# A tourist's guide to intersection theory on moduli spaces of curves

Norman Do*


## Abstract

In the past few decades, moduli spaces of curves have become increasingly prominent and important in mathematics. In fact, the study of moduli spaces lies at the centre of a rich confluence of rather disparate areas such as geometry, combinatorics and string theory. Starting from baby principles, I will describe exactly what a moduli space is and motivate the study of its intersection theory. This scenic tour will guide us towards a discussion of Kontsevich's combinatorial formula, including a description of a new proof.


## What is a moduli space and what is it good for?

Let us begin our journey into the world of moduli spaces with the following two statements.

- A moduli space parametrises a family of geometric objects.
- Different points in a moduli space represent different geometric objects and nearby points represent objects with similar structure.

Rather than elaborate on these cryptic remarks, perhaps the best way forward is to consider the following example. Although of no great interest in itself, it will give us a taste of what a moduli space is and what it is good for.

**Toy example.** Let $\mathcal{M}_\triangle$ be the moduli space of triangles whose vertices are labelled $A$, $B$ and $C$. Therefore, every point in $\mathcal{M}_\triangle$ should correspond to a particular triangle. If we denote the side lengths by $a = BC$, $b = CA$ and $c = AB$, then the triangle can be uniquely described by the triple $(a, b, c)$. However, not all triples of positive real numbers give rise to a triangle. Indeed, a necessary and sufficient condition is that the three numbers must satisfy the triangle inequalities. So we can describe the moduli space of triangles as follows.

$$\mathcal{M}_\triangle = \{(a,b,c) \in \mathbb{R}^3_+ \mid a+b>c,\ b+c>a,\ \text{and } c+a>b\}.$$

Standing at a point in the moduli space corresponds to thinking about a particular triangle. On the other hand, moving through the moduli space corresponds to

Received 29 January 2008; accepted for publication 18 March 2008.

*Department of Mathematics and Statistics, The University of Melbourne, VIC 3010.
E-mail: N.Do@ms.unimelb.edu.au

Norman Do was joint winner of the B.H. Neumann Prize for best student talk at the Annual Meeting of the AustMS, held in Melbourne in September 2007.

continuously deforming the triangle[1]. We will use this toy example of a moduli space to consider a baby problem from enumerative geometry.

> How many triangles with vertices labelled $A$, $B$ and $C$ are isosceles, have at least one side of length 5, and have at least one side of length 7?

I have called this a baby problem because it is really quite simple. What is interesting, however, is the way in which a moduli space enthusiast would think about it. We start by defining the following three subsets of $\mathcal{M}_\triangle$.

- Let $X_{\text{iso}} \subseteq \mathcal{M}_\triangle$ be the locus of isosceles triangles.
- Let $X_5 \subseteq \mathcal{M}_\triangle$ be the locus of triangles with at least one side of length 5.
- Let $X_7 \subseteq \mathcal{M}_\triangle$ be the locus of triangles with at least one side of length 7.

Given these definitions, our baby problem can be rephrased as follows.

> How many points are in the set $X_{\text{iso}} \cap X_5 \cap X_7$?

So we have translated our original question about counting triangles to one concerning subsets of the moduli space of triangles. What we have gained from doing so is not an easier path to the solution, but a deeper geometric perspective on the matter. And it is this perspective that has motivated the exploration of moduli spaces in general, and moduli spaces of curves in particular.

## Intuitive intersection theory

Playing with our toy example led us to consider intersections of certain subsets lying in a larger space. Such matters belong to the realm of geometry known, unsurprisingly, as intersection theory. Despite arising from such simple considerations, intersection theory is both deep and technical. In this article, we will content ourselves with a brief review of the underlying intuition.

We begin by defining the *codimension* of a $D$-dimensional subset of an $N$-dimensional space to be the number $N - D$. For example, a line in the plane has codimension 1, while a line in space has codimension 2. Now we can state one of the most fundamental facts about intersection theory.

