hexagon:2611.00529v1

THREE PROPERTIES OF MARKOV–HURWITZ EQUATIONS

QIYUE TANG

Abstract. We prove Laurent positivity for pure Markov–Hurwitz mutations, density-one uniqueness of positive integral solutions by their maximum for the symmetric generalized equation, and Vieta connectivity of the pure equation on points with at most one zero coordinate over sufficiently large prime fields.

1. Introduction

Markoff's study of indefinite binary quadratic forms led to the equation

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

The positive integral solutions are generated from $(1, 1, 1)$ by permutations and Vieta involutions, such as $z \mapsto 3xy - z$. In the theory of indefinite forms, the largest entry of a triple determines the corresponding normalized minimum above $1/3$ [9, 12]. Hurwitz extended the descent to equations in arbitrarily many variables [8]. For the equation $\sum_{i=1}^{n} x_i^2 = a \prod_{i=1}^{n} x_i$, the involution in slot i replaces x_i by $a \prod_{j \neq i} x_j - x_i$. The coefficient $a = n$ gives the all-one solution. We study the associated rational recurrence, the multiplicity of the largest entry among integral solutions, and the graph of solutions over prime fields.

1.1. **Laurent coefficients.** With indeterminate initial coordinates, replacing one root by the other can be written as a division. For the pure equation the numerator is the sum of the squares of the other coordinates. A succession of such divisions can introduce intermediate denominators that are sums of monomials. Fomin and Zelevinsky proved that the final denominators are Laurent monomials, for this recurrence and more generally for a quadratic form with cross terms [6, Theorem 1.10]. Guéritaud obtained the Laurent property for a family with arbitrary square-free terms in the exchange polynomials and conjectured positivity of the Laurent coefficients [16, Proposition 11 and Conjecture 12]. His Theorem 1 proves the specialization with three variables and zero extra coefficients. Theorem 1 gives this specialization in every dimension. Its proof tracks matchings between factors of a conserved Laurent polynomial and subwords of the mutation sequence.

1.2. **Multiplicity of the largest entry.** Frobenius asked whether a positive Markoff triple, up to permutation, is determined by its largest entry [7]. Zagier proved an asymptotic of order $(\log B)^2$ for the number of sorted triples with largest entry at most B [12]. This counts triples with their multiplicities over a largest entry. Chen's density-one theorem shows that the repeated largest entries account for a vanishing proportion in the classical equation [5]. Tang extended the density statement to the symmetric three-variable equation with a nonnegative integral coefficient of the three pairwise products [18].

For pure Hurwitz equations, Baragar studied the growth of the number of solutions by height and established the growth exponent [2, 3]. Gamburd, Magee, and Ronan subsequently obtained a counting asymptotic in the nonexceptional orbits of the Markoff–Hurwitz varieties by comparing

Date: November 16, 2026.
2020 *Mathematics Subject Classification.* Primary 11D45; Secondary 11D79.
Key words and phrases. Markov–Hurwitz equation, Laurent positivity, Vieta involution.

the dynamics in logarithmic coordinates with a linear renewal problem [15]. The analogous density-one statement for the pure equation with coefficient n was published as a preprint in [10]. Theorem 2 proves density-one uniqueness in the family with a common nonnegative integral coefficient k on the products omitting one variable. The proof reduces equal maxima to collisions on rational graphs of integer labels, then to common continuations of Vieta moves from two fixed seeds.

1.3. **Prime-field graphs.** Baragar asked whether the graph of nonzero Markoff triples over a prime field is connected [1]. Bourgain, Gamburd, and Sarnak showed that for large primes all but a small number of vertices lie in a single orbit [4]. Chen's divisibility theorem rules out small orbits for the classical equation outside a finite set of primes [13]. Martin subsequently gave a proof of the needed divisibility using a weight whose values cancel under the three Vieta involutions [17].

In higher dimension Vyugin constructed a large component using cubic sections and an action allowing coordinate permutations and paired sign changes [11]. With only Vieta involutions as edges, a solution with two or more zero coordinates can only undergo sign changes. Such solutions obstruct connectivity on all nonzero points. Theorem 3 proves connectivity for fixed $n \geq 4$ and all sufficiently large prime fields on the solutions having at most one zero coordinate. The argument joins the large orbits in the cubic sections, bounds the number of residual points independently of the dimension in the exponent of the prime, and applies a higher-dimensional version of Martin's weighted congruence to rule out any residual orbit.

2. Laurent positivity for pure mutations

Theorem 1. *Let $n \geq 3$ and let $x_1, \dots, x_n$ be algebraically independent. Starting with $X_i = x_i$, apply any finite sequence of replacements*

$$X_i \longmapsto \frac{\sum_{j \neq i} X_j^2}{X_i},$$

each of which fixes the other coordinates. Every resulting coordinate belongs to $\mathbb{Z}_{\geq 0}[x_1^{\pm 1}, \dots, x_n^{\pm 1}]$.

Proof. The replacements are involutions, so we work with words whose consecutive directions are distinct. Set $w_i = x_i^2$, $L_i = X_i/x_i$, and $c_i = (\sum_j w_j)/w_i$. The invariant is the quotient of the sum of squares by the product of the coordinates. In fact, if $X_i' = (\sum_{j \neq i} X_j^2)/X_i$, then

$$\frac{X_i^2 + \sum_{j \neq i} X_j^2}{\prod_j X_j} = \frac{X_i + X_i'}{\prod_{j \neq i} X_j} = \frac{(X_i')^2 + \sum_{j \neq i} X_j^2}{X_i' \prod_{j \neq i} X_j}.$$

Equating this quotient with its value at the initial seed and dividing the resulting identity for $X_i + X_i'$ by x_i gives

$$L_i' = c_i \prod_{j \neq i} L_j - L_i. \tag{1}$$

In particular, the coordinates are integral Laurent polynomials in every seed, since the same recurrence can be started at any seed and the inverse replacements are of the same form. Each replacement and its inverse are rational, so the coordinates in every seed are algebraically independent and generate the initial rational function field.

Records and their words. Attach to each coordinate a record of rows, initially empty. Each row has a color in $\{1, \dots, n\}$ and carries an interval ending at that row. At a replacement in direction i, concatenate the other records in decreasing order of length, keeping their colors and intervals at the copied positions, and append a row of color i. The interval ending at this row starts immediately after the end of the old i-record. The current records form a chain of prefixes, because the new record extends the previous longest record, which belongs to a direction different

from i. The latest record is uniquely longest. If its length is d_h, then the sum of all current lengths is $2d_h - 1$: the new longest length is the sum of the unchanged lengths plus one. Thus at the next replacement in direction i the appended interval has length $2(d_h - d_i)$, where d_i is the old length in direction i.

We assign to each interval a word with the same color multiplicities as the rows of the interval. Call a family coherent if the assigned words of any intersecting intervals differ somewhere in their intersection. A word on a record is allowed if it does not equal any assigned word on its interval. The normalized weight of a word on rows of colors $a_1, \dots, a_m$ is $\prod_{v=1}^{m} w_{u_v}/w_{a_v}$, where u_v is its letter in row v. A word is balanced if its normalized weight is 1. Empty words have weight 1.

A coherent assignment gives L_i as the sum of the weights of the allowed words. Indeed, the sum before imposing the newest interval restriction, called the raw sum, is $c_i \prod_{j \neq i} L_j$, since the appended row can carry any letter. If the new pattern is q, the excluded words are precisely the allowed old i-words followed by q: coherence guarantees avoidance of each interval that meets q, and the restrictions entirely in the prefix are exactly those of the old i-record. Their sum is L_i, by balance. This proves the claim by (1). The same deletion rule applies to a suffix of a record when one retains only intervals contained in the suffix. If the new interval starts before the cut there is no deletion. We will construct the assignments inductively.

First activation. For a Laurent polynomial, its Newton polytope is the convex hull of its exponent vectors with nonzero coefficient. For weights on the seed coordinates, its support function is the maximum scalar product of a weight vector with an exponent vector. A subtraction-free expression for a nonzero Laurent polynomial has tropical evaluation equal to this support function, where multiplication adds weights, division subtracts them, and addition takes their maximum. Indeed, write the expression as P/Q with P, Q positive Laurent polynomials, and use the Minkowski sum identity for the Newton polytope of a product on $P = (P/Q)Q$. The identity follows because the leading parts in every weight have nonzero product.

A changed coordinate cannot contain any initial coordinate as a monomial in its support. Otherwise the support function of the initial coordinate is bounded by that of the changed one for every seed weight. This inequality transports to the chart containing the changed coordinate, by tropical evaluation of the subtraction-free coordinate changes. In that chart the changed coordinate has singleton Newton polytope, so the initial coordinate is an integer multiple of it. In the initial chart the reciprocal multiple is also integral, and positivity on the positive torus makes both multiples 1. To see that the coordinates are distinct, evaluate the sequence at the all-one initial point. The first replacement gives $n - 1 > 1$. Every subsequent replacement changes a coordinate strictly smaller than the unique largest one, say X_h. The new value is at least $X_h^2/X_i > X_h$, so it becomes strictly largest. Each coordinate that has been replaced therefore has value greater than 1.

When i is first activated, the raw sum is $1 + L_i'$, whose constant coefficient is 1 by the preceding support argument applied to the initial monomial x_i in X_i'. Thus there is exactly one balanced allowed word on the raw record. The new interval is the entire record, so assigning that word gives coherence.

A fresh letter and the first crossing of a cut. In a coherent history, add one coordinate of a new color a and leave it unchanged throughout the history. In dimension $n + 1$ the additional record remains empty. Thus the records and patterns in the other directions are unchanged, while allowed words may now use the new letter. The old patterns are coherent on this larger alphabet, and the first-activation argument shows that the concatenation of the current records and a free last letter has a unique balanced allowed word when the last row has color a. The count is independent of the order of the blocks. Write Z_r for the part in record r. The word contains exactly one a.

Under a known coherent replacement in direction j, with new pattern q_j, let H be the old fresh word with its Z_j block removed and the other blocks arranged in the order used in the new j-record. The last free letter stays at the end. Then H is allowed in the new record. In fact, if $H = Uq_j$ with U an allowed old j-word, then Z_jUq_j and UZ_jq_j both occur in the old fresh language arranged with the j-block first. They have the same content and are distinct because exactly one of Z_j, U contains a. This also rules out the case of empty Z_j. Hence HZ_jq_j is the new fresh word, by balance and uniqueness. After the first replacement it begins with a, and thereafter H begins with the preceding longest block. Inductively only the latest block contains a, at its first position. In particular a first-activation pattern begins with its own new color.

We record the behavior of the ordinary blocks Z_r. By the r-ray in a current seed we mean $X_r = t$ and the other coordinates equal to 1, with t tending to infinity. Equivalently, its tropical weights are 1 at r and 0 elsewhere. The weight of a word is the exponent of t in the leading term of its normalized weight after pulling the ray back to the initial seed. On the fresh ray the raw first-activation sum is $1 + (n/t^2)$ in dimension $n + 1$. Its top term 1 is represented by the unique balanced word, so each Z_r is the unique leading word of its block.

In any open cone with a unique dominant coordinate, a reverse replacement at a different coordinate of degree R, when the maximum degree is M, gives degree $2M - R > M$. Thus on a strict dominant cone a reduced reverse sequence whose first direction is different from the dominant coordinate has fixed max selectors and is invertible linear in the weights. A positive word sum whose support function is linear on the image of such an open cone has a single leading exponent throughout it. The reverse evaluations have leading coefficient 1 because each exchange has a unique largest summand. Hence in an allowed-word sum the leading exponent is represented by a single word. Dividing an actual coordinate by its initial variable shifts every weight by the same amount.

Let h be the latest direction and $j \neq h$ the next one. Undoing h on either the fresh ray or the old j-ray puts the weights in the same strict h-dominant cone. For all $r \neq h$ the coordinate is unchanged in this step. The inverse cone argument shows that Z_r is the unique leading word of its record on the old j-ray. On undoing j from the new fresh ray, the old j-weight is 2, the fresh weight is 1, and the other weights are 0. This lies in the same strict j-dominant cone as the old j-ray, so for $r \neq j$ the new fresh block is the old j-ray leading word. Consequently the blocks Z_r with $r \notin \{h, j\}$ are unchanged by the replacement. The new block Z_h is the unique leading word of the old longest record on the old j-ray. From HZ_jq_j it is the prefix of length d_h of Z_jq_j.

Fix a current nonempty record in direction i from its most recent birth until its next replacement. An interval crosses its cut if it starts in the record and ends beyond it. Its head is the part of its pattern before the cut and its tail is the part after the cut. Every later longest record starts with the preceding longest record. Thus all crossing intervals are created by the successive appended rows in these longest records. The first crossing occurs in the direction j replaced immediately after the birth of the fixed record. Its tail, denoted by T, has length $d_i - d_j$, where d_j is the length just before the crossing. The old j-block followed by q_j equals Z_iT, by the transport formula. In particular the head of this first crossing matches Z_i.

Every later crossing tail starts with T. To prove this, label each nonlatest direction by a different original color, initially its own color after the first replacement. For a nonempty nonlatest record the label is the first letter of Z_r. At a replacement, the label of the replaced direction transfers to the previous latest direction by the formula Z_jq_j for its new prefix; when the old j-block is empty, q_j starts with j. Other labels stay in place by block stability. At a later crossing in direction r, the new block of the previous latest record starts with the old label of r, different from that of i. In this leading old r-ray word replace its prefix of length d_i by Z_i. The weight increases strictly, since Z_i is the unique leading word of that prefix. The resulting word must match some old forbidden interval, as it cannot be an allowed word of the previous latest record. The matched

interval crosses the cut, since both parts separately avoid intervals contained in them. Its tail is a prefix of the new tail: after the cut the original leading word agrees up to its end with the corresponding part of q_r. By induction that old tail starts with T. This proves the assertion.

Thus a suffix is compatible with every allowed prefix of the fixed record exactly when it avoids the intervals internal to the suffix and does not start with T. Necessity follows by taking prefix Z_i for the first crossing. Sufficiency follows since every crossing tail starts with T. Write $A = L_i$ for the fixed value, $C = L_j$ just before the first crossing, and z, c for the normalized weights of Z_i after the crossing and the old Z_j, respectively. Balance of q_j gives the weight $\mu = c/z$ of T. In addition,

$$A\mu \;\geq\; C \;\geq\; c \tag{2}$$

where coefficientwise comparison means comparison at each Laurent monomial. Indeed, adjoining the head of q_j to any allowed old j-word gives an allowed word of the fixed record, by coherence. The head has weight μ^{-1}, and the map is injective. The second inequality holds since the old Z_j contains no fresh letter.

A conserved quantity at the cut. For a current record extending beyond the cut, let E_r and F_r be the allowed-word sums on its suffixes after the cut and after the endpoint of the first crossing, respectively, imposing only the intervals contained in the indicated suffix. Then $K_r = E_r - \mu F_r$ counts the suffixes compatible with every prefix. To check this, a word counted by E_r starting with T is exactly T followed by a word counted by F_r. Intervals internal to the first suffix that meet T conflict with its letters by coherence with the first crossing. Conversely, no other intervals constrain the remaining letters.

For shorter records set $K_r = 0$, and set $K_i = 1$. Put $D_r = L_r - AK_r$. After the first crossing these are positive word sums for the complements, including the zero sum. In particular $\rho_r = D_r/L_r$ satisfies $0 \leq \rho_r \leq 1$ on the positive torus and $\rho_i = 0$. For a replacement in direction r with latest direction h, set $a_{rh} = c_r \prod_{\ell \neq r,h} L_\ell$. The suffix deletion rule gives

$$K'_r = a_{rh}K_h - K_r, \qquad D'_r = a_{rh}D_h - D_r.$$

For clarity, for both suffix cuts the raw sum comes from appending the other blocks and the free letter to the suffix of the latest record. The deleted sum comes from the suffix of the old r-record when its endpoint lies at or beyond the cut; if it lies before the cut the new interval is not internal and there is no deletion. After the first crossing, each endpoint other than the fixed one is either shorter than the fixed cut or at least as long as the endpoint of that crossing. Thus the rule applies at both cuts with the prescribed values. At the first crossing itself the recurrence holds if, only at its source, one takes $K_j = \mu$ and $D_j = C - A\mu$: the compatible suffix sum is $a_{ji} - \mu$, since its only crossing tail ends at the appended row. All other source values have the usual definitions.

From (1) the ratio $L_r/(L_r + L'_r)$ is $X_r^2/\sum_s X_s^2$. Dividing the defect recurrence gives

$$(\sum_{s \neq r} X_s^2)(\rho_h - \rho'_r) = X_r^2(\rho_r - \rho_h).$$

It follows that

$$I = \sum_r X_r^2 \rho_r - (\sum_r X_r^2)\rho_h$$

is invariant through the intervening replacements. Indeed, the new expression is $\sum_{s \neq r} X_s^2(\rho_s - \rho'_r)$, and subtracting the old one gives zero by the preceding identity. At the source just before the first crossing, $h = i$ and all ρ_s for $s \neq i$ equal 1 except $\rho_j = 1 - \mu A/C$. If Y denotes the physical coordinates immediately before the birth of the fixed record, then $\sum_{s \neq i} X_s^2 = Y_i x_i A$ at the source by the birth exchange, and $X_j = Y_j = x_j C$. Therefore

$$I = x_i A\big(Y_i - (x_j/x_i)\mu Y_j\big). \tag{3}$$

Decay on the revisit ray. At the target just before revisiting i, take the i-ray and pull it back through the known history. Degrees are degrees at infinity of the rational functions of t. Every reverse step gives a new unique maximum and preserves leading coefficient 1. The degree in direction i is odd and the other degrees are even. After undoing the intervening segment up to the birth of the fixed record, direction j has even degree $D \geq 2$. Direction i still has degree 1, and all other directions have degree at most $D-2$. Undoing the birth gives

$$\deg Y_i = 2D-1, \qquad \deg Y_j = D, \qquad \deg Y_s \leq D-2 \quad (s \neq i,j).$$

At a further reverse replacement in direction r with dominant coordinate h, its exact exchange has the factor $1+\sum_{s\neq r,h}(X_s/X_h)^2$ multiplying X_h^2/X_r. Transporting a Laurent monomial from the state before the reverse step into the state after it replaces the exponent v_r by $-v_r$, increases v_h by $2v_r$, and contributes this factor to the power v_r. The X_s in the factor are the unchanged coordinates, so it has the same value on either side.

Start with Y_i/Y_j. While the reverse directions stay in $\{i,j\}$, they alternate and the exponent vector is always $\pm(e_i-e_j)$, where e_s is the s-th unit vector. The error terms in the factors involve only coordinates of degree at most $D-2$ divided by a coordinate of degree at least $2D-1$, and therefore are $O(t^{-2D-2})$. At the first direction outside $\{i,j\}$ the exponent is zero. That step makes the gap between the two largest degrees at least $D+1$. This gap cannot decrease: if the previous two largest degrees were $M > L$ and the changed degree was $R \leq L$, the new gap is $M-R \geq M-L$. All subsequent errors are again $O(t^{-2D-2})$. Consequently

$$Y_i/Y_j = q(x)\big(1+O(t^{-2D-2})\big)$$

for the initial Laurent monomial $q(x)$ obtained by the exponent transport. This includes the case of an empty reverse prefix.

