)$ can be at most $p+1$ (see §2.2) and thus the character sums in the above two theorems are of size $O(p^3(\log p)^4)$.

1.2. **Non-abelian Gauss sums.** The analogue of the Gauss sums for $GL(n, \mathbb{F}_p)$ was introduced by Lamprecht ([La]). Let ρ be an irreducible, complex representation of the group $GL(n, \mathbb{F}_p)$ and let χ_ρ be its character. By $e(z)$ we shall denote $e^{2\pi iz}$ for a complex number z and by $e_p(z)$ we shall denote $e(z/p)$ throughout. Then, for integers x, the map $x \mapsto e_p(x)$ defines an additive character (denoted again by e_p) on the finite field $\mathbb{F}_p$ identified with $\mathbb{Z}/p\mathbb{Z}$. The bilinear pairing $(A, X) \mapsto e_p(\mathrm{Tr}(AX))$ on $M(n, \mathbb{F}_p)$, yields an identification of $M(n, \mathbb{F}_p)$ with its dual group of characters. Following Lamprecht ([La]), define the (matrix valued) Gauss sum attached to ρ and A as:

$$G(\rho, A) = \sum_{X \in G} \rho(X) e_p(\mathrm{Tr}(AX)). \tag{1.5}$$

It is easy to verify that for $A \in GL(n, \mathbb{F}_p)$,

$$G(\rho, A) = \rho(A)^{-1} G(\rho, I_d), \tag{1.6}$$

where $d = d(\rho)$. By Schur's lemma, it follows that $G(\rho, I_d)$ is a scalar matrix,

$$G(\rho, I_d) = g(\rho) I_d, \tag{1.7}$$

for some constant $g(\rho)$.

The characters of the irreducible complex representations of $GL(n, \mathbb{F}_p)$ were obtained explicitly by Green [Gr] in terms of the 'dual data' consisting of the conjugacy classes of elements in $GL(n, \mathbb{F}_p)$. Using Green's work, Kondo [Ko] obtained the following estimate for the size of the above Gauss sums:

Theorem 1.5 (Kondo)**.** *Let ρ be an irreducible, complex representation of $GL(n, \mathbb{F}_p)$. Then,*

$$|g(\rho)| = p^{(n^2 - k(\rho))/2}, \tag{1.8}$$

where $k(\rho)$ is the generalized multiplicity of the eigenvalue 1 in the conjugacy class attached to ρ by the Green correspondence.

Remark 1.6. More precisely, Kondo proves that up to a power of p, the non-abelian Gauss sum $g(\rho)$ is actually an 'abelian' Gauss sum, attached to a character of a maximal torus T of $GL(n)$. Kondo's result was also proved by Braverman and Kazhdan ([BK, Theorem 1.3]), using the construction of irreducible representations of $GL(n, \mathbb{F}_p)$ by Deligne and Lusztig ([DL]) and the theory of character sheaves due to Lusztig. We recall this result now.

Let T be a maximal torus of $GL(n)$ over $\mathbb{F}_p$, and $\theta : T(\mathbb{F}_p) \to \bar{Q}_\ell^*$ be a character. Associated to this data, Deligne and Lusztig construct a virtual representation $R_T(\theta)$ of $GL(n, \mathbb{F}_p)$, and show that every irreducible representation ρ of $GL(n, \mathbb{F}_p)$ is an irreducible constituent of some $R_T(\theta)$. Consider the abelian Gauss sum

$$g(\theta) = \sum_{X \in T(\mathbb{F}_p)} \theta(X) e_p(\mathrm{Tr}(X)).$$

Braverman and Kazhdan have shown ([BK, Theorem 1.3]) that

$$g(\rho) = p^{(n^2-n)/2} g(\theta).$$

Remark 1.7. The foregoing result allows us to specify $k(\rho)$. With the notation of Section 2.2,

$$k(\rho) = \begin{cases} 2 & \text{if } \rho \simeq St, \\ 1 & \text{if } \rho \simeq I_{\chi,1} \text{ and } \chi \text{ is non-trivial}, \\ 0 & \text{otherwise} \end{cases}$$

1.3. **Singular non-abelian Gauss sums.** Equations (1.6) and (1.7), gives an estimate for the trace of $G(\rho, A)$, provided A is a *non-singular* matrix:

$$|\mathrm{Tr}(G(\rho, A))| \leq d(\rho) p^{n^2/2}, \tag{1.9}$$

where $d(\rho)$ is the dimension of ρ. If ρ is not abelian, $d(\rho)$ is either $p-1, p$ or $p+1$. from the classification of irreducible representations of $GL(2, \mathbb{F}_p)$ (see §2.2). Thus $d(\rho) \leq p+1$ for any representation ρ and $d(\rho)$ is of order p unless ρ is abelian.

However, for the purpose of establishing an analogue of the Polya-Vinogradov inequality, we need to estimate Gauss sums attached to all (additive) characters $M(n, \mathbb{F}_p)$; in paricular, we need to estimate the *singular* Gauss sums, by which we mean the trace of $G(\rho, A)$ where A is a singular matrix in $M(n, \mathbb{F}_p)$.

It is easy to see that the trace of $G(\rho, A)$ depends only on the conjugacy class of A. Let

$$A_a = \begin{pmatrix} a & 0 \\ 0 & 0 \end{pmatrix}, \ a \neq 0 \quad \text{and} \quad N = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

A non-zero singular matrix in $M(2, \mathbb{F}_p)$ is conjugate to either A_a or N. One of our main results is Theorem 1.8 below which gives the explicit values of the singular Gauss sums when $n = 2$. We restrict to the case $n = 2$ for simplicity and this case is already quite involved. We expect that a similar result for general n should hold.

Theorem 1.8. *Let ρ be a complex irreducible representation of $G = GL(2, \mathbb{F}_p)$, and let A be a non-zero singular matrix in $M(2, \mathbb{F}_p)$. Then the following statements hold:*

(1) *Suppose ρ is not isomorphic to either the trivial representation 1_G, or the Steinberg representation St or the principal series representation $I_{\chi,1}$ with χ a non-trivial character of $\mathbb{F}_p^*$. Then,*

$$\mathrm{Tr}(G(\rho, A)) = 0.$$

(2) *For the trivial representation 1_G,*

$$G(1_G, A) = -p(p-1).$$

(3) *If $\rho \simeq I_{\chi,1}$ with χ a non-trivial character of $\mathbb{F}_p^*$, then*

$$\mathrm{Tr}(G(I_{\chi,1}, A_a)) = p(p-1)\overline{\chi(a)}G(\chi)$$
$$\mathrm{Tr}(G(I_{\chi,1}, N)) = p(p-1)G(\chi),$$

where

$$G(\chi) = \sum_{a \in \mathbb{F}_p^*} \chi(a) e_p(ax)$$

is the usual classical Gauss sum.

(4) *For the Steinberg representation St,*

$$\mathrm{Tr}(G(St, A_a)) = -p(p-1).$$
$$\mathrm{Tr}(G(St, N)) = p^2(p-1).$$

Remark 1.9. It follows from the orthogonality of characters, that $\mathrm{Tr}(G(\rho, 0))$ vanishes when ρ is a non-trivial irreducible representation of $GL(n, \mathbb{F}_p)$, and equal to $|GL(n, \mathbb{F}_p)|$ if $\rho = 1_G$, the trivial representation.

As a consequence of the above result and Kondo's estimate for non-singular Gauss sums given by Eq. (1.9), the following general theorem is immediate after one applies the Gauss estimate for the classical Gauss sum: $|G(\chi)| = \sqrt{p}$ and recalls the fact that the dimensions of $I_{\chi,1}$ and St are, respectively, $p+1$ and p (see §2).

Theorem 1.10. *Let $p > 2$ be a prime and let ρ be a complex irreducible representation of $GL(2, \mathbb{F}_p)$ and let A be a non-zero matrix in $M(2, \mathbb{F}_p)$. Then,*

$$|\mathrm{Tr}(G(\rho, A))| \leq d(\rho)p^2. \tag{1.10}$$

1.4. **Applications of the $GL(2)$ Polya-Vinogradov inequality.** We first describe the general plan for applications here. Let ϕ be a $GL(2, \mathbb{F}_p)$ conjugacy-invariant function on $M(2, \mathbb{F}_p)$. Consider the sum,

$$S(\phi, x) = \sum_{h(A) \leq x} \phi(\overline{A}), \tag{1.11}$$

where $\overline{A}$ denotes $A (\mathrm{mod}\ p)$. Decomposing ϕ as a Fourier series in terms of the irreducible characters of G, we write

$$\phi = \sum_{\rho \in \hat{G}} c_\phi(\rho)\chi_\rho,$$

where $\hat{G}$ is the collection of complex irreducible representations of G up to isomorphism and

$$c_\phi(\rho) = \frac{1}{|G|} \sum_{g \in G} \phi(g)\overline{\chi_\rho(g)}$$

is the Fourier coefficient of ϕ with respect to the character χ_ρ. From Theorem 1.2, upon singling out the contribution from the trivial representation of G as the 'main term', we obtain the estimate

$$\sum_{h(A)\leq x}\phi(A)=c_\phi(1_G)\sum_{h(A)\leq x}\chi_1(A)+O\left(d(\rho)p^2(\log p)^4\left|\sum_{\rho\in\hat{G}}c_\phi(\rho)\right|\right), \tag{1.12}$$

where for simplicity of notation, we write χ_1 to denote the trivial character of G. By Lemma 6.5, the contribution of the trivial character is,

$$\sum_{h(A)\leq x}\chi_1(A)=16\gamma_p x^4+O(x^3), \tag{1.13}$$

where $\gamma_p=1-\frac{1}{p}-\frac{1}{p^2}+\frac{1}{p^3}$.

Thus we obtain the general formula

$$\sum_{h(A)\leq x}\phi(A)=16c_\phi(1_G)\gamma_p x^4+O(c_\phi(1_G)x^3)+O\left(d(\rho)p^2(\log p)^4\left|\sum_{\rho\in\hat{G}}c_\phi(\rho)\right|\right). \tag{1.14}$$

1.4.1. *Counting elements in a conjugacy class.* We now consider the case where ϕ is the characteristic function of a conjugacy class C in $G=GL(2,\mathbb{F}_p)$. We want to count the number of matrices A in $M(2,\mathbb{Z})$ with height bounded by x that reduces modulo p to an element lying in C.

Let δ_C we denote the indicator function of the subset C of G. By orthogonality of characters,

$$\delta_C=\frac{|C|}{|G|}\sum_{\rho\in\hat{G}}\overline{\chi_\rho(c)}\chi_\rho,$$

for any $c\in C$. Therefore, proceeding as before, we have,

$$\begin{aligned}S(\delta_C,x)=\sum_{h(A)\leq x}\delta_C(A)&=\frac{|C|}{|G|}\sum_{\rho\in\hat{G}}\overline{\chi_\rho(c)}\sum_{h(A)\leq x}\chi_\rho(A)\\&=\frac{|C|}{|G|}\sum_{h(A)\leq x}\chi_1(A)+\frac{|C|}{|G|}\sum_{\rho\neq 1_G}\overline{\chi_\rho(c)}\sum_{h(A)\leq x}\chi_\rho(A);\end{aligned}$$

and we obtain the following general statement:

Proposition 1.11. *Suppose C is a conjugacy class in $G=GL(2,\mathbb{F}_p)$ and $c\in C$ is any element, we have the equality*

$$S(\delta_C,x)=\frac{16|C|\gamma_p}{|G|}x^4+O\left(\frac{|C|}{|G|}\left(x^3+p^2(\log p)^4\sum_{\rho\neq 1_G}d(\rho)|\overline{\chi_\rho(c)}|\right)\right).$$

Remark 1.12. The above result is of limited use as the inner sum $\sum_{\rho \neq 1_G} d(\rho)|\overline{\chi_\rho(c)}|$ can be quite large in general. However, for certain conjugacy classes this simple approach already gives a non-trivial result. See the next subsection for an example.

1.4.2. *Elliptic elements.* An element in $GL(2, \mathbb{F}_p)$ is said to be *elliptic* if its characteristic polynomial is irreducible over $\mathbb{F}_p$. We shall call an integer matrix elliptic if its reduction modulo p is elliptic. The problem of finding an elliptic element of the least height can be considered in analogy with the classical problem of finding the least quadratic non-residue for a prime p (see [Mo2]). It follows from (1.1) that for any $\varepsilon > 0$, there is a positive integer $\tau = O(p^{\frac{1}{2}+\varepsilon})$ that is a quadratic non-residue for the prime p and the matrix $\left(\begin{smallmatrix} 0 & \tau \\ 1 & 0 \end{smallmatrix}\right)$ is an elliptic element of height $O(p^{\frac{1}{2}+\varepsilon})$. Henceforth, we shall follow the standard custom of using the symbol ε to denote a positive real number which will be assumed to be as small as we please and the value of ε may differ from one occurrence to the other.

Now, suppose we want to count the elliptic elements of height up to x. Let Ω_e denote the set of elliptic elements in $GL(2, \mathbb{F}_p)$. Therefore, we need to estimate the size of $S(\delta_{\Omega_e}, x)$. Following the proof of Prop. 1.11, we can easily obtain a result of the form

$$S(\delta_{\Omega_e}, x) = 8\left(1 - \frac{2}{p} + \frac{1}{p^2}\right)x^4 + O(x^3 + p^{3+\varepsilon}),$$

which shows that asymptotically half of all matrices reduce to elliptic elements modulo p as soon as $x \gg p^{3/4+\varepsilon}$. However, by a direct and simple argument using the classical Polya-Vinogradov bound for characters of $\mathbb{F}_p^*$, we establish the following easy result which shows that it is enough to take $x \gg p^{1/2+\varepsilon}$:

Proposition 1.13. *With notation as above,*

$$S(\delta_{\Omega_e}, x) = 8\left(1 - \frac{2}{p} + \frac{1}{p^2}\right)x^4 + O(x^3\sqrt{p}\log p).$$

This theorem is used in the problem of estimating the growth of the number of primitive elements of height up to x described in the next section.

