hexagon:2611.00216v3

ALMOST ALL K-MARKOV NUMBERS ARE UNIQUE

QIYUE TANG

Abstract. For every fixed nonnegative integer k, we prove that almost every positive integral solution of $x^2+y^2+z^2+k(xy+xz+yz) = 3(k+1)xyz$ is uniquely determined, up to permutation, by its largest coordinate.

Contents

1. Introduction 1
2. Descent, Farey labels, and polynomial dynamics 4
3. The convex norm and exponential Farey estimates 6
4. Rational linear isometries 12
5. Zero density on a fixed rational graph 12
6. Uniform summation over rational graphs 16
7. Completion of the proof 21
References 21

1. Introduction

The classical Markov equation

$$x^2 + y^2 + z^2 = 3xyz$$

arose in Markoff's study of indefinite binary quadratic forms [15, 16]. His work connects its positive integral solutions with the discrete part of the Markov spectrum and with extremal phenomena in Diophantine approximation. The solutions are generated from $(1, 1, 1)$ by permutations and the Vieta involution $(x, y, z) \mapsto (x, y, 3xy - z)$. This recursive description gives a tree of triples. Frobenius [11] conjectured in 1913 that each triple, after its coordinates have been put in increasing order, is determined by its largest coordinate.

Two lines of work on this problem concern arithmetic restrictions on the largest coordinate and the geometry of the solution tree. Baragar [3], Schmutz [20], and Button [5] obtained uniqueness results in special arithmetic cases. Lang and Tan [14] gave a short proof of uniqueness when the largest coordinate is a prime power, combining congruences with the geometry of the modular

Date: November 8, 2026.
2020 *Mathematics Subject Classification.* 11D25, 11H06, 37P20.
Key words and phrases. Generalized Markov equation, uniqueness, Farey tree, convex norm, lattice orbit, p-adic interpolation.

torus. A useful arithmetic invariant of a classical triple (a, b, m) is a root of $u^2+1 \equiv 0 \pmod{m}$, selected by the relation $au \equiv \pm b \pmod{m}$. The uniqueness of an appropriate representative of this root underlies several of these results.

The geometric interpretation goes back to Cohn [8, 9]. A classical Markov number m occurs as one third of the trace of a simple closed geodesic on the modular torus; the corresponding length is $2\operatorname{arcosh}(3m/2)$. The uniqueness problem is thereby related to the multiplicity of simple lengths, with the symmetries of the torus taken into account. McShane and Rivin [17] extended the simple length function to a norm on the first homology of a punctured torus and counted simple geodesics by primitive lattice points in its unit ball. In the arithmetic formulation, Zagier [22] had obtained an asymptotic of order $(\log H)^2$ for the number of sorted Markov triples with largest coordinate at most H. Risager [19] recently improved the error in this classical triple count to $o(\log H)$. These asymptotics count each triple. To obtain the same leading term for distinct largest coordinates, one must also control repeated values in the tree.

Gyoda and Matsushita [13] introduced the family

$$x^2 + y^2 + z^2 + k_1yz + k_2zx + k_3xy = (3 + k_1 + k_2 + k_3)xyz, \qquad k_1, k_2, k_3 \in \mathbb{Z}_{\geq 0},$$

in the setting of generalized cluster algebras. They proved that all positive integral solutions are generated from $(1, 1, 1)$ by the three Vieta involutions. Their work also shows that unequal parameters can give distinct triples with the same largest coordinate. The equal-parameter family retains the permutation symmetry of the classical equation and is the subject of the present paper:

$$x^2 + y^2 + z^2 + k(xy + xz + yz) = 3(k + 1)xyz, \qquad k \in \mathbb{Z}_{\geq 0}. \tag{1.1}$$

A positive integer appearing in one of its positive integral solutions is called a k-Markov number.

Gyoda and Maruyama [12, Conjecture 1.8] formulated the largest-coordinate uniqueness conjecture for this family. Their generalized Cohn matrices lead to a criterion involving the roots of $u^2 + ku + 1 \equiv 0 \pmod{m}$, and to uniqueness results for prime and certain prime-power largest coordinates. The same uniqueness problem appears in Chen and Huang [6, Conjecture 7.3], in their study of mutation-preserving generalized cluster algebras and Laurent mutation invariants. Combinatorial studies include Banaian's work on orderings of k-Markov numbers [1], the study by Banaian and Huang of monotonicity along lines for the extended labels [2], and parameter comparisons in [21]. These studies compare values at specified labels or parameters. The counting question considered here asks how often any two distinct triples share their largest coordinate for one fixed parameter.

For the classical equation, Chen [7] proved that almost all represented largest coordinates are unique. His proof uses the geometry of the homology norm to reduce equal lengths to rational linear relations, then combines polynomial dynamics, p-adic interpolation, and lattice orbit counting. Doan [10] developed this approach for simple lengths on punctured tori and tori with geodesic boundary, and transferred the result to four-cusped spheres. The latter transfer corresponds to the parameter $k = 2$ in (1.1).

There is also a direct arithmetic explanation of this special parameter. Gyoda and Matsushita [13, Theorem 11] proved that the 2-Markov triples are exactly the coordinatewise squares of classical Markov triples. Indeed, squaring $a^2+b^2+c^2 = 3abc$ gives (1.1) for $(x, y, z) = (a^2, b^2, c^2)$ and $k = 2$, and

$$(3ab - c)^2 = 9a^2b^2 - 2a^2 - 2b^2 - c^2$$

identifies the two mutation rules. Generation from $(1, 1, 1)$ gives the converse. Hence the classical density theorem also gives the result for $k = 2$. The new parameter range in the following theorem is $k = 1$ and $k \geq 3$; the statement and proof treat every fixed nonnegative integer k together.

For a positive integer m and a real number $H \geq 1$, put

$$r_k(m) = \#\{(x, y) \in \mathbb{Z}^2 : 1 \leq x \leq y \leq m,\ (x, y, m) \text{ satisfies (1.1)}\},$$

and define

$$T_k(H) = \sum_{m \leq H} r_k(m), \qquad D_k(H) = \sum_{m \leq H} \mathbf{1}_{r_k(m)>0}. \tag{1.2}$$

Here $\mathbf{1}_E$ denotes the indicator of a condition E. Thus T_k counts sorted triples and D_k counts their distinct largest coordinates. All limits are taken with k fixed.

Theorem 1.1. *For every fixed integer $k \geq 0$, there is a constant $C_k > 0$ such that*

$$T_k(H) \sim D_k(H) \sim C_k(\log H)^2, \qquad T_k(H) - D_k(H) = o_k((\log H)^2). \tag{1.3}$$

Consequently,

$$\#\{m \leq H : r_k(m) \geq 2\} = o_k(D_k(H)), \tag{1.4}$$

$$\sum_{\substack{m \leq H \\ r_k(m) \geq 2}} r_k(m) = o_k(T_k(H)). \tag{1.5}$$

The constant is

$$C_k = \frac{\operatorname{area}(\mathcal{B}_k)}{12\zeta(2)}, \tag{1.6}$$

where $\zeta(2) = \pi^2/6$ and $\mathcal{B}_k$ is the unit ball of the norm constructed in Section 3.

The first density statement concerns the represented largest coordinates; the second counts all solutions in nonunique fibers, with their multiplicities. The square correspondence above gives $T_2(H) = T_0(\sqrt{H})$ and $D_2(H) = D_0(\sqrt{H})$, so $C_2 = C_0/4$.

We construct the required norm from the Vieta recursion and Farey labels. Its positive corners at rational directions determine its rational linear isometry group, and its exponentially small variation between consecutive Farey rays confines equal-length pairs to finitely many rational graph planes at each scale. On a fixed graph, integral polynomial dynamics and the interpolation argument of Poonen [18] yield a zero-density statement. The lattice orbit theorem of Burrin–Nevo–Rühr–Weiss [4], with the cusp-width normalization given in Theorem 5.4, converts the finite congruence counts into real counts. A second Farey partition bounds the sum over graphs of growing complexity. Primitive lattice counting then gives (1.3).

2. Descent, Farey labels, and polynomial dynamics

Throughout the proof, k is fixed. The notation $|v|$ denotes the Euclidean norm. Positive constants c, C may depend on k and may change from line to line. Put

$$\kappa = k(k+2), \qquad \nu = k^2(2k+3), \qquad a_0 = 2k+3. \tag{2.1}$$

The symbol $\mathbb{Z}^2_{\mathrm{prim}}$ denotes the set of nonzero primitive vectors in $\mathbb{Z}^2$, with both signs included. Set

$$e_1 = (1,0), \quad e_2 = (0,1), \qquad \mathcal{C} = \{(p,q) \in \mathbb{R}^2 : p \geq q \geq 0\}.$$

Lemma 2.1 (Descent and the label tree)**.** *Every positive integral solution of (1.1) descends by Vieta mutation to $(1,1,1)$. The coordinates of every solution are pairwise coprime. There are integer labels $M_v = M_{-v} \geq 1$, indexed by $v \in \mathbb{Z}^2_{\mathrm{prim}}$, such that*

$$M_{u+v} = 3(k+1)M_uM_v - k(M_u + M_v) - M_{u-v} \qquad (|\det(u,v)| = 1), \tag{2.2}$$

with $M_{e_1} = M_{e_2} = M_{e_1-e_2} = 1$. Primitive vectors in $\mathcal{C}$ are in bijection with sorted positive solutions, and M_v is the largest coordinate of the corresponding solution.

