Can Transformers Do Enumerative Geometry?

Baran Hashemi[1] R. Corominas[2] A. Giacchetto[3]

[1]ORIGINS Cluster, Munich, Germany
[2]Harvard University, Cambridge, USA
[3]ETH Zurich, Zurich, Switzerland

arXiv: 2408.14915

ICLR 2025

Motivation

- Imagine that you have a problem in Enumerative geometry, e.g., count curves over a finite space.

Motivation

- Imagine that you have a problem in Enumerative geometry, e.g., count curves over a finite space.
- This computation is intractable

Motivation

- Imagine that you have a problem in Enumerative geometry, e.g., count curves over a finite space.
- This computation is intractable
- Given one finds a generating function, the computation is recursive with exponential or factorial blow up

d	n_d
1	2875
2	609250
3	317206375
4	242467530000
5	229305888887625
6	248249742118022000
7	295091050570845669250
8	375632160937476603550000
9	503804510141698524364510625O
10	704288164978454686113488294750

Number of rational curves of degree d that lie on a 4-dimensional complex projective space.

Motivation

- Imagine that you have a problem in Enumerative geometry, e.g., count curves over a finite space.
- This computation is intractable
- Given one finds a generating function, the computation is recursive with exponential or factorial blow up
- Is it possible to surrogate this computation or a **Neural Enumerative Reasoning** model?

Motivation

- Imagine that you have a problem in Enumerative geometry, e.g., count curves over a finite space.
- This computation is intractable
- Given one finds a generating function, the computation is recursive with exponential or factorial blow up
- Is it possible to surrogate this computation or a **Neural Enumerative Reasoning (NER)** model?
- IF so, what can we learn from the what the NER network learned? Is it possible to extract knowledge from mathematical "World Model" of the network?

"Experimental Mathematics"

Moduli Spaces of Algebraic Curves

Let g, n be non-negative integers satisfying the **stability condition**

$$2g - 2 + n > 0.$$

Moduli Spaces of Algebraic Curves

Let g, n be non-negative integers satisfying the **stability condition**

$$2g - 2 + n > 0.$$

Let's denote by $\overline{\mathcal{M}}_{g,n}$ the **Deligne–Mumford moduli space** of stable algebraic curves of genus g with n distinct marked points. This is a complex, compact, smooth orbifold of dimension

$$\dim \overline{\mathcal{M}}_{g,n} = 3g - 3 + n.$$

ψ Classes

One can associate with each marked point a **cotangent line bundle** $\mathcal{L}_j$ on $\overline{\mathcal{M}}_{g,n}$ for $j = 1, \ldots, n$. The ψ**-class**, ψ_j, is defined as the first Chern class of $\mathcal{L}_j$:

$$\psi_j = c_1(\mathcal{L}_j).$$

Intersection of ψ Classes

One can associate with each marked point a **cotangent line bundle** $\mathcal{L}_j$ on $\overline{\mathcal{M}}_{g,n}$ for $j = 1, \ldots, n$. The ψ**-class**, ψ_j, is defined as the first Chern class of $\mathcal{L}_j$:

$$\psi_j = c_1(\mathcal{L}_j).$$

Consider the integrals of products of ψ-classes over $\overline{\mathcal{M}}_{g,n}$:

$$\langle \tau_{d_1} \cdots \tau_{d_n} \rangle_{g,n} := \int_{\overline{\mathcal{M}}_{g,n}} \psi_1^{d_1} \cdots \psi_n^{d_n} \in \mathbb{Q}, \quad \text{where} \quad \sum_n d_n = 3g - 3 + n.$$

Witten's Analysis

- Witten '91 found a theory of two-dimensional topological quantum gravity, where the **intersection numbers** in this context are interpreted as correlators of the theory.
- He conjectured that the **exponential generating function**

$$Z(\mathbf{x};\hbar) = \exp\left(\sum_{\substack{g\geq 0\\ n>0}} \frac{\hbar^{2g-2+n}}{n!} \sum_{\substack{d_1,\ldots,d_n\geq 0\\ d_1+\cdots+d_n=3g-3+n}} \left\langle \tau_{d_1}\cdots\tau_{d_n}\right\rangle_{g,n} x^{d_1}\cdots x^{d_n} \right)$$

satisfies an infinite tower of quadratic partial differential equations known as the **Korteweg–de Vries (KdV) hierarchy**.

Kontsevich's Proof

- Witten '91 found a theory of two-dimensional topological quantum gravity, where the **intersection numbers** in this context are interpreted as correlators of the theory.
- He conjectured that the **exponential generating function**

$$Z(\mathbf{x};\hbar) = \exp\left(\sum_{\substack{g\geq 0\\ n>0}} \frac{\hbar^{2g-2+n}}{n!} \sum_{\substack{d_1,\ldots,d_n\geq 0\\ d_1+\cdots+d_n=3g-3+n}} \left\langle \tau_{d_1}\cdots\tau_{d_n}\right\rangle_{g,n} x^{d_1}\cdots x^{d_n}\right)$$

  satisfies an infinite tower of quadratic partial differential equations known as the **Korteweg–de Vries (KdV) hierarchy**.
- Kontsevich '92 provided a rigorous mathematical proof of Witten's conjecture by relating intersection numbers on $\overline{\mathcal{M}}_{g,n}$ to the combinatorics of ribbon graphs and matrix models.

