STRONG APPROXIMATION AND DIOPHANTINE PROPERTIES OF MARKOFF TRIPLES

JEAN BOURGAIN, ALEXANDER GAMBURD, AND PETER SARNAK

ABSTRACT. Transitivity properties of the group of morphisms generated by Vieta involutions on the solutions to the Markoff equation modulo primes are established, yielding forms of strong approximation for the Markoff surface. These are applied to show that almost all Markoff numbers are highly divisible.

1. Introduction

1.1. A deep interaction between arithmetic and dynamics is a hallmark of Markoff's oeuvre. While he is most widely known today for the chains named after him, it is in the context of his early work [Mar79, Mar80] on the minima of binary quadratic forms and badly approximable numbers[1] that the following equation, now bearing his name, was born

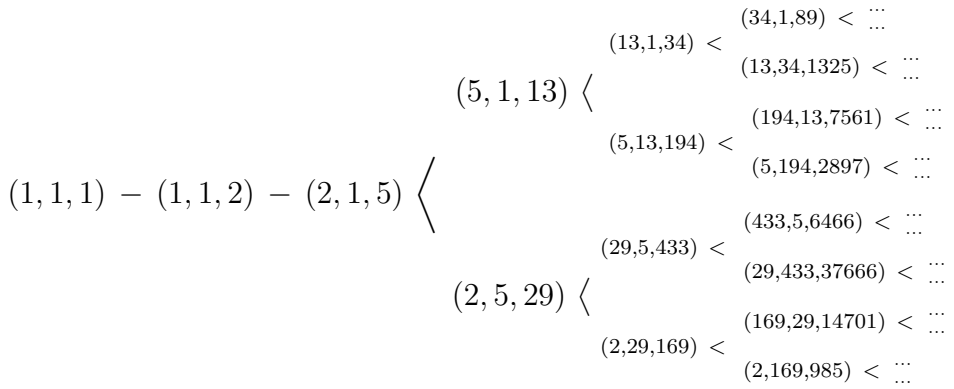
$$x_1^2 + x_2^2 + x_3^2 - 3x_1x_2x_3 = 0. \tag{1}$$

It describes a surface in $\mathbb{A}^3$ that we denote by X. *Markoff triples* $\mathcal{M}$ are natural number solutions of (1); the *Markoff numbers* $\mathbb{M} \subset \mathbb{N}$ are obtained as coordinates of elements of $\mathcal{M}$. The *Markoff sequence* $\mathbb{M}^s$ is the (multi-)set of largest coordinates of an $m \in \mathcal{M}$ counted with multiplicity. The uniqueness conjecture of Frobenius [Fro13] asserts that $\mathbb{M} = \mathbb{M}^s$.

All elements of $\mathcal{M}$ are obtained from the root solution $r = (1, 1, 1)$ by repeated applications of the Vieta involutions R_1, R_2, R_3 of $\mathbb{A}^3$, with $R_1(x_1, x_2, x_3) = (3x_2x_3 - x_1, x_2, x_3)$ and R_2, R_3 defined similarly. Denoting by Γ the **nonlinear** group of automorphisms of $\mathbb{A}^3_{\mathbb{Z}}$ generated by R_1, R_2, R_3, the set of Markoff triples, $\mathcal{M}$ can be identified with the **orbit** of the root r under the action of Γ, that is to say, $\mathcal{M} = \Gamma \cdot r$,

[1]See [Bom07, Cas49] for a masterful presentation of this work of Markoff and [Aig13, Per02] for a survey of the subsequent appearances of his equation in a variety of different contexts.

giving rise to the **Markoff tree**:



The first few members of $\mathbb{M}$ are

$$1, 2, 5, 13, 29, 34, 89, 169, 194, 233, 433, 610, 985, \cdots$$

The sequence $\mathbb{M}^s$ is **sparse**, as shown by Zagier [Zag82]:

$$\sum_{\substack{m\in\mathbb{M}^s\\ m\leq T}} 1 \sim c(\log T)^2 \quad \text{as} \quad T\to\infty, (c>0). \tag{2}$$

1.2. Our diophantine analysis of the Markoff equation began with attempts to execute a sieve on $\mathcal{M}$ (see [Sar10] and [Gam23] for accounts). It led to a sieve in the setting where the group of morphisms is affine linear [BGS10]. That theory is now well developed thanks to many works (see [SS13], [Gam23] and references therein). In [BGS16] we returned to the nonlinear setting of surfaces of Markoff type and announced various results. The aim in this paper is to give a full account of the Markoff surface part of the announcement.

The first issue that one faces in any sieve is to understand the image of the reduction of $X(\mathbb{Z})$ to $X(\mathbb{Z}/p\mathbb{Z})$ for primes p. If Δ is the group of automorphisms of $\mathbb{A}^3$ generated by Γ together with the (linear) involutions which replace two of the coordinates of x by their negatives, then $X(\mathbb{Z})$ consists of the two Δ orbits, namely the orbit of $(0,0,0)$ and the orbit of $(1,1,1)$. The actions of Γ and Δ descend to a permutation actions on the finite sets $X(\mathbb{Z}/p\mathbb{Z})$ for any prime p. The orbits of this action will be referred to as the *components*. Clearly the Δ action on $X(\mathbb{Z}/p\mathbb{Z})$ consisting of the two orbits of $(0,0,0)$ and its complement $X^*(\mathbb{Z}/p\mathbb{Z})$, which is often denoted simply by $X^*(p)$, is equivalent to the following

Conjecture 1 (Strong Approximation Conjecture)**.** *For any prime p, the reduction of $X(\mathbb{Z})\backslash\{(0,0,0)\}$ (or of $\mathcal{M}$) to $X^*(\mathbb{Z}/p\mathbb{Z})$ is onto.*

Remarks:

(a) This question was raised in [Bar91] and the conjecture has been verified for $p \leq 3000$ in [DL20].

(b) A consequence of the Conjecture is that the only prime congruence obstruction to being a Markoff number m is the one noted in [Fro13]:

$$m \neq \pm\frac{2}{3}, 0 \bmod p \text{ if } p = 3 \bmod 4, p \neq 3. \tag{3}$$

For ease of exposition we will assume henceforth that $p > 3$ and in fact is large. Our first result asserts that there is always a giant connected component.

Theorem 1. *Fix $\varepsilon > 0$. Then for p large enough (depending on ε) there is a Γ orbit $\mathcal{C}(p)$ in $X^*(p)$ for which*

$$|X^*(p)\backslash\mathcal{C}(p)| \leq p^{\varepsilon} \tag{4}$$

(note that $|X^*(p)| \sim p^2$), and any Γ orbit $\mathcal{D}(p)$ satisfies

$$|\mathcal{D}(p)| \gg (\log p)^{\frac{1}{3}}. \tag{5}$$

We are able to prove Conjecture 1 unless $p^2 - 1$ is a very smooth number; see (89). In particular, we show that the set of primes for which the Conjecture fails is very small.

Theorem 2. *Let E be the set of primes for which Conjecture 1 fails. For $\varepsilon > 0$ the number of primes $p \leq T$ with $p \in E$ is at most T^{ε}, for T large.*

Theorems 1 and 2 are used together with some simple sieving to show that almost all Markoff numbers are composite[2] (in fact highly divisible; see Theorem 19).

Theorem 3. *Almost all Markoff numbers are composite, that is*

$$\sum_{\substack{p \in \mathbb{M}^s \\ p \text{ prime}, p \leq T}} 1 = o\Big(\sum_{\substack{m \in \mathbb{M}^s \\ m \leq T}} 1\Big).$$

Remark: It is worth contrasting this result with the state of knowledge regarding the sequence $H_n = 2^n + b$, which is a little sparser than the sequence of Markoff numbers. By (2) we have that $M_n \sim A^{\sqrt{n}}$. Even assuming the generalized Riemann Hypothesis, which allowed Hooley [Hoo67] to give a conditional proof of Artin's primitive root conjecture was not sufficient to establish that almost all the members of the sequence

[2]We remark that in [CZ06] Corvaja and Zannier showed that the greatest prime factor of xy for a Markoff triple (x, y, z) tends to infinity.

H_n are composite. The conditional proof in [Hoo76] necessitated postulating an additional "Hypothesis A".

1.3. In a recent breakthrough making use of the interpretation of the Γ action on $X^*(\mathbb{Z}/p\mathbb{Z})$ in terms of connectivity properties of moduli spaces of elliptic curves with $\mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})$ structure, William Chen [Che21] showed that

(6) The size of every Γ orbit $|\mathcal{D}(p)|$ is divisible by p.

Combining (6) with Theorem 1 establishes Conjecture 1 for p sufficiently large.

In our followup to the present paper [BGS24] we complete the strong approximation picture for $X(\mathbb{Z}_p)$, where $\mathbb{Z}_p$ are the p-adic integers. It is clear from the definition of the Vieta generators that $X(\mathbb{Z}_p)$ is invariant under Γ. In fact, the compact set $X^*(\mathbb{Z}_p)$ consisting of all $\mathbf{x} \in X(\mathbb{Z}_p)$ for which at least one of x_1, x_2, x_3 is not divisible by p is Γ-invariant.

Theorem 4. *Assume Conjecture 1 holds for p. Then the action of Γ on $X^*(\mathbb{Z}_p)$ is minimal, that is every orbit is dense.*

Remark: It is instructive to put this Theorem in the more general context of relative character varieties of representations of the fundamental group of a surface into SL_2 (see [Wha20a]). X corresponds to the once punctured torus and trace equal to -2 and Γ to the mapping class group. The action of the mapping class group on the real points of these character varieties has been studied extensively (see [Thu88], [Gol03]). All of these affine varieties V are defined over $\mathbb{Z}$ ([Wha20b]) and one gets a natural action of the mapping class group on the p-adic points for any p. Salehi and Tamam [ST23] show even more generally that the sets corresponding to $X^*(\mathbb{Z}_p)$ consist of finitely many minimal sets. What Conjecture 1 and Theorem 4 point to is that the Γ action on $\mathbb{Z}_p$ points is as minimal as possible.

Our followup paper [BGS24] is mainly concerned with the more general surfaces, where in (1) the right hand side is replaced by an integer $k \in \mathbb{Z}$. With suitable adjustments (see [BGS16]) to Conjecture 1, the analogues of Theorems 1, 2 and 4 can be established. In particular, the closely related conjecture of McCullough and Wanderley [MW13] concerning t-systems of pairs of generators of $G = \mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})$ is proven for all but very few p's, as in Theorem 2. While there are results, even quantitative ones ([Gil77], [BG10b]) for t-systems with more than 2 generators, as far as we know this is the first such for two generators (cf. [GP03]).

1.4. We end the introduction by outlining the rest of the paper. In Section 2 we define the fundamental rotations in Γ that are associated to an $\mathbf{x} \in X^*(p)$ and one of its coordinates. These act on the conic sections gotten by intersecting $X^*(p)$ with the plane corresponding to the particular coordinate. Some basic properties of the incidence graph of the intersections of the conic sections are established.

In Section 3, which we call the *endgame*, we define the *cage* $\mathcal{C}(p)$ which is shown to be a large component of $X^*(p)$. Specifically, any $\mathbf{x} \in X^*(p)$ for which the rotation associated to one of its coordinates (see Section 2 for definitions) has order at least $p^{\frac{1}{2}+\delta_0}$ ($\delta_0 > 0$) is shown to be in $\mathcal{C}(p)$.

In Section 4 , *the middle game*, the last statement is extended to $\mathbf{x}$'s for which the corresponding rotation has order p^{ε_0} ($\varepsilon_0 > 0$ any fixed small number).

The methods used in Sections 3 and 4 rely on nontrivial upper bounds for the number of points lying on curves over finite fields. In Section 3 Weil's Riemann Hypothesis [Wei41] is a key tool, but this is not strong enough when the order is less than $p^{\frac{1}{2}+\delta_0}$. In its stead we use Stepanov's auxiliary polynomial method [Ste69] in the Appendix, or the $\gcd(u-1, v-1)$ bounds of Corvaja-Zannier [CS13], or the combinatorial method based on the projective Szemeredi-Trotter Theorem [Bou12] developed in section 4.

Section 5, *the opening*, deals with $\mathbf{x}$'s for which the orders of the associated rotations are very small (for example, being uniformly bounded). This is done by lifting the equations to characteristic zero which leads to an equation in $(\bar{\mathbb{Q}})^3$ in roots of unity. In the more general setting of [BGS24] we invoke Lang's G_m conjecture at this point (proven in [Lau83], for example); however for the special case at hand one can show directly that $\mathbb{X}^*(\bar{\mathbb{Q}})$ has no finite Γ orbits[3]. This $\bar{\mathbb{Q}}$ analysis leads to part 2 of Theorem 1.