Remark 1.14. If we use the Burgess bound (see [Bur, Bur2, Bur3]) instead of the Polya-Vonogradov bound, then it is possible to obtain a superior result but that does not lead to any improvement in the final application towards counting primitive elements.

1.5. **Application to counting Primitive elements.** Given a prime p, assumed to be large, a classical problem is to estimate the size of the smallest positive primitive root g_p (i.e., a generator for the cyclic group $\mathbb{F}_p^*$). This can be reduced to a question of estimation of character sums and by the celebrated bound of Burgess

[Bur] on character sums, one can show that (see [Mo2])

$$g_p \ll_\varepsilon p^{\frac{1}{4\sqrt{e}}+\varepsilon}.$$

For $G = GL(2, \mathbb{F}_p)$, we consider the generators of the subgroup $\mathbb{F}_{p^2}^*$ as analogue of the primitive roots for $\mathbb{F}_p^*$. Such elements are the elliptic semisimple elements (see §5) of order $p^2 - 1$, which is the maximum possible order in G. We shall refer to them as *primitive elements*.

Let Ω_{prim} denote the set of primitive elements in G. One has (see §2.1)

$$\frac{|\Omega_{prim}|}{|G|} = \frac{\phi(p^2-1)}{2(p^2-1)}. \tag{1.15}$$

By the observation that 4 divides p^2-1 and by the lower bound $\frac{\phi(n)}{n} \gg (\log\log n)^{-1}$ (see [RS, Thm 15]), we have the following bounds for the above ratio:

$$(\log\log p)^{-1} \ll \frac{|\Omega_{prim}|}{|G|} \leq \frac{1}{4}.$$

Here ϕ denotes the Euler ϕ-function. We have used the representation theory of $GL(2, \mathbb{F}_p)$, Theorem 1.4, and the classical Polya-Vinogradov estimate to prove the following theorem which gives an asymptotic formula for the number of elements in the set $\{A \in M(2, \mathbb{Z}) : h(A) \leq x\}$ that reduce to primitive elements modulo p.

Theorem 1.15. *For any $\varepsilon > 0$, we have*

$$S(\delta_{\Omega_{prim}}, x) = \frac{8\phi(p^2-1)}{(p^2-1)}(1-2/p+1/p^2)x^4 + O(x^3\sqrt{p}\log p) + O(x^2 p\log p) + O(p^{2+\varepsilon}) \tag{1.16}$$

The following is immediate:

Corollary 1.16. *Given a sufficiently large but fixed prime p and any $x \gg p^{1/2+\varepsilon}$, a positive proportion of the set of matrices of height up to x reduce to primitive elements of $GL(2, \mathbb{F}_p)$. In particular, there is a matrix of height $O(p^{1/2+\varepsilon})$ that reduces to a primitive element of $GL(2, \mathbb{F}_p)$.*

Remark 1.17. An interesting question is whether one can prove the existence of primitive elements in a one-parameter family of the form $\mathcal{A} = \{B+nI : 1 \leq n \leq x\}$, where B is some suitable fixed matrix and $x > 0$ is a parameter that we want to make as small as possible relative to p (for example, $x = p^{1/2+\varepsilon}$ would be a natural choice). In other words, we would like to know whether there is an integer n which is not too large such that the eigenvalues of the matrix $B + nI$ are primitive roots for p^2 (i.e., generators of the cyclic group $\mathbb{F}_{p^2}^*$).

Assume that the characteristic polynomial of B is not reducible over $\mathbb{F}_p$ and that θ_1 and θ_2 are the eigenvalues of B. Then the eigenvalues of $B + nI$ are $\theta_1 + n$ and $\theta_2 + n$, and thus we are led to the following general question:

Suppose $q = p^m$, and θ is an element of $\mathbb{F}_q$ such that $\mathbb{F}_p(\theta) = \mathbb{F}_q$. Is there some element $a \in \mathbb{F}_p$ such that $\theta + a$ is a primitive root for q and if so, how small can we take a to be (identifying the elements of $\mathbb{F}_p$ with integers from 0 to $p-1$)?

The study of such questions was initiated by Davenport [Dav2] and there have many works subsequently, e.g., [Da-Le] and [Bur2], to name a few. In [PS], Perel'muter and Shparlinski count the number of primitive roots for q in a set of the form $\{\theta+n : 0 \leq n \leq X\}$. It follows from their result that there are integers $n = O(p^{1/2+\varepsilon})$ such that $\theta + n$ is a primitive root for q. This proves the existence of primitive matrices in one-parameter families of the form $B + nI$ with $n = O(p^{1/2+\varepsilon})$, provided that the characteristic polynomial of B is irreducible over $\mathbb{F}_p$.

Now, using Prop. 1.13 and the work of [PS], it is possible by a careful analysis to give an alternative proof of Theorem 1.15 and we have carried it out in §6.6. The error term we get by this method is a little different but there is no substantive change in the strength of the result.

We emphasize here that the result in [PS] depends crucially on the Riemann Hypothesis for curves over finite field proved by Weil, whereas the first proof of Theorem 1.15 we have given in §6 using representation theory requires no tool form Algebraic Geometry. A curious feature of the representation-theoretic proof is that the main term results not from the contribution of the trivial representation alone and both the trivial representation and the Steinberg representation have to be considered together to obtain the main term.

Remark 1.18. In view of Theorem 1.4, one can replace the sum $S(\delta_{\Omega_{prim}}, x)), x)$ by the sum $S(\delta_{\Omega_{prim}}, x)), A_0, x) := \sum_{h(A-A_0)\leq x} \delta_{\Omega_{prim}}(A)$ and arrive at a similar estimate, where A_0 is some chosen base matrix.

Remark 1.19. When x is small, namely if $x < p$, the cofficient of the main term in Prop. 1.11 (resp. Theorem 1.13, Theorem 1.15) can be taken to be $16|C|/|G|$ (resp. 8, $8\phi(p^2-1)/(p^2-1)$). The correction factor γ_p (resp. $(1-2/p+1/p^2)$, $(1-2/p+1/p^2)$) arises for larger x, due to the contribution from matrices that reduce to singular matrices modulo p.

1.6. **Some general remarks.** 1. It will be interesting to extend our results to $GL(2, \mathbb{Z}/q\mathbb{Z})$ for an arbitrary positive integer q. If q is square-free, this group is a product of groups of the form $GL(2, \mathbb{F}_p)$ for primes p dividing q, and the irreducible representations of $GL(2, \mathbb{Z}/q\mathbb{Z})$ is a tensor product of the irreducible representations of $GL(2, \mathbb{F}_p)$

2. In the case of a Dirichlet character $\chi(\text{mod } q)$, the Polya-Vinogradov bound indicates cancellations as soon as the length X of the sum $\sum_{n\leq X} \chi(n)$ is somewhat

larger than $\sqrt{q}\log q$. However, cancellations do take place in sums of much shorter length and cancellations in such shorter sums correspond to strong bounds on the Dirichlet L-function. Indeed, showing cancellations in a sum of length $O(q^{1/2-\delta})$ for any $\delta > 0$ amounts to proving a subconvex estimate for $L(s,\chi)$ (see [IK, Chap. 5]) and the greater the value of δ we can take, the stronger is the bound on the L-function. In particular, Lindelöf Hypeothesis on $L(s,\chi)$ corresponds to cancellations in extremely short sums of length $O(q^{\varepsilon})$ for any $\varepsilon > 0$. It will be very interesting to develop of a theory of L-function attached to a representations of $GL(n,\mathbb{F}_p)$ in order to study the sums we are considering. It is not clear within what height we should expect to find cancellations in the sums over matrices and, in particular, whether the analogue of Lindeöf hypotehsis should hold. Any theory, even a conjectural one, for making a deeper analysis of these sums will be welcome.

3. There are several natural choices for a height functions other than the one considered here; e.g., the operator norm or the L^2-norm of a matrix. It would be interesting to investigate whether one could obtain similar results with other height functions.

1.7. **Main ideas behind the proofs and the structure of the paper.** The proof of Theorem 1.4 follows the usual approach for proving the classical Polya-Vinogradov inequality. The periodicity of χ_ρ allows one to consider the sum

$$S(\chi_\rho, \mathbf{I}) := \sum_{A\in\mathbf{I}} \chi_\rho(A),$$

as an inner product $< \chi_\rho, \delta_{\bar{\mathbf{I}}} >$ on the group $M(n,\mathbb{Z}/p\mathbb{Z})$, where $\bar{\mathbf{I}}$ is the image of $\mathbf{I}$ under the natural projection map from $M(n,\mathbb{Z})$ to $M(n,\mathbb{Z}/p\mathbb{Z})$. Applying the isometry of the Fourier transform on $M(n,\mathbb{Z}/p\mathbb{Z})$, the problem reduces to that of estimating two kinds of sums: sums of additive characters that lead to finite geometric sums, and the matrix Gauss sums, including the singular Gauss sums, that occur as Fourier transforms of χ_ρ with respect to the characters of $M(n,\mathbb{Z}/p\mathbb{Z})$.

For the non-singular Gauss sums, the formula of Kondo, namely Eq. (1.8) suffices but we need to analyze the singular Gauss sums as well. After collecting some background material on conjugacy classes and representations of $GL(2,\mathbb{F}_p)$ in §2, we analyze these singular Gauss sums for $GL(2,\mathbb{F}_p)$ and prove the main result for them, namely Theorem 1.8, in §3. In §4, we carry out the analytic part of the proof of Theorem 1.4, thus completing the proof.

The next sections are on applications. Theorem 1.13 is proved in §5 and to obtain the specific error term, we use the classical Polya-Vinogradov bound together with a counting argument. The proof of Theorem 1.15 is given in §6. A natural idea here would be to first expand the indicator function of the set Ω_{prim} in terms of the characters and then to apply Theorem 1.2 and estimate the sum of the Fourier

coefficients. This is done in §6.1 after obtaining bounds for the sum of the Fourier coefficients (see Lemma 6.3) and we obtain a weaker result, namely, Prop. 6.4.

Note that the problematic term $O(p^{3+\varepsilon})$ in Prop. 6.4 arises from Theorem 1.10 and the bound $d(\rho) \leq p+1$. In order to improve upon this, we need to carefully analyze and accordingly utilize the instances where the estimate in Theorem 1.10 can be improved to $O(p^{2+\varepsilon})$. The one-dimensional representations do not pose a problem, and there is no contribution from the principal series as their characters vanish on Ω_{prim}. The improvement arises from two crucial observations. One is the striking fact that $|\mathrm{Tr}(\rho(A))| \leq 2$ for non-central elements of $GL(2,\mathbb{F}_p)$ (see Prop. 6.6), which allows one to improve the estimate in Theorem 1.10 by a factor of p when A is non-singular. The second observation is that the trivial and the Steinberg representations are related. Their contributions can be clubbed together as the main term, allowing one to avoid the problems arising from the contributions of the singular Gauss sums attached to the Steinberg reprsentation which are of order p^3. An appeal to Prop. 1.13 finishes the proof of Theorem 1.15.

The proof of Prop. 1.13 rests only on the classical Polya-Vinogradov theorem, whereas that of Theorem 1.15 makes use of the non-abelian versiod developed in this paper. Thus, the proof of Theorem 1.15, involves both the $GL(1)$ and $GL(2)$-versions of the Polya-Vinogradov type theorems.

Finally, in §6.6, we explain an alternative approach towards the problem of counting primitive elements using older results on exponential sums that depend crucially on the work of Weil on the Riemann Hypothesis for curves over finite fields.

Acknowledgement. This work was started when the second author visited ISI, Kolkata in March, 2016. Both the authors thank ISI and TIFR, Mumbai where much of the work was carried out for excellent working condition. The second author thanks MPIM, Bonn for two visits during May of 2018 and 2019, for an excellent working environment allowing the authors to make progress on these questions. It is a pleasure to acknowledge J.-M. Deshouillers, É. Fouvry, E. Ghate, H. Iwaniec, F. Jouve, D. Prasad, O. Ramaré, D.S. Ramana, S. Sen, S. Varma for their interest, suggestions and encouragement.

2. Conjugacy classes and representations of $GL(2,\mathbb{F}_p)$

2.1. **Conjugacy classes in $GL(2,\mathbb{F}_p)$.** Let p be an odd prime. We recall the classification of conjugacy classes in $GL(2,\mathbb{F}_p)$ (see [FH]):

Central elements. The central elements given by scalar matrices. These have order dividing $(p-1)$.

Non-semisimple classes. The non-semisimple elements are conjugate to a matrix of the form $\left(\begin{smallmatrix} x & 1 \\ 0 & x \end{smallmatrix}\right)$, with $x \in \mathbb{F}_p^*$. The order of these elements divides $p(p-1)$.

Split semisimple classes. The non-central split semisimple elements are those whose characteristic polynomials have distinct roots in $\mathbb{F}_p$. These are conjugate to a matrix of the form $\left(\begin{smallmatrix} x & 0 \\ 0 & y \end{smallmatrix}\right)$, with $x,\ y \in \mathbb{F}_p^*,\ x \neq y$. These elements have order dividing $(p-1)$.

