

Partition stabilization and semigroup gaps for ternary Parikh matrices

Christopher Yamba

October 2026

Abstract

We study words with fixed positive content (a, b, c) over $1 < 2 < 3$, jointly by their Parikh-matrix entries and ordinary inversions. Words with at most b inversions correspond bijectively to pairs of restricted partitions. This gives a bivariate stable product, exact matrix-class sizes in a corner chamber, and explicit first corrections in both inversion number and the deficit $abc - |w|_{123}$. Below the sharp threshold $(b-1)\min(a, c) + a + c - 1$, an integer deficit is attainable exactly when it belongs to the semigroup generated by a and c . For coprime endpoints and an explicit sufficient bound on b , all missing triple counts are the reflections of the Sylvester gaps. We also evaluate the first exterior matrix fibres and classify the first words containing a reverse triple. The proofs use endpoint gaps, identical-letter shuffles and constructive profile realizations. We distinguish the results proved in Lean from the larger collection established by human-readable proofs.

Keywords: Parikh matrix; multiset permutation; scattered subword; restricted partition; inversion; numerical semigroup.

Mathematics Subject Classification (2020): 05A05, 05A15, 68R15.

1 Introduction

A Parikh matrix records occurrences of consecutive increasing subwords of an ordered alphabet. When every letter occurs once, its entries are Boolean: the relative order of consecutive values determines the entire matrix. With repeated letters, the entries become genuine counting statistics, and adjacent-pair counts need no longer determine longer subword counts. The resulting problem is both geometric and enumerative: which matrices occur, and how many words have each matrix?

This paper treats one precise part of that problem. We fix the ternary content (a, b, c) and study the region near the maximum abc of the scattered-123 count. This region admits two complementary descriptions. Endpoint gaps give a partition-pair bijection and exact multiplicities; their truncated sums give a numerical-semigroup description of support. The same encoding explains the first failure of the bijection, including the extra shuffles of indistinguishable letters.

Parikh matrices were introduced by Mateescu, Salomaa, Salomaa and Yu [3]. Their subword identities and the subword-history framework of Mateescu, Salomaa and Yu [4] supply basic algebra used here. Salomaa [5, §5, Definition 6] calls the cardinality of a matrix fibre its *degree of M -ambiguity*; his discussion also includes a partition description of binary fibres. We use this existing notion, evaluating specified positive-content ternary fibres rather than introducing a new ambiguity invariant.

The permutation-statistics viewpoint comes from the author's 2006 thesis [8]. Its Theorem 3.5, printed pp. 19–20 (PDF pp. 23–24), describes completeness on permutations, and its Theorems 6.12 and 6.15, printed pp. 31–34 (PDF pp. 35–38), relate the associated statistic

to inverse descents. These are translations of the classical permutation dictionary. The binary inversion generating function in the thesis, printed p. 36 (PDF p. 40), is MacMahon's Gaussian-binomial enumeration; we use Stanley's account [7, Proposition 1.7.1]. Our question concerns exact ternary distributions with repetitions.

Partition-pair stabilization itself has precedents. Claesson, Jelínek and Steingrímsson [1, §3, Proposition 15] obtain a two-partition description for 1324-avoiding permutations with few inversions; Linusson and Verkama [2, Theorem 1.2] study corrections beyond that stable regime. Fixed-content avoidance is studied by Savage and Wilf [6], and Yang [9] enumerates words with exactly one increasing triple. Our enumerations retain the literal ternary pair counts when evaluating matrix fibres. We state the exact results and credit these ingredients without an assertion of historical primacy.

Theorem 1 is the organizing bijection. Its scalar restriction and fixed-size form yield stable triple multiplicities and the strict matrix chamber. Theorems 2 and 3 evaluate the chamber boundary and the first scalar and inversion corrections. Theorems 4 and 5 describe support, while Theorems 6 and 7 classify the first exterior and reverse-pattern layers. All proofs are given below. Section 6 states the exact scope of the accompanying computation and formalization.

2 Endpoint gaps and a partition bijection

2.1 Words, matrices and deficits

Let $\mathcal{W}(a, b, c)$ be the set of words with a copies of 1, b copies of 2 and c copies of 3, where $a, b, c \geq 1$ throughout unless stated otherwise. An occurrence of $u = u_1 \cdots u_\ell$ in w is an increasing sequence of positions whose letters are $u_1, \dots, u_\ell$. All subwords in this paper are scattered. The ternary Parikh matrix is

$$\Psi(w) = \begin{pmatrix} 1 & a & x & z \\ 0 & 1 & b & y \\ 0 & 0 & 1 & c \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad x = |w|_{12}, \quad y = |w|_{23}, \quad z = |w|_{123}. \quad (2.1)$$

At fixed content write

$$\rho_{a,b,c}(x, y, z) = \#\{w \in \mathcal{W}(a, b, c) : (|w|_{12}, |w|_{23}, |w|_{123}) = (x, y, z)\}.$$

Two words with the same matrix are *M-equivalent* in the Parikh-matrix sense. Thus ρ is a matrix-class size, not a count of distinct matrices. We use the complementary pair counts and triple deficit

$$\begin{aligned} j &= ab - x = |w|_{21}, & i &= bc - y = |w|_{32}, & d &= abc - z, \\ m &= \min(a, c), & M &= \max(a, c). \end{aligned} \quad (2.2)$$

The ordinary inversion number is

$$k = \text{inv}(w) = |w|_{21} + |w|_{31} + |w|_{32} = i + j + |w|_{31}. \quad (2.3)$$

Reverse the word and interchange 1 with 3. This *reverse-complement* bijection interchanges a, c and i, j , while preserving k, z and the counts of both 123 and 321.

2.2 The gap encoding

The b twos determine $b + 1$ gaps, numbered $0, \dots, b$ from left to right. For each 1, record the number of preceding twos; list these as $p_1 \leq \dots \leq p_a$. For each 3, record the number of following twos; list these as $q_1 \geq \dots \geq q_c$. Then

$$0 \leq p_r, q_s \leq b, \quad \sum_r p_r = j, \quad \sum_s q_s = i. \quad (2.4)$$

Each list, with its zeros omitted, is an integer partition; their opposite display orders merely follow their occurrences in the word. Put $p_* = \max_r p_r$ and $q_* = \max_s q_s$.