We identify $q(x)$ as a monomial in the initial variables, using the full exponent vectors. Every known coordinate has a unique leading Laurent monomial on the pulled-back ray: its positive expansion and the leading coefficient 1 preclude a second term of the same degree. In each reverse exchange the product of the leading monomials in the changed coordinate before and after the exchange equals the square of the dominant leading monomial. Thus the exponent transport for $q(x)$ is precisely the transport of the quotient of the two leading monomials in Y_i/Y_j. By uniqueness of the leading word Z_i, $A/z \to 1$; by (2), $1 \leq C/c \leq A/z$. Hence the leading monomials in x_iA and Y_j are x_iz and x_jc. The birth exchange now gives $q(x) = (x_j/x_i)c/z = (x_j/x_i)\mu$. Since $x_iA = t$, (3) yields $I = O(t^{-2})$.

Completion of the induction. At the target the conserved expression reads

$$I = \sum_{r\neq i,h} \rho_r - (t^2+n-2)\rho_h.$$

Hence $\rho_h = O(t^{-2})$. For the raw next i-record let B be its raw word sum and let K be the sum of the weights of its compatible suffixes. Raw concatenation gives $K = a_{ih}K_h = (B/A)(1-\rho_h)$. The physical exchange gives $B/A = 1+(n-1)/t^2$, so $K \to 1$. Since K has nonnegative integer coefficients, and each initial variable has leading coefficient 1, this means K has a unique top monomial of weight zero, with coefficient 1. Moreover $AK \leq B$ coefficientwise by concatenating each allowed prefix with each compatible suffix. Block stability identifies z with the weight of the current fresh i-block. The top monomial of B is z, with coefficient 1, since $B = A + L_i'$ and the actual old and new i-coordinates have degrees 1 and -1 on this ray. The top monomial of AK must therefore equal z, forcing the top monomial of K to be 1. The resulting balanced suffix does not start with T, whereas every old pattern crossing the cut has a tail starting with T. It therefore differs from each crossing pattern on the overlap. It also avoids all forbidden intervals contained in the suffix, by the definition of K. Assigning this suffix to the new interval preserves coherence. Induction completes the proof. □

3. Density-one uniqueness of the maximum

Fix integers $n \geq 3$ and $k \geq 0$. For $B \geq 1$, let $T(B)$ count the sorted positive integral solutions of

$$\sum_{i=1}^{n} x_i^2 + k \sum_{i=1}^{n} \prod_{j \neq i} x_j = n(k+1) \prod_{i=1}^{n} x_i, \qquad x_1 \leq \cdots \leq x_n,$$

with $x_n \leq B$, and let $D(B)$ count the distinct values of their last coordinate.

Theorem 2. *For each fixed $n \geq 3$ and $k \geq 0$,*

$$T(B) - D(B) = o(T(B)) \qquad (B \to \infty).$$

The case $n = 3$ follows from [18, Theorem 1.1]. We prove the result for $n \geq 4$ and every $k \geq 0$. The higher-dimensional case $k = 0$ also appears in [10, Theorem 1].

A maximum attained by m solutions contributes $m - 1$ to $T(B) - D(B)$. Since $m \leq 2(m-1)$ for $m \geq 2$, the limit gives density-one uniqueness among both the solutions and their distinct maxima.

3.1. **Descent and labels.** For $1 \leq i \leq n$, the Vieta involution replaces x_i by

$$n(k+1) \prod_{j \neq i} x_j - k \sum_{\ell \neq i} \prod_{j \notin \{i,\ell\}} x_j - x_i.$$

Here and below the product over an empty set is one. This formula is the sum of the two roots minus the original root, so the new coordinate is integral. The product of the roots is $\sum_{j \neq i} x_j^2 + k \prod_{j \neq i} x_j > 0$, so the new coordinate is positive.

For a sorted solution write its first $n-2$ entries as $a_1, \ldots, a_{n-2}$ and its last two entries as b, c. Let $f(t)$ be the monic quadratic obtained by replacing c by t and subtracting the right-hand side of the equation from the left. Then

$$f(b) = \sum_{i=1}^{n-2} a_i^2 + 2b^2 - nb^2 \prod_{i=1}^{n-2} a_i + kb \left(\prod_{i=1}^{n-2} a_i \right) \left(2 + b \sum_{i=1}^{n-2} a_i^{-1} - nb \right).$$

The first three terms are at most $nb^2(1 - \prod_{i=1}^{n-2} a_i)$, and they vanish only if all $a_i = b = 1$. The last term is at most $2kb(1-b) \prod_{i=1}^{n-2} a_i \leq 0$. Thus, unless all the first $n-1$ entries are one, $f(b) < 0$. The roots then lie on opposite sides of b, and the root c lies above b. If the first $n-1$ entries are one, the roots are 1 and $n+k-1$, since their sum is $n+k$ and one is a root. Consequently mutating the maximum of a nonconstant solution strictly decreases it, and repeating gives the all-one solution.

For a solution with a unique maximum, every mutation in a nonmaximum slot produces a new strict maximum. Indeed, if the old entries in that slot and in the maximum slot are $x_i < x_n$, the product of the two Vieta roots exceeds x_n^2, since $n \geq 3$ and an additional square occurs. The new root therefore exceeds x_n.

Put $d = n - 1$. Starting from $(1, \ldots, 1, n+k-1)$, the ascending move with index $i \in \{1, \ldots, d\}$ mutates slot i and then exchanges slots i and n, so the last slot holds the new maximum. Reversing the descending sequence constructs such a word for each sorted solution other than the all-one solution: at each step the coordinate to be increased is among the first d entries, and the permutation of those entries is recorded by the choice of its slot. If several words represent the same sorted solution, use the lexicographically first of minimum length, with the order $1 < \cdots < d$. The all-one solution contributes at most one to the counts and can be omitted from the estimates below.

For a word $w = i_1 \cdots i_m$, let

$$A_i = I + (\mathbf{1} - e_i) e_i^{\mathsf{T}}, \qquad C_A(w) = A_{i_1} \cdots A_{i_m}, \qquad v(w) = C_A(w)\mathbf{1}.$$

Here I is the d-dimensional identity, e_i is its i-th column, and $\mathbf{1}$ has all entries one. The product for the empty word is I. This is chronological matrix order. Each label is primitive, meaning that its integer entries have greatest common divisor one, because each A_i is integral with determinant one. The unique minimum of a nonempty label identifies its first letter, and subtracting this minimum from the other entries recovers the remaining label. The label of a solution x, denoted by $v(x)$, is the label of its chosen word. Distinct solutions have distinct labels.

3.2. **Reduction to quantitative estimates.** The counting bounds follow from Lemmas 6 and 7 and Proposition 11. The partitions by prefixes and their error bounds are proved in Appendix B, in particular (9).

Write $L = \log B$, and use Euclidean and induced operator norms. The counting estimates give $T(B) \asymp L^{\beta}$, where $\beta > 1$ is fixed, and $\|v(x)\| = O(L)$ for each solution counted. For some fixed $\eta > 0$, the proportion of labels in a strip

$$|\ell(v)| \le t\|\ell\|\|v\|$$

is $O(t^{\eta} + L^{-\eta})$, uniformly in nonzero linear forms ℓ and $0 < t < 1$. In particular, proper subspaces contain $O(T(B)L^{-\eta})$ labels. For $G \in \mathrm{GL}_d(\mathbb{Q})$, define

$$D_G = [\mathbb{Z}^d : \{v \in \mathbb{Z}^d : Gv \in \mathbb{Z}^d\}],$$
$$K(G) = \max(2, D_G, \|G\|, \|G^{-1}\|).$$

The number of pairs with $v(y) = Gv(x)$, both counted by $T(B)$, is

$$O\big(T(B)(K(G)^{-\eta} + L^{-\eta})\big).$$

If G is determined by d independent such pairs, Cramer's rule gives $K(G) = O(L^d)$. By decreasing η the pair bound then becomes

$$O(T(B)K(G)^{-\eta}).$$

For a solution x, put

$$h(x) = \log\big((n(k+1))^{1/(n-2)}(x_n - k/(n(k+1)))\big).$$

The excision estimates give finite partitions by ascending prefixes after deleting exceptional vertices. Each part p has a linear form λ_p. For an integer-vector norm cutoff $M > 0$, the constants $\tau, \delta > 0$ are common to the parts and satisfy

$$|\lambda_p(b)| > \tau \quad (0 < \|b\| \le M,\ b \in \mathbb{Z}^d), \qquad |h(x) - \lambda_p(v(x))| \le \delta\|v(x)\|.$$

For any fixed exponent A one may take $\delta \le \tau/(100M^A)$. With M a sufficiently large fixed multiple of L^{d+1}, the number of parts is $L^{o(1)}$ and the deleted proportion is $O(\epsilon) + o(1)$ for any fixed $\epsilon > 0$. There are also partitions for integer bands j with $M = \exp(\kappa j)$, for any fixed $\kappa > 0$. They have $P_j = \exp(o(j))$ parts and deletion proportion $O((j+1)^{-2} + L^{-\eta_0})$ for some $\eta_0 > 0$, uniformly up to $j \le a \log_2 L$ for fixed small $a > 0$. These estimates use a fixed cutoff on initial cusp runs. The relative count omitted by this cutoff tends to zero as the cutoff tends to infinity.

3.3. **Fixed graphs.** For each fixed $G \in \mathrm{GL}_d(\mathbb{Q})$, the number of off-diagonal equal-maximum pairs on its graph is $o(T(B))$. To see the passage from the boundary theorem to the finite claim, let $B_i = A_i^{\mathsf{T}}$, let J be the all-one matrix, and put $U = J - I$. Direct multiplication gives $A_i = UB_iU^{-1}$. Define $C_B(w)$ by replacing each A_i in $C_A(w)$ by B_i. The graph in the new coordinates is given by $U^{-1}GU$.

The accelerated coding groups consecutive copies of one letter into runs and records the ordering of the first d labels. We first fix a cutoff on the initial run excluded from the accelerated alphabet. Among retained words, condition on each of the finitely many initial runs and initial orderings at the two endpoints. For respective initial words w_0, u_0, the rational graph for the continuing directions is $C_B(u_0)^{-1}U^{-1}GUC_B(w_0)$. The boundary and prefix estimates then apply from the

corresponding starting states. At the end the cutoff tends to infinity, using the initial-run estimate in Appendix A. The prefix estimate bounds the normalized count of finite words having their first q runs in any measurable prefix event by a constant times its Gibbs probability, with the constant independent of q. For fixed q, words shorter than q runs have normalized count tending to zero. The Gibbs transitions have a uniform summable tail in the run exponent. Represent the direction of a finite word w by $C_B(w)\mathbf{1}/\|C_B(w)\mathbf{1}\|$. Directions sharing q fixed runs differ by $O(\rho^{q-1})$ for some $\rho < 1$. Indeed, evaluate their products from right to left, starting from the two tails after the shared runs. After applying the last shared run, their simplex representatives both have its index as a largest coordinate. Their joining segment has that property as well. Each of the remaining shared runs chooses a different label from its predecessor and therefore uses a nonlast slot. Subdivide the segment according to the order of its coordinates, breaking ties by a fixed order that places the active largest coordinate last. On each piece the sequence of ordered branches is fixed, because each switch moves the selected label to the end. The uniform contraction from Appendix A gives the bound on each piece and hence on the entire segment. This proves uniform convergence and continuity of the boundary directions in the run coordinates.

Suppose the normalized collision count stays bounded below along a subsequence. Extend both words in each pair with legal runs of exponent one and give each pair mass $1/T(B)$. The two marginals satisfy the prefix count because a fixed invertible graph has at most one partner at either endpoint. For each position select a finite set containing the exponent-one alphabet so that the probability of its complement is at most $\epsilon 2^{-q}$ at position q. A word outside that set at position q must have originally had at least q runs. The prefix bound therefore shows tightness on the product of these finite sets, at both endpoints. Pass to a nonzero joint weak limit. On each fixed cylinder the finite extensions from shorter words have negligible mass. Hence each marginal is dominated by a constant times the corresponding Gibbs law, first on cylinders and then on Borel sets. The finite set of initial orderings can be handled separately.

In the joint limit the two boundary directions obey the positively oriented graph equality. Indeed, for fixed q the short words at either endpoint have vanishing mass, and on the remaining pairs both directions are uniformly approximated by the finite directions, with error tending to zero as $q \to \infty$. Theorem 13 now gives, almost everywhere, an input prefix p, an output word u, a permutation matrix P, and $c > 0$ satisfying

$$(U^{-1}GU)C_B(p) = cC_B(u)P.$$

Here p, u include their fixed initial words, restored by multiplying the terminal relation for the continuing graph on the left by $C_B(u_0)$. Since U commutes with P, this is $GC_A(p) = cC_A(u)P$. For an input word with prefix p and finite continuation t, the output label is consequently $cC_A(u)C_A(P(t))\mathbf{1}$, where $P(t)$ permutes the letters of t. Both unscaled label vectors are primitive. An integral linear combination of the coordinates of the output before scaling is one, so c is integral. The scaled output is primitive as well, giving $c = 1$. Decoding gives output word $uP(t)$. Permuting the first d coordinates of the output seed by P^{-1} makes the continuation identical at both ends. Equal aligned seeds produce identical sorted solutions. For a fixed distinct pair of aligned seeds, Proposition 14 gives $o(T(B))$ equal-maximum pairs. A fixed raw prefix is a clopen condition in run coordinates, since it is decided by finitely many run lengths and letters. Thus every terminal input cylinder has zero limiting mass. The countably many terminal cylinders cover almost all the limiting mass by boundary termination, a contradiction.

3.4. **Summing the graph estimates.**

Proof of Theorem 2. It remains to sum over varying graphs. To count the excess, order each collision class and pair each solution with its successor in that class. This gives exactly one fewer pair than there are solutions in the class, and there is at most one occurrence of each solution in each projection. Apply the graph estimates to this chosen collection throughout. Use the coarse

partition above. In each ordered pair of parts take the span of the paired labels of the equal-maximum solutions. Both coordinate projections of this span are injective. Otherwise a basis of one projection and an extra independent pair give, by integer Cramer's rule, a nonzero integral combination with first block zero and second block $b \neq 0$. Its coefficients are $O(L^d)$, so $\|b\| = O(L^{d+1})$. Summing the equal-height relations with the same coefficients gives $|\lambda_q(b)| = O(\delta L^{d+1})$ in the second part q, contradicting the excision inequality. The other projection is handled by interchanging the parts.

If the span has dimension below d, discard the input labels in its proper projected subspace. The total discarded count is $o(T(B))$ by the strip bound and the subpower bound for the number of parts. Otherwise the span is the graph of a rational invertible matrix determined by d actual pairs. In particular there are at most $L^{o(1)}$ graphs, and those with $K(G) \geq L^a$ contain altogether $o(T(B))$ pairs for any fixed $a > 0$.

For the remaining graphs use the band partitions. For a fixed starting band, delete the union of all exceptional vertex sets for the bands from that starting band up to the cutoff. This makes the band estimates apply at both endpoints of every retained pair. The number of matrices with $K(G) < 2^{j+1}$ is $O(2^{(2d^2+1)j})$, since $D_G G$ is integral with entries $O(2^{2j})$. Fix $H > 2d^2 + 3$. Call an incidence between a graph in $[2^j, 2^{j+1})$ and an ordered pair of parts large if it contains at least $2^{-Hj}T(B)$ collision pairs. For parts p, q, every such pair satisfies

$$|(\lambda_p - \lambda_q G)(v)| \ll \delta 2^j \|v\|.$$

The strip estimate for a large incidence gives $\|\lambda_p - \lambda_q G\| \ll \delta 2^{(1+H/\eta)j}$, taking $aH < \eta/2$ to absorb the finite-size error. If two distinct graphs G_1, G_2 were large for those parts, some column b of $D_{G_1} D_{G_2}(G_1 - G_2)$ would be nonzero integral, with norm $O(2^{3j})$ and $|\lambda_q(b)| \ll \delta 2^{(3+H/\eta)j}$. Choose $\kappa > 3 \log 2$ and the flatness exponent A sufficiently large. The excision inequalities exclude such a column for all sufficiently large j.

Thus the large incidences in band j contribute $O(P_j^2 2^{-\eta j} T(B))$ pairs by the graph estimate. The other incidences contribute

$$O(P_j^2 2^{-(H-2d^2-1)j} T(B))$$

pairs. Both bounds are summable in j after division by $T(B)$. The band deletions above any fixed starting band also have summable relative cost, up to $O(L^{-\eta_0} \log L)$. Each deleted vertex removes at most two of the chosen pairs.

Below the starting band there are only finitely many rational matrices, so the fixed-graph result applies. For any error tolerance, first choose the initial cusp cutoff, then the coarse deletion budget and the starting band to make the three deletion and tail bounds smaller than that tolerance. Letting B tend to infinity gives $T(B) - D(B) = o(T(B))$. □

4. Connectivity over prime fields

Theorem 3. *Fix an integer $n \geq 4$. For all sufficiently large primes p, the graph on the solutions of*

$$\sum_{i=1}^{n} x_i^2 = n \prod_{i=1}^{n} x_i \quad \text{in } \mathbb{F}_p^n$$

having at most one zero coordinate is connected. The edges are the labelled Vieta moves, which replace x_i by $n \prod_{j \neq i} x_j - x_i$ and leave the other coordinates fixed.

4.1. **A uniform orbit on cubic sections.** For $K \in \mathbb{F}_p$, let M_K be the cubic surface $x^2 + y^2 + z^2 - xyz = K$. We use the group generated by its Vieta involutions and coordinate permutations. A point is called an axis point if at least two of its coordinates vanish. At such a point, a Vieta involution acts by changing the sign of the active coordinate, since the product of the other two

coordinates vanishes. Thus the set of axis points is invariant under the entire group. For a trace $x \in \mathbb{F}_p$ with $x \neq \pm 2$, define its order to be the multiplicative order of either root of $U^2 - xU + 1$. The two roots are inverse to each other, and the order divides $p-1$ or $p+1$. The traces ± 2 are called parabolic and are assigned order p.

Lemma 4. *Fix $0 < \delta < 1/4$. For all sufficiently large primes p, uniformly in $K \neq 4$, one orbit on $M_K(\mathbb{F}_p)$ contains every nonaxis point with a coordinate of order at least p^δ. Its complement has at most $6 + 2(p^\delta + 1)^4$ points.*

Proof. Fix $x \neq \pm 2$ and choose λ with $x = \lambda + \lambda^{-1}$. The linear coordinates

$$y = a + b, \qquad z = \lambda a + \lambda^{-1} b$$

are unique, and the cubic equation becomes $ab = (x^2 - K)/(x^2 - 4)$. The rotation $(y, z) \mapsto (z, xz - y)$ is a Vieta move followed or preceded by a transposition, and sends (a, b) to $(\lambda a, \lambda^{-1} b)$. At a nonaxis point at least one of a, b is nonzero. Therefore the mth iterate fixes the point exactly when $\lambda^m = 1$, and the orbit has length $r = \operatorname{ord}(\lambda)$. The values of y in this orbit are $at + b/t$ for $t \in \langle \lambda \rangle$. If λ is split, a, b belong to $\mathbb{F}_p$ and the ambient group of parameters t is $\mathbb{F}_p^\times$. Otherwise $b = a^p$ and the ambient group is the group of norm-one elements of $\mathbb{F}_{p^2}^\times$. We view the latter as the rational points of the norm-one torus over $\mathbb{F}_p$. In its split coordinates the descent exchanges a and b and inverts t, so $at + b/t$ is a map over $\mathbb{F}_p$.

