HOW MANY DYCK WORDS DO THE DIVISORS OF AN INTEGER MAKE? II. GENERIC RIGIDITY AND PERSISTENT AMBIGUITY

CARLES MARÍN
Independent researcher

Abstract

For integers $a, b \geq 1$ let $W(a,b)$ be the number of distinct Dyck words of Rodríguez Caballero attached to the integers $2^{e-1} r^a s^b$ ($r \neq s$ odd primes), and $R(a,b)$ the number of cells of the line arrangement that produces them. We study the excess $E(a,b) = R(a,b) - W(a,b)$, whose growth for $a \geq 2$ was left open in [I] (doi:10.5281/zenodo.24541288).

General theory. A semilinear dimension criterion characterises $E(a,b) = O(b)$; a finite calculation with affine observation vectors gives a sufficient test for each fixed a. We derive $R(a,b) = C_a b(b+1) + (a+1)b + \pi_a(b)$, where $C_a = \sum_{q \leq 2a} (2a+1-q)\varphi(q)/q$ and π_a has period dividing the product of the odd primes $\leq 2a$. Crossing a wall $\Lambda + h\theta = \rho$ raises the sum of the word-encoding multiset by $(a+1-|h|)(b+1-|\rho|) > 0$. Thus cells with one word form an antichain of at most $4ab+1$ cells, and a number of power sums depending only on a determines the word.

Rigidity. Exact affine certificates for $a = 2, 3, 4, 5$, combined with a written reduction, yield $E(a,b) = \Theta_a(b)$ and $W(a,b) = C_a b^2 + O_a(b)$, including $W(5,b) \sim \frac{7657}{210} b^2$. The reduction and program have not been checked by a human referee. Ambiguous cells form a fraction $\Theta_a(1/b)$ of all cells. For every $a \geq 2$, the mean ambiguous proportion of vertical sections over $I \supseteq [\frac{2a-1}{2a(a+1)}, \frac{1}{a}]$ has $\liminf_b \geq 3/(4a(a+1)|I|) > 0$; the limit is open. A written theorem for every $a \geq 2$ characterises equal signatures of pieces with distinct vector formulas in the all-positive chamber by constant differences of window sums and equal multisets of level constants. Their parameter shift is a non-integral multiple of $1/(a+1)$, explained by the smallest element of the multiset, so that chamber yields no two-parameter family of coincidences. Non-integrality and the four-element fibre for $a = 3$ are among the claims checked in Isabelle/HOL.

The case $a = 3$. A further computer-assisted step (Lemma D) confines coincidences of cells with at least two free coordinates to a bounded neighbourhood of 25 lines when $b \geq 317$. Nine explicit one-parameter families give proved coincidences. Together with a periodic remainder fitted to data, they suggest a formula for $E(3,b)$ that agrees with exact enumeration for $5 \leq b \leq 170$ and selected $b \leq 400$; completeness remains conjectural. We also report exact intersections of distinct two-prime shapes and sampled ambiguity among integers.

1. Introduction

For an integer $n \geq 1$, Rodríguez Caballero's word $\langle\langle n \rangle\rangle_2$ records how the divisors of n interleave with the doubled divisors: one reads $D_n \triangle 2D_n$ in increasing order and writes an opening bracket for elements of D_n and a closing bracket for elements of $2D_n$ [6, 7]. The language and the questions are Rodríguez Caballero's: he asked which Dyck words arise in this way and whether membership is decidable [4]. The reduction to the odd part used below is implicit in his 2020 construction from the divisors of $2n$ [4]. In [I] the question was reduced, for the integers whose odd part has two prime factors $r^a s^b$, to the cells of an arrangement of

Date: Preprint, 27 November 2026. Companion blueprint.

lines, and studied through

$$W(a,b) = \#\{\langle\langle n\rangle\rangle_2 : \ n = 2^{e-1} r^a s^b, \ r \neq s \text{ odd primes}, \ e \geq 1\}, \qquad R(a,b) = \#\{\text{cells}\}.$$

We recall what [I] proves. $R(a,b)$ has an exact formula and is a quasipolynomial of degree two with leading coefficient

$$C_a = \sum_{q=1}^{2a} (2a+1-q)\frac{\varphi(q)}{q}, \qquad C_1, \ldots, C_5 = \tfrac{5}{2}, \ \tfrac{22}{3}, \ \tfrac{73}{5}, \ \tfrac{1709}{70}, \ \tfrac{7657}{210}$$

[I, Theorem 5.1 and Corollary 5.2]. The count itself is classical: by [I, Proposition 5.3] it is an affine rescaling of Alekseyev's exact count of two-dimensional threshold functions on a grid [1] (see also Acketa and Žunić [2] and Žunić [3]). By a theorem of Woods on Presburger counting functions [8], $W(a,b)$ is eventually a quasipolynomial of degree exactly two [I, Theorem 6.1 and Corollary 6.4]. For $a = 1$ and $a = 2$ it is known in closed form: $W(1,b) \sim \frac{9}{4}b^2$ [I, Theorem 8.12] and $W(2,b) = (44b^2 + 19b + e_{b \bmod 6})/6$ [I, Theorem 9.4, computer-assisted]. So the excess $E(a,b) = R(a,b) - W(a,b)$, the number of cells lost to coincidences of words, is quadratic for $a = 1$ ($E \sim b^2/4$) and linear for $a = 2$ ($E \sim 43b/6$). Question 11.1 of [I] asks whether $E(a,b) = \Theta_a(b)$ for every $a \geq 2$, which would give $W(a,b) \sim C_a b^2$, and where the cells that share their word lie; [I, Question 11.3] records a fitted quasipolynomial for $W(3,b)$ with period 60. This paper is the companion announced in [I, §11].

Results. The paper has three levels, which differ in their status; Table 1 lists each result with its status.

Level 1: general theory, valid for every a.

(1) *A dimension criterion* (Proposition 4.1). The set B_a of pairs (b,c) such that the cell vector c shares its word with a lexicographically smaller cell is semilinear, and $E(a,b) = O(b)$ if and only if its linear components have rank at most two. In particular rigidity is decidable for each fixed a (Remark 4.3).

(2) *A finite reduction* (Theorem 5.1). For each fixed a, $E(a,b) = O(b)$ follows from two conditions on finitely many affine objects, one per *piece* of the parameter space. The conditions can be decided exactly.

(3) *The linear term of $R(a,b)$* (Proposition 2.3), an elementary consequence of Alekseyev's exact formula through [I, Proposition 5.3]. $R(a,b) = C_a b^2 + D_a b + \pi_a(b)$ with

$$D_a = C_a + a + 1, \qquad \text{i.e.} \qquad R(a,b) = C_a\, b(b+1) + (a+1)\, b + \pi_a(b),$$

$D_1, \ldots, D_5 = \frac{9}{2}, \frac{31}{3}, \frac{93}{5}, \frac{2059}{70}, \frac{8917}{210}$, where π_a is periodic from $b = 1$ with period dividing the product P_a of the odd primes $\leq 2a$; for $a \leq 6$ the minimal period is exactly P_a (computation). This is an explicit fixed-width expansion of the classical threshold-function count of Acketa–Žunić and Alekseyev, a routine consequence of their formula (Remark 2.4).

(4) *Wall weights and fibres* (Section 3). Let $\Sigma(c)$ be the sum of the multiset $Z(c)$. Crossing the wall $\Lambda + h\theta = \rho$ upwards (towards $\Lambda + h\theta > \rho$) raises Σ by exactly

$$\omega(h,\rho) = (a+1-|h|)(b+1-|\rho|) > 0.$$

Consequently adjacent cells have different words, every fibre (set of vectors with one word) is an antichain for the coordinatewise order, and a coincidence forces the weights of the walls crossed in the two directions to balance. With a decomposition into vertical strips this bounds every fibre by $1 + \sum_{q \leq 2a} \Phi_q(2b) \leq 4ab + 1$. Further: $E \leq \mathrm{Amb} \leq 2E$ for the number $\mathrm{Amb}(a,b)$ of cells whose word is shared; a number of power sums of $Z(c)$

depending on a only, polynomial in a ($O(a^4)$), determines the word (in the data, three suffice in most cases); and for $a = b$ the exchange of the two primes gives $W(a,a) \le (R(a,a) + 2a + 1)/2$.

Level 2: computer-assisted rigidity for $a \le 5$.

(5) *The certificate* (Theorem 6.1, computer-assisted). For $a = 2, 3, 4, 5$ the conditions hold according to an exact computation (the program and the reduction have not been checked by a human referee; its output was re-checked by independent scripts, Appendix A). Hence, for b large, $E(a,b) = \Theta(b)$ and

$$W(a,b) = C_a b^2 + O(b) \qquad (a = 2, 3, 4, 5).$$

For $a = 1$ the same computation detects the quadratic family of [I] (the exchange of progressions), which is a positive control. Figure 4 shows both behaviours in the exact data. The ambiguous cells are then a proportion $\Theta_a(1/b)$ of the cells (Corollary 6.2), but not of the parameter space: for every $a \ge 2$ two numerations of $0, \dots, K-1$ give thin cells of positive total section length near $\theta = \frac{1}{a+1}$, and the mean ambiguous proportion of the vertical sections over an interval of slopes has a positive $\liminf$ (Proposition 6.3 and Corollary 6.4, written proofs).

(6) *The denominator $a+1$* (Proposition 7.1; notation of Section 5: $c = g(n, m, \varepsilon)$ with $n = \lfloor \Lambda \rfloor$, $m = \lfloor \theta \rfloor$). For $a = 2, \dots, 5$ every affine candidate for a two-parameter family, and for $a = 1$ the one candidate that is not integer-compatible, lies in the chamber where all coordinates are free and positive, keeps $\lfloor \theta \rfloor$ and shifts $\lfloor \Lambda \rfloor$ by $-k/(a+1)$, $0 < k < a+1$ (109 pairs; the eight integer-compatible pairs of $a = 1$ are the exchange, and lie elsewhere). The smallest element of the multiset $Z(c)$ that encodes the word, $\tau(c) = (a+1)(n+1) - \frac{a(a+1)}{2} m + \sum_{h \le 0} \varepsilon_h$, explains the shift exactly in these 109 candidates: it is $(S_-(\varepsilon) - S_-(\varepsilon'))/(a+1)$, and the obstruction to integrality is the congruence $S_-(\varepsilon) \not\equiv S_-(\varepsilon') \pmod{a+1}$.

(7) *The all-positive chamber, for every $a \ge 2$* (Theorem 7.6, written proof). In C_+, $\mathcal{Q}(z) \sum_{v \in Z(c)} z^{T-v}$ splits into clusters, one per level $r \le a$, placed at $T_r m$ with $T_r = r(r+1)/2$ (Proposition 7.4). From this:

- every piece of C_+ has an observation vector of rank two (Lemma 7.5), so (C1) holds there;
- two pieces of C_+ with different vector formulas have the same signature if and only if their window sums differ by one constant and their multisets of level constants agree;
- the map between them keeps m and shifts n by $(S_-(\varepsilon) - S_-(\varepsilon'))/(a+1)$;
- this shift is never an integer (Theorem 7.10): if the window sums of two distinct templates differ by a constant D, then $a+1 \nmid D$. The proof uses only that templates are floors of affine functions.

So, for every $a \ge 2$, no pair of pieces inside C_+ violates (C2) (Corollary 7.11). For $a \le 6$ this reproduces exactly the candidates of the finite computations. The computation also shows, for $a \le 6$, that no candidate lies outside C_+ (Proposition 7.8; no regression for $a = 6$). Whether every candidate lies in C_+ for every a is Conjecture 7.12.

(8) *Lemma D for $a = 3$* (Theorem 8.1, computer-assisted). There is a constant C such that, for $b \ge 317$, a realisable vector with at least two free coordinates whose cell reaches distance $> C$ from a fixed set $\mathcal{L}_{\deg}$ of 25 lines shares its word with no other realisable

result	status
Level 1	
Proposition 2.3 (linear term D_a of $R(a,b)$; period divides P_a)	written proof, a consequence of the published count (checked exactly, $a \le 6$, every class mod $\mathrm{lcm}(1,\dots,2a)$); minimal period $= P_a$: exact enumeration $a \le 6$
Proposition 3.1 (wall weights)	Isabelle/HOL (a), (c) and $\omega \ge 1$; the crossing argument (b): written proof
Corollary 3.2 (antichains)	Isabelle/HOL (ii), (iii); (i): written proof
Proposition 3.5 (fibres $\le 4ab+1$)	written proof (finite checks, $a \le 3$, $b \le 15$)
Lemma 3.6 ($E \le \mathrm{Amb} \le 2E$)	Isabelle/HOL (the counting lemma)
Theorem 3.7, Proposition 3.8 (finitely many power sums)	written proof; the minimal number: exact enumeration $a \le 3$, $b \le 20$
Proposition 3.9 ($W(a,a) \le (R(a,a)+2a+1)/2$)	Isabelle/HOL the counting step ($2W \le R+f$); the geometric step: written proof
Proposition 3.11 (one pair of $a=2$, every $b \ge 2$)	written proof (checked $b \le 300$)
Proposition 4.1 (dimension criterion)	written proof
Theorem 5.1 (finite reduction)	written proof; the encoding tested on exact data, $a \le 3$, $b \le 200$ (Appendix A)
Level 2	
Theorem 6.1: $E(a,b) = O(b)$, $W(a,b) = C_a b^2 + O(b)$ for $a = 2,3,4,5$	computer-assisted, conditional on the correctness of the program and of Theorem 5.1
Corollary 6.2 ($\mathrm{Amb}/R = \Theta_a(1/b)$ for $a = 2,\dots,5$)	computer-assisted: a written consequence of Theorem 6.1 and Lemma 3.6 (data in Table 3)
Proposition 6.3 (two numerations of one progression)	Isabelle/HOL (i), and (ii) as a line of the arrangement that separates the two points; (iii), (iv): written proof
Corollary 6.4 (ambiguity of positive section measure, every $a \ge 2$)	written proof (segments checked in exact arithmetic, $a \le 6$, $b \le 60$)
Corollary 6.5 (no coincidence between two deep cells, $a = 3,4,5$)	computer-assisted
Proposition 7.1 (the identity for τ)	Isabelle/HOL; Corollary 7.2 (the 109 candidates): computer-assisted
Proposition 7.4 (levels of C_+, every a)	Isabelle/HOL
Lemma 7.5, Theorem 7.6 (all-positive chamber, every $a \ge 2$)	written proof (checked: Corollary 7.9, $a \le 6$; all templates, $a \le 15$)

Table 1. Status of the results, row by row (Levels 1 and 2; the categories are explained at the end of the table).

vector. The proof reduces the statement to a finite exact computation (about 3.6×10^6 linear systems, no alarm); C is not explicit.

Level 3: a conjectural classification for $a = 3$ (Section 9).

(9) We list the empty-graph fibre and nine one-parameter families: the one-prime pairs $O(s)$, the two reflections of [I], five further families $F_1, \dots, F_5$ (we know of no earlier list of coincidences of these words) that live next to the upper clipping, and a family S_1 at the exact point $\theta = m/5$. We prove by hand that each listed pair gives one word (Theorem 9.7). In the exact data every other coincidence is sporadic, in number constant in b. If that is so for all b (Conjecture 9.8; the constant g is fitted, not derived), then

$$E(3,b) = 7b + 2\left\lfloor \tfrac{b+1}{4} \right\rfloor + 2\left\lfloor \tfrac{b+2}{4} \right\rfloor + 2\left\lfloor \tfrac{b+1}{3} \right\rfloor + 3\left\lfloor \tfrac{b+1}{4} \right\rfloor + \left\lfloor \tfrac{b-3}{2} \right\rfloor + \left\lfloor \tfrac{b-1}{4} \right\rfloor + g_{b \bmod 4},$$

$(g_0, g_1, g_2, g_3) = (2, 4, 2, 17)$. This formula agrees with exact enumeration for $5 \le b \le 170$ and for $b \in \{180, 190, \dots, 290, 299, 300, 331, 400\}$, and with $R(3,b)$ it reproduces the fitted quasipolynomial of [I, Question 11.3] for $b \le 1000$; its period 12 and the period 15 of $R(3,b)$ account for the "12 plus 5" structure of that fit, and its slope $\frac{61}{6}$ together with $D_3 = \frac{93}{5}$ gives the fitted linear coefficient $\frac{93}{5} - \frac{61}{6} = \frac{253}{30}$ of $W(3,b)$ (a conjectural coefficient: it holds if the list is complete).

Literature status. We searched the literature result by result (arXiv, OEIS, the papers of Rodríguez Caballero, Alekseyev, Acketa–Žunić and Haukkanen–Merikoski, the Sturmian-word literature, Woods on Presburger counting functions); the log is in `NOVEDAD_26.md`. The searches found no earlier treatment of the questions of this paper, and a search that finds nothing is only an upper bound on novelty. In this paper:

- *known or standard:* the template count and its Sturmian interpretation (Section 5); the count $R(a,b)$ and its quasipolynomial structure (Alekseyev, Acketa–Žunić; Proposition 2.3 is an elementary consequence of their formula); the Vandermonde step of Theorem 3.7 and Proposition 3.8 (Hajós's lemma), whose power-sum bounds are problem-specific consequences of this known machinery; the polyhedral decomposition in Lemma 8.4 and Hoffman's bound in Theorem 5.1; the mechanism of Proposition 6.3 (i), (ii), which is the one-prime construction of [I] (two radix numerations of one progression, as in the refinement theorem [I, Theorem 3.1]), here for every $a \ge 2$;
- *elementary or routine, no novelty claimed:* Lemmas 2.1, 2.2, 3.6, 7.5, 8.3, Corollary 3.2, Propositions 3.1, 3.5, 3.9, 3.11, 7.1, 10.6, 10.10, Corollary 6.2 and Remark 3.3 (immediate consequences of results above), and Proposition 4.1 (a short consequence of the Ginsburg–Spanier theorem);
- *ours as far as we could see:* the finite reduction and its certificate, Theorem 8.1, the all-positive-chamber analysis (Propositions 7.4, 7.8, Theorems 7.6, 7.10), the quantitative section lengths and the persistence of ambiguity (Proposition 6.3 (iii), (iv), Corollary 6.4), and the families of Section 9.

Context. Three kinds of link with other work are separated here: *exact* (the same objects), *structural* (the same mechanism, no equivalence proved) and *only an analogy. Exact.* (i) The count of finite Sturmian words, $1 + \sum_{i \le n}(n+1-i)\varphi(i)$, is attributed to Lipatov (1982, in Russian), before the independent discovery of Berenstein, Kanal, Lavine and Olson (1987), Mignosi's independent proof of 1991 and the geometric proof of Berstel and Pocchiola, by Frid [18] and by Frid and Jamet [19], whom we follow for the history; we read both, not Lipatov's paper [17]. (ii) As in Rodríguez Caballero [7], the maximal height of the Dyck path

of $\langle\!\langle n \rangle\!\rangle_2$ is the generalised Hooley function $\Delta_2(n) = \max_u \#\{d \mid n : \ u < \log_2 d \le u+1\}$, the base-2 case of the function Δ_λ treated there (the classical Δ of Hooley [20] has base e): the height after reading up to t is $\#\{d \mid n : \ t/2 < d \le t\}$ (the opening brackets $d \le t$ minus the closing brackets $2d \le t$; the elements of $D_n \cap 2D_n$ cancel). It is not used below. By [I, Appendix A] the first half of these heights, after merging equal neighbours, is row n of A249223, the widths of the symmetric representation of $\sigma(n)$ [26]. Three further entries follow from this identification, and we claim none of them: the number of returns of the path to height 0 is A237271(n) (the number of parts of the symmetric representation of $\sigma(n)$), the maximal height is A250068(n) (the maximum width of a region), and the integers whose word differs from the words of all smaller integers are the terms of A342594 (first occurrences of a new width pattern, listed in A342592). We checked the first two for $n \le 10^4$ and the third on its 64 terms in the b-file. (iii) Counting digital straight lines on an $N \times N$ grid is counting linear dichotomies of the grid [21], the count that Acketa–Žunić and Alekseyev treat (we did not open [21]). *Structural.* In a cut-and-project set the patterns correspond to acceptance domains in the window, and one often replaces them by cut regions, the cells cut out by lattice translates of the hyperplanes bounding the window; Koivusalo and Walton [22] (we read the abstract and the introduction) show that the two partitions agree, up to a rescaling of the radius, under a *quasicanonical* condition and not in general; they report that an earlier proof of Julien [23] assumed too weak a condition (we did not open [23]). Our cells play the part of cut regions, the sets of parameters with a given word that of acceptance domains, and $E = R - W$ is the gap between the two counts. This is a parallel, not an instance: there the scheme is fixed and the patches are those of one pattern, here the slope θ and the shift Λ both vary over finite configurations, and [22] does not count the gap. Likewise Woods's theorem [8] gives the quasipolynomial structure used above, and McAllister and Woods [24] show (abstract read) that the least common multiple of the denominators is only an upper bound for the minimal period of a quasipolynomial, which is why the minimal period of $W(a,b)$ is a question (Section 10) and not an inference.

What is not proved. The completeness of the list for $a = 3$ (the direction ($\Rightarrow$) of a classification) is not proved for any b beyond the enumerations: Lemma D confines the coincidences to cells near $\mathcal{L}_{\deg}$ for $b \ge 317$, but the coincidences near $\mathcal{L}_{\deg}$ are not classified, the sporadic classes have no written proof, and between 171 and 316 only some values of b have been enumerated (Section 10). The constants in $O(b)$ and the constant C of Lemma D are not made explicit. Rigidity is not claimed for $a \ge 6$. For $a = 6$ the program's output satisfies the conditions of the reduction, but without the regression that checks its formulas. Rigidity for all a remains open: Theorems 7.6 and 7.10 settle the pairs of pieces inside the all-positive chamber, but not the other chambers, where (C1) and (C2) are only checked for $a \le 6$.

Formalisation. Fifteen numbered results are formalised in Isabelle/HOL, wholly or in part, with no `sorry` and with negative controls that fail as they should: Theorem 7.10 (the whole statement, templates defined as floors of affine functions), Proposition 7.4 (both parts, for every a and b) and the first sentence of Corollary 9.5 (the four vectors have the same multiset $Z(c)$; the equivalence of $Z(c)$ with the word is taken from [I]); and, in the session `Caballero_Setting`, for every a and b: Lemma 2.1 (also for several coordinates at once), Lemma 2.2, Proposition 3.1 (a), (c) and the weight $\omega \ge 1$ of (b), Corollary 3.2 (ii), (iii), Proposition 7.1 (the minimum of $Z(c)$, the formula for τ and the consequence $(a+1)(n'-n) = S_-(\varepsilon) - S_-(\varepsilon')$), and the counting parts of Lemma 3.6, Proposition 3.9 (the bound $2W \le R + f$ for an involution that preserves words and has f fixed cells; that

σ permutes the cells and fixes exactly $2a+1$ of them is the geometric part, taken as a hypothesis) and Theorem 9.7 (the bound $E \geq |\mathcal{U}| - |\pi_0(\mathcal{G}_b)|$ once each listed set has one word, which is the content of Proposition 9.1 (formalised, below), Theorem 9.2 (formalised, below), Theorem 9.4 (only its witnesses, below) and Corollary 9.5, not formalised except the first sentence of the last). In these statements *realisable* is replaced by what the proofs use (non-decreasing, coordinates in $[-b-1, b]$; the vector of a point has both properties, also checked), so they hold for more vectors. Also in `Caballero_Setting` (theory `Caballero_Twonum`): Proposition 6.3 (i) for $a \geq 2$ (the points $P_1(k)$ and $P_2(k)$ lie in the open quadrant off the lines and both have $Z = \{\{\min(K, N+k+1)\}\}$, through the two radix numerations of $0, \ldots, K-1$) and (ii) in the form of a line of the arrangement that separates the points (an integer $\rho \in [-b, b]$ between the values of $\Lambda - a\theta$ at the two points, so that the coordinates c_{-a} of their vectors differ), and Proposition 9.1 (the vectors $O_1(s)$ and $O_2(s)$ are those of the two points, $Z(O_1(s)) = Z(O_2(s)) = O(s)$, and they differ iff $s \leq 4b+1$ for $b \geq 2$). Theory `Caballero_Families` of the same session: Theorem 9.2 for the five families, for every b, i and ε in the stated ranges (both vectors are vectors of points of the open quadrant with no $\Lambda + h\theta$ an integer, by the affine witnesses of Table 6; they are distinct; and $Z(P) = Z(Q)$ as multisets, through an involution of the grid $\{0, \ldots, 3\} \times \{0, \ldots, b\}$ that swaps blocks of points between rows) and, of Theorem 9.4, the paragraph on witnesses (for $m \in \{2, 3\}$, $\rho \in \{0, 1\}$ and $n \geq 1$, c and G are realisable and differ in c_{-1}). Part (iii) of Proposition 6.3 (segments, lengths and the bound) is not formalised, nor is the identification of a cell with a connected region: separation by a line of the arrangement is what is proved. Not formalised: the words $Z(c)$ and $Z(G)$ of Theorem 9.4, their equality and the count of pairs, and the counts of pairs of Theorem 9.2, the geometry of walls (part (b) of Proposition 3.1, part (i) of Corollary 3.2), the framework of chambers, pieces and signatures (Theorems 5.1 and 7.6, Lemma 7.5) and the computer-assisted results, which are certificates (Section 6.6), not proofs in a proof assistant. The sessions (`Caballero_Nonint`, `Caballero_Levels`, `Caballero_Fibre4`, `Caballero_Setting`) are public at `https://github.com/karlesmarin/formalizations/tree/main/Isabelle/CaballeroDyckWords`.

Organisation. Level 1: Section 2 fixes notation, recalls two tools of [I] and proves the formula for D_a and the period; Section 3 gives the wall weights and their consequences for fibres; Section 4 proves the dimension criterion and Section 5 the finite reduction. Level 2: Section 6 describes the computation and its results for $a \leq 5$, Section 7 explains the denominator $a+1$, and Section 8 proves Lemma D for $a = 3$. Level 3: Section 9 treats $a = 3$ family by family. Section 10 lists the open questions, with two short results that answer parts of them (Propositions 10.6 and 10.10). Appendix A collects the finite checks, the independent re-computations and the data.

result	status
Proposition 7.8 (candidates only in C_+; one row forces $m' = m$)	computer-assisted ($a \leq 6$; $a = 6$ without regression)
Theorem 7.10 (the shift is never an integer)	Isabelle/HOL (the whole statement)
Corollary 7.11 ((C1), (C2) inside C_+)	written proof, from written and formalised results
Conjecture 7.12 (candidates only in C_+)	conjecture (checked $a \leq 6$)
Theorem 8.1 (Lemma D for $a = 3$, $b \geq 317$)	computer-assisted (Lemmas 8.3 and 8.4: written proof)
Level 3	
Proposition 9.1 ($O(s)$ for $a = 3$)	Isabelle/HOL
Theorem 9.2 (families $F_1, \ldots, F_5$)	Isabelle/HOL the identities and the witnesses; the counts: written proof
Theorem 9.4 (family S_1)	Isabelle/HOL the witnesses; the words, their equality and the count: written proof
Corollary 9.5 (a fibre with at least four cells)	Isabelle/HOL the equality of the four multisets; that the fibre has exactly four cells: observed
Theorem 9.7 (lower bound by the coincidence graph)	Isabelle/HOL the counting implication; its premises are the results above
Conjecture 9.8: the list is complete ($\Rightarrow$), formula for $E(3, b)$	conjecture exact enumeration $b \leq 170$, some $b \leq 400$
Section 10	
Proposition 10.6 (fibre counts of fixed size are quasipolynomials of degree ≤ 1)	written proof, conditional on rigidity (so computer-assisted for $a = 2, \ldots, 5$)
Proposition 10.10 (wall weights for k primes)	written proof (checked on the shape $(2, 2, 3)$)

Table 1. Status of the results (continued). *Written proof*: a mathematical argument given in the paper. *Isabelle/HOL*: the stated part is checked by the proof assistant (Section 1, Formalisation). *Computer-assisted*: a written reduction plus a finite exact computation in rational arithmetic that the result needs; no proof assistant was used for these, and the programs have not been checked by a human referee. *Exact enumeration*: exact data, with no conclusion for all b. *Conjecture*: not proved. The finite checks in parentheses, and the independent re-computations by an AI agent, are described in Appendix A. Row colours: written proof, written proof with a part checked in Isabelle/HOL, computer-assisted, conjecture (two shades alternate within a category).

2. Setting, two tools, and the linear term of $R(a,b)$

We use the notation of [I, Part II]. Fix $a,b \ge 1$. In units of $\log s$ put $\Lambda = \log t/\log s$ and $\theta = \log r/\log s$, where $t = 2^e$. For (Λ,θ) in the open quadrant, off the lines $\Lambda + h\theta = \rho$ ($|h| \le a$, $\rho \in \mathbb{Z} \cap [-b,b]$), put

$$c_h = \mathrm{clip}\lfloor \Lambda + h\theta \rfloor \quad (-a \le h \le a), \qquad \mathrm{clip}(v) = \min(b, \max(-b-1, v)),$$

$$f(v) = \min(b+1, \max(0, v+1)), \qquad H_i(j) = \sum_{l=0}^{a} f(c_{i-l} + j) \quad (0 \le i \le a,\ 0 \le j \le b),$$

and let $Z(c) = \sum_{i,j} \delta_{H_i(j)}$ be the multiset of the $K = (a+1)(b+1)$ values $H_i(j)$, written as a formal sum (δ_v is the multiset $\{v\}$). By [I, Section 6, Steps 1 and 2], the word of n is determined by $Z(c)$ and conversely; a vector c is *realisable* if it comes from a point of the open quadrant off the lines, and every realisable vector is realised by primes [I, Lemma 2.3]. So $R(a,b)$ is the number of realisable vectors, $W(a,b)$ the number of distinct multisets $Z(c)$, and two realisable vectors are *equivalent* if $Z(c) = Z(d)$. A *witness* for c is a point (Λ,θ) with $\Lambda,\theta > 0$ and $\Lambda + h\theta \notin \mathbb{Z}$ for $|h| \le a$ whose vector is c; it proves that c is realisable. A coordinate c_h is *free* if $-b \le c_h \le b-1$, and *clipped* otherwise. We write $T = K = (a+1)(b+1)$ for the largest value.

Each term $f(c_h + j)$, $0 \le j \le b$, is 0 if $c_h = -b-1$; it is $\max(0, j-\beta)$ with $\beta = -c_h - 1$ if $-b \le c_h \le -1$; it is $c_h + 1 + \min(j,\beta)$ with $\beta = b - c_h$ if $0 \le c_h \le b-1$; and it is $b+1$ if $c_h = b$ [I, Supplement, (F2)]. The two tools of [I, Supplement, §2] hold for every a with the same proofs.