Elliptic semisimple classes. The elliptic (or non-split) semisimple conjugacy classes are those whose characteristic polynomials are irreducible over $\mathbb{F}_p$. Let τ be a non-square in $\mathbb{F}_p$, and $\tau' \in \mathbb{F}_{p^2}$ be a squareroot of τ. The roots of the characteristic polynomial can be written as $\zeta_{x,y} = x + \tau' y$ and $\zeta_{x,y}^p = x - \tau' y$. The matrix $d_{x,y} = \left(\begin{smallmatrix} x & \tau y \\ y & x \end{smallmatrix}\right)$, with $x,\ y,\ \tau \in \mathbb{F}_p,\ y \neq 0$ is a representative for the conjugacy class determined by $\{\zeta_{x,y},\ \zeta_{x,y}^p\}$. These elements have order dividing (p^2-1).

The action of $\mathbb{F}_{p^2}$ on itself by multiplication gives an embedding of $\mathbb{F}_{p^2}^*$ into $GL(2,\mathbb{F}_p)$ and thus a generator for the cyclic group $\mathbb{F}_{p^2}^*$ yields an element of order (p^2-1) in $GL(2,\mathbb{F}_p)$. The matrix $d_{x,y}$ is the matrix of the transformation given by multiplication by $\zeta_{x,y} = x + \tau' y$ on $\mathbb{F}_{p^2}$ with respect to the basis $(1, \tau')$ of $\mathbb{F}_{p^2}$ over $\mathbb{F}_p$. The determinant of $d_{x,y}$ is $N(\zeta_{x,y}) = \zeta_{x,y}^{p+1}$, where $N : \mathbb{F}_{p^2}^* \to \mathbb{F}_p^*$ is the norm map.

Let Ω_e denote the set of elliptic semisimple elements in $G = GL(2,\mathbb{F}_p)$. The centralizer of an elliptic element $d_{x,y}$ is the group $\mathbb{F}_{p^2}^*$. Hence the number of elements in the conjugacy class is $p^2 - p$. Since the elliptic classes are parametrized by pairs of elements of the form $\{\zeta, \zeta^p\}$, with $\zeta \in \mathbb{F}_{p^2}^* \backslash \mathbb{F}_p^*$, the number of elliptic conjugacy classes is $(p^2-p)/2$. Thus the cardinality of Ω_e is $(p^2-p)^2/2$.

Let Ω_{prim} be the subset of Ω_e consisting of elements of order p^2-1. From the description of the conjugacy classes we note that these are the elements with maximum order in $GL(2,\mathbb{F}_p)$ and can be thought of as two-dimensional analogues of primitive roots; i.e., (elliptic) generators of $\mathbb{F}_{p^2}^*$. The number of such classes is $(p^2-p)\phi(p^2-1)/2$, where ϕ denotes the Euler ϕ-function. The proportion of these classes in G is given by,

$$\frac{|\Omega_{prim}|}{|G|} = \frac{\phi(p^2-1)(p^2-p)/2}{(p^2-1)(p^2-p)} = \frac{\phi(p^2-1)}{2(p^2-1)}.$$

2.2. **Irreducible representations of $GL(2,\mathbb{F}_p)$.** The irreducible complex representations of $GL(2,\mathbb{F}_p)$ were classified by Schur. Green ([Gr]) constructed the irreducible characters of $GL(n,\mathbb{F}_p)$ parametrized by the conjugacy classes in $GL(n,\mathbb{F}_p)$. We recall the classification of the irreducible complex representations of $G = GL(2,\mathbb{F}_p)$ (see [FH]).

One dimensional representations. The one dimensional representations U_χ, corresponding to the scalar matrices, defined by $U_\chi(A) = \chi(\mathrm{Det}(A))$, where χ is character of $\mathbb{F}_p^*$. There are $(p-1)$ isomorphism classes, and

$$\chi(d_{x,y}) = \chi(N(\zeta_{x,y})). \tag{2.1}$$

Irreducible Principal series. Given a subgroup H of a finite group G, and a representation θ of H on V, a model for the induced representation $\rho = I_H^G(\theta)$ can be taken as follows:

$$I_H^G(\theta) = \{f : G \to V \mid f(gh) = \theta(h)^{-1} f(g) \quad \forall h \in H\}. \tag{2.2}$$

The group G acts on the left: $(\rho(g_0)f)(g) = f(g_0^{-1}g)$ for $g_0, g \in G$.

Let P (resp. P', U, U') denote the subgroups of G consisting of lower triangular (resp. upper triangular, unipotent lower triangular, unipotent upper triangular) matrices in $GL(2, \mathbb{F}_p)$. The principal series representations $I_{\chi,\eta}$ are indexed by pairs of distinct characters χ, η of $\mathbb{F}_p^*$, and correspond to the non-central split semisimple conjugacy classes. Via the exact sequence,

$$1 \to U' \to P' \to (\mathbb{F}_p^*)^2 \to 1,$$

$\chi \oplus \eta$ defines a representation of P', and $I_{\chi,\eta}$ is defined to be the induced representation $I_{P'}^G(\chi \oplus \eta)$. We have isomorphisms $I_{\chi,\eta} \simeq I_{\eta,\chi}$. The dimension of these representations is $p+1$, and the character of these representations vanish on the set of elliptic semisimple conjugacy classes.

Twists of Steinberg. Given a character χ of $\mathbb{F}_p^*$, there is a decomposition,

$$I_{P'}^G(\chi \circ \mathrm{Det}) = St_\chi \oplus \chi \circ \mathrm{Det}.$$

The Steinberg representation St corresponds to the trivial character $1_{P'}$ of P'. The induced representation $I_{P'}^G(1_{P'})$ is the regular action of G on the space of functions on the projective line $\mathbb{P}^1 = G/P'$. Given two functions f_1, f_2 on $\mathbb{P}^1$, an invariant inner product is,

$$\langle f_1, f_2 \rangle = \sum_{x \in \mathbb{P}^1} f_1(x)\overline{f_2(x)}.$$

The Steinberg St is the orthogonal complement of the trivial representation in $I_{P'}^G(1_{P'})$. The underlying space V_{St} for the Steinberg is,

$$V_{St} = \{f : \mathbb{P}^1 \to \mathbb{C} \mid \sum_{x \in \mathbb{P}^1} f(x) = 0\}. \tag{2.3}$$

We have $St_\chi = St \otimes \chi \circ \mathrm{Det}$. These representations correspond to the non-semisimple conjugacy classes. The dimension of these representations is p, and there are $(p-1)$ representations upto isomorphism. The character St_χ on an elliptic semisimple element is given by

$$\mathrm{Tr}\,(St_\chi(d_{x,y})) = -\chi(N(\zeta_{x,y})). \tag{2.4}$$

Cuspidal representations. The cuspidal representations X_ϕ are indexed by characters ϕ of $\mathbb{F}_{p^2}^*$ satisfying $\phi \neq \phi^p$. They are defined by the property that the invariants with respect to the subgroup U' is trivial, and correspond to the elliptic conjugacy classes. The dimension of these representations is $p-1$, and there are $(p^2-p)/2$ distinct cuspidal representations. The character of X_ϕ vanishes on the split semisimple conjugacy classes, and on elliptic conjugacy classes its value is,

$$\mathrm{Tr}\,(X_\phi(d_{x,y})) = -(\phi(\zeta_{x,y}) + \phi(\zeta_{x,y}^p)). \tag{2.5}$$

3. Singular Gauss sums

In this section we compute the trace of $G(\rho, A)$, where A is a singular matrix in $M(2, \mathbb{F}_p)$ and prove Theorem 1.8. We refer to $Tr\,(G(\rho, A))$ as *singular Gauss sums.* When A is the zero matrix,

$$G(\rho, A) = \sum_{X \in G} \rho(X) = \begin{cases} 0 & \text{if } \rho \text{ is irreducible, non-trivial,} \\ |G| & \text{if } \rho \text{ is trivial.} \end{cases}$$

Suppose now A is a non-zero singular matrix. For any $Z \in G = GL(2, \mathbb{F}_p)$

$$G(\rho, ZAZ^{-1}) = \sum_{X \in G} \rho(X) e_p(tr(ZAZ^{-1}X)) = \rho(Z) G(\rho, A) \rho(Z^{-1}).$$

Therefore, as far as determination of the trace of $G(\rho, A)$ is concerned, it is enough to consider the matrices A up to conjugacy:

Semisimple case: $A_a := \begin{pmatrix} a & 0 \\ 0 & 0 \end{pmatrix},\ a \neq 0.$

Nilpotent case: $N = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$

3.1. **A decomposition for the singular Gauss sum.** The calculation of the singular Gauss sums uses a Bruhat type decomposition of $GL(2, \mathbb{F}_p)$.

Lemma 3.1. *Let P (resp. P') and U (resp. U') denote the subgroups of lower triangular (resp upper triangular) and lower unipotent (resp. upper unipotent) matrices in $GL(2, \mathbb{F}_p)$. Then*

$$GL(2, \mathbb{F}_p) = PU' \sqcup Pw = PU' \sqcup wP' \quad \text{and} \quad GL(2, \mathbb{F}_p) = U'wP' \sqcup P', \tag{3.1}$$

where $w = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.

Proof. The second decomposition is the Bruhat decomposition. The first decomposition can be obtained from the Bruhat decomposition $GL(2, \mathbb{F}_p) = PwU \sqcup P$ by multiplying on the right by w, and using the fact that $wUw = U',\ wPw = P'$. □

The group P of upper triangular matrices factorizes as a product $P = U \times M \times L$, where

$$M = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & m \end{pmatrix} : m \neq 0 \right\} \text{ and } L = \left\{ \begin{pmatrix} l & 0 \\ 0 & 1 \end{pmatrix} : \ell \neq 0 \right\}.$$

We shall write an element $X \in PU'$ as

$$X = x_u x_l x_m x_{u'}, \tag{3.2}$$

where

$$x_u = \begin{pmatrix} 1 & 0 \\ u & 1 \end{pmatrix}, \quad x_m = \begin{pmatrix} 1 & 0 \\ 0 & m \end{pmatrix}, \quad x_l = \begin{pmatrix} l & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad x_{u'} = \begin{pmatrix} 1 & u' \\ 0 & 1 \end{pmatrix}.$$

Note that such a representation is unique. We also note that x_l and x_m commute. Similarly we shall write an element $X \in wP'$ as

$$X = w x_l x_m x_{u'}.$$

Corresponding to the first decomposition given in the foregoing lemma, we write

$$G(\rho, A) = G_1(\rho, A) + G_2(\rho, A),$$

where

$$G_1(\rho, A) = \sum_{X \in PU'} \rho(X) e\left(\frac{tr(AX)}{p}\right) \quad \text{and} \quad G_2(\rho, A) = \sum_{X \in wP'} \rho(X) e\left(\frac{tr(AX)}{p}\right). \tag{3.3}$$

We now compute the traces. For $A_a := \begin{pmatrix} a & 0 \\ 0 & 0 \end{pmatrix}$, a semisimple singular matrix,

$$\mathrm{Tr}(A_a x_u x_l x_m x_{u'}) = al \quad \text{and} \quad \mathrm{Tr}(A_a w x_l x_m x_{u'}) = 0. \tag{3.4}$$

When $N = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, the traces are,

$$\mathrm{Tr}(N x_u x_l x_m x_{u'}) = ul \quad \text{and} \quad \mathrm{Tr}(N w x_l x_m x_{u'}) = l. \tag{3.5}$$

3.2. **Vanishing criteria for the singular Gauss sums.** Given a representation $\rho : G \mapsto GL(V)$ and a subgroup H of G, the projection operator $Pr_H \in \mathrm{End}(V)$

$$Pr_H(v) = \left(\frac{1}{|H|} \sum_{h \in H} \rho(h)\right)(v),$$

maps V to the subspace V^H of vectors invariant under H. The operator satisfies the projection property $Pr_H^2 = Pr_H$.

The reason for splitting the singular Gauss sums in terms of the Bruhat decomposition are the following expressions for G_1 and G_2 in terms of projection operators:

$$G_1(\rho, A_a) = \sum_{X \in PU'} e(al/p)\rho(x_u x_l)\rho(x_m x_{u'}) = |MU'| \sum_{\substack{u \in \mathbb{F}_p \\ l \in \mathbb{F}_p^*}} e(al/p)\rho(x_u x_l) Pr_{MU'} \tag{3.6}$$

$$G_2(\rho, A_a) = \sum_{X \in P'} \rho(w)\rho(X) = |P'|\rho(w) Pr_{P'} \tag{3.7}$$

$$G_1(\rho, N) = \sum_{X \in PU'} e(ul/p)\rho(x_u x_l)\rho(x_m x_{u'}) = |MU'| \sum_{\substack{u \in \mathbb{F}_p \\ l \in \mathbb{F}_p^*}} e(ul/p)\rho(x_u x_l) Pr_{MU'} \tag{3.8}$$

$$G_2(\rho, N) = \sum_{X \in P'} e(l/p)\rho(w)\rho(x_l)\rho(x_m x_{u'}) = |MU'|\rho(w) \sum_{l \in \mathbb{F}_p^*} e(l/p)\rho(x_l) Pr_{MU'}. \tag{3.9}$$

As all the above sums involve the projection operator $Pr_{MU'}$, we observe the following easy consequence:

Proposition 3.2. *Let A be a non-zero singular matrix in $M(2, \mathbb{F}_p)$. Suppose ρ is a non-trivial irreducible representation of $GL(2, \mathbb{F}_p)$ acting on the space V_ρ. Then, the singular Gauss sums $G(\rho, A)$ vanish if $V_\rho^{MU'} = (0)$.*

Further, if $V_\rho^{P'} = (0)$, then $G_2(\rho, A_a)$ vanishes.

