ARITHMETIC AND DYNAMICS ON VARIETIES OF MARKOFF TYPE

ALEXANDER GAMBURD

ABSTRACT

The Markoff equation $x^2 + y^2 + z^2 = 3xyz$, which arose in his spectacular thesis in 1879, is ubiquitous in a tremendous variety of contexts. After reviewing some of these, we discuss Hasse principle, asymptotics of integer points, and, in particular, recent progress towards establishing forms of strong approximation on varieties of Markoff type, as well as ensuing implications, diophantine and dynamical.

MATHEMATICS SUBJECT CLASSIFICATION 2020

Primary 14G12; Secondary 11N36, 11D45, 37P55

KEYWORDS

Markoff triples, strong approximation, nonlinear dynamics

INTERNATIONAL CONGRESS
OF MATHEMATICIANS
2022 JULY 6–14

Proc. Int. Cong. Math. 2022, Vol. 3, pp. 1800–1836
DOI 10.4171/ICM2022/191

Important though the general concepts and propositions may be with which the modern industrious passion for axiomatizing and generalizing has presented us, in algebra perhaps more than anywhere else, nevertheless I am convinced that the special problems in all their complexity constitute the stock and core of mathematics; and to master their difficulties requires on the whole the harder labor.

Hermann Weyl, The Classical Groups, 1939

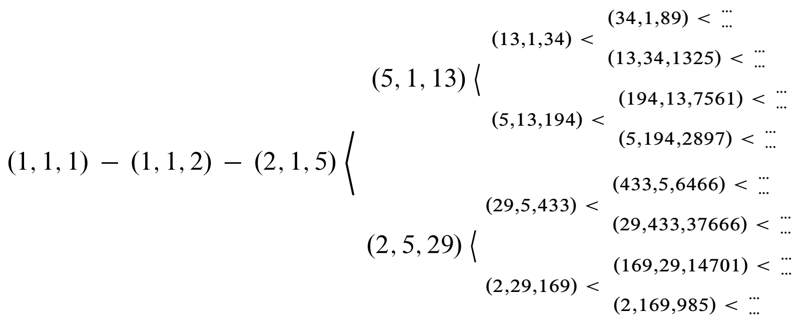
1. INTRODUCTION

1.1. Andrei Andreevich Markov is one of the towering peaks of the illustrious Saint Petersburg school of number theory, alongside with Chebyshev and Linnik. A singular characteristic of this school is a deep, often subterranean, interaction between arithmetic/combinatorics and probability/dynamics. While Markov is perhaps most widely known today for the chains named after him, it is in the context of his arguably deepest work on the minima of binary quadratic forms and badly approximable numbers[1] that the following equation, now bearing his name, was born:

$$x_1^2 + x_2^2 + x_3^2 = 3x_1x_2x_3, \tag{1.1}$$

describing a Markoff surface $X \subset \mathbb{A}^3$. *Markoff triples* $\mathcal{M}$ are the solutions of (1.1) with positive integral coordinates. *Markoff numbers* $\mathbb{M} \subset \mathbb{N}$ are obtained as coordinates of elements of $\mathcal{M}$. The *Markoff sequence* $\mathbb{M}^s$ is the set of largest coordinates of an $m \in \mathcal{M}$ counted with multiplicity; the uniqueness conjecture of Frobenius [62] asserts that $\mathbb{M} = \mathbb{M}^s$.

All elements of $\mathcal{M}$ are gotten from the root solution $r = (1, 1, 1)$ by repeated applications of an element in a set S, consisting of $\sigma \in \Sigma_3$, the permutations of the coordinates of (x_1, x_2, x_3), and 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 affine morphisms of $\mathbb{A}^3$ generated by S, 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, $\mathcal{M} = \Gamma \cdot r$, giving rise to the *Markoff tree* [8]:



[1] This work of Markoff and some of the subsequent appearances of his equation in a tremendous variety of different contexts are briefly discussed in Section 2.

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

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

The sequence $\mathbb{M}^s$ is sparse, as shown by Zagier [147]:

$$\sum_{\substack{m \in \mathbb{M}^s \\ m \leq T}} 1 \sim c(\log T)^2 \quad \text{as } T \to \infty \; (c > 0). \tag{1.2}$$

1.2. The origins of investigations which underlie "the stock and core" of this report date back to August of 2005 and involve a "special problem" pertaining to Markoff numbers; here is Peter Sarnak's recollection [126]: "For me the starting point of this investigation was in 2005 when Michel and Venkatesh asked me about the existence of poorly distributed closed geodesics on the modular surface. It was clear that Markov's constructions of his geodesics using his Markov equation provided what they wanted but they preferred quadratic forms with square free discriminant. This raised the question of sieving in this context of an orbit of a group of (nonlinear) morphisms of affine space. The kind of issues that one quickly faces in attempting to execute such a sieve are questions of the image of the orbit when reduced mod q and interestingly whether certain graphs associated with these orbits are expander families.[2] Gamburd in his thesis had established the expander property in some simpler but similar settings and he and I began a lengthy investigation into this sieving problem in the simpler setting when the group of affine morphisms acts linearly (or what we call now the affine linear sieve)."

The question posed by Michel and Venkatesh arose in the course of their joint work with Einsiedler and Lindenstrauss [58,59] on generalizations of Duke's theorem [57]; formulated in terms of Markoff numbers, it leads to the following:

Conjecture 1. *There are infinitely many square-free Markoff numbers.*

As detailed in [21], an application of sieve methods in the setting of affine orbits leads to and demands an affirmative answer to the question as to whether *Markoff graphs*, obtained as a modular reduction of the Markoff tree,[3] form a family of expanders. Numerical experiments by de Courcy-Ireland and Lee [55], as well as results detailed in Section 2.5, are compelling in favor of the following *superstrong approximation conjecture for Markoff graphs*:

Conjecture 2. *The family of Markoff graphs $X^*(p)$ forms a family of expanders.*

Before attacking this conjecture, asserting high connectivity of Markoff graphs, one has to confront the question of their connectivity, that is to say, the issue of the *strong approximation for Markoff graphs*:

[2] See [125] and [81] for definition and properties of expanders.

[3] Let p be a large prime and denote by $X^*(p) = X(p) \backslash (0,0,0)$ the solutions of (1.1) modulo p with the removal of $(0,0,0)$. The Markoff graphs are obtained by joining each x in $X^*(p)$ to $R_j(x)$, $j = 1,2,3$. They were considered first by Arthur Baragar in his thesis [3].

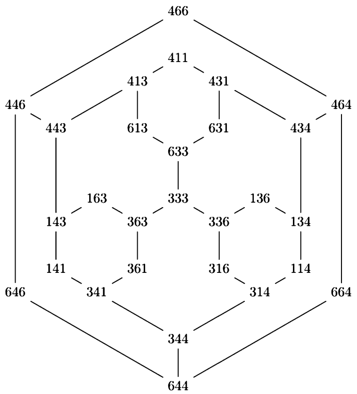


FIGURE 1
Markoff graph mod 7. In [54] it is proved that the Markoff graphs are not planar for primes greater than 7.

Conjecture 3. *The map $\pi_p : \mathcal{M} \to X^*(p)$ is onto, that is to say, Markoff graphs $X^*(p)$ are connected.*

While Conjectures 1 and 2 have withstood our protracted attack over the past 17 years, much progress has been made on parallel questions in the case of affine linear maps. We will return to the recent resolution of Conjecture 3, and resulting progress on diophantine properties of Markoff numbers in Section 1.5.

1.3. Before describing the general setting of Affine Linear Sieve, it is instructive to briefly examine an example which is in many ways parallel to the Markoff situation, namely integral Apollonian packings [63,127]. A theorem of Descartes asserts that $x_1, x_2, x_3, x_4 \in \mathbb{R}^4$ are the curvatures of four mutually tangent circles in the plane if

$$2(x_1^2 + x_2^2 + x_3^2 + x_4^2) = (x_1 + x_2 + x_3 + x_4)^2. \tag{1.3}$$

Given an initial configuration of 4 such circles, we fill in repeatedly the lune regions with the unique circle which is tangent to 3 sides (which is possible by a theorem of Apollonius). In this way we get a packing of the outside circle by circles giving an Apollonian packing. The interesting diophantine feature is that if the initial curvatures are integral then so are the curvatures of the entire packing.

The numbers in the circles in Figure 2 indicate their curvatures; note that by convention the outer circle has negative curvature. Viewing equation (1.3) as a quadratic equation in x_1, we see that the two solutions are related as $x_1 + x_1' = 2x_2 + 2x_3 + 2x_4$, the crucial point being that the Vieta involutions in this case are given by linear maps A_1, A_2, A_3, A_4 where $A_j(e_k) = -3e_k + 2(e_1 + e_2 + e_3 + e_4)$ if $k = j$ and $A_j(e_k) = e_k$ if $k \neq j$ (e_1, e_2, e_3, e_4 are the standard basis vectors). The configurations of 4 mutually tangent circles in the packing with initial configuration $a = (a_1, a_2, a_3, a_4)$ consist of points x in the orbit $\mathcal{O} = \Lambda \cdot a$ where $\Lambda = \langle A_1, A_2, A_3, A_4 \rangle$ is the Apollonian group. The elements A_j

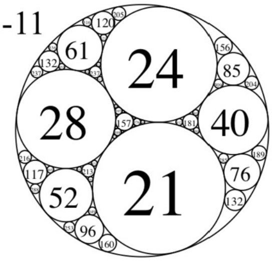


FIGURE 2
Integral Apollonian packing $(-11, 21, 24, 28)$.

preserve F given by

$$F(x_1, x_2, x_3, x_4) = 2(x_1^2 + x_2^2 + x_3^2 + x_4^2) - (x_1 + x_2 + x_3, +x_4)^2,$$

and hence $\Lambda \leq O_F(\mathbb{Z})$. The group Λ is Zariski dense in O_F, but it is thin in $O_F(\mathbb{Z})$. For example, $|\{\gamma \in O_F(\mathbb{Z}) : ||\gamma|| \leq T\}| \sim c_1 T^2$ as $T \to \infty$, while $|\{\gamma \in \Lambda : ||\gamma|| \leq T\}| \sim c_1 T^\delta$, where[4] $\delta = 1.3\ldots$ is the Hausdorff dimension of the limit set of Λ.

The general setting of Affine Linear Sieve, introduced in [20,21], is as follows. For $j = 1, 2, \ldots, k$, let A_j be invertible integer coefficient polynomial maps from $\mathbb{Z}^n$ to $\mathbb{Z}^n$ (here $n \geq 1$ and the inverses of A_j's are assumed to be of the same type). Let Λ be the group generated by $A_1, \ldots, A_k$ and let $\mathcal{O} = \Lambda b$ be the orbit of some $b \in \mathbb{Z}^n$ under Λ. Given a polynomial $f \in Q[x_1, ..., x_n]$ which is integral on $\mathcal{O}$, the aim is to show that there are many points $x \in \mathcal{O}$ at which $f(x)$ has few or even the least possible number of prime factors, in particular that such points are Zariski dense in the Zariski closure, $\mathrm{Zcl}(\mathcal{O})$ of $\mathcal{O}$. Let $\mathcal{O}(f, r)$ denote the set of $x \in \mathcal{O}$ for which $f(x)$ has at most r prime factors. As $r \to \infty$, the sets $\mathcal{O}(f, r)$ increase and potentially at some point become Zariski dense. Define the *saturation number* $r_0(\mathcal{O}, f)$ to be the least integer r such that $\mathrm{Zcl}(\mathcal{O}(f, r)) = \mathrm{Zcl}(\mathcal{O})$. It is by no means obvious that it is finite or even if one should expect it to be so, in general. If it is finite, we say that the pair $(\mathcal{O}, f)$ saturates. In the case of linear maps, the theory by now is quite advanced and the basic result pertaining to the finiteness of the saturation number in all cases where it is expected to hold, namely in the case of the Levi factor of $G = \mathrm{Zcl}(\Lambda)$ being semisimple,[5] has been established [123]. Both strong and superstrong approximation, particularly for thin

[4] This result can be deduced from the work of Lax and Phillips [93]. A beautiful overview of striking developments pertaining to dynamics on geometrically finite hyperbolic manifolds with applications to Apollonian circle packings (and beyond) is contained in Hee Oh's ICM report [114].

[5] On the other hand, as detailed in [21, 85, 123], when torus intervenes, the saturation most likely fails. Tori pose particularly difficult problems, in terms of sparsity of elements in an orbit, strong approximation and diophantine properties: see [104] for a discussion of Artin's Conjecture in the context of strong approximation.

groups such as the Apollonian group, are crucial ingredients in executing Brun combinatorial sieve in this setting.