Proof. For a sorted solution $x \leq y \leq z$, the other root in the last coordinate is

$$z' = 3(k+1)xy - k(x+y) - z = \frac{x^2 + y^2 + kxy}{z} > 0.$$

Let P be the monic quadratic in that coordinate. Then

$$P(y) = x^2 + 2kxy + \{k + 2 - 3(k+1)x\}y^2.$$

If $x = 1$, this equals $-(y-1)((2k+1)y+1)$. If $x \geq 2$, the inequality $x \leq y$ gives $P(y) \leq 3(k+1)(1-x)y^2 < 0$. Thus $0 < z' < y < z$ unless $x = y = 1$. In the latter case $P(Z) = (Z-1)(Z-k-2)$, so the two possibilities are $(1,1,1)$ and $(1,1,k+2)$. The mutation of either of the other coordinates is greater than z: for example, $y' = (x^2 + z^2 + kxz)/y > z$. Hence each noninitial sorted solution has a unique descending parent, and descent terminates because the sum of the coordinates decreases.

Pairwise coprimality is preserved along an edge. Indeed, if a prime p divides both x and z', the mutation formula gives $z \equiv -ky \pmod{p}$. Substitution into the equation modulo p then gives $y^2 \equiv 0 \pmod{p}$, contradicting $\gcd(x,y) = 1$. The other retained coordinate is treated identically. Starting at $(1,1,1)$ proves the claim. In particular, equal coordinates greater than 1 cannot occur.

The dual graph of the Farey triangulation is a tree: adjacent unimodular pairs (u,v) are refined by their mediant $u+v$. Starting with the triangle of unoriented directions $e_1, e_2, e_1 - e_2$, assign its three labels the value 1 and propagate across its edges by (2.2). This gives every label unambiguously. Equivalently, in $\mathcal{C}$ begin with the boundary vectors $e_1, e_1 + e_2$, of labels $1, k+2$, and insert mediants. For an explicit description of the parents of (p,q), where $0 < q < p$ and $\gcd(p,q) = 1$, choose the unique $r \in \{1, \ldots, p-1\}$ with $rq \equiv 1 \pmod{p}$, and put $s = (rq-1)/p$.

Then

$$u = (r, s), \qquad v = (p - r, q - s)$$

belong to $\mathcal{C} \cap \mathbb{Z}^2_{\text{prim}}$, have determinant one, and sum to (p, q). Their slopes lie on opposite sides of q/p, and these properties uniquely determine r, s. Every integer vector strictly between their rays is a positive integer combination of u, v, so its first coordinate is at least p. They are therefore consecutive before q/p is inserted. Passage to a parent decreases the first coordinate and ends at e_1 or $e_1 + e_2$.

At a sorted solution with distinct coordinates, the two upward mutations give distinct children, each with the original triple as its unique descending parent. The triple $(1, 1, k+2)$ has one distinct upward child. These rules identify successive mediant insertions with the sorted mutation tree. The two boundary vectors correspond to $(1, 1, 1)$ and $(1, 1, k + 2)$; interior mediants correspond to all remaining solutions. □

Define

$$R_0(p, q) = (p + q, -q), \qquad R_1(p, q) = (q, p), \qquad K = \langle R_0, R_1 \rangle \subset \mathrm{GL}_2(\mathbb{Z}).$$

The group K has order 12, preserves the six vectors $\{\pm e_1, \pm e_2, \pm(e_1 - e_2)\}$, and has $\mathcal{C}$ as a closed fundamental cone. These assertions follow by multiplying the two displayed matrices: R_0R_1 has order 6 and its cube is $-I$. Both generators preserve the initial unoriented Farey triangle. Uniqueness of the recursion therefore gives

$$M_{gv} = M_v \quad (g \in K). \tag{2.3}$$

Every K-orbit in $\mathbb{Z}^2_{\text{prim}}$ has exactly one representative in $\mathcal{C}$.

Lemma 2.2 (An ambient polynomial action)**.** *Let $\Gamma = \mathrm{SL}_2(\mathbb{Z})$ and $o = (a_0, a_0, a_0)$. There is a right action of Γ on $\mathbb{A}^3$, by integral polynomial automorphisms, and polynomials $P_v \in \mathbb{Z}[a, b, c]$ for $v \in \mathbb{Z}^2_{\text{prim}}$, such that*

$$P_v(P * h) = P_{hv}(P), \qquad P_v(o) = X_v := 3(k + 1)M_v - k, \qquad P_{-v} = P_v. \tag{2.4}$$

The action preserves

$$F_k(a, b, c) = a^2 + b^2 + c^2 - abc + \kappa(a + b + c) + \nu, \tag{2.5}$$

and $F_k(o) = 0$.

Proof. Writing $a = 3(k + 1)x - k$, $b = 3(k + 1)y - k$, and $c = 3(k + 1)z - k$ gives the identity

$$F_k(a, b, c) = 9(k + 1)^2\big(x^2 + y^2 + z^2 + k(xy + xz + yz) - 3(k + 1)xyz\big).$$

Define

$$S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \qquad R = \begin{pmatrix} 0 & -1 \\ 1 & 1 \end{pmatrix},$$

and set

$$(a, b, c) * S = (b, a, ab - \kappa - c), \qquad (a, b, c) * R = (b, c, a). \tag{2.6}$$

The two polynomial maps have orders 2 and 3 and preserve F_k. The presentation $\mathrm{PSL}_2(\mathbb{Z}) \cong (\mathbb{Z}/2\mathbb{Z}) * (\mathbb{Z}/3\mathbb{Z})$, with factors generated by the classes of S and R, defines the claimed action; $-I$ acts trivially. The three initial coordinate directions are $e_1, e_2, e_1 - e_2$.

For $T = S^{-1}R = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$, direct composition gives

$$(a, b, c) * T = (a, ab - \kappa - c, b), \qquad (a, b, c) * T^{-1} = (a, c, ac - \kappa - b). \tag{2.7}$$

In particular, every power of T fixes the first coordinate identically. Choose $g_v \in \Gamma$ with $g_v e_1 = v$, and put $P_v(P) = \pi_1(P * g_v)$, where $\pi_1(a, b, c) = a$. Any other choice is $g_v T^n$, so this is well-defined. Replacing g_v by hg_v proves equivariance; triviality of $-I$ proves $P_{-v} = P_v$. Finally the polynomial identity across a Farey edge is $P_{u+v} + P_{u-v} = P_u P_v - \kappa$. At o this is precisely the affine transform of (2.2), with the same initial labels. □

3. The convex norm and exponential Farey estimates

For $v \in \mathbb{Z}^2_{\mathrm{prim}}$, let

$$\ell_v = \operatorname{arcosh}(X_v/2) > 0. \tag{3.1}$$

We construct a norm $N = N_k$ with $N(v) = \ell_v$ for every primitive integer vector. Its unit ball will be denoted by $\mathcal{B} = \mathcal{B}_k$.

3.1. **Fan recurrences and convexity.** Fix a primitive vector u, and choose $B \in \mathbb{Z}^2$ with $\det(u, B) = 1$. Set $a = X_u$ and $z_n = X_{B+nu}$. The Farey recursion gives

$$z_{n+1} = a z_n - z_{n-1} - \kappa, \qquad z_n = d_a + A_+ \lambda_a^n + A_- \lambda_a^{-n}, \tag{3.2}$$

where

$$\lambda_a = \frac{a + \sqrt{a^2 - 4}}{2} = e^{\ell_u}, \qquad d_a = \frac{\kappa}{a - 2}.$$

Substitution into (2.5), or elimination of z_0, z_1, gives

$$A_+ A_- = R_k(a) := \frac{(a + k)^2 (a + k^2 - 2)}{(a - 2)(a^2 - 4)} = \frac{a^2 + \kappa a + \nu + \kappa^2/(a - 2)}{a^2 - 4} > 1. \tag{3.3}$$

Indeed, using $\lambda_a + \lambda_a^{-1} = a$ in (3.2) gives

$$\begin{aligned} z_0^2 + z_1^2 - a z_0 z_1 + \kappa(z_0 + z_1) &= (2 - a) d_a^2 + 2\kappa d_a - (a^2 - 4) A_+ A_- \\ &= \frac{\kappa^2}{a - 2} - (a^2 - 4) A_+ A_-. \end{aligned}$$

The cubic identity therefore gives $(a^2 - 4) A_+ A_- = a^2 + \kappa a + \nu + \kappa^2/(a - 2)$, while

$$(a - 2)(a^2 + \kappa a + \nu) + \kappa^2 = (a + k)^2 (a + k^2 - 2).$$

Since

$$a_0(a_0 - 2) - 4\kappa = 3, \qquad 0 \le d_a < a_0/4, \tag{3.4}$$

and $z_0 \ge a_0 \ge 3$, the sum $A_+ + A_- = z_0 - d_a$ is positive. Thus both coefficients in (3.2) are positive.