For the trivial representation 1_G, the singular Gauss sums are equal to $-p(p-1)$.

Proof. Only the part about the trivial representation needs to be proved. We have,

$$\begin{aligned} G_1(1_G, A_a) &= |MU'| \sum_{\substack{u \in \mathbb{F}_p \\ l \in \mathbb{F}_p^*}} e(al/p) = -|MU'|p \\ &= -p^2(p-1). \\ G_2(1_G, A_a) &= |P'| = p(p-1)^2. \end{aligned}$$

Hence, $G(1, A_a) = -p^2(p-1) + p(p-1)^2 = -p(p-1)$.

Similarly,

$$G_1(1, N) = |MU'| \sum_{\substack{u \in \mathbb{F}_p \\ l \in \mathbb{F}_p^*}} e(ul/p) = 0.$$

$$\begin{aligned} \text{and} \quad G_2(1, N) &= |MU'|\rho(w) \sum_{l \in \mathbb{F}_p^*} e(l/p) = -|MU'| \\ &= -p(p-1). \end{aligned}$$

□

3.3. **Vanishing of certain singular Gauss sums.** We now classify those irreducible representations of $GL(2,\mathbb{F}_p)$ whose MU'-invariants are non-zero:

Lemma 3.3. *Let ρ be a non-trivial irreducible representation of $GL(2,\mathbb{F}_p)$ acting on the space V_ρ. Then the invariant space $V_\rho^{MU'}$ is at most one dimensional.*

If the space $V_\rho^{MU'}$ is non-zero, then ρ is isomorphic either to the Steinberg representation St, or one of the irreducible principal series representations $I_{\chi,1}$ with χ a non-trivial character of $\mathbb{F}_P^*$.

The space $V_\rho^{P'}$ is non-zero only for the Steinberg representation.

Proof. Given a representation η of a subgroup H of a finite group G and a representation ρ of G, Frobenius reciprocity gives an isomorphism,

$$\mathrm{Hom}_G(I_H^G(\eta),\rho) \simeq \mathrm{Hom}_H(\eta, \mathrm{Res}_G^H(\rho)), \tag{3.10}$$

where $\mathrm{Res}_G^H(\rho)$ denotes the restriction of ρ to H.

To say that $V_\rho^{MU'}$ is non-zero means that the trivial representation $1_{MU'}$ occurs in the restriction of ρ to MU'. By Frobenius reciprocity, this is equivalent to ρ being a subrepresentation of $I_{MU'}^G(1_{MU'})$. Inducing in stages to P' and then to $G = GL(2,\mathbb{F}_p)$ we have,

$$I_{MU'}^G(1_{MU'}) = I_{P'}^G(I_{MU'}^{P'}(1_{MU'})).$$

Let χ be a character of $\mathbb{F}_p^*$. Consider $\chi \otimes 1_M$, as a character of P' with its M component being trivial, defined by the formula $\chi(x_l x_m x_{u'}) = \chi(l)$. By definition, these characters are trivial on MU'. By Frobenius reciprocity applied to $MU' \subset P'$, these appear as constituents in $I_{MU'}^{P'}(1_{MU'})$. Since the index of MU' in P' is $(p-1)$, dimension count yields an isomorphism,

$$I_{MU'}^{P'}(1_{MU'}) = \oplus_{\chi\in\hat{L}}\chi\otimes 1_M.$$

Hence,

$$I_{MU'}^G(1_{MU'}) = \oplus_{\chi\in\hat{L}} I_{P'}^G(\chi\otimes 1_M).$$

From the classification of irreducible representations of G, we obtain

$$I_{MU'}^G(1_{MU'}) = \oplus_{\chi\in\hat{L},\chi\neq 1_L} I_{\chi,1} \oplus St \oplus 1_G. \tag{3.11}$$

Among these representations, only St and the trivial representation of G have a non-zero subspace of P'-fixed vectors.

As a consequence of Frobenius reciprocity and the fact that the decomposition given by Eq. (3.11) is multiplicity free, it follows that the space of invariant vectors under MU' is at most one-dimensional. □

From Prop. 3.2, Lemma 3.3 and the classification of representations, we conclude the following proposition, proving in particular, Part (1) of Theorem 1.8:

Proposition 3.4. *Let ρ be a non-trivial irreducible representation of $GL(2,\mathbb{F}_p)$ not isomorphic to the Steinberg or to $I_{\chi,1}$ for a non-trivial character χ of $\mathbb{F}_p^*$. Then for any non-zero singular matrix A, $G(\rho,A)=0$.*

For a non-trivial irreducible representation of $GL(2,\mathbb{F}_p)$, $G_2(\rho,A_a)$ vanishes unless ρ is isomorphic to the Steinberg.

3.4. **Invariant elements in induced representations.** In order to calculate the traces of the singular Gauss sums, we calculate explicitly the invariant element and the projection to the space of invariants with respect to the action of MU'.

Given a character θ of P', a model for the induced representation $\rho = I_{P'}^G(\theta)$ is given as follows:

$$I_{P'}^G(\theta) = \{f : G \to \mathbb{C} \mid f(gp') = \theta(p')^{-1} f(g)\}. \tag{3.12}$$

The group G acts on the left: $(\rho(g_0)f)(g) = f(g_0^{-1}g)$ for $g_0, g \in G$. From the Bruhat decomposition $G = U'wP' \sqcup P'$ a collection of left coset representatives for P' in G is given by $U'w$ and the identity element e of G. Thus an element of ρ is determined by its values on $U'w$ and e.

The natural action of $GL(2,\mathbb{F}_p)$ on $\mathbb{F}_p^2$ induces a transitive action of $GL(2,\mathbb{F}_p)$ on the projective line $\mathbb{P}^1(\mathbb{F}_p)$ consisting of the lines through the origin in $\mathbb{F}_p^2$. The identity coset eP' of P' is the isotropy group of the point at 'infinity' given by the line defined by the vector $\begin{pmatrix}1\\0\end{pmatrix}$ in $\mathbb{F}_p^2$. The group U' can be identified with its orbit through the point 'zero' given by $\begin{pmatrix}0\\1\end{pmatrix} = w\begin{pmatrix}1\\0\end{pmatrix}$. This is precisely the affine line $\mathbb{A}^1(\mathbb{F}_p)$. The Weyl element w switches the points zero and infinity of $\mathbb{P}^1(\mathbb{F}_p)$.

It follows that the restriction of ρ to U' splits as a direct sum of two representations:

$$\rho\,|_{U'} \simeq R_{U'} \oplus 1_{U'}, \tag{3.13}$$

where $R_{U'}$ is the regular representation of U' on the space of functions on U'. The trivial representation of U' corresponds to the element of ρ 'supported' at infinity.

Lemma 3.5. *(a) Let χ be a non-trivial character of $\mathbb{F}_p^*$ and $\rho = I_{P'}^G(\chi \otimes 1_M)$ be the irreducible representation of $GL(2,\mathbb{F}_p)$ with the model given by Eq. (3.12).*

Consider the function δ_{in} of G defined by,

$$\delta_{in}(g) = \begin{cases} 0 & \text{if } g \notin P' \\ \chi(l)^{-1} & \text{if } g = x_l x_m x_{u'} \in P'. \end{cases}$$

The function δ_{in} belongs to the space underlying ρ, and spans the one dimensional space of MU'-invariants of ρ. For an element $f \in I_{P'}^G(\chi \oplus 1_M)$,

$$Pr_{MU'}(f) = \left(\frac{1}{|MU'|} \sum_{x \in MU'} \rho(x) \right)(f) = f(e)\delta_{in}. \tag{3.14}$$

(b) The space of MU'-invariant elements of the Steinberg for the model given by Eq. (2.3) is the space spanned by the function $\delta_{in} = \delta_\infty - \frac{1}{p}\delta_{\mathbb{A}^1}$, where δ_∞ is the function supported at 'infinity' with value 1, and $\delta_{\mathbb{A}^1}$ is the characteristic function of $\mathbb{A}^1$.

Given a function $f \in V_{St}$, the projection to the space of MU'-invariants is given by,

$$Pr_{MU'}(f) = f(\infty)\delta_\infty - \frac{(\sum_{x \in \mathbb{A}^1} f(x))}{p}\delta_{\mathbb{A}^1}.$$

In other words, essentially the lemma says that the invariant element is the element in the induced model 'supported' at infinity, where for the Steinberg we need to take the projection to the Steinberg of the function supported at infinity.

Proof. (a) Since MU' respects the Bruhat decomposition $G = U'wP' \sqcup P'$ it follows that $\rho(x_m x_{u'})\delta_{in}$ is supported at the coset P'. From the definition of δ_{in},

$$(\rho(x_m x_{u'})\delta_{in})(e) = \delta_{in}(x_{u'}^{-1} x_m^{-1}) = \delta_{in}(x_m^{-1} x_{u'/m}) = 1.$$

This proves the invariance of δ_{in} under the action of MU'.

To prove the formula for the projection operator, it is sufficient to show that for any function f supported in the finite part $U'wP'$ of G, the projection is zero. Given an element $x_{v'} \in U'$,

$$\begin{aligned} \sum_{m,u'} \rho(x_m x_{u'})(f)(x_{v'} w) &= \sum_{m,u'} f(x_{u'}^{-1} x_m^{-1} x_{v'} w) = \sum_{m,u'} f(x_{u'}^{-1} x_{mv'} x_m^{-1} w) \\ &= \sum_{m,u'} f(x_{mv'-u'} w w x_m^{-1} w) \\ &= \sum_{m,u'} \chi(m) f(x_{mv'-u'} w) = 0. \end{aligned}$$

(b) For the Steinberg, the calculation is immediate given that it is a permutation action of G on $G/P' = \mathbb{P}^1$.

□

3.5. **A formula for the trace.** Equations (3.6, .., 3.9) express the partial Gauss sums G_1 and G_2 as operators of the form TQ, where $Q^2 = Q$ is a projection operator. For such operators, the trace of TQ is computed by restricting the action of T to the image of Q:

Lemma 3.6. *Suppose V is a finite dimensional vector space and $T,\ Q \in End(V)$, where $Q^2 = Q$. Then*

$$Tr(TQ) = Tr(QTQ).$$

Proof.

$$Tr(QTQ) = Tr(TQQ) = Tr(TQ).$$

□

We apply this lemma in the context of Lemma 3.5 and the projection operator $Pr_{MU'}$:

Corollary 3.7. *With notation as in Lemma 3.5, let T be an operator on the space underlying the representation ρ. Then,*

$$\mathrm{Tr}(TPr_{MU'}) = T(\delta_{in})(e),$$

where ρ is as in Part (a) of Lemma 3.5.

When ρ is the Steinberg representation,

$$\mathrm{Tr}(TPr_{MU'}) = T(\delta_{in})(\infty).$$

Proof. The projection operator $Pr_{MU'}$ projects onto the one dimensional space of invariants spanned by δ_{in}. Thus the trace is equal to the multiple of δ_{in} in $Pr_{MU'}TPr_{MU'}(\delta_{in})$.

For the Steinberg, we observe that this multiple is as given in the equation. □

3.6. **Proof of Theorem 1.8.** We now apply Corollary 3.7, to compute the traces of the singular Gauss sums for the principal series representations and Steinberg.

3.6.1. *Irreducible principal series: semisimple case.* Suppose ρ is an irreducible principal series representation $I_{\chi,1}$ with χ a non-trivial character of $\mathbb{F}_p^*$.

$$\mathrm{Tr}(G(\rho, A_a)) = \mathrm{Tr}(G_1(\rho, A_a)) = |MU'| \sum_{u\in\mathbb{F}_p, l\in\mathbb{F}_p^*} e(al/p)\rho(x_u x_l)(\delta_{in})(e)$$
$$= |MU'| \sum_{u\in\mathbb{F}_p, l\in\mathbb{F}_p^*} e(al/p)\delta_{in}(x_l^{-1}x_u^{-1}).$$

Since U acts simply transitively on $\mathbb{P}^1\backslash\{0\}$, only the term $u = 0$ corresponds to the point at infinity and contributes to the trace. Hence,

$$\begin{aligned}\mathrm{Tr}(G(\rho, A_a)) &= |MU'| \sum_{l\in\mathbb{F}_p^*} e(al/p)\chi(l)\\ &= p(p-1)\overline{\chi(a)}G(\chi).\end{aligned}$$

3.6.2. *Irreducible principal series: nilpotent case.* We now consider the case of irreducible principal series and nilpotent conjugacy class N. We calculate first the G_1 term:

$$\begin{aligned}\mathrm{Tr}(G_1(\rho, N)) &= |MU'| \sum_{u\in\mathbb{F}_p, l\in\mathbb{F}_p^*} e(ul/p)(\rho(x_u x_l)(\delta_{in}))(e)\\ &= |MU'| \sum_{u\in\mathbb{F}_p, l\in\mathbb{F}_p^*} e(ul/p)\delta_{in}(x_l^{-1}x_u^{-1}).\end{aligned}$$

Again, only the $u = 0$ contributes to the trace. The sum becomes,

$$\mathrm{Tr}(G_1(\rho, N)) = |MU'| \sum_{l\in\mathbb{F}_p^*} \chi(l) = 0.$$

Similarly, the G_2-term can be calculated:

$$\begin{aligned}\mathrm{Tr}(G_2(\rho, N)) &= |MU'| \sum_{l\in\mathbb{F}_p^*} e(l/p)\rho(x_l)(\delta_{in})(e) = |MU'| \sum_{l\in\mathbb{F}_p^*} e(l/p)\delta_{in}(x_l^{-1})\\ &= |MU'| \sum_{l\in\mathbb{F}_p^*} e(l/p)\chi(l) = p(p-1)G(\chi).\end{aligned}$$