Quantum Airy Structures

A **Quantum Airy Structure** (Kontsevich et al. '17) on a complex vector space V, dually generated by $(x^i)_{i\in I}$, is a family of differential operators $(L_i)_{i\in I}$ in the Weyl algebra over V of the form

$$L_i = \hbar\partial_i - \hbar^2 \sum_{a,b\in I} \left(\frac{1}{2} A_{i,a,b}\, x^a x^b + B^b_{i,a}\, x^a \partial_b + \frac{1}{2} C^{a,b}_i\, \partial_a \partial_b \right) - \hbar^2 D_i\,, \tag{1}$$

where:

- $\partial_i = \partial_{x^i}$
- $A_{i,a,b} = A_{i,b,a}$
- $C^{a,b}_i = C^{b,a}_i$
- $B^b_{i,a}$, $C^{a,b}_i$, and D_i are scalars indexed by elements in I

Existence and Uniqueness of the Partition Function

Theorem (Kontsevich–Soibelman '17)

If $(L_i)_{i \in I}$ is a quantum Airy structure on V, there exists a unique formal function Z of the form

$$Z(\boldsymbol{x};\hbar) = \exp\left(\sum_{\substack{g \geq 0 \\ n > 0}} \frac{\hbar^{2g-2+n}}{n!} \sum_{d_1,\ldots,d_n \in I} F_{g;d_1,\ldots,d_n} \, x^{d_1} \cdots x^{d_n} \right)$$

with $F_{0;d_1} = F_{0;d_1,d_2} = 0$, $F_{g;d_1,\ldots,d_n}$ symmetric in $d_1,\ldots,d_n \in I$, and such that

$$L_i Z = 0 \qquad \forall i \in I\,.$$

The coefficients $F_{g;d_1,\ldots,d_n}$, called *quantum Airy structure amplitudes*, are determined recursively based on the initial data (A, B, C, D).

Topological Recursion

For general values of g and n satisfying $2g-2+n>0$, the recursion formula satisfies *Topological Recursion*, is given by

$$F_{g;d_1,d_2,\ldots,d_n} = \sum_{m=2}^{n}\sum_{a\in I} B^{a}_{d_1,d_m} F_{g;a,d_2,\ldots,\widehat{d_m},\ldots,d_n}$$

$$+\frac{1}{2}\sum_{a,b\in I} C^{a,b}_{d_1}\left(F_{g-1;a,b,d_2,\ldots,d_n} + \sum_{\substack{g_1+g_2=g\\ I_1\sqcup I_2=\{d_2,\ldots,d_n\}}} F_{g_1;a,I_1}\,F_{g_2;b,I_2}\right).$$

Topological Recursion and ψ-Class Intersections

Given that the differential operators $(L_i)_{i\geq 0}$ are determined by the tensors:

$$A_{i,j,k} := \delta_{i,j,k}\,, \qquad B^k_{i,j} := \delta_{i+j,k+1}\,\frac{(2k+1)!!}{(2i+1)!!(2j-1)!!}\,,$$

$$C^{j,k}_i := \delta_{i,j+k+2}\,\frac{(2j+1)!!(2k+1)!!}{(2i+1)!!}\,, \qquad D_i := \frac{\delta_{i,1}}{24}\,.$$

the associated amplitudes coincide with ψ-class intersection numbers:

$$F_{g;d_1,\ldots,d_n} = \begin{cases} \langle \tau_{d_1}\cdots\tau_{d_n}\rangle_{g,n} & \text{if } d_1+\cdots+d_n = 3g-3+n\,, \\ 0 & \text{otherwise.} \end{cases}$$

Computability: Challenges and Ideas

- **Challenges**
  - Computation Complexity of of $\mathcal{O}(g! \cdot 2^n)$.

Computability: Challenges and Ideas

- **Challenges**
  - Computation Complexity of of $\mathcal{O}(g! \cdot 2^n)$.
  - Extremely challenging to find asymptotic closed-form expression $\rightarrow$ Only recently for large genus limit, where $n = \mathcal{O}(\sqrt{g})$ (Aggarwal '20). (**Large n limits is still open problem.**)

Computability: Challenges and Ideas

- **Challenges**
  - Computation Complexity of of $\mathcal{O}(g! \cdot 2^n)$.
  - Extremely challenging to find asymptotic closed-form expression → Only recently for large genus limit, where $n = \mathcal{O}(\sqrt{g})$ (Aggarwal '20). (**Large n limits is still open problem.**)
  - Very hard to find new identities to make the computation more feasible

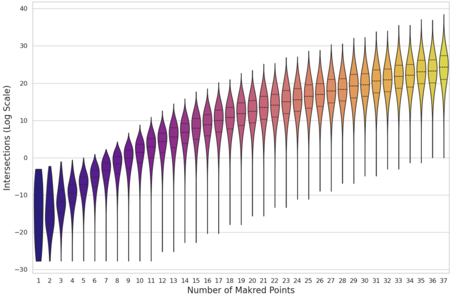
Computability: Challenges and Ideas

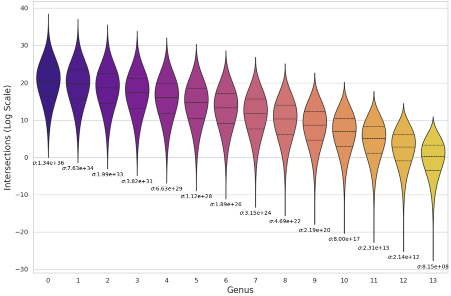
- **Challenges**
  - Computation Complexity of of $\mathcal{O}(g! \cdot 2^n)$.
  - Extremely challenging to find asymptotic closed-form expression $\rightarrow$ Only recently for large genus limit, where $n = \mathcal{O}(\sqrt{g})$ (Aggarwal '20). (**Large n limits is still open problem.**)
  - Very hard to find new identities.
- **Simple Idea:**
  - Let's surrogate the computation, given Quantum Airy structure initial data $\rightarrow$ aligned with AI-Mathematician collaboration paradigm (Davies et al. '21).
  - Oh, but one can only wish this were simple!

