PATTERNMINER: AUTOMATED MATHEMATICAL DISCOVERY THROUGH LLM-GUIDED TRAINING AND INTERPRETATION OF SMALL MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

A large body of work in AI for math focuses on two approaches to mathematical discovery: the use of large language models (LLMs) to resolve well-specified conjectures and the use of small, narrow models for more open-ended, human-in-the-loop exploration. LLMs draw on broad mathematical knowledge to produce human-readable arguments, while small models support example-driven discovery by detecting patterns in computed examples. We present *PatternMiner*, a system in which an LLM agent generates exact datasets for a chosen problem, trains small models, and interprets what the models learn. The agent formulates conjectures, searches for counterexamples that prompt revisions, and uses hard-negative mining to select training examples that satisfy the current necessary conditions but fail the target property. The agent attempts proofs of conjectures that survive these tests. Applied to five problems in algebraic combinatorics, PatternMiner proves Schubert positivity and projection theorems, a character-row evaluation formula, restrictions on Kazhdan–Lusztig combinatorial invariance, and Schur positivity for a family of immanant coefficients; it also recovers classical Kronecker bounds.

1 INTRODUCTION

Large language models (LLMs) are now capable of solving challenging mathematics problems (Lample et al., 2022; Novikov et al., 2025). However, formulating conjectures is often just as important as proving them. Mathematical conjectures arise through interplay between theory-based reasoning and the identification of patterns within examples. Current AI-based approaches to conjecture generation emphasize one of these, with LLM-based approaches concentrating on theory-based reasoning (Long et al., 2026). One family of example-based approaches to conjecture generation uses interpretation of small models trained on highly curated datasets of examples (Davies et al., 2021; Chau et al., 2025). Interpreting these models can reveal patterns that theory does not suggest, but it is done by hand and is labor-intensive.

PatternMiner combines the two approaches: an LLM agent designs and runs the experiments, drawing on its mathematical knowledge, and small models trained on exact data detect patterns. The agent designs a training task in line with the area of inquiry. The agent then generates data related to this task using mathematical software like SageMath. The agent then trains and interprets a small model such as a linear scorer, decision stump, boosted trees, or neural network and uses insights gained from this analysis to formulate a precise conjecture. Next, the agent searches for counterexamples and revises the conjecture if any are found. Hard-negative mining supplies training examples that satisfy the current necessary conditions but fail the target property, together with positive examples satisfying those conditions, for the next round of training and interpretation. When the counterexample search is unsuccessful, the agent attempts to prove the conjecture.

We apply PatternMiner to five problems from Richard Stanley's survey of positivity problems in algebraic combinatorics (Stanley, 2000); Appendix A states the original questions. PatternMiner proves new results on four of these problems, although it resolves none of the original questions. For Problem 11, which involves finding a combinatorial interpretation of Schubert polynomial structure constants, PatternMiner gives a recursive construction showing that projection tests retaining a fixed number of flag dimensions cannot detect every vanishing structure constant. It also proves a

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

new equivariant positivity inequality that resolves Conjectures A and B of Chen et al. (2026a) on Schubert nonvanishing. For Problem 12, it derives a generating-function formula and an efficient algorithm for character-table row sums (Theorem B.3.2). For Problem 16, it proves restrictions on a possible failure of Kazhdan–Lusztig combinatorial invariance at length 13 (Theorem B.4.5). For Problem 18, which involves Schur positivity of monomial immanants, PatternMiner extends the known nonnegativity of Schur coefficients indexed by hook partitions (Lesnevich, 2024; Hikita, 2025) to partitions whose rows below the first have length at most two. We provide a more detailed description of PatternMiner's results on these and other problems in Section 4 along with a summary in Table 1.

Our contributions are as follows.

- The PatternMiner system for conjecture generation through model training, interpretation, and hard-negative mining (Section 3).
- New Schubert positivity and projection theorems, a character-row evaluation formula, restrictions on Kazhdan–Lusztig invariance, and Schur positivity for a family of immanant coefficients (Theorems B.1.3, B.1.5, B.3.2, B.4.5, and B.5.2).
- A cross-problem analysis of where interpretation helped, where it failed, and where the authors intervened (Section 5).

Code and experiment records are available in the repository:

https://anonymous.4open.science/r/patternminer-8B1C.

2 Background and Related Work

Learning and conjecture generation Many conjectures begin as patterns observed in computed examples. Classical automated conjecturing searches for relations among prescribed numerical invariants (Larson & Van Cleemput, 2016). Supervised learning can suggest nonlinear dependencies among these quantities. Davies et al. (2021) trained predictors and used gradient attribution to identify relevant features, leading to a theorem relating knot invariants and a conjecture about Kazhdan–Lusztig polynomials. PatternMiner automates this loop, which Davies et al. ran with mathematicians choosing the data, models and interpretation.

Small models have also been used to study quiver mutation (He et al., 2025) and predict intersection numbers in enumerative geometry (Hashemi et al., 2025). The Algebraic Combinatorics Dataset Repository provides nine tasks for conjecture generation from partitions, permutations, and partial orders (Chau et al., 2025).

Learned models also guide searches for constructions and counterexamples. Wagner (2021) used the cross-entropy method to refute combinatorial conjectures. PatternBoost alternates local search with transformer training (Charton et al., 2024). FunSearch and AlphaEvolve evolve LLM-generated programs, obtaining new cap-set constructions and improved algorithms (Romera-Paredes et al., 2024; Novikov et al., 2025).

Language models can also propose conjectures and proof steps. Neural proof search constructs derivations of supplied statements (Lample et al., 2022), and AlphaGeometry combines language-model proposals with symbolic deduction (Trinh et al., 2024). MECA addresses conjecture generation by revising candidate statements together with supporting arguments drawn from mathematical literature (Long et al., 2026). MECA organizes this search around mechanisms: principles, such as invariants or reductions, that suggest hypotheses.

Mechanistic interpretability Mechanistic interpretability studies how a trained network implements a particular computation (Olah et al., 2020). One step is to identify the information encoded in its activations. Geometric analyses describe representations as directions or higher-dimensional configurations (Park et al., 2025); InversionView generates inputs with similar activations to reveal the information they share (Huang et al., 2024).

An *intervention* changes a network's computation, for example by zeroing an attention head's output to measure its effect on the prediction. Interchange interventions replace an activation with the value

computed on another input, then compare the resulting prediction with that of a proposed causal model (Geiger et al., 2021).

Circuit analysis identifies components and connections that account for a specified model behavior. Automated circuit discovery removes connections using activation-patching tests (Conmy et al., 2023); edge attribution patching estimates their effects through a gradient approximation (Syed et al., 2024). Activation-patching results depend on the input perturbation and output metric (Zhang & Nanda, 2024). Evaluation therefore includes tests of how well a circuit preserves the behavior, what remains after its removal, and whether its components are redundant (Shi et al., 2024). MIB compares circuit and causal-variable localization methods across shared tasks and models (Mueller et al., 2025).

Interpretations of small models have revealed mathematical structure in their computations and yielded explicit algorithms. Analyses of modular addition and multiplication for finite-groups and monoids have recovered computations using Fourier features and group representations (Nanda et al., 2023; Chughtai et al., 2023; Chen et al., 2026b). Huang et al. (2026) interpreted a small transformer, derived a traversal algorithm on Dyck paths and proved its equivalence to the zeta map under their labeling convention.

3 PATTERNMINER

PatternMiner alternates model training and interpretation with exact tests of mathematical conjectures (Figure 1). An LLM agent chooses a prediction task and uses a small model's learned features or computations to propose a conjecture. Counterexamples prompt revisions to the conjecture, and hard-negative mining supplies examples for the next round of training and interpretation.

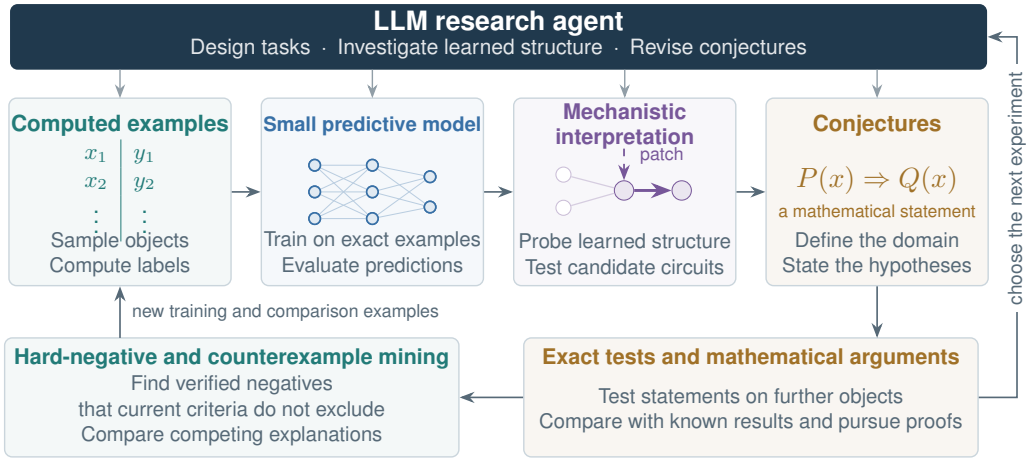


Figure 1: The agent trains and interprets small models, tests the resulting conjectures by exact computation, and mines hard negatives for the next round.

Formulating a mathematical prediction task The agent translates the mathematical question into a prediction task with exactly computable outputs. It selects a family of objects, an encoding, and a target to predict: a numerical invariant, a property to decide, or a valid operation to choose. A learned relation between numerical invariants may suggest an identity or bound; a decision rule may suggest a criterion; a rule for choosing operations may suggest an algorithm. The agent generates a supervised dataset of these objects and their exact target values using mathematical software such as SageMath.

Mining hard negatives For a decision task, a *hard negative* is an object that fails the target property while satisfying the current necessary conditions. The agent searches among objects satisfying these conditions and computes their exact labels. In later rounds, it prioritizes objects that the current model predicts to be positive.

The agent adds these negatives, together with positives that satisfy the same conditions, to the training data. Interpreting the retrained model can then suggest further necessary conditions. A hard negative shows the current conditions are not sufficient. A counterexample is a positive object that violates a proposed condition, and so refutes its necessity.

Training and evaluating small models The agent fits linear models, decision trees, gradient-boosted trees, or neural networks to the generated data, using validation errors to adjust the model and training examples. The agent favors smaller models when they achieve comparable accuracy. The agent evaluates on held-out mathematical objects, keeping equivalent presentations in the same split, and reports errors separately on difficult examples and new sizes or families.

Interpreting the learned computation For trees and linear models, the agent examines split thresholds, feature importance, and fitted coefficients to identify mathematical quantities associated with the predictions. It tests the resulting candidate relations on further objects. For neural networks, it compares inputs that produce similar activations at a chosen neuron or layer, proposes a mathematical property they share, and fits a probe to predict that property from the activations.

The agent next predicts how changing an activation should affect the answer. Activation patching tests this prediction: the agent replaces the activation on one input with the activation computed on another and recomputes the output (Conmy et al., 2023; Zhang & Nanda, 2024). The agent uses these results to propose a rule for the network's computation and assign its steps to neurons or attention heads. Patching their connections tests the proposed circuit. Appendix B.1.3 specifies the score and controls for the Schubert experiments.

Formulating and testing conjectures The agent states the proposed rule as a conjecture about the mathematical objects, specifying its domain and hypotheses. It checks whether the statement follows from known results, then tests it by exact computation on new objects. Counterexamples prompt a revised conjecture. For conjectures that survive these tests, the agent attempts a proof.

4 Experiments

We apply PatternMiner to five problems from Stanley's survey of positivity in algebraic combinatorics (Stanley, 2000, Problems 10, 11, 12, 16, and 18). A session groups related experiments and proof attempts addressing the same question and reaching a recorded conclusion or next decision. Table 1 gives the counts; Section 5 explains the accounting differences. Four of the five investigations produced novel results with no counterparts in the literature, to the best of our knowledge. Table 1 summarizes all outcomes, and Appendix B records the data selection, models, controls, exact checks, and proofs behind each.

4.1 Schubert structure constants (Stanley's Problem 11)

Background A *flag* is a nested sequence of subspaces of a fixed vector space. For permutations u, v, w, the *Schubert structure constant* $c_{u,v}^{w}$ is a nonnegative integer that counts flags satisfying incidence conditions. Equivalently, these integers are the coefficients in the expansion of a product of *Schubert polynomials*, which are polynomial representatives of cohomology classes of Schubert varieties, indexed by permutations:

$$\mathfrak{S}_u \cdot \mathfrak{S}_v = \sum_w c_{u,v}^{w} \mathfrak{S}_w.$$

Finding a combinatorial interpretation of these numbers is a central open problem in Schubert calculus. Figure 2 illustrates a coefficient equal to one and a coefficient equal to zero. We study two weaker questions: when do these coefficients vanish, and which operations on permutations preserve nonvanishing? For a comprehensive introduction, we refer the reader to Billey et al. (2025).

Results A projection test retains selected flag dimensions and asks whether the projected Schubert incidence conditions have a common solution; an empty intersection certifies vanishing.

Table 1: Mathematical outcomes and research sessions by problem. Generated conjectures include statements subsequently proved or refuted.

Problem	Sessions	New results	Rediscovered results	Generated conjectures
Schubert (11)	61	Projection obstruction and equivariant positivity (Thms B.1.3, B.1.5); insufficient candidate statistics (Prop. B.1.8)	Descent and essential-rank conditions (Props B.1.1, B.1.2)	Proved: arbitrarily large projection obstructions (Thm. B.1.3)
Kronecker (10)	17	—	First-row inequalities and hook bounds (Thms B.2.1, B.2.2)	Refuted: two-row triangle inequalities (Prop. B.2.4); empirical: distinct-parts ordering (Table 13)
Row sums (12)	33	Fixed-tail formula and evaluation algorithm (Thm. B.3.2)	Positive formulas for hook partitions (App. B.6.1)	Refuted: injectivity of learned tableau map (Table 16)
Kazhdan–Lusztig (16)	62	Restrictions on length-13 failures (Prop. B.4.2; Thm. B.4.5; Prop. B.4.6)	Invariance through length 12 (Thm. B.4.1)	Open: two coefficient bounds (Conjs B.4.7, B.6.1)
Monomial immanants (18)	32	Schur positivity for $\theta_2 \leq 2$ (Thm. B.5.2)	—	Proved: nonnegative character decomposition (Thm. B.5.2)

- Transformers trained on exact Schubert products suggested the classical descent condition: $c^w_{u,v} > 0$ implies $\mathrm{Des}(w) \subseteq \mathrm{Des}(u) \cup \mathrm{Des}(v)$, where $\mathrm{Des}(x) = \{i : x(i) > x(i+1)\}$ (Knutson, 2003).
- Rank-matrix feature searches on triples passing the existing tests suggested a second classical condition: if the three *permutation rank matrices* agree at an *essential position* of w, then $c^w_{u,v} = 0$. This is a consequence of the Littlewood–Richardson rule (Fulton, 1997); see Proposition B.1.2.
- Hard-negative mining led to a recursive construction: for every $n \geq 5$, there exist $u, v, w \in S_n$ with $c^w_{u,v} = 0$ and $|\mathrm{Des}(u) \cup \mathrm{Des}(v)| = n - 3$, yet every projection to a *partial flag variety* retaining a proper subset of these dimensions has a nonempty generic intersection. Thus no fixed number of retained dimensions suffices for these projection tests (Theorem B.1.3).
- Exact computations suggested a positivity inequality for the *equivariant structure constants* $c^w_{u,v}(t)$, whose constant terms are the ordinary structure constants. We prove

$$c^w_{u,v}(t) - \alpha_i c^w_{s_i u,v}(t) \in \mathbb{Z}_{\geq 0}[\alpha_1, \ldots, \alpha_{n-1}] \qquad \text{if } s_i u < u \leq w, \tag{1}$$

  where $\alpha_i = t_{i+1} - t_i$, s_i swaps the values $i, i+1$, and $\leq$ is *Bruhat order*. Nonvanishing is therefore preserved when the first factor increases in *left weak order* below w. This proves Conjectures A and B of Chen et al. (2026a) on equivariant nonvanishing (Theorem B.1.5 and Corollary B.1.6).
- A six-layer transformer learned to select j with $c^w_{u,s_j v} > 0$ given $c^w_{s_i u,v} > 0$, under the hypotheses in Appendix B.1.2, where Corollary B.1.6 guarantees a valid j. The agent refined its training data by mining prediction errors. Activation patching indicates that the model relies partly on the positions of values outside the candidate pairs $(j, j+1)$ to choose its answer; Appendix B.1.3 reports the measurements and controls.

Discussion The agent needed human redirection at one decisive point. In the projection tests above, the agent repeatedly sought zeros requiring five, six, and then seven retained flag dimensions for the intersection of the projected Schubert varieties to be empty. We directed it toward a construction valid for every $n \geq 5$; Theorem B.1.3 shows no fixed number of retained dimensions suffices. The agent also trained a classifier whose output depended on the order of u and v, although $c^w_{u,v} = c^w_{v,u}$. Representation choices also mattered: sparse-autoencoder features gave AUC 0.9854, compared with 0.9855 for random linear projections of the same activations (Table 4).

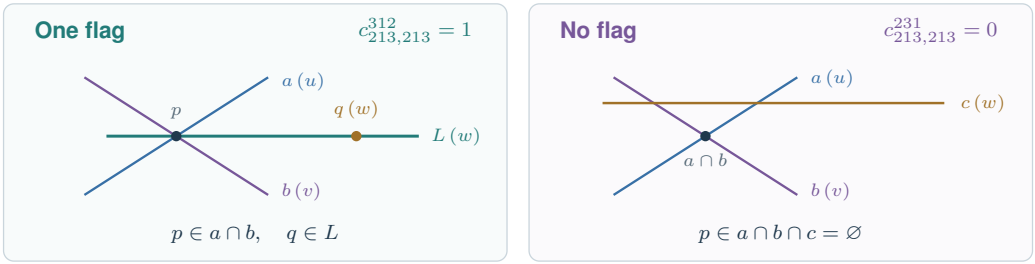


Figure 2: **Schubert structure constants count flags satisfying incidence conditions.**

4.2 Kronecker coefficients (Stanley's Problem 10)

Background The *irreducible representations* of the *symmetric group* S_n are indexed by *partitions* $\lambda \vdash n$. The *Kronecker coefficient* $g_{\lambda,\mu,\nu}$ is the nonnegative integer multiplicity of V^ν in the *tensor product* $V^\lambda \otimes V^\mu$. Stanley's Problem 10 asks for a combinatorial interpretation of these multiplicities. Figure 3 shows a tensor product for S_3 in which each of three irreducible representations occurs once.

