Trace Fibres of Marked Interval Heaps

Marc Craig
Independent researcher, London, United Kingdom

30 August 2026
Version 3

Abstract

Let a marked interval pass append one letter to each site in a nonempty interval of a finite path. Two histories are identified when adjacent passes with disjoint supports are interchanged. A history therefore defines a Cartier–Foata trace, or equivalently a Viennot heap, and determines at each site an ordered word of letters. We call the resulting tuple of words a stratigraphy.

For a fixed stratigraphy Λ, let $Q_\Lambda(t)$ count trace classes that induce Λ, graded by their number of passes. We prove an exact factorisation of Q_Λ into polynomials associated with adjacent pairs of words. The coefficients of the local factors count matching pairs of common-subsequence embeddings. In particular, if $\Lambda = (w_1, \ldots, w_n)$ has total length M, then the minimum number of passes is

$$M - \sum_{i=1}^{n-1} \mathrm{LCS}(w_i, w_{i+1}).$$

The same argument gives a corresponding product over the edges of a tree, for passes supported on connected subtrees.

We also determine the support and endpoint coefficients of Q_Λ and give an infinite binary family for which its coefficients are not log-concave. An exhaustive computation shows that the least mass of a binary counterexample is 14; up to exchanging the two letters, the unique example of that mass is $(abababa, abababa)$.

1 Introduction

Cartier–Foata traces and Viennot heaps encode sequences of partially commuting operations [3, 11]. When the pieces are intervals of a path and disjoint intervals commute, the underlying model is the classical heap-of-segments model [2, 8]. We consider an inverse enumeration problem for this model.

Each interval carries a letter. A heap is observed locally by reading, from bottom to top, the letters of the pieces meeting each site. This observation forgets which occurrences at different sites belong to the same interval piece. For prescribed local words, we enumerate the fibre of compatible heaps by number of pieces.

For words u and v, let $\mathrm{CS}_k(u, v)$ be the number of pairs of k-subsets of positions that induce the same subword. These are the matching joint embeddings studied by Elzinga, Rahmann and Wang [5]. Define

$$R(u, v; t) = \sum_{k \geq 0} \mathrm{CS}_k(u, v)\, t^{|v|-k}.$$

Our main result is the factorisation

$$Q_\Lambda(t) = t^{|w_1|} \prod_{i=1}^{n-1} R(w_i, w_{i+1}; t). \tag{1}$$

It reduces the complete fibre enumeration to independent common-subsequence calculations at adjacent boundaries. Proposition 4.3 shows that the same product formula holds over the edges of a tree, when passes are supported on connected subtrees; the resulting polynomial does not depend on the choice of root.

The numbers $\mathrm{CS}_k(u,v)$ are known, but the present use of their generating polynomial as a local factor for a prescribed heap fibre does not appear in the literature cited above. The distinction from Chase's theorem [4] is important: Chase counts distinct subsequence words of one word, whereas $\mathrm{CS}_k(u,v)$ counts pairs of embeddings in two words.

The paper is organised as follows. In section 2 we define the model. In section 3 we identify trace classes with consistent partitions of a cell diagram. The factorisation is proved in section 4, together with its extension to tree-supported passes. Its principal consequences are recorded in section 5. In section 6 we give an infinite family of failures of log-concavity. The finite computation establishing the least binary counterexample is described in section 7.

2 Marked interval heaps and stratigraphies

Fix a finite alphabet $\mathcal{A}$ and the path of sites $[n] = \{1, \dots, n\}$.

Definition 2.1. A *pass* is a pair (I, a), where $I = [\ell, r] \subseteq [n]$ is a nonempty interval and $a \in \mathcal{A}$. A *history* is a finite sequence of passes.

Two passes are *independent* if their supports are disjoint. Histories are *trace-equivalent* if one can be obtained from the other by interchanging adjacent independent passes. A *trace* is an equivalence class of histories.

A history H determines a word w_x at each site x: read in chronological order the letters of all passes whose supports contain x. The tuple

$$\Lambda(H) = (w_1, \dots, w_n)$$

is its *stratigraphy*. We assume throughout that every w_x is nonempty. Write

$$d_x = |w_x|, \qquad M = |\Lambda| = \sum_{x=1}^{n} d_x.$$

The map $H \mapsto \Lambda(H)$ is constant on trace classes.