Hence,

$$\begin{aligned}\mathrm{Tr}(G(\rho, N)) &= \mathrm{Tr}(G_1(\rho, N)) + \mathrm{Tr}(G_2(\rho, N))\\ &= p(p-1)G(\chi).\end{aligned}$$

3.6.3. *Steinberg: semisimple case.* We consider now the Steinberg representation. By Corollary 3.7,

$$\mathrm{Tr}(G_1(St, A_a)) = |MU'| \sum_{u\in\mathbb{F}_p, l\in\mathbb{F}_p^*} e(al/p)(\rho(x_u x_l)(\delta_\infty - \frac{1}{p}\delta_{\mathbb{A}^1})(\infty).$$

The group U fixes 0 of $\mathbb{P}^1$ and acts by translations on $\mathbb{P}^1\backslash\{0\}$. Hence for the δ_∞ term, only $u = 0$ contributes non-trivially. Hence,

$$|MU'| \sum_{u\in\mathbb{F}_p, l\in\mathbb{F}_p^*} e(al/p)(\rho(x_u x_l)(\delta_\infty)(\infty) = |MU'| \sum_{l\in\mathbb{F}_p^*} e(al/p) = -p(p-1).$$

Similarly, for the δ_{A^1}-term, the contribution comes from non-zero u. Taking infinity to be given by the column vector $\begin{pmatrix}1\\0\end{pmatrix}$, the calculation becomes,

$$\begin{aligned}|MU'|\sum_{u\in\mathbb{F}_p,l\in\mathbb{F}_p^*}e(al/p)(\rho(x_ux_l)\frac{1}{p}\delta_{A^1}(\infty) &= |M|\sum_{u,l\in\mathbb{F}_p^*}e(al/p)\delta_{\mathbb{A}^1}\left(\begin{pmatrix}l^{-1}\\-u\end{pmatrix}\right)\\ &= |M|\sum_{u,l\in\mathbb{F}_p^*}e(al/p) = -(p-1)^2.\end{aligned}$$

Hence,

$$\mathrm{Tr}(G_1(St,A_a)) = -p(p-1)+(p-1)^2 = -(p-1).$$

By Eq. (3.7), the second sum becomes,

$$\mathrm{Tr}(G_2(St,A_a)) = |P'|\rho(w)\delta_{in}(\infty) = |P'|\delta_{in}(0) = -\frac{|P'|}{p}\delta_{\mathbb{A}^1}(0) = -(p-1)^2.$$

Hence

$$\begin{aligned}\mathrm{Tr}(G(St,A_a)) &= \mathrm{Tr}(G_1(St,A_a)) + \mathrm{Tr}(G_2(St,A_a))\\ &= -(p-1)-(p-1)^2\\ &= -p(p-1).\end{aligned}$$

3.6.4. *Steinberg: nilpotent case.* When the conjugacy class of A is nilpotent, we argue as above in the semisimple case, considering the sum over $u=0$ and u non-zero separately. From Eq. (3.8) and Corollary 3.7, $\mathrm{Tr}(G_1(St,N))$ is equal to

$$\begin{aligned}|MU'|\sum_{l\in\mathbb{F}_p^*}(\rho(x_l)(\delta_\infty)(\infty) - |M|\sum_{u,l\in\mathbb{F}_p^*}e(ul/p)(\rho(x_ux_l)(\delta_{\mathbb{A}^1})(\infty)\\ = |LMU'| - |M|\sum_{u,l\in\mathbb{F}_p^*}e(ul/p) = p(p-1)^2+(p-1)^2\\ = (p+1)(p-1)^2.\end{aligned}$$

From Eq. (3.9), the second sum becomes,

$$\begin{aligned}\mathrm{Tr}(G_2(St,N)) &= |MU'|\rho(w)\sum_{l\in\mathbb{F}_p^*}e(l/p)\rho(x_l)\delta_{in}(\infty)\\ &= |MU'|\sum_{l\in\mathbb{F}_p^*}e(l/p)\rho(x_l)\delta_{in}(0)\\ &= -\frac{|MU'|}{p}\sum_{l\in\mathbb{F}_p^*}e(l/p)\rho(x_l)\delta_{\mathbb{A}^1}(0)\\ &= -|M|\sum_{l\in\mathbb{F}_p^*}e(l/p)\\ &= (p-1).\end{aligned}$$

Hence

$$\begin{aligned}\mathrm{Tr}(G(St,N)) &= \mathrm{Tr}(G_1(St,N)) + \mathrm{Tr}(G_2(St,N)) \\ &= (p-1) + (p-1)^2(p+1) \\ &= p^2(p-1).\end{aligned}$$

This proves Theorem 1.8.

Remark 3.8. It will be interesting to figure out the nature of these singular traces for general $GL(n, \mathbb{F}_p)$. To try to make sense of these values in terms of the parametrization of the representations by the conjugacy classes, we make two definitions:

Definition 3.9. An irreducible representation ρ of $GL(2, \mathbb{F}_p)$ to be of *unit class* if the semisimple part of the conjugacy class parametrizing it has 1 as an eigenvalue.

Definition 3.10. The *unit multiplicity* $k(\rho)$ of an irreducible representation ρ is defined to be the multiplicity of the eigenvalue 1 in the semisimple part of the conjugacy class parametrizing it.

The unit multiplicity appears as a 'defect' term in Kondo's estimate for the non-abelian Gauss sum:

$$|g(\rho)| = p^{(n^2 - k(\rho))/2}.$$

From the classification given by Theorem 1.8, we see that the non-trivial unit class representations of $GL(2, \mathbb{F}_p)$ is isomorphic to either the trivial or Steinberg or to the prinicipal series representation $I_{\chi,1}$ for some non-trivial character χ of $\mathbb{F}_p^*$. These are precisely the representations that occur in the induced representation $I_{MU'}^G(1)$. Theorem 1.8 says that the singular Gauss sums does not vanish precisely for the representations of unit class.

4. Proof of the $GL(2)$ Polya-Vinogradov theorem

As we have already obtained the bound for Gauss sums, what remains in order to prove Theorem 1.4 is the Fourier analytic part which we develop here. First we recall some basic facts from Fourier Analysis on finite abelian groups and then we proceed as in the standard proofs of the classical Polya-Vinogradov Theorem. We consider the case of general $n \times n$ matrices until the point when we need to apply the Gauss sum bound.

4.1. **Fourier analysis on finite groups.** Let G be a finite group. Let $\hat{G}$ denote the set of isomorphism classes of irreducible complex representations of G. For $\rho \in \hat{G}$, let χ_ρ denote its character. The space of complex valued functions on G carries an inner product,

$$\langle f_1, f_2 \rangle = \frac{1}{|G|} \sum_{x \in G} f_1(x)\overline{f_2(x)},$$

where f_1, f_1 are complex valued functions on G, and $|G|$ denotes the cardinality of G. With respect to this inner product, the characters of G form an orthonormal basis for the conjugation invariant functions on G. On the space of functions on $\hat{G}$, define the inner product

$$\langle \phi_1, \phi_2 \rangle = \sum_{\rho \in \hat{G}} \phi_1(\rho)\overline{\phi_2(\rho)},$$

where ϕ_1, ϕ_1 are complex valued functions on $\hat{G}$. For a conjugacy invariant function f on G, its Fourier transform $\hat{f}$ is a function on $\hat{G}$, defined by $\hat{f}(\rho) = \langle f, \chi_\rho \rangle$. With these normalizations, the Fourier transform $f \mapsto \hat{f}$ is an isometry from conjugacy invariant functions on G to functions on $\hat{G}$.

4.2. **Dual of $M(n, \mathbb{Z}/p\mathbb{Z})$.** We specialize the foregoing discussion to the case when $G = M(n, \mathbb{Z}/p\mathbb{Z})$, where p is a prime number. Denote by e the exponential function $e(x) = \exp(2\pi i x),\ x \in \mathbb{C}$. From the identification of the finite field $\mathbb{F}_p$ with $\mathbb{Z}/p\mathbb{Z}$, we have an additive character e_p of $\mathbb{F}_p$ given by $e_p(x) = e(x/p)$. Let n be a positive integer. For each matrix $A \in M(n, \mathbb{Z}/p\mathbb{Z})$, consider the character $\psi_A(X) = e_p(\mathrm{Tr}(AX))$. We have,

Lemma 4.1. *The map $A \mapsto \psi_A$ yields an isomorphism of $M(n, \mathbb{Z}/p\mathbb{Z})$ with its dual group $\widehat{M(n, \mathbb{Z}/p\mathbb{Z})}$.*

Proof. Since for each non-zero matrix $A \in M(n, \mathbb{Z}/p\mathbb{Z})$, there exists a matrix X with $\mathrm{Tr}(AX) \neq 0$, the map $A \mapsto \psi_A$ is injective. Hence the lemma follows by comparing the cardinalities. □

For functions $\phi, \phi' : M(n, \mathbb{Z}/p\mathbb{Z}) \longrightarrow \mathbb{C}$, the isometry of Fourier transform translates to the following Plancherel formula:

$$\frac{1}{p^{n^2}} \sum_{A \in M(n, \mathbb{Z}/p\mathbb{Z})} \phi(A)\overline{\phi'(A)} = \sum_{A \in M(n, \mathbb{Z}/p\mathbb{Z})} \widehat{\phi}(A)\overline{\widehat{\phi'}(A)}. \tag{4.1}$$

4.3. **A general estimate for box sums.** Let $\mathbf{I}$ be an n^2-dimentional matrix interval in $M(n, \mathbb{Z})$; i.e., $\mathbf{I}$ is the Cartesian product of n^2 many intervals $I_{ij}, 1 \leq i, j \leq n$ for each entry, where each I_{ij} is an interval in $\mathbb{Z}$. We may assume without loss of generality that the length of the interval I is at most p. Let ϕ be a complex

valued function on $M(n, \mathbb{Z}/p\mathbb{Z})$. The following proposition gives an estimate of the general sum

$$S(\phi, \mathbf{I}) = \sum_{A \in \mathbf{I}} \phi(A).$$

Proposition 4.2. *Suppose each component interval I_{ij} has length $|I_{ij}| \leq cp$, where $c > 0$ is a constant. Then we have the estimate*

$$S(\phi, \mathbf{I}) \ll ||\hat{\phi}||_\infty p^{n^2} (\log p)^{n^2}$$

Moreover, for $p \geq 11$, the implied constant can be taken to be $\left(\frac{c+3}{2}\right)^{n^2}$.

By Eq. (4.1), we have

$$p^{-n^2} S(\phi, \mathbf{I}) = \sum_{B \in M(n, \mathbb{F}_p)} \hat{\phi}(B) \overline{\hat{\delta}_{\mathbf{I}}}(B), \tag{4.2}$$

which yields the bound

$$|S(\phi, \mathbf{I})| \leq p^{n^2} ||\hat{\phi}||_\infty \sum_{B \in M(n, \mathbb{F}_p)} \left| \hat{\delta}_{\mathbf{I}}(B) \right|. \tag{4.3}$$

Hence we need to bound the above sum over B and this will be done in the next few lemmas.

Lemma 4.3. *For any real number α, we have the bound*

$$\sum_{1 \leq n \leq N} e(n\alpha) \leq min\left(N, \frac{1}{2||\alpha||}\right), \tag{4.4}$$

where $||\alpha||$ is the distance of α from the nearest integer.

Proof. This is quite standard. See, e.g., [Mo1, Chap. 3]. □

Now we prove a lemma that gives an estimate for $\widehat{\delta_{\mathbf{I}}}(B)$:

Lemma 4.4.

$$|\widehat{\delta_{\mathbf{I}}}(B)| \leq p^{-n^2} \prod_{1 \leq i,j \leq n} min\left(cp, \frac{1}{||b_{ij}/p||}\right).$$

Proof.

$$\begin{aligned} \widehat{\delta_{\mathbf{I}}}(B) &= \frac{1}{p^{n^2}} \sum_{X \in \mathscr{S}_p} \delta_{\mathbf{I}}(X) \psi_X(-B) \\ &= \frac{1}{p^{n^2}} \sum_{X \in \mathbf{I}} e\left(\frac{-\mathrm{Tr}(BX)}{p}\right). \end{aligned}$$

Now the sum over X factors as

$$\prod_{i,j} \sum_{x_{ji} \in \mathbf{I}_{ji}} e\left(\frac{b_{ij} x_{ji}}{p}\right).$$

Since for every (i,j) the interval $\bar{\mathbf{I}}_{ji}$ is of length at most cp, an application of Eq. (4.4) yields the bound

$$\begin{aligned}\sum_{x_{ji}\in\bar{\mathbf{I}}_{ji}} e\left(\frac{b_{ij}x_{ji}}{p}\right) &\leq min\left(|\bar{\mathbf{I}}_{ji}|, \frac{1}{||b_{ij}/p||}\right)\\ &\leq min\left(cp, \frac{1}{||b_{ij}/p||}\right).\end{aligned}$$

The lemma follows by taking product over all the entries. □

We now consider the sum over B.

Lemma 4.5.

$$\sum_{B\in M(n,\mathbb{F}_p)} \left|\widehat{\delta_{\bar{\mathbf{I}}}}(B)\right| \ll (\log p)^{n^2}.$$

For $p \geq 11$, the implied constant can be taken to be $\left(\frac{c+3}{2}\right)^{n^2}$.