Results

- Gradient-boosted trees trained to predict vanishing concentrated feature importance on the largest parts, recovering Murnaghan's necessary condition: $g_{\lambda,\mu,\nu} > 0$ implies $n - \lambda_1, n - \mu_1, n - \nu_1$ satisfy the triangle inequalities (Mishna et al., 2021, Equation (22)). Exact counterexamples ruled out the analogous condition on the first two rows.
- Hooks were overrepresented among correctly predicted zeros satisfying Murnaghan's inequalities. Follow-up classifiers for $g \leq 1$ and $g \leq 2$ led the agent to Rosas's bound $g_{\lambda,\mu,\nu} \leq 2$ when two partitions are hooks (Rosas, 2001, Corollary 15). A later classifier respected this bound on all 5,236 applicable test triples.
- The total number of distinct parts across the three partitions predicts relative coefficient size, with reversal rates of 0.44% ($n = 10, 11$) and 1.41% ($n = 15$) when totals differ by at least four (Table 13). At $n = 10, 11$, its Spearman correlation with coefficient size is 0.6421, compared with 0.4857 for the total *Durfee-square size*, which controls growth in Pak–Panova's bounds (Pak & Panova, 2023) (Table 14).

Discussion The investigation recovered classical bounds and identified an empirical relation between the number of distinct parts and coefficient size. Appendix B.2 gives sampling conditions and the full experimental results.

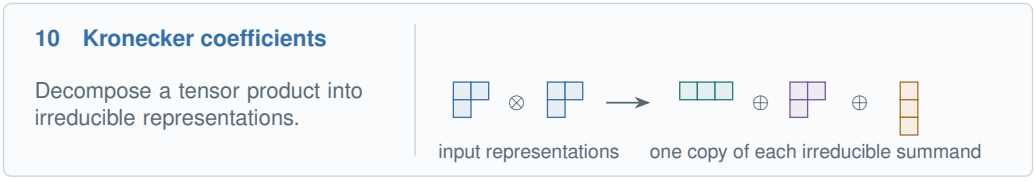


Figure 3: **Kronecker coefficients give tensor-product multiplicities.**

4.3 Character-table row sums (Stanley's Problem 12)

Background The *character table* of a finite group encodes much of its representation theory. For S_n, the entry $\chi^\lambda(\mu)$ is the irreducible character indexed by $\lambda \vdash n$, evaluated on permutations of cycle type $\mu \vdash n$. The *character-table row sums* $R(\lambda) = \sum_{\mu \vdash n} \chi^\lambda(\mu)$ are nonnegative even though individual entries can be negative. Stanley's Problem 12 asks for a combinatorial interpretation of

these numbers. Column sums, by contrast, have a classical interpretation: the column sum at $g \in S_n$ counts the $\sigma \in S_n$ with $\sigma^2 = g$ (Sagan, 2001). Figure 4 shows the sign-character row of S_3, whose entries $1, -1, 1$ sum to one.

Results

- We prove that, for partitions $(n-k, \nu)$ with fixed $\nu \vdash k$, the row-sum generating series is the partition generating function times a rational function. Using the stable character polynomials of Garsia & Goupil (2009), the resulting exact evaluation algorithm takes $\exp(O(\sqrt{k}))\,\mathrm{poly}(n)$ bit operations. Thus row sums can be evaluated in polynomial time when $k = O((\log n)^2)$, whereas direct summation uses $p(n) = \exp(\Theta(\sqrt{n}))$ conjugacy classes. Appendix B.3.2 gives the formula and complexity argument.
- A linear model trained to pair negative tableaux with positive tableaux suggested a sign-reversing injection, but collisions on held-out partitions refuted injectivity (Appendix B.3.1).

Discussion We found that the hardest part was turning a rule learned from finite data into a uniform combinatorial construction. Exact row-sum calculations did not reveal what the sums count. To make further progress, we need a uniform cancellation rule whose injectivity can be proved beyond the observed data.

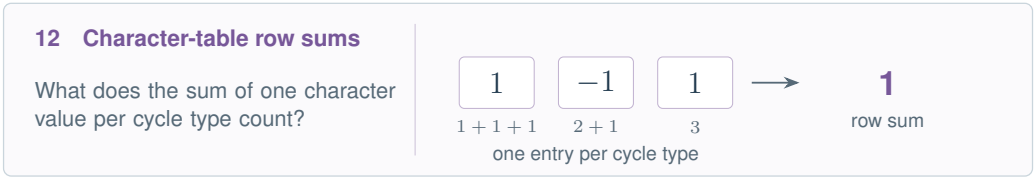


Figure 4: **Character-table row sums combine positive and negative entries.**

4.4 Kazhdan–Lusztig coefficients (Stanley's Problem 16)

Background *Kazhdan–Lusztig polynomials* $P_{u,w}(q)$ are indexed by pairs of permutations $u \le w$ in *Bruhat order* and have nonnegative integer coefficients. For example, $P_{1234,3412}(q) = 1 + q$ (Figure 5). Stanley's Problem 16 asks for a combinatorial interpretation of these coefficients. We study the related conjecture of *combinatorial invariance*, due to Lusztig and Dyer: whether the poset isomorphism type of $[u, w]$ determines $P_{u,w}$ (Dyer, 1987; Blundell et al., 2022). Previous work used neural prediction and saliency analysis to guide progress on this conjecture (Davies et al., 2021). The interval length is $\ell(w) - \ell(u)$, where ℓ counts inversions. Invariance is known for type-A intervals of length at most 12 (Esposito et al., 2026, Corollary 1.3); our investigation focuses on length 13.

Results

- Decision stumps were trained to classify individual recurrence terms as positive or zero. The two best stumps (tied, 30/33) both failed on a withheld pair, ruling out the selected single-feature thresholds (Appendix B.4.1).
- We prove that a failure of combinatorial invariance at length 13 would require an interval $[u, w]$ in S_{12} such that $u^{-1}w$ has exactly three cycles and $[q^6]P_{u,w} \ge 126$. The argument combines a relation among the possible coefficient differences, cancellation of common permutation entries, and Warrington's computations through S_{11} (Warrington, 2011); see Theorem B.4.5.
- Iterating the Kazhdan–Lusztig recurrence shows that such a failure would force the coefficient of q^6 to decrease by at least five at one of at most 26 common-descent steps (Proposition B.4.6). A bound of four for every decrease in the specified class would therefore establish invariance at length 13. This bound remains open (Conjecture B.4.7).

Discussion Counterexamples directed the investigation from predicting individual terms toward controlling their combined contribution in the recurrence. The failed thresholds exposed a limitation

of the selected features. A later exact example also ruled out a simpler coefficient bound that omitted the recurrence's correction terms (Table 18). The resulting reductions isolate a precise inequality for further study; they follow from computations and algebraic arguments. Appendix B.4 gives the proofs and the remaining conjectures.

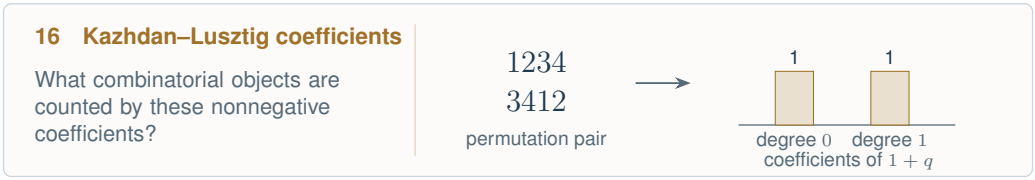


Figure 5: **Kazhdan–Lusztig polynomials have nonnegative coefficients.**

4.5 Monomial immanants (Stanley's Problem 18)

Background An *immanant* generalizes the determinant by replacing its signs with other permutation weights. Stanley's Problem 18 asks whether *monomial immanants* of *Jacobi–Trudi matrices* $H_{\mu/\nu}$ are *Schur-positive*: whether their expansions in Schur symmetric functions s_θ have nonnegative coefficients. These matrices are indexed by skew shapes μ/ν; Figure 6 gives an example whose expansion is $2s_{(3)}$. For hook partitions θ, whose diagrams consist of one row and one column meeting at a corner, Lesnevich decomposed the associated immanant character into a nonnegative integer combination of Stanley–Stembridge characters (Lesnevich, 2024). Hikita's positivity theorem for these characters implies nonnegativity of the corresponding Schur coefficients (Hikita, 2025). We investigate how far this decomposition extends beyond hooks.

Results

- We prove that the coefficient of s_θ is nonnegative in every monomial immanant of $H_{\mu/\nu}$ whenever $\theta_2 \leq 2$. This permits arbitrarily many rows of length two below the first row and places no restriction on the skew shape. The proof decomposes the associated immanant character into a nonnegative integer combination of *Stanley–Stembridge characters*; Hikita's theorem supplies positivity for each summand (Hikita, 2025). Theorem B.5.2 gives the precise statement.
- Exact computations verified nonnegativity of 776,891 decomposition coefficients across 1,587 shape instances, together with the auxiliary identities used in the proof. Appendix B.5.2 records the finite checks and their conventions.

Discussion The key step was to express the decomposition through *Kostka numbers*, which count tableaux of prescribed shape and content. Extending the hook calculation led to a generating function for coefficient partitions with every row below the first of length at most two. Factoring the resulting alternating sums then gave a proof for the entire family. The finite computations checked the formulas and searched for negative coefficients; the general positivity result rests on the symbolic factorization in Appendix B.5. A later classifier tested whether character-data vectors came from an enumerated collection of skew shapes (Appendix B.7).

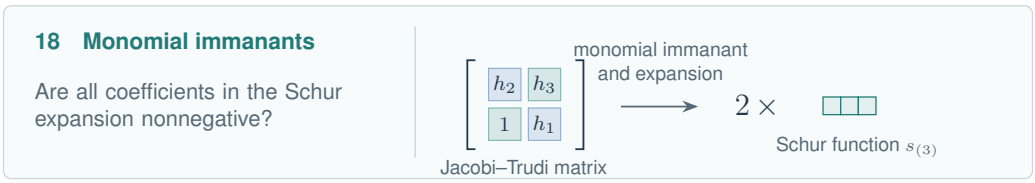


Figure 6: **Monomial immanants are expanded in Schur functions.**

5 Analysis of the discovery process

Across the five studies, PatternMiner proposed and tested conjectures using model interpretation and exact computation. We examine how conjectures developed, what model tests revealed, and where human guidance changed the search.

From patterns to conjectures The agent turned feature associations into explicit conjectures, then examined examples those conjectures left unexplained. In the Kronecker study, feature importance highlighted the largest partition parts, leading to Murnaghan's first-row inequalities. Zeros satisfying these inequalities prompted a two-row extension, which an exact counterexample refuted (Proposition B.2.4). The prevalence of hooks among the remaining zeros suggested a more specific family to study; coefficient-threshold classifiers then led to Rosas's bounds (Theorem B.2.2).

A special case provided a proof idea for a larger family. In the immanant study, the agent extended the known hook formula with a correction term. Exact computation found a negative coefficient in its decomposition. The revised conjecture explained how the hook contribution cancels the negative terms. Proving this cancellation settled the first non-hook case; further extensions led to the theorem for all coefficient partitions with $\theta_2 \leq 2$ (Appendix B.6.3).

Interpreting model performance Model tests distinguished useful clues from complete rules. The selected Schubert attention heads had stronger effects than matched random heads, yet retaining only those attention heads reduced correct confirmation pairs from $248/249$ to $80/249$: other components were still needed (Table 6). The Kazhdan–Lusztig stumps classified 30 of 33 training terms correctly but assigned the same label to a withheld pair with different exact labels. This pair exposed a distinction the proposed thresholds could not express (Appendix B.4.1).

Baselines showed how much of an apparent success a simpler predictor could explain. The learned row-sum map had five excess test collisions, compared with a median of four for random linear scores (Table 16). Kronecker predictions satisfied the semigroup inequality in 78.1% of cases, versus 74.2% for a size-matched baseline: an improvement of 3.9 percentage points (Table 11).

Iteration and human guidance The Schubert search kept producing finite examples without reaching a general explanation. PatternMiner repeatedly sought examples requiring more retained flag dimensions. Human guidance shifted the goal to a construction valid at every size, leading to the theorem that no fixed number of retained dimensions suffices (Theorem B.1.3). The intervention changed what counted as progress, from a larger example to a general result.

Limitations We have not run a controlled LLM-only comparison. Such a comparison would use the same initial context, exact-computation tools, and budget, evaluating the mathematical statements obtained with and without learning.

The session counts in Table 1 were not all obtained the same way: the Schubert count was reconstructed retrospectively by grouping related experiments and proof attempts by question and recorded outcome, while the other four are drawn directly from per-problem experiment logs. This difference limits direct comparison across problems; applying a single accounting scheme from the start of each investigation would make such comparisons meaningful.

Statement of AI use

We used LLM agents to propose experiments and conjectures, implement exact computations, train and analyze small models, and develop proofs. LLMs also assisted with literature searches, manuscript drafting and revision, and figure preparation. The authors set the research objectives, approved computational resources, redirected the Schubert search toward a construction valid for arbitrary size, and guided manuscript revisions. The authors checked the proofs in the Schubert structural constants section. Appendix B reports the model evaluations, exact checks, and proofs supporting the results.

REFERENCES

Christine Bessenrodt and Christopher Bowman. Multiplicity-free Kronecker products of characters of the symmetric groups. *Advances in Mathematics*, 322:473–529, 2017. doi: 10.1016/j.aim.2017.10.009. URL https://arxiv.org/abs/1609.03596.

Sara C. Billey, Yibo Gao, and Brendan Pawlowski. Introduction to the cohomology of the flag variety. arXiv preprint arXiv:2506.21064, 2025. URL https://arxiv.org/abs/2506.21064.

Anders Björner and Francesco Brenti. *Combinatorics of Coxeter Groups*, volume 231 of *Graduate Texts in Mathematics*. Springer, 2005. doi: 10.1007/3-540-27596-7. URL https://link.springer.com/book/10.1007/3-540-27596-7.

Charles Blundell, Lars Buesing, Alex Davies, Petar Veličković, and Geordie Williamson. Towards combinatorial invariance for Kazhdan–Lusztig polynomials. *Representation Theory*, 26:1145–1191, 2022. doi: 10.1090/ert/624. URL https://arxiv.org/abs/2111.15161.

François Charton, Jordan S. Ellenberg, Adam Zsolt Wagner, and Geordie Williamson. PatternBoost: Constructions in mathematics with a little help from AI. arXiv preprint arXiv:2411.00566, 2024. URL https://arxiv.org/abs/2411.00566.

Herman Chau, Helen Jenne, Davis Brown, Jesse He, Mark Raugas, Sara C. Billey, and Henry Kvinge. Machine learning meets algebraic combinatorics: A suite of datasets capturing research-level conjecturing ability in pure mathematics. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 7515–7554. PMLR, 2025. URL https://proceedings.mlr.press/v267/chau25a.html.

Yiming Chen, Neil J. Y. Fan, Rui Xiong, and Ming Yao. Non-vanishing of single, double, and triple Schubert structure constants, 2026a. URL https://arxiv.org/abs/2608.17378.

Zitong Andrew Chen, Junaid Hasan, Akhil Srinivasan, Hemkesh Bandi, and Jarod Alper. Multiplication beyond groups: Stratified Fourier mechanisms in transformer circuits, 2026b. URL https://arxiv.org/abs/2607.07066.

Bilal Chughtai, Lawrence Chan, and Neel Nanda. A toy model of universality: Reverse engineering how networks learn group operations. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 6243–6267. PMLR, 2023. URL https://proceedings.mlr.press/v202/chughtai23a.html.

Arthur Conmy, Augustine N. Mavor-Parker, Aengus Lynch, Stefan Heimersheim, and Adrià Garriga-Alonso. Towards automated circuit discovery for mechanistic interpretability. In *Advances in Neural Information Processing Systems*, volume 36, pp. 16318–16352, 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/34e1dbe95d34d7ebaf99b9bcaeb5b2be-Abstract-Conference.html.

Alex Davies, Petar Veličković, Lars Buesing, Sam Blackwell, Daniel Zheng, Nenad Tomašev, Richard Tanburn, Peter Battaglia, Charles Blundell, András Juhász, Marc Lackenby, Geordie Williamson, Demis Hassabis, and Pushmeet Kohli. Advancing mathematics by guiding human intuition with AI. *Nature*, 600:70–74, 2021. doi: 10.1038/s41586-021-04086-x.

Matthew John Dyer. *Hecke Algebras and Reflections in Coxeter Groups*. PhD thesis, University of Sydney, 1987. URL https://academicweb.nd.edu/~dyer/papers/dyerthesis.pdf.

Francesco Esposito, Mario Marietti, and Salvatore Stella. Flip combinatorial invariance and Weyl groups. arXiv preprint arXiv:2509.16433v2, 2026. URL https://arxiv.org/abs/2509.16433v2. Version 2, 17 September 2026.

William Fulton. *Young Tableaux: With Applications to Representation Theory and Geometry*, volume 35 of *London Mathematical Society Student Texts*. Cambridge University Press, 1997. URL https://www.cambridge.org/core/books/young-tableaux/A7570B10D82AE7233E25E5D6F70A07B6.

Adriano M. Garsia and Alain Goupil. Character polynomials, their q-analogs and the Kronecker product. *The Electronic Journal of Combinatorics*, 16(2):R19, 2009. doi: 10.37236/85. URL https://www.combinatorics.org/ojs/index.php/eljc/article/view/v16i2r19.

Atticus Geiger, Hanson Lu, Thomas Icard, and Christopher Potts. Causal abstractions of neural networks. In *Advances in Neural Information Processing Systems*, volume 34, pp. 9574–9586, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/4f5c422f4d49a5a807eda27434231040-Abstract.html.

William Graham. Positivity in equivariant Schubert calculus. *Duke Mathematical Journal*, 109(3): 599–614, 2001. doi: 10.1215/S0012-7094-01-10935-6. URL https://arxiv.org/abs/math/9908172.

Baran Hashemi, Roderic G. Corominas, and Alessandro Giacchetto. Can transformers do enumerative geometry? In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/aee2f03ecb2b2c1ea55a43946b651cfd-Abstract-Conference.html.

Jesse He, Helen Jenne, Herman Chau, Davis Brown, Mark Raugas, Sara C. Billey, and Henry Kvinge. Machines and mathematical mutations: Using GNNs to characterize quiver mutation classes. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 22434–22455. PMLR, 2025. URL https://proceedings.mlr.press/v267/he25g.html.

Tatsuyuki Hikita. A proof of the Stanley–Stembridge conjecture. arXiv preprint arXiv:2410.12758v2, 2025. URL https://arxiv.org/abs/2410.12758v2. Originally posted 2024; revised December 2025.

Xiaoyu Huang, Blake Jackson, and Kyu-Hwan Lee. Discovering a Zeta map algorithm on Dyck paths via mechanistic interpretability. arXiv preprint arXiv:2605.30482, 2026. URL https://arxiv.org/abs/2605.30482.