1.4. The strong approximation for $\mathrm{SL}_n(\mathbb{Z})$, asserting that the reduction π_q modulo q is onto, is a consequence of the Chinese remainder theorem; its extension to arithmetic groups is far less elementary but well understood [118]. If S is a finite symmetric generating set of $\mathrm{SL}_n(\mathbb{Z})$, strong approximation is equivalent to the assertion that the Cayley graphs $\mathcal{G}(\mathrm{SL}_n(\mathbb{Z}/q\mathbb{Z}), \pi_q(S))$ are connected. The quantification of this statement, asserting that they are in fact highly-connected, that is to say, form a family of expanders, is what we mean by superstrong approximation. The proof of the expansion property for $\mathrm{SL}_2(\mathbb{Z})$ has its roots in Selberg's celebrated lower bound [131] of $\frac{3}{16}$ for the first eigenvalue of the Laplacian on the hyperbolic surfaces associated with congruence subgroups of $\mathrm{SL}_2(\mathbb{Z})$. The generalization of the expansion property to $G(\mathbb{Z})$ where G is a semisimple matrix group defined over $\mathbb{Q}$ is also known thanks to developments towards the general Ramanujan conjectures that have been established; this expansion property is also referred to as *property* τ for congruence subgroups [133].

Let Γ be a finitely generated subgroup of $\mathrm{GL}_n(\mathbb{Z})$ and let $G = \mathrm{Zcl}(\Gamma)$. The discussion of the previous paragraph applies if Γ is of finite index in $G(\mathbb{Z})$. However, if Γ is *thin*, that is to say, of infinite index in $G(\mathbb{Z})$, then $\mathrm{vol}(G(\mathbb{R})\backslash\Gamma) = \infty$ and the techniques used to prove both of these properties do not apply. It is remarkable that, under suitable natural hypothesis, strong approximation continues to hold in this thin context, as proved by Matthews, Vaserstein, and Weisfeiler in 1984 [105,143]. That the expansion property might continue to hold for thin groups was first suggested by Lubotzky and Weiss in 1993 [101]; for $\mathrm{SL}_2(\mathbb{Z})$, the issue is neatly encapsulated in the following 1–2–3 question of Lubotzky [99]. For a prime $p \geq 5$ and $i = 1, 2, 3$, let us define $S_p^i = \left\{\left(\begin{smallmatrix} 1 & i \\ 0 & 1 \end{smallmatrix}\right), \left(\begin{smallmatrix} 1 & 0 \\ i & 1 \end{smallmatrix}\right)\right\}$. Let $\mathcal{G}_p^i = \mathcal{G}(\mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z}), S_p^i)$, the Cayley graph of $\mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})$ with respect to S_p^i. By Selberg's theorem, $\mathcal{G}_p^1$ and $\mathcal{G}_p^2$ are families of expander graphs. However, the group $\langle\left(\begin{smallmatrix} 1 & 3 \\ 0 & 1 \end{smallmatrix}\right), \left(\begin{smallmatrix} 1 & 0 \\ 3 & 1 \end{smallmatrix}\right)\rangle$ has infinite index in $\mathrm{SL}_2(\mathbb{Z})$ and thus does not come under the purview of Selberg's theorem.

In my thesis [66], extending the work of Sarnak and Xue [129], [128] for cocompact arithmetic lattices, a generalization of Selberg's theorem for infinite index "congruence" subgroups of $\mathrm{SL}_2(\mathbb{Z})$ was proved; for such subgroups with a high enough Hausdorff dimension of the limit set, a spectral gap property was established. Following the groundbreaking work of Helfgott [77] (which builds crucially on sum–product estimate in $\mathbb{F}_p$ due to Bourgain, Katz, and Tao [27]), Bourgain and Gamburd [13] gave a complete answer to Lubotzky's question. The method introduced in [12,13] and developed in a series of papers [14–19] became known as "Bourgain–Gamburd expansion machine"; thanks to a number of major developments by many people [22,28,35,82,91,115,120,122,124], the general superstrong approximation for thin groups is now known. The state-of-the-art is summarized in *Thin groups and superstrong approximation* [36] which contains an expanded version of most of the invited lectures from the eponymous MSRI "Hot Topics" workshop, in the surveys by Breuillard [33] and Helfgott [78], and in the book by Tao *"Expansion in finite simple groups of Lie type"* [140].

1.5. We return to the progress on Conjecture 3 [23–26]. Our first result [25] asserts that there is a very large orbit.

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

$$|X^*(p) \backslash \mathcal{C}(p)| \leq p^{\varepsilon} \tag{1.4}$$

(note that $|X^*(p)| \sim p^2$), and any Γ orbit $\mathcal{D}(p)$ satisfies[6]

$$|\mathcal{D}(p)| \gg (\log p)^{\frac{1}{3}}. \tag{1.5}$$

The proof, discussed in section 3, establishes the strong approximation conjecture, unless $p^2 - 1$ is a very smooth number. In particular, the set of primes for which the strong approximation conjecture fails is very small.

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

Very recently, in a remarkable breakthrough, using geometric techniques involving Hurwitz stacks, degeneration, and some Galois theory, William Chen [45] proved the following result:

Theorem 3. *Every Γ orbit $\mathcal{D}(p)$ has size divisible by p.*

Combining Theorems 1 and 3 establishes Conjecture 3 for all sufficiently large primes; in combination with the following result established in [26], namely

Theorem 4. *Assume that $X^*(\mathbb{Z}/p\mathbb{Z})$ is connected. Then $X^*(\mathbb{Z}/p^k\mathbb{Z})$ is connected for all k.*

it yields

Theorem 5. *For all sufficiently large primes p, the group Γ acts minimally on $X^*(\mathbb{Z}_p)$.*

We remark that Theorem 5 is not true for $X^*(\mathbb{R})$; cf. section 4.1. While Conjecture 1 remains out of reach, the progress on strong approximation allows us to establish the following result on the diophantine[7] properties of Markoff numbers [25]:

Theorem 6. *Almost all Markoff numbers are composite, that is,*

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

It is worth contrasting this result with the state of knowledge regarding the sequence $H_n = 2^n + b$, which is just a little more sparse than the sequence of Markoff numbers, for which, by Zagier's result (1.2), we have $M_n \sim A^{\sqrt{n}}$. Even assuming the generalized Riemann Hypothesis, which allowed Hooley [79] to give a conditional proof of Artin's conjecture (cf. footnote 5), was not sufficient to establish that almost all members of the sequence H_n are composite: the conditional proof in [80] necessitated postulating additional "Hypothesis A."

[6] The exponent $\frac{1}{3}$ in (1.5) has been improved to $\frac{7}{9}$ in [87].

[7] We remark that in [52] Corvaja and Zannier showed that the greatest prime factor of xy for a Markoff triple (x, y, z) tends to infinity.

1.6. The methods of proof of Theorems 1, 2, 4 discussed in Section 3 are robust enough to enable handling their extension to more general Markoff-type cubic surfaces, namely

$$X_k : \Phi(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2 - x_1 x_2 x_3 = k, \tag{1.6}$$

where the real dynamics was studied by Goldman [73], as discussed in Section 4.1; the family of surfaces $S_{A,B,C,D} \subset \mathbb{C}^3$ given by

$$x_1^2 + x_2^2 + x_3^2 + x_1 x_2 x_3 = A x_1 + B x_2 + C x_3 + D, \tag{1.7}$$

where the real dynamics was studied by Cantat [38], as discussed in Section 4.2; those in [60] and even the general such nondegenerate cubic surface

$$Y = Y(\alpha, \beta, \gamma, \delta) : \sum_{i,j=1}^{3} \alpha_{ij} x_i x_j + \sum_{j=1}^{3} \beta_j x_j + \gamma = \delta x_1 x_2 x_3, \tag{1.8}$$

with $\alpha_{ij}, \beta_j, \gamma, \delta$ being integers.

The group Γ_Y is again generated by the corresponding Vieta involutions R_1, R_2, R_3. For such a Y and action Γ_Y, one must first show that there are only finitely many finite orbits in $Y(\bar{\mathbb{Q}})$, and that these may be determined effectively. The analogue of Conjecture 1 for Y is that for p large, Γ_Y has one big orbit on $Y(\mathbb{Z}/p\mathbb{Z})$ and that the remaining orbits, if there are any, correspond to one of the finite $\bar{\mathbb{Q}}$ orbits determined above.

The determination of the finite orbits of Γ on $X_k(\bar{\mathbb{Q}})$ and on $S_{A,B,C,D}(\bar{\mathbb{Q}})$ has been carried out in [56] and [96], respectively. Remarkably for these, the Γ action on affine 3-space corresponds to the (nonlinear) monodromy group for Painlevé VI equations on their parameter spaces. In this way the finite orbits in question turn out to correspond bijectively to those Painlevé VI's which are algebraic functions of their independent variable.

In this setting of the more general surfaces Y in (1.8), strong approximation for $Y(\mathbb{Z}_S)$, where S is the set of primes dividing $\alpha_{11}, \alpha_{22}, \alpha_{33}$ (so that Γ_Y preserves the S-integers $\mathbb{Z}_S$), will follow from Conjecture 1 for Y (and the results we can prove towards it, as in Theorem 2) once we have a point of infinite order in $Y(\mathbb{Z}_S)$. If there is no such point, we can increase S or replace $\mathbb{Z}$ by $\mathcal{O}_K$, the ring of integers in a number field $K/\mathbb{Q}$, to produce such a point and with it strong approximation for $Y\big((\mathcal{O}_K)_S\big)$.

Vojta's conjectures and the results proven towards them [51,141] assert that cubic and higher-degree affine surfaces typically have few S-integral points. In the rare cases where these points are Zariski dense, such as tori (e.g., $N(x_1, x_2, x_3) = k$ where N is the norm form of a cubic extension of $\mathbb{Q}$), strong approximation fails. So these Markoff surfaces appear to be rather special affine cubic surfaces not only having a Zariski dense set of integral points, but also a robust strong approximation.

1.7. Zagier's result (1.2) can be viewed as a statement about asymptotic growth of integral points on the Markoff variety, $|X(\mathbb{Z}) \cap B(T)| \sim (\log T)^2$. In Section 5 we discuss the work in [68], establishing an asymptotic formula for the number of integer solutions to the Markoff–Hurwitz equation

$$x_1^2 + x_2^2 + \cdots + x_n^2 = a x_1 x_2 \cdots x_n + k, \tag{1.9}$$

giving an interpretation of the exponent of growth, which for $n > 3$ is not integral, in terms of the unique parameter for which there exists a certain conformal measure on a projective space.

1.8. The issue of the existence of a *single* integral solution to (1.9) for general a and k, even for $n = 3$, is quite subtle; see [112,130]. In the work of Ghosh and Sarnak [71], the Hasse principle is established to hold for Markoff-type cubic surfaces $X(k)$ given by (1.6) for almost all k, but it also fails to hold for infinitely many k; this work is discussed in Section 6.

1.9. Regrettably, the space/time constraints prevented us from covering cognate results pertaining to arithmetic and dynamics on K3 surfaces; see [37,65,106,108,109,135] and references therein. The Markoff equation over quadratic imaginary fields is studied in [134]. Potential cryptographic applications of Markoff graphs are discussed in [64].

1.10. To conclude this introduction, let us note that X_k is the relative character variety of representations of the fundamental group of a surface of genus 1 with one puncture to SL_2. The action of the mapping class group is that of Γ. More generally, the (affine) relative character variety V_k of representation of $\pi_1(\Sigma_{g,n})$, a surface of genus g with n punctures, into SL_2 is defined over $\mathbb{Z}$, and one can study the diophantine properties of $V_k(\mathbb{Z})$. In the work of Whang [144–146], it was shown that V_k has a projective compactification relative to which V_k is "log-Calabi–Yau." According to the conjectures of Vojta, this places V_k as being in the same threshold setting as affine cubic surfaces. Moreover, $V_k(\mathbb{Z})$ has a full descent in that the mapping class group acts via nonlinear morphisms on $V_k(\mathbb{Z})$ with finitely many orbits. These and more general character varieties connected with higher Teichmüller theory offer a rich family of threshold affine varieties for which one can approach the study of integral points.

2. THE UNREASONABLE(?) UBIQUITY OF MARKOFF EQUATION

Markoff equation and numbers appear in a surprising variety of contexts: see, for example, [1] (subtitled *Mathematical Journey from Irrational Numbers to Perfect Matchings*) and the references therein.

2.1. The Markoff chain. Equation (1.1) was discovered by Markoff in 1879 in his work on badly approximable numbers. As the sentiment[8] expressed by Frobenius [62] in 1913 seems to remain true today, we briefly review the context and statement of Markoff's theorem.

Let α be an irrational number. A celebrated theorem of Hurwitz asserts that α admits infinitely many rational approximations p/q such that $|\alpha - \frac{p}{q}| < \frac{1}{\sqrt{5}q^2}$, and, moreover, that if α is $\mathrm{GL}_2(\mathbb{Z})$-equivalent to the Golden Ratio $\theta_1 = (1+\sqrt{5})/2$, in the sense that $\alpha = \frac{a\theta_1+b}{c\theta_1+d}$

[8] "*Trotz der außerordentlich merkwürdigen und wichtigen Resultate scheinen diese schwiergen Untersuchungen wenig bekannt zu sein*" [In spite of the extraordinarily noteworthy and important results these difficult investigations seem to be little known]

for some integers a, b, c, d with $ad - bc = \pm 1$, the above result is sharp and the constant $\frac{1}{\sqrt{5}}$ cannot be replaced by any smaller.