Lemma 1 (Gap identities and shuffle multiplicity). *For endpoint data as above,*

$$z = \sum_{r=1}^a \sum_{s=1}^c (b - p_r - q_s)_+, \quad (2.5)$$

$$d = \sum_{r=1}^a \sum_{s=1}^c \min(p_r + q_s, b) = cj + ai - |w|_{321}. \quad (2.6)$$

Conversely, every two multisets of the indicated lengths with entries in $[0, b]$ are simultaneously realizable. If $\alpha_g = \#\{r : p_r = g\}$ and $\gamma_g = \#\{s : q_s = b - g\}$, their number of realizing words is

$$\prod_{g=0}^b \binom{\alpha_g + \gamma_g}{\alpha_g}. \quad (2.7)$$

Proof. The selected 1 lies in gap p_r , the selected 3 in gap $b - q_s$. If the 1 is in an earlier gap, the number of intervening twos is $b - p_r - q_s$; in the same gap it is zero, and in reverse gap order the pair contributes no 123. This proves (2.5). Subtracting from abc gives the first deficit sum. The identity $\min(p + q, b) = p + q - (p + q - b)_+$ gives its second form, because $(p_r + q_s - b)_+$ counts the twos between a selected 3 and a later 1. Summing counts all 321 occurrences.

To realize the data, put the prescribed ones in their gaps p_r and the prescribed threes in their gaps $b - q_s$. In gap g one chooses a binary shuffle of α_g identical ones and γ_g identical threes. These choices are independent, and deleting the twos recovers the choice in each gap. This proves both existence and (2.7). $\square$

The affine identity in (2.6) is classical subword-shuffle algebra [4, §§2–5]. In matrix coordinates it reads

$$z - |w|_{321} = cx + ay - abc. \quad (2.8)$$

In particular the usual lower bound $z \geq cx + ay - abc$ is an equality exactly when w avoids 321.

Lemma 2 (Canonical realization). *Two zero-padded partitions of lengths a, c have a realization without 31 if and only if $p_* + q_* \leq b$. This realization is unique, has $\text{inv} = i + j$, and has deficit $d = cj + ai$.*

Proof. If $p_* + q_* > b$, either a part exceeds b or the extreme 3-gap lies strictly before the extreme 1-gap. Both preclude the required word. If $p_* + q_* \leq b$, all parts are in $[0, b]$, and every 1-gap is before or equal to every 3-gap. In any shared gap put all ones before all threes. This gives the unique word without 31. It also avoids 321, so (2.3) and (2.6) give its statistics. $\square$

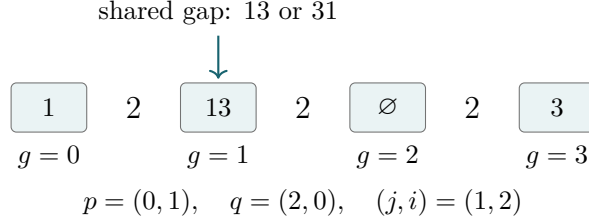


Figure 1: The canonical word 1213223 of content $(2, 3, 2)$. The shared gap has two shuffles, producing 1213223 and 1231223. Their inversion numbers are 3 and 4, but both have $z = 6$. The other partition pair with the same sums has $q = (1, 1)$ and produces 1212332. Thus this boundary matrix class has three words.

Let $p_r(n)$ denote the number of partitions of n into at most r parts, with $p_r(0) = 1$. Here the function $p_r(n)$ is distinguished from the gap-list entry p_r by its argument. Put

$$H_{a,b,c}(t, q) = \sum_{w \in \mathcal{W}(a,b,c)} t^{d(w)} q^{\text{inv}(w)}, \quad (2.9)$$

$$\mathcal{P}_{a,c}(t, q) = \prod_{u=1}^a (1 - q^u t^{cu})^{-1} \prod_{v=1}^c (1 - q^v t^{av})^{-1}. \quad (2.10)$$

All series are formal; no analytic convergence is used.

Theorem 1 (Stable partition bijection). *Words in $\mathcal{W}(a, b, c)$ with at most b inversions are in bijection with pairs of partitions having at most a and at most c parts and total size at most b . Under this bijection the partition sizes j, i give $\text{inv} = i + j$ and $d = cj + ai$. For every $0 \leq k \leq b$ and every $d \geq 0$,*

$$\#\{w : d(w) = d, \text{inv}(w) = k\} = \sum_{\substack{i+j=k \\ ai+cj=d}} p_a(j) p_c(i). \quad (2.11)$$

Equivalently,

$$[q^k] H_{a,b,c}(t, q) = [q^k] \mathcal{P}_{a,c}(t, q) \quad (0 \leq k \leq b). \quad (2.12)$$

Proof. For a word with $k \leq b$, equation (2.3) gives $i + j \leq b$. If $i + j < b$, then $p_* + q_* \leq i + j < b$, so every one precedes every three and $|w|_{31} = 0$. If $i + j = b$, the inequality $k \leq b$ forces the same conclusion. Thus the word is the canonical realization of its pair of partitions, and $k = i + j$. Conversely a partition pair with $i + j = k \leq b$ satisfies $p_* + q_* \leq b$; Lemma 2 constructs its unique word with the required statistics. These two operations are inverse. Counting by the two sizes proves (2.11).