At a mediant insertion inside $\mathcal{C}$, let a, b be the traces of the two parents, c the new trace and d the old trace. Inductively $d \le \max(a,b)$ and $c + d = ab - \kappa$. If $a \le b$, then

$$c - b \ge b(a-2) - \kappa \ge a_0(a_0-2) - \kappa = 3(k+1)^2 > 0.$$

Consequently c is the larger root. Since $4\kappa < a_0^2 \le ab$,

$$c \ge \frac{ab-\kappa}{2} \ge \frac{3}{8}ab, \qquad X_{(p,q)} \ge \frac{8}{3}\gamma^p \quad ((p,q) \in \mathcal{C} \cap \mathbb{Z}^2_{\text{prim}}), \qquad \gamma = \frac{3a_0}{8} > 1. \tag{3.5}$$

For the second inequality, multiply traces by $3/8$; the first inequality becomes a multiplicative lower bound, and the first coordinates add under mediants. The boundary traces are at least a_0, which starts the induction.

For a convex positively homogeneous function N and $x \neq 0$, define its set of supporting functionals by

$$\partial N(x) = \{\alpha \in (\mathbb{R}^2)^* : \alpha(y) \le N(y) \text{ for all } y \in \mathbb{R}^2, \quad \alpha(x) = N(x)\}.$$

We identify linear forms with Euclidean vectors. The set $\partial N(x)$ is a segment in the line $\alpha(x) = N(x)$. Denote its endpoints by $\nabla N(x)^-$ and $\nabla N(x)^+$, ordered so that their difference is a nonnegative multiple of $\mathcal{R}x$, where $\mathcal{R}(p,q) = (-q,p)$. They are the extreme supporting gradients on the clockwise and counterclockwise sides of the ray through x.

Proposition 3.1. *There is a K-invariant strictly convex norm N on $\mathbb{R}^2$ taking the values (3.1) on primitive integer vectors. It satisfies $c|v| \le N(v) \le C|v|$. At a primitive vector v, the difference of the extreme gradients, in counterclockwise order, is*

$$\nabla N(v)^+ - \nabla N(v)^- = J_k(X_v)\mathcal{R}v, \qquad J_k(a) = \log R_k(a) > 0. \tag{3.6}$$

Proof. For a reduced rational $q/p \in [0,1]$, define $f(q/p) = \ell_{(p,q)}/p$. Writing $\eta = \frac{1}{2}\log(A_+/A_-)$ gives

$$z_n = d_a + 2\sqrt{R_k(a)}\cosh(n\ell_u + \eta) > 2\cosh(n\ell_u + \eta),$$

and hence

$$\ell_{B+nu} > |n\ell_u + \eta|. \tag{3.7}$$

Suppose V, U, W are consecutive vectors in the Farey sequence of a fixed order. Then $V + W = mU$ for a positive integer m. Expressing $V = -B + nU$ and $W = B + (m-n)U$ and applying (3.7) on the two sides of the fan gives

$$\ell_V + \ell_W > |n\ell_U - \eta| + |(m-n)\ell_U + \eta| \ge m\ell_U.$$

Writing $V = (p_V, q_V)$, $U = (p_U, q_U)$, and $W = (p_W, q_W)$, the difference between the right and left slopes of the polygonal interpolant at q_U/p_U is

$$p_U(\ell_V + \ell_W) - (p_V + p_W)\ell_U = p_U(\ell_V + \ell_W - m\ell_U) > 0.$$

Thus the interpolant at every Farey order is convex. A finer interpolant agrees at all previous vertices and, by convexity, lies below each previous chord. The interpolants decrease and are bounded between zero and the endpoint chord. Their limit is finite, convex, and continuous on

$(0,1)$. Every interpolant of order at least p takes the value $\ell_{(p,q)}/p$ at the reduced fraction q/p, so the limit retains this value. The fan expansions at e_1 and e_1+e_2 give the endpoint limits along $1/n$ and $n/(n+1)$ and finite limits of the corresponding secant slopes. Convexity then gives continuity and finite one-sided derivatives at both endpoints. At $u=(p,q)$ in the interior, with $B=(r,s)$ and $\det(u,B)=1$, it gives

$$\ell_{B+nu}=n\ell_u+\log A_+ +o(1), \qquad \ell_{nu-B}=n\ell_u+\log A_- +o(1).$$

The two one-sided derivatives therefore satisfy

$$f'_+(q/p)-f'_-(q/p)=p\log(A_+A_-)=pJ_k(X_u). \tag{3.8}$$

For instance, the first derivative equals $p\log A_+ - r\ell_u$, as is seen by evaluating $f((nq+s)/(np+r))$; the second equals $-p\log A_- - r\ell_u$.

Set $N(0)=0$ and

$$N(p,q)=pf(q/p) \qquad (p>0,\ 0\le q\le p).$$

Extend by K-invariance. The unique representative in the closed fundamental cone makes the extension well-defined. It is positively homogeneous and even, since $-I\in K$. We check convexity at the boundary of the cone. At either boundary vector, choose the adjacent vector $B=e_2$. The fan traces obey $z_0=z_{-1}$ and hence $z_n=z_{-n-1}$, by the root symmetries. Thus $A_-=A_+/\lambda_a$. The traces at e_1 and $(1,1)$ are a_0 and $a_1=3k^2+8k+6$, respectively. At e_1 the inward derivative is $f'_+(0)=\log A_+$; at e_1+e_2 it is $f'_-(1)=-\log A_-$. Since $\log A_+ - \log A_- = \ell_u$, these identities give

$$2f'_+(0)-f(0)=J_k(a_0)>0, \qquad f(1)-2f'_-(1)=J_k(a_1)>0. \tag{3.9}$$

At e_1, the gradients on the two sides are

$$(f(0),f(0)-f'_+(0)), \qquad (f(0),f'_+(0)),$$

because the boundary reflection is $(p,q)\mapsto(p+q,-q)$. At e_1+e_2, they are

$$(f(1)-f'_-(1),f'_-(1)), \qquad (f'_-(1),f(1)-f'_-(1)).$$

Their counterclockwise differences are respectively $J_k(a_0)\mathcal{R}e_1$ and $J_k(a_1)\mathcal{R}(e_1+e_2)$. The same property holds at the other cone boundaries by K-invariance. On a line of direction w crossing the ray of v, the derivative jumps by $J_k(X_v)|\det(v,w)|\ge 0$. The restriction of N to a line not through the origin is convex between its finitely many cone crossings and at each crossing, so it is convex on that line. On a line through the origin, $N(tv)=|t|N(v)$ is convex. Thus N is globally convex.

The lower bound (3.5) and the endpoint chord give $c|v|\le N(v)\le C|v|$ on rational directions and then everywhere by continuity. Convexity and homogeneity give the triangle inequality, so N is a norm. Formula (3.8), the boundary identities, and unimodularity of K give (3.6). Finally a line segment on the unit sphere would make N linear on an open cone. That cone contains a rational ray, contradicting its strictly positive gradient jump. The norm is strictly convex. □

3.2. **The expanding coefficient.** We next estimate the full change of gradient within an open Farey sector.

Lemma 3.2. *Let a, b be the parent traces of a mediant inside $\mathcal{C}$, and let c be the new trace. In the fan centered at the first parent, with $z_0 = b, z_1 = c$, denote the expanding coefficient by $A_{a,b}$. Then*

$$\frac{1}{8}\lambda_b \le A_{a,b} \le \lambda_b, \qquad 0 \le \log \lambda_b - \log A_{a,b} \le \frac{C}{ab}. \tag{3.10}$$

Proof. Set

$$\widehat{c} = 2\cosh(\log \lambda_a + \log \lambda_b) = \frac{ab + \sqrt{(a^2-4)(b^2-4)}}{2}.$$

If d is the old trace, evaluating the quadratic with roots c, d at $\widehat{c}$ gives

$$(\widehat{c} - c)(\widehat{c} - d) = 4 + \kappa(a + b + \widehat{c}) + \nu. \tag{3.11}$$

Here $\widehat{c}$ is to the right of the quadratic's vertex. Its positive value implies $\widehat{c} > c$, and

$$\widehat{c} - d \ge \tfrac{1}{2}\sqrt{(a^2-4)(b^2-4)} \ge \tfrac{1}{2}(1 - 4/a_0^2)ab.$$

Since $\widehat{c} \le ab$ and $a, b \ge a_0 \ge 3$, it follows that $0 < \widehat{c} - c \le C$. Solving the two fan initial equations yields

$$\begin{aligned} A_{a,b} &= \frac{c - b/\lambda_a - d_a(1 - 1/\lambda_a)}{\sqrt{a^2-4}} \\ &= \lambda_b - \frac{\widehat{c} - c + d_a(1 - 1/\lambda_a)}{\sqrt{a^2-4}}. \end{aligned} \tag{3.12}$$

In particular $0 \le \lambda_b - A_{a,b} \le C/a$.

For a lower bound, use $c \ge 3ab/8$, $\lambda_a \ge a - 1$, and $d_a < a_0/4 \le b/4$. The numerator in the first line of (3.12) is at least

$$b\left(\frac{3a}{8} - \frac{1}{a-1} - \frac{1}{4}\right) > 0 \qquad (a \ge 3).$$

We may therefore replace its denominator by the larger number a:

$$A_{a,b} \ge b\left(\frac{3}{8} - \frac{1}{a(a-1)} - \frac{1}{4a}\right) \ge \frac{b}{8} \ge \frac{\lambda_b}{8}.$$

The middle inequality follows from

$$\frac{3}{8} - \frac{1}{a(a-1)} - \frac{1}{4a} - \frac{1}{8} = \frac{(a-3)(a+1)}{4a(a-1)} \ge 0.$$

Finally $\lambda_b \ge b - 1 \ge 2b/3$, and $-\log t \le 8(1-t)$ on $1/8 \le t \le 1$. These facts prove (3.10). □

3.3. **Finite fans and small integral normals.** Let $\mathcal{F}_H$ be the fan obtained by taking the Farey fractions of order $\lfloor H \rfloor$ in $[0, 1]$ and their K-images. Its rays are generated by primitive vectors; an open sector means the open cone between consecutive rays. Increasing the order refines the fan. Every supporting functional $\alpha \in \partial N(x)$ satisfies $c \le |\alpha| \le C$.

