Institut de mathématiques de Jussieu - Paris Rive Gauche

École doctorale de sciences mathématiques de Paris centre

Thèse de doctorat

Discipline : Mathématiques

présentée par

Mingkun Liu

Random multi-geodesics on hyperbolic surfaces

Multi-géodésiques aléatoires sur les surfaces hyperboliques

---

dirigée par Vincent Delecroix, Bram Petri, and Anton Zorich

Soutenue le 27 septembre 2022 devant le jury composé de :

M. Nicolas Curien	Université Paris-Saclay	rapporteur
M. Vincent Delecroix	Université de Bordeaux	invité
Mme Viveka Erlandsson	University of Bristol	examinatrice
Mme Ursula Hamenstädt	Universität Bonn	examinatrice
M. Carlos Matheus	École polytechnique	examinateur
M. Bram Petri	IMJ-PRG	invité
M. Kasra Rafi	University of Toronto	examinateur
M. Alex Wright	University of Michigan	rapporteur
M. Anton Zorich	IMJ-PRG	directeur

Institut de mathématiques
de Jussieu - Paris Rive gauche.
UMR 7586.
Boite courrier 247
4 place Jussieu
75 252 Paris Cedex 05

Université Paris Cité.
École doctorale de sciences
mathématiques de Paris centre.
Boite courrier 290
4 place Jussieu
75 252 Paris Cedex 05

Remerciements

Une thèse est une expérience aléatoire. Dans cette expérience, j'ai eu énormément de chance d'avoir rencontré mes trois directeurs.

Merci Anton ! Merci de m'avoir confié ce sujet de thèse à la fois amusant et profond, et de m'avoir appris bien plus que les mathématiques. Ta présence et ton soutien aux moments décisifs m'ont été inestimables. Je n'aurais pu rêver meilleur directeur.

Merci Bram ! Merci d'avoir pris le temps de m'avoir expliqué tant de mathématiques fascinantes à de si nombreuses reprises, et de d'avoir su mettre de la rigueur dans mon enthousiame un peu trop empressé. Merci pour les moments dans le bureau 526, dans le couloir 16-26, dans le Jardin des plantes, dans la rue Linné, la rue des Écoles, la rue de la Roquette, et la rue Crozatier.

Merci Vincent ! En plus des mathématiques passionnantes sur lesquelles nous avons travaillé ensemble, merci de m'avoir démontré que « les mathématiques sont une science expérimentale » par tes simulations numériques ingénieuses. Les discussions avec toi sont toujours fort éclairantes, et tes nombreux conseils avisés me sont précieux. C'est toujours un plaisir de passer à Bordeaux.

Je tiens à remercier vivement Nicolas Curien et Alex Wright qui m'ont fait l'honneur d'être rapporteurs de cette thèse, ainsi que Viveka Erlandsson, Ursula Hamenstädt, Carlos Matheus, et Kasra Rafi d'avoir bien voulu participer à ce jury.

Je souhaite remercier les jeunes chercheurs avec qui j'ai pu parler de mathématiques (mais pas que). Merci à Simon Barazer, Etienne Bonnafoux, Ivan Yakovlev, Charles Fougeron, Laura Monk, Alessandro Giacchetto, Elba Garcia-Failde, Yiming Zang, Eduard Duryev, Baptiste Louf, Adrien Sauvaget, Thomas Budzinski, Valentin Baillard, Amol Aggarwal, Aaron Fenyes, Aaron Calderon, Francisco Arana-Herrera, Sam Freedman, Kai Fu, Juan Marshall... et un remerciement spécial à Julien Boulanger, Jérôme Carrand, Pablo Montealegre, Magali Jay, et Mélanie Theillière pour la relecture.

Je suis également très reconnaissant aux chercheurs un tout petit peu moins jeunes que j'ai rencontrés et qui m'ont encouragés et soutenus pendant cette thèse : Erwan Lanneau, Federica Fanoni, Samuel Lelièvre, Alba Málaga, Elise Goujard, Hugo Parlier, Daniel Massart, Samuel Tapie, Pascal Hubert, Dimitri Zvonkine, Julien Courtiel, Anna Lenzhen, Sébastien Labbé, François Ledrappier, Dawei Chen, Frank Herrlich. J'aimerais remercier Maryam Mirzakhani que je n'ai pas pu avoir l'occasion de rencontrer.

Je tiens à exprimer ma profonde gratitude à Philippe Ciarlet sans qui je n'aurais probablement jamais pu commencer mes études en France, condition sine qua non de cette belle aventure qui est le doctorat. Je souhaite également remercier mes professeurs de l'X, dont Sébastien Boucksom, Anna Cadoret, Charles Favre, et srtout, mon tuteur de stage Luc Pirio.

Un grand merci aux chercheurs de l'IMJ-PRG (dans un ordre aléatoire comme d'habitude) : Marie-Claude Arnaud, Bassam Fayad, Laurent Charles, Hussein Mourtada, Julien Marché, Maxime Wolff, Antonin Guilloux, Gregor Masbaum, Elisha Falbel, Olivier Debarre, Xiaonan Ma, Frédéric Naud, Romain Tessera, Gilles Courtois, Claire Debord, Thibault Lefeuvre, Loïc Merel, Serguei Barannikov, Andrés Sambarino. J'aimerais également remercier tous les membres de l'administration de l'IMJ-PRG, notamment Amina Hariti et Emilie Jacquemot.

Je n'oublie pas les doctorants de l'IMJ-PRG. Merci à Ratko (et ton mantra "La porte est là". Les soirs qu'on a passé ensemble avec Oussama dans le bureau 751 me manquent déjà), Hao et Wille (pour les jours à la Collégiale), Tommaso et Antoine (les deux beaux gosses du 7e), Mahsa (pour le safran), Zahraa (pour son courage), Rahman (le plus grand logicien que je connaisse), Jérôme (c'est toujours un plaisir d'écouter notre doctorant émérite au ministère !), Maxime (pour les repas au Crous qu'on a dégustés ensemble), Jérémie (pour les discussions impressionnantes sur la physique), Tal (grâce à qui le 7e a retrouvé sa vitalité), Marcel, Matias, et Maud (pour leur accompagnement au bureau 750), Anna, Thomas, Pietro et Irene, Yi, Haowen, Chenyu, Zhuchao, Alessandro, Jules, Romain, Farouk, Emmanuel, Julien, Skander, Razvan, Dorian, Théo, Yanbo, Keyu, Qiaochu, Frank, Alex...

Merci à Enguerrand, à Anass, Méline, et Soufiane, pour de magnifiques vacances ensemble.

Merci à Marion et Clémence, qui m'ont appris le français, pour les livres, les BDs, et tant d'années d'amitié.

Merci à ma famille (presque) à Villeneuve : Hélène, Michel, Lucile, Manu, Lïa, et Yann. Et à Vincennes : Coralie, Thierry, Camille, Louis, et Yaku.

Merci à Grégoire, à Oussama, à Daniele, à Jieao, et à Laura, à travers qui j'ai pu avoir un lien plus fort avec ce monde, et sans qui les jours les plus difficiles auraient été insurmontables.