Suppose next that α is not $\mathrm{GL}_2(\mathbb{Z})$-equivalent to θ_1. Then the result of Markoff's doctoral advisors, Korkine and Zolotareff, [88] asserts that α admits infinitely many rational approximations p/q such that $|\alpha - \frac{p}{q}| < \frac{1}{\sqrt{8}q^2}$, and, moreover, that the constant $\frac{1}{\sqrt{8}}$ is sharp if and only if α is $\mathrm{GL}_2(\mathbb{Z})$-equivalent $\theta_2 = 1 + \sqrt{2}$.

The general result found by Markoff in his *Habilitation* and published in 1879 and 1880 in *Mathematische Annalen* is as follows.

Markoff's Theorem. *Let $\mathbb{M} = \{1, 2, 5, 13, 29, 34, 89, 169, 194, \dots\}$ be the sequence of Markoff numbers. There is a sequence of associated quadratic irrationals $\theta_i \in \mathbb{Q}(\sqrt{\Delta_i})$, where $\Delta_i = 9m_i^2 - 4$ and m_i is the ith element of the sequence, with the following property. Let α be a real irrational, not $GL_2(\mathbb{Z})$-equivalent to any of the numbers θ_i whenever $m_i < m_j$. Then α admits infinitely many rational approximations p/q with $\left|\alpha - \frac{p}{q}\right| < \frac{m_j}{\sqrt{\Delta_j}\, q^2}$; the constant $m_j/\sqrt{\Delta_j}$ is sharp if and only if α is $GL_2(\mathbb{Z})$-equivalent to θ_h, for some h such that $m_h = m_j$.*

2.2. Continued fractions and binary quadratic forms.

The first paper by Markoff [102] used the theory of continued fractions, while the second memoir [103] was based on the theory of binary indefinite quadratic forms, with the final result stated as a theorem on minima of binary indefinite quadratic forms.

The alternative approach based on indefinite binary quadratic forms was the subject of an important memoir by Frobenius [62] and complete details were finally provided by Remak [121] and much simplified by Cassels [39, 40].

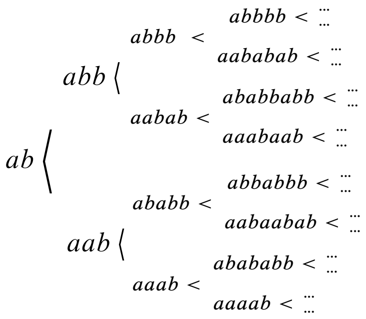
2.3. The geometry of Markoff numbers.

A third way of looking at the problem, via hyperbolic geometry, was introduced by Gorshkov [74] in his thesis of 1953, but published only in 1977. The connection with hyperbolic geometry was rediscovered, in a somewhat different way, by Cohn [46]. The paper by Caroline Series [132] contains a beautiful exposition of the problem in this context.

2.4. Cohn tree and Nielsen transformations.

Cohn is also credited for the interpretation of the problem [47] in the context of primitive words in F_2, the free group on two generators. Its automorphism group $\Phi_2 = \mathrm{Aut}(F_2)$ is generated by the following *Nielsen transformations:* $(a, b)^P = (b, a)$, $(a, b)^\sigma = (a^{-1}, b)$, $(a, b)^U = (a^{-1}, ab)$. Let $V = \sigma U$. Then $(a, b)^V = (a, ab)$.

The *Cohn tree* is a binary tree with root ab, branching to the top with U and to the bottom with V,



Markoff numbers are obtained from the Cohn tree by taking a third of the trace of the matrix obtained by substituting the matrices $A = \left(\begin{smallmatrix} 5 & 2 \\ 2 & 1 \end{smallmatrix}\right)$ and $B = \left(\begin{smallmatrix} 2 & 1 \\ 1 & 1 \end{smallmatrix}\right)$ in place of a, b and performing the matrix multiplication.

2.5. Nielsen systems and product replacement graphs. Conjecture 3 is a special case of *Conjecture Q* made by McCullough and Wanderley [107] in the context of Nielsen systems and product replacement graphs.

Given a group G, the *product replacement graph* $\Gamma_k(G)$ introduced in [42] in connection with computing in finite groups is defined as follows. The vertices of $\Gamma_k(G)$ consist of all k-tuples of generators $(g_1, \dots, g_k)$ of the group G. For every (i, j), $1 \le i$, $j \le k$, $i \ne j$, there is an edge corresponding to transformations $L^{\pm}_{i,j}$ and $R^{\pm}_{i,j}$, where $R^{\pm}_{i,j} : (g_1, \dots, g_i, \dots, g_k) \to (g_1, \dots, g_i \cdot g_j^{\pm 1}, \dots, g_k)$ and $L^{\pm}_{i,j}$ defined similarly. The graphs $\Gamma_k(G)$ are regular, of degree $4\,k\,(k-1)$, possibly with loops and multiple edges. The connectivity of $\Gamma_k(G)$ has been the subject of intensive recent investigations; for $G = \mathrm{SL}_2(p)$ and $k \ge 3$, it was established by Gilman in [72].

In the case of the free group F_k, the moves $L^{\pm}_{i,j}$ and $R^{\pm}_{i,j}$ defined above correspond to Nielsen moves on $\Gamma_k(F_k)$. For every group G, the set $\Gamma_k(G)$ can be identified with $E = \mathrm{Epi}(F_k, G)$, the set of epimorphisms from F_k onto G, and the group $A = \mathrm{Aut}(F_k)$ acts on E in the following way: if $\alpha \in A$ and $\varphi \in E$, $\alpha(\varphi) = \varphi \cdot \alpha^{-1}$. A long-standing problem is whether $\mathrm{Aut}(F_k)$ has property (T) for $k \ge 4$; in [100] Lubotzky and Pak observed that a positive answer to this problem implies the expansion of $\Gamma_k(G)$ for all G and proved that $\Gamma_k(G)$ are expanders when G is nilpotent of class l and both k and l are fixed. Property (T) for $\mathrm{Aut}(F_k)$ for $k \ge 5$ was recently established in [84].[9] Note that $\mathrm{Aut}(F_2)$ and $\mathrm{Aut}(F_3)$ do not satisfy property (T), while the problem is still open for $k = 4$.

In a joint work with Pak [69], we established a connection between the expansion coefficient of the product replacement graph $\Gamma_k(G)$ and the minimal expansion coefficient of a Cayley graph of G with k generators, and, in particular, proved that for $k > 3$ the product

[9] The proof stems from the groundbreaking observation by Ozawa [116] that G satisfies Kazhdan's property (T) if there exist $\lambda > 0$ and finitely many elements ξ_i of $\mathbb{R}[G]$ such that $\Delta^2 - \lambda\Delta = \sum_i \xi_i^* \xi_i$ where Δ is the Laplacian of the finite symmetric generating set of G.

replacement graphs $\Gamma_k(\mathrm{SL}(2,p))$ form an expander family under assumption of strong uniform expansion of $\mathrm{SL}(2,p)$ on k generators. In a joint work with Breuillard [34], combining the "expansion machine" [13] with the uniform Tits Alternative[10] established by Breuillard [32], we proved that Cayley graphs of $\mathrm{SL}(2,p)$ are strongly uniformly expanding for infinitely many primes of density one. Consequently, the following form of *nonlinear superstrong approximation* is obtained:

Theorem 7. *Let $k > 3$. The family of product replacement graphs $\{\Gamma_k(\mathrm{SL}(2,p_n))\}_n$ forms a family of expanders for infinitely many primes p_n of density one.*

As detailed in [107], the situation is different for the product replacement graph of $\mathrm{SL}(2,\mathbb{F}_p)$ on 2 generators, due to Fricke identity for 2×2 matrices M and N:

$$\operatorname{tr}(M)^2+\operatorname{tr}(N)^2+\operatorname{tr}(MN)^2=\operatorname{tr}(M)\operatorname{tr}(N)\operatorname{tr}(MN)+\operatorname{tr}([M,N])+2. \tag{2.1}$$

Letting $x_1=\operatorname{tr}(M)$, $x_2=\operatorname{tr}(N)$, $x_3=\operatorname{tr}(MN)$, the Q conjecture[11] in [107] amounts to the assertion of the strong approximation for the surfaces

$$X_k : \Phi(x_1,x_2,x_3)=k, \tag{2.2}$$

$$\Phi(x_1,x_2,x_3)=x_1^2+x_2^2+x_3^2-x_1x_2x_3, \tag{2.3}$$

and $k=\operatorname{tr}([M,N])+2$, with Markoff surface[12] being the special case corresponding to $\operatorname{tr}([M,N])=-2$.

3. STRONG APPROXIMATION

We give a brief overview of the methods and tools used in the proof of Theorems 1 and 2 and some comments about their extensions to the setting of more general surfaces of Markoff type. 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 [73] where it is shown that the action of Γ on the compact real components of the relative character variety of the mapping class group of the once punctured torus is ergodic. As in [73] our proof makes use of the rotations $\tau_{ij}\circ R_j$, $i\neq j$, where τ_{ij} permutes x_i and x_j. These preserve the conic sections gotten by intersecting $X^*(p)$ with the planes $y_k=x_k$ (k different from i and j). If $\tau_{ij}\circ R_j$ has order t_1 (here $t_1|p(p-1)(p+1)$), then x and these t_1 points of the conic section are connected (i.e., are in the same Γ orbit). If t_1 is maximal (i.e., is p, $p-1$, or $p+1$), then this entire conic section is connected and such conic sections in different planes which intersect are also connected. This leads to a large component which we denote by $\mathcal{C}(p)$.

[10] This states that if the subgroup of $\mathrm{GL}_d(K)$ (where K is an algebraic number field) generated by F is not virtually solvable, then there is some $N\in\mathbb{N}$, depending only on d, such that $(F\cup F^{-1}\cup\{1\})^N$ contains two elements that generate a nonabelian free group.

[11] See the paper of Will Chen [45] for the discussion of the relation between this conjecture and the connectivity properties of the moduli spaces of elliptic curves with $G=\mathrm{SL}(2,p)$ structures.

[12] Note that the congruence $x^2+y^2+z^2\equiv xyz \pmod 3$ has no nontrivial solutions.

If our starting rotation has order t_1 which is not maximal, then the idea is to ensure that among the t_1 points to which it is connected, at least one has a corresponding rotation of order $t_2 > t_1$, and then to repeat. To ensure that one can progress in this way, a critical equation over $\mathbb{F}_p$ intervenes:

$$\begin{cases} x + \dfrac{b}{x} = y + \dfrac{1}{y}, \quad b \neq 1, \\ x \in H_1, y \in H_2 \text{ with } H_1, H_2 \text{ subgroups of } \mathbb{F}_p^* \text{ (or } \mathbb{F}_{p^2}^*). \end{cases} \tag{3.1}$$

If $t_1 = |H_1| \geq p^{1/2+\delta}$ (with δ small and fixed), one can apply the proven Riemann Hypothesis for curves over finite fields [142] to count the number of solutions to (3.1). Together with a simple inclusion/exclusion argument, this shows that one of the t_1 points connected to our starting x has a corresponding maximal rotation and hence x is connected to $\mathcal{C}(p)$.

If $|H_1| \leq p^{1/2+\delta}$ then RH for these curves is of little use (their genus is too large), and we have to proceed using other methods. We assume that $|H_1| \geq |H_2|$ so that the trivial upper bound for the number of solutions to (3.1) is $2|H_2|$. What we need is a power saving in this upper bound in the case that $|H_2|$ is close to $|H_1|$, that is, a bound of the form $C_\tau |H_1|^\tau$, with $\tau < 1$, $C_\tau < \infty$ (both fixed).

In the prime modulus case, there are several ways to proceed. The first and second methods are related to "elementary" proofs of the Riemann Hypothesis for curves. One can use auxiliary polynomials as in Stepanov's proof [137] of the Riemann Hypothesis for curves to give the desired power saving with an explicit τ (cf. [76] which deals with $x + y = 1$ and $|H_1| = |H_2|$ in (3.1)). The second method, giving the best upper bound, namely $20 \max\{(|H_1|.|H_2|)^{1/3}, \frac{|H_1|.|H_2|}{p}\}$, is due to Corvaja and Zannier [53]. It uses their method for estimating the greatest common divisor of $u - 1$ and $v - 1$ in terms of the degrees of u and v and their supports, as well as (hyper) Wronskians.

The third method is based on Szemerédi–Trotter theorem for modular hyperbolas [11], whose proof uses crucially expansion and L^2-flattening lemma in $\mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})$ [16].

Theorem 8. *Let* $\Phi : \mathbb{F}_p \to \mathrm{Mat}_2(\mathbb{F}_p)$ *be such that* $\det \Phi$ *does not vanish identically and* $\mathrm{Im}\Phi \cap PGL_2(\mathbb{F}_p)$ *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*

$$1 \ll |A| < p^{1-\varepsilon}, \tag{3.2}$$

$$\log |A| < r \log |L|, \tag{3.3}$$

then

$$|\{(x, y, t) \in A \times A \times L; y = \tau_{\Phi(t)}(x)\}| < |A|^{1-\delta} |L|, \tag{3.4}$$

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

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 8 for $\mathbb{Z}/p^n\mathbb{Z}$, which follows from expansion in $\mathrm{SL}_2(\mathbb{Z}/p^n\mathbb{Z})$, established[13] in [16], plays crucial role in the proof of Theorem 4 in [26].

