An asymptotic formula for integer points on Markoff-Hurwitz varieties

Alex Gamburd*, Michael Magee†, and Ryan Ronan

July 17, 2019

Abstract

We establish an asymptotic formula for the number of integer solutions to the Markoff-Hurwitz equation

$$x_1^2 + x_2^2 + \ldots + x_n^2 = ax_1x_2\ldots x_n + k.$$

When $n \geq 4$ the previous best result is by Baragar (1998) that gives an exponential rate of growth with exponent β that is not in general an integer when $n \geq 4$. We give a new interpretation of this exponent of growth in terms of the unique parameter for which there exists a certain conformal measure on projective space.

Contents

*A. Gamburd was supported in part by NSF award DMS-1603715.

†M. Magee was supported in part by NSF award DMS-1701357.

1 Introduction

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

$$x_1^2 + x_2^2 + \ldots + x_n^2 = ax_1x_2\ldots x_n + k. \tag{1.1}$$

We call this the generalized[1] *Markoff-Hurwitz equation.* In this paper we count solutions to (1.1) in integers, which we we call *Markoff-Hurwitz tuples.* More precisely, let V be the affine subvariety of $\mathbf{C}^n$ cut out by (1.1). We are interested in the asymptotic size of the set

$$V(\mathbf{Z}) \cap B(R)$$

where $B(R)$ is the ball of radius R in the ℓ^∞ norm on $\mathbf{R}^n \subset \mathbf{C}^n$.

When $n = 3, a = 3$ and $k = 0$ solutions to (1.1) in positive integers are called Markoff triples, and the numbers that appear therein are called Markoff numbers[2]. The Markoff numbers are intimately connected with Diophantine properties of the rationals via the Markoff spectrum [Mar79, Mar80] (see also [Bom07] for an excellent exposition), and also with hyperbolic geometry and free groups [Aig13].

The question of counting $|V(\mathbf{Z}) \cap B(R)|$ for Markoff triples was first investigated in the thesis of Gurwood [Gur76] who established an asymptotic formula using the correspondence between Markoff and Farey trees. An improved error term was obtained by Zagier in [Zag82, pg. 711], and a very clean proof of a slightly weaker result can be found in Belyi [Bel01]. The current best result is due to McShane and Rivin [MR95]:

Theorem 1 (McShane-Rivin)**.** *The number $M(R)$ of Markoff triples (x, y, z) with $x \leq y \leq z \leq R$ is given by*

$$M(R) = C(\log R)^2 + O(\log R \log\log R)$$

as $R \to \infty$, with $C > 0$.

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 [Bar94a], who subsequently in [Bar98] obtained the following result

[1] Normally $k = 0$ is considered.

[2] A long standing conjecture of Frobenius asserts that each Markoff number appears as the maximal entry of only one triple, up to reordering. If one assumes this conjecture, then the problems of counting Markoff triples and numbers are the same.

Theorem 2 (Baragar)**.** *There is a number $\beta = \beta(n)$ such that when $k = 0$, if $V(\mathbf{Z}) - \{(0, 0, \ldots, 0)\}$ is nonempty then*

$$|V(\mathbf{Z}) \cap B(R)| = (\log R)^{\beta + o(1)} \tag{1.2}$$

as $R \to \infty$.

In [Bar98] the following bounds for the exponents $\beta(n)$ were also obtained

$$\begin{aligned} \beta(3) &= 2, \\ \beta(4) &\in (2.430, 2.477), \\ \beta(5) &\in (2.730, 2.798), \\ \beta(6) &\in (2.963, 3.048), \end{aligned} \tag{1.3}$$

and in general

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

In 1995 [Sil95], it was asked by Silverman whether in the setting of $k = 0$

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

2. furthermore, $\beta(n)$ is irrational?

The irrationality of β remains a tantalizing open question and one may wonder whether it is even algebraic. On the other hand, our methods do give some further insight into the nature of this mysterious number (cf. Theorem 10 below). The main goal of this paper is to extend Baragar's exponential rate of growth estimate to a true asymptotic formula[3].

When $k > 0$ there are certain exceptional families of solutions to (1.1) that have a different quality of growth. We describe these families in Definition 15 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 3. *For each (n, a, k) with $V(\mathbf{Z}) - \mathcal{E}$ infinite, there is a positive constant $c = c(n, a, k)$ such that*

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

Here β is the same constant as Theorem 2.

Remark 4. We explain in Section 2.1 that removing $\mathcal{E}$ is necessary in Theorem 3 since the exceptional families have $|\mathcal{E} \cap B(R)| \geq cR$, $c > 0$ for $R \geq R_0(n, a, k)$ when they are non-empty. On the other hand, $\mathcal{E}$ is non-empty only when $k - n + 2$ or $k - n - 1$ is a square.

Remark 5. The issue of the existence and infinitude of integral solutions for general a, k, even for $n = 3$, is quite subtle: see [Mor53, SM57]. In recent work of Ghosh and Sarnak [GS17], the Hasse principle is established to hold for Markoff-type cubic surfaces $x_1^2 + x_2^2 + x_3^2 - x_1x_2x_3 = k$ for almost all k, but also fails to hold for infinitely many k.

[3] The techniques in [Bar98] "were inspired in part by Boyd's work on the Apollonian packing problem [Boy71, Boy73, Boy82]." Boyd's result was extended to a true asymptotic formula in the work of Kontorovich and Oh [KO11].

As such, we are not able to give an explicit list of n, a, k for which Theorem 3 is valid. However, it is not difficult to generate examples for any given n. Let us discuss the case that $k = 0$, when there are no exceptional solutions. Then it follows from Baragar's Theorem 2 that if there is an element of $V(\mathbf{Z})$ with positive coordinates, then there are infinitely many elements of $V(\mathbf{Z})$. In the paper [Bar94b], Baragar characterizes all pairs (a, n) with $a \geq 2(n-1)^{1/2}$ for which $V(\mathbf{Z})$ has an element with positive coordinates, and this characterization gives explicit examples to which Theorem 3 applies, including the classical example of $a = n$.

Remark 6. Our proof of Theorem 3 makes important use of Baragar's Theorem 2. Our analysis leads to a dynamical system with a certain critical parameter. We use Theorem 2 to prove this critical parameter coincides with β in Section 4.3. We can make this argument even though Theorem 2 applies only to $k = 0$, as our dynamical system only depends on n.

As a consequence, in Theorem 10 we give a new characterization of β as the unique parameter for which there exists a conformal measure for the action of a linear semigroup on projective space.

Our counting arguments, as in [Zag82] and [Bar94a, Bar98], depend on an infinite descent for solutions to (1.1) that goes back to Markoff [Mar80] in the case of Markoff triples and Hurwitz [Hur07] in the higher dimensional setting of $n > 3$, $k = 0$. In Section 2.1 we explain how the counting problem for $V(\mathbf{Z})$ can be related to the analogous one for $V(\mathbf{Z}_+)$, where $\mathbf{Z}_+$ are the positive integers.

Given $x \in V(\mathbf{Z}_+)$, fixing all of the coordinates of x except x_j and viewing (1.1) as a quadratic polynomial in x_j, the other root is given by

$$x_j' = a \prod_{i \neq j} x_i - x_j.$$

Therefore for each j one has the *Markoff-Hurwitz move*

$$m_j(x_1, x_2, \ldots, x_n) = (x_1, x_2, \ldots, \underbrace{a \prod_{i \neq j} x_i - x_j}_{j}, \ldots, x_n)$$

that preserves solutions to (1.1). Infinite descent for the Markoff-Hurwitz equation says that any unexceptional tuple in $V(\mathbf{Z}_+)$ can be reduced to one in a compact set $K_0 = K_0(n, a, k)$ by a sequence of Markoff-Hurwitz moves (cf. Corollary 19).

After renormalizing (1.1), which allows us to set $a = 1$ (see Section 2.2), and rearranging entries, Markoff-Hurwitz moves $\{m_j\}$ induce the moves

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

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

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

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

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

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

$$\gamma_j(y_1, y_2, \ldots, y_n) = \left(y_1, \ldots, \widehat{y_j}, \ldots, y_n, \sum_{i \neq j} y_i \right) \tag{1.4}$$

on ordered n-tuples $(y_1, \ldots, y_n)$. Indeed, this is the approach of Zagier [Zag82] in the setting of Markoff triples and Baragar [Bar94a] for general n, a with $k = 0$. Let

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

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

An important idea in this work that explains why we are able to make progress on the counting problem is that we replace[4] the generators of Γ with the countably infinite generating set

$$T_\Gamma = \left\{ \gamma_{n-1}^A \gamma_j \;:\; A \in \mathbf{Z}_{\geq 0},\ 1 \leq j \leq n-2 \right\}$$

and then consider the semigroup

$$\Gamma' = \langle\, T_\Gamma \,\rangle_+.$$

Both Γ and Γ' are freely generated by their respective generating sets[5]. Notice that Γ and Γ' preserve the nonnegative ordered hyperplane

$$\mathcal{H} \equiv \left\{ (y_1, \ldots, y_n) \in \mathbf{R}^n_{\geq 0} \;:\; y_1 \leq y_2 \leq \ldots \leq y_n,\ \sum_{j=1}^{n-1} y_j = y_n \right\} \subset \mathbf{R}^n_{\geq 0} \tag{1.5}$$

and that 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 $\mathcal{H}$.

Example 7. When $n = 3$, the linear map $\sigma : \mathcal{H} \to \mathcal{H}$ defined by

$$\sigma(a, b, a+b) = \mathsf{order}\,(b - a, a, b)\,, \tag{1.6}$$

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 $a \leq b \in \mathbf{Z}$ 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. This is exploited in Zagier's paper [Zag82].

[4]See our discussion in Section 3.1 about the benefits of this replacement. It is inspired by the 'Time Acceleration Machine' described by Zorich in [Zor06, Section 5.3].

[5]This follows from a similar argument to the proof of Lemma 20 we give below.

We can use the basis

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

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

Example 8. When $n = 3$ the semigroup Γ' is generated by the

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

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

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

Example 9. When $n = 4$ the semigroup Γ acts in the basis given by the e_i as

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

This semigroup appears naturally in different areas of mathematics. In most situations that this semigroup appears, as will also be the case in this paper, the dynamics of the projective linear action of Γ on $\mathbf{R}_+^3/\mathbf{R}_+$ 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 $\mathbf{R}_+^3/\mathbf{R}_+$ has a fractal attracting set that is known as the *Rauzy gasket.*

The Rauzy gasket first appears in the literature in a paper of Levitt [Lev93] in connection with the dynamics of partially defined rotations of the circle. The Rauzy gasket has been rediscovered by different groups of mathematicians, including De Leo and Dynnikov [DLD09] in connection to a conjecture of Novikov [Nov82] on triply periodic surfaces, Arnoux and Starosta [AS13] (wherein the Rauzy gasket was given its name) in relation to generalizations of Sturmian words to three letters and the 'fully subtractive' continued fractions algorithm, and now, in this paper, in connection to Diophantine geometry.

The Rauzy gasket was proven by Avila, Hubert, and Skripchenko [AHS16b] to have Hausdorff dimension less than 2, answering a question of Arnoux. The acceleration, replacing Γ by Γ', that we perform here is also carried out (in the context of iterated function systems) by Arnoux and Starosta [AS13] and Avila, Hubert, and Skripchenko [AHS16b], where the acceleration is viewed as analogous to Zorich's acceleration (see [Zor06, Section 5.3]) of Rauzy-Veech induction that is well known in Teichmüller dynamics.

It is also worth pointing out that higher dimensional versions of the Rauzy gasket have been defined [AS13, De 08], and the branches of the corresponding iterated function system, after the same simple modifications as before, match with our Γ for $n > 4$.

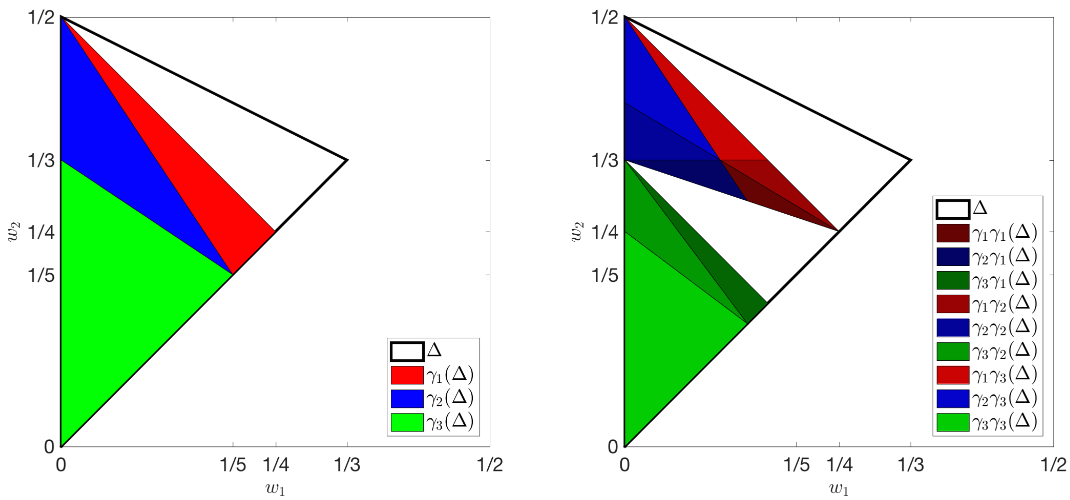

Figure 1: When $n = 4$, the semigroup elements map $\Delta = \mathcal{H}/\mathbf{R}_+$ into a strictly smaller subset. After iteration this leads to more and more empty space (see also Figure 2). This doesn't occur when $n = 3$, as one can also see from the picture: the action of the group elements γ_2 and γ_3 on the vertical coordinate axis is a copy of the $n = 3$ dynamics.

Some of our technical results in Sections 4 and 5 can be closely compared to, intersect with, or generalize, results obtained by Avila, Hubert, and Skripchenko for the Rauzy gasket in [AHS16a, AHS16b]. We point out these intersections throughout the paper.

So our semigroups Γ and Γ' are natural extensions of the Euclidean algorithm and continued fractions semigroup to higher dimensions[6]. We write $\Delta = \mathcal{H}/\mathbf{R}_+$ and we can view Δ as a subset of $\mathbf{R}^{n-2}$ (see Section 5 for details). 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 $\mathcal{H}/\mathbf{R}_+$ has 'holes' as we illustrate in Figure 1.

We get a new characterization of the parameter β in terms of the action of Γ' on $\mathcal{H}/\mathbf{R}_+$.

Theorem 10. *The β from Theorem 2 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.

Remark 11. 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 [Pat76, Sul79, Sul84]. In our setting, the

[6] See [Zor06] for the discussion of such an extension in the context of translation surfaces.

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.

De Leo has conjectured in [DL15, Conjecture 1] that if δ is the Hausdorff dimension of the Rauzy gasket (see Example 9 and the remark below) then $\delta \geq \frac{2}{3}\beta(4)$. By Baragar's estimate (1.3), this conjecture would imply $\delta > 1.62$. De Leo and Dynnikov [DLD09] have numerically estimated the box-counting dimension of the Rauzy gasket to be in the range $[1.7, 1.8]$, which implies $\delta \leq 1.8$.

Remark 12. In the case of $n = 4$, the measure ν_β is essentially the same as the measure obtained for the Rauzy gasket by Avila, Hubert, and Skripchenko in [AHS16a, Theorem 1] in the context of a problem of Novikov [Nov82] on triply periodic surfaces.

In Section 3.4 we reduce Theorem 3 to a counting theorem for orbits of the semigroup Γ'. The relevant counting quantity is defined by

$$N(y,r) \equiv \sum_{\gamma \in \Gamma' \cup \{e\}} \mathbf{1}\{\log(\gamma.y)_n - \log(y)_n \leq r\} \tag{1.7}$$

for $y \in \mathcal{H} - 0$ and $r \geq 0$. Here we use the notation $(\gamma.y)_n$ for the nth entry of the vector $\gamma.y$. We prove

Theorem 13. *There is a positive bounded C^1 function h on $\mathcal{H}$ that is invariant under the action of $\mathbf{R}_+$ and such that*

$$N(y,r) = h(y)e^{\beta r}(1 + o_{r\to\infty}(1))$$

for all $y \in \mathcal{H} - 0$, where the implied function in the small o does not depend on y. Moreover, h satisfies the recursion

$$\sum_{\gamma \in T_\Gamma} \left(\frac{(\gamma.y)_n}{y_n}\right)^{-\beta} h(\gamma.y) = h(y). \tag{1.8}$$

The constant β is the same as in Theorem 2.

Remark 14. The embedding of the $(n-1)$-dimensional version of $\mathcal{H}$ inside the n-dimensional version implies by Theorem 13 that $\beta(n) \geq \beta(n-1)$ and in particular that $\beta(n) \geq 2$ for all $n \geq 3$.

1.1 Connection to simple closed curves and character varieties

Theorem 1 can be rephrased as a counting result for the number of simple[7] closed geodesics of length $\leq \log R$ on the modular torus. This is the topological once-punctured torus that is uniformized by the quotient of the hyperbolic plane by the group

$$\left\langle \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}, \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} \right\rangle \leq \mathrm{PSL}_2(\mathbf{R}).$$

McShane and Rivin [MR95] actually obtain the analogous counting result to Theorem 1 for simple closed geodesics on arbitrary hyperbolic once punctured tori, by use of a special norm

[7]This means there are no self crossings.

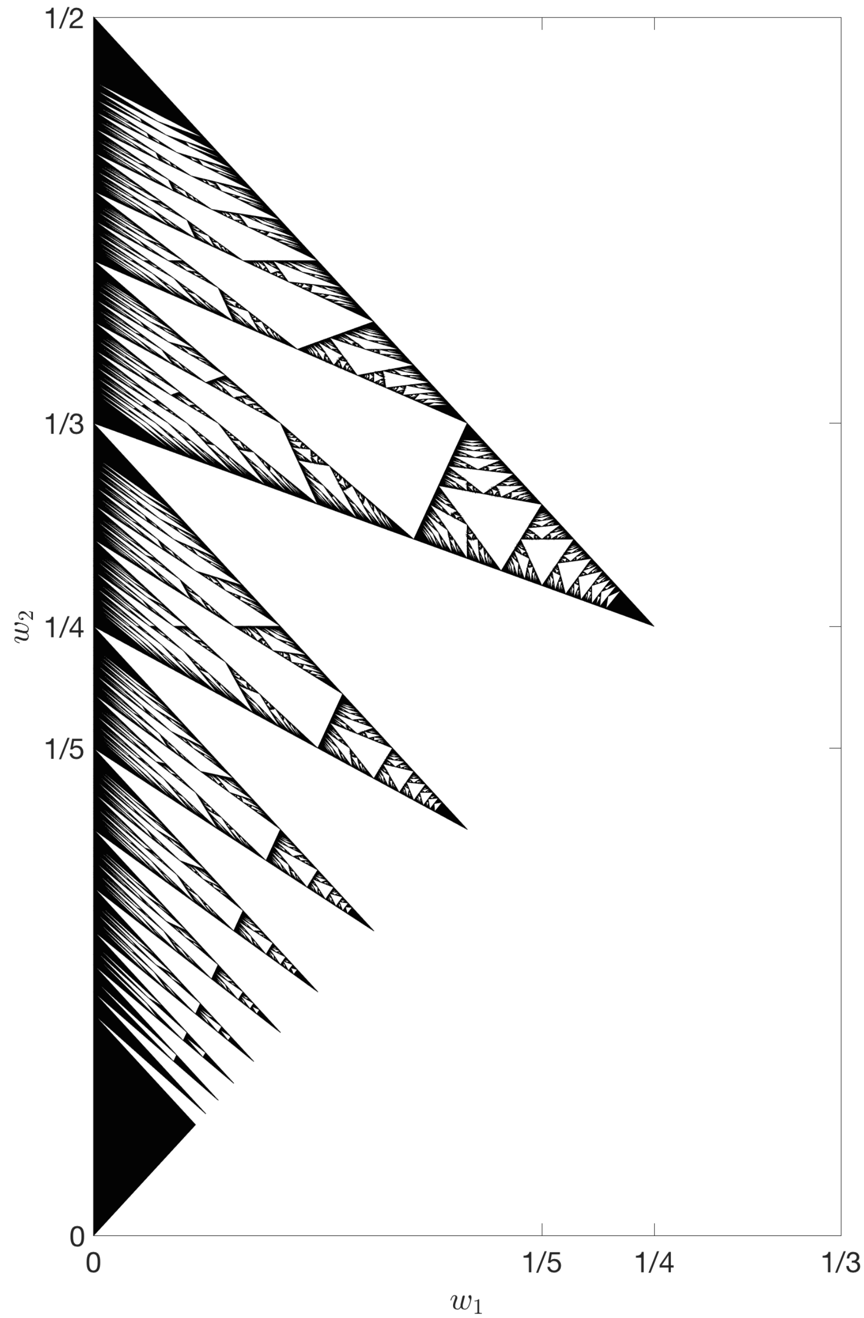

Figure 2: In the same setting ($n = 4$) of Figure 1, we show in black the images of Δ under the action of all words of length 10 in the generators $\{\gamma_1, \gamma_2, \gamma_3\}$.

on the first homology of the surface. Mirzakhani proved in [Mir08] an asymptotic counting result, without explicit error term, for simple closed geodesics on any finite area complete Riemann surface. These asymptotics have recently been extended by Mirzakhani [Mir16] to more general orbits of the mapping class group. In Mirzakhani's results the exponents of growth are dimensions of Teichmüller spaces. It is interesting to compare this to our characterization of Theorem 10.