最后，感谢我的爸爸妈妈，我的奶奶，和去世的姥姥姥爷。谢谢你们一直以来的理解，宽容，支持，和爱。我将这篇论文献给你们。

Résumé

Une multi-géodésique est une union disjointe de géodésiques fermées sans auto-intersections. Après avoir expliqué comment choisir au hasard une multi-géodésique sur une surface hyperbolique, on détermine la loi de la partition de la longueur totale d'une multi-géodésique aléatoire d'un type topologique fixe sur une surface hyperbolique en utilisant les méthodes développés par Margulis dans sa thèse et le théorème d'équidistribution des horosphères dû à Mirzakhani. Cette loi admet une densité polynomiale dont les coefficients s'écrivent explicitement en termes de nombres d'intersection des classes de psi sur la compactification de Deligne–Mumford de l'espace de modules de courbes complexes lisses, et en particulier, elle ne dépend pas de la métrique hyperbolique de la surface. On montre ensuite que, lorsque le genre de la surface tend vers l'infinie, la distribution de la partition de la longueur totale d'une multi-géodésique aléatoire générale (sans imposer de contraintes topologiques) converge en loi vers le processus de Poisson–Dirichlet de paramètre $1/2$. En particulier, les longueurs moyennes des trois composantes connexes les plus longues d'une multi-géodésique aléatoire sur une surface hyperbolique de grand genre sont approximativement $75,8\%$, $17,1\%$, $4,9\%$, respectivement, de la longueur totale. Les éléments clés de la preuve sont l'analyse asymptotique des nombres d'intersection et des volumes de Masur–Veech effectuée par Aggarwal, et les travaux de Delecroix–Goujard–Zograf–Zorich sur la topologie des multi-géodésiques aléatoire.

Mots-clés

multi-courbes, surfaces hyperboliques, surfaces à petits carreaux, théorie de Teichmüller, loi de Poisson–Dirichlet, théorie d'intersection sur les espaces de modules.

Abstract

A multi-geodesic is a disjoint union of closed geodesics without self-intersections. After explaining how to randomly pick a multi-geodesic on a hyperbolic surface, we determine the distribution of the length partition of a random multi-geodesic with fixed topological type on a hyperbolic surface using the methods of Margulis' thesis and Mirzakhani's equidistribution theorem for horospheres. This distribution admits a polynomial density, whose coefficients can be expressed explicitly in terms of intersection numbers of psi-classes on the Deligne–Mumford compactification of the moduli space of smooth complex curves, and in particular it does not depend on the hyperbolic metric of the surface. We then show that, as the genus of the surface goes to infinity, the distribution of the length partition of a general multi-geodesic (with no topological constraints) converges in law to the Poisson–Dirichlet process of parameter $1/2$. In particular, the average lengths of the first three largest connected components of a random multi-geodesic on a large genus hyperbolic surface are approximately, 75.8%, 17.1%, 4.9%, respectively, of its total length. The key ingredients of the proof are the asymptotic analysis of intersection numbers and Masur–Veech volumes performed by Aggarwal, and the work of Delecroix–Goujard–Zograf–Zorich on the topology of random multi-geodesics.

Keywords

multicurves, hyperbolic surfaces, square-tiled surfaces, Teichmüller theory, Poisson–Dirichlet distribution, intersection theory on moduli spaces.

Contents

Chapter 1

Introduction

This thesis attempts to answer the following question:

What does a random curve on a large genus surface look like?

More precisely, we are interested in random multi-geodesics on hyperbolic surfaces, particularly when the genus of the surface becomes large.

1.1 Random hyperbolic multi-geodesics

Closed *geodesics* have fundamental importance in geometry. Those that do not intersect themselves, called *simple* closed geodesic, have been central in the study of hyperbolic surfaces and their moduli spaces for a long time and for various reasons. For example, a surface can be cut nicely along a simple closed geodesic (see e.g. [Mir07b], [ABO17]); it should perhaps also be mentioned that the curve complex, which will not at all be treated here, plays a crucial role in the study of the Teichmüller space and of mapping class group.

A *multi-geodesic* is a disjoint union of simple closed geodesics which are not necessarily primitive and whose orientations shall be ignored (a closed geodesic is said to be *primitive* if it traces out on its image exactly once). A multi-geodesic can be written as a formal sum $m_1\gamma_1 + \cdots + m_k\gamma_k$ where m_i's are positive integers and γ_i's are disjoint primitive simple closed geodesics. Each m_i is called the *multiplicity* of γ_i, and $m_i\gamma_i$ is called a *connected component*, or simply a *component*, of γ. On a closed hyperbolic surface, every closed curve not freely homotopic to a point is freely homotopic to a unique closed geodesic, and for any $R > 0$, there are only finitely many closed geodesics of lengths at most R.

The starting point of this thesis is the following result that Mirzakhani proved in her thesis:

Theorem 1.1.1 ([Mir08b]). *Let X be a closed hyperbolic surface of genus $g \geq 2$. Given a multi-geodesic γ on X, define $s_X(\gamma, R)$ to be the set of multi-geodesics on X of length at most R with the same topological type than that of γ (two curves are said to have the same topological type if they lie in the same mapping class group orbit). Then*

$$|s_X(\gamma, R)| \sim c(\gamma) \cdot \frac{B(X)}{b_g} \cdot R^{6g-6} \tag{1.1}$$

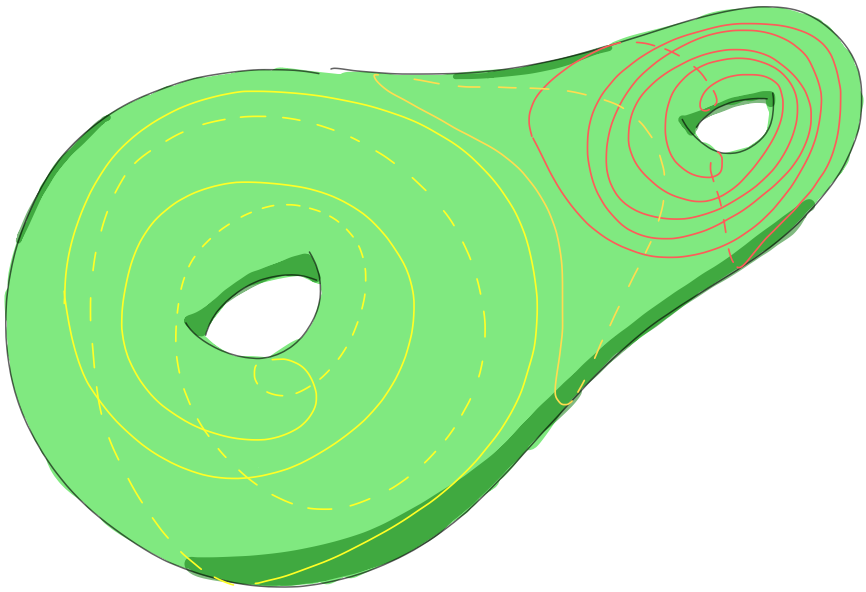

Figure 1.1: A multi-geodesic

as $R \to \infty$, where $c(\gamma)$, $B(X)$, and b_g are explicit positive constants depending only on γ, X, and g respectively.