The standard restricted-partition identity is $\sum_{n \geq 0} p_r(n) v^n = \prod_{h=1}^r (1 - v^h)^{-1}$: conjugate partitions to bound their part sizes and choose the multiplicity of each size. Applying this to each partition gives (2.12). $\square$

Lemma 3 (Scalar truncation). *If $d < bm$, then $\text{inv}(w) < b$.*

Proof. If a 3 precedes a 1 with gap values p, q , then $p + q \geq b$. The union of their row and column in (2.6), counting the intersection once, gives

$$d \geq b + (c - 1)p + (a - 1)q \geq b + (m - 1)(p + q) \geq bm. \quad (2.13)$$

Indeed every other entry in the selected row is at least p , and every other entry in its column is at least q . Thus $d < bm$ excludes 31. Now $d = cj + ai \geq m(i + j) = m \text{inv}(w)$, proving the assertion. $\square$

Corollary 1 (Scalar stable multiplicities). *For $0 \leq d < bm$,*

$$[t^d]H_{a,b,c}(t, 1) = [t^d]\mathcal{P}_{a,c}(t, 1) = \sum_{ai+cj=d} p_a(j)p_c(i). \quad (2.14)$$

In this range the coefficient is positive if and only if $d \in \langle a, c \rangle = \{ai + cj : i, j \geq 0\}$.

Proof. Lemma 3 puts every word in Theorem 1. On the partition side $m(i + j) \leq ai + cj = d < bm$ also forces $i + j < b$. Sum (2.11) over k . Since a, c are positive, restricted partition numbers are positive at every nonnegative argument; this proves the support assertion. $\square$

Corollary 2 (Congruence and matrix reconstruction). *Suppose $a \neq c$, $0 \leq k \leq b$, and $d \geq 0$. The joint count is nonzero if and only if*

$$mk \leq d \leq Mk, \quad d \equiv ck \pmod{|a - c|}. \quad (2.15)$$

When nonzero, it is $p_a(j)p_c(i)$, where

$$i = \frac{d - ck}{a - c}, \quad j = \frac{ak - d}{a - c}. \quad (2.16)$$

Consequently, at fixed unequal-endpoint content, $(\text{inv}(w), |w|_{123})$ determines the complete Parikh matrix for every word with $\text{inv}(w) \leq b$.

Proof. Solve $i + j = k$ and $ai + cj = d$. The conditions in (2.15) are exactly integrality and nonnegativity of the solution. Then Theorem 1 proves existence and the count. Finally recover $x = ab - j$, $y = bc - i$ and $z = abc - d$ in (2.1). $\square$

This conclusion determines a matrix, not an individual word. For example, in content $(3, 4, 2)$ the conditions $(d, k) = (7, 3)$ force $(i, j) = (1, 2)$, but the class consists of the two words

$$112212323, \quad 121122323, \quad (x, y, z) = (10, 7, 17). \quad (2.17)$$

When $a = c = m$, the two linear constraints are dependent instead: for $k \leq b$ the deficit is mk and the count is $\sum_{i+j=k} p_m(j)p_m(i)$.

3 Exact chamber sizes and first corrections

The stable bijection counts one canonical word for each partition pair. A matrix fibre can contain additional words when ones and threes share a gap. The following theorem evaluates exactly that first effect.

Theorem 2 (The full-matrix corner chamber). *Let $i, j \geq 0$ and $i + j \leq b$, and put $z_0 = abc - cj - ai$. Then*

$$\rho_{a,b,c}(ab - j, bc - i, z) = \begin{cases} p_a(j)p_c(i) + \varepsilon_b(i, j), & z = z_0, \\ 0, & z \neq z_0, \end{cases} \quad (3.1)$$

where $\varepsilon_b(i, j) = 0$ for $i + j < b$, and for $i + j = b$,

$$\varepsilon_b(i, j) = \begin{cases} 1, & i, j > 0, \\ a, & j = 0, \\ c, & i = 0. \end{cases} \quad (3.2)$$

Proof. Every endpoint pair satisfies $p_r + q_s \leq j + i \leq b$. Thus every word has $|w|_{321} = 0$, giving the forced value z_0 . If $i + j < b$, all words are canonical and the fixed-size part of Theorem 1 gives the count.

If $i + j = b$, the same theorem still provides $p_a(j)p_c(i)$ canonical words. A shared gap requires $p_* + q_* = b = j + i$. Equality in $p_* \leq j$, $q_* \leq i$ forces every positive sum to be concentrated in one part. Exactly one partition pair has this property. If both sums are positive, the shared gap contains one 1 and one 3, giving two words instead of one. If $j = 0$, it contains all a ones and one 3, giving $a + 1$ words. If $i = 0$, it contains one 1 and all c threes, giving $c + 1$ words. Subtracting the already counted canonical word proves (3.2). Since $b \geq 1$, the three boundary cases are disjoint. $\square$

The chamber is a sufficient condition for a forced triple coordinate; it is not claimed to be the largest such region. In particular, its boundary can lie outside the scalar range $d < bm$.

Theorem 3 (The first scalar and inversion corrections). *For positive a, b, c ,*

$$[t^{bm}](H_{a,b,c}(t, 1) - \mathcal{P}_{a,c}(t, 1)) = \begin{cases} m, & a \neq c, \\ b - 1 + 2m, & a = c = m \geq 2, \\ \binom{b+2}{2}, & a = c = 1. \end{cases} \quad (3.3)$$

Thus bm is exactly the first scalar degree of failure. In the inversion direction the first correction is

$$[q^{b+1}](H_{a,b,c}(t, q) - \mathcal{P}_{a,c}(t, q)) = \sum_{i+j=b} t^{ai+cj} - \sum_{i+j=b+1} t^{ai+cj}. \quad (3.4)$$

All sums are over nonnegative integers; equal exponents keep their multiplicities.

Proof. First suppose $M > 1$. If some endpoint sum exceeds b , select $p + q \geq b + 1$; both are positive. Assume $a \leq c$ by symmetry. The same row and column argument as (2.13) gives

$$d \geq b + (c - 1)p + (a - 1)q = b + (a - 1)(p + q) + (c - a)p \geq ab + c - 1 = bm + M - 1 > bm. \quad (3.5)$$

Hence every word at deficit bm is untruncated, with $cj + ai = bm$ and $i + j \leq b$. Every partition pair of that weighted size is feasible, so Theorem 2 evaluates the difference from the product. Only $i + j = b$ contributes. If $a < c$, these two equations force $j = 0, i = b$, giving the excess $a = m$. The other unequal case is symmetric. If $a = c = m \geq 2$, all $b + 1$ splits contribute: the $b - 1$ interior splits contribute one each and the two ends contribute m each.