In Section 6 we assemble the various stages of our argument, explicated in Sections 3, 4, 5 and give a proof of Theorem 1 in a form from which strong approximation follows if $p^2 - 1$ is not very smooth: see (89). To prove Theorem 2 we combine the above with a variant of the results in [CKSZ14] concerning the multiplicative orders of coordinates of points of curves on $\mathbb{A}^2$ over $\mathbb{F}_p$.

In section 7 we discuss strong approximation for certain moduli which are products of primes congruent to 1 mod 4. For such p's Meiri and Puder [MP18] show that if Conjecture 1 holds then the permutation action of Γ on $X^*(p)$ is either the symmetric or alternating group. This together with Goursat (disjointness) lemma leads to the

[3]The absence of finite Γ orbits is a necessary condition for strong approximation, in the form of Conjecture 1, to hold.

transitivity of the Γ action on products of $X^*(p_j)$'s. This together with Mirzakhani's orbit counting and equidistribution theorem [Mir16] and Theorem 2 are combined to prove Theorem 3 in the stronger form that almost all Markoff numbers are highly composite; see Theorem 19.

There is a natural cubic graph that one can attach to the Vieta moves acting on $X^*(p)$. The vertices of the graph are the points in $X^*(p)$ and $\mathbf{x}$ and $\mathbf{x}'$ are joined if one of R_1, R_2 or R_3 takes one to the other. Conjecture 1 is equivalent to this graph being connected. A pregnant question (given its centrality in the theory of affine linear sieve) is whether these graphs form an expander family[4]. Such an input would be powerful in any further sieving in this context and it would also ensure that there is a short path ($O(\log |X^*(p)|)$) between any two solutions in $X^*(p)$. Numerical experiments for p's up to 3000 point to these graphs being expanders [DL20] and in [deC22] it is shown that these graphs are not planar for $p > 7$. An algorithm to navigate these graphs has recently been devised and implemented [Sil22]. It should allow one to check Conjecture 1 for much larger p. Whether it is feasible to bridge the gap and verify Conjecture 1 for all primes is an interesting question.

2. Preliminaries

For ease of exposition, in Sections 2-6 we consider the equation $x^2+y^2+z^2 = xyz$. Transferring the results obtained in these Sections to the equation $x^2+y^2+z^2 = 3xyz$, which is pertinent to the diophantine analysis in Section 7, is straightforward (see e.g. Section 5.5. in [Che21]).

2.1. Analysis of the conic sections. Theorem 1 in the weaker form that $|\mathcal{C}(p)| \sim |X^*(p)|$ as $p \to \infty$, can be viewed as the finite field analogue of [Gol03], where it is shown that the action of Γ on the compact real components of the character variety of the mapping class group of the once punctured torus is ergodic. As in [Gol03], our proof makes use of the rotations $\tau_{ij} \circ R_i, i \neq j$ where τ_{ij} permutes x_i and x_j. For example,

$$\tau_{2,3} \circ R_2(x_1, x_2, x_3) = (x_1, x_3, x_1x_3 - x_2),$$

so the action on (x_2, x_3) for fixed x_1 is given by the rotation $\mathrm{rot}(x_1)$

$$\mathrm{rot}(x_1) \begin{pmatrix} x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_3 \\ x_1x_3 - x_2 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & x_1 \end{pmatrix} \begin{pmatrix} x_2 \\ x_3 \end{pmatrix}. \tag{7}$$

[4]See [Sar04] and [HLW06] for definition and properties of expanders.

This rotation preserves the conic section obtained by intersecting $X^*(p)$ with the plane defined by the first coordinate being equal to the value of x_1; in general, we define the conic section $C_j(a)$ as follows:

$$C_j(a) = \{x_j = a\} \cap X^*(p). \tag{8}$$

We give an explicit description of this action, depending on whether $x = x_1$ is *parabolic* ($x^2 - 4 = 0$, that is $x = \pm 2 \mod p$), *hyperbolic* ($\left(\frac{x^2-4}{p}\right) = 1$) or *elliptic* ($\left(\frac{x^2-4}{p}\right) = -1$) with $\left(\frac{\cdot}{\cdot}\right)$ being the Legendre symbol.

Lemma 5. *Let $x = \pm 2$. If $\left(\frac{-1}{p}\right) = -1$, that is if $p \equiv -1 \mod 4$, then $C_1(x)$ is empty. If $\left(\frac{-1}{p}\right) = 1$, that is if $p \equiv 1 \bmod 4$, then*

$$C_1(2) = \{(2, t, t \pm 2i) \mid t \in \mathbb{Z}/p\mathbb{Z}\} \tag{9}$$

where $i^2 \equiv -1 \mod p$;

$$C_1(-2) = \{(-2, t, -t \pm 2i) \mid t \in \mathbb{Z}/p\mathbb{Z}\}, \tag{10}$$

which are pairs of disjoint lines. The action of $\rho_1 = \mathrm{rot}(x)$ is

$$\rho_1(2, t, t \pm 2i) = (2, t \pm 2i, t \pm 4i), \tag{11}$$

$$\rho_1(-2, t, -t \pm 2i) = (-2, -t \pm 2i, -t \mp 4i), \tag{12}$$

so $\mathrm{rot}(2)$ *preserves each line and* $\mathrm{rot}(-2)$ *interchanges them.*

Now when $x \neq \pm 2 \mod p$ we write

$$x = \chi + \chi^{-1},$$

where $\chi \in \mathbb{F}_p$ if $\left(\frac{x^2-4}{p}\right) = 1$ and $\chi \in \mathbb{F}_{p^2}$ if $\left(\frac{x^2-4}{p}\right) = -1$.

Note that

$$\mathrm{rot}(x) = \begin{pmatrix} 1 & 1 \\ \chi & \frac{1}{\chi} \end{pmatrix} \begin{pmatrix} \chi & 0 \\ 0 & \frac{1}{\chi} \end{pmatrix} \begin{pmatrix} \frac{1}{\chi} & -1 \\ -\chi & 1 \end{pmatrix} \left(\frac{1}{\chi} - \chi\right)^{-1} = \begin{pmatrix} 1 & 1 \\ \chi & \frac{1}{\chi} \end{pmatrix} \begin{pmatrix} \chi & 0 \\ 0 & \frac{1}{\chi} \end{pmatrix} \begin{pmatrix} 1 & 1 \\ \chi & \frac{1}{\chi} \end{pmatrix}^{-1}$$

and consequently

$$\mathrm{rot}(x)^\ell = \begin{pmatrix} 1 & 1 \\ \chi & \frac{1}{\chi} \end{pmatrix} \begin{pmatrix} \chi^\ell & 0 \\ 0 & \frac{1}{\chi^\ell} \end{pmatrix} \begin{pmatrix} 1 & 1 \\ \chi & \frac{1}{\chi} \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 1 \\ \chi & \frac{1}{\chi} \end{pmatrix} \begin{pmatrix} \chi^\ell & 0 \\ 0 & \frac{1}{\chi^\ell} \end{pmatrix} \begin{pmatrix} \frac{1}{\chi} & -1 \\ -\chi & 1 \end{pmatrix} \left(\frac{1}{\chi} - \chi\right)^{-1}$$

and

$$\langle \mathrm{rot}(x) \rangle = \left(\frac{1}{\chi} - \chi\right)^{-1} \left\{ \begin{pmatrix} \frac{\chi_1}{\chi} - \frac{\chi}{\chi_1} & \frac{1}{\chi_1} - \chi_1 \\ \chi_1 - \frac{1}{\chi_1} & \frac{1}{\chi\chi_1} - \chi\chi_1 \end{pmatrix} ; \chi_1 \in \langle \chi \rangle \right\}.$$

Consequently $C_1(x)$ contains, as its second and third components, all elements of the form

$$\left(\left(\chi-\frac{1}{\chi}\right)^{-1}\left(\left(x_3-\frac{x_2}{\chi}\right)\chi_1+(\chi x_2-x_3)\frac{1}{\chi_1}\right),\left(\chi-\frac{1}{\chi}\right)^{-1}\left((\chi x_3-x_2)\chi_1+\left(x_2-\frac{x_3}{\chi}\right)\frac{1}{\chi_1}\right)\right)$$

with $\chi_1\in\langle\chi\rangle$.

Note that

$$\mathrm{Proj}_{x_2}(C_1(x))\supset\left\{a\chi_1+\frac{b}{\chi_1};\chi_1\in\langle\chi\rangle\right\}$$

where

$$a=\left(\chi-\frac{1}{\chi}\right)^{-1}\left(x_3-\frac{x_2}{\chi}\right)\ \text{ and }\ b=\left(\chi-\frac{1}{\chi}\right)^{-1}(\chi x_2-x_3)$$

satisfy

$$\sigma=ab=\left(\frac{x}{\chi-\frac{1}{\chi}}\right)^2=\left(\frac{\chi+\frac{1}{\chi}}{\chi-\frac{1}{\chi}}\right)^2\neq 1. \tag{13}$$

Denoting by ρ a primitive root of $\mathbb{F}_p$, a hyperbolic element x can be written in the form

$$x=\rho^j+\rho^{-j}. \tag{14}$$

For a hyperbolic element we let $\mathrm{ord}(x)=\frac{p-1}{j}$ (primitive root ρ is chosen depending on x, so that the exponent j divides $p-1$).

An elliptic element x can be written in the form

$$x=\xi^j+\xi^{-j}, \tag{15}$$

where ξ is an element in $\mathbb{F}_{p^2}$, $\xi=(\tilde{\rho})^{p-1}$, where $\tilde{\rho}$ is a generator for the multiplicative group of $\mathbb{F}_{p^2}$. For an elliptic element we let $\mathrm{ord}(x)=\frac{p+1}{j}$.

Lemma 6. *Let x be hyperbolic, $x\neq 0(p)$; write*

$$x=w+w^{-1}, \tag{16}$$

where $w=\rho^j\in\mathbb{F}_p$. Then $C_1(x)$ is a hyperbola with $p-1$ points. Set

$$\kappa(x)=\frac{x^2}{x^2-4}. \tag{17}$$

Let

$$H(x)=\{\left(t,\frac{\kappa(x)}{t}\right)\mid t\in{\mathbb{F}_p}^*\}. \tag{18}$$

Then we have the following bijective map from $H(x)$ to $C_1(x)$:

$$\left(t,\frac{\kappa(x)}{t}\right)\to\left(x,t+\frac{\kappa(x)}{t},tw+\frac{\kappa(x)}{tw}\right). \tag{19}$$

In these coordinates

$$\rho_1\left(t, \frac{\kappa(x)}{t}\right) = \left(tw, \frac{\kappa(x)}{tw}\right). \tag{20}$$

Lemma 7. *Let x be elliptic, $x \neq 0(p)$; write*

$$x = v + v^p, \tag{21}$$

where $v \in \mathbb{F}_{p^2} - \mathbb{F}_p$, $v^{p+1} = 1$. Then $C_1(x)$ is an ellipse with $p+1$ points. Set

$$\kappa(x) = \frac{x^2}{x^2 - 4}. \tag{22}$$

Let

$$E(x) = \{(t, t^p) \,|\, t \in \mathbb{F}_{p^2}, t^{p+1} = \kappa(x)\}. \tag{23}$$

Then we have the following bijective map from $E(x)$ to $C_1(x)$:

$$(t, t^p) \to \left(x, t + \frac{\kappa(x)}{t}, tv + \frac{\kappa(x)}{tv}\right). \tag{24}$$

In these coordinates

$$\rho_1\left(t, \frac{\kappa(x)}{t}\right) = \left(tv, \frac{\kappa(x)}{tv}\right). \tag{25}$$

2.2. Incidence graph for the conic sections. We restrict ourselves to the case of $p \equiv 3(\mathrm{mod}4)$ (the case of $p \equiv 1(\mathrm{mod}4)$ is simpler because of the special point in Lemma 5). Let $X^*(p)$ be the $\mathrm{mod}\, p$ Markoff triples; ξ any coordinate of a triple. By the observation (3), due to Frobenius [Fro13], for $p \equiv 3(\mathrm{mod}4)$ we have $\xi \neq 0, \pm 2$.

For $j \neq k$ and ξ, η in $\mathbb{F}_p$

$$|C_j(\xi) \cap C_k(\eta)| \in \{0, 1, 2\}. \tag{26}$$

To determine which it is, the intersection consists of all z's such that

$$\xi^2 + \eta^2 + z^2 = \xi\eta z, \tag{27}$$

which has a solution if and only if

$$\xi^2\eta^2 - 4(\xi^2 + \eta^2)$$

is a square in $\mathbb{F}_p$. In terms of Legendre's symbol

$$|C_j(\xi) \cap C_k(\eta)| = 1 + \left(\frac{\xi^2\eta^2 - 4(\xi^2 + \eta^2)}{p}\right). \tag{28}$$

By Jacobstahl's formula

$$\sum_{\eta \in \mathbb{F}_p{}^*} \left(\frac{\xi^2\eta^2 - 4(\xi^2 + \eta^2)}{p}\right) = -\left(\frac{\xi^2 - 4}{p}\right).$$

So each $C_j(\xi)$ meets $\frac{p+1}{2}$ conic sections $C_k(\eta)$'s if ξ is hyperbolic and $\frac{p-1}{2}$ conic sections $C_k(\eta)$'s if ξ is elliptic.