In [HN17], Huang and Norbury showed that when $n = a = 4$ and $k = 0$, $V(\mathbf{R}_+)$ is a parametrization of the Teichmüller space of finite area hyperbolic structures on $\mathbf{R}P^2$ minus three points, and moreover the coordinates of points on $V(\mathbf{R}_+)$ are functions of the lengths of *one-sided*[8] simple closed geodesics in the relevant hyperbolic structure. From these facts they deduce from Baragar's Theorem 2 that the number $n_J^{(1)}(L)$ of one sided simple closed geodesics of length $\leq L$ in a hyperbolic structure J on $\mathbf{R}P^2$ minus three points satisfies

$$\lim_{L\to\infty} \frac{\log n_J^{(1)}(L)}{\log L} = \beta(4).$$

The second author (Magee) of this paper has recently shown [Mag18] that the methods here can be extended to prove that $n_J^{(1)}(L)$ is asymptotic to cL^{β}, for some $c = c(J) > 0$, somewhat in analogy to Mirzakhani's results.

We also mention the recent work of Hu, Tan and Zhang [HPZ18] that describes some regions in $\mathbf{C}^n$ where the group of automorphisms of (1.1) acts properly discontinuously. This extends previous work of Goldman [Gol03] that describes ranges of k in the case of $n = 3$ where the group $\mathrm{Aut}(V)$ act ergodically or properly discontinuously (or some combination thereof, on different components of the variety). Quite strikingly, for certain ranges of k the action of $\mathrm{Aut}(V)$ is ergodic on $V(\mathbf{R})$ yet preserves the infinite discrete subset $V(\mathbf{Z})$. In [HPZ18] the authors also prove a 'McShane identity' that gives an expression for the constant 1 in terms of an infinite sum over any orbit of the semigroup; see [McS91, McS98] for McShane's original identity.

1.2 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 3. It is illuminating to recall the methods used by Lalley in [Lal89] where the action of a Schottky subgroup G of $\mathrm{SL}_2(\mathbf{R})$ on the hyperbolic upper half plane $\mathbb{H}$ is considered. Lalley obtains in [Lal89, Theorem 9] 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 $\partial\mathbb{H}$.

[8] This means a thickening of the geodesic is homeomorphic to a Möbius band.

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 the $\mathcal{N}^*(y^*, r')$ are obtained using the renewal method and spectral estimates for transfer operators. This gives asymptotic formulas for the $\mathcal{N}(y, r')$. 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 the $\mathcal{N}(y, r')$ to obtain an asymptotic formula for $\mathcal{N}(x, r)$. This is done using the recursive formulas obtained in the previous step.

Now let us explain our methods at a high level, comparing them to Lalley's technique. In Section 2.1 we describe the passage from $V(\mathbf{Z})$ to $V(\mathbf{Z}_+)$ and describe in full the action of the Markoff-Hurwitz moves on $V(\mathbf{Z}_+)$. We also explain in Section 2.1 that outside a large compact region of $V(\mathbf{Z}_+)$, there are finitely many orbits of the group generated by Markoff-Hurwitz moves and each of these orbits are well understood. Following this, in Section 2.3 we reduce the proof of Theorem 3 to the problem of counting in an orbit of the non-linear semigroup Λ on ordered tuples of positive real numbers satisfying (2.5).

To try 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 $\mathcal{H}$ discussed in the Introduction; we call this set Δ. We compare points in the orbit of Λ 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 $\alpha(z)$ defined in (3.2) 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 by 'acceleration' as mentioned in the Introduction, 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 off. 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. We need to control the size of such an intersection, which is done in Lemma 25. The second issue is that the original Λ-orbit breaks up 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. The recombination phase of our argument for Λ' takes place in Section 3.5. The extra summation is dealt with earlier in Section 3.1.

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 from the Introduction. 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 $y \in \mathcal{H}$, 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. A precise version of this statement is given in Lemma 28.

By the end of Section 3 we have 'passed to the boundary' by reducing counting in Λ'-orbits to the counting estimate of Theorem 13 for the linear semigroup Γ', and hence have proved Theorem 3 modulo the deferred proof of Theorem 13.

In Section 4, we prove Theorems 10 and 13. The proofs rely 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, which we state precisely in Proposition 45. This result was established for $n = 4$ by Avila, Hubert, and Skripchenko in [AHS16b]. We explain the proof of the result for general $n \geq 4$ in Section 5. 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. This is established in Proposition 42. Finally, but importantly, to obtain the statement of Theorem 13, which was the input into the recombination phase of the argument, and as such must have uniformity over $y \in \mathcal{H}$, we must obtain spectral estimates for transfer operators acting on $C^1(\Delta)$. This means that we cannot work with a symbolic model as was done in [AHS16a] for $n = 4$, and rather, we follow Liverani's approach to spectral estimates from [Liv95].

1.3 Notation

For the reader's convenience we describe the notation we use in this paper. We will use $\mathbf{1}$ for an indicator function. A vector with an entry $\hat{\bullet}$ with a hat means that that entry is omitted. We use Vinogradov notation $O, o, \ll, \gg$ in the standard way. Any implied constants may depend on n, a, k that we view as fixed throughout much of the paper. If there is any dependence of an implied constant on a variable we denote this as a subscript e.g. $\ll_\epsilon$, and we also use subscripts to indicate which variable is tending to a limit, e.g. $o_{a\to\infty}(1)$. For the sake of convenience, we take the liberty of applying functions to vectors, which means we apply the function component-wise, and we write inequalities between vectors to mean that the inequality holds at every component. For a set S in a semigroup we may write $S^{(k)}$ for the k-fold product of the set with itself. We also write $\mathbf{R}_+, \mathbf{R}_{\geq 0}$ for the sets of positive (resp. nonnegative) real numbers, and similar for integers. We write $\{x\}$ for the fractional part of a real number x, that is, $x = n + \{x\}$ for $n \in \mathbf{Z}$ and $0 \leq \{x\} < 1$. For $N \in \mathbf{N}$ we use the notation $[N]$ for the set $\{1, 2, \ldots, N\}$.

Acknowledgements

We would like to thank Peter Sarnak, Giulio Tiozzo, and Peter Whang for helpful conversations about this work. We are grateful to the anonymous referees for their comments that have improved the paper.

2 Markoff-Hurwitz tuples and moves

2.1 Basic properties of the Markoff-Hurwitz equation

The automorphism group

By an automorphism of V we mean a polynomial automorphism of $V(\mathbf{C})$. We write $\mathrm{Aut}(V)$ for the group of all such maps. By results of Horowitz [Hor75] when $n = 3$ and Hu, Tan and Zhang [HPZ18, Theorem 1.1] for $n \geq 4$, one has

$$\mathrm{Aut}(V) = \mathcal{G} \rtimes (N \rtimes S_n)$$

where

1. N is the group of transformations that change the sign of an even number of variables. Hence $|N| = 2^{n-1}$.
2. S_n is the symmetric group on n letters that acts by permuting the coordinates of $\mathbf{C}^n$.
3. $\mathcal{G}$ is the nonlinear group generated by the Markoff-Hurwitz moves m_j discussed in the Introduction.

One important corollary of this classification is that $V(\mathbf{Z})$ is invariant under $\mathrm{Aut}(V)$.

Exceptional solutions

For $a = 1$ and $a = 2$ there are certain exceptional families of points in $V(\mathbf{Z})$ whose growth rate is different from the points we wish to count[9]. These appear only for certain values of k and we describe them now.

Definition 15. We say that $x \in V(\mathbf{Z}_+)$ is a *fundamental exceptional solution* if it belongs to one of the following two families

1. One has $a = 1$ and after reordering the coefficients of x so that $x_1 \leq x_2 \leq \ldots \leq x_n$
$$x_1 = x_2 = \ldots = x_{n-3} = 1, \quad x_{n-2} = 2.$$
In this case x is a Markoff-Hurwitz tuple if and only if
$$(x_{n-1} - x_n)^2 = k - n - 1. \tag{2.1}$$
2. One has $a = 2$ and after reordering the coefficients of x so that $x_1 \leq x_2 \leq \ldots \leq x_n$
$$x_1 = x_2 = \ldots = x_{n-2} = 1.$$
In this case x is a Markoff-Hurwitz tuple if and only if
$$(x_{n-1} - x_n)^2 = k - n + 2. \tag{2.2}$$

[9]See Silverman [Sil89] for a discussion of a phenomenon of surfaces containing curves that have many more integral points than one would expect from the surface as a whole.

We say that $x \in V(\mathbf{Z})$ is an *exceptional solution* if x is in the $\mathrm{Aut}(V)$-orbit of a fundamental exceptional solution. We write $\mathcal{E}$ for the collection of exceptional solutions in $V(\mathbf{Z})$. If $x \in V(\mathbf{Z})$ is not an exceptional solution we say it is an *unexceptional solution*.

Note that if (2.1) or (2.2) occur then they occur in an infinite family for that given n, a, k. In either case, all sufficiently large positive integers appear as the maximal entry of some fundamental exceptional solution and this maximal entry determines the tuple up to reordering. Therefore for some $c > 0$ there are $cR + O(1)$ fundamental exceptional solutions with maximal entry $\leq R$. It is also clear, but useful to note, that the property of being exceptional (respectively, unexceptional) in $V(\mathbf{Z})$ is $\mathrm{Aut}(V)$-invariant.

The following two interesting examples concerning exceptional solutions were pointed out to us by an anonymous referee, to whom we are grateful.

Example 16. When $n = 3, a = 1$, and $k = 4$, V is *Cayley's cubic surface* [Cay69]. In this case, one has a parametric family of fundamental exceptional solutions given by

$$\mathcal{C}(t) = (2, t, t)$$

that corresponds to a rational curve $\mathcal{C} \subset V$. Letting T_n denote the nth Chebyshev polynomial of the first kind, we have $\mathcal{C}(t) = \left(2T_0\left(\frac{t}{2}\right), 2T_1\left(\frac{t}{2}\right), 2T_1\left(\frac{t}{2}\right)\right)$. Let us view this as a point in $V(\mathbf{Z}[t])$. More generally, for $h, i, j \in \mathbf{Z}_{\geq 0}$ such that the largest of h, i, j is equal to the sum to the other two, we have

$$\left(2T_h\left(\frac{t}{2}\right), 2T_i\left(\frac{t}{2}\right), 2T_j\left(\frac{t}{2}\right)\right) \in V(\mathbf{Z}[t])$$

by the identity $\cos^2(A)+\cos^2(B)+\cos^2(A+B) = 2\cos(A)\cos(B)\cos(A+B)+1$. One can check using further trigonometric identities that the Markoff-Hurwitz moves preserve these points. Hence it follows that the polynomials that appear in the orbit of $\mathcal{C}(t)$ under Markoff-Hurwitz moves are all of the form $2T_i\left(\frac{t}{2}\right)$ for $i \in \mathbf{Z}_{\geq 0}$.

Example 17. When $n = 4, a = 2$, and $k = 2$, again V has a parametric family of fundamental exceptional solutions given by

$$\mathcal{C}'(t) = (1, 1, t, t) \in V(\mathbf{Z}[t]).$$

It was pointed out to us by an anonymous referee that using the methods of the current paper, it is possible to prove that the number of points in $V(\mathbf{Z}[t])$ in the orbit of $\mathcal{C}'(t)$ under the Markoff-Hurwitz moves, all of whose coordinate polynomials have degree $\leq D$, is asymptotic to $cD^{\beta(4)}$ for some $c > 0$. This fact, its generalizations, and its detailed proof, will be pursued elsewhere.

Passage from $V(\mathbf{Z})$ to $V(\mathbf{Z}_+)$

We now describe the relationship between asymptotic counting of $V(\mathbf{Z}) - \mathcal{E}$ and $V(\mathbf{Z}_+) - \mathcal{E}$. Recall that $n \geq 3$, $a \geq 1$ and k are fixed integers, and N is the group of automorphisms of $V = V_{n,a,k}$ that change the sign of an even number of the coordinates. We decompose the action of N on $V(\mathbf{Z}) - \mathcal{E}$ as follows.

Let X_0 be the elements of $V(\mathbf{Z}) - \mathcal{E}$ with at least one coordinate equal to 0. If $k < 0$ then X_0 is empty, and if $k \geq 0$ then one obtains for $(x_1, \ldots, x_n) \in X_0$ the equation

$$x_1^2 + \ldots + x_n^2 = k$$

from which it is apparent that X_0 is finite, with a bound on its size depending on n and k. To indicate this we write $|X_0| = O_{n,k}(1)$.

Now let $X(R) = (V(\mathbf{Z}) - \mathcal{E} - X_0) \cap B(R)$, the unexceptional elements of $V(\mathbf{Z})$ with norm $\leq R$ and no zero coordinate. The group N acts freely on $X(R)$. Therefore

$$2^{n-1}|N\backslash X(R)| = |X(R)|.$$

The orbits of N on $X(R)$ fall into two categories, according to which we decompose

$$N\backslash X(R) = Y_+(R) \sqcup Y_-(R)$$

where $Y_+(R)$ are orbits with a unique representative with all coordinates positive, and $Y_-(R)$ the remaining orbits, which have a unique representative with $x_1 < 0$ and $x_i > 0$ for $i \geq 2$.

We now argue that $|Y_-(R)|$ is bounded independently of R. To see this, consider $N.x \in Y_-(R)$, where x is the representative described before with x_1 the only negative coordinate. Let $\tilde{x}_1 = -x_1$ and $\tilde{x}_i = x_i$ for $i \geq 2$ be the coordinates of $\tilde{x}$. The parametrization $x \to \tilde{x}$ is obviously 1:1 and

$$\tilde{x}_1^2 + \ldots \tilde{x}_n^2 + a\tilde{x}_1\tilde{x}_2 \ldots \tilde{x}_n = k.$$

Because all the $\tilde{x}_i > 0$ and $a \geq 1$, this equation has no solutions when $k \leq 0$ and only finitely many when $k > 0$, with a bound depending only on n and k. In any case, this shows $|Y_-(R)| = O_{n,k}(1)$.

Since $Y_+(R)$ is parametrized 1:1 by $(V(\mathbf{Z}_+) - \mathcal{E}) \cap B(R)$, the previous arguments combine to show

$$\begin{aligned} |(V(\mathbf{Z}) - \mathcal{E}) \cap B(R)| = |X(R)| + |X_0 \cap B(R)| &= 2^{n-1}|N\backslash X(R)| + O_{n,k}(1) \\ &= 2^{n-1}(|Y_+(R)| + |Y_-(R)|) + O_{n,k}(1) \\ &= 2^{n-1}|(V(\mathbf{Z}_+) - \mathcal{E}) \cap B(R)| + O_{n,k}(1). \end{aligned}$$

Infinite descent

The following proposition says that outside of a compact set, the effects of the moves m_i on the maximal entries of unexceptional Markoff-Hurwitz tuples are at least somewhat predictable. This is a very special feature of the Diophantine equation (1.1) that will allow us to count solutions.

Proposition 18. *Suppose $k \in \mathbf{Z}$. There is a compact set $K_0 = K_0(n, a, k)$ such that for unexceptional $x \in V(\mathbf{Z}_+) - K_0$ the following hold:*

1. *If x_j is the largest coordinate of x then the largest entry of $m_j(x)$ is smaller than x_j, that is, $(m_j(x))_i < x_j$ for all i.*

2. *The largest entry of x appears in exactly one coordinate.*

3. *If x_j is not the largest coordinate of x then it becomes the largest after the move m_j, that is, $(m_j(x))_j > (m_j(x))_i$ for all $i \neq j$. (This property holds for all $x \in V(\mathbf{Z}_+)$.)*

4. *If x_j is not the largest coordinate of x, then the number of distinct entries of $m_j(x)$ is at least the number of distinct entries of x. In particular, if x has distinct entries then $m_j(x)$ has distinct entries.*

5. *Every move m_j maps $V(\mathbf{Z}_+) - K_0$ into $V(\mathbf{Z}_+)$.*

The compact K_0 can be taken to be a closed ball about the origin in the ℓ^∞ norm on $\mathbf{R}^n$, and the result still holds after increasing the radius of K_0.

Proof of Proposition 18. Part 1. Suppose without loss of generality that $x_1 \leq x_2 \leq \ldots \leq x_{n-1} \leq x_n$. Adapting a proof of Cassels from [Cas57, pg. 27], consider the quadratic polynomial in x_n given by

$$f(T) = T^2 - ax_1x_2\ldots x_{n-1}T + x_1^2 + x_2^2 + \ldots + x_{n-1}^2 - k.$$

Then f has roots at x_n and x_n' where x_n' is the last entry of $m_n(x)$. The conclusion of Part 1 holds unless

$$x_{n-1} \leq x_n \leq x_n'$$

or

$$x_n' < x_{n-1} = x_n.$$

In either case, since the coefficient of T^2 is positive it follows that $f(x_{n-1}) \geq 0$. Then

$$\begin{aligned} 0 \quad &\leq \quad f(x_{n-1}) = -ax_1x_2\ldots x_{n-1}^2 + x_1^2 + x_2^2 + \ldots + 2x_{n-1}^2 - k. \\ &\leq \quad (n - ax_1x_2\ldots x_{n-2})x_{n-1}^2 - k \end{aligned}$$

implying

$$ax_1x_2\ldots x_{n-2} \leq n - \frac{k}{x_{n-1}^2} \leq n + |k|.$$

This means there are a finite number of possibilities for $x_1, x_2, \ldots, x_{n-2}$.

In the case $x_n' \geq x_n$ one has

$$ax_1x_2\ldots x_{n-2}x_{n-1} - x_n \geq x_n$$

so

$$ax_1x_2\ldots x_{n-2}x_{n-1}x_n \geq 2x_n^2.$$

Then from (1.1)

$$x_n^2 \leq x_1^2 + \ldots + x_{n-2}^2 + x_{n-1}^2 - k$$

and it follows that

$$(x_n + x_{n-1})(x_n - x_{n-1}) \leq x_1^2 + \ldots + x_{n-2}^2 - k.$$

If $x_n - x_{n-1} > 0$ then the finite number of possibilities for $x_1, x_2, \ldots, x_{n-2}$ yield a finite number of possible x.

The alternative is that $x_n = x_{n-1}$, and the following logic also applies to the case $x'_n < x_{n-1} = x_n$. Then x_n is a root of one of finitely many quadratic polynomials

$$(2 - ax_1 \dots x_{n-2})x_n^2 + x_1^2 + \dots + x_{n-2}^2 - k = 0.$$

Again, this yields finitely many possibilities for x aside from those having $x_1, \dots, x_{n-2}$ such that $2 - ax_1 \dots x_{n-2} = 0$ and $x_1^2 + \dots + x_{n-2}^2 - k = 0$. Note that if $k \leq 0$ we have exhausted the possibilities. Otherwise we must have either $a = 1$ and $k = (n-3)1 + 4$ in which case

$$x_1 = x_2 = \dots = x_{n-3} = 1, \quad x_{n-2} = 2,$$

or $a = 2$ and $k = n - 2$, in which case

$$x_1 = x_2 = \dots = x_{n-2} = 1.$$

These are precisely the fundamental exceptional solutions that are ruled out by hypothesis. Therefore for any given n, a, k only finitely many unexceptional x do not satisfy Part 1 of the Proposition.