To compare orders on rotation orbits, apply Corvaja–Zannier's bound [14, Corollary 2] to

$$at + b/t = s + s^{-1}.$$

In the two-dimensional torus this curve is geometrically irreducible and is not a translate of a subtorus. Indeed, when $ab \neq 0$ the discriminant in s, after multiplication by t^2, is $a^2 t^4 + (2ab - 4)t^2 + b^2$. Comparing coefficients of a polynomial square shows that it can be a square only if $ab = 1$, which would imply $K = 4$. If exactly one of a, b vanishes, the equation is linear in t or t^{-1} after writing it as a Laurent polynomial. In all cases the irreducible Laurent equation has at least three monomials. Its bidegree in the two projective coordinates is at most $(2, 2)$, so the normalization has genus at most one and at most eight points where a coordinate has a zero or pole. Both coordinate maps have degree at most two and are separable in odd characteristic. Thus the bound for pairs with $t^r = s^m = 1$ is

$$O\big((rm)^{1/3} + rm/p\big)$$

with an absolute constant. The estimate uses characteristic p and applies to the curve over the algebraic closure, including the nonsplit case.

Suppose the largest coordinate order of a point is $r \in [p^\delta, p^{3/4}]$. Use that coordinate for x. For each $m \leq r$ dividing $p-1$ or $p+1$, one has $rm/p \leq (rm)^{1/3}$. Hence the number of rotation parameters whose output y is nonparabolic and has order at most r is

$$O\big(r^{2/3}(\tau(p-1) + \tau(p+1))\big) = o(r),$$

where τ counts positive divisors and $\tau(N) = N^{o(1)}$. The estimate is uniform because $r \geq p^\delta$. A rotation therefore produces a coordinate of order larger than the previous maximum. Iteration reaches a parabolic coordinate or an order at least $p^{3/4}$.

Call a trace regular primitive if it has the form $s + s^{-1}$ with s a generator of $\mathbb{F}_p^\times$ and its square differs from K. We show that the large-order points can reach such a trace. First assume $ab \neq 0$. Write e for the index of $\langle \lambda \rangle$ in its ambient group, and let d divide $p-1$. Consider the curve

$$au^e + bu^{-e} = v^d + v^{-d},$$

with u in the same torus as t and v in the split torus. Over the algebraic closure the two maps to the common value are tame Galois covers of degrees $2e$ and $2d$. The first has deck transformations

generated by $u \mapsto \zeta u$ for $\zeta^e = 1$ and $u \mapsto \beta/u$ for $\beta^e = b/a$. Its two finite branch values are $\pm 2\sqrt{ab}$, with inertia order two, and its inertia order at infinity is e. The second has finite branch values ± 2 and inertia order d at infinity. The finite branch sets are disjoint since $ab \neq 1$. The intersection of the two Galois function fields can ramify only at infinity and is tame. A degree $q > 1$ extension with that property would have $2g-2 \leq -2q+(q-1) < -2$ by Riemann–Hurwitz. The fields are therefore linearly disjoint.

The composite cover has degree $4ed$ and four finite branch values with inertia order two. The tame inertia order at infinity is $\operatorname{lcm}(e, d)$, since the local extensions over an algebraically closed residue field are tame cyclic extensions. Riemann–Hurwitz gives genus $2ed - 2\gcd(e, d) + 1$. The affine fiber product is smooth because above any finite value at least one of the two covering maps is unramified. Locally the fiber product is then étale over the source of the other map. Its projective normalization adds at most $4ed$ points, so Hasse–Weil gives $p + O(ed\sqrt{p})$ rational points on the affine curve. The maps $u \mapsto u^e$ and $v \mapsto v^d$ on rational points have fibers of sizes e and d and images $\langle \lambda \rangle$ and $(\mathbb{F}_p^\times)^d$. Thus the number of pairs (t, s) on the rotation curve with $s \in (\mathbb{F}_p^\times)^d$ is $p/(ed) + O(\sqrt{p})$.

For a cyclic group of order $p-1$, the generator indicator at s is

$$\sum_{\substack{d \mid p-1 \\ s \in (\mathbb{F}_p^\times)^d}} \mu(d),$$

where μ is the Möbius function. Consequently the number of pairs with s primitive is

$$\frac{p}{e}\frac{\varphi(p-1)}{p-1} + O(\tau(p-1)\sqrt{p}),$$

where φ is Euler's totient. As $p/e \asymp r$ and $\varphi(N)/N \geq 1/\tau(N)$, this count tends to infinity whenever $r \geq p^{5/8}$. Each forbidden value $y^2 = K$ accounts for at most four pairs, since each of the two maps $t \mapsto at + b/t$ and $s \mapsto s + s^{-1}$ has degree at most two.

When $ab = 0$ and $r \geq p^{3/4}$, the rotation gives r distinct output traces, since exactly one of a, b is nonzero. The number of nonparabolic traces of order at most $p^{5/8}$ is at most $p^{5/8}(\tau(p-1) + \tau(p+1))$. One can therefore reach a nonparabolic trace of order greater than $p^{5/8}$ whose square differs from K, and then apply the preceding count on its fiber. If $x = \pm 2$, the fiber satisfies $(y \mp z)^2 = K - 4$. Each rational point belongs to a line $y \mp z = c$ with $c \neq 0$ and the square of the rotation preserves that line. Its first coordinate is $xz - y$, equal to $y - 2c$ for either $x = 2$ or $x = -2$. Its orbit gives every value of y, including regular primitive traces for large p.

It remains to connect the regular primitive fibers. For each such x, its rotation is transitive on the fiber, since $ab = (x^2 - K)/(x^2 - 4)$ is nonzero and the split rotation group is all of $\mathbb{F}_p^\times$. Take regular primitive x, x' with $x^2 \neq (x')^2$, and let κ, κ' be their respective values of $(x^2 - K)/(x^2 - 4)$. They are distinct since $K \neq 4$, and neither is 0 or 1. The fibers admit a common coordinate y if the equations $v^2 = y^2 - 4\kappa$ and $w^2 = y^2 - 4\kappa'$ both have rational solutions. Indeed, on the first split fiber $y = a + b$ and $ab = \kappa$, so a, b are roots of $T^2 - yT + \kappa$. They exist over $\mathbb{F}_p$ precisely when $y^2 - 4\kappa$ is a square, with zero allowed. The same argument applies to κ'. Impose also $y = s^d + s^{-d}$ for $d \mid p - 1$. These three covers of the y-line are Galois and have pairwise disjoint finite branch sets. The same argument at infinity, applied first to the two quadratic covers and then to their composite and the third cover, gives geometric irreducibility. The degree is $8d$, with six finite inertia orders two and infinity inertia order d. The genus is $8d - 3$. Hasse–Weil gives $p + O(d\sqrt{p})$ rational points on the affine fiber product. Dividing by d and applying the generator indicator gives

$$p\frac{\varphi(p-1)}{p-1} + O(\tau(p-1)\sqrt{p})$$

solutions with $y = s + s^{-1}$ for a generator s. The count tends to infinity, so we can exclude $y^2 = K$, whose values yield at most eight solutions each after the power map. The two fibers connect through this regular primitive y-fiber, by permuting coordinates as needed. If $x^2 = (x')^2$, insert a third regular primitive trace with a different square. There are $\varphi(p-1)/2$ primitive split traces for large p, so the bounded exclusions permit this choice.

The estimates are uniform in $K \neq 4$. Axis points form an invariant set of size at most six. Outside the orbit just found, every nonaxis point has three coordinates of order less than p^δ. There are at most $(p^\delta+1)^2$ such traces, by summing the number of roots of $U^m = 1$ over positive $m < p^\delta$. After two coordinates are fixed, the third satisfies a monic quadratic. This gives the stated bound. □

4.2. **Gluing cubic sections.**

Proof of Theorem 3. Write X° for the vertex set in the theorem, H for its group of Vieta involutions R_i, and G for the group obtained by adjoining coordinate permutations. Each R_i preserves X°: at a point with a zero in another coordinate it acts on x_i by negation, and elsewhere at most the updated coordinate can become zero.

First, extend Martin's weight argument [17] to all n coordinates. For $x \in X^\circ$ and $1 \le i \le n$ define

$$f_i(x) = \begin{cases} x_i/(n\prod_{j\neq i} x_j), & \prod_{j\neq i} x_j \neq 0, \\ 1/2, & \prod_{j\neq i} x_j = 0. \end{cases}$$

Here and throughout take p odd and $p \nmid n$. Then $f_i(x) + f_i(R_i x) = 1$. In the nonzero-denominator case this follows from the sum of the two Vieta roots, and in the zero-denominator case both terms are $1/2$. For a nonempty H-invariant subset C of X°, summation over C gives $\sum_{x\in C} f_i(x) = |C|/2$ in $\mathbb{F}_p$. At a point with no zero coordinate $\sum_i f_i(x) = 1$ by the defining equation, and at a point with one zero coordinate it is $(n-1)/2$. If T and Z count the two kinds of points in C, respectively, it follows that $p \mid (n-2)T + Z$. This is a positive integer at most $(n-2)|C|$, and therefore

$$|C| \ge p/(n-2).$$

Fix $\delta = 1/(8n)$ and set $B = 6 + 2(p^\delta+1)^4$, the bound in Lemma 4. For a fixed triple of active coordinates, freeze all others at nonzero values. Multiplying the three active coordinates by $c = n\prod x_j$, where the product is over frozen coordinates, identifies the slice with M_K for $K = -c^2\sum x_j^2$, with the same frozen indices in the sum. The map conjugates each active Vieta involution and commutes with active coordinate permutations. If $K \neq 4$, the orbit from Lemma 4 lies in X° because the cubic axis points form an invariant set and the orbit contains a nonaxis point.

To compare slices in a block of four coordinates, freeze the rest at nonzero values. The resulting equation has the form

$$z_1^2 + z_2^2 + z_3^2 + z_4^2 - Az_1z_2z_3z_4 = b,$$

where A is n times the product of the fixed coordinates and b is minus the sum of their squares. The empty product and sum are 1 and 0. On fixing two distinct coordinates at values $s, t \neq 0$, the other two, u, v, satisfy

$$(2u - Astv)^2 - ((Ast)^2 - 4)v^2 = 4(b - s^2 - t^2).$$

If both $b - s^2 - t^2$ and $(Ast)^2 - 4$ are nonzero, the conic has $p-1$ or $p+1$ points according as $(Ast)^2 - 4$ is a square or a nonsquare. This follows from the product equation in the split case and from the norm map of $\mathbb{F}_{p^2}^\times$ in the nonsplit case. Removing at most four points with $uv = 0$ and at most $2B$ points outside the two cubic orbits still leaves a point for large p. Thus the two orbits lie in the same G-orbit whenever both cubic slices are non-Cayley, meaning their levels differ from 4, and the conic meets these conditions.

For any two non-Cayley cubic slices in this four-coordinate block, choose a coordinate fixed by neither of them and freeze it at a nonzero value w. The new slice is non-Cayley unless $A^2w^2(b-w^2)=4$. If one original slice freezes a value s, the conditions on its conic with the new slice exclude the roots of $(Asw)^2-4$ and $b-s^2-w^2$. The second original slice excludes at most four further values in the same way. These exclusions are bounded independently of p, so the third slice connects the first two through the conic intersections.

For a fixed active triple, admissible frozen vectors are the elements of $(\mathbb{F}_p^\times)^{n-3}$ with level $K\neq 4$. With all other entries fixed, the equation

$$n^2\Big(\prod_j x_j\Big)^2\sum_j x_j^2+4=0$$

for a forbidden vector is a quartic in the remaining entry with nonzero leading coefficient. Given two admissible vectors, choose a uniformly random intermediate vector in $(\mathbb{F}_p^\times)^{n-3}$. Replace the entries one at a time to pass from the first vector to the intermediate vector and then to the second. Each state other than the endpoints contains at least one independent uniformly random entry of the intermediate vector, so it is inadmissible with probability at most $4/(p-1)$. There are at most $2(n-3)$ such states. For large p there is a choice with no inadmissible state. Consecutive slices lie in a common four-coordinate block and hence their cubic orbits belong to the same G-orbit.

The all-one frozen vector has level $-n^2(n-3)\neq 4$ for sufficiently large p, since $n^2(n-3)+4=(n-2)^2(n+1)$. Permutations identify these reference slices for all choices of active triple. Therefore there is a single G-orbit $\mathcal{O}$ containing the cubic orbits for all non-Cayley slices.

For an all-nonzero point $x\in X^\circ\setminus\mathcal{O}$, define the product trace $M_{ij}=n\prod_{\ell\neq i,j}x_\ell$ for distinct i,j. For any $k\notin\{i,j\}$ this is the rescaled kth coordinate of the active triple $\{i,j,k\}$. If its order is at least p^δ, then every one of these triples has level 4, by Lemma 4. In terms of the frozen coordinates these conditions say

$$M_{ij}^2\sum_{\ell\neq i,j}x_\ell^2=(M_{ij}^2-4)x_k^2\qquad(k\neq i,j).$$

If $M_{ij}^2\neq 4$, all the squares x_k^2 on the right are equal and nonzero, giving $(n-3)M_{ij}^2=-4$. Thus each M_{ij} belongs to a set of at most $(p^\delta+1)^2+4$ traces, consisting of low orders and roots of the two exceptional quadratic equations.

The n values $M_{12},M_{13},M_{23},M_{24},\dots,M_{2n}$ determine all coordinate ratios to x_1, by $x_2/x_1=M_{13}/M_{23}$ and $x_j/x_1=M_{12}/M_{2j}$ for $3\leq j\leq n$. Substitution into $M_{12}=n\prod_{j=3}^n x_j$ fixes x_1^{n-2}, so at most $n-2$ points realize any given list of these traces. The all-nonzero part of the complement has size at most $(n-2)((p^\delta+1)^2+4)^n$. For each fixed i, the involution R_i injects the part with $x_i=0$ into the all-nonzero part of the complement. The entire complement thus has size $O_n(p^{1/4})$. It is H-invariant, so the weight bound forces it to be empty for sufficiently large p.

Finally, coordinate permutations normalize H because conjugation by σ sends R_i to $R_{\sigma(i)}$. The all-one vector belongs to X° and is fixed by every permutation. Its G-orbit is consequently its H-orbit. Since the former is all of X°, the labelled graph is connected. □

Proposition 5. *For every integer $n\geq 4$ and prime $p\equiv 1 \pmod 4$, the pure Vieta graph on all nonzero solutions of*

$$\sum_{i=1}^n x_i^2=n\prod_{i=1}^n x_i\quad\text{in }\mathbb{F}_p^n$$

is disconnected. Here a nonzero solution is any solution other than the zero vector.

Proof. Since $4\mid p-1$, the cyclic group $\mathbb{F}_p^\times$ contains an element ι of order four. Its square is -1, so $(1,\iota,0,\dots,0)$ is a nonzero solution. If a solution has at least two zero coordinates, then

$\prod_{j\neq i} x_j = 0$ for every i. Each Vieta move therefore only changes a coordinate's sign and preserves the set of zero coordinates. The all-one solution lies outside the orbit of $(1, \iota, 0, \dots, 0)$. □

Appendix A. Counting and arithmetic estimates

Fix integers $n \geq 4$ and $k \geq 0$, and put $d = n - 1$. Use the ascending words and matrices $A_i = I + (\mathbf{1} - e_i)e_i^{\mathsf{T}}$ of the density section. For a finite word $w = i_1 \cdots i_m$, set $C(w) = A_{i_1} \cdots A_{i_m}$ and denote its scalar size by $N(w) = \mathbf{1}^{\mathsf{T}} C(w)\mathbf{1}$. All norms on vectors below are Euclidean unless the sum of coordinates is explicitly used.

Lemma 6. *Let $x(w)$ be the solution obtained by applying the ascending word w to $(1, \dots, 1, n + k - 1)$. Then*

$$\frac{\log(n+k-1)}{n-2}(N(w)-1) \leq \log x_n(w) \leq \frac{\log(n(k+1))}{n-2}(N(w)-1).$$

Each sorted solution other than the all-one solution occurs from at least one and at most $d!$ ascending words.

Proof. In an ascending move let P be the product of the retained coordinates and let x_i be the replaced coordinate. The new entry is

$$\left(n(k+1) - k \sum_{j\neq i} x_j^{-1} \right) P - x_i.$$

The sum is at most $n - 1$, and $x_i \leq x_n \leq P$, so the new entry is between $(n + k - 1)P$ and $n(k+1)P$. Thus at each forward vertex $\log x_n$ lies between the sum of the first d logarithms plus $\log(n + k - 1)$ and the same sum plus $\log(n(k + 1))$. The inequalities hold also at the starting seed.

For the lower bound add $\log(n + k - 1)/(n - 2)$ to each of the first d logarithms. In move i, the shifted entry in slot i becomes the shifted old maximum, which is at least the sum of the old shifted entries. The other slots stay fixed. Hence the new column is at least A_i^{T} times the old column, coordinate by coordinate. All matrix entries are nonnegative, so the word bounds that column below by $C(w)^{\mathsf{T}}\mathbf{1}$ times the initial shift. The shifted maximum is at least the sum of the shifted parent entries. This gives the lower bound. With the upper shift the same inequalities reverse, giving the upper bound. Descent determines the chain of sorted parents. Every forward step creates a new strict maximum, so the only repetitions of entries come from unused initial ones. Choosing among these equal entries multiplies the word count by at most $d(d-1)\cdots 1$. □

A.1. **Accelerated words and height counts.** An accelerated letter consists of a nonlast slot of the current ordered parent column and a positive run length a. At each run, the chosen slot is moved to the end of the ordering. The ordered state is $w = (w_1, \dots, w_d)$, with $0 \leq w_1 \leq \cdots \leq w_d$ and $\sum_i w_i = 1$. Keep a permutation of the fixed parent labels as part of the state to recover the corresponding index i of A_i. For an accelerated letter choosing position $j < d$, its state map first replaces w_j by $w_j + a(1 - w_j)$, moves it to the end, and divides all coordinates by $1 + a(1 - w_j)$. Its roof is $\tau = \log(1 + a(1 - w_j))$, the logarithm of that factor. A word's total roof S_m is the sum of the roofs along its first m successive state updates, with $S_0 = 0$.

Gamburd, Magee, and Ronan [15, Theorem 39, Proposition 43, Proposition 45 and Section 4.3] give $\beta > 1$, a positive C^1 eigenfunction h of the roof transfer operator at eigenvalue one, and constants $D > 0$ and $0 < \rho < 1$ such that every composition of m normalized maps has Lipschitz constant at most $D\rho^m$. The probability of choosing letter a, j from state w is

$$e^{-\beta\tau(w)} \frac{h(f(w))}{h(w)},$$

where f is its normalized map. The eigenfunction equation gives total probability one. Compactness gives positive upper and lower bounds on h. Since $w_j \le 1/2$, every roof is at least $\log(3/2)$, and the conditional expectation of $e^{\varkappa\tau}$ is bounded uniformly for $0 < \varkappa < \beta - 1$, by comparison with $\sum_{a\ge1} a^{-\beta+\varkappa}$.