Lemma 2.1 Lemma M: elementary move. *Let c' agree with c except $c'_h = c_h + \sigma$, $\sigma \in \{1,-1\}$, with $c'_h \in [-b-1, b]$. Put $J = \{j \in [0,b] : 0 \le c_h + 1 + j \le b\}$ if $\sigma = 1$ and $J = \{j \in [0,b] : 0 \le c_h + j \le b\}$ if $\sigma = -1$. Then $H'_i(j) = H_i(j) + \sigma$ if $0 \le i - h \le a$ and $j \in J$, and $H'_i(j) = H_i(j)$ otherwise. If several coordinates move, the changes of one row add, computed on the original vector.*

Proof. c_h enters row i exactly when $0 \le i - h \le a$, and $f(v+1) - f(v) = [-1 \le v \le b-1]$. Different coordinates enter a row through different terms. □

So if the net change of row i is $\sigma_i(j)$, then $Z(c') - Z(c) = \sum_i \sum_j (\delta_{H_i(j)+\sigma_i(j)} - \delta_{H_i(j)})$. A sum $\sum_{j=j_1}^{j_2} (\delta_{\alpha+j+1} - \delta_{\alpha+j})$ *telescopes* to $\delta_{\alpha+j_2+1} - \delta_{\alpha+j_1}$ (also when $j_2 = j_1 - 1$, where both sides are 0).

Lemma 2.2 Lemma P: exact point. *Let $\Lambda,\theta > 0$ with $\Lambda + h\theta \notin \mathbb{Z}$ for $|h| \le a$, let c be its vector and $\mathcal{P}$ the multiset of the positions $i\theta + j$ ($0 \le i \le a$, $0 \le j \le b$). Then $H_{i'}(j') = \#\{p \in \mathcal{P} : p < \Lambda + i'\theta + j'\}$, so $Z(c)$ is the multiset of the counts $\#\{p \in \mathcal{P} : p < p_0 + \Lambda\}$, $p_0 \in \mathcal{P}$.*

Proof. For one row i of positions and $h = i' - i$, $\#\{j : i\theta + j < \Lambda + i'\theta + j'\} = f(\lfloor \Lambda + h\theta \rfloor + j')$, and clipping does not change $f(v + j')$ for $0 \le j' \le b$ [I, eq. (2)]. □

2.1. The linear term and the period of $R(a,b)$.

Put $\Phi_q(x) = \#\{1 \le p \le x : \gcd(p,q) = 1\}$ and, following Alekseyev [1],

$$V(m,n) = \sum_{\substack{1 \le i \le m,\ 1 \le j \le n \\ \gcd(i,j) = 1}} (m+1-i)(n+1-j), \qquad N(m,n) = (2m+1)(2n+1) + 1 + 4V(m,n);$$

$N(m,n)$ is the number of two-dimensional threshold functions on the grid $\{0,\dots,m\}\times\{0,\dots,n\}$ [1, Theorem 3], and $(N(m,n)-2)/2$ the number of its linear partitions, studied by Acketa and Žunić [2]; see also Žunić [3]. By [I, Proposition 5.3], which identifies the cells of the quadrant with the regions of Alekseyev's arrangement in one octant,

$$R(a,b) = 1+a+b+2ab+\tfrac{1}{2}V(2b,2a) = \frac{N(2b,2a)+4a+4b+6}{8}. \tag{1}$$

Grouping the coprime pairs (p,q) by $q\le 2a$, $V(2b,2a)=\sum_{q=1}^{2a}(2a+1-q)S_q(b)$ with $S_q(b)=\sum_{p\le 2b,\ (p,q)=1}(2b+1-p)=\sum_{x=1}^{2b}\Phi_q(x)$, since p is counted once for each $x\in[p,2b]$. [I, Corollary 5.2] extracts the leading coefficient C_a. The next term and the period are as follows.

Proposition 2.3. *Let $a\ge 1$ and let P_a be the product of the odd primes $p\le 2a$ ($P_1=1$).*

(a) For all $b\ge 1$, $R(a,b)=C_ab^2+D_ab+\pi_a(b)$ with π_a periodic and

$$\boxed{D_a = C_a+a+1,} \qquad \textit{that is,} \qquad R(a,b) = C_a\,b(b+1)+(a+1)\,b+\pi_a(b).$$

In particular $D_1,\dots,D_6=\frac{9}{2},\ \frac{31}{3},\ \frac{93}{5},\ \frac{2059}{70},\ \frac{8917}{210},\ \frac{44707}{770}$.

(b) The period of π_a (equivalently, of the quasipolynomial $R(a,\cdot)$) divides P_a.

Proof. By Möbius inversion $\Phi_q(x)=\sum_{d\mid q}\mu(d)\lfloor x/d\rfloor$, so $S_q(b)=\sum_{d\mid q}\mu(d)G_d(2b)$ with $G_d(X)=\sum_{x=1}^{X}\lfloor x/d\rfloor$. Write $X=dk+r$, $0\le r<d$. Since $\lfloor x/d\rfloor=(x-(x \bmod d))/d$ and the residues of $1,\dots,X$ sum to $k\,d(d-1)/2+r(r+1)/2$,

$$G_d(X)=\frac{X^2-(d-2)X}{2d}+\gamma_d(r), \qquad \gamma_d(r)=\frac{r(d-2-r)}{2d}, \qquad r=X \bmod d,$$

for every $X\ge 0$. Only squarefree d occur ($\mu(d)\ne 0$). For $X=2b$, $r=2b \bmod d$ depends only on b modulo $d/\gcd(d,2)$, which for squarefree $d\mid q\le 2a$ is the product of the odd primes dividing d, a divisor of P_a. So each $S_q(b)$, hence by (1) $R(a,b)$, is a polynomial of degree 2 in b plus a function of period dividing P_a, for every $b\ge 1$; this is (b). The coefficients of b^2 and b in this decomposition are unique (a periodic function is bounded); the constant term is not, unless the periodic part is required to have mean zero, and it is not used.

Its coefficients follow from the polynomial parts. For $q=1$, $S_1(b)=b(2b+1)=2b^2+b$. For $q\ge 2$, the polynomial part of $\sum_{d\mid q}\mu(d)\frac{(2b)^2-(d-2)2b}{2d}$ is $\frac{\varphi(q)}{q}\,2b^2-\sum_{d\mid q}\mu(d)\frac{d-2}{d}\,b=\frac{\varphi(q)}{q}(2b^2+2b)$, because $\sum_{d\mid q}\mu(d)=0$ and $\sum_{d\mid q}\mu(d)/d=\varphi(q)/q$. Hence the polynomial part of $\frac{1}{2}V(2b,2a)$ is

$$\tfrac{1}{2}\cdot 2a(2b^2+b)+\sum_{q=2}^{2a}(2a+1-q)\frac{\varphi(q)}{q}(b^2+b) = 2ab^2+ab+(C_a-2a)(b^2+b) = C_ab^2+(C_a-a)b,$$

and adding $1+a+b+2ab$ gives the coefficient $C_a-a+1+2a=C_a+a+1$ of b. □

For $a=1$, $R(1,b)=(5b+4)(b+1)/2$ [I, Lemma 8.3] has linear coefficient $\frac{9}{2}=D_1$. For $a=2$, $W(2,b)=(44b^2+19b+e_{b \bmod 6})/6$ [I, Theorem 9.4] then gives $E(2,b)=(D_2-\frac{19}{6})b+O(1)=\frac{43}{6}b+O(1)$, the slope recorded in [I].

Remark 2.4 (literature, and what is checked). The exact count (1) is Alekseyev's [1, Theorem 3], through [I, Proposition 5.3], and the quasipolynomial structure follows from it by the elementary argument above; we claim no novelty for either. The leading term is in [1, Theorem 25] ([I, Remark 5.4]). The explicit linear term $D_a=C_a+a+1$ and the period

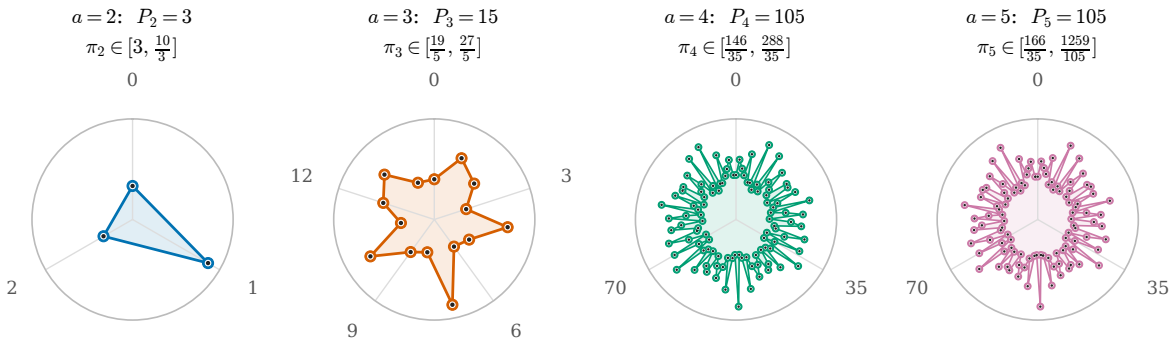

Figure 1. The periodic part of the count of cells. For $a = 2, 3, 4, 5$ each clock has P_a hours and shows $\pi_a(b) = R(a,b) - C_a b^2 - D_a b$ for $b = 1, \dots, 3P_a$, at angle $2\pi b/P_a$ and at a radius that is an affine function of $\pi_a(b)$ (scaled per panel; the range of π_a is printed above each clock). The three laps, $b \le P_a$ (large coloured rings), $P_a < b \le 2P_a$ (white rings) and $2P_a < b \le 3P_a$ (black dots), fall on the same points, as Proposition 2.3(b) predicts, and no proper divisor of P_a is a period. Values of $R(a,b)$ from (1), which agree with the exact enumerations of cells used in Figure 4.

bound P_a are not stated in [1], which we read in full, nor in the OEIS entries cited in [I]; we have not read [2] or [3] in full; the closed formula of [2, Theorem 3], $t(m,n) = f(m,n) + 2$ with $f(m,n) = \sum_{-m<i<m,\,-n<j<n,\,(i,j)=1} (m-|i|)(n-|j|)$, is seen as quoted in [13, eq. (1)]. It is a sum over coprime pairs of exactly the kind treated above, so the Möbius computation of Proposition 2.3 is a routine consequence of it. Haukkanen and Merikoski's asymptotics of the threshold-function count (read in full) give $t(m,n) = \frac{6}{\pi^2}(mn)^2 + O(mn^2)$ for $m \le n$, which with one side fixed leaves the linear term inside the error. So what may be new here is at most this explicit form. The bound P_a sharpens the period $\operatorname{lcm}(\operatorname{lcm}(1,\dots,a),\ \text{odd } q \le 2a)$ given in the proof of [I, Corollary 5.2]: through (1) the boundary terms of [I, Theorem 5.1], which carry the factor $\operatorname{lcm}(1,\dots,a)$, add up to $2ab$. By exact computation the minimal period of π_a is exactly P_a for $a \le 6$: $1, 3, 15, 105, 105, 1155$ (checked for $b \le 60, 60, 200, 800, 800, 3600$). With the linear coefficient $C_a + a$ in place of D_a, the remainder is not periodic in any of the six cases. Equality for every a is not proved. Figure 1 shows the periodic part π_a for $a = 2, \dots, 5$.

3. Wall weights and fibres

This section holds for every a. For a realisable c put

$$\Sigma(c) = \sum_{v \in Z(c)} v = \sum_{i=0}^{a} \sum_{j=0}^{b} H_i(j), \qquad \omega(h,\rho) = (a+1-|h|)(b+1-|\rho|) \quad (|h| \le a,\ |\rho| \le b),$$

and order vectors coordinatewise: $c \le d$ if $c_h \le d_h$ for all h. The *fibre* of a word is the set of realisable vectors with that word, and e_h is the unit vector of coordinate h. Cells and realisable vectors correspond one to one (Section 2), so we speak of cells and vectors interchangeably.

3.1. The wall-crossing formula.

Proposition 3.1 wall weights. *(a) Let $c \in [-b-1, b]^{2a+1}$ (realisable or not), $|h| \le a$ and $c_h \le b-1$. Then*

$$\Sigma(c + e_h) - \Sigma(c) = (a + 1 - |h|)(b + 1 - |c_h + 1|).$$

(b) Crossing the wall $\ell_{h,\rho} : \Lambda + h\theta = \rho$ ($|\rho| \le b$) at a point on no other line, from $\Lambda + h\theta < \rho$ to $\Lambda + h\theta > \rho$, changes only c_h, from $\rho - 1$ to ρ, and raises Σ by exactly $\omega(h, \rho) \ge 1$.

(c) For realisable c, d, $\Sigma(d) - \Sigma(c) = \sum_{|h| \le a} (a + 1 - |h|)(F(d_h) - F(c_h))$, where $F(v) = \sum_{j=0}^{b} f(v + j)$ and $F(v) - F(u) = \sum_{\rho=u+1}^{v} (b + 1 - |\rho|)$ for $u \le v$: the signed sum of the weights of the walls that separate the two cells, independent of the path.

Proof. The coordinate c_h enters row i exactly when $0 \le i - h \le a$, that is, in $a + 1 - |h|$ rows, always through the term $f(c_h + j)$. So $\Sigma(c) = \sum_h (a + 1 - |h|) F(c_h)$. Since $f(v+1) - f(v) = [-1 \le v \le b-1]$ (proof of Lemma 2.1), $F(v+1) - F(v) = \#\{j \in [0, b] : -1 \le v + j \le b - 1\}$, which is $b - v = b + 1 - |v + 1|$ if $v \ge -1$ and $v + b + 2 = b + 1 - |v + 1|$ if $v \le -2$; this is (a), and (c) follows by summing. For (b): at such a point only the floor $\lfloor \Lambda + h\theta \rfloor$ changes, from $\rho - 1$ to ρ, and since $-b - 1 \le \rho - 1 < \rho \le b$ clipping leaves both values unchanged; (a) with $c_h = \rho - 1$ gives $\omega(h, \rho)$, which is ≥ 1 because $|h| \le a$ and $|\rho| \le b$. □

Corollary 3.2. *(i) Two cells separated by a single wall (adjacent cells) have different words.*

(ii) If $c \le d$ and $c \ne d$ then $\Sigma(c) < \Sigma(d)$. In particular every fibre is an antichain for $\le$.

(iii) If $Z(c) = Z(d)$, the weights of the walls crossed upwards (towards $\Lambda + h\theta > \rho$) and downwards between the two cells balance: $\sum_h (a+1-|h|)(F(d_h) - F(c_h)) = 0$.

Proof. Σ is a function of $Z(c)$, hence of the word. (i) is Proposition 3.1(b), (ii) follows from (a) because every step weight is ≥ 1, and (iii) is (c) with $\Sigma(c) = \Sigma(d)$. □

Remark 3.3 (what E counts: the arrangement, and generic words). (a) *E does not depend on a choice of presentation.* A word is a function of the cell (Lemma 2.2: the counts depend only on the signs of $\Lambda + h\theta - \rho$), so no line is missing; and by Corollary 3.2(i) every wall between two cells separates two different words, so no line can be erased, and no two cells of one fibre are adjacent. The cells are convex, and two cells meet at most along walls and at vertices, which are off the set of generic parameters. Hence the connected components of the open set of generic parameters whose word is w are exactly the cells of the fibre of w, and $E = \sum_w (\#\pi_0(U_w) - 1)$ where U_w is that set: there is no excess from over-subdivision, only from words that have several disconnected realisations. (If one adds lines that change no word, R and E grow together; here none is superfluous.) The proof is the one of the cited results; it is stated here because it is what makes E an intrinsic count of the word map. (b) *Only generic words occur.* No integer lies on a line: $\Lambda + h\theta = \rho$ means $t\, r^h = s^\rho$ with $t = 2^e$, $e \ge 1$, and r, s odd primes, which is impossible because the 2-adic valuation of the left side is $e \ge 1$ and that of the right side is 0. So $W(a, b)$ counts the words of integers, all of which lie in open cells, and the words of singular parameters (where two projected values coincide and brackets cancel) never occur.

Figure 2 draws Σ as a staircase for $a = 2$, $b = 3$: every fibre sits on one level, and its cells are never adjacent nor comparable.

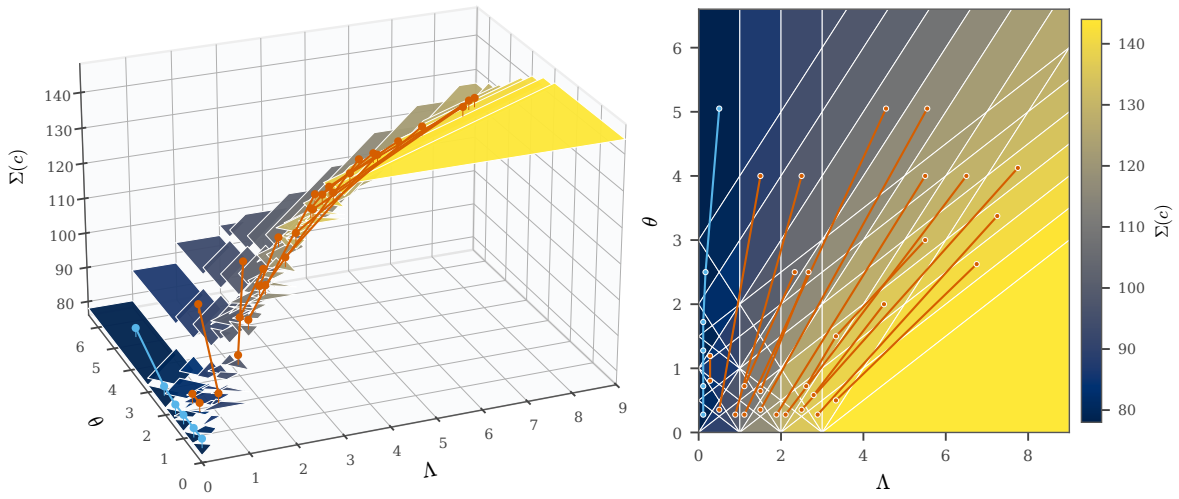

Figure 2. Wall weights for $a = 2$, $b = 3$. Left: the cells of the arrangement in the window $0 < \Lambda < 9$, $0 < \theta < 6.6$ (99 of the $R(2,3) = 100$ cells; the missing one shares its word with no other cell), each drawn as a horizontal terrace at height $\Sigma(c)$; right: the same cells seen from above. Crossing a wall upwards raises Σ by $\omega(h,\rho)$ (Proposition 3.1). The 15 non-trivial fibres ($E(2,3) = 20$) are joined by bridges at their common height: the fibre of $O(0)$ (light blue, 6 cells near $\Lambda = 0$) and 14 others (vermilion). Every bridge is horizontal and joins cells that are neither adjacent nor comparable (Corollary 3.2). Equal height is necessary, not sufficient, for equal words. Cells computed exactly as intersections of half-planes; fibres by exact enumeration.

Remark 3.4 (how much the balance filters). Corollary 3.2(iii) is necessary, not sufficient, and is cheap to test before comparing observation vectors. For $a \le 3$ and $b \le 15$ we measured it on all pairs of cells. Few pairs pass it: 4011 of the 6.36×10^6 pairs for $a = 3$, $b = 15$ (0.06%). Among those that pass, the share of genuine coincidences falls with b: from 75% ($b = 2$) to 40% ($b = 15$) for $a = 1$, from 74% to 21% for $a = 2$, and from 57% ($b = 3$) to 17% for $a = 3$.

3.2. A linear bound for every fibre.

Proposition 3.5. *Every fibre has at most* $1 + \sum_{q=1}^{2a} \Phi_q(2b) \le 4ab + 1$ *vectors.*

Proof. Two lines $\ell_{h,\rho}$, $\ell_{h',\rho'}$ with $h \neq h'$ meet at $\theta = (\rho - \rho')/(h - h')$, and lines with the same h are parallel. So every vertex of the arrangement in the open quadrant has θ in the set $\mathcal{T}$ of positive fractions p/q in lowest terms with $q \le 2a$ and $p \le 2b$, and $|\mathcal{T}| = \sum_{q \le 2a} \Phi_q(2b)$. The set $\mathcal{T}$ also contains the abscissa $\theta = \rho/h$ at which a line $\ell_{h,\rho}$ crosses the boundary line $\ell_{0,0} = \{\Lambda = 0\}$, so inside each strip below every line lies entirely above or entirely below $\Lambda = 0$. The vertical lines $\theta \in \mathcal{T}$ cut the quadrant into $|\mathcal{T}| + 1$ open vertical strips. Inside one strip no two lines cross, so the lines are graphs $\Lambda = \rho - h\theta$ over its θ-interval, totally ordered, and every cell that meets the strip meets each vertical line $\theta = \theta_0$ of it. Fix such a θ_0: along it every coordinate $\mathrm{clip}\lfloor \Lambda + h\theta_0 \rfloor$ is non-decreasing in Λ, so any two cells meeting the strip have comparable vectors. By Corollary 3.2(ii) a fibre meets each strip in at most one cell, and every cell, being open, meets some strip. Finally $\Phi_q(2b) \le 2b$. □

The strip argument is the usual vertical decomposition of a line arrangement, with the Farey-type set $\mathcal{T}$ of abscissas of the vertices; only its combination with Corollary 3.2(ii) is specific to this problem, and it is elementary. In the data ($a \le 3$, $b \le 15$) the bound holds with a wide margin: the largest fibre is always that of $O(0)$, of size $1 + \sum_{h \le a} \Phi_h(b)$ [I, Lemma 7.1] (34 for $a = 3$, $b = 15$, against 115 and $4ab+1 = 181$). So the order $O_a(b)$ is right and the constant is not. A bound on fibres does not bound E: for $a = 1$ most non-trivial fibres have size two, and $E(1, b)$ is quadratic. The relevant count is that of the cells whose word is shared,

$$\mathrm{Amb}(a, b) = \#\{c \text{ realisable} : \ Z(d) = Z(c) \text{ for some realisable } d \neq c\}.$$

Lemma 3.6. *$E(a, b) \le \mathrm{Amb}(a, b) \le 2E(a, b)$. Hence $E(a, b) = O_a(b)$ if and only if $\mathrm{Amb}(a, b) = O_a(b)$.*

Proof. A fibre of size $s \ge 2$ contributes $s - 1$ to E and s to Amb, and $s - 1 \le s \le 2(s - 1)$; fibres of size 1 contribute to neither. □

In the 42 cases $a \le 3$, $2 \le b \le 15$ the ratio Amb/E lies between 1.75 and 1.92, usually near 1.78: most non-trivial fibres are pairs. So rigidity is a statement on the size of the set of ambiguous cells, not on the size of fibres (Question 10.4).

3.3. Power sums. For a realisable c put $I_j(c) = \sum_{v \in Z(c)} v^j$; $I_0 = (a + 1)(b + 1)$ is constant and $I_1 = \Sigma$. The word determines all I_j. Conversely:

Theorem 3.7. *Let $L_a = \mathrm{lcm}(1, \dots, a + 1)$ and $t_a = 2(a + 1)(2a + 3)L_a$. If c, d are realisable for the same b and $I_j(c) = I_j(d)$ for $0 \le j \le 2t_a - 2$, then $Z(c) = Z(d)$. So the number of power sums needed depends on a only.*

Proof. Let $M_c(z) = \sum_{i,j} z^{H_i(j)}$. By [I, Section 6, Step 3], $(1 - z^{L_a})M_c$ is a sum of at most t_a monomials for every c, so $P = (1 - z^{L_a})(M_c - M_d) = \sum_{k=1}^{s} \kappa_k z^{e_k}$ with distinct exponents $e_k \ge 0$ and $s \le 2t_a$. Let $\vartheta = z\frac{d}{dz}$. Then $\vartheta^j(M_c - M_d)(1) = I_j(c) - I_j(d) = 0$ for $j \le 2t_a - 2$. Since $z^i \frac{d^i}{dz^i} = \vartheta(\vartheta - 1) \cdots (\vartheta - i + 1)$, the derivatives of $M_c - M_d$ of order $\le 2t_a - 2$ vanish at $z = 1$, so $M_c - M_d$ vanishes there to order $\ge 2t_a - 1$, and P, with the simple zero of $1 - z^{L_a}$, to order $\ge 2t_a \ge s$. By the same relation, $\vartheta^j P(1) = \sum_k \kappa_k e_k^j = 0$ for $0 \le j \le s - 1$. The matrix $(e_k^j)_{k,j}$ is a Vandermonde matrix on distinct nodes, hence invertible (this is Hajós's lemma [14, 15]: a nonzero polynomial with s monomials has no root $z \neq 0$ of multiplicity $\ge s$), so every $\kappa_k = 0$, $P = 0$, and $M_c = M_d$ because $\mathbb{Z}[z]$ is a domain. □

The argument is an application of Hajós's lemma to the sparse representation of [I, Section 6, Step 3]; we claim no novelty for it, only for the explicit bound t_a through [I]. The cyclotomic multiplier of Section 5 gives a bound polynomial in a.

Proposition 3.8 written proof. *Let $d_a = \sum_{k=1}^{a+1} \varphi(k)$ and $s_a = 2(a + 1)(a + 2)(d_a + 1)$. If c, d are realisable for the same b and $I_j(c) = I_j(d)$ for $0 \le j \le 2s_a - 2$, then $Z(c) = Z(d)$. So $2s_a - 1 = O(a^4)$ power sums determine the word.*

Proof. Let $\mathcal{Q} = \prod_{k \le a+1} \Phi_k$, of degree d_a, divisible by $1 - z^s$ for $1 \le s \le a + 1$, with a simple zero at $z = 1$. Each term $f(c_h + j)$ of a row changes slope at most once for $j \in [0, b]$: if $c_h < 0$ only the lower clipping point can occur there, and if $c_h \ge 0$ only the upper one can. Thus H_i is affine on at most $a + 2$ intervals, with slope $s \in \{0, \dots, a + 1\}$. An interval of slope $s \ge 1$, first value u and N terms contributes $(\mathcal{Q}/(1 - z^s))(z^u - z^{u+sN})$ to $\mathcal{Q}M_c$, at most

$2(d_a - s + 1) \le 2(d_a + 1)$ monomials, and one of slope 0 contributes $Nz^u\mathcal{Q}$, at most $d_a + 1$. So $\mathcal{Q}M_c$ has at most $(a+1)(a+2)\cdot 2(d_a+1) = s_a$ monomials, and $P = \mathcal{Q}(M_c - M_d)$ at most $2s_a$. The proof of Theorem 3.7 applies with $\mathcal{Q}$ in place of $1 - z^{L_a}$: $M_c - M_d$ vanishes at 1 to order $\ge 2s_a - 1$, so P to order $\ge 2s_a$, and Hajós's lemma gives $P = 0$. Finally $d_a \le \binom{a+2}{2}$, so $s_a = O(a^4)$. □

For $a = 2, \dots, 6$, $s_a = 120, 280, 660, 1092, 2128$ against $t_a = 252, 864, 6600, 9360, 88200$. The bound is very large ($2t_2 - 2 = 502$ for $a = 2$). In the data a handful of power sums suffice. Let $J(a,b)$ be the least J such that $(I_0, \dots, I_J)$ separates all the words of (a,b) (computed by exact enumeration): $J(1,b) = 3$ for $2 \le b \le 20$; $J(2,b) = 3$ for $b \le 20$ except $J = 5$ at $b = 6, 12, 16$; $J(3,b) = 3$ for $b \le 14$ and $J(3,15) = 4$. In each exceptional case $(I_0, \dots, I_3)$ confuses a single pair of words (for $a = 2$, $b = 6$: 287 classes against $W = 288$). The sum I_1 alone separates about half of the words (1585 classes of 3399 for $a = 3$, $b = 15$). See Question 10.5.

3.4. Square patterns. For $a = b$ the integer $2^{e-1}r^a s^a$ does not change when r and s are exchanged.

Proposition 3.9. *Let $a = b$. The map $\sigma(\Lambda, \theta) = (\Lambda/\theta, 1/\theta)$ permutes the cells, preserves words and fixes exactly $2a + 1$ cells. Hence*

$$W(a,a) \le \frac{R(a,a) + 2a + 1}{2}.$$

Proof. σ is an involution of the open quadrant. A point of $\ell_{h,\rho}$ goes to a point with $\Lambda' - \rho\theta' = (\Lambda - \rho)/\theta = -h$, so $\sigma(\ell_{h,\rho}) = \ell_{-\rho,-h}$, a line of the arrangement because $a = b$; hence σ permutes the cells. The point $(\log t/\log s, \log r/\log s)$ of $n = 2^{e-1}r^a s^a$ goes to $(\log t/\log r, \log s/\log r)$, the point of the same n with r and s exchanged; every cell contains points of this kind [I, Lemma 2.3] and the word is constant on cells, so σ preserves words. In homogeneous coordinates $(\Lambda, \theta) = (X/Z, Y/Z)$, σ is the linear map $(X, Y, Z) \mapsto (X, Z, Y)$; it maps the open octant $X, Y, Z > 0$ to itself, and the cone $\{(Z\Lambda, Z\theta, Z) : (\Lambda, \theta) \in C,\ Z > 0\}$ to the cone over $\sigma(C)$. If $\sigma(C) = C$, the cone over C is therefore convex and σ-invariant, so $x + \sigma(x)$ lies in it for any x in it, and C contains a fixed point, that is, a point with $\theta = 1$; conversely a cell meeting $\theta = 1$ is fixed. The lines meet the half-line $\theta = 1$, $\Lambda > 0$ at $\Lambda = \rho - h$, that is, at $\Lambda = 1, \dots, 2a$, and cut it into $2a + 1$ open intervals; each lies in one cell, and no two lie in the same cell, which is convex and avoids the lines. So $2a + 1$ cells are fixed, and σ has $(R(a,a) - (2a+1))/2 + (2a+1)$ orbits, each inside one fibre. □

This is the orbit count of an involution (Burnside's lemma); we claim no novelty beyond identifying the fixed cells. For $a = 1, 2, 3, 4$ the orbits number $6, 29, 99, 262$ and $W(a,a) = 6, 28, 95, 255$: the bound is attained for $a = 1$, and the differences $0, 1, 4, 7$ count the coincidences that the exchange of the primes does not explain.

Remark 3.10 (exchanging the primes when $a \ne b$; written proof for the first part). For $a \ne b$ the same map σ sends $\ell_{h,\rho}$ to $\ell_{-\rho,-h}$, that is, the arrangement of the box (a,b) onto that of (b,a). The point σP of $n = 2^{e-1}r^a s^b$ is its point when the primes are exchanged, the exponents then being (b,a) in the box; every cell is realised by primes [I, Lemma 2.3] and the word is constant on cells, so σ is a bijection from the cells of (a,b) onto those of (b,a) that preserves words. Hence

$$R(a,b) = R(b,a), \qquad W(a,b) = W(b,a).$$

(Exact values of $R(a,b)$ and $W(a,b)$ for $(a,b) = (1,2)$: $21, 16$; $(1,5)$: $87, 69$; $(2,3)$: $100, 80$; $(2,4)$: $162, 136$; $(2,5)$: $238, 200$; $(2,6)$: $329, 288$; $(3,4)$: $313, 268$; $(3,5)$: $462, 410$; each equals the value at (b,a).) But σ does not permute the cells of one arrangement unless $a = b$, so the integers $2^{e-1}r^a s^b$ and $2^{e-1}s^a r^b$ need not have the same word. They do exactly when σP lies in a cell of the fibre of P. Computed on cells for $a = 2$, $b = 12$, over the 149 736 triples (r, s, e) with $r \neq s$ odd primes ≤ 200 and $1 \leq e \leq \min(200, e_{\max})$, where $e_{\max} = \lfloor (b + a\max(\theta, 1/\theta) + 1) \log s / \log 2 \rfloor + 1$, excluding the triples for which P or σP lies in the trivial cell (all $c_h = b$) or within 10^{-9} of a line (floating point, the cell of each point then read exactly): the words coincide for 22.0% of the triples; in 21.2% the point σP is in the *same* cell as P (this is the effect of σ fixing $\theta = 1$: in this sample the same-cell proportion is 0 for $\min(\theta, 1/\theta) < 0.8$, and the coincidence of words is 0 below 0.5), and in 0.85% it is in another cell of the fibre, which happens for only three pairs of cells. The most frequent is the pair of Proposition 3.11 below; an instance is $11^2\,7^{12}$ and $7^2\,11^{12}$, whose words (78 symbols) coincide. So in this sample there is no symmetry $p^a q^b \leftrightarrow q^a p^b$ for $a \neq b$, only $\sigma P \in$ cell of P near $\theta = 1$ and three exceptional pairs of cells.