Remark 1.1.2. A closed geodesics is very rarely simple. By a famous theorem due to Huber [Hub59] and Selberg called *prime geodesic theorem*, the number of (oriented) primitive closed geodesic of length at most R is asymptotically equivalent to $\exp(R)/R$ as R tends to infinite. Similar results holds for hyperbolic manifolds, and is generalised to compact negatively curved manifolds by Margulis in his thesis [Mar04], in a completely different approach (dynamic systems). Before the breakthrough work of Mirzakhani, progress in simple closed geodesics counting had been made in [BS85], [MR95a], [MR95b], and [Riv01]. Counting geodesics on hyperbolic surfaces keeps being a subject of active research. For some more recent developments, see e.g., [AH20a] (filling curves), [Mir16], [ES16] (non-simple multicurves), [ES20] (on orbifolds), [EMM21], [AH21a] (effective counting), [EPS20] (general metric and length), [RS19] (geodesic current). To those readers wishing to learn more about this topic, we strongly recommend the expository survey [AH22] and the book [ES22].

Theorem 1.1.1 provides a model of *random* multi-geodesics.

Generally speaking, a model of a random structure is basically a probability measure on the moduli space M of objects under consideration. If one believes that "all multi-geodesic are born equal", then the uniform probability measure seems to be the most natural choice. However, the set of multi-geodesics on a hyperbolic surface is a countably infinite set, and no uniform probability measures exist on such sets, just like picking a positive integer "uniformly at random" is not possible. Nevertheless, a computer has no difficult in generating an integer between one and, say, a billion. The general idea is to consider certain *complexity function*, say, $h\colon M \to \mathbb{R}$, such that for any $x \in M$, $h(x) < \infty$, and for any given number n (which can be arbitrarily large), $|\{x \in M : h(x) \leq n\}| < \infty$. Thus, we obtain a family of uniform probability measures, indexed by n, supported on a larger and larger subset of M, and we are interested in

the asymptotic behaviors of random variables (measurable functions) on M, as n goes to infinity.

The *length function* seems to be a perfect choice. Let X be a closed hyperbolic surface, and R be a positive real number. The set $s_X(R)$ of multi-geodesics on X of length not exceeding R, is finite. From now on, we endow $s_X(R)$ with the uniform probability measure to make it a probability space. A multi-geodesic $m_1\gamma_1 + \cdots + m_k\gamma_k$ on X is said to be *non-separating* if $X \smallsetminus (\gamma_1 \cup \cdots \cup \gamma_k)$ is connected. Denote by $p_{X,R}$ the probability that a multi-geodesic chosen uniformly at random is non-separating. A natural question to ask is:

Question 1.1.3. What is the limit (if exists) of $p_{X,R}$ as $R \to \infty$? In other words, what is the probability of a random multi-geodesic on X is non-separating?

Another one is:

Question 1.1.4. How many components does a random multi-geodesic on X have?

The two questions above concern only the topology of a multi-geodesic, and ultimately comes down to the following one:

Question 1.1.5. What is the probability that a random multi-geodesic has a given topological type?

The product structure of the constant before R^{6g-6} in the formula (1.1) leads immediately to the following surprising fact: $\lim_{R\to\infty} |s_X(\gamma, R)|/|s_X(R)|$ exists, and is equal to $c(\gamma)/b_g$. This answers Question 1.1.5: the probability that a random multi-geodesic has the same topological type than that of γ is equal to $c(\gamma)/b_g$, and this probably does not depend on the hyperbolic metric X. So we may talk about "random multicurves" rather than random multi-geodesics. Thus, all above-mentioned questions boil down to the these *frequencies* $c(\gamma)$. For example, Mirzakhani managed to compute some of these numbers and proved

Proposition 1.1.6 ([Mir08b]). *On a closed hyperbolic surface of genus* 2*, the probability that a random simple closed geodesic is separating equals to* 1/49*.*

In general, for any $g \geq 2$, there exists $p_g \in \,]0, 1[$ and a probability distribution K_g on $\{1, \dots, 3g-3\}$ such that a random multi-geodesic on a hyperbolic of genus g has a p_g chance of being non-separating and the number of its components is distributed according to K_g.

Nevertheless, the calculation of p_g and K_g quickly becomes almost unmanageable as g increases. However, although things are super complicated when g is large, but it turns out that when g is super large, things become much simpler. In particular, people have found pretty beautiful answers (Theorem 1.4.4) to the two following questions:

Question 1.1.7. What is the probability that a random multi-geodesic on a large genus hyperbolic surface is separating? More precisely, what can we say about p_g as $g \to \infty$?

Question 1.1.8. How many components does a random multi-geodesic on a large genus hyperbolic surface? More precisely, what can we say about K_g when $g \to \infty$?

What about the *geometry* of a random multi-geodesic? The decomposition of a multi-geodesic $\gamma = m_1\gamma_1 + \cdots + m_k\gamma_k$ into components $m_i\gamma_i$, $1 \leq i \leq k$, allows us to write the total length $\ell_X(\gamma)$ of γ as the sum of the lengths of its components. Thus, γ defines a vector, on the standard infinite-dimensional simplex $\Delta_1^\infty \coloneqq \{(x_1, x_2, \dots) \in [0, 1]^{\mathbb{Z}_{\geq 1}} : x_1 + x_2 + \cdots = 1\}$,

$$\hat{\ell}_X^{\downarrow}(\gamma) \coloneqq \frac{1}{\ell_X(\gamma)} \left(m_1\ell_X(\gamma_1), \dots, m_k\ell_X(\gamma_k)\right)^{\downarrow} \in \Delta_1^\infty$$

where $(x_1, x_2, \dots)^{\downarrow}$ stands for the rearrangement of $(x_1, x_2, \dots)$ in descending order; for example, $(1, 0, 2, 4)^{\downarrow} = (4, 2, 1, 0)$. Note that $\hat{\ell}_X^{\downarrow}(\gamma)$ does not depend on the labelling of the components of γ by $1, \dots, k$.

Question 1.1.9. What is the limiting distribution (if exists) of $\hat{\ell}_{X,R}^{\downarrow}$ as $R \to \infty$?

The limiting distribution of $\hat{\ell}_{X,R}^{\downarrow}$ depends a priori on X. Now let us sample a random hyperbolic surface X of genus g (in certain way), then pick a random multi-geodesic on X, and obtain a random variable $\hat{\ell}_g^{\downarrow}$ representing the shape of a random multi-geodesic on a random hyperbolic surface.

Question 1.1.10. What is the distribution of $\hat{\ell}_g^{\downarrow}$? What happens when $g \to \infty$?