**Fundamental fact.** A generic intersection between two subsets with codimensions $D_1$ and $D_2$ has codimension $D_1 + D_2$. It follows that a generic intersection between $m$ subsets with codimensions $D_1, D_2, \ldots, D_m$ has codimension $D_1 + D_2 + \cdots + D_m$.

For example,

- a line in the plane (codimension 1) will generically intersect another line (codimension 1) in a point (codimension 2);
- a line in space (codimension 2) will generically intersect a plane (codimension 1) in a point (codimension 3); and
- a plane in space (codimension 1) will generically intersect another plane (codimension 1) in a line (codimension 2).

[1] At this point in time, the keen reader may like to consider why we bothered to label the vertices of the triangle $A$, $B$ and $C$. What would happen if we left the vertices unlabelled?

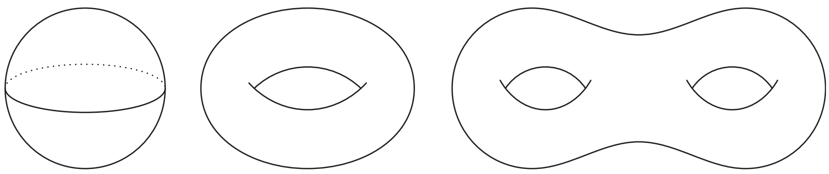
By now, you might be wondering what I mean by the word 'generic'. I have carefully used it to deal with certain cases which would otherwise render our fundamental fact blatantly false. For example, two copies of the same line in the plane do not meet in a point, as expected. One way to get around this fact is simply to 'jiggle the picture'. A slight perturbation of the subsets will then leave us with a generic intersection with the expected codimension[2].

One important case is when we have $m$ subsets of an $N$-dimensional space with codimensions $D_1 + D_2 + \cdots + D_m = N$. Then their intersection is a set of points and the number of these points is called an *intersection number*. We will use the following non-standard notation for intersection numbers.

$$X_1 \cdot X_2 \cdots X_m = \text{the number of points in the set } X_1 \cap X_2 \cap \ldots \cap X_m \, .$$

## An introduction to moduli spaces of curves

A long time ago, topologists showed that (closed, connected, orientable) surfaces are classified by their genus, or more informally, their number of holes.




However, someone interested in geometry, rather than topology, would want to know more. They would consider the question, 'What can one do to a surface?' If you asked an algebraic geometer, they would be interested in putting an algebraic structure on a given surface. In other words, they would represent it as the complex solution set to a polynomial in two variables.

$$\mathcal{C} = \{(x, y) \in \mathbb{C}^2 \mid P(x, y) = 0\}.$$

Such sets can be considered up to a certain equivalence relation and the resulting equivalence classes are known as algebraic curves.

On the other hand, a mathematician interested in hyperbolic geometry, complex analysis or differential geometry would consider putting other structures on a surface. The following table summarises the different approaches and the three resulting objects — namely, algebraic curves, hyperbolic surfaces and Riemann surfaces.

A surprising fact is that no matter which of these structures you choose to put on a surface, you essentially get the same result. More precisely, there is a natural dictionary correspondence between the set of algebraic curves[3], the set of hyperbolic surfaces and the set of Riemann surfaces!

[2] This is analogous to the geometric notion of points being in general position. For example, a finite set of points in the plane is said to be in general position if no three lie on a line. A set of points which is not in general position can be made so by a slight perturbation.
[3] To be more precise still, we should really be talking about *smooth* algebraic structures and *smooth* algebraic curves.

| Area | Structure | Equivalence | Class |
|---|---|---|---|
| algebraic geometry | algebraic structure | isomorphism | algebraic curve |
| hyperbolic geometry | hyperbolic structure | isometry | hyperbolic surface |
| complex analysis | complex structure | biholomorphism | Riemann surface |
| differential geometry | Riemannian metric | conformal equivalence | Riemann surface |

We are now ready to define moduli spaces of curves as follows[4].

$$\begin{aligned}\mathcal{M}_{g,n} &= \text{the moduli space of algebraic curves with genus } g \text{ and } n \text{ marked points,}\\ &= \text{the moduli space of hyperbolic surfaces with genus } g \text{ and } n \text{ cusps,}\\ &= \text{the moduli space of Riemann surfaces with genus } g \text{ and } n \text{ marked points.}\end{aligned}$$

Standing at a point in $\mathcal{M}_{g,n}$ corresponds to thinking about a particular algebraic curve with genus $g$ and some marked points on it labelled from 1 up to $n$. On the other hand, moving through $\mathcal{M}_{g,n}$ corresponds to continuously deforming the algebraic curve and/or continuously moving these marked points. Unfortunately, there are three somewhat technical issues we must deal with before progressing.