Lemma 7. *The number of accelerated words of total roof at most t lies between $ce^{\beta t}$ and $Ce^{\beta t}$ for $t \ge 0$, uniformly in the initial state. If a set of length-q prefixes has probability at most ε, the number of words of total roof at most t beginning with such prefixes is at most $C\varepsilon e^{\beta t}$, with C independent of q. Words of run length less than q number at most $C_s e^{st} K_s^q$ for fixed $1 < s < \beta$ and suitable $K_s > 1$.*

Proof. The probability of a word of length m telescopes to $e^{-\beta S_m} h(w_m)/h(w_0)$. Thus the normalized count $e^{-\beta t}$ times the number of words is the expected sum over m with $S_m \le t$ of

$$e^{-\beta(t-S_m)} h(w_0)/h(w_m).$$

For any $b > 0$, the sum of $e^{-b(t-S_m)}$ over these times is bounded by $(1 - e^{-b\log(3/2)})^{-1}$ on every path. This proves the upper bound. For the lower bound cross $t - M$ at the first time J with $S_J \ge t - M$, where M is fixed and $t \ge M$. A crossing from time j with $S_j < t - M$ overshoots t only if the next increment exceeds $M + (t - M - S_j)$. The conditional exponential moment and the preceding geometric bound show that $\mathbb{P}(S_J > t) \le Ce^{-\varkappa M}$. Choose M with this upper bound at most $1/2$. The contribution at J on $S_J \le t$ is at least a fixed positive constant. For $t < M$ the empty word suffices. To count words in a prefix event, sum the uniform upper tail count after each eligible prefix p, weighted by $e^{-\beta S(p)}$. The cylinder probability formula bounds that sum of weights by $C\varepsilon$. Finally the sum of $e^{-s\tau}$ over the single letters is uniformly bounded. Iterating this bound and summing lengths below q gives the stated short-word bound. □

At the barycenter $\mathbf{1}/d$ the total roof of an accelerated word satisfies $e^{S_m} = N(w)/d$: the unnormalized parent dynamics multiply successively by $A_i^{\mathsf T}$, and the product of the normalization factors is its final coordinate sum. An unaccelerated word decomposes into a possible initial run of the initially last label followed by accelerated letters. An initial run of length $A \ge 0$ at the barycenter gives factor $1 + A(d-1)/d$. The sum of these factors to power $-s$ is finite for every $s > 1$. Lemmas 6 and 7 therefore give $T(B) \asymp (\log B)^\beta$, and normalized bad-prefix estimates $O(\varepsilon + e^{-ct})$ for $q \le c't$, with $t = \log\log B$ and positive constants c, c'. Here the short-word bound in Lemma 7 is summed with the initial-run weights at $s < \beta$. The relative count lost by cutting the initial runs at $A = R$ is at most $C\sum_{A>R}(1 + A(d-1)/d)^{-\beta}$. For any fixed initial run, the proof of prefix domination gives the same bound with its resulting initial state, up to a fixed constant after normalization by $T(B)$.

A.2. **Escape from matrix subspaces.** Write $U = J - I$, where J is the all-one $d \times d$ matrix. We first prove that the group Γ generated by the A_i is of finite index in $\mathrm{SL}_d(\mathbb{Z})$. Put $u_i = \mathbf{1} - e_i$. For distinct i, j, l, direct multiplication gives

$$A_l A_j A_l^{-1} = I + (u_l + u_j)(e_j - e_l)^{\mathsf T}.$$

The right-hand matrix fixes u_i, and right multiplication of $e_i^{\mathsf T}$ by its inverse changes that row by $-2(e_j - e_l)^{\mathsf T}$. Its commutator with A_i is therefore $I - 2u_i(e_j - e_l)^{\mathsf T}$. This and A_i^2 give $I + 2u_i f$ for every integral row f annihilating u_i, since the rows $e_i^{\mathsf T}$ and $(e_j - e_l)^{\mathsf T}$ with $j, l \ne i$ generate the annihilator. For $i \ne j$ take $f = \mathbf{1}^{\mathsf T} - (d-1)e_j^{\mathsf T}$. Then $fU = (d-1)e_j^{\mathsf T}$ and $Ue_i = u_i$, so $U^{-1}\Gamma U$ contains $I + 2(d-1)e_i e_j^{\mathsf T}$. It lies in $\mathrm{SL}_d(\mathbb{Z})$ because $U^{-1}A_i U = A_i^{\mathsf T}$. The elementary level subgroup has finite index by the Tits–Vaserstein theorem [19, Proposition 3.18]. Rational conjugation yields commensurable integral groups, so Γ has finite index as well. In particular it is Zariski dense and has property (T) [20].

Lemma 8. *For every proper real linear subspace W of the matrix space, the probability that the product of m accelerated letters is in W is at most Ce^{-cm}, uniformly in the initial state and W. This holds for the chronological matrix product and for its reversal.*

Proof. The derivative of the one-letter log probability is uniformly bounded: the derivative of $\log(1 + a(1 - w_j))$ is bounded independently of a, and the eigenfunction ratios have uniformly bounded derivatives by the one-step contraction bound and the positive minimum of h. Summing the variations under successive contraction gives

$$|\log \mathbb{P}_x(p) - \log \mathbb{P}_y(p)| \le C\|x - y\|$$

for any word p and starting columns x, y with the same ordering.

In the finite permutation graph, every state reaches every other because the move-to-last cycles of the final two through the final d positions generate the symmetric group. There are loops of lengths two and three at every state, obtained by repeating the move at the second-last and third-last position, respectively. Hence there is a fixed core length ℓ such that for each pair of states a connecting core of length ℓ can finish with one move of each label in a prescribed order giving the terminal state. To see this, first connect to any state whose last label differs from the first prescribed label, pad by combinations of lengths two and three, then apply the prescribed sequence. Use exponent one throughout, and form d alternatives by increasing one exponent of the final sequence from one to two. The matrix product of the base core cancels the initial connector in taking right differences with its alternatives. For a final sequence $i_1, \dots, i_d$, those differences are the conjugates of A_{i_j} by $A_{i_{j+1}} \cdots A_{i_d}$, with inverse on the left. Starting at the last position and successively stripping known conjugators shows that the differences generate Γ.

Split the word into cores of length ℓ, each followed by a buffer of a length b to be fixed, and one bounded final remainder. Condition on all buffers, all core endpoint states, and the final remainder. The allowed cores form a Cartesian product. For each core after the first, approximate its input by applying the preceding fixed buffer to a fixed reference column with the correct ordering. The distortion estimate bounds the log probability error of the whole following core and buffer by $O(\rho^b)$. Use the actual initial column for the first core. The final remainder incurs at most a fixed additional distortion factor. Normalizing the conditional weights therefore dominates the conditional law by independent core laws times $C\exp(Cm_0\rho^b)$, where m_0 is the number of cores. Each witness core has probability at least a fixed positive constant in its reference law. Indeed the probability of each fixed witness from any start has a uniform positive lower bound. The weights of the following buffer from the different admissible end columns have bounded ratios by the word distortion estimate. Their common scale cancels in normalization, leaving a witness lower bound independent of b.

Property (T) gives a uniform upper bound $\theta < 1$ on the operator norm of the average of the reference core matrices in any unitary representation with no invariant vectors. In fact the difference between the squared norm of a vector and the squared norm of its average is one half the weighted sum of the squared distances between its translates by two possible cores. Restrict to pairs consisting of the base core and one of its alternatives. The witness lower bound and the Kazhdan constant for their differences give a fixed lower bound times the squared norm of the vector. There are only finitely many possible witness sets. Multiplication by fixed buffers on either side and inversion preserve the operator bound.

For a proper complex algebraic subgroup H of SL_d, its intersection with Γ has infinite index by Zariski density and the irreducibility of SL_d. Hence the representation on square-summable functions on the cosets has no invariant vectors. Taking matrix coefficients of the product of independent averages gives a bound θ^{m_0} for hitting a coset of H, with fixed initial or final factors allowed. For an irreducible proper subvariety V, split the independent product as XY into nearly

equal halves. Cauchy–Schwarz, applied before averaging over X, gives

$$\mathbb{P}(XY \in V)^2 \le \mathbb{P}(VY^{-1} = V(Y')^{-1}) + \sup_{y,y':,Vy^{-1}\neq V(y')^{-1}} \mathbb{P}(X \in Vy^{-1} \cap V(y')^{-1}),$$

where Y' is an independent copy of Y. The first term is a right-stabilizer hitting probability for the product $Y^{-1}Y'$, and the stabilizer is proper. Each intersection in the second term has dimension strictly below $\dim V$ and degree at most $\deg(V)^2$. The second term is omitted when all pairs of translates coincide. Induction on dimension gives a bound $2\deg(V)e^{-c_j m_0}$ for dimension at most j, allowing c_j to decrease at each step. For dimension zero one applies the coset bound to each point. For the induction one can use $c_j \le c_{j-1}/6$, since both half lengths are at least $m_0/3$ for $m_0 \ge 2$, and sum over components using the degree bound. The cases of zero or one factors are covered by decreasing constants. The complex span of SL_d is all matrix space. Subtracting the identity from elementary matrices gives the off-diagonal matrix units. Multiplying $I + e_i e_j^{\mathsf{T}}$ by $I + e_j e_i^{\mathsf{T}}$ and subtracting the identity and the two off-diagonal units gives the diagonal unit in slot i. Thus the complexification of a proper real matrix subspace intersects SL_d in a proper variety of bounded degree, and the independent estimate applies. Choose b so that the exponential comparison loss is less than half the escape rate per core. Now m_0 is proportional to the full run length up to a bounded error. The same proof for the reverse product conditions chronologically and multiplies the independent factors in the opposite order. □

A.3. **Lattices, strips, and rational graphs.**

Lemma 9. *Let M_m be the chronological or reverse product of m accelerated letters. For a primitive integer column v and a subgroup Λ of $\mathbb{Z}^d$ of finite index D_Λ,*

$$\mathbb{P}(M_m v \in \Lambda) \le C\exp[-c\min(m, \log(2 + D_\Lambda))].$$

For the reverse product, any nonzero real column v, and any real linear form $\ell \neq 0$, one has

$$\mathbb{P}\big(|\ell(M_m v)| \le \epsilon\|\ell\|\|M_m v\|\big) \le C\exp[-c\min(m, \log(1/\epsilon))] \quad (0 < \epsilon < 1/2).$$

The constants are uniform in the initial state and the objects in these events.

Proof. The maximum row sum of A_i^a equals $1 + a \le 2e^\tau$. Hence the absolute entries of a product of m letters are bounded by $2^m e^{S_m}$. Iterating the conditional moment bound gives $\mathbb{E}e^{\varkappa S_m} \le M_\varkappa^m$ for a constant $M_\varkappa$. Markov's inequality gives probability at most e^{-m} that some entry exceeds e^{Km}, for large fixed K.

If the integral matrices of entry size at most R sending v into Λ span matrix space, select d^2 independent ones. The matrix of their vectorizations has nonzero integer determinant b with $|b| \le (dR)^{d^2}$ by the row-norm determinant bound. Its adjugate shows that their integer span contains b times every integer matrix. A primitive column has coordinates of gcd one, so integer matrices send it onto all of $\mathbb{Z}^d$. Thus $b\mathbb{Z}^d \subseteq \Lambda$, giving $D_\Lambda \le (dR)^{d^3}$. For $R = e^{Km}$ and larger D_Λ, the truncated hitting matrices lie in a proper subspace and Lemma 8 applies.

For the strip, normalize the matrix linear functional $g \mapsto \ell(gv)$ to coefficient norm one by dividing by $\|\ell\|\|v\|$. On hitting matrices of entry size at most R, its values have magnitude at most $dR\epsilon$. If the matrices span, solve for the coefficients of the functional by inverting the square vectorization matrix of d^2 independent ones. Every cofactor is bounded by $C_d R^{d^2-1}$ and the determinant has absolute value at least one. The coefficient norm forces $1 \le C_d \epsilon R^{d^2}$. For smaller ϵ, the truncated hitting matrices again lie in a proper subspace.

For a word of arbitrary length m, condition before its final block of length $m' \le m$. In the chronological product, the fixed prefix replaces the target lattice by its inverse image under an integral unimodular matrix, preserving index. In the reverse product it replaces the input by another primitive column. For a strip in the reverse product it replaces only the nonzero input. Choose m' comparable to the minimum in the appropriate assertion with a sufficiently small

proportionality constant, so the preceding thresholds apply. Bounded minima are absorbed by increasing C. □

Lemma 10. *The number of accelerated words of total roof at most t satisfying the lattice event in Lemma 9 is at most*

$$Ce^{\beta t}\exp[-c\min(t,\log(2+D_\Lambda))].$$

For the strip event with the reverse product the same conclusion holds with $\log(1/\epsilon)$ *replacing* $\log(2+D_\Lambda)$.

Proof. Let Q denote the corresponding logarithm. Choose an integer N proportional to $\min(t,Q)$ and put $u=L_0N$, with L_0 large and the constant in N small enough that $t\ge 2u$. Bounded N are covered by Lemma 7. The normalized word count is bounded by a constant times the expected sum of $e^{-\beta(t-S_m)}$ at eligible endpoints $S_m\le t$. Let J be the first time with $S_J\ge t-u$. Earlier endpoints contribute $O(e^{-\beta u})$ by the geometric occupation estimate. After J there are at most $1+u/\log(3/2)$ feasible offsets. For each offset at least N, condition at J and apply Lemma 9 to the continuation, obtaining $O(e^{-cN})$.

An overshoot with $S_J>t-u/2$ has probability at most $Ce^{-\varkappa u/2}$, by summing the one-increment overshoot bound against the geometric occupation estimate below $t-u$. The probability that $S_{J+N}-S_J>u/4$ is at most $M_\varkappa^N e^{-\varkappa u/4}$ by conditioning at J. On the complement of both events, the offsets less than N have $S_m\le t-u/4$, hence discount at most $e^{-\beta u/4}$. Multiplying by $O(1+u)$, choosing L_0 large, and decreasing c gives the asserted bound. □

Proposition 11. *Choose one ascending word for each sorted nonroot solution and let $v(x)=C(w)\mathbf{1}$ be its label. With $L=\log B$ for $B\ge 2$, there is $\eta>0$ such that, uniformly in the target,*

$$\frac{\#\{x:x_n\le B,\ v(x)\in\Lambda\}}{T(B)}\ll D_\Lambda^{-\eta}+L^{-\eta},$$

$$\frac{\#\{x:x_n\le B,\ |\ell(v(x))|\le\epsilon\|\ell\|\|v(x)\|\}}{T(B)}\ll \epsilon^{\eta}+L^{-\eta}.$$

Here Λ is any subgroup of $\mathbb{Z}^d$ of finite index D_Λ, ℓ is a nonzero linear form, and $0<\epsilon<1/2$. In particular a proper real subspace contains $O(T(B)L^{-\eta})$ labels.

For $G\in\mathrm{GL}_d(\mathbb{Q})$ put $D_G=[\mathbb{Z}^d:\{v\in\mathbb{Z}^d:Gv\in\mathbb{Z}^d\}]$ and $K(G)=\max(2,D_G,\|G\|,\|G^{-1}\|)$. The number of pairs of solutions with both maxima at most B and $v(y)=Gv(x)$, divided by $T(B)$, is $O(K(G)^{-\eta}+L^{-\eta})$. If G is determined by d independent pairs of such labels, this is $O(K(G)^{-\eta'})$ for a fixed $\eta'>0$.

Proof. For lattices split an unaccelerated word into the initial run and accelerated part. The initial prefix preserves the index of the inverse-image lattice. If its roof is at most half the total cutoff t, Lemma 10 gives a common exponential saving in $\min(t,\log(2+D_\Lambda))$, and the initial-run weights sum to a bounded factor. Larger initial runs contribute at most $Ce^{\beta t}e^{-(\beta-1)t/2}$, since the factors are linear in the initial run length and their inverse β-powers sum to that tail bound. Apply the nonlinear comparison and $T(B)\asymp L^\beta$ with $t=\log L+O(1)$.

For strips, reverse each finite word. If C is its chronological product and C' the reverse product, the identity $U^{-1}A_iU=A_i^{\mathsf T}$ gives $C'=UC^{\mathsf T}U^{-1}$. Because the row and column sums of U equal $d-1$, the scalar size is unchanged by reversal. The old chronological vector is the reverse-product vector of the newly indexed word. For each nonempty reversed run word choose an initial ordering in which its first label is nonlast. These finitely many orderings cover all words. At the barycenter the scalar roof remains $\log(N(w)/d)$ for each word. Apply the reverse-product strip bound of Lemma 10 to the reindexed upper count; the empty word is absorbed by decreasing η. Letting ϵ tend to zero gives the hyperplane estimate, and any proper subspace is contained in a hyperplane.

For the graph estimate, input and output each determine the other by label injectivity and invertibility. Apply the lattice estimate with index D_G. Every label has norm at most CL by Lemma 6. For $\|G\| \geq 2$, the inputs of norm at most $CL\|G\|^{-1/2}$ number at most $C'(L^\beta \|G\|^{-\beta/2} + 1)$ by the scalar word bound, since their positive coordinate sums are comparable to their norms. For the other inputs, take a row of G of norm at least $\|G\|/\sqrt{d}$. Its value on the input has absolute size at most CL, so these inputs lie in a strip of relative width $O(\|G\|^{-1/2})$. This gives the bound in $\|G\|$, and switching endpoints gives that in $\|G^{-1}\|$. Use the strongest of the three bounds. Finally let X, Y be matrices with the independent input and output labels as columns, so $G = YX^{-1}$. The input lattice contains $X\mathbb{Z}^d$ and hence $D_G \leq |\det X| \ll L^d$. The cofactor formulas and the nonzero integer determinants give $\|YX^{-1}\| + \|XY^{-1}\| \ll L^d$. Thus the time error $L^{-\eta}$ is absorbed after decreasing the exponent. □

Appendix B. Quantitative separation of integer normals

Fix the integers $n \geq 4$ and $k \geq 0$ of the density problem and put $d = n - 1$ and $C = n(k + 1)$. This section works with sorted tuples. Start from a fixed nonroot positive integral solution. In a forward move $j \leq d$, mutate the jth entry, remove it from its old position, and place the new maximum last. A block (A, j), where $A \geq 0$ and $j < d$, consists of this move followed by A moves in position d. Under the Gibbs law in Appendix A, after any finite history,

$$c_*(A + 1)^{-\beta} \leq \Pr((A, j) \mid \text{history}) \leq C_*(A + 1)^{-\beta}, \qquad \beta > 1, \tag{4}$$

where $c_*, C_* > 0$ are fixed. In the notation of Lemma 7, this block has run length $a = A + 1$. Its active label is the label currently in sorted position j. The GMR transition probabilities give (4), and Lemma 7 transfers finite prefix probabilities to height counts.

B.1. **Height approximation on a prefix.** Put $t = k/C$, $s = C^{1/(n-2)}$, and $z_i = s(x_i - t)$ for every entry of a tuple. Each $z_i > 1$: indeed $t < 1/n$ and $s \geq n^{1/(n-2)}$, while $(n/(n-1))^{n-2} < e < n$. Write $e_r(z)$ for the sum of the products of r distinct entries, with $e_0 = 1$. The equation becomes

$$\prod_{i=1}^{n} z_i = \sum_{i=1}^{n} z_i^2 + \sum_{r=0}^{n-2} a_r e_r(z), \qquad a_r \geq 0.$$

In fact, in the variables $y_i = x_i - t$, the coefficient of $e_r(y)$ from the product terms is $k(n - r)t^{n-1-r} - Ct^{n-r} = k(n - r - 1)t^{n-1-r}$ for $0 \leq r \leq n - 1$. It vanishes for $r = n - 1$. The square terms add nt^2 and $2t$ to the constant and linear terms, respectively. Substituting $y_i = z_i/s$ and multiplying by s^2 gives the displayed identity.