Part 2. If the largest entry of x is not unique then performing the move at one of the largest entries does not decrease the largest entry, contradicting Part 1.

Part 3. Suppose $x_1 \leq x_2 \leq \dots < x_n$ and let $x' = (x'_1, \dots, x'_n) = m_j(x)$ with $j < n$. The coefficient x'_j satisfies

$$x'_j - x_n = a \prod_{i \neq j} x_i - x_j - x_n = x_n \left(a \prod_{i \neq j,n} x_i - 1 \right) - x_j.$$

If $a \geq 2$ then the right hand side is $\geq x_n - x_j > 0$ so we are done. If $a = 1$ and $x_{n-2} \geq 2$ then we are also done by a similar argument.

The remaining scenario is $a = 1$ and $x_1 = x_2 = \dots = x_{n-2} = 1$. In this case x satisfies the equation

$$x_{n-1}^2 + x_n^2 - x_{n-1}x_n = k - n + 2.$$

The form on the left hand side is positive definite so only finitely many possible solutions exist for (x_{n-1}, x_n) given n and k. Add these to the compact set of Part 1.

Part 4. This follows from Part 3 since if $x' = m_j(x)$ as in the Proposition, then all the entries of x'_i with $i \neq j$ are distinct, but x'_j is larger than all of these.

Part 5. By Part 3 it suffices to check that we can increase the radius of K_0 so that for $x \in V(\mathbf{Z}_+) - K_0$ with $x_1 \leq x_2 \leq \dots \leq x_n$, $m_n(x)_n > 0$. If not, one obtains $ax_1 \dots x_{n-1} - x_n \leq 0$ from which it follows $ax_1 x_2 \dots x_n \leq x_n^2$. The Markoff-Hurwitz equation then gives

$$x_1^2 + \dots + x_{n-1}^2 \leq k. \tag{2.3}$$

By an easy argument (cf. Section 2.4) it is possible to increase the radius of K_0 so that for $x \in V(\mathbf{Z}_+) - K_0$ ordered as we assume, $x_{n-1} \geq \left(\frac{x_n}{2a}\right)^{\frac{1}{n-1}}$. In particular, we can increase the radius of K_0 so that under the ongoing assumptions on x, $x_{n-1}^2 > |k|$. It follows then that (2.3) cannot occur outside of K_0. $\square$

Corollary 19 (Infinite descent)**.** *Any unexceptional Markoff-Hurwitz tuple can be algorithmically reduced to one in the compact set K_0 by a sequence of Markoff-Hurwitz moves that strictly decrease maximal entries.*

Corollary 19 was established by Markoff [Mar80] in the case $n = a = 3$ and $k = 0$. In that case, every Markoff triple can be reduced to $(1, 1, 1)$ by a series of Markoff moves. Hurwitz [Hur07] showed the analogous result for $n = a > 3$ and $k = 0$ and showed more generally that when $k = 0$, the Markoff-Hurwitz tuples can be reduced to a finite set of fundamental solutions. These fundamental solutions were characterized by Baragar in [Bar94b] whenever $a \geq 2(n-1)^{1/2}$; he also presented two different constructions yielding sequences of equations whose sets of fundamental solutions grow without bound.

In the case $n = a = 3$, recent work [GS17] of Ghosh and Sarnak gives much more refined information than Corollary 19 for a wide range of k. For example, when $n = a = 3$ and $k < 0$ with k not congruent to 4 or 5 modulo 9, Ghosh and Sarnak prove [GS17, Theorem 1.1(ii)] that there is an explicit compact *fundamental set* $\mathcal{I}_k \subset \mathbf{R}^3$ such that every orbit of $\mathrm{Aut}(V)$ on $V(\mathbf{Z})$ contains a unique element of $\mathcal{I}_k \cap \mathbf{Z}^3$. They also prove a similar statement for arbitrary $k \geq 5$ [GS17, Theorem 1.1(i)].

2.2 The polynomial semigroup

We now perform a normalization that allows us to treat all parameters a, k with a semigroup action that only depends on n. For $x \in V(\mathbf{Z}_+)$ let

$$z = z(x) = a^{\frac{1}{n-2}} x \tag{2.4}$$

Note that $a^{\frac{1}{n-2}} \geq 1$ with equality if and only if $a = 1$. Then $z = (z_1, \ldots, z_n)$ satisfies the equation

$$z_1^2 + z_2^2 + \ldots + z_n^2 = z_1 z_2 \ldots z_n + k' \tag{2.5}$$

where

$$k' = k a^{\frac{2}{n-2}}.$$

Say that z is exceptional/unexceptional if x has the corresponding property. We will also work with ordered tuples z so that

$$z_1 \leq z_2 \leq \ldots \leq z_n.$$

Write $\mathcal{M}$ for the set of all such ordered tuples $z \in a^{\frac{1}{n-2}} \mathbf{Z}_+^n$ satisfying (2.5). Counting

$$\mathcal{M} \cap B(R)$$

is not equivalent to counting $V(\mathbf{Z}_+) \cap B(a^{-\frac{1}{n-2}} R)$ due to the presence of elements with duplicate entries. We will return to treat this point in Section 2.3. Let

$$K \quad = \quad a^{\frac{1}{n-2}} K_0 \tag{2.6}$$

where K_0 is the compact set from Proposition 18.

The Markoff-Hurwitz moves $\{m_j\}$ induce the moves

$$\lambda_j(z_1,\ldots,z_n) = \left(z_1,\ldots,\widehat{z_j},\ldots,z_n,\prod_{i\neq j} z_i - z_j\right), \quad 1\leq j\leq n-1, \tag{2.7}$$

where $\widehat{\bullet}$ denotes omission[10]. Since K is a closed ball about 0 in the ℓ^∞ norm, Part 3 of Proposition 18 implies that the $\{\lambda_j\}$ preserve $\mathcal{M}-K$. Let $\Lambda=\Lambda(n)$ denote the semigroup of piecewise polynomial self-maps of $\mathbf{C}^n$ generated by the λ_j. In Section 2.3 we will reduce Theorem 3 to an orbital counting estimate. For $z_0\in\mathcal{M}-K$ let

$$\Lambda.z_0\subset\mathcal{M}-K$$

denote the orbit of z_0 under Λ.

Lemma 20. *If $z_0\in\mathcal{M}-K$ has distinct entries then the map $\Lambda\to\mathcal{M}-K$ given by*

$$\lambda\mapsto\lambda(z_0)$$

is injective. It follows that the semigroup Λ is free[11] *on the generators $\{\lambda_j\}$.*

Proof. For the first part, if the map is not injective then at some point there must be $\lambda_1\in\Lambda$ and some $j_1\neq j_2$ such that

$$\lambda_{j_1}\lambda_1(z_0)=\lambda_{j_2}\lambda_1(z_0). \tag{2.8}$$

Since by Proposition 18, Part 4 the entries of $\lambda_1(z_0)$ are distinct we find $z=\lambda_1(z_0)$ with distinct entries so that $\lambda_{j_1}z=\lambda_{j_2}z$. But this cannot be the case since e.g. the sets $\{z_1,\ldots,\widehat{z_{j_1}},\ldots,z_n\}$ and $\{z_1,\ldots,\widehat{z_{j_2}},\ldots,z_n\}$ are not the same.

For the second part it is enough to find some a and k so that there is a point z_0 in $\mathcal{M}-K$ with all entries distinct. Given this point, the freeness of Λ follows from the first part of the proof applied to z_0. To give an explicit example of a point with these properties, given n, if we let $k=\sum_{j=1}^{n-1} j^2$, and $a=2$, then $x=(1,2,\ldots,n-1,2(n-1)!)$ is in $V(\mathbf{Z}_+)$. Let $z=z(x)$. For sufficiently large A, $z_0=\lambda_{n-1}^A(z)$ is in $\mathcal{M}-K$ with distinct entries. □

2.3 Multiplicities

In the rest of the paper we will count in orbits of the free semigroup Λ. It is extremely useful to be able to work with a fixed free semigroup for each n. The cost of this, however, is that Λ acts on ordered tuples. Since the original problem was to count points in $V(\mathbf{Z}_+)$ we therefore need to take into account the multiplicity of the order map $V(\mathbf{Z}_+)\to\mathcal{M}$.

This is best done in relation to the moves m_j. Given $x\in V(\mathbf{Z}_+)-K_0$, we say that a sequence

$$j_1,j_2,j_3,\ldots,j_l,\ldots$$

is *admissible* for x if for all l, j_l is not the index of the largest coordinate of

$$x^{(l-1)}=m_{j_{l-1}}m_{j_{l-2}}\ldots m_{j_2}m_{j_1}x.$$

[10] The reason we now have $n-1$ moves instead of n is that we never perform the move that will decrease the maximal entry, therefore moving us towards K. This eliminates backtracking from our 'random walk'.

[11] As a semigroup of polynomial maps.

Notice then that by Proposition 18, Part 3, the largest entries of $x^{(l)}$ are increasing in l and therefore $x^{(l)} \in V(\mathbf{Z}_+) - K_0$ for all $l \geq 1$. Also, a sequence is admissible if and only if j_1 is not the largest coordinate index of x and $j_l \neq j_{l-1}$ for any $l \geq 2$. Write $\Sigma^*(x)$ for the set of all finite admissible sequences for x.

Lemma 21. *Given $x \in V(\mathbf{Z}_+) - K_0$ the map $\phi_x : \Sigma^*(x) \to V(\mathbf{Z}_+)$ given by*

$$\phi_x(j_1, j_2, j_3, \ldots, j_l) = m_{j_l} m_{j_{l-1}} m_{j_{l-2}} \ldots m_{j_2} m_{j_1} x$$

is injective. Note that this is regardless of whether x has duplicate entries. Moreover, for any x, $x' \in V(\mathbf{Z}_+) - K_0$, the images of ϕ_x and $\phi_{x'}$ are disjoint unless either $x' \in \mathrm{image}(\phi_x)$ or $x \in \mathrm{image}(\phi_{x'})$.

Proof. It is clear from Proposition 18, Part 3 that the $m_{j_1} x$ with j_1 admissible are distinct. To show that ϕ_x is injective, it is enough to show that there are no $x \neq x' \in V(\mathbf{Z}_+) - K_0$ and j, j' admissible for the respective x, x' so that $m_j(x) = m_{j'}(x')$. But since $m_j(x)$ has a distinct largest entry by Proposition 18 Part 2, it has to be the case that $j = j'$. Then applying m_j gives $x = x'$.

Now suppose $x' \notin \mathrm{image}(\phi_x)$ and $x \notin \mathrm{image}(\phi_{x'})$. If $\mathrm{image}(\phi_x) \cap \mathrm{image}(\phi_{x'}) \neq \emptyset$ then at some point there must have been $x^{(3)} \neq x^{(4)} \in V(\mathbf{Z}_+) - K_0$ and j, j' admissible for $x^{(3)}, x^{(4)}$ respectively so that $m_j(x^{(3)}) = m_{j'}(x^{(4)})$. But we have already established this cannot happen. $\square$

Lemma 22. *Let $x \in V(\mathbf{Z}_+) - K_0$ and $z = \mathsf{order}(a^{\frac{1}{n-2}} x)$ the corresponding element of $\mathcal{M} - K$. There exists a bijection*

$$\Theta_x : \Sigma^*(x) \to \Lambda$$

that is an intertwiner for the map $x' \mapsto z(x') = \mathsf{order}(a^{\frac{1}{n-2}} x')$ in the sense that

$$\Theta_x(j_1, j_2, \ldots, j_l).z(x) = z(\phi_x(j_1, j_2, j_3, \ldots, j_l))$$

for all $(j_1, \ldots, j_l) \in \Sigma^*(x)$.

Proof. We'll show for all x' there is a one to one correspondence between the admissible sequences (j) of length 1 and $\{\lambda_j : 1 \leq j \leq n-2\}$ so that $\Theta_x(j).z(x) = z(\phi_{x'}(j))$. This is clear if $x'_1 \leq x'_2 \leq \ldots < x'_n$ is ordered (send $j \mapsto \lambda_j$). Otherwise pick an ordering of x'. The general result follows by repeating this process. $\square$

Lemma 21 implies that the set $V(\mathbf{Z}_+)$ decomposes into the finite set K_0 and a finite number of orbits of the form

$$\phi_{x^{(0)}}(\Sigma^*(x^{(0)})).$$

Each one of these orbits has either all its points exceptional or unexceptional. Since we assume throughout the rest of the paper that $V(\mathbf{Z}) - \mathcal{E}$ is infinite, it follows that the collection $\mathcal{U}$ of unexceptional basepoints $x^{(0)}$ is finite and nonempty. Summing up,

$$V(\mathbf{Z}_+) - \mathcal{E} - K_0 = \coprod_{x^{(0)} \in \mathcal{U}} \phi_{x^{(0)}}(\Sigma^*(x^{(0)})),$$

so

$$\begin{aligned} |(V(\mathbf{Z}_+) - \mathcal{E}) \cap B(R)| &= O_{n,a,k}(1) + \sum_{x^{(0)} \in \mathcal{U}} \sum_{s \in \Sigma^*(x^{(0)})} \mathbf{1}\{\max(\phi_{x^{(0)}}(s)) \le R\} \\ &= O_{n,a,k}(1) + \sum_{x^{(0)} \in \mathcal{U}} \sum_{s \in \Sigma^*(x^{(0)})} \mathbf{1}\left\{z(\phi_{x^{(0)}}(s))_n \le a^{\frac{1}{n-2}} R\right\}. \end{aligned}$$

Applying Lemma 22 to the above sum, one obtains

$$|(V(\mathbf{Z}_+) - \mathcal{E}) \cap B(R)| = O_{n,a,k}(1) + \sum_{x^{(0)} \in \mathcal{U}} \sum_{\lambda \in \Lambda} \mathbf{1}\left\{(\lambda.z(x^{(0)}))_n \le a^{\frac{1}{n-2}} R\right\}.$$

Therefore, Theorem 3 will follow from asymptotic estimates for the quantity

$$\sum_{\lambda \in \Lambda} \mathbf{1}\left\{(\lambda.z^{(0)})_n \le R\right\} \tag{2.9}$$

where $z^{(0)} \in z(\mathcal{U}) \subset \mathcal{M} - K$. These estimates are taken up in the next section. We draw the reader's attention to the fact that the count is over Λ and not over $\mathcal{M}$.

2.4 Increasing the size of K

Before we begin the count we increase the size of K. Recall that K and K_0 are balls with center 0 in the ℓ^∞ norm with radii coupled by (2.6) and that we are free to increase their radii (maintaining the relationship (2.6)). The following can be thought of as regularizing the dynamics of $\mathcal{M}$ at a fixed scale depending on n, a, k. We state our requirements in terms of $z = (z_1, \ldots, z_n)$.

First we make sure z_{n-1} is reasonably large compared to z_n. Suppose $z_{n-1} \le c z_n^{\frac{1}{n-1}}$. Then $z_1 \le z_2 \le \ldots \le z_{n-1} \le c z_n^{\frac{1}{n-1}}$. Then (2.5) gives

$$z_n^2 \le c^{n-1} z_n^2 + k'$$

which is a contradiction for $c < 1$ and z_n large enough depending on k'. We increase the radius of K so that

$$z_{n-1} \ge \frac{1}{2} z_n^{\frac{1}{n-1}} \tag{2.10}$$

for all $z \in \mathcal{M} - K$.

Now, in preparation for later, we wish to make sure various inequalities hold. If $z_n \ge \left(\frac{2}{2-\sqrt{3}}\right)^{n-1}$ then

$$\frac{(n-1)\log(1 - 2z_n^{-1/(n-1)}) - (n-1)\log 2}{\log z_n} \ge -1/2, \tag{2.11}$$

$$z_n \ge 10, \tag{2.12}$$

and

$$z_{n-1} > 2 \tag{2.13}$$

all follow. We increase the radius of K if necessary to be at least $\left(\frac{2}{2-\sqrt{3}}\right)^{n-1}$ so that (2.11), (2.12), and (2.13) all hold for $z \in \mathcal{M} - K$. Furthermore by increasing the radius of K, using (2.10) we can also ensure

$$z_1^2 + \ldots + z_{n-1}^2 - k' \geq 0 \tag{2.14}$$

for $z \in \mathcal{M} - K$.

3 Converting the linear count to the nonlinear count

3.1 Acceleration

In the last Section 2 we reduced our Main Theorem 3 to obtaining an asymptotic formula for the count

$$\sum_{\lambda \in \Lambda} \mathbf{1}\left\{(\lambda.z^{(0)})_n \leq R\right\} \tag{3.1}$$

where $z^{(0)}$ is one of a finite set of unexceptional points in $\mathcal{M} - K$. For the rest of the paper we view $z^{(0)}$ as fixed.

There is a general framework in which to count over the tree-like Λ, called the *renewal method.* This was first used in counting problems by Parry and Pollicott [PP83] to establish an analog of the prime number theorem for Axiom A flows. It was further developed by Lalley [Lal89] to perform lattice point counting in Fuchsian groups. The essence of the method is a recursion over Λ. One feature of the current work that makes matters more complicated than for the original expositions of the renewal method is that we perform what we call *acceleration.* A similar acceleration technique has appeared in works of Pollicott [Polb] in the context of counting circles in an Apollonian packing, and also in work of Pollicott and Urbański [PU17] for more general conformal graph directed Markov systems.

Concretely, we replace the generators $\{\lambda_j : 1 \leq j \leq n-1\}$ of Λ with the countably infinite set of generators

$$S = S_\Lambda = \left\{\lambda_{n-1}^A \lambda_j \;:\; A \in \mathbf{Z}_{\geq 0},\, 1 \leq j \leq n-2\,\right\}.$$

It is easy to see that S_Λ are free generators for the subsemigroup

$$\Lambda' = \cup_{j=1}^{n-2} \Lambda.\lambda_j \subset \Lambda$$

that contains the words beginning with λ_j, $1 \leq j \leq n-2$. This acceleration is crucial for our method and has two advantages:

1. The quality of our fitting the nonlinear count for Λ to a linear count to Γ depends on the size of the quantity

$$\alpha(z) = \prod_{j=1}^{n-2} z_j, \tag{3.2}$$

cf. Lemma 27 below. This quantity can be small for long words with respect to the generators $\{\lambda_j\}$, because λ_{n-1} does not alter $\alpha(z)$. On the other hand, we prove in Lemma 26 that $\alpha(z)$ grows doubly exponentially in the word length with respect to the generators S_Λ.

2. When we eventually arrive at the dynamics of Γ' on $P(\mathbf{R}^n_{\geq 0})$ (throughout the paper we use P to denote the projective version of an object), the unaccelerated system would be non-uniformly contracting and therefore we could not expect there to be a finite invariant measure for this system. On the other hand, the acceleration we perform leads to uniformly contracting dynamics (cf. Proposition 45) and in turn to a nice description of the invariant measure and leading eigenfunction for the transfer operator in the Ruelle-Perron-Frobenius Theorem (Theorem 39).

Now, the orbit $\Lambda.z^{(0)}$ breaks up into the countable union of orbits

$$\Lambda.z^{(0)} = \bigcup_{A_0=0}^{\infty} \Lambda'.\lambda_{n-1}^{A_0} z^{(0)}. \tag{3.3}$$

It is clear that an asymptotic formula for (3.1) is equivalent to an asymptotic formula for

$$M_0(z,r) \equiv \sum_{\lambda \in \Lambda \cup \{e\}} \mathbf{1}\{\log\log(\lambda.z)_n - \log\log z_n \leq r\} \tag{3.4}$$

when $z = z^{(0)}$. Note that here we intend to take $r = \log\log R - \log\log(z^{(0)})_n$ which tends to ∞ as $R \to \infty$. On the other hand, our methods can prove an asymptotic formula for the following quantity

$$M(z,r) = \sum_{\lambda \in \Lambda' \cup \{e\}} \mathbf{1}\{\log\log(\lambda.z)_n - \log\log z_n \leq r\} \tag{3.5}$$

for arbitrary unexceptional $z \in \mathcal{M} - K$. Precisely, we will obtain the following proposition.