Proposition 3.11 written proof. *Let $a = 2$ and $b \geq 2$. The vectors $c = (-3, -1, 0, 1, 2)$ and $d = (-2, -1, 0, 1, 1)$ are realisable, distinct, and $Z(c) = Z(d)$.*

Proof. The points $(\frac{3}{10}, \frac{6}{5})$ and $(\frac{3}{10}, \frac{4}{5})$ are witnesses: $\Lambda + h\theta$ is $-\frac{21}{10}, -\frac{9}{10}, \frac{3}{10}, \frac{3}{2}, \frac{27}{10}$ and $-\frac{13}{10}, -\frac{1}{2}, \frac{3}{10}, \frac{11}{10}, \frac{19}{10}$ for $h = -2, \ldots, 2$, with floors c and d, unchanged by clipping since $b \geq 2$. From c to d, c_{-2} rises by 1 and c_2 falls by 1. By Lemma 2.1, $c_{-2} = -3$ enters only row 0, on $J = [2, b]$, and $c_2 = 2$ only row 2, on $J = [0, b-2]$. On $[2, b]$ row 0 of c is $f(j) + f(j-1) + f(j-3) = (j+1) + j + (j-2) = 3j - 1$, and on $[0, b-2]$ row 2 is $f(j+2) + f(j+1) + f(j) = 3j + 6$. So

$$Z(d) - Z(c) = \sum_{j=2}^{b} (\delta_{3j} - \delta_{3j-1}) + \sum_{j=0}^{b-2} (\delta_{3j+5} - \delta_{3j+6}) = 0,$$

since the second sum is the first with the opposite sign ($j \mapsto j + 2$). □

4. The dimension criterion

Let $\leq_{\text{lex}}$ be the lexicographic order on $\mathbb{Z}^{2a+1}$ and

$$B_a = \{(b, c) : \ b \geq 1, \ c \text{ realisable for } b, \ \exists d <_{\text{lex}} c \text{ realisable for } b \text{ with } Z(d) = Z(c)\}.$$

Each class of size k contributes $k - 1$ vectors to the section $B_a(b)$, so $E(a, b) = |B_a(b)|$. By [I, Section 6, Steps 2 and 3], realisability and $Z(c) = Z(d)$ are quantifier-free Presburger conditions on (b, c, d), so B_a is Presburger-definable. After the shift $c \mapsto c + (b+1)\mathbf{1}$ it lies in $\mathbb{N}^{2a+2}$, and by the theorem of Ginsburg and Spanier [5] it is *semilinear*: a finite union of linear sets $L = x_0 + \mathbb{N}v_1 + \cdots + \mathbb{N}v_k$. The *rank* of L is the rank of $\{v_1, \ldots, v_k\}$.

Proposition 4.1 dimension criterion. *Fix a and a decomposition of B_a into linear sets. Then $E(a, b) = O(b)$ if and only if every component has rank at most 2. In particular the condition does not depend on the decomposition.*

Proof. Zero generators can be deleted without changing L, so assume $v_i \neq 0$ for all i; a component of rank 0 is the single point x_0 and contributes at most one vector, to one value of b.

($\Leftarrow$) Let $L = x_0 + \sum \mathbb{N} v_i$ be a component. Each $v_i \neq 0$ has positive b-coordinate: otherwise the points $x_0 + tv_i$, $t \in \mathbb{N}$, would be infinitely many points of B_a with the same b, while $B_a(b) \subset [0, 2b+1]^{2a+1}$ is finite. So the span V of the v_i contains a vector with nonzero b-coordinate, and the section of L at $b = b_0$ lies in the affine space $(x_0 + V) \cap \{b = b_0\}$, of dimension $\operatorname{rank} L - 1 \leq 1$. A point contributes one vector. A line lies in $\{b = b_0\}$, so some c-coordinate varies along it, and it meets the cube $[0, 2b_0+1]^{2a+1}$ in at most $2b_0 + 2$ integer points. Summing over the finitely many components gives $E(a, b) = O(b)$.

($\Rightarrow$) Suppose a component has rank ≥ 3, and take independent v_1, v_2, v_3 among its generators, with b-coordinates $\beta_i > 0$. Put $\ell = \operatorname{lcm}(\beta_1, \beta_2, \beta_3)$ and $w_i = (\ell/\beta_i)v_i$. The points $x_0 + k_1 w_1 + k_2 w_2 + k_3 w_3$ with $k_i \in \mathbb{N}$, $k_1 + k_2 + k_3 = t$, lie in L, have $b = b_0 + t\ell$, and are pairwise distinct because the w_i are independent. So $E(a, b_0 + t\ell) \geq \binom{t+2}{2}$, which is not $O(b)$. □

Remark 4.2. $E(a, b) = R(a, b) - W(a, b)$ is eventually a quasipolynomial, since both terms are [I, Corollary 5.2 and Theorem 6.1]. With $E(a, b) = O(b)$ and the linear lower bound of [I, Lemmas 7.1 and 7.5] for $a \geq 2$, every constituent then has degree exactly one. Proposition 4.1 is not used below: the reduction bounds $E(a, b)$ directly. It explains what is being certified: the absence of two-parameter families of coincidences inside one value of b. Woods's theorem [8] gives that $E(a, b)$ is eventually a quasipolynomial, not its degree in terms of the ranks of the components; we did not find the criterion in this form, but it is a short consequence of the Ginsburg–Spanier theorem and we do not present it as a deep result.

Remark 4.3 (effectivity). A semilinear representation of a Presburger-definable set can be computed from its defining formula (quantifier elimination for Presburger arithmetic and the effective form of the Ginsburg–Spanier theorem [5]), and the ranks of the components are then linear algebra. So, for each fixed a, whether $E(a, b) = O(b)$ is decidable. Moreover the proof of ($\Leftarrow$) gives, if the decomposition has N components all of rank ≤ 2, $E(a, b) \leq N(2b + 2)$ for *every* $b \geq 1$. This is a decision procedure in principle only: N is not known and the general algorithms are far too costly. The finite reduction of Section 5 is a practical sufficient condition, and a failure of its conditions for some a would not by itself disprove rigidity.

5. The finite reduction

Fix a. All constants below depend on a only, and "for b large" means $b \geq b_0(a)$.

5.1. Templates. Write a point of the quadrant off the lines $\Lambda + h\theta \in \mathbb{Z}$ as $\Lambda = n + u$, $\theta = m + v$ with $n, m \in \mathbb{Z}_{\geq 0}$ and $u, v \in (0, 1)$. Then $\lfloor \Lambda + h\theta \rfloor = n + hm + \varepsilon_h$ with $\varepsilon_h = \lfloor u + hv \rfloor$, so

$$c = g(n, m, \varepsilon) := \big(\operatorname{clip}(n + hm + \varepsilon_h)\big)_{|h| \leq a}.$$

The vector $\varepsilon = (\varepsilon_h)$, the *template*, depends only on the cell of (u, v) in the arrangement of the lines $u + hv = k$ ($|h| \leq a$, $k \in \mathbb{Z}$) in the open unit square; there are finitely many templates (4, 14, 36, 76 for $a = 1, \ldots, 4$). This is classical: since $\varepsilon_0 = 0$, a template is determined by its gap word $(\varepsilon_{h+1} - \varepsilon_h)_{-a \leq h < a} \in \{0, 1\}^{2a}$, which is a factor of length $2a$ of a lower mechanical word of slope v, that is, a finite Sturmian (balanced) word [12, Ch. 2]. The bijection between these factors and the cells of such an arrangement in the unit square is Berstel and Pocchiola's geometric proof [11] of Mignosi's count $1 + \sum_{i=1}^{2a} (2a + 1 - i)\varphi(i)$ [10] (OEIS A005598), which gives $4, 14, 36, 76, 136, 224, 346$ for $a = 1, \ldots, 7$, as our sweep does. Conversely, every triple (n, m, ε) with $n, m \geq 0$ gives a realisable vector, witnessed by $(n + u, m + v)$ for any (u, v) in

the cell of ε. Every realisable vector has a representative (n, m, ε), because its cell is open and so contains points with $\Lambda, \theta \notin \mathbb{Z}$; representatives need not be unique, since the lines $\theta \in \mathbb{Z}$ are not lines of the arrangement and a cell may contain points with different $\lfloor\theta\rfloor$. The counts below are counts of representatives, which can only over-count vectors. The templates are found by a complete sweep. Let $0 = v_0 < v_1 < \cdots < v_r = 1$ be the critical values of v (the v-coordinates of the vertices of the arrangement in the square and of its crossings with $u = 0$ and $u = 1$). A cell is open and convex, so its projection to the v-axis is an open interval whose ends are critical values; it therefore meets the vertical line through the midpoint $\bar{v}_k$ of some interval (v_{k-1}, v_k). On that line the arrangement cuts $(0, 1)$ into finitely many open u-intervals, each lying in exactly one cell. For *each* $\bar{v}_k$ the program takes the midpoint of *every* one of these u-intervals; one point per interval of v would not suffice, since a vertical line meets several cells.

5.2. The box and the thin vectors. If c has two free coordinates $c_h, c_{h'}$, $h \neq h'$, then $\Lambda + h\theta$ and $\Lambda + h'\theta$ lie in $(-b, b)$, so $|h - h'|\theta < 2b$; hence $m = \lfloor\theta\rfloor \leq 2b - 1$, and $\Lambda < b + a\theta$ gives $n < (2a+1)b$. So every representative of such a c lies in the box $\mathcal{B}_b = [0, (2a+2)b] \times [0, 3b]$. Vectors with at most one free coordinate (*thin* vectors) are $O(b)$ in number: c is non-decreasing, so it is fixed by the position of the free coordinate, its value, and the clipped values on either side.

5.3. Chambers and strips. In the normalised coordinates $(\lambda, \mu) = (n/b, m/b)$ the box is $[0, 2a + 2] \times [0, 3]$. A *chamber* is one of finitely many convex polygons covering it, obtained from the lines $\lambda + h\mu \in \{-1, 0, 1\}$ ($|h| \leq a$), the lines $\lambda = 0$, $\mu = 0$ and the edges of the box, refined as described below. Every sign used in the deduction of the next subsection is the sign of an affine form $\phi = Ab + Bn + Cm + D$ with integer coefficients, and the deduction uses the sign of its homogeneous part $A + B\lambda + C\mu$ on the interior of the chamber. If that part changes sign in a chamber (tested exactly at its vertices), the chamber is cut along its zero line. The deduction uses strict signs. If $(n/b, m/b)$ is interior to the chamber but $\phi(b, n, m)$ does not have the strict sign of its homogeneous part (it has the other sign or vanishes), then $|Ab + Bn + Cm| \leq |D|$: the point lies in a *strip* of bounded width. A strip with $(B, C) \neq (0, 0)$ contains $O(b)$ integer points of $\mathcal{B}_b$; one with $(B, C) = (0, 0)$ only matters for $b \leq |D|/|A|$; a constant form ($A = B = C = 0$) has the sign of D everywhere, and that sign is used directly. Points on chamber edges lie in the strip of width 0 of that edge. The refinement stops: for fixed a there are finitely many clipping states, orders of break points, row decompositions and resulting affine forms, so only finitely many zero lines can be added.

A *piece* $P = (C_P, \varepsilon)$ is a chamber together with a template. A triple (n, m, ε), or a point (b, n, m) of the piece, is *generic* if $(n/b, m/b)$ lies in the interior of C_P and outside all the strips of P. Pieces are formed only in the chambers whose interior points have at least two free coordinates. A vector with at least two free coordinates whose representatives all fall in the other chambers has (n, m) within $O(1)$ of a line $\lambda + h\mu = \pm 1$: the clipping state of c_h changes only across these lines, which are chamber edges. So it lies in a strip. A vector is *generic* if it has a generic representative and *exceptional* otherwise. By the above,

$$\#\{\text{exceptional vectors for } b\} = O(b). \tag{2}$$

5.4. The observation vector. At the generic points of a piece, the clipping state of every coordinate, the order of the break points of every row (the numbers $-c_h - 1$ and $b - c_h$, together with 0 and b) and the slopes of every row are fixed. Multiply the generating polynomial

$M_c(z) = \sum_{i,j} z^{H_i(j)}$ by

$$\mathcal{Q}(z) = \prod_{d \le a+1} \Phi_d(z),$$

the product of the cyclotomic polynomials, which is divisible by $1 - z^s$ for every $s \le a + 1$. As in [I, Section 6, Step 3], a piece of a row with slope $s \ge 1$, first exponent u and N terms contributes $(\mathcal{Q}(z)/(1 - z^s))(z^u - z^{u+sN})$, a piece of slope 0 contributes $Nz^u\mathcal{Q}(z)$, and so $K_c = \mathcal{Q}M_c$ is a sum of boundedly many monomials whose exponents and coefficients are affine in (b, n, m). The sparse representation is made canonical in three steps. (a) Monomials whose exponent forms are *identical* (as affine forms) are merged, adding their coefficients. (b) Coefficients that vanish identically are deleted; every remaining non-constant coefficient vanishes only on a strip (discarded), so at generic points every listed coefficient is nonzero. (c) The distinct exponent forms are sorted at generic points: two forms with the same homogeneous part differ by a nonzero constant and are ordered by it; two forms with different homogeneous parts are ordered by the sign of their difference, which is one of the recorded forms (strict at generic points). The order is checked for every consecutive pair. So at a generic point the listed exponents are pairwise distinct and increasing, and the list is the sparse representation of K_c itself. This gives the *observation vector*

$$V_P(b, n, m) = (e_1, \kappa_1, e_2, \kappa_2, \dots), \qquad \text{affine in } (b, n, m),$$

the sorted sparse representation of K_c. Since $\mathbb{Z}[z]$ is a domain, $M_c = M_d$ if and only if $K_c = K_d$; so for generic $c \in P$, $d \in P'$,

$$Z(c) = Z(d) \iff V_P(b, n, m) = V_{P'}(b, n', m'). \tag{3}$$

5.5. Signatures and the count. Suppose that the (n, m)-part of V_P has rank 2. Then $(n, m) \mapsto V_P(b, n, m)$ is injective, and $A_P = \{(b, V_P(b, n, m))\}$ is an affine subspace of dimension 3 of $\mathbb{Q}^{1+|V_P|}$. Its canonical form (reduced row-echelon form of the three directions and the reduced base point) is the *signature* of P.

Theorem 5.1 finite reduction. *Fix a. Suppose that*

(C1) every piece of a chamber that contains vectors with at least two free coordinates has observation vector of (n, m)-rank 2; and

(C2) for every pair of pieces P, P' with the same signature, writing $V_{P'}(b, \phi(b, n, m)) = V_P(b, n, m)$ for the affine map ϕ so defined, and ϕ_{hom} for the affine map of the normalised plane with $\phi(b, n, m) = (b,\, b\,\phi_{\mathrm{hom}}(n/b, m/b) + k)$, k constant: either $c_{P'} \circ \phi = c_P$ as affine maps, or no integer point is mapped by ϕ to an integer point, or the set $\{(\lambda, \mu) \in C_P : \phi_{\mathrm{hom}}(\lambda, \mu) \in C_{P'}\}$ has area 0 (area measured in the plane of C_P).

Then $E(a, b) = O(b)$.

Proof. $E(a, b)$ is at most the number of realisable c that have a partner $d \ne c$ with $Z(d) = Z(c)$. We bound four kinds.

(i) c thin or exceptional: $O(b)$, by the box and (2).

(ii) c generic and d thin or exceptional. Here "generic" refers only to pieces of chambers whose vectors have at least two free coordinates, so c lies in such a piece P and (C1) applies to it; d need not have an observation vector. By (C1) and (3), for a fixed d and a fixed piece there is at most one generic (n, m) with $V_P(b, n, m)$ equal to the sparse representation

of K_d. So the number of such c is at most the number of pieces times the number of thin or exceptional d, which is $O(b)$ by the box (thin vectors are $O(b)$) and (2).

(iii) $c \in P$, $d \in P'$ generic, with different signatures. Then $S = \{(b, x, y) : V_P(b, x) = V_{P'}(b, y)\}$ is isomorphic to $A_P \cap A_{P'}$, a proper affine subspace of A_P, so $\dim S \leq 2$. If b is constant on S, it concerns one value of b. Otherwise each section $S \cap \{b = b_0\}$ has dimension ≤ 1, its x-projection is injective (by (C1)) and lies on a line or a point, and contains $O(b)$ integer points of the box.

(iv) $c \in P$, $d \in P'$ generic, with the same signature. Then $A_P = A_{P'}$, so $y = \phi(b, x)$ for the affine rational map ϕ of (C2), which preserves b. In the first alternative of (C2), $d = c$. In the second, there is no pair. In the third, a pair gives a point $(\lambda, \mu) = (n/b, m/b)$ of C_P with $\phi_{\hom}(\lambda, \mu) + k/b \in C_{P'}$. So (λ, μ) satisfies, up to an error $O(1/b)$, the finitely many linear inequalities that define the closed polygon $X = \overline{C_P} \cap \phi_{\hom}^{-1}(\overline{C_{P'}})$ (open or closed chambers give the same area). By Hoffman's error bound for systems of linear inequalities [9], such a point lies within $O(1/b)$ of X if $X \neq \emptyset$. If $X = \emptyset$ there is no such point for b large: otherwise points of the compact polygon $\overline{C_P}$ with errors tending to 0 would have a limit point in X. A convex polygon of area 0 is a segment or a point, and its $O(1/b)$-neighbourhood scaled by b contains $O(b)$ integer points.

There are finitely many pieces and pairs of pieces, so the four bounds add up to $O(b)$. □

Remark 5.2. A positive area in (C2) does not by itself produce coincidences. A pair of pieces gives a genuine two-parameter family of coincidences inside each b, and then $E(a, b) \geq cb^2$ on an arithmetic progression of values of b, when in addition (1) the map ϕ sends some integer point to an integer point (integer compatibility), (2) the vectors $c_P(b, x)$ and $c_{P'}(b, \phi(b, x))$ are distinct at these points, not merely given by different affine formulas, and (3) a positive proportion of these integer points stays generic for both pieces as b grows (persistence of generic points under the strips and the conditions of the two pieces). Integrality of a rational affine map is a congruence condition, so the integer points mapped to integer points form a union of cosets of a full-rank sublattice of $\mathbb{Z}^3$; they may meet only some residue classes of b, which is why only a progression of values of b is obtained. Since the strips have bounded width, (3) should follow from (1) and a positive area, removing $O(b)$ points; we have not written this out. For $a = 1$ all three hold (Section 6); for $a = 2, \ldots, 5$ every candidate already fails (1).

The tests of every step of this reduction on exact data are described in Appendix A.

6. The certificate for $a \leq 5$

6.1. The computation. The program builds the templates, the chambers and the pieces, and for every generic piece the observation vector together with the list of every affine form whose sign it used. All arithmetic is exact (integers and rationals).

- *Chambers.* Starting from the lines $\lambda + h\mu \in \{0, \pm 1\}$ and the box, a chamber is cut whenever a form used in the deduction changes sign inside it. The only lines added were $q\mu = 1$ with $q \leq a$ (for $a = 3$: $\mu = 1, \frac{1}{2}, \frac{1}{3}$); no line coming from the order of exponents was needed. The final numbers of chambers are $10, 25, 47, 76, 112$ for $a = 1, \ldots, 5$. Each piece records, as its conditions of genericity, the union of the forms used by all the templates of its chamber. This only enlarges the exceptional set.

- *Templates*: complete sweep of the unit square as described in Section 5 (all u-intervals on the vertical line through the midpoint of every interval between critical values of v).
- *Observation vectors*: canonical sparse form, (a)–(c) of Section 5.
- *Signatures* are compared exactly (a hash only groups candidates; equality of the full signatures decides).
- *Integrality* in (C2). Write $\phi(x) = Mx+k$ with M, k rational and D the least common denominator of their entries. Then $\phi(x + Dy) - \phi(x) = DMy \in \mathbb{Z}^3$ for $y \in \mathbb{Z}^3$, so whether $\phi(x) \in \mathbb{Z}^3$ depends only on the class of $x \in \mathbb{Z}^3$ modulo D; equivalently, writing $\phi(x) = (Ax+k')/D$ with A, k' integral, one solves $Ax+k' \equiv 0 \pmod{D}$. The program runs over all D^3 integer representatives $x \in \{0, \dots, D-1\}^3$ of $(\mathbb{Z}/D\mathbb{Z})^3$, which is exhaustive.
- *Areas* are computed by exact polygon clipping.

6.2. Results. Table 2 summarises the output. Condition (C1) holds for $a \le 5$ (no piece of rank < 2).

a	templ.	chambers	generic pieces	rank < 2	signatures	same sign., c differs	integer-compatible	area > 0
1	4	10	20	0	7	9	8	8
2	14	25	238	0	83	6	0	0
3	36	47	1296	0	401	15	0	0
4	76	76	4712	0	1320	42	0	0
5	136	112	12 920	0	3452	45	0	0

Table 2. Output of the certificate. The last three columns concern (C2): pairs of pieces with the same signature and different vector formulas; those with an integer point mapped to an integer point; those with positive area. Running times: 57 s for $a = 4$, 167 s for $a = 5$. $a = 1$, the positive control (a quadratic family is found); $a = 2, \dots, 5$, the computer-assisted cases.

- $a = 1$ (*positive control*). The 8 integer-compatible pairs have area $\frac{1}{8}$ each, on two sheets: they are the exchange of [I, Lemma 7.2]. The map is $n' = b - m$, $m' = 2b - 2n - m$ (up to $0, \pm 1$ by template), with $c_P = (n - m - 1, n, n + m)$ and $c_{P'} = (2n-b-1, b-m, b)$, which is the map T of [I, Lemma 7.2] with $d = m$. The two sheets give $b^2/4$, the known growth of $E(1, b)$. One more pair is not integer ($n' = n - \frac{1}{2}$).
- $a = 2$. The 6 candidates have $n' = n - \frac{1}{3}$ or $n - \frac{2}{3}$, all in the chamber where every coordinate is positive; none is integer.
- $a = 3$. The 15 candidates all lie in the chamber $0 < \lambda - 3\mu$, $\lambda + 3\mu < 1$, with $(n', m') = (n - k/4, m)$: 6, 5 and 4 pairs for $k = 1, 2, 3$. None is integer.
- $a = 4$. The 42 candidates have $n' = n - k/5$: 16, 12, 8, 6 pairs for $k = 1, \dots, 4$. None is integer.
- $a = 5$. The 45 candidates have $m' = m$ and $n' = n - k/6$: 10, 12, 7, 6, 10 pairs for $k = 1, \dots, 5$. None is integer.

For $a = 2, \dots, 5$ all the candidates, and for $a = 1$ the one that is not integer-compatible, lie in the all-positive chamber, keep m and shift n by a non-integral multiple of $1/(a+1)$; Section 7 explains this. The split by k is not intrinsic (the direction of each pair depends on the order inside its signature group). For $a \ge 2$ the totals $6, 15, 42, 45$ are the numbers of

pairs of templates singled out by Theorem 7.6 (Corollary 7.9). We have no closed formula for them.

Figure 3 shows the non-integrality alternative for one actual candidate with $a = 2$.

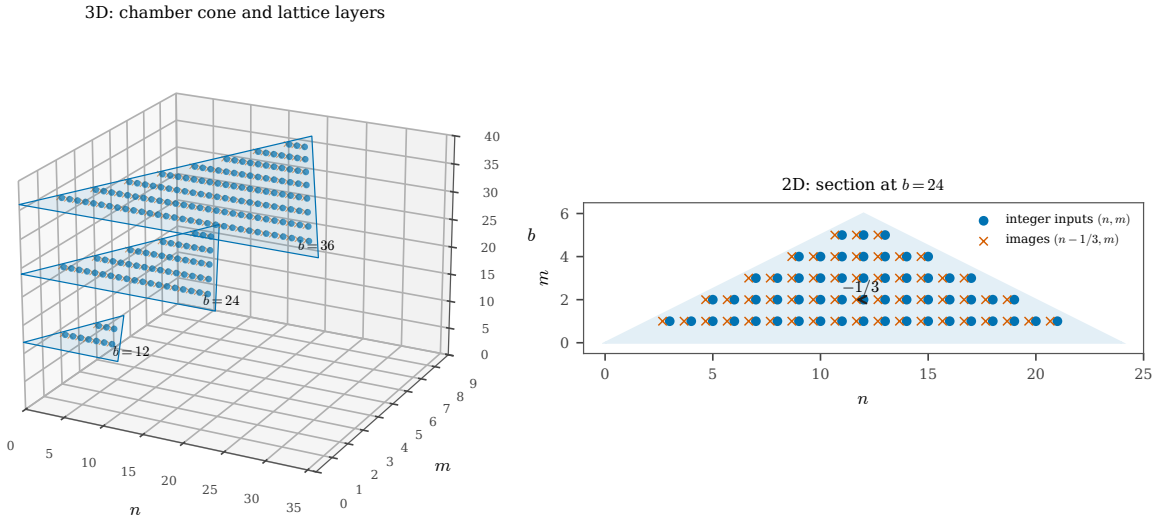

Figure 3. An equal-signature candidate with no integer solutions, from the portable $a = 2$ certificate. Two pieces share the all-positive triangular chamber and their observation vectors agree under $(b, n, m) \mapsto (b, n - \frac{1}{3}, m)$. Left (3D): the chamber cone at three values of b; blue dots are chamber lattice points and orange crosses their shifted images. Right (2D): the section $b = 24$ shows the two disjoint translates of the integer lattice. The drawing illustrates the integrality obstruction; points are selected by chamber inequalities only, without asserting piece-specific genericity.

Theorem 6.1 rigidity for $a \le 5$; computer-assisted. *Assume Theorem 5.1 (written proof) and the correctness of the program* `certificado_rigidez.py` *(not checked by a human referee; its output re-checked by independent scripts, §6.6; run unchanged for $a = 5$). For $a \in \{2, 3, 4, 5\}$ there are $K_a, b_0(a)$ such that $E(a, b) \le K_a b$ for $b \ge b_0(a)$; the vectors with at most one free coordinate are included in this count (they are $O(b)$). With the lower bound, $E(a, b) = \Theta(b)$. Consequently*

$$W(2,b) = \tfrac{22}{3}b^2 + O(b), \quad W(3,b) = \tfrac{73}{5}b^2 + O(b),$$
$$W(4,b) = \tfrac{1709}{70}b^2 + O(b), \quad W(5,b) = \tfrac{7657}{210}b^2 + O(b),$$

and for these a every constituent of the quasipolynomial $W(a, b)$ ($b \ge N_a$) has leading coefficient C_a, and every constituent of $E(a, b)$ has degree one. By Proposition 2.3, $W(a, b) = C_a b^2 + (D_a - \eta_a(b))b + O(1)$, where $\eta_a(b)$ is the slope of the constituent of $E(a, b)$.

Proof. The upper bound is Theorem 5.1 with the output above. The lower bound $E(a, b) \ge \sum_{h \le a} \#\{j \le b : \gcd(j, h) = 1\} + \sum_{2 \le j \le a} \lfloor b/j \rfloor$ is [I, Remark 7.6]. Then $W = R - E$ and $R(a, b) = C_a b^2 + O(b)$ in every residue class [I, Corollary 5.2]. □

a	b	$R(a,b)$	$\mathrm{Amb}(a,b)$	Amb/R	$b\,\mathrm{Amb}/R$
2	20	3143	260	0.0827	1.65
2	30	6913	381	0.0551	1.65
2	100	74370	1278	0.0172	1.72
3	20	6216	361	0.0581	1.16
3	30	13702	540	0.0394	1.18
3	80	94932	1451	0.0153	1.22
4	20	10359	496	0.0479	0.96
4	30	22860	734	0.0321	0.96
4	40	40246	975	0.0242	0.97
5	20	15440	631	0.0409	0.82
5	30	34095	903	0.0265	0.80

Table 3. Cells whose word is shared by another cell, from exact enumeration (all values of b in $2 \le b \le 100$ for $a = 2$, $3 \le b \le 80$ for $a = 3$, $4 \le b \le 40$ for $a = 4$, $5 \le b \le 30$ for $a = 5$; $E \le \mathrm{Amb} \le 2E$ holds in every one of the 240 cases). For $b \ge 2a$ the product $b\,\mathrm{Amb}/R$ stays between 1.14 and 1.73 ($a = 2$), 0.97 and 1.29 ($a = 3$), 0.90 and 1.12 ($a = 4$), 0.75 and 1.02 ($a = 5$). exact enumeration; the two shades separate the values of a.

For $a = 2$ this recovers the leading term of [I, Theorem 9.4] by another route. The value $C_5 = \frac{7657}{210}$ was recomputed with exact rationals. Figure 4 shows the theorem in the exact data.

Corollary 6.2 asymptotic faithfulness for $a \le 5$. *Under the hypotheses of Theorem 6.1, for $a \in \{2,3,4,5\}$ and $b \ge b_0(a)$, $\mathrm{Amb}(a,b) \le 2K_a b$ and*

$$\frac{\mathrm{Amb}(a,b)}{R(a,b)} = \Theta_a\Big(\frac{1}{b}\Big),$$

so the proportion of cells whose word is shared with another cell tends to 0, at the exact rate $1/b$.

Proof. By Lemma 3.6, $E \le \mathrm{Amb} \le 2E \le 2K_a b$, and $R(a,b) = C_a b^2 + O(b)$ with $C_a > 0$ [I, Corollary 5.2]. For the lower bound use $\mathrm{Amb} \ge E \ge c\,b$ (the lower bound of Theorem 6.1). □

That $R - W$ is $O(b)$ does not by itself say how many cells are ambiguous, since a fibre with s cells costs $s - 1$ in E but s in Amb; Lemma 3.6 shows the two differ by at most a factor 2, so no positive proportion of cells can hide behind a small E. The proportion is taken over the *set of cells*, with counting measure: it is not a statement about a measure on the parameters (Λ, θ) (the cells have very different areas), nor about the density of the integers n whose word is shared. It says nothing for a and b growing together (see the diagonal in Section 10). Table 3 gives the exact values.