For the quadratic in z_n, write $P = \prod_{i=1}^{d} z_i$,

$$E = \sum_{r=1}^{n-2} a_r e_{r-1}(z_1, \ldots, z_d), \qquad Q = \sum_{i=1}^{d} z_i^2 + \sum_{r=0}^{n-2} a_r e_r(z_1, \ldots, z_d).$$

At every nonroot sorted solution z_n is the larger root of $Z^2 - (P - E)Z + Q = 0$, by descent. Put

$$u(x) = (\log z_1, \ldots, \log z_d)^{\mathsf{T}}, \quad h(x) = \log z_n, \quad \Delta(x) = \sum_{i=1}^{d} \log z_i - h(x).$$

The larger-root formula gives

$$\Delta(x) = -\log(1 - E/P) - \log\left(\frac{1 + \sqrt{1 - 4(Q/P^2)/(1 - E/P)^2}}{2}\right).$$

This is nonnegative and increasing in each of E/P and Q/P^2 on the larger-root domain. Every monomial in these two ratios has nonpositive exponents. Forward moves increase the entries of

the sorted tuple coordinatewise, so Δ never increases. If $min_i x_i \geq Y$, both ratios are $O(Y^{-2})$: a monomial of E/P omits at least two slots and the terms z_i^2/P^2 each have at least two inverse factors. The terms e_r/P^2 satisfy the same bound. For Y sufficiently large the root formula thus gives $\Delta = O(Y^{-2})$.

Let

$$M_j u = (u_1, \dots, \widehat{u}_j, \dots, u_d, \textstyle\sum_{i=1}^d u_i)^{\mathsf{T}}.$$

Here a hat denotes omission. The exact log recurrence is $u(x') = M_j u(x) - \Delta(x) e_d$, where e_d is the last standard vector. The matrix of a chronological sequence $w = j_1 \cdots j_m$ is $M_w = M_{j_m} \cdots M_{j_1}$, with empty product the identity. Write $v_w = M_w^{\mathsf{T}} \mathbf{1}$, with $\mathbf{1}$ the all-one column. For $b = (0, \dots, 0, 1, 1)^{\mathsf{T}}$, inspection gives $M_j b - b \geq e_d$ coordinatewise. Along a tail W from x_p, nonnegativity of the matrices and monotonicity of the defect give by induction

$$M_W(u(x_p) - \Delta(x_p) b) + \Delta(x_p) b \leq u(x_{pW}) \leq M_W u(x_p).$$

Summing and subtracting the terminal defect gives an error between zero and $\Delta(x_p) \mathbf{1}^{\mathsf{T}} M_W b$. Thus with $\lambda_p = M_p^{-1} u(x_p)$,

$$|h(x_{pW}) - \lambda_p \cdot v_{pW}| \leq \sqrt{2}\, \Delta(x_p) \|M_p^{-1}\| \|v_{pW}\|. \tag{5}$$

All norms hereafter are Euclidean or induced operator norms. In the slot convention the first d entries are kept in place except that the previous maximum is inserted in the mutated slot. The corresponding forward log matrix is $I + e_i(\mathbf{1} - e_i)^{\mathsf{T}}$ in slot i. If P_w is the permutation carrying the first-slot order to the sorted order after w, each sorted matrix is the corresponding slot matrix multiplied on the left by the new permutation and on the right by the inverse of the preceding one. Thus the sorted product is P_w times the slot product times $P_\varnothing^{-1}$. Taking transposes and applying to $\mathbf{1}$ cancels P_w and leaves the fixed initial permutation $P_\varnothing$.

B.2. **Almost-sure limits of pulled-back logs.** A move $j < d$ gives a new maximum at least $(n + k - 1)$ times the product of the two largest entries before that move. Indeed the retained product includes both entries, the coefficient of that product in Vieta's formula is at least $n + k$, and the removed entry is at most the retained product. At the end of a block the second-largest entry is at least the preceding block's maximum. The logarithms of successive block maxima therefore satisfy a Fibonacci lower bound. In particular, if $\varphi = (1 + \sqrt{5})/2$, the log maximum at depth q is at least $c\varphi^q$. If the initial minimum is Y sufficiently large, this bound holds with $c \log Y$ in place of c, for a constant $c > 0$ independent of the seed. A complete block of d consecutive $(0, 1)$ moves raises the minimum to at least the maximum at its start. If one occurs in the last half of the first q blocks, then $\Delta(x_q) \leq C_1 \exp(-c_1 \varphi^{q/2})$, with an additional factor $\log Y$ in the exponent in the thick-seed case.

The norm of one block matrix and of its inverse is at most $C_2(A+1)$, since $M_d = I + e_d(\mathbf{1} - e_d)^{\mathsf{T}}$ and the added matrix squares to zero. The log increment across a block, before pullback, is bounded by $C_3(A + 1)\Delta(x_q)$. To see this, use the coordinatewise sandwich with block matrix $M_d^A M_j$ and bound its norm. Pulling back multiplies this error by at most the accumulated inverse norm. By (4), large runs with $\log(A_q + 1) > aq$ have summable probabilities for any $a > 0$. Also, failure of a complete block in the last half at depth q has probability $O(e^{-cq})$, by testing disjoint windows, each with conditional success probability at least c_*^d. Consequently, almost surely the pulled-back logs converge to a vector Λ_x. The convergence follows from summing errors of size $\exp(O(q^2) - c\varphi^{q/2})$. If p is a prefix and ω a continuation, then

$$\Lambda_x(p\omega) = M_p^{-1} \Lambda_{x_p}(\omega). \tag{6}$$

The same limit is obtained by using the vector $u^0(x) = (\log x_i + \log s)_{i=1}^d$. Their difference is $O(1/\min_i x_i)$, which vanishes after pullback along the same regular paths by the complete-block bound.

For a seed of minimum at least Y sufficiently large we will use the uniform estimate

$$\Pr(\|\Lambda_x - u^0(x)\| > K/Y) \le KY^{-\nu} \tag{7}$$

with $K, \nu > 0$ independent of the seed. For any fixed $0 < \theta < \beta - 1$, the conditional θ-moment of $C_2(A+1)$ is bounded. Iterating conditional expectations and summing Markov bounds over $q \ge 1$ gives accumulated matrix and inverse norms at most $e^{Dq}Y^b$ outside an event of size $O(Y^{-\theta b})$, for fixed large D and any fixed $b > 0$. The same bounds can be imposed on the products of the individual norm upper bounds. For q below $R = \lfloor a \log Y \rfloor$, defect monotonicity bounds the error sum by $O(Y^{-2}Y^{2b}\sum_{q\le R} e^{2D(q+1)}) = O(Y^{-3/2})$ on taking a, b sufficiently small. For every $q \ge R$ require a complete block in its last half, at total cost $O(e^{-cR})$. The error sum from these depths is bounded using $\Delta(x_q) \le C_1 \exp(-c_1(\log Y)\varphi^{q/2})$ and is $O(Y^{-3/2})$. The difference $u - u^0$ at the initial state is $O(Y^{-1})$, proving (7).

B.3. **Uniform avoidance of a prescribed hyperplane.** Write $X = x_n$. Fix $i < d$, freeze the entries with indices in $S = \{1, \dots, d\} \setminus \{i\}$, and let P_i be their product. For $T \subseteq S$ put

$$E_T = C - k\sum_{j\in T} x_j^{-1}, \qquad D_T = \log(C/E_T), \qquad A_i = P_i E_S.$$

In the two active slots, $b_0 = x_i, b_1 = X$ and $b_{m+1} = A_i b_m - kP_i - b_{m-1}$. If $\lambda > 1$ is the root of $\lambda + \lambda^{-1} = A_i$ and $q = kP_i/(A_i - 2)$, then $b_m = q + a\lambda^m + b\lambda^{-m}$. Here $A_i \ge (n+2k)P_i$, so $q < 1/2$. To compute ab, the active conic is

$$b_m^2 + b_{m+1}^2 - A_i b_m b_{m+1} + kP_i(b_m + b_{m+1}) + \sum_{j\in S} x_j^2 = 0.$$

After shifting each active entry by q, the constant term is $\sum_{j\in S} x_j^2 + (kP_i)^2/(A_i - 2)$. Substituting the two powers of λ leaves the mixed term $-(A_i^2 - 4)ab$. Hence $ab > 0$ and is bounded uniformly, as the frozen squares sum to at most $(n-2)P_i^2$. Since X is the larger root when x_i is retained, $X - q \ge A_i(x_i - q)/2$, so $a \ge b > 0$ and $a \ge x_i/4$. Also $x_i^2 \le \prod_{j=1}^d x_j$ since $i < d$. The upper Vieta bound $X \le C\prod_{j=1}^d x_j$ then gives $X \le K_1\lambda^2$. The lower Vieta bound $X \ge (n + k - 1)\prod_{j=1}^d x_j$ gives $A_i \le K_2 X$. Constants here and below may depend on n, k and (4).

Make m moves in the active pair and then clear the frozen slots in some order $j_1, \dots, j_{d-1}$. The first move is noncusp and the next $m - 1$ are cusp moves. Every clearing is noncusp because an active entry greater than each frozen entry now precedes the maximum. Thus this specifies exactly d blocks, with first exponent $m - 1$ and all other exponents zero.

Write $g = \log(A_i/\lambda)$, and let S_l be the frozen set left after l clearings, so $S_0 = S$. In initial slot coordinates, the pulled-back vector using u^0 approaches

$$\xi = u^0(x) - \log(x_i/a)e_i - ge_{j_1} - \sum_{l=1}^{d-1}(D_{S_{l-1}} - D_{S_l})e_{j_l}. \tag{8}$$

Here is the computation. In the slot convention, the active move has matrix $I + e_i(\mathbf{1} - e_i)^{\mathsf{T}}$. The total of the frozen normalized logs is $\log(CP_i) = \log\lambda + g + D_S$. Since $\log b_m = m\log\lambda + \log a + O(\lambda^{-m})$, before the first clearing the pulled-back vector equals $u^0(x) - \log(x_i/a)e_i - m(g + D_S)e_i + O(\lambda^{-m})$. The raw defect at that time is $\log(CP_i b_m/b_{m+1}) = g + D_S + O(\lambda^{-m})$. When the first clearing acts, the inverse of the previous matrix product sends its error direction to $e_{j_1} - me_i$. This cancels the linear term in m. For clearing $l \ge 2$, the newest inverse subtracts $e_{j_{l-1}}$ from e_{j_l}. All earlier inverses preserve that difference because its sum is zero and its support excludes their slots. The raw defect before this clearing is $D_{S_{l-1}} + O(1/b_m + X/b_m^2)$. Indeed the current maximum arose from multiplying the other entries by $E_{S_{l-1}}$ with a relative error of that size: the large reciprocal terms cost $O(1/b_m)$, the removed entry was at most X, and the product includes

both large active entries. Summing the error directions telescopes to (8). The inverse norm for the whole prescribed word is $O(m+1)$, so the finite error in this computation is

$$O\big((m+1)(\lambda^{-m}+X\lambda^{-2m})\big).$$

After the clearings the minimum is at least $b_m \geq \lambda^m/4$. By (7), the random continuation adds $O((m+1)\lambda^{-m})$ to the pulled-back error except with conditional probability $O(\lambda^{-\nu m})$. Both error estimates divided by X^{-2} tend to zero uniformly, using $X \ll \lambda^2$ and the fixed lower bound on $\lambda > 1$.

For two orders that interchange the first two slots j, j', subtracting (8) gives $e_{j'} - e_j$ times

$$g + \log \frac{(E_S + k/x_j)(E_S + k/x_{j'})}{E_S(E_S + k/x_j + k/x_{j'})}.$$

The logarithm is nonnegative, and $g = \log(1+\lambda^{-2}) \geq cA_i^{-2} \geq c'X^{-2}$. For a vector v, let $m(v)$ be the largest absolute pairing with the finitely many candidates (8), ranging over all $i < d$ and all orders. For any j, j' not equal to some $i < d$, the difference of the corresponding coordinates of v is $O(X^2 m(v))$. These edges form a connected graph on the slots: with $d \geq 3$, the stars through the slot d obtained using the other available omitted indices suffice. All coordinates therefore differ from their average by $O(X^2 m(v))$. Each candidate has norm $O(1 + \log X)$ and coordinate sum $\log(\lambda a) + \log s \gg 1 + \log X$. The latter follows from $X = q + \lambda a + \lambda^{-1} b$, $q < 1/2$, and $b \leq a$. Pairing with any one candidate now bounds the average by $O(X^2 m(v))$ as well. Thus $m(v) \geq cX^{-2}\|v\|$.

Lemma 12. *For every $0 < \epsilon < 1/4$, there is $B_\epsilon \leq \exp(C_4[\log(2/\epsilon)]^{\beta+2})$ such that, uniformly in the nonroot seed x, the law satisfying (4), and $v \neq 0$,*

$$\Pr\big(|v \cdot \Lambda_x| \leq X^{-B_\epsilon}\|v\|\big) \leq \epsilon.$$

Proof. Choose $m = O(\log(2/\epsilon))$ sufficiently large. A candidate for which the pairing is maximal specifies a cylinder of d blocks. Except on a conditional set of probability at most $\epsilon/4$, the pairing there is at least $cX^{-2}\|v\|$ by the error estimates. By (4) its probability is at least $c'(m+1)^{-\beta}$. Remove that cylinder, and repeat the construction at every surviving prefix of the next depth divisible by d, using the transported normal $M_p^{-\mathsf{T}} v$. After $R = O([\log(2/\epsilon)]^{\beta+1})$ rounds the remaining probability is at most $\epsilon/4$. The selected cylinders are disjoint, so their bad conditional fractions total at most $\epsilon/4$. Choose a cutoff J with $\log(J+1) = O(\log(2/\epsilon))$ such that the probability of any run beyond J in the first dR blocks is at most $\epsilon/4$. This follows directly by summing the transition tails. At every selected prefix obeying the cutoff, $\|M_p\|$ is at most $[C_2(J+1)]^{dR}$ and the log of the current maximum is at most a constant times $[C_2(J+1)]^{dR}(1+\log X)$, by iterating Vieta's upper bound, or the log sandwich. The transported normal has norm at least $\|v\|/\|M_p\|$. Thus (6) gives the claimed threshold on choosing B_ϵ to dominate a fixed multiple of $[C_2(J+1)]^{dR}$. The three exceptional probabilities sum to at most $3\epsilon/4$. □

B.4. **Finite prefix partitions.** Fix $a > 0$ and a requested exponent $A_0 > 0$. For a large normal cutoff M put $K = (\log M)^a$ and $z = \log\log M$. At depth $q_0 = \lceil C_0 z \rceil$, take C_0 sufficiently large in terms of a, n, k, A_0 and the law bounds. For estimates on the probability of finite events, use the reference event that the run at position l satisfies $A_l + 1 \leq K\exp(l)$ for all l, and there is a complete block in the last half of each depth $q \geq q_0$. Its complementary probability is

$$O\big(K^{1-\beta} + e^{-cq_0}\big)$$

by summing the tail and complete-block failure estimates. The accumulated matrix and inverse norms on this event are bounded by $\exp(O(q\log(K+2) + q^2))$. For a small fixed $c_1 > 0$ put

$e_q = \exp(-c_1\varphi^{q/2})$. Since $\varphi^{q_0/2}$ dominates $q_0\log(K+2)+q_0^2$ when C_0 is large, the bound on pulled-back errors proves

$$\|\Lambda_x - \lambda_p\| \le e_q$$

on the reference event for the depth-q prefix p, simultaneously for $q \ge q_0$. Independently of this reference event, every finite prefix of depth $q \ge q_0$ with runs at most K and a complete block in its last half has the all-tail error (5) with coefficient at most e_q.

For the finite construction, keep only prefixes with runs at most K and a complete block in the last half of every tested depth. For any such prefix at depth q the matrix and inverse norms are at most $R(q) = [C_2(K+2)]^q$ and $\log X_p \le H(q)$, where $H(q)$ is a constant times $R(q)$ by the log recurrence for the fixed starting seed. At depth q_0 retain as candidates the integer normals b with $0 < \|b\| \le M$ and $|\lambda_p \cdot b| \le 10Me_{q_0}$. Their span has dimension at most $d-1$. In fact, d independent candidates form an integer column matrix of determinant of absolute value at least one. Its cofactors give $\|\lambda_p\| = O(M^d e_{q_0})$, whereas $\|\lambda_p\| \ge \|u(x_p)\|/R(q_0) \gg R(q_0)^{-1}$. These bounds contradict each other for large M on increasing C_0.

Fix $0 < \epsilon < 1/4$, and use B_ϵ from Lemma 12. At a retained prefix of depth q_s with nonempty candidate set, designate one candidate b_0. By (6) and Lemma 12, the conditional probability that

$$|\Lambda_x \cdot b_0| \le \sigma_s\|b_0\|, \qquad \sigma_s = \exp(-B_\epsilon H(q_s))/R(q_s),$$

is at most ϵ. Indeed $\|M_p^{-\mathsf{T}} b_0\| \ge \|b_0\|/R(q_s)$. Refine every current prefix to the least common depth $q_{s+1} > q_s$ with $e_{q_{s+1}} \le \sigma_s/(C_5 M^d)$, for a sufficiently large fixed C_5. For each child at that depth require that its center differ from its center at each previous tested depth q_t by at most $2e_{q_t}$, and, when b_0 was designated, require its pairing to exceed $\sigma_s\|b_0\|/2$ in absolute value. On the reference event the center comparisons always pass. On that same event the last test fails on a subset of the conditional exceptional event of size at most ϵ. Summation over parents costs at most ϵ, since their cylinders are disjoint.

The child's candidate set is the intersection of the parent's candidates with the set pairing with its center by at most $10Me_{q_{s+1}}$. If the rank is still the parent's positive rank r, write b_0 as a linear combination of r independent child candidates. By choosing an invertible integer coordinate minor, Cramer's rule bounds each coefficient by $O(\|b_0\|M^{r-1})$. Thus the child's pairing with b_0 is at most $O(\|b_0\|M^r e_{q_{s+1}})$, smaller than $\sigma_s\|b_0\|/2$ by the choice of C_5. This contradicts retention. After $d-1$ refinements all candidate sets are empty.

Write $Q = q_{d-1}$. If a normal of length at most M paired with the depth-Q center by at most $5Me_Q$, its pairing with each earlier tested center would be at most $5Me_Q + 2Me_{q_t} \le 7Me_{q_t}$. It would have survived every candidate intersection. Thus all pairings of nonzero such normals at depth Q exceed $5Me_Q$. Refine once more to the least depth $q_* > Q$ with $e_{q_*} \le 2Me_Q/(100M^{A_0})$ and keep the finite tests for runs, complete blocks, and center comparisons, including the comparison with depth Q. Then for every retained prefix p and every finite descendant,

$$\begin{aligned} |\lambda_p \cdot b| &> \tau \qquad && (0 < \|b\| \le M,\ b \in \mathbb{Z}^d), \\ |h(x_{pW}) - \lambda_p \cdot v_{pW}| &\le \delta\|v_{pW}\|, && \delta \le \tau/(100M^{A_0}), \end{aligned} \tag{9}$$

where $\tau = 2Me_Q$ and $\delta = e_{q_*}$.

Taking logarithms twice in the minimal depth inequality gives

$$q_{s+1} \ll \log(2+B_\epsilon) + q_s\log(K+2) + \log\log M.$$

The last refinement can be made at $Q+1$ for large M, since e_{Q+1}/e_Q is smaller than any prescribed power of M^{-1}. Therefore $q_* \ll (z+\log(2+B_\epsilon))z^{d-1}$, and the log of the number of retained prefixes is at most a constant times $(z+\log(2+B_\epsilon))z^d$, by counting at most $(d-1)(\lfloor K\rfloor+1)$ choices at each depth. The exceptional probability is at most

$$(d-1)\epsilon + O(q_* K^{1-\beta} + e^{-cq_0}).$$

This charges the reference failure once, the run failures by their union bound, and the selected normal failures once per depth. The retained partition uses only finite prefixes and (9) holds for every tail.