The above leads to a proof of part 1 of Theorem 1. To continue, one needs to deal with t_1 which is very small (here $|H_1| = t_1$ which divides $p^2 - 1$).

To handle these, we lift to characteristic zero and examine the finite orbits of Γ in $X(\bar{\mathbb{Q}})$. In fact, by the Chebotarev Density Theorem, a necessary condition for Conjecture 3 to hold is that there are no such orbits other than $\{0\}$. Again using the rotations in the conic sections by planes, one finds that any such finite orbit must be among the solutions with t_j's roots of unity to

$$(t_1 + t_1^{-1})^2 + (t_2 + t_2^{-1})^2 + (t_3 + t_3^{-1})^2 = (t_1 + t_1^{-1})(t_2 + t_2^{-1})(t_3 + t_3^{-1}). \tag{3.5}$$

For this particular surface X, one can show using the inequality between the geometric and arithmetic means, that (3.5) has no nontrivial solutions for complex numbers with $|t_j| = 1$. For the more general surfaces X_k, $S_{A,B,C,D}$, and those in (1.8), there is a variety of solutions with $|t_j| = 1$. However, Lang's $\mathbb{G}_m$ Conjecture which is established effectively (see [2, 92]) yields that there are only finitely many solutions to these equations in roots of unity. This allows for an explicit determination of the finite orbits of Γ_Y in $Y(\bar{\mathbb{Q}})$ (as noted earlier for the cubic surfaces $S_{A,B,C,D}$, the long list of these orbits [96] correspond to the algebraic Painlevé VI's). This $\bar{\mathbb{Q}}$ analysis leads to part 2 of Theorem 1 and, combined with the discussion above, it yields a proof of Conjecture 3, at least if $p^2 - 1$ is not very smooth. To prove Theorem 2, we need to show that there are very few primes for which the above arguments fail. This is done by extending the arguments and results in [43] and [44] 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 $\mathbb{F}_p^*$).

The proof of Theorem 6 in the stronger form that all Markoff numbers are highly composite, that is, for every $\nu \geq 1$, as $T \to \infty$,

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

makes use of counting points on $X^*(\mathbb{Z})$ of height at most T and, in particular, Mirzakhani's orbit equidistribution [111], as well as the transitivity properties of Γ on $X^*(q)$ for q a product of suitable primes p. The latter are provided by the results of Meiri and Puder [110]. 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 $p_1 < p_2 < \cdots < p_k$.

[13] The proof of this expansion result, in turn, builds crucially on Bourgain's sum-product theorem in $\mathbb{Z}/p^n\mathbb{Z}$ in [10], which is intimately realted to his discretized sum-product theorem [9]; the origins, nature and impact of the latter are discussed in [67].

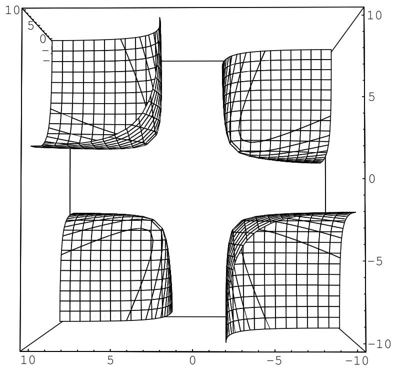


FIGURE 3
Level set $\kappa = -2.1$.

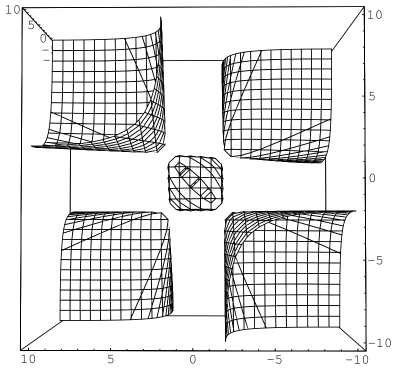


FIGURE 4
Level set $\kappa = -1.9$.

4. REAL DYNAMICS ON SURFACES OF MARKOFF TYPE

4.1. In this section we discuss the work of Goldman [73] pertaining to modular group action on real SL(2)-characters of a one-holed torus. The fundamental group π of the one-holed torus is the free group of rank two. The mapping class group of the 1-holed torus is isomorphic to the outer automorphism group $\mathrm{Out}(\pi) \cong \mathrm{GL}(2, \mathbb{Z})$ of π and acts on the moduli space of equivalence classes of $\mathrm{SL}(2, \mathbb{C})$-representations of π; this moduli space identifies naturally with affine 3-space $\mathbb{C}^3$, using the traces of two generators of π and of their product as coordinates. In these coordinates, the trace of the commutator of the two generators (representing the boundary curve of the torus) is given by $\kappa(x, y, z) = x^2 + y^2 + z^2 - xyz - 2$, which is preserved under the action of $\mathrm{Out}(\pi)$, and the action of $\mathrm{Out}(\pi)$ on $\mathbb{C}^3$ is commensurable with the action of the group Γ of polynomial automorphisms of $\mathbb{C}^3$ which preserve κ. Figures 3–8 show level sets of κ.

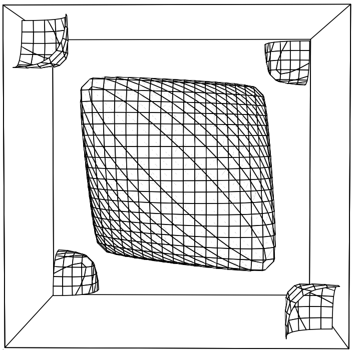


FIGURE 5
Level set $\kappa = 1.9$.

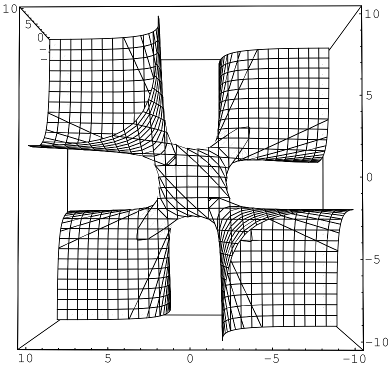


FIGURE 6
Level set $\kappa = 2.1$.

In [73] Goldman studied the dynamics of the Γ-action on the set of real points of this moduli space, and more precisely on the level sets $\kappa^{-1}(t) \cap \mathbb{R}^3$, for $t \in \mathbb{R}$. The action of Γ preserves a Poisson structure defining a Γ-invariant area form on each level set. It is shown that for $t < 2$ the Γ-action is properly discontinuous on the four contractible components of each level set and ergodic on the compact component (which is empty if $t < -2$); the contractible components correspond to Teichmüller spaces of complete hyperbolic structures on a one-holed torus if $t \leq -2$, and of a torus with a single cone point singularity if $-2 < t < 2$. For $t = 2$, the level set consists of characters of reducible representations and comprises two ergodic components, for $2 < t \leq 18$ the action of Γ on a level set is ergodic, and for $t > 18$ the moduli space contains characters of discrete representations uniformizing a three-holed sphere and the action is ergodic on the complement.

4.2. The main objective of [38] is the dynamical description of elements of the mapping class group of the four-punctured sphere acting on two-dimensional slices of its

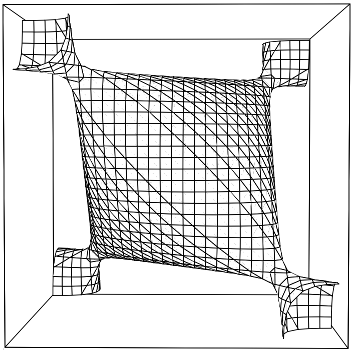


FIGURE 7
Level set $\kappa = 2.1$.

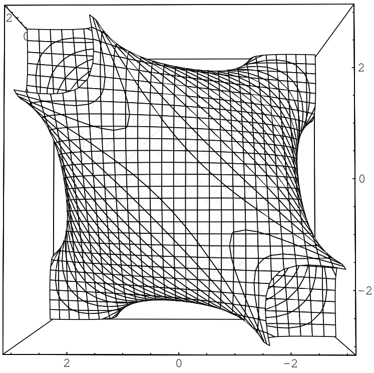


FIGURE 8
Level set $\kappa = 4$.

character variety. It also contains three striking applications of this analysis to the dynamics of the mapping class group on the character variety, to the spectrum of certain discrete Schrödinger equations, and to Painlevé sixth equation. Cantat considers the space of representations of the free group given by the presentation $F_3 = \langle \alpha, \beta, \gamma, \delta | \alpha\beta\gamma\delta = 1 \rangle$ into $\mathrm{SL}(2, \mathbb{C})$ modulo conjugacy. By fixing the trace of the images of the four generators, one obtains a space that is naturally parameterized by a cubic surface $S_{A,B,C,D} \subset \mathbb{C}^3$ given by $x^2 + y^2 + z^2 + xyz = Ax + By + Cz + D$ for some parameters $A, B, C, D \in \mathbb{C}$. This surface admits three natural involutions s_x, s_y, s_z which fix two out of the three coordinates and transform the last to the other root of the quadratic. These involutions generate a group Γ of affine automorphisms. Automorphisms of F_3 act by composition on the space of representations by preserving the trace, and the group of outer automorphisms of F_3 acts on $S_{A,B,C,D}$ in such a way that its image contains Γ as a finite index subgroup.

An element $f \in \Gamma$ is called hyperbolic if it corresponds to a pseudo-Anosov automorphism in the mapping class group, or, equivalently, if it is not conjugated to the product of

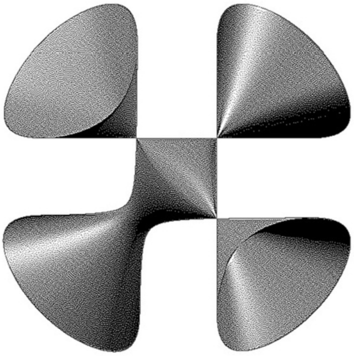


FIGURE 9

$S_{(-0.2,-0.2,-0.2,4.39)}$.

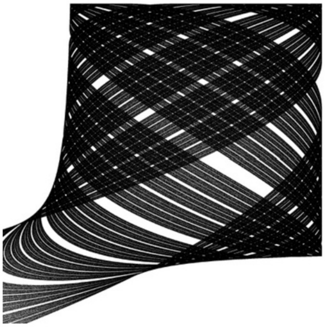


FIGURE 10

Projection of the stable manifold.

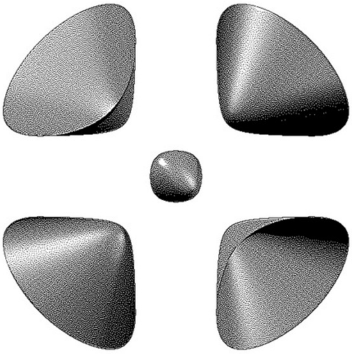


FIGURE 11

$S_{(0,0,0,3)}$.

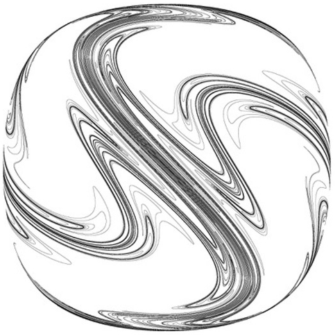


FIGURE 12

Projection of the intersection of the stable manifold with the upper part of the surface.

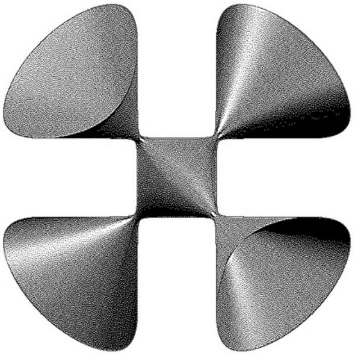


FIGURE 13

$S_{(0,0,0,4.1)}$.

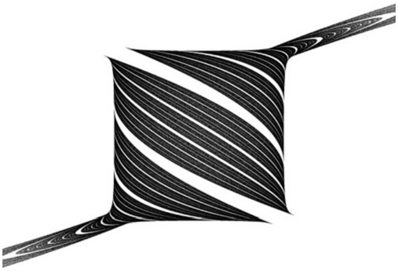


FIGURE 14

Projection of the intersection of the stable manifold with the upper part of the surface.

one or two involutions as above. Take $f \in \Gamma$ of hyperbolic type, and compactify $S_{A,B,C,D}$ by taking its closure in $\mathbb{P}^3$. The divisor at infinity is a cycle of three rational curves. By conjugating f in a suitable way, one can make it algebraically stable in the sense of Fornaess and Sibony [61], so that it contracts all curves at infinity to a single superattracting fixed point. One can then prove that at that point the map is locally conjugated to a monomial map whose spectral radius $\lambda(f)$ is greater than 1. This enables one to define the Green functions $G_{\pm} f = \lim_n \lambda(f)^{-n} \log^+ |f^{\pm n}|$, and show that they are plurisubharmonic, continuous, and possess natural invariance properties. It follows that the positive measure $\mu_f = dd^c G^+ f \wedge dd^c G^- f$ is well defined and f-invariant. Moreover, μ_f turns out to be mixing and the unique measure of maximal entropy equal to $\log \lambda(f)$. All these properties are reminiscent of the dynamics of Hénon mappings in the complex plane, and are proved analogously. Next, assume all coefficients A, B, C, D are real, and suppose the real part $S_{A,B,C,D}(\mathbb{R})$ is connected (in which case it is homeomorphic to the sphere minus four points). The main theorem of the paper states that the support of the measure μ_f is then included in $S_{A,B,C,D}(\mathbb{R})$ and that the induced map on $S_{A,B,C,D}(\mathbb{R})$ is uniformly hyperbolic on its nonwandering set. The proof of this striking theorem uses deep results by Bedford and Smillie [6] on the characterization of nonhyperbolic real Hénon maps having the same entropy as their complexification and relies on a delicate geometrical analysis of the possibilities for the intersection of stable and unstable manifolds in $S_{A,B,C,D}(\mathbb{R})$.