6.3. Faithfulness is a statement about cells, not about integers. Corollary 6.2 says that a cell chosen with the counting measure is almost always faithful. Two other measures are natural: a measure on the parameters (Λ, θ), and counting the integers $2^{e-1} r^a s^b$ themselves. Table 4 compares the three at moderate b. The *integer sample* has odd primes $r, s \le 2000$ and all e with $\Lambda \le b + a\theta + 1$: a finite sample of primes and exponents, not a density of the

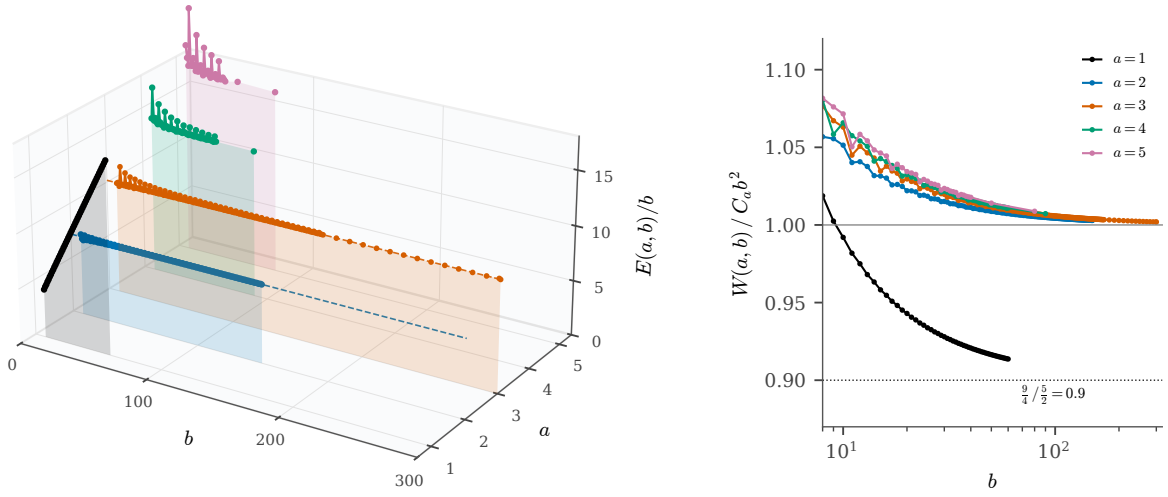

Figure 4. Rigidity in the exact data. Left: $E(a,b)/b$ over the (b,a)-plane, $8 \le b \le 300$, from exact enumeration of cells and words ($a=1$: $b \le 60$; $a=2$: $b \le 150$; $a=3$: $b \le 170$ and $b = 180, 190, \dots, 290, 299, 300$; $a=4$: $b \le 60$ and $b=90$; $a=5$: $b \le 40$ and $b = 50, 80$; isolated values are drawn as dots). For $a=1$ it grows like $b/4$, the quadratic exchange family of [I]; for $a=2,\dots,5$ it stays bounded, as Theorem 6.1 asserts. The theorem gives boundedness only, not a single limit slope (each residue class of b may have its own slope η_a); the data suggest a single slope. Dashed: the slopes $\frac{43}{6}$ ($a=2$, [I, Theorem 9.4]) and $\frac{61}{6}$ ($a=3$, Conjecture 9.8). Right: $W(a,b)/(C_a b^2)$, which tends to 1 for $a=2,\dots,5$ (Theorem 6.1) and to $\frac{9}{4}/\frac{5}{2} = 0.9$ for $a=1$.

integers ordered by size. The *section measure* takes θ uniform in $I = [0.14, 6.9]$, the range of the integer sample, and then Λ uniform in $(0, b + a\theta + 1)$, that is, the probability measure

$$d\mu_{a,b} = \frac{d\Lambda\, d\theta}{|I|\,(b + a\theta + 1)} \qquad \text{on } \{(\Lambda,\theta) : \theta \in I,\ 0 < \Lambda < b + a\theta + 1\};$$

it is not normalised planar Lebesgue measure on that domain, although on a fixed compact I the two differ by $O_I(1/b)$ in total variation, since the density ratio is $1 + O_I(1/b)$ uniformly. They do not agree: the section column is two to five times the cell column, and for $a=2$ and $a=5$ the integer column is larger than the cell column too (by a factor 1.7 and 2.3 at the larger b).

The excess has a mechanism, and it is elementary. For $a=3$ it is Proposition 9.1: the point $O_1(s)$ at $\theta = \frac{1}{4}$ and $O_2(s)$ at $\theta = b+1$ carry the same word. The argument works for every $a \ge 2$, and carries a quantitative statement.

Proposition 6.3 two numerations of one progression; written proof. *Let $a \ge 2$, $b \ge 1$, $K = (a+1)(b+1)$, $q = (a+1)b + a = K - 1$, and for $0 \le k \le K-2$ let $P_1(k) = (\frac{k+1/2}{a+1}, \frac{1}{a+1})$ and $P_2(k) = (k + \frac{1}{2}, b+1)$.*

(i) $P_1(k)$ and $P_2(k)$ are off the lines and $Z(P_1(k)) = Z(P_2(k)) = \{\{\min(K, N+k+1) : 0 \le N \le K-1\}\}$.

a	b	$R(a,b)$	cells, Amb $/R$	integer sample	section measure
2	20	3143	0.083	0.082	0.203
2	40	12150	0.042	0.071	0.136
3	20	6216	0.058	0.052	0.182
3	40	24109	0.030	0.032	0.121
5	20	15440	0.041	0.080	0.168
5	30	34095	0.026	0.059	0.141

Table 4. Proportion of ambiguous cells (the word is shared with another cell), counted three ways, rounded to three decimals (so 0.026 here is the 0.0265 of Table 3). *Integer sample*: the integers $2^{e-1}r^a s^b$ with odd primes $r \neq s \leq 2000$ (about 10^7 per row), computed with exact integer words; a word is ambiguous if its cell has a fibre of at least two cells. *Section measure*: the measure $\mu_{a,b}$ of the text; the choice of θ-range is a design choice, not a canonical measure. Numerical, not proved. The shading only separates the values of a.

(ii) *If $k \leq K-3$, then $P_1(k)$ and $P_2(k)$ lie in distinct cells, so these cells have the same word and form a fibre of at least two cells.*

(iii) *Let $0 \leq d < \frac{1}{a(a+1)}$, $u = a(a+1)d$ and $\theta = \frac{1}{a+1} + d$. For each $0 \leq k \leq K-3$ the cell of $P_1(k)$ contains a segment of the line $\{\theta\} \times (0, b+a\theta)$ of length $\frac{1-u}{a+1}$; these segments are pairwise disjoint. Hence the proportion $A_{a,b}(\theta)$ of that line lying in cells whose word is shared with another cell satisfies*

$$A_{a,b}(\theta) \;\geq\; \frac{(q-1)(1-u)}{q+u} \;\geq\; F_a(d) - \frac{1}{q}, \qquad F_a(d) = 1 - a(a+1)d,$$

and the last bound holds trivially also at $d = \frac{1}{a(a+1)}$.

(iv) *Let $0 \leq d < \frac{1}{2a(a+1)}$, $u = a(a+1)d$ and $\theta = \frac{1}{a+1} - d$. For each $0 \leq k \leq K-3$ the cell of $P_1(k)$ contains the segment $|\Lambda - \frac{k+1/2}{a+1}| < \frac{1}{2(a+1)} - ad$ of the line $\{\theta\} \times (0, b+a\theta)$; these segments are pairwise disjoint, and*

$$A_{a,b}(\theta) \;\geq\; \frac{(q-1)(1-2u)}{q-u}.$$

Proof. (i) At P_2, $\Lambda + h\theta = k + \frac{1}{2} + h(b+1)$ is a half-integer; at P_1 it is $\frac{2(k+h)+1}{2(a+1)}$, not an integer because its numerator is odd. By Lemma 2.2, Z is the multiset of the counts $\#\{p \in \mathcal{P} : p < p_0 + \Lambda\}$. At P_1 the positions $i\theta + j = \frac{i+(a+1)j}{a+1}$ $(0 \leq i \leq a,\ 0 \leq j \leq b)$ are $N/(a+1)$ with $N = i+(a+1)j$ running once through $0, \ldots, K-1$ (division with remainder), and $p < p_0 + \Lambda$ reads $N' < N + k + \frac{1}{2}$, that is $N' \leq N + k$: the count is $\min(K, N+k+1)$. At P_2 the positions $i(b+1)+j$ are again $N = 0, \ldots, K-1$ each once, and $\Lambda = k + \frac{1}{2}$ gives the same counts.

(ii) Put $x_1 = \Lambda_1 - a\theta_1 = \frac{2k+1-2a}{2(a+1)}$ and $x_2 = \Lambda_2 - a\theta_2 = k + \frac{1}{2} - a(b+1)$ (values of $\Lambda + h\theta$ at $h = -a$). Then $x_1 - x_2 = a\,\frac{K-k-3/2}{a+1} \geq \frac{3a}{2(a+1)} \geq 1$ for $a \geq 2$, $k \leq K-3$. Neither x_i is an integer, so the open interval (x_2, x_1) contains an integer ρ. Moreover $x_1 > -1$ and $x_1 \leq b + 1 - \frac{a+5/2}{a+1} < b$, so ρ can be taken in $[-b, b]$ (take $\rho = -b$ if $x_2 < -b$): the line $\ell_{-a,\rho}$ of the arrangement separates the two points.

(iii) Write $\Lambda = \frac{k+1/2}{a+1} + \lambda$. Then $\Lambda + h\theta = \frac{k+h+1/2}{a+1} + \lambda + hd$, and its floor stays $\lfloor (k+h)/(a+1) \rfloor$ (so c_h is that of $P_1(k)$, clipped or not) iff $-\frac{r_h+1/2}{a+1} < \lambda + hd < \frac{a+1/2-r_h}{a+1}$, where $r_h = (k+h) \bmod (a+1)$. Let $h_0 \in [-a, 0]$ be the h with $r_{h_0} = 0$; then $r_h = a$ for $h = h_0 + a \in [0, a]$. We claim that for $0 \le d \le \frac{1}{a(a+1)}$ the lower constraint of h_0 and the upper one of $h_0 + a$ imply all the others. For the lower constraint of h, $r_h/(a+1) + (h - h_0)d \ge 0$ is clear if $h \ge h_0$; if $h = h_0 - j$, $1 \le j \le a$, then $r_h = a + 1 - j \ge 1$ and $jd \le ad \le \frac{1}{a+1} \le \frac{r_h}{a+1}$. The upper constraints are symmetric (for $h = h_0 + a + j$, $r_h = j - 1$, and $\frac{a+1-j}{a+1} \ge jd$ in the same way). So the cell contains all λ in the interval from $-\frac{1}{2(a+1)} - h_0 d$ to $\frac{1}{2(a+1)} - (h_0 + a)d$, of length $\frac{1}{a+1} - ad$. For consecutive k the upper end of the interval of k is $\frac{k+1}{a+1} - (h_0(k) + a)d$ and the lower end of that of $k+1$ is $\frac{k+1}{a+1} - h_0(k+1)d$, with $h_0(k+1) \le h_0(k) + a$ (it is $h_0(k) - 1$, or $0 = h_0(k) + a$ if $h_0(k) = -a$); so they do not overlap. All lie in $(0, b + a\theta)$ since $h_0 \le 0$. The sum of the lengths is $(q-1)\frac{1-u}{a+1}$ with $u = a(a+1)d$, while the line has length $\frac{q+u}{a+1}$. Finally $\frac{(q-1)(1-u)}{q+u} \ge (1-u) - \frac{1}{q}$, since $\frac{(1-u)(1+u)}{q+u} \le \frac{1}{q}$. By (ii) every one of these segments lies in a cell of a fibre of at least two cells, and they are disjoint, so their lengths add. For $d = \frac{1}{a(a+1)}$ the right-hand side $F_a(d) - \frac{1}{q}$ is negative.

(iv) Now $\Lambda + h\theta = \frac{k+h+1/2}{a+1} + \lambda - hd$, and its floor stays $\lfloor (k+h)/(a+1) \rfloor$ iff $-\frac{r_h+1/2}{a+1} < \lambda - hd < \frac{a+1/2-r_h}{a+1}$. Since $0 \le r_h \le a$, both margins are at least $\frac{1}{2(a+1)}$, and $|hd| \le ad$; so every λ with $|\lambda| < \frac{1}{2(a+1)} - ad$ keeps every floor, and the point has the vector of $P_1(k)$. The segments have length $\frac{1-2u}{a+1} < \frac{1}{a+1}$ and centres $\frac{k+1/2}{a+1}$ spaced by $\frac{1}{a+1}$, so they are disjoint; they lie in $(0, b + a\theta)$, because $\frac{k}{a+1} + ad > 0$ and $\frac{k+1}{a+1} - ad \le \frac{K-2}{a+1} < b + a\theta = \frac{q-u}{a+1}$. There are $q - 1$ of them, each in a cell of a fibre of at least two cells by (ii), and the line has length $\frac{q-u}{a+1}$. □

The factor F_a is a tent of base $1/(a(a+1))$ to the right of $\theta = 1/(a+1)$, with half height at $d = \frac{1}{2a(a+1)}$. On the left, part (iv) proves the weaker bound $1 - 2a(a+1)d$ up to $d = \frac{1}{2a(a+1)}$. The exact computation with rational numbers ($a \le 5$) gives more, the same tent $1 - a(a+1)|d|$ for $|d| \le \frac{1}{2a(a+1)}$ and a larger value beyond; this is not proved for every a. (The full width at half height is then $\frac{1}{a(a+1)}$, not independent of a.) The cells of (iii) and (iv) are those of the points $P_1(k)$; the exact cell sections can be longer than the segments, so these are lower bounds for $A_{a,b}$, not its value.

Corollary 6.4 ambiguity of positive measure; written proof. *Let $a \ge 2$, $q = (a+1)b + a$, let $I \subset (0, \infty)$ be a compact interval, and let $P_{a,b}(I)$ be the mean over $\theta \in I$ of $A_{a,b}(\theta)$, the proportion of the line $\{\theta\} \times (0, b + a\theta)$ lying in cells whose word is shared with another cell.*

(a) If $I \supseteq [\frac{1}{a+1}, \frac{1}{a}]$, then for every $b \ge 1$

$$P_{a,b}(I) \;\ge\; \frac{q-1}{a(a+1)|I|}\Big[(q+1)\log\Big(1 + \frac{1}{q}\Big) - 1\Big],$$

so

$$\liminf_{b \to \infty} P_{a,b}(I) \;\ge\; \frac{1}{2a(a+1)\,|I|}.$$

(b) If $I \supseteq [\frac{2a-1}{2a(a+1)}, \frac{1}{a}]$, then for every $b \geq 1$

$$P_{a,b}(I) \;\geq\; \frac{q-1}{a(a+1)|I|}\Big[(q+1)\log\Big(1+\frac{1}{q}\Big) - (2q-1)\log\Big(1+\frac{1}{2q-1}\Big)\Big],$$

so

$$\boxed{\liminf_{b\to\infty} P_{a,b}(I) \;\geq\; \frac{3}{4a(a+1)\,|I|} > 0.}$$

Proof. Write $u = a(a+1)d$, so $d\theta = du/(a(a+1))$. By Proposition 6.3(iii), the right band $\theta \in [\frac{1}{a+1}, \frac{1}{a}]$ contributes at least $\frac{1}{a(a+1)}\int_0^1 \frac{(q-1)(1-u)}{q+u}\,du = \frac{q-1}{a(a+1)}\int_0^1\Big(\frac{q+1}{q+u} - 1\Big)du = \frac{q-1}{a(a+1)}\big[(q+1)\log\frac{q+1}{q} - 1\big]$ to $|I|\,P_{a,b}(I)$, and by (iv) the left band $\theta \in [\frac{2a-1}{2a(a+1)}, \frac{1}{a+1}]$ contributes at least $\frac{1}{a(a+1)}\int_0^{1/2}\frac{(q-1)(1-2u)}{q-u}\,du = \frac{q-1}{a(a+1)}\big[1-(2q-1)\log\frac{2q}{2q-1}\big]$; the rest of I contributes ≥ 0. As $q\to\infty$, $(q-1)[(q+1)\log(1+\frac{1}{q})-1] \to \frac{1}{2}$ and $(q-1)[1-(2q-1)\log(1+\frac{1}{2q-1})] \to \frac{1}{4}$. □

Figure 5 shows the tent: the ridge at $\theta = \frac{1}{a+1}$ has height exactly $1-\frac{1}{q}$ and rises with b, while the proportion of ambiguous cells falls like $1/b$.

For $a = 2$ and $I = [0.14, 6.9]$ the one-sided bound (a) is 0.0123 and the two-sided bound (b) is 0.0185; for $a = 3$, 0.0062 and 0.0092.

This does not contradict Corollary 6.2: the cells of the band are few ($O(b)$) but they are long strips in Λ, and the integer sample and the section measure see lengths. It also means that the proportion over cells and the proportion over θ-uniform sections are different invariants. What we do *not* know is the limit of $P_{a,b}(I)$. The exact means over $\theta \in [0.14, 6.9]$ (exact segmentation of the line at each θ, 600 equally spaced values of θ; the result is a quadrature, not a closed form) are

b	8	20	40	80	120	160	320
$a=2$	0.307	0.211	0.139	0.094	0.077	0.068	0.053
$a=3$	0.257	0.188	0.125	0.078	–	0.050	0.034

They are above the bound of Corollary 6.4 and still decreasing at the largest b computed. The successive drops for $b = 40 \to 80 \to 160 \to 320$ ($0.0453, 0.0264, 0.0142$ for $a = 2$; $0.0463, 0.0283, 0.0158$ for $a = 3$) roughly halve each time, which is what a limit $L + c/b$ would give, with $L \approx 0.04$ ($a = 2$) and $L \approx 0.02$ ($a = 3$); with three values this is an extrapolation, not a result, and a slower decay cannot be excluded.

6.4. Words shared by two shapes with the same number of divisors. So far the shape (a, b) was fixed. Two shapes (a, b) and (a', b') with $(a+1)(b+1) = (a'+1)(b'+1) = K$ give words of the same semilength K. Write $L(a, b)$ for the set of words of the integers $2^{e-1}r^a s^b$ (so $|L(a, b)| = W(a, b)$) and $L(K-1)$ for the set of words of the integers $2^{e-1}p^{K-1}$, which has K elements [I, Corollary 3.3(a)]. The shape $(K-1)$ is the only common coarsening of two distinct two-prime shapes with K divisors, and neither of these refines the other; by the maximal refinement theorem [I, Theorem 3.1], $L(K-1) \subseteq L(a, b) \cap L(a', b')$. In the integers up to 10^6 the intersection is larger: $287496 = 2^3 \cdot 3^3 \cdot 11^3$ (shape $(3, 3)$) and $122472 = 2^3 \cdot 3^7 \cdot 7$ (shape $(7, 1)$) have the same word

$$\texttt{((((()(()(())(())(())(())())())))),}$$

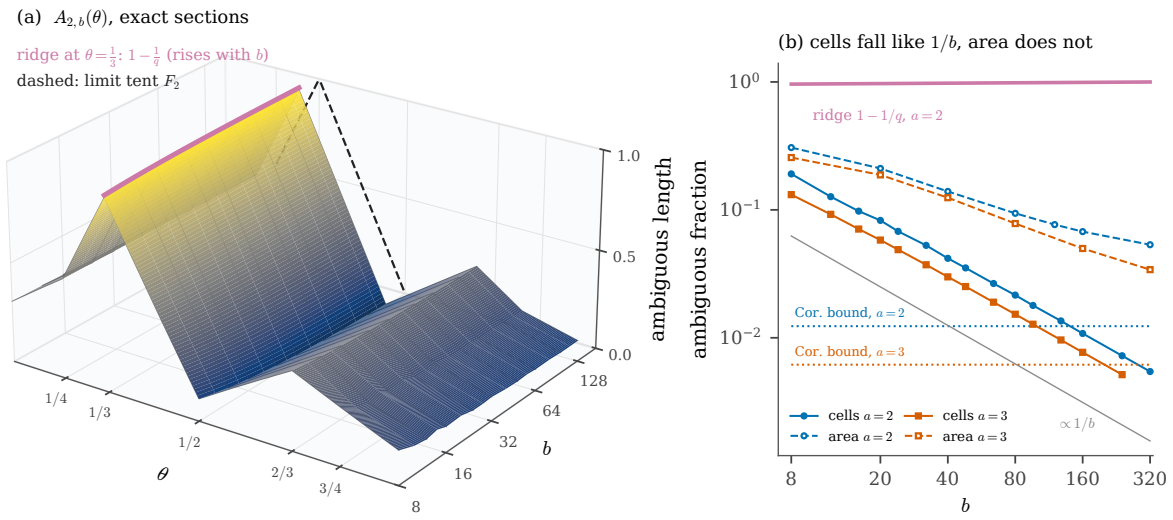

Figure 5. A non-shrinking tent-shaped lower bound for the ambiguity profile. (a) $A_{2,b}(\theta)$, the proportion of the line $\{\theta\}\times(0, b+2\theta)$ lying in cells whose word is shared with another cell, for $b = 8, \dots, 128$ (exact sections, $\theta = k/720$). The ridge at $\theta = \frac{1}{3}$ (reddish purple) has height $1 - \frac{1}{q}$, $q = 3b + 2$, and rises with b; the dashed curve on the back wall is the limiting lower bound F_2 on the right (Proposition 6.3(iii)) and the computed profile on the left (where (iv) proves the weaker $1 - 12d$). The profile itself can be higher. Away from the ridge the surface does not vanish either: at $b = 128$ it is about 0.40 at $\theta = 0.2$ and 0.34 at $\theta = \frac{2}{3}$ (against 0.40 and 0.39 at $b = 8$); among the observed values it is near 0 only around $\theta = \frac{1}{2}$ (0.111, 0.030, 0.008 at $b = 8, 32, 128$). (b) The same arrangement counted two ways, $a = 2$ (circles) and $a = 3$ (squares): the proportion of ambiguous cells Amb/R (solid) falls like $1/b$; the mean $P_{a,b}(I)$ of the proportion of the line over $\theta \in I = [0.14, 6.9]$ (dashed) falls more slowly and stays above the one-sided lower bound $1/(2a(a+1)|I|)$ of Corollary 6.4(a) (dotted). Numerical values exact up to the sampling of θ; the limit of $P_{a,b}(I)$ is not known.

which is not in $L(15)$. Table 5 gives the exact intersections for the 32 pairs of two-prime shapes with $12 \le K \le 48$.

The common coarsening explains between 21 % and 48 % of the shared words, so the intersection is about two to five times K. The rest is explained by [I] as well. The proof of [I, Theorem 3.1] embeds each shape, by a mixed-radix substitution, in the parameter space of the finest shape with K divisors (for $K = 12$, the shape $(2, 1, 1)$), and the arrangement of the coarser shape is the trace of the finer one, so every cell of (a, b) lies in a single cell of the finer arrangement and carries its word. If one fine cell meets both embedded shapes, its word lies in both languages: this accounts for 2143 of the 2852 shared words outside $L(K - 1)$ (all radix embeddings and all permutations of coordinates with equal exponents tried, on rational witnesses of the cells). The other 709 are reached through different fine cells with the same word, that is, through fibres of the finer arrangement; their words repeat a few types, such as $()^k(())()^k$, which is shared by every pair with $K = 2k + 2$ even, and nested words. The criterion "one fine cell" is sufficient, not necessary, so the split between the last two columns depends on the embeddings tried. A second code (a rational segment between the two witnesses that crosses no line of the finer arrangement, with a constant word at 41

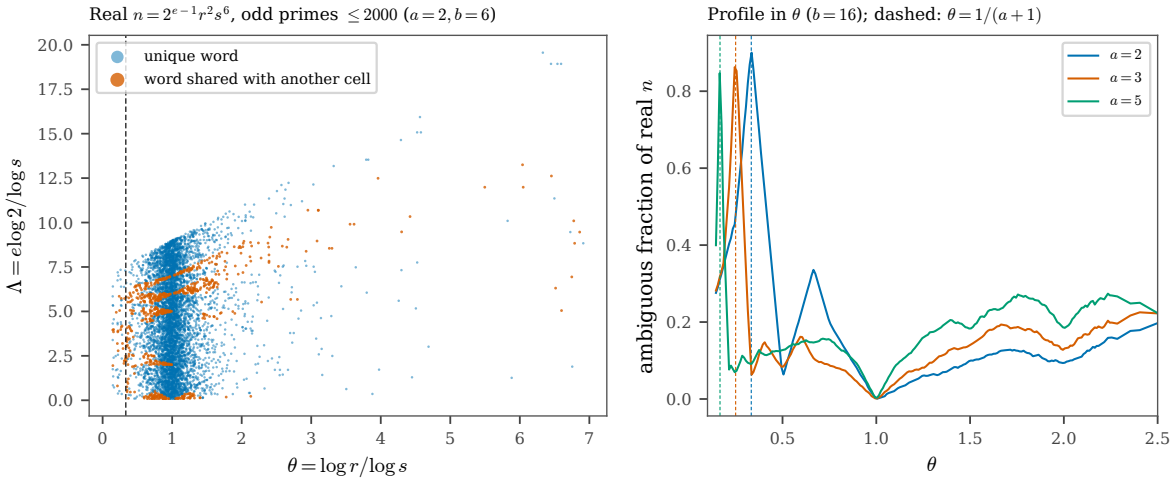

Figure 6. Integers and ambiguity. Left: 6000 integers $2^{e-1}r^2s^6$ with odd primes ≤ 2000 in the plane (θ, Λ); orange: the word is shared with another cell. Right: proportion of ambiguous integers as a function of θ ($b = 16$, $a = 2, 3, 5$); the dashed lines are $\theta = 1/(a+1)$. The sampled peaks at $1/(a+1)$ align with the cell-level ridge of Proposition 6.3 (their height is lower than $1 - 1/q$; the sample includes the trivial strip and a discrete set of e, which we did not analyse further); the minimum at $\theta = 1$ (cells meeting $\theta = 1$ have no partner for $a \leq 5$, $b \leq 16$, checked) and the second peak at $\theta \approx \frac{2}{3}$ for $a = 2$ are observed, not proved.

points along it) gives the same counts 15, 17, 27 for $K = 12, 16, 18$; on 40 of the 68 remaining words of the pair $(7, 2)$, $(5, 3)$ it confirms 31. The integers $n \leq 10^6$, which also have shapes with three and four primes, contain 127 pairs (word, two incomparable shapes) whose word lies in no language of a common coarsening; with the same criterion applied to at most 60 integers per shape (floating point with a watched margin), 114 of 129 such pairs (these 127 and two more) lie in one fine cell, and for the other 15 the criterion is inconclusive.

The counts are new as far as we know (we did not search the literature for them beyond [I] and the OEIS), and checked with two codes; the mechanism is not: it is the refinement theorem, the convexity of cells and the fibres of [I].

6.5. Pairs with different signatures, and deep cells. Condition (C2) concerns equal signatures; pairs of pieces with different signatures only cost $O(b)$ by case (iii). A second program, `pares_rigidez.py`, solves them exactly. For $a = 3$ there are 27 232 pairs of pieces *of equal length*; 6879 of them have the same signature, and the remaining 20 353 pairs *with different signatures* split into 16 527 incompatible ones, 3028 that meet in dimension 2 with b varying and 798 in dimension 1. For $a = 4$ the numbers are 170 308 pairs of equal length, 42 157 with the same signature, and 128 151 with different signatures: 106 277 incompatible, 18 126 in dimension 2 and 3748 in dimension 1. For $a = 5$: 828 008 pairs of equal length, 171 762 with the same signature, and 656 246 with different signatures: 571 974 incompatible, 72 238 in dimension 2 and 12 034 in dimension 1 (925 s). In every compatible case the asymptotic location of the solutions touches the boundary of the chambers, and none lies strictly outside; the consequence used below is that, for every coincidence between two generic cells with different signatures, *at least one* of the two cells is at distance $O(1)$ from a chamber boundary, hence from $\mathcal{L}_{\text{deg}}$. *Independent check:* for $a \leq 5$ a second exact solver reproduced the pair-by-pair classification, including all 828 008 pairs for $a = 5$, with no discrepancies.

K	pairs	$\lvert L(a,b)\cap L(a',b')\rvert$	share of $L(K-1)$ (%)	one fine cell	no common cell found
12	1	28	43	15	1
16	1	33	48	17	0
18	1	47	38	27	2
20	1	56	36	32	4
24	3	68–92	26–35	133	30
28	1	94	30	60	6
30	3	91–113	27–33	152	54
32	1	101	32	68	1
36	6	83–151	24–43	336	133
40	3	130–189	21–31	255	84
42	3	139–169	25–30	230	95
44	1	169	26	110	15
45	1	188	24	90	53
48	6	166–212	23–29	618	231
total	32	3979	1127 words	2143	709

Table 5. Words shared by two distinct two-prime shapes with K divisors (all pairs, $12 \le K \le 48$; ranges over the pairs with the same K). *Share of* $L(K-1)$: $K/\lvert L(a,b)\cap L(a',b')\rvert$, the part explained by the common coarsening. Of the remaining words, *one fine cell*: one cell of the arrangement of the finest shape with K divisors meets both embedded shapes; *no common cell found*: the word is not explained by a common fine cell under the embeddings tested, so the two shapes reach it, as far as tested, through different fine cells. Exact computation in rational arithmetic; a numerical observation, not a theorem. The shading only alternates the values of K.

For $a = 5$ all 84 272 compatible pairs with different signatures outside the open chambers touch their closures. An integer contrast for $a = 2, 3$ up to $b = 601$ found no coincidence between two generic cells.

For $a = 3$, the boundaries of the final chambers are the lines $\lambda + h\mu \in \{0, \pm 1\}$ ($|h| \le 3$), $\mu = 1, \frac{1}{2}, \frac{1}{3}$, $\lambda = 0$, $\mu = 0$ and the box. In the original coordinates these are the lines

$$\mathcal{L}_{\mathrm{deg}} = \{\Lambda + h\theta \in \{0, b, -b\} : |h| \le 3\} \cup \{k\theta = b : k = 1, 2, 3\} \cup \{\theta = 0\},$$

25 distinct lines ($\Lambda = 0$ is the case $h = 0$ of the first set), of which 18 meet the closed quadrant in more than a point (the lines $\Lambda + h\theta = -b$ with $h \ge 0$ miss it, and $\Lambda + h\theta = 0$ with $h \ge 1$ meet it only at the origin): the lines along which two break points of the macroscopic profile of the word meet. The strips of the nine forms whose zero lines are not in $\mathcal{L}_{\mathrm{deg}}$ meet their chamber only within $O(1)$ of a vertex or an edge, because these lines do not cross the interior of the chamber.

Corollary 6.5 no coincidence between deep cells; computer-assisted. *Let $a = 3$. There is a constant C such that, for b large, two distinct realisable vectors with at least two free coordinates whose cells both contain a point at distance $> C$ from $\mathcal{L}_{\mathrm{deg}}$ have different words. The same holds for $a = 4$ and $a = 5$ with the corresponding sets of lines.*

Sketch. A cell with a point at distance $> C$ from $\mathcal{L}_{\mathrm{deg}}$ has a generic representative, for C large. Two generic vectors with the same word fall into case (iii) or (iv) of the proof of Theorem 5.1. Case (iv) has no integer solutions (Table 2). In case (iii), by the output of

`pares_rigidez.py`, at least one of the two cells lies within $O(1)$ of a chamber boundary, hence of $\mathcal{L}_{\deg}$; this is enough, since both were assumed deep. That output was reproduced pair by pair by an independent solver ($a \le 5$), with an integer contrast for $a = 3$ up to $b = 601$ finding no generic coincidence. □

This is the easy half of Lemma D: it says nothing about a deep cell and a cell near $\mathcal{L}_{\deg}$, generic or not. That half is Theorem 8.1 below, for $a = 3$. Cells with at most one free coordinate can be far from $\mathcal{L}_{\deg}$ (for instance the cell of $(-b-1, -b-1, -b-1, s, b, b, b)$ contains points with $\theta \to \infty$), and they do share words with cells near $\mathcal{L}_{\deg}$ (Remark 8.2), which is why they are excluded.