Proposition 23. *For all unexceptional $z \in \mathcal{M} - K$ there is a positive constant $c_\star$ such that as $r \to \infty$,*

$$M(z,r) = e^{\beta r}(c_\star(z) + o(1)),$$

where $\beta > 1$ is the constant from Theorem 2 and the rate of decay in the small o does not depend on z. Moreover, the $c_\star(z)$ have a uniform bound depending only on n.

The proof of Proposition 23 will occupy the rest of this Section. Before beginning, we show how Proposition 23 implies our main Theorem 3. This passage relies on the following elementary lemma.

Lemma 24. *For unexceptional $z \in \mathcal{M} - K$ we have*

$$(\lambda_{n-1}^A z)_n \geq (\alpha(z) - 1)^A z_n \geq 2^A z_n,$$

where $\alpha(z)$ is the quantity defined in (3.2).

Proof. One can calculate easily that for $z = (z_1, \ldots, z_n)$, $\lambda_{n-1}^A z$ is obtained by A applications of the matrix

$$g_{\alpha(z)} = \begin{pmatrix} 0 & 1 \\ -1 & \alpha(z) \end{pmatrix}$$

to the last two entries of z. If we let $Z_A = (\lambda_{n-1}^A z)_n$ then the Z_A satisfy the recurrence

$$Z_{A+1} = \alpha(z) Z_A - Z_{A-1} \geq (\alpha(z) - 1) Z_A.$$

Therefore $(\lambda_{n-1}^A z)_n \geq (\alpha(z)-1)^A z_n$. This proves the first inequality.

If $z = z(x) = a^{\frac{1}{n-2}} x$ as in (2.4) with $x_1 \leq x_2 \leq \ldots \leq x_n$ then

$$\alpha(z) = a x_1 x_2 \ldots x_{n-2} \in \mathbf{Z}_+.$$

If $\alpha(z) = 1$ then the matrix $g_{\alpha(z)}$ is torsion and this contradicts the maximal entries of $\lambda_{n-1}^A z$ growing with A (since $z \in \mathcal{M} - K$). If $\alpha(z) = 2$ then z must be a fundamental exceptional solution. Otherwise $\alpha(z) \geq 3$ and this proves the second inequality. □

Proof of Theorem 3 given Proposition 23. By our previous discussion it suffices to prove an asymptotic formula for $M_0(z^{(0)}, r)$ for a fixed $z^{(0)}$. But using (3.3) gives

$$M_0(z^{(0)}, r) = \sum_{A_0=0}^{\infty} M(\lambda_{n-1}^{A_0} z^{(0)}, r - \log\log(\lambda_{n-1}^{A_0} z^{(0)})_n + \log\log z_n^{(0)}). \tag{3.6}$$

By using Lemma 24, the value $A_0 = A_{\max}$ where $r - \log\log(\lambda_{n-1}^{A_0} z^{(0)})_n + \log\log z_n^{(0)}$ first becomes negative is bounded by

$$A_{\max} \leq \frac{\log z_n^{(0)} e^r}{\log 2}.$$

Let the small o term in Proposition 23 be bounded in absolute value by a positive function $F(r)$ that tends to 0 as $r \to \infty$. Let κ be a small positive constant to be chosen. The A_0 such that $r - \log\log(\lambda_{n-1}^{A_0} z^{(0)})_n + \log\log z_n^{(0)} \geq \kappa r$ contribute

$$\left(\log z_n^{(0)}\right)^{\beta} e^{\beta r} \sum_{A_0 : r - \log\log(\lambda_{n-1}^{A_0} z^{(0)})_n + \log\log z_n^{(0)} \geq \kappa r} \frac{c_\star(\lambda_{n-1}^{A_0} z^{(0)})}{(\log(\lambda_{n-1}^{A_0} z^{(0)})_n)^{\beta}} (1 + O(\sup_{r' \geq \kappa r} F(r')).$$

to (3.6) by Proposition 23. Furthermore, by Lemma 24,

$$\sum_{A_0 : r - \log\log(\lambda_{n-1}^{A_0} z^{(0)})_n + \log\log z_n^{(0)} \geq \kappa r} \frac{c_\star(\lambda_{n-1}^{A_0} z^{(0)})}{(\log(\lambda_{n-1}^{A_0} z^{(0)})_n)^{\beta}} \leq \sum_{A_0} \frac{c_\star(\lambda_{n-1}^{A_0} z^{(0)})}{(A_0 \log 2)^{\beta}}$$

converges to some limit $c_\infty(z^{(0)})$ as $r \to \infty$, using $\beta > 1$. Therefore the terms we have discussed so far give a contribution of

$$\left(\log z_n^{(0)}\right)^{\beta} c_\infty(z^{(0)}) e^{\beta r} (1 + o(1))$$

to $M_0(z^{(0)}, r)$ via (3.6).

For the remaining A_0 such that $r - \log\log(\lambda_{n-1}^{A_0} z^{(0)})_n + \log\log z_n^{(0)} < \kappa r$ we use Proposition 23 in a coarser way to get $M(z, r) \leq C e^{\beta r}$ for some constant C, uniformly over unexceptional $z \in \mathcal{M} - K$. Then any remaining A_0 contributes at most $C e^{\beta\kappa r}$ to (3.6). Therefore the remaining contributions are in total at most

$$A_{\max} C e^{\beta\kappa r} \leq \frac{\log z_n^{(0)} C e^{(1+\beta\kappa) r}}{\log 2}$$

which is negligible when $1 + \beta\kappa < \beta$, and we can find such a κ since $\beta > 1$. □

3.2 The renewal equation for M

We now take up the proof of Proposition 23. While the statement of Proposition 23 is uniform over all unexceptional $z \in \mathcal{M} - K$, our previous arguments show that the unexceptional elements of $\mathcal{M} - K$ break up into finitely many orbits of Λ. Therefore it is sufficient for us to establish Proposition 23 for $z = \lambda_0 z^{(0)}$, where $z^{(0)} \in z(\mathcal{U})$ is a fixed unexceptional basepoint and λ_0 is an arbitrary element of Λ. We therefore view $z^{(0)}$ as fixed from now on, and we will prove Proposition 23 for $z = \lambda_0 z^{(0)}$, with uniformity over $\lambda_0 \in \Lambda$.

We now describe the renewal equation, for which we need some new concepts. Define the *shift* s by

$$s(\lambda_{n-1}^{A_l}\lambda_{j_l}\lambda_{n-1}^{A_{l-1}}\lambda_{j_{l-1}}\dots\lambda_{n-1}^{A_2}\lambda_{j_2}\lambda_{n-1}^{A_1}\lambda_{j_1}) \equiv \lambda_{n-1}^{A_{l-1}}\lambda_{j_{l-1}}\dots\lambda_{n-1}^{A_2}\lambda_{j_2}\lambda_{n-1}^{A_1}\lambda_{j_1}.$$

Now extend this definition so that $s(\lambda\lambda_0) = s(\lambda)\lambda_0$ for all $\lambda \in \Lambda'$ and $\lambda_0 \in \Lambda \cup \{e\}$. We define the *distortion function* $\tau_\star : \Lambda'.(\Lambda \cup \{e\}) \to \mathbf{R}_{\geq 0}$ by

$$\tau_\star(\lambda) \equiv \log\log(\lambda.z^{(0)})_n - \log\log(s(\lambda).z^{(0)})_n.$$

This depends on the constant $z^{(0)}$. One also has the iterated version of distortion

$$\tau_\star^N(\lambda) = \sum_{p=0}^{N-1} \tau_\star(s^p(\lambda)) = \log\log(\lambda.z^{(0)})_n - \log\log(s^N(\lambda).z^{(0)})_n. \tag{3.7}$$

for any $\lambda \in s^{-N}(\Lambda)$. The *renewal equation* for M is then

$$M(\lambda z^{(0)}, r) = \sum_{\lambda' \in S_\Lambda} M(\lambda'\lambda z^{(0)}, r - \tau_\star(\lambda'\lambda)) + \mathbf{1}\{0 \leq r\} \tag{3.8}$$

for all $\lambda \in \Lambda$. Note that the summation above is finite since the λ' act to strictly increase maximal entries in $\mathcal{M}$.

3.3 Iteration

The eventual goal is to compare the asymptotics of $M(\lambda z^{(0)}, r)$ to those of an analogous quantity for the linear semigroup Γ introduced in the Introduction. Before this happens, a regularization must occur. In our approach[12], the quality of the comparison to the linear semigroup depends on the size of

$$\alpha(z_1, \dots, z_n) = \prod_{j \leq n-2} z_j.$$

It is clear that no $\lambda \in \Lambda$ decreases $\alpha(z)$. To pass to the case that $\alpha(\lambda'.z^{(0)})$ is large, we iterate the renewal equation (3.8) L times. This yields

$$M(\lambda z^{(0)}, r) = \sum_{\lambda' : s^L(\lambda') = \lambda} M(\lambda' z^{(0)}, r - \tau_\star^L(\lambda')) + \sum_{l=1}^{L-1} \sum_{\lambda' : s^l(\lambda') = \lambda} \mathbf{1}\left\{\tau_\star^l(\lambda') \leq r\right\} + \mathbf{1}\left\{0 \leq r\right\}, \tag{3.9}$$

[12]In Zagier's approach in [Zag82] for the case $n = a = 3$, there is a special mapping arising from the close connection between the Markoff equation and hyperbolic geometry. This mapping offers a much better fit to the linear semigroup count than is available in general. See footnote 16 for more on this.

recalling the definition of $\tau_\star^L$ from (3.7). We now show that for suitable L the last two summations in (3.9) are negligible. The following lemma is used at several points in the rest of the paper.

Lemma 25. *There are constants c_0 and c_1 depending only on n such that for all $L \in \mathbf{N}$, $x \geq 0$*

$$\sum_{\lambda' : s^L(\lambda') = \lambda} \mathbf{1}\left\{ \tau_\star^L(\lambda') \leq x \right\} \leq c_1^L (c_0 + x)^L e^x. \tag{3.10}$$

As a consequence, for any $\delta > 0$, there is $c = c(\delta) > 0$ such that when $L = \left\lceil \frac{cr}{\log r} \right\rceil$ one has

$$\sum_{l=1}^{L-1} \sum_{\lambda' : s^l(\lambda') = \lambda} \mathbf{1}\left\{ \tau_\star^l(\lambda') \leq r \right\} = O(e^{(1+\delta)r}) \tag{3.11}$$

and

$$c_1^L (c_0 + x)^L \leq e^{\delta x} \tag{3.12}$$

for all $x \geq r/2$.

Proof. For the first part of this proof, let $\tilde{\lambda}$ denote an arbitrary element of Λ', and $z := \tilde{\lambda}.z^{(0)}$. By applying Lemma 24 to $\lambda_j z$ we obtain

$$\begin{aligned} \tau_\star(\lambda_{n-1}^A \lambda_j \tilde{\lambda}) &= \log\log(\lambda_{n-1}^A \lambda_j z)_n - \log\log z_n \\ &\geq \log\log((\alpha(\lambda_j z) - 1)^A (\lambda_j z)_n) - \log\log z_n. \end{aligned}$$

Since $z_j \geq x_j \geq 1$, we have for $1 \leq j \leq n-2$

$$\alpha(\lambda_j z) = z_{n-1} \prod_{\substack{1 \leq i \leq n-2 \\ i \neq j}} z_i \geq z_{n-1} \geq \frac{1}{2} z_n^{\frac{1}{n-1}}$$

where the last inequality used (2.10). Hence using $(\lambda_j z)_n \geq z_n$,

$$\begin{aligned} \tau_\star(\lambda_{n-1}^A \lambda_j \tilde{\lambda}) &\geq \log\log\left(\frac{1}{2^A} z_n^{A/(n-1)} (1 - 2 z_n^{-1/(n-1)})^A z_n\right) - \log\log z_n \\ &\geq \log\left(1 + \frac{A}{n-1}\left(1 + \frac{(n-1)\log(1 - 2z_n^{-1/(n-1)}) - (n-1)\log 2}{\log z_n}\right)\right) \\ &\geq \log\left(1 + \frac{A}{2(n-1)}\right), \end{aligned} \tag{3.13}$$

where the last inequality is by the previously prepared (2.11). Now, if
$\lambda = \lambda_{n-1}^{A_l} \lambda_{j_l} \lambda_{n-1}^{A_{l-1}} \lambda_{j_{l-1}} \ldots \lambda_{n-1}^{A_2} \lambda_{j_2} \lambda_{n-1}^{A_1} \lambda_{j_1}$ then by l applications of (3.13) we get

$$\tau_\star^l(\lambda) = \sum_{p=0}^{l-1} \tau_\star(s^p(\lambda)) \geq \sum_{q=1}^{l} \log\left(1 + \frac{A_q}{2(n-1)}\right).$$

Therefore the number of λ' that can contribute to (3.10) is bounded by the size of the set

$$\left\{ (A_1, A_2, A_3, \ldots, A_L) \in \mathbf{Z}_{\geq 0}^L : \sum_{q=1}^{L} \log\left(1 + \frac{A_q}{2(n-1)}\right) \leq x \right\}. \tag{3.14}$$

times the number of possible choices for $j_1, \dots j_L$. The latter can be crudely bounded by $(n-2)^L$.

Claim. The size of the set in (3.14) is bounded by $(2(n-1)(c_0+x))^L e^x$ for some positive constant c_0.

Proof of Claim. We prove this by induction on L. The base case ($L=1$) is clear. For the induction, after choosing the first A_1 the remaining $A_2, \dots, A_L$ must satisfy

$$\sum_{q=2}^{L} \log\left(1+\frac{A_q}{2(n-1)}\right) \le x - \log\left(1+\frac{A_1}{2(n-1)}\right).$$

So the size of the set in (3.14) is bounded by

$$\begin{aligned}&\sum_{A_1=1}^{\lfloor 2(n-1)e^x\rfloor} (2(n-1))^{L-1}\left(c_0+x-\log\left(1+\frac{A_1}{2(n-1)}\right)\right)^{L-1} e^x \frac{1}{1+\frac{A_1}{2(n-1)}}\\ \le&(2(n-1))^L(c_0+x)^{L-1}e^x \sum_{A_1=1}^{\lfloor 2(n-1)e^x\rfloor} \frac{1}{2(n-1)+A_1}.\end{aligned}$$

The final sum is within a constant c_0 of x. This completes the proof of the Claim.

So in total we obtain that the sum in (3.10) is bounded by $c_1^L(c_0+x)^L e^x$ with $c_1 = 2(n-2)(n-1)$. As for the stated consequence, we get

$$\sum_{l=1}^{L-1} \sum_{\lambda' : s^l(\lambda')=\lambda} \mathbf{1}\left\{\tau_\star^l(\lambda') \le r\right\} \ll c_1^L(c_0+r)^L e^r.$$

If we choose $L \approx cr/\log(1+r)$ with c small enough depending on δ we obtain our result. $\square$

Since we expect $M(\lambda z^{(0)}, r) \approx e^{\beta r}$ with $\beta = \beta(n) > 1$, choosing parameters as in Lemma 25 with $L \approx r/\log r$ gives

$$M(\lambda z^{(0)}, r) = \sum_{\lambda' : s^L(\lambda')=\lambda} M(\lambda' z^{(0)}, r-\tau_\star^L(\lambda')) + O(e^{(1+\delta)r}) \tag{3.15}$$

and the big O term is truly an error term when δ is small. The benefits to our iteration in (3.15) can be quantified by the following result.

Lemma 26. *There is some $C>0$ such that for all $\lambda \in \Lambda \cup \{e\}$ and λ' such that $s^L(\lambda') = \lambda$, we have both*

$$\alpha(\lambda' z^{(0)}) \ge \frac{1}{2}\exp(C\phi^L) \tag{3.16}$$

and

$$(\lambda' z^{(0)})_n \ge \exp(C\phi^L) \tag{3.17}$$

where $\phi = \frac{1+\sqrt{5}}{2} > 1$ is the golden ratio.

Proof. For $1 \leq j \leq n-2$

$$(\lambda_j z)_n = \prod_{i \neq j} z_i - z_j = z_n z_{n-1} \prod_{1 \leq i \leq n-2, i \neq j} z_i - z_j \geq (z_n - 1) z_{n-1}$$

since $z_i \geq 1$ for all i and $z_{n-1} \geq z_j$. So then for any $A \geq 0$

$$(\lambda_{n-1}^A \lambda_j z)_n \geq (\lambda_j z)_n \geq (z_n - 1) z_{n-1}.$$

Then

$$\begin{aligned}
(\lambda_{n-1}^{A_2} \lambda_{j_2} \lambda_{n-1}^{A_1} \lambda_{j_1} z)_n &\geq ((\lambda_{n-1}^{A_1} \lambda_{j_1} z)_n - 1)(\lambda_{n-1}^{A_1} \lambda_{j_1} z)_{n-1} \\
&\geq ((\lambda_{n-1}^{A_1} \lambda_{j_1} z)_n - 1) z_n
\end{aligned}$$

using the inequality $(\lambda z)_{n-1} \geq z_n$ for any $\lambda \in \Lambda$. Therefore the numbers

$$Z_p = (\lambda_{n-1}^{A_p} \lambda_{j_p} \ldots \lambda_{n-1}^{A_2} \lambda_{j_2} \lambda_{n-1}^{A_1} \lambda_{j_1} z)_n \geq 10$$

(cf. (2.12)) satisfy the two stage recursive estimate $Z_p \geq (Z_{p-1} - 1) Z_{p-2}$ for $p \geq 2$. Then an elementary argument gives the existence of C such that

$$Z_p \geq \exp(C \phi^p).$$

This gives the required (3.17).

On the other hand

$$\alpha(\lambda_{n-1}^A \lambda_j z) \geq \alpha(\lambda_j z) \geq z_{n-1} \geq \frac{1}{2} z_n^{\frac{1}{n-1}}$$

where the last inequality is by (2.10). The result (3.16) now follows after replacing C with a suitable smaller constant. □

In the sequel we choose

$$L = \left\lceil c \frac{r}{\log r} \right\rceil \tag{3.18}$$

so that (3.15) and (3.12) hold with[13]

$$\delta = \min\left(\frac{1}{10}, \frac{\beta - 1}{2}\right). \tag{3.19}$$

Then for all $\lambda' z^{(0)}$ appearing in (3.15) we have

$$\alpha(\lambda' z^{(0)}) \geq \frac{1}{2} \exp(C \phi^{cr/\log r}) \tag{3.20}$$

by Lemma 26.

[13] We know by Remark 14 that $\beta \geq 2$.

3.4 Comparison to the linear count

Now we relate the terms $M(\lambda' z^{(0)}, r - \tau_\star^L(\lambda'))$ appearing in (3.15) to orbital counting for Γ, the linear semigroup defined in the Introduction. We begin with the expression for $M(\lambda' z^{(0)}, r - \tau_\star^L(\lambda'))$ in (3.5). Denoting by $S^{(N)}$ the N-fold product[14] of the countable generating set S for Λ', then we can write

$$M(\lambda' z^{(0)}, r - \tau_\star^L(\lambda')) = \sum_{N=0}^{\infty} \sum_{\lambda^{(2)} \in S^{(N)}} \mathbf{1}\{\tau_\star^N(\lambda^{(2)}\lambda') \leq r - \tau_\star^L(\lambda')\}. \tag{3.21}$$

We will proceed by

1. Matching $\lambda' z^{(0)}$ with some element of $\mathcal{H} \subset \mathbf{R}_+^n$ that is very close to[15] $\log(\lambda' z^{(0)})$.
2. Matching each $\lambda^{(2)}$ with an element $\gamma^{(2)}$ of Γ in the obvious way.

With Part 1 in mind, we define for $z \in \mathcal{M}$

$$f(z) \equiv (\log z_1, \log z_2, \ldots, \log z_{n-1}, \sum_{j=1}^{n-1} \log z_i).$$

The reason to use this map over just taking log of coordinates is that we expect $\log(z)$ to be very close to the hyperplane $\mathcal{H}$ defined in (1.5), so we just go ahead and fit $\log(z)$ to this plane. The following lemma (cf. Lemma 2 in [Zag82]) says that when $\alpha(z)$ is big, $f(z)$ is a good[16] fit to $\log(z)$. In this paper, we write inequalities between vectors to mean they hold at every coordinate.