5. AN ASYMPTOTIC FORMULA FOR INTEGER POINTS ON MARKOFF–HURWITZ VARIETIES

For integer parameters $n \geq 3$, $a \geq 1$, and $k \in \mathbb{Z}$, consider the Diophantine equation

$$x_1^2 + x_2^2 + \cdots + x_n^2 = a x_1 x_2 \cdots x_n + k. \tag{5.1}$$

We call this the generalized[14] *Markoff–Hurwitz equation.* In this section we count solutions to (5.1) in integers, which we call *Markoff–Hurwitz tuples*. More precisely, let V be the affine subvariety of $\mathbb{C}^n$ cut out by (5.1). In a joint work with Magee and Ronan [68], we investigated the asymptotic size of the set $V(\mathbb{Z}) \cap B(R)$ where $B(R)$ is the ball of radius R in the ℓ^∞-norm on $\mathbb{R}^n \subset \mathbb{C}^n$. Perhaps somewhat surprisingly, the asymptotic growth for $n \geq 4$ is not of the order $(\log R)^{n-1}$, as was first noticed by Baragar [4], who subsequently in [5] proved that there is a number $\beta = \beta(n)$ such that when $k = 0$, if $V(\mathbb{Z}) - \{(0, 0, \dots, 0)\}$ is nonempty then

$$|V(\mathbb{Z}) \cap B(R)| = (\log R)^{\beta + o(1)} \tag{5.2}$$

as $R \to \infty$.

[14] Hurwitz [83] considered the case $k = 0$.

In [5] the following bounds for the exponents $\beta(n)$ were also obtained:

$$\begin{aligned} &\beta(3) = 2, \\ &\beta(4) \in (2.430, 2.477), \\ &\beta(5) \in (2.730, 2.798), \\ &\beta(6) \in (2.963, 3.048), \end{aligned} \tag{5.3}$$

and, in general,

$$\frac{\log(n-1)}{\log 2} < \beta(n) < \frac{\log(n-1)}{\log 2} + o(n^{-0.58}).$$

The following problems were posed by Silverman in 1995 [136] (in the setting of $k = 0$):

Problem 1. Is there is a true asymptotic formula for $|V(\mathbb{Z}) \cap B(R)|$ with main term proportional to $\log(R)^\beta$?

Problem 2. Is $\beta(n)$ irrational?

In [68] a complete answer to Problem 1 was obtained by extending Baragar's exponential rate of growth estimate to a true asymptotic formula.[15]

When $k > 0$, there are certain exceptional families of solutions to (5.1) that have a different quality of growth and, for fixed k, a, n, we write $\mathcal{E}$ for the set of exceptional tuples. We obtain the following theorem for the asymptotic number of Markoff–Hurwitz tuples:

Theorem 9. *For each (n, a, k) with $V(\mathbb{Z}) - \mathcal{E}$ infinite, there is a positive constant $c = c(n, a, k)$ such that*

$$|(V(\mathbb{Z}) - \mathcal{E}) \cap B(R)| = c(\log R)^\beta + o((\log R)^\beta).$$

Here β is the same constant as in (5.2).

After renormalizing (5.1), which allows us to set $a = 1$, and rearranging entries, Markoff–Hurwitz transformations induce the moves

$$\lambda_j(z_1, \dots, z_n) = \left(z_1, \dots, \widehat{z_j}, \dots, z_n, \left(\prod_{i \neq j} z_i\right) - z_j\right), \quad 1 \le j \le n-1, \tag{5.4}$$

on ordered tuples of real numbers. Above, $\widehat{\bullet}$ denotes omission. If sufficiently many of the z_i are large, the move λ_j can be approximated by

$$z \mapsto (z_1, \dots, \widehat{z_j}, \dots, z_n, \prod_{i \neq j} z_i)$$

[15] The techniques in [5] "were inspired in part by Boyd's work on the Apollonian packing problem [29–31]". Boyd's result was extended to a true asymptotic formula in the work of Kontorovich and Oh [86].

to high accuracy relative to the largest entries of z. When the z_i are positive, at the level of logarithms this corresponds to

$$(\log z_1, \log z_2, \dots, \log z_n) \mapsto (\log z_1, \dots, \widehat{\log z_j}, \dots, \log z_n, \sum_{i \neq j} \log z_i).$$

Thus one is naturally led to study the linear semigroup generated by linear maps

$$\gamma_j(y_1, y_2, \dots, y_n) = (y_1, \dots, \widehat{y_j}, \dots, y_n, \sum_{i \neq j} y_i) \tag{5.5}$$

on ordered n-tuples $(y_1, \dots, y_n)$.

Let

$$\Gamma = \langle \gamma_1, \dots, \gamma_{n-1} \rangle_+,$$

where we have written a "+" to indicate we are generating a semigroup, not a group.

An important idea in [68] that explains why we are able to make progress on the counting problem is that we replace the generators of Γ with the countably infinite generating set $T_\Gamma = \{\gamma_{n-1}^A \gamma_j : A \in \mathbb{Z}_{\geq 0}, 1 \leq j \leq n-2\}$ and then consider the semigroup $\Gamma' = \langle T_\Gamma \rangle_+$.

Both Γ and Γ' preserve the nonnegative ordered hyperplane

$$\mathcal{H} \equiv \left\{ (y_1, \dots, y_n) \in \mathbf{R}^n_{\geq 0} : y_1 \leq y_2 \leq \dots \leq y_n, \sum_{j=1}^{n-1} y_j = y_n \right\} \subset \mathbf{R}^n_{\geq 0}; \tag{5.6}$$

any element of Γ maps ordered tuples in $\mathbf{R}^n_{\geq 0}$ into $\mathcal{H}$. Therefore the study of orbits of Γ and Γ' on ordered tuples boils down to the study of orbits in $\mathcal{H}$. We can use the basis

$$e_j = (0, \dots, 0, \underbrace{1}_{j}, 0, \dots, 0, 1)$$

for the subspace spanned by $\mathcal{H}$. This basis clarifies the action of Γ'.

When $n = 3$, the linear map $\sigma : \mathcal{H} \to \mathcal{H}$ defined by

$$\sigma(a, b, a+b) = \mathsf{order}(b-a, a, b), \tag{5.7}$$

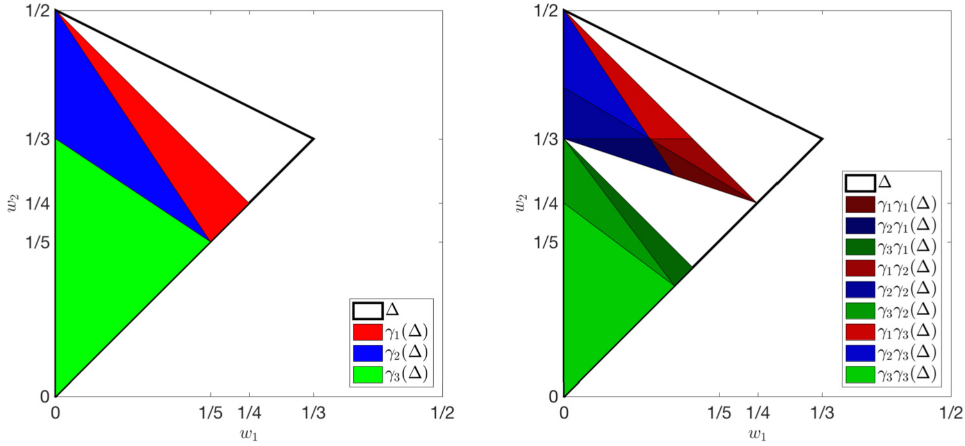
where order puts a tuple in ascending order from left to right, is such that for $j = 1, 2$ we have $\sigma\gamma_j.y = y$ for all $y \in \mathcal{H}$. Repeatedly applying the map σ to a triple $(a, b, a+b)$ with $a \leq b \in \mathbb{Z}$ performs the Euclidean algorithm on a, b. However, one application of σ corresponds in general to less than one step of the algorithm. Replacing Γ with Γ' corresponds to speeding this up so one whole step of the Euclidean algorithm corresponds to one semigroup generator. As for counting, the orbit of $(0, 1, 1)$ under Γ is precisely those $(a, b, a+b)$ with $(a, b) = 1$ and thus can be counted by elementary methods.

When $n = 3$, the semigroup Γ' is generated by

$$g_A := \gamma_2^A \gamma_1 = \begin{pmatrix} 0 & 1 \\ 1 & A+1 \end{pmatrix}$$

with respect to the basis $\{e_1, e_2\}$. These generators are classically connected with continued fractions by the formulae

$$\begin{pmatrix} 0 & 1 \\ 1 & A_1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & A_2 \end{pmatrix} \cdots \begin{pmatrix} 0 & 1 \\ 1 & A_k \end{pmatrix} = \begin{pmatrix} \star & b \\ \star & d \end{pmatrix}, \quad \frac{b}{d} = \cfrac{1}{A_1 + \cfrac{1}{A_2 + \ddots \frac{1}{A_k}}}.$$



1/2
1/3
w_2 1/4
1/5
0
0
1/5 1/4
1/3
1/2
w_1
Δ
$\gamma_1(\Delta)$
$\gamma_2(\Delta)$
$\gamma_3(\Delta)$
1/2
1/3
w_2 1/4
1/5
0
0
1/5 1/4
1/3
1/2
w_1
Δ
$\gamma_1\gamma_1(\Delta)$
$\gamma_2\gamma_1(\Delta)$
$\gamma_3\gamma_1(\Delta)$
$\gamma_1\gamma_2(\Delta)$
$\gamma_2\gamma_2(\Delta)$
$\gamma_3\gamma_2(\Delta)$
$\gamma_1\gamma_3(\Delta)$
$\gamma_2\gamma_3(\Delta)$
$\gamma_3\gamma_3(\Delta)$

FIGURE 15

When $n = 4$, the semigroup elements map $\Delta = \mathbb{H}/\mathbb{R}_+$ into a strictly smaller subset. After iteration, this leads to more and more empty space (see also Figure 16). This does not occur when $n = 3$, as one can also see from the picture: the action of the group elements γ_2 and γ_3 on the vertical coordinate axis is a copy of the $n = 3$ dynamics.

When $n = 4$, the semigroup Γ acts in the basis given by the e_i as

$$\gamma_1 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \quad \gamma_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \quad \gamma_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}.$$

This semigroup appears naturally in different areas of mathematics. In most situations that this semigroup appears, as is the case in [68], the dynamics of the projective linear action of Γ on $\mathbb{R}^3_+/\mathbb{R}_+$ becomes relevant. Up to the minor modification of possibly multiplying the generators on the left or right by permutation matrices, the iterated function system given by the projective linear action of Γ on $\mathbb{R}^3_+/\mathbf{R}_+$ has a fractal attracting set that is known as the *Rauzy gasket* [95].

So the semigroups Γ and Γ' are natural extensions of the Euclidean algorithm and continued fractions semigroup to higher dimensions. Writing $\Delta = \mathcal{H}/\mathbb{R}_+$, we can view Δ as a subset of $\mathbb{R}^{n-2}$. The key distinction that appears when $n \geq 4$ is that

$$\Delta \neq \bigcup_{j=1}^{n-1} \gamma_j(\Delta)$$

and so the induced dynamics on $\mathcal{H}/\mathbb{R}_+$ has "holes" as we illustrate in Figure 15.

Structure of the proof and the difficulties that arise. Here we highlight some of the main difficulties that must be overcome during the proof of Theorem 9. It is illuminating to recall the methods used by Lalley[16] in [90] where the action of a Schottky subgroup G

[16] See Mark Policott's ICM report [119] for an overview of recent developments pertaining to dynamical zeta functions and thermodynamic formalism.

of $\mathrm{SL}_2(\mathbf{R})$ on the hyperbolic upper half-plane $\mathbb{H}$ is considered. Lalley obtains that, for any $x \in \mathbb{H}$, the number $\mathcal{N}(x, r)$ of elements γ of G such that

$$d_{\mathbb{H}}(i, \gamma x) - d_{\mathbb{H}}(i, x) \leq r,$$

where $d_{\mathbb{H}}$ is hyperbolic distance, satisfies $\mathcal{N}(x, r) \approx Ce^{\delta r}$, where $\delta = \delta(G)$ is the Hausdorff dimension of the limit set of G, and $C = C(G, x) > 0$. Lalley's proof incorporates at various stages the following arguments:

Shell argument. By repeated application of a "renewal equation," the quantity $\mathcal{N}(x, r)$ is related to a sum of $\mathcal{N}(y, r')$, where the sum is over y on a shell of radius $\approx cr$ in a Cayley tree of G, and r' is a translate of r that corrects for the passage between x and y. The purpose of this shell argument is that now, the points y lie close to $\partial\mathbb{H}$.