Define *the incidence graph* $I(p)$ of $X^*(p)$ to have vertices $C_j(\xi)$'s with the number of edges between $C_j(\xi)$ and $C_k(\eta)$ being $|C_j(\xi) \cap C_k(\eta)|$.

Proposition 8. *For p large ($p > 10$) the incidence graph $I(p)$ is connected and in fact* $\mathrm{diam}(I(p)) = 2$.

Proof. Fix ξ_1, ξ_2 and i, j, say $i, j \in \{1, 2\}$. We seek $y \in \mathbb{F}_p$ such that $C_3(y) \cap C_i(\xi_1) \neq \emptyset$ and $C_3(y) \cap C_j(\xi_2) \neq \emptyset$. This amounts to solving the pair of equations:

$$\begin{cases} (\xi_1^2 - 4)y^2 - \lambda^2 = 4\xi_1^2 \\ (\xi_2^2 - 4)y^2 - \mu^2 = 4\xi_2^2 \end{cases} \tag{29}$$

for $y, \lambda, \mu \in \mathbb{F}_p$.

If $\xi_1^2 = \xi_2^2$ then (29) reduces to the first equation (take $\lambda = \mu$) and since $\xi_1^2 - 4 \neq 0$ and $\xi_1^2 \neq 0$, it defines a conic section. Thus for p large it has a solution and provides us with our y.

If $\xi_1^2 \neq \xi_2^2$ then (29) defines an absolutely irreducible curve in $\mathbb{A}^3$ of genus one, as proved in Lemma 9 below. Thus again for p large it has solutions over $\mathbb{F}_p$ providing us with our desired y. It follows that the distance in $I(p)$ between any two points is at most 2. On the other hand, $C_i(\xi)$ and $C_i(\eta)$ are not joined if $\xi \neq \eta$ nor is $C_i(\xi)$ joined to $\frac{p\pm1}{2}$ of the $C_j(\eta)$'s for $j \neq i$. Hence $\mathrm{diam}(I(p)) = 2$. □

Lemma 9. *If $\xi_1^2 \neq \xi_2^2$ then* (29) *defines an absolutely irreducible curve of genus one.*

Proof. Our curve C in $\mathbb{A}^3$ over $\overline{\mathbb{F}_p}$ is of the form

$$\begin{cases} u^2 - A_1 v^2 = B_1 \\ u^2 - A_2 w^2 = B_2, \end{cases} \tag{30}$$

where our condition $\xi_1^2 \neq \xi_2^2$ implies that A_1, A_2, B_1, B_2 and $B_1 - B_2$ are all not zero. Changing variable $v = \lambda v'$ with $\lambda^2 = A_1$ (over $\overline{\mathbb{F}_p}$) the curve C becomes C' given by

$$\begin{cases} u^2 - v^2 = B_1 \\ u^2 - w^2 = B_2, \end{cases} \tag{31}$$

with B_1, B_2 and $B_1 - B_2$ all not zero. We claim that C' is irreducible and of genus one.

Eliminating u we are led to the plane conic

$$v^2 + B_1 = w^2 + B_2. \tag{32}$$

Set $v - w = \xi$, $v + w = \eta$; then $\xi\eta = B_2 - B_1$, so

$$v = \frac{1}{2}\left(\xi + \frac{B_2 - B_1}{\xi}\right) \tag{33}$$

and (31) becomes the plane curve

$$u^2 = \frac{1}{4}\left(\xi + \frac{B_2 - B_1}{\xi}\right)^2 + B_1, \tag{34}$$

which is equivalent to

$$u^2 = (\xi^2 + B_2 - B_1)^2 + 4B_1\xi^2 = \xi^4 + 2(B_1 + B_2)\xi^2 + (B_2 - B_1)^2, \tag{35}$$

that is the plane curve

$$u^2 = P(\xi), \tag{36}$$

where P is of degree four and has the four roots $\pm\sqrt{-(B_1 + B_2) \pm 2\sqrt{B_1B_2}}$. Under the condition that $B_1, B_2, B_1 - B_2$ are all not zero, one checks that the four roots are all distinct. Hence (36) is an absolutely irreducible plane curve of genus one. □

3. Endgame

By the order of a point $\mathbf{x} = (x_1, x_2, x_3) \in X^*(p)$ we mean $\max(\mathrm{ord}(\mathrm{rot}(x_j)))$. A point $\mathbf{x} \in X^*(p)$ is called maximal if $\mathrm{ord}(\mathrm{rot}(x_j))$ is maximal for some j. Note that the condition that the order of $\mathrm{rot}(x_j)$ be maximal depends only on x_j and not on the other coordinates of x (since it depends on the order of λ, where $\lambda + \lambda^{-1} = x_1$ in $\mathbb{F}_p^*$ of $\mathbb{F}_{p^2}^*$).

3.1. Use of Weil's bound.

We begin with the following

Proposition 10. *If $x = (x_1, x_2, x_3)$ is in $X^*(p)$ and for some $j \in \{1, 2, 3\}$ the order of the induced rotation $\mathrm{rot}(x_j)$ is at least $p^{\frac{1}{2}+\delta}$ ($\delta > 0$ fixed) then $\mathbf{x}$ is joined to a point in $X^*(p)$, one of whose induced rotations is of maximal order (for large enough p depending on δ).*

Proof. Consider first the case that x_1 (say $j = 1$) is hyperbolic. In light of the discussion in section 2, $\mathbf{x} = (x_1, x_2, x_3)$ is connected to the points in $X^*(p)$ of the form

$$(x_1, \alpha_1 t + \alpha_2 t^{-1}, \alpha_3 t + \alpha_4 t^{-1}) \tag{37}$$

with $t \in H$, a cyclic subgroup of ${\mathbb{F}_p}^*$. Here $|H| \mid (p-1)$; we set $e_H = \frac{p-1}{|H|}$. Our aim is to produce t's in H for which there is a primitive root $y \in {\mathbb{F}_p}^*$ satisfying

$$\alpha_1 t + \alpha_2 t^{-1} = y + y^{-1}. \tag{38}$$

Let $P(H)(= P_{\alpha_1,\alpha_2}(H))$ denote the number of such solutions.

A subgroup K of ${\mathbb{F}_p}^*$ is determined by its order $|K|$ which divides $p-1$; let $d_K = \frac{p-1}{|K|}$. Let $f_H(K) = f_H(d_K)$ be the number of solutions to

$$\alpha_1 t + \alpha_2 t^{-1} = y + y^{-1},\ t \in H,\ y \in K \tag{39}$$

(note that the traces of the matrices that we produce, namely the common values of the left- and right-hand side of (39), are hit with multiplicity 2 in both t and y in our counting). Clearly

$$f_H(K) \leq 2\min(|K|, |H|). \tag{40}$$

We can estimate $f_H(K)$, at least if $|H| \geq p^{\frac{1}{2}+\delta}$, using Weil's Riemann Hypothesis for curves over finite fields. The map

$$\xi \to \xi^{d_K},\ \eta \to \eta^{e_H}$$

sends solutions of

$$C_{\alpha_1,\alpha_2} :\ \alpha_1\eta^{e_H} + \alpha_2\eta^{-e_H} = \xi^{d_K} + \xi^{-d_K} \tag{41}$$

to solutions of (39) and it is $e_H d_K$ to 1. Hence if $N(C_{\alpha_1,\alpha_2})$ is the number of solutions to (41) then

$$f_H(K) = \frac{N(C_{\alpha_1,\alpha_2})}{e_H d_K}. \tag{42}$$

As we prove below (see Lemma 11) the curve C_{α_1,α_2} is absolutely irreducible. Since its genus is $O(e_H d_K)$ (see e.g. Proposition 2.3 in [Pak09]), applying Weil's bound yields

$$N(C_{\alpha_1,\alpha_2}) = p + O(\sqrt{p} e_H d_K). \tag{43}$$

Hence

$$f_H(K) = \frac{p}{e_H d_K} + O(\sqrt{p}). \tag{44}$$

By inclusion/exclusion

$$P(H) = \sum_{d|(p-1)} \mu(d) f_H(d_K), \tag{45}$$

where μ is the Mobius function. Hence

(46)
$$P(H) = \sum_{d|(p-1)} \mu(d) \left(\frac{|H|}{d} + O(\sqrt{p}) \right) = |H| \sum_{d|(p-1)} \frac{\mu(d)}{d} + O_\varepsilon(p^{\frac{1}{2}+\varepsilon}) = |H| \frac{\varphi(p-1)}{p-1} + O_\varepsilon(p^{\frac{1}{2}+\varepsilon}).$$

Here φ is the Euler function and it satisfies $\varphi(n) \gg_\varepsilon n^{1-\varepsilon}$ and hence from (46) we deduce that $P(H) > 1$ under the assumption that $|H| \geq p^{\frac{1}{2}+\delta}$. This proves Proposition 10 in the hyperbolic case.

Now consider the case of $x = x_1$ elliptic. Let D be a non-square element in $\mathbb{F}_p$. Then $\mathbb{F}_{p^2} = \mathbb{F}_p[\sqrt{D}]$ and we can parametrize a subgroup $H_1 \in \mathbb{F}_{p^2}$ as follows

$$\{(\xi + \sqrt{D}\eta)^{d_1}; \xi, \eta \in \mathbb{F}_p; \xi^2 - D\eta^2 = 1\}, \tag{47}$$

where $d_1 = \frac{p+1}{|H_1|}$. The conic section $C_1(x)$ is an ellipse which can be parametrized as

$$\alpha^2 - D\beta^2 = \kappa(x), \tag{48}$$

where $\kappa(x) = \frac{x^2}{x^2-4}$ and $\alpha, \beta \in \mathbb{F}_p$. We seek α which can be written as $\alpha = u + u^{-1}$ with u a primitive root in ${\mathbb{F}_p}^*$.

Now

$$(\xi + \sqrt{D})^n = g_n(\xi) + h_n(\xi)\sqrt{D}, \tag{49}$$

where g_n, h_n are integral polynomials

$$g_n(\xi) = \sum_{i=0}^{\lfloor n/2 \rfloor} \binom{n}{2i} D^i \xi^{n-2i}, \tag{50}$$

$$h_n(\xi) = \sum_{i=0}^{\lfloor \frac{n-1}{2} \rfloor} \binom{n}{2i+1} D^i \xi^{n-2i-1}. \tag{51}$$

Let

$$g_n(\xi, \eta) = \sum_{i=0}^{\lfloor n/2 \rfloor} \binom{n}{2i} D^i \xi^{n-2i} \eta^{2i}, \tag{52}$$

$$h_n(\xi, \eta) = \sum_{i=0}^{\lfloor \frac{n-1}{2} \rfloor} \binom{n}{2i+1} D^i \xi^{n-2i-1} \eta^{2i+1}. \tag{53}$$

Consequently we have

$$(\xi + \sqrt{D}\eta)^n = g_n(\xi, \eta) + h_n(\xi, \eta)\sqrt{D}. \tag{54}$$

Now we seek to bound $f_{H_1}(K)$ with K subgroup of $\mathbb{F}_p{}^*$ with $d_2 = \frac{p-1}{|K|}$. As in the hyperbolic case this is given by $\frac{M(d_1,d_2)}{d_1 d_2}$ where $M(d_1, d_2)$ now counts the number of points on the following curve in $\mathbb{F}_p{}^3$:

$$\begin{cases} \xi^2 - D\eta^2 = \kappa(x) \\ g_{d_1}(\xi, \eta) = \mu^{d_2} + \mu^{-d_2} \end{cases} \tag{55}$$

This is a curve of genus $O(d_1 d_2)$ and we can apply Weil bound and inclusion-exclusion as in the hyperbolic case to produce the primitive u. This completes the proof of Proposition 10. □

Lemma 11. *Suppose α_1, α_2 are not both zero and $\alpha_1\alpha_2 \neq 1 \mod p$. Then the curve*

$$\alpha_1 y^e + \alpha_2 y^{-e} = x^d + x^{-d}$$

is absolutely irreducible.

Proof. Consider

$$P(x, y) = \alpha_1 x^d y^{2e} + \alpha_2 x^d - x^{2d} y^e - y^e \in \overline{\mathbb{F}_p}[X, Y].$$

For $d = 1$, P is clearly irreducible. Let $d > 1$ and assume that P is not irreducible and $f(x, y) = \sum a_{jk} x^j y^k \in \overline{\mathbb{F}_p}[X, Y]$ an irreducible factor. Assume $d \geq e$ and u a d-th root of unity. Since for $0 \leq s \leq d$, $P(x, y) = P(u^s x, y)$, also

$$f_s(x, y) = f(u^s x, y) = \sum a_{jk} u^{sj} x^j y^k$$

is an irreducible component of P. Thus either f_s and $f_{s'}$ are coprime or $f_s \sim f_{s'}$. Since $a_{0k} \neq 0$ for some k (otherwise x would be a factor of $f(x, y)$), it follows that $f_s = f_{s'}$ if $f_s \sim f_{s'}$.