Xinting Huang, Madhur Panwar, Navin Goyal, and Michael Hahn. InversionView: A general-purpose method for reading information from neural activations. In *Advances in Neural Information Processing Systems*, volume 37, pp. 137903–137964, 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/f94cfd15db3f16ee7789b6b7e91ec476-Abstract-Conference.html.

Allen Knutson. A Schubert calculus recurrence from the noncomplex W-action on G/B, 2003. URL https://arxiv.org/abs/math/0306304.

Allen Knutson. Schubert calculus and quiver varieties. In *Proceedings of the International Congress of Mathematicians 2022*, volume 6, pp. 4582–4605, 2023. doi: 10.4171/ICM2022/118.

Allen Knutson and Ezra Miller. Gröbner geometry of Schubert polynomials. *Annals of Mathematics*, 161(3):1245–1318, 2005. doi: 10.4007/annals.2005.161.1245. URL https://annals.math.princeton.edu/2005/161-3/p04.

Guillaume Lample, Timothee Lacroix, Marie-Anne Lachaux, Aurelien Rodriguez, Amaury Hayat, Thibaut Lavril, Gabriel Ebner, and Xavier Martinet. HyperTree proof search for neural theorem proving. In *Advances in Neural Information Processing Systems*, volume 35, pp. 26337–26349, 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/a8901c5e85fb8e1823bbf0f755053672-Abstract-Conference.html.

C.E. Larson and N. Van Cleemput. Automated conjecturing I: Fajtlowicz's Dalmatian heuristic revisited. *Artificial Intelligence*, 231:17–38, 2016. doi: 10.1016/j.artint.2015.10.002. URL https://doi.org/10.1016/j.artint.2015.10.002.

Nathan R. T. Lesnevich. Hook-shape immanant characters from Stanley–Stembridge characters. *Algebraic Combinatorics*, 7(1):137–157, 2024. doi: 10.5802/alco.331. URL https://alco.centre-mersenne.org/articles/10.5802/alco.331/.

Wentao Long, Yunfei Zhang, Chenyi Li, and Zaiwen Wen. MECA: A mechanism-centered agent for constructing well-specified and valuable mathematical conjectures. arXiv preprint arXiv:2607.27709, 2026. URL https://arxiv.org/abs/2607.27709.

Marni Mishna, Mercedes Rosas, and Sheila Sundaram. Vector partition functions and Kronecker coefficients. *Journal of Physics A: Mathematical and Theoretical*, 54(20), 2021. doi: 10.1088/1751-8121/abf45b. URL https://arxiv.org/abs/1811.10015v4.

Aaron Mueller, Atticus Geiger, Sarah Wiegreffe, Dana Arad, Iván Arcuschin, Adam Belfki, Yik Siu Chan, Jaden Fried Fiotto-Kaufman, Tal Haklay, Michael Hanna, Jing Huang, Rohan Gupta, Yaniv Nikankin, Hadas Orgad, Nikhil Prakash, Anja Reusch, Aruna Sankaranarayanan, Shun Shao, Alessandro Stolfo, Martin Tutek, Amir Zur, David Bau, and Yonatan Belinkov. MIB: A mechanistic interpretability benchmark. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 45069–45108. PMLR, 2025. URL https://proceedings.mlr.press/v267/mueller25a.html.

Neel Nanda, Lawrence Chan, Tom Lieberum, Jess Smith, and Jacob Steinhardt. Progress measures for grokking via mechanistic interpretability. In *International Conference on Learning Representations*, 2023. URL https://arxiv.org/abs/2301.05217.

Alexander Novikov, Ngân Vũ, Marvin Eisenberger, Emilien Dupont, Po-Sen Huang, Adam Zsolt Wagner, Sergey Shirobokov, Borislav Kozlovskii, Francisco J. R. Ruiz, Abbas Mehrabian, M. Pawan Kumar, Abigail See, Swarat Chaudhuri, George Holland, Alex Davies, Sebastian Nowozin, Pushmeet Kohli, and Matej Balog. AlphaEvolve: A coding agent for scientific and algorithmic discovery. arXiv preprint arXiv:2506.13131, 2025. URL https://arxiv.org/abs/2506.13131.

Chris Olah, Nick Cammarata, Ludwig Schubert, Gabriel Goh, Michael Petrov, and Shan Carter. Zoom in: An introduction to circuits. *Distill*, 5(3):e00024.001, 2020. doi: 10.23915/distill.00024.001. URL https://distill.pub/2020/circuits/zoom-in/.

Igor Pak and Greta Panova. Durfee squares, symmetric partitions and bounds on Kronecker coefficients. *Journal of Algebra*, 629:358–380, 2023. URL https://arxiv.org/abs/2207.02561.

Kiho Park, Yo Joong Choe, Yibo Jiang, and Victor Veitch. The geometry of categorical and hierarchical concepts in large language models. In *International Conference on Learning Representations*, pp. 76441–76463, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/be7430d22a4dae8516894e32f2fcc6db-Abstract-Conference.html.

Leonardo Patimo. A combinatorial formula for the coefficient of q in Kazhdan–Lusztig polynomials. *International Mathematics Research Notices*, 2021(5):3203–3223, 2021. doi: 10.1093/imrn/rnz255. URL https://doi.org/10.1093/imrn/rnz255.

Kevin Purbhoo. Vanishing and non-vanishing criteria in Schubert calculus. *International Mathematics Research Notices*, 2006:1–38, 2006. URL https://arxiv.org/abs/math/0304070. Article ID 24590.

Edward Richmond. A multiplicative formula for structure constants in the cohomology of flag varieties. *Michigan Mathematical Journal*, 61(1):3–17, 2012. URL https://arxiv.org/abs/0812.1856.

Bernardino Romera-Paredes, Mohammadamin Barekatain, Alexander Novikov, Matej Balog, M. Pawan Kumar, Emilien Dupont, Francisco J. R. Ruiz, Jordan S. Ellenberg, Pengming Wang, Omar Fawzi, Pushmeet Kohli, and Alhussein Fawzi. Mathematical discoveries from program search with large language models. *Nature*, 625:468–475, 2024. doi: 10.1038/s41586-023-06924-6.

Mercedes H. Rosas. The Kronecker product of Schur functions indexed by two-row shapes or hook shapes. *Journal of Algebraic Combinatorics*, 14:153–173, 2001. URL https://arxiv.org/abs/math/0001084.

Bruce E. Sagan. *The Symmetric Group: Representations, Combinatorial Algorithms, and Symmetric Functions*, volume 203 of *Graduate Texts in Mathematics*. Springer, 2 edition, 2001. doi: 10.1007/978-1-4757-6804-6. URL https://link.springer.com/book/10.1007/978-1-4757-6804-6.

Claudia Shi, Nicolas Beltran-Velez, Achille Nazaret, Carolina Zheng, Adrià Garriga-Alonso, Andrew Jesson, Maggie Makar, and David M. Blei. Hypothesis testing the circuit hypothesis in LLMs. In *Advances in Neural Information Processing Systems*, volume 37, pp. 94539–94567, 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/abccb8a90b30d45b948360ba41f5a20f-Abstract-Conference.html.

Richard P. Stanley. Positivity problems and conjectures in algebraic combinatorics. In V. I. Arnold, M. F. Atiyah, P. D. Lax, and B. Mazur (eds.), *Mathematics: Frontiers and Perspectives*, pp. 295–319. American Mathematical Society, 2000. URL https://math.mit.edu/~rstan/pubs/pubfiles/116.pdf.

Sheila Sundaram. The conjugacy action of S_n and modules induced from centralisers. *Journal of Algebraic Combinatorics*, 48(2):179–225, 2018. doi: 10.1007/s10801-017-0796-9. URL https://arxiv.org/abs/1603.05898.

Aaquib Syed, Can Rager, and Arthur Conmy. Attribution patching outperforms automated circuit discovery. In *Proceedings of the 7th BlackboxNLP Workshop: Analyzing and Interpreting Neural Networks for NLP*, pp. 407–416. Association for Computational Linguistics, 2024. doi: 10.18653/v1/2024.blackboxnlp-1.25. URL https://aclanthology.org/2024.blackboxnlp-1.25/.

Trieu H. Trinh, Yuhuai Wu, Quoc V. Le, He He, and Thang Luong. Solving olympiad geometry without human demonstrations. *Nature*, 625(7995):476–482, 2024. doi: 10.1038/s41586-023-06747-5. URL https://doi.org/10.1038/s41586-023-06747-5.

Adam Zsolt Wagner. Constructions in combinatorics via neural networks. arXiv preprint arXiv:2104.14516, 2021. URL https://arxiv.org/abs/2104.14516.

Gregory S. Warrington. Equivalence classes for the μ-coefficient of Kazhdan–Lusztig polynomials in S_n. *Experimental Mathematics*, 20(4):457–466, 2011. doi: 10.1080/10586458.2011.565260. URL https://doi.org/10.1080/10586458.2011.565260.

Fred Zhang and Neel Nanda. Towards best practices of activation patching in language models: Metrics and methods. In *International Conference on Learning Representations*, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/06a52a54c8ee03cd86771136bc91eb1f-Abstract-Conference.html.

A PROBLEM STATEMENTS

We state the five questions from Stanley (2000) in the notation of this paper, followed by the tasks pursued in our investigations. A combinatorial interpretation means an explicit family of objects counted by the quantity in question. Its nonnegativity or exact numerical evaluation alone does not supply that interpretation.

Problem 10: Kronecker coefficients For partitions $\lambda, \mu, \nu \vdash n$, define $g_{\lambda,\mu,\nu}$ by the decomposition of irreducible characters

$$\chi^\lambda \chi^\mu = \sum_{\nu \vdash n} g_{\lambda,\mu,\nu} \chi^\nu.$$

The problem asks for a combinatorial interpretation of these coefficients. Our prediction tasks concern vanishing, small values and magnitude; recovered bounds and empirical comparisons do not give a general counting rule.

Problem 11: Schubert structure constants For Schubert polynomials indexed by permutations with finite support, write

$$\mathfrak{S}_u \mathfrak{S}_v = \sum_w c_{u,v}^w \mathfrak{S}_w.$$

The problem asks for a combinatorial interpretation of $c_{u,v}^w$. We study necessary conditions for nonvanishing, limits of projection tests, and operations preserving nonvanishing. The learned selector addresses one such operation. These tasks are narrower than constructing a general counting rule.

Problem 12: character-table row sums For $\lambda \vdash n$, put

$$R(\lambda) = \sum_{\rho \vdash n} \chi^\lambda(\rho).$$

The sum has one term per conjugacy class, without class-size weights. The problem asks for a combinatorial interpretation of $R(\lambda)$. We derive a fixed-tail evaluation formula and test a learned injection between negative and positive tableaux in a restricted character expansion. That finite injection task does not represent every row sum.

Problem 16: Kazhdan–Lusztig coefficients For a finite or affine Weyl group W and $x \le w$ in Bruhat order, the problem asks for a combinatorial interpretation of each coefficient $[q^k]P_{x,w}(q)$, especially for $W = S_n$. Our investigation instead focuses on the related conjecture of *combinatorial invariance* in type A:

$$[x, w] \cong [x', w'] \quad \Longrightarrow \quad P_{x,w} = P_{x',w'}.$$

Here the isomorphism preserves the partial order, without permutation labels. We learn rules for individual recurrence terms and derive restrictions on a possible length-thirteen failure. Neither invariance nor the original counting problem is resolved in general by this work.

Problem 18: monomial immanants Let $D = \mu/\nu$ be a skew shape with at most n rows, and set $H_D = (h_{\mu_i - \nu_j - i + j})_{i,j=1}^n$, where h_k is the complete symmetric function, $h_0 = 1$ and $h_k = 0$ for $k < 0$. For $\lambda \vdash n$, let ϕ^λ be the virtual character with Frobenius characteristic m_λ. The problem asks whether

$$[s_\theta] \operatorname{Imm}_{\phi^\lambda}(H_D) \ge 0 \quad \text{for every } \theta \vdash |D|, \qquad \operatorname{Imm}_{\phi^\lambda}(A) = \sum_{w \in S_n} \phi^\lambda(w) \prod_{i=1}^n a_{i,w(i)}.$$

We prove this positivity for coefficient partitions with $\theta_2 \le 2$, through a nonnegative character decomposition. The immanant index λ, coefficient partition θ and skew shape D play different roles. The separate classifier experiment tests which character-data vectors occur in an enumerated shape collection; it does not decide the full positivity problem.

The detailed experiments and proofs follow in Appendix B.

B MATHEMATICAL AND EXPERIMENTAL DETAILS

The label *LLM-generated* identifies statements and proofs proposed by an LLM agent during the investigation.

B.1 SCHUBERT STRUCTURE CONSTANTS

B.1.1 MATHEMATICAL SETTING

Let $\mathrm{Fl}_n(\mathbb{C})$ be the *complete flag variety*, and let σ_w be its *Schubert class* of complex codimension $\ell(w)$, where ℓ is the *inversion number*. We use the conventions of Knutson (2023); Fulton (1997) gives the background on flag varieties and *Schubert polynomials*. The ordinary structure constants satisfy

$$\sigma_u \sigma_v = \sum_{w \in S_n} c_{u,v}^w \sigma_w, \qquad c_{u,v}^w \in \mathbb{Z}_{\geq 0}.$$

For $\ell(w) = \ell(u) + \ell(v)$, the coefficient counts the intersection of the *Schubert varieties* indexed by $u, v, w_0 w$ relative to three general flags. Here w_0 is the longest permutation. We write $\mathrm{Des}(w) = \{i : w(i) > w(i+1)\}$ for *right descents* and $\mathrm{Des}_L(w) = \{i : w^{-1}(i) > w^{-1}(i+1)\}$ for *left descents*. Right multiplication by $s_i = (i, i+1)$ swaps positions, and left multiplication swaps values.

The examples in Figure 2 are $c_{213,213}^{312} = 1$ and $c_{213,213}^{231} = 0$. A flag in $\mathbb{C}^3$ is a point p on a line L in $\mathbf{P}^2$. Two general conditions on p determine their intersection point, and a condition that L contain another general point determines L. Three general conditions on p have no solution. The same counts follow from $\mathfrak{S}_{213}^2 = x_1^2 = \mathfrak{S}_{312}$.

Automatically generated by the PatternMiner system

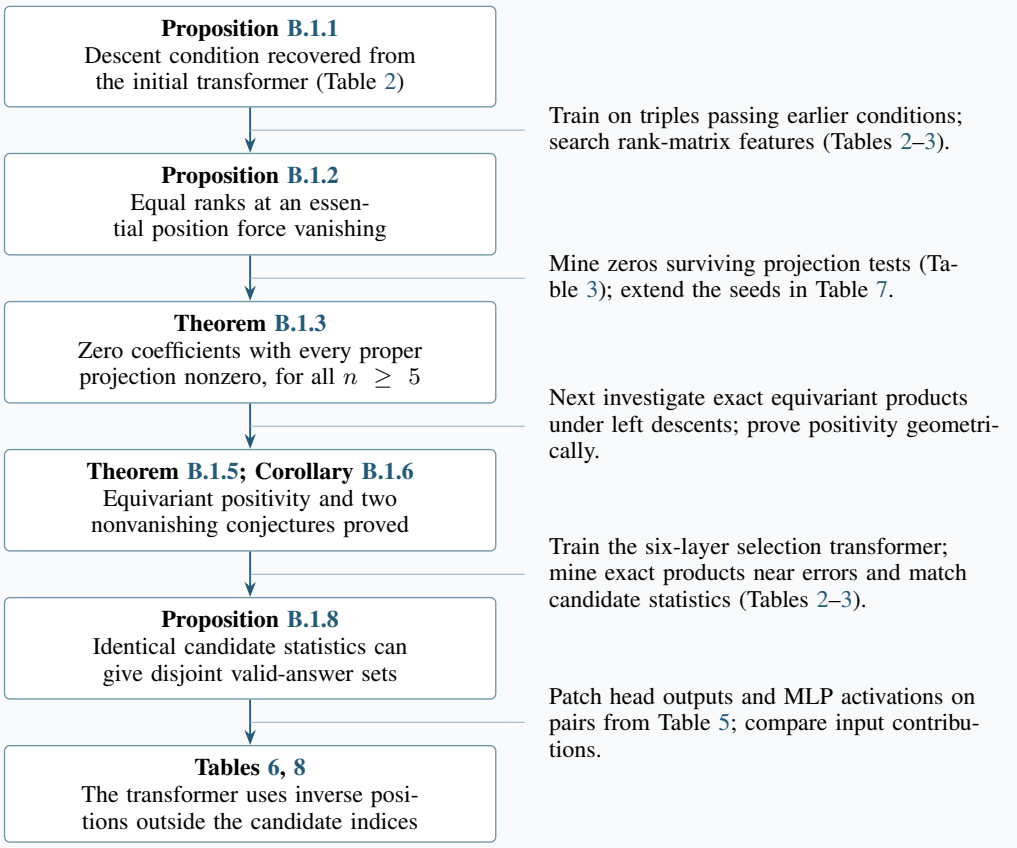


Figure 7: Schubert discoveries and the experiments leading to them.

B.1.2 Models and hard-negative experiments

The first task predicts whether $c^w_{u,v} > 0$ from a degree-compatible triple. A hard negative is an exactly computed zero that passes the necessary conditions used to select its dataset. For the model trained on the remaining triples, those conditions were degree compatibility, $u, v \leq w$ in Bruhat order, the descent condition, and the zero-rank case of Proposition B.1.2. Permutation products, rather than individual triples from the same product, determine the training and evaluation split.

Table 2: Schubert prediction models.

Task	Model and training data	Recorded evaluation
$c^w_{u,v} > 0$	Transformer: width 192, six heads, four encoder layers and two cross-attention layers; 1,272,767 training triples	Product-disjoint test: F1 0.992, Matthews correlation 0.980
$c^w_{u,v} > 0$ after necessary conditions	Transformer trained on triples passing the conditions above	Matthews correlation 0.961 on the corresponding evaluation population
Select j with $c^w_{u,s_j v} > 0$	Transformer: six layers, width 384, sixteen heads per layer; 17,994,585 training examples	Separate S_{10} assessment: 99.799% per supplied input; 1,874/1,902 difficult pairs correct

The selection task uses $\ell(w) = \ell(u) + \ell(v) - 1$ and a supplied $i \in \mathrm{Des}_L(u)$ with $c^w_{s_i u,v} > 0$. A correct answer belongs to

$$\mathcal{V}(u, v, w) = \{j \in \mathrm{Des}_L(v) : c^w_{u,s_j v} > 0\}. \tag{B.1}$$

PatternMiner system output (continued)

We retain inputs with $\mathcal{V}(u, v, w) \neq \varnothing$; $u \leq w$ guarantees this by Corollary B.1.6. The final architecture discards the supplied index. Its inputs are inverse positions, descent indicators, rank-matrix entries, and images of simple roots under relative permutations. An auxiliary prediction of $\log(1 + c^{w}_{u,s_j v})$ was combined with the selection score into a single linear output head. Exact coefficient values supply supervision only. Factor labels S_7–S_{10} specify the sampled factors; stable products can have output permutations in a larger symmetric group, here of size at most seventeen.