If $a = c = 1$, deficit b means $z = 0$. Words with the 1 before the 3 must place them in a common 2-gap; there are $b + 1$ such words, equal to $[t^b](1 - t)^{-2}$. With 3 before 1, distribute the b twos before, between and after the endpoints. This gives $\binom{b+2}{2}$ extra words. Corollary 1 and the positivity of these corrections prove the first assertion.

For (3.4), let $r = |w|_{31}$ and $\text{inv} = b + 1$. The case $i + j < b$ is impossible, since it forces $r = 0$. There are two types:

- (A) $i + j = b + 1$ and $r = 0$. Of the $p_a(j)p_c(i)$ partition pairs, exactly one is not canonically realizable. Indeed failure requires $p_* + q_* > b$, which, since the total sum is $b + 1$, forces every positive sum into a single part. Conversely that pair fails. This includes a zero sum on one side: its other side then has the invalid part $b + 1$. The remaining pairs each give one word of deficit $ai + cj$.
- (B) $i + j = b$ and $r = 1$. A shared gap again forces both positive sums to concentrate. With both positive the shared gap is 31; with one zero it contains a single opposite letter

among the identical endpoint letters. Exactly one of its shuffles has one inversion. Thus each split gives exactly one additional word. No 2 intervenes in its reversed endpoint pair, so its 321 count is zero and its deficit is $ai + cj$.

Type (A) removes the second sum and type (B) supplies the first. $\square$

The signs in (3.4) compare enumerations; the right side is not itself a word-count polynomial. For $(a, b, c) = (3, 4, 2)$,

$$[q^5](H_{3,4,2} - \mathcal{P}_{3,2}) = t^8 + t^9 - t^{13} - t^{14} - t^{15}. \quad (3.6)$$

At $t = 1$ the correction is -1 , as is already implied by the classical Gaussian multinomial [7, Proposition 1.7.1]:

$$H_{a,b,c}(1, q) = \frac{\prod_{s=1}^{a+c} (1 - q^{b+s})}{\prod_{u=1}^a (1 - q^u) \prod_{v=1}^c (1 - q^v)}. \quad (3.7)$$

The numerator starts with $1 - q^{b+1}$; the refined formula identifies where that discrepancy occurs in triple deficit.

Example 1 (Scalar coefficients). For $(a, b, c) = (2, 4, 3)$,

$$\mathcal{P}_{2,3}(t, 1) = \frac{1}{(1 - t^2)(1 - t^3)(1 - t^4)(1 - t^6)^2}.$$

Its coefficients in degrees $0, \dots, 7$ are $1, 0, 1, 1, 2, 1, 5, 2$, the actual word counts at triple values $24, 23, \dots, 17$. At deficit 8 the product predicts 6 words and the actual count is 8; the first correction is $m = 2$.

Proposition 1 (Sharp next-degree reconstruction criterion). *Suppose $a \neq c$, and put $g = \gcd(a, c)$. Two words of content (a, b, c) can have $\text{inv} = b + 1$ and the same z , but different pair coordinates, if and only if*

$$|a - c| = g \quad \text{and} \quad b \geq m/g. \quad (3.8)$$

Proof. Use the two types in the preceding proof. Both have $d = ai + cj$, and within a type the pair (i, j) is determined by d , since $a \neq c$. By reverse-complement assume $a > c$. A type-(A) pair (i, j) and a type-(B) pair (i', j') have equal deficit exactly when $(a - c)(i' - i) = c$. Thus $a - c$ divides c , equivalent to $a - c = g$, and $b \geq i' \geq c/g$.

Conversely suppose these conditions hold. The type-(A) pair $(0, b + 1)$ has $p_a(b + 1) - 1 \geq 1$ words: $a \geq 2$ and the partition $(b, 1)$ is allowed. The type-(B) pair $(c/g, b - c/g)$ has one word by the preceding classification. They have equal deficit and distinct pair coordinates. $\square$

The inversion bound in Corollary 2 is therefore uniformly sharp, although a particular content may remain reconstructible at $b + 1$. Content $(1, 1, 2)$ gives 1332 and 2313, both with $(\text{inv}, z) = (2, 0)$ but pair coordinates $(1, 0)$ and $(0, 2)$. Content $(2, 2, 1)$ gives 21213 and 12312, both with $(\text{inv}, z) = (3, 1)$ and different pair coordinates.

4 Semigroup windows and the complete support

Write

$$S(a, c) = \{ai + cj : i, j \in \mathbb{N}\}, \quad \mathbb{N} = \{0, 1, 2, \dots\}, \quad \theta = (b - 1)m + a + c - 1 = bm + M - 1.$$

When the endpoints are coprime, $S(a, c)$ is a numerical semigroup. The gap encoding relates its small elements to triple deficits.

Lemma 4 (An untruncated realization criterion). *For nonnegative i, j , there are endpoint data of sums (j, i) with $p_* + q_* \leq b$ if and only if*

$$\lceil j/a \rceil + \lceil i/c \rceil \leq b. \quad (4.1)$$

For these data the triple deficit is $cj + ai$.

Proof. A list of a nonnegative integers of sum j has maximum at least $\lceil j/a \rceil$. This lower bound is attained by distributing j as evenly as possible among the a entries. Do the same for the c entries of sum i . The two lower bounds prove necessity, and these balanced lists prove sufficiency by Lemma 1. No summand is truncated when the two maxima sum to at most b , so (2.6) gives the deficit. $\square$

In matrix coordinates, put $Q = \lfloor x/a \rfloor$ and $T = \lfloor y/c \rfloor$. Since $\lceil j/a \rceil = b - Q$ and $\lceil i/c \rceil = b - T$, criterion (4.1) is equivalently $Q + T \geq b$. It characterizes when the affine bound (2.8) is attained at these pair counts. Thus the two formulations require only this one proof.