Case 1 The f_s are not pairwise coprime.

The $f(x, y) = f_s(x, y)$ for some $0 < s < d$, implying that $u^{s_j} = 1$, i.e. $s_j \equiv 0 (\mod d)$ if $a_{jk} \neq 0$. It follows that d has a divisor $d_1 > 1$ such that $d_1 | j$ if $a_{jk} \neq 0$ and hence $f(x, y)$ has form $f(x, y) = g(x^{d_1}, y)$. The polynomial $g(x, y)$ is therefore a factor of $Q(x, y) = \alpha_1 x^{d_2} y^{2e} + \alpha_2 x^{d_2} - x^{2d_2} y^e - y^e$ with $d_2 = \frac{d}{d_1}$ and we lowered d.

Case 2 The f_s $(0 \leq s \leq d)$ are mutually coprime. Define

$$P_1(x, y) = \prod_{s=0}^{d-1} f_s(x, y),$$

which divides P. Degree considerations show that

$$2d \geq d \deg_x f,\ 2e \geq d \deg_y f,\ 2d + e \geq d \deg f.$$

Case 2.1 $\deg_x f > 1,\ \deg_y f > 1$.

It follows that $\deg_x f = 2$, $\deg_y f = 2$, $e = d$, $\deg f = 3$, and $P(x,y) = P_1(x,y)$. With u as above, $\varphi(x,y) = f(x,uy)$ is an irreducible factor of $P(x,y)$. Therefore for some $0 \le s \le d$

$$\varphi(x,y) = \sum_{j,k\le 2} a_{jk}u^k x^j y^k \sim f_s(x,y) = \sum a_{jk}u^{sj}x^j y^k.$$

Consequently, there is some $0 \le l \le d$ such that $k - sj \equiv l(\mod d)$ if $a_{jk} \neq 0$. Since

$$\alpha_1 x^d y^{2d} + \alpha_2 x^d - x^{2d}y^d - y^d = \prod_{0\le s\le d} f_s(x,y),$$

clearly $a_{0,1} \neq 0$, $a_{1,0} \neq 0$, and therefore $1 \equiv l \equiv -s(\mod d)$, i.e. $k + j = 1(\mod d)$ if $a_{jk} \neq 0$. Since $\deg f = 3$, $2 \equiv 0(\mod d)$, hence $d = 2$ and $a_{1,1} = a_{2,0} = a_{0,2} = 0$. Thus

$$\begin{aligned}\alpha_1 x^2y^4 + ga_1x^2 - x^4y^2 - y^2 &\sim (a_{21}x^2y + a_{12}xy^2 + a_{10}x + a_{01}y)(a_{21}x^2y - a_{12}xy^2 - a_{10}x + a_{01}y)\\ &\sim y^2(a_{21}x^2 + a_{01})^2 - x^2(a_{12}y^2 + a_{10})^2.\end{aligned}$$

Setting $a_{0,1} = 1$ gives $-y^2(a_{21}x^2+1)^2+x^2(a_{12}y^2+a_{10})^2$ and $a_{21}^2 = 1$, $a_{21}-a_{12}a_{10} = 0$, $a_{12}^2 = \alpha_1$, $a_{10}^2 = \alpha_2$. But this contradicts the assumption $\alpha_1\alpha_2 \neq 1$.

Case 2.2 $\deg_x f = 1$ or $\deg_y f = 1$.

Assume $\deg_y f = 1$, say. Then there are coprime $a(x), b(x) \in \overline{\mathbb{F}_p}[X]$ such that $P(x, \frac{a(x)}{b(x)}) = 0$, that is

$$\alpha_1 x^d a(x)^{2e} + \alpha_2 x^d b(x)^{2e} - x^{2d}a(x)^e b(x)^e - a(x)^e b(x)^e = 0.$$

Since $a(x), b(x)$ are coprime, it follows that $a(x)^e | x^d$, $b(x)^e | x^d$, hence $a(x)$ or $b(x)$ is constant. If, say, $b(x)$ is constant, previous equation implies $x^d | a(x)^e$, hence $a(x)^e = \gamma x^d$ and

$$\alpha_1\gamma^2 x^{2d} + \alpha_2 b^{2e} - \gamma b^2 x^{2d} - \gamma b^2 = 0.$$

It follows that $\alpha_1\gamma = b^2$, $\alpha_2 b^2 = \gamma$, hence $\alpha_1\alpha_2 = 1$, contradicting the assumptions that $\alpha_1\alpha_2 \neq 1$. □

3.2. The Cage. By the *cage* we mean the set of maximal elements in $X^*(p)$. We claim that the cage is connected, that is to say if $\mathbf{x}$ and $\mathbf{y}$ are in the cage then $\mathbf{x}$ is connected to $\mathbf{y}$. Let ξ be a coordinate of maximal order of $\mathbf{x}$ and η be a coordinate of maximal order of $\mathbf{y}$, so that $\mathbf{x}$ is connected to all points in $C_j(\xi)$ and similarly $\mathbf{y}$ is connected to all the points in $C_k(\eta)$.

Now keeping in mind Proposition 8, we can apply inclusion/exclusion argument (as in Proposition 10) to obtain a y of maximal order, such that

$$\begin{cases} P \in C_j(\xi) \cap C_l(y) \\ Q \in C_k(\eta) \cap C_l(y) \end{cases} \tag{56}$$

provided we establish absolute irreducibility of the curve in question. In light of Lemma 9 the latter is of the form (for $l \geq 1$)

$$\begin{cases} u = t^l + t^{-l} \\ u^2 - A_1 v^2 = B_1 \\ u^2 - A_2 w^2 = B_2, \end{cases} \tag{57}$$

which becomes, after the change of variables

$$\begin{cases} u = t^l + t^{-l} \\ u^2 = \dfrac{1}{4}\left(\xi + \dfrac{B_2 - B_1}{\xi}\right)^2 + B_1. \end{cases} \tag{58}$$

We now have

$$(t^l + t^{-l})^2 = \frac{1}{4}\left(\xi + \frac{B_2 - B_1}{\xi}\right)^2 + B_1$$

or

$$t^{2l} + t^{-2l} = \frac{(B_2 - B_1)^2}{4\xi^2} + \left(B_1 + \frac{B_2 - B_1}{2} - 2\right) + \frac{\xi^2}{4} = \frac{(B_2 - B_1)^2}{4\xi^2} + \frac{\xi^2}{4} + \left(\frac{B_1 + B_2}{2} - 2\right),$$

leading to the curve of the form

$$t^{2l} + t^{-2l} = \alpha_1 \xi^2 + \alpha_2 \xi^{-2} + \alpha_3, \tag{59}$$

whose irreducibility follows from the following generalization of Lemma 11 for which we give a geometric group-theoretic proof, cf. [Cas70], [Pak09] (over the complex numbers) and [LMT93, Chapter 6], [Fri99] for fiber products and modifications to tame extensions over $\overline{\mathbb{F}_p}$.

Lemma 12. *For p odd, e and d integers less than p (in particular prime to p) and $\alpha_1, \alpha_2, \alpha_3 \in \overline{\mathbb{F}_p}$ satisfying $\alpha_1, \alpha_2, \alpha_3$ are not all zero and $\pm 2\sqrt{\alpha_1 \alpha_2} + \alpha_3 \neq \pm 2$,*

$$\alpha_1 y^e + \alpha_2 y^{-e} + \alpha_3 = x^d + x^{-d}$$

is irreducible over $\overline{\mathbb{F}_p}$.

Proof. Let $f(x) = x^d + x^{-d}$, $g(y) = \alpha_1 y^l + \alpha_2 y^{-l} + \alpha_3$. The curve $X_{f,g}$, defined by $f(x) - g(y) = 0$ can be viewed as a fiber product $\mathbb{P}^1_x \times_{\mathbb{P}^1_w} \mathbb{P}^1_y$, where the covers are given by

$$\begin{aligned} f &: \mathbb{P}^1_x \to \mathbb{P}^1_w \quad f(x) - w = 0; \\ g &: \mathbb{P}^1_y \to \mathbb{P}^1_w \quad g(y) - w = 0. \end{aligned}$$

The branch points for f are $\{-2, 2, \infty\}$ with branch cycles given by

$$\begin{cases} \sigma_{-2} = (12)(34)\dots(2d-1\,2d) \\ \sigma_2 = (12d)(23)\dots(2d-2\,2d-1) \\ \sigma_\infty = (135\dots 2d-1)(246\dots 2d). \end{cases} \tag{60}$$

The branch points for g are $\{-2\sqrt{\alpha_1\alpha_2} + \alpha_3, 2\sqrt{\alpha_1\alpha_2} + \alpha_3, \infty\}$ with branch cycles given by

$$\begin{cases} \tau_{-2\sqrt{\alpha_1\alpha_2}+\alpha_3} = (12)(34)\dots(2l-1\,2l) \\ \tau_{2\sqrt{\alpha_1\alpha_2}+\alpha_3} = (12l)(23)\dots(2l-2\,2l-1) \\ \tau_\infty = (135\dots 2l-1)(246\dots 2l). \end{cases} \tag{61}$$

Now the absolute irreducibility of $f(x) = g(y)$ is equivalent to the product of monodromy groups $\mathrm{Mon}(f)$, given by (60), and $\mathrm{Mon}(g)$, given by (61) acting transitively on the $2d \cdot 2l$ sheeted covering $\mathbb{P}^1_x \times_{\mathbb{P}^1_w} \mathbb{P}^1_y$, which is immediate in the case $\pm 2\sqrt{\alpha_1\alpha_2} + \alpha_3 \neq \pm 2$.

□

Keeping in mind (29), (30) the condition $\pm 2\sqrt{\alpha_1\alpha_2} + \alpha_3 \neq \pm 2 \mod p$ amounts to $\xi^2 \neq 0, \frac{1}{2} \mod p$, so Lemma 12 establishes the connectivity of the bulk of the cage, namely those points for which the coordinate of maximal order is not $\pm\frac{1}{\sqrt{2}} \mod p$. For the rest, the connectivity to the bulk follows from Proposition 10.

4. Middle Game

In the endgame (section 3) we connected any $\mathbf{x}$ of order $l \geq p^{\frac{1}{2}+\delta_0}$ ($\delta_0 > 0$) to the cage in one step. In this section we allow any number of moves to do the connecting. In particular, any $\mathbf{x}$ of order at least p^ε is shown to be in the giant component. As in section 3, the $\mathbf{y}$'s which are joined to a given $\mathbf{x}$ whose order is l_1 via the corresponding rotation and which have orders l_2 (here l_1 and l_2 divide $p-1$ or $p+1$) correspond to solutions of an equation (with $\sigma \in \mathbb{F}_p$, $\sigma \neq 1$ – see (13)):

$$\left.\begin{aligned} & h_1 + \frac{\sigma}{h_1} = h_2 + \frac{1}{h_2}, \sigma \neq 1 \\ & \text{with } h_1 \in H_1, h_2 \in H_2 \text{ with } H_1, H_2 \text{ subgroups of } {\mathbb{F}_p}^* \text{ or } {\mathbb{F}_{p^2}}^*. \end{aligned}\right\} \tag{62}$$

In (62) we have $|H_1| = l_1$, $|H_2| = l_2$.

If we have an upper bound on the number of solutions to (62) so that on summing over all $l_2 \leq l_1$ with l_2 dividing $p-1$ or $p+1$, yields a quantity which is less than l_1, then there is at least one h_1 for which the corresponding y will have order bigger than l_1. We then repeat this procedure replacing x by this y and so on, until the order of the element is at least $p^{\frac{1}{2}+\delta_0}$. At that point we are in the endgame and can finish.

The key therefore is a suitable upper bound to the number of solutions to (62).

Our original treatment used Stepanov's technique of auxiliary polynomials in his elementary proof of this Riemann Hypothesis for curves. This yields explicit and reasonably sharp estimates which are ample for our application. We carry this out in the Appendix partly to illustrate the flexibility of this method.

Subsequently the recent powerful technique for estimating from above the g.c.d. of $(u-1, v-1)$, of Corvaja and Zannier [CS13] yields sharper bounds. This is relevant for the purpose of giving effective bounds on p after which Theorem 1 takes effect. The precise upper bound to (62) established by [CS13] is

$$20 \max \Big\{ (|H_1|.|H_2|)^{1/3}, \frac{|H_1|.|H_2|}{p} \Big\}.$$

The third treatment, and the one which we develop in this section, while special to (62), is robust in that the upper bound requires little further structure and it is suitable for generalization for more general moduli as will demonstrated in [BGS24]. It is based on the following projective Szemeredi-Trotter theorem (Proposition 2 in [Bou12]).