Lemma 27. *There are constants C_1 and C_2 depending only on n such that when $z \in \mathcal{M} - K$ with $\alpha(z) > C_1$*

$$\log(z) \leq f(z) \leq \log(z) + C_2 \alpha(z)^{-2}(0, 0, 0, \ldots, 0, 1). \tag{3.22}$$

Proof. Since z satisfies the equation (2.5), and z_n is always the larger of the two quadratic roots of the resulting quadratic in z_n, we have

$$z_n = \frac{A(z)}{2}\left(1 + \sqrt{1 - 4\frac{C(z) - k'}{A(z)^2}}\right)$$

where

$$A(z) = \prod_{i=1}^{n-1} z_i, \quad C(z) = \sum_{i=1}^{n-1} z_i^2$$

and $k' \geq 0$ is the constant from (2.5). Now the first inequality of (3.22) follows from (2.14).

[14] That is, $S^{(N)}$ is the set of elements of Λ' that are a product of N generators. We extend this definition to $S^{(0)} = \{e\}$.

[15] When we write log of a vector we always mean take log of each coordinate.

[16] Although our f is not even close to being as good as Zagier's function f from [Zag82]: the quality of fit of Zagier's f improves with the size of z_{n-1} whereas we need z_{n-2} to be big. This is one reason we must accelerate.

For the second inequality we estimate

$$\frac{C(z)-k'}{A(z)^2} \leq \sum_{i=1}^{n-1} \frac{z_{n-1}^2}{\prod_{j\neq n} z_i^2} \leq (n-1)\frac{1}{\prod_{j\leq n-2} z_j^2} = (n-1)\alpha(z)^{-2}.$$

We can then choose C_1 large enough so that when $\alpha(z) > C_1$ we have

$$z_n = A(z)(1 + O_n(\alpha(z)^{-2}));$$

by increasing C_1 again if necessary we obtain

$$\log(z_n) = \log(A(z)) + O_n(\alpha(z)^{-2}) = f(z)_n + O_n(\alpha(z)^{-2}).$$

□

The following adapts an idea of Zagier from [Zag82, Proof of Lemma 3] to our setting. While the strength of approximation is different, we take the same approach in noting that if $f(z)$ is close to y then $f(\lambda_j z)$ will be close to $\gamma_j y$. Of course this is designed to be iterated.

Lemma 28. *There are C_1, C_2 depending only on n such that for all $\epsilon > 0$, for $z \in \mathcal{M} - K$, $\alpha(z) > \max(C_1, 2C_2^{1/2}\epsilon^{-1/2})$, and for $y^{(1)}, y^{(2)} \in \mathcal{H}$, if*

$$y^{(1)} + \epsilon(0,0,\dots 0, \frac{1}{2}, \frac{1}{2}, 1) < f(z) \leq y^{(2)} \tag{3.23}$$

then

$$\gamma_j y^{(1)} + \epsilon(0,0,\dots 0, \frac{1}{2}, \frac{1}{2}, 1) < f(\lambda_j z) \leq \gamma_j y^{(2)} \tag{3.24}$$

for all $1 \leq j \leq n-1$.

Proof. We first prove the upper bound for $f(\lambda_j z)$ in (3.24). The inequality $f(z) \leq y^{(2)}$ implies that $\log(z_i) \leq y_i^{(2)}$ for $i \leq n-1$. By Lemma 27 we get $\log(z_n) \leq f(z)_n \leq y_n$ as well. Then $f(\lambda_j z) \leq \gamma_j y^{(2)}$ follows.

For the other inequality, $f(z) > y^{(1)} + \epsilon(0,0,\dots 0, 1/2, 1/2, 1)$ implies $\log(z_i) > y_i^{(1)}$ for all $i \leq n-3$ and $\log(z_i) > y_i^{(1)} + \epsilon/2$ for $i = n-2, n-1$. By Lemma 27, $\log(z_n) \geq f(z)_n - C_2\alpha(z)^{-2} \geq y_n^{(1)} + \epsilon - C_2\alpha(z)^{-2}$. Since $\alpha(z) > 2C_2^{1/2}\epsilon^{-1/2}$ we get

$$\log(z_n) \geq y_n^{(1)} + 3\epsilon/4.$$

When $i \leq n-3$ we have $f(\lambda_j z)_i \geq (\gamma_j y^{(1)})_i$ quite clearly. If $j \leq n-2$ we have $f(\lambda_j z)_{n-2} = \log z_{n-1} \geq y_{n-1}^{(1)} + \epsilon/2 = (\gamma_j y^{(1)})_{n-2} + \epsilon/2$ and if $j = n-1$ then $f(\lambda_j z)_{n-2} = \log z_{n-2} \geq y_{n-2}^{(1)} + \epsilon/2 = (\gamma_j y^{(1)})_{n-2} + \epsilon/2$. At the $(n-1)$st coordinate we have $f(\lambda_j z)_{n-1} = \log z_n \geq y_n^{(1)} + 3\epsilon/4 = (\gamma_j y^{(1)})_{n-1} + 3\epsilon/4$ which is sufficient. It remains to check the last coordinate. Here,

$$f(\lambda_j z)_n = \sum_{i\neq j} \log z_i \geq \sum_{i\neq j} y_i^{(1)} + 5\epsilon/4 = (\gamma_j y^{(1)})_n + 5\epsilon/4.$$

The inequality above is due to the fact that at least one of $\log z_{n-2}, \log z_{n-1}$ appear on the left hand side (giving $\epsilon/2$) and $\log z_n$ also appears (giving $3\epsilon/4$). □

We can now accomplish Parts 1 and 2 of our plan above. Recall we have some fixed $z^{(0)} \in \mathcal{M} - K$. For each given $\lambda' \in \Lambda$ (in particular, those that occur in (3.15)) we define

$$y(\lambda') = f(\lambda' z^{(0)}).$$

We choose our parameters as follows: let C_2 be the constant from Lemma 28 and set

$$\epsilon = \epsilon(r) = 16C_2 \exp(-2C\phi^{cr/\log r}). \tag{3.25}$$

so that by (3.20)

$$4C_2\alpha(\lambda' z^{(0)})^{-2} \leq \epsilon$$

for all λ' appearing in (3.15).

Lemma 29 (Completing Part 1)**.** *For any $\lambda \in \Lambda \cup \{e\}$ and $\lambda' \in S^{(L)}\lambda$ with L as in (3.18) we have*

$$(1-\epsilon)y(\lambda') + \epsilon(0, 0, \ldots 0, \frac{1}{2}, \frac{1}{2}, 1) < f(\lambda' z^{(0)}) = y(\lambda').$$

Proof. Since $L > 0$ and $\lambda' \in S^{(L)}(\Lambda \cup e)$ we have

$$(\lambda' z^{(0)})_{n-2} \geq (z^{(0)})_{n-1} > 2,$$

using (2.13). Therefore $f(\lambda' z^{(0)})_{n-2} \geq \log(2) > 1/2$. Since $f(\lambda' z^{(0)}) \in \mathcal{H}$ it follows that

$$\epsilon f(\lambda' z^{(0)}) \geq \epsilon(0, 0, \ldots 0, \frac{1}{2}, \frac{1}{2}, 1),$$

from which the lemma is a direct consequence. $\square$

Now for each

$$\lambda^{(2)} = \lambda_{n-1}^{A_N}\lambda_{j_N}\lambda_{n-1}^{A_{N-1}}\lambda_{j_{N-1}} \ldots \lambda_{n-1}^{A_2}\lambda_{j_2}\lambda_{n-1}^{A_1}\lambda_{j_1} \in S^{(N)}, \quad 1 \leq j_i \leq n-2 \;\; \forall i$$

appearing in (3.21), we set

$$\gamma^{(2)} = \gamma^{(2)}(\lambda^{(2)}) = \gamma_{n-1}^{A_N}\gamma_{j_N}\gamma_{n-1}^{A_{N-1}}\gamma_{j_{N-1}} \ldots \gamma_{n-1}^{A_2}\gamma_{j_2}\gamma_{n-1}^{A_1}\gamma_{j_1} \in \Gamma' \cup \{e\}. \tag{3.26}$$

This is the matching of Part 2. Since Λ' and Γ' are free, this gives a bijective correspondence.

Now we claim we can reasonably compare each of the $M(\lambda' z^{(0)}, r - \tau_\star^L(\lambda'))$ from (3.15) to $N(y(\lambda'), r')$ defined in (1.7) with r' very close to $r - \tau_\star^L(\lambda')$.

Lemma 30. *For any $\lambda \in \Lambda \cup \{e\}$ and $\lambda' \in S^{(L)}\lambda$ with L as in (3.18) we have*

$$N(y(\lambda'), r - \tau_\star^L(\lambda') - 2\epsilon) \leq M(\lambda' z^{(0)}, r - \tau_\star^L(\lambda')) \leq N(y(\lambda'), r - \tau_\star^L(\lambda') + 2\epsilon).$$

Proof. Consider the expression (3.21) for $M(\lambda' z^{(0)}, r - \tau_\star^L(\lambda'))$. The key point now is that by iterating Lemma 28 we obtain for all coupled $\lambda^{(2)}, \gamma^{(2)}$,

$$(1-\epsilon)\gamma^{(2)}.y(\lambda') + \epsilon(0, 0, \ldots 0, \frac{1}{2}, \frac{1}{2}, 1) < f(\lambda^{(2)}\lambda' z^{(0)}) \leq \gamma^{(2)}.y(\lambda')$$

where we have used the linearity of the action of Γ to pull out the factor of $(1-\epsilon)$. Using Lemma 27 we get

$$\log(\lambda^{(2)}\lambda' z^{(0)})_n \leq f(\lambda^{(2)}\lambda' z^{(0)})_n \leq (\gamma^{(2)}.y(\lambda'))_n$$

and

$$\log(\lambda^{(2)}\lambda' z^{(0)})_n \geq f(\lambda^{(2)}\lambda' z^{(0)})_n - \frac{\epsilon}{4} \geq (1-\epsilon)(\gamma^{(2)}.y(\lambda'))_n.$$

Then taking logarithms gives

$$\log\log(\lambda^{(2)}\lambda' z^{(0)})_n \leq \log(\gamma^{(2)}.y(\lambda'))_n \leq \log\log(\lambda^{(2)}\lambda' z^{(0)})_n + 2\epsilon \tag{3.27}$$

using $2\epsilon + \log(1-\epsilon) > 0$ for $\epsilon \ll 1$. Note that (3.27) also holds when $\gamma^{(2)} = e$, $\lambda^{(2)} = e$.

Let $r' = r - \tau_\star^L(\lambda') \pm 2\epsilon$. We write out

$$N(y, r') = \sum_{\gamma^{(2)} \in \Gamma' \cup \{e\}} \mathbf{1}\{\log(\gamma^{(2)}.y(\lambda'))_n - \log y(\lambda')_n \leq r'\}$$

and compare to

$$M(\lambda' z^{(0)}, r - \tau_\star^L(\lambda')) = \sum_{\lambda^{(2)} \in \Lambda' \cup \{e\}} \mathbf{1}\left\{\log\log(\lambda^{(2)}\lambda' z^{(0)})_n) - \log\log(\lambda' z^{(0)})_n \leq r - \tau_\star^L(\lambda')\right\}$$

term by term, matching $\gamma^{(2)}$ with $\lambda^{(2)}$ as in (3.26). By (3.27) we have

$$\log(\gamma^{(2)}.y(\lambda'))_n - \log y(\lambda')_n \leq \log\log(\lambda^{(2)}\lambda' z^{(0)})_n - \log\log(\lambda' z^{(0)})_n + 2\epsilon$$

and

$$\log\log(\lambda^{(2)}\lambda' z^{(0)})_n) - \log\log(\lambda' z^{(0)})_n - 2\epsilon \leq \log(\gamma^{(2)}.y(\lambda'))_n - \log y(\lambda')_n$$

from which the result follows. □

3.5 Using the linear semigroup count to prove Proposition 23

We now use Theorem 13, whose proof will be deferred to Section 4. Let $y' = y(\lambda') = f(\lambda' z^{(0)})$.

Lemma 31. *Let δ be the small constant from (3.19). We have*

$$\begin{aligned} M(\lambda z^{(0)}, r) &= (1+o(1))\, e^{\beta r} \sum_{\lambda' : s^L(\lambda') = \lambda} e^{-\beta \tau_\star^L(\lambda')} h(y') \\ &+ O\left(\exp(\beta r^\delta + (1+\delta) r)\right). \end{aligned}$$

The big and small o terms have implied constant and decay rates that are independent of $\lambda z^{(0)}$.

Proof. Using Lemma 30 in the expression (3.15) gives that up to a negligible $O(e^{(1+\delta)r})$,

$$\sum_{\lambda' : s^L(\lambda') = \lambda} N(y(\lambda'), r - 2\epsilon - \tau_\star^L(\lambda')) \leq M(\lambda z^{(0)}, r) \leq \sum_{\lambda' : s^L(\lambda') = \lambda} N(y(\lambda'), r + 2\epsilon - \tau_\star^L(\lambda')) \tag{3.28}$$

where $y(\lambda') = f(\lambda' z^{(0)})$.

We want to carefully use Theorem 13 that says that along with h, β there is some function $F(r)$ such that

$$|N(y,r) - e^{\beta r}h(y)| \le F(r)e^{\beta r}h(y)$$

and $F(r) \to 0$ as $r \to \infty$. The minor problem with using this in (3.28) is that there may be terms with $r' = r \pm 2\epsilon - \tau_\star^L(\lambda')$ close to zero, or less than zero. Letting δ be the same small parameter as before, we note that if $r' \le r^\delta$ then there is some constant $C_3 \ge 1$ such that

$$|N(y,r') - e^{\beta r'}h(y)| \le C_3 e^{\beta r'}$$

which follows from Theorem 13 when $0 \le r' \le r^\delta$ and is trivial when $r' < 0$ since then $N(y, r') = 0$.

Therefore, working with the right hand inequality of (3.28) we get

$$M(\lambda z^{(0)}, r) \le \sum_{\lambda' : s^L(\lambda') = \lambda} \left(e^{\beta r'} h(y') + \mathbf{1}\{r' \le r^\delta\} C_3 e^{\beta r'} + \mathbf{1}\{r' > r^\delta\} F(r') e^{\beta r'} h(y') \right)$$

where we write $r' = r'(\lambda') = r + 2\epsilon - \tau_\star^L(\lambda')$ and $y' = y(\lambda')$. Therefore

$$M(\lambda z^{(0)}, r) \le \left(1 + \sup_{b \ge r^\delta} F(b)\right) \sum_{\lambda' : s^L(\lambda') = \lambda} e^{\beta r'} h(y') + C_3 \sum_{\substack{\lambda' : s^L(\lambda') = \lambda \\ r' \le r^\delta}} e^{\beta a'}. \qquad (3.29)$$

For the first term in (3.29) note that by using (3.25)

$$\begin{aligned} \sum_{\lambda' : s^L(\lambda') = \lambda} e^{\beta r'} h(y') &= e^{\beta r} \sum_{\lambda' : s^L(\lambda') = \lambda} e^{2\beta\epsilon} e^{-\beta \tau_\star^L(\lambda')} h(y') \\ &= (1 + O(\exp(-2C\phi^{\frac{cr}{\log r}})))e^{\beta r} \sum_{\lambda' : s^L(\lambda') = \lambda} e^{-\beta \tau_\star^L(\lambda')} h(y'). \end{aligned}$$

The last term in (3.29) can be bounded by

$$\ll e^{\beta r} \sum_{\substack{\lambda' : s^L(\lambda') = \lambda \\ \tau_\star^L(\lambda') \ge r + 2\epsilon - r^\delta}} e^{-\beta \tau_\star^L(\lambda')}.$$

The contributions to the sum above from $M - 1 \le \tau_\star^L(\lambda') \le M$ are bounded by

$$\sum_{\lambda : s^L(\lambda') = \lambda} \mathbf{1}\{M \ge \tau_\star^L(\lambda') \ge M - 1\} e^{-\beta \tau_\star^L(\lambda')} \le c_1^L (c_0 + M)^L e^M e^{-\beta(M-1)}$$

by Lemma 25, equation (3.10). Summing this quantity over natural numbers from $M_0 = \lfloor r - r^\delta - 1 \rfloor$ to infinity, using the bound (3.12) to replace $c_1^L(c_0 + M)^L$ by $e^{\delta M}$, gives

$$\sum_{\substack{\lambda' : s^L(\lambda') = \lambda \\ \tau_\star^L(\lambda') \ge r + 2\epsilon - r^\delta}} e^{-\beta \tau_\star^L(\lambda')} \ll e^{-(\beta - 1 - \delta)(r - r^\delta)};$$

so we get for the last term in (3.29)

$$\sum_{\substack{\lambda':s^L(\lambda')=\lambda \\ r'\le r^\delta}} e^{\beta r'} \ll \exp((\beta-1-\delta)r^\delta + (1+\delta)r).$$

Therefore it can be absorbed into the error stated in the lemma. The lower bound for $M(\lambda z^{(0)}, r)$ is similar. Notice that our constants and rates of decay do not depend on $\lambda z^{(0)}$. $\square$

Proposition 23 will now follow from Lemma 31 and the following proposition.

Proposition 32. *For fixed λ and $z^{(0)}$ there is a constant $c_\star(\lambda z^{(0)})$ such that*

$$a_L(\lambda z^{(0)}) := \sum_{\lambda':s^L(\lambda')=\lambda} h(y(\lambda'))e^{-\beta\tau_\star^L(\lambda')} = c_\star(\lambda z^{(0)}) + o(1)$$

as $L\to\infty$, with a rate of decay that is independent of λ. The values $c_\star(\lambda z^{(0)})$ are bounded by some constant independent of λ.

Proof. We are going to prove the sequence is Cauchy with a very fast rate. Consider the difference of consecutive terms. Again we write $y' = y(\lambda')$. For $\lambda'' \in S_\Lambda$ we write $y'' = y''(\lambda'', \lambda') = f(\lambda''\lambda' z^{(0)})$. We suppress the dependence of these variables on others to improve readability.

We obtain

$$\begin{aligned}
a_{L+1} - a_L &= \sum_{\lambda^{(2)}=\lambda''\lambda':s^{L+1}(\lambda^{(2)})=\lambda} h(y'')e^{-\beta\tau_\star^{L+1}(\lambda^{(2)})} - \sum_{\lambda':s^L(\lambda')=\lambda} h(y')e^{-\beta\tau_\star^L(\lambda')} \\
&= \sum_{\lambda':s^L(\lambda')=\lambda} e^{-\beta\tau_\star^L(\lambda')}\left(\left(\sum_{\lambda''\in S_\Lambda} h(y'')e^{-\beta(\tau_\star^{L+1}(\lambda''\lambda')-\tau_\star^L(\lambda'))}\right) - h(y')\right) \\
&= \sum_{\lambda':s^L(\lambda')=\lambda} e^{-\beta\tau_\star^L(\lambda')}\left(\left(\sum_{\lambda''\in S_\Lambda} h(y'')e^{-\beta\tau_\star(\lambda''\lambda')}\right) - h(y')\right) \\
&= \sum_{\lambda':s^L(\lambda')=\lambda} e^{-\beta\tau_\star^L(\lambda')}\left(\left(\sum_{\lambda''\in S_\Lambda} h(y'')\left(\frac{\log(\lambda' z^{(0)})_n}{\log(\lambda''\lambda' z^{(0)})_n}\right)^\beta\right) - h(y')\right)
\end{aligned} \tag{3.30}$$

The point is that the terms in parentheses should be close to zero by the recursion (1.8) satisfied by h over Γ'. We will use Lemma 27 which gives a bound when $\alpha(\lambda' z^{(0)}) > C_1$. On the other hand by Lemma 26 there is some L_0 such that when $L \ge L_0$ and $s^L(\lambda') = \lambda$ then $\alpha(\lambda' z^{(0)}) > C_1$.