Proof. By the above lemma,

$$\sum_{B\in M(n,\mathbb{F}_p)} \left|\widehat{\delta_{\bar{\mathbf{I}}}}(B)\right| \leq \frac{1}{p^{n^2}} \sum_{B\in M(n,\mathbb{F}_p)} \prod_{i,j} min\left(cp, \frac{1}{||b_{ij}/p||}\right).$$

Since B is varying over the set of all $n\times n$ matrices over $\mathbb{F}_p$, for each (i,j), b_{ij} varies form 0 to $p-1$ and hence the above sum of products can be written as a product of sums as follows:

$$\sum_{B\in M(n,\mathbb{F}_p)} \prod_{i,j} min\left(cp, \frac{1}{||b_{ij}/p||}\right) = \prod_{i,j} \sum_{0\leq b_{ij}\leq p-1} min\left(cp, \frac{1}{||b_{ij}/p||}\right).$$

Now we bound the individual sums. We have,

$$\begin{aligned}\sum_{0\leq b_{ij}\leq p-1} min\left(cp, \frac{1}{||b_{ij}/p||}\right) &\leq cp + p\sum_{1\leq b\leq p-1}\frac{1}{b}\\ &\leq cp + p(1+\log p)\\ &\leq \left(\frac{c+3}{2}\right) p\log p,\end{aligned}$$

provided that $\log p \geq 2$; i.e., $p \geq 11$.

Hence, for $p \geq 11$,

$$\sum_{B\in M(n,\mathbb{F}_p)} \left|\widehat{\delta_{\bar{\mathbf{I}}}}(B)\right| \leq ((c+1)\log p)^{n^2}.$$

For smaller primes, a similar bound holds with a different constant. □

The proof of Prop. 4.2 is now clear from Lemma 4.5 and Eq. (4.3).

4.4. **Estimate for $\widehat{\chi_\rho}$.** Suppose ρ is an irreducible complex representation of $GL(2, \mathbb{F}_P)$. Extend the character χ_ρ of ρ to a function on $M(n, \mathbb{Z}/p\mathbb{Z})$ by defining it to be zero on singular matrices. Then,

$$\begin{aligned}\widehat{\chi_\rho}(A) = \langle \chi_\rho, \psi_A \rangle &= \frac{1}{p^4} \sum_{X \in M(2,\mathbb{Z}/p\mathbb{Z}} \chi_\rho(X)\overline{\psi_A(X)} \\ &= \frac{1}{p^4}\mathrm{Tr}(G(\rho, -A)).\end{aligned}$$

As a consequence of Theorem 1.10, we have:

Proposition 4.6. *Let ρ be a non-trivial irreducible complex representation of $GL(2, \mathbb{F}_P)$ and A a non-zero matrix. Then*

$$|\widehat{\chi_\rho}(A)| \leq d(\rho)p^{-2},$$

where $d(\rho)$ is the dimension of ρ.

4.5. **Proof of Theorem 1.4.** We now prove Theorem 1.4. Let ρ be an irreducible, complex representation of $GL(2, \mathbb{F}_p)$. In the foregoing notation the sum we want to estimate is,

$$S(\chi_\rho, \mathbf{I}) = \sum_{A \in \mathbf{I}} \chi_\rho(A),$$

where $\mathbf{I}$ is a matrix interval of the form

$$I = \prod_{ij} I_{ij},$$

where, for each pair (i, j) I_{ij} is an interval of length $|I_{ij}| \leq cp$. By Prop. 4.2, we get

$$S(\chi_\rho(A), \mathbf{I}) \leq ||\widehat{\chi}_\rho||_\infty p^4 \left((\left(\frac{c+3}{2} \right) \log p \right)^4, \tag{4.5}$$

and by Theorem 4.6,

$$||\widehat{\chi}_\rho||_\infty \leq p^{-2}d(\rho).$$

This proves Theorem 1.4.

5. Growth of elliptic elements: Proof of Theorem 1.13

In this section, we give an estimate for the function $S(\delta_{\Omega_e}, x)$ that countins the number of integer matrices of height up to x that reduce to elliptic elements modulo p. In other words, we need to count integer matrices of height up to x for which the characteristic polynomials are irreducible over $\mathbb{F}_p$; i.e., integer matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ of height up to x such that $(\mathrm{Tr})^2 - 4(\mathrm{Det}) = (a-d)^2 + 4bc$ is not a quadratic residue modulo p.

Let χ denote the Legendre symbol modulo p and let Δ denote the collection of elements in $M(2,\mathbb{F}_p)$ that have characteristic polynomials with discriminant divisible by p. Consider the sum

$$S=\frac{1}{2}\sum_{0\leq|a|,|b|,|c|,|d|\leq x}\sum\sum\sum\sum\left\{1-\chi\left((a-d)^2+4bc\right)\right\}. \tag{5.1}$$

When p divides $\mathrm{Det}(A)$, then the discriminant is always a square modulo p. Hence,

$$S=S(\delta_{\Omega_e},x)+\frac{1}{2}S(\delta_\Delta,x). \tag{5.2}$$

Now, from Eq. (5.1),

$$S=\frac{1}{2}(2[x]+1)^4-\frac{1}{2}S', \tag{5.3}$$

where

$$S'=\sum_{0\leq|a|,|b|,|c|,|d|\leq x}\sum\sum\sum\sum\chi((a-d)^2+4bc).$$

When p divides b, then $\chi((a-d)^2+4bc)$ is identically 1, unless $a\equiv d(\mathrm{mod}\ p)$ when it vanishes. Thus the contribution of terms with $p|b$ to S' is:

$$(2x/p+O(1))(2x)^3+O(x^3/p)-(2x/p+O(1))(2x)^2(2x/p+O(1))=16\left(\frac{1}{p}-\frac{1}{p^2}\right)x^4+O(x^3).$$

When b is invertible in $\mathbb{F}_p$, we pull it out in order to obtain a sum over c varying in an interval which can be estimated by the classical Polya-Vinogradov bound (1.1). The sum over the other three variables is bounded trivially. Thus the contribution of terms with $b\not\equiv 0(\mathrm{mod}\ p)$ is

$$\begin{aligned}
&\sum_{\substack{0\leq|a|,|b|,|c|,|d|\leq x\\ b\not\equiv 0(\mathrm{mod}\ p)}}\sum\sum\sum\sum\chi((a-d)^2+4bc)\\
&=\sum_{0<|a|,|d|\leq x}\sum\sum_{0<|b|\leq x,b\not\equiv 0(\mathrm{mod}\ p)}\sum\chi(4b)\sum_{0<|c|\leq x}\chi((4b)^{-1}(a-d)^2+c)\\
&\leq\sum_{0\leq|a|,|d|\leq x}\sum\sum_{0\leq|b|\leq x,b\not\equiv 0(\mathrm{mod}\ p)}\sum|\chi(4b)|\left|\sum_{0<|c|\leq x}\chi((4b)^{-1}(a-d)^2+c)\right|\\
&\ll\sum_{0<|a|,|b|,|d|\leq x}\sum\sum\sum\sqrt{p}\log p\\
&\ll x^3\sqrt{p}\log p.
\end{aligned}$$

Hence,

$$S=8x^4-\frac{8x^4}{p}+\frac{8x^4}{p^2}+O(x^3\sqrt{p}\log p). \tag{5.4}$$

It remains to estimate $S(\delta_\Delta,x)$ which is the content of the next Lemma.

Lemma 5.1. *The number of matrices* $\begin{pmatrix}a&b\\c&d\end{pmatrix}$ *of height up to* x *and with* $(a-d)^2+4bc\equiv 0(\mathrm{mod}\ p)$ *is* $\frac{16x^4}{p}+O(x^3)$.

Proof. We need to count 4-tuples (a, b, c, d) such that $-x \leq a, b, c, d \leq x$ and $(a-d)^2 \equiv -4bc (\mathrm{mod}\ p)$. First we note that the number of integers in the interval $[-x, x]$ is $2[x] + 1 = 2x + O(1)$ and the number of integers in this interval that are divisible by p or lies in a fixed residue class modulo p is $2[x/p] + 1 = 2x/p + O(1)$. The number of pairs (a, d) with $a \equiv d (\mathrm{mod}\ p)$ is, therefore,

$$(2x + O(1))(2x/p + O(1)) = 4x^2/p + O(x),$$

and for each such a pair, the number of possible pairs (b, c), i.e., with the property $bc \equiv 0 (\mathrm{mod}\ p)$ is

$$2(2x - 2x/p + O(1))(2x/p + O(1)) + (2x/p + O(1))^2 = 4x^2(2/p - 1/p^2) + O(x).$$

On the other hand, the number of pairs (a, d) with $a \not\equiv d (\mathrm{mod}\ p)$ is

$$(2x + O(1))(2x - 2x/p + O(1)) = 4x^2(1 - 1/p) + O(x).$$

For each such pair, fixing any $b \not\equiv 0 (\mathrm{mod}\ p)$ will determine c modulo p. Hence, for each pair (a, d) with $a \not\equiv d (\mathrm{mod}\ p)$ there is a total of

$$(2x - 2x/p + O(1))(2x/p + O(1)) = 4x^2(1/p - 1/p^2) + O(x)$$

many pairs (b, c). Hence the total number we want is

$$\begin{aligned}(4x^2/p + O(x))(4x^2(2/p - 1/p^2) + O(x)) + (4x^2(1 - 1/p) + O(x))(4x^2(1/p - 1/p^2) + O(x)) \\ = 16x^4/p + O(x^3).\end{aligned}$$

□

From Eq. (5.4), (5.2) and Lemma 5.1, we have,

$$\begin{aligned}S(\delta_{\Omega_e}, x) &= S - \frac{1}{2}S(\delta_\Delta, x) \\ &= 8x^4 - \frac{8x^4}{p} + \frac{8x^4}{p^2} + O(x^3\sqrt{p}\log p) - \frac{8x^4}{p} + O(x^3) \\ &= 8\left(1 - \frac{2}{p} + \frac{1}{p^2}\right)x^4 + O(x^3\sqrt{p}\log p).\end{aligned}$$

This proves Prop. 1.13.

6. Growth of primitive elements: Proof of Theorem 1.15

In this section our principal interest is in the elliptic semisimple conjugacy classes (see §2.1). Our goal is to count integer matrices of height up to X that reduces to a primitive element modulo p.

6.1. **Fourier expansion of $\delta_{\Omega_{prim}}$.** In order to estimate $S(\delta_{\Omega_{prim}}, x)), x)$, we begin by following the method given in §1.4. First we expand the characteristic function of $\delta_{\Omega_{prim}}$ in a finite Fourier series. Denoting c_{χ_ρ} by c_ρ for ease of notation, we write

$$\delta_{\Omega_{prim}} = \sum_{\rho \in \hat{G}} c_\rho \cdot \chi_\rho, \tag{6.1}$$

where ρ varies over the set of irreducible representations of G and the Fourier coefficients c_ρ are given by

$$c_\rho = \langle \delta_{\Omega_{prim}}, \chi_\rho \rangle = \frac{1}{|G|} \sum_{\omega \in \Omega_{prim}} \overline{\chi_\rho(\omega)}.$$

Let T be the collection of conjugacy classes consisting of primitive element in G. Each conjugacy class $t \in T$ is of size $(p^2 - p)$ and is defined by a pair $\{\zeta_t,\ \zeta_t^p\}$, where ζ_t generates $\mathbb{F}_{p^2}^*$. Also, recall that $|G| = (p^2 - p)(p^2 - 1)$. Thus we have the following formula for the Fourier coefficients:

$$c_\rho = \frac{1}{p^2 - 1} \sum_{t \in T} \overline{\chi_\rho(\zeta_t)}. \tag{6.2}$$

The next proposition gives estimates for the Fourier coefficients for different types of characters.

Proposition 6.1. *(i) For the one-dimensional representation $\rho = U_\eta$, where $\eta : \mathbb{F}_p^* \mapsto \mathbb{C}^*$ is a character,*

$$c_{U_\eta} = \frac{1}{2} \sum_{\substack{d | p^2 - 1 \\ ord(\eta) | d}} \frac{\mu(d)}{d}; \tag{6.3}$$

in particular, for the trivial character 1_G, the corresponding Fourier coefficient is given by

$$c_1 = c_{1_G} = \frac{|\Omega_{prim}|}{|G|}.$$

(ii) For the Steinberg representation St and its twists by characters St_η, we have

$$c_{U_\eta} = -c_{St_\eta} = -\frac{1}{2} \sum_{\substack{d | p^2 - 1 \\ ord(\eta) | d}} \frac{\mu(d)}{d}. \tag{6.4}$$

(iii) For the principal series representation $I_{\chi,\eta}$,

$$c_{I_{\chi,\eta}} = 0. \tag{6.5}$$

(iv) For the cuspidal representation X_ϕ,

$$c_{X_\phi} = \sum_{\substack{d | p^2 - 1 \\ ord(\phi) | d}} \frac{\mu(d)}{d}. \tag{6.6}$$

Before proving this, we recall a lemma expressing the characteristic function of the set of generators of a cyclic group in terms of characters of the group (see, e.g., [Sh, Eq. (8.5.3), page 302]).

Lemma 6.2. *Let m be a natural number and let C_m be the cylic group of order m. Let P be the subset consisting of generators of C_m. Then*

$$\delta_P = \sum_{d|m} \frac{\mu(d)}{d} \sum_{\chi^d=\chi_0} \chi, \tag{6.7}$$

where $\chi : C_m \to \mathbb{C}^*$ are characters of C_m, and χ_0 is the trivial character.