Challenges: Heteroscedasticity

- Intersection numbers exhibit extreme high-variance → **Heteroscedasticity**

- *Heteroscedasticity* occurs when the standard deviation of a predicted variable varies across different values of an independent variable





Challenges: Recursive Behavior with Factorial Blow-Up

- **Multi-Dimensional Recursion:**
  - Extends beyond a single linear progression and spans across multiple dimensions, (g, n).

Challenges: Recursive Behavior with Factorial Blow-Up

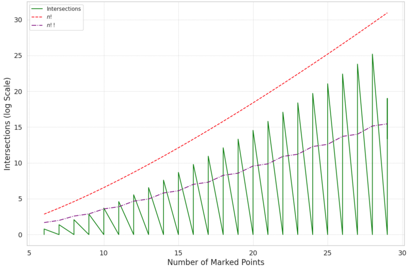
- **Multi-Dimensional Recursion:**
  - Extends beyond a single linear progression and spans across multiple dimensions, (g, n).
- **No Fixed Order:**
  - Unlike typical recursions with a fixed number of preceding terms, e.g Fibonacci sequence:

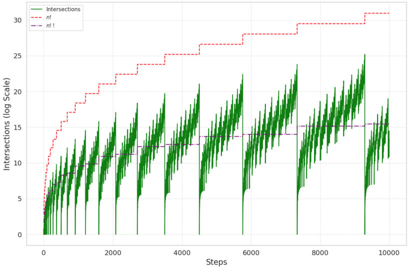
$$a_n = a_{n-1} + a_{n-2}$$

  - There is not fixed recurrence order. To compute a point at (g, n), one needs information from multiple points such as: $(g-1, n+1)$, $(g_1, n_1) \wedge (n_2, n_2)$ s.t $g_1 + g_2 = g$ and $n_1 + n_2 = n$

Challenges: How do they Look like?

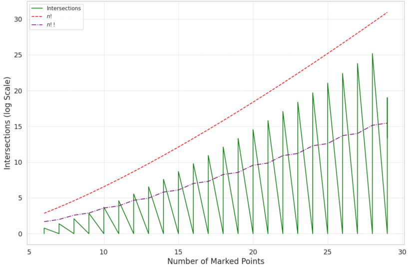
- Multi-Dimensional Recursive behavior with factorial blow-up

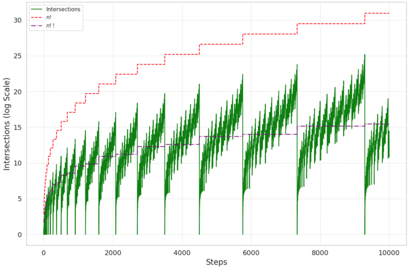




Challenges: How do they Look like?

- (Multi-Dimensional) Recursive behavior with factorial blow-up



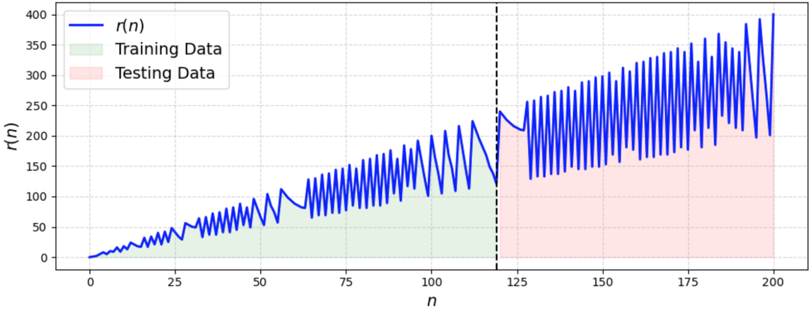


- **But**, how does ML perform against recursive function?

Simple Example

Recursive function $r(n)$:

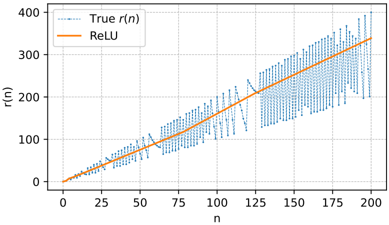
$$r(n) = n + (n \,\&\, r(n-1))$$

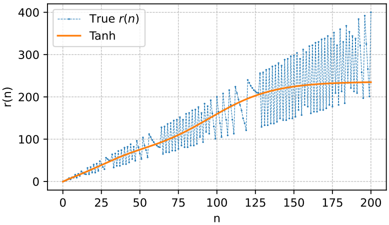


Simple Example

Recursive function $r(n)$:

$$r(n) = n + (n \,\&\, r(n-1))$$

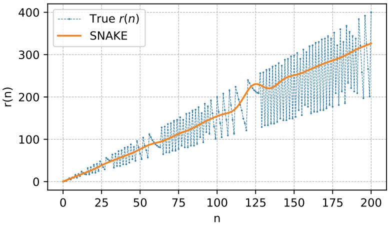


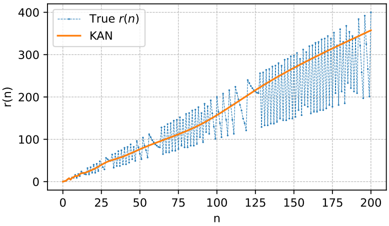


Simple Example

Recursive function $r(n)$:

$$r(n) = n + (n \,\&\, r(n-1))$$





Dynamic Range Activator $\mathfrak{DRA}$

- Activation function is important for capturing different functional behaviors.