Definition 2.2. Let $\mathcal{T}(\Lambda)$ be the set of trace classes inducing Λ. The *trace-fibre polynomial* is

$$Q_\Lambda(t) = \sum_{T \in \mathcal{T}(\Lambda)} t^{|T|},$$

where $|T|$ is the number of passes in any representative of T.

The observation map from traces to stratigraphies is not injective. For example, the stratigraphy (a, a) is induced either by one pass on $[1, 2]$ or by two singleton passes. Hence $Q_{(a,a)}(t) = t + t^2$.

Proposition 2.3. *For every stratigraphy of mass M,*

$$\deg Q_\Lambda = M, \qquad [t^M] Q_\Lambda(t) = 1.$$

Proof. Every pass contributes at least one local letter, so no realisation has more than M passes. Equality forces all passes to have singleton support. Their order at each site is fixed by w_x, while passes at distinct sites commute. Thus there is exactly one trace with M passes. □

3 A partition model

For $\Lambda = (w_1, \dots, w_n)$ define its set of cells by

$$\mathrm{Cells}(\Lambda) = \{(x, j) : 1 \le x \le n,\ 1 \le j \le d_x\}.$$

The cell (x, j) has letter $w_x[j]$.

Definition 3.1. A *piece* is a set containing exactly one cell from each site of a nonempty interval $I \subseteq [n]$, all carrying the same letter. Its *support* is I.

A partition $\mathcal{P}$ of $\mathrm{Cells}(\Lambda)$ into pieces is *consistent* if, whenever two pieces meet at more than one site, their vertical order is the same at every common site.

Theorem 3.2. *Trace classes inducing Λ are in bijection with consistent partitions of* $\mathrm{Cells}(\Lambda)$ *into pieces. The number of passes equals the number of parts.*

Proof. A history partitions the cells according to the pass that created them. Each block is a piece. Two passes meeting at several sites have the same chronological order at all those sites, so the partition is consistent. Interchanging disjoint passes does not alter it.

Conversely, let $\mathcal{P}$ be a consistent piece partition. Form a directed graph on its pieces: if P and Q meet, orient $P \to Q$ when the cell of P lies below the cell of Q at a common site. Consistency makes this orientation well defined.

The directed graph is acyclic. Otherwise choose a shortest directed cycle. The intersection graph of intervals is chordal. A cycle of length at least four therefore has a chord; either orientation of the chord, together with one of the directed paths on the cycle, gives a shorter directed cycle. A shortest directed cycle must consequently be a triangle. Its three interval supports pairwise intersect, and intervals have the Helly property, so they share a site. At that site the three corresponding cells are totally ordered, which excludes a directed triangle.

Choose a linear extension of the resulting partial order and perform the pieces in that order. The local order at each site is exactly the order prescribed by Λ. Any two linear extensions differ by interchanges of adjacent incomparable elements. Overlapping pieces are comparable, so such interchanges involve disjoint supports and preserve the trace. Thus the partition determines one trace class. The two constructions are inverse. $\square$

4 Adjacent-boundary factorisation

If $u = u_1 \cdots u_p$ and $A = \{a_1 < \cdots < a_k\} \subseteq [p]$, write $u|_A = u_{a_1} \cdots u_{a_k}$.

Definition 4.1. For words u and v, put

$$\mathrm{CS}_k(u, v) = \#\{(A, B) : |A| = |B| = k,\ u|_A = v|_B\}$$

and

$$R(u, v; t) = \sum_{k=0}^{\min(|u|,|v|)} \mathrm{CS}_k(u, v)\, t^{|v|-k}. \tag{2}$$

Equivalently,

$$\mathrm{CS}_k(u, v) = \sum_{z \in \mathcal{A}^k} \mathrm{emb}(z, u)\ \mathrm{emb}(z, v),$$

where $\mathrm{emb}(z, u)$ denotes the number of embeddings of z as a subsequence of u.