Table 3: Exact tests used to refine the Schubert tasks.

Test	Population or selection rule	Result
Equal-rank condition	Zeros passing the earlier necessary conditions	$133/45{,}000$ excluded
Soundness of that condition	Separate nonzero coefficients	$0/59{,}979$ violations
Neighboring factor pairs	Exact products near prediction errors, excluding reserved factor orbits	4,683,591 new examples and 27,352 difficult pairs in 952 factor groups
Proper projections	The stated seeds in S_5 and S_6	Two zero coefficients; all eight nonempty proper projections nonzero

The representation comparison in Table 4 selects the best feature on the same 59,441 examples. The circuit tests below use a separate confirmation sample.

Table 4: Feature comparison on the same Schubert examples.

Representation	Best zero-label AUC
1,024 sparse-autoencoder features	0.9854
1,024 random linear projections	0.9855

B.1.3 Circuit interventions

Activation patching replaces specified activations on a recipient input with those computed on a donor, then recomputes the affected downstream activations (Conmy et al., 2023; Zhang & Nanda, 2024). For disjoint recipient and donor answer sets A, B and output logits z, we measure the change in

$$\log \sum_{j \in B} e^{z_j} - \log \sum_{j \in A} e^{z_j}. \tag{B.2}$$

A positive change shifts relative probability toward the donor's valid answers. Both directions are evaluated for every pair. To form factor orbits, remove common initial and terminal fixed points, then identify factor pairs under exchange and the simultaneous involution $(u, v) \mapsto (w_0 u w_0, w_0 v w_0)$. Means weight these orbits equally; bootstrap intervals resample them. Identity replacements and agreement with the original forward computation check the implementation.

Table 5: Samples for the fixed Schubert transformer.

Sample	Pairs	Factor orbits	Both answers correct
Component selection	320	29	316
Confirmation	249	22	248
Earlier pairs matching the candidate statistics	8	4	8

The first two samples have disjoint factor orbits and were excluded from model training. Each pair shares its factors and has different output permutations with disjoint valid-answer sets. The eight earlier pairs also agree in the candidate statistics of Proposition B.1.8. Component selection scanned all 96 attention heads and six multilayer perceptrons. The selected heads are $0, 8$ in layer 1 and $3, 11$ in layer 3, using zero-based indices. Random controls match their layers and the head pairing inherited from widening the model.

PatternMiner system output (continued)

Table 6: Confirmation results for Schubert interventions.

Intervention	Mean score change
Replace the four selected head outputs	15.738
Twenty matched random head sets, averaged	4.115
First-MLP activations outside the candidate set, through layer-1 keys and values	11.658
Same replacement restricted to 32 previously selected hidden-unit pairs	8.471
Replace only the inverse-position contribution	10.124
Replace only the rank-matrix contribution	0.211
Replace only the descent-indicator contribution	0.084
Replace only the relative-root contribution	0.470
Hold later MLP and attention outputs fixed during the 32-pair replacement	0.435
Ablation	**Both answers correct**
No ablation	248/249
Zero the four selected heads	215/249
Zero all other attention heads	80/249

The selected-head effect exceeds the matched-control mean by 11.623, with a descriptive 95% bootstrap interval $[8.826, 14.920]$. A hidden-unit pair consists of units $k, k+384$, inherited from widening the MLP; there are 384 such pairs. The input contributions are replaced before the activation function and subsequent layers, so their effects are not additive. The ablations leave MLPs and residual connections active. Together, the interventions show a contribution from inverse positions outside the candidate set and a dependence on later computation.

B.1.4 Rediscovered necessary conditions

Proposition B.1.1 (Knutson, 2003)**.** *If $c^w_{u,v} > 0$, then $\mathrm{Des}(w) \subseteq \mathrm{Des}(u) \cup \mathrm{Des}(v)$.*

The earlier model experiments suggested this descent condition, which the system then used to select hard negatives for further training.

For $x \in S_n$, let $r_x(i,j) = \#\{a \le i : x(a) \le j\}$. The *Rothe diagram* is $D(x) = \{(i,j) : j < x(i),\ i < x^{-1}(j)\}$, and its *essential set* consists of the southeast corners. The following condition is a consequence of Grassmannian projection and the Littlewood–Richardson rule (Fulton, 1997).

Proposition B.1.2 (Fulton, 1997)**.** *If $(i,j) \in \mathrm{Ess}(w)$ and $r_u(i,j) = r_v(i,j) = r_w(i,j)$, then $c^w_{u,v} = 0$.*

A model trained on triples passing the earlier conditions prompted a search over rank-matrix features, yielding the equal-rank condition above. It excluded 133 of 45,000 hard negatives, with no violations among 59,979 nonzero coefficients (Table 3).

B.1.5 Vanishing under proper projections

For $E \subseteq \{1, \dots, n-1\}$, let $\mathrm{Fl}(E; \mathbb{C}^n)$ retain the subspaces with dimensions in E. Write $p_E(x)$ for the minimal representative obtained by sorting the entries of x within the position blocks determined by E. The generic projected intersection is nonempty exactly when

$$\sigma_{p_E(u)} \sigma_{p_E(v)} \sigma_{p_E(w_0 w)} \neq 0.$$

These are pullbacks of partial-flag Schubert classes. If $D = \mathrm{Des}(u) \cup \mathrm{Des}(v)$ and $\mathrm{Des}(w) \subseteq D$, the product at $E = D$ is nonzero exactly when $c^w_{u,v} > 0$: the first two Schubert conditions pull back from this partial flag variety, and the projection of the third is indexed by $p_D(w_0 w)$.

Theorem B.1.3 (LLM-generated)**.** *For every $n \ge 5$, there exist $u, v, w \in S_n$ with $\ell(w) = \ell(u) + \ell(v)$, $c^w_{u,v} = 0$, and*

$$\mathrm{Des}(w) \subseteq D = \mathrm{Des}(u) \cup \mathrm{Des}(v), \qquad |D| = n - 3,$$

such that every projection retaining a proper subset $E \subsetneq D$ has a nonempty generic intersection.

Starting with $(u, v, w) = (215346, 135264, 256134)$, let f be the factor beginning with 1 and g the other factor. To form f' and w', prepend 2 to f and w and increase each old value at least 2 by one. To form g', prepend 1 to g and increase every old value by one. Preserve the labels of the two factors at each step. This recursion supplies the examples for every $n \ge 6$; $(14235, 24153, 45123)$ supplies $n = 5$.

PatternMiner system output (continued)

The projected intersection matrix For three generic upper unitriangular matrices ρ_1, ρ_2, ρ_3, form the matrix F_E whose columns are

$$\rho_i e_{ab} \rho_i^{-1}, \qquad a < b,\ x_i(a) > x_i(b), \qquad (x_1, x_2, x_3) = (p_E(u), p_E(v), p_E(w_0 w)).$$

Rows are indexed by upper triangular positions (s, t), and the entry in column (i, a, b) is $\rho_i[s, a]\rho_i^{-1}[b, t]$. Purbhoo's criterion identifies nonemptiness of the generic projected intersection with full column rank of F_E (Purbhoo, 2006). All subsequent ranks and kernels are taken over the rational function field in the matrix entries.

Set $q = (p_D(w_0 w))^{-1}(n)$, the position of the maximum after sorting within the parabolic blocks. For the seed in S_6, $p_D(w_0 w) = (5, 1, 2, 4, 6, 3)$ and $q = 5$.

Lemma B.1.4 (LLM-generated)**.** *Suppose $(u, v, w) \in S_n^3$ has $\ell(w) = \ell(u) + \ell(v)$, has empty generic intersection on D, and has nonempty generic intersection on every proper subset of D, with $\mathrm{Des}(w) \subseteq D$ and $|D| = k$. Assume that $1 \in D$, $w(1) = 2$, and exactly one of $u(1), v(1)$ equals 1. Suppose F_D has column nullity one, with kernel vector κ, and, for the index q defined above, at least one of*

$$\Psi_{i,a} = \sum_{b > a} \kappa_{(i,a,b)} (\rho_i^{-1} \rho_3)[b, q], \qquad i \in \{1, 2\}, \tag{B.3}$$

is a nonzero rational function. Coordinates absent from the matrix are zero. Every t-fold application of this recursion then has empty generic intersection on its $k + t$ descent dimensions in S_{n+t} and nonempty generic intersection on every proper subset.

Proof. Write f for the factor beginning with 1 and g for the other factor. The construction sends (f, g, w) to (f', g', w'), with the two factor labels preserved. It increases the union of descents to $D' = \{1\} \cup (D + 1)$. The first two positions are singleton blocks for D', and the hypotheses on the first entries persist.

For any $E \subseteq D$, let $E' = \{1\} \cup (E + 1)$ and extend the generic flags as $\rho_i' = \left[\begin{smallmatrix} 1 & r_i \\ 0 & \rho_i \end{smallmatrix}\right]$, where the r_i have independent entries. Sorting within blocks commutes with this extension. Order rows and columns with first position 1 first. The matrix takes the form

$$F_{E'} = \begin{pmatrix} X & Y \\ 0 & F_E \end{pmatrix}. \tag{B.4}$$

There are n rows and n columns involving the new first position. The square block X represents the projective-space intersection with codimensions $1, 0, n - 1$, whose degree is one. Therefore X is generically invertible. This is the degree-one base intersection in parabolic factorization (Richmond, 2012). Equation (B.4) preserves full column rank and column nullity. In particular it preserves vanishing for the full flag problem, and it preserves feasibility for every proper projection containing the newly retained dimension.

It remains to check $M = D + 1$, which omits that new dimension. Every other proper projection omitting it is a projection of M. Passing from D' to M sorts only positions 1 and 2. In f' these entries are 2 and 1; in the third factor they are n and $n - 1$. Their values are consecutive, so sorting removes the inversion $(1, 2)$ and changes no other inversion. The factor g' is already increasing at those positions. Thus

$$F_M = F_{D'} \text{ with columns } (f, 1, 2) \text{ and } (3, 1, 2) \text{ deleted.} \tag{B.5}$$

Put $F = F_D$. Its kernel lifts to $\kappa' = (-X^{-1} Y \kappa, \kappa)$. On row $(1, 2)$, the two columns to be deleted both have entry one, all other columns involving the first position have entry zero, and $Y\kappa$ has entry zero. Hence their kernel coordinates α_f and α_3 are opposite. Since the kernel is one-dimensional, Equation (B.5) has full column rank exactly when $\alpha_f \neq 0$.

The position of the maximum in the new third minimal representative is $q' = q + 1$. Its inversions involving the first position include $(1, j)$ for all $j \geq 2$ except $j = q'$. If $\alpha_f = \alpha_3 = 0$, the remaining coordinates involving the first position must solve the equations on rows $(1, j)$, $j \geq 3$, using the rows of $\rho_3'^{-1}$ indexed by $j \geq 3$, $j \neq q'$. Their annihilator is the q'-th column of the corresponding lower-right block of ρ_3'. It follows that the equations are solvable exactly when

$$\Phi = \sum_{i=1}^{2} \sum_{(a,b) \in \mathrm{Inv}(x_i)} \kappa_{(i,a,b)}\, \rho_i'[1, a + 1](\rho_i^{-1} \rho_3)[b, q] = 0. \tag{B.6}$$

PatternMiner system output (continued)

There is no contribution from the third factor: such a term would require an inversion ending at the position of its maximum. By Equation (B.3), the coefficient of the fresh variable $\rho_i'[1, a+1]$ is $\Psi_{i,a}$. At least one is nonzero, so Φ is not identically zero and F_M is generically of full column rank.
Finally, the nonzero rational function persists at every extension. Normalize each lifted kernel so that its restriction to old columns equals the previous kernel. After t extensions a seed column (i, a, b) has become $(i, a+t, b+t)$, and its coordinate is unchanged. The corresponding lower-right blocks of $\rho_i^{-1}\rho_3$ are also unchanged, while $q_t = q+t$. Every newly added column has first position at most t; therefore none contributes to the sum with first position $a+t > t$. Consequently $\Psi^{(t)}_{i,a+t} = \Psi^{(0)}_{i,a}$. The same fresh-variable argument proves feasibility of M at every stage. Together with the proper projections containing the new dimension, this proves the assertion by induction. □

Exact seed calculations For the S_6 seed, divided differences give $c^{256134}_{215346,135264} = 0$. Table 7 supplies a nonzero term for every nonempty proper projection. In each row, $c^{z}_{p_E(u),p_E(v)} = 1$ and $z \le w_0 p_E(w_0 w)$, so the projected triple product is nonzero. The same calculation applies to the S_5 seed, whose ordinary coefficient is also zero.

Table 7: Nonzero terms in the proper projections of the two Schubert seeds.

n	Retained dimensions E	z	$w_0 p_E(w_0 w)$
6	$\{1\}$	213456	265431
6	$\{3\}$	146235	652431
6	$\{5\}$	123564	653214
6	$\{1,3\}$	246135	265431
6	$\{1,5\}$	213564	265314
6	$\{3,5\}$	236154	652314
5	$\{2\}$	35124	54321
5	$\{4\}$	12453	54213

The rank and kernel conditions for the S_6 seed can be checked at the following three upper unitriangular matrices. Listing their strictly upper triangular entries in lexicographic order of (a, b) gives

$$(\rho_1[a,b])_{a<b} = (32, 25, 46, 17, 25, 81, 60, 45, 33, 2, 24, 62, 15, 65, 28),$$
$$(\rho_2[a,b])_{a<b} = (53, 56, 78, 26, 29, 97, 53, 91, 54, 35, 51, 11, 6, 94, 54),$$
$$(\rho_3[a,b])_{a<b} = (95, 10, 36, 64, 23, 65, 34, 7, 11, 16, 25, 47, 37, 98, 47).$$

Order rows of F_D by (a, b) and columns by (i, a, b), both lexicographically. Deleting rows $(2,3), (3,5), (4,5)$ and column $(3,5,6)$ gives a 12×12 minor of determinant -349758000. Vanishing bounds the generic rank of this 15×13 matrix by twelve, so its generic column nullity is one. Normalize the kernel by $\kappa_{(3,5,6)} = 1$. At this specialization, $q = 5$ and $\Psi_{1,3} = -191797/254 \neq 0$. Cramer's rule expresses this normalized kernel near the specialization, proving that $\Psi_{1,3}$ is not identically zero. Lemma B.1.4 now proves Theorem B.1.3 for $n \ge 6$, and the separate seed proves it for $n = 5$.

B.1.6 Equivariant positivity and nonvanishing

The *double and triple Schubert structure constants* are defined by

$$\mathfrak{S}_u(x;t)\mathfrak{S}_v(x;t) = \sum_w c^w_{u,v}(t)\mathfrak{S}_w(x;t),$$
$$\mathfrak{S}_u(x;y)\mathfrak{S}_v(x;t) = \sum_w c^w_{u,v}(t;y)\mathfrak{S}_w(x;t).$$

The first coefficients are the structure constants of $H^*_T(\mathrm{Fl}_n;\mathbb{Z})$ when $u, v, w \in S_n$. The second multiplication uses two independent parameter sets and need not be symmetric in u, v. Set $\alpha_i = t_{i+1} - t_i$ and $\Phi^+ = \{\alpha_a + \cdots + \alpha_b : 1 \le a \le b < n\}$. Graham positivity gives $c^w_{u,v}(t) \in \mathbb{Z}_{\ge 0}[\alpha_1, \ldots, \alpha_{n-1}]$ (Graham, 2001).

PatternMiner system output (continued)

Theorem B.1.5 (LLM-generated). *For $u, v, w \in S_n$ with $s_i u < u \leq w$ in Bruhat order, the difference*

$$c^w_{u,v}(t) - \alpha_i c^w_{s_i u, v}(t)$$

is a nonnegative integer combination of products of distinct roots in $\Phi^+ \setminus \{\alpha_i\}$.

Corollary B.1.6 (LLM-generated). *For $u, v, w \in S_\infty$,*

$$c^w_{u,v}(t) \neq 0 \iff u \leq w \text{ and } c^w_{u,v}(t; y) \neq 0, \tag{B.7}$$

$$\iff c^w_{u,v}(t; y) \neq 0 \text{ and } c^w_{v,u}(t; y) \neq 0. \tag{B.8}$$

Consequently Conjectures A and B of Chen et al. (2026a) hold.

Proof. By Theorem 2.2 of Chen et al. (2026a), $c^w_{u,v}(t; y) \neq 0$ if and only if there is a length-additive factorization $u = az$ with $c^w_{z,v} > 0$. If $u \leq w$, every element of the associated chain in left weak order from z to u lies below w. The coefficient $c^w_{z,v}(t)$ is the positive integer $c^w_{z,v}$, since its degree is zero. Iterating Theorem B.1.5 gives a nonzero monomial lower bound for $c^w_{u,v}(t)$. Conversely, specialization $y = t$ and the necessary Bruhat bound prove the reverse implication in Equation (B.7). The same description of triple nonvanishing gives $c^w_{v,u}(t; y) \neq 0 \Rightarrow u \leq w$. Equation (B.7) therefore proves the reverse implication in Equation (B.8); specialization and symmetry of double multiplication prove the forward implication. This is Conjecture B. By Theorem 1.1 of Chen et al. (2026a), it is equivalent to Conjecture A, which states that $c^w_{u,v}(t) \neq 0$ exactly when there are length-additive factorizations $u = u_1 u_2$ and $v = v_1 v_2$ with $c^w_{u,v_2} > 0$ and $c^w_{u_2,v} > 0$. □

Geometric setup for the proof Let $G = \mathrm{GL}_n(\mathbb{C})$, let B, B^- be opposite *Borel subgroups*, and let $T = B \cap B^-$. Choose the identification of the *character lattice* with the parameters t so that the weights of $U = R_u(B)$ are Φ^+. For upper triangular matrices, this means that the a, b entry, $a < b$, has weight $t_b - t_a$. Put $N = \dim G/B$ and

$$C_w = BwB/B, \qquad X_w = \overline{C_w}, \qquad X^u = \overline{B^- uB/B}, \qquad \sigma_u = [X^u]_T.$$

We use *equivariant Chow classes*, identified here with equivariant cohomology classes. The coefficient can be extracted by

$$c^w_{u,v}(t) = \int_{X_w} \sigma_u \sigma_v. \tag{B.9}$$

Indeed, $\int_{X_w} \sigma_y = \delta_{y,w}$. If $\ell(y) < \ell(w)$, the integral has negative degree. Otherwise $X^y \cap X_w$ is empty unless $y = w$, in which case the intersection is the transverse fixed point.