Dynamic Range Activator $\mathfrak{DRA}$

- Activation function is important for capturing different functional behaviors.
- Current non-linear activations fail miserably on (even first-order) recursive patterns.

Dynamic Range Activator $\mathfrak{DRA}$

- We introduced *Dynamic Range Activator* ($\mathfrak{DRA}$)

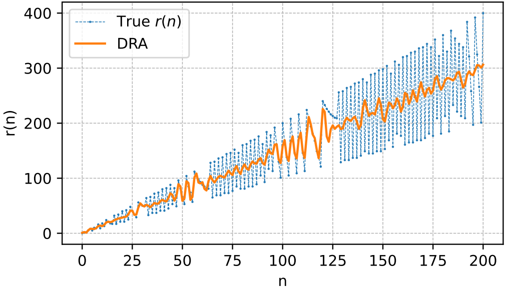
$$\mathfrak{DRA}(x) := x \; + \; b \sin^2(ax) \; + \; c \cos(ax) \; + \; d \tanh(bx)$$

- Learnable parameters $(a, b, c, d) \in \mathbb{R}$ to shape the activation's curvature, frequency, and amplitude.
- The harmonic part quasi-stable cycles and hyperbolic part leads to saturations or stable attractors.

Dynamic Range Activator $\mathfrak{DRA}$

- We introduced *Dynamic Range Activator* ($\mathfrak{DRA}$)

$$\mathfrak{DRA}(x) := x \;+\; b\sin^2(ax) \;+\; c\cos(ax) \;+\; d\tanh(bx)$$

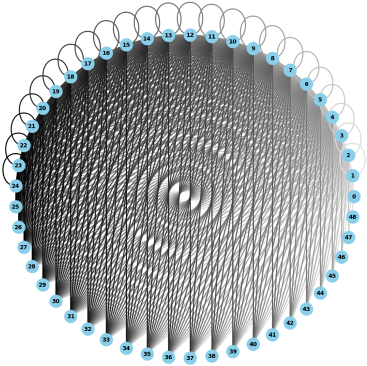


- There is hope to capture the recursive ψ-class intersection...

Let's be One with the Data

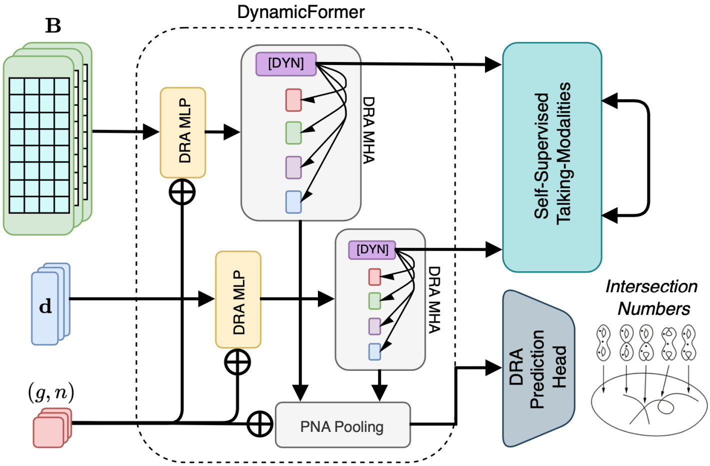
```
g,n,d,Fgn
0,6,"(1, 1, 1, 0, 0, 0)",6.0
0,6,"(2, 1, 0, 0, 0, 0)",3.0
0,6,"(3, 0, 0, 0, 0, 0)",1.0
0,7,"(1, 1, 1, 1, 0, 0, 0)",24.0
0,7,"(2, 1, 1, 0, 0, 0, 0)",12.0
0,7,"(2, 2, 0, 0, 0, 0, 0)",6.0
0,7,"(3, 1, 0, 0, 0, 0, 0)",4.0
0,7,"(4, 0, 0, 0, 0, 0, 0)",1.0
0,8,"(1, 1, 1, 1, 1, 0, 0, 0)",120.0
0,8,"(2, 1, 1, 1, 0, 0, 0, 0)",60.0
0,8,"(2, 2, 1, 0, 0, 0, 0, 0)",30.0
0,8,"(3, 1, 1, 0, 0, 0, 0, 0)",20.0
0,8,"(3, 2, 0, 0, 0, 0, 0, 0)",10.0
0,8,"(4, 1, 0, 0, 0, 0, 0, 0)",5.0
0,8,"(5, 0, 0, 0, 0, 0, 0, 0)",1.0
0,9,"(1, 1, 1, 1, 1, 1, 0, 0, 0)",720.0
0,9,"(2, 1, 1, 1, 1, 0, 0, 0, 0)",360.0
0,9,"(2, 2, 1, 1, 0, 0, 0, 0, 0)",180.0
0,9,"(2, 2, 2, 0, 0, 0, 0, 0, 0)",90.0
0,9,"(3, 1, 1, 1, 0, 0, 0, 0, 0)",120.0
0,9,"(3, 2, 1, 0, 0, 0, 0, 0, 0)",60.0
0,9,"(3, 3, 0, 0, 0, 0, 0, 0, 0)",20.0
0,9,"(4, 1, 1, 0, 0, 0, 0, 0, 0)",30.0
0,9,"(4, 2, 0, 0, 0, 0, 0, 0, 0)",15.0
0,9,"(5, 1, 0, 0, 0, 0, 0, 0, 0)",6.0
0,9,"(6, 0, 0, 0, 0, 0, 0, 0, 0)",1.0
0,10,"(1, 1, 1, 1, 1, 1, 1, 0, 0, 0)",5040.0
0,10,"(2, 1, 1, 1, 1, 1, 0, 0, 0, 0)",2520.0
0,10,"(2, 2, 1, 1, 1, 0, 0, 0, 0, 0)",1260.0
```