Theorem 4.2 (Factorisation theorem)**.** *For every stratigraphy $\Lambda = (w_1, \dots, w_n)$,*

$$Q_\Lambda(t) = t^{d_1} \prod_{x=1}^{n-1} R(w_x, w_{x+1}; t). \tag{3}$$

Proof. Let $\mathcal{P}$ be a consistent piece partition. At the boundary between sites x and $x+1$, record the cells belonging to pieces that cross that boundary. If there are k_x such pieces, their cells determine subsets

$$A_x \subseteq [d_x], \qquad B_{x+1} \subseteq [d_{x+1}], \qquad |A_x| = |B_{x+1}| = k_x.$$

Consistency forces the matching of the two subsets to preserve their orders, and constancy of the letter on a piece gives $w_x|_{A_x} = w_{x+1}|_{B_{x+1}}$. There are therefore $\mathrm{CS}_{k_x}(w_x, w_{x+1})$ choices at boundary x.

Conversely, choose independently at every boundary a pair of subsets satisfying this equality, and match them in increasing order. Join the corresponding cells across adjacent sites. Each cell has at most one edge to its left and at most one edge to its right. Every connected component is consequently a path meeting one cell at each site of a contiguous interval. Its letters agree, so it is a piece. Order preservation at every boundary makes the resulting partition consistent.

These constructions are inverse. The cell graph has M vertices and $\sum_x k_x$ edges and is a disjoint union of paths. Its number of components is therefore

$$M - \sum_{x=1}^{n-1} k_x = d_1 + \sum_{x=1}^{n-1} (d_{x+1} - k_x).$$

Summing t to this power over the independent choices at all boundaries gives (3). $\square$

Proposition 4.3 (Tree extension). *Let G be a finite tree rooted at r. Replace interval passes by marked passes supported on nonempty connected subtrees of G, and define local words and trace equivalence as above. Then*

$$Q_\Lambda^G(t) = t^{|w_r|} \prod_{(p(v),v)\in E(G)} R(w_{p(v)}, w_v; t), \tag{4}$$

where every edge is oriented from a parent $p(v)$ to its child v.

Proof. Families of subtrees of a tree have the Helly property, and their intersection graphs are chordal [6]. The proof of theorem 3.2 therefore applies verbatim to the static piece partitions.

Choose an order-preserving common-subsequence matching independently across each oriented edge of G. In the resulting cell graph a cell has at most one incident matching edge over any fixed edge of G; consequently a walk that arrives at a cell along a tree edge e cannot immediately return along e. A path joining two cells over the same vertex, or a cycle in the cell graph, would therefore project to a nontrivial closed walk in a tree without an immediate backtrack, which is impossible. Hence each component meets every tree vertex at most once and has connected support. It is therefore a valid subtree piece. Order preservation on the edges gives consistency, and the inverse construction is the same as in theorem 4.2.

If k_e cells are matched across edge e, the number of components is $M - \sum_e k_e$. Since

$$M - \sum_e k_e = |w_r| + \sum_{(p(v),v)\in E(G)} \big(|w_v| - k_{p(v),v}\big),$$

summing over the independent edge choices gives (4). $\square$

Remark 4.4 (Root independence). The right-hand side of (4) does not depend on the choice of root. It suffices to treat a move of the root from r to an adjacent vertex s, since any two vertices of a tree are joined by a path. Such a move reverses the orientation of the edge $\{r, s\}$ and leaves every other parent–child orientation unchanged. The two expressions therefore differ only in the factors $t^{|w_r|}R(w_r, w_s; t)$ and $t^{|w_s|}R(w_s, w_r; t)$, which are equal by the identity used in corollary 5.3.

4.1 Computation of a local factor

Let

$$F_{i,j}(z) = \sum_{k\geq 0} \mathrm{CS}_k(u_1\cdots u_i,\, v_1\cdots v_j)\, z^k.$$

Then $F_{i,0} = F_{0,j} = 1$ and

$$F_{i,j} = F_{i-1,j} + F_{i,j-1} - F_{i-1,j-1} + \mathbf{1}_{u_i=v_j}\, z\, F_{i-1,j-1}. \tag{5}$$

Thus all coefficients of $R(u, v; t)$ are computable by a standard length-graded common-subsequence dynamic programme [5].