Theorem 13. *Let $\Phi : \mathbb{F}_p \to \mathrm{Mat}_2(\mathbb{F}_p)$ be a polynomial map such that $\det \Phi$ does not vanish identically and $\mathrm{Im}\Phi \cap \mathrm{GL}_2(\mathbb{F}_p)$ is not contained in a set of the form $\mathbb{F}_p^* \cdot gH$ for some $g \in \mathrm{SL}_2(\mathbb{F}_p)$ and H a proper subgroup of $\mathrm{SL}_2(\mathbb{F}_p)$. Then the following holds.*

Given $\varepsilon > 0, r > 1$, there is $\delta > 0$ such that if $A \subset P^1(\mathbb{F}_p)$ and $L \subset \mathbb{F}_p$ satisfy

$$1 \ll |A| < p^{1-\varepsilon} \tag{63}$$

$$\log |A| < r \log |L|. \tag{64}$$

Then

$$|\{(x, y, t) \in A \times A \times L \,:\, y = \tau_{\Phi(t)}(x)\}| < |A|^{1-\delta}|L|, \tag{65}$$

where for $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, $\tau_g(x) = \frac{ax+b}{cx+d}$.

While producing poor exponents τ, this method is robust and works in the generality that the superstrong approximation for $\mathrm{SL}_2(\mathbb{Z}/q\mathbb{Z})$ has been established; in particular the analogue of Theorem 13 for $\mathbb{Z}/p^n\mathbb{Z}$ which follows from expansion in $\mathrm{SL}_2(\mathbb{Z}/p^n\mathbb{Z})$ established in [BG08d], plays crucial role in the proof of Theorem 4.

Using Theorem 13 we prove the following:

Proposition 14. *Given $\delta > 0$ there is $\tau < 1$ and C_τ depending on δ such that if $p^\delta < |H_1| < p^{1-\delta}$ then the number of solutions to (62) is at most $C_\tau |H_1|^{1-\tau}$.*

Proof. For $h \in H$, a subgroup of $\mathbb{F}_p^*$ or $\mathbb{F}_{p^2}^*$, denote by $\tilde{h} = h + h^{-1}$. Similarly we denote by $\widetilde{H} = \{\tilde{h} \,|\, h \in H\}$.

Suppose that (62) has T solutions. Then

$$\left.\begin{aligned} h_1 + \frac{\sigma}{h_1} &= u \\ h_1 t + \frac{\sigma}{h_1 t} &= v, \end{aligned}\right\} \tag{66}$$

where $h_1, t \in H_1$ and $u, v \in \widetilde{H_2}$, has at least T^2 solutions.

Elimination of h_1 in (66) yields

$$u^2 + v^2 - \Big(t + \frac{1}{t}\Big)uv + \sigma\Big(t - \frac{1}{t}\Big)^2 = 0 \tag{67}$$

which, by assumption, has at least T^2 solutions in $(t, u, v) \in H_1 \times \widetilde{H_2} \times \widetilde{H_2}$.

Next, let $u = \tilde{f}_1$, $v = \tilde{f}_2$ with f_1 and $f_2 \in H_2$ and define the following elements $x, y \in \widetilde{H_2}$:

$$x = \widetilde{(f_1 f_2)} = f_1 f_2 + \frac{1}{f_1 f_2},$$
$$y = \widetilde{(f_1 f_2^{-1})} = \frac{f_1}{f_2} + \frac{f_2}{f_1}.$$

Thus $uv = x + y, u^2 + v^2 = xy + 4$ and equation (67) gets transformed into

$$xy - \tilde{t}(x + y) + \sigma(\tilde{t})^2 + 4(1 - \sigma) = 0. \tag{68}$$

Denoting

$$\alpha = \tilde{t} \quad \text{and} \quad \beta = \sigma(\tilde{t})^2 + 4(1 - \sigma) \tag{69}$$

we obtain

$$y = \frac{\alpha x - \beta}{x - \alpha} = \tau_g(x) \tag{70}$$

with

$$g = \begin{pmatrix} \alpha & \beta \\ 1 & -\alpha \end{pmatrix} = g(\tilde{t})$$

and τ_g the Mobius transformation.

Equation (70) has at least T^2 solutions in $(x, y, \tilde{t}) \in \widetilde{H}_2 \times \widetilde{H}_2 \times \widetilde{H}_1$.

Now we apply Theorem 13 taking

$$\Phi(t) = \begin{pmatrix} t & -\sigma t^2 - 4(1-\sigma) \\ 1 & -t \end{pmatrix}, \tag{71}$$

and $A = \widetilde{H}_2, L = \widetilde{H}_1$.

We verify the assumption on Φ. Since $\sigma \neq 1$, $\det \Phi(t) = (1-\sigma)(4-t^2)$ does not vanish identically. It remains to show that

$$\Big\{ \Phi(s)^{-1}\Phi(t) \, \frac{\det \Phi(s)}{\det \Phi(t)}; s, t \in \mathbb{F}_p \Big\}$$

is not contained in a proper subgroup H of $\mathrm{SL}_2(p)$.

By (71)

$$\Phi(s)^{-1}\Phi(t) = \begin{pmatrix} -s & \sigma s^2 + 4(1-\sigma) \\ -1 & s \end{pmatrix} \begin{pmatrix} t & -\sigma t^2 - 4(1-\sigma) \\ 1 & -t \end{pmatrix}$$

$$= \begin{pmatrix} -st + \sigma s^2 + 4(1-\sigma) & (s-t)(4(1-\sigma) - \sigma st) \\ s - t & -st + \sigma t^2 + 4(1-\sigma)) \end{pmatrix}$$

Taking

$$s = \sigma t + \frac{4(1-\sigma)}{t}$$

gives

$$(1-\sigma)\Big(\frac{4}{t} - t\Big) \begin{pmatrix} \sigma(1+\sigma)t + \frac{4(1-\sigma)}{t} & -\sigma^2 t^2 + 4(1-\sigma)^2 \\ 1 & 0 \end{pmatrix} = (1-\sigma)\Big(\frac{4}{t} - t\Big) g_t. \tag{72}$$

As the proper subgroups of $\mathrm{SL}(2, \mathbb{F}_p)$ have trivial second commutator [Suz82], it suffices to show that

$$(g_{t_1} g_{t_2} g_{t_1}^{-1} g_{t_2}^{-1})(g_{t_3} g_{t_4} g_{t_3}^{-1} g_{t_4}^{-1})(g_{t_2} g_{t_1} g_{t_2}^{-1} g_{t_1}^{-1})(g_{t_4} g_{t_3} g_{t_4}^{-1} g_{t_3}^{-1}) \tag{73}$$

is not identically one for $t_1, t_2, t_3, t_4 \in \mathbb{F}_p^*$. If this were the case, the same would be true for t_1, t_2, t_3, t_4 taken in an extension field of $\mathbb{F}_p$ so as to make

$$t^2 = \frac{4(1-\sigma)^2 - \varepsilon}{\sigma^2} \; (\varepsilon = \pm 1) \tag{74}$$

solvable.

Taking $t = \pm\kappa$ satisfying (74), we get

$$g_{\pm\kappa} = \begin{pmatrix} \pm\frac{\sigma}{\kappa}[(1+\sigma)\frac{4(1-\sigma)^2-\varepsilon}{\sigma^2} + 4(1-\sigma)] & \varepsilon \\ 1 & 0 \end{pmatrix}. \tag{75}$$

We choose $\varepsilon = \pm 1$ as to ensure that

$$(1+\sigma)(4(1-\sigma)^2 - \varepsilon) + 4\sigma^2(1-\sigma) \neq 0$$

and obtain matrices

$$g_{\pm} = \begin{pmatrix} \pm\eta & \varepsilon \\ 1 & 0 \end{pmatrix} \tag{76}$$

that clearly generate $\mathrm{SL}_2(p)$.

Consequently, Theorem 13 is applicable, yielding the bound $T^2 \ll |H_2|^{1-\tau}|H_1|$. □

Proposition 14 yields the following corollary:

Corollary 15. *Every $\mathbf{x} \in X^*(p)$ whose order is at least p^ε is in the giant component $\mathcal{C}(p)$.*

5. Opening

The analysis of the previous sections shows that we can connect $\mathbf{x} \in X^*(p)$ whose order is at least p^ε (or smaller if the divisors of p^2-1 are not too numerous) to the cage. To deal with all x's and in particular ones whose orders are uniformly bounded (independent of p) we lift to characteristic zero. In this connection we observe first that if the action of Γ on $X^*(\bar{\mathbb{Q}})$ has a finite orbit F then the strong approximation conjecture cannot hold. To see this consider more generally any finite orbit F of the Γ action on $\mathbb{A}^3(\mathbb{C})$. Any coordinate of any ξ in such an F must lie in a cyclotomic field $L_n = \mathbb{Q}(\zeta_n)$, where ζ_n is a primitive n-th root of 1. For if $\xi = (\xi_1, \xi_2, \xi_3)$ then $\mathrm{ord}(\mathrm{rot}(\xi_j))$ must be finite and hence

$$\xi_j = t_j + t_j^{-1} \tag{77}$$

with t_j a root of unity. If n is the least common multiple of all the orders of all the t_j's corresponding to the ξ's in F then $\xi_j \in L_n$ and hence $F \subset \mathbb{A}^3(L_n)$. In fact the ξ_j's are all integral, so that $F \subset \mathbb{A}^3(\mathcal{O}_{L_n})$ where $\mathcal{O}_{L_n}$ is the ring of integers in L_n. If p is a rational prime ($p \neq 3$) which splits completely in L_n and P is a prime of L_n with $P|(p)$ then $\mathcal{O}_{L_n}/P \cong \mathbb{F}_p (\cong \mathbb{Z}/p\mathbb{Z})$. The Γ action of $\mathbb{A}^3(\mathcal{O}_{L_n})$ factors through the

reduction π mod P

$$\begin{array}{ccc} \mathbb{A}^3(\mathcal{O}_{L_n}) & \xrightarrow{\Gamma} & \mathbb{A}^3(\mathcal{O}_{L_n}) \\ \downarrow \pi & & \downarrow \pi \\ \mathbb{A}^3(\mathcal{O}_L/P) & \xrightarrow{\bar{\Gamma}} & \mathbb{A}^3(\mathcal{O}_L/P) \end{array}$$

and hence

$$\bar{F} = \pi(F) \subset \mathbb{A}^3(\mathbb{F}_p), \text{ is } \bar{\Gamma} - \text{invariant.} \tag{78}$$

Since Γ preserves the level sets X_k:

$$x_1^2 + x_2^2 + x_3^2 - 3x_1x_2x_3 = k, \tag{79}$$

any such F is contained in $X_k(\mathcal{O}_L^{(3)})$ for a suitable k. Thus for any such F, there is a positive density of p's which split completely in L_n, and hence for which $\bar{F} \subset \mathbb{X}_k(\mathbb{F}_p)$ is a fixed size $\bar{\Gamma}$ -orbit ($|\bar{F}| \leq |F|$). That is, the finite Γ-orbits in $\mathbb{A}^3(\bar{\mathbb{Q}})$ must be part of any description of the $\bar{\Gamma}$-orbits on $\mathbb{A}^3(\mathbb{F}_p)$, for p large.

In our setting of this paper, $k = 0$ and we have (we thank E. Bombieri for this simple proof)

Proposition 16. *$X^*(\bar{\mathbb{Q}})$ has no finite Γ-orbit.*

Proof. As in the discussion above, if F is such an orbit and $\xi \in F$ then the ξ_j satisfy (77) with t_j an l_j-th root of one. The Markoff equation for t_1, t_2, t_3 becomes

$$(t_1 + t_1^{-1})^2 + (t_2 + t_2^{-1})^2 + (t_3 + t_3^{-1})^2 - (t_1 + t_1^{-1})(t_2 + t_2^{-1})(t_3 + t_3^{-1}) = 0. \tag{80}$$

Now (80) has no solutions with $|t_j| = 1$ (let alone being roots of unity) except for $t_j = \pm i$, $j = 1, 2, 3$. To see this note that if

$$a = t_1 + t_1^{-1} (= t_1 + \bar{t_1}),\ b = t_2 + t_2^{-1},\ c = t_3 + t_3^{-1}$$

then a, b, c, lie in $[-2, 2]$ and by the inequality of the geometric and arithmetic means

$$0 \leq a^2 + b^2 + c^2 = |abc| \leq \frac{|a|^3 + |b|^3 + |c|^3}{3} \leq \frac{2}{3}(a^2 + b^2 + c^2). \tag{81}$$

Hence the only solutions to (80) correspond to $a = b = c = 0$ or $t_j = \pm i$. In terms of ξ_j this gives $\xi = (0, 0, 0)$, which is the only invariant set for the action of Γ on $\mathbb{X}(\bar{\mathbb{Q}})$. □

We remark that in the context of the general surfaces that are studied in [BGS24], for example the surfaces $\mathbb{X}_k$ in (79) with $k \neq 0$, there can be a continuum of solutions to the analogue of equation (80) with $|t_j| = 1$. However the solutions with t_j a root of unity (with unspecified order) are still restricted to a finite number of nondegenerate solutions. This follows from Lang's G_m conjecture, see [Lau83] and [SA94] for proofs

which give the solutions effectively. In various special cases these finite $\bar{\mathbb{Q}}$ orbits for the Γ-action correspond to the determination of the algebraic Painleve VI solutions ([DM00], [LT14]); we leave the details to [BGS24] .