- **g**: genus (Integer)
- **n**: number of marked points (Integer)
- **d**: Partitions (Vector of length n)
- **B**: Graph in COO format of variable size

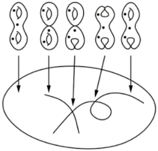


DynamicFormer

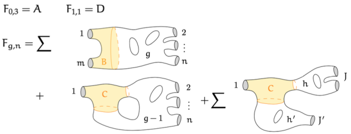
- Multi-Modal Transformer model with $\mathfrak{DRA}$ activation function



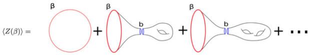
Bird's Eye View



Compute Geometric Objects (Amplitudes/ Intersections)

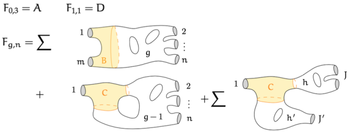


Analytical Methods (recursive)

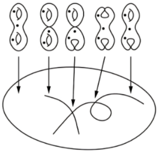


Large Genus Asymptotic Open Question?

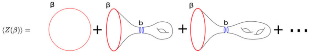
Bird’s Eye View



Analytical Methods (recursive)



Compute Geometric Objects (Amplitudes/ Intersections)



Large Genus Asymptotic Open Question?

$F_{g,n} \approx f_\theta(g, n; ABCD)$

Results

- **In Distribution (ID):** Same **g** as training, but higher **n**

Results

- **In Distribution (ID):** Same $\mathbf{g}$ as training, but higher $\mathbf{n}$
- **Out-of-Distribution (OOD):** higher $\mathbf{g}$ than training.
- Uncertainty quantification with *Conformal Procedure.*

Results

- **In Distribution (ID):** Same **g** as training, but higher **n**
- **Out-of-Distribution (OOD):** higher **g** than training.
- Uncertainty quantification with *Conformal Procedure.*

(g, n)	$R^2 \uparrow$	Coverage	CW
$(1, 35)$	99.8	90.35	1.03
$(2, 33)$	99.6	83.60	0.84
$(3, 31)$	99.9	74.79	0.76
$(4, 29)$	98.7	95.66	1.11
$(5, 27)$	99.1	92.66	1.03
$(6, 25)$	99.3	91.18	0.80
$(7, 23)$	99.1	93.05	0.68
$(8, 21)$	99.8	90.88	0.76
$(9, 19)$	99.9	96.71	0.91
$(10, 17)$	99.9	90.01	1.04
$(11, 15)$	99.8	91.97	0.87
$(12, 13)$	99.6	89.08	1.30
$(13, 11)$	99.9	95.90	0.97

(a) R^2 and conformal uncertainty estimation results with $\alpha = 0.1$ (90% target coverage) in the ID setting.

(g, n)	$R^2 \uparrow$	Coverage	CW
$(14, [1, 9])$	99.6	93.82	0.93
$(15, [1, 7])$	95.9	84.27	0.91
$(16, [1, 5])$	94.1	89.60	3.55
$(17, [1, 3])$	93.8	95.27	8.30

(b) R^2 and conformal uncertainty estimation results with $\alpha = 0.1$ (90% target coverage) in the OOD setting.

Results

- **In Distribution (ID):** Same **g** as training, but higher **n**
- **Out-of-Distribution (OOD):** higher **g** than training.
- Uncertainty quantification with *Conformal Procedure*.

ReLU		GLU		Snake		DRA	
$R^2\uparrow$	CW $\downarrow$	$R^2\uparrow$	CW $\downarrow$	$R^2\uparrow$	CW $\downarrow$	$R^2\uparrow$	CW $\downarrow$
71.5	9.73	74.7	8.34	82.9	6.55	**95.8**	**3.42**

How does the Model do Enumerative Geometry?

- Do Transformers actually understand the underlying enumerative geometry?

How does the Model do Enumerative Geometry?

- Do Transformers actually understand the underlying enumerative geometry?
- Is it possible to extract useful knowledge or hints toward a possible discovery of a theorem?

How does the Model do Enumerative Geometry?

- Do Transformers actually understand the underlying enumerative geometry?
- Is it possible to extract useful knowledge or hints toward a possible discovery of a theorem?
- Interpretability Approaches
  1. Self-Similarity

How does the Model do Enumerative Geometry?

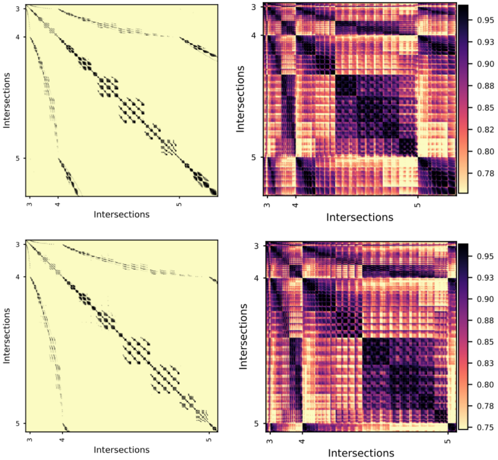
- Do Transformers actually understand the underlying enumerative geometry?
- Is it possible to extract useful knowledge or hints toward a possible discovery of a theorem?
- Interpretability Approaches
  1. Self-Similarity
  2. Causal Tracing

How does the Model do Enumerative Geometry?

- Do Transformers actually understand the underlying enumerative geometry?
- Is it possible to extract useful knowledge or hints toward a possible discovery of a theorem?
- Interpretability Approaches
    1. Self-Similarity
    2. Causal Tracing
    3. Abductive Hypothesis Testing

Self-Similarity

- Internal representation of the last layer before the prediction head. What is this pattern?