5 Consequences

Corollary 5.1 (Minimum number of passes). *For $M = \sum_x |w_x|$,*

$$\min\{|T| : T \in \mathcal{T}(\Lambda)\} = M - \sum_{x=1}^{n-1} \mathrm{LCS}(w_x, w_{x+1}). \tag{6}$$

This number is computable in $O\big(\sum_x d_x d_{x+1}\big)$ time.

Proof. In (3), minimising the exponent is equivalent to maximising the selected common-subsequence length at every boundary. The largest k for which $\mathrm{CS}_k(u, v) > 0$ is $\mathrm{LCS}(u, v)$. □

For two sites, (6) is the familiar shortest-common-supersequence identity $|u|+|v|-\mathrm{LCS}(u, v)$. For contrast, the shortest common supersequence problem for an unbounded number of input words is NP-complete even over a binary alphabet [9].

Corollary 5.2 (Support and endpoint coefficients). *Set $L_x = \mathrm{LCS}(w_x, w_{x+1})$ and let μ_Λ denote the quantity in (6). Then*

$$\operatorname{supp} Q_\Lambda = \{\mu_\Lambda, \mu_\Lambda + 1, \ldots, M\},$$

and

$$[t^{\mu_\Lambda}]Q_\Lambda(t) = \prod_{x=1}^{n-1} \mathrm{CS}_{L_x}(w_x, w_{x+1}), \qquad [t^M]Q_\Lambda(t) = 1, \qquad Q_\Lambda(1) = \prod_{x=1}^{n-1}\sum_{k\geq 0} \mathrm{CS}_k(w_x, w_{x+1}).$$

Proof. If a common subsequence of length L exists, deleting letters from it gives a common subsequence of every length from 0 to L. Each factor in (3) therefore has contiguous positive support. The formulas follow by taking the appropriate endpoint terms and by setting $t = 1$. □

Corollary 5.3 (Symmetries). *The polynomial Q_Λ is invariant under a common relabelling of the alphabet, simultaneous reversal of all words, and reflection of the path.*

Proof. The first two operations preserve every CS_k. Reflection follows from $\mathrm{CS}_k(u, v) = \mathrm{CS}_k(v, u)$ and

$$t^{|u|}R(u, v; t) = t^{|v|}R(v, u; t).$$

□

Proposition 5.4 (Visible complexity does not bound stratigraphic complexity). *Let V_Λ be the word formed by the last letter of each stack, and let $\pi(V)$ be the minimum number of ordinary interval overwrites needed to produce a visible word V from a blank path. For every $m \geq 1$ there is a stratigraphy Λ_m on $2m$ sites such that*

$$\pi(V_{\Lambda_m}) = 1, \qquad \mu_{\Lambda_m} = 2m + 1.$$

Proof. Take

$$\Lambda_m = (ba,\ aa,\ ba,\ aa,\ \ldots,\ ba,\ aa).$$

Its mass is $4m$, and every adjacent pair of stacks has longest common subsequence of length 1. Hence corollary 5.1 gives $\mu_{\Lambda_m} = 4m - (2m-1) = 2m+1$. Every final letter is a, so one full-width overwrite produces the visible word. □

Remark 5.5 (Chain interpretation)**.** Let

$$P(u,v) = \{(i,j) : u_i = v_j\},$$

ordered by $(i,j) < (i',j')$ if $i < i'$ and $j < j'$. Then $\mathrm{CS}_k(u,v)$ is the number of k-element chains in $P(u,v)$. If both words are permutations of the same alphabet, these are the increasing-subsequence counts of the associated permutation. This per-permutation statistic is different from the distribution, over all permutations, of the length of a longest increasing subsequence studied in [1].