- Problem: The space $\mathcal{M}_{g,n}$ does not always exist.
  Solution: There are just a handful of problem cases which we must exclude: $(g, n) = (0, 0)$, $(0, 1)$, $(0, 2)$ and $(1, 0)$. However, all other non-negative integer values for $g$ and $n$ do give us bona fide moduli spaces.
- Problem: The space $\mathcal{M}_{g,n}$ is not compact.
  Solution: Compact spaces are generally more tractable than those which are not. Fortunately, there are many ways to compactify $\mathcal{M}_{g,n}$ by throwing in some extra points[5]. The most natural compactification is due to David Mumford and Pierre Deligne and involves considering $\overline{\mathcal{M}}_{g,n}$, the moduli space of *stable curves* with genus $g$ and $n$ marked points. Stable curves are constructed by picking a set of smooth algebraic curves with marked points and gluing them together at pairs of marked points. From this definition, it should be clear that $\mathcal{M}_{g,n} \subseteq \overline{\mathcal{M}}_{g,n}$.
- Problem: The space $\overline{\mathcal{M}}_{g,n}$ is not a manifold.
  Solution: We are forced to treat the space $\overline{\mathcal{M}}_{g,n}$ as an orbifold. Although the definition of an orbifold is quite technical, we need only keep in mind that there are certain subsets of the space which correspond to 'fractional points'. In particular, there may be points which correspond to $\frac{1}{2}$ points, $\frac{1}{3}$ points, $\frac{1}{4}$ points, and so on. The upshot of all this is that when counting intersection numbers, we must expect them to be rational, rather than integral. Despite seeming counterintuitive, this is the only way that we can accommodate an intersection theory which obeys the rules that are consistent with our intuition.

One of the earliest results on moduli spaces of curves is the following, which was essentially known to Riemann himself.

[4] Note that a cusp is basically a puncture on a hyperbolic surface which has been 'pulled out to infinity'. In the natural dictionary correspondence between algebraic curves and hyperbolic surfaces, a marked point translates to a cusp.

[5] As an example of compactification, note that the moduli space of triangles discussed earlier can be compactified by including degenerate triangles formed by three collinear points and triangles with sides of length 0.

**Fact.** The dimension of the moduli space $\overline{\mathcal{M}}_{g,n}$ is $6g - 6 + 2n$.

This shows that moduli spaces of curves can be of arbitrarily high, though necessarily even, dimension. As a result, they can be very difficult to visualise. The difficulty is compounded by the fact that the structure of $\overline{\mathcal{M}}_{g,n}$ is also extremely complicated in general.

### Intersection numbers on moduli spaces of curves

Since we are interested in the intersection theory of moduli spaces of curves, we need to find some particular subsets of $\overline{\mathcal{M}}_{g,n}$ to intersect with each other. It turns out that on $\overline{\mathcal{M}}_{g,n}$, there are $n$ natural cohomology classes, one for each of the marked points.

$$\psi_1, \psi_2, \ldots, \psi_n \in H^2(\overline{\mathcal{M}}_{g,n}, \mathbb{Q}) .$$