Passage to the boundary. Each of the resulting $\mathcal{N}(y, r')$ is compared to an analogous quantity $\mathcal{N}^*(y^*, r')$ where y^* is a point in $\partial\mathbb{H}$ close to y. Because each y is close to $\partial\mathbb{H}$, the errors incurred are acceptable.

Transfer operator techniques. Asymptotic formulas for $\mathcal{N}^*(y^*, r')$ are obtained using the renewal method and spectral estimates for transfer operators. This gives asymptotic formulas for $\mathcal{N}(y, r')$. The main terms of the asymptotic formulas satisfy recursive relationships between different y.

Recombination. One finally has to recombine all the asymptotic formulas obtained for $\mathcal{N}(y, r')$ to obtain an asymptotic formula for $\mathcal{N}(x, r)$. This is done using the recursive formulas obtained in the previous step.

Trying to follow the method outlined above for this orbital counting problem, we first need a suitable replacement for $\partial\mathbb{H}$. Our idea is to use the projectivization of the hyperplane $\mathcal{H}$; we call this set Δ. We compare points in the orbit of Λ (generated by λ_j in (5.4) to points in Δ by taking logarithms of all coordinates and then projectivizing. This process does not necessarily lead to a point in Δ; there is an important parameter $\alpha(z) = \prod_{j=1}^{n-2} z_j$ that appears throughout the paper and measures how good the fit is. If $\alpha(z)$ is large, then one can, in analogy with Lalley's setting, think of z as being "close to the boundary."

For Lalley, the word length of γ is roughly proportional to the quantity $d_{\mathbb{H}}(i, \gamma x) - d_{\mathbb{H}}(i, x)$ with respect to which he counts. This implies, during the shell argument, that all the elements of the shell are roughly the same distance from $\partial\mathbb{H}$. However, for us, there are arbitrarily long words in the generators of Λ for which $\alpha(z)$ is small. We solve this problem using "acceleration," by replacing Λ by Λ', and instead aim to follow Lalley's argument for orbits of Λ'. This has the immediate benefit that we can guarantee that elements z of shells of radius L, with respect to Λ', have large $\alpha(z)$, if we make L appropriately large.

However, the acceleration also has some costs to be paid. The first issue arising is that now Λ' has countably many generators, so shells for word length on Λ' are not finite. Instead of using shells, we use intersections of shells with the elements of the Λ'-orbit whose coordinates are not too large. The second issue is that the original Λ-orbit breaks up

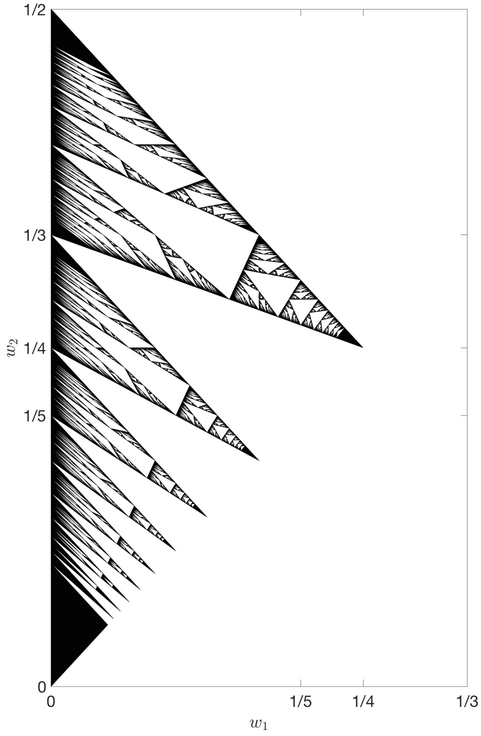


FIGURE 16

In the same setting ($n = 4$) of Figure 15, we show in black the images of Δ under the action of all words of length 10 in the generators $\{\gamma_1, \gamma_2, \gamma_3\}$.

into countably many Λ'-orbits. So we not only have to perform the recombination argument for Λ', but then have to perform an extra summation over the countably many Λ'-orbits.

After setting up our shell argument appropriately, we must perform the passage to the boundary (i.e., Δ). To this end, we compare orbits of Λ' to orbits of Γ', where Γ' is the linear semigroup. To get this to work, we must exploit the following "shadowing" feature of the map log that takes logarithms of all entries of a vector. It says (roughly) that if $\log(z)$ is within ϵ of $y \in \mathscr{H}$, with ϵ on the scale of $\alpha(z)^{-2}$, then for all $\lambda \in \Lambda'$, $\log(\lambda(z))$ is within ϵ of $\gamma(\log(z))$, where $\gamma \in \Gamma'$ is matched with λ in a natural way.

The completion of the proof relies on spectral estimates for transfer operators associated to the projective linear action of Γ' on Δ. There are three key issues arising here. First, to obtain the spectral estimates we need, we must establish that the action of Γ' on Δ is uniformly contracting; it is important to note that this argument would not work if the acceleration had not been performed previously. Secondly, we need to establish that the relevant "log-Jacobian" cocycle over the dynamical system is not cohomologous to a lattice cocycle. Finally, but importantly, we must obtain spectral estimates for transfer operators acting on $C^1(\Delta)$ which is accomplished by adapting Liverani's approach to spectral estimates from [97]. See section 4 of [68], and references therein, for the discussion of Gauss map and Gauss measure [70,89] in this context.

The question of whether β is irrational (Problem 2) remains a tantalizing open question, and one may wonder whether it is even algebraic. Our methods do give some partial insight into the nature of this mysterious number in terms of the action of Γ' on $\mathcal{H}/\mathbf{R}_+$.

Theorem 10. *The number β is the unique parameter in $(1,\infty)$ such that there exists a probability measure ν_β on $\Delta = \mathcal{H}/\mathbf{R}_+$ with the property*

$$\int_{w\in\Delta} f(w)\,d\nu_\beta(w) = \sum_{\gamma\in T_\Gamma}\int_{w\in\Delta} f(\gamma.w)|\mathrm{Jac}_w(\gamma)|^{\frac{\beta}{n-1}}\,d\nu_\beta(w)$$

for all $f \in C^0(\Delta)$. We call ν_β a conformal measure.

Theorem 10 can be viewed as a partial analog of the connection between the exponent of growth of a finitely generated Fuchsian group and the Hausdorff dimension of its limit set as a result of Patterson–Sullivan theory [117,138,139]. In our setting, the lack of any symmetric space means the parameter β is not in any obvious way connected to the Hausdorff dimension of the compact Γ'-invariant subset of Δ.

The issue of the existence of a *single* integral solution for general a and k is very subtle, even for $n = 3$, as discussed in the next section.

6. HASSE PRINCIPLE ON SURFACES OF MARKOFF TYPE

Little is known about the values at integers assumed by *affine cubic forms*[17] F in three variables. For $k \neq 0$, set

$$V_{k,F} = \{\mathbf{x} = (x_1,x_2,x_3) : \ F(\mathbf{x}) = k\}. \tag{6.1}$$

The basic question is for which k is $V_{k,F}(\mathbb{Z}) \neq \emptyset$, or, more generally, infinite or Zariski dense in $V_{k,F}$?

A prime example is $F = S$, the sum of three cubes,

$$S(x_1,x_2,x_3) = x_1^3 + x_2^3 + x_3^3. \tag{6.2}$$

There are obvious local congruence obstructions, namely that $V_{k,S}(\mathbb{Z}) = \emptyset$ if $k \equiv 4, 5 \pmod 9$, but beyond that, it is possible that the answers to all three questions are yes for all the other k's, which we call the *admissible* values (see [50,113]). It is known that strong approximation in its strongest form fails for $V_{k,S}(\mathbb{Z})$; the global obstruction coming from an application of cubic reciprocity [41,49,75]). Moreover, the authors of [94] and [7] show that $V_{1,S}(\mathbb{Z})$ is Zariski dense in $V_{1,S}$.

In [71] Ghosh and Sarnak investigate the Markoff form $F = M$,

$$M(\mathbf{x}) = x_1^2 + x_2^2 + x_3^2 - x_1x_2x_3. \tag{6.3}$$

[17] By an *affine form* f in n variables we mean $f \in \mathbb{Z}[x_1,\dots,x_n]$ whose leading homogeneous term f_0 is nondegenerate and such that $f - k$ is (absolutely) irreducible for all constants k.

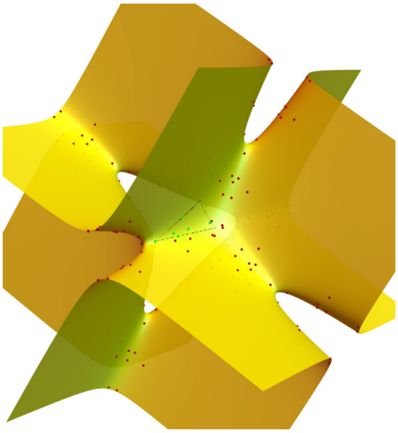


FIGURE 17
Lattice points and fundamental set for $k = 3685$.

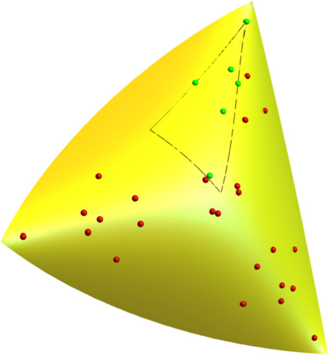


FIGURE 18
Closeup of fundamental set for $k = 3685$.

Except for the case of the Cayley cubic with $k = 4$, $V_{k;M}(\mathbb{Z})$ decomposes into a finite number $\mathfrak{h}_M(k)$ of Γ-orbits. For example, if $k = 0$, then $\mathfrak{h}_M(k) = 2$ corresponds to the orbits of $(0,0,0)$ and $(3,3,3)$. In order to study $\mathfrak{h}_M(k)$ both theoretically and numerically, they give an explicit reduction (descent) for the action of Γ on $V_{k,M}(\mathbb{Z})$. For this purpose, it is convenient to remove an explicit set of special admissible k's, namely those for which there is a point in $V_{k,M}(\mathbb{Z})$ with $|x_j| = 0,\ 1$ or 2. These k's take the form (i) $k = u^2 + v^2$, (ii) $4(k-1) = u^2 + 3v^2$, or (iii) $k = 4 + u^2$. The number of these special k's (referred to as *exceptional*) with $0 \le k \le K$ is asymptotic to $C' \frac{K}{\sqrt{\log K}}$. The remaining admissible k's are called *generic* (all negative admissible k's are generic). For them Ghosh and Sarnak give the following elegant reduced forms:

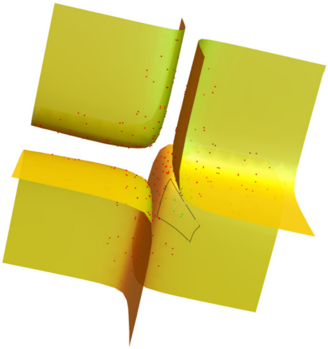


FIGURE 19
Lattice points and fundamental set for $k = -3691$.

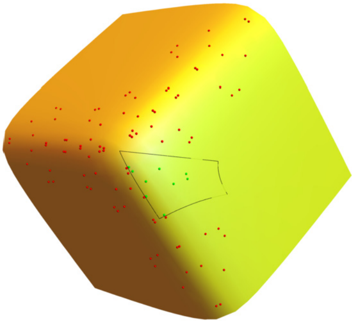


FIGURE 20
Closeup of fundamental set for $k = -3691$.

Proposition 11. (1) *Let $k \geq 5$ be generic and consider the compact set*

$$\mathfrak{F}_k^+ = \{\mathbf{u} \in \mathbb{R}^3 : 3 \leq u_1 \leq u_2 \leq u_3\,,\ u_1^2 + u_2^2 + u_3^2 + u_1u_2u_3 = k\}.$$

The points in $\mathfrak{F}_k^+(\mathbb{Z}) = \mathfrak{F}_k^+ \cap \mathbb{Z}^3$ are Γ-inequivalent, and any $\mathbf{x} \in V_{k,M}(\mathbb{Z})$ is Γ-equivalent to a unique point $\mathbf{u}' = (-u_1, u_2, u_3)$ with $\mathbf{u} = (u_1, u_2, u_3) \in \mathfrak{F}_k^+(\mathbb{Z})$.