The random variable $\hat{\ell}_g^{\downarrow}$ is rather intricate since the topology of multi-geodesics becomes tricky when the genus gets larger and larger (the number of topological types of multicurves on Σ_g grows super-exponentially as a function of g). Let us begin with a simpler question: what if we already know the topological type of the random multi-geodesic? More precisely, given an ordered multi-geodesic $(m_1\gamma_1, \dots, m_k\gamma_k)$, its *normalized length vector* is defined to be

$$\hat{\ell}_X(m_1\gamma_1, \dots, m_k\gamma_k) = \frac{1}{\ell_X(\gamma)} \cdot (m_1\ell_X(\gamma_1), \dots, m_k\ell_X(\gamma_k)) \in \Delta_1^{k-1}$$

where $\Delta_1^{k-1} := \{(x_1, \dots, x_k) \in \mathbb{R}_{\geq 0}^k : x_1 + \cdots + x_k = 1\}$ is the standard simplex of dimension $k-1$. Let us denote by $s_X(\gamma, R)$ be the set of multi-geodesics of the same topological type that γ on X of length at most R, and we equip it with the uniform probability measure. The random variable $\hat{\ell}_X$ on $s_X(\gamma, R)$ depends on γ and R, although the underlying map does not. We denote it by $\hat{\ell}_{X,\gamma,R}$ to emphasize these dependencies.

Question 1.1.11. What is the limiting distribution of $\hat{\ell}_{X,\gamma,R}$ as $R \to \infty$?

The study of the limiting distribution of $\hat{\ell}_{X,\gamma,R}$ was initiated by Mirzakhani in [Mir16], where she proves the following theorem:

Theorem 1.1.12 ([Mir16, Theorem 1.2])**.** *If $\gamma = (\gamma_1, \dots, \gamma_{3g-3})$ gives a pants decomposition of X, then $\hat{\ell}_{X,\gamma,R}$ converges in law to the Dirichlet distribution of order $3g-3$ with parameters $(2, \dots, 2)$, i.e., the limit distribution of $\hat{\ell}_{X,\gamma,R}$ admits density function $(6g-7)! \cdot x_1 \cdots x_{3g-3}$ with respect to the Lebesgue measure on the standard simplex Δ_1^{3g-4}. In other words, for any open subset U of Δ_1^{3g-4},*

$$\lim_{R\to\infty} \mathbb{P}(\hat{\ell}_{X,\gamma,R} \in U) = (6g-7)! \int_U x_1 \cdots x_{3g-3}\, \lambda(dx).$$

where λ is the Lebesgue measure on Δ_1^{3g-4}.

So the limiting distribution of $\hat{\ell}_{X,\gamma,R}$ does not depend on the hyperbolic metric when γ is a pants decomposition, and in fact, it never does; see Section 1.5. As a result, in Question 1.1.10, the procedure of choosing a random surface is redundant since the every hyperbolic surface has exactly the same statistics.

Remark 1.1.13. The study of random surfaces has a long history in physics and mathematics. The discrete combinatorial model, which goes under names as *maps*, *ribbon graphs*, *graphs embedded into* (or *drawn on*) *a surface*, has been extensively studied in the last years, especially in the

planar case, see e.g. [LG19] and references contained therein. One can construct a random hyperbolic surface by gluing ideal triangles [BM04], [Pet17], [BCP21], [SW22]; or fix a surface then look at its covers [MP20], [MNP22], [HM21]. Pioneered by [GPY11] [Mir13], the study of random hyperbolic surfaces sampled with respect to the Weil–Petersson measure has received increasing attention in recent years. Much effort has been deployed to investigate the large genus regime; see for instance [MP19], [NWX20], [PWX21], [DGZZ20b]. And in particular from spectral aspect; see e.g. [GLMST21], [Mon20], [MS20], [WX22], [LW21], [Hid21], [Rud22], and we recommend strongly [Mon21, Introduction] for an expository survey about this topic.

1.2 Random square-tiled surfaces

In this section, we consider random geometric (almost combinatoric) objects of different nature: random square-tiled surfaces.

A *square-tiled* surface (or an *origami*) is a connected closed surface obtained by gluing squares. Here, every square is isometric to the standard Euclidean square $[0,1]^2 \subset \mathbb{R}^2$; its four sides have names "up", "left", "down" and "right" in anticlockwise; the two parallel sides "left" and "right" are said to be *horizontal*, and the two sides "up" and "down" are said to be *vertical*. See Figure 1.2. When gluing squares, horizontal (resp. vertical) sides are glued together two-by-two, and we

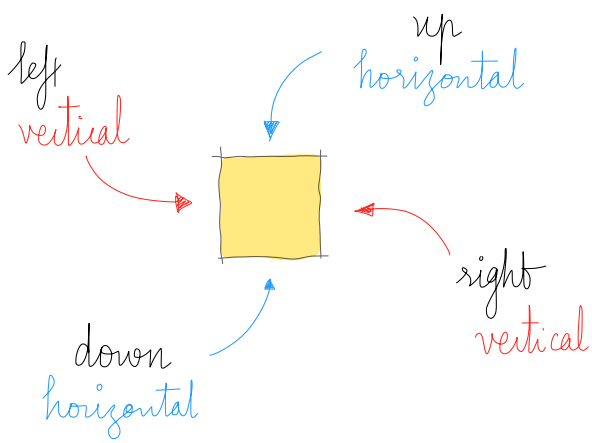

Figure 1.2: A square

assume that there are 4 or 6 squares around each vertex on the obtained surface. Those vertices surrounded by 6 squares are called *singularities*. Figure 1.3 (and Figure 1.4) is an example of a square-tiled surface of genus 2 with 4 singularities.

A square-tiled surface is foliated by horizontal flat geodesics. We say that a horizontal flat geodesic is *singular* if it goes through at at least a singularity, and *regular* if it does not. A square-tiled surface can be decomposed into singular leaves and *maximal horizontal cylinders* consisting of regular parallel horizontal flat geodesics. The cylinder decomposition of a square-tiled surface q can be encoded into a *weighted stable graph* (which is basically a graph with weighed edges),

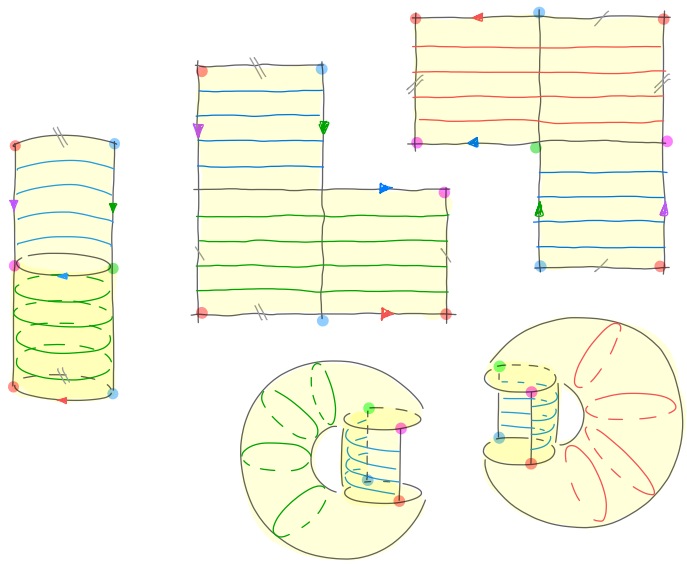

Figure 1.3: A square-tiled surface of genus 2 with 4 singularities and 3 cylinders

and each singular leaf of q defines a *ribbon graph*. See Figure 1.5 for the stable graph (on the left) of the square-tiled surface given by Figure 1.3 and Figure 1.5, and the two ribbon graphs (on the right) corresponding to the two vertices of the stable graph. See Section 3.6.1 for a precise description of this construction. We say that two square-tiled surface has the same *combinatorial type* if they share the same weight stable graphs.