B.5. **Transfer to height counts.** Let $L = \log B$ for a maximum cutoff B. The counting estimates give a total comparable to L^β, and the number of words through an event at depth q is at most a constant times L^β times its Gibbs probability, uniformly in q. Words of depth below q are bounded in count by $O(L^{s_0} C_{s_0}^q)$ for some fixed $1 < s_0 < \beta$ and constant C_{s_0}. In particular for $q = o(\log L)$ their normalized count is at most $O(L^{-\eta_0})$ for some $\eta_0 > 0$. These statements also hold for each of the finitely many initial seeds obtained on imposing a fixed cutoff on the initial cusp run. The discarded proportion for that initial cutoff tends to zero as the cutoff increases. If a fixed initial word has matrix M_0, a local label v becomes $M_0^{\mathsf{T}} v$ and a local center λ becomes $M_0^{-1}\lambda$. A global integer normal b of length at most M gives the local integer normal $M_0^{-\mathsf{T}} b$ of length at most $\|M_0^{-1}\| M$, since M_0 is unimodular. The flatness coefficient is multiplied by at most $\|M_0^{-1}\|$. A fixed larger local cutoff and fixed stronger flatness ratio therefore give the same assertions in the global labels. Take the maximum of the height and matrix bounds over the finitely many initial seeds before choosing any of the tested depths. All constructions then have the same tested depths and hence common separation and flatness thresholds.

For M a fixed multiple of L^{d+1} and ϵ fixed, the depth is a fixed power of $\log\log L$ and the number of parts is $L^{o(1)}$. The normalized count discarded after the initial cusp cutoff is $O(\epsilon) + o(1)$. For integer bands $j \geq 2$, take $M = \exp(\kappa j)$ and $\epsilon = (j+1)^{-2}$, with fixed $\kappa > 0$. Now $\log(2 + B_\epsilon) = O((\log(j+1))^{\beta+2})$, so the depth and log of the number of parts are fixed powers of $\log(j+1)$, with the number of parts $\exp(o(j))$. Choose $a(\beta - 1) > 4$ and enlarge C_0 so the complete-block error is $O(j^{-4})$. For all sufficiently large j the exceptional symbolic mass is $O((j+1)^{-2})$. Uniformly for j up to any fixed multiple of $\log L$ the depths are $o(\log L)$. The normalized integer loss in band j is therefore $O((j+1)^{-2} + L^{-\eta_0})$, with the separation and flatness in (9).

Appendix C. Termination of rational overlaps

Fix an integer $r \geq 3$. All vectors are columns except covectors, which are rows. Let $\mathbf{1}$ denote the all-one column, e_i the i-th standard column, $J = \mathbf{1}\mathbf{1}^{\mathsf{T}}$, and

$$B_i = I + e_i(\mathbf{1} - e_i)^{\mathsf{T}} \qquad (1 \leq i \leq r).$$

For a finite word $w = i_1 \cdots i_m$, write $B_w = B_{i_1} \cdots B_{i_m}$, with the identity for the empty word. We use Euclidean vector, covector, and induced operator norms unless otherwise indicated. The simplex Δ_r consists of nonnegative vectors with coordinate sum one. We represent a nonnegative projective direction by its representative in Δ_r. The Rauzy gasket $\mathcal{R}$ is the intersection, over $m \geq 0$, of the unions of the projective simplices $B_w \Delta_r$ for all words of length m.

C.1. **The law of the input word.** The ordered state space is the part of Δ_r where $x_1 \leq \cdots \leq x_r$. Let Q_j be the permutation matrix moving position j to the end and preserving the order of the others. The accelerated letter (a, j), with $a \geq 1$ and $1 \leq j < r$, has matrix $\gamma_{a,j} = Q_j B_j^a = B_r^a Q_j$ and projective map $f_{a,j}$. Let h be the positive C^1 eigenfunction for the critical exponent $\beta > 1$ of the Gamburd–Magee–Ronan operator, normalized so that its transition probabilities are

$$p_{a,j}(x) = \big(1 + a(1 - x_j)\big)^{-\beta} \frac{h(f_{a,j}(x))}{h(x)}.$$

We use the existence of this eigenfunction and the uniform exponential shrinking of ordered compositions from [15, Theorems 10, 13, and 39 and Proposition 43]. Thus there are $D > 0$ and $0 < \rho < 1$ such that each ordered composition of m letters has image diameter at most $D\rho^m$. The probabilities sum to one. Their logarithms are uniformly Lipschitz: the logarithmic derivative

of $1 + a(1 - x_j)$ is bounded because $x_j \le 1/2$, and the projective maps have uniformly bounded derivatives on the ordered simplex. The latter follows by differentiating $B_j^a x/(1+a(1-x_j))$, whose numerator and derivative are $O_r(1+a)$. Also $p_{a,j}(x) \le C_r a^{-\beta}$, and $\max(\|\gamma_{a,j}\|, \|\gamma_{a,j}^{-1}\|) = O_r(1+a)$. Consequently the cost

$$\ell(a, j) = 1 + \log \max(\|\gamma_{a,j}\|, \|\gamma_{a,j}^{-1}\|)$$

has a uniform conditional exponential moment. Throughout, C_r denotes a positive constant depending only on r, whose value can change.

Write P_t for the permutation taking the actual coordinate order to the ordered positions after t runs. An actual run uses the actual index $i = P_t^{-1}(j)$ for the exponent a, then changes P_t to $Q_j P_t$. Thus the next index can be any index other than the one in the final position. If C_m is the ordered-in-time product of the actual B-matrices after m runs and L_m is the corresponding ordered-in-time product of $\gamma_{a,j}^{\mathsf{T}}$, multiplication and transposition give

$$C_m = P_0^{-1}(J - I)^{-1} L_m (J - I) P_m.$$

Indeed the forward product of the $\gamma_{a,j}$, latest factor on the left, is P_m times the corresponding latest-first product of actual B-matrices times P_0^{-1}. Its transpose reverses the order and turns B_i into $(J - I)B_i(J - I)^{-1}$. This also proves the formula for continuations from every current state.

We establish a probabilistic input needed below. For a chronological suffix w of m ordered letters define f_w by applying them in chronological order and set

$$q_m(b \mid w) = \inf_x p_b(f_w(x)), \qquad \alpha_m = \inf_{|w|=m} \sum_b q_m(b \mid w).$$

At depth zero take $f_w = I$. A longer suffix gives a smaller possible state set, so the envelopes increase when another letter is supplied on the left. In particular α_m is nondecreasing. Uniform logarithmic distortion gives $\alpha_m \ge e^{-LD\rho^m}$, where L is a common Lipschitz constant. For each letter partition a uniform random variable's interval into the successive envelope increments and the residual at the current finite history. The ordering of the intervals is by depth and then by a fixed enumeration of letters. Their lengths telescope to the original probabilities. If the uniform variable is smaller than α_m, the selected letter depends only on the last m letters.

Let $U_0, U_1, \dots$ be the independent uniforms used for this construction. Call t a cut when $U_{t+s} < \alpha_s$ for every $s \ge 0$. Its probability is $b = \prod_m \alpha_m > 0$. On a cut at t, the condition that an earlier $j < t$ be a cut reduces to its uniform restrictions before t, because α_m is nondecreasing. Therefore the event of the first cut at t factors into an event in the uniforms before t and the event of a cut at t. Across consecutive cut intervals the same factorization shows that the cut-to-cut marked words are independent and identically distributed under conditioning on a cut at their initial time. The words are marks of the intervals because the update within each interval depends on no earlier letters. The initial word is independent of the sequence of marked words, whose law is independent of the initial state.

Here is an exponential tail justification. Under the cut law, the probability of another cut at time m is $u_m = \prod_{j=0}^{m-1} \alpha_j$, with $u_0 = 1$. For $m \ge 1$ put $f_m = u_{m-1} - u_m$. Its generating function $F(z) = \sum_{m \ge 1} f_m z^m$ has radius of convergence greater than one and $F(1) = 1 - b < 1$. The renewal generating function $\sum_{m \ge 0} u_m z^m$ equals $(1 - F(z))/(1 - z)$. Its gap distribution therefore has generating function $1 - (1 - z)/(1 - F(z))$, analytic on a disk of radius greater than one. The cut set may also be constructed with uniforms indexed by all integers. It is a stationary renewal set, of intensity b, and the first forward waiting time has the residual gap tail, which is exponential. These identities are also the renewal construction of [22, Corollary 4.18 and §8]. The log-cost of the initial word and an ordinary marked word has an exponential tail. To prove this, split according to whether the word is longer than $\lfloor \delta s \rfloor$, for a small fixed $\delta > 0$, and on the other event use the exponential moment of the accumulated costs of that many letters and Markov's

inequality at cost s. The Palm conditioning costs at most $1/b$. The iteration of the conditional exponential moments is at a deterministic time. No cut is used as a stopping time for the input filtration.

Let μ be the law of the products of transposed ordered matrices of one marked word. Every ordered branch occurs in this support: fixing the level-zero interval for a branch at time zero and a cut at time one has conditional probability $q_0(b) > 0$ given a cut at zero. The quotients $\gamma_{2,j}\gamma_{1,j}^{-1} = B_r$ recover B_r, and then each Q_j. The first and last Q_j generate all permutations, so conjugation gives all B_i. For clarity, the group generated by all B_i is Zariski dense in SL_r. To see this by transposing, let $u_i = \mathbf{1} - e_i$, $M_{ij} = u_i e_j^{\mathsf{T}}$, and $N_i = M_{ii}$. The powers of $I + N_i$ give N_i in the Lie algebra, and

$$[N_i, N_j] = M_{ij} - M_{ji}, \qquad [N_k, [N_i, N_j]] = M_{kj} + M_{jk} - M_{ki} - M_{ik}$$

for distinct indices. Since $J - I$ is invertible, the M_{ij} form a basis of matrices. These brackets, together with the N_i, span all trace-zero matrices: the second set spans the differences of symmetric off-diagonal pairs by connectivity of the graph of pairs sharing an index. This proves the claim. The transposed branch law has the same Zariski-density conclusion. The standard and second exterior representations are strongly irreducible because their restrictions to connected SL_r are irreducible. A product using all actual indices is strictly positive, and supplies a proximal matrix in the standard representation after absorbing its final permutation by taking a repeated complete ordered pattern with trivial final permutation.

The projective stationary measure of μ satisfies a uniform power bound on hyperplane strips. In applying [23, Proposition 2.56], one can pass to the product at the first return of the determinant sign to $+1$. The sign-return time has exponential tail, and its product has an exponential log-cost moment by the same length splitting argument. Its support generates the determinant-one kernel: a positive atom, a pair of negative atoms, and a pair of negative atoms enclosing a positive atom generate the index-two kernel by the usual two-coset rewriting. The stationary measure is unchanged, since successive returns group the same infinite independent products. The right-product limit under consideration exists: a strictly positive finite product recurrently contracts Hilbert distance, and the other nonnegative matrices are nonexpanding. The uniqueness of the stationary law is part of the strongly irreducible proximal theorem used in [23]. The original future-direction law is a mixture of pushforwards by the independent initial word, a starting permutation, and the fixed conjugacy above. For strip width $s \in (0, 1)$, restrict the condition number of the initial matrix to have both norms at most $s^{-1/4}$. The strip then pulls back to width $O_r(s^{1/2})$, and the exceptional cost is a positive power of s. This proves a bound Cs^c, with $C, c > 0$ uniform over the actual current state and every hyperplane through zero.

We will also use log-norm estimates for actual deterministic run counts. Let $\lambda_1 > \lambda_2$ be the two largest exponents of the marked block matrix law, and let $\bar{t}$ be the expected block length. The largest exponent is simple by standard proximal and strong irreducibility. Set $\Lambda_i = \lambda_i/\bar{t}$ for $i = 1, 2$. Uniformly in the current state and every nonzero covector v,

$$\mathbb{E}\log\frac{\|vC_m\|}{\|v\|} = m\Lambda_1 + o(m),\ \mathbb{E}\log\|\wedge^2 C_m\| \le m(\Lambda_1 + \Lambda_2) + o(m). \tag{10}$$

Indeed, [21, Theorem 3.1] gives exponential large deviations for the vector norm on the transposed standard representation and the operator norm on the exterior-square representation. The latter has top exponent $\lambda_1 + \lambda_2$ by the singular value formula. Fix a small positive loss ϵ. By the exponential length tail and Chernoff's inequality, the number of full marked blocks in m runs lies within $O(\epsilon m)$ of $m/\bar{t}$, except with exponentially small probability. Apply the large-deviation inequalities for every deterministic count in that interval and take their union. The initial and final partial costs exceed ϵm with exponentially small probability: the final cost is at most the largest cost of the first $m + 1$ marked blocks. Condition on the independent initial word before

using the uniform vector estimate. The conjugacy and the initial and final partial weights change logarithmic norms by only $O(\epsilon m + 1)$. Finally the total raw log-cost has second moment $O(m^2)$, so Cauchy–Schwarz bounds the expectation on the exceptional event by $o(m)$. Letting ϵ decrease proves (10).

C.2. **Cylinders and reduced matrices.** The first projective simplex for letter i is the set whose i-th coordinate is at least $1/2$. Two distinct first simplices meet only at their edge midpoint. The projective map of B_i is $x_i \mapsto 1/(2 - x_i)$, $x_j \mapsto x_j/(2 - x_i)$ for $j \neq i$. Its derivative on the simplex has ℓ^1-norm at most $1/(2 - x_i)$: it suffices to test the extremal tangent vectors $(e_j - e_k)/2$. For $j, k \neq i$ the ratio is $1/(2 - x_i)$; for indices i, j it is $(2 - x_i - x_j)/(2 - x_i)^2$. Hence the maps are nonexpanding, and inside a different first simplex the factor is at most $2/3$. A constant run of length a has diameter $2/(a + 1)$, by its simplex vertices. Thus cylinders shrink uniformly as raw word length grows, and their diameters decrease exponentially in run count. The gasket meets a first simplex in the corresponding image of the gasket. This follows from the first-level union and the shared midpoints, which are images of vertices from both cells. Iteration gives the analogous cylinder identity. The entire edge between any two vertices belongs to the gasket, since the two corresponding first maps split that edge into its two halves. Strictly positive points are dense in the gasket: each finite cylinder contains such a point by following it with a complete positive word.

We need the following classification. Suppose a nonzero nonnegative matrix acts projectively on every positive point of $\mathcal{R}$ and maps it into $\mathcal{R}$. The map extends to every point with nonzero image by continuity. Each pair of nonzero column directions joins by a segment in the gasket. Any segment not contained in a first simplex crosses between two first simplices through their common midpoint. Since all coordinates are nonnegative, each coordinate zero at such an interior point vanishes on the entire segment. Thus the segment lies on the original supporting edge. Stripping common first simplices and using uniform cylinder shrinking shows that every nondegenerate segment in the gasket lies on an edge of some finite cylinder.

If a triple of noncollinear points has its three sides in the gasket, either the three points lie in one first simplex, they are three original vertices, or they are three pairwise contact midpoints. To verify this, associate to each point its set of one or two incident first simplices. If all pairs intersect but the total intersection is empty, the three sets must be $\{i, j\}, \{j, k\}, \{i, k\}$. Otherwise a pair with disjoint sets lies on an original edge. A nonvertex point on that edge cannot connect to a point off the edge in a different first simplex. A point off the edge connecting to both endpoints of the chosen pair would then have to be in both their first simplices, impossible off the edge. Thus both members of that pair are vertices; connecting an off-edge point to the two vertices now requires that it too be a vertex.

For rank at least three, strip common first simplices as long as possible. This means applying B_i^{-1} on the left, preserving nonnegativity and the gasket mapping by the cylinder identity. Uniform shrinking ensures that the stripping terminates. If the remaining column directions all have pairwise intersecting membership sets, three contact midpoints must occur, since the total intersection is empty. The relative interior of their triangle is in the first-level hole, whereas a positive point in the corresponding input three-coordinate gasket face maps into this relative interior. The gasket on a coordinate face is the gasket for its supported indices, directly from the first-level formula; such positive points exist there. Otherwise choose a disjoint-membership pair and a third direction off its line. The three directions are original vertices, and every other nonzero column must also point along a vertex by applying the same triangle fact with a noncollinear pair of these vertices. The multipliers of any three distinct coordinate images must agree. Indeed, if $a \leq b, c$ with at least one strict inequality, the point $(1/2, c/(2(b + c)), b/(2(b + c)))$ lies on an edge of the first simplex for coordinate one, and its image under $\mathrm{diag}(a, b, c)$ lies in the hole since $2bc > a(b + c)$. Thus all nonzero columns have a common multiplier. After dividing by it,

duplicate columns e_p in positions a, d, together with columns e_q, e_s in positions b, c of distinct images, give the columns $3e_p, 3e_p + e_q, 3e_p + e_s$ in positions a, b, c after right multiplication by B_aB_d. On that face stripping B_p^3 contradicts diagonal rigidity. A zero column at d becomes a duplicate at d after right multiplication by B_a, where column a is e_p, and then left multiplication by B_p^{-1}. The classification now gives a scalar permutation preceded by the stripped word, hence full rank. At rank two the entire image is the segment joining two extreme column directions and is contained in a finite-cylinder edge. At rank one the image point belongs to $\mathcal{R}$.

For a nonnegative invertible matrix H with rational entries, an output letter is legal if $B_i^{-1}H \geq 0$ entrywise. Two different letters cannot be legal simultaneously because every column would then belong to the intersection of two first cones, giving rank at most one. Full stripping terminates because each legal letter decreases the entry sum by at least the reciprocal of a fixed common denominator. A matrix with no legal letter is called reduced. If a reduced H is not a positive scalar permutation, no further finite input and output can make it one. In fact, the i-th column of HB_i is its unique componentwise least column: every other column is the sum of that column and a nonzero column of H. Thus an equality $HB_v = cB_uP$ with a positive scalar c and a permutation P can be right-cancelled after moving P to the left of the input word by conjugation. The rightmost letters agree successively. The right side cannot run out first because a permutation has no componentwise least column. It follows that H was a scalar word-permutation; reducedness forces the word to be empty.

Starting at a nonnegative remainder H, at every run append the actual input on the right and strip all legal outputs on the left. Let H_m denote the remainders, and let E be the event that the initial matrix sends the future boundary direction into the gasket with positive orientation. Write p_m for the conditional probability of E at the current input time. The cylinder identity shows that this is the probability the current remainder sends the original, unconditioned future direction into the gasket. Thus $p_m \to 1$ almost surely on E. If a nonterminal reduced remainder preserved an entire permitted first corner of the gasket, the classification applied to HB_i would give a scalar word-permutation, in contradiction with right cancellation. The allowed initial run index excludes only the last active index, and all run lengths at that allowed index, and all allowed continuations, have positive probability. Consequently an open hole inside a permitted corner gives an escaping finite cylinder of uniformly positive probability over the ordered state. Since bounded rational remainders with a fixed denominator form finite sets, nonterminal survival implies that their entry sums tend to infinity. Their determinants are fixed and nonzero, so normalized limits are singular. The same cylinder argument shows that along a subsequence with fixed finite ordering, every normalized limit preserves each allowed first corner. Apply the classification after right multiplication by an allowed B_i, which does not change rank or range. Such a limit has rank one with image point in the gasket, or rank two with range a finite-cylinder edge line.