6.6. Checks.

- *Regression.* For every generic piece, at one integer point near the centre of its chamber (the first $b \in \{400, 2000, 10000\}$, resp. $\{400, 2000\}$ for $a = 4$, at which the piece is valid), the vector and K_c computed directly from $(\Lambda, \theta) = (n + u, m + v)$ agree with the formulas: 20, 238, 1296, 4712, 12 920 pieces, no failure. One point per piece is only a control.
- *Multi-point regression and affine generation.* For every generic piece and each $b \in \{300, 1000, 5000\}$, six valid integer points (half of them within distance 3 to $\sqrt{b}$ of a vertex of the chamber); c and K_c recomputed directly from $(\Lambda, \theta) = (n + u, m + v)$. Result: 360, 4284, 23 328, 84 816 and 232 543 points for $a = 1, \dots, 5$, no failure, no piece without points, and in *every* piece the valid points have rank 4 in the rows $(1, b, n, m)$, that is, they contain four affinely independent points of $\mathbb{Z}^3$. Negative control: with a corrupted coefficient of $\mathcal{Q}/(1 - z)$ in the construction, all 2856 points of $a = 2$ fail. Consequently, *if* the true K_c is given on each piece by the same kind of affine sparse representation (exponents and coefficients affine in (b, n, m), fixed order), the formulas of every piece are proved, not only controlled: two affine maps that agree on four affinely independent points agree everywhere. That premise is the *affinity argument*: in a piece the clipping state, the order of the break points and the slopes of every row are fixed, so each piece of each row contributes monomials with affine exponents and coefficients. It is written by hand (a written proof, not checked by a human referee); it is the remaining condition for this step.
- *Independent checks of the output* (separate scripts that import the program without modifying it). The vector formula agrees with direct computation at *all* generic integer points of the box for $a = 3$, $b = 60, 120, 200$ (3.8×10^6 points at $b = 200$), and K_c at 12 points per piece and value of b; for $a = 4$, 37 696 random points in all 4712 pieces with $b \in \{300, 1000, 5000\}$. The templates and the box are complete ($a \le 3$ and $a = 5$, $b = 7, 20, 40$: the triples (n, m, ε) give exactly the realisable vectors of an independent enumerator by critical points). Signatures are invariant under reparametrisation and under reordering of the input, and grouping by signature agrees with grouping by an independent rank computation ($a \le 5$). The integrality test agrees with brute force on 300 random maps ($a \le 5$). Every non-constant coefficient of an observation vector vanishes only on a line that does not cross its chamber.
- *Negative controls.* With one coefficient of $\mathcal{Q}/(1 - z)$ corrupted, the regression fails in all 238 pieces of $a = 2$. The integrality test returns false for $n - \frac{1}{4}$ and true for $n + (b + m)/2$.
- *Contrast with exact enumeration.* For $a = 1$ the method finds the exchange in the real data (pairs of generic cells with equal signature growing quadratically: $64, \dots, 132$ for

$b = 23, \dots, 30$). For $a = 2$ ($b \le 30$), $a = 3$ ($b \le 30$ and $b = 40, 60, 80, 120, 150, 300$), $a = 4$ ($b = 30, 60, 90$) and $a = 5$ ($b = 6, \dots, 16$ and $30, 50, 80$), every real coincidence is exceptional on both sides; for $a \le 4$ the fraction of generic cells grows with b, so this is not vacuous. For $a = 5$ the contrast is weak: no cell has a valid generic representative for $b \le 30$ (the check there cannot fail), and only 20 of 3010 and 98 of 3036 sampled cells at $b = 50$ and $b = 80$ do.

What the computer-assisted step trusts. (i) The written reduction of Section 5, which has not been checked by a human referee; (ii) the correctness of the program, which was not formally verified. For $a = 5$ its portable output was fully accepted by a separate Excovery verifier, and SageMath independently checked the rank of all 12 920 generic pieces and the 45 declared non-integral pairs; neither check proves completeness of the written reduction. (iii) the affinity argument above, if the formulas of the pieces are to be proved rather than controlled (proved by hand for the all-positive chamber, Proposition 7.4); (iv) the results of [I] used: realisation by primes, the Presburger encoding and the formula for $R(a, b)$.

7. The denominator $a + 1$

All the candidates of Table 2 for $a = 2, \dots, 5$, and the one candidate of $a = 1$ that is not integer-compatible ($1 + 6 + 15 + 42 + 45 = 109$ pairs of pieces; with the eight exchange pairs of $a = 1$ the table has 117), have the same shape: both pieces lie in the all-positive chamber, $m' = m$ and $n' = n + \kappa$ with $\kappa \in \frac{1}{a+1}\mathbb{Z} \setminus \mathbb{Z}$. The denominator has a one-line explanation. For a template ε put

$$S_-(\varepsilon) = \sum_{h=-a}^{0} \varepsilon_h.$$

Proposition 7.1. *For every realisable c, the smallest element of $Z(c)$ is $\tau(c) = H_0(0) = \sum_{h=-a}^{0} f(c_h)$. If $c = g(n, m, \varepsilon)$ and $-1 \le n + hm + \varepsilon_h \le b$ for $-a \le h \le 0$, then*

$$\tau(c) = (a+1)(n+1) - \frac{a(a+1)}{2}\, m + S_-(\varepsilon).$$

Consequently, if $c = g(n, m, \varepsilon)$ and $d = g(n', m, \varepsilon')$ both satisfy this condition and $Z(c) = Z(d)$, then

$$(a+1)(n' - n) = S_-(\varepsilon) - S_-(\varepsilon'),$$

and in particular $S_-(\varepsilon) \equiv S_-(\varepsilon') \pmod{a+1}$.

Proof. c is non-decreasing in h (floors of the increasing function $h \mapsto \Lambda + h\theta$, then clipped) and f is non-decreasing, so $H_i(j) = \sum_{l=0}^{a} f(c_{i-l} + j) \ge \sum_{l=0}^{a} f(c_{-l}) = H_0(0)$ for all $i, j \ge 0$. Under the condition, clipping leaves $c_h = n + hm + \varepsilon_h$ unchanged and $f(c_h) = c_h + 1$ for $-a \le h \le 0$; summing, and using $\sum_{h=-a}^{0} h = -a(a+1)/2$, gives the formula. Equal words have equal multisets, hence equal minima. □

In the all-positive chamber the condition holds at every generic point. The computation gives:

Corollary 7.2 computer-assisted. *In each of the 109 candidates of Table 2 ($a \le 5$) the shift found by the certificate is exactly $\kappa = (S_-(\varepsilon) - S_-(\varepsilon'))/(a+1)$, and $S_-(\varepsilon) \not\equiv S_-(\varepsilon')$*

(mod $a+1$). *So in every case the obstruction to integrality is this congruence. The residues of $S_-(\varepsilon)-S_-(\varepsilon')$ modulo $a+1$ that occur are all the nonzero ones, for each a.*

Remark 7.3. Proposition 7.1 explains the denominator, not the phenomenon. Three facts would remove case (iv) of the reduction for every a (Question 10.1):

(1) every candidate of (C2) lies in the all-positive chamber;
(2) every candidate keeps m;
(3) $S_-(\varepsilon)\not\equiv S_-(\varepsilon')$ whenever two templates there have the same signature and different vector formulas.

Proposition 7.1 proves none of them. Theorem 7.6 below proves the second for every $a\ge 2$, for the candidates inside the all-positive chamber, and Theorem 7.10 proves the third for every a. The first is checked for $a\le 6$ (Conjecture 7.12).

7.1. The all-positive chamber. Let C_+ be the chamber $0<\lambda+h\mu<1$ ($|h|\le a$). For $a\le 6$ the program finds it as a single chamber. At its generic points every coordinate is free and $0\le c_h\le b-1$. For a template put

$$\eta_h=\varepsilon_{h+1}-\varepsilon_h\in\{0,1\}\quad(-a\le h<a),\qquad W_i=\{i-a,\dots,i\},\qquad T_r=\frac{r(r+1)}{2},$$

$$\kappa_{i,r}(\varepsilon)=\sum_{t=0}^{r-1}(t+1)\,\eta_{i-a+t}\qquad(0\le i\le a,\ 1\le r\le a),$$

so that $0\le\kappa_{i,r}\le T_r$. Here $\eta_h\in\{0,1\}$ because $v\in(0,1)$, and $\varepsilon_0=\lfloor u\rfloor=0$. Further notation:

- $\mathcal{R}_s=\mathcal{Q}/(1-z^s)$ for $s\le a+1$ (a polynomial; not to be confused with the count $R(a,b)$);
- $B_r=z^{r+1}\mathcal{R}_{r+1}-z^r\mathcal{R}_r$ for $1\le r\le a$, a nonzero polynomial;
- $G_c(z)=\sum_{v\in Z(c)}z^{T-v}$, with $T=(a+1)(b+1)$, the generating polynomial of the *deficits* $T-v$.

Since $\mathcal{Q}(z)=-z^{\deg\mathcal{Q}}\mathcal{Q}(1/z)$ (every Φ_d with $d\ge 2$ is palindromic, and $\Phi_1=z-1$), $K_c(z)=-z^{T+\deg\mathcal{Q}}(\mathcal{Q}G_c)(1/z)$. So $Z(c)=Z(d)$ if and only if $G_c=G_d$, and the exponents of K_c are those of $\mathcal{Q}G_c$ reflected.

Proposition 7.4 levels; written proof. *Let c be non-decreasing with $0\le c_h\le b-1$ for all h. Put $S=\sum_{h=-a}^{0}c_h$ and $w_{i,r}=\sum_{t=0}^{r-1}(t+1)(c_{i-a+t+1}-c_{i-a+t})$. Then*

$$\mathcal{Q}G_c=(S+a+1)\mathcal{Q}+(a+1)z\mathcal{R}_1+\sum_{r=1}^{a}\Big(\sum_{i=0}^{a}z^{w_{i,r}}\Big)B_r-\sum_{i=0}^{a}z^{\,T-\sum_{h\in W_i}c_h}\mathcal{R}_{a+1}.$$

If $c=g(n,m,\varepsilon)$ with no coordinate clipped, then

$$w_{i,r}=T_rm+\kappa_{i,r}(\varepsilon),\qquad T-\sum_{h\in W_i}c_h=(a+1)(b+1-n)-(a+1)\big(i-\tfrac{a}{2}\big)m-\sum_{h\in W_i}\varepsilon_h.$$

Proof. For $c_h\ge 0$ and $0\le j\le b$, $f(c_h+j)=b+1-(e_h-j)_+$ with $e_h=b-c_h$. So the deficits of row i are $T-H_i(j)=\sum_{h\in W_i}(e_h-j)_+$.

Put $y=e_{i-a}-j$ and $q_r=c_{i-a+r}-c_{i-a}$ ($0\le r\le a$). Then $e_h-j=y-q_r$ for $h=i-a+r$, and as j runs over $[0,b]$, y runs over $[-c_{i-a},Y]$ with $Y=b-c_{i-a}$. The deficit $\Psi(y)=\sum_{r=0}^{a}(y-q_r)_+$:

- vanishes for $y\le 0$, which gives $c_{i-a}+1$ values;

- has slope r on $[q_{r-1}, q_r]$;
- has slope $a+1$ on $[q_a, Y]$, where $q_a \le Y$ because $c_i \le b$.

On each segment the sum is a geometric progression: $\sum_{y=q_{r-1}+1}^{q_r} z^{\Psi(y)} = (z^{\Psi(q_{r-1})+r} - z^{\Psi(q_r)+r})/(1-z^r)$, also when $q_{r-1} = q_r$; the same holds on $[q_a, Y]$ with $r = a+1$.

Multiply by $\mathcal{Q}$ and collect the terms at each kink:

- $\Psi(0) = 0$ gives $z\mathcal{R}_1$;
- $\Psi(q_r) = \sum_{r'<r}(q_r - q_{r'}) = w_{i,r}$ gives $z^{w_{i,r}} B_r$;
- $\Psi(Y) + a + 1 = (a+1)(Y+1) - \sum_r q_r = T - \sum_{h\in W_i} c_h$ gives the last term.

The zero deficits of all rows give $(S+a+1)\mathcal{Q}$. For $c = g(n,m,\varepsilon)$ the gaps are $c_{h+1} - c_h = m + \eta_h$ and $\sum_{h\in W_i} h = (a+1)(i - \frac{a}{2})$. □

So in C_+ the polynomial $\mathcal{Q}G_c$ is made of clusters. One sits at 0, one at $T_r m$ for each level r, and $a+1$ more, which carry n, sit above all of them. The clusters at the levels depend on the template only through the multisets of the $\kappa_{i,r}$; the $a+1$ upper clusters depend on it through the window sums $\sum_{h\in W_i} \varepsilon_h$, and the cluster at 0 through $S_-(\varepsilon)$.

Lemma 7.5 formal sums and rank two; written proof. *Let $a \ge 1$, let P be a generic piece of C_+ with template ε, and let F_ε be the right-hand side of Proposition 7.4 for $c = g(n,m,\varepsilon)$, read as a formal sum of terms κz^e whose exponents and coefficients are affine forms in $x = (b,n,m)$, with identical exponent forms merged and zero coefficients deleted.*

(a) After the reflection $e \mapsto T + \deg \mathcal{Q} - e$ of the exponents (and the change of sign), the terms of F_ε and the entries of V_P correspond one to one as affine forms.

(b) V_P has (n,m)-rank 2. So (C1) holds for every piece of C_+, for every $a \ge 1$.

Proof. (a) At an integer point of P that avoids finitely many planes (where two distinct exponent forms of F_ε agree or a coefficient form vanishes), the value of F_ε is the sparse form of $\mathcal{Q}G_c$, and reflected it is the list $V_P(x)$. So at each such point some bijection between the terms of F_ε and the entries of V_P matches exponents and coefficients in value. If a bijection pairs two different affine forms, it works only on the plane where they agree. There are finitely many bijections, and the points in question are not contained in finitely many planes (for b large they include a positive proportion of the integer points of bC_P). So some bijection pairs only identical forms.

(b) The cluster of F_ε at level $r = 1$ is $z^m P_1 B_1$ with $P_1 = \sum_i z^{\kappa_{i,1}} \ne 0$ and $B_1 \ne 0$, so F_ε has a term whose exponent has linear part m. The upper clusters are monomials times $-\mathcal{R}_{a+1} \ne 0$, so it has a term whose exponent has linear part $(a+1)(b-n) - (a+1)(i-\frac{a}{2})m$. By (a), V_P has two exponent entries with (n,m)-parts $(0,-1)$ and $(a+1, (a+1)(i-\frac{a}{2}))$, whose determinant is $a+1 \ne 0$. □

Theorem 7.6 pairs in the all-positive chamber; written proof. *Let $a \ge 2$. Let P, P' be generic pieces of C_+ with templates $\varepsilon, \varepsilon'$ and the same signature, and let ϕ be the map of (C2). If $c_{P'} \circ \phi \ne c_P$, then*

$$m' = m, \qquad n' = n + \nu, \qquad (a+1)\,\nu = S_-(\varepsilon) - S_-(\varepsilon'),$$

and the templates satisfy

(N2) $\sum_{h\in W_i}(\varepsilon_h - \varepsilon'_h) = S_-(\varepsilon) - S_-(\varepsilon')$ for $0 \le i \le a$;

(N3) for $1 \le r \le a$, the multisets $\{\kappa_{i,r}(\varepsilon)\}_{0\le i\le a}$ and $\{\kappa_{i,r}(\varepsilon')\}_{0\le i\le a}$ are equal.

Conversely, two templates $\varepsilon \neq \varepsilon'$ that satisfy (N2) and (N3) give pieces $P = (C, \varepsilon)$, $P' = (C, \varepsilon')$ of any chamber $C \subseteq C_+$ (for $a \leq 6$, $C = C_+$) with the same signature, related by this ϕ, and with $c_{P'} \circ \phi \neq c_P$.

Proof. Formal sums. With F_ε as in Lemma 7.5, part (a) of the lemma shows that $V_P = V_{P'} \circ \phi$ means $F_\varepsilon = F_{\varepsilon'} \circ \phi$ as formal sums.

Linear parts. Group the terms by the linear part of the exponent (the exponent form minus its constant), and call the sum of the terms with a given linear part its *cluster*. F_ε has exactly $2a+2$ nonzero clusters:

- linear part 0, with cluster $(S+a+1)\mathcal{Q}+(a+1)z\mathcal{R}_1$, where $S = (a+1)n - T_a m + S_-(\varepsilon)$;
- linear part $T_r m$ for $1 \leq r \leq a$, with cluster $z^{T_r m} P_r B_r$, where $P_r = \sum_i z^{\kappa_{i,r}(\varepsilon)}$ has positive coefficients;
- linear part $\ell_i = (a+1)(b-n) - (a+1)(i - \frac{a}{2})m$ for $0 \leq i \leq a$, with cluster a monomial times $-\mathcal{R}_{a+1}$.

Each cluster is nonzero because $\mathbb{Z}[z]$ is a domain. As points of the space of linear forms in (b, n, m), the first $a+1$ linear parts lie on the line $\mathbb{R}m$ through 0. The other $a+1$ lie on the parallel line $(a+1)(b-n) + \mathbb{R}m$, which misses 0.

Let $\mu_{\text{lin}}, \nu_{\text{lin}}$ be the linear parts of $m' \circ \phi$ and $n' \circ \phi$ (not to be confused with the constant ν below). Then $\mu_{\text{lin}} \neq 0$: for fixed b, ϕ is a bijection of the (n, m)-plane, since both observation vectors are injective by Lemma 7.5(b) and have the same image. The linear parts of $F_{\varepsilon'} \circ \phi$ are 0, $T_r \mu_{\text{lin}}$ and $(a+1)(b - \nu_{\text{lin}}) - (a+1)(i - \frac{a}{2})\mu_{\text{lin}}$. If two of them coincided, $F_{\varepsilon'} \circ \phi$ would have at most $2a+1$ nonzero clusters. So the two sets of linear parts are equal.

Each set lies on two parallel lines, each carrying $a+1 \geq 3$ points, and no other line carries three of them. So the line through 0 is the same: $\mu_{\text{lin}} = tm$, and $\{T_1, \ldots, T_a\} = \{tT_1, \ldots, tT_a\}$ forces $t = 1$. On the other line, $\nu_{\text{lin}} = n + sm$, and matching its $a+1$ points forces $s = 0$. Hence $m' \circ \phi = m + \delta$ and $n' \circ \phi = n + \nu$ with constants δ, ν. (For $a = 1$ each line carries only two points, and indeed the exchange has $m' = 2b - 2n - m$.)

Clusters. Equal linear parts carry equal clusters. At level r, cancelling $z^{T_r m} B_r$ in the domain $\mathbb{Z}[z^{1/N}]$ (N a common denominator) gives $P_r = z^{T_r \delta} P'_r$, with $P'_r = \sum_i z^{\kappa_{i,r}(\varepsilon')}$. For $r = 1$ the exponents $\kappa_{i,1} = \eta_{i-a}$ lie in $\{0, 1\}$ on both sides, so $\delta \in \{-1, 0, 1\}$.

Suppose $\delta = 1$. Then $\{\kappa_{i,r}(\varepsilon)\} = \{\kappa_{i,r}(\varepsilon') + T_r\}$, and since $0 \leq \kappa \leq T_r$, every $\kappa_{i,r}(\varepsilon) = T_r$ and every $\kappa_{i,r}(\varepsilon') = 0$. At $r = a$ this means $\eta \equiv 1$ and $\eta' \equiv 0$ on $[-a, a-1]$, that is, $\varepsilon_h = h$ and $\varepsilon'_h = 0$. The clusters at ℓ_i give $\sum_{W_i} c_h = \sum_{W_i} c'_h \circ \phi$, where $c_h = n + h(m+1)$ and $c'_h \circ \phi = n + \nu + h(m+1)$. So $\nu = 0$ and $c' \circ \phi = c$, which is excluded. The case $\delta = -1$ is the same.

Hence $\delta = 0$ and $P_r = P'_r$, which is (N3). The clusters at ℓ_i give $\sum_{W_i}(c_h - c'_h \circ \phi) = 0$, that is, $\sum_{W_i}(\varepsilon_h - \varepsilon'_h) = (a+1)\nu$ for every i. The case $i = 0$ is the value of ν, and all of them together are (N2). The cluster at 0 then agrees, because its S is the case $i = 0$.

Converse. Under (N2) and (N3), with $\phi(b, n, m) = (b, n + \nu, m)$, the clusters of F_ε and $F_{\varepsilon'} \circ \phi$ agree, so $F_\varepsilon = F_{\varepsilon'} \circ \phi$ as formal sums. We compare the observation vectors as lists of affine forms, without evaluating them at points: since $\nu \notin \mathbb{Z}$ in general (Theorem 7.10), ϕ maps no integer point to an integer point, and no argument through integer points generic for both pieces is available. By Lemma 7.5(a), V_P is the list of the terms of F_ε (reflected), and $V_{P'}$ that of $F_{\varepsilon'}$, each put in canonical form by steps (a)–(c) of Section 5 with respect to the common chamber C. These steps commute with composition by ϕ: ϕ is an affine bijection that fixes b, so it preserves identity and distinctness of exponent forms (step (a)) and the identical vanishing of coefficient forms (step (b)), and the reflection $e \mapsto T + \deg \mathcal{Q} - e$

depends on b only. For step (c), let e_1, e_2 be two exponent forms of $F_{\varepsilon'}$. The linear part of ϕ is the identity, so $(e_1 - e_2) \circ \phi$ has the same homogeneous part as $e_1 - e_2$. If that part is 0, then $e_1 - e_2$ is a constant whose n-coefficient is 0, and $(e_1 - e_2) \circ \phi = e_1 - e_2$: the same constant decides the order. Otherwise the order is decided by the sign of the same homogeneous part on the interior of the same chamber C. So the canonical form of $F_{\varepsilon'} \circ \phi$ is the canonical form of $F_{\varepsilon'}$ composed with ϕ, that is, $V_{P'} \circ \phi$; and it is the canonical form of F_ε, that is, V_P. Hence $V_P = V_{P'} \circ \phi$ as affine maps, and $A_P = A_{P'}$: the signatures are equal. Moreover $c'_h \circ \phi - c_h = \nu + \varepsilon'_h - \varepsilon_h$; at $h = 0$ this is ν, so it vanishes for every h only if $\nu = 0$ and $\varepsilon' = \varepsilon$. □

Corollary 7.7. *In the situation of Theorem 7.6, the entries of V_P and $V_{P'}$ that come from level r coincide entry by entry. Each exponent is $(a+1)b - T_r m + \kappa$, with the same constant κ in both, because the cluster $P_r B_r$ is the same. A single such entry already forces $m' = m$. The shift of n is that of Proposition 7.1.*

Theorem 7.6 uses, like the rest of the reduction, that V_P is the sparse form of K_c on the piece. For C_+ that premise is Proposition 7.4, so the affinity argument of Section 6.6 is proved there.

7.2. Checks beyond the certificate. The next statements were found by computation, and the first two are explained by Theorem 7.6.

Proposition 7.8 computer-assisted. *Let $2 \le a \le 6$.*

(a) Every pair of generic pieces with the following four properties lies in C_+ on both sides:
- *observation vectors of the same length;*
- *the same directions (the system $V_P(b, n, m) = V_{P'}(b, n', m')$ has rank 2 in (n, m, n', m'));*
- *a consistent system (as affine equations in the five variables (b, n, m, n', m'): the augmented matrix has the same rank as the matrix);*
- *different vector formulas.*

Hence every candidate of (C2) lies in C_+.

(b) In every candidate pair, a single row of the system implies $m' = m$. It is an exponent entry $(a+1)b - T_k m + \kappa$ with the same constant κ in both pieces.

(c) For $a = 6$ the program finds:
- *224 templates, 155 chambers and 30 240 generic pieces, none of rank < 2;*
- *7707 signatures and 132 candidate pairs;*
- *every candidate in C_+, with $m' = m$ and $n' = n - k/7$, $1 \le k \le 6$; none is integer.*

Checks and controls.

- *(a).* The grouping by directions was compared with the rank of the system computed pair by pair for $a \le 3$ (62, 1923 and 27 232 pairs). For $a = 1$ the eight pairs of the exchange have rank 2, different vectors and a chamber other than C_+, so the test is not vacuous.
- *(b).* The same search never certifies the row $n - n'$, as expected since $n - n' = k/(a+1)$ on the solutions. For $a = 1$ only 1 of the 9 candidates has $m' = m$ certified.
- *(c).* The program for $a \le 5$ does not fit in 6 GB for $a = 6$. The run uses a mode that does not store the conditions of genericity, and so has *no regression*: the formulas

of the 30 240 pieces were not re-checked. We therefore do *not* claim $E(6,b) = O(b)$, although the output satisfies (C1) and (C2).

Corollary 7.9 computer-assisted. *For $a = 2, \dots, 6$ the candidate pairs of the certificate are exactly the unordered pairs of distinct templates that satisfy (N2) and (N3):* 6, 15, 42, 45 *and* 132 *pairs. In each of them the linear part of ϕ is the identity on (n, m), $m' = m$, and $n' - n = (S_-(\varepsilon) - S_-(\varepsilon'))/(a+1)$.*

Two checks are combined here: the pairs with (N2) and (N3) come from an enumeration of templates that does not use the certificate, and the two sets of pairs are then compared.

Checks and controls.

- *Without a certificate* ($a = 7, \dots, 15$), the same count gives 139, 238, 245, 572, 351, 970, 725, 942 and 1149 pairs inside C_+, none with $S_-(\varepsilon) \equiv S_-(\varepsilon') \pmod{a+1}$ (as Theorem 7.10 below requires). Two enumerations of templates agree ($a \le 8$).
- *An observation, not proved.* For every $a \le 15$, (N2) and (N3) single out the same pairs of templates (Question 10.2).
- *Controls.* With (N3) required only for $r = 1$, or with windows of length a in (N2), the counts differ (22 and 10 pairs for $a = 2$, against 6). For $a = 1$, where Theorem 7.6 does not apply, the pairs with (N2) and (N3) do not match the candidates.

7.3. The shift is never an integer.

Theorem 7.10 written proof. *Let $a \ge 1$, $L = a + 1$, and let $\varepsilon \ne \varepsilon'$ be templates whose window sums differ by a constant: $\sum_{h \in W_i}(\varepsilon_h - \varepsilon'_h) = D$ for $0 \le i \le a$. Then $L \nmid D$. In particular two distinct templates that satisfy (N2) have $S_-(\varepsilon) \not\equiv S_-(\varepsilon') \pmod{a+1}$, and a class of templates whose window sums agree up to a constant has at most $a + 1$ elements.*

Proof. Write $\varepsilon_h = \lfloor u + hv \rfloor$ and $\varepsilon'_h = \lfloor u' + hv' \rfloor$ with $u, v, u', v' \in (0,1)$, and suppose $D = kL$ with $k \in \mathbb{Z}$. Put $d_h = \varepsilon_h - \varepsilon'_h - k$ for $|h| \le a$. Every window of L consecutive indices in $[-a, a]$ has $\sum d_h = 0$, and the difference of two consecutive windows gives $d_{h+L} = d_h$ for $-a \le h \le -1$.

Suppose some $d_h \ne 0$. By periodicity some d_h with $-a \le h \le 0$ is nonzero, and since $\sum_{h=-a}^{0} d_h = 0$ the window $[-a, 0]$ contains an index p with $d_p > 0$ and an index q with $d_q < 0$. Every index of $[-a, -1]$ occurs again L places later; only 0 does not. So one of p, q, say p, is in $[-a, -1]$, and the $L - 1$ indices strictly between p and $p + L$ meet every residue class modulo L except that of p; one of them, q', has $q' \equiv q$, so $d_{q'} = d_q < 0$. (If $q \in [-a, -1]$ instead, exchange the signs.) This gives $h_1 < h_2 < h_3$ with d_{h_1}, d_{h_3} of one sign and d_{h_2} of the other.

Now $d_h = \lfloor u + hv \rfloor - \lfloor u' + k + hv' \rfloor$, and $\lfloor x \rfloor > \lfloor y \rfloor$ implies $x > y$. So the sign of d_h, when it is not 0, is the sign of $f(h) = (u - u' - k) + h(v - v')$. An affine function of h cannot take signs $+, -, +$ or $-, +, -$ at $h_1 < h_2 < h_3$. Hence $d \equiv 0$, that is, $\varepsilon_h - \varepsilon'_h = k$ for every h; at $h = 0$ both templates are $\lfloor u \rfloor = \lfloor u' \rfloor = 0$, so $k = 0$ and $\varepsilon = \varepsilon'$, a contradiction.

For (N2), $D = \sum_{h=-a}^{0}(\varepsilon_h - \varepsilon'_h) = S_-(\varepsilon) - S_-(\varepsilon')$. Having window sums that agree up to a constant is an equivalence relation, and inside one class the residues of S_- modulo $a + 1$ are pairwise distinct, so a class has at most $a + 1$ elements. □

The proof uses that ε and ε' are floors of affine functions. In the language of Sturmian words (Section 5), the hypothesis says, after taking differences of consecutive windows ($\varepsilon_{h+a+1} - \varepsilon_h =$

$\varepsilon'_{h+a+1} - \varepsilon'_h$), that the gap words η, η', two finite Sturmian words of length $2a$, have equal sums over *corresponding* windows of $a+1$ consecutive gaps, position by position (a condition on the a windows of each word at the same places, not an equality of factor languages), and the theorem bounds the resulting shift; we did not find this statement in the literature on balanced and mechanical words that we searched (Lothaire [12, Ch. 2], abelian complexity and balance of Sturmian words), and its two ingredients (the sign of a difference of floors and the periodicity forced by equal window sums) are elementary. For arbitrary sequences with $\varepsilon_0 = 0$ and gaps in $\{0, 1\}$ the statement fails from $a = 3$ on: for instance $(-1, -1, -1, 0, 1, 1, 1)$ and $(-2, -1, 0, 0, 0, 1, 2)$ have the same window sums of length 4 (so $D = 0$), and among all such sequences there are 2, 16 and 132 pairs with $D \equiv 0 \pmod{a+1}$ for $a = 3, 4, 5$. In the enumerations ($a \leq 15$) the bound $a + 1$ on the size of a class is attained for every a.

Corollary 7.11. *Let $a \geq 2$. Every piece of C_+ satisfies (C1), and every pair of pieces of C_+ with the same signature satisfies (C2): either $c_{P'} \circ \phi = c_P$, or $\phi(b, n, m) = (b, n+\nu, m)$ with $\nu \notin \mathbb{Z}$, which maps no integer point to an integer point.*

Proof. Lemma 7.5(b), Theorem 7.6 and Theorem 7.10 with $(a+1)\nu = S_-(\varepsilon) - S_-(\varepsilon')$. □

Conjecture 7.12. *Let $a \geq 2$. Every candidate pair of (C2) lies in C_+ (checked for $a \leq 6$, Proposition 7.8).*

With (C1) outside C_+ and this conjecture, Corollary 7.11 would give (C2), and with Theorem 5.1, $E(a, b) = O(b)$ for every $a \geq 2$. A weaker statement suffices (Question 10.1).