We use the natural bijection

$$S_\Lambda \to T_\Gamma, \quad \lambda'' \mapsto \gamma(\lambda'').$$

When $L > L_0$, repeating the arguments of the previous section leading up to (3.27) gives the bounds

$$\log(\lambda' z^{(0)})_n \le y'_n \le (1 + O(\alpha(\lambda' z^{(0)})^{-2}))\log(\lambda' z^{(0)})_n \tag{3.31}$$

and

$$\log(\lambda''\lambda' z^{(0)})_n \leq (\gamma(\lambda'').y')_n \leq (1 + O(\alpha(\lambda' z^{(0)})^{-2})) \log(\lambda''\lambda' z^{(0)})_n \tag{3.32}$$

where the implied constants depend only on n. Moreover, using Lemma 27 gives

$$\log(\lambda''\lambda' z^{(0)}) \leq y'' \leq \log(\lambda''\lambda' z^{(0)}) + C_2\alpha(\lambda' z^{(0)})^{-2}(0, 0, \ldots, 0, 1) \tag{3.33}$$

whenever $L > L_0$.

Suppose $L > L_0$. We must estimate the cost of replacing y'' by $\gamma(\lambda'')y'$ and $\left(\frac{\log(\lambda' z^{(0)})_n}{\log(\lambda''\lambda' z^{(0)})_n}\right)^\beta$ by $\left(\frac{y'_n}{(\gamma(\lambda'')y')_n}\right)^\beta$ in (3.30). Since using (3.32) and (3.33) gives that y'' is within $O(\alpha(\lambda' z^{(0)})^{-2} \log(\lambda''\lambda' z^{(0)})_n)$ of $\gamma(\lambda'').y'$ and h is C^1, we get

$$h(y'') = h(\gamma(\lambda'')y') + O(\alpha(\lambda' z^{(0)})^{-2} \log(\lambda''\lambda' z^{(0)})_n).$$

Using (3.31) and (3.32) gives

$$\begin{aligned} \left(\frac{y'_n}{(\gamma(\lambda'')y')_n}\right)^\beta (1 + O(\alpha(\lambda' z^{(0)})^{-2}))^{-\beta} &\leq \left(\frac{\log(\lambda' z^{(0)})_n}{\log(\lambda''\lambda' z^{(0)})_n}\right)^\beta \\ &\leq \left(\frac{y'_n}{(\gamma(\lambda'')y')_n}\right)^\beta (1 + O(\alpha(\lambda' z^{(0)})^{-2}))^\beta. \end{aligned}$$

Using that h and $\left(\frac{\log(\lambda' z^{(0)})_n}{\log(\lambda''\lambda' z^{(0)})_n}\right)^\beta$, $\left(\frac{y'_n}{(\gamma(\lambda'')y')_n}\right)^\beta$ are bounded we get

$$\begin{aligned} \sum_{\lambda'' \in S_\Lambda} h(y'') \left(\frac{\log(\lambda' z^{(0)})_n}{\log(\lambda''\lambda' z^{(0)})_n}\right)^\beta &= \sum_{\gamma'' \in T_\Gamma} h(\gamma(\lambda'')y') \left(\frac{y'_n}{(\gamma(\lambda'')y')_n}\right)^\beta + O(\alpha(\lambda' z^{(0)})^{-2}) \\ &= h(y') + O(\alpha(\lambda' z^{(0)})^{-2}) \end{aligned}$$

where the last equality uses the recursion (1.8). Therefore for $L \geq L_0$

$$|a_{L+1} - a_L| \ll \left(\sup_{\lambda' : s^L(\lambda') = \lambda} \alpha(\lambda' z^{(0)})^{-2}\right) \sum_{\lambda' : s^L(\lambda') = \lambda} e^{-\beta\tau_\star^L(\lambda')}.$$

It is possible to use a fortiori estimates to prove the sum above is universally bounded, for example by using the work of Baragar [Bar94a] in the case of $k = 0$. To keep things self contained, since we only need a coarse bound we instead use Lemma 25 to prove

$$\sum_{\lambda' : s^L(\lambda') = \lambda} e^{-\beta\tau_\star^L(\lambda')} \ll \exp(C_4 L^{1+\eta}) \tag{3.34}$$

for some constant C_4 and small η. However, $\sup_{\lambda' : s^L(\lambda') = \lambda} \alpha(\lambda' z^{(0)})^{-2}$ is much smaller than this: by Lemma 26 we have for any λ' with $s^L(\lambda') = \lambda$ that $\alpha(\lambda' z^{(0)})^{-2} \ll \exp(-2C\phi^L)$ where $\phi > 1$. So not only is

$$|a_{L+1} - a_L| \ll \exp(C_4 L^{1+\eta} - 2C\phi^L)$$

very small but we can sum the differences to get a Cauchy sequence. Indeed $C_4 L^{1+\eta} - 2C\phi^L \leq C_5 - C_6\phi^L$ for some $C_5, C_6 > 0$. Therefore for $L_1 \geq L_0$

$$\sum_{L=L_1}^{\infty} |a_{L+1} - a_L| \ll \sum_{L=L_1}^{\infty} \exp(-C_6\phi^L) = o_{L_1\to\infty}(1) \tag{3.35}$$

so the sequence converges at a uniform rate to its limit $c_\star(\lambda z^{(0)})$. The uniform boundedness of $c_\star(\lambda z^{(0)})$ will follow from the uniform boundedness of $a_{L_0}(\lambda z^{(0)})$ given (3.35), and $a_{L_0}(\lambda z^{(0)})$ is uniformly bounded by using that h is bounded and the already established (3.34). This finishes the proof. □

Putting Proposition 32 and Lemma 31 together proves Proposition 23 given Theorem 13. In the rest of the paper we prove Theorem 13.

4 The linear semigroup count

4.1 Renewal (again)

Now we discuss renewal for the quantity $N(y, r)$ that appears in Theorem 13. The renewal equation for $N(y, r)$ says

$$N(y, r) = \sum_{\gamma\in T_\Gamma} N(\gamma.y, r - \log(\gamma.y)_n + \log y_n) + \mathbf{1}\{0 \leq r\}. \tag{4.1}$$

Notice from its definition in (1.7) that the function $N(y, r)$ is invariant under multiplication of the y variable by $\mathbf{R}_+$. With this in mind, we are going to consider

$$P(\mathbf{R}^n_{\geq 0}) = \mathbf{R}^n_{\geq 0}/\mathbf{R}_+,$$

the quotient of $\mathbf{R}^n_{\geq 0}$ by the multiplicative action of positive real numbers. Let $\Delta \subset P(\mathbf{R}^n_{\geq 0})$ denote the projection of $\mathcal{H}$. We will from now on use a coordinate

$$w = (w_1, w_2, \ldots, w_{n-1}, 1)$$

with $w_1 \leq w_2 \leq \ldots \leq w_{n-1}$ and $\sum_{j=1}^{n-1} w_j = 1$ to uniquely represent a point in Δ. We now view $N(w, r)$ as a function on $\Delta\times\mathbf{R}_{\geq 0}$. Note that equation (4.1) descends to $(w, r) \in \Delta\times\mathbf{R}_{\geq 0}$.

Now, for the first time in the paper, we start the full argument of the renewal method[17]. This begins with taking a Laplace transform which we define for general f of suitable decay by

$$\hat{f}(s) = \int_{-\infty}^{\infty} e^{-sx} f(x)dx.$$

The outcome of taking a Laplace transform of the renewal equation (4.1) in the r variable, ignoring issues of convergence[18], is that

$$\hat{N}(w, s) = \sum_{\gamma\in T_\Gamma} \left(\frac{w_n}{(\gamma.w)_n}\right)^s \hat{N}(\gamma.w, s) + \frac{1}{s} \tag{4.2}$$

[17] Previously we just used an iteration of a renewal equation to perform a linearization.

[18] These issues are worked out in Lemma 35.

for all $w \in \Delta$, where $\hat{N}(w,s)$ is the Laplace transform $\widehat{N(w,\bullet)}$ in the r variable. Thus s is a frequency parameter dual to the counting parameter r. Notice that the function

$$(\gamma, w) \mapsto \frac{w_n}{(\gamma.w)_n}$$

descends from $T_\Gamma \times \mathcal{H}$ to a well defined real valued function on $T_\Gamma \times \Delta$.

Now we introduce the transfer operator that will play a crucial role in this section. For a function f on Δ we define

$$\mathcal{L}_s[f](w) = \sum_{\gamma \in T_\Gamma} \left(\frac{w_n}{(\gamma.w)_n} \right)^s f(\gamma.w) \tag{4.3}$$

whenever the sum is pointwise absolutely convergent on Δ. Then (4.2) can be rephrased as

$$\hat{N}(\bullet, s) = s^{-1}(1 - \mathcal{L}_s)^{-1}\mathbf{1}, \tag{4.4}$$

whenever the resolvent operator $(1-\mathcal{L}_s)^{-1}$ exists in such a way that it can act on the constant function $\mathbf{1}$.

There is a procedure due to Lalley to convert (4.4) together with a sufficiently complete description of the spectrum of $\mathcal{L}_s$ on a suitable Banach space into Theorem 13. More specifically we will appeal to the perturbation theory and Fourier analysis developed in [Lal89, Sections 7 and 8]. In the next section we will lay out the necessary spectral theory of $\mathcal{L}_s$. Before that, let us calculate explicitly the sum in (4.3).

Lemma 33. *An element $\gamma_{n-1}^A \gamma_j$ of T_Γ acts on Δ by*

$$\begin{aligned} &\gamma_{n-1}^A \gamma_j.[w_1, \ldots, w_{n-1}, 1] \\ &\qquad = \quad [w_1, \ldots, \widehat{w_j}, \ldots, w_{n-1}, 1 + A(1 - w_j), 1 + (A+1)(1 - w_j)]; \end{aligned} \tag{4.5}$$

in particular,

$$(\gamma_{n-1}^A \gamma_j.(w_1, \ldots w_{n-1}, 1))_n = 1 + (A+1)(1 - w_j). \tag{4.6}$$

Proof. This is a direct calculation. $\square$

Example 34 (Gauss map)**.** When $n = 3$, the only inverse branches are of the form

$$\gamma_2^A \gamma_1(w_1, w_2, 1) = (w_2, 1 + Aw_2, 1 + (A+1)w_2).$$

With the change of variables $x = w_1/w_2$, these are precisely the inverse branches of the Gauss map $x \mapsto \{\frac{1}{x}\}$:

$$\gamma_2^A \gamma_1 : x \mapsto \frac{1}{x + A + 1}, \quad A \in \mathbf{Z}_{\geq 0}.$$

4.2 Spectral theory of the transfer operator

In this section, we give a full account of the spectral theory of $\mathcal{L}_s$. A good reference for the spectral theory of transfer operators is the book of Baladi [Bal00]. We begin with the following lemma.

Lemma 35. *When $\Re(s) > 1$ the summation in the defining equation (4.3) of $\mathcal{L}_s$ is absolutely and uniformly convergent on Δ and so gives a well defined continuous map of Banach spaces*[19]

$$\mathcal{L}_s : C^0(\Delta) \to C^0(\Delta).$$

Proof. Substituting Lemma 33, equation (4.6) in the Definition (4.3), the summation amounts to

$$\mathcal{L}_s[f](w) = \sum_{j\in[n-2]} \sum_{A\in\mathbf{Z}_{\geq 0}} \frac{1}{(1+(A+1)(1-w_j))^s} f(\gamma_{n-1}^A \gamma_j . w). \tag{4.7}$$

Here and henceforth we use the notation $[N]$ for the set $\{1, 2, \ldots, N\}$. Since $w_j \leq 1/2$ for $j \in [n-2]$ and f is bounded, each sum in L converges uniformly absolutely on Δ for $\Re(s) > 1$. The limit is then continuous and bounded by a constant multiple, depending on s, of $\|f\|_\infty$. □

We obtain the following consequence of Lemma 35 by a standard application of the Schauder-Tychonoff Theorem.

Corollary 36 (Existence of eigenmeasures)**.** *Let $\mathcal{L}_s^*$ denote the dual of $\mathcal{L}_s$. For each real $s > 1$ there is a number $\lambda_s > 0$ and a probability measure ν_s such that $\mathcal{L}_s^* \nu_s = \lambda_s \nu_s$.*

Example 37 (Transfer operator for the Gauss map)**.** Let $n = 3$. Carrying on from Example 34, we have in the coordinate $x = w_1/w_2$

$$\mathcal{L}_s[f](x) = \sum_{A\in\mathbf{Z}_{\geq 0}} \frac{(x+1)^s}{(x+A+2)^s} f\left(\frac{1}{x+1+A}\right).$$

This is not the usual transfer operator for the Gauss map. However, letting $M_{(x+1)^s}$ denote the operator of multiplication by $(x+1)^s$, we get

$$M_{(x+1)^s}^{-1} \mathcal{L}_s M_{(x+1)^s}[f](x) = \sum_{A\in\mathbf{Z}_{\geq 0}} \frac{1}{(x+A+1)^s} f\left(\frac{1}{x+1+A}\right) = \mathcal{L}_s^{\text{Gauss}}[f](x),$$

the classical transfer operator for the Gauss map. This coincides with the Perron-Frobenius operator for the Gauss map when $s = 2$. The leading eigenfunction of $\mathcal{L}_2^{\text{Gauss}}$ corresponds to a multiplicity 1 eigenvalue 1 and eigenfunction

$$h(x) = \frac{1}{1+x}.$$

This eigenfunction was known to Gauss [Gau], and its invariance property was formally proved by Kuzmin [Kuz32]. Correspondingly, the leading eigenfunction of $\mathcal{L}_2$ is $[M_{(x+1)^2} h](x) = (x+1) = \frac{w_1}{w_2} + 1 = \frac{1}{w_2}$ with eigenvalue 1.

Our functional analysis takes place on the Banach space $C^1(\Delta)$ which consists of continuously differentiable functions on Δ with the norm

$$\|f\|_{C^1} = \|f\|_\infty + \|\nabla f\|_\infty.$$

We use the standard Euclidean metric on Δ given by the coordinates $w_1, \ldots, w_{n-1}$.

[19] C^0 is the Banach space of continuous functions with the supremum norm.

Lemma 38. *In the region $\Re(s) > 1$, the mapping $s \mapsto \mathcal{L}_s$ gives a holomorphic family of bounded operators on the Banach space $C^1(\Delta)$. In particular, for $\Re(s) > 1$, $\mathcal{L}_s$ is bounded on $C^1(\Delta)$.*

We will prove the following version of the Ruelle-Perron-Frobenius Theorem.

Theorem 39 (Ruelle-Perron-Frobenius)**.** *Let $s \in (1, \infty)$ be a real parameter for the transfer operator $\mathcal{L}_s : C^1(\Delta) \to C^1(\Delta)$.*

1. *The eigenvalue λ_s is multiplicity one and the rest of the spectrum of $\mathcal{L}_s$ in contained in a ball of radius $R(s)$ strictly less than λ_s. For any compact interval $I \subset (1, \infty)$ there is an $\epsilon(I) > 0$ such that $\lambda_s - R(s) \geq \epsilon$ for $s \in I$.*

2. *There is a unique probability measure ν_s such that $\mathcal{L}_s^* \nu_s = \lambda_s \nu_s$.*

3. *The unique eigenfunction $h_s \in C^1(\Delta)$ for the eigenvalue λ_s with $\nu_s(h_s) = 1$ is positive.*

In the case of the Gauss map, a version of Theorem 39 was first proved by Wirsing [Wir74]. In the case of $n = 4$, when there is a close connection between the Rauzy gasket and the dynamics of Γ' on Δ as explained in Example 9, a version of Theorem 39 was proved by Avila, Hubert, and Skripchenko in [AHS16a, Proof of Theorem 22]. There are slight differences; in [AHS16a] the authors work in a symbolic setting, so their function space is not the same as ours, whereas we need to know that $h \in C^1(\Delta)$, for example, in order to state Theorem 13.

It is well-known that Theorem 39 follows from eventually contracting dynamics for example, by the use of Birkhoff cones and contraction of a Hilbert projective metric as in the paper of Liverani [Liv95]. The only thing that is possibly nonstandard about our setting is the presence of both countably many branches and a semigroup action for which we expect the invariant set to have non full Hausdorff dimension (cf. Figures 1 and 2). We explain the proof of Lemma 38 and Theorem 39 in Section 4.4.

These proofs depend crucially on our dynamics being uniformly contracting, which we make precise in Proposition 45. We freely make use of this property henceforth. Let $T_\Gamma^{\mathbf{Z}_+}$ denote the set of all positively indexed sequences $(\gamma^{(1)}, \gamma^{(2)}, \ldots)$ with each $\gamma^{(j)} \in T_\Gamma$. Because the elements of T_Γ uniformly contract Δ, one obtains for any fixed $w_0 \in \Delta$ a map

$$\mathrm{limit} : T_\Gamma^{\mathbf{Z}_+} \to \Delta, \quad \mathrm{limit}(\gamma^{(1)}, \gamma^{(2)}, \ldots) := \lim_{j \to \infty} \gamma^{(1)} \ldots \gamma^{(j)}.w_0;$$

in fact, this map does not depend on the choice of w_0. The image of this map is the attractor of the iterated function system given by the elements of T_Γ, which we also call the *limit set* of Γ', and denote it by $\mathfrak{K}(\Gamma')$. Then $\mathfrak{K}(\Gamma')$ is a compact Γ'-invariant subset of Δ.

The Ruelle-Perron-Frobenius Theorem is not enough for input to Lalley's framework of complex analysis. One must also know that there is some non trivial spectral bound for $\mathcal{L}_s$ on the vertical line $s = \beta + it$, the trivial bound being that the spectral radius is no greater than λ_β. In the context of subshifts of finite type, this was investigated by Pollicott in [Pol84] who found a cohomological criterion for a nontrivial spectral bound. We make the following definition as in Pollicott [Pol84, pg. 139], adapted to the current setting.

Definition 40. We say that a function $f = u + iv$ with

$$u, v : T_\Gamma \times \Delta \to \mathbf{R}$$

is *regular* if there is no $r \in \mathbf{R}$ and bounded[20] function $G : \mathfrak{K}(\Gamma') \to \mathbf{R}$ such that

$$v(\gamma, w) - G(\gamma.w) + G(w) - r \in 2\pi\mathbf{Z}$$

for all $\gamma \in T_\Gamma$ and $w \in \mathfrak{K}(\Gamma')$. In other words, there is no $r \in \mathbf{R}$ so that $v - r$ is *cohomologous* on $\mathfrak{K}(\Gamma')$ to a $2\pi\mathbf{Z}$-valued function.

The following theorem can be viewed as an an extension of a result of Wielandt [Wie50] on the spectrum of finite dimensional complex matrices. It was proved by Pollicott [Pol84, Theorem 2] in the context of shifts of finite type in symbolic dynamics. The proof goes through perfectly well in our context[21] to give

Theorem 41 (Wielandt's Theorem, after Pollicott)**.** *If*

$$F_s(\gamma, w) \equiv -s \log\left(\frac{(\gamma.w)_n}{w_n}\right) \in C^1(\Delta; \mathbf{C}) \tag{4.8}$$

is regular, $\Im(s) \neq 0$, and $\Re(s) > 1$ then the spectral radius of the operator $\mathcal{L}_s : C^1(\Delta) \to C^1(\Delta)$ is strictly less than $\lambda_{\Re(s)}$.

This is applicable in the present setting:

Proposition 42. *For all $s \in \mathbf{C} - \mathbf{R}$, the function in (4.8) is regular.*

Proof. It is enough to show that for

$$\tau(\gamma, w) = \log\left(\frac{(\gamma.w)_n}{w_n}\right) = \log(\gamma.w)_n - \log w_n$$

there is no bounded G on $\mathfrak{K}(\Gamma')$ such that the values of

$$\tau'(\gamma, w) := \tau(\gamma, w) - G(\gamma.w) + G(w)$$

for $(\gamma, w) \in T_\Gamma \times \mathfrak{K}(\Gamma')$ are contained in a translate of a discrete subgroup of $\mathbf{R}$. So it is also enough to show that for any such τ', the gaps between distinct values of τ' are not bounded below.