C.3. **Deterministic area estimates.** For a nonnegative matrix H, let $N(H)$ be its entry sum and $A(H)$ the sum of the absolute values of its two-by-two minors. Put $R(H) = A(H)/N(H)^2$ and $Q(H) = -\log R(H)$ for invertible H. The bound on each row wedge by the product of the row masses gives $Q(H) \geq \log 2$. For an invertible matrix C write $K(C) = \max(1, \|C\|, \|C^{-1}\|)$. Comparison with the operator norms of H and $\wedge^2 H$ shows that a right input changes Q by at most $C_r + 4\log K(C)$ in absolute value.

Adding a nonnegative multiple of the sum of the inactive rows to the active row does not decrease A. To prove this, restrict to a pair of columns. If the active row becomes $a + qb$, where $b = \sum_j b_j$ is the sum of the other rows, the contribution involving it is $\sum_j |\det(a + qb, b_j)|$. Order the directions of the b_j's in the positive quadrant. The slopes in q of the determinants switch sign at the direction of b and sum to zero. The determinants themselves switch sign at the direction of $a + qb$. Hence on at least one entire side of the direction of b, the slopes agree in sign with the determinants, contributing the sum of their absolute values to the derivative. This sum is as large

as the sum of the absolute values of the opposite slopes, which bounds any negative contribution on the other side. The derivative is thus nonnegative wherever it exists, and zero rows follow by continuity. Sum over the column pairs.

For probability rows p, q, the kernel

$$1 - \sum_{i<j} |p_i q_j - p_j q_i| = \sum_i p_i q_i + 2 \sum_{i<j} \min(p_i q_j, p_j q_i)$$

is positive semidefinite. The minimum term is the integral over $t \geq 0$ of the product of $p_j \mathbf{1}_{t \leq p_i/p_j}$ and $q_j \mathbf{1}_{t \leq q_i/q_j}$, with boundary values by continuity. Its unit Hilbert feature vectors realize the determinant distance as half their squared distance. For two independent draws of common law their expected squared distance is at most twice the expected squared distance from a fixed feature vector, by expansion about the mean. Weight the normalized inactive rows by their masses. If the active row has mass a and the others total mass b, their area A_0 and the area A_i involving the active row satisfy $A_i \geq (a/b)A_0$. For $H = B_i^q H'$ let u be the inactive mass fraction of H'. The wedge triangle inequality gives $A(H) \leq A(H') + 2qA_0$, and $N(H) = (1 + qu)N(H')$, so

$$\frac{R(H)}{R(H')} \leq \frac{1 + 2qu}{(1 + qu)^2} \leq 1.$$

If the run is followed by a different legal output, that row dominates the sum of all remaining rows, so $u \geq 1/2$. Each such completed run increases R by at least $9/8$. A run with mass ratio e^d increases $\log R$ by at least $2d - \log(2e^d - 1)$, which is at least $d/5$ for $d \geq \log(3/2)$ by its derivative and its value at the endpoint. The final run gives at least $d/5 - \log(3/2)/5$. Therefore, for any legal sequence D with $H = DH'$,

$$\log(N(H)/N(H')) \leq 5(Q(H) - Q(H')) + \log(3/2).$$

The column-sum vector of D belongs to $(J - I)\mathbb{R}^r_{\geq 0}$, invariant under the transposed generators, so all but at most one of its coordinates are at least half their maximum. The mass of H' outside any row is at least $R(H')N(H')$, by summing the wedge bound by row masses. Thus $N(H)$ is at least half the maximum column sum of D times $R(H')N(H')$. Since $\det D = 1$, the inverse norm is bounded by a constant times $\|D\|^{r-1}$. Consequently $\log K(D) \leq C_r(1 + Q(H))$. During an excursion from depth at most Q_0 with accumulated input cost S, move all the already read input factors to the right before stripping all the already performed outputs. Each intermediate output remains legal by entrywise nonnegativity of the later inputs. The last bound gives $\log K(D) \leq C_r(1+Q_0+S)$ also in the middle of a run, and the input distortion bound gives $Q \leq Q_0+C_r(1+S)$.

C.4. **Rank-one drift.** There is an absolute second singular value bound $\sigma_2(C_m) \leq Ce^{-cm}$ for an input of m runs outside an event of probability Ce^{-cm}, uniformly in the initial state. Here is a direct proof. Write $A_i = B_i^{\mathsf{T}}$. On the cone $(J - I)\mathbb{R}^r_{\geq 0}$, a vector scales as $v = \mathbf{1} - \xi$ with $\xi \in \Delta_r$. For $v^{\mathsf{T}} z = 0$, the map A_i^{T} replaces coordinate i of z by $\mathbf{1}^{\mathsf{T}} z = \xi^{\mathsf{T}} z$, a convex combination, and maps the annihilator of v into that of $A_i^{-1} v$. Start with v in the image of the cone under a word of A_i's. In two consecutive cycles each using every index once with exponent one, the future complete cycle ensures that every ξ-coordinate in the first cycle is uniformly positive, since a product of the corresponding B_i's is strictly positive. Every row gets updated, so the product of the stochastic updates has strictly positive entries uniformly in these choices. On the unit sup-norm sphere in the annihilator, its norm is strictly less than one, since the initial v is uniformly positive and a convex average with positive weights can have absolute value one only when all coordinates equal the same signed unit. Compactness gives a uniform contraction factor. All other steps are nonexpanding. Multiplying along the word and using min–max yields an absolute second singular value bound by a dimension constant times a fixed power less than one per disjoint good cycle. Transpose and reverse the finite word to get the bound for B_i's,

since reversed double cycles have the same property. At each block of $2r$ runs one such admissible witness with exponents one has conditional probability at least $p > 0$, uniformly. For its indicator X_j, $\mathbb{E}(e^{-X_j} \mid \text{past}) \le 1-p(1-e^{-1})$. Iterating and applying Markov's inequality gives exponentially small probability of fewer than $pq/2$ witnesses in q blocks, proving the claim.

Suppose $H/N(H)$ has small inactive mass fraction δ outside an active row a, and let b be the sum of the other rows. For a strictly positive input C, put $B = \|bC\|_1$, $W = \|aC \wedge bC\|_1$, and $\kappa = B/\min_j (bC)_j$, where the exterior ℓ^1-norm sums absolute coordinates. The maximal legal active run has length $q = \lfloor \min_j (aC)_j/(bC)_j \rfloor$. If j_0 attains this minimum and f is its fractional part, the new active entry at j is

$$\frac{(aC)_j (bC)_{j_0} - (aC)_{j_0}(bC)_j}{(bC)_{j_0}} + f(bC)_j.$$

The resulting matrix H' satisfies $N(H') \le \kappa W/B + 2B$ and $A(H') \ge W$, the latter by summing the active-row wedges before taking absolute values. Also $A(HC) \le (1+2q)A(H')$ by the wedge triangle inequality for the reverse operation, and $q \le N(HC)/B$. Thus $A(H') \ge BA(HC)/(3N(HC))$.

If $W \ge B^2$, combine $N(H') \le (\kappa+2)W/B$ with the bounds $B \ge m(C)\|b\|_1$, $W \le \| \wedge^2 C\|_1 N(H)\|b\|_1$, and $R(H) \le \|b\|_1/N(H)$, where $m(C) = \min_i \sum_j C_{ij}$. This gives

$$R(H')/R(H) \ge m(C)^2/((\kappa+2)^2 \| \wedge^2 C\|_1).$$

If $W < B^2$, the estimate $N(H') \le (\kappa+2)B$ gives

$$R(H')/R(HC) \ge \big(3(\kappa+2)^2 B/N(HC)\big)^{-1}.$$

Use buffers of $\lfloor \zeta m \rfloor$ runs at each end, for small fixed $\zeta > 0$. Their products L, L' are strictly positive except with exponentially small probability by the witness estimate, and have entry sums at most $\exp(K_0 \zeta m)$ except with exponentially small probability by the conditional cost moment, for large fixed K_0. In $C = LML'$, positivity of integer entries gives $m(C) \ge N(C)/N(L)$ and $\kappa \le N(L')$. Indeed each row of LML' has mass at least $N(ML')$ and each coordinate of $bLML'$ is at least $\|bLM\|_1$. Since $\| \wedge^2 C\|_1 \le C_r N(C)\sigma_2(C)$, the first case gives exponential gain when $4K_0\zeta < c$. In the second, $B/N(HC) \le N(C)\delta$ by $C \ge I$, and $R(HC)/R(H) \ge c_r K(C)^{-4}$. The gain is thus at least $c_r e^{-K_1 m}/\delta$ outside exponentially small probability. Further stripping only increases R. For gain $g > 1$ and failure tolerance $\epsilon > 0$, a fixed time $m = O(1 + \log g + \log(1/\epsilon))$ therefore gives the gain uniformly for sufficiently small δ, after choosing the threshold for δ at that time.

To extend the vertex estimate, first note that a gasket point has a unique address until it meets a first-cell intersection after stripping a common prefix. At such an intersection it has exactly two addresses, because the inverse maps in the two cells send it to vertices, each with a unique constant address. At any fixed finite depth a neighborhood of the point within the gasket is contained in the union of its address cylinders at this depth: the other cylinders form a finite compact union excluding it. Accordingly a point with infinitely many runs in its address has a unique arbitrarily long prefix, while a point with an eventually constant address has one or two prefixes sending it to a vertex.

A uniform sign test allows a proposed output prefix to be made legal. For a fixed matrix M, if $M\theta \ge 0$ at the future direction θ but MC_m has a negative entry, then a nonzero row has opposite signs on θ and one of the normalized columns of C_m. Both directions lie in the same cylinder, of diameter exponentially small in m. Its probability is therefore at most $C\rho_0^m$, for a fixed $0 < \rho_0 < 1$, by hyperplane regularity and a union over rows. Apply this with $M = B_w^{-1}H$ for a fixed proposed output w. If all columns are nonnegative after the full proposed strip, each earlier strip is legal by multiplying by its nonnegative suffix. Hyperplane regularity applied to each coordinate also places θ in a fixed positive compact subset of the simplex outside probability $O(\epsilon)$.

Fix $g > 1$ and $\epsilon > 0$. Use a first block of $m_1 = O(1 + \log(1/\epsilon))$ runs for the sign test, with $K(C_{m_1}) \le L_1$ outside probability $O(\epsilon)$ and $\log L_1 = O(1 + \log(1/\epsilon))$. Use a second block of $m_2 = O(1 + \log g + \log(1/\epsilon))$ runs for the vertex estimate with gain gL_1^4/c_r. Its condition bound L_2 can be chosen with logarithm of the same order and probability of failure $O(\epsilon)$. Let a normalized nonnegative rank-one map have image $u \in \mathcal{R}$. If u has infinitely many runs, take a fixed prefix with enough completed runs that their factor $9/8$ compensates the worst distortions of both input blocks and leaves gain g. By shrinking the matrix neighborhood, on survival and the positive compact core the output direction lies in that prefix cylinder. Indeed its image is uniformly close to u on the core and the other cylinders are separated from u. The sign test makes that prefix legal after the first block with the prescribed exception probability. Its completed runs provide the claimed gain, and further stripping helps. If u has an eventually constant address, use the one or two prefixes sending it to a vertex. One of these applies on survival in a sufficiently small matrix neighborhood. Use the sign test for both, then select a legal candidate from the first-block matrices alone. After stripping it and before any other output, the result is uniformly close, in relative row masses, to a vertex matrix. In fact the limiting matrix is sent to a vertex by this fixed prefix, and first-block multiplication has bounded norm and does not decrease the limiting mass because its matrix is entrywise at least I. Thus the neighborhood can be shrunk to meet the vertex threshold. The second block now has its original conditional law and supplies the gain. Extra outputs can be deferred until after the next right input, since their legality persists under right multiplication and the full stripping is unique. Crucially the two block lengths were fixed before shrinking the neighborhoods. Compactness gives a neighborhood $\mathcal{U}$ of the rank-one preserving stratum with a single time $M = m_1 + m_2$ and bound

$$\mathbb{P}(E \cap \{R(H_M) < gR(H_0)\}) \le C\epsilon \qquad \text{if } H_0/N(H_0) \in \mathcal{U}.$$

Use $g = 2$. When the current conditional probability of E is at least $1 - \epsilon$, the conditional gain failure probability in the original law is $O(\epsilon)$. The positive change in Q is bounded by $C_r + 4 \log K(C_M)$, which has conditional second moment $O(M^2)$. A variable dominating the depth change, equal to $-\log 2$ on gain and this bound on failure, has uniformly negative mean for small fixed ϵ, by Cauchy–Schwarz and $M = O(1 + \log(1/\epsilon))$. It also has uniform exponential moments at this fixed M. Taylor expansion and Hölder's inequality give $s, \vartheta > 0$ and $q_* < 1$ such that

$$\mathbb{E} \exp\{sS_M + \vartheta(Q(H_M) - Q(H_0))\} \le q_*, \tag{11}$$

where S_M is the total input cost in the block. The estimate holds in the original conditional law at each actual time in this neighborhood with survival probability at least $1 - \epsilon$.

C.5. **Rational planes and carrier transport.** A rational two-dimensional subspace P is positive if its intersection with the nonnegative cone spans it. Its height is the exterior ℓ^1-norm of a primitive integer generator of $\wedge^2 P$. If $B_i^{-1} P$ is positive, its height is at most that of P. Choose two independent nonnegative vectors in the inverse plane and form the matrix of these columns. The area monotonicity already proved applies to this rectangular matrix. Applying B_i maps its exterior product to one with the same scalar factor relative to the primitive vector of the target, because the exterior map is unimodular. Cancelling this scalar proves the height comparison. Thus any finite family of positive rational planes can be enlarged to a finite family closed forward under all inverse outputs that preserve positivity, by taking all positive planes up to the maximum of its heights.

Every directed cycle in this family acts trivially on the annihilator. To see this, let a positive word matrix D preserve P. On the active indices, those that occur in the word, its induced block is strictly positive if there are at least two active indices. Its second eigenvalue in absolute value is less than one. In fact the transpose is a word of A_i's in reverse order, and its Perron vector

belongs to the cone $(J-I)\mathbb{R}^l_{\geq 0}$ in the active dimension l, by cone invariance and positivity. The moving annihilator argument above gives strict contraction on that Perron vector's annihilator under one period: every stochastic replacement has all coefficients positive because its Perron continuation lies in the interior of the cone. For each cyclic continuation the Perron vector lies in the interior of that cone: by conjugation with $J-I$, it is the image of the strictly positive Perron vector for the corresponding complete cyclic product of the B_i. This includes $l=2$, where $J-I$ is the swap. The active block has determinant one. Each monic integral factor of its characteristic polynomial omitting the Perron root would have nonzero constant term of absolute value less than one, by this contraction. Hence the polynomial is irreducible over $\mathbb{Q}$. The full matrix has the same expanding Perron eigenvalue and a left Perron vector positive on every coordinate: on the inactive coordinates the full matrix has an upper coupling block with a positive entry in each column, since every active row update adds all other rows. Any nonzero nonnegative vector in P has nonzero Perron projection, so invariance puts the Perron eigenvector in P by iterating and scaling. Since P is rational and two-dimensional, the Perron eigenvalue has degree at most two. It follows that $l=2$ and P is the active plane, obtained as the rational invariant span supporting that irreducible quadratic. More explicitly the full quotient by the active plane is the identity, so the quadratic factor annihilates P and its value at one is nonzero, forcing the quotient image of P to vanish. If only index i is active, applying $D-I$ to a nonnegative vector of P with nonzero mass outside i gives a nonzero multiple of e_i. Such a vector exists because P is positive and two-dimensional. In both cases the annihilator action by right multiplication is the identity. Removing cycles from paths therefore gives, for each plane P in the finite family, a finite set $\mathcal{N}_P$ spanning its annihilator, closed under the forward rule $q \mapsto qB_i$ to the inverse-image plane. Start with bases and transport along all simple paths to obtain these sets.

For a plane in this family define

$$d_P(H) = \max_{q\in\mathcal{N}_P} \|qH\|, \qquad f_P(H) = \log \frac{\sqrt{A(H)}}{d_P(H)}.$$

The denominator is comparable to the Frobenius norm of the projections of the columns normal to P, uniformly over this finite family. If two distinct planes P, T are in the family, then

$$\sigma_2(H) \leq C(d_P(H) + d_T(H)). \tag{12}$$

Indeed, the map to the pair of normal projections is injective on the orthogonal complement of $P \cap T$, giving a uniform distance bound to their intersection. Projecting the matrix into that intersection gives a rank-at-most-one approximation. For the finitely many possible first failed planes $P' = B_i^{-1}P$ which are not positive, the cone $P' \cap \mathbb{R}^r_{\geq 0}$ has dimension at most one. Hoffman's error bound [25] applied to the system of nonnegativity inequalities and normal equalities of P' bounds the distance of a nonnegative column to this cone by a constant times its distance to P'. Thus $\sigma_2(H)$ is bounded by a constant times the projected Frobenius norm for the failed plane as well.

Write T_{in} for the one-run input-only expectation operator, acting on both the matrix and the current input state. For any fixed P, choose a maximizing initial covector in $d_P(H)$. Submultiplicativity of the exterior norm and use of this one covector after a right input give, up to a fixed dimensional additive constant,

$$f_P(HC_m) - f_P(H) \leq \tfrac{1}{2}\log \| \wedge^2 C_m\| - \log \frac{\|qHC_m\|}{\|qH\|}.$$

The mean bounds (10) give $T_{\text{in}}^m f_P - f_P \leq -cm$ for one fixed large m and $c>0$, uniformly. Set

$$\psi_P = \frac{1}{m}\sum_{j=0}^{m-1} T_{\text{in}}^j f_P.$$

The absolute input distortion of f_P is at most $C_r + 2\log K(C)$, by operator and inverse norm bounds. Therefore $\psi_P - f_P$ is uniformly bounded and $T_{\mathrm{in}}\psi_P - \psi_P \le -c$ by telescoping. Its one-step change has a uniform exponential moment. A small positive exponential power gives constants $\delta > 0$, $\rho < 1$ for which $\exp(\delta\psi_P)$ contracts by factor ρ in one input expectation. If stripping D sends P to $T = D^{-1}P$ and all intervening planes are positive, the forward normal sets give $d_T(D^{-1}H) \ge d_P(H)$ and area decreases. The same comparison holds after any future right inputs, which do not change the legality of D or the future law when only the left matrix is changed. Hence $\psi_T(D^{-1}H) \le \psi_P(H)$. A tracked plane transported along outputs thus has a contracting nonnegative exponential supermartingale, killed at zero on first plane loss.

C.6. **Returns at bounded depth.** Choose Q_0 so large that every singular preserving matrix of entry sum one with area at most $2e^{-Q_0}$ belongs to $\mathcal{U}$. Here preservation may mean preservation of one allowed initial corner. The finite union of these preservation sets is compact: the condition that $MB_i\theta$ belongs projectively to the gasket is closed for every positive gasket point θ, since the image is nonzero for every nonnegative entry-sum-one M. The zero-area part is rank one with image in the gasket. Compactness proves the choice. On nonterminal survival all sufficiently late states at multiples of M runs consequently lie in the domain

$$p \ge 1 - \epsilon, \qquad Q \le Q_0 \quad \text{or} \quad H/N(H) \in \mathcal{U},$$

where the probability condition is required in both cases. Kill the process when it exits this domain or becomes terminal. Starting at $Q \le Q_0$, end an excursion at the next such sampled time with $Q \le Q_0$, or at killing. After its first block, (11) makes $\exp(sS + \vartheta Q)$ a supermartingale while the process is at high depth in the domain. The first block has a uniform cost moment and its depth increase is bounded by $C_r + 4\log K(C_M)$. Decrease s, ϑ by a common factor if needed, using Hölder's inequality in (11), to bound the first-block expectation uniformly. As $Q \ge \log 2$, Fatou's lemma gives a uniform exponential moment of excursion cost, in particular a finite excursion time. Starting at any finite depth gives a finite first entrance to bounded depth or killing by the same argument, with a bound depending on that depth.