8. Lemma D for $a = 3$

Fix $a = 3$ and the 25 lines $\mathcal{L}_{\deg}$ of Section 6. The *depth* of a realisable vector is the supremum of the distances to $\mathcal{L}_{\deg}$ of the points of its cell. A vector is *thin* if it has at most one free coordinate and *exceptional* if it has at least two and no generic representative.

Theorem 8.1 Lemma D; computer-assisted. *Let $a = 3$. There is a constant C such that, for every $b \geq 317$, a realisable vector with at least two free coordinates and depth $> C$ shares its word with no other realisable vector.*

Remark 8.2. The restriction to at least two free coordinates is necessary. For $0 < x < b$ and $s = 3b + 3 + x$, the one-prime pair of Proposition 9.1 is $O_2(s) = (x, b, b, b, b, b, b)$, a thin vector whose cell contains the half-line $\Lambda - 3\theta = x + \frac{1}{2}$, $\theta > b$, at distance $\min(x + \frac{1}{2}, b - x - \frac{1}{2})/\sqrt{10}$ from the parallel lines $\Lambda - 3\theta \in \{0, b\}$ and far from the others; its partner $O_1(s)$ lies at $\theta = \frac{1}{4}$, within $\frac{1}{4}$ of the line $\theta = 0$. Equal words checked directly for $b = 100, 300, 1000, 3001$ (depths about $15.8, 47.4, 158.1, 474.5$), with the control $x \mapsto x + 1$ giving different words.

Figure 7 shows both phenomena at $b = 40$: the cells with a shared word and at least two free coordinates lie along the lines of $\mathcal{L}_{\deg}$, while the cells of the thin vectors $O_2(s)$ are wide wedges.

The two halves. Let c be deep with at least two free coordinates and $Z(d) = Z(c)$, $d \neq c$. For C large, c has a generic representative in some piece P: the strips of the forms whose zero lines are not in $\mathcal{L}_{\deg}$ meet their chamber only within $O(1)$ of a vertex or an edge (Section 6).

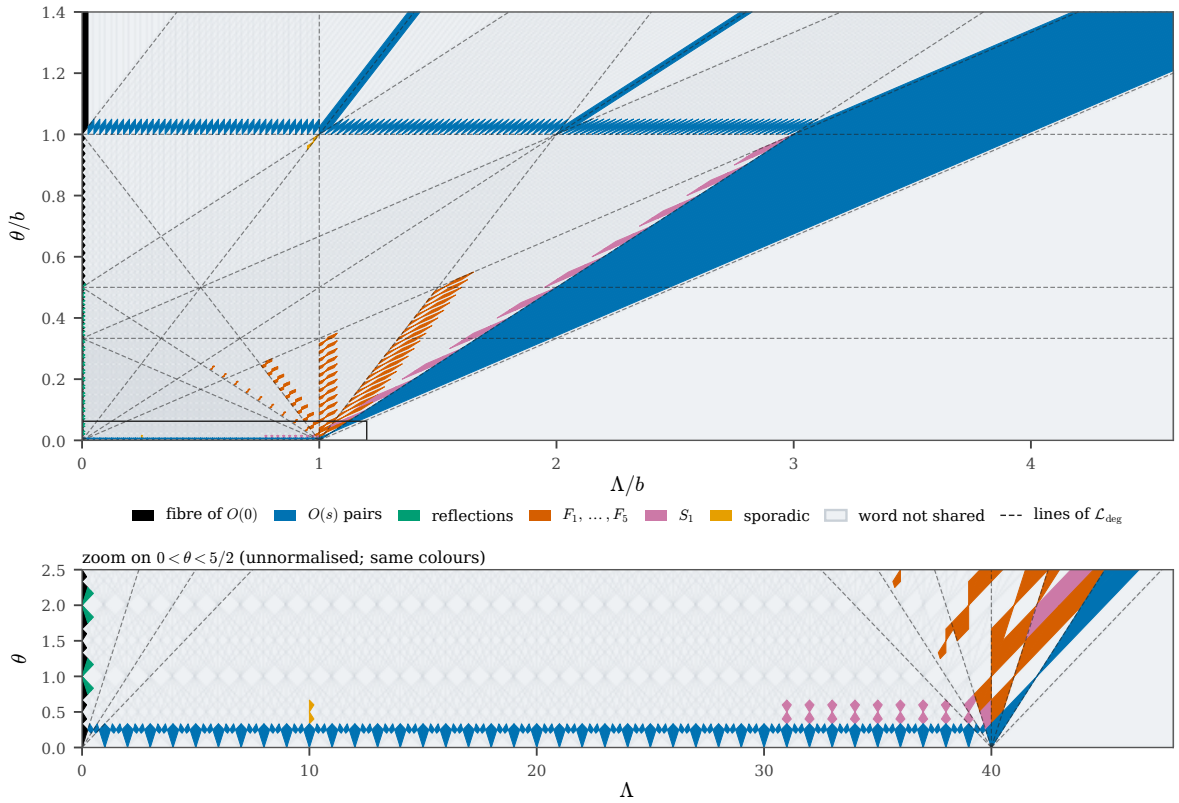

Figure 7. Where the coincidences live, $a = 3$, $b = 40$. Top: the exact cells in the window $0 < \Lambda < 4.6b$, $0 < \theta < 1.4b$ (23 134 of the 24 109 cells, in normalised coordinates). A cell is coloured when its word is shared (all 722 such cells meet the window), by the list of Section 9 that contains it, or as sporadic; dashed: the lines of $\mathcal{L}_{\mathrm{deg}}$. In this picture the coloured cells with at least two free coordinates lie along lines of $\mathcal{L}_{\mathrm{deg}}$ (Theorem 8.1 proves that they stay within a bounded distance of $\mathcal{L}_{\mathrm{deg}}$ for $b \geq 317$); the large blue wedges are cells of the thin vectors $O_2(s)$, which reach far from $\mathcal{L}_{\mathrm{deg}}$ (Remark 8.2). Bottom: the strip $0 < \theta < \frac{5}{2}$, $0 < \Lambda < 48$ (the box in the top panel), not normalised: the partners $O_1(s)$ on $\theta = \frac{1}{4}$, the vectors $G_{m/5}(s)$ of S_1 on $\theta = \frac{2}{5}, \frac{3}{5}$, and the reflections and the fibre of $O(0)$ along $\Lambda = 0$. Fibres from the exact enumeration.

- *d generic.* For every *ordered* pair of generic pieces (P, Q) of equal length, with the chamber of P open and that of Q closed, the program `aud4_pares_asim.py` computes the solution set of $V_P(b, n, m) = V_Q(b, n', m')$ and its asymptotic directions, and raises an alarm if a direction keeps the side of P strictly inside its chamber with an integer point and $c \neq d$. For $a = 3$: 7652 compatible pairs, all with the side of P on the boundary; 0 alarms (also 0 for $a = 1, 2$). With the chamber of P closed the same program raises 3960 alarms (420 for $a = 2$), so the check is not vacuous. Corollary 6.5 does not suffice here: it looks only at the interior of the product of the two chambers.
- *d exceptional or thin.* This is the case (ii) of Theorem 5.1, treated by the two lemmas below and a second computation.

Lemma 8.3 Lemma C: covering; written proof. *There is a finite explicit set of* families $\mathcal{F}$, *each with integer parameters, an affine form giving b and an affine observation vector $V_{\mathcal{F}}$,*

such that for $b \geq b_1$ every exceptional or thin realisable vector d is the vector $c_{\mathcal{F}}$ of some family at a valid *point (all recorded signs correct, all listed coefficients nonzero), with K_d given by $V_{\mathcal{F}}$ there. The families are of three kinds:* strips *(a plane $\ell(b,n,m) = t$, ℓ primitive with $(B,C) \neq (0,0)$, with an integral parametrisation by (β, s), a template and a closed interval of $\sigma = s/\beta$),* thin families *($((-b-1)^{h_0+a}, s, b^{a-h_0})$ with parameters (b,s)) and* rays *(a line inside a strip or thin family, with one integer parameter).*

Sketch. An exceptional d has a representative (n,m,ε) in the box; $(n/b, m/b)$ lies in a closed chamber. On an edge, it satisfies the equation of that edge. Inside, a recorded form $f = g\ell + D$ has the wrong sign or a coefficient vanishes; since the homogeneous part has a strict sign on the interior, $0 < |\ell(b,n,m)| \leq |D|/g$, so d lies on one of finitely many planes $\ell = t$; forms $Ab + D$ with $(B,C) = (0,0)$ only fail for $b \leq |D|/|A|$. This includes the chambers whose pieces have fewer than two free coordinates: if d has two, a clipping sign recorded there is wrong, and the same bound applies. All recorded forms and coefficients of all pieces (generic or not), and all edges, are taken. On a plane the deduction of the certificate is repeated in one variable $\sigma = s/\beta$, cutting the interval at every sign change; a form whose homogeneous part vanishes on the plane is constant there, so it is either right on the whole family or wrong on all of it (and the family is empty). A point that is still not valid lies on one of finitely many lines $X\beta + Ys = t'$ (a ray) or has β bounded. On a ray every form is $X\gamma + D$, with the sign of X for γ large. Thin vectors are treated in the same way with (b,s) and $\sigma = s/b \in [-\frac{5}{4}, \frac{5}{4}]$; this range contains $s = -b-1$ and $s = b$, so the vectors with no free coordinate are among them. □

Lemma 8.4 Lemma S: recession; written proof. *Let P be a generic piece and $\mathcal{F}$ a family, let S be the rational affine solution set of $V_P(b,n,m) = V_{\mathcal{F}}(x)$, $b = b_{\mathcal{F}}(x)$, and let $\Pi \subset S$ be the closed polyhedron cut out by the interval of σ (for strips), by $(n/b, m/b) \in C_P$ and by $b \geq 1$ (so every recession direction of Π has $b(v) \geq 0$). Let $f_e(b,n,m) \geq 0$ be the inequalities of the edges e of C_P and $g = \min_e f_e$. If no direction v of S satisfies $b(v) > 0$, the interval condition on σ, and $f_{e,\mathrm{hom}}(v) > 0$ for every e, then g is bounded above on Π.*

Proof. If $\Pi = \emptyset$ there is nothing to prove. Otherwise, by the decomposition theorem for polyhedra (Minkowski–Weyl in Motzkin's form, valid for every nonempty polyhedron, pointed or not; V need not consist of vertices; textbook material, e.g. [16, Ch. 7]), $\Pi = \mathrm{conv}(V) + \mathrm{rec}(\Pi)$ with V finite. For $x \in \mathrm{conv}(V)$ and $r \in \mathrm{rec}(\Pi)$, $g(x+r) \leq \max_e f_e(x) + \min_e f_{e,\mathrm{hom}}(r)$. Let $r \neq 0$ be a recession direction with $b(r) = 0$. For a strip, $b = q\beta + r_0$ gives $\beta(r) = 0$, and $lo{\cdot}\beta \leq s \leq hi{\cdot}\beta$ gives $s(r) = 0$; for a thin family the parameters are (b,s) with $lo{\cdot}b \leq s \leq hi{\cdot}b$, so again $s(r) = 0$; for a ray with b non-constant, $b(r) = 0$ forces $\gamma(r) = 0$ (rays with b constant are the pairs with b constant on S, which concern one value of b). So $V_{\mathcal{F}}$ is constant along r, and by the injectivity of $V_P(b,\cdot)$ (condition (C1)) $r = 0$, a contradiction. Since $b(r) \geq 0$ on $\mathrm{rec}(\Pi)$, $b(r) > 0$, and by hypothesis $\min_e f_{e,\mathrm{hom}}(r) \leq 0$. Hence $g \leq \max_{x \in \mathrm{conv}\, V} \max_e f_e(x)$. Equivalently, g is bounded on Π if and only if the linear program "maximise t subject to $t \leq f_e$ for every e, on Π" is bounded; that program also gives the constant. □

Proof of Theorem 8.1. The program `lema_d.py` builds the families of Lemma C for $a = 3$ (83 556 strip families on 333 planes, 28 thin families, and 967 920 ray families in two passes), and for every pair (generic piece, family) that passes two necessary filters (equal length of the observation vectors; equal constant coefficients) it solves the linear system exactly and decides the hypothesis of Lemma S by an exact feasibility test on the recession cone. Of the 3 624 618 pairs solved, 1 398 370 are incompatible, 359 673 have b constant on S (all with $b \leq 165$), 1 543 823 satisfy Lemma S (bounded depth), 322 752 violate it only because $c_P = c_{\mathcal{F}}$

identically on S (the same vector, described by a strip that crosses another chamber), none violates it without integer points, and *none* is an alarm (unbounded depth, integer points, $c \neq d$). So for $b \geq b_0$ every coincidence of a deep c with an exceptional or thin d has bounded depth, uniformly over the finitely many pairs, and with the ordered generic check above this proves the theorem with $C = \max(C_0, \max_{P,\mathcal{F}} C_{P\mathcal{F}}) + \sqrt{53}$, where $\sqrt{53}$ is a uniform bound on the diameter of a cell with at least two free coordinates: if $c_h, c_{h'}$ $(h \neq h')$ are free, two points of the cell have $|\Delta\Lambda + h\Delta\theta| < 1$ and $|\Delta\Lambda + h'\Delta\theta| < 1$, so $|\Delta\theta| < 2/|h - h'| \leq 2$ and $|\Delta\Lambda| < 1 + 2a = 7$, and the diameter is at most $\sqrt{(2a+1)^2 + 4} = \sqrt{53}$. The thresholds computed by the program give $b_0 = 317$. By source: forms $Ab + D$, 1; the margin of σ on the strip planes, 82 $(= b_1)$; sign failures off the rays on the strips, 66.7 (with the wide rays; 580 with the basic rays alone, which is why the wide rays are added); all signs of a ray equal to their asymptotic signs, 317 (basic rays) and 270 (wide rays); pairs with b constant on S, $b \leq 165$. The maximum, 317, comes from the basic rays. □

Controls and contrast. *Positive:* for $a = 1$ a plane that crosses the interior of the chamber of the exchange produces 16 genuine alarms, confirmed directly at $b \approx 3000$ (2028 integer points with $c \neq d$ and equal words); for $a = 3$ the solver finds the near–near family $O(s)$ (432 pairs, 7290 points checked at $b \approx 400$). *Negative:* with the chamber of P closed, 4816 alarms for $a = 2$; a corrupted coefficient of $\mathcal{Q}/(1 - z)$ in the families gives 5653 mismatches against the direct K at $b = 30$. The family code reproduces the certificate on all pieces (40, 350, 1692 pieces for $a = 1, 2, 3$). *Coverage:* at $b = 60$ and $b = 90$ every exceptional or thin vector (43 872 and 77 191) is covered by a strip or a ray with K of the formula equal to the direct K (264 vectors are uncovered at $b = 40$, below the threshold $b_1 \geq 82$ of the planes). *Above b_0:* an independent enumerator (not importing `lema_d.py`) finds, at $b = 331$ and $b = 400$, no vector with at least two free coordinates and depth > 3 in a coincidence; the largest depth in a coincidence is 3.0.

What Theorem 8.1 trusts, and its limits. It rests on: the trust base of Theorem 6.1 (pieces and observation vectors); Lemmas 8.3 and 8.4, written proofs (the independent re-reading of Appendix A required narrowing the argument to the ordered generic check, which is done above); and the programs `lema_d.py` and `aud4_pares_asim.py`, which have not been read by a human. Its limits: the constant C is not computed (Lemma S gives a bound per pair, obtainable by one linear program per pair, which has not been run; the data suggest $C \approx 3$); the threshold $b_0 = 317$ lies above the range of complete enumeration with classification; the coverage of Lemma C has been contrasted with exact enumeration only up to $b = 90$; the only positive control of the open branch is for $a = 1$. Theorem 8.1 does not classify the coincidences near $\mathcal{L}_{\text{deg}}$.

9. The case $a = 3$: the families of coincidences

Throughout this section $a = 3$, $b \geq 3$, $T = 4b + 4$, and $c = (c_{-3}, \ldots, c_3)$. The coordinate c_h enters the rows i with $0 \leq i - h \leq 3$: c_{-3} only row 0, c_{-2} rows $0, 1$, c_{-1} rows $0, 1, 2$, c_0 all rows, and so on. We write B^k for k final coordinates equal to b, and $[\alpha, \gamma]$ for $\sum_{v=\alpha}^{\gamma} \delta_v$.

9.1. The classes known from [I].

- *The empty graph.* The fibre of the word $O(0) = [1, T - 1] + \delta_T$ has $1 + b + \lceil b/2 \rceil + (b - \lfloor b/3 \rfloor)$ vectors [I, Lemma 7.1].

- *Reflections* [I, Lemma 7.5] with $p = 1, 2$: $\lfloor b/2 \rfloor + \lfloor b/3 \rfloor$ pairs, in distinct fibres. The rational reflections of [I, Lemma 7.7] need $(p+1)q \le 3$, hence $q = 1$: there are none beyond these.
- *One-prime words.* For $0 \le s \le 4b+3$ let $O(s) = [s+1, 4b+3] + (s+1)\delta_T$ and

$$O_1(s) = \left(\mathrm{clip}\lfloor (s+h)/4 \rfloor\right)_h, \qquad O_2(s) = \left(\mathrm{clip}(s + h(b+1))\right)_h.$$

Proposition 9.1 $O(s)$ for $a = 3$. *For $0 \le s \le 4b+3$, $O_1(s)$ is the vector of $(\frac{2s+1}{8}, \frac{1}{4})$, $O_2(s)$ that of $(s+\frac{1}{2}, b+1)$, and $Z(O_1(s)) = Z(O_2(s)) = O(s)$. For $b \ge 2$, $O_1(s) \ne O_2(s)$ if and only if $s \le 4b+1$. For $s = 0$ both vectors lie in the fibre of $O(0)$.*

Proof. At $\theta = \frac{1}{4}$ the positions are $N/4$, $N = i + 4j$, which runs once over $0, \dots, 4b+3$. With $4\Lambda = s + \frac{1}{2}$ the values $\Lambda + h/4 = (s + h + \frac{1}{2})/4$ are not integers and have floor $\lfloor (s+h)/4 \rfloor$. By Lemma 2.2 the count at $N/4$ is $\#\{N' \le 4b+3 : N' < N + s + \frac{1}{2}\} = \min(T, N+s+1)$, which takes the values $s+1, \dots, 4b+3$ once and the value T for the $s+1$ largest N. At $\theta = b+1$ the positions $i(b+1)+j$ are again the integers $0, \dots, 4b+3$, and with $\Lambda = s + \frac{1}{2}$ the counts are the same; the vector is $\mathrm{clip}(s + h(b+1))$.

Distinctness. If $s \le b$: $O_2(s)$ has $c_1 = b$, while $O_1(s)$ has $c_1 = \lfloor (s+1)/4 \rfloor < b$. If $b < s \le 2b+1$: $c_0 = b$ against $\lfloor s/4 \rfloor < b$. If $2b+2 \le s \le 3b+2$: $c_{-1} = \mathrm{clip}(s-b-1) = b$ against $\lfloor (s-1)/4 \rfloor \le \lfloor (3b+1)/4 \rfloor < b$ $(b \ge 2)$. If $3b+3 \le s \le 4b+3$: $O_2(s) = (s-3b-3, B^6)$, and $O_1(s)$ has $c_{-2} = \lfloor (s-2)/4 \rfloor$, which is b iff $s \ge 4b+2$; for $s \in \{4b+2, 4b+3\}$ also $c_{-3} = \lfloor (s-3)/4 \rfloor = s - 3b - 3$, so $O_1(s) = O_2(s)$. Finally $O(0)$ is the word of the empty graph. □

So the one-prime classes contribute $4b+1$ pairs $(1 \le s \le 4b+1)$. The proof is that of [I, Supplement, Lemma 2.3(a)] and works for every a, with $\theta = 1/(a+1)$.

9.2. Five families next to the upper clipping.

Theorem 9.2 the families $F_1, \dots, F_5$. *In each line below, P and Q are distinct realisable vectors with $Z(P) = Z(Q)$, for all integers $i \ge 0$ and $\varepsilon \in \{0, 1\}$ in the stated range.*

F_1: $P = (x, y, z, B^4)$, $Q = (x+1, y, z-1, B^4)$; $x = b-3-3i$, $y = b-2-2i+\varepsilon$, $z = b-i$; $x \ge -1$.

F_2: $P = (x, y, z, w, B^3)$, $Q = (x+1, y, z, w-1, B^3)$; $x = b-4-4i$, $y = b-3-3i+\varepsilon$, $z = b-2-2i+\varepsilon$, $w = b-i$; $x \ge -1$.

F_3: $P = (x, y, B^5)$, $Q = (x+1, y-1, B^5)$; $x = b-6-2i$, $y = b-2-i$; $x \ge -1$.

F_4: $P = (x, y, z, w, B^3)$, $Q = (x+1, y+1, z-1, w-1, B^3)$; $x = b-4-4i$, $y = b-3-3i$, $z = b-1-2i$, $w = b-i$; $x \ge -1$.

F_5: $P = (x, y, z, w, v, B^2)$, $Q = (x, y+1, z, w-1, v, B^2)$; $x = b-7-5i$, $y = b-6-4i$, $z = b-4-3i$, $w = b-2-2i$, $v = b-1-i$; $y \ge -1$.

The numbers of pairs are $2\lfloor (b+1)/3 \rfloor$, $2\lfloor (b+1)/4 \rfloor$, $\lfloor (b-3)/2 \rfloor$, $\lfloor (b+1)/4 \rfloor$ and $\lfloor (b-1)/4 \rfloor$.

Proof. Realisability and distinctness. Table 6 gives a witness for each vector. In $F_1, \dots, F_4$ the vectors differ in c_{-3}, in F_5 in c_{-2}.

In $F_1, \dots, F_4$ all coordinates are ≥ -1 in the stated ranges. In F_5 all coordinates but x are ≥ -1; $x = y - 1 - i \ge (-b-3)/4 > -b-1$ is free, it enters only row 0, and on the interval $[3+2i, 5+4i]$ where row 0 changes, $x + j + 1 \ge b - 3 - 3i \ge i + 2 > 0$, so there $f(x+j) = x + j + 1$. For $v \ge -1$ and $j \ge 0$, $f(v+j) = \min(b+1, v+j+1)$, which equals $v + j + 1$ up to the *cut* $j = b - v$ and $b+1$ after. In each case we move from P

vector	Λ	θ	$\Lambda+h\theta-c_h$ (free h, from $h=-3$)	$\Lambda+h\theta-b$ (other h)
F_1: P, $\varepsilon=0$	$b+\frac{7}{4}$	$i+\frac{3}{2}$	$\frac{1}{4},\frac{3}{4},\frac{1}{4}$	$\geq\frac{7}{4}$
F_1: Q, $\varepsilon=0$	$b+\frac{1}{6}$	$i+\frac{2}{3}$	$\frac{1}{6},\frac{5}{6},\frac{1}{2}$	$\geq\frac{1}{6}$
F_1: P, $\varepsilon=1$	$b+\frac{9}{4}$	$i+\frac{3}{2}$	$\frac{3}{4},\frac{1}{4},\frac{3}{4}$	$\geq\frac{9}{4}$
F_1: Q, $\varepsilon=1$	$b+\frac{1}{4}$	$i+\frac{1}{2}$	$\frac{3}{4},\frac{1}{4},\frac{3}{4}$	$\geq\frac{1}{4}$
F_2: P, $\varepsilon=0$	$b-i+\frac{1}{6}$	$i+\frac{4}{3}$	$\frac{1}{6},\frac{1}{2},\frac{5}{6},\frac{1}{6}$	$\geq\frac{3}{2}$
F_2: Q, $\varepsilon=0$	$b-i-\frac{5}{8}$	$i+\frac{3}{4}$	$\frac{1}{8},\frac{7}{8},\frac{5}{8},\frac{3}{8}$	$\geq\frac{1}{8}$
F_2: P, $\varepsilon=1$	$b-i+\frac{5}{6}$	$i+\frac{4}{3}$	$\frac{5}{6},\frac{1}{6},\frac{1}{2},\frac{5}{6}$	$\geq\frac{13}{6}$
F_2: Q, $\varepsilon=1$	$b-i-\frac{1}{6}$	$i+\frac{2}{3}$	$\frac{5}{6},\frac{1}{2},\frac{1}{6},\frac{5}{6}$	$\geq\frac{1}{2}$
F_3: P	$b+i+\frac{13}{2}$	$i+4$	$\frac{1}{2},\frac{1}{2}$	$\geq\frac{5}{2}$
F_3: Q	$b+i+\frac{11}{4}$	$i+\frac{5}{2}$	$\frac{1}{4},\frac{3}{4}$	$\geq\frac{1}{4}$
F_4: P	$b-i+\frac{3}{4}$	$i+\frac{3}{2}$	$\frac{1}{4},\frac{3}{4},\frac{1}{4},\frac{3}{4}$	$\geq\frac{9}{4}$
F_4: Q	$b-i-\frac{1}{2}$	$i+\frac{2}{3}$	$\frac{1}{2},\frac{1}{6},\frac{5}{6},\frac{1}{2}$	$\geq\frac{1}{6}$
F_5: P	$b-2i-\frac{11}{6}$	$i+\frac{5}{3}$	$\frac{1}{6},\frac{5}{6},\frac{1}{2},\frac{1}{6},\frac{5}{6}$	$\geq\frac{3}{2}$
F_5: Q	$b-2i-\frac{13}{6}$	$i+\frac{4}{3}$	$\frac{5}{6},\frac{1}{6},\frac{1}{2},\frac{5}{6},\frac{1}{6}$	$\geq\frac{1}{2}$

Table 6. Witnesses for $F_1,\dots,F_5$. The fourth column lists $\Lambda+h\theta-c_h\in(0,1)$ for the coordinates that are written out in Theorem 9.2 (when one of them equals b, as $z=b$ for $i=0$ in F_1, the value $b+\text{offset}$ still clips to b); the fifth gives a lower bound, valid for all $i\geq 0$, of $\Lambda+h\theta-b$ for the coordinates equal to b. All offsets are affine and do not depend on (b,i), the fractional parts of Λ and θ make every $\Lambda+h\theta$ a non-integer, and $\Lambda,\theta>0$ in the stated ranges. The shading only separates the families.

to Q by Lemma 2.1: a coordinate $c_h\geq -1$ raised by 1 changes the rows $0\leq i-h\leq 3$ on $J=[0,b-c_h-1]$, lowered by 1 on $J=[0,b-c_h]$. Below, "row r is $\alpha+\sigma j$ on $[j_1,j_2]$" refers to P, and $\pm$ is the net change of the row there.

F_3. Moves $c_{-3}\colon x\mapsto x+1$ (row 0, $J=[0,5+2i]$) and $c_{-2}\colon y\mapsto y-1$ (rows $0,1$, $J=[0,2+i]$). Net: row 0 is $+1$ on $[3+i,5+2i]$, row 1 is -1 on $[0,2+i]$. Cuts: $b-y=2+i$, $b-x=6+2i$. Row $0=2(b+1)+f(y+j)+f(x+j)$ is $4b-2-2i+j$ on $[3+i,5+2i]$; row $1=3(b+1)+f(y+j)$ is $4b+2-i+j$ on $[0,2+i]$. The two contributions telescope to $\delta_T-\delta_{4b+1-i}$ and $\delta_{4b+1-i}-\delta_T$. This is the family E of [I, Supplement, Proposition 4.7] with one more row.

F_1. Moves $c_{-3}\colon x\mapsto x+1$ (row 0, $J=[0,2+3i]$) and $c_{-1}\colon z\mapsto z-1$ (rows $0,1,2$, $J=[0,i]$). Net: row 0 is $+1$ on $[i+1,2+3i]$; rows $1,2$ are -1 on $[0,i]$. Cuts: $b-z=i$, $b-y=2+2i-\varepsilon$, $b-x=3+3i$. On P:

$$\text{row } 0=\begin{cases}4b-1-5i+\varepsilon+2j, & [i+1,\ 2+2i-\varepsilon],\\ 4b+1-3i+j, & [3+2i-\varepsilon,\ 2+3i],\end{cases}$$

$$\text{row } 1=4b+2-3i+\varepsilon+2j,\qquad \text{row } 2=4b+4-i+j\qquad \text{on } [0,i].$$

The slope-1 pieces give $(\delta_T-\delta_{4b+4-i-\varepsilon})+(\delta_{4b+3-i}-\delta_T)$. With $j=i+1+k$ the slope-2 piece of row 0 gives $\sum_{k=0}^{1+i-\varepsilon}(\delta_{4b+2-3i+\varepsilon+2k}-\delta_{4b+1-3i+\varepsilon+2k})$, and row 1 gives $\sum_{k=0}^{i}(\delta_{4b+1-3i+\varepsilon+2k}-\delta_{4b+2-3i+\varepsilon+2k})$. For $\varepsilon=1$ all four cancel. For $\varepsilon=0$ the slope-2 sums leave the term $k=i+1$, $\delta_{4b+4-i}-\delta_{4b+3-i}$, which cancels the remainder $\delta_{4b+3-i}-\delta_{4b+4-i}$ of the slope-1 pieces.

F_2. Moves $c_{-3}\colon x\mapsto x+1$ (row 0, $J=[0,3+4i]$) and $c_0\colon w\mapsto w-1$ (rows $0,\dots,3$, $J=[0,i]$). Net: row 0 is $+1$ on $[i+1,3+4i]$; rows $1,2,3$ are -1 on $[0,i]$. Cuts: i, $2+2i-\varepsilon$,

$3+3i-\varepsilon$, $4+4i$. On P:

$$\text{row } 0 = \begin{cases} 4b-5-9i+2\varepsilon+3j, & [i+1,\, 2+2i-\varepsilon], \\ 4b-3-7i+\varepsilon+2j, & [3+2i-\varepsilon,\, 3+3i-\varepsilon], \\ 4b-4i+j, & [4+3i-\varepsilon,\, 3+4i], \end{cases}$$

and on $[0,i]$: row $1 = 4b-1-6i+2\varepsilon+3j$, row $2 = 4b+2-3i+\varepsilon+2j$, row $3 = 4b+4-i+j$. Slope 1: row 0 gives $\delta_T-\delta_{4b+4-i-\varepsilon}$ and row 3 gives $\delta_{4b+3-i}-\delta_T$. Slope 3: with $j=i+1+k$, row 0 gives $\sum_{k=0}^{1+i-\varepsilon}(\delta_{4b-1-6i+2\varepsilon+3k}-\delta_{4b-2-6i+2\varepsilon+3k})$, and row 1 cancels its terms $k\le i$. Slope 2: with $j=3+2i-\varepsilon+k$ $(0\le k\le i)$, row 0 gives $\sum_k(\delta_{4b+4-3i-\varepsilon+2k}-\delta_{4b+3-3i-\varepsilon+2k})$, and row 2 gives $\sum_{k=0}^{i}(\delta_{4b+1-3i+\varepsilon+2k}-\delta_{4b+2-3i+\varepsilon+2k})$. For $\varepsilon=1$ everything cancels. For $\varepsilon=0$ the two slope-2 sums telescope against each other to $(\delta_{4b+4-i}-\delta_{4b+3-i})+(\delta_{4b+1-3i}-\delta_{4b+2-3i})$; the slope-3 remainder $(k=i+1)$ is $\delta_{4b+2-3i}-\delta_{4b+1-3i}$, and the slope-1 remainder is $\delta_{4b+3-i}-\delta_{4b+4-i}$. The total is 0.