6 Failure of log-concavity

A nonnegative sequence (a_i) is *log-concave* if $a_i^2 \geq a_{i-1}a_{i+1}$ at every internal index. Let

$$w_k = (ab)^k a, \qquad N = |w_k| = 2k+1, \qquad \Lambda_k = (w_k, w_k).$$

Lemma 6.1. *For $k \geq 1$,*

$$\mathrm{CS}_N(w_k,w_k) = 1, \qquad \mathrm{CS}_{N-1}(w_k,w_k) = N, \qquad \mathrm{CS}_{N-2}(w_k,w_k) = 6k^2 - k.$$

Proof. The first identity is immediate. Deleting one position from an alternating odd word gives a word that determines the deleted position: an internal deletion is identified by the resulting doubled adjacent letter, and an endpoint deletion by the changed endpoint. Thus equal subsequences of length $N-1$ arise only from equal deletion positions. This gives N pairs.

For length $N-2$, the $N-1 = 2k$ adjacent pairs of deleted positions all leave the same alternating word and hence contribute $(2k)^2$ ordered pairs. Each nonadjacent deletion pair leaves a distinct word and contributes only its diagonal pair. Therefore

$$\mathrm{CS}_{N-2} = (N-1)^2 + \binom{N}{2} - (N-1) = 6k^2 - k. \qquad \square$$

Theorem 6.2. *For every $k \geq 3$, the coefficients of $Q_{\Lambda_k}(t)$ are not log-concave.*

Proof. By theorem 4.2, $Q_{\Lambda_k}(t) = t^N R(w_k, w_k; t)$. Its first three nonzero coefficients are, by lemma 6.1,

$$1, \qquad 2k+1, \qquad 6k^2 - k.$$

The first log-concavity margin is

$$(2k+1)^2 - (6k^2 - k) = -2k^2 + 5k + 1,$$

which is negative for every integer $k \geq 3$. □

At $k = 3$ this gives

$$Q_{(abababa,abababa)}(t) = t^7 + 7t^8 + 51t^9 + 135t^{10} + 197t^{11} + 117t^{12} + 25t^{13} + t^{14}. \tag{7}$$

The first inequality fails since $7^2 < 51$.

7 The least binary counterexample

The next statement is computer-assisted.

Theorem 7.1. *Over a binary alphabet, every stratigraphy of mass at most* 13 *has a log-concave trace-fibre polynomial. At mass* 14 *there are exactly two labelled failures,*

$$(abababa, abababa) \qquad \text{and} \qquad (bababab, bababab).$$

Consequently (7) *is the unique least binary counterexample up to exchanging the two letters.*

Exhaustive range

For fixed mass M and n nonempty sites, a binary stratigraphy is specified by a composition of M into n positive parts and a binary label on every cell. Its number is $\binom{M-1}{n-1}2^M$. Summing over n gives 2^{2M-1} stratigraphies of mass M. The computation evaluated all

$$\sum_{M=1}^{13} 2^{2M-1} = 44{,}739{,}242$$

stratigraphies of mass at most 13, with no failure, and all $2^{27} = 134{,}217{,}728$ stratigraphies of mass 14, with exactly the two failures stated above. No symmetry reduction or heuristic pruning was used.

As a second exhaustive calculation, every ordered pair of nonempty binary words with total length at most 14 was evaluated. All 180,228 pairs of total length at most 13 had log-concave local factors. Among the 212,992 pairs of total length 14, only the two alternating pairs failed. Products of nonnegative log-concave sequences without internal zeros are log-concave [7, 10]. Moreover, in a mass-14 stratigraphy with at least three sites, every adjacent pair has total length at most 13. The factor scan therefore proves the entire conclusion of theorem 7.1, not merely the lower bound: a mass-14 failure must have two sites and must be one of the two alternating pairs.

Cross-checks

Four implementations were compared on overlapping complete ranges:

(M1) direct history enumeration followed by canonical trace reduction;

(M2) dynamic programming over Cartier–Foata layers;

(M3) subset dynamic programming on consistent cell partitions;

(M4) the product formula (3) with recurrence (5).

All four agreed on 890 stratigraphies with $n \leq 3$ and $M \leq 6$. Methods M2–M4 agreed on 71,690 stratigraphies with $n \leq 4$ and $M \leq 9$. An independently written Python layer computation and a C product computation agreed on every (n, M) ledger cell through mass 8. Five deliberately corrupted variants of the recurrences were rejected by the checkpoint suite.