Now it doesn't matter if you don't know what cohomology is because, translated into the language of intuitive intersection theory, what this means is that $\psi_1, \psi_2, \ldots, \psi_n$ correspond to subsets of $\overline{\mathcal{M}}_{g,n}$ of codimension two. Therefore, if we pick non-negative integers $a_1 + a_2 + \cdots + a_n = \frac{1}{2} \dim \overline{\mathcal{M}}_{g,n} = 3g - 3 + n$, then we can form the intersection number

$$\psi_1^{a_1} \cdot \psi_2^{a_2} \cdots \psi_n^{a_n} \in \mathbb{Q}.$$

This means that we should take $a_1$ copies of $\psi_1$, $a_2$ copies of $\psi_2$, and so on, that we should jiggle everything so that the subsets intersect generically, and that we should count the number of points of intersection, keeping in mind that some points are fractional. We call $\psi_1, \psi_2, \ldots, \psi_n$ *psi-classes* and refer to the numbers $\psi_1^{a_1} \cdot \psi_2^{a_2} \cdots \psi_n^{a_n}$ as *intersection numbers of psi-classes.* For example, the intersection number $\psi_1 \cdot \psi_2$ on $\overline{\mathcal{M}}_{0,5}$ is 2 while the intersection number $\psi_1$ on $\overline{\mathcal{M}}_{1,1}$ is $\frac{1}{24}$. In general, these intersection numbers of psi-classes are very tough to calculate.

But before we get too carried away with moduli spaces of curves and intersection numbers of psi-classes, it would be remiss not to mention why anyone would want to study them. From the way that we have motivated this discussion, one can see that moduli spaces of curves are of fundamental importance in algebraic geometry, hyperbolic geometry and topology. A deeper look reveals interesting connections with seemingly unrelated areas such as combinatorics, integrable systems and matrix models. However, one of the most amazing places where moduli spaces can be found is in string theory. In fact, the area of moduli spaces provides us with one of the most fascinating illustrations of the symbiosis between pure mathematics and theoretical physics.

One of the landmark results on moduli spaces of curves is a conjecture first posed by Edward Witten. The conjecture involves the intersection numbers of psi-classes and, amazingly, arose from the study of a particular model of two-dimensional quantum gravity. The following is a very brief history of Witten's conjecture and some of the various proofs that have emerged over the years.

- In 1991, Edward Witten gave a conjectural recursive formula to generate all intersection numbers of psi-classes. More precisely, he claimed that a natural generating function for the intersection numbers of psi-classes satisfies a series of differential equations known as the KdV hierarchy[6] [5].

- In 1992, Maxim Kontsevich found a formula which relates intersection numbers of psi-classes with combinatorial objects known as ribbon graphs. He then deduced Witten's conjecture from this by using the analysis of a related matrix model [2].
- In 2001, Andrei Okounkov and Rahul Pandharipande found another proof of Witten's conjecture by relating intersection numbers of psi-classes with Hurwitz numbers. Hurwitz numbers arise in the enumeration of branched covers of the sphere and in counting factorisations of permutations into transpositions [4].
- In 2004, Maryam Mirzakhani found yet another proof of Witten's conjecture by relating intersection numbers of psi-classes with volumes of moduli spaces of hyperbolic surfaces [3].
- My own contribution to the area consists of a new proof of Kontsevich's combinatorial formula using the volumes of moduli spaces of hyperbolic surfaces considered by Mirzakhani. This gives a new path to Witten's conjecture by drawing together two previously unrelated proofs [1].

Further testament to the importance of moduli spaces of curves is the fact that five of the mathematicians we have mentioned — namely Mumford, Deligne, Witten, Kontsevich and Okounkov — are all recipients of the prestigious Fields medal.

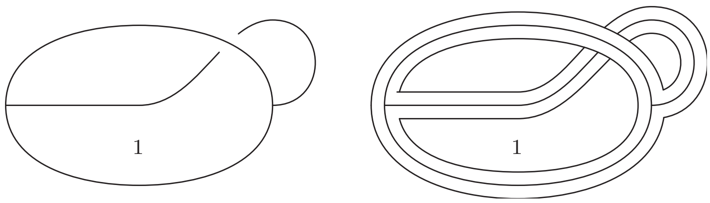
## Kontsevich's combinatorial formula explained

We have already mentioned that Kontsevich's proof of Witten's conjecture relies on a combinatorial formula involving ribbon graphs. We define a *ribbon graph of type* $(g, n)$ to be a graph with a cyclic ordering of the edges meeting at every vertex which can be thickened to give a surface with genus $g$ and $n$ boundaries labelled from 1 up to $n$.