Theorem 4 (The semigroup window). *For every integer $0 \leq d \leq abc$ with $d < \theta$,*

$$abc - d \in \{|w|_{123} : w \in \mathcal{W}(a, b, c)\} \iff d \in S(a, c). \quad (4.2)$$

Every word in this window avoids 321.

Proof. If a word contains 321, some endpoint pair satisfies $p_r + q_s \geq b + 1$. The row and column estimate (3.5) gives $d \geq \theta$. Consequently a word in the stated window has $d = cj + ai$, proving necessity.

Conversely write $d = cj + ai$ with $i, j \geq 0$. We show that (4.1) holds. If $i, j > 0$ and $u = \lceil j/a \rceil$, $v = \lceil i/c \rceil$ have $u + v \geq b + 1$, then

$$d \geq c(a(u - 1) + 1) + a(c(v - 1) + 1) \geq ac(b - 1) + a + c > \theta.$$

If one index is zero, failure of (4.1) forces respectively $j > ab$ or $i > bc$, hence $d > abc$. Both alternatives contradict the hypotheses. Lemma 4 now realizes the deficit. $\square$

Example 2 (Why the window is strict). For $(a, b, c) = (3, 1, 5)$ one has $\theta = 7$ and $7 \notin S(3, 5)$. The word 113213333 nevertheless has $z = 8 = 15 - 7$. It contains a reverse triple. Thus the strict endpoint in Theorem 4 cannot be uniformly included.

4.1 Constructing the rest of the spectrum

We need a second view of the same word. For its t th two let A_t be the number of preceding ones and C_t the number of following threes. Then

$$0 \leq A_1 \leq \cdots \leq A_b \leq a, \quad c \geq C_1 \geq \cdots \geq C_b \geq 0, \quad z = \sum_{t=1}^b A_t C_t. \quad (4.3)$$

Every such pair of profiles is realizable: place the increments of A in the successive gaps as ones, and the decrements of C as threes, putting the remaining ones in the last gap and the remaining threes in the first gap. The orders within a gap do not change the profiles. Appending a two at the very end preserves z . Thus any count realized with b_0 twos is also realized with $b \geq b_0$ twos at the same endpoints.

Lemma 5 (A long untruncated interval). *Assume $a \leq c$. If $d \in S(a, c)$ and $d \leq (b - 1)ac + c$, then $abc - d$ is attainable by untruncated endpoint data.*

Proof. Choose $d = ai + cj$ with $i, j \geq 0$ and replace (i, j) by $(i + c, j - a)$ as often as possible. This leaves d unchanged and gives $0 \leq j < a$. If $j = 0$, then $i \leq bc$, so (4.1) holds. If $j > 0$, then

$$ai = d - cj \leq (b - 1)ac,$$

so $\lceil i/c \rceil \leq b - 1$ and $\lceil j/a \rceil = 1$. Apply Lemma 4. $\square$

Lemma 6 (Filling the bottom interval). *Assume $2 \leq a \leq c$. Every integer $0 \leq z < (a - 1)c$ is attainable under any one of these sufficient conditions:*

- (i) $b \geq 2$ and $c \geq 2a - 3$;
- (ii) $b \geq a - 1$;
- (iii) $b \geq a - 2 \geq 3$.

Proof. We construct profiles (4.3). Extra twos can be appended, so it is enough to use the least indicated value of b .

Case (i). Use two twos. Write $z = qc + r$ with $0 \leq q \leq a - 2$ and $0 \leq r < c$, and let t be the residue of r modulo $q + 1$, in $[0, q]$. Set

$$(A_1, A_2) = (q, q + 1), \quad (C_1, C_2) = \left(c - t, \frac{r + qt}{q + 1} \right).$$

The second coordinate is an integer, since $q \equiv -1 \pmod{q + 1}$. It is nonnegative, and $C_2 \leq C_1$ follows from

$$r + (2q + 1)t \leq c - 1 + (2q + 1)q \leq (q + 1)c,$$

using $c \geq 2q + 1$. The dot product is $qc + r = z$.

Case (ii). Put $A_t = 1$ for all t . Any integer $0 \leq z \leq bc$ is a sum of a nonincreasing b -tuple in $[0, c]$: write $z = bu + s$ with $0 \leq s < b$ and take s entries $u + 1$ and $b - s$ entries u . This covers the required interval when $b \geq a - 1$.

Case (iii). Work with $b = a - 2 \geq 3$. The all-ones profile of case (ii) covers $0 \leq z \leq bc$. For the remaining values write $z = bc + r$, $1 \leq r \leq c - 1$. If $r \leq c - 2$, choose $k' \in \{c, c - 1\}$ of the same parity as r , and put

$$A = (1^{b-1}, 2), \quad C = \left(c^{b-2}, k', \frac{2c + r - k'}{2} \right).$$

Exponents here mean repetition. The last entry is integral and nonnegative. It is at most k' : for $k' = c$ this follows from $r \leq c$; for $k' = c - 1$, parity excludes $r = c - 2$, so $r \leq c - 3$. The dot product is $bc + r$.

It remains to obtain $r = c - 1$. Take $A = (1^{b-2}, 2, 2)$. For odd c set

$$C = (c^{b-1}, (c - 1)/2),$$

and for even c set

$$C = (c^{b-3}, c - 1, c - 1, c/2 + 1).$$

These are nonincreasing profiles; in the even case $c \geq a \geq 5$ ensures $c/2 + 1 \leq c - 1$. Both dot products equal $(b + 1)c - 1$, as required. $\square$

Theorem 5 (Reflected Sylvester gaps). *Suppose $\gcd(a, c) = 1$, and define the sufficient threshold*

$$B(a, c) = \max \left(2, \left\lfloor \frac{(m-2)M+1}{m} \right\rfloor \right). \quad (4.4)$$

If $b \geq B(a, c)$, the set of unattainable triple counts in $[0, abc]$ is

$$\{abc - g : g \in \mathbb{N} \setminus S(a, c)\}. \quad (4.5)$$

It has exactly $(a-1)(c-1)/2$ elements. If $m \leq 2$, the same conclusion holds for every $b \geq 1$. The bound $B(a, c)$ is sufficient; no optimality is asserted.