The fundamental simple fact we use is that for $\gamma \in T_\Gamma$ and w such that $\gamma.w = w$, (from which it follows $w \in \mathfrak{K}(\Gamma')$)

$$\tau'(\gamma, w) = \tau(\gamma, w) - G(\gamma.w) + G(w) = \tau(\gamma, w).$$

Then it remains to show that gaps between distinct values of τ on the fixed points of $\gamma \in T_\Gamma$ are not bounded below. We compute that

$$\gamma_{n-1}^A \gamma_{n-2} = \begin{pmatrix} 1 & & & & & 0 \\ & \ddots & & & & \vdots \\ & & 1 & 0 & 0 & 0 \\ 0 & \cdots & 0 & 0 & 1 & 0 \\ A & \cdots & A & 0 & A & 1 \\ A+1 & \cdots & A+1 & 0 & A+1 & 1 \end{pmatrix},$$

[20] It is possible to impose more regularity on G in this definition but it is not necessary for our purposes.

[21] The main point is that our definition of regular function is strong enough to rule out $\mathcal{L}_s$ having an eigenvalue of modulus $\lambda_{\Re(s)}$. This fact is supplemented by compactness arguments relying on the Ionescu Tulcea-Marinescu type inequality that we establish in Lemma 46.

so (using the block lower triangular structure)

$$\det(\gamma_{n-1}^{A}\gamma_{n-2} - TI_n) = (1-T)^{n-3}(-T)(T^2 - (A+1)T - 1).$$

Consequently, the eigenvalues aside from 0 and 1 are

$$T = \frac{A+1 \pm \sqrt{(A+1)^2+4}}{2}.$$

Let T_+ be the largest, that is, $T_+ = \frac{A+1+\sqrt{(A+1)^2+4}}{2} > A+1$. One can find an eigenvector v_+ for T_+ where

$$v_+ = (0, 0, \ldots, 0, 1, T_+, T_+(T_+ - A)) > 0,$$

moreover, $v_+ \in \mathcal{H}$. Now, $(\gamma.v_+)_n/(v_+)_n = T_+$ and so $\tau(\gamma_{n-1}^{A}\gamma_{n-2}, [v_+]) = \log T_+$. Writing $T_+ = T_+(A)$, since $T_+(A) = A(1 + O(A^{-1}))$, it follows that $\log T_+(A+1) - \log T_+(A) = O(A^{-1}) \to 0$ as $A \to \infty$. On the other hand, the quantities $\log T_+(A+1) - \log T_+(A)$ are easily seen to be non-zero. This completes the proof. $\square$

The contour shifting argument of Lalley hinges on the behavior of the eigenvalue λ_s and, in particular, on the location of the possible real value β such that $\lambda_\beta = 1$. Since our dynamics is suitably uniformly contracting, if such a value exists it is unique:

Proposition 43. *The eigenvalue λ_s is a real analytic function of s that is strictly decreasing on $(1, \infty)$. We have $\lambda_s < 1$ for sufficiently large s. As such, any value $\beta_0 \in (1, \infty)$ such that $\lambda_{\beta_0} = 1$ is unique, and if no such β_0 exists then $\lambda_s < 1$ for all $s \in (1, \infty)$.*

As we will discuss momentarily, such a β_0 does exist, and it coincides with Baragar's β from Theorem 2. Note that when $s = \beta$ we obtain from Theorem 39 a unique measure such that $\mathcal{L}_\beta^* \nu_\beta = \nu_\beta$. Then we will show ν_β is the *conformal measure* of Theorem 10. Proposition 43 will be proved in Section 4.5.

4.3 Proofs of Theorem 10 and 13 given the spectral theorems

Here we make a sketch of the passage from the spectral theory outlined in Section 4.2 to Theorems 10 and 13 via (4.4) and the techniques of Lalley from [Lal89]. Firstly, if there is no value $\beta_0 > 1$ such that $\lambda_{\beta_0} = 1$ then Proposition 43 together with Lemma 38 imply that the resolvent $(1 - \mathcal{L}_s)^{-1}$ exists as a holomorphic family of bounded operators on $C^1(\Delta)$ in the region $\Re(s) > 1$. This would imply by standard contour shifting arguments in combination with (4.4) that for any $\eta > 0$

$$N(w, r) = O(e^{(1+\eta)r}). \tag{4.9}$$

But this can be used along with the arguments of Section 3 to show for some z in an infinite orbit of Λ that $M(z, r) = O(e^{(1+\eta)r})$, in contradiction to Baragar's result (Theorem 2) when η is small. Here we use the fact that for any n, there is an infinite orbit in $V(\mathbf{Z}_+)$ when $n = a$ and $k = 0$ coming from the tuple $(1, 1, \ldots, 1)$. In fact, for small η, (4.9) is already in contradiction to some of Baragar's results from [Bar94a] on orbits of the linear semigroup Γ.

Hence there must be $\beta_0 > 1$ with $\lambda_{\beta_0} = 1$ as in Proposition 43. Lalley's method of proof of his analog of Theorem 13 is by a contour shifting argument involving control on the meromorphic behavior of $(1 - \mathcal{L}_s)^{-1}$ in the following two ways:

1. By standard results in Linear Perturbation Theory [Kat76, Sections 4.3 and 7.1], Lemma 38 and Part 1 of Theorem 39 imply that the functions

$$s \mapsto \lambda_s,\ s \mapsto h_s,\ s \mapsto \nu_s$$

extend to holomorphic functions on a neighborhood of the real line segment $(1, \infty)$ in $\Re(s) > 1$ such that

$$\lambda_s \neq 0,\ \mathcal{L}_s h_s = \lambda_s h_s,\ \mathcal{L}_s^* \nu_s = \lambda_s \nu_s,\ \nu_s(h_s) = 1.$$

By suitable spectral decomposition of $\mathcal{L}_s$, one finds a neighborhood U of $s = \beta_0$ and an operator $\mathcal{L}'_s$ such that $(1 - \mathcal{L}'_s)^{-1}$ is a holomorphic family of bounded operators on $C^1(\Delta)$ for $s \in U$ and moreover

$$(1 - \mathcal{L}_s)^{-1} g = (1 - \lambda_s)^{-1} \nu_s(g) h_s + (1 - \mathcal{L}'_s)^{-1} g$$

for $s \in U - \{\beta_0\}$. This is the analog of [Lal89, Proposition 7.2].

2. By use of Theorem 41 along with its supplement Proposition 42, we obtain that

$$s \mapsto (1 - \mathcal{L}_s)^{-1}$$

is holomorphic in a neighborhood of every s with $\Re(s) = \beta_0$, with the exception of $s = \beta_0$.

The outcome of Lalley's argument is that

$$N(w, r) = h_{\beta_0}(w) e^{\beta_0 r} + o(e^{\beta_0 r})$$

where the decay in the small o does not depend on w. Our argument of Section 3.4 converts this into a version of Theorem 3 with β replaced by β_0. Finally, this contradicts Baragar's Theorem 2 unless $\beta = \beta_0$. Then Theorem 13 is proved, assuming the theorems of Section 4.2.

Theorem 10 is now a direct consequence of the following fact:

Lemma 44. *For all $\gamma \in \Gamma$ we have*

$$\frac{(\gamma.w)_n}{w_n} = |\mathrm{Jac}_w(\gamma)|^{-\frac{1}{n-1}}$$

where $|\mathrm{Jac}_w(\gamma)|$ is the absolute value of the Jacobian determinant of γ acting on $\Delta = \mathcal{H}/\mathbf{R}_+$ at the point w.

This can be checked by a direct calculation on general grounds as in [Pola, Lemma 2.1], or by using explicit formulae that appear later in this paper, e.g. by calculating the determinants of total derivatives we calculate in Section 5.

4.4 Consequences of uniformly contracting dynamics

The spectral theorems of the previous section all rely on the action of Γ' on Δ being by contractions. That can be summarized in the following proposition.

Proposition 45. *There are constants $D > 0$ and $\rho < 1$ such that for all $\gamma^{(1)}, \gamma^{(2)}, \ldots, \gamma^{(N)} \in T_\Gamma$ we have*

$$\|d_w[\gamma^{(1)}\gamma^{(2)}\ldots\gamma^{(N)}]\|_{\mathrm{op}} \leq D\rho^N.$$

Here we view $\gamma^{(1)}\gamma^{(2)}\ldots\gamma^{(N)}$ as self-maps of Δ, using the fixed Euclidean metric on Δ, d_w is the total derivative of the map at $w \in \Delta$, and $\|\bullet\|_{\mathrm{op}}$ is the operator norm of the map between tangent spaces (using the ℓ^2 norms coming from the metric).

When $n = 4$, modulo translation between the Rauzy gasket and our dynamical system, a proof of Proposition 45 was outlined by Arnoux and Starosta in [AS13, Lemma 2] and given in more detail by Avila, Hubert, and Skripchenko in [AHS16b, Lemma 13].

We will prove Proposition 45 for all $n \geq 4$ in Section 5. The dynamical Proposition 45 gets brought into play by the following two-norm inequality with origins in the work of Ionescu Tulcea and Marinescu [ITM50].

Lemma 46. *There is $C > 0$ such that for any $Q \in \mathbf{N}$ and $\Re(s) > 1$*

$$\|\nabla\mathcal{L}_s^Q[f](w)\|_2 \leq C|s|\mathcal{L}_s^Q[|f|](w) + D\rho^Q\mathcal{L}_s^Q[\|\nabla f\|_2](w)$$

for all $w \in \Delta$. We write $\|\bullet\|_2$ for the pointwise ℓ^2 norm in an individual tangent space fiber.

Proof. This is standard given Proposition 45: it essentially boils down to the chain rule. The only things to take care of are the infinite sums that arise, but these are all absolutely convergent when $\Re(s) > 1$. $\square$

We can now prove Lemma 38.

Proof of Lemma 38. We are proving $s \mapsto \mathcal{L}_s$ is a holomorphic mapping to bounded operators on $C^1(\Delta)$. If we truncate the summation going into the expression (4.7) for $\mathcal{L}_s$ at some fixed B to form

$$\mathcal{L}_s^{(B)} = \sum_{j\in[n-2]}\sum_{A\leq B}\frac{1}{(1+(A+1)(1-w_j))^s}f(\gamma_{n-1}^A\gamma_j.w);$$

the resulting $\mathcal{L}_s^{(B)}$ is easily seen to be holomorphic by taking a complex derivative. So it remains to show that $\mathcal{L}_s^{(B)} \to \mathcal{L}_s$ uniformly on compact sets, say in the norm topology. But the tail consists of $n-2$ terms of the form

$$(\mathcal{L}_s - \mathcal{L}_s^{(B)})[f](w) = \sum_{A>B}\frac{1}{(1+(A+1)(1-w_j))^s}f(\gamma_{n-1}^A\gamma_j.w).$$

Then $\|\mathcal{L}_s - \mathcal{L}_s^{(B)}\|_{C^0} \to 0$ as $B \to \infty$ and this is uniform for s in W, a compact subset of $\Re(s) > 1$. On the other hand, the proof of Lemma 46 also applies to $\mathcal{L}_s - \mathcal{L}_s^{(B)}$, so applying it when $Q = 1$ gives

$$\|\nabla(\mathcal{L}_s - \mathcal{L}_s^{(B)})[f]\|_\infty \leq C|s|\|(\mathcal{L}_s - \mathcal{L}_s^{(B)})[|f|]\|_\infty + D\rho\|(\mathcal{L}_s - \mathcal{L}_s^{(B)})[\|\nabla f\|_2]\|_\infty.$$

This implies

$$\|\mathcal{L}_s - \mathcal{L}_s^{(B)}\|_{C^1(\Delta)} \ll_W \|\mathcal{L}_s - \mathcal{L}_s^{(B)}\|_{C^0(\Delta)},$$

which we've established goes to zero uniformly on W. $\square$

The proof of the Ruelle-Perron-Frobenius Theorem 39 now proceeds either via use of Birkhoff cones as in Liverani's paper [Liv95] or by a more direct approach as in Pollicott [Pola, Lemma 2.3]. The classical proof of this Theorem for subshifts of finite type can be found in [PP90, Theorem 2.2]. In any approach Lemma 46 is the key input. The uniform spectral gap stated in Part 1 of Theorem 39 is a consequence of the uniformity of Lemma 46 for s in a fixed compact subinterval of $(1, \infty)$.

4.5 Behavior of the eigenvalue

In this section we prove Proposition 43. The statement that λ_s is real analytic on $(1, \infty)$ follows from the fact we noted in the previous Section 4.2 that by perturbation theory in combination with Theorem 39 Part 1

$$s \mapsto \lambda_s$$

is holomorphic in a neighborhood of $(1, \infty)$ in $\Re(s) > 1$. Notice that we have the bound

$$\begin{aligned} \mathcal{L}_s[f](w) &= \sum_{j \in [n-2]} \sum_{A \in \mathbf{Z}_{\geq 0}} \frac{1}{(1 + (A+1)(1-w_j))^s} f(\gamma_{n-1}^A \gamma_j . w) \\ &\leq (n-2)\|f\|_\infty \sum_{A \in \mathbf{Z}_{\geq 0}} \frac{1}{(1 + \frac{1}{2}(A+1))^s} \leq 2(n-2)\|f\|_\infty \sum_{A \in \mathbf{Z}_{\geq 0}} \frac{1}{(3+A)^s}. \end{aligned}$$

Letting $f = h_s$ and w such that $h_s(w) = \|h_s\|_\infty$ gives

$$\lambda_s \leq 2(n-2) \sum_{A \in \mathbf{Z}_{\geq 0}} \frac{1}{(3+A)^s}$$

so $\lambda_s \to 0$ as $s \to \infty$.

It remains to show that λ_s is strictly decreasing in s. Let I be a fixed compact subinterval of $(1, \infty)$. By Theorem 39 $\lambda_s^{-N}\mathcal{L}_s^N 1$ converges in C^1 norm to h_s and this convergence is uniform for $s \in I$. This implies

$$\log \lambda_s = \frac{\log\left(\mathcal{L}_s^N[1](w)\right)}{N} + o(1) \tag{4.10}$$

as $N \to \infty$, where the error is uniform in $s \in I$ and $w \in \Delta$. We calculate

$$\mathcal{L}_s^N[1](w) = \sum_{\gamma \in (T_\Gamma)^N} \left(\frac{(\gamma.w)_n}{w_n}\right)^{-s}, \quad \frac{d}{ds}\mathcal{L}_s^N[1](w) = \sum_{\gamma \in (T_\Gamma)^N} -\log\left(\frac{(\gamma.w)_n}{w_n}\right)\left(\frac{(\gamma.w)_n}{w_n}\right)^{-s}.$$

Now we make the *Claim*: There is some $c > 0$ such that

$$\log\left(\frac{(\gamma.w)_n}{w_n}\right) \geq cN.$$

for all $\gamma \in (T_\Gamma)^N$. Assuming the Claim we get

$$\frac{d}{ds}\mathcal{L}_s^N[1](w) \le -cNL_s^N[1](w)$$

and hence

$$\frac{d}{ds}\log \mathcal{L}_s^N[1](w) \le -cN.$$

This means $\log \lambda_s$ is a uniform limit of functions with derivatives bounded above by a negative constant, so λ_s must be strictly decreasing as required.

To prove the Claim it is enough to show (by expanding $\log(\gamma.w)_n - \log w_n$ as a telescoping sum) that for all $w \in \Delta$ and $\gamma' = \gamma_{n-1}^A \gamma_j \in T_\Gamma$

$$\frac{(\gamma'.w)_n}{w_n} = 1 + (A+1)(1-w_j) \ge \frac{3}{2},$$

since $w_j \le 1/2$. This completes the proof of Proposition 43.

5 Proof of uniform contraction

In this section we prove Proposition 45 asserting that the elements of T_Γ eventually uniformly contract Δ.

5.1 Setup

We define the sets

$$\Delta \equiv \{(w_1, w_2, \ldots, w_{n-2}, w_{n-1}) : 0 \le w_1 \le w_2 \le \ldots \le w_{n-2} \le w_{n-1} \le 1, \sum_{i\in[n-1]} w_i = 1\},$$

$$\Delta_{\text{core}} \equiv \{(w_1, w_2, \ldots, w_{n-2}, w_{n-1}) \in \Delta : 0 \le w_{n-1} - \sum_{j\in[n-2]} w_j \le w_{n-2}\},$$

and

$$\Delta_{\text{cusp}} \equiv \{(w_1, w_2, \ldots, w_{n-2}, w_{n-1}) \in \Delta : w_{n-1} - \sum_{j\in[n-2]} w_j \ge w_{n-2}\}$$

where we continue to use the notation $[N] = \{1, 2, \ldots, N\}$. We also define the set

$$\Delta_0 \equiv \Delta_{\text{core}} \cup \Delta_{\text{cusp}}.$$

Recall that the elements of T_Γ are all of the form $\gamma = \gamma_{n-1}^A \gamma_i$ where $A \in \mathbf{Z}_{\ge 0}$ and $i = 1, 2, \ldots, n-2$. Note that for each $w \in \Delta$, we have $\gamma_i(w) \in \Delta_{\text{core}}$ for $i = 1, 2, \ldots, n-2$ and $\gamma_{n-1}(w) \in \Delta_{\text{cusp}}$. In particular, $\gamma(w) \in \Delta_0$ for all $\gamma \in T_\Gamma$ and $w \in \Delta$.

From now on, we choose to use $n-2$ coordinates in Δ instead of $n-1$, using the relationship $w_{n-1} = 1 - \sum_{i\in[n-2]} w_i$.

Note that on Δ_0, $w_{n-1} \ge \frac{1}{2}$ so that $\sum_{j\in[n-2]} w_j \le \frac{1}{2}$. On Δ_{core}, we have $w_{n-j} \le \frac{1}{2(j-1)}$ for $2 \le j \le n-1$; in particular $w_{n-2} \le \frac{1}{2}$ and $w_{n-3} \le \frac{1}{4}$. Next, on Δ_{cusp}, we have $w_{n-1} > (1 + w_{n-2})/2$, so that $w_{n-j} \le \frac{1}{2j-1}$ for $j \ge 2$. In particular, on Δ_{cusp} we have $w_{n-2} \le \frac{1}{3}$ and $w_{n-3} \le \frac{1}{5}$.

Remark 47. It is clear that it is sufficient to prove Proposition 45 with the local ℓ^2 operator norms replaced by local ℓ^1 norms, since the norms are equivalent. This is the approach we take below.

5.2 Overview of the proof of Proposition 45

We will now prove Proposition 45 (the ℓ^1 norm variant). We will appeal to the following bounds.

$$\|d\gamma_i\|_1 \leq \frac{2}{2-w_i} \leq \begin{cases} \frac{6}{5} & \text{on } \Delta_{\text{cusp}} \\ \frac{4}{3} & \text{on } \Delta_{\text{core}} \end{cases}, 1 \leq i \leq n-2 \tag{5.1}$$

$$\|d\gamma_{n-1}\|_1 = \frac{1+2(w_1+w_2+\ldots w_{n-2})-2w_1}{(1+w_1+w_2+\ldots w_{n-2})^2} \leq 1 \text{ on } \Delta_0 \tag{5.2}$$

$$\|d(\gamma_i \circ \gamma_j)\|_1 \leq \frac{2}{4-2w_j-w_i} \leq \frac{4}{5} \text{ on } \Delta_0, 1 \leq i < j \leq n-2 \tag{5.3}$$

$$\|d(\gamma_i \circ \gamma_j)\|_1 \leq \frac{2}{4-2w_j-w_{i+1}} \leq \frac{4}{5} \text{ on } \Delta_0, 1 \leq j \leq i < n-2 \tag{5.4}$$

$$\|d(\gamma_{n-2} \circ \gamma_j)\|_1 \leq \frac{4+2(w_1+\ldots+w_{n-2})-2w_1-3w_j}{3+(w_1+\ldots+w_{n-2})-2w_j} \leq \frac{4}{5} \text{ on } \Delta_0, 1 \leq j \leq n-2 \tag{5.5}$$

$$\|d(\gamma_{n-1} \circ \gamma_i)\|_1 \leq \frac{2}{3-2w_i} \leq \begin{cases} \frac{10}{13} & \text{on } \Delta_{\text{cusp}} \\ \frac{4}{5} & \text{on } \Delta_{\text{core}} \end{cases}, 1 \leq i \leq n-3 \tag{5.6}$$

$$\|d(\gamma_{n-1} \circ \gamma_{n-2})\|_1 \leq \frac{2}{3-2w_{n-2}} \leq \begin{cases} \frac{6}{7} & \text{on } \Delta_{\text{cusp}} \\ 1 & \text{on } \Delta_{\text{core}} \end{cases} \tag{5.7}$$

$$\|d(\gamma_i \circ \gamma_{n-1} \circ \gamma_{n-2})\|_1 \leq \frac{2}{6-4w_{n-2}-w_i} \leq \frac{4}{7} \text{ on } \Delta_0, 1 \leq i \leq n-3 \tag{5.8}$$

$$\|d(\gamma_{n-2} \circ \gamma_{n-1} \circ \gamma_{n-2})\|_1 \leq \frac{7+2(w_1+\ldots+w_{n-2})-2w_1-6w_{n-2}}{5+(w_1+\ldots+w_{n-2})-4w_{n-2}} \leq \frac{32}{49} \text{ on } \Delta_0 \tag{5.9}$$

$$\|d(\gamma_{n-1} \circ \gamma_{n-1} \circ \gamma_{n-2})\|_1 \leq \frac{2}{4-3w_{n-2}} \leq \begin{cases} \frac{2}{3} & \text{on } \Delta_{\text{cusp}} \\ \frac{4}{5} & \text{on } \Delta_{\text{core}} \end{cases} \tag{5.10}$$

These bounds can be proved by direct calculation, we will explain (5.1) in detail below to illustrate the method. The full proofs of (5.2)-(5.10) can be found in the arXiv posting [GMR18]. Using these bounds we can prove the following result for any $n \geq 3$ which implies Proposition 45 via Remark 47.