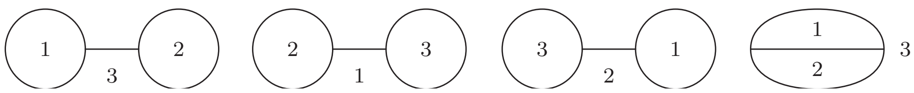
For example, consider the following diagram which shows a ribbon graph of type $(1, 1)$ on the left. It has three edges which meet in the same cyclic orientation at the two vertices. To thicken the ribbon graph, imagine that it is a railing being held on to by an incompetent ice skater. As they work their way around the ribbon graph, they carve out the edges of the thickened graph, as shown on the right. Gluing ribbons onto the edges of the graph in the manner shown yields a surface with exactly one boundary. Although it is not immediately obvious, the surface actually has genus one.

As another example, we have the following four ribbon graphs. Note that they are all trivalent, meaning that all vertices have degree three. In fact, they are the only trivalent ribbon graphs of type $(0, 3)$. The three boundary components are labelled with the numbers 1, 2, 3 and come from the fact that this time we require three incompetent ice skaters to traverse both sides of every edge. To understand

[6] The KdV (Korteweg–de Vries) equation is a non-linear partial differential equation which first appeared in the study of shallow water waves and has generated a tremendous amount of mathematical interest over the past few decades. It seems to me both interesting and surprising that this equation from classical physics is central to Witten's conjecture, which originated from theoretical physics.




the effect of the cyclic orientations of the edges, consider the ribbon graph on the far right. As a graph, it is isomorphic to the example above since both contain three edges connecting two vertices. However, it is a different ribbon graph since after being thickened, one obtains a surface with genus zero and three boundaries.




Now without further ado, let me introduce Kontsevich's combinatorial formula, which at first glance appears to be an obfuscated mess of mathematical symbols.

**Kontsevich's combinatorial formula.** The intersection numbers of psi-classes on $\overline{\mathcal{M}}_{g,n}$ satisfy the following equation:

$$\sum_{|a|=3g-3+n} \psi_1^{a_1} \cdot \psi_2^{a_2} \cdots \psi_n^{a_n} \prod_{k=1}^{n} \frac{(2a_k-1)!!}{s_k^{2a_k+1}} = \sum_{G \in \mathcal{R}_{g,n}} \frac{2^{2g-2+n}}{|\mathrm{Aut}(G)|} \prod_{e \in G} \frac{1}{s_{\ell(e)} + s_{r(e)}}.$$

We will attempt to unravel this seemingly complicated formula and make sense of what it has to offer.

- The left-hand side is a polynomial in $\frac{1}{s_1}, \frac{1}{s_2}, \ldots, \frac{1}{s_n}$ and its coefficients store all intersection numbers of psi-classes on $\overline{\mathcal{M}}_{g,n}$.
- The right-hand side is a rational function in $s_1, s_2, \ldots, s_n$ obtained by a strange enumeration.
  - On the outside is a sum over all trivalent ribbon graphs $G$ of type $(g, n)$.
  - On the inside is a product over the edges of $G$ — here, $\ell(e)$ and $r(e)$ denote the labels of the boundaries on each side of $e$.
  - In between is a constant involving $|\mathrm{Aut}(G)|$, which denotes the number of automorphisms of the ribbon graph $G$.

Therefore, if we wanted to calculate the psi-class intersection numbers on $\overline{\mathcal{M}}_{g,n}$, we simply need to list all of the trivalent ribbon graphs of type $(g, n)$ and write down the corresponding terms in the summation on the right-hand side of Kontsevich's combinatorial formula. The result should be a polynomial from which we can simply read off the psi-class intersection numbers. A priori, it seems that Kontsevich's combinatorial formula could not possibly be true since the left-hand side is polynomial in nature, while the right-hand side does not appear to be a polynomial at all! Seeing is believing, so let us consider a particular example.