Effective cycles and multidegrees Root coordinates identify U with an affine T-representation, with one coordinate of each vector weight $\beta \in \Phi^+$. For a closed T-invariant subvariety Y of such a representation V, write $\mathrm{mdeg}^V_T(Y) = [Y]_T \in A^*_T(V) = A^*_T(\mathrm{pt})$. We omit the superscript when the ambient representation is specified.

Lemma B.1.7 (Knutson & Miller, 2005, Theorem D). *An effective T-invariant cycle in a representation with positive-root weights has multidegree equal to a nonnegative integer combination of products of distinct coordinate weights.*

This classical lemma supplies the multidegree positivity used in the proof of Theorem B.1.5. In particular, every nonempty invariant subvariety has nonzero multidegree: a product of positive roots is nonzero and has nonnegative simple-root coefficients.

Intersection over a minimal parabolic subgroup Fix i, let P_i be the *minimal parabolic* containing B, and put $K_i = R_u(P_i)$ and $d = N - 1$. Its coordinate weights are precisely $\Phi^+ \setminus \{\alpha_i\}$. A representative n_i of s_i normalizes K_i.

Suppose $s_i z > z$. The P_i-orbit $\Omega_z = C_z \sqcup C_{s_i z}$ maps to $P_i/B \simeq \mathbb{P}^1$. The stabilizer of zB in P_i lies in B and contains TU_{α_i}, so K_i acts transitively on each fiber, of dimension $\ell(z)$. Label the point whose fiber is C_z by ∞ and the other fixed point by 0. Then $C_{s_i z}$ lies over $\mathbb{A}^1 = \mathbb{P}^1 \setminus \{\infty\}$, with fiber $n_i C_z$ at 0, and n_i interchanges 0 and ∞.

PatternMiner system output (continued)

For any h, the image of $X^h \cap \Omega_z$ is

$$\begin{cases} \mathbb{P}^1, & h \le z, \\ \{0\}, & h \le s_i z,\ h \not\le z, \\ \varnothing, & h \not\le s_i z. \end{cases} \tag{B.10}$$

Choose a cocharacter that vanishes on α_i and is positive on all other positive roots. Its limit contracts each fiber to the Levi fixed component. The variety X^h is closed and T-stable. Its intersection with this component is B_i^--stable and has fixed points given by Bruhat order, which gives the three stated possibilities. The zero section supplies every point of the indicated image.

We use the standard Richardson-variety fact that $X^h \cap C_y$, when nonempty, is irreducible of dimension $\ell(y) - \ell(h)$. If $h \le z$, the affine variety $X^h \cap C_{s_i z}$ surjects onto $\mathbb{A}^1$. Every fiber has dimension $\ell(z) - \ell(h)$: it is cut out in an integral affine variety by one nonconstant regular function minus a scalar. The fiber at ∞, namely $X^h \cap C_z$, has the same dimension. If $h \le s_i z$ but $h \not\le z$, the whole patch lies over 0 and has dimension $\ell(z) + 1 - \ell(h)$.

Consider the closed incidence scheme

$$\mathcal{Z} = \{(g, F) \in K_i \times X_w : F \in n_i X^u, \quad g^{-1} F \in X^v\}. \tag{B.11}$$

For its part $\mathcal{Z}_y$ over C_y, put $D_y = d + \ell(y) - \ell(u) - \ell(v)$. The nonempty strata have dimensions

	conditions	$\dim \mathcal{Z}_y$
$s_i y > y$	$u \le y,\ v \le y$	D_y
$s_i y > y$	$u \not\le y,\ u \le s_i y,\ v \le y$	$D_y + 1$
$s_i y < y$	$u \le s_i y,\ v \le y$	D_y.

(B.12)

All other strata are empty. Here is the parameter count. If $s_i y > y$, the cell is one K_i-orbit. Translation by n_i identifies $n_i X^u \cap C_y$ with the fiber at 0 of $X^u \cap C_{s_i y}$. Its dimension is respectively $\ell(y) - \ell(u)$ or $\ell(y) + 1 - \ell(u)$ in the first two rows. For each such F, the choices of g have dimension $d - \ell(v)$.

If $s_i y < y$, put $z = s_i y$. When $u \not\le z$, the image of X^u in Equation (B.10) is empty or $\{0\}$, whereas $n_i C_y$ avoids that fiber; hence the stratum is empty. When $u \le z$, every fiber of $n_i X^u \cap C_y$ has dimension $\ell(y) - \ell(u) - 1$. If $v \le z$, the choices of g have dimension $d - \ell(v)$ and the base point varies in $\mathbb{A}^1$. If $v \not\le z$ but $v \le y$, only base point 0 is allowed and the choices of g have dimension $d + 1 - \ell(v)$. Both give the last row.

Under the hypothesis $u \le w$, a stratum in the second row has $y < w$. Thus $\ell(y) + 1 \le \ell(w)$, and every stratum has dimension at most $d + \ell(w) - \ell(u) - \ell(v)$. This is the expected dimension in Equation (B.11). The intersection dimension inequality supplies the reverse bound at every nonempty component. Consequently the three subvarieties defining this incidence scheme in the smooth variety $K_i \times G/B$ meet properly.

Proof of Theorem B.1.5. The proper intersection above defines an effective intersection cycle. The Schubert varieties and their indicated inverse images are Cohen–Macaulay; the expected codimension ensures positive intersection multiplicities at the generic points of the components. Projection $\pi : K_i \times X_w \to K_i$ is proper. The action map and the projection $K_i \times G/B \to G/B$ induce the same pullback in equivariant Chow theory. In type A an explicit equivariant homotopy is given by $g \mapsto 1 + t(g - 1)$ in root coordinates, followed by the action on the flag variety. Also the left Demazure identity gives

$$[n_i X^u]_T = s_i \sigma_u = \sigma_u - \alpha_i \sigma_{s_i u} \qquad (s_i u < u).$$

Therefore

$$\pi_* [\mathcal{Z}]_{\text{int}} = \int_{X_w} (s_i \sigma_u) \sigma_v = c^w_{u,v}(t) - \alpha_i c^w_{s_i u, v}(t).$$

Proper pushforward preserves effective cycles: components with positive-dimensional generic fibers contribute zero, and the others contribute their positive generic degrees. Apply Lemma B.1.7 to the representation K_i. □

PatternMiner system output (continued)

B.1.7 A LIMITATION OF THE STATISTICS AT CANDIDATE INDICES

Proposition B.1.8 (LLM-generated)**.** *For every $N \geq 10$, there exist $u, v, w, w' \in S_N$ with a common supplied left descent and nonempty disjoint answer sets, for which the following data agree at every $j \in \mathrm{Des}_L(v)$:*

$$z^{-1}(j),\ z^{-1}(j+1), \qquad r_z(a,j),\ r_z(a,j+1) \quad (1 \leq a \leq N),$$
$$(xy^{-1})(e_{j+1} - e_j) \quad \textit{for ordered distinct } x, y \in \{u, v, z\},$$

where the comparison is between $z = w$ and $z = w'$. No deterministic rule depending only on these data and the common factors can choose a valid answer for both inputs.

Proof. Take

$$u = (1,3,8,2,6,4,7,5,9,10), \qquad v = (1,6,2,3,7,5,8,4,9,10),$$
$$w = (1,10,5,2,6,4,8,3,7,9), \quad w' = (2,10,5,1,6,4,7,3,8,9).$$

Their lengths are $9, 8, 16, 16$, respectively, and $\mathrm{Des}_L(v) = \{4, 5\}$. Exact Schubert multiplication gives

z	$c^z_{s_5u,v}$	c^z_{u,s_4v}	c^z_{u,s_5v}
w	1	0	2
w'	1	1	0

so $\mathcal{V}(u, v, w) = \{5\}$ and $\mathcal{V}(u, v, w') = \{4\}$. The displayed statistics agree by direct substitution. The same inputs to a deterministic rule give the same output, which cannot belong to both sets. Adjoining fixed points preserves the coefficients and all stated equalities, proving the assertion for every $N \geq 10$. □

The model distinguishes inputs with identical candidate statistics by using other positions. Table 8 records the predictions when first-MLP activations outside the candidate set are transferred through layer-1 keys and values.

Table 8: Interventions on inputs with identical candidate statistics.

Input or intervention	Model answer
Unmodified w, w'	5, 4
Transfer from w to w'	$4 \to 5$
Transfer from w' to w	$5 \to 5$
All eight pairs with identical candidate statistics	Donor-valid answer in 12/16 directions

B.2 KRONECKER COEFFICIENTS

B.2.1 MATHEMATICAL SETTING

Partitions $\lambda \vdash n$ index the irreducible complex representations V^λ of S_n. The Kronecker coefficient is the tensor-product multiplicity

$$V^\lambda \otimes V^\mu \cong \bigoplus_{\nu \vdash n} (V^\nu)^{\oplus g_{\lambda,\mu,\nu}}.$$

We follow the character and symmetric-function conventions of Sagan (2001). If $\rho = (1^{m_1} 2^{m_2} \cdots)$ and $z_\rho = \prod_i i^{m_i} m_i!$, character orthogonality gives the exact formula

$$g_{\lambda,\mu,\nu} = \sum_{\rho \vdash n} \frac{\chi^\lambda(\rho)\chi^\mu(\rho)\chi^\nu(\rho)}{z_\rho}.$$

The implementation evaluates characters by the *Murnaghan–Nakayama rule* and checks them against the *hook-length formula* and orthogonality. Figure 3 shows $V^{(2,1)} \otimes V^{(2,1)} \cong V^{(3)} \oplus V^{(2,1)} \oplus V^{(1,1,1)}$.

Automatically generated by the PatternMiner system

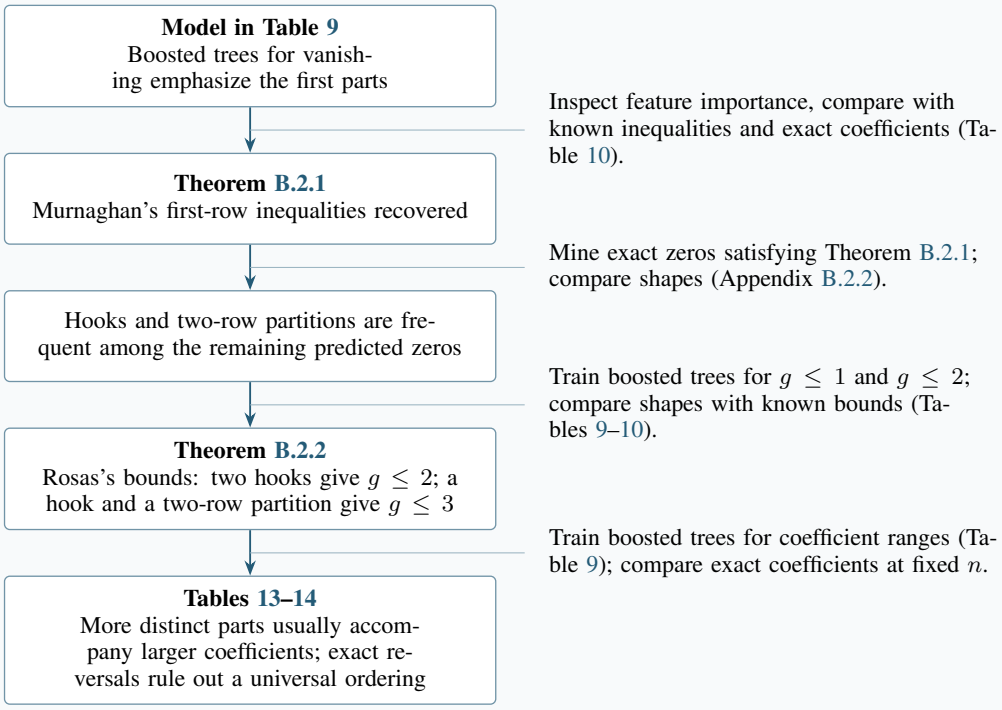


Figure 8: Kronecker rediscoveries and the subsequent empirical comparison.

B.2.2 Models and hard-negative experiments

Each input is an unordered triple of partitions, sorted lexicographically and encoded by its parts padded with zeros, together with n. We predict vanishing $(g = 0)$, the thresholds $g \leq 1$ and $g \leq 2$, and a seven-class range of g: $g = 0,\ g = 1,\ g = 2,\ \ldots,\ g = 5,\ g \geq 6$.

Training uses all 1,095 such triples for $3 \leq n \leq 7$. Evaluation uses 6,984 triples at $n = 8, 9$ and 44,100 at $n = 10, 11$. Padding is extended before training a model evaluated at larger n. The retained gradient-boosting implementation uses 200 trees of maximum depth three. Table 9 reports classification accuracy.

Table 9: Kronecker classification accuracy.

Target	Evaluation sizes	Logistic regression	Boosted trees
$g = 0$	$8, 9$	77.51%	89.09%
$g \leq 1$	$8, 9$	76.92%	87.41%
$g \leq 1$	$10, 11$	72.11%	79.27%
$g \leq 2$	$8, 9$	80.70%	87.41%
$g \leq 2$	$10, 11$	69.84%	72.22%
$g \in \{0, 1, \ldots, 5, \geq 6\}$	$10, 11$	30.94%	46.46%

Hard negatives for vanishing are zeros satisfying Murnaghan's inequalities in Theorem B.2.1. The first search also excluded triples containing (n) or (1^n), whose coefficients follow from character orthogonality. Among 5,600 remaining test triples at $n = 8, 9$, the classifier correctly identified 1,145 zeros that passed those inequalities. Their first and second parts suggested testing the analogous inequalities after removing two rows. Exact counterexamples rejected that extension; Proposition B.2.4 gives a small example.

The same strategy applies to Rosas's bounds in Theorem B.2.2: both bounds hold on every sampled triple, and classifier predictions for the magnitude tasks respect them almost everywhere. Table 10 summarizes all four checks.

PatternMiner system output (continued)

Table 10: Kronecker tests of necessary conditions and coefficient bounds.

Condition or prediction	Population	Violations
Murnaghan's inequalities on exact coefficients	$3 \le n \le 9$; 1,479 exclusions	0
Triangle inequalities after deleting two rows	Same sizes; 4,014 exclusions	1,344 positive coefficients
Two hooks imply $g \le 2$	2,308 exact triples	0
A hook and a two-row partition imply $g \le 3$	2,262 exact triples	0
Classifier prediction respects the two-hook bound	5,236 triples at $n = 10, 11$	0
Classifier prediction respects the hook/two-row bound	5,232 triples at $n = 10, 11$	15

A further check asked whether classifier predictions for the magnitude task respect the semigroup property of Theorem B.2.3. The theorem was confirmed on 35 sampled pairs with positive coefficients. Predictions for the 1,614 sampled sums satisfied this inequality more often than a size-matched baseline, with a small improvement (78.1% vs. 74.2%) (Table 11). Much of the agreement reflects the baseline effect that predictions increase with n, and a sum always has a larger n than either summand.

Table 11: Semigroup-inequality agreement for magnitude predictions.

	Agreement
Classifier predictions	78.1%
Size-matched baseline	74.2%

Returning to Rosas's bounds directly, after removing the triples covered by either bound, the dataset leaves 348 training triples and 35,730 evaluation triples at $n = 10, 11$. Table 12 reports retraining on this population. Agreement with a coefficient bound in Table 10 measures a weaker requirement than predicting the coefficient itself.

Table 12: Kronecker prediction outside the two special-shape bounds.

Predictor for $g \le 2$ after removing Rosas's cases	Accuracy
Retrained boosted trees	59.72%
Constant majority prediction	53.23%

B.2.3 Rediscovered coefficient bounds

Theorem B.2.1 (Mishna et al., 2021, Equation (22)). *If $g_{\lambda,\mu,\nu} > 0$, then*

$$n - \lambda_1 \le (n - \mu_1) + (n - \nu_1),$$

and the two inequalities obtained by permuting λ, μ, ν also hold.

Feature importance in the boosted trees emphasized the first parts, prompting comparison with Murnaghan's inequalities. The system then mined zeros satisfying these inequalities to investigate the remaining predictions.

Theorem B.2.2 (Rosas, 2001, Corollaries 15 and 17). *For $\lambda, \mu, \nu \vdash n$, the following bounds hold, with the third partition arbitrary:*

$$g_{\lambda,\mu,\nu} \le 2 \quad \textit{if two partitions are hooks},$$
$$g_{\lambda,\mu,\nu} \le 3 \quad \textit{if two partitions are a hook and a two-row partition}.$$

A hook is a partition $(n - r, 1^r)$.

Hard-negative mining found an excess of hooks and two-row partitions among correctly predicted zeros. Subsequent coefficient-threshold experiments led the agent to Rosas's bounds, which it checked by exact computation and against the model predictions (Table 10).

PatternMiner system output (continued)

Theorem B.2.3 (Bessenrodt & Bowman, 2017, Proposition 2.4)**.** *If $g_{\lambda,\mu,\nu} > 0$ and $g_{\alpha,\beta,\gamma} > 0$, then*

$$g_{\lambda+\alpha,\mu+\beta,\nu+\gamma} \geq \max(g_{\lambda,\mu,\nu}, g_{\alpha,\beta,\gamma}).$$

The system used this known inequality to test the magnitude predictions on sums of partition triples (Table 11).

Proposition B.2.4 (LLM-generated)**.** *The analogous triangle condition on $n-\lambda_1-\lambda_2$, $n-\mu_1-\mu_2$, and $n-\nu_1-\nu_2$ is false.*

Proof. For $\lambda = (1,1,1)$ and $\mu = \nu = (2,1)$, these three quantities are $1, 0, 0$. The tensor-product decomposition above gives $g_{\lambda,\mu,\nu} = 1$. □

B.2.4 Empirical comparison by the number of distinct parts

For each triple, we compare the coefficient with

$$\#\{\lambda_i : \lambda_i > 0\} + \#\{\mu_i : \mu_i > 0\} + \#\{\nu_i : \nu_i > 0\}.$$

Pairs of triples are compared within the same n, omitting equal values of this statistic. A reversal means that the triple with the larger statistic has the smaller coefficient. The $n = 10, 11$ experiment sampled 30,223 comparable pairs; the later experiment sampled 6,000 triples at $n = 15$.

Table 13: Reversals in the comparison by distinct parts.

Difference	Pairs $n = 10, 11$	Reversals $n = 10, 11$	Reversals $n = 15$
1	16,170	18.86%	25.48%
2	9,329	7.49%	12.04%
3	3,599	2.53%	5.47%
≥ 4	1,125	0.44%	1.41%

The comparisons in Table 13 concern individual triples. At a gap of one, reversals occur in roughly one in five comparisons at $n = 10, 11$ and one in four at $n = 15$; at a gap of at least four, the respective rates are 0.44% and 1.41%. These reversals nonetheless preclude an unconditional ordering by the displayed statistic.

Table 14: Association with Kronecker coefficients at $n = 10, 11$.