Lemma 48. *Given the bounds (5.1)-(5.10),* $\left\| d(\gamma_{n-1}^A \circ \gamma_i \circ \gamma_{n-1}^B \circ \gamma_j)\big|_{\Delta_0} \right\|_1 \leq \frac{24}{25}$ *for each* $A, B \in \mathbf{Z}_{\geq 0}$*, and each* $i, j = 1, 2, \ldots, n-2$*.*

Proof. Throughout this proof, we repeatedly use the fact that $\gamma_k(w) \in \Delta_{\text{core}}$ for $k = 1, 2, \ldots, n-2$ and $\gamma_{n-1}(w) \in \Delta_{\text{cusp}}$. We distinguish 3 cases.
Case I: $A \geq 1, B \geq 1$**:**

Using equations (5.2), (5.6) and (5.7), we have

$$\left\| d(\gamma_{n-1}^{A} \circ \gamma_i \circ \gamma_{n-1}^{B} \circ \gamma_j)|_{\Delta_0} \right\|_1$$
$$\leq \left\| d\gamma_{n-1}^{A-1}|_{\Delta_{\text{cusp}}} \right\|_1 \left\| d(\gamma_{n-1} \circ \gamma_i)|_{\Delta_{\text{cusp}}} \right\|_1 \left\| d\gamma_{n-1}^{B-1}|_{\Delta_{\text{cusp}}} \right\|_1 \left\| d(\gamma_{n-1} \circ \gamma_j)|_{\Delta_0} \right\|_1 \leq 1 \cdot \frac{6}{7} \cdot 1 \cdot 1 < \frac{24}{25}.$$

Case II: $A \geq 0, B = 0$**:**
Using equations (5.2), (5.3), (5.4), (5.5), we have

$$\left\| d(\gamma_{n-1}^{A} \circ \gamma_i \circ \gamma_j)|_{\Delta_0} \right\|_1 \leq \left\| d\gamma_{n-1}^{A}|_{\Delta_{\text{core}}} \right\|_1 \left\| d(\gamma_i \circ \gamma_j)|_{\Delta_0} \right\|_1 \leq 1 \cdot \frac{4}{5} < \frac{24}{25}.$$

Case III: $A = 0, B \geq 1$**:**
We first suppose that $j \leq n-3$. Then by equations (5.1), (5.2), (5.6) we have

$$\left\| d(\gamma_i \circ \gamma_{n-1}^{B} \circ \gamma_j)|_{\Delta_0} \right\|_1 \leq \left\| d\gamma_i|_{\Delta_{\text{cusp}}} \right\|_1 \left\| d\gamma_{n-1}^{B-1}|_{\Delta_{\text{cusp}}} \right\|_1 \left\| d(\gamma_{n-1} \circ \gamma_j)|_{\Delta_0} \right\|_1 \leq \frac{6}{5} \cdot 1 \cdot \frac{4}{5} = \frac{24}{25}.$$

Finally, if $j = n-2$ we are left with two subcases. If $B = 1$, then by equations (5.8) and (5.9) we have

$$\left\| d(\gamma_i \circ \gamma_{n-1} \circ \gamma_{n-2})|_{\Delta_0} \right\|_1 \leq \frac{32}{49} < \frac{24}{25}.$$

Otherwise, we have $B \geq 2$ and by equations (5.1), (5.2), (5.10) we have

$$\begin{aligned} \left\| d(\gamma_i \circ \gamma_{n-1}^{B} \circ \gamma_{n-2})|_{\Delta_0} \right\|_1 &\leq \left\| d\gamma_i|_{\Delta_{\text{cusp}}} \right\|_1 \left\| d\gamma_{n-1}^{B-2}|_{\Delta_{\text{cusp}}} \right\|_1 \left\| d(\gamma_{n-1} \circ \gamma_{n-1} \circ \gamma_{n-2})|_{\Delta_0} \right\|_1 \\ &\leq \frac{6}{5} \cdot 1 \cdot \frac{4}{5} = \frac{24}{25}. \end{aligned}$$

□

In the remainder of this section we will prove equation (5.1); the remaining (5.2)-(5.10) can be proved similarly. In the following, we define

$$w \equiv (w_1, w_2, \ldots, w_{n-2}, 1 - \sum_{k=1}^{n-2} w_k, 1),$$

and

$$\beta(w) \equiv \sum_{k=1}^{n-2} w_k.$$

Also recall that the $\|\cdot\|_1$ of a matrix is equal to the maximum over columns of the matrix of the sum of the absolute values of the column. From now on, we call such a sum an *absolute column sum*.

Proof of equation (5.1)

For $i = 1, 2, \ldots, n-2$ we have

$$\gamma_i(w) = (w_1, w_2, \ldots, \widehat{w_i}, \ldots, w_{n-2}, 1-\beta(w), 1, 2-w_i),$$

which, after projectivizing and removing the placeholder components, gives

$$\gamma_i(w) = \left(\frac{w_1}{2-w_i}, \frac{w_2}{2-w_i}, \ldots, \widehat{\frac{w_i}{2-w_i}}, \ldots, \frac{w_{n-2}}{2-w_i}, \frac{1-\beta(w)}{2-w_i} \right)$$

which is a function in $(n-2)$ variables with $(n-2)$ components. The $(n-2)\times(n-2)$ total derivative $d\gamma_i$ is given by the following matrix

$$\begin{array}{cc}
 & \begin{array}{cccccccccc} 1 & 2 & 3 & \ldots & i-1 & i & i+1 & \ldots & n-3 & n-2 \end{array} \\
\begin{array}{c} 1 \\ 2 \\ 3 \\ \vdots \\ i-1 \\ i \\ \vdots \\ n-3 \\ n-2 \end{array} &
\begin{bmatrix}
\frac{1}{2-w_i} & 0 & 0 & \cdots & 0 & \frac{w_1}{(2-w_i)^2} & 0 & \cdots & 0 & 0 \\
0 & \frac{1}{2-w_i} & 0 & \cdots & 0 & \frac{w_2}{(2-w_i)^2} & 0 & \cdots & 0 & 0 \\
0 & 0 & \frac{1}{2-w_i} & \cdots & 0 & \frac{w_2}{(2-w_i)^2} & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & \frac{1}{2-w_i} & \frac{w_{i-1}}{(2-w_i)^2} & 0 & \cdots & 0 & 0 \\
0 & 0 & 0 & \cdots & 0 & \frac{w_{i+1}}{(2-w_i)^2} & \frac{1}{2-w_i} & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & 0 & \frac{w_{n-2}}{(2-w_i)^2} & 0 & \cdots & 0 & \frac{1}{2-w_i} \\
\frac{-1}{2-w_i} & \frac{-1}{2-w_i} & \frac{-1}{2-w_i} & \cdots & \frac{-1}{2-w_i} & \frac{-1+w_i-\beta(w)}{(2-w_i)^2} & \frac{-1}{2-w_i} & \cdots & \frac{-1}{2-w_i} & \frac{-1}{2-w_i}
\end{bmatrix}
\end{array}$$

where the row and column indices are indicated to the left and above respectively. Each of these partial derivatives is immediate, except for the $(n-2, i)$ entry which follows from an application of the quotient rule. Note that the sign of entry $(n-2, i)$ is negative on Δ_0. The signs of the other entries are self-evident.

The absolute column sum for each column k with $k \neq i$ is

$$C_k = \frac{2}{2-w_i}.$$

For column $k = i$ the absolute column sum is

$$C_i = \frac{1 + 2\beta(w) - 2w_i}{(2-w_i)^2}.$$

We must compute which absolute column sum is maximal on Δ_0. Note that on Δ_0 we have $\beta(w) \leq \frac{1}{2}$. Furthermore, we have the following equivalences:

$$C_i \leq C_k, k \neq i \qquad \Leftrightarrow \qquad 1 + 2\beta(w) - 2w_i < 4 - 2w_i \qquad \Leftrightarrow \qquad \beta(w) < \frac{2}{3}.$$

Every column $k \neq i$ is maximal and $\|d\gamma_i\|_1 \leq \dfrac{2}{2-w_i}$ on Δ_0. For each i, we have $w_i \leq \frac{1}{3}$ on Δ_{cusp}, and $w_i \leq \frac{1}{2}$ on Δ_{core}. This gives the bound that $\|d\gamma_i\|_1 \leq \frac{6}{5}$ on Δ_{cusp} and $\leq \frac{4}{3}$ on Δ_{core}, proving equation (5.1).

References

[AHS16a] A. Avila, P. Hubert, and A. Skripchenko. Diffusion for chaotic plane sections of 3-periodic surfaces. *Invent. Math.*, 206(1):109–146, 2016.

[AHS16b] A. Avila, P. Hubert, and A. Skripchenko. On the Hausdorff dimension of the Rauzy gasket. *Bull. Soc. Math. France*, 144(3):539–568, 2016.

[Aig13] M. Aigner. *Markov's theorem and 100 years of the uniqueness conjecture*. Springer, Cham, 2013. A mathematical journey from irrational numbers to perfect matchings.

[AS13] P. Arnoux and Š. Starosta. The Rauzy gasket. In *Further developments in fractals and related fields*, Trends Math., pages 1–23. Birkhäuser/Springer, New York, 2013.

[Bal00] V. Baladi. *Positive transfer operators and decay of correlations*, volume 16 of *Advanced Series in Nonlinear Dynamics*. World Scientific Publishing Co., Inc., River Edge, NJ, 2000.

[Bar94a] A. Baragar. Asymptotic growth of Markoff-Hurwitz numbers. *Compositio Math.*, 94(1):1–18, 1994.

[Bar94b] A. Baragar. Integral solutions of Markoff-Hurwitz equations. *J. Number Theory*, 49(1):27–44, 1994.

[Bar98] A. Baragar. The exponent for the Markoff-Hurwitz equations. *Pacific J. Math.*, 182(1):1–21, 1998.

[Bel01] G.V. Belyiĭ. Markov's numbers and quadratic forms. *J. Math. Sci., New York*, 106(4):3087–3097, 2001.

[Bom07] E. Bombieri. Continued fractions and the Markoff tree. *Expo. Math.*, 25(3):187–213, 2007.

[Boy71] D.W. Boyd. The disk-packing constant. *Aequationes Math.*, 7:182–193, 1971.

[Boy73] D.W. Boyd. Improved bounds for the disk-packing constant. *Aequationes Math.*, 9:99–106, 1973.

[Boy82] D.W. Boyd. The sequence of radii of the Apollonian packing. *Math. Comp.*, 39(159):249–254, 1982.

[Cas57] J.W.S. Cassels. *An introduction to Diophantine approximation*. Cambridge Tracts in Mathematics and Mathematical Physics, No. 45. Cambridge University Press, New York, 1957.

[Cay69] A. Cayley. A memoir on cubic surfaces. *Philosophical Transactions of the Royal Society of London*, 159:231–326, 1869.

[De 08] R. De Leo. On a generalized Sierpinski fractal in RP^n. *arXiv:0804.1154*, April 2008.

[DL15] R. De Leo. A conjecture on the Hausdorff dimension of attractors of real self-projective iterated function systems. *Exp. Math.*, 24(3):270–288, 2015.

[DLD09] R. De Leo and I.A. Dynnikov. Geometry of plane sections of the infinite regular skew polyhedron $\{4, 6 \mid 4\}$. *Geom. Dedicata*, 138:51–67, 2009.

[Gau] C.F. Gauss. Brief an Laplace vom 30 Jan. 1812 Werke X_1 pp 371-374.

[GMR18] A. Gamburd, M. Magee, and R. Ronan. An asymptotic formula for integer points on Markoff-Hurwitz varieties. *arXiv:1603.06267v3*, June 2018.

[Gol03] W.M. Goldman. The modular group action on real SL(2)-characters of a one-holed torus. *Geom. Topol.*, 7:443–486, 2003.

[GS17] A. Ghosh and P. Sarnak. Integral points on Markoff type cubic surfaces. *arXiv://1706.06712 v1*, October 2017.

[Gur76] C. Gurwood. *Diophantine approximation and the Markov chain.* PhD thesis, New York University, 1976.

[HN17] Y. Huang and P. Norbury. Simple geodesics and Markoff quads. *Geom. Dedicata*, 186:113–148, 2017.

[Hor75] R.D. Horowitz. Induced automorphisms on Fricke characters of free groups. *Transactions of the American Mathematical Society*, 208:41–50, 1975.

[HPZ18] H. Hu, S. Peow Tan, and Y. Zhang. Polynomial automorphisms of $\mathbf{C}^n$ preserving the Markoff-Hurwitz polynomial. *Geom. Dedicata*, 192:207–243, 2018.

[Hur07] A. Hurwitz. Über eine Aufgabe der unbestimmten Analysis. *Archiv. Math. Phys.*, 3:185–196, 1907. Also: Mathematisch Werke, Vol. 2, Chapt. LXX (1933 and 1962), 410–421.

[ITM50] C.T. Ionescu Tulcea and G. Marinescu. Théorie ergodique pour des classes d'opérations non complètement continues. *Ann. of Math. (2)*, 52:140–147, 1950.

[Kat76] T. Kato. *Perturbation theory for linear operators.* Springer-Verlag, Berlin-New York, second edition, 1976. Grundlehren der Mathematischen Wissenschaften, Band 132.

[KO11] A. Kontorovich and H. Oh. Apollonian circle packings and closed horospheres on hyperbolic 3-manifolds. *J. Amer. Math. Soc.*, 24(3):603–648, 2011. With an appendix by Oh and Nimish Shah.

[Kuz32] R.O. Kuzmin. On a problem of Gauss. *Atti del Congresso Internazionale dei Matematici, Bologna*, 6:83–89, 1932.

[Lal89] S.P. Lalley. Renewal theorems in symbolic dynamics, with applications to geodesic flows, non-Euclidean tessellations and their fractal limits. *Acta Math.*, 163(1-2):1–55, 1989.

[Lev93] G. Levitt. La dynamique des pseudogroupes de rotations. *Invent. Math.*, 113(3):633–670, 1993.

[Liv95] C. Liverani. Decay of correlations. *Ann. of Math. (2)*, 142(2):239–301, 1995.

[Mag18] M. Magee. Counting One-Sided Simple Closed Geodesics on Fuchsian Thrice Punctured Projective Planes. *International Mathematics Research Notices*, page rny112, 2018.

[Mar79] A. Markoff. Sur les formes binaires indéfinies. *Math. Ann.*, 15:381–409, 1879.

[Mar80] A. Markoff. Sur les formes quadratiques binaires indéfinies. *Math. Ann.*, 17(3):379–399, 1880.

[McS91] G. McShane. *A remarkable identity for lengths of curves*. ProQuest LLC, Ann Arbor, MI, 1991. Thesis (Ph.D.)–University of Warwick (United Kingdom).

[McS98] G. McShane. Simple geodesics and a series constant over Teichmuller space. *Invent. Math.*, 132(3):607–632, 1998.

[Mir08] M. Mirzakhani. Growth of the number of simple closed geodesics on hyperbolic surfaces. *Ann. of Math. (2)*, 168(1):97–125, 2008.

[Mir16] M. Mirzakhani. Counting Mapping Class group orbits on hyperbolic surfaces. *arXiv:1601.03342*, January 2016.

[Mor53] L.J. Mordell. On the integer solutions of the equation $x^2 + y^2 + z^2 + 2xyz = n$. *J. London Math. Soc.*, 28:500–510, 1953.

[MR95] G. McShane and I. Rivin. A norm on homology of surfaces and counting simple geodesics. *Internat. Math. Res. Notices*, (2):61–69 (electronic), 1995.

[Nov82] S.P. Novikov. The Hamiltonian formalism and a multivalued analogue of Morse theory. *Uspekhi Mat. Nauk*, 37(5(227)):3–49, 248, 1982.

[Pat76] S.J. Patterson. The limit set of a Fuchsian group. *Acta Math.*, 136(3-4):241–273, 1976.

[Pola] M. Pollicott. Apollonian circle packings. Available at https://homepages.warwick.ac.uk/~masdbl/apollo-29Dec2014.pdf.

[Polb] M. Pollicott. Statistical properties of the Rauzy-Veech-Zorich map. Available at http://homepages.warwick.ac.uk/~masdbl/teichmuller-asip.pdf.

[Pol84] M. Pollicott. A complex Ruelle-Perron-Frobenius theorem and two counterexamples. *Ergodic Theory Dynam. Systems*, 4(1):135–146, 1984.

[PP83] W. Parry and M. Pollicott. An analogue of the prime number theorem for closed orbits of Axiom A flows. *Ann. of Math. (2)*, 118(3):573–591, 1983.

[PP90] W. Parry and M. Pollicott. Zeta functions and the periodic orbit structure of hyperbolic dynamics. *Astérisque*, (187-188):268, 1990.

[PU17] M. Pollicott and M. Urbanski. Asymptotic Counting in Conformal Dynamical Systems. *arXiv e-prints*, page arXiv:1704.06896, April 2017.

[Sil89] J.H. Silverman. Integral points on curves and surfaces. In *Number theory (Ulm, 1987)*, volume 1380 of *Lecture Notes in Math.*, pages 202–241. Springer, New York, 1989.

[Sil95] J.H. Silverman. Counting integer and rational points on varieties. *Astérisque*, (228):4, 223–236, 1995. Columbia University Number Theory Seminar (New York, 1992).

[SM57] H. Schwartz and H.T. Muhly. On a class of cubic Diophantine equations. *J. London Math. Soc.*, 32:379–382, 1957.

[Sul79] D. Sullivan. The density at infinity of a discrete group of hyperbolic motions. *Inst. Hautes Études Sci. Publ. Math.*, (50):171–202, 1979.

[Sul84] D. Sullivan. Entropy, Hausdorff measures old and new, and limit sets of geometrically finite Kleinian groups. *Acta Math.*, 153(3-4):259–277, 1984.

[Wie50] H. Wielandt. Unzerlegbare, nicht negative Matrizen. *Math. Z.*, 52:642–648, 1950.

[Wir74] E. Wirsing. On the theorem of Gauss-Kusmin-Lévy and a Frobenius-type theorem for function spaces. *Acta Arith.*, 24:507–528, 1973/74. Collection of articles dedicated to Carl Ludwig Siegel on the occasion of his seventy-fifth birthday, V.

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

[Zor06] A. Zorich. Flat surfaces. In *Frontiers in number theory, physics, and geometry. I*, pages 437–583. Springer, Berlin, 2006.

Alex Gamburd,
CUNY Graduate Center,
New York, NY, USA
agamburd@gc.cuny.edu

Ryan Ronan,
Baruch College (CUNY),
New York, NY, USA
ryan.ronan@baruch.cuny.edu

Michael Magee,
Durham University,
Durham, UK
michael.r.magee@durham.ac.uk