**Example.** In the very simple case of $g = 0$ and $n = 3$, the left-hand side contains only one term.

$$\mathrm{LHS} = \psi_1^0 \cdot \psi_2^0 \cdot \psi_3^0 \ \frac{1}{s_1 s_2 s_3}.$$

For the right-hand side, consider the leftmost of the four trivalent ribbon graphs of type $(0,3)$ pictured earlier. One of its edges is adjacent to the boundary labelled 1 and the boundary labelled 3. Therefore, Kontsevich's combinatorial formula tells us to write down the expression $1/(s_1 + s_3)$. Doing this for all three edges and multiplying the corresponding expressions together gives us the term $1/(2s_3(s_1 + s_3)(s_2 + s_3))$.

The next step is to add up these expressions over all four trivalent ribbon graphs of type $(0,3)$, each multiplied by the appropriate constant.

$$\begin{aligned}\text{RHS} &= \frac{2}{2s_3(s_1+s_3)(s_2+s_3)} + \frac{2}{2s_1(s_2+s_1)(s_3+s_1)} \\ &\quad + \frac{2}{2s_2(s_3+s_2)(s_1+s_2)} + \frac{2}{(s_1+s_2)(s_2+s_3)(s_3+s_1)} \\ &= \frac{s_1s_2(s_1+s_2) + s_2s_3(s_2+s_3) + s_3s_1(s_3+s_1) + 2s_1s_2s_3}{s_1s_2s_3(s_1+s_2)(s_2+s_3)(s_3+s_1)} \\ &= \frac{(s_1+s_2)(s_2+s_3)(s_3+s_1)}{s_1s_2s_3(s_1+s_2)(s_2+s_3)(s_3+s_1)} \\ &= \frac{1}{s_1s_2s_3}\,.\end{aligned}$$

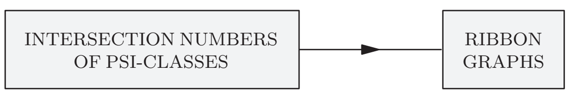
Finally, equating the left- and right-hand sides allows us to conclude that $\psi_1^0 \cdot \psi_2^0 \cdot \psi_3^0 = 1$.

Unfortunately, all this hard work has been for nought since the intersection number $\psi_1^0 \cdot \psi_2^0 \cdot \psi_3^0$ represents an intersection of zero subsets and is equal to one by definition! However, Kontsevich's combinatorial formula did give us the correct answer, and one can observe that there was some algebraic magic required to show that the left-hand side, which was inherently polynomial, was equal to the right-hand side, which was not. You can take my word for it that Kontsevich's combinatorial formula continues to hold for larger values of $g$ and $n$ and that the algebraic magic is even more striking.

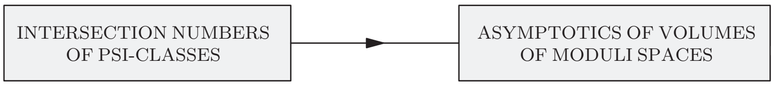
## A new proof of Kontsevich's combinatorial formula

An accurate blow-by-blow account of the proof of Kontsevich's combinatorial formula would inflate the size of this already lengthy article several times (cf. my PhD thesis). Given the constraints and nature of this exposition, let us content ourselves with understanding how intersection numbers of psi-classes are related to ribbon graphs.

### Kontsevich's combinatorial formula simplified



**Very sketchy proof: Step 1**

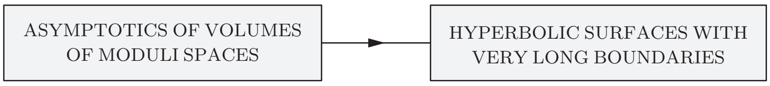



Earlier it was noted that we have some choice in whether we want to consider the moduli space of algebraic curves, Riemann surfaces or hyperbolic surfaces. One of the distinct advantages of working with hyperbolic surfaces is that we can open up each cusp to give a boundary whose length can be any positive real number. This allows us to define $\mathcal{M}_{g,n}(\ell_1, \ell_2, \ldots, \ell_n)$ to be the moduli space of hyperbolic surfaces with genus $g$ and $n$ boundaries of lengths $\ell_1, \ell_2, \ldots, \ell_n$.