Proposition 3.3 (Farey estimates)**.** *There are positive constants $c, C, \rho, C_{\mathrm{sep}}$, depending only on k, with the following properties for $H \geq 1$.*

(i) *The fan $\mathcal{F}_H$ has $O(H^2)$ rays, generated by vectors of length at most CH.*
(ii) *In each open sector, all supporting functionals differ by at most $Ce^{-\rho H}$. The sum over all open sectors of their total normal angular variations is at most $Ce^{-\rho H}$.*
(iii) *The full normal angle at each boundary ray is at least $ce^{-C_{\mathrm{sep}}H}$.*
(iv) *The total angular width of any m sectors is at most $C\sqrt{m}/H$.*
(v) *If α is a supporting functional at any point of an open sector, then throughout that sector*

$$|N(v) - \alpha(v)| \leq C|v|e^{-\rho H}. \tag{3.13}$$

Proof. For adjacent fractions $q/p < s/r$ in the basic cone, let their traces be a, b. The one-sided derivatives computed above show that the total variation of f' in the open interval is exactly

$$p(\log \lambda_b - \log A_{a,b}) + r(\log \lambda_a - \log A_{b,a}) \leq C\frac{p+r}{ab}. \tag{3.14}$$

To check the identity, the inward derivative at the left endpoint is $p \log A_{a,b} - r \log \lambda_a$, whereas the inward derivative at the right endpoint is $p \log \lambda_b - r \log A_{b,a}$; subtract them. This is the entire variation of the monotone derivative, including any continuous part. Since $p+r > \lfloor H \rfloor$, $p, r \leq H$, and $ab \geq (8/3)^2 \gamma^{p+r}$, summing over the $O(H^2)$ gaps gives $CH^3\gamma^{-H}$, up to an absolute factor. For an open gap $I = (x, y)$, the positive Lebesgue–Stieltjes measure df'_+ has mass $f'_-(y) - f'_+(x)$. The support curve on the slice $(1, t)$ satisfies the vector-valued measure identity

$$d(f - tf'_+, f'_+) = (-t, 1)\, df'_+,$$

because f is Lipschitz and $df = f'_+(t)\, dt$. At each jump, include the full segment between the two extreme supports. The complete gradient variation on I is therefore at most

$$\int_I \sqrt{1+t^2}\, df'_+(t) \leq \sqrt{2}(f'_-(y) - f'_+(x)).$$

The lower bound on support lengths gives the same estimate, up to a fixed factor, for normal angular variation. The finitely many fixed linear maps in K change these bounds only by constant factors.

For a boundary vector v, the absolute determinant of the extreme supports equals $N(v)J_k(X_v)$, by (3.6). From (3.3),

$$R_k(a) - 1 = \frac{\kappa a + \nu + 4 + \kappa^2/(a-2)}{a^2 - 4}.$$

For $k = 0$ this is $4a^{-2} + O(a^{-4})$; for $k > 0$ it is $\kappa a^{-1} + O_k(a^{-2})$. Using $\log(1+s) = s + O(s^2)$ and $2\cosh t = e^t(1 + e^{-2t})$ gives

$$J_k(2\cosh t) \sim \begin{cases} 4e^{-2t}, & k = 0, \\ k(k+2)e^{-t}, & k > 0. \end{cases} \qquad (t \longrightarrow \infty). \tag{3.15}$$

Together with $N(v) \leq CH$, positivity on bounded intervals, and $c \leq |\alpha| \leq C$, they imply (iii).

For (iv), let r be the smaller denominator of a Farey gap. Its width is $1/(pr) = O(1/(rH))$, because $p + r > \lfloor H \rfloor$. There are $O(r)$ gaps with smaller denominator r. Those with $r \le \sqrt{m}$ have total width $O(\sqrt{m}/H)$; any m of the others have the same bound. Angular width is comparable to slope width on $\mathcal{C}$, and the finite K-images preserve the estimate. Assertion (i) is the elementary Farey count, and (v) follows by comparing α with a supporting functional at v. □

Strict convexity means that every nonzero linear form β has a unique contact point on $\partial\mathcal{B}$: the point where it attains its maximum. Its positive ray is called the contact direction of β. The normal directions of supporting functionals run monotonically around the circle; the normals at a corner fill the arc between its two extreme supports. For nonzero vectors x, y, write

$$\angle(x, \{y, -y\}) = \arccos \frac{|x \cdot y|}{|x|\,|y|} \in [0, \pi/2]$$

for the smaller angle from x to the unoriented line through y.

Lemma 3.4 (Excision of small integral normals)**.** *Let $L, M \ge 1$ and $H \ge C \log(2 + M)$. From $\{v \in \mathbb{Z}^2_{\mathrm{prim}} : N(v) \le L\}$ remove the rays of $\mathcal{F}_H$ and every open sector whose closure contains the contact direction of some $0 \ne \beta \in \mathbb{Z}^2$ with $|\beta| \le M$. The number removed is at most*

$$C\left(\frac{\log(2 + M)}{H} L^2 + L \log^2(2 + M) + H^2\right). \tag{3.16}$$

For every retained sector, each of its supporting functionals α satisfies

$$\angle(\alpha, \{\beta, -\beta\}) \ge c e^{-C_{\mathrm{sep}} H} \quad (0 < |\beta| \le M,\ \beta \in \mathbb{Z}^2). \tag{3.17}$$

Proof. Distinct nonparallel integer normals of length at most M have angular separation at least cM^{-2}. Choose a preliminary order $H_0 = C_{\mathrm{pre}} \log(2+M)$, with C_{pre} sufficiently large, so that the normal variation inside any of its open sectors is smaller than this separation. Each such sector then contains at most one distinct contact direction for the specified normals. On a boundary ray several normals may have the same contact, but there are only $O(H_0^2)$ such rays. Thus the number of distinct contact directions is $O(\log^2(2 + M))$.

At order H, each contact direction belongs to the closures of at most two sectors. Proposition 3.3(iv) bounds their total angular width by $C \log(2 + M)/H$. The intersection of each of these sectors with $L\mathcal{B}$ is convex, has perimeter $O(L)$, and has area at most CL^2 times its angular width. The elementary lattice bound $\#(\mathbb{Z}^2 \cap D) \le \mathrm{area}(D) + C\,\mathrm{perimeter}(D) + C$ for a planar convex set proves the first two terms of (3.16). Each boundary ray contains only one primitive integer vector, giving the $O(H^2)$ term.

By deletion of every sector whose *closure* contains a contact, there is a full boundary corner between a retained sector and each specified contact direction, in either direction around the circle. Monotonicity of the normal direction and the complete corner bound in Proposition 3.3(iii) yield (3.17). Normals of both signs are included. □

4. Rational linear isometries

Proposition 4.1. *The rational linear isometry group of N is exactly K:*

$$\{G \in \mathrm{GL}_2(\mathbb{Q}) : N \circ G = N\} = K.$$

Proof. The inclusion of K follows from Proposition 3.1. Conversely, suppose $G \in \mathrm{GL}_2(\mathbb{Q})$ preserves N, and write $G = A/d$ with A integral and d a positive integer. For $v \in \mathbb{Z}^2_{\mathrm{prim}}$, write $Av = gw$ with $w \in \mathbb{Z}^2_{\mathrm{prim}}$ and $g > 0$. Since $\mathrm{adj}(A)Av = (\det A)v$, we have $g \mid |\det A|$. Thus the factor $r = g/d$ in $Gv = rw$ belongs to a fixed finite set.

Let $t = N(w)$, so $N(v) = rt$. Positive homogeneity and $N \circ G = N$ give

$$\partial N(v) = G^{\mathsf{T}} \partial N(Gv) = G^{\mathsf{T}} \partial N(w).$$

The endpoint orders agree when $\det G > 0$ and are interchanged when $\det G < 0$. Since $w = Gv/r$ and $G^{\mathsf{T}}\mathcal{R}G = (\det G)\mathcal{R}$, formula (3.6) yields

$$\begin{aligned} J_k(2\cosh(rt))\mathcal{R}v &= \mathrm{sgn}(\det G) G^{\mathsf{T}} \big(J_k(2\cosh t)\mathcal{R}w\big) \\ &= \frac{|\det G|}{r} J_k(2\cosh t)\mathcal{R}v. \end{aligned}$$

Cancelling the nonzero vector $\mathcal{R}v$ gives

$$J_k(2\cosh(rt)) = \frac{|\det G|}{r} J_k(2\cosh t). \tag{4.1}$$

If a fixed $r \neq 1$ occurred for infinitely many primitive vectors, their associated t would tend to infinity along a subsequence, whereas

$$\frac{r J_k(2\cosh(rt))}{J_k(2\cosh t)} = |\det G|.$$

By (3.15), the left side tends to zero for $r > 1$ and to infinity for $r < 1$, a contradiction. Hence all sufficiently large primitive vectors have $r = 1$, and (4.1) then gives $|\det G| = 1$.