Returning to the Markoff surface $\mathbb{X}$, let $\xi = (\xi_1, \xi_2, \xi_3) \in X^*(p)$ with $\mathrm{ord}(\mathrm{rot}(\xi_j)) = l_j$ for $j = 1, 2, 3$. Let $n = \mathrm{lcm}(l_1, l_2, l_3)$ and $L_n = \mathbb{Q}(\zeta_n)$ and let $\zeta_{l_1}, \zeta_{l_2}, \zeta_{l_3}$ be primitive roots of one respectively. Let

$$\eta = (\zeta_{l_1}+\zeta_{l_2}^{-1})^2+(\zeta_{l_2}+\zeta_{l_2}^{-1})^2+(\zeta_{l_3}+\zeta_{l_3}^{-1})^2-(\zeta_{l_1}+\zeta_{l_2}^{-1})(\zeta_{l_2}+\zeta_{l_2}^{-1})(\zeta_{l_3}+\zeta_{l_3}^{-1}) \in \mathcal{O}_{L_n}. \tag{82}$$

According to the proof of Proposition 16, unless $l_1 = l_2 = l_3 = 4$ (i.e. $\zeta_{l_j} = \pm i$), $\eta \neq 0$. Now $|\eta| \leq 20$ and hence

$$\mathrm{Norm}(\eta) \leq 20^{\phi(n)} \leq 20^n. \tag{83}$$

If P is a prime in $\mathcal{O}_{L_n}$ and $\eta \in P$, then $P|(\eta)$ and hence

$$\mathrm{N}(P) \leq \mathrm{Norm}(\eta) \leq 20^n. \tag{84}$$

Put differently, if

$$\log_{20} \mathrm{N}(P) > n$$

then

$$\eta \neq 0(\bmod P). \tag{85}$$

For our point ξ in $X^*(p)$, $\xi_j = \lambda_j + \lambda_j^{-1}$ with λ_j in $\mathbb{F}_p$ or $\mathbb{F}_{p^2}$ and λ_j an l_j-th root of 1, and $(l_1, l_2, l_3) \neq (4, 4, 4)$ since $\xi \neq (0, 0, 0)$. If all the λ_j's are in $\mathbb{F}_p$ then $\mathbb{Q}(\zeta_n)$ splits completely at p, that is there is a prime P dividing (p) such that

$$\mathcal{O}_{L_n}/P \cong \mathbb{F}_p \ , \ \mathrm{N}(P) = p$$

and $\pi(\zeta_{l_j}) = \lambda_j$ in $\mathcal{O}_{L_n}/P$ and $\eta \equiv 0(\mod P)$. Hence from (85) we conclude that

$$\log_{20} p \leq n. \tag{86}$$

If the field generated by λ_j's (over $\mathbb{F}_p$) is $\mathbb{F}_{p^2}$ then there is a prime P of $\mathcal{O}_{L_n}$ dividing (p) such that

$$\mathcal{O}_{L_n}/P \cong \mathbb{F}_{p^2} \ , \ \mathrm{N}(P) = p^2$$

and $\pi(\zeta_{l_j}) = \lambda_j$ in $\mathcal{O}_{L_n}/P$ and $\eta \equiv 0(\bmod P)$. Hence again from (85) we conclude that

$$2\log_{20} p \leq n. \tag{87}$$

Hence in either case $n \geq \log_{20} p$ where $n = \mathrm{lcm}(l_1, l_2, l_3)$, and hence

$$\max(l_1, l_2, l_3) \geq (\log_{20} p)^{\frac{1}{3}}. \tag{88}$$

We have proven

Proposition 17. *Let $\xi \in X^*(p)$, and let $l = \max(l_1, l_2, l_3)$ with $l_j = \mathrm{ord}(\mathrm{rot}(\xi_j))$; then $l \geq (\log_{20} p)^{\frac{1}{3}}$. In particular, any component F of $X^*(p)$ satisfies*

$$|F| \geq (\log_{20} p)^{\frac{1}{3}}.$$

6. Proofs of Theorems 1 and 2

The first part of Theorem 1 follows from Corollary 15, combined with the fact that there are at most $p^{\varepsilon'}$ elements $a \in {\mathbb{F}_p}^*$ (or in $\mathbb{F}_{p^2}^*$ satisfying $a^{p+1} = 1$) of order less than p^{ε}. Proposition 17 establishes the second part of Theorem 1 and, combined with the analysis in sections 3 and 4, yields a proof of the strong approximation conjecture if $p^2 - 1$ is not very smooth. For example, the strong approximation conjecture is true for $X^*(p)$ if the prime p satisfies

$$\sum_{\substack{(\log p)^{\frac{1}{3}} \leq d \leq y \\ d | (p^2-1)}} d^{\frac{2}{3}} < y; \text{ for any } y. \tag{89}$$

The proof proceeds by using the arguments and results in [Cha13] and [CKSZ14] concerning points (x, y) on irreducible curves over $\mathbb{F}_p$ for which $\mathrm{ord}(x) + \mathrm{ord}(y)$ is small (here $\mathrm{ord}(x)$ is the order of x in $\mathbb{F}_p^*$).

Theorem 18. *Fix $d \in \mathbb{Z}_+$ and $\delta > 0$. There is an $\varepsilon > 0$, $\varepsilon = \varepsilon(d, \delta)$, such that for all primes $p \leq z$ (z sufficiently large) with the exception of at most z^{δ} of them, the following property holds. Let $f(x, y) \in \mathbb{F}_p[x, y]$ be of degree at most d and not divisible by any non-constant polynomial of the form $\rho x^{\alpha} y^{\beta} - 1$ or $\rho y^{\beta} - x^{\alpha}$ for any $\rho \in \bar{\mathbb{F}}_p$ and integers α and β. Then all solutions $(x, y) \in (\bar{\mathbb{F}}_p \times \bar{\mathbb{F}}_p)^*$ of $f(x, y) = 0$ satisfy*

$$\mathrm{ord}(x) + \mathrm{ord}(y) \geq p^{\varepsilon} \tag{90}$$

except for at most $11d^3 + d$ of them.

Proof. Theorem 1.2 in [CKSZ14] establishes what we want except that p^{ε} in (90) is replaced by the stronger bound $p^{\alpha(d)}$, with

$$\alpha(d) = \frac{2}{89d^2 + 3d + 14},$$

while the exceptional set of primes is of zero density. For our purpose the exponent in (90) is allowed to be small and in exchange we want the exceptional set to be much smaller. To this end we follow verbatum the discussion in Section 4 of [CKSZ14]. For d fixed and T a large parameter they show that there is a $U = U(d, T)$ which has at

most $O(T^{\frac{1}{\alpha(d)}})$ prime factors (their $\log T$ in the denominator is irrelevant for us) with the property: If p does not divide U and f as in Theorem 18 and $f(x,y)=0$ in $\mathbb{F}_p$, then

$$\mathrm{ord}(x)+\mathrm{ord}(y)\geq T \tag{91}$$

except for at most $11d^3+d$ such (x,y) in ${\mathbb{F}_p}^*\times{\mathbb{F}_p}^*$. For our given $\delta>0$ and large parameter z choose T to be

$$T=z^{\delta\alpha(d)}. \tag{92}$$

Then the number of primes p with $p|U$ is $O(z^\delta)$, and if p does not divide U then Theorem 18 holds with (91) and (92), that is with $\varepsilon=\delta\alpha(d)$.

□

To prove Theorem 2 we apply Theorem 18 to the curves $f_\sigma(x,y)$ given by equation

$$x+\frac{\sigma}{x}=y+\frac{1}{y}$$

with $\sigma\neq 1$. If $(\log_{20}p)^{\frac{1}{3}}>1000$, then according to Proposition 17, for any $\xi\in X^*(p)$ we have $\mathrm{ord}(\mathrm{rot}(\xi_{j_0}))$ is at least 1000 for j_0 one of the j's in $\{1,2,3\}$. Hence if p is not in the exceptional set in Theorem 18 with $d=3$, then in the typical equation $x+\frac{\sigma}{x}=y+\frac{1}{y}$, corresponding to the orders of the rotations in the $\mathrm{rot}(\xi_{j_0})$ orbit, there is (x,y) which is not one of the exceptional $11d^3+d<1000$ possible points. For such (x,y) the induced rotation has order at least p^ε and hence ξ is joined to the cage by the middle game.

Our methods fall short of dealing with all p, specifically for those rare p's for which p^2-1 is very smooth. The following hypothesis which is a strong variant of conjectures of M.C. Chang and B. Poonen [Cha13, Vol10] would suffice to deal with all large p's.

Hypothesis: *Given $d\in\mathbb{N}$, there is $\delta>0$ and $K=K(d)$ such that for p large and $f(x,y)$ absolutely irreducible over $\mathbb{F}_p$ and of degree d and $f(x,y)=0$ is not a translate of a subtorus of $(\bar{\mathbb{F}}_p^*)^2$, the set of $(x,y)\in(\mathbb{F}_p^*)^2$ for which $f(x,y)=0$ and $\max(\mathrm{ord}x,\mathrm{ord}y)\leq p^\delta$, is at most K.*

7. Proof of Theorem 3

We turn to the proof of Theorem 3 and establish the following slight strengthening thereof.

Theorem 19. *Almost all Markoff numbers are highly composite, that is for every $\nu \geq 1$, as $T \to \infty$*

$$\sum_{\substack{m \in \mathbb{M}^s, m \leq T \\ m \text{ has at most} \\ \nu \text{ distinct prime factors}}} 1 = o\Big(\sum_{\substack{m \in \mathbb{M}^s \\ m \leq T}} 1 \Big).$$

The proof makes use of counting points on $X^*(\mathbb{Z})$ of height at most T and in particular Mirzakhani's orbit equidistribution [Mir16], as well as the transitivity properties of Γ on $X^*(q)$ for q a product suitable primes p. Our original treatment established that for large primes $p_1 \leq p_2$, $p_j \equiv 1(4)$, for which Conjecture 1 holds, Γ acts transitively on $X^*(p_1 p_2)$. This was combined with Theorem 2 and a simple sieving argument in section 7 of [Hoo76] to deduce Theorem 3. The details of these arguments can be found in version 2 of [BGS16a].

We proceed here more directly using subsequent results of Meiri and Puder [MP18]. For $p \equiv 1(4)$ for which the induced permutation action of Γ on $X^*(p)$ is transitive, they show that the resulting permutation group is essentially the full symmetric or alternating group on $X^*(p)$. Applying Goursat's (disjointness) Lemma leads to the Γ-action on $X^*(p_1 p_2 \cdots p_k)$ being transitive for any such primes $p_1 < p_2 \cdots < p_k$.

In what follows we let Q be be a set of primes $p \equiv 1(4)$ for which Conjecture 1 holds. We estimate the sizes of various subsets of $X^*(q)$ where $q = \prod_{p \in Q} p$. Let

$$L = L_Q = \sum_{p \in Q} \frac{1}{p}, \tag{93}$$

which we assume is large.

$$X^*(q) \cong \prod_{p \in Q} X^*(p),$$

and since $p \equiv 1(4)$ one checks that

$$|X^*(p)| = p^2 + 3p. \tag{94}$$

Thus

$$|X^*(q)| = \prod_{p \in Q} |X^*(p)| = \prod_{p \in Q} (p^2 + 3p). \tag{95}$$

In what follows we denote elements of X^* by (x, y, z).

$$|\{\alpha \in X^*(p) \,:\, x(\alpha) \equiv 0(p)\}| = |\{y^2 + z^2 \equiv 0(p), (y, z) \neq (0, 0)\}| = 2p - 2. \tag{96}$$

Hence

$$\begin{aligned}&\{\alpha \in X^*(q) \,:\, (x(\alpha), q) = 1\}| = |\{\alpha \in X^*(q) \,|\, (x(\alpha), p) = 1 \,\text{for all}\, p \in Q\}| \\ &= \prod_{p\in Q}(|X^*(p)| - (2p-2)) = |X^*(q)| \prod_{p\in Q}(1 - \frac{2p-2}{p^2+3p}).\end{aligned} \tag{97}$$

Now

$$\log \prod_{p\in Q}(1 - \frac{2p-2}{p^2+3p}) = -2\sum_{p\in Q}\frac{1}{p} + O(1)$$

(the implied constant being absolute).