For all choices of $\ell_1, \ell_2, \ldots, \ell_n$, these spaces are naturally endowed with a symplectic structure known as the Weil–Petersson symplectic form. What this means is that we can assign a volume to these moduli spaces, a calculation which was recently accomplished by Mirzakhani. Essentially, she proved that the volume $V_{g,n}(\ell_1, \ell_2, \ldots, \ell_n)$ of $\mathcal{M}_{g,n}(\ell_1, \ell_2, \ldots, \ell_n)$ is a polynomial in $\ell_1, \ell_2, \ldots, \ell_n$ of degree $6g - 6 + 2n$. Furthermore, its top degree coefficients store all of the intersection numbers of psi-classes on $\overline{\mathcal{M}}_{g,n}$, the very numbers that we wish to know.

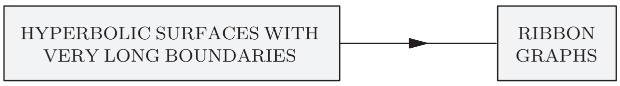
So how does one access the top degree coefficients of a polynomial? The answer is to look at its asymptotics — more precisely, we consider the behaviour of $V_{g,n}(T\ell_1, T\ell_2, \ldots, T\ell_n)$ as $T$ approaches infinity.

**Very sketchy proof: Step 2**

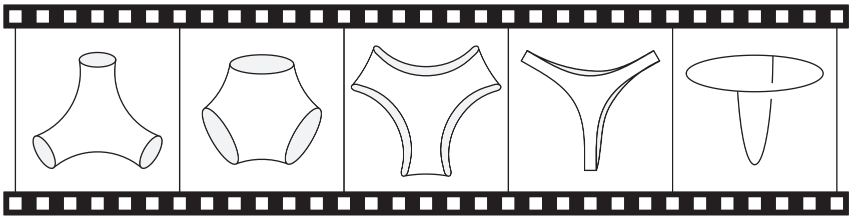



The connection here is simple. To understand $V_{g,n}(T\ell_1, T\ell_2, \ldots, T\ell_n)$ for very large values of $T$, of course, one must first understand hyperbolic surfaces with boundaries of lengths $T\ell_1, T\ell_2, \ldots, T\ell_n$ for very large values of $T$.

**Very sketchy proof: Step 3**




The final piece of the puzzle is to relate hyperbolic surfaces with very long boundaries and ribbon graphs. By the Gauss–Bonnet theorem, a fundamental tool in geometry, all hyperbolic surfaces with genus $g$ and $n$ boundaries have the same surface area. If we begin to 'stretch out' the boundaries to make them longer, the surface area condition forces the surface to become skinny. Now suppose that while we are stretching out the boundaries, we also simultaneously 'zoom out' so that the surface remains in our field of vision. After performing this stretch-and-zoom ad infinitum, we obtain a very skinny surface indeed. And what does a very skinny surface look like? A ribbon graph, of course!



## References


[1] Do, N. Intersection theory on moduli spaces of curves from a hyperbolic viewpoint. PhD thesis, The University of Melbourne, Australia (in progress).

[2] Kontsevich, M. (1992). Intersection theory on the moduli space of curves and the matrix Airy function. *Communications in Mathematical Physics* **47,** 1–23.

[3] Mirzakhani, M. (2007). Weil–Petersson volumes and Intersection theory on the moduli space of curves. *J. Amer. Math. Soc.* **20,** 1–23.

[4] Okounkov, A. and Pandharipande, R. (2001). Gromov–Witten theory, Hurwitz numbers, and matrix models, I. Preprint math.AG/0101147v1.

[5] Witten, E. (1991). Two-dimensional quantum gravity and intersection theory on moduli space. *Surveys in Diff. Geom.* **1,** 243–310.