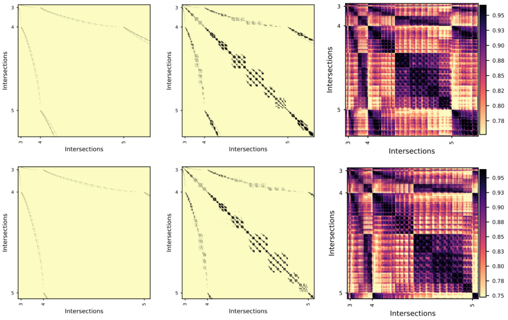


Self-Similarity

- **Dilaton Equations:** the intersection numbers involving a τ_1 term can be reduced to a simpler intersection number with one fewer marked point.

$$\langle \tau_1, \tau_{d_1}, \ldots, \tau_{d_n} \rangle_{g,n+1} = (2g - 2 + n) \langle \tau_{d_1}, \ldots, \tau_{d_n} \rangle_{g,n} .$$

- τ_1 represents the first power of the ψ-class at a marked point.

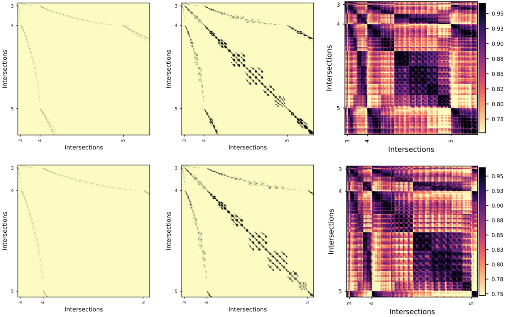


Self-Similarity

- **Dilaton Equations:** the intersection numbers involving a τ_1 term can be reduced to a simpler intersection number with one fewer marked point.

$$\langle \tau_1, \tau_{d_1}, \ldots, \tau_{d_n} \rangle_{g,n+1} = (2g - 2 + n) \langle \tau_{d_1}, \ldots, \tau_{d_n} \rangle_{g,n} \,.$$

- New Identities: decreased the search space from $\mathcal{O}(10^7)$ to $\mathcal{O}(10^3)$!



Causal Tracing

- Let's understand the model's decision-making process

Causal Tracing

- Specifically identifying which input modalities are more influential when predicting the intersections → we want to uncover how the network causally comprehends the underlying mathematics.

Causal Tracing

- Let the input modalities be $m[i] = \{n[i], B[i], d[i]\}$ for each input instance i

Causal Tracing

- Let the input modalities be $m[i] = \{n[i], B[i], d[i]\}$ for each input instance i
- There are then two steps:
  1. **Clean run.** This run records the model's prediction on a regular input. With this run, we record auto-correlation of the predicted intersection numbers across different number of marked points.

Causal Tracing

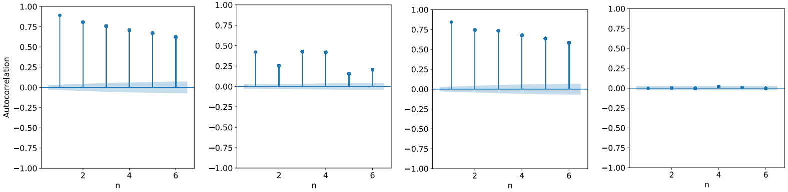
- Let the input modalities be $m[i] = \{n[i], B[i], d[i]\}$ for each input instance i
- There are then two steps:
  1. **Clean run.** This run records the model's prediction on a regular input. With this run, we record auto-correlation of the predicted intersection numbers across different number of marked points.
  2. **Counter-factual intervention.** We perform interventions by modifying instances in one modality while keeping others unchanged to observe how these changes affect the model's predictions, e.g., associating the partition $d = (29, 10, 5, 3, 1, 0)$ with $n = 3$ rather than of $n = 6$.

Causal Tracing

- Let the input modalities be $m[i] = \{n[i], B[i], d[i]\}$ for each input instance i
- There are then two steps:
  1. **Clean run.** This run records the model's prediction on a regular input. With this run, we record auto-correlation of the predicted intersection numbers across different number of marked points.
  2. **Counter-factual intervention.** We perform interventions by modifying instances in one modality while keeping others unchanged to observe how these changes affect the model's predictions.
- This replacement results in a miss-assignment of features to a sample, which alters the target prediction, *i.e.*, $m[i] \to m[j]$ where $i \neq j$, to obtain $F_{g,n}(\langle d \rangle_{g,n} | \mathrm{do}(m[i]))$.

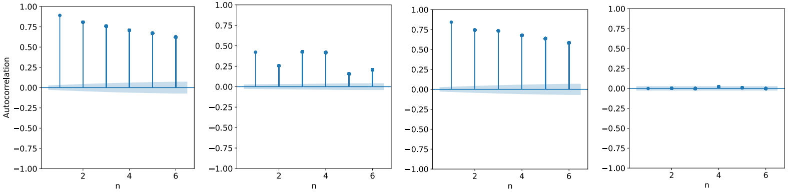
Causal Tracing: Results

Auto-correlation across different number of marked points n for clean (leftmost) and intervened runs (from left to right) for n, $\boldsymbol{B}$, and $\boldsymbol{d}$ respectively.