There is only a finite number of square-tiled surfaces of genus g tiled by at most N squares. Denote the set of all such surfaces by $\mathcal{ST}_g(N)$. For a square-tiled surface q chosen uniformly at random from $\mathcal{ST}_g(N)$, let us address the following questions:

Question 1.2.1. How likely is it that q has one single singular leaf?

Question 1.2.2. How many cylinders does q have?

Question 1.2.3. What is the area of its biggest cylinder of q?

The last question can be formalized more precisely as follows. Let us consider the random variable $\hat{a}^{\downarrow}_{g,N} \colon \mathcal{ST}_g(N) \to \Delta_1^{\infty}$ that maps each square-tiled surface to the sequence whose i-th entry is the area of its i-th largest horizontal cylinder divided by the total area N.

Question 1.2.4. What is the limiting distribution (if exists) of $\hat{a}^{\downarrow}_{g,N}$ as $N \to \infty$? And what happens when $g \to \infty$?

1.3 Random permutations

A *permutation* of n letters is a bijection from $\{1, 2, \ldots, n\}$ to itself. The most straight-forward way to represent a permutation τ is to write down the map in the following table

i	1	2	3	4	5	6	7
$\tau(i)$	1	7	2	4	6	5	3

(1.2)

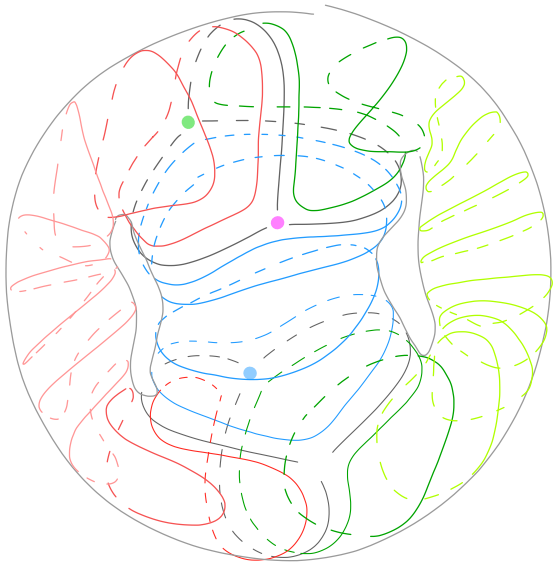

Figure 1.4: The square-tiled surface in Figure 1.3

A permutation can also be written as a product of disjoint *cycles*. For example,

$$\tau = (1)\,(2\,7\,3)\,(4)\,(5\,6) \tag{1.3}$$

where $1 \mapsto 1$, $2 \mapsto 7 \mapsto 3 \mapsto 2$, etc.

Let us denote by S_n the set of permutations of n letters, and consider random permutations. The most straight-forward (and the most fundamental) model of random permutation is the uniform measure on S_n, with respect to which $\mathbb{P}_n(\sigma) = 1/n!$ for all $\sigma \in S_n$. We shall consider a more general family of probability measures $\mathbb{P}_{\theta,n}$ parameterized by $\theta \in \mathbb{R}_{>0}$, defined by

$$\mathbb{P}_{\theta,n}(\sigma) := \frac{\theta^{K_n(\sigma)}}{Z_{\theta,n}}, \qquad \text{where} \quad Z_{\theta,n} := \sum_{\sigma \in S_n} \theta^{K_n(\sigma)} = \theta(\theta+1)\cdots(\theta+n-1)$$

where $K_n(\sigma)$ denotes the number of cycles in σ. This model, called *Ewens model*, was initially introduced by Ewens in the context of population genetics [Ewe72]. Note that we recover the uniform probability measure on S_n by taking $\theta = 1$, and in a random permutation of parameter θ, we expect find many (small) cycles if $\theta > 1$ is big, and few (big) cycles if $\theta < 1$ is close to 0.

Two natural questions to ask are

Question 1.3.1. When $n \to \infty$, how many cycles does a random Ewens permutation have?

Question 1.3.2. How big is its biggest cycle?

The second question can be refined as follows. Given a permutation σ of n letters, denote by $\hat{c}^{\downarrow}_{\theta,n}(\sigma)$ the n-vector whose i-th entry is the length of the i-th longest cycle in σ divided by n. For the permutation defined in the beginning of this section, we have $\hat{c}^{\downarrow}_{\theta,n}(\tau) = (3/7, 2/7, 1/7, 1/7)$.

Question 1.3.3. What is the limiting distribution (if exists) of $\hat{c}^{\downarrow}_{\theta,n}$ as $n \to \infty$?

The Ewens model of random permutations has been widely studied. For a detailed treatment,

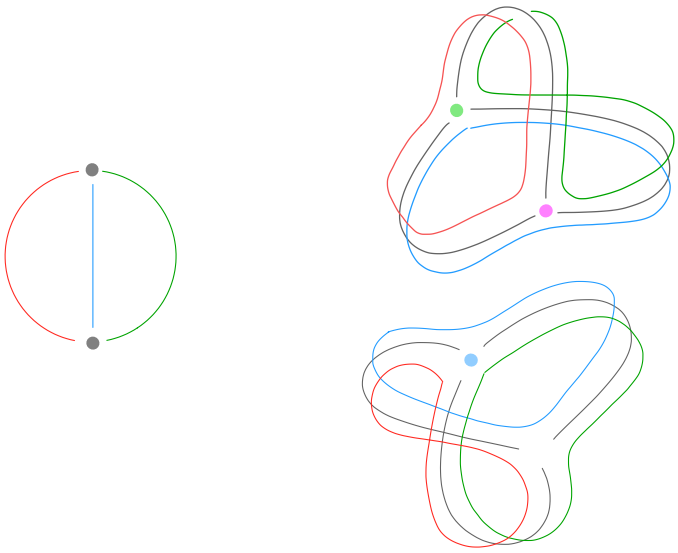

Figure 1.5: Associated stable graph and ribbon graphs

we refer the reader to [ABT03, Example 2.19, Chapter 4], [Ols11], and references therein.