Every preserving normalized limit with $Q \le Q_0$ has rank two and image interval of diameter bounded below in terms of Q_0, also after composition with one permitted B_i. Indeed its normalized second singular value is bounded below by a dimensional multiple of the area, and each of the finitely many right factors changes singular values by bounded factors. If the image diameters tended to zero, compactness would yield a rank-one limit. Each interval lies on a cylinder edge, so uniform cylinder shrinking bounds the raw depth of its cylinder. Its plane is spanned by two columns of that integral cylinder matrix and is positive. Thus all limiting planes lie in one finite base family, enlarged by primitive height as above.

At each bounded-depth return set $V = \max_P \exp(\delta\psi_P)$ over this enlarged family, and track one maximizing plane. During its excursion, the cumulative input and output matrices C, D satisfy $\log K(C) + \log K(D) \le C_0(1 + Q_0 + S)$ by the output budget. At an intermediate positive plane, a target normal q' pulls back to $q'D^{-1}$, a source normal of size at most $K(D)\|q'\|$. Its coefficients in a fixed source basis are bounded by a constant times this size. Since the matrix has become $D^{-1}HC$, its denominator is at most a constant times the starting denominator times $K(D)K(C)$. This also holds at first plane loss using a fixed basis of the failed plane, because there are only finitely many first failed planes. The area is bounded below by its initial value divided by the exterior norm distortion of C, D. The tracked f is therefore at least its starting value minus $C_1(1 + Q_0 + S)$ even at first plane loss.

At first loss, Hoffman's bound gives denominator at least $c\sigma_2(H)$, hence $f \le Q/2 + C_2$, by $A(H) \asymp N(H)\sigma_2(H)$. Here $Q \le Q_0 + C_3(1 + S)$ by the input distortion bound. If the tracked plane stays positive but a different plane maximizes at the next return, (12) gives $f \le Q/2 + C_4$ for the tracked plane there: the competitor's f is at least the tracked f minus the bounded

corrector error, so both denominators cannot be small compared with $\sigma_2(H)$. Either failure thus requires excursion cost at least $c \log V - C_5$. The exponential cost moment bounds the conditional probability of such a bad reset by $C_6 V^{-\eta}$ for some $\eta > 0$. On good returns the actual height equals the tracked height, and optional stopping gives $\mathbb{E}(V_{\mathrm{next}} \mathbf{1}_{\mathrm{good}} \mid \mathrm{past}) \le \rho V$. Apply the strict one-step contraction before optional stopping because the return takes at least one input. At killing set height zero and mark the transition good.

These inequalities yield recurrence even with arbitrary reset amplitudes. For a large threshold $R_* > 0$, kill visits to $V \le R_*$ at zero and mark them good. Above R_*, normalize the good kernel by its mass. This conditional good chain contracts expected height by $\rho' \le \rho/(1 - C_6 R_*^{-\eta}) < 1$ and therefore reaches the threshold almost surely. In the band $[2^j R_*, 2^{j+1} R_*)$ its expected number of visits is at most $2/(1-\rho')$: wait for the first entry below the band's upper end and bound the expected total subsequent killed height by $2^{j+1} R_*/(1-\rho')$. Couple the original chain to this good chain until the first bad transition. The conditional bad-step probabilities have expected sum at most

$$\frac{2C_6}{1-\rho'} \sum_{j \ge 0} (2^j R_*)^{-\eta} < 1$$

for large R_*, uniformly over initial heights. Restart after each bad reset at its actual finite height. The probability of avoiding a threshold hit through arbitrarily many bad resets is zero by the resulting geometric bound. With finitely many resets the last good chain hits almost surely. Restart upon each exit from the bounded-height set. These arguments apply to the conditional laws here by carrying the full past in the state, and exclude a surviving return chain whose heights tend to infinity.

Theorem 13. *Let G be an invertible rational r-by-r matrix. For the future input direction θ under the ordered law, almost surely on the event that $G\theta$ lies with positive orientation in $\mathcal{R}$, there are an input prefix v, a finite output word u, a permutation matrix P, and $c > 0$ such that*

$$GB_v = cB_u P.$$

Proof. Start at a nonnegative reduced remainder. On nonterminal survival its size tends to infinity and it eventually stays in the sampled domain. From any deterministic sampled start, the entrance and return bounds show that a path staying in the domain has infinitely many bounded-depth returns. All their normalized limits have rank two with range in the finite family. Along a subsequence converging to such a limit the normalized area stays bounded below and the normal residual of that plane divided by $N(H)$ tends to zero. Hence the corresponding ψ_P tends to infinity. The maximized height tends to infinity along all bounded-depth returns, since otherwise a bounded-height subsequence would admit a normalized convergent further subsequence. This contradicts recurrence. The countable union over deterministic sampled starts covers eventual residence.

For arbitrary G, the sign test and summability show that almost surely on survival, GB_v is eventually nonnegative for finite input prefixes v. At each such prefix strip to a reduced remainder. Conditional on any fixed prefix the future law is the original law at that state. Applying the preceding argument at each of the countably many prefixes yields terminal scalar permutation remainder almost surely on survival. Multiplying by the stripped output words proves the formula. □

Appendix D. Collisions along a common suffix

Proposition 14. *For the density equation in Section 3, let $n \ge 4$. Fix two distinct labelled positive integral solutions reached by ascending words, allowing a permutation of their first $d = n-1$ slots. The number of common ascending words for which the resulting last coordinates agree and are at most B is $o((\log B)^\beta)$, with β the linear counting exponent.*

D.1. **Observability on a finite-index orbit.** Let F_i be the ascending move in slot $i \le d$, regarded as a polynomial automorphism of the ambient space over $\mathbb{Z}$. Its inverse reverses the swap and the Vieta operation. Both preserve positive integral solutions. Write H for the group generated by these maps, acting diagonally on pairs. In a run of F_i, put $P = \prod_{j \notin \{i,n\}} x_j$ and

$$a = n(k+1)P - k \sum_{j \notin \{i,n\}} P/x_j.$$

Each quotient is the product omitting the given frozen slot. On the two changing coordinates $(u, v) = (x_i, x_n)$, the move is $(u, v) \mapsto (v, av - u - kP)$. The frozen coordinates are positive, so $a \ge (n + 2k)P > 2$ and $0 \le kP/(a-2) < 1/2$. With $\lambda > 1$ the larger root of $t^2 - at + 1$, the last coordinate after $t \in \mathbb{Z}$ moves is

$$kP/(a-2) + c_+\lambda^t + c_-\lambda^{-t}.$$

Both coefficients are positive. A negative coefficient is incompatible with positivity at the growing end, and a vanishing coefficient gives a limit below $1/2$ at that end, incompatible with positive integrality. If two sequences agree at all nonnegative multiples of a positive integer m, consecutive ratios identify λ^m, hence λ, and linear independence of the constant and two exponential sequences identifies the remaining coefficients. The values at times 0 and -1 give equality in slots n and i. Every finite-index subgroup of H contains a positive power of every F_i. Consequently, on the orbit of a distinct pair under such a subgroup, the difference of the last coordinates cannot vanish identically.

D.2. **Congruence collision probabilities.** Let H_9 be the kernel of the action of H on $(\mathbb{Z}/9\mathbb{Z})^n$. At every residue representative b and for $g \in H_9$, the evaluations at b and at $b + 3e_j$ for each standard basis vector e_j of $\mathbb{Z}^n$ show that the derivative of g at b is the identity modulo 3. Thus in coordinates $x = b + 3z$ on the corresponding 3-adic ball, $(g(b + 3z) - b)/3$ is the identity modulo 3 coefficientwise. Indeed the constant displacement is divisible by 9, and the terms of degree at least two retain a factor 3 after division. Poonen's interpolation theorem [24, Theorem 1] therefore gives an analytic family of iterates with time in $\mathbb{Z}_3$, since the coefficient valuation is at least $1 > 1/(3-1)$. This applies on each ball, also for the diagonal action.

On the closure of an H_9-orbit of a pair, take the probability measure obtained by inverse limit of the uniform measures on its finite congruence orbits. It has full support. Each finite orbit is transitive, and the conditional expectations of an invariant function on these finite levels are constant. Their convergence to the function proves ergodicity. The analytic time maps preserve this measure by approximation with nonnegative integer powers, using joint continuity on a compact polydisk. For each pair, its last-coordinate difference along an analytic flow is identically zero or has finitely many zeros by Strassmann's theorem. If E is the set where the difference vanishes, the integral over $t \in \mathbb{Z}_3$ of the absolute difference of its indicators at times t and $t+1$ is zero. By Fubini and invariance under the time maps, the indicator is invariant almost everywhere under g. This holds for each $g \in H_9$. Ergodicity makes the measure of E zero or one. Measure one, together with closedness and full support, contradicts observability on the orbit of a distinct pair.

For $q = 3^a$ with $a \ge 1$, let Ω_q be the diagonal H-orbit of the two seeds modulo q. Its inverse-limit measure is the uniform average of the translates of the H_9-orbit measure by coset representatives. The conclusion on every translate and continuity from above imply

$$\lim_{a \to \infty} \frac{\#\{(x, y) \in \Omega_{3^a} : x_n = y_n\}}{\#\Omega_{3^a}} = 0,$$

where equality in the numerator is in $\mathbb{Z}/3^a\mathbb{Z}$.

D.3. **Renewal with a finite permutation action.** Use the accelerated linear system of Gamburd, Magee, and Ronan [15, Sections 4–5]. In its ordered coordinates the base consists of

$$0 \le w_1 \le \cdots \le w_d, \qquad \sum_{i=1}^{d} w_i = 1.$$

A branch indexed by (a, j) with $a \ge 1$, $1 \le j < d$, deletes coordinate w_j, appends $1 + (a - 1)(1 - w_j)$, and divides each of the resulting d entries by $1 + a(1 - w_j)$. Denote this map by $f_{a,j}$ and its roof by $\tau_{a,j}(w) = \log(1 + a(1 - w_j))$. A word's roof is the sum of consecutive roofs evaluated at the consecutive states. The scalar transfer operator is the sum over branches of $e^{-s\tau_{a,j}}$ times composition by $f_{a,j}$. We use its positive C^1 eigenfunction h_β, eigenvalue 1 at $s = \beta > 1$, and the uniform exponential contraction of branch compositions from [15, Sections 4.2–4.4 and Proposition 45].

For fixed q, the finite state consists of a permutation of the d labels and a point in Ω_q. A branch (a, j) moves the j-th label to the last position and applies F_i^a to the residue pair for that label i. This operation is a permutation: the last label identifies the moved slot and the residue operation has an inverse. Let m be a common multiple of all residue orders of F_i. Runs of length m change only the ordering, and these move-to-last permutations generate the ordering group. One can move any label to a nonlast slot, apply a positive power realizing any desired residue power, and restore the order. It follows that the branches generate a transitive group on the joint finite state.

Consider the operator with the same base weights and branch maps on this finite state. Divide by h_β so that the weights at $s = \beta$ sum to one. At $s = \beta + it$, $t \in \mathbb{R}$, the weights differ by phases. In the Euclidean norm on the finite-state vector the branch permutations are isometries. The normalized operator is power bounded in the supremum norm. Each differentiated iterate is bounded by a constant times the supremum norm plus $O(\rho^m)$ times the derivative norm after m steps, for some $\rho < 1$. To see this, the derivative of the argument is bounded by the uniform contraction estimate, and the derivative of the total roof is bounded by the sum of a geometric series: each one-branch roof has bounded derivative because $w_j \le 1/2$. The bounded derivative of the logarithm of h_β gives the same bound for the normalized weights. Compact inclusion of C^1 in C^0 gives essential spectral radius less than one as in [15, Lemmas 38 and 46]. The defining series is locally holomorphic for $\Re(s) > 1$; its C^1 bound follows by summing $O((1 + a/2)^{-\Re(s)})$ with the logarithmic factors from parameter derivatives.

A peripheral eigenfunction at $\beta + it$ has maximal Euclidean norm on the common attractor of the branches. More precisely, equality in the triangle inequality at a point of maximum norm implies equality at each of its word images. The closure of these images contains the limit of every growing sequence of nested branch cylinders, because the diameter of a depth-m cylinder is $O(\rho^m)$. In particular equality holds at every branch fixed point and along each of its images. If the eigenvalue is $e^{i\theta}$, at the fixed point of branch $(a, d - 1)$ it follows that

$$M\left(\theta + t \log \frac{a + \sqrt{a^2 + 4}}{2}\right) \in 2\pi\mathbb{Z},$$

where M is a common multiple of the finite-state permutation orders. In fact the earlier coordinates of this fixed point are zero, because they are unchanged before division by a factor greater than one. The unnormalized update of the last two is $(x, y) \mapsto (y, x + ay)$, giving the indicated roof. For $t \ne 0$, the consecutive differences of the displayed logarithms are nonzero and tend to zero. This excludes peripheral eigenvalues. At $t = 0$ and eigenvalue one, apply the equality relation to a common branch word from two attractor points. The same permutation occurs at both points, so contraction and continuity imply equality of the eigenfunction values. The resulting constant vector is invariant under the transitive finite action. Iteration extends this constant to

all base points, since trajectories approach the attractor uniformly. Power boundedness excludes a Jordan block. The spectral projection at one is therefore the scalar projection on the uniform fiber average, because fiber averaging commutes with the scalar transfer.

For threshold u, the number of words with roof at most u and terminal state in a subset E has Laplace transform obtained by applying $s^{-1}(I - L_s)^{-1}$ to its indicator, where L_s is the unnormalized finite-state operator. The argument above gives a simple pole at β with residue equal to the scalar residue times the uniform fraction of E, and local holomorphy at every other point of the critical line. The renewal argument of [15, Section 4.3] consequently gives, at the fixed initial barycenter and for any initial ordering, an asymptotic of the form

$$c_\beta h_\beta(\mathbf{1}/d)\frac{\#E}{d!\,\#\Omega_q}e^{\beta u}.$$

Here the scalar coefficient is independent of q, and the error is taken with q fixed. The conclusion for an empty E is the identically zero count.

D.4. **Passage to actual heights.**

Proof of Proposition 14. Lemma 6 supplies an upper bound $O(\log B)$ on $\mathbf{1}^{\mathsf{T}}C_A(s)\mathbf{1}$ for a continuation s from either of the two fixed seeds, where $C_A(s)$ is the chronological product defined in Section 3, with final maximum at most B. Indeed the full label height of the associated word from the base seed is at least this linear height since the prefix matrix is entrywise at least the identity. A permutation of the parent slots preserves the scalar linear height. At the barycenter the exponential of the total linear roof along s is exactly $\mathbf{1}^{\mathsf{T}}C_A(s)\mathbf{1}/d$, by telescoping the normalizations of the successive transposed matrices. The finitely many initial orderings cover all run words. Apply the finite-state count at a fixed modulus q, with E the pairs of equal last coordinates and all orderings. An exact collision is counted by this congruence condition. The upper limit of the number of exact collisions divided by $(\log B)^\beta$ is bounded by a constant independent of q times the collision fraction in Ω_q. Letting $q = 3^a$ tend to infinity proves Proposition 14. □

AI contribution

AI agents assisted in developing and checking the arguments in the research materials. OpenAI Codex and its subagents assisted in reconstructing the proofs, auditing their scope, and preparing the manuscript.

References

[1] A. Baragar, *The Markoff equation and equations of Hurwitz*, Ph.D. thesis, Brown University, 1991.
[2] A. Baragar, *Asymptotic growth of Markoff–Hurwitz numbers*, Compos. Math. 94 (1994), 1–18.
[3] A. Baragar, *The exponent for the Markoff–Hurwitz equations*, Pacific J. Math. 182 (1998), 1–21.
[4] J. Bourgain, A. Gamburd, and P. Sarnak, *Markoff triples and strong approximation*, C. R. Math. Acad. Sci. Paris 354 (2016), 131–135.
[5] W. Y. Chen, *Almost all Markoff numbers are unique*, Hexagon preprint 2610.00281v2 (2026).
[6] S. Fomin and A. Zelevinsky, *The Laurent phenomenon*, Adv. Appl. Math. 28 (2002), 119–144.
[7] G. Frobenius, *Über die Markoffschen Zahlen*, Sitzungsber. Preuss. Akad. Wiss. (1913), 458–487.
[8] A. Hurwitz, *Über eine Aufgabe der unbestimmten Analysis*, Arch. Math. Phys. (3) 11 (1907), 185–196.
[9] A. Markoff, *Sur les formes quadratiques binaires indéfinies*, Math. Ann. 15 (1879), 381–406, and 17 (1880), 379–399.
[10] Q. Tang, *Density-one uniqueness for Markoff–Hurwitz solutions*, Hexagon preprint 2611.00490v1 (2026).
[11] I. V. Vyugin, *On the structure of solutions graph of the Markoff–Hurwitz equation and recurrence sequences*, arXiv:2509.19289v1.
[12] D. Zagier, *On the number of Markoff numbers below a given bound*, Math. Comp. 39 (1982), 709–723.
[13] W. Y. Chen, *Nonabelian level structures, Nielsen equivalence, and Markoff triples*, Ann. of Math. (2) 199 (2024), 301–443.

[14] P. Corvaja and U. Zannier, *Greatest common divisors of $u-1$, $v-1$ in positive characteristic and rational points on curves over finite fields*, J. Eur. Math. Soc. 15 (2013), 1927–1942.
[15] A. Gamburd, M. Magee, and R. Ronan, *An asymptotic formula for integer points on Markoff–Hurwitz varieties*, Ann. of Math. (2) 190 (2019), 751–809.
[16] F. Guéritaud, *Formal Markoff maps are positive*, arXiv:math/0605735.
[17] D. E. Martin, *A new proof of Chen's theorem for Markoff graphs*, Invent. Math. 241 (2025), 623–626, doi:10.1007/s00222-025-01346-9.
[18] Q. Tang, *Almost all k-Markov numbers are unique*, Hexagon preprint 2611.00216v1 (2026).
[19] Y. Benoist, *Arithmeticity of discrete subgroups*, lecture notes, author's version.
[20] B. Bekka, P. de la Harpe, and A. Valette, *Kazhdan's Property (T)*, New Mathematical Monographs 11, Cambridge University Press, 2008.
[21] E. Breuillard, *A non concentration estimate for random matrix products*, author's version.
[22] F. Comets, R. Fernández, and P. A. Ferrari, Processes with long memory: regenerative construction and perfect simulation, *Ann. Appl. Probab.* 12 (2002), 921–943.
[23] J. Li, *Fourier decay, renewal theorem and spectral gaps for random walks on split semisimple Lie groups*, Ann. Sci. Éc. Norm. Supér. (4) 55 (2022), 1613–1686; arXiv:1811.06484v2.
[24] B. Poonen, p-adic interpolation of iterates, *Bull. Lond. Math. Soc.* 46 (2014), 525–527.
[25] A. J. Hoffman, On approximate solutions of systems of linear inequalities, *J. Research Nat. Bur. Standards* 49 (1952), 263–265.

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