Proof. By symmetry assume $a \leq c$. If $a = 1$, the all-ones profile $A = (1, \dots, 1)$ realizes every $0 \leq z \leq bc$, and $S(1, c) = \mathbb{N}$. Hence assume $a \geq 2$, so coprimality gives $a < c$.

We recall the two-generator gap calculation. The numbers cj , $0 \leq j < a$, occupy all residue classes modulo a . In the class of cj , the nonnegative representable numbers begin at cj and then proceed in steps of a : reduction of the coefficient of c modulo a proves minimality. The gaps in this class are the nonnegative numbers $cj - a, cj - 2a, \dots$. Consequently the largest gap is

$$F = ac - a - c,$$

and the total number of gaps is

$$\sum_{j=0}^{a-1} \lfloor cj/a \rfloor = \frac{c(a-1)}{2} - \frac{a-1}{2} = \frac{(a-1)(c-1)}{2}.$$

The middle equality uses that the fractional parts of cj/a are $0, 1/a, \dots, (a-1)/a$ in some order.

First suppose $b \geq B(a, c)$. Since $b \geq \lfloor ((a-2)c+1)/a \rfloor$, we have

$$\theta = ab + c - 1 > ac - a - c = F. \quad (4.6)$$

Every semigroup gap therefore falls in the window of Theorem 4 and gives a missing count $abc - g$. The same theorem realizes every other count with deficit less than θ .

All integers $d > F$ belong to the semigroup. Lemma 5 therefore realizes the entire interval

$$(a-1)c \leq z \leq abc - F - 1.$$

Together with the window, this covers all counts from $(a-1)c$ upwards except precisely the reflected gaps. It remains to fill the bottom. If $c \geq 2a-3$, use Lemma 6(i). Otherwise $a < c \leq 2a-4$, so $a \geq 5$. If $c = a+1$, formula (4.4) gives $B = a-2$ and Lemma 6(iii) applies. If $c \geq a+2$, then

$$(a-2)c+1 \geq a^2-3 \geq a(a-1),$$

so $b \geq a-1$ and Lemma 6(ii) applies. This completes the asserted support description for $b \geq B$.

For $a = 2$ and any $b \geq 1$, the same argument works without the bound $b \geq B$: here $\theta = 2b+c-1 > F = c-2$, and Lemma 6(ii) fills the bottom because $b \geq a-1$. The preceding $a = 1$ argument covers the remaining small endpoint case. $\square$

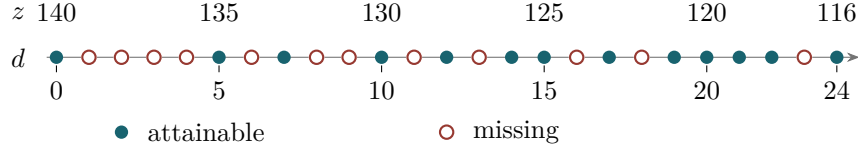


Figure 2: The top of the spectrum for $(a, b, c) = (5, 4, 7)$. Here $B = 4$, $abc = 140$ and the twelve gaps of $\langle 5, 7 \rangle$ are 1, 2, 3, 4, 6, 8, 9, 11, 13, 16, 18, 23. A hollow point at deficit d means the count $z = 140 - d$ is missing. Every triple count below this pictured range is attained.

4.2 A noncoprime contrast

Coprimality in Theorem 5 is substantial. Truncation can create residues absent from $S(a, c)$, but does not in general fill the whole interval. The smallest symmetric example has an exact description.

Proposition 2 (The complete $(2, b, 2)$ spectrum). *For every $b \geq 0$,*

$$\{|w|_{123} : w \in \mathcal{W}(2, b, 2)\} = \{0, 0 + 1, \dots, 2b\} \cup \{2t : 0 \leq t \leq 2b\}. \quad (4.7)$$

Proof. For $0 \leq t \leq 2b$, put $u = \lfloor t/2 \rfloor$, $v = \lceil t/2 \rceil$, and form

$$U_t = 2^u 3 2^{v-u} 3 2^{b-v}.$$

This binary word has $|U_t|_{23} = u + v = t$. The words $1U_t1$ and $11U_t$ realize respectively t and $2t$.

Conversely use (4.3) with $A_t, C_t \in \{0, 1, 2\}$. If z is odd, at least one product $A_t C_t$ is odd, hence equals 1. Before this position $A_s \leq 1$ and after it $C_s \leq 1$. Thus all other products are at most 2, giving $z \leq 2b - 1$. An even count is at most $4b$. These restrictions are exactly the right-hand side of (4.7). For $b = 0$ the sum is empty and the same conclusion holds. $\square$

5 The first exterior fibres and reverse triples

5.1 The first layer outside the chamber

The chamber $i + j \leq b$ has a single triple coordinate for each pair (i, j) . One step outside it, precisely one exceptional word moves to a second coordinate. We retain endpoint multiplicities at least two in the following statement.

Theorem 6 (First exterior fibres). *Let $a, c \geq 2$, $i, j \geq 2$, and $b = i + j - 1$. Set $z_0 = abc - cj - ai$. For the fixed pair counts $x = ab - j$, $y = bc - i$,*

$$\rho_{a,b,c}(x, y, z) = \begin{cases} p_a(j)p_c(i) + 1 + \mathbf{1}_{\{j=2\}} + \mathbf{1}_{\{i=2\}}, & z = z_0, \\ 1, & z = z_0 + 1, \\ 0, & \text{otherwise.} \end{cases} \quad (5.1)$$

The unique word at $z_0 + 1$ is

$$1^{a-1} 2^{j-1} 3 2 1 2^{i-1} 3^{c-1}. \quad (5.2)$$

Proof. Every partition pair of sums j, i has entries at most $b - 1$, so is valid endpoint data. Put $E = \sum_{r,s} (p_r + q_s - b)_+ = |w|_{321}$. Since the total endpoint sum is $b + 1$, a positive summand forces all other endpoint entries to vanish. There is precisely one such partition

pair: one positive entry j on the one side and one i on the three side. Its excess is 1, and its occupied gaps are $0, j-1, j, b$. They are distinct because $i, j \geq 2$, so it gives exactly one word, namely (5.2), at $z_0 + 1$.