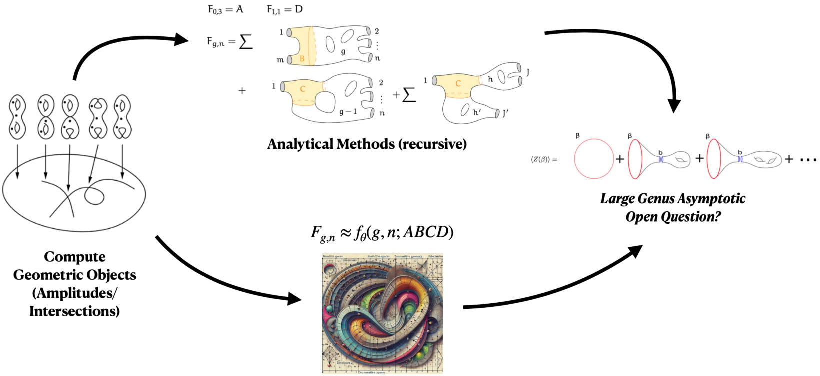
Causal Tracing: Results

Auto-correlation across different number of marked points n for clean (leftmost) and intervened runs (from left to right) for n, $\boldsymbol{B}$, and $\boldsymbol{d}$ respectively.



The model is predicting the intersections mainly using the partitions and the number of marked points!

How does the Model do Enumerative Geometry?



Abductive Hypothesis Testing

- We have a network that is predicting the values of Intersection numbers.

Abductive Hypothesis Testing

- We have a network that is predicting the values of Intersection numbers.
- We know that the model is learning the underlying constraints and geometry.

Abductive Hypothesis Testing

- We have a network that is predicting the values of Intersection numbers.
- We know that the model is learning the underlying constraints and geometry.
- Mathematicians with less data, but with intuitional hints, conjecture closed-form expressions for the Intersections.

Abductive Hypothesis Testing

- We have a network that is predicting the values of Intersection numbers.
- We know that the model is learning the underlying constraints and geometry (from both casual tracing and self-similarity)
- Mathematicians with less data, but with intuitional hints, conjecture closed-form expressions for the Intersections.
- **Question:** *In scenarios where the form (e.g parameters) of the asymptotic closed-form formulas are unknown, can one use AI-based abduction to infer and provide evidence for them?*

Abductive Hypothesis Testing

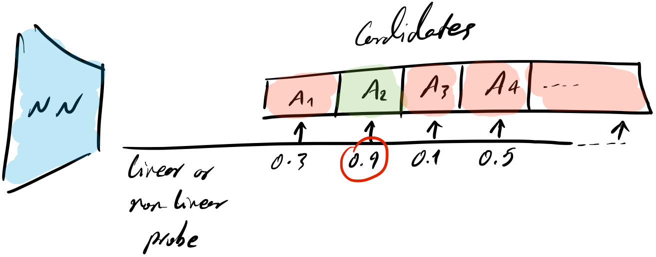
- Conjectural form of the large genus asymptotic ψ-class Intersections:

$$\langle \boldsymbol{d} \rangle_{g,n} \prod_{i=1}^{n} (2d_i + 1)!! \approx \frac{2^n}{4\pi} \frac{1}{(\mathfrak{A})^{2g-2+n}} \left(1 + \alpha_1 + \alpha_2 + \cdots\right)$$

Abductive Hypothesis Testing

- Conjectural form of the large genus asymptotic ψ-class Intersections:

$$\langle \boldsymbol{d} \rangle_{g,n} \prod_{i=1}^{n} (2d_i+1)!! \approx \frac{2^n}{4\pi} \frac{1}{(\mathfrak{A})^{2g-2+n}} \left(1 + \alpha_1 + \alpha_2 + \cdots \right)$$



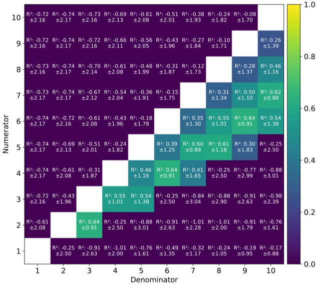
Abductive Hypothesis Testing

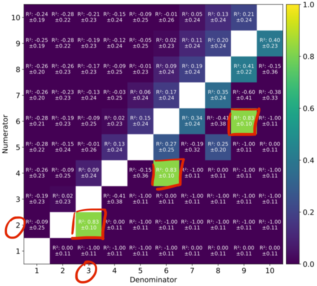
- Conjectural form of the large genus asymptotic ψ-class Intersections:

$$\langle \boldsymbol{d} \rangle_{g,n} \prod_{i=1}^{n} (2d_i + 1)!! \approx \frac{2^n}{4\pi} \frac{1}{(\mathfrak{A})^{2g-2+n}} (1 + \alpha_1 + \alpha_2 + \cdots)$$

- Given that $\mathfrak{A}$ is a rational number, let's grid-probe the it's values, such that the internal representation of the model best predicts RHS.

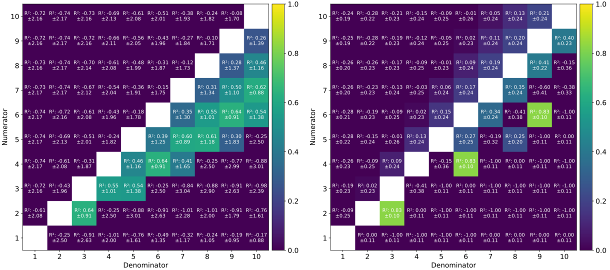
Linear probe





Non-linear probe

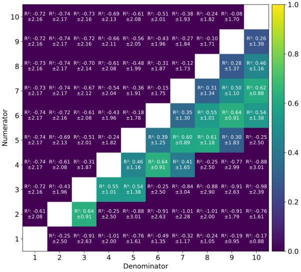
Abductive Hypothesis Testing

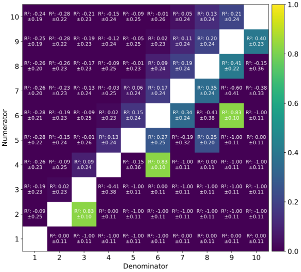


$\Rightarrow$ The network's internal representation of the polynomiality phenomenon (Guo et al. '22; Eynard et al. '23) does not have a simple linear form, but a non-linear representation.