These comparisons do not constitute four independent confirmations of theorem 4.2. Methods M1 and M2 work from the trace definition and Cartier–Foata normal form, respectively, and are structurally independent of the adjacent-boundary product. Methods M3 and M4 both use the boundary decomposition and are therefore coupled checks of its implementation.

The full polynomial (7) was reproduced by the applicable methods. The verification package contains source code, complete logs, cell ledgers and mutation tests. These checks support the finite calculation but do not replace independent mathematical review.

8 Concluding remarks

The factorisation theorem solves the prescribed-stratigraphy fibre problem for marked interval heaps on a path. It gives both the complete distribution by number of pieces and a polynomial-time formula for the minimum. It also reduces coefficient questions for arbitrary stratigraphies to the two-word matching-embedding polynomials $R(u, v; t)$.

The tree extension is given by proposition 4.3. Cycles are not covered by that argument: the proof of theorem 3.2 uses both chordality of the intersection graph and the Helly property of the supports, and neither holds for arcs of a cycle. Further problems include structural criteria for log-concavity of $(\mathrm{CS}_k(u, v))_k$ and extremal estimates for its failure.

Data and code

The source code and logs for theorem 7.1 accompany this manuscript. The verification archive has SHA-256 digest

`0326f08c5a3ab6911bc527663133b101de0497a4d107abb80df0920d10618c02`.

An external verifier should, at minimum:

(i) inspect independently the proofs of theorems 3.2 and 4.2;

(ii) rebuild the local factors without reusing recurrence (5);

(iii) reproduce the two failures at mass 14 and the absence of failures below mass 14;

(iv) verify the full polynomial (7); and

(v) confirm that the exhaustive loops cover each positive depth composition and each binary cell labelling exactly once.

Declaration of generative AI and AI-assisted technologies

During the preparation of this work the author used generative-AI systems to assist with conjecture generation, code development, adversarial checking and language revision. All mathematical claims were reconstructed as explicit proofs or identified as computer-assisted statements. AI systems are not authors and are not treated as mathematical authorities. Independent human verification is pending.

References

[1] M. Bóna, M.-L. Lackner and B. E. Sagan, *Longest increasing subsequences and log concavity*, Annals of Combinatorics **21** (2017), 535–549, doi:10.1007/s00026-017-0365-x.

[2] M. Bousquet-Mélou and X. G. Viennot, *Empilements de segments et q-énumération de polyominos convexes dirigés*, Journal of Combinatorial Theory, Series A **60** (1992), 196–224.

[3] P. Cartier and D. Foata, *Problèmes combinatoires de commutation et réarrangements*, Lecture Notes in Mathematics, vol. 85, Springer, Berlin, 1969.

[4] P. J. Chase, *Subsequence numbers and logarithmic concavity*, Discrete Mathematics **16** (1976), 123–140.

[5] C. Elzinga, S. Rahmann and H. Wang, *Algorithms for subsequence combinatorics*, Theoretical Computer Science **409** (2008), 394–404, doi:10.1016/j.tcs.2008.08.035.

[6] F. Gavril, *The intersection graphs of subtrees in trees are exactly the chordal graphs*, Journal of Combinatorial Theory, Series B **16** (1974), 47–56, doi:10.1016/0095-8956(74)90094-X.

[7] S. G. Hoggar, *Chromatic polynomials and logarithmic concavity*, Journal of Combinatorial Theory, Series B **16** (1974), 248–254.

[8] C. Krattenthaler, *The theory of heaps and the Cartier–Foata monoid*, appendix to the electronic edition of Cartier and Foata, Problèmes combinatoires de commutation et réarrangements, 2006.

[9] K.-J. Räihä and E. Ukkonen, *The shortest common supersequence problem over binary alphabet is NP-complete*, Theoretical Computer Science **16** (1981), 187–198, doi:10.1016/0304-3975(81)90075-X.

[10] R. P. Stanley, *Log-concave and unimodal sequences in algebra, combinatorics, and geometry*, Annals of the New York Academy of Sciences **576** (1989), 500–535.

[11] G. X. Viennot, *Heaps of pieces, I: Basic definitions and combinatorial lemmas*, in Combinatoire énumérative, Lecture Notes in Mathematics, vol. 1234, Springer, Berlin, 1986, pp. 321–350.