Every remaining pair has $E = 0$ and gives $z = z_0$. A mixed gap requires $p + q = b$, with both entries positive. The mass outside these entries is 1. There are exactly two possibilities:

$$p = (j-1, 1, 0, \dots), \quad q = (i, 0, \dots), \quad \text{or} \quad p = (j, 0, \dots), \quad q = (i-1, 1, 0, \dots),$$

where the display suppresses the chosen ordering of each multiset. In the first case, the shuffle count is 2 if $j > 2$, and 3 if $j = 2$; its excess over the baseline count one is $1 + \mathbf{1}_{\{j=2\}}$. The other case contributes $1 + \mathbf{1}_{\{i=2\}}$. Starting with $p_a(j)p_c(i)$ pairs, remove the single crossing pair and add these two shuffle corrections. This gives the first line of (5.1). No other triple coordinate is possible. Finally $z_0 \geq 0$: at $a = c = 2$ it equals $2b - 2 > 0$, and increasing a or c changes it by $bc - i > 0$ or $ab - j > 0$, respectively. $\square$

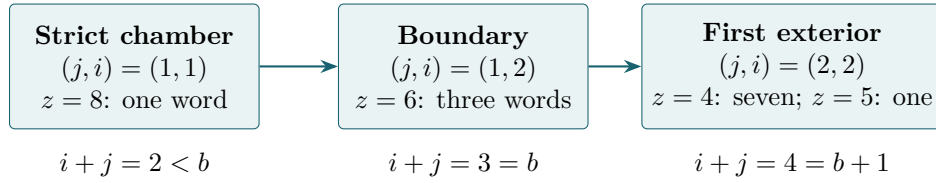


Figure 3: A chamber transition at content $(2, 3, 2)$. A shared-gap shuffle first increases a fibre on the boundary; outside, a crossing endpoint pair separates a second triple count. Table 1 lists every word in these three examples.

j	i	z	Words of content $(2, 3, 2)$
1	1	8	1212323
1	2	6	1212332, 1213223, 1231223
2	2	4	1221332, 1223132, 1223312, 2112332, 2113223, 2131223, 2311223
2	2	5	1232123

Table 1: Complete fibres used in Figure 3. No factorial factor for repeated letters is present: every displayed word occurs once.

5.2 Minimizing the deficit of a reverse triple

Theorem 7 (The first reverse triples in deficit). *Let $a, c \geq 2$ and $b \geq 1$. Among words containing 321, the minimum triple deficit is $\theta = (b-1)m + a + c - 1$. Every word attaining this minimum contains exactly one 321. The number of minimizing words is*

$$\begin{cases} ac, & b = 1, \\ m, & b \geq 2, \quad a \neq c, \\ b + 2m - 2, & b \geq 2, \quad a = c. \end{cases} \quad (5.3)$$

Proof. Assume $a \leq c$. A chosen reverse triple has an endpoint pair p, q with $p + q \geq b + 1$. Its selected row and column in the deficit sum give

$$d \geq b + (c-1)p + (a-1)q \geq b + (a-1)(b+1) = \theta.$$

Equality forces $p + q = b + 1$ because $a - 1 > 0$, and forces $p = 1, q = b$ if $a < c$. It also forces every contribution outside the selected row and column to vanish. Since $a, c \geq 2$, there is at least one endpoint on each side outside the selected pair. Their off-row, off-column intersection has zero contribution, forcing both entries to zero; the remaining such entries must likewise vanish. Thus the only positive entries are one $p \in [1, b]$ and one $q = b + 1 - p$. Conversely each of these concentrated data has deficit θ when $a = c$, or when $p = 1$ if $a < c$. Their reverse-triple count is $p + q - b = 1$.

If $b = 1$, the data are $p = q = 1$. Gap 0 contains $a - 1$ ones and one three, with a shuffles; gap 1 contains one one and $c - 1$ threes, with c shuffles. This gives ac words. For $b \geq 2$, if $a < c$, only $p = 1, q = b$ is allowed, with $a = m$ shuffles in gap 0 and all other gaps unmixed. If $a = c = m$, the choices $p = 1$ and $p = b$ each give m words. Each of the $b - 2$ interior choices has distinct occupied gaps and gives one word. This proves (5.3); symmetry handles $c < a$. $\square$

Remark 1 (Endpoint multiplicity one). The same minimum θ and a unique minimizing word hold when $m = 1 < M$ and $b \geq 2$. For $a = 1 < c$, the row estimate is $d \geq b + (c - 1)p$, whose minimum forces $p = 1$ and a selected $q = b$. Equality requires all other $q_s = 0$: their summands $\min(1 + q_s, b)$ must equal 1, and $b \geq 2$. The resulting word is $3212^{b-1}3^{c-1}$, with no mixed gap. Reverse-complement gives the other endpoint case.

The exclusions for $b = 1$ or $a = c = 1$ are real. With content $(1, 1, c)$, every word containing 321 has $z = 0$, and there are $c(c + 1)/2$ such words. Indeed, with $h \geq 1$ threes before the two, put the single one after that two; the remaining $c - h$ threes have $c - h + 1$ possible splits around the one. With content $(1, b, 1)$, reverse triples require the three to precede the one with $h \geq 1$ twos between them, giving $\sum_{h=1}^b (b - h + 1) = b(b + 1)/2$ words, again all with $z = 0$. For example, content $(1, 1, 2)$ has 3213, 3231, 3321; the last word contains two reverse triples.