Remark 1.3.4. In fact, we wish to consider a slightly more general measure defined as follows. Let $\theta = (\theta_i)_{i=1}^{\infty}$ be a sequence of non-negative real numbers, and let $K_{n,i}(\sigma)$ denotes the number of cycles of length i in $\sigma \in S_n$. Define

$$\mathbb{P}_{\theta,n}(\sigma) = \frac{1}{Z_{\theta,n}} \theta_1^{K_{n,1}(\sigma)} \cdots \theta_n^{K_{n,n}(\sigma)}, \qquad Z_{\theta,n} := \sum_{\sigma \in S_n} \theta_1^{K_{n,1}(\sigma)} \cdots \theta_n^{K_{n,n}(\sigma)}.$$

For us, the most interesting case of the sequence θ is $(\zeta(2i)/2)_{i\geq 1}^{\infty}$. However, $(\zeta(2i)/2)_{i=1}^{\infty}$ and $(1/2)_{i=1}^{\infty}$ give the same answers to all questions that we ask in this section, because, roughly speaking, the Riemann zeta function $\zeta(x)$ goes to 1 very rapidly as $x \to +\infty$.

1.4 Asymptotic random geometry

In this section we shall see that the behaviors of

1. a random multi-geodesic on a hyperbolic surface of genus g,
2. a random square-tiled surface of genus g,
3. an Evens(1/2) random permutation of g letters,

are similar as $g \to \infty$.

The first point is: random multi-geodesics on a hyperbolic surface and random square-tiled surfaces are very much the same thing, as we now explain.

Square-tiled surface can be seen as integral points in the principal stratum $\mathcal{Q}(1^{4g-4})$ of the moduli space of *holomorphic quadratic differentials* of genus g, and the *Masur–Veech volume*

$\mathrm{vol}_{\mathrm{MV}}(\mathcal{Q}(1^{4g-4}))$ of this stratum is closely related to the asymptotic counting of square-tiled surfaces. See Section 3.3. More precisely, Delecroix–Goujard–Zograf–Zorich showed in [DGZZ21] that $\mathrm{vol}_{\mathrm{MV}}(\mathcal{Q}(1^{4g-4}))$ can be written as a sum over all weighted stable graphs, and the contribution from a particular graph can be interpreted as the probability that a random square-tiled surface has the combinatorial type given by this graph (up tp a normalizing constant). On the other hand, the same graph represents also the topological type of a certain multi-geodesic γ, and it follows from the work of Mirzakhani [Mir08a], Monin–Telpukhovskiy [MT19], Arana-Herrera [AH20b], Erlandsson–Souto [ES22], Delecroix–Goujard–Zograf–Zorich [DGZZ21] that the contribution of the Masur–Veech volume from the graph is equal to $c(\gamma)$ up to an explicit normalizing constant depending only on the genus; see Chapter 4. The probability that a random multi-geodesic has the topological type determined by a given weighted graph is equal to the probability that a random square-tiled surface has the combinatorial type given by the same graph.

multi-geodesic $m_1\gamma_1 + \cdots + m_k\gamma_k$	square-tiled surface
component $m_i\gamma_i$ of the multi-geodesic	cylinder C_i
multiplicity m_i	height of C_i
length $\ell_X(\gamma_i)$	perimeter of C_i
length $m_i\ell_X(\gamma_i)$	area of C_i
a connected component in $X \smallsetminus \{\gamma_1, \ldots, \gamma_k\}$	a singular leaf
non-separating	all singularities sit on the same leaf

Table 1.1: Multi-geodesics versus square-tiled surfaces

Remark 1.4.1. The Masur–Veech volumes play an important role in the study of the dynamics and the geometry of billiards in rational polygons, interval exchange transformations, etc. In particular, it is closely related to the Siegel–Veech constants, and the Lyapunov exponents of the Hodge bundle along the Teichmüller geodesic flow. See, e.g., [EMZ03], [MZ08], [EKZ14], [Gou15], [AEZ16], [DZ20]. The approach for the calculation of Masur–Veech volumes via square-tiled surfaces counting was proposed independently by Eskin–Masur, and Kontsevich–Zorich [Zor02], and these volumes has been extensively studied in recent years from different points of views; see [EO01], [EO06], [EOP08], [GM20] (Hurwitz theory, quasi-modularity), [Sau18], [Sau21], [CMSZ20], [CMS+19] (intersection theory), [AEZ16], [Agg20], [Agg21], [DGZZ20a], [DGZZ21] (combinatorics) [ABC+19] (topological recursion), etc.

Now the problem is: how can we compute these numbers ($c(\gamma)$ or Masur–Veech volumes)? Quite surprisingly, it turns out that these numbers can be expressed in terms of *intersection numbers between psi-classes*, which classical objects in algebraic geometry.

Remark 1.4.2. Based on her work about the Weil–Petersson volume of the moduli space of hyperbolic surfaces with totally geodesic boundary components [Mir07c], Mirzakhani gave in [Mir08b] an explicit formula of $c(\gamma)$ involving intersection numbers of psi-classes. In [DGZZ21], the authors computed the Masur–Veech volume of the moduli space of quadratic differentials contributed from a stable graph, and got almost the same expression; counting results for metric ribbon graphs due to Kontsevich [Kon92] and Norbury [Nor13] play a crucial role in their work.

Roughly speaking, the i-th *tautological line bundle* $\mathcal{L}_i \to \overline{\mathcal{M}}_{g,n}$ over $\overline{\mathcal{M}}_{g,n}$ at $(C; x_1, \ldots, x_n) \in \overline{\mathcal{M}}_{g,n}$ is the cotangent space of C at the i-th marked point x_i, where $\overline{\mathcal{M}}_{g,n}$ denotes the Deligne–

Mumford compactification of the moduli space of smooth complex curves of genus g with n marked points. The i-th *psi-class*, denoted by ψ_i, is the first Chern class $c_1(\mathcal{L}_i) \in H^2(\overline{\mathcal{M}}_{g,n}, \mathbb{Q})$ of $\mathcal{L}_i$. Motivated by the equivalence of different models for 2-dimensional quantum gravity, Witten formulated a striking conjectured in [Wit91] claiming that a generating function of these intersection numbers obeys the *KdV equation* (named after Korteweg and de Vries, the KdV equation is a mathematical model of shallow-water waves), and as a consequence, all these intersection numbers can be completely determined recursively. Witten conjecture was soon proved by Kontsevich [Kon92], and several different proofs have been found later, e.g., [OP09], [KL07], [Mir07c], and [ABC+20]. Nevertheless, these frightening recursive relations are delicate to analyse.

The hope arises at infinity: in [ADG+20] the authors conjectured that as the genus goes to infinite, these elusive numbers admit a simple closed asymptotic expression. Several months after, the conjecture was turned into a theorem by Aggarwal [Agg21].

Based on Aggarwal's work, Delecroix–Goujard–Zograf–Zorich proved in [DGZZ20b] a series of results about large genus random square-tiled surfaces.

Theorem 1.4.3 ([DGZZ20b]). *Informally, as $g \to \infty$,*

1. *the probability that every maximal horizontal cylinder of a random square-tiled surface is primitive, goes to $1/\sqrt{2}$.*

2. *the probability that all conical singularities of a random square-tiled surfaces sit on the same leaf of the horizontal foliation and the same leaf of the vertical foliation, goes to 1.*

3. *the distribution of the number of cylinders of a random square-tiled surface is very well approximated by the Poisson distribution of parameter $\log(g)/2$.*

The dictionary between multi-geodesics frequencies and Masur–Veech volumes allows them to translate the results above into the following equivalent description about random multi-geodesics.