For two sufficiently large consecutive vectors $(n, 1), (n+1, 1)$, both images under G are integral. Their difference shows that Ge_1 is integral, and then Ge_2 is integral as well. Therefore $G \in \mathrm{GL}_2(\mathbb{Z})$. The shortest primitive vectors for N are precisely $\pm e_1, \pm e_2, \pm(e_1 - e_2)$, by Lemma 2.1. The full group of linear maps permuting this hexagon is K: there are six choices for the image of e_1 and two compatible choices for the adjacent image of e_2, and the twelve elements of K realize them. Thus $G \in K$. □

5. Zero density on a fixed rational graph

Proposition 5.1. *For every fixed $G \in \mathrm{GL}_2(\mathbb{Q}) \setminus K$,*

$$\#\{v \in \mathbb{Z}^2_{\mathrm{prim}} : Gv \in \mathbb{Z}^2_{\mathrm{prim}},\ N(v) = N(Gv) \le L\} = o_{k,G}(L^2). \tag{5.1}$$

5.1. **Finite orbits and a paired action.** Let $\Gamma = \mathrm{SL}_2(\mathbb{Z})$ and $H = \Gamma \cap G^{-1}\Gamma G$. For $m \ge 1$, write

$$\Gamma(m) = \{g \in \Gamma : g \equiv I \pmod m\}.$$

For a nonzero integral vector $b = (b_1, b_2)$, put $\mathrm{cont}(b) = \gcd(|b_1|, |b_2|)$. Write $G = A/d$, where d is a positive integer and A is a nonsingular integral matrix, and set $\delta = |\det A|$. Then $\Gamma(\delta) \subset H$: for $h = I + \delta B$, $AhA^{-1} = I + \mathrm{sgn}(\det A)AB\,\mathrm{adj}(A)$ is integral. If v is primitive, $\mathrm{cont}(Av)$ divides both coordinates of $\mathrm{adj}(A)Av = (\det A)v$, hence divides δ. It therefore equals $\gcd((Av)_1, (Av)_2, \delta)$ and is determined by v modulo δ. Thus

$$Gv \in \mathbb{Z}^2_{\mathrm{prim}} \quad \Longleftrightarrow \quad \mathrm{cont}(Av) = d$$

is a residue condition. If $d \nmid \delta$, there are no feasible vectors.

Primitive vectors in any one residue class modulo δ form one $\Gamma(\delta)$-orbit. To see this, complete v, w to matrices $B_v, B_w \in \Gamma$. If $v \equiv w \pmod{\delta}$, then $B_v^{-1}B_w \equiv T^b \pmod{\delta}$ for some b. Choose $n \equiv -b \pmod{\delta}$. The matrix $B_w T^n B_v^{-1} \in \Gamma(\delta)$ sends v to w. The feasible primitive vectors are therefore a finite union of H-orbits.

Fix an odd prime p, for example $p = 3$. Write $\mathbb{Z}_p$ for the ring of p-adic integers and $\mathbb{Q}_p$ for its fraction field. Let v_p be the valuation normalized by $v_p(p) = 1$, and set $|x|_p = p^{-v_p(x)}$ for $x \neq 0$ and $|0|_p = 0$. Lemma 2.2 gives a right action of H on the entire $\mathbb{Z}_p^6$:

$$(P, P') \star h = (P * h, P' * (GhG^{-1})). \tag{5.2}$$

Take J to be the kernel of this action on $(\mathbb{Z}/p^2\mathbb{Z})^6$. This is a finite-index normal subgroup of H; it is finitely generated and contains $-I$. The feasible primitive vectors are still a finite union of J-orbits. Fix one such orbit Ju, put $w = Gu$, and define

$$z_0 = (o, o), \qquad \mathcal{P}(P, P') = P_u(P) - P_w(P'). \tag{5.3}$$

Equivariance gives the exact identity

$$\mathcal{P}(z_0 \star j) = X_{ju} - X_{Gju} \quad (j \in J). \tag{5.4}$$

5.2. **Analytic interpolation and a zero–one law.** The Tate algebra $\mathbb{Z}_p\langle x_1, \ldots, x_d\rangle$ is the algebra of power series with integral coefficients tending to zero p-adically; it carries the coefficient supremum norm.

Lemma 5.2 (Interpolation)**.** *Suppose an integral polynomial automorphism g fixes every point of $(\mathbb{Z}/p^2\mathbb{Z})^d$, where p is odd. On each residue ball $a + p\mathbb{Z}_p^d$, its iterates extend to a jointly analytic action of $\mathbb{Z}_p$.*

Proof. Choose $a \in \mathbb{Z}^d$ representing the residue ball and use coordinates $x = a + py$. The normalized map $g_a(y) = (g(a + py) - a)/p$ has the form

$$g_a(y) = y + pR(y), \qquad R \in \mathbb{Z}_p[y]^d. \tag{5.5}$$

In fact, $g(a) - a \equiv 0 \pmod{p^2}$. Applying the pointwise congruence also to $a + pb$, for every $b \in (\mathbb{Z}/p\mathbb{Z})^d$, and expanding to first order gives $Dg(a)b \equiv b \pmod{p}$ for all b, so $Dg(a) \equiv I \pmod{p}$. The integral Taylor coefficients of degree $j \geq 2$ acquire a factor p^{j-1} after normalization. This proves coefficientwise congruence in (5.5).

Let $\Delta f = f \circ g_a - f$ on the Tate algebra. By (5.5), Δ maps p^m times this algebra into p^{m+1} times it. Therefore

$$\Phi_g(s,y) = \sum_{n\geq 0} \binom{s}{n} \Delta^n y \tag{5.6}$$

converges in the Tate algebra in (s,y): the coefficient valuations of its nth summand are at least $n - v_p(n!) \geq n(1-1/(p-1))$, which tend to infinity for odd p. For every nonnegative integer s the binomial identity gives $\Phi_g(s,y) = g_a^s(y)$. Density of the nonnegative integers in $\mathbb{Z}_p$ and continuity give the group law $\Phi_g(s+t,y) = \Phi_g(s,\Phi_g(t,y))$. This is Poonen's interpolation argument [18, Theorem 1], here with coefficient congruence exponent $1 > 1/(p-1)$. Transport the action back by $x = a + py$, and use Φ_g for the resulting action on $a + p\mathbb{Z}_p^d$. □

We also recall the elementary zero statement used below. A nonzero one-variable Tate series $f(s) = \sum a_n s^n$ has only finitely many zeros in $\mathbb{Z}_p$ (Strassmann's theorem). One proof takes the largest index m at which $|a_m|_p$ is maximal. If $b \in \mathbb{Z}_p$ is a zero, division by $s-b$ produces a Tate series whose last maximal index is $m-1$: its coefficients are $c_n = \sum_{j\geq n+1} a_j b^{j-n-1}$, so $|c_{m-1}|_p = |a_m|_p$ and $|c_n|_p < |a_m|_p$ for $n \geq m$. Induction gives at most m zeros, counting multiplicity. Scaling coefficients proves the same statement over $\mathbb{Q}_p$.

Let Y be the closure of $z_0 \star J$ in $\mathbb{Z}_p^6$. Each finite residue image of this orbit is a transitive finite J-set. Its uniform probability measures are compatible, since the transition maps have equal-size fibers. Their inverse limit is a J-invariant probability measure μ of full support on Y. This action is ergodic: if an integrable function is J-invariant, its conditional expectation on every finite residue partition is constant, by transitivity. The residue partitions have diameters tending to zero, so their locally constant functions are dense in $L^1(Y,\mu)$. The conditional expectations therefore converge in L^1 to the original function, which is constant almost everywhere.

Lemma 5.3 (Zero–one law)**.** *For any integral polynomial Q on the ambient space,*

$$\mu\{y \in Y : Q(y) = 0\} \in \{0,1\}.$$

Proof. Let $E = Y \cap \{Q = 0\}$ and let g be a generator of J. The interpolation of Lemma 5.2 preserves Y: each p-adic time is a uniform limit of integer times on the compact residue ball. It also preserves μ, first for continuous test functions by that uniform limit, then for all Borel sets. For each $y \in Y$, the function $s \mapsto Q(\Phi_g(s,y))$ is analytic on $\mathbb{Z}_p$. Its zero set is either all of $\mathbb{Z}_p$ or finite. In either case, with normalized Haar measure ds,

$$\int_{\mathbb{Z}_p} |\mathbf{1}_E(\Phi_g(s+1,y)) - \mathbf{1}_E(\Phi_g(s,y))|\, ds = 0.$$

Integrating over Y, using Fubini, the group law, and preservation of μ by $\Phi_g(s,\cdot)$, gives

$$\begin{aligned} 0 &= \int_{\mathbb{Z}_p} \int_Y |\mathbf{1}_E(\Phi_g(s+1,y)) - \mathbf{1}_E(\Phi_g(s,y))|\, d\mu(y)\, ds \\ &= \int_{\mathbb{Z}_p} \int_Y |\mathbf{1}_E(\Phi_g(1,x)) - \mathbf{1}_E(x)|\, d\mu(x)\, ds = \mu(E \triangle g^{-1}E). \end{aligned}$$