5.3 Minimizing inversions instead

Proposition 3 (The first reverse triples in inversions). *For all positive a, b, c , a word containing 321 has $\text{inv} \geq b + 2$. There are exactly b words attaining this bound:*

$$1^{a-1}2^{j-1}3212^{i-1}3^{c-1}, \quad i, j \geq 1, \quad i + j = b + 1. \quad (5.4)$$

Each contains exactly one 321, and their deficit enumerator is

$$\sum_{\substack{w \in \mathcal{W}(a,b,c) \\ |w|_{321} > 0, \text{ inv}(w) = b+2}} t^{abc - |w|_{123}} = \sum_{i=1}^b t^{ai + c(b+1-i)-1}. \quad (5.5)$$

For $a, c \geq 2$, these inversion-minimizing words overlap the Theorem 7 minimizers in b words if $a = c$, and in one word otherwise.

Proof. A reverse triple gives $p_r + q_s \geq b + 1$, so $i + j \geq b + 1$. Also $|w|_{31} \geq 1$, and (2.3) gives the lower bound. Equality requires $i + j = b + 1$ and $|w|_{31} = 1$. The chosen endpoint pair exhausts the entire sum of all endpoint entries: there is one positive $p = j$, one positive $q = i$, and all others vanish. Because $p + q = b + 1$, the selected three and one occupy adjacent gaps, with exactly one two between them. The other ones are in gap 0 and the other threes in gap b . If a boundary gap is mixed, its order is uniquely fixed by $|w|_{31} = 1$. These observations give exactly (5.4). Its deficit is $ai + cj - 1$ by (2.6); the b values of i give distinct words. The displayed polynomial follows. Finally this deficit is minimized over i at a unique endpoint if $a \neq c$, and is constantly θ if $a = c$. $\square$

For example, with $(a, b, c) = (3, 4, 2)$ the inversion-minimal reverse words have $k = 6$ and deficit polynomial $t^{10} + t^{11} + t^{12} + t^{13}$. Minimizing inversions and maximizing the triple count subject to a reverse triple are therefore different optimization problems.

6 Verification and formal scope

The arguments above prove the statements for arbitrary parameters in their specified domains. Finite computation serves two further purposes: it tests boundary interpretations and makes the examples reproducible. The accompanying standard-library Python checks enumerate actual words, compute subword counts directly, and compare them with the partition and profile formulas. A separate article check verifies every displayed word example, the complete fibres in Table 1, and the semigroup picture by an independent enumeration of monotone profiles. These checks are not substitutes for the proofs.

The repository also contains a Lean 4 certificate using the standard library. Its scope is narrower than the present paper. It formalizes the strict-chamber actual-word bijection and cardinality, scalar separation, finite strict scalar and joint convolutions, and strict linear reconstruction. It also proves the complete $(2, b, 2)$ spectrum of Proposition 2, including $b = 0$, and the lower bound $\text{inv} \geq b + 2$ for a word containing 321. The full $k \leq b$ bijection, the boundary and exterior corrections, the general semigroup window and eventual support theorem, the sharp next-degree reconstruction criterion, and the equality classifications have human-readable proofs here but are not all formalized in Lean. Nor does the certificate establish the infinite formal-power-series identities as Lean theorems. This distinction matters: a finite coefficient identity and an infinite product identity are different formal objects, even when the mathematical passage between them is standard.

7 Further directions

Three questions arise directly from these proofs. First, the sufficient middle-multiplicity threshold in (4.4) need not be the least one. Determining the smallest b for which only reflected semigroup gaps remain would sharpen the support theorem. The constructive profiles in Lemma 6 give a starting point.

Second, at larger $i + j - b$ more than one endpoint pair can cross, and mixed gaps can interact. Exact fibre polynomials in those layers would extend (5.1). The inversion correction in Theorem 3 suggests retaining both d and k rather than discarding inversions at the outset.

Third, noncoprime endpoints require a description combining semigroup restrictions near the maximum with residue creation by truncation. Proposition 2 is a complete test case, but its interval plus progression has no asserted general analogue. Four-letter words, including ordered encodings of DNA, add coupled higher subword counts; the present paper makes no biological discrimination claim. A useful extension would first establish an exact four-letter analogue of the endpoint encoding and identify which joint statistics still factor.

Acknowledgements and disclosure

Generative AI tools, including OpenAI Codex and Anthropic Claude, were used to assist mathematical exploration, drafting, computational checks and the development of formal proofs. Mathematical responsibility and authorship rest with Christopher Yamba. The use of these tools does not constitute an independent human peer review.

References

- [1] A. Claesson, V. Jelínek and E. Steingrímsson, *Upper bounds for the Stanley–Wilf limit of 1324 and other layered patterns*, arXiv:1111.5736, 2011. arxiv.org/abs/1111.5736.
- [2] S. Linusson and E. Verkama, *Enumerating 1324-avoiders with few inversions*, arXiv:2408.15075, 2024. arxiv.org/abs/2408.15075.
- [3] A. Mateescu, A. Salomaa, K. Salomaa and S. Yu, A sharpening of the Parikh mapping, *RAIRO – Theoretical Informatics and Applications* **35** (2001), no. 6, 551–564.
- [4] A. Mateescu, A. Salomaa and S. Yu, *Subword histories and Parikh matrices*, TUCS Technical Report 442, Turku Centre for Computer Science, February 2002. TUCS publication record.
- [5] A. Salomaa, Parikh matrices: subword indicators and degrees of ambiguity, in H.-J. Böckenhauer, D. Komm and W. Unger (eds.), *Adventures Between Lower Bounds and Higher Altitudes*, Lecture Notes in Computer Science **11011**, Springer, 2018, 100–112.
- [6] C. D. Savage and H. S. Wilf, *Pattern avoidance in compositions and multiset permutations*, arXiv:math/0504310, 2005. arxiv.org/abs/math/0504310.
- [7] R. P. Stanley, *Enumerative Combinatorics*, vol. 1, second edition, Cambridge University Press, Cambridge, 2012.
- [8] C. Yamba, *Parikh Matrices and Permutation Statistics*, Examensarbeten i matematik 2006:8, Matematiska institutionen, Stockholms universitet, 2006.
- [9] M. Yang, *Enumeration of words that contain the pattern 123 exactly once*, arXiv:1712.08662, 2017. arxiv.org/abs/1712.08662.