Theorem 1.4.4 ([DGZZ20b]). *Informally, on a closed hyperbolic surface of genus g, as $g \to \infty$,*

1. *the probability that a random multi-geodesic is primitive, goes to $1/\sqrt{2}$.*

2. *the probability that a random multi-geodesic is separating, goes to 0.*

3. *the distribution of the number of components of a random multi-geodesic is very well approximated by the Poisson distribution of parameter $\log(g)/2$.*

See Section 4.4 for more precise statements.

On the other hand, it is also known that, roughly speaking, the distribution of the number of cycles in a random permutation chosen according to the Ewens probability measure of parameter θ can be very well approximated by the Poisson distribution of parameter θn.

This thesis provides some further evidence that both a random multi-geodesic on a hyperbolic surface of genus g and a random square-tiled of genus g, behave like a Ewens $1/2$ random permutation of g elements.

1.5 Main results

The first step is to prove the following generation of Theorem 1.1.12 to multi-geodesics of arbitrary topological type.

Theorem 1.5.1 ([AH20a], [Liu19]). *Let $\gamma = (m_1\gamma_1, \dots, m_k\gamma_k)$ be an ordered multi-geodesic on X. The random variable $\hat{\ell}_{X,\gamma,R}$ converges in law to a random variable which admits a polynomial density with respect to the Lebesgue measure on Δ^{k-1} given by, up to a normalizing constant,*

$$(x_1, \dots, x_k) \mapsto \bar{F}_\gamma(x_1/m_k, \dots, x_k/m_k)$$

where $\bar{F}_\gamma$ is top-degree (homogeneous) part of the graph polynomial F_γ associated to γ defined by (3.7).

Remark 1.5.2. The density function $\bar{F}_\gamma$ can be expressed explicitly in terms of intersection numbers of psi-classes on $\overline{\mathcal{M}}_{g,n}$, and in particular it does not depends on the hyperbolic metric X. Furthermore, the work of Erlandsson, Parlier, and Souto [EPS20] suggests that the same statistics probably holds even for a large class of metrics.

Remark 1.5.3. This result was motivated the equivalent version for random square-tiled surfaces given in [DGZZ21]. Results very close to our was proved independently by Arana-Herrera [AH21b, AH20a].

The next step is to extend this result to a general random multi-geodesic (without any knowledge of its topological type), and study its large genus asymptotics. In a joint work with Vincent Delecroix, we proved

Theorem 1.5.4 ([DL22]). *The random variable $\hat{\ell}^{\downarrow}_{X,R}$ converges in distribution to a random variable L_g as $R \to \infty$, and L_g converges in distribution to the Poisson–Dirichlet distribution of parameter $1/2$ as $g \to \infty$.*

Remark 1.5.5. The same convergence holds for random *primitive* multi-geodesics.

The random square-tiled surfaces counterpart of this theorem is

Theorem 1.5.6 ([DL22]). *The random variable $\hat{a}^{\downarrow}_{g,N}$ converges in distribution to L_g as $N \to \infty$, and L_g converges in law to the Poisson–Dirichlet distribution of parameter $1/2$ as $g \to \infty$.*

The Poisson–Dirichlet distribution is a well-known probability distribution which arises naturally in a variety of contexts including arithmetic, combinatorics, Bayesian statistics, and population genetics; see [ABT03] and [PY97] and references therein. In particular, if a random sequence $V = (V_1, V_2, \dots) \in \Delta_1^\infty$ follows the Poisson–Dirichlet distribution of parameter θ, then the n-th moment of V_i is given explicitly by the following formula [SL66]

$$\mathbb{E}(V_i^n) = \frac{\Gamma(\theta+1)}{\Gamma(\theta+n)} \int_0^\infty \frac{\theta E_1(x)^{i-1}}{(i-1)!} x^{n-1} e^{-x-\theta E_1(x)}\, dx$$

where

$$E_1(x) := \int_x^\infty \frac{e^{-y}}{y}\, dy.$$

Thus Theorem 1.5.4 implies that, roughly speaking, the average lengths of the first three largest components of a random multi-geodesic on a large genus hyperbolic surface is approximately,

75.8%, 17.1%, and 4.9%, respectively, of the total length; and the variance of the length of the largest component is about 0.0037.

Remark 1.5.7. It is known that the random variable $\hat{c}^{\downarrow}_{\theta,n}$ representing the length spectrum of a random Ewens(θ) permutation defined in Section 1.3 converges in distribution to the Poisson–Dirichlet distribution of parameter θ as $n \to \infty$; see [ABT03, Section 5.4]. The case when $\theta = 1$ is due to Kingman [Kin77] and Vershik–Shmidt [VS77].

Remark 1.5.8. The same Poisson–Dirichlet pattern can be found in the integer factorization. Every integer $k \geq 2$ can be uniquely decomposed into a product of prime factors. Write $p_i(k)$ for the i-th largest prime factor in the k. For example,

$$744 = 31 \times 3 \times 2^3, \qquad p_1(744) = 31,\ p_2(744) = 3,\ p_3(744) = p_4(744) = p_5(744) = 2$$

and $p_6(744) = p_7(744) = \cdots = 1$ by convention. Now consider the uniform probability measure on $\{2, 3, \ldots, n\}$, and define a random variable by

$$\hat{\ell}^{\downarrow}_n : \{2, 3, \ldots, n\} \to \Delta^{\infty}_1, \qquad k \mapsto \frac{1}{\log(k)} \left(\log p_1(k), \log p_2(k), \ldots\right).$$

Billingsley proved in [Bil72] that $\hat{\ell}^{\downarrow}_n$ converges in distribution to the Poisson–Dirichlet law of parameter 1 as $n \to \infty$.

Remark 1.5.9. The convergence towards a Poisson–Dirichlet distribution says nothing about the small components (of length o(1)). In a work in progress with Vincent Delecroix, we are trying to prove that the number of small components (of order g^{-1}) in a random multi-geodesic on a hyperbolic surface of genus g, converges as $g \to \infty$ to a Poisson process with intensity (on $\mathbb{R}_{>0}$)

$$\frac{e^{-x}}{x} \sum_{n=1}^{\infty} (\cosh(x/n) - 1) = \frac{e^{-x}}{x} \sum_{n=1}^{\infty} \frac{\zeta(2n)}{(2n)!} x^{2n}. \tag{1.4}$$

In particular, for any $a, b \in \mathbb{R}_{\geq 0}$ with $a < b$, the number of components of (normalized) length in $[a/(6g-6), b/(6g-6)]$ converges to a Poisson distribution of parameter

$$\int_a^b \frac{e^{-x}}{x} \sum_{n=1}^{\infty} (\cosh(x/n) - 1) \tag{1.5}$$

as $g \to \infty$. Note that for a random permutation sampled according to the Ewens measure of parameter $\theta = (\zeta(2k)/2)_{k=1}^{\infty}$, the number of cycles of length between a and b (so the normalized length is in $[a/n, b/n]$) follows a Poisson distribution of parameter

$$\sum_{k=a}^{b} \frac{\theta_k}{k} = \sum_{k=a}^{b} \frac{\zeta(2k)}{2k},$$

which is very close to (1.5) (but different). Finally, it should be observed that the intensity of the Poisson process that appears in [MP19] (short primitive geodesics counting on a large genus random Weil–Petersson hyperbolic surface) and in [JL21] (short primitive loop counting in a large genus random unicellular map) are somewhat similar to the first term in (1.4). The presence of the higher-order terms is due to the fact that our multi-geodesics are not necessarily

primitive.