Statistic summed over the three partitions	Spearman correlation
Number of distinct parts	0.6421
Durfee-square size	0.4857

Table 14 compares two partition statistics on the same $n = 10, 11$ population, motivated by Pak–Panova's polynomial growth theorem, which assumes bounded Durfee-square sizes (Pak & Panova, 2023). The number of distinct parts correlates more strongly with the Kronecker coefficient than the Durfee-square size does. Pak–Panova's theorem bounds coefficient growth when Durfee-square sizes are bounded; the present observation compares individual coefficients empirically. However, it shows that, for the specific task of comparing two triples, the statistic in Table 13 is a better-performing choice than the one their proof uses.

B.3 CHARACTER-TABLE ROW SUMS

For a partition $\lambda \vdash n$, let χ^λ be the corresponding *irreducible character* of S_n and define

$$R(\lambda) = \sum_{\rho \vdash n} \chi^\lambda(\rho).$$

There is one summand per conjugacy class. Stanley's Problem 12 asks for a combinatorial interpretation of this integer; the representation-theoretic setting is described by Sagan (2001) and Sundaram (2018). For the sign character of S_3, the three entries are $1, -1, 1$, giving the sum 1 in Figure 4.

Automatically generated by the PatternMiner system

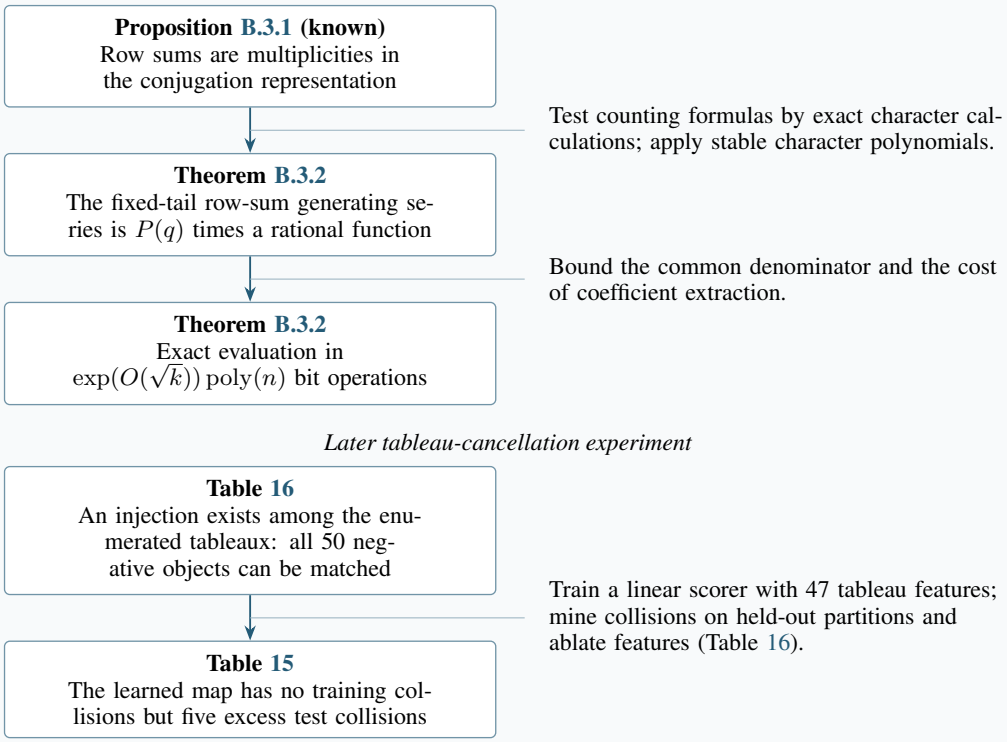


Figure 9: Character-row evaluation and the later search for a cancellation map.

Proposition B.3.1 (Sundaram, 2018)**.** *The number $R(\lambda)$ is the multiplicity of the irreducible representation indexed by λ in the conjugation action of S_n on $\mathbb{C}[S_n]$. In particular, $R(\lambda) \geq 0$.*

This classical interpretation supplies the character decomposition underlying the tableau-cancellation experiments below.

B.3.1 MODELS AND CANCELLATION EXPERIMENTS

The learning experiment sought an injection between finite sets of tableaux arising in the character decomposition. We used the expression

$$(s_{(2,2)} + s_{(2,1,1)})h_2[s_{(2,1)}] - s_{(1)}h_3[s_{(2,1)}], \tag{B.13}$$

where brackets denote *plethysm*. For $\nu \vdash 10$ with $\nu_1 \leq 5$, we checked its s_ν coefficient against $[s_{(5,\nu)}]h_3[s_{(2,2,1)}]$. Such plethysms occur in the *Frobenius characteristic* of the conjugation representation through its decomposition by cycle type (Sundaram, 2018). Expanding the positive and negative terms of (B.13) gives the two sets to be matched; an injection would cancel the negative contribution.

A linear model scored each possible image using 47 integer-valued features of the tableaux, the added cell, and the summand contributing the source. Each source selected its highest-scoring image independently. We optimized the weights by 600 coordinate proposals from each of three

PatternMiner system output (continued)

initializations and selected the model on validation data. Output partitions were disjoint across the three splits, although the same constituent tableaux could occur in more than one split. The finite family had been explored symbolically before training; this split evaluates the learned scorer on held-out output partitions. We counted excess collisions as the number of sources minus the number of distinct images. The geometric baseline scores an image by $4a + 2b - c + e$, where a counts shared cells, b counts agreeing entries, c counts disagreeing entries, and e indicates inclusion of the added cell. The lexicographic baseline chooses the first valid image in a fixed order.

Table 15: Independent image selection in the tableau cancellation experiment.

	Sources	Output	Excess collisions		
Split		partitions	Learned	Geometric	Lexicographic
Training	22	10	0	8	12
Validation	11	4	1	1	7
Test	17	7	5	6	10

The remaining nine output partitions have no negative sources and require no matching. The selected rule was injective on every training output and on three of the seven test outputs. Its five test collisions exceeded the median of four from random linear scores (Table 16). The random scores are an unselected reference distribution. Removing the feature measuring entry disagreement reduced test collisions while increasing validation collisions. Thus this experiment supplied explicit failures of the proposed map and no improvement over the random-score comparison.

Table 16: Exact checks and feature ablations for the cancellation experiment.

Calculation	Result
Exact tableau enumeration	97 positive and 50 negative objects over 30 output partitions.
Lookup injection	All 50 negative objects matched; 47 unmatched positive objects over 22 outputs.
Littlewood–Richardson enumeration	Two enumerations agree on all 240 ordered families.
Ambient plethysm coefficients	All 30 signed counts agree with Murnaghan–Nakayama and Jacobi–Trudi extraction.
100 random linear scores	Minimum, median, and maximum test collisions: $2, 4, 8$.
Entry-disagreement feature removed	Validation collisions $1 \to 4$; test collisions $5 \to 4$.
Added-cell entry feature removed	Test collisions $5 \to 6$.

The two Littlewood–Richardson enumerations used separate tableau predicates; the ambient coefficient calculations shared power-sum arithmetic. A concrete collision occurs at output $(4, 3, 3)$: sources obtained by adding cells $(3, 3)$ and $(1, 4)$ to shapes $(4, 3, 2)$ and $(3, 3, 3)$ select the same image. Both sources arise by adding a horizontal strip of three cells to $(3, 3)$ before adding the final cell, and select the same positive tableau with inner shape $(2, 2, 2)$, content $(2, 1, 1)$, and reading word 1123. The exhaustive matching in Table 16 establishes that an injection exists on this finite domain, while the learned rule fails to construct one.

Follow-up development experiments We subsequently reused all 50 negative objects as development data, training a 47-parameter linear scorer and a 784-parameter tanh MLP with three seeds each. One linear-model seed and two MLP seeds achieved collision-free assignments. An integer rule with 22 nonzero coefficients, extracted from the successful linear scorer, preserved all 50 selected images. These results concern the same 50 objects used for development. A subsequent task audit identified a coordinated rank matching for unrestricted pairing within each output partition. The tested restriction allowing a one-cell change of the inner diagram admitted a greedy matching through nested neighbor sets. Thus the development fits do not establish a learning advantage over elementary combinatorial baselines.

B.3.2 Evaluation for partitions with a fixed tail

The stable character polynomial gives an evaluation formula with complexity controlled by the number of boxes below the first row. Fix $\nu \vdash k$, $k \geq 1$, and put $P(q) = \prod_{i \geq 1}(1 - q^i)^{-1}$ and

PatternMiner system output (continued)

$D_k(q) = \prod_{i=1}^{k}(1-q^i)$. Let $a_{\nu,n}$ be the sum over conjugacy classes of the virtual character with Frobenius characteristic $s_{(n-k,\nu)}$, defined by Jacobi–Trudi with $h_0 = 1$ and $h_j = 0$ for $j < 0$. For $n \geq k + \nu_1$, this is the ordinary row sum $R((n-k,\nu))$.

Theorem B.3.2 (LLM-generated)**.** *There is an effectively computable polynomial $N_\nu(q) \in \mathbb{Z}[q]$ of degree at most $k(k+1)/2$ such that*

$$\sum_{n\geq 0} a_{\nu,n} q^n = P(q)\frac{N_\nu(q)}{D_k(q)}.$$

The numerator can be computed in $\exp(O(\sqrt{k}))\,\mathrm{poly}(k)$ bit time and space. The coefficients through degree n can be computed in $\exp(O(\sqrt{k}))\,\mathrm{poly}(n)$ bit time. In particular, evaluation takes polynomial time in n when $k = O((\log n)^2)$.

Proof. For $\gamma = (1^{m_1}2^{m_2}\cdots) \vdash k$, put $z_\gamma = \prod_i i^{m_i} m_i!$. The stable character polynomial of Garsia & Goupil (2009, Proposition I.1) is

$$\sum_{\gamma\vdash k} \frac{\chi^\nu(\gamma)}{z_\gamma} \prod_i \sum_{a=0}^{m_i} (-1)^{m_i-a} i^a \binom{m_i}{a} (X_i)_a,$$

where X_i is the number of i-cycles and $(X)_a = X(X-1)\cdots(X-a+1)$. This polynomial also gives the generalized-Schur character below the ordinary range. Indeed, expanding the first row of Jacobi–Trudi gives an integer combination of permutation characters induced from Young subgroups valid for every n. It agrees with the stable polynomial for all sufficiently large X_1 and every nonnegative choice of the other cycle counts. Polynomial identity then gives equality in every degree.

Summing over cycle types amounts to substituting

$$(X_i)_a \longmapsto \frac{a!q^{ia}}{(1-q^i)^a}$$

after extracting the factor $P(q)$. Define

$$B_{i,m}(q) = \sum_{a=0}^{m} (-1)^{m-a} i^a \binom{m}{a} a! q^{ia}(1-q^i)^{m-a}, \qquad G_\gamma(q) = \prod_i (1-q^i)^{m_i}.$$

Then

$$N_\nu(q) = \sum_{\gamma\vdash k} \frac{\chi^\nu(\gamma)}{z_\gamma} \frac{D_k(q)}{G_\gamma(q)} \prod_i B_{i,m_i}(q).$$

The multiplicity of the hth cyclotomic polynomial in G_γ is $\sum_{h|i} m_i \leq \lfloor k/h \rfloor$, so G_γ divides D_k in $\mathbb{Z}[q]$. Since $\deg B_{i,m} \leq im$, each summand has degree at most $k(k+1)/2$. Integrality follows from the Jacobi–Trudi expansion in binomial polynomials: $\binom{X_i}{a}$ maps to $q^{ia}/(1-q^i)^a$, and every denominator divides D_k by the same argument. Computation can therefore use integer arithmetic after multiplying by $k!$, followed by exact division.

Memoized Murnaghan–Nakayama computes the required character values using at most $(\sum_{j\leq k} p(j))^2$ pairs of partitions and polynomially many operations per pair. The bound $p(k) = \exp(O(\sqrt{k}))$ gives the stated number of operations. Character values, factorial scales, and polynomial coefficients have polynomial bit length in k: for example, $\|B_{i,m}\|_1 \leq m!(i+2)^m$, and all polynomial degrees are $O(k^2)$. Writing $\delta = k(k+1)/2$, the coefficients of $G_\gamma(q)^{-1}$ through degree δ are bounded by $\binom{\delta+k}{k}$. Since $\|D_k\|_1 \leq 2^k$, the coefficients of D_k/G_γ also have polynomial bit length in k. This proves the numerator bound. Multiplying by $P(q)/D_k(q)$ through degree n requires only $O(n^2)$ additions of integers with polynomial bit length, proving the coefficient bound. □

B.4 KAZHDAN–LUSZTIG COEFFICIENTS

For $x \le w$ in Bruhat order on S_n, let $P_{x,w}(q)$ be the Kazhdan–Lusztig polynomial, normalized by $P_{w,w} = 1$ and $\deg P_{x,w} \le (\ell(w) - \ell(x) - 1)/2$ for $x < w$. Stanley's Problem 16 asks for a combinatorial interpretation of its coefficients. We use the conventions of Björner & Brenti (2005, Chapter 5). Figure 5 shows $P_{1234,3412}(q) = 1 + q$.

Put $d = \ell(w) - \ell(x)$ and normalize the *R-polynomial* by

$$R_{x,w}(q) = q^{d/2} \widetilde{R}_{x,w}(q^{1/2} - q^{-1/2}).$$

The polynomial $\widetilde{R}_{x,w}$ is monic of degree d, has nonnegative coefficients, and contains only powers of the same parity as d. For odd $d > 0$, write $\mu(x, w) = [q^{(d-1)/2}]P_{x,w}$; set $\mu(x, w) = 0$ otherwise. Combinatorial invariance asserts that $P_{x,w}$ depends only on the isomorphism type of the interval $[x, w]$ as a partially ordered set.

Automatically generated by the PatternMiner system

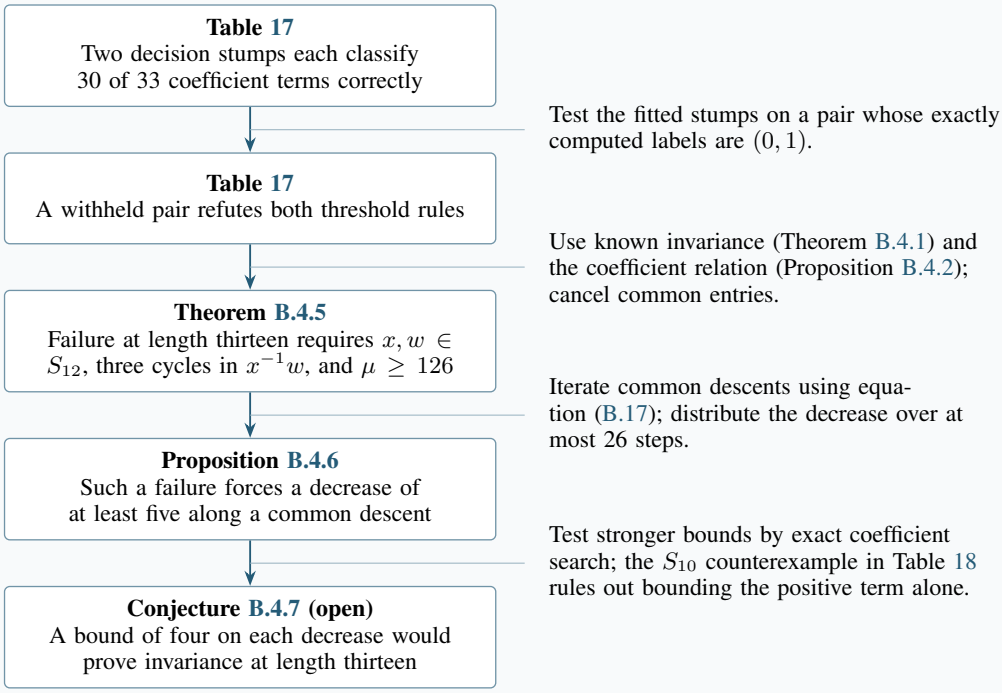


Figure 10: Kazhdan–Lusztig counterexamples, proved reductions, and the remaining conjecture.

B.4.1 MODELS AND COUNTEREXAMPLE EXPERIMENTS

The learning experiment classified summands in a coefficient calculation as positive or zero. It used 33 terms from five permutation pairs after cancellation of common entries. Exact integer Kazhdan–Lusztig recurrence supplied the labels; Bruhat-order and descent tests selected eligible examples. Decision stumps, each using one feature threshold, were fitted to these terms and tested on a further pair with different exact labels (Table 17). The two best-fitting rules both failed to distinguish that pair.

Table 17: Decision stumps for vanishing of individual terms.

Experiment	Result
Training labels	16 positive and 17 zero terms from five permutation pairs.
Best decision stumps	Two tied rules, each with 3 errors among 33 terms.
Withheld pair	Exact labels $(0, 1)$; both rules predict $(1, 1)$.

PatternMiner system output (continued)

B.4.2 KNOWN INVARIANCE AND THE POSSIBLE COEFFICIENT DIFFERENCE

Theorem B.4.1 (Esposito et al., 2026, Corollary 1.3)**.** *For isomorphic type-A Bruhat intervals I, I', one has $\widetilde{R}_I \equiv \widetilde{R}_{I'} \pmod{z^8}$. If their length is at most twelve, then $P_I = P_{I'}$.*

This known invariance theorem determines the possible coefficient difference at length thirteen in Proposition B.4.2.

Proposition B.4.2 (LLM-generated)**.** *For isomorphic type-A intervals I, I' of length thirteen, there is an integer Δ such that*

$$\widetilde{R}_I(z) - \widetilde{R}_{I'}(z) = \Delta z^9, \tag{B.14}$$

$$P_I(q) - P_{I'}(q) = \Delta(q^2 - 9q^3 + 36q^4 - 84q^5 + 126q^6). \tag{B.15}$$

Proof. The coefficients below degree eight agree by Theorem B.4.1. The coefficient of q in P is invariant by Patimo (2021, Corollary 6.2). Inversion gives $[z^{d-2}]\widetilde{R}_{x,w} = d + [q]P_{x,w} - \#\{y \in [x,w] : \ell(y) = \ell(w) - 1\}$, so this coefficient is invariant as well. Monicity and parity leave only degree nine. Every proper subinterval has length at most twelve, so subtracting the two Kazhdan–Lusztig inversion identities gives

$$q^{13}D(q^{-1}) - D(q) = \Delta q^2(q-1)^9, \qquad D(q) = P_I(q) - P_{I'}(q).$$

Since $\deg D \le 6$, comparing coefficients through degree six proves (B.15). □

B.4.3 REDUCTION TO TWELVE POSITIONS

Cancellation removes entries that do not affect the interval. For $1 \le i, j \le n$, define

$$D_{x,w}(i,j) = \#\{k < i : w(k) \ge j\} - \#\{k < i : x(k) \ge j\}.$$

If $x(i) = w(i) = j$ and $D_{x,w}(i,j) = 0$, delete position i and value j from both permutations and relabel in increasing order. This preserves the interval length and $P_{x,w}$ (Warrington, 2011, Fact 8). The induced interval identification preserves the same data on every subinterval, and hence also $R_{x,w}$ by inversion. Following Warrington, call a pair *uncompressible* if no such deletion is possible.