Hence

$$\prod_{p\in Q}(1 - \frac{2p-2}{p^2+3p}) \ll e^{-2L_Q} \tag{98}$$

and

$$|\{\alpha \in X^*(q) \,:\, (x(\alpha), q) = 1\}| \ll |X^*(q)|e^{-2L_Q}. \tag{99}$$

More generally if $R \subset Q$ with $|R| = t$ then

$$\begin{aligned}&|\{\alpha \in X^*(q) \,:\, x(\alpha) \equiv 0(p) \,\text{for}\, p \in R, \,\text{and}\, x(\alpha) \neq 0(p) \,\text{for}\, p \notin R\}| \\ &= |X^*(q)| \prod_{p\notin R}(1 - \frac{2p-2}{p^2+3p}) \prod_{p\in R}\frac{2p-2}{p^2+3p} \\ &\ll |X^*(q)| 2^t e^{-2L_{Q\setminus R}} \prod_{p\in R}\frac{1}{p} \\ &\leq |X^*(q)|(2e^2)^t e^{-2L_Q} \prod_{p\in R}\frac{1}{p}.\end{aligned} \tag{100}$$

Thus

$$\begin{aligned}&|\{\alpha \in X^*(q) \,:\, x(\alpha) \equiv 0(p) \,\text{for exactly}\, t \,p\text{'s in}\, Q\}| \\ &= \sum_{\substack{R\subset Q \\ |R|=t}} |\{\alpha \in X^*(q) \,:\, x(\alpha) \equiv 0(p), p \in R \,\text{and}\, x(\alpha) \neq 0(p), p \notin R\}| \\ &\ll |X^*(q)|(2e^2)^t e^{-2L_Q} \sum_{\substack{R\subset Q \\ |R|=t}} \prod_{p\in R}\frac{1}{p} \\ &\leq |X^*(q)|(2e^2)^t e^{-2L_Q} \left(\sum_{p\in Q}\frac{1}{p}\right)^t \\ &= |X^*(q)|(2e^2 L_Q)^t e^{-2L_Q}.\end{aligned} \tag{101}$$

For $\nu \geq 0$ let $X^*_\nu(q)$ be the set of $\alpha \in X^*(q)$ for which $x(\alpha) \equiv 0(p)$ for at most ν p's in Q. Then

$$|X^*_\nu(q)| \leq \sum_{t=0}^{\nu} |\{\alpha \in X^*(q) \,:\, x(\alpha) \equiv 0(p) \text{for exactly } t \text{ } p\text{'s in } Q\}|,$$

which by (101)

$$\ll |X^*(q)| e^{-2L_Q} \sum_{t=0}^{\nu} (2e^2 L_Q)^t.$$

This leads to our main estimate

$$|X^*_\nu(q)| \ll |X^*(q)|(\nu+1)(2e^2 L_Q)^\nu e^{-2L_Q}. \tag{102}$$

We apply (102) in Mirzakhani's result (Corollary 3.1 in [Mir16]; see also [ES22] for another treatment) to count points in $X^*(\mathbb{Z})$ with $|m|_\infty \leq T$ using the orbit of Δ_q, the stabilizer in Γ of $(1,1,1)$ in $X^*(q)$.

For $\alpha \in X^*(q)$, as $T \to \infty$

$$\sum_{\substack{m \in X^*(\mathbb{Z}) \\ |m|_\infty \leq T \\ m \equiv \alpha(q)}} 1 \sim \begin{cases} \frac{1}{|\mathcal{O}_{(1,1,1)}(q)|} \sum_{\substack{m \in X^*(\mathbb{Z}) \\ |m|_\infty \leq T}} 1 \text{ if } \alpha \in \mathcal{O}_{(1,1,1)}(q) \\ 0 \text{ otherwise} \end{cases}. \tag{103}$$

Here $\mathcal{O}_{(1,1,1)}(q)$ is the Γ-orbit of $(1,1,1)$ in $X^*(q)$.

In particular for our choice of Q and the transitivity of Γ on $X^*(q)$ we have that for Q and ν fixed, as $T \to \infty$

$$\sum_{\substack{m \in X^*(\mathbb{Z}) \\ |m|_\infty \leq T \\ \bar{m} \in X^*_\nu(q)}} 1 \sim \frac{|X^*_\nu(q)|}{|X^*(q)|} \sum_{\substack{m \in X^*(\mathbb{Z}) \\ |m|_\infty \leq T}} 1, \tag{104}$$

where $\bar{m}$ denotes reduction modulo q.

If $m \in X^*(\mathbb{Z})$ and $x(m)$ has at most ν distinct prime factors then it projects into $X^*_\nu(q)$ and hence, denoting by $w(n)$ the number of distinct prime factors of n, we have from (104) and (102)

$$\limsup_{T \to \infty} \frac{\sum_{\substack{|m|_\infty \leq T \\ w(m) \leq \nu}} 1}{\sum_{|m|_\infty \leq T} 1} \leq (\nu+1)(2e^2 L_Q)^\nu e^{-2L_Q}. \tag{105}$$

Since according to Theorem 2 we can choose Q so as to make L_Q as large as we please, and also the transitivity of Γ on $X^*(q)$ holds by [MP18], it follows that the the left hand side of (105) is equal to 0.

The above arguments apply equally well with $x(\alpha)$ replaced by $y(\alpha)$ or $z(\alpha)$ and hence to $\max(x(\alpha), y(\alpha), z(\alpha))$. Hence for ν fixed we have that as $T \to \infty$

$$\sum_{\substack{m \in \mathbb{M}^s \\ m \leq T \\ w(m) \leq \nu}} 1 = o\Big(\sum_{\substack{m \in \mathbb{M}^s \\ m \leq T}} 1 \Big).$$

This completes the proof of Theorem 19.

Appendix **A**

Stepanov's auxiliary polynomial method [Ste69] for bounding the number of solutions to equations like (62) is quite flexible. We demonstrate this for some special cases (the general case can be handled similarly). The proposition below is an extension of the approach and bounds in [HK00] (where $S(x) = x$, $T(x) = 1 - x$ and $t_1 = t_2$).

In what follows $S(x)$ and $T(x)$ are rational functions in $\mathbb{F}_p(x)$ of total degree d_1 and d_2 respectively and with disjoint divisors; $e = d_1 + d_2$ is fixed.

Proposition 20. *For p a large prime, t_1, t_2 dividing $p-1$, $t_1 \geq t_2$, let*

$$Y = \{y \in \mathbb{F}_p \,:\, S(y)^{t_1} = T(y)^{t_2} = 1\}.$$

Then if $t_1 \ll_e p^{1-\frac{1}{2e}}$,

$$|Y| \ll_e \min\{t_2, t_1 t_2^{-\frac{1}{4e}}\}.$$

Remarks:

(1) The trivial bound is $O(t_2)$ so the Proposition gives an improvement (power saving) if $t_2 \geq t_1^{\frac{4e}{4e-1}}$.
(2) If $h(\xi, \eta) = 0$ is a plane curve of genus 0 over $\mathbb{F}_p$, then the Proposition gives an upper bound on the number of solutions with $\xi^{t_1} = \eta^{t_2} = 1$ (cf. [CS13]).

Applying Proposition 20 with $t_1 = t_2$, $S(y) = y$, $T(y) = \frac{ay+b}{cy+d}$ yields

Corollary 21. *For p large prime, $t|(p-1)$, $t \leq p^{\frac{3}{4}}$ and $U_t = \{y \in \mathbb{F}_p \,:\, y^t = 1\}$ the t-th roots of 1,*

$$|\sigma(U_t) \cap U_t| \ll t^{\frac{3}{4}}$$

for $\sigma \in PGL_2(\mathbb{F}_p)$, $\sigma \neq 1$.

Corollary 22. *For $t|(p-1)$, $t \leq p^{\frac{3}{4}}$, $b \in \mathbb{F}_p$, $b \neq 1$,*

$$|\{w, \rho \in \mathbb{F}_p \,:\, w + w^{-1} = \rho + b\rho^{-1}, w^t = \rho^t = 1\}| \ll t^{\frac{3}{4}}.$$

Proof. Put $\rho w = \xi, \frac{w}{\rho} = \eta$, then $\xi^t = \eta^t = 1$ and each such solution with $\xi = \frac{b\eta - 1}{\eta - 1}$ corresponds to at most two solutions (w, ρ) above. Applying Corollary 21 yields Corollary 22. □

Proof of Proposition 20: First we need a generalization of Proposition 3.2 in [VS12] where their common t is replaced by $t_0, t_1, \ldots, t_n$. The result is the following Lemma, whose proof is the same

Lemma 23. *Let $t_0, t_1, \ldots, t_n$, as well as B and J be integers, p a large prime, and $\alpha_1, \ldots, \alpha_n$ distinct elements in $\mathbb{F}_p{}^*$. Assume that*

$$\min(t_0, \ldots, t_n) \geq \frac{1}{2}(n-1)B^{2n} + JB$$

and that

$$p \geq (2nB + 2)\max(t_0, t_1, \ldots, t_n).$$

Then

$$x^{a_i} x^{t_0 b_{0,i}} (x - \alpha_1)^{t_1 b_{1,i}} \ldots (x - \alpha_n)^{t_n b_{n,i}}$$

with $a_j \leq J$ and $b_{0,i}, \ldots b_{n,i} \leq B$ are linearly independent in $\mathbb{F}_p[x]$.

Let $\alpha_1, \ldots, \alpha_k \in \mathbb{F}_p$ be distinct and $\nu_1, \nu_2, \ldots \nu_k \in \mathbb{Z}$; set

$$R_\nu(x) = (x - \alpha_1)^{\nu_1} \ldots (x - \alpha_k)^{\nu_k}.$$

For $m \geq 1$,

$$\frac{d^m}{dx^m}\left[R_\nu(x)\right] = \sum_{j_1 + \ldots j_k = m} \binom{m}{j} \frac{d^{j_1}}{dx^{j_1}}\left[(x - \alpha_1)^{\nu_1}\right] \ldots \frac{d^{j_k}}{dx^{j_k}}\left[(x - \alpha_1)^{\nu_k}\right] =$$

$$\sum_{j_1 + j_2 + \ldots j_k = m} B_{m,j} (x - \alpha_1)^{\nu_1 - j_1} \ldots (x - \alpha_k)^{\nu_k - j_k}.$$

Hence

$$[(x - \alpha_1)(x - \alpha_2) \ldots (x - \alpha_k]^m \frac{d^m}{dx^m} R_\nu(x) = R_\nu(x) P_{m,\nu}(x), \tag{106}$$

where $P_{m,\nu}$ is a polynomial of degree at most km.

Stepanov's polynomial method is based on constructing a polynomial which vanishes to high order on Y. Let λ_{a,b_1,b_2} be in $\mathbb{F}_p$ with $0 \leq a \leq J$ and $0 \leq b_j \leq B$ and form

$$\phi(x) = \sum \lambda_{a,b_1,b_2} x^a (S(x))^{t_1 b_1} (T(x))^{t_2 b_2}. \tag{107}$$

Write $S(x), T(x)$ in the form (we assume that both factor into linear factors $\mathbb{F}_p[x]$):

$$S(x) = \frac{A(x-\alpha_1)\dots(x-\alpha_t)}{(x-\beta_1)\dots(x-\beta_\tau)}, \tag{108}$$
$$T(x) = \frac{B(x-\gamma_1)\dots(x-\gamma_\mu)}{(x-\delta_1)\dots(x-\delta_\nu)}.$$

For simplicity we assume that $S(x)$ and $T(x)$ are square-free and we are assuming that the $\alpha, \beta, \gamma, \delta$'s are all distinct. The constants A and B can be absorbed into the λ's, so without loss of generality we can take $A = B = 1$. For $m \geq 0$

(109)
$$[(x-\alpha_1)\dots(x-\alpha_t)(x-\beta_1)\dots(x-\beta_\tau)\dots(x-\delta_\nu)]^m \frac{d^m}{dx^m}\left[x^a(S(x))^{t_1b_1}(T(x))^{t_2b_2}\right] = x^a S(x)^{t_1b_1}T(x)^{t_2b_2}P_{m,a,b_1,b_2}(x)$$

with P_m of degree at most em. Hence for $x = y \in Y$ and $m \leq M$

(110)
$$\frac{d^m}{dx^m}\phi(x)|_{x=y} = \sum \lambda_{a,b_1,b_2}\frac{d^m}{dx^m}\left[x^a S(x)^{t_1b_1}T(x)^{t_2b_2}\right]_{x=y} = \sum_{a,b_1,b_2} \lambda_{a,b_1,b_2} y^a P_{m,a,b_1,b_2}(y),$$

by the definition of Y. We can make (110) equal to 0 for all y in Y by noting that $y^a P_m(y)$ is a polynomial of degree $J + em$. So (110) can be made 0 with not all of the λ_{a,b_1,b_2}'s equal to 0 and for all $m \leq M$ as long as

$$(J + eM)M < B^2 J. \tag{111}$$

Assuming that this is satisfied, we have $\phi(x)$ which is not identically zero and has degree (as a rational function) at most $J + eBt_1$.