This holds for all generators. Ergodicity proves the assertion. □

Apply this lemma to $\mathcal{P}$ from (5.3). If its zero set has measure one, closedness and full support imply that it vanishes throughout Y. At every actual integral orbit point, (5.4) is an equality of integers. Injectivity of $\mathbb{Z} \longrightarrow \mathbb{Z}_p$ then gives

$$N(ju) = N(Gju) \qquad (j \in J).$$

The directions of Ju are dense in the projective line. Indeed, for a primitive a the parabolic $v \mapsto v + \det(a, v)a$ belongs to Γ, and some positive power belongs to the finite-index subgroup J. If u is not parallel to a, its iterates approach the direction of a. Omitting the single parallel direction leaves a dense set of rational directions; $-I \in J$ supplies both orientations. Continuity and homogeneity of N would give $N \circ G = N$, contrary to Proposition 4.1. Its zero set therefore has measure zero. The sets $\{y \in Y : \mathcal{P}(y) \equiv 0 \pmod{p^n}\}$ decrease to $\{y \in Y : \mathcal{P}(y) = 0\}$. Continuity of the probability measure from above gives

$$\mu\{y \in Y : \mathcal{P}(y) \equiv 0 \pmod{p^n}\} \longrightarrow 0. \tag{5.7}$$

5.3. **Lattice orbit counting and cusp widths.**

Theorem 5.4 (Finite-index orbit counting)**.** *Let $V \leq \Gamma$ have finite index and contain $-I$, and let $a \in \mathbb{Z}^2_{\mathrm{prim}}$. Choose $g_a \in \Gamma$ with $g_a e_1 = a$ and let $\omega(a; V)$ be the least positive integer ω such that $g_a T^\omega g_a^{-1} \in V$. Let*

$$\Omega = \{r(\cos\theta, \sin\theta) : 0 \leq r \leq r_\Omega(\theta),\ 0 \leq \theta < 2\pi\},$$

where r_Ω is a fixed nonnegative 2π-periodic Lipschitz function. Then

$$\#(Va \cap L\Omega) = \frac{\omega(a; V)}{[\Gamma : V]} \frac{\mathrm{area}(\Omega)}{\zeta(2)} L^2 + o_{V,a,\Omega}(L^2). \tag{5.8}$$

This follows from [4, Theorem 2.7 and Section 5.2], which applies to every finite-index subgroup of Γ. To compute the constant, set $\omega = \omega(a; V)$ and

$$\sigma = g_a \begin{pmatrix} \sqrt{\omega} & 0 \\ 0 & 1/\sqrt{\omega} \end{pmatrix}.$$

Then $\sigma T \sigma^{-1} = g_a T^\omega g_a^{-1}$ and $\sigma e_1 = \sqrt{\omega}\, a$. The cusp-normalized disk constant in [4, Section 3.5] is $2/\operatorname{covol}(V)$. Rescaling $\sqrt{\omega}\, a$ to a multiplies this constant by ω. Since $\operatorname{covol}(V) = [\Gamma : V]\pi/3$, the density per unit Euclidean area is

$$\frac{2\omega}{\pi \operatorname{covol}(V)} = \frac{\omega}{[\Gamma : V]\zeta(2)}.$$

In the orbit-map coordinates of [4, Section 5.2], integration over a fixed stabilizer interval makes the main term proportional to

$$\int_0^{2\pi} \int_0^{Lr_\Omega(\theta)} r\, dr\, d\theta = L^2\, \mathrm{area}(\Omega).$$

The disk constant therefore gives the stated density for every such region. These constants count both signs, consistently with $-I \in V$. The norm ball $\mathcal{B}$ has a Lipschitz radial function: on the Euclidean unit circle it is $1/N$, a positive Lipschitz function.

Proof of Proposition 5.1. Continue with the fixed feasible orbit Ju. For each n, let U_n be the kernel of the action of J on the whole ambient space modulo p^n, and write $Q_n = J/U_n$. The group U_n is normal in J and contains $-I$. The stabilizer of the oriented vector u in Γ is $g_u\langle T\rangle g_u^{-1}$, so its stabilizer in J is generated by $s = g_u T^{\omega} g_u^{-1}$, where $\omega = \omega(u; J)$. Write $\bar{j}$ for the image of $j \in J$ in Q_n, and let r_n be the order of $\bar{s}$. For $j_1, j_2 \in J$,

$$\begin{aligned} U_n j_1 u = U_n j_2 u &\iff j_2^{-1} h j_1 \in \langle s\rangle \text{ for some } h \in U_n \\ &\iff \bar{j}_1 \in \bar{j}_2\langle \bar{s}\rangle. \end{aligned}$$

Thus the U_n-orbits correspond to the orbits of right multiplication by $\langle \bar{s}\rangle$ in Q_n, each of size r_n. For ju, use the completion jg_u. Normality shows that its cusp width for U_n is the least $m > 0$ with $g_u T^m g_u^{-1} \in U_n$. Such an m is a multiple ωb, and membership is equivalent to $s^b \in U_n$, or $r_n \mid b$. Every width is therefore ωr_n. Each U_n-orbit contains both signs, since $-I \in U_n$.

Set

$$E_n = \{\bar{j} \in Q_n : \mathcal{P}(z_0 \star j) \equiv 0 \pmod{p^n}\}.$$

This set is invariant under right multiplication by $\bar{s}$: $su = u$ and $(GsG^{-1})w = w$, so equivariance fixes the two coordinate polynomials in (5.3). Moreover,

$$\frac{|E_n|}{|Q_n|} = \mu\{\mathcal{P} \equiv 0 \pmod{p^n}\}. \tag{5.9}$$

Indeed, the orbit map from Q_n to the finite residue orbit of z_0 has equal-size fibers; its uniform group measure pushes forward to uniform orbit measure.

Every genuine collision belongs to one of the $|E_n|/r_n$ selected U_n-orbits. For a fixed n, apply Theorem 5.4 to each of these finitely many orbits, with $\Omega = \mathcal{B}$. Their total leading coefficient is

$$\frac{|E_n|}{r_n} \frac{\omega r_n}{[\Gamma : J]|Q_n|} \frac{\operatorname{area}(\mathcal{B})}{\zeta(2)}.$$

The factor r_n cancels. In particular, if $C_u(L)$ counts collisions in Ju, then

$$\limsup_{L\to\infty} \frac{C_u(L)}{L^2} \le \frac{\omega}{[\Gamma : J]} \frac{\operatorname{area}(\mathcal{B})}{\zeta(2)} \frac{|E_n|}{|Q_n|}.$$

First take $L \to \infty$ for this fixed n, and then take $n \to \infty$. Equations (5.7) and (5.9) give a zero upper bound. Summing over the finitely many feasible J-orbits proves (5.1). □

6. Uniform summation over rational graphs

We prove that the excess of K-orbits over distinct norm values is $o(L^2)$. All constants in this section depend only on the fixed norm and its Farey estimates, except where additional subscripts are displayed.

6.1. Graph complexity and a population bound.

For $G \in \mathrm{GL}_2(\mathbb{Q})$, put

$$\Lambda_{\mathrm{in}}(G) = \mathbb{Z}^2 \cap G^{-1}\mathbb{Z}^2, \qquad \Lambda_{\mathrm{out}}(G) = G\Lambda_{\mathrm{in}}(G),$$
$$I_{\mathrm{in}}(G) = [\mathbb{Z}^2 : \Lambda_{\mathrm{in}}(G)], \qquad I_{\mathrm{out}}(G) = [\mathbb{Z}^2 : \Lambda_{\mathrm{out}}(G)],$$

and define

$$\mathfrak{J}(G) = \max\{1, I_{\mathrm{in}}(G), I_{\mathrm{out}}(G), h(G)\}, \qquad h(G) = \max_{i,j} |G_{ij}|. \tag{6.1}$$

Lemma 6.1. *Let $G \in \mathrm{GL}_2(\mathbb{Q})$ and $L \geq 1$, and let $\mathscr{P}$ be a set of integral graph points (v, Gv) with $|v|, |Gv| \leq CL$. Then, uniformly in G,*

$$\#\mathscr{P} \leq C(L^2/\mathfrak{J}(G) + L + 1). \tag{6.2}$$

For $Q \geq 1$, there are $O(Q^9)$ rational matrices with $\mathfrak{J}(G) \leq Q$. Each of their graph planes has two spanning integer vectors of length $O(Q^2)$.

Proof. If the inputs span a space of dimension at most one, they lie in $\mathbb{Z}u$ for some $u \in \mathbb{Z}^2_{\mathrm{prim}}$, taking $u = e_1$ if all inputs are zero. An input determines its graph point, so

$$\#\mathscr{P} \leq 2\left\lfloor \frac{CL}{|u|} \right\rfloor + 1 \leq 2CL + 1.$$

This proves (6.2) in this case. Otherwise choose two independent inputs. Rounding coordinates in their basis shows that their generated sublattice has covering radius at most half the sum of their lengths, hence $O(L)$. The larger lattice $\Lambda_{\mathrm{in}}(G)$ has no greater covering radius. The Voronoi cells centered at the selected inputs have disjoint interiors, area $I_{\mathrm{in}}(G)$, and lie in a disk of radius $O(L)$. Thus $\#\mathscr{P} \leq CL^2/I_{\mathrm{in}}(G)$. The same argument on the output lattice gives $CL^2/I_{\mathrm{out}}(G)$. If a row of G has length comparable to $h(G)$, the condition $|Gv| \leq CL$ confines the inputs to a strip of width $O(L/h(G))$. Its intersection with the input disk has area $O(L^2/h(G))$ and perimeter $O(L)$. The planar lattice bound gives $C(L^2/h(G) + L + 1)$. Taking the strongest of these bounds, and the trivial disk bound when necessary, proves (6.2).