Proof. We work with $C_m \simeq \mathbb{Z}/m\mathbb{Z}$. A set of representatives for P is given by the natural numbers n up to m and coprime to m. From the properties of Möbius μ-function,

$$\delta_P(n) = \sum_{d|(n,m)} \mu(d).$$

Let ξ_d be the indicator function:

$$\xi_d(n) = \begin{cases} 1 & \text{if } d|n \\ 0 & \text{otherwise.} \end{cases}$$

From orthogonality of characters,

$$\xi_d(n) = \frac{1}{d} \sum_{\chi^d=\chi_0} \chi(n).$$

Hence,

$$\delta_P(n) = \sum_{d|m} \mu(d)\xi_d(n) = \sum_{d|m} \frac{\mu(d)}{d} \sum_{\chi^d=\chi_0} \chi(n).$$

□

Now we prove Prop. 6.1.

Proof. Let $N : \mathbb{F}_{p^2}^* \to \mathbb{F}_p$ be the norm map. Then, by (6.2),

$$\begin{aligned} c_{U_\eta} &= \frac{1}{p^2-1} \sum_{t\in T} \overline{\eta}(N(\zeta_t)) \\ &= \frac{1}{2(p^2-1)} \sum_{\langle\zeta\rangle=\mathbb{F}_{p^2}^*} \overline{\eta}(N(\zeta)), \end{aligned}$$

where the factor $1/2$ is to account for the fact that the same conjugacy class is generated by both ζ and ζ^p. By Lemma 6.2,

$$\begin{aligned} c_{U_\eta} &= \frac{1}{2(p^2-1)} \sum_{\zeta \in \mathbb{F}_p^*} \Big(\sum_{d|p^2-1} \frac{\mu(d)}{d} \sum_{\substack{\chi \in \widehat{\mathbb{F}_{p^2}^*} \\ \chi^d = \chi_0}} \chi(\zeta) \Big) \overline{\eta \circ N}(\zeta) \\ &= \frac{1}{2(p^2-1)} \sum_{d|p^2-1} \frac{\mu(d)}{d} \sum_{\chi^d=\chi_0} \sum_{\zeta \in \mathbb{F}_p^*} (\chi \, \overline{\eta \circ N})(\zeta) \\ &= \frac{1}{2} \sum_{d|p^2-1} \frac{\mu(d)}{d} \sum_{\chi^d=\chi_0} \sum_{\zeta \in \mathbb{F}_p^*} \delta_{\chi = \eta \circ N} \\ &= \frac{1}{2} \sum_{\substack{d|p^2-1 \\ \mathrm{ord}(\eta)|d}} \frac{\mu(d)}{d}, \end{aligned}$$

by orthogonality of characters and the observation that $\mathrm{ord}(\eta \circ N) = \mathrm{ord}(\eta)$. Note that the condition that η^d is the trivial character translates to the condition that the order of η divides d. This proves part (i).

Now we consider part (iii). The character of a representation induced from the Borel subgroup (say, upper triangular matrices P') in G is supported on the conjugacy classes which intersect P'. By definition, the elliptic classes cannot be conjugated into P'. This proves (iii).

Part (ii) follows from the fact that $Ind_{P'}^G(\eta \oplus \eta_{P'}) = \eta \circ \mathrm{Det} \oplus St_\eta$, where η is a character of $\mathbb{F}_p^*$, and $\eta \oplus \eta$ is considered as a character of P' via the projection $P' \to \mathbb{F}_p^* \oplus \mathbb{F}_p^*$. Hence $c_{U_\eta} = -c_{St_\eta}$.

Now we prove part (iv). We have

$$c_{X_\phi} = \frac{1}{p^2-1} \sum_{t \in T} -(\overline{\phi(\zeta_t)} + \overline{\phi(\zeta_t^p)}) = -\frac{1}{(p^2-1)} \sum_{\zeta} \overline{\phi(\zeta)},$$

where the last sum runs over all generators ζ of $\mathbb{F}_{p^2}^*$. From Lemma 6.2,

$$c_{X_\phi} = -\frac{1}{(p^2-1)} \sum_{\zeta \in \mathbb{F}_{p^2}^*} \sum_{d|p^2-1} \frac{\mu(d)}{d} \sum_{\chi^d=\chi_0} \chi\overline{\phi}(\zeta).$$

Interchanging the order of summation, we get

$$c_{X_\phi} = -\frac{1}{(p^2-1)} \sum_{d|p^2-1} \frac{\mu(d)}{d} \sum_{\chi^d=\chi_0} \sum_{\zeta \in \mathbb{F}_{p^2}^*} \chi\overline{\phi}(\zeta).$$

By orthogonality, the last sum is zero unless $\chi = \phi$ and this proves (iv). □

We give now an estimate for the sum of the Fourier coefficients.

Lemma 6.3. *We have the estimates*

$$(i)\sum_{\eta}|c_{U_\eta}| \leq \tau(p^2-1);$$

$$(ii)\sum_{\eta}|c_{St_\eta}| \leq \tau(p^2-1);$$

$$(iii)\sum_{X_\phi}|c_{X_\phi}| \leq \tau(p^2-1).$$

Here $\tau(n)$ denotes the number of divisors of n.

Proof. Note that (ii) follows from (i) because $c_{U_\eta} = -c_{St_\eta}$. For part (i), we partition the sum according to the orders of the characters η and apply the above proposition and estimate the sum as follows:

$$\begin{aligned}\sum_{\alpha}|c_{U_\alpha}| &\leq \frac{1}{2}\sum_{m|p-1}\phi(m)\sum_{\substack{d|p^2-1\\ m|d}}\frac{1}{d}\\ &= \sum_{d|p^2-1}\frac{1}{d}\sum_{m|(d,p-1)}\phi(m)\\ &\leq \sum_{d|p^2-1}\frac{1}{d}\sum_{m|d}\phi(m)\\ &= \tau(p^2-1),\end{aligned}$$

where ϕ above denotes the Euler ϕ-function and we have used elementary result $\sum_{m|d}\phi(m) = d$. For (iii) we recall that cuspidal representations are parametrized by characters ϕ of $\mathbb{F}_{p^2}^*$ satisfying $\phi \neq \phi^p$. Suppose ϕ_1 is a generator of the group of all characters of $\mathbb{F}_{p^2}^*$. Then ${\phi_1}^j$ for $j = 1, 2, \cdots, p^2-1$ are all the characters. For estimating the sum in question, we first enlarge the set to include all the characters and then divide the sum according to the order of the characters. Note that the number of characters of order m is $\phi(m)$. Thus we obtain,

$$\begin{aligned}\sum_{X_\phi}|c_{X_\phi}| &\leq \sum_{m|p^2-1}\phi(m)\sum_{\substack{d|p^2-1\\ m|d}}\frac{1}{d}\\ &= \sum_{d|p^2-1}\frac{1}{d}\sum_{m|d}\phi(m),\\ &= \tau(p^2-1).\end{aligned}$$

□

6.2. **Application of the $GL(2)$ Polya-Vinogradov estimate.** At this stage a direct application of Theorem 1.2 and Lemma 6.3 easily gives us the following:

Proposition 6.4.

$$S(\delta_{\Omega_{prim}}, x) = \frac{8\gamma_p \phi(p^2-1)}{(p^2-1)} x^4 + O(x^3) + O\left(p^{3+\varepsilon}\right), \tag{6.8}$$

where $\gamma_p = 1 - 1/p - 1/p^2 + 1/p^3$.

Proof. We are interested in the sum

$$S(\delta_{\Omega_{prim}}, x) := \sum_{h(A)\leq x} \delta_{\Omega_{prim}}(\overline{A}), \tag{6.9}$$

which, after an application of (6.1) and interchange of summation, becomes

$$\sum_{\rho} c_\rho \sum_{h(A)\leq x} \chi_\rho(A), \tag{6.10}$$

from which we isolate the contribution of the trivial character. Thus we obtain

$$\begin{aligned} S(\delta_{\Omega_{prim}}, x) &= \frac{|\Omega_{prim}|}{|G|} \sum_{h(A)\leq x} \chi_1(A) + \sum_{\chi\neq\chi_1} c_\rho \sum_{h(A)\leq x} \chi_\rho(A) \\ &= \frac{|\Omega_{prim}|}{|G|} 16\gamma_p x^4 + O(x^3) + O\left(p^3 (\log p)^4 \sum_{\rho} |c_\rho|\right) \\ &= \frac{8\gamma_p \phi(p^2-1)}{(p^2-1)} x^4 + O(x^3) + \left(p^{3+\varepsilon}\right), \end{aligned}$$

where we have estimated the sum over A for non-trivial characters by Theorem 1.2, we have appealed to Lemma 6.5 proved below for the sum corresponding to the trivial character, we have applied Lemma 6.3 for estimating the sum over Fourier coefficients, and finally we have applied the standard bounds: $\tau(n),\ \log n = O\left(n^\varepsilon\right)$ for any $\varepsilon > 0$. □

The contribution of the trivial character is given by the following lemma:

Lemma 6.5.

$$\sum_{h(A)\leq x} \chi_1(A) = 16\gamma_p x^4 + O(x^3), \tag{6.11}$$

where $\gamma_p = 1 - \frac{1}{p} - \frac{1}{p^2} + \frac{1}{p^3}$.

Proof. First we note that

$$\sum_{h(A)\leq x} \chi_1(A) = \#\{A \in M(2,\mathbb{Z}) : \mathrm{Det}(A) \not\equiv 0 (\mathrm{mod}\ p), h(A) \leq x\}$$

We count the complimentary set, i.e., matrices of height up to x that are singular modulo p and this amounts to counting 4-tuples (a,b,c,d) such that $-x \leq a,b,c,d \leq$

x and $ad - bc \equiv 0 (\mathrm{mod}\ p)$. An elementary argument as in the proof of Lemma 5.1 shows that this number is

$$\begin{aligned}(4x^2(1-1/p)^2 + O(x))(4x^2(1/p - 1/p^2) + O(x)) + (4x^2(2/p - 1/p^2) + O(x))^2 \\ = 16x^4(1/p + 1/p^2 - 1/p^3) + O(x^3).\end{aligned}$$

Upon subtracting this from $(2[x]+1)^4 = 16x^4 + O(x^3)$, the total number of matrices of height up to x, the lemma follows. □

6.3. **First steps towards the proof of Theorem 1.15.** In order to prove Theorem 1.15, we need to improve upon the term $O(p^{3+\varepsilon})$ in Eq. (6.8) above to $O(p^{2+\varepsilon})$. The estimate $O\left(p^{3+\varepsilon}\right)$ arises from the estimate $tr(G(\rho, A)) \leq d(\rho)p^2$ for the Gauss sums. Below we make a deeper analysis of the Gauss sums depending on whether A is singular or non-singular and also depending on what type of representation ρ we have.

Recall that we have (see Eq. (6.10))

$$S(\delta_{\Omega_{prim}}, x) = \sum_{\rho} c_\rho \sum_{h(A) \leq x} \chi_\rho(A) \tag{6.12}$$

The first observation is that in the Fourier expansion of $\delta_{\Omega_{prim}}$ given by Eq. (6.1), the irreducible principal series do not occur as $c_\rho = 0$ for these representations (see Prop. 6.1). Also, for the representations U_η where η is a non-trivial character of $\mathbb{F}_p^*$, we note that the dimension $d(U_\eta) = 1$ and hence Theorem 1.2 gives the bound

$$\sum_{h(A) \leq x} \chi_{U_\eta}(A) \ll p^2 (\log p)^4, \tag{6.13}$$

which is good enough for our purpose. Therefore, it is enough to consider the trivial representation, the Steinberg representation St, the non-trivial twists of St, and the cuspidal representations X_ϕ. Now observe that if A is non-singular, by Eq. (1.6),

$$\mathrm{Tr}(G(\rho, A)) = g(\rho)\mathrm{Tr}(\rho(A^{-1}),$$

where $|g(\rho)| \leq p^2$. A *striking fact* about the values of irreducible characters of $GL(2, \mathbb{F}_p)$ that can be read off the character table for $GL(2, \mathbb{F}_p)$ (see [FH, Page 70, Section 5.2]) is the following:

Proposition 6.6. *Suppose A is a 2×2 integer matrix that reduces modulo p to a non-singular matrix which is not central. Then for any non-trivial representation ρ of G, we have the bound*

$$|\widehat{\chi_\rho}(A)| \leq 2p^{-2}. \tag{6.14}$$

This suggests that we should isolate the contribution of the scalar matrices after an application of the Plancherel formula

$$\sum_{h(A)\leq x} \chi_\rho(A) = p^4 \sum_{B\in M(n,\mathbb{F}_p)} \widehat{\chi_\rho}(B)\overline{\widehat{\delta}_{\mathbf{I}}(B)}, \tag{6.15}$$

where $\mathbf{I}$ is the interval

$$\mathbf{I} = \{A \in M(2,\mathbb{Z}) : h(A) \leq x\}.$$

Accordingly, we subdivide the resulting sum over B into three parts: (i) over singular matrices, (ii) over scalar non-singular matrices and (iii) over non-singular matrices that are not scalar. However, we do this only for the cuspidal representation and the non-trivial twists of the Steinberg representation. We treat the trivial and the Steinberg representation together in §6.5 as they both contribute to the main term.