Hence if $\phi(x)$ is not identically zero, then its order of vanishing on Y is at least M and hence

$$M|Y| \leq J + eBt_1$$

or

$$|Y| \leq \frac{J + eBt_1}{M}. \tag{112}$$

We now check that under suitable constraints on the sizes of parameters, $\phi(x)$ does not vanish identically. We have

$$\phi(x) = \sum \lambda_{a,b_1,b_2} x^a \frac{(x-\alpha_1)^{t_1b_1}\dots(x-\alpha_t)^{t_1b_1}}{(x-\beta_1)^{t_1b_1}\dots(x-\beta_\tau)^{t_1b_1}} \frac{(x-\gamma_1)^{t_2b_2}\dots(x-\gamma_\mu)^{t_2b_2}}{(x-\delta_1)^{t_2b_2}\dots(x-\delta_\nu)^{t_2b_2}},$$

consequently

$$(x-\beta_1)^B\dots(x-\beta_\tau)^B(x-\delta_1)^B\dots(x-\delta_\nu)^B\phi(x)=$$
$$\sum\lambda_{a,b_1,b_2}x^a(x-\alpha_1)^{t_1b_1}\dots(x-\alpha_t)^{t_1b_1}(x-\beta_1)^{(B-b_1)t_1}\dots(x-\beta_\tau)^{(B-b_1)t_1}\cdot$$
$$\cdot(x-\gamma_1)^{t_2b_2}\dots(x-\gamma_\mu)^{t_2b_2}(x-\delta_1)^{(B-b_2)t_2}\dots(x-\delta_\nu)^{(B-b_2)t_2}.$$

Now the monomials appearing in the last expression are linearly independent over $\mathbb{F}_p[x]$ according to Lemma 23 as long as

$$\begin{aligned} p &\geq (2eB+2)t_1 \\ t_2 &\geq \frac{1}{2}eB^{2e}+JB^e. \end{aligned} \tag{113}$$

Thus, if (113) and (111) hold, so does (112).

Choose $J\leq M$ and

$$M^2=C_eB^2J.$$

Then

$$M=\sqrt{C_e}B\sqrt{J}$$

and $M\geq J$ iff $J\ll_e B^2$. Now choose $B+t_1^{\frac{1}{2e}}$ and if $t_1\ll p^{1-\frac{1}{2e}}$ then (113) and (111) hold and

$$|Y|\ll_e t_1t_2^{-\frac{1}{4e}}.$$

This completes the proof of Proposition 20.

Acknowledgements: The authors thank Joseph Silverman for his careful reading of the paper and insightful remarks; their gratitude is also extended to Alex Lopez and the anonymous referees, whose comments considerably improved the paper. The authors were supported, in part, by the following NSF DMS awards: 1301619 (Bourgain), 1603715 (Gamburd), 1302952 (Sarnak).

References

[Aig13] Aigner, Martin, Markov's Theorem and 100 years of the Uniqueness Conjecture, Springer 2013.

[Bar91] A. Baragar, The Markoff equation and equations of Hurwitz, Thesis, Brown University, 1991

[BS02] F. Beukers and C. J. Smyth, Cyclotomic points on curves, Number theory for the millenium (Urbana, Illinois, 2000), I, A.K. Peters, (2002), 67-85.

[Bom07] E. Bombieri, Continued fractions and the Markoff tree, Expo. Math. 25 (2007), no 3, 187–213.

[Bou12] J. Bourgain, A modular Szemeredi-Trotter theorem for hyperbolas, C.R. Acad. Sci. Paris Ser 1, 350 (2012), 793–796.

[BG08d] Bourgain, Jean; Gamburd, Alex Expansion and random walks in $\mathrm{SL}_d(\mathbb{Z}/p^n\mathbb{Z})$. I. J. Eur. Math. Soc. (JEMS) 10 (2008), no. 4, 987-1011.

[BGS10] J. Bourgain, A. Gamburd and P. Sarnak, Affine linear sieve, expanders and sum product, Invent. Math. 179 (2010), 559–644.

[BGS16] J. Bourgain, A. Gamburd, P. Sarnak, Markoff triples and strong approximation, C. R. Math. Acad. Sci. Paris 354 (2016), no. 2, 131-135.

[BGS16a] J. Bourgain, A. Gamburd, P. Sarnak, Markoff surfaces and strong approximation: 1, `arXiv:1607.01530`

[BGS24] J. Bourgain, A. Gamburd, P. Sarnak, Strong approximation for varieties of Markoff type, in preparation.

[BG10b] E. Breuillard and A. Gamburd, Strong uniform expansion in $SL_2(\mathbb{F}_p)$, Geometric and Functional Analysis, 20 (5), 2010, 1201–1209.

[Cas49] Cassels, The Markoff chain, Ann. of Math. (2) 50 (1949) 676–685.

[Cas70] Cassels, J. W. S., Factorization of polynomials in several variables Cassels, J. W. S. Lecture Notes in Math., Vol. 118 Springer-Verlag, Berlin-New York, 1970, pp. 1–17.

[Cas78] J. W. S. Cassels, *Rational Quadratic Forms*, Academic Press, 1978.

[Cha13] Chang, Mei-Chu Elements of large order in prime finite fields. Bull. Aust. Math. Soc. 88 (2013) 169-176.

[CKSZ14] M-C. Chang, B. Kerr, I. Shparlinski and U. Zannier, Elements of large orders on varieties over prime finite fields, J. Theor. Nombres Bordeaux 26 (2014) 579-594.

[Che21] W. Chen Nonabelian level structures, Nielsen equivalence, and Markoff triples, `arXiv:2011.12940`; Annals of Mathematics, **199** (2024) 301-443.

[CZ06] Corvaja, Pietro; Zannier, Umberto On the greatest prime factor of Markov pairs. Rend. Sem. Mat. Univ. Padova 116 (2006), 253–260.

[CS13] P. Corvaja, U. Zannier, Greatest common divisors of $u-1, v-1$ in positive characteristic and rational points on curves over finite fields, J. Eur. Math. Soc. (JEMS) 15 (2013) 1927-1942.

[DKS13] C. D'Andrea, T. Krick and M. Sombra, Heights of varieties in multiprojective spaces and arithmetic Nullstellensatz, Annales Sci. de l'ENS, 46, (2013),549-627.

[deC22] Matthew de Courcy-Ireland, Non-planarity of Markoff graphs mod p, `arXiv:2105.12411v3`

[DL20] Matthew de Courcy-Ireland and Seungjae Lee (2020) Experiments with the Markoff Surface, Experimental Mathematics, `DOI: 10.1080/10586458.2019.1702123`

[DM00] B. Dubrovin, M. Mazzocco, Monodromy of certain Painleve-VI transcendents and reflection groups, Invent. Math. 141 (2000) 55-147.

[ES22] Erlandsson, Viveka; Souto, Juan, Mirzakhani's curve counting and geodesic currents, Progr. Math., 345 Birkhäuser/Springer, Cham, 2022

[Fri99] Fried, Michael, Variables separated polynomials, the genus 0 problem and moduli spaces, Number theory in progress, Vol. 1 (Zakopane-Kościelisko, 1997), 169–228, Walter de Gruyter, Berlin, 1999.

[Fro13] G. Frobenius, Über die Markoffschen Zahlen, Akad. Wiss. Berlin, (1913), 458–487.

[Gam23] Gamburd, Alexander, Arithmetic and dynamics on varieties of Markoff type. Proc. Int. Cong. Math. 2022, Vol. 3, pp. 1800–1836. EMS Press, Berlin, 2023.

[Gil77] R. Gilman, Finite quotients of the automorphism group of a free group, Canad. J. Math. 29 (1977), 541–551.

[Gol03] W. Goldman, The modular group action on real $SL(2)$-characters of a one-holed torus, Geom. and Top. Vol. 7 (2003), 443–486.

[GP03] Guralnick, Robert, Pak, Igor, On a question of B. H. Neumann, Proc. Amer. Math. Soc. 131 (2003), no. 7, 2021–2025.

[HK00] R. Heath-Brown and S. Konyagin, New bounds for Gauss sums derived from k-th powers and for Heilbronn's exponential sum, QUAR. J MATH (2000), 221–235.

[Hoo67] Hooley, Christopher On Artin's conjecture. J. Reine Angew. Math. 225 1967

[Hoo76] C. Hooley, *Applications of sieve methods to the theory of numbers*, Cambridge Tracts in Mathematics, Vol 70, (1976).

[HLW06] S. Hoory, N. Linial, A. Wigderson, Expander graphs and their applications, Bull. AMS, 2006

[Lau83] M. Laurent, Exponential diophantine equations, CR Acad Sc. 296 (1983), 945–947.

[LMT93] Lidl, R.; Mullen, G. L.; Turnwald, G., Dickson polynomials. Pitman Monogr. Surveys Pure Appl. Math., 65, 1993.

[LT14] O. Lisovyy and Y. Tykhyy, Algebraic solutions of the sixth Painleve equation, Journal of Geometry and Physics, **85** (2014) 124-163.

[Mar79] A. Markoff, Sur les formes quadratiques binaires indéfinies, Math. Ann. 15 (1879) 381–409.

[Mar80] A. Markoff, Sur les formes quadratiques binaires indéfinies, Math. Ann. 17 (1880) 379–399.

[MW13] D. McCullough and M. Wanderley, Nielsen equivalence of generating pairs in $SL(2, q)$, Glasgow Math. J. 55 (2013), 481–509.

[MP18] C. Meiri and D. Puder, The Markoff group of transformations in prime and composite moduli, Duke Math. J. Volume 167, Number 14 (2018), 2679-2720.

[Mir16] M. Mirzakhani, Counting mapping class group orbits on hyperbolic surfaces, `arXiv:1601.03342v1`.

[Per02] S. Perrine, La Théorie de Markoff et ses développements, Tessier Ashpool, 2002.

[Pak09] Pakovich, F., Prime and composite Laurent polynomials Bull. Sci. Math. 133 (2009), no. 7, 693–732

[SS13] A. Salehi Golsefidy and P. Sarnak, The affine sieve, JAMS (2013), no 4, 1085–1105.

[ST23] A. Salehi Golsefidy and N. Tamam, Closure of orbits of the pure mapping class group in the character variety, in preparation.

[Sar04] P. Sarnak, What is an expander?, Notices of the American Mathematical Society, **51**, 2004, 762-763.

[Sar10] P. Sarnak, Affine sieve lecture slides, 2010, `http://publications.ias.edu/sarnak/paper/508`

[SA94] P. Sarnak and S. Adams, Betti numbers of congruence groups, with an appendix by Z. Rudnick, Israel J. Math, 88, 1994, 31-72.

[Sch04] Schmidt, W. M *Equations over finite fields: an Elementary approach*, Kendrick Press 2004.

[Sil22] J. Silverman, A heuristic subexponential algorithm to find paths in Markoff graphs over finite fields, `arXiv:2211.08511`.

[Ste69] S.A. Stepanov, The number of points of a hyperelliptic curve over a prime field, MATH USSR-IZV 3:5 (1969), 1103–1114.

[Suz82] M. Suzuki, *Group Theory I*, Springer-Verlag, Berlin-Heidelberg-New York, 1982.

[Thu88] W.P. Thurston, On the geometry and dynamics of diffeomorphisms of surfaces, Bull. Amer. Math. Soc. (N.S.) 19(2) (1988) 417–431.

[Vol10] Voloch, Elements of high order on finite fields from elliptic curves, Bull. Aust. Math. Soc. 81 (2010), 425-429.

[VS12] Vyugin, I. V., Shkredov, I. D., On additive shifts of multiplicative subgroups, Sb. Math. 203(2012), 844-863.

[Wei41] A. Weil, On the Riemann Hypothesis in function fields, Proc. Nat. Acad. Sci. USA 27 (1941), 345–347.

[Wha20a] Whang, Junho Peter Global geometry on moduli of local systems for surfaces with boundary. Compos. Math. 156 (2020), no. 8, 1517–1559.

[Wha20b] Whang, Junho Peter Nonlinear descent on moduli of local systems. Israel J. Math. 240 (2020), no. 2, 935–1004.

[Zag82] D. Zagier, *On the number of Markoff numbers below a given bound*, Math of Comp. 39, 160 (1982), 709–723.

IAS

The Graduate Center, CUNY
Email address: `agamburd@gc.cuny.edu`

IAS and Princeton University
Email address: `sarnak@math.princeton.edu`