F_4. Four moves: $x\mapsto x+1$ (row 0, $J=[0,3+4i]$), $y\mapsto y+1$ (rows $0,1$, $J=[0,2+3i]$), $z\mapsto z-1$ (rows $0,1,2$, $J=[0,1+2i]$), $w\mapsto w-1$ (rows $0,\dots,3$, $J=[0,i]$). Net:

	$[0,i]$	$[i+1,1+2i]$	$[2+2i,2+3i]$	$[3+3i,3+4i]$
row 0	0	$+1$	$+2$	$+1$
row 1	-1	0	$+1$	0
row 2	-2	-1	0	0
row 3	-1	0	0	0

Cuts: i, $1+2i$, $3+3i$, $4+4i$. On the intervals with nonzero net change, P has: row $0 = 4b-4-9i+3j$, $4b-3-7i+2j$, $4b-4i+j$ on the three intervals; row $1 = 4b-6i+3j$ on $[0,i]$ and $4b+1-3i+j$ on $[2+2i,2+3i]$; row $2 = 4b+3-3i+2j$ on $[0,i]$ and $4b+3-2i+j$ on $[i+1,1+2i]$; row $3 = 4b+4-i+j$ on $[0,i]$. Slope 3: row 0 on $[i+1,1+2i]$ gives $\sum_{k=0}^{i}(\delta_{4b-6i+3k}-\delta_{4b-1-6i+3k})$, cancelled by row 1 on $[0,i]$. Slope 2, with the net change ±2: row 0 gives $\sum_{k=0}^{i}(\delta_{4b+3-3i+2k}-\delta_{4b+1-3i+2k})$, cancelled by row 2 on $[0,i]$. Slope 1: row 0 on $[3+3i,3+4i]$ and row 1 on $[2+2i,2+3i]$ each give $\delta_T-\delta_{4b+3-i}$; row 2 on $[i+1,1+2i]$ and row 3 each give $\delta_{4b+3-i}-\delta_T$. The total is 0.

F_5. Moves $c_{-2}\colon y\mapsto y+1$ (rows $0,1$, $J=[0,5+4i]$, as $y\ge-1$) and $c_0\colon w\mapsto w-1$ (rows $0,\dots,3$, $J=[0,2+2i]$). Net: rows $0,1$ are $+1$ on $[3+2i,5+4i]$; rows $2,3$ are -1 on $[0,2+2i]$. Cuts: $b-v=1+i$, $b-w=2+2i$, $b-z=4+3i$, $b-y=6+4i$. On P:

$$\begin{aligned} &\text{row } 0 = 4b-13-12i+3j\ [3+2i,4+3i], && 4b-9-9i+2j\ [5+3i,5+4i], \\ &\text{row } 1 = 4b-6-7i+2j\ [3+2i,4+3i], && 4b-2-4i+j\ [5+3i,5+4i], \\ &\text{row } 2 = 4b-3-6i+3j\ [0,1+i], && 4b-2-5i+2j\ [2+i,2+2i], \\ &\text{row } 3 = 4b+1-3i+2j\ [0,1+i], && 4b+2-2i+j\ [2+i,2+2i]. \end{aligned}$$

Reindex each piece from its first point ($k=0,1,\dots$). The slope-3 piece of row 0 takes the values $4b-4-6i+3k$ $(0\le k\le1+i)$ and raises each by 1; the slope-3 piece of row 2 takes $4b-3-6i+3k$ (same k) and lowers each by 1: the two contributions cancel term by term. In the same way the slope-2 piece of row 0 on $[5+3i,5+4i]$ (values $4b+1-3i+2k$, raised) cancels that of row 2 on $[2+i,2+2i]$ (values $4b+2-3i+2k$, lowered), $0\le k\le i$; and the slope-2 piece of row 1 on $[3+2i,4+3i]$ (values $4b-3i+2k$, raised) cancels that of row 3 on $[0,1+i]$ (values $4b+1-3i+2k$, lowered), $0\le k\le1+i$. The slope-1 pieces give $\delta_T-\delta_{4b+3-i}$ (row 1) and $\delta_{4b+3-i}-\delta_T$ (row 3).

Counts. The ranges are $0 \le i \le (b-2)/3$, $(b-3)/4$, $(b-5)/2$, $(b-3)/4$, $(b-5)/4$, which give the stated numbers (F_1 and F_2 have two values of ε). □

Remark 9.3. The five families live next to the upper clipping ($c_h = b$ for large h), at θ close to an integer plus $\frac{3}{2}$, $\frac{4}{3}$ or 4 (Table 6). Heuristically they are clipped variants of reflections around m/q: for $q \in \{2,3\}$, [I, Lemma 7.7] needs $(p+1)q \le 3$, which fails for every $p \ge 1$, and for $q = 1$ the reflections exist ([I, Lemma 7.5], $p = 1, 2$) but not with these clipped vectors; the clipping truncates the levels and plays the role of the missing room. We have not made this precise. F_3 is the family E of $a = 2$ with the same formula, and F_1 is the analogue of the family D. At the first parameter i beyond the range, the first moving coordinate takes a value between -2 and -5 (depending on the family); there the two vectors are still realisable but the words differ, which recalls the condition $\beta' \ge -1$ of the exchange lemma [I, Lemma 7.2].

9.3. The family S_1: an exact point of denominator five. For $m \in \{2,3\}$ and $s \in \mathbb{Z}$ let $G_{m/5}(s) = \left(\mathrm{clip}\lfloor (s+mh)/5 \rfloor\right)_h$.

Theorem 9.4 the family S_1. *Let $m \in \{2,3\}$, $\rho \in \{0,1\}$ and $n \ge 1$ with $x = b - 4n + \rho \ge -1$, and put $s = 5b+4-5n+\rho$ and $y = b+1-m$. Then $c = (x, y, B^5)$ and $G = G_{m/5}(s)$ are distinct realisable vectors and $Z(c) = Z(G)$; precisely, with $\sigma = 3b+3+x$ for $m = 2$ and $\sigma' = 3b+2+x$ for $m = 3$,*

$$m = 2: \quad Z = \delta_\sigma + [\sigma + 2, 4b+3] + \delta_{4b+3} + \sigma\,\delta_T,$$
$$m = 3: \quad Z = \delta_{\sigma'} + \delta_{\sigma'+2} + [\sigma' + 4, 4b+3] + \delta_{4b+2} + \delta_{4b+3} + \sigma'\,\delta_T.$$

There are $2\big(\lfloor (b+1)/4 \rfloor + \lfloor (b+2)/4 \rfloor\big)$ such pairs. The pairs with $(n,\rho) = (1,1)$ are the pairs of F_1 with $i = 0$ ($m = 2$ for $\varepsilon = 1$, $m = 3$ for $\varepsilon = 0$).

Proof. Witnesses. For c take $\theta = y - x + \frac{1}{2}$ and $\Lambda = 3y - 2x + \frac{7}{4}$; then $\Lambda - 3\theta = x + \frac{1}{4}$, $\Lambda - 2\theta = y + \frac{3}{4}$ and $\Lambda - \theta = 2y - x + \frac{5}{4} \ge b + \frac{1}{4}$ (as $y \ge b-2$ and $x \le b-3$), and every $\Lambda + h\theta$ has fractional part $\frac{1}{4}$ or $\frac{3}{4}$. For G take $(\Lambda, \theta) = (\frac{2s+1}{10}, \frac{m}{5})$: $\Lambda + h\theta = (s + mh + \frac{1}{2})/5$ is not an integer and has floor $\lfloor (s+mh)/5 \rfloor$. The vectors differ in c_{-1}: G has $c_{-1} = \lfloor (s-m)/5 \rfloor \le b - n < b$.

The word of c. Rows 2, 3 are constant, $= T$. For $m = 2$ ($y = b-1$): $f(b-1+j)$ is b at $j = 0$ and $b+1$ after, so row $1 = 3(b+1) + f(b-1+j)$ is $4b+3$ at $j = 0$ and T after, and row $0 = 2(b+1) + f(b-1+j) + f(x+j)$ is $3b+3+x = \sigma$ at $j = 0$, $3b+4+x+j$ for $1 \le j \le b-x-1$ (the values $\sigma+2, \dots, 4b+3$), and T for the $x+1$ values $j \ge b-x$. Hence $Z(c) = \delta_\sigma + [\sigma+2, 4b+3] + \delta_{4b+3} + (x+1+b+2b+2)\delta_T$, and $x + 3b + 3 = \sigma$. For $m = 3$ ($y = b-2$): row 1 is $4b+2$, $4b+3$, then T; row 0 is $3b+2+x = \sigma'$, $3b+4+x$, then $3b+4+x+j$ for $2 \le j \le b-x-1$ (the values $\sigma'+4, \dots, 4b+3$), then T ($x+1$ times); the number of T is $x + 1 + (b-1) + 2(b+1) = \sigma'$.

The word of G. At $\theta = m/5$ the positions are $N/5$ with $N = mi + 5j$; as mi runs over four distinct residues modulo 5, the multiset of positions is a set S_m:

$$S_2 = \{N \in [0, 5b+6] : N \not\equiv 3\} \setminus \{1, 5b+5\},$$
$$S_3 = \{N \in [0, 5b+9] : N \not\equiv 2\} \setminus \{1, 4, 5b+5, 5b+8\}$$

(congruences modulo 5). With $5\Lambda = s + \frac{1}{2}$, Lemma 2.2 gives $Z(G) = \{\mathrm{cnt}(N+s) : N \in S_m\}$, where $\mathrm{cnt}(M) = \#\{N' \in S_m : N' \le M\}$. Below the top, S_m meets every five consecutive integers in four elements; precisely

$$\mathrm{cnt}(M+5) = \mathrm{cnt}(M) + 4 \qquad (\text{for } 1 \le M \ (m=2),\ 4 \le M \ (m=3),\ \text{and } M+5 \le 5b+4),$$

while at the top $\mathrm{cnt}(5b+5) = 4b+3$ ($m=2$), $\mathrm{cnt}(5b+5) = 4b+2$ and $\mathrm{cnt}(5b+6) = \mathrm{cnt}(5b+7) = \mathrm{cnt}(5b+8) = 4b+3$ ($m=3$), and $\mathrm{cnt} = T$ from $5b+6$ ($m=2$), resp. $5b+9$ ($m=3$), on. For $M \le 5b+4$ one has $\mathrm{cnt}(M) = M - \lfloor (M+2)/5 \rfloor$ ($m=2$, $M \ge 1$) and $\mathrm{cnt}(M) = M - 1 - \lfloor (M+3)/5 \rfloor$ ($m=3$, $M \ge 4$). Note $s \ge 5b+4-\frac{5}{4}(b+2) > 4$.

$m=2$. Here $s \equiv 4+\rho \pmod 5$ and $\sigma = 3b+3+x = 4b+3-4n+\rho$; one checks $\mathrm{cnt}(s) = s - \lfloor (s+2)/5 \rfloor = \sigma$ for $\rho = 0, 1$. Among $s+1, \dots, s+4$ exactly one integer is $\equiv 3$ (namely $s+4$ if $\rho=0$, $s+3$ if $\rho=1$) and none is 1, so $\mathrm{cnt}(s+e) = \sigma + \kappa(e)$ for $e \in \{0,1,2,4\}$, with $\kappa(0,1,2,4) = (0,1,2,3)$. Write $N \in S_2$ as $N = 5t+e$, $e \in \{0,1,2,4\}$ (and $N \ne 1$). If $N+s \le 5b+4$, then $\mathrm{cnt}(N+s) = \sigma + 4t + \kappa(e)$. This condition reads $N \le 5n-\rho$. Here $t=0$ contributes $\sigma, \sigma+2, \sigma+3$ ($N = 0, 2, 4$) and each $t \ge 1$ the block $\sigma+4t+\{0,1,2,3\}$, so the elements with $t \le n-1$ give $\{\sigma\} \cup [\sigma+2, \sigma+4n-1]$, each once. For $\rho=0$, $\sigma+4n-1 = 4b+2$, and the condition also admits $N = 5n$, which gives $\sigma+4n = 4b+3$. For $\rho=1$, $\sigma+4n-1 = 4b+3$ and nothing else is admitted. In both cases these N give $\{\sigma\} \cup [\sigma+2, 4b+3]$, each once. The next element of S_2, $N = 5n+1-\rho$, has $N+s = 5b+5$ and gives $4b+3$. All larger N give T, and there are $4b+4-\mathrm{cnt}(5n+1-\rho) = 4b+4-(4n+1-\rho) = \sigma$ of them. This is the stated word.

$m=3$. Here $\sigma' = 3b+2+x = 4b+2-4n+\rho$ and $\mathrm{cnt}(s) = s-1-\lfloor (s+3)/5 \rfloor = \sigma'$. Among $s+1, \dots, s+4$ exactly one integer is $\equiv 2$ ($s+3$ if $\rho=0$, $s+2$ if $\rho=1$), so $\mathrm{cnt}(s+e) = \sigma' + \kappa(e)$ for $e \in \{0,1,3,4\}$, $\kappa(0,1,3,4) = (0,1,2,3)$. Write $N \in S_3$ as $N = 5t+e$ with $e \in \{0,1,3,4\}$; for $t=0$ only $N = 0, 3$ occur, with values σ', $\sigma'+2$. If $N+s \le 5b+4$, i.e. $N \le 5n-\rho$, then $\mathrm{cnt}(N+s) = \sigma'+4t+\kappa(e)$, and the elements with $t \le n-1$ give $\{\sigma', \sigma'+2\} \cup [\sigma'+4, \sigma'+4n-1]$, each once. We separate the two values of ρ.

- $\rho = 0$: $\sigma' = 4b+2-4n$, so $\sigma'+4n-1 = 4b+1$; the condition also admits $N = 5n$, which gives $\sigma'+4n = 4b+2$. The part below $5b+5$ is $\{\sigma', \sigma'+2\} \cup [\sigma'+4, 4b+2]$. Above it, $N = 5n+1, 5n+3, 5n+4$ have $N+s = 5b+5, 5b+7, 5b+8$ and give $4b+2, 4b+3, 4b+3$.
- $\rho = 1$: $\sigma' = 4b+3-4n$, so $\sigma'+4n-1 = 4b+2$, reached by $N = 5n-1$ (for $n=1$ the interval $[\sigma'+4, 4b+2]$ is empty and the part below $5b+5$ is $\{4b-1, 4b+1\}$). Nothing else is admitted, and the part below $5b+5$ is again $\{\sigma', \sigma'+2\} \cup [\sigma'+4, 4b+2]$. Above it, $N = 5n, 5n+1, 5n+3$ have $N+s = 5b+5, 5b+6, 5b+8$ and give $4b+2, 4b+3, 4b+3$.

In both cases all larger N give T, and there are $4b+4-(4n+2-\rho) = \sigma'$ of them. The values below T are therefore $\{\sigma', \sigma'+2\} \cup [\sigma'+4, 4b+2] \cup \{4b+2, 4b+3, 4b+3\} = \{\sigma', \sigma'+2\} \cup [\sigma'+4, 4b+3] \cup \{4b+2, 4b+3\}$, which is the stated word.

Count and overlap. For each m, $\rho=0$ gives $\lfloor (b+1)/4 \rfloor$ values of n and $\rho=1$ gives $\lfloor (b+2)/4 \rfloor$. For $(n,\rho) = (1,1)$, $c = (b-3, b+1-m, B^5)$ and $G_{m/5}(5b)$ is $(b-2, b-1, b-1, B^4)$ for $m=2$ and $(b-2, b-2, b-1, B^4)$ for $m=3$, which are the pairs of F_1 with $i=0$. □

Corollary 9.5 a fibre containing at least four cells. *For $b \ge 5$ the four distinct vectors*

$$(b-4, b-1, B^5), \quad (b-3, b-2, B^5), \quad (b-2, b-2, b-1, B^4), \quad (b-2, b-1, b-1, b-1, B^3)$$

have the same word. Among the pairs of Theorem 9.4, a pair with $m=2$ and a pair with $m=3$ have the same word only for $(n,\rho) = (1,0)$ and $(1,1)$ respectively, which are these four vectors.

Proof. The first vector is c for $(n,\rho,m) = (1,0,2)$, with $\sigma = 3b+3+x = 4b-1$; its partner is $G_{2/5}(5b-1) = (b-2, b-1, b-1, b-1, B^3)$. The second is c for $(n,\rho,m) = (1,1,3)$, with $\sigma' = 3b+2+x = 4b-1$; its partner is $G_{3/5}(5b) = (b-2, b-2, b-1, B^4)$. With $\sigma = \sigma' = 4b-1$ both formulas of Theorem 9.4 read $\delta_{4b-1} + \delta_{4b+1} + \delta_{4b+2} + 2\delta_{4b+3} + (4b-1)\delta_T$. Conversely,

the two formulas have the same total weight on T only if $\sigma = \sigma'$, and then they differ by $\delta_{\sigma+3} - \delta_{4b+2}$, which vanishes only for $\sigma = 4b-1$, that is, $4n - \rho = 4$ ($m = 2$) and $4n - \rho = 3$ ($m = 3$). □

Remark 9.6. For $a = 2$ the analogous family F of [I] pairs $(x, b-1, b, b, b)$ with the exact point $\theta = \frac{2}{3}$, of denominator $3 = a + 1$. For $a = 3$ the denominator is 5, not $a + 1$; this is consistent with the factor 5 in the period of $R(3, b)$, but we have not proved a link.

9.4. The structural formula.

Theorem 9.7 the listed classes for $a = 3$. *Let $b \geq 5$. Each of the following is a set of realisable vectors with a single word: the fibre of $O(0)$; the pairs $\{O_1(s), O_2(s)\}$, $1 \leq s \leq 4b+1$; the reflection pairs with $p = 1, 2$; the pairs of $F_1, \ldots, F_5$; the pairs of S_1. Let $\mathcal{G}_b$ be the graph whose vertices are the vectors of these sets, $\mathcal{U}_b$, with an edge between two vectors whenever they lie in a common listed set. Then*

$$E(3, b) \;\geq\; |\mathcal{U}_b| - |\pi_0(\mathcal{G}_b)|,$$

where $\pi_0(\mathcal{G}_b)$ is the set of connected components of $\mathcal{G}_b$. The same holds for the graph $\mathcal{G}_b^+$ obtained by adding the set of Corollary 9.5.

Proof. By [I, Lemmas 7.1 and 7.5], Proposition 9.1, Theorems 9.2 and 9.4 and Corollary 9.5, each listed set has one word, so vectors in one connected component have one word, and $E(3, b)$ counts, for each word, its number of vectors minus one. □

The sum of the sizes of the lists, minus one per class, is

$$\underbrace{b + \lceil \tfrac{b}{2} \rceil + b - \lfloor \tfrac{b}{3} \rfloor}_{O(0)} + \underbrace{4b+1}_{O(s)} + \underbrace{\lfloor \tfrac{b}{2} \rfloor + \lfloor \tfrac{b}{3} \rfloor}_{\text{refl.}} + \underbrace{2\lfloor \tfrac{b+1}{4} \rfloor + 2\lfloor \tfrac{b+2}{4} \rfloor}_{S_1}$$
$$+ \underbrace{2\lfloor \tfrac{b+1}{3} \rfloor}_{F_1} + \underbrace{2\lfloor \tfrac{b+1}{4} \rfloor}_{F_2} + \underbrace{\lfloor \tfrac{b-3}{2} \rfloor}_{F_3} + \underbrace{\lfloor \tfrac{b+1}{4} \rfloor}_{F_4} + \underbrace{\lfloor \tfrac{b-1}{4} \rfloor}_{F_5}.$$

By Theorem 9.4 two pairs are listed twice. For $5 \leq b \leq 300$ we checked by computer that these are the only overlaps between the lists, that the union is a disjoint union of the $O(0)$ fibre and of pairs, and that each component has one word. Two quantities must be kept apart: the *graph count* $E_{\text{graph}}(b) = |\mathcal{U}_b| - |\pi_0(\mathcal{G}_b)|$, the right-hand side of Theorem 9.7, which is a lower bound for $E(3, b)$ for every $b \geq 5$; and the *floor formula* $E_{\text{floor}}(b)$, the sum above minus 2, an explicit sum of floor functions. The computation shows $E_{\text{graph}}(b) = E_{\text{floor}}(b)$ for $5 \leq b \leq 300$; for larger b the equality is not proved. Floor arithmetic simplifies the explicit sum to

$$E_{\text{floor}}(b) = 9b + 2\left\lfloor \tfrac{b+1}{3} \right\rfloor + \left\lfloor \tfrac{b-1}{2} \right\rfloor - r_{b \bmod 4}, \qquad (r_0, r_1, r_2, r_3) = (3, 4, 4, 1).$$

Comparing with the exact values:

Conjecture 9.8 structural formula; checked. *For $b \geq 5$, $E(3, b) = E_{\text{floor}}(b) + g^*_{b \bmod 4}$ with $(g^*_0, g^*_1, g^*_2, g^*_3) = (3, 5, 3, 18)$, that is,*

$$E(3, b) = 9b + 2\left\lfloor \tfrac{b+1}{3} \right\rfloor + \left\lfloor \tfrac{b-1}{2} \right\rfloor + h_{b \bmod 4},$$

$(h_0, h_1, h_2, h_3) = (0, 1, -1, 17)$, and $W(3, b) = R(3, b) - E(3, b)$ with $R(3, b)$ of [I, Theorem 5.1]. The slope of $E(3, b)$ is $\frac{61}{6}$.

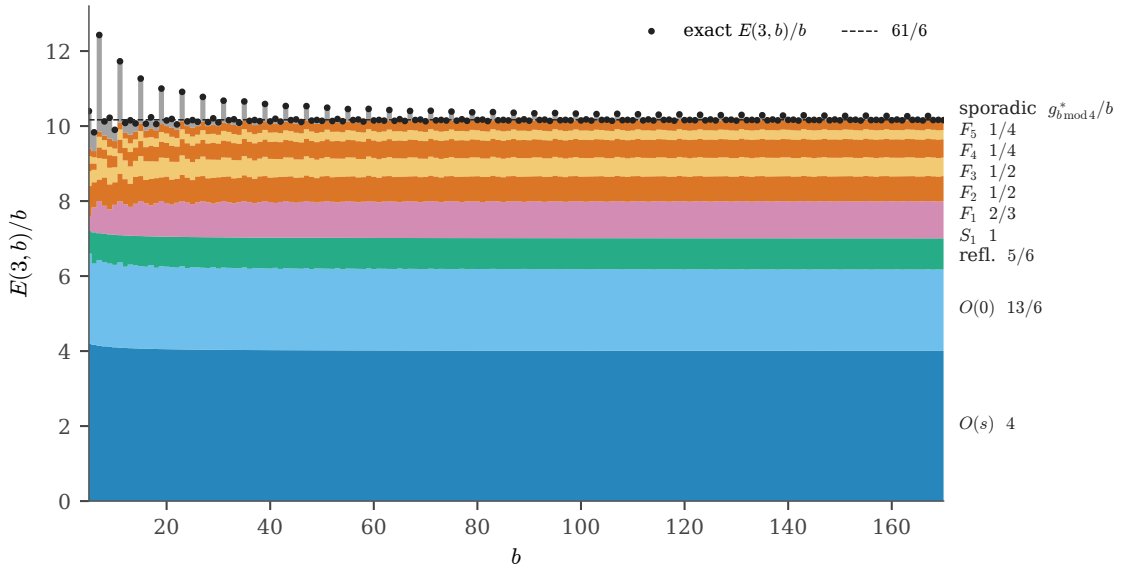

Figure 8. The listed classes account for the conjectured slope of $E(3,b)$. For $5 \le b \le 170$, the contribution of each list of Theorem 9.7 to $E(3,b)$, divided by b and stacked (the two pairs listed both in S_1 and in F_1 are removed from S_1). Each band tends to the slope written at its right, and these slopes add up to $\frac{61}{6}$. The grey top band is the sporadic remainder $E(3,b) - E_{\text{floor}}(b)$, exact value minus listed classes: in the data it is $g^*_{b \bmod 4} = 3, 5, 3, 18$, hence the comb at $b \equiv 3 \pmod 4$ that fades like $18/b$. Dots: exact $E(3,b)/b$.

The displayed formula is the floor sum above minus the two listed overlaps, plus g^*; collecting the terms for each residue of b modulo 4 gives the shorter form. This is an algebraic identity between two proposed formulas, not a proof that either counts every coincidence.

Status. The constant g^* is fitted to the data, not derived. The equality holds for $5 \le b \le 170$ and for $b \in \{180, 190, \dots, 290, 299, 300\}$ by exact enumeration of all cells and words (it fails for $b \le 4$), and also for $b = 331$ and $b = 400$ ($E(3,331) = 3381$, $E(3,400) = 4065$) with the independent enumerator used to check Lemma D. The resulting $W(3,b)$ coincides with the fitted quasipolynomial of [I, Question 11.3] for all $5 \le b \le 1000$, so the two are the same quasipolynomial. In particular, with Proposition 2.3, the linear coefficient of that fit is $D_3 - \frac{61}{6} = \frac{93}{5} - \frac{61}{6} = \frac{253}{30}$. The two coefficients of $W(3,b)$ have different status: the quadratic one, $\frac{73}{5}$, is a theorem (Theorem 6.1, computer-assisted); the linear one, $\frac{253}{30}$, comes from the counts of the families without fitting, but it is a coefficient of $W(3,b)$ only if the remaining coincidences contribute $O(1)$, that is, it is conjectural. Three levels should be kept apart: (1) every listed pair has one word (proved, Theorem 9.7); (2) the lists contribute exactly $E_{\text{floor}}(b)$ (this needs the overlap count, checked by computer for $b \le 300$ and not written by hand); (3) the remaining coincidences contribute exactly g^* (Conjecture 9.8). The proved quantity is $E_{\text{graph}}(b)$, the number of vertices minus the number of connected components of the *coincidence graph* whose edges are the proved coincidences; counting this way, overlaps between lists and fibres of size > 2 are handled automatically, and adding an edge raises the bound by one exactly when it joins two components. Figure 8 shows how the listed classes contribute to the conjectured slope $\frac{61}{6}$.

The sporadic coincidences. In the data, $g^*_{b \bmod 4} = 1+(2,4,2,8)_{b \bmod 4}+9\,[b \equiv 3\ (4)]$. The term 1 is proved: it is Corollary 9.5, which merges the S_1 pair with $(n,\rho,m) = (1,0,2)$ and the S_1 pair with $(n,\rho,m) = (1,1,3)$ into one fibre. Since these two pairs are different components of $\mathcal{G}_b$ for $5 \le b \le 300$ (computation above), Theorem 9.7 with $\mathcal{G}_b^+$ gives $E(3,b) \ge E_{\text{floor}}(b)+1$ there; for every $b \ge 5$ the same extra unit would need a proof that the two pairs lie in different components of $\mathcal{G}_b$, which we have only computed. That this fibre has exactly four elements, and is the only fibre of size > 2 outside $O(0)$, is only observed ($6 \le b \le 80$, and $b = 331, 400$). The term $(2,4,2,8)$ consists of clipped prolongations of S_1, with $c_{-3} \in \{-b-1, -b, -b+1\}$ depending on b mod 4 and small s. The term $9[b \equiv 3]$ consists of nine pairs near $\theta = (b+1)/4$ when $4 \mid b+1$, an integer reflection with Λ close to an integer, the analogue of the family B of $a = 2$ (which needs $3 \mid b+1$). Their number is constant in b in the data; we have not written proofs for them.

The period. $R(3,b)$ is a quasipolynomial on all $b \ge 1$ with period dividing $P_3 = 15$ (Proposition 2.3(b); [I, Corollary 5.2] gives 60), so its minimal period is decided by finitely many exact values; it is 15 (a finite exact check, so rigorous). The formula of Conjecture 9.8 has minimal period 12. Minimal periods 15 and 12 of the two terms only guarantee that the minimal period of their difference is a multiple of 20 dividing 60; that the minimal period of $R(3,b)$ minus the formula is exactly 60 is a finite exact computation (both are explicit quasipolynomials, and 12, 20 and 30 are not periods), so it is proved for the formula; for $W(3,b)$ itself it holds only if Conjecture 9.8 does. This is the "12 plus 5" structure observed in [I, Question 11.3]: the residues modulo 12 come from the families, those modulo 5 from the count of cells (Figure 9).

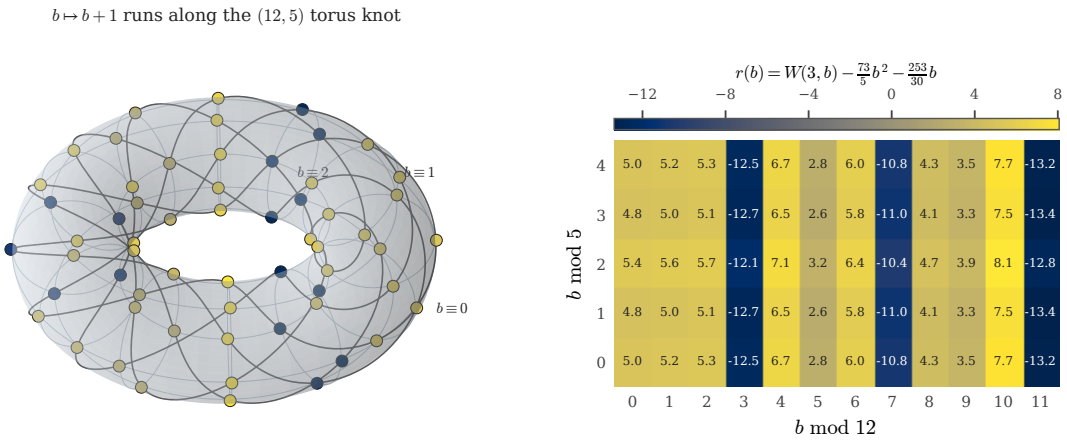

Figure 9. The residues of $W(3,b)$ modulo 60 on a torus: an observation on the exact values, and a property of the conjectural formula, not a theorem about $W(3,b)$. Since $\mathbb{Z}/60 = \mathbb{Z}/12 \times \mathbb{Z}/5$, the residue of b is a point of a torus with 12 steps around one circle and 5 around the other, and $b \mapsto b+1$ runs along the $(12,5)$ torus knot. Each of the 60 points is coloured by $r(b) = W(3,b) - \frac{73}{5}b^2 - \frac{253}{30}b$; right: the same values on the flat torus. In the 180 exact values of $W(3,b)$ for $5 \le b \le 300$, $r(b)$ depends only on b mod 60, and not only on b modulo 12, 15, 20 or 30; the formula of Conjecture 9.8 gives the same 60 values for all $5 \le b \le 1000$. The columns carry the period 12 of the proposed formula for $E(3,b)$ (the three dark columns are $b \equiv 3$ (mod 4), where $h_{b \bmod 4}$ jumps to 17); the small changes along each column come from $R(3,b)$.

10. Open questions

The certificates settle rigidity one value of a at a time. The questions below ask for what the certificates suggest but do not prove.