(2) *Let $k < 0$ be admissible and consider the compact set*

$$\mathfrak{F}_k^- = \{\mathbf{u} \in \mathbb{R}^3 : 3 \leq u_1 \leq u_2 \leq u_3 \leq \frac{1}{2}u_1u_2\,,\ u_1^2 + u_2^2 + u_3^2 - u_1u_2u_3 = k\}.$$

The points in $\mathfrak{F}_k^-(\mathbb{Z}) = \mathfrak{F}_k^- \cap \mathbb{Z}^3$ are Γ-inequivalent, and any $\mathbf{x} \in V_{k,M}(\mathbb{Z})$ is Γ-equivalent to a unique point $\mathbf{u} = (u_1, u_2, u_3) \in \mathfrak{F}_k^-(\mathbb{Z})$.

Some consequences of this are as follows: As $k \to \pm\infty$, we have

$$\mathfrak{h}_M(k) \ll_\varepsilon |k|^{\frac{1}{3}+\varepsilon}.$$

This follows from the fact that when considering the values taken by the corresponding indefinite quadratic form in the y and z variables, for each fixed x, the units are bounded in number due to the restrictions imposed by the fundamental sets.

Let $\mathfrak{h}_M^{\pm}(k) = |\mathfrak{F}_k^{\pm}(\mathbb{Z})|$ where $\pm = \operatorname{sgn}(k)$, this being defined for any k. Then for generic k, $\mathfrak{h}_M^{\pm}(k) = \mathfrak{h}_M(k)$ while otherwise $\mathfrak{h}_M(k) \le \mathfrak{h}_M^{\pm}(k)$. We have

$$\sum_{\substack{k\neq 4\\ |k|\le K}} \mathfrak{h}_M^{\pm}(k) \sim C^{\pm} K(\log K)^2, \tag{6.4}$$

where $C^{\pm} > 0$ and $K \to \infty$.

So on average the numbers $\mathfrak{h}_M(k)$ are small. The explicit fundamental domains allow for the numerical computations; these indicate that

$$\sum_{\substack{0<k\le K\\ k \text{ admissible}\\ \mathfrak{h}_M(k)=0}} 1 \sim C_0 K^{\theta}, \tag{6.5}$$

with $C_0 > 0$ and $\theta \approx 0.8875\ldots.$

The main result in [71] concerns the values assumed by M and the Hasse failures in (6.5):

Theorem 12. (i) *There are infinitely many Hasse failures. More precisely, the number of* $0 < k \le K$ *and* $-K \le k < 0$ *for which the Hasse principle fails is at least* $\sqrt{K}(\log K)^{-\frac{1}{4}}$ *for K large.*

(ii) *Fix* $t \ge 0$. *Then as* $K \to \infty$,

$$\#\{|k| \le K : k \text{ admissible}, \ \mathfrak{h}_M(k) = 0\} = o(K).$$

Hasse failures are produced by an obstruction via quadratic reciprocity. They come in two types: one via direct use of reciprocity and the second also incorporating the descent group. Recently Colliot-Thélène, Wei, and Xu [48] and, independently, Loughran and Mitankin [98] have shown that the obstruction of the first (but not the second type) can be explained in terms of integral Brauer–Manin obstruction. For example, if $k = 4 + 2\nu^2$, with ν having all of its prime factors congruent to $\pm 1 \pmod 8$ and ν congruent to $0, \pm 3, \pm 4 \pmod 9$, then k is admissible but $V_{k,M}(\mathbb{Z}) = \emptyset$.

Part (ii) of the theorem is proved by comparing the number of points on $V_{k,M}(\mathbb{Z})$ in certain tentacled regions gotten by special plane sections, with the expected number of solutions according to a product of local densities; the crucial point being that the variance of this comparison goes to zero on averaging $|k| \le K$. As detailed in [71], this moving plane quadric method applies to more general cubic surfaces including those that do not carry morphisms.

ACKNOWLEDGMENTS

I would like to thank Jean Bourgain, Emmanuel Breuillard, Michael Magee, Igor Pak, Ryan Ronan, and Peter Sarnak. My various collaborations with them form the core of this report.

Figure 1 is courtesy of Matthew de Courcy-Ireland.
Figure 2 is courtesy of Elena Fuchs.
Figures 3–8 are courtesy of William Goldman.
Figures 9–14 are courtesy of Serge Cantat.
Figures 17–20 are courtesy of Amit Ghosh.

FUNDING

The author was supported, in part, by NSF award DMS-1603715.

REFERENCES

[1] M. Aigner, *Markov's Theorem and 100 years of the Uniqueness Conjecture*. Springer, 2013.

[2] I. Aliev and C. Smyth, Solving Algebraic equations in roots of unity. *Forum Math.* **24** (2012), 641–645.

[3] A. Baragar, *The Markoff equation and equations of Hurwitz*. PhD Thesis, Brown University, RI, 1991.

[4] A. Baragar, Asymptotic growth of Markoff–Hurwitz numbers. *Compos. Math.* **94** (1994), no. 1, 1–18.

[5] A. Baragar, The exponent for the Markoff–Hurwitz equations. *Pacific J. Math.* **182** (1998), no. 1, 1–21.

[6] E. Bedford and J. Smillie, Real polynomial diffeomorphisms with maximal entropy: Tangencies. *Ann. of Math. (2)* **160** (2004), no. 1, 1–26.

[7] F. Beukers, Ternary form equations. *J. Number Theory* **54** (1995), 113–133.

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

[9] J. Bourgain, On the Erdös-Volkmann and Katz-Tao ring conjecture, *Geom. Funct. Anal.* **13** (2003), 334–365.

[10] J. Bourgain, The sum-product theorem in $\mathbb{Z}_q$ with q arbitrary. *J. Anal. Math.* **106**, 1–93 (2008)

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

[12] J. Bourgain and A. Gamburd, New results on expanders. *C. R. Math. Acad. Sci. Paris* **342** (2006).

[13] J. Bourgain and A. Gamburd, Uniform expansion bounds for Cayley graphs of $SL_2(\mathbb{F}_p)$. *Ann. of Math.* **167** (2008), 625–642.

[14] J. Bourgain and A. Gamburd, Random walks and expansion in $\mathrm{SL}_d(\mathbb{Z}/p^n\mathbb{Z})$. *C. R. Math. Acad. Sci. Paris* **346** (2008), no. 11–12, 619–623.

[15] J. Bourgain and A. Gamburd, On the spectral gap for finitely-generated subgroups of SU(2). *Invent. Math.* **171** (2008), no. 1, 83–121.

[16] J. Bourgain and A. Gamburd, 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.

[17] J. Bourgain and A. Gamburd, Expansion and random walks in $\mathrm{SL}_d(\mathbb{Z}/p^n\mathbb{Z})$. II. With an appendix by Bourgain. *J. Eur. Math. Soc. (JEMS)* **11** (2009), no. 5, 1057–1103.

[18] J. Bourgain and A. Gamburd, Spectral gaps in $\mathrm{SU}(d)$. *C. R. Math. Acad. Sci. Paris* **348** (2010), no. 11–12, 609–611.

[19] J. Bourgain and A. Gamburd, A spectral gap theorem in $\mathrm{SU}(d)$. *J. Eur. Math. Soc. (JEMS)* **14** (2012), no. 5, 1455–1511.

[20] J. Bourgain, A. Gamburd, and P. Sarnak, Sieving and expanders. *C. R. Acad. Sci. Paris, Sér. I* **343** (2006).

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

[22] J. Bourgain, A. Gamburd, and P. Sarnak, Generalization of Selberg's $\frac{3}{16}$ theorem and affine sieve. *Acta Math.* **207** (2011), no. 2, 255–290.

[23] J. Bourgain, A. Gamburd, and P. Sarnak, Markoff surfaces and strong approximation: 1. 2016, arXiv:1607.01530.

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

[25] J. Bourgain, A. Gamburd, and P. Sarnak, Strong approximation and diophantine properties of Markoff numbers, preprint.

[26] J. Bourgain, A. Gamburd, and P. Sarnak, Strong approximation for varieties of Markoff type, preprint.

[27] J. Bourgain, N. Katz, and T. Tao, A sum-product estimate in finite fields, and applications. *Geom. Funct. Anal.* **14** (2004), no. 1, 27–57.

[28] J. Bourgain and P. P. Varjú, Expansion in $SL_d(Z/qZ)$, q arbitrary. *Invent. Math.* **188** (2012).

[29] D. W. Boyd, The disk-packing constant. *Aequationes Math.* **7** (1971), 182–193.

[30] D. W. Boyd, Improved bounds for the disk-packing constant. *Aequationes Math.* **9** (1973), 99–106.

[31] D. W. Boyd, The sequence of radii of the Apollonian packing. *Math. Comp.* **39** (1982), no. 159, 249–254.

[32] E. Breuillard, A strong Tits alternative. 2008, arXiv:0804.1395.

[33] E. Breuillard, Approximate subgroups and superstrong approximation. In *Groups St Andrews 2013*, pp. 1–50, London Math. Soc. Lecture Note Ser. 422, Cambridge University Press, Cambridge, 2015.

[34] E. Breuillard and A. Gamburd, Strong uniform expansion in $SL_2(\mathbb{F}_p)$. *Geom. Funct. Anal.* **20** (2010), no. 5, 1201–1209.

[35] E. Breuillard, B. Green, and T. Tao, Approximate subgroups of linear groups. *Geom. Funct. Anal.* **21** (2011).

[36] E. Breuillard and H. Oh, *Thin groups and superstrong approximation*. Math. Sci. Res. Inst. Publ. 61, Cambridge University Press, Cambridge, 2014.

[37] S. Cantat, Dynamique des automorphismes des surfaces K3. *Acta Math.* **187** (2001), no. 1, 1–57

[38] S. Cantat, Bers, Hénon, Painlevé and Schrödinger. *Duke Math. J.* **149** (2009), no. 3, 411–460.

[39] J. W. S. Cassels, The Markoff chain. *Ann. of Math. (2)* **50** (1949), 676–685.

[40] J. W. S. Cassels, *An Introduction to Diophantine Approximation*. Cambridge Tracts in Mathematics and Mathematical Physics 45, Cambridge University Press, New York, 1957, x+166 pp.

[41] J. W. S. Cassels, A note on the Diophantine equation $x^3 + y^3 + z^3 = 3$. *Math. Comp.* **44** (1985), 265–266.

[42] F. Celler, C. R. Leedham-Green, S. Murray, A. Niemeyer, and E. A. O'Brien, Generating random elements of a finite group. *Comm. Algebra* **23** (1995), 4931–4948.

[43] M.-C. Chang, Elements of large order in prime finite fields. *Bull. Aust. Math. Soc.* **88** (2013).

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

[45] W. Chen, Nonabelian level structures, Nielsen equivalence, and Markoff triples. 2021, arXiv:2011.12940.

[46] H. Cohn, Approach to Markoff's minimal forms through modular functions. *Ann. of Math.* **61** (1955), 1–12.

[47] H. Cohn, Markoff forms and primitive words. *Math. Ann.* **196** (1972), 8–22.

[48] J.-L. Colliot-Thélène, D. Wei, and F. Xu, Brauer–Manin obstruction for Markoff surfaces. Annali della Scuola Normale Superiore di Pisa – Classe di Scienze Volume XXI/2021, no. 3 série V (2021).

[49] J. Colliot-Thélène and O. Wittenberg, Groupe de Brauer et points entiers de deux familles de surfaces cubiques affines. *Amer. J. Math.* **134** (2012), no. 5, 1303–1327.

[50] W. Conn and L. N. Vaserstein, *On sums of three integral cubes*. pp. 285–294, Contemp. Math. 166, Amer. Math. Soc., 1994.

[51] P. Corvaja and U. Zannier, On integral points on surfaces. *Ann. of Math. (2)* **160** (2004).

[52] P. Corvaja and U. Zannier, On the greatest prime factor of Markov pairs. *Rend. Semin. Mat. Univ. Padova* **116** (2006), 253–260.

[53] P. Corvaja and 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).

[54] M. de Courcy-Ireland, Non-planarity of Markoff graphs mod p. 2022, arXiv:2105.12411v3.

[55] M. de Courcy-Ireland and S. Lee, Experiments with the Markoff surface. *Exp. Math.* (2020). DOI 10.1080/10586458.2019.1702123.

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

[57] W. Duke, Hyperbolic distribution problems and half-integral weight Maass forms. *Invent. Math.* **92** (1988).

[58] M. Einsiedler, E. Lindenstrauss, P. Michel, and A. Venkatesh, Distribution of periodic torus orbits on homogeneous spaces. *Duke Math. J.* **148** (2009).

[59] M. Einsiedler, E. Lindenstrauss, P. Michel, and A. Venkatesh, The distribution of closed geodesics on the modular surface, and Duke's theorem. *Enseign. Math. (2)* **58** (2012).

[60] M. H. El-Huti, Cubic surfaces of Markov type. *Math. USSR, Sb.* **22** (1974), no. 3, 333–348.

[61] J. Fornaess and N. Sibony, Complex dynamics in higher dimension. II. In: *Modern methods in complex analysis (Princeton, NJ, 1992).* Ann. of Math. Stud. 137, Princeton University Press, Princeton, NJ, 1995.