6.4. **Cuspidal representations and non-trivial twists of the Steinberg representation.** The result we want to prove here is:

Proposition 6.7. *Suppose ρ is either a cuspidal representation X_ϕ or a non-trivial twist of the Steinberg representation St_η. Then we have the bound*

$$\sum_{h(A)\leq x} \chi_\rho(A) \ll x^2 p\log p + p^2(\log p)^4$$

Proof. We apply Eq. (6.15) and and split the sum on the right hand side into three parts as described at the end of the previous subsection. The contribution of part (i) is zero by part (1) of Theorem 1.8. For part (iii), i.e., when B is non-singular and not scalar, we have the bound $\widehat{\chi_\rho}(B) \ll p^{-2}$ by Prop. 6.6. Also, recall that by Lemma 4.5 we have the bound

$$\sum_{B\in M(n,\mathbb{F}_p)} \widehat{\delta}_{\mathbf{I}}(B) \ll (\log p)^4.$$

This gives the bound $O(p^2(\log p)^4)$ for the sum over non-singular and non-scalar matrices.

For part (ii), we need to consider the sum over non-singular scalar matrices for characters coming from St_η and X_ϕ. For St_η, its character takes the value $p\eta(a^2)$, and for X_ϕ, its character takes the value $(p-1)\phi(a)$ on the central elements $\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}$. By I we will denote the identity matrix in $GL(2,\mathbb{F}_p)$ and by X we will denote a variable matrix $\begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix}$. We recall that (see Equations (1.7) and (1.6)) for non-singular B,

$$\widehat{\chi_\rho}(B) = g(\rho)\chi_\rho(B^{-1}).$$

Therefore, the sum we need to estimate is

$$p^4 \sum_{B=bI, b\not\equiv 0 (\text{mod } p)} \widehat{\chi_\rho}(B)\overline{\widehat{\delta_{\bar{\mathbf{I}}}}(B)} = p^{-4} g(\rho) \sum_{b\in\mathbb{F}_p^*} \chi_\rho(b^{-1}I) \sum_{X\in\bar{\mathbf{I}}} e\left(\frac{b(x_{11}+x_{22})}{p}\right),$$

for $\rho = X_\phi$ or St_η. First we consider the case of cuspidal representations X_ϕ attached to a character $\phi : \mathbb{F}_{p^2}^* \to \mathbb{C}^*$ satisfying $\phi \neq \phi^p$. The above sum becomes

$$p^{-4} g(X_\phi) \sum_{b\in\mathbb{F}_p^*} (p-1)\overline{\phi(b)} \sum_{X\in\bar{\mathbf{I}}} e\left(\frac{b(x_{11}+x_{22})}{p}\right).$$

Now we factor the above exponential sum and the sums involving x_{11} and x_{22} are estimated by Lemma 4.4, while the sums over x_{12} and x_{21} are bounded trivially. Thus the above sum is

$$\begin{aligned} p^{-4}(p-1)g(X_\phi) \sum_{b\in\mathbb{F}_p^*} \phi(b) \sum_{x_{12}} \sum_{x_{21}} \sum_{x_{11}} e\left(\frac{-bx_{11}}{p}\right) \sum_{x_{22}} e\left(\frac{-bx_{22}}{p}\right) \\ \ll p^{-4}(p-1)p^2x^2 \sum_{b\in\mathbb{F}_p^*} |\phi(b)|||b/p||^{-2} \\ \ll x^2 p \log p, \end{aligned}$$

where we have used (1.8) to bound $g(X_\phi)$.
For the characters associated to the representations of the type St_η, the treatment is similar. In this case, $\chi_{St_\eta}(bI) = \eta(b^2) = \eta^2(b)$ and we obtain the sum

$$p^{-4}(p-1)g(St_\eta) \sum_{X\in\bar{\mathbf{I}}} \sum_{b\in\mathbb{F}_p^*} \eta^2(b) e\left(\frac{-b(x_{11}+x_{22})}{p}\right).$$

Proceeding as before we find that this sum is also $O(x^2 p \log p)$. □

6.5. **The main term.** We still have to consider the trivial representation 1_G and the Steinberg representation St. They are 'closely related', in that they are the components of the representation parabolically induced from the trivial representation of the Borel subgroup. The character values of the trivial and the Steinberg represention are equal on split semisimple conjugacy classes, and equal but of opposite sign at the elliptic semisimple conjugacy classes. This suggests that not just the trivial character, but both the trivial character and the Steinberg character contribute to the main term. This is the reason we have postponed the treatment of these two representations thus far and we shall now analyze their contribution.

From Eq. (6.12), we write

$$S(\delta_{\Omega_{prim}}, x) = c_1 \sum_{h(A)\leq x} \chi_1(A) + c_{St} \sum_{h(A)\leq x} \chi_{St}(A) + \sum_\rho c_\rho \sum_{h(A)\leq x} \chi_\rho(A), \quad (6.16)$$

where we recall that c_1 is the Fourier coefficient for the trivial representation; i.e., $c_1 = c_{1_G}$ and ρ runs over representations that are not isomorphic to 1_G or to St.

We recall that by part (ii) of Prop. 6.1, $-c_{St} = c_1 = |\Omega_{prim}|/|G|$. Note that the Steinberg character vanishes for non-semisimple conjugacy classes and the character values of 1_G and St are equal on split semisimple conjugacy classes, and equal but of opposite sign at the elliptic semisimple conjugacy classes. Also, we recall that on the central elements, the value of the character χ_{St} is p.

Using the above facts, the total contribution of 1_G and St to the sum in Eq. (6.16) is given by

$$c_1 \sum_{h(A)\leq x} \chi_1(A) + c_{St} \sum_{h(A)\leq x} \chi_{St}(A) = \frac{2|\Omega_{prim}|}{|G|} S(\delta_{\Omega_e}, x) + \frac{|\Omega_{prim}|}{|G|}(1-p)x$$
$$= \frac{|\Omega_{prim}|}{|G|} 16(1-2/p+1/p^2)x^4 + O(x^3\sqrt{p}\log p) + O(xp),$$

by Prop. 1.13. Combining this estimate with Prop. 6.7, Eq. (6.13), estimates on Fourier coefficients of $\delta_{\Omega_{prim}}$ given by Lemma 6.3, and arguing as in the proof of Prop. 6.4, we obtain the asymptotic formula

$$S(\delta_{\Omega_{prim}}, x) = c_1 \sum_{h(A)\leq x} \chi_1(A) + c_{St} \sum_{h(A)\leq x} \chi_{St}(A) + \sum_{\rho} c_\rho \sum_{h(A)\leq x} \chi_\rho(A),$$
$$= \frac{|\Omega_{prim}|}{|G|} 16(1-2/p+1/p^2)x^4 + O(x^3\sqrt{p}\log p) + O(x^2 p\log p) + O(p^{2+\varepsilon}),$$

from which the theorem follows.

Remark 6.8. Note that Prop. 1.13 is an ingredient in the proof of Theorem 1.15. Thus, both the $GL(1)$ version (i.e., the classical one) and the $GL(2)$-analogue of the Polya-Vinogradov inequality have been used in the proof of Theorem 1.15.

It is to be expected that for similar applications for $GL(n)$ or more generally for a reductive group G, one will have to invoke Polya-Vinogradov type results attached to Levi components of parabolics in G.

6.6. **A different approach towards counting primitive elements.** In [PS], Perel'muter and Shparlinski proves the following theorem.

Theorem 6.9. *Suppose $\theta \in \mathbb{F}_{p^n}$ is such that $\mathbb{F}_p(\theta) = F_{p^n}$. Then the number of integers $m, 0 \leq m \leq x$ such that $\theta + m$ is a generator of the cyclic group $\mathbb{F}_{p^n}^*$ is*

$$\frac{\phi(p^n-1)}{p^n-1} x + O(p^{1/2+\varepsilon}).$$

We briefly describe how this is proved. After expanding the indicator function of the set of primitive roots in terms of character sums (see Lemma 6.2) and collecting the main term arising from the trivial character, all we need is the following bound

for non-trivial characters χ of $\mathbb{F}_q^*$:

$$\sum_{0\leq m\leq x} \chi(\theta+m) \ll \sqrt{p}\log p, \tag{6.17}$$

where the implied constant depends only on n (in fact, one can take the constant to be n). This bound, in turn, follows by standard analytic methods (see [IK, Chap. 12]) from the bound on the complete exponential sum given below:

$$\sum_{m\in\mathbb{F}_p} \chi(\theta+m)e_p(ma) \ll \sqrt{p}, \tag{6.18}$$

for any $a \in \mathbb{F}_p^*$. This beautiful result due to Perel'muter and Shparlinski is an ingenious application of the Riemann Hypothesis for curves over finite fields proved by Weil. See [Ka] for a different approach for a special case of the above sum where the additive character is trivial.

We now give a different proof of Theorem 1.15 using Theorem 6.9 and Prop. 1.13. Let us consider the set of $S(\Omega_{prim}, x)$ all integer matrices of height up to x that reduces to primitive elements modulo p. We partition this set according to the equivalence relation given by $A_1 \sim A_2$ if and only if $A_1 - A_2$ is an integer multiple of the identity matrix. Now, each equivalence class is of the form $\{B+nI : n \in \mathbb{Z}, h(B+nI) \leq x\}$, where B is some fixed elliptic element. For $x < p$, every element in an equivalence class is elliptic and every such class has $2[x]+1$ elements. Since the total number of elliptic elements of height up to x is $8(1-2/p+1/p^2)x^4 + O(x^3\sqrt{p}\log p)$ by Prop. 1.13, it follows that the number of equivalence class is

$$4(1-2/p+1/p^2)x^3 + O(x^2\sqrt{p}\log p).$$

Now, by Theorem 6.9, every equivalence class has

$$\frac{\phi(p^2-1)}{p^2-1}(2x) + O(p^{1/2+\varepsilon})$$

many primitive elements. Therefore, after multiplication, we obtain a result of the same strength as Theorem 1.16, the difference being in the precise shape of the error term.

References

[BK] A. Braverman and D. Kazhdan, *γ-sheaves on reductive groups*, Studies in memory of Issai Schur (Chevaleret/Rhovot 2000), Progress in Mathematics 210, 2003, 27–47.

[Bur] D. A. Burgess, *On character sums and primitive roots*, Proc. London Math. Soc. (3), vol. 12, 1962, 179–192,

[Bur2] D. A. Burgess, *Character sums and primitive roots in finite fields*, Proc. London Math. Soc. (3), vol. 17, 1967, 11–25.

[Bur3] D.A. Burgess, *On character sums and L-series*, Proc. London Math. Soc. (3), vol. 12, 1962, 193–206.

[Bur4] D.A. Burgess, *On character sums and L-series. II*, Proc. London Math. Soc. (3), vol. 13, 1963, 524–536.

[Dav] H. Davenport, *Multiplicative Numebr Theory*, Third edition. Revised and with a preface by Hugh L. Montgomery. Graduate Texts in Mathematics, 74. Springer-Verlag, New York, 2000. xiv+177 pp.

[Dav2] H. Davenport, *On primitive roots in finite fields*, Q. J. Math. vol. 8, 1937, 308–312.

[Da-Le] H. Davenport and D. J. Lewis, *Character sums and primitive roots in finite fields*, Rend. Circ. Mat. Palermo (2), vol. 12, 1963, 129–136.

[DL] P. Deligne and G. Lusztig, *Representations of a reductive groups over finite fields*, Annals of Math. 103 (1976) 103-161.

[FH] W. Fulton and J. Harris, *Representation theory, A first course*, Graduate Texts in Mathematics, **129**. Readings in Mathematics. Springer-Verlag, New York, 1991.

[Gr] J. A. Green, *The characters of the finite general linear groups*, Trans. Amer. Math. Soc. 80 (1955), 402-447.

[IK] H. Iwaniec and E. Kowalski, *Analytic number theory*, American Mathematical Society Colloquium Publications, American Mathematical Society, Providence, RI, 2004, xii+615 pp.

[Ka] N. M. Katz, *An Estimate for Character Sums* J. Amer. Math. Soc., Vol. 2, No. 2, 1989, 197–200.

[Ko] T. Kondo, *On Gaussian sums attached to the general linear groups over finite fields*, J. Math. Soc. Japan (15) 1963. 244–255.

[La] E. Lamprecht, *Struktur und Relationen allgemeiner Gaussacher Summen in endlichen Ringen I, II*, J. Reine Angew. Math. **197** (1957) 1-48.

[Mo1] H. L. Montgomery, *Ten lectures on the interface between analytic number theory and harmonic analysis*, CBMS No. 84, Amer. Math. Soc., Providence, 1994.

[Mo2] H. L. Montgomery, *Topics in Multiplicative Number Theory* Springer L.N. 227 (1971).

[MM] M. Ram Murty and W. Kumar Murty, *Non-vanishing of L-functions and applications, Progress in Mathematics*, vol. 157, Birkhauser (Boston) 1997.

[PS] G.I. Perel'muter and I. Shparlinski, *Distribution of primitive roots in finite fields*, Russian Math. Surveys 45 (1990), no. 1, 223–224

[RS] J. B. Rosser and L. Schoenfeld, *Approximate formulas for some functions of prime numbers*, Illinois J. Math. 6, 1962, 64–94.

[Sh] Harold N. Shapiro, Introduction to the theory of numbers. Pure and Applied Mathematics. A Wiley-Interscience Publication. John Wiley and Sons, Inc., New York, 1983.

[Shp] I. E. Shparlinski, Finite fields: theory and computation. The meeting point of number theory, computer science, coding theory and cryptography. Mathematics and its Applications, 477. Kluwer Academic Publishers, Dordrecht, 1999.

Indian Statistical Institute, Kolkatta 700108, INDIA.

Email address: sgisical@gmail.com

Tata Institute of Fundamental Research, Homi Bhabha Road, Bombay - 400 005, INDIA.

Email address: rajan@math.tifr.res.in