Lemma B.4.3 (LLM-generated)**.** *Label a saturated chain from x to w by its position transpositions and form a graph on $[n]$ with these transpositions as edges. Its connected components are independent of the chain. They give a product decomposition of the interval and of its P- and $\widetilde{R}$-polynomials. If (x,w) is uncompressible, $d \ge 3$ is odd, and $\mu(x,w) > 0$, the graph is connected.*

Proof. Saturated chains in a Bruhat interval are connected by changes across rank-two diamonds (Björner & Brenti, 2005, Chapter 2). In type A, the two transposition factorizations in a diamond involve either three positions or two disjoint pairs; the connected components are unchanged. Thus every vertex of the interval belongs to the coset of the reflection subgroup generated by transpositions within those components. Restricting the permutations to each component and relabeling gives the product of its symmetric groups. The induced Bruhat graph on the coset is the Bruhat graph of this reflection subgroup (Esposito et al., 2026, Theorem 2.2). Its interval is therefore a product. Increasing paths in the factors have a unique increasing interleaving for the restricted reflection order, so Dyer's path formula gives the product for $\widetilde{R}$ (Björner & Brenti, 2005, Theorem 5.3.4). Inversion gives the product for P.
Two factors of positive length would force $\deg P_{x,w} \le (d-2)/2$, contradicting $\mu(x,w) > 0$. An isolated position i is fixed throughout every saturated chain. A cover not involving i cannot move a value across both position i and value $x(i)$, since that would violate the type-A cover criterion. Hence the prefix count at $(i, x(i))$ is unchanged along the chain, so $D_{x,w}(i, x(i)) = 0$. This permits cancellation and contradicts uncompressibility. □

The number of cycles in the relative permutation gives a second constraint. For a word of L transpositions on m positions, let k be the number of connected components of its graph and c the number of cycles of its product, including fixed points. Then

$$G = \frac{L - m - c + 2k}{2} \in \mathbb{Z}_{\ge 0}. \tag{B.16}$$

PatternMiner system output (continued)

Indeed, a transposition joining graph components decreases both k and c by one, a transposition splitting a cycle increases c by one, and a transposition joining two cycles within a component decreases c by one. These operations change G by $0, 0, 1$, respectively. Reversing the word preserves G, so deleting either end cannot increase it. Extending a saturated chain proves $G(a, b) \le G(x, w)$ for every subinterval; Lemma B.4.3 makes these quantities independent of the chains.

Lemma B.4.4 (LLM-generated). *Suppose an uncompressible pair (x, w) in S_m has odd length difference $d \ge 3$ and $\mu(x, w) > 0$. If $[z^{d-2r}]\widetilde{R}_{x,w} > 0$, then $m \le d - r + 1$.*

Proof. The graph is connected by Lemma B.4.3, so (B.16) gives $c \le d - m + 2$. Dyer's formula supplies a path of $d-2r$ transpositions with product $x^{-1}w$. Any such product requires at least $m-c$ transpositions. Thus $m - c \le d - 2r$, and adding the two inequalities gives the bound. □

Theorem B.4.5 (LLM-generated). *If combinatorial invariance fails for type-A intervals of length thirteen, there is an uncompressible pair $x \le w$ in S_{12} with connected transposition graph such that*

$$\ell(w) - \ell(x) = 13, \qquad \#\,\mathrm{cycles}(x^{-1}w) = 3, \qquad [q^6]P_{x,w} \ge 126.$$

For this pair, $G(x, w) = 0$.

Proof. Orient a pair of counterexamples so that $\Delta \ge 1$ in Proposition B.4.2. Nonnegativity gives $\mu \ge 126\Delta$ and $[z^9]\widetilde{R} \ge \Delta$ for the first interval. After cancellation, Lemma B.4.4, with $d = 13$ and $r = 2$, gives $m \le 12$. The largest leading coefficient through S_{11} is 28 (Warrington, 2011, Theorem 1), so $m = 12$. The two inequalities in that lemma now give $c \le 3$ and $c \ge 3$. Substitution in (B.16) gives $G = 0$. □

B.4.4 A SUFFICIENT INEQUALITY ALONG COMMON DESCENTS

For a common right descent s of x, w and odd $d \ge 3$, the Kazhdan–Lusztig recurrence gives

$$\mu(x, w) - \mu(xs, ws) = [q^{(d-3)/2}]P_{x,ws} - \sum_{\substack{x<z\le ws,\ zs<z \\ \ell(z)-\ell(x)\ \text{odd}}} \mu(x, z)\mu(z, ws). \tag{B.17}$$

Here $P_{u,v} = 0$ when $u \not\le v$. To obtain this identity, take the coefficient of $q^{(d-1)/2}$ in the usual recurrence. The term $P_{xs,ws}$ contributes $\mu(xs, ws)$, the term $qP_{x,ws}$ contributes the displayed coefficient, and the degree bounds retain exactly the displayed correction terms.

Proposition B.4.6 (LLM-generated). *If combinatorial invariance fails at length thirteen, there is an uncompressible pair in S_{12} with length difference thirteen, connected transposition graph, three cycles in $x^{-1}w$, and $\mu(x, w) > 0$, together with a common right descent s, for which*

$$\mu(x, w) - \mu(xs, ws) \ge 5.$$

Proof. Simultaneously applying a common descent preserves uncompressibility. For $s = (a, a+1)$, the prefix counts at common entries outside these two positions are unchanged. A common entry j at position a moves to $a+1$; both entries moved before it are less than j, so its cancellation condition is unchanged. A common entry at $a+1$ moves to a; both removed preceding entries exceed j, again leaving the difference unchanged. The relative permutation is conjugated by s, so its cycle count is preserved.

Start with the pair in Theorem B.4.5. Longest-element symmetry sends it to (w_0w, w_0x) and preserves μ (Warrington, 2011, Fact 6), uncompressibility, and the relative cycle count. The two lower lengths sum to $\binom{12}{2} - 13 = 53$, so choose the orientation with $\ell(x) \le 26$. Whenever $\mu(x, w) > 0$ and $d \ge 3$, every right descent of w is a descent of x (Warrington, 2011, Fact 5). Apply any such common descent. Uncompressibility and the cycle count persist, and Lemma B.4.3 gives connectedness at each positive step.

Each step lowers $\ell(x)$ by one. The leading coefficient must vanish by the time x is the identity, since the identity has no descent. Thus a sum of at most 26 successive differences in (B.17) is at least 126. One difference is at least $\lceil 126/26 \rceil = 5$. □

Conjecture B.4.7 (LLM-generated). *For every uncompressible pair $x \le w$ in S_{12} with length difference thirteen, connected transposition graph, three cycles in $x^{-1}w$, and $\mu(x, w) > 0$, every common right descent s satisfies $\mu(x, w) - \mu(xs, ws) \le 4$.*

PatternMiner system output (continued)

Proposition B.4.6 shows that this conjecture would imply combinatorial invariance at length thirteen. A coefficient search ruled out a simpler bound on the first term of (B.17). The interval in Table 18 has connected transposition graph and $G = 0$, but its highest possible coefficient is eight. Therefore those two conditions alone do not bound this coefficient by four.

Table 18: A counterexample to a coefficient bound based on connectedness and (B.16).

Quantity	Exact value
Lower permutation x	$(1, 5, 4, 3, 2, 9, 8, 7, 6, 10)$
Upper permutation v	$(5, 7, 8, 1, 9, 10, 2, 3, 4, 6)$
Length difference	10
Cycles of $x^{-1}v$	2
Transposition graph	Connected; $G(x, v) = 0$.
Polynomial	$P_{x,v} = 1 + 10q + 29q^2 + 29q^3 + 8q^4$.

B.5 Monomial immanants

Let h_r be the *complete homogeneous symmetric function*, with $h_0 = 1$ and $h_r = 0$ for $r < 0$. For partitions $\nu \subseteq \mu$ padded to length n, the Jacobi–Trudi matrix of $D = \mu/\nu$ is

$$H_D = (h_{\mu_i - \nu_j - i + j})_{1 \le i,j \le n}.$$

Let ϕ^λ be the class function of S_n with Frobenius characteristic m_λ. The monomial immanant is

$$\operatorname{Imm}_{\phi^\lambda}(H_D) = \sum_{w \in S_n} \phi^\lambda(w) \prod_{i=1}^{n} h_{\mu_i - \nu_{w(i)} - i + w(i)}.$$

Stanley's Problem 18 asks whether its Schur coefficients are nonnegative. We follow Lesnevich (2024) for immanant characters and Sagan (2001) for symmetric functions. In Figure 6, $D = (2, 1)$ and $\lambda = (2)$; the class function has values $0, 2$, so the immanant is $2h_3 = 2s_{(3)}$.

B.5.1 Character formulation

The Schur coefficients can be expressed through Kostka numbers. Put $N = |D|$, fix $\theta \vdash N$, and define

$$B_i = \mu_i + n - i, \qquad C_j = \nu_j + n - j, \qquad \alpha_j(w) = B_{w(j)} - C_j.$$

Write $K_{\theta,\alpha}$ for the number of *semistandard tableaux* of shape θ and content α, and set it to zero if any entry of α is negative. Permuting the content leaves the Kostka number unchanged. For a conjugacy class $C_\rho \subseteq S_n$, define

$$\Gamma^\theta_D(w) = \frac{n!}{|C_\rho|} \sum_{v \in C_\rho} K_{\theta,\alpha(v)} \qquad (w \in C_\rho).$$

This is Lesnevich's immanant character, with normalization chosen so that

$$[s_\theta] \operatorname{Imm}_{\phi^\lambda}(H_D) = \langle \Gamma^\theta_D, \phi^\lambda \rangle \tag{B.18}$$

for the usual character inner product (Lesnevich, 2024, Definition 2.8 and Lemma 2.9).

For a weakly increasing function $\eta : [n] \to [n]$ with $\eta(j) \ge j$, let $\widehat{\eta}(w) = \mathbf{1}\{w(j) \le \eta(j) \text{ for all } j\}$. Its conjugacy average defines the Stanley–Stembridge character

$$\Gamma_\eta(w) = \frac{n!}{|C_\rho|} \sum_{v \in C_\rho} \widehat{\eta}(v) \qquad (w \in C_\rho).$$

Our decomposition is first proved for the functions $K_{\theta,\alpha(w)}$ and $\widehat{\eta}(w)$, before averaging over conjugacy classes.

Automatically generated by the PatternMiner system

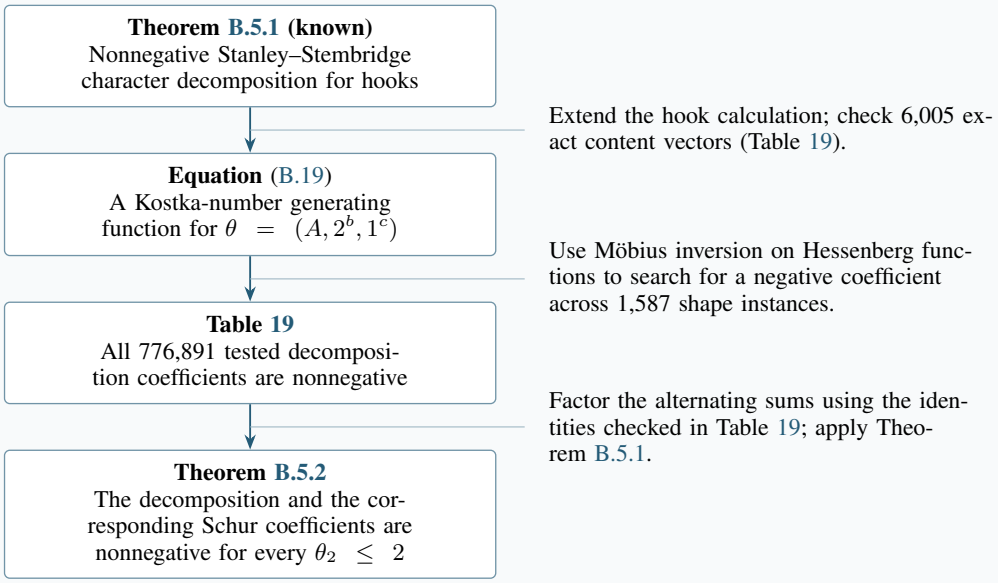


Figure 11: Exact decomposition searches leading to immanant positivity.

B.5.2 Exact decomposition experiments

The experiments enumerated Kostka numbers and tested expansions in the functions $\widehat{\eta}$. Möbius inversion on the partially ordered set of Hessenberg functions determines the coefficients a_η in

$$K_{\theta,\alpha(w)} = \sum_{\eta} a_\eta \widehat{\eta}(w).$$

A negative a_η would refute this proposed decomposition. Table 19 records the coefficient search and the checks of the identities used in the proof.

Table 19: Exact checks for the immanant decomposition.

Exact test	Cases and outcome
Content formula (B.19)	6,005 content vectors; all agree with Kostka numbers.
Path identities (B.21)	28 identities verified.
Möbius coefficients a_η	776,891 coefficients over 1,587 shape instances; all nonnegative.
Exponent in (B.22)	14,424 cases; all satisfy the stated bound.
Tableau-descent formula (B.23)	526 cases; all agree.
$D = (2^6), \theta = (8, 2, 2)$	38 positive coefficients a_η, with $\sum_\eta a_\eta = 50$.

The final row counts distinct Hessenberg indices; their associated class functions need not be distinct. The identities checked here are used in the proof for every θ with $\theta_2 \leq 2$.

B.5.3 Known positivity and its extension

Theorem B.5.1 (Lesnevich, 2024, Theorem 1.4; Hikita, 2025)**.** *If θ is a hook, then Γ_D^θ is a nonnegative integer combination of Stanley–Stembridge characters. Each Stanley–Stembridge character is a nonnegative integer combination of characters induced from the trivial representations of Young subgroups.*

The two assertions are due to Lesnevich and Hikita, respectively. The induced characters have Frobenius characteristics h_λ, which are dual to the monomial functions m_λ. Thus (B.18) is nonnegative whenever such a character decomposition exists.

PatternMiner system output (continued)

Theorem B.5.2 (LLM-generated). *Let $D = \mu/\nu$ be a nonempty skew shape and $\theta \vdash |D|$ with $\theta_2 \le 2$, taking a missing second part to be zero. Then Γ_D^θ is a nonnegative integer combination of Stanley–Stembridge characters. Consequently,*

$$[s_\theta]\,\mathrm{Imm}_{\phi^\lambda}(H_D) \ge 0 \qquad \text{for every } \lambda \vdash n.$$

If D has no empty rows, there are integers $a_\eta \ge 0$ satisfying the stronger identity $K_{\theta,\alpha(w)} = \sum_\eta a_\eta \widehat{\eta}(w)$ for every $w \in S_n$.

Proof. The column case follows from Theorem B.5.1; write the remaining partitions as $\theta = (A, 2^b, 1^c)$ with $A \ge 2$. We may assume D has no empty rows. Removing empty rows and inducing back to the original symmetric group preserves a nonnegative Stanley–Stembridge decomposition (Lesnevich, 2024, Propositions 3.1 and 3.2).

A generating function for the Kostka numbers For a nonnegative content vector α of total size $A + 2b + c$, the required Kostka number depends only on the number r of positive entries and the number o of entries equal to one. Set $U = (1+x)(1+y)$, $V = 1 + x + y$, $W = xy$. Then

$$K_{(A,2^b,1^c),\alpha} = [x^{b+c}y^b](1 - y/x)U^{r-o-1}V^o. \tag{B.19}$$

Indeed, for $\beta = (2^b, 1^c)$, Bernstein creation gives $s_{(A,\beta)} = \sum_{j\ge 0}(-1)^j h_{A+j} e_j^\perp s_\beta$. Dual Cauchy gives

$$\sum_{\gamma:\gamma_1\le 2} (e_j^\perp s_\gamma)(X) s_{\gamma'}(x,y) = h_j(x,y) \prod_i (1 + X_i x)(1 + X_i y).$$

After extracting X^α, multiplication by the appropriate complete function contributes one for each monomial bounded coordinatewise by α. The factors truncate to $1, V, U$ when α_i is zero, one, or at least two. Since $\sum_{j\ge 0}(-1)^j h_j(x,y) = U^{-1}$, the coefficient of $s_{(b+c,b)}(x,y)$ is that in $U^{r-o-1}V^o$. In two variables this is the extraction in (B.19). A negative exponent means a formal power series with invertible constant term; only finitely many degrees contribute.

Prefix maxima and Möbius inversion To express the Kostka number as a sum of Hessenberg indicator functions, group permutations by their prefix maxima. Write $p(w)_j = \max_{q\le j} w(q)$. At a nonrecord position j, an earlier record ℓ of value $m > w(j)$ gives

$$B_{w(j)} - C_j = (B_m - C_\ell) + (B_{w(j)} - B_m) + (C_\ell - C_j) \ge B_m - C_\ell + 2.$$

Thus nonnegativity is determined by the record positions and values; every nonrecord entry is at least two when the content is feasible. The prefix maxima therefore determine the numbers of zero and unit entries. Let $F(\eta)$ be the common Kostka number for permutations with $p(w) = \eta$, and set $F(\eta) = 0$ for infeasible content. Since $\widehat{\eta}(w) = \mathbf{1}\{p(w) \le \eta\}$, the desired coefficients satisfy $F(k) = \sum_{\eta\ge k} a_\eta$. The distributive lattice of Hessenberg diagrams gives the local Möbius formula

$$a_\eta = \sum_{S\subseteq \mathcal{A}(\eta)} (-1)^{|S|} F(\eta + \mathbf{1}_S), \qquad \mathcal{A}(\eta) = \{i < n : \eta_i < \eta_{i+1}\}. \tag{B.20}$$

Feasibility is $\eta_j \le H_j$, where $H_j = \max\{i : B_i \ge C_j\}$. If η is infeasible its entire upper interval is infeasible. Otherwise retain only the corners in $\mathcal{A}_{\mathrm{legal}}(\eta) = \{i \in \mathcal{A}(\eta) : \eta_i + 1 \le H_i\}$; every subset of these increases is feasible.

Factorization over paths We factor the alternating sum in (B.20) by recording how each allowed increase changes the content. Let R, O be the initial numbers of positive and unit entries, and put $b_0 = R - O$, the number of entries at least two. At a legal corner i, increasing η_i changes the selected value at i and, if the next record disappears, at that record. It either creates a zero, creates a one without creating a zero, or changes neither statistic. An increase that leaves both counts unchanged cancels the terms in (B.20) in pairs. Let z count the corners at which the increase creates a unit entry.