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

[63] E. Fuchs, Counting problems in Apollonian packings. *Bull. Amer. Math. Soc. (N.S.)* **50** (2013), no. 2, 229–266.

[64] E. Fuchs, K. Lauter, M. Litman, A. Tran, A Cryptographic Hash Function from Markoff Triples. 2021, arXiv:2107.10906.

[65] E. Fuchs, M. Litman, J. Silverman, A. Tran, Orbits on K3 Surfaces of Markoff Type. 2022, arXiv:2201.12588.

[66] A. Gamburd, Spectral gap for infinite index "congruence" subgroups of $\mathrm{SL}_2(\mathbb{Z})$. *Israel J. Math.* **127** (2002).

[67] A. Gamburd, Singular adventures of Baron Bourgain in the labyrinth of the continuum. *Notices Amer. Math. Soc.* **67** (2020), no. 11, 1716–1733.

[68] A. Gamburd, M. Magee, and R. Ronan, An asymptotic formula for integer points on Markoff–Hurwitz varieties. *Ann. of Math. (2)* **190** (2019), no. 3, 751–809.

[69] A. Gamburd and I. Pak, Expansion of product replacement graphs. *Combinatorica* **26** (2006), no. 4, 411–429.

[70] C. F. Gauss, Brief an Laplace vom 30 Jan. 1812, Werke X1, 1812, pp. 371–374.

[71] A. Ghosh and P. Sarnak, Integral points on Markoff type cubic surfaces. 2022, arXiv:1706.06712v1.

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

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

[74] D. S. Gorshkov, Geometry of Lobachevskii in connection with certain questions of arithmetic. *Zap. Nauch. Sem. Lenin. Otd. Math. Inst. V.A. Steklova AN SSSR* **67** (1977), 39–85. English translation in *J. Sov. Math.* **16** (1981) 788–820.

[75] D. R. Heath-Brown, The density of zeros of forms for which weak approximation fails. *Math. Comp.* **59** (1992), no. 200, 613–623.

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

[77] H. A. Helfgott, Growth and generation in $\mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})$. *Ann. of Math. (2)* **167** (2008), no. 2, 601–623.

[78] H. A. Helfgott, Growth in groups: ideas and perspectives. *Bull. Amer. Math. Soc. (N.S.)* **52** (2015), no. 3, 357–413.

[79] C. Hooley, On Artin's conjecture. *J. Reine Angew. Math.* **225** (1967).

[80] C. Hooley, *Applications of sieve methods to the theory of numbers*. Cambridge Tracts in Math. 70, Cambridge, 1976.

[81] S. Hoory, N. Linial, and A. Wigderson, Expander graphs and their applications. *Bull. Am. Math. Soc.* (2006).

[82] E. Hrushovski, Stable group theory and approximate subgroups. *J. Amer. Math. Soc.* **25** (2012).

[83] A. Hurwitz, Über eine Aufgabe der unbestimmten Analysis. *Arch. Math. Phys.* **11** (1907), no. 3, 185–196.

[84] M. Kaluba, D. Kielak, and P. W. Nowak, On property (T) for $\mathrm{Aut}(F_n)$ and $\mathrm{SL}_n(\mathbb{Z})$. *Ann. of Math. (2)* **193** (2021), no. 2, 539–562.

[85] A. Kontorovich and J. Lagarias, On the expected number of prime factors in the affine sieve with toral Zariski closure, *Exp. Math.* **30** (2021), no. 4, 575–586.

[86] A. Kontorovich and H. Oh, Apollonian circle packings and closed horospheres on hyperbolic 3-manifolds. *J. Amer. Math. Soc.* **24** (2011), no. 3, 603–648. With an appendix by Oh and Nimish Shah.

[87] S. V. Konyagin, S. V. Makarychev, I. E. Shparlinski, and I. V. Vyugin, On the structure of graphs of Markoff triples. *Q. J. Math.* **71** (2020), no. 2, 637–648.

[88] M. A. Korkine and G. Zolotareff, Sur les formes quadratiques. *Math. Ann.* **6** (1873), 279–284.

[89] R. O. Kuzmin, On a problem of Gauss. *Atti Congr. Int. Mat., Bologna* **6** (1932), 83–89.

[90] S. P. Lalley, Renewal theorems in symbolic dynamics, with applications to geodesic flows, non-Euclidean tessellations and their fractal limits. *Acta Math.* **163** (1989), no. 1–2, 1–55.

[91] M. J. Larsen and R. Pink, Finite subgroups of algebraic groups. *J. Amer. Math. Soc.* **24** (2011).

[92] M. Laurent, Exponential diophantine equations. *C. R. Acad. Sci.* **296** (1983), 945–947.

[93] P. D. Lax and R. S. Phillips, The asymptotic distribution of lattice points in Euclidean and non-Euclidean spaces. *J. Funct. Anal.* **46** (1982), no. 3, 280–350.

[94] D. H. Lehmer, On the Diophantine equation $x^3 + y^3 + z^3 = 1$. *J. Lond. Math. Soc.* **s1-31** (1956), no. 3, 275–280.

[95] G. Levitt, La dynamique des pseudogroupes de rotations. *Invent. Math.* **113** (1993), no. 3, 633–670.

[96] O. Lisovyy and Y. Tykhyy, Algebraic solutions of the sixth Painlevé equation. *J. Geom. Phys.* **85** (2014), 124–163.

[97] C. Liverani, Decay of correlations. *Ann. of Math. (2)* **142** (1995), no. 2, 239–301.

[98] D. Loughran and V. Mitankin, Integral Hasse principle and strong approximation for Markoff surfaces. *Int. Math. Res. Not. IMRN* **18** (2021), 14086–14122.

[99] A. Lubotzky, Cayley graphs: Eigenvalues, expanders and random walks. In *Surveys in Combinatorics*, edited by P. Rowbinson, London Math. Soc. Lecture Note Ser. 218, Cambridge University Press, 1995.

[100] A. Lubotzky and I. Pak, The product replacement algorithm and Kazhdan's property (T). *J. Amer. Math. Soc.* **14** (2001).

[101] A. Lubotzky and B. Weiss, Groups and Expanders. In *DIMACS Series in Disc. Math. and Theor. Comp. Sci.* 10, edited by J. Friedman, pp. 95–109, 1993.

[102] A. Markoff, Sur les formes quadratiques binaires indéfinies. *Math. Ann.* **15** (1879).

[103] A. Markoff, Sur les formes quadratiques binaires indéfinies. *Math. Ann.* **17** (1880).

[104] C. R. Matthews, Counting points modulo p for some finitely generated subgroups of algebraic groups. *Bull. Lond. Math. Soc.* **14** (1982), 149–154.

[105] C. Matthews, L. Vaserstein, and B. Weisfeiler, Congruence properties of Zariski dense groups. *Proc. Lond. Math. Soc.* **48** (1984), 514–532.

[106] Mazur, Barry The topology of rational points. *Experiment. Math.* **1** (1992), no. 1, 35–45.

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

[108] C. T. McMullen, Dynamics on K3 surfaces: Salem numbers and Siegel disks. *J. Reine Angew. Math*. **545** (2002), 201–233

[109] C. T. McMullen, Automorphisms of projective K3 surfaces with minimum entropy. *Invent. Math*. **203** (2016), no. 1, 179–215

[110] C. Meiri and D. Puder, The Markoff group of transformations in prime and composite moduli. *Duke Math. J.* **167** (2018), no. 14, 2679–2720.

[111] M. Mirzakhani, Counting mapping class group orbits on hyperbolic surfaces. 2016, arXiv:1601.03342v1.

[112] L. J. Mordell, On the integer solutions of the equation $x^2 + y^2 + z^2 + 2xyz = n$. *J. Lond. Math. Soc.* **28** (1953), 500–510.

[113] L. J. Mordell, *Diophantine equations*. Academic Press, London, New York, 1969.

[114] H. Oh, Dynamics on geometrically finite hyperbolic manifolds with applications to Apollonian circle packings and beyond. In *Proceedings of the International Congress of Mathematicians* III, pp. 1308–1331, Hindustan Book Agency, New Delhi, 2010.

[115] H. Oh and D. Winter, Uniform exponential mixing and resonance free regions for convex cocompact congruence subgroups of $\mathrm{SL}_2(\mathbb{Z})$. *J. Amer. Math. Soc.*, **29** (2016), 1069–1115

[116] N. Ozawa, Aut(F_5) has property (T). *J. Inst. Math. Jussieu* **15** (2016), no. 1, 85–90.

[117] S. J. Patterson, The limit set of a Fuchsian group. *Acta Math.* **136** (1976), no. 3–4, 241–273.

[118] V. Platonov and A. Rapinchuk, *Algebraic groups and number theory*. Pure Appl. Math. 139, Academic Press Inc., Boston, MA, 1994.

[119] M. Pollicott, Zeta functions for Anosov flows. In *Proceedings of the International Congress of Mathematicians—Seoul 2014. Vol. III*, pp. 661–681, Kyung Moon Sa, Seoul, 2014.

[120] L. Pyber and E. Szabo, Growth in finite simple groups of Lie type. *J. Amer. Math. Soc.* **29** (2016), no. 1, 95–146.

[121] R. Remak, Über indefinite binare quadratische Minimalformen. *Math. Ann.* **92** (1924), 155–182.

[122] A. Salehi Golsefidy, Super-approximation, I: $\mathfrak{p}$-adic semisimple case. *Int. Math. Res. Not. IMRN* **23** (2017), 7190–7263.

[123] A. Salehi Golsefidy and P. Sarnak, The affine sieve. *J. Acad. Mark. Sci.* **4** (2013), 1085–1105.

[124] A. Salehi Golsefidy and P. P. Varju, Expansion in perfect groups. *Geom. Funct. Anal.* **22** (2012).

[125] P. Sarnak, What is an expander? *Not. Amer. Math. Soc.* **51** (2004), 762–763.

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

[127] P. Sarnak, Integral Apollonian packings. *Amer. Math. Monthly* **118** (2011), no. 4, 291–306.

[128] P. C. Sarnak, Diophantine problems and linear groups. In *Proceedings of the International Congress of Mathematicians, Vol. I, II (Kyoto, 1990)*, pp. 459–471, Math. Soc. Japan, Tokyo, 1991.

[129] P. Sarnak and X. X. Xue, Bounds for multiplicities of automorphic representations. *Duke Math. J.* **64** (1991), no. 1, 207–227.

[130] H. Schwartz and H. T. Muhly, On a class of cubic Diophantine equations. *J. Lond. Math. Soc.* **32** (1957), 379–382.

[131] A. Selberg, On the estimation of Fourier coefficients of modular forms. *Proc. Sympos. Pure Math.* **VII** (1965), 1–15.

[132] C. Series, The geometry of Markoff numbers. *Math. Intelligencer* **7** (1985), no. 3, 20–29.

[133] Y. Shalom, The algebraization of Kazhdan's property (T). In *International Congress of Mathematicians. Vol. II*, pp. 1283–1310, Eur. Math. Soc., Zürich, 2006.

[134] J. H. Silverman, The Markoff equation $X^2 + Y^2 + Z^2 = aXYZ$ over quadratic imaginary fields. *J. Number Theory* **35** (1990), no. 1, 72–104.

[135] J. H. Silverman, Rational points on K3 surfaces: a new canonical height. *Invent. Math.* **105** (1991), no. 2, 347–373.

[136] J. H. Silverman, Counting integer and rational points on varieties. In *Columbia University Number Theory Seminar*, pp. 4, 223–236, Astérisque 228, Math. Soc. France, Paris, 1992.

[137] S. A. Stepanov, The number of points of a hyperelliptic curve over a prime field. *Math. USSR, Izv.* **3** (1969), no. 5, 1103–1114.

[138] D. Sullivan, The density at infinity of a discrete group of hyperbolic motions. *Publ. Math. Inst. Hautes Études Sci.* **50** (1979), 171–202.

[139] D. Sullivan, Entropy, Hausdorff measures old and new, and limit sets of geometrically finite Kleinian groups. *Acta Math.* **153** (1984), no. 3–4, 259–277.

[140] T. Tao, *Expansion in finite simple groups of Lie type*. Grad. Stud. Math. 164, American Mathematical Society, Providence, RI, 2015.

[141] P. A. Vojta, generalization of theorems of Faltings and Thue–Siegel–Roth–Wirsing. *J. Amer. Math. Soc.* **25** (1992).

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

[143] B. Weisfeiler, Strong approximation for Zariski-dense subgroups of semisimple algebraic groups. *Ann. of Math. (2)* **120** (1984), no. 2, 271–315.

[144] J. P. Whang, Arithmetic of curves on moduli of local systems. *Algebra Number Theory* **14** (2020), no. 10, 2575–2605.

[145] J. P. Whang, Global geometry on moduli of local systems for surfaces with boundary. *Compos. Math.* **156** (2020), no. 8, 1517–1559.

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

[147] D. Zagier, On the number of Markoff numbers below a given bound. *Math. Comp.* **39** (1982), no. 160, 709–723.

ALEXANDER GAMBURD

The Graduate Center, CUNY, New York, NY, USA, agamburd@gmail.com