Rigidity for every a. The cost grows quickly (4, 14, 36, 76, 136 templates and 20, 238, 1296, 4712, 12 920 generic pieces for $a = 1, \dots, 5$), so certificates alone will not reach all a. Yet the output has the same shape for $a = 2, \dots, 5$, Proposition 7.1 accounts for its arithmetic, and Corollary 7.11 settles the all-positive chamber for every a. What is left lies outside C_+.

Question 10.1. Is $W(a, b) = C_a b^2 + O_a(b)$ for every $a \geq 2$? More concretely:

(i) Does (C1) hold outside C_+: does every full-dimensional generic piece with at least two free coordinates have an observation vector of (n, m)-rank 2?

(ii) Does every pair of pieces with the same signature, not both in C_+, satisfy one of the three alternatives of (C2): $c_{P'} \circ \phi = c_P$, no integer point mapped to an integer point, or an overlap of area 0?

Part (ii) is weaker than Conjecture 7.12, which asks that such pairs do not exist at all; a pair outside C_+ that fails all three alternatives would refute it and, if its generic points persist (Remark 5.2), give a quadratic family of coincidences. For part (i), one route is through the macroscopic profile of the rows, $t \mapsto \sum_{h \in W_i} \min(1, \max(0, \lambda + h\mu + t))$ for $0 \leq t \leq 1$, whose break points and masses are piecewise affine in (λ, μ): can two independent affine invariants of (λ, μ) be read from the aggregated distribution of the rows in every chamber? Contributions of different rows may cancel after aggregation, so a rank-one chamber with two free coordinates for this macroscopic data would close this route without refuting (C1). Which entries of the observation vector force $m' = m$ is answered for the all-positive chamber: any entry of a level $r \geq 1$ (Corollary 7.7). Outside C_+ some coordinates are clipped and the level structure of Proposition 7.4 is broken; for $a \leq 6$, in every chamber other than C_+, no two pieces with the same directions and different vectors are compatible.

Question 10.2. For every $a \geq 1$, if two templates have window sums that differ by a constant (condition (N2)), are their multisets of level constants equal (condition (N3)): $\{\kappa_{i,r}(\varepsilon)\}_i = \{\kappa_{i,r}(\varepsilon')\}_i$ for $1 \leq r \leq a$?

This holds for every pair of templates for $a \leq 15$ (1149 pairs with (N2) for $a = 15$), and fails for arbitrary sequences with gaps in $\{0, 1\}$ (2, 34, 290, 2006 pairs with (N2) and not (N3) for $a = 2, 3, 4, 5$), so a proof must use that templates are floors of affine functions, as in Theorem 7.10. The identity $\kappa_{i,r} = r\varepsilon_{i-a+r} - \sum_{t=0}^{r-1} \varepsilon_{i-a+t}$ may help. If true, the all-positive classification depends on the window sums alone.

Where the coincidences live. For $a \geq 2$ let $\mathcal{L}_{\deg}(a)$ be the lines $\Lambda + h\theta \in \{0, \pm b\}$ ($|h| \leq a$), $k\theta = b$ ($1 \leq k \leq a$) and $\theta = 0$; for $a \leq 5$ these are the boundaries of the final chambers.

Question 10.3. Is there, for every $a \geq 2$, a constant $K(a)$ such that every realisable vector with at least two free coordinates that shares its word with another vector has its cell within distance $K(a)$ of $\mathcal{L}_{\deg}(a)$?

For $a = 3$ and $b \geq 317$ this is Theorem 8.1; for $a = 4, 5$, Corollary 6.5 gives the half in which both cells are deep. It fails for $a = 1$ (the exchange of [I, Lemma 7.2] fills a region of positive area), and it fails without the condition on free coordinates (Remark 8.2). It would say not only that coincidences are few but where they are: the word of a cell far from the degenerate lines determines the cell.

An explicit Lemma D. Theorem 8.1 gives a constant C but not its value. Lemma S yields a bound for each pair by a rational linear program (maximise a depth variable $t \le f_e$ over the polyhedron; dual certificates make the bound checkable), and the data suggest $C \approx 3$. The threshold 317 comes from the homogeneous chambers; a partition by the signs of the complete affine forms (constants included) would be valid from $b = 1$ at a higher computational cost. With an explicit C and a threshold inside the enumerated range, Lemma D would hold for every $b \ge 5$, the small values being covered by the enumerations. Can this be done, and is $C = 3$ the true value for $a = 3$?

The classification for $a = 3$. Theorem 9.7 gives the lower bound; is Conjecture 9.8 true for all $b \ge 5$? Theorem 8.1 does not make it a theorem. It confines, for $b \ge 317$, every coincidence of a vector with at least two free coordinates to cells near $\mathcal{L}_{\deg}$; a proof of the conjecture would still need (1) an exhaustive description of the coincidences near $\mathcal{L}_{\deg}$, uniform in b (the families F_i, S_1, $O(s)$ and the sporadic pairs live there), (2) the proof that their coincidence graph has exactly the listed components, (3) proofs for the sporadic classes, which give the constant g^*, and (4) all b below the threshold, where the enumerations cover $5 \le b \le 170$ and some values up to 400 only. For (1) the solver of Lemma D applied to pairs of strip families, as in its positive control for $O(s)$, is a candidate; the SMT method of [I, Section 9.2], which proved the analogous statement for $a = 2$, is another, and for $a = 3$ its number of branches has not been measured. A statement stronger than the formula, and harder to satisfy by accident, is a partition: after adding to $\mathcal{G}_b^+$ the clipped prolongations of S_1 and the nine pairs near $\theta = (b+1)/4$, are the connected components exactly the non-singleton fibres, for every $b \ge 5$? Two errors in a classification can cancel in a count, but not in a partition; a union-find on the proposed edges against the exact fibres tests it, best on every b with $171 \le b \le 330$ rather than on isolated values.

The ambiguous cells as a semilinear set. Let $\mathcal{A}_a$ be the set of pairs (b, c) with c realisable for b and $Z(d) = Z(c)$ for some realisable $d \ne c$; its section at b has $\mathrm{Amb}(a, b)$ elements. It is the projection to (b, c) of a quantifier-free Presburger relation in (b, c, d) (realisability, equal words, $c \ne d$), hence semilinear, and by Lemma 3.6 and the argument of Proposition 4.1, $E(a, b) = O(b)$ if and only if every linear component of $\mathcal{A}_a$ has rank at most 2. Working with $\mathcal{A}_a$ rather than with B_a removes the choice of a representative in each fibre and the comparison of whole fibres: only the cells that have a partner are counted.

Question 10.4. For fixed $a \ge 2$, is $\mathcal{A}_a$ a finite union of affine families each of which, for fixed b, has at most one free integer parameter (plus finitely many sporadic points per b)? Can a decomposition of $\mathcal{A}_a$, separating its bounded pieces from its unbounded families, be computed for $a = 3$ without the full classification of Section 9?

By the argument of Proposition 4.1 applied to $\mathcal{A}_a$, the first part, read with Presburger-definable families, is equivalent to rigidity: a linear component of rank at most 2 is, for fixed b, a point or a set of points on a line. Its content is in an explicit decomposition, with simple ranges for the parameter. We expect a positive answer: for $a = 2$ it follows from the classification of the coincidence classes in [I, Lemma 9.3] (computer-assisted), and for $a = 3$ every coincidence found in the data belongs to the one-parameter families of Section 9 or to a sporadic class of constant size. For $a = 1$ it fails, through the two-parameter exchange of [I, Lemma 7.2]. Corollary 3.2(iii) gives a cheap necessary condition to prune candidates in any such computation.

How many power sums? Proposition 3.8 bounds the number of power sums that determine the word by $2s_a - 1$, which is 239 for $a = 2$ (503 with Theorem 3.7), while the data need at most 6 $(I_0, \dots, I_5)$ for $a \le 3$ and $b \le 20$.

Question 10.5. Is the least number $J(a, b)$ of Section 3 bounded by a small constant for fixed a, for instance $J(2, b) \le 5$ for all b? Which values of b need more than $I_0, \dots, I_3$, and why are they 6, 12 and 16 for $a = 2$ in the range computed?

The sparsity bound s_a is far from the data: the largest number of monomials of $\mathcal{Q}M_c$ over all realisable c is 37 for $a = 2$ ($b = 30, 60$) and 79 for $a = 3$ ($b = 40$), against $s_2 = 120$ and $s_3 = 280$, and it does not grow with b, while with the multiplier $z - 1$ alone it grows (34, 94, 184 for $a = 2$, $b = 10, 30, 60$).

Fibres. [I, Question 11.5] asks whether the fibres other than $O(0)$ have bounded size, perhaps 4. Proposition 3.5 bounds every fibre by $1 + \sum_{q \le 2a} \Phi_q(2b) \le 4ab + 1$, which has the right order for the fibre of $O(0)$ but says nothing on the others. For $a = 3$ the data agree with a bound 4: at $b = 40$, 331 and 400 the fibres outside $O(0)$ are 315, 2660 and 3195 pairs and one fibre of size 4, that of Corollary 9.5. The families $F_1, \dots, F_5$ and S_1 are mechanisms beyond the three listed in [I] (clipped reflections, and an exact point of denominator $5 \ne a + 1$). The distribution of fibre sizes has been computed for $a = 2, \dots, 5$ in the ranges of Table 3 (all b there). Let $N_j(a, b)$ be the number of fibres of size j other than that of $O(0)$. For $2a \le b$ in these ranges the fibres outside $O(0)$ have size at most 4 and N_3, N_4 do not depend on b:

a	2	3	4	5
(N_3, N_4)	$(1, 0)$	$(0, 1)$	$(3, 1)$	$(0, 4)$

while $N_2(a, b)$ grows linearly in b (for $a = 3$: 43 at $b = 6$, 636 at $b = 80$). So, outside $O(0)$, almost all fibres are pairs and a bounded number have 3 or 4 cells; for $a = 3$ the fibre of size 4 is that of Corollary 9.5. Fibres of size larger than 2 exist outside $O(0)$ for every $a \le 5$ (four of size four for $a = 5$); we have not examined their origin beyond $a = 3$. The second half of the question of degree has a short answer under rigidity.

Proposition 10.6 written proof. *Fix $a \ge 1$ and $j \ge 1$. Then $N_j(a, b)$ is, for b large, a quasipolynomial in b. If $E(a, b) = O_a(b)$ and $j \ge 2$, every constituent has degree at most one; so for $a = 2, 3, 4, 5$ this holds by Theorem 6.1 (computer-assisted).*

Proof. Let $N'_j(a, b)$ count all fibres of size exactly j. A fibre is counted once through its lexicographically smallest vector c, and "c is realisable, no realisable $d <_{\text{lex}} c$ has $Z(d) = Z(c)$, there are realisable $d_1 <_{\text{lex}} \cdots <_{\text{lex}} d_{j-1}$, all $>_{\text{lex}} c$, with $Z(d_i) = Z(c)$, and every realisable d with $Z(d) = Z(c)$ is one of $c, d_1, \dots, d_{j-1}$" is a Presburger formula in (b, c) for fixed j, since realisability and $Z(c) = Z(d)$ are [I, Section 6, Steps 2 and 3]. Its sections are finite, so by Woods's theorem [8], as in [I, Theorem 6.1], $N'_j(a, b)$ is eventually a quasipolynomial. The fibre of $O(0)$ has $1 + \sum_{h \le a} \Phi_h(b)$ vectors [I, Lemma 7.1], more than j for b large, so $N_j = N'_j$ eventually. A fibre of size $j \ge 2$ contributes $j - 1 \ge 1$ to E, so $(j - 1)N_j(a, b) \le E(a, b)$; a quasipolynomial that is $O(b)$ has every constituent of degree ≤ 1. □

For $j = 1$ the degree is two (the singletons are almost all cells). The proposition does not bound the size of the largest fibre.

Question 10.7. For fixed $a \ge 2$, is the size of the largest fibre other than $O(0)$ eventually constant in b?

Degree one for every a, and a regression test. By Woods's theorem [8], applied in [I, Theorem 6.1], $E(a,b) = R(a,b) - W(a,b)$ is eventually a quasipolynomial of degree at most two for every a; what is open is that its degree is at most one, which is the same as $E(a,b) = O_a(b)$: Theorem 6.1 for $a \le 5$, and Question 10.1 beyond. Stated as a test: if M is a period of $E(a,\cdot)$ from b_0 on and its degree is at most one, then

$$D_M(a,b) = E(a,b+2M) - 2E(a,b+M) + E(a,b) = 0 \qquad (b \ge b_0),$$

and one nonzero value beyond the claimed threshold refutes that period and threshold; a finite table cannot refute eventual quasipolynomiality itself, only a family with $E \ge cb^2$ would. For $a = 3$ and the exact values $E(3,b)$ with $b \le 110$ and $111 \le b \le 170$, $b \in \{180, 190, \dots, 290, 299\}$ (the ones available), $D_{60}(3,b) = 0$ for all 58 values of $b \ge 5$ with the three terms known, and $D_{12}(3,b) = 0$ for all 143; the only nonzero values are $D_M(3,1), \dots, D_M(3,4) = -1, 1, 49, 6$, so $b_0 = 5$ is the true threshold of the data. (A control that changes $E(3,150)$ by one gives three nonzero values, so the test is not vacuous.) This is a regression test of the enumeration and of Conjecture 9.8 (the formula has period $12 \mid 60$), not an independent confirmation of it: it uses the same values.

The ambiguity profile in θ. Let $A_{a,b}(\theta)$ be the proportion of the line $\{\theta\} \times (0, b+a\theta)$ lying in cells of fibre at least 2 (Section 6.3). Proposition 6.3 bounds it from below near $\theta = 1/(a+1)$. For fixed a, b the ambiguous length of the section is piecewise affine in θ: between two consecutive abscissas of vertices of the arrangement (all rational) the same cells cross the vertical line, and the ends of their sections lie on lines of the arrangement. So $A_{a,b}$ is piecewise linear-fractional, $A_{a,b}(\theta) = (\alpha + \beta\theta)/(b + a\theta)$ on each such interval, with rational α, β, and not piecewise linear in general; for rational ends of I the mean $P_{a,b}(I)$ is therefore of the form $r_0 + \sum_j r_j \log s_j$ with rational r_j and positive rational s_j (an antiderivative of $(\alpha+\beta\theta)/(b+a\theta)$ is $\frac{\beta}{a}\theta + \frac{\alpha - \beta b/a}{a}\log(b + a\theta)$). We have not computed these exact values; the means quoted in Section 6.3 are quadratures. On a compact interval of slopes the computations ($b = 12$–120; numerical second differences) suggest a piecewise linear limit with kinks only at the Farey fractions of order $2a$ (for $a = 2$: $\frac{1}{4}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, 1$).

Question 10.8. (i) Does $A_{a,b}$ converge, as $b \to \infty$, to a piecewise linear limit profile with kinks at these fractions, uniformly on compact intervals avoiding them, and is the limit of the mean $P_{a,b}(I)$ positive? A sharper test than comparing values at $b = 80, 160, 320$: for $a = 2$, compute the exact profile in each residue class of b modulo the period and compare the coefficients of b in the ambiguous length; different leading coefficients in two classes on the same interval of slopes would refute a single limit. (The exact means decrease like $L + c/b$ with $L \approx 0.04$ ($a = 2$), 0.02 ($a = 3$) between $b = 80$ and $b = 320$, a fit to three values.) (ii) For $a = 2$ the profile has a second peak at $\theta = \frac{2}{3}$ whose height, computed exactly, is $0.409, 0.375, 0.355, 0.345$ for $b = 6, 12, 24, 48$: does it tend to $\frac{1}{3}$ with a defect of order $\frac{1}{2b}$? We have a pairing of cells that accounts for it for $b = 3, 5, 8, 12, 20$, by computation only, and no proof. (iii) The cells meeting $\theta = 1$ have a fibre of size one for $a \le 5$, $b \le 16$ (exact check): is this so for all a, b?

The diagonal. Proposition 3.9 gives $E(a,a) \ge \frac{1}{2}(R(a,a) - 2a - 1)$, so $\liminf_n E(n,n)/R(n,n) \ge \frac{1}{2}$: rigidity for fixed a says nothing about $a = b \to \infty$, where at least half of the cells share their word with another. In the exact data $E(a,a)/R(a,a) = 0.472,\ 0.503,\ 0.505,\ 0.508,\ 0.506$ for $a = 2, \dots, 6$, and $W(a,a)$ is below the bound $(R + 2a + 1)/2$ by $1, 4, 7, 15, 20$.

Question 10.9. Does $E(n,n)/R(n,n)$ have a limit, is it $\frac{1}{2}$, and what is the growth of $(R(n,n)+2n+1)/2 - W(n,n)$? For $a \neq b$ with both large, which ratios b/a keep the proportion of ambiguous cells away from 0?

A value of $E(n,n)/R(n,n)$ clearly above $\frac{1}{2}$ for some n (we computed $n \leq 6$) would show another source of coincidences on the diagonal besides the exchange of the primes.

Two free parameters, and more primes. A family of coincident cells with two independent parameters at fixed b would be a linear component of rank three of the semilinear set B_a of Proposition 4.1; the certificates exclude it for $a = 2, \dots, 5$ and detect it for $a = 1$ (the exchange of progressions). So the search proposed by the companion for such families is the certificate itself; for $a = 6$ the run exists without its regression (Section 6), and a regression for $a = 6$ would settle whether the pattern continues. For $n = p_1^{a_1} \cdots p_k^{a_k}$ the same construction projects a box of exponents with $k-1$ slopes and a shift, and gives an arrangement of hyperplanes in dimension k. Write $B = \prod_{i=1}^{k} \{0, \dots, a_i\}$ for the box of exponents, $\vartheta = (\theta_1, \dots, \theta_{k-1}, 1)$ for the slopes, $\langle x, \vartheta \rangle$ for the projected position of $x \in B$, and, as in Lemma 2.2, $Z(\Lambda, \vartheta)$ for the multiset of the counts $\#\{y \in B : \langle y, \vartheta \rangle < \langle x, \vartheta \rangle + \Lambda\}$, $x \in B$; its sum is $\Sigma(\Lambda, \vartheta) = \#\{(x,y) \in B^2 : \langle y - x, \vartheta \rangle < \Lambda\}$. For $k = 2$ this is the Σ of Section 3.

Proposition 10.10 written proof. *Crossing the wall $\Lambda = \langle h, \vartheta \rangle$, $h \in B - B$, in the direction of increasing Λ, at a point on no other wall, raises Σ by exactly $\omega(h) = \prod_{i=1}^{k} (a_i + 1 - |h_i|) \geq 1$. Consequently $\Sigma = \sum_{h:\ \langle h, \vartheta \rangle < \Lambda} \omega(h)$, two cells separated by one wall have different multisets Z, and the cells with one multiset form an antichain for inclusion of the sets $\{h : \langle h, \vartheta \rangle < \Lambda\}$.*

Proof. Only the pairs with $y - x = h$ change status, and for each i there are $a_i + 1 - |h_i|$ pairs (x_i, y_i) in $\{0, \dots, a_i\}^2$ with $y_i - x_i = h_i$; the choices are independent. Summing over the walls below Λ gives the formula for Σ, and all weights are positive. □

For $k = 2$, $a_1 = a$, $a_2 = b$, the wall of $(-h, \rho) \in B - B$ is $\Lambda + h\theta = \rho$ and its weight is $\omega(h, \rho)$: this is Proposition 3.1. That Z encodes the word, as for two primes [I, Section 6], we have not checked for $k \geq 3$.

Question 10.11. For $k = 3$ and fixed a_1, a_2, is the number of cells whose word is shared by another cell of lower order than the number of cells, as $a_3 \to \infty$?

We have not examined this case, and we do not know that the reduction of Section 5 extends to it.

Effective constants and the threshold. The reduction gives $E(a,b) \leq K_a b$ for $b \geq b_0(a)$, with K_a and $b_0(a)$ computable from the certificate (strip widths, number of pieces), but not computed here. Explicit constants would not by themselves bound the threshold N_a beyond which $W(a,b)$ is a quasipolynomial: a function can be linearly bounded and still have arbitrarily late exceptions. Bounding N_a needs information on the semilinear decomposition of B_a (base points and periods), or a classification as in Conjecture 9.8. This is [I, Question 11.4] for the threshold.

The period. Is the minimal period of $R(a, \cdot)$ equal to the product P_a of the odd primes $\leq 2a$ for every a? Proposition 2.3(b) proves that it divides P_a, and equality holds for $a \leq 6$. For $a = 3$, if Conjecture 9.8 holds, the minimal period of $W(3,b)$ is 60 (both 15 for $R(3,b)$ and 60 for $R(3,b)$ minus the formula are finite exact computations); the conjecture is not proved. To prove $\mathrm{per}(\pi_a) = P_a$ it suffices that no odd prime $p \leq 2a$ disappears, that is, that $\pi_a(b + P_a/p) = \pi_a(b)$ fails for some b; grouping the Möbius expression of the proof of

Proposition 2.3 by the squarefree divisor d may allow a direct argument, the difficulty being cancellation among the multiples of p. Is the minimal period of $W(a,b)$ always the least common multiple of that of $R(a,b)$ and that of $E(a,b)$, and does the period of $E(a,b)$ come from the step d of one-parameter families $0 \le i \le \lfloor (b-r)/d \rfloor$, as the periods 2, 3 and 4 of the families for $a = 3$ do? For $a = 4$ the period of $E(4,b)$ is not known.

Acknowledgements

The problem was posed by José Manuel Rodríguez Caballero; I thank him for it and for his explanations.

Use of AI. I worked with an AI assistant.

Appendix A. Reproducibility and independent checks

Categories. We use the five categories of Table 1: *written proof* (an argument given in the paper), *formalised* (the stated part checked in Isabelle/HOL), *computer-assisted* (a finite exact computation that the result needs), *computed* (exact or numerical data with no conclusion for all b) and *conjecture*. The checks below are re-computations; they raise confidence in a written or computer-assisted step, but they do not have the status of a formal proof or of a referee's verification. Several of them were carried out by an AI agent other than the one that wrote the corresponding proof or program, re-deriving each step and recomputing it on exact data; none has been checked by a human referee. We say *certificate* for the output of the program of Section 6 together with its written reduction. A separate Excovery verifier, which imports neither the producer nor the paper's code, accepts the portable certificates for $a = 1, \dots, 5$ in full mode; the $a = 1$ positive control has 8 two-parameter families. For $a = 5$ it checks 15 232 pieces and all 45 non-integral candidates, with no errors. This checks the finite output, not the written reduction or the affinity of the true K_c on each piece; the certificate is not a formal proof object for those premises.

The finite reduction. The reduction of Section 5 was re-derived step by step by an AI agent other than the one that wrote it. No step had to be changed. Four points were made explicit:

- representatives are not unique;
- signs are strict, and vectors in chambers without pieces are exceptional;
- the definition of ϕ_{hom};
- the error bound in case (iv).

Every intermediate statement was compared with exact data for $a = 1, 2, 3$ and b up to 200, each comparison with a negative control that fails:

- the box (the bound $m \le 2b - 1$ is attained);
- the number of thin vectors;
- the formulas for c and K_c at generic points;
- (3), as an equality of the partitions by observation vector and by word;
- the injectivity used in (ii);
- the growth of the number $E_x(b)$ of exceptional vectors.

$E_x = O(b)$ by the strip argument above; the numerical data do not establish an eventually affine formula. For $a = 2$, $E_x(2b) - 2E_x(b) = 2595, 2667, 2640$ at $b = 25, 50, 100$, respectively. For $a = 3$ the increments over steps of 40 are 44 610, 44 276 and 43 249 for $80 \le b \le 200$,

while those of the generic vectors grow. In these data every coincidence for $a = 2, 3$ is of kind (i). Kinds (ii)–(iv) occur only for $a = 1$, where (iv) grows quadratically: $4, 84, 544, 2664$ at $b = 10, 20, 40, 80$, the exchange. So the bounds for (ii)–(iv) are tested on data only for $a = 1$. The same kind of re-derivation was made for Lemmas 8.3 and 8.4 (it required narrowing the argument to the ordered generic check of Section 8), for the proofs of Section 9 and for the output of the certificate (Section 6.6).

Scripts. Every explicit formula of Section 9 is checked by a script for $3 \le b \le 60$, with negative controls that fail as they should: Lemma M for $a = 3$ (all non-decreasing vectors, $b \le 5$), the rows of each family by pieces, the net changes, the words, the witnesses of Table 6 and Theorem 9.4, and the words of $O(s)$. The count of S_1, the two cases of its proof, Proposition 2.3 (exactly, $a \le 6$, every residue class) and the identity of Proposition 7.1 on the 109 candidates are checked by one more script with negative controls. The statements of Section 3 and the period of Proposition 2.3 are checked against the exact enumeration of cells and words by two further scripts, with negative controls. Theorem 7.10, Question 10.2 and Corollary 9.5 are checked by a separate script with its own enumeration of templates ($a \le 15$) and direct computation of words ($b \le 80$), with negative controls. Table 3, the fibre sizes, the diagonal values and the test $D_M = 0$ for $a = 3$ (with a control that fails) are computed by the scripts `V385_multiplicidades.py`, `V385_extra.py` and `V385_regresion_cuasipol.py`. Proposition 6.3 (i)–(iii) is checked in exact rational arithmetic ($a = 2, \dots, 5$, every k; words and widths, with controls that fail) by `n26_auditoria.py`; Table 4, the exact means and Figure 6 are produced by the scripts of the folder `numeros/` (`n02_medir.py`, `n23_media_theta.py`). The script `V387_comprobaciones.py` checks, in exact arithmetic, Proposition 3.11 ($b \le 300$), Proposition 6.3 (iii), (iv) (every segment for $a \le 6$, $b \le 60$, against the exact cell sections, with a wider segment as a control that fails) and both bounds against exact profiles with exact fibres, the numbers of Proposition 3.8, Proposition 10.10 on the shape $(2, 2, 3)$ (with a control that fails) and the diameter bound used in the proof of Theorem 8.1. All scripts use Python (3.14 for the last runs) with exact integers and rationals except where floating point is said; the Isabelle sessions are in the public repository named in Section 1.

Data. The certificate and its logs, the programs of Lemma D, the independent checks and the exact enumerations are not yet in a permanent archive; they are to be archived with a fixed version of this preprint.

References

[I] C. Marín, *How Many Dyck Words Do the Divisors of an Integer Make? Exact Counts and Quasipolynomial Growth*, preprint, Zenodo, 2026, version 2, doi:10.5281/zenodo.24541288. Cited as [I].

[1] M. A. Alekseyev, *On the number of two-dimensional threshold functions*, SIAM J. Discrete Math. **24** (2010), no. 4, 1617–1631, doi:10.1137/090750184; arXiv:math/0602511.

[2] D. M. Acketa and J. D. Žunić, *On the number of linear partitions of the (m, n)-grid*, Inform. Process. Lett. **38** (1991), no. 3, 163–168, doi:10.1016/0020-0190(91)90240-I.

[3] J. Žunić, *Note on the number of two-dimensional threshold functions*, SIAM J. Discrete Math. **25** (2011), no. 3, 1266–1268, doi:10.1137/110822037.

[4] J. M. Rodríguez Caballero, *A decision problem about balanced parentheses*, message to the FOM mailing list, 17 March 2020, https://fomarchive.ugent.be/2020-March/022069.html.

[5] S. Ginsburg and E. H. Spanier, *Semigroups, Presburger formulas, and languages*, Pacific J. Math. **16** (1966), no. 2, 285–296.

[6] J. M. Rodríguez Caballero, *Elementary number-theoretical statements proved by language theory*, arXiv:1709.09617, 2017.

[7] J. M. Rodríguez Caballero, *Symmetric Dyck paths and Hooley's Δ-function*, in: Combinatorics on Words (WORDS 2017), Lecture Notes in Comput. Sci., Springer, 2017, 252–261, doi:10.1007/978-3-319-66396-8_23.

[8] K. Woods, *Presburger arithmetic, rational generating functions, and quasi-polynomials*, J. Symbolic Logic **80** (2015), 433–449, doi:10.1017/jsl.2015.4.

[9] A. J. Hoffman, *On approximate solutions of systems of linear inequalities*, J. Research Nat. Bur. Standards **49** (1952), 263–265.

[10] F. Mignosi, *On the number of factors of Sturmian words*, Theoret. Comput. Sci. **82** (1991), no. 1, 71–84, doi:10.1016/0304-3975(91)90172-X.

[11] J. Berstel and M. Pocchiola, *A geometric proof of the enumeration formula for Sturmian words*, Internat. J. Algebra Comput. **3** (1993), no. 3, 349–355, doi:10.1142/S0218196793000238.

[12] M. Lothaire, *Algebraic Combinatorics on Words*, Encyclopedia Math. Appl. 90, Cambridge Univ. Press, 2002, Chapter 2 (J. Berstel and P. Séébold, *Sturmian words*).

[13] P. Haukkanen and J. K. Merikoski, *Asymptotics of the number of threshold functions on a two-dimensional rectangular grid*, arXiv:1110.3566v2, 2012.

[14] G. Hajós, *Solution to problem 41* (in Hungarian), Mat. Lapok **4** (1953), 40–41.

[15] H. W. Lenstra Jr., *Finding small degree factors of lacunary polynomials*, in: Number Theory in Progress, Vol. 1, de Gruyter, 1999, 267–276 (Proposition 3.2).

[16] A. Schrijver, *Theory of Linear and Integer Programming*, Wiley, 1986, Ch. 7.

[17] E. P. Lipatov, *A classification of binary collections and properties of homogeneity classes*, Problemy Kibernet. **39** (1982), 67–84 (in Russian).

[18] A. Frid, *Quelques méthodes pour les mots sturmiens*, in: J. Chalopin and P. Guillon (eds), Informatique mathématique: une photographie en 2019, CNRS Éditions, 2019, 35–64; arXiv:1901.01952.

[19] A. E. Frid and D. Jamet, *The number of binary rotation words*, RAIRO Theor. Inform. Appl. **48** (2014), 453–465, doi:10.1051/ita/2014019; arXiv:1302.3722.

[20] R. R. Hall and G. Tenenbaum, *Divisors*, Cambridge Tracts in Mathematics **90**, Cambridge Univ. Press, 1988.

[21] J. Koplowitz, M. Lindenbaum and A. Bruckstein, *The number of digital straight lines on an $N \times N$ grid*, IEEE Trans. Inform. Theory **36** (1990), no. 1, 192–197, doi:10.1109/18.50392.

[22] H. Koivusalo and J. J. Walton, *Cut and project sets with polytopal window I: Complexity*, Ergodic Theory Dynam. Systems **41** (2021), no. 5, 1431–1463, doi:10.1017/etds.2020.10; arXiv:1907.13382.

[23] A. Julien, *Complexity and cohomology for cut-and-projection tilings*, Ergodic Theory Dynam. Systems **30** (2010), no. 2, 489–523.

[24] T. B. McAllister and K. M. Woods, *The minimum period of the Ehrhart quasi-polynomial of a rational polytope*, arXiv:math/0310255, 2003.

[25] T. Zaslavsky, *Facing up to arrangements: face-count formulas for partitions of space by hyperplanes*, Mem. Amer. Math. Soc. **1** (1975), no. 154, doi:10.1090/memo/0154.

[26] OEIS Foundation Inc., *The On-Line Encyclopedia of Integer Sequences*, https://oeis.org, sequences A237271 (O. E. Pol, 2014), A249223 and A250068 (H. F. W. Hoft, 2014), A342592 and A342594 (H. F. W. Hoft, 2021).