Abductive Hypothesis Testing

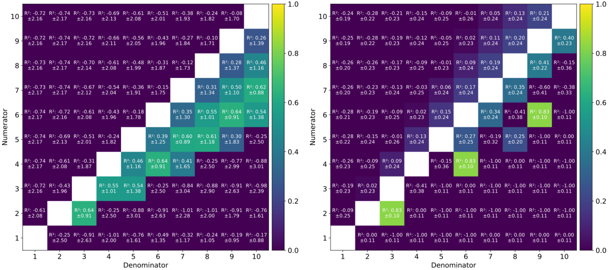
	No Intervention	n	$\boldsymbol{B}$	$\boldsymbol{d}$
R^2	0.96	-12.6	0.54	-52.2
R^2_{probe}	0.83	0.52	-2.77	0.43





Abductive Hypothesis Testing

	No Intervention	n	$\boldsymbol{B}$	$\boldsymbol{d}$
R^2	0.96	-12.6	0.54	-52.2
R^2_{probe}	0.83	0.52	-2.77	0.43



⇒ The model is implicitly learning the asymptotic closed-form expression of Intersection numbers!

Updated history of Intersection Theory of ψ-classes

- **1991** – **E.Witten** proposes a recursive formula for generating all intersection numbers of ψ-classes, asserting that their generating function satisfies the **KdV hierarchy** of differential equations.

Updated history of Intersection Theory of ψ-classes

- **1991** – **E.Witten** proposes a recursive formula for generating all intersection numbers of ψ-classes, asserting that their generating function satisfies the **KdV hierarchy** of differential equations.
- **1992** – **M. Kontsevich** proves Witten's conjecture by relating intersection numbers of ψ-classes to **ribbon graphs** and utilizing a matrix model approach.

Updated history of Intersection Theory of ψ-classes

- **1991** – **E.Witten** proposes a recursive formula for generating all intersection numbers of ψ-classes, asserting that their generating function satisfies the **KdV hierarchy** of differential equations.
- **1992** – **M. Kontsevich** proves Witten's conjecture by relating intersection numbers of ψ-classes to **ribbon graphs** and utilizing a matrix model approach.
- **2004** – **M. Mirzakhani** offers a novel proof of Witten's conjecture by linking intersection numbers of ψ-classes to **volumes of moduli spaces** of hyperbolic surfaces.

Updated history of Intersection Theory of ψ-classes

- **1991** – **E.Witten** proposes a recursive formula for generating all intersection numbers of ψ-classes, asserting that their generating function satisfies the **KdV hierarchy** of differential equations.
- **1992** – **M. Kontsevich** proves Witten's conjecture by relating intersection numbers of ψ-classes to **ribbon graphs** and utilizing a matrix model approach.
- **2004** – **M. Mirzakhani** offers a novel proof of Witten's conjecture by linking intersection numbers of ψ-classes to **volumes of moduli spaces** of hyperbolic surfaces.
- **2007** – **B. Eynard** and **N. Orantin** introduce **Topological Recursion**, generalizes Mirzakhani's recursive framework, **Topological Recursion** for computing intersection numbers.

Updated history of Intersection Theory of ψ-classes

- **1991** – **E.Witten** proposes a recursive formula for generating all intersection numbers of ψ-classes, asserting that their generating function satisfies the **KdV hierarchy** of differential equations.
- **1992** – **M. Kontsevich** proves Witten's conjecture by relating intersection numbers of ψ-classes to **ribbon graphs** and utilizing a matrix model approach.
- **2004** – **M. Mirzakhani** offers a novel proof of Witten's conjecture by linking intersection numbers of ψ-classes to **volumes of moduli spaces** of hyperbolic surfaces.
- **2007** – **B. Eynard** and **N. Orantin** introduce **Topological Recursion**, generalizes Mirzakhani's recursive framework, **Topological Recursion** for computing intersection numbers.
- **2018** – **M. Kontsevich** and **Y. Soibelman** develop **Quantum Airy Structures** (qAs), offering an *algebraic* approach to calculating intersection numbers with Topological Recursion.

Updated history of Intersection Theory of ψ-classes

- **1991** – **E.Witten** proposes a recursive formula for generating all intersection numbers of ψ-classes, asserting that their generating function satisfies the **KdV hierarchy** of differential equations.
- **1992** – **M. Kontsevich** proves Witten's conjecture by relating intersection numbers of ψ-classes to **ribbon graphs** and utilizing a matrix model approach.
- **2004** – **M. Mirzakhani** offers a novel proof of Witten's conjecture by linking intersection numbers of ψ-classes to **volumes of moduli spaces** of hyperbolic surfaces.
- **2007** – **B. Eynard** and **N. Orantin** introduce **Topological Recursion**, generalizes Mirzakhani's recursive framework, **Topological Recursion** for computing intersection numbers.
- **2018** – **M. Kontsevich** and **Y. Soibelman** develop **Quantum Airy Structures** (qAs), offering an *algebraic* approach to calculating intersection numbers with Topological Recursion.
- **2024** – **B. H.**, **R. Corominas** and **A. Giacchetto**, **Neural Enumerative Reasoning** for highly recursive ψ-class intersection numbers, and abductive hypothesis testing to recover the closed-form formula.