Only the increases that create zeros can share an affected unit entry. Each removes at most two initial unit entries: the entry at i, if it was one, and the next record entry. The latter can be one only when it is at $i+1$ and $C_i - C_{i+1} = 1$. Consequently these sets overlap only for adjacent corners, in one entry, and no entry belongs to three sets. Their overlap graph is a disjoint union of paths.

PatternMiner system output (continued)

A component with s corners has $s-1$ shared ones and $e \in \{0,1,2\}$ unshared endpoint ones. Denote by c_0, c_2 the numbers of components with $e=0$ and $e=2$, and by s_{tot} the total number of corners at which the increase creates a zero. The number of initial ones outside these sets is $f = O - s_{\mathrm{tot}} + c_0 - c_2 \geq 0$.
For a component let $\mathrm{cov}(S)$ count its initial ones removed by a selected subset S of corners, and put $h_j = h_j(x,y)$. The alternating contribution from this component is

$$M_{s,e} = \sum_{S \subseteq [s]} (-1)^{|S|} U^{s-|S|+\mathrm{cov}(S)} V^{s-1+e-\mathrm{cov}(S)}.$$

These polynomials satisfy

$$M_{s,0} = U^{s-1}(h_s + W h_{s-1}), \qquad M_{s,1} = U^s h_s,$$
$$M_{s,2} = U^s(h_s + h_{s+1}). \tag{B.21}$$

To verify them for every s, replace S by its complement T. Write $a, b \in \{0,1\}$ for the indicators for an unshared unit entry at each endpoint, so $e = a+b$. Then $M_{s,e} = (-1)^s U^{s-1+e} Z_s$, where Z_s sums $(-U)^{|T|}(V/U)^{\#\text{uncovered ones}}$. Keeping the two states for membership of the last vertex in T gives transfer matrix $\left(\begin{smallmatrix} 1 & 1 \\ -U & -V \end{smallmatrix}\right)$, initial column $(1, -U(V/U)^a)^{\mathsf{T}}$, and terminal row $(1, (V/U)^b)$. Its characteristic polynomial is $(t+x)(t+y)$. Thus $(-1)^s Z_s$ obeys the complete-function recurrence; checking $s = 1, 2$ gives the identities at every length.
Each of the z increases creating a unit entry contributes $1 - V/U = W/U$. Combining components factors the generating function in the alternating sum as

$$W^z U^{b_0-1-z-c_0} V^f \prod_{e=0} (h_s + W h_{s-1}) \prod_{e=1} h_s \prod_{e=2} (h_s + h_{s+1}), \tag{B.22}$$

where each factor uses its own component size. All factors are Schur-positive in two variables if $b_0 - 1 - z - c_0 \geq 0$: $U = 1 + s_{(1)} + s_{(1,1)}$, $V = 1 + s_{(1)}$, $W = s_{(1,1)}$, and products preserve Schur positivity.

Nonnegativity of the remaining exponent It remains to prove that the exponent of U in (B.22) is nonnegative when $b_0 > 0$. Each corner that creates a one, and the first corner of each component with $e = 0$, uses a distinct initial entry at least two. The first corner of a component with $e = 0$ cannot have its own initial one, since it would be unshared. Hence $b_0 \geq z + c_0$.
One further entry is unused when $b_0 > 0$. Choose the rightmost position j whose initial entry is at least two in a representative permutation. Every later position is a record, since every nonrecord has content at least two. Backward induction from the last record gives $w(\ell) = \eta_\ell = \ell$ for $\ell > j$. If j is not a legal corner, its entry is unused in this assignment. Otherwise $\eta_j = j$, and the next initial entry $d = B_{j+1} - C_{j+1}$ is zero or one. Feasibility of the raise forces $d = 1$, $C_j - C_{j+1} = 1$, and raised entry zero. Thus j removes the following one and does not create a one. If this unit entry is shared with the next corner, repeat the argument at that corner. The chain terminates at an unshared one because the last position has no corner. Its component therefore has $e \geq 1$, so j is not the assigned first corner of a component with $e = 0$ either. This supplies one unused entry at least two and proves $b_0 \geq z + c_0 + 1$. Equation (B.22) is consequently Schur-positive whenever $b_0 > 0$.

The case of unit content When $b_0 = 0$, the coefficient has a direct description by standard tableaux. All content entries are then zero or one. Since there are n nonempty rows, $N \geq n$, whereas $N = R = O \leq n$. Hence $N = n$, all content entries are one, and all positions are records: $\eta = w = \mathrm{id}$ and $B_i = C_i + 1$. A legal corner i has $C_i - C_{i+1} = 1$. Raising a subset S merges consecutive chosen boundaries into blocks; the forward cyclic representative of each block has zero entries followed by its block length. After removing zeros the content is the composition with precisely those internal block boundaries S.
Write $\mathrm{Des}(T) = \{i : i+1 \text{ lies in a lower row of } T \text{ than } i\}$. Standardization identifies the Kostka count for this composition with standard tableaux whose descent set avoids S. Inclusion–exclusion gives

$$a_{\mathrm{id}} = \#\{T \in \mathrm{SYT}(\theta) : \mathcal{A}_{\mathrm{legal}}(\mathrm{id}) \subseteq \mathrm{Des}(T)\} \geq 0. \tag{B.23}$$

Together with (B.22), this proves $a_\eta \geq 0$ for every Hessenberg function η. Averaging over conjugacy classes proves the character decomposition. The nonnegativity of the immanant coefficient follows from (B.18) and Theorem B.5.1. □

B.6 Recovered results and generated conjectures

B.6.1 Recovering known statements

The descent and essential-rank conditions in Propositions B.1.1 and B.1.2 were formulated during the Schubert investigations and identified as consequences of established theory. Similarly, the Kronecker experiments recovered Murnaghan's inequalities and Rosas's bounds. The semigroup inequality was instead supplied as a known relation against which to test predictions. These are different uses of prior mathematics.

Hook row sums The row-sum investigation reconstructed a positive formula for hook partitions. For a cycle type ρ, let Z_ρ be the centralizer of a permutation of that type and let Std_n be the standard representation. The conjugation interpretation gives

$$R((n-r,1^r)) = \sum_{\rho \vdash n} \dim\Big(\bigwedge^r \mathrm{Std}_n\Big)^{Z_\rho}.$$

On each cycle, a Fourier basis identifies invariant exterior monomials by subsets whose indices sum to zero modulo the cycle length. Taking invariants under permutations of equal-length cycles and removing the global trivial summand gives the positive counting formula. The implementation agrees with all 55 hook row sums through $n = 10$ and with direct centralizer calculations for all 240 hook multiplicities over the 44 cycle types through $n = 7$. This reconstructs known hook formulas; the classical representation-theoretic background is described by Sundaram (2018). These checks are separate from the fixed-tail evaluation theorem and the learned tableau map.

Invariance through length twelve The Kazhdan–Lusztig investigation also derived the bound now stated in Theorem B.4.1. The low-degree congruence modulo z^8, parity, monicity, and invariance of $[z^{d-2}]\widetilde{R}$ determine all coefficients through interval length eleven. The last invariance follows from the invariant linear coefficient of P and the coatom count, as in the proof of Proposition B.4.2. For an even interval length $d > 0$, the standard inversion identity gives

$$2R_{x,w} = -\sum_{x<z<w} (-1)^{\ell(z)-\ell(x)} R_{x,z} R_{z,w}.$$

Every factor on the right belongs to a proper subinterval. Hence invariance through length eleven implies invariance of R at length twelve; Kazhdan–Lusztig inversion and the degree bound then give invariance of P. This rederives the established invariance result. The length-thirteen restrictions in Theorem B.4.5 are separate deductions.

B.6.2 Two remaining Kazhdan–Lusztig conjectures

In addition to the corrected-descent conjecture, the investigation proposed a stronger coefficient bound. Use the uncompressibility and connected transposition graph defined in Appendix B.4.3. If the interval has m positions, odd length d, and c cycles in $x^{-1}w$, including fixed points, put

$$G(x,w) = \frac{d-m-c+2}{2}.$$

The transposition argument in (B.16) gives $G \geq 0$.

Conjecture B.6.1. *For every uncompressible type-A interval $[x,w]$ with connected transposition graph and odd length d, if $G(x,w) = 0$, then*

$$[q^{(d-1)/2}]P_{x,w}(q) \leq 1.$$

This concerns the coefficient at the largest permitted degree, which may be zero even when the polynomial has a nonzero coefficient in a smaller degree. The hypothetical obstruction in Theorem B.4.5 has $(m,d,c) = (12,13,3)$, so this conjecture would exclude it. Conjecture B.4.7 instead bounds a corrected difference on that specific length-thirteen class. Both remain open. The even-length coefficient-eight example in Table 18 refutes the broader bound that omits recurrence corrections. Bounded examples and the standard small-rank zero–one results do not establish the new conjecture.

B.6.3 From the first non-hook immanant case to width two

For a feasible content vector α of total size $N \geq 4$, let r count its positive entries and let t count entries at least two. The first non-hook calculation gives

$$K_{(N-2,2),\alpha} = \binom{r-1}{2} + t - 1.$$

The binomial term is the known hook weight for $(N-2,1,1)$. Exact searches suggested that the complete weight admits a nonnegative pointwise expansion in the indicators $\widehat{\eta}$. A stronger attempted shortcut, requiring the correction $\delta = t - 1$ to have such an expansion by itself, fails for the skew shape $(3,3,3,1)/(1,1)$. With $\delta = 0$ on infeasible contents, its unique expansion is

$$\delta = \widehat{(1,3,4,4)} + \widehat{(2,2,4,4)} - \widehat{(2,3,4,4)} + \widehat{(3,3,4,4)}.$$

Here $\widehat{(\eta_1,\ldots,\eta_4)}$ denotes the indicator associated with the displayed Hessenberg function, using one-based indices. The negative coefficient led to a revised cancellation conjecture: each negative correction coefficient is -1 and occurs at an index where the hook contribution is at least one. A local analysis of record positions proved this cancellation, establishing the $(N-2,2)$ case.

Subsequent extensions to $(A,2,1^c)$ suggested the general statement: for every skew shape and every coefficient partition θ with $\theta_2 \leq 2$, the full Kostka weight has a nonnegative pointwise expansion in Hessenberg indicators. This generated conjecture is now Theorem B.5.2, proved by the content series, path factorization and boundary argument in Appendix B.5. It allows arbitrarily many parts equal to two. Lesnevich's hook theorem was the starting input, and positivity of the Stanley–Stembridge summands converts the decomposition into Schur positivity. The conjecture-to-proof trajectory used exact computation and symbolic reasoning; the separate classifier below addressed another question.

B.7 Additional cross-problem analysis

Schubert predicate extraction without a model A retrospective comparison repeated the historical predicate searches on the recorded data using 37 curated features, without model predictions. It recovered the descent condition and the two essential-rank predicates. Among single-feature candidates, the both-tight essential-rank condition ranked first by the number of excluded zeros. Validation on 83,689 separate positive examples removed the spurious candidate regions. This comparison conditions on features developed during the earlier investigation; it does not measure discovery from an independently specified initial representation. The random-feature comparison in Table 4 likewise uses activations of the trained model for both representations.

Later row-sum matching experiments Appendix B.3.1 reports the fits on all 50 negative objects and the coordinated matching constructions found in the subsequent task audit.

Classifying enumerated skew-shape targets Each example is a vector of unscaled sums of Kostka weights over conjugacy classes of S_6, with identity coordinate 24. The two inputs are its coordinates for cycle types $(2,1^4)$ and $(5,1)$. Positive examples come from the enumerated skew shapes; negative examples are constructed vectors absent from that collection. All examples have positive five-cycle coordinate. The labels concern membership in this collection. The stump fits 550/550 training examples and scores 279/279 on an ordinary holdout containing 268 vectors from enumerated shapes and 11 constructed negative vectors. It scores 0/15 on a mined challenge containing only constructed negative vectors. After inspection of that challenge, adding 12 training examples from different parents and fitting a depth-two tree gives 562/562 on training, 278/279 on the ordinary holdout, and 14/15 on the reused challenge. Both models fail the recorded zero-error criterion. The changed training data and depth are not a controlled architecture comparison, and this study is separate from the proof of Theorem B.5.2.

B.8 Small-model analyses for Problems 16 and 18

Both investigations combined LLM-guided mathematical arguments with small-model experiments. The plots below show how fitted rules, hard negatives, and branch tests informed the model analyses. The recorded data and plotting code are available in the Problem 16 figures and Problem 18 figures directories.

Automatically generated by the PatternMiner system

Kazhdan–Lusztig recurrence terms The two fitted stumps in Table 17 proposed conditions for a recurrence term to be positive. One used the number of atoms of the associated Bruhat interval; the other tested whether the cover transposition exchanges adjacent positions or adjacent values. The withheld pair has the same values of these two features and different exact labels, refuting both conditions. Figure 12 compares their predictions with the training-majority baseline.

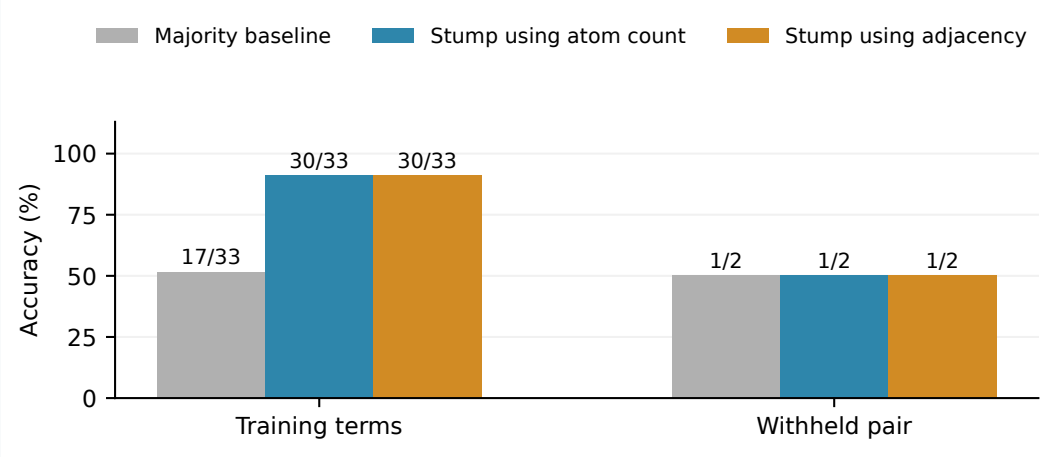


Figure 12: **Fitted stumps and the withheld pair.** Both stumps classify 30 of 33 training terms correctly, compared with 17 for the majority baseline. Each predicts one of the two withheld labels correctly. The pair was selected to test the learned conditions, so its accuracy does not estimate population performance.

Character data from skew shapes The classifier in Appendix B.7 tested membership in an enumerated collection of character-data vectors. Hard-negative mining found 15 constructed vectors that the initial stump classified as members. Adding 12 training examples and fitting a depth-two tree corrected 14 of these errors (Figure 13). The same 15 vectors were reused after inspection. These experiments studied the learned membership rule; Theorem B.5.2 follows from the exact decomposition calculations and symbolic argument in Appendix B.5.

PatternMiner system output (continued)

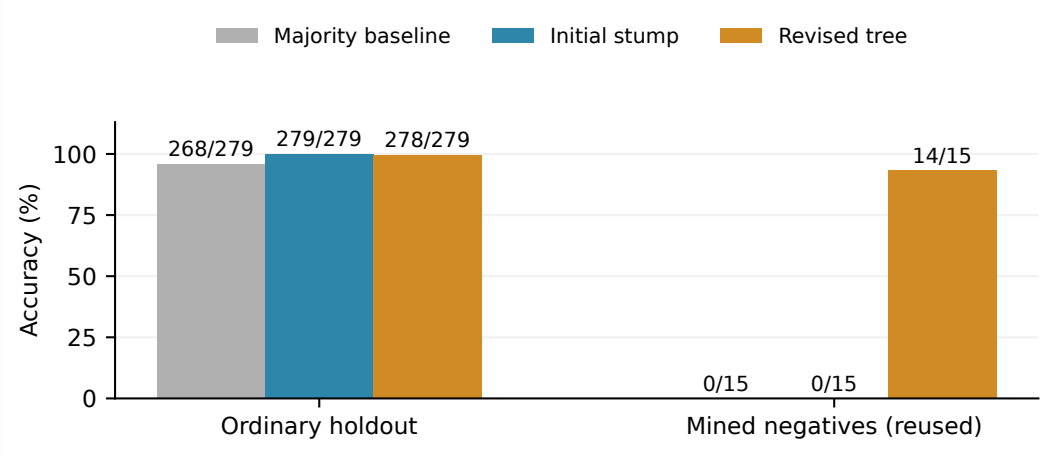


Figure 13: **Hard negatives expose errors in shape-membership predictions.** The ordinary holdout contains 268 vectors from enumerated shapes and 11 constructed negatives. The majority baseline always predicts membership. Training data and tree depth both changed in the revision.

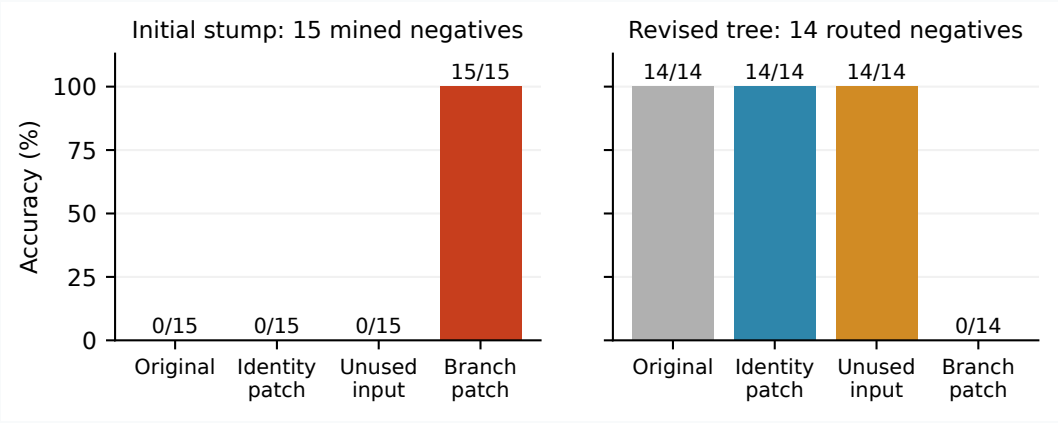


Figure 14: **Branch tests identify which decisions control the predictions.** Replacing the stump's branch decision with that from a negative training example corrects all 15 mined cases. Replacing the revised tree's child decision with that from a positive training example makes all 14 cases reaching that node incorrect. Identity patches and changes to unused inputs preserve the original predictions. The two panels use different donor classes and sample counts.