Let q be the least positive integer for which qG is integral. It is the exponent of the finite group $\mathbb{Z}^2/\Lambda_{\mathrm{in}}(G)$, since $q\mathbb{Z}^2 \subset \Lambda_{\mathrm{in}}(G)$ is equivalent to integrality of qG. Hence $q \mid I_{\mathrm{in}}(G)$. If $\mathfrak{J}(G) \leq Q$, write $G = A/q$ with $q \leq Q$ and the four entries of A bounded by Q^2. There are $O(Q^9)$ such choices. The integer vectors (qe_i, Ae_i), $i = 1, 2$, span its graph plane and have length $O(Q^2)$. □

6.2. **A first partition and rank reduction.** A collision will mean a pair $v, w \in \mathbb{Z}^2_{\mathrm{prim}}$ satisfying $N(v) = N(w) \leq L$ and $w \notin Kv$. Fix a sufficiently large constant A, and choose a larger constant B such that

$$\rho B > C_{\mathrm{sep}} A + 5. \tag{6.3}$$

Use the coarse fan of order $H_c = \lceil A \log L \rceil$ and the fine fan of order $H_f = \lceil B \log L \rceil$. Apply Lemma 3.4 with $M = C_{\mathrm{rank}} L^3$ at the coarse order, for a fixed sufficiently large C_{rank}, and also remove the fine fan rays. The number of primitive vectors removed is at most

$$CL^2/A + O_A(L \log^2 L + \log^2 L). \tag{6.4}$$

The constants B and C_{rank} can be fixed once A is fixed.

Consider collisions whose two endpoints have survived, in a fixed pair of fine open sectors. Choose supporting functionals λ, λ' at interior points of these sectors, and put $\ell = (\lambda, -\lambda') \in$

$(\mathbb{R}^4)^*$. Every such collision $z = (v, w)$ satisfies

$$|z| \le CL, \qquad |\ell(z)| \le CLe^{-\rho H_f}, \qquad c \le |\ell| \le C. \tag{6.5}$$

Every fine sector containing a retained point lies wholly in a retained coarse sector. Thus the chosen supports themselves satisfy the coarse normal separation (3.17).

We will use the following elementary estimate. If $z_1, \dots, z_d$ are linearly independent integral vectors of length at most R, their d-dimensional volume is at least one: its square is the sum of the squares of the integral $d \times d$ minors. The largest $d-1$ singular values of their column matrix are at most $\sqrt{d}\,R$, so the smallest is at least $(\sqrt{d}\,R)^{-(d-1)}$. Consequently,

$$|\ell(z_i)| \le \varepsilon \ (1 \le i \le d) \quad \Longrightarrow \quad \|\ell|_{\mathrm{span}_{\mathbb{R}}(z_1,\dots,z_d)}\| \le d^{d/2} R^{d-1} \varepsilon. \tag{6.6}$$

Lemma 6.2 (Rank reduction)**.** *For all sufficiently large L, the collision vectors in one such pair of fine sectors span a space of dimension at most two.*

Proof. Four independent integral collision vectors would have determinant of absolute value at least one. Taking the first coordinate along $\ell/|\ell|$, estimate (6.5) instead bounds it by $CL^4 e^{-\rho H_f} < 1$, a contradiction.

Suppose three independent vectors span W. Their exterior product is a nonzero integral vector, so their three-dimensional volume is at least one. Since their lengths are $O(L)$, their smallest singular value is at least cL^{-2}. It follows that

$$\|\ell|_W\| \le CL^3 e^{-\rho H_f}. \tag{6.7}$$

A primitive integral normal to W has the form (α, β), of length at most $C_{\mathrm{rank}} L^3$ after fixing C_{rank} large enough. If $\beta = 0$, the output coordinate plane is contained in W, and (6.7) contradicts $|\lambda'| \ge c$. Otherwise write $\ell = t(\alpha, \beta) + r$, with $r \in W$ and $|r| = \|\ell|_W\|$. Projecting the output component onto $\beta^\perp$ gives $|\operatorname{proj}_{\beta^\perp} \lambda'| \le |r|$. Since $|\lambda'| \ge c$, this yields

$$\angle(\lambda', \{\beta, -\beta\}) \le CL^3 e^{-\rho H_f}.$$

This contradicts the lower bound $ce^{-C_{\mathrm{sep}} H_c}$ from the coarse excision, by (6.3). Three independent vectors are therefore impossible as well. □

If a space of dimension at most two has a projection of rank at most one onto one coordinate plane, there are at most two possible primitive vectors at that endpoint. For each, the other endpoint lies on a convex norm circle of radius $O(L)$, with $O(L+1)$ integer points. Indeed, order these points cyclically along the convex boundary. The arc between consecutive points has length at least their Euclidean distance, which is at least one. Their number is therefore bounded by the perimeter, apart from the harmless cases of zero or one point. There are $O_A(\log^4 L)$ fine sector pairs, so these degenerate blocks contribute $O_A(L \log^4 L) = o_A(L^2)$ collisions.

Every remaining two-dimensional space has both projections of rank two. It is the graph $w = Gv$ of a unique rational invertible matrix G, since it is spanned by integral vectors. Necessarily $G \notin K$. Lemma 6.1 shows that the blocks with $\mathfrak{J}(G) > (\log L)^5$ have total population at most

$$O_A(L^2/\log L + L \log^4 L) = o_A(L^2). \tag{6.8}$$

6.3. **A second partition for complexity bands.** For integers $j \geq 1$, put $Q_j = 2^j$ and

$$\mathcal{G}_j = \{G \in \mathrm{GL}_2(\mathbb{Q}) \setminus K : Q_j < \mathfrak{J}(G) \leq 2Q_j\}.$$

We only need $j \leq j_{\max} = \lceil 5 \log_2 \log L \rceil$. There are at most CQ_j^9 candidates in a band. By Lemma 6.1, if the sum of two distinct graph planes in this band is three-dimensional, it has a primitive integral normal of length at most

$$M_j = C_{\mathrm{band}} Q_j^6. \tag{6.9}$$

Here C_{band} is a fixed positive constant. Indeed, choose three independent vectors from their integer spanning pairs of length $O(Q_j^2)$ and take their exterior product.

Fix a sufficiently large integer j_0. For each $j_0 \leq j \leq j_{\max}$, apply the excision lemma at coarse order $H_j = j^3$ with $M = M_j$, and remove also the rays of a fine fan of order $\lceil B_1 j^3 \rceil$. Here $B_1 > 1$ is fixed with

$$\rho B_1 > C_{\mathrm{sep}} + 1. \tag{6.10}$$

Since $\log(2 + M_j) = O(j)$, all these extra deletions together cost at most

$$CL^2/j_0 + O\big(L(\log\log L)^3 + (\log\log L)^7\big). \tag{6.11}$$

This follows by summing $C(L^2/j^2 + Lj^2 + j^6)$. As before, each fine sector with retained points lies in a retained coarse sector at that band.

For this second partition, include all retained collisions on each candidate graph; this contains the blocks assigned by the first partition. For a fixed band and a pair of its fine sectors, call the retained collision block in the graph of $G \in \mathcal{G}_j$ large if its population is at least

$$L^2/Q_j^{20}. \tag{6.12}$$

In the required range $Q_j \leq C(\log L)^5$,

$$\sup_{j \leq j_{\max}} \frac{Q_j^{20}}{L} \leq C \frac{(\log L)^{100}}{L} \longrightarrow 0.$$

Thus one sufficiently large L makes the threshold exceed every fixed multiple of L, simultaneously for all these bands.

Lemma 6.3 (A large block contains a well-spread pair)**.** *The inputs of a large block contain two vectors with*

$$|\det(v_1, v_2)| \geq L^2/(8Q_j^{20}), \tag{6.13}$$

for all sufficiently large L, uniformly in $j \leq j_{\max}$.

Proof. Among the inputs choose a pair maximizing the absolute determinant D. A collinear primitive set has at most two points, so $D > 0$. Every other input has coordinates in $[-1, 1]^2$ in this basis, by maximality of D. Thus all inputs lie in a parallelogram of area $4D$ and perimeter $O(L)$. Their number is at most $4D + CL + C$. In view of (6.12) and its uniform domination of L, inequality (6.13) follows. □

Lemma 6.4 (One large graph per sector pair)**.** *For all sufficiently large fixed j_0 and then all sufficiently large L, each band $j_0 \le j \le j_{\max}$ and each pair of its fine sectors contain at most one large graph block.*

Proof. Choose supports λ, λ' in the two fine sectors and write $\ell = (\lambda, -\lambda')$. For a large graph G, put $\theta_G = \lambda - \lambda' \circ G$. At the two points of Lemma 6.3, the fine approximation gives $|\theta_G(v_i)| \le CLe^{-\rho B_1 j^3}$. Their dual basis has length at most CQ_j^{20}/L, so

$$|\theta_G| \le CQ_j^{20} e^{-\rho B_1 j^3}. \tag{6.14}$$

If two different large graphs occurred, their sum W would have dimension $d = 3$ or 4. From their integer spanning vectors of length $O(Q_j^2)$ choose a basis for W. On each of these vectors the value of ℓ has absolute value at most $CQ_j^{22} e^{-\rho B_1 j^3}$, by (6.14). The integer d-volume is at least one, so the smallest singular value is at least $cQ_j^{-2(d-1)}$. Consequently

$$\|\ell|_W\| \le CQ_j^{28} e^{-\rho B_1 j^3} =: \varepsilon_j. \tag{6.15}$$

If $d = 4$, this contradicts $|\ell| \ge c$ for sufficiently large j. If $d = 3$, choose its primitive integral normal (α, β) as in (6.9). Here $\beta \ne 0$: otherwise a linear functional depending only on the input would vanish on a graph with full input projection and hence be zero. Projection onto this normal line and $|\lambda'| \ge c$ give

$$\angle(\lambda', \{\beta, -\beta\}) \le C\varepsilon_j.$$

The band excision instead gives the lower bound $ce^{-C_{\mathrm{sep}} j^3}$. These contradict one another since

$$2^{28j} e^{-(\rho B_1 - C_{\mathrm{sep}}) j^3} \longrightarrow 0.$$

One choice of j_0, depending only on the fixed geometry and B_1, suffices. □

There are $O(j^{12})$ fine sector pairs in band j. Summing the small-block bound over its $O(Q_j^9)$ candidate graphs gives $CL^2 j^{12} Q_j^{-11}$. By Lemmas 6.4 and 6.1, the large blocks contribute at most $Cj^{12}(L^2 Q_j^{-1} + L + 1)$. Therefore all bands with $j \ge j_0$ together contribute at most

$$CL^2\tau(j_0) + O(L(\log\log L)^{13}), \qquad \tau(j_0) = \sum_{j \ge j_0} j^{12}(2^{-11j} + 2^{-j}) \longrightarrow 0. \tag{6.16}$$

The set

$$\{G \in \mathrm{GL}_2(\mathbb{Q}) \setminus K : \mathfrak{J}(G) \le 2^{j_0+1}\}$$

is finite. Proposition 5.1 gives a total contribution $o_{k,j_0}(L^2)$ from its members.

6.4. **From collision bounds to excess.** Let

$$\mathscr{A}(L) = \#(\{v \in \mathbb{Z}^2_{\mathrm{prim}} : N(v) \le L\}/K), \qquad \mathscr{D}(L) = \#\{N(v) : v \in \mathbb{Z}^2_{\mathrm{prim}},\ N(v) \le L\}.$$

Define the excess $\mathscr{E}(L) = \mathscr{A}(L) - \mathscr{D}(L)$. Discard every complete K-orbit that meets the deleted set. If $d(L)$ primitive vectors were deleted, this discards at most $d(L)$ orbits. Deleting one orbit reduces the excess by at most one. All the remaining vectors satisfy the required geometric conditions.

Within a retained norm fiber containing r different K-orbits, the excess $r-1$ is bounded by the number of non-K collision pairs. Consequently

$$\mathscr{E}(L) \le d(L) + \#\{\text{retained collisions}\}.$$

The preceding estimates give

$$\limsup_{L\to\infty} \frac{\mathscr{E}(L)}{L^2} \le C/A + C/j_0 + C\tau(j_0). \tag{6.17}$$

Here $L \to \infty$ with A, j_0 fixed, and the constant on the right is independent of A, j_0. Letting $A, j_0 \to \infty$ gives

$$\mathscr{A}(L) - \mathscr{D}(L) = o_k(L^2). \tag{6.18}$$

7. Completion of the proof

The standard elementary primitive lattice count in a bounded convex body gives

$$\#(\mathbb{Z}^2_{\text{prim}} \cap L\mathcal{B}) = \frac{\text{area}(\mathcal{B})}{\zeta(2)} L^2 + O_k(L\log(2+L)). \tag{7.1}$$

Indeed, count all nonzero lattice points in $t\mathcal{B}$ as $\text{area}(\mathcal{B})t^2 + O_k(t+1)$, and apply Möbius inversion to their common coordinate divisor. With μ denoting the Möbius function, the sums $\sum_{d\le CL} \mu(d)/d^2 = 1/\zeta(2) + O(1/L)$ and $\sum_{d\le CL} 1/d = O(\log(2+L))$ give (7.1).

Apart from finitely many primitive vectors, the K-action is free: any nonidentity element has its fixed vectors on finitely many lines, and each rational ray has just one primitive vector. Consequently

$$\mathscr{A}(L) = \frac{\text{area}(\mathcal{B})}{12\zeta(2)} L^2 + o_k(L^2). \tag{7.2}$$

Combining this with (6.18) proves the same asymptotic for $\mathscr{D}(L)$.

By Lemma 2.1, a K-orbit corresponds to exactly one sorted positive solution, with largest coordinate M_v. The map

$$m \longmapsto \operatorname{arcosh}\big((3(k+1)m - k)/2\big)$$

is strictly increasing. Thus, with

$$L_k(H) = \operatorname{arcosh}\big((3(k+1)H - k)/2\big) = \log H + \log(3(k+1)) + o(1), \tag{7.3}$$

we have exactly $T_k(H) = \mathscr{A}(L_k(H))$ and $D_k(H) = \mathscr{D}(L_k(H))$. Equations (6.18)–(7.3) prove (1.3) and (1.6). Finally,

$$\#\{m \le H : r_k(m) \ge 2\} \le T_k(H) - D_k(H), \qquad \sum_{\substack{m\le H \\ r_k(m)\ge 2}} r_k(m) \le 2\big(T_k(H) - D_k(H)\big),$$

because $r \le 2(r-1)$ for integers $r \ge 2$. This proves both density statements in Theorem 1.1. □

References

[1] E. Banaian, *Orderings on k-Markov numbers*, Ramanujan J. **71** (2026), Article 12. doi:10.1007/s11139-026-01468-5.

[2] E. Banaian and M. Huang, *Orderings of generalized k-Markov numbers*, preprint (2026), arXiv:2604.17445v1.
[3] A. Baragar, *On the unicity conjecture for Markoff numbers*, Canad. Math. Bull. **39** (1996), no. 1, 3–9.
[4] C. Burrin, A. Nevo, R. Rühr, and B. Weiss, *Effective counting for discrete lattice orbits in the plane via Eisenstein series*, Enseign. Math. **66** (2020), no. 3–4, 259–304. The numbering cited here is that of arXiv:1905.01493v1.
[5] J. O. Button, *The uniqueness of the prime Markoff numbers*, J. London Math. Soc. (2) **58** (1998), no. 1, 9–17.
[6] Z. Chen and Y. Huang, *Mutation-preserving generalized cluster algebras and Laurent mutation invariants*, preprint (2026), arXiv:2608.10551v3.
[7] W. Y. Chen, *Almost all Markoff numbers are unique*, preprint (2026), Hexagon:2610.00281v2.
[8] H. Cohn, *Approach to Markoff's minimal forms through modular functions*, Ann. of Math. (2) **61** (1955), no. 1, 1–12.
[9] H. Cohn, *Representation of Markoff's binary quadratic forms by geodesics on a perforated torus*, Acta Arith. **18** (1971), 125–136.
[10] N. M. Doan, *Almost all simple lengths are unique up to isometry*, preprint (2026), Hexagon:2610.00298v1.
[11] G. Frobenius, *Über die Markoffschen Zahlen*, Sitzungsber. Königl. Preuss. Akad. Wiss. Berlin (1913), 458–487.
[12] Y. Gyoda and S. Maruyama, *Uniqueness theorem of generalized Markov numbers that are prime powers*, Integers **26** (2026), A91. doi:10.5281/zenodo.21678095.
[13] Y. Gyoda and K. Matsushita, *Generalization of Markov Diophantine equation via generalized cluster algebra*, Electron. J. Combin. **30** (2023), no. 4, P4.10, 20 pp. doi:10.37236/11420.
[14] M. L. Lang and S. P. Tan, *A simple proof of the Markoff conjecture for prime powers*, Geom. Dedicata **129** (2007), 15–22.
[15] A. Markoff, *Sur les formes quadratiques binaires indéfinies*, Math. Ann. **15** (1879), 381–406.
[16] A. Markoff, *Sur les formes quadratiques binaires indéfinies*, Math. Ann. **17** (1880), 379–399.
[17] G. McShane and I. Rivin, *A norm on homology of surfaces and counting simple geodesics*, Int. Math. Res. Not. (1995), no. 2, 61–69.
[18] B. Poonen, *p-adic interpolation of iterates*, Bull. Lond. Math. Soc. **46** (2014), no. 3, 525–527. Author's version.
[19] M. S. Risager, *Counting Markov triples and simple geodesics*, preprint (2026), arXiv:2610.36069v1.
[20] P. Schmutz, *Systoles of arithmetic surfaces and the Markoff spectrum*, Math. Ann. **305** (1996), 191–203.
[21] Q. Tang, *Parameter monotonicity for generalized k-Markov numbers*, preprint (2026), Hexagon:2611.00023v1.
[22] D. Zagier, *On the number of Markoff numbers below a given bound*, Math. Comp. **39** (1982), no. 160, 709–723.

Zhili College, Tsinghua University, Beijing 100084, China
Email address: tangqy24@mails.tsinghua.edu.cn