Remark 1.5.10. Our random multi-geodesics model does not see short multi-geodesics. For example, removing a finite number of multi-geodesics does not perturb the limiting distribution $\hat{\ell}_{g,\gamma}$ of $\hat{\ell}_{X,\gamma,R}$, and this somewhat explains why $\hat{\ell}_{g,\gamma}$ does not depend on the hyperbolic metric X. It would be interesting to study other models for random multi-geodesics (on a random hyperbolic surface), for instance, the one where the probability of picking γ is proportional to $\exp(-\ell_X(\gamma) \cdot t)$, where t is a parameter.

Remark 1.5.11. Our model of random square-tiled surfaces can be straightforwardly modified to study random square-tiled surfaces in other strata (quadratic and Abelian). Very recently, progress has been made in the minimum stratum $\mathcal{H}(2g-2)$ of Abelian differentials [Yak22] which generalizes a result in [Sau18]. The proof has a combinatorial nature where the bijection between unicellular maps and plane trees with distinguished vertices established in [Cha11] plays a crucial role. Yet it is not known if the counterparts of square-tiled surfaces in any stratum other than $\mathcal{Q}(1^{4g-4})$ exists in hyperbolic geometry. Another model of random square-tiled surfaces was studied by Shrestha in [Shr22] using representation theory of the symmetric group S_n.

Therefore, a typical multi-geodesic on a large genus hyperbolic surface looks like the following: it is non-separating, has about $\log(g)/2$ components, but we can only see about three components if we do not look carefully, since the three largest components take already about 98% of the total length of the multi-geodesic.

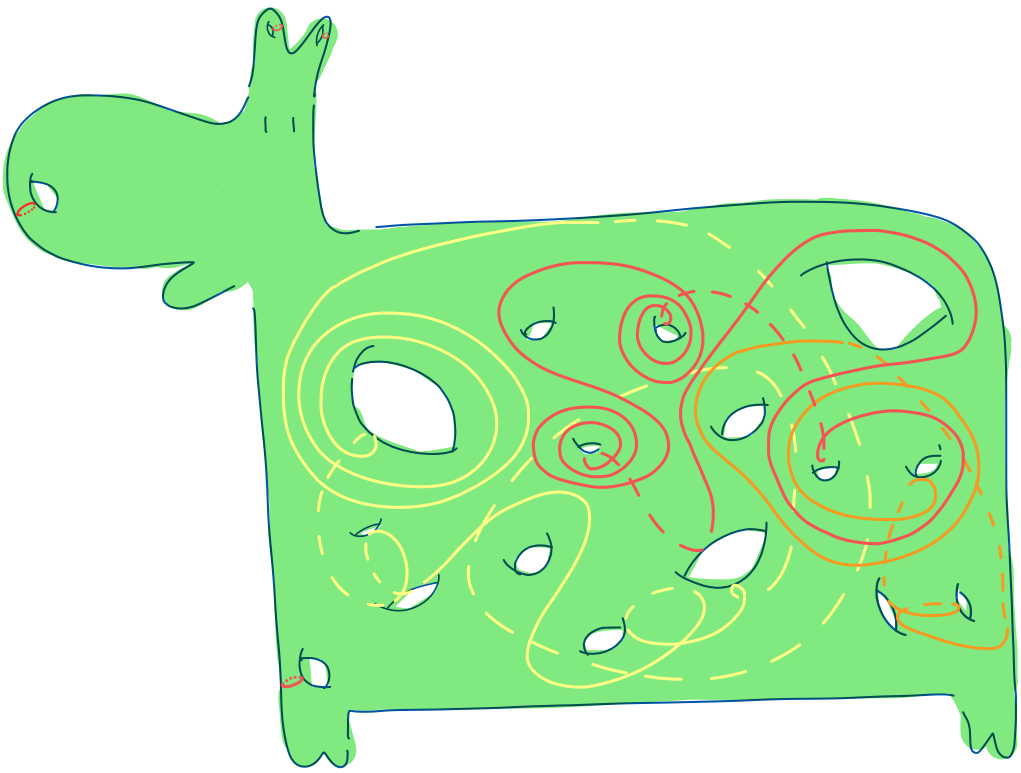

Figure 1.6: A typical multi-geodesic on a large genus hyperbolic surface

Chapter 2

Introduction (en français)

Cette thèse a pour objectif de tenter d'apporter une réponse à la question suivante :

À quoi ressemble-t-elle une courbe aléatoire sur une surface ?

Plus précisément, on s'intéresse aux multi-géodésiques aléatoires sur une surface hyperbolique, surtout lorsque le genre de la surface est grand.

2.1 Multi-géodésiques hyperboliques aléatoires

Géodésiques fermées sont d'une importance fondamentale en géométrie. Celles qui n'ont pas d'auto-intersection, appelées géodésiques fermées *simples*, sont centrales dans l'étude des surface hyperboliques et leur espace de modules depuis longtemps et pour diverses raisons. Par exemple, une surface se coupe proprement le long d'une géodésique fermée simple (voir e.g. [Mir07b], [ABO17]) ; il faudrait peut-être également mentionner que le complexe des courbes, qui ne sera pas du tout traité ici, joue un rôle crucial dans l'étude de l'espace de Teichmüller et du groupe modulaire.

Une *multi-géodésique* est une union disjointe de géodésiques fermées simples qui ne sont pas nécessairement primitives (dire qu'une courbe fermée est *primitive* signifie qu'elle n'est pas la n-ième itérée d'une courbe fermée avec $n > 1$), et dont les orientations sont ignorées. Une multi-géodésique s'écrit comme somme formelle $m_1\gamma_1 + \cdots + m_k\gamma_k$ où les m_i sont des entiers positifs et et les γ_i sont des géodésiques fermées simples primitives disjointes. Chaque m_i s'appelle la *multiplicité* de γ_i; et on dit que $m_i\gamma_i$ est une *composante connexe* or simplement une *composante* de γ_i ($k \leq 3g-3$ sur une surface close de genre g). Sur une surface hyperbolique close, toute courbe fermée non librement homotope à une point est librement homotope à une unique géodésique fermée, et pour tout $R > 0$, il n'existe qu'un nombre fini de géodésiques fermées de longueurs au plus R.

Le point de départ de cette thèse est le résultat suivant que Mirzakhani a démontré dans sa thèse

Théorème 2.1.1 ([Mir08b]). *Soit X une surface hyperbolique close de genre $g \geq 2$. Étant donné une multi-géodésique sur X, On note $s_X(\gamma, R)$ l'ensemble des multi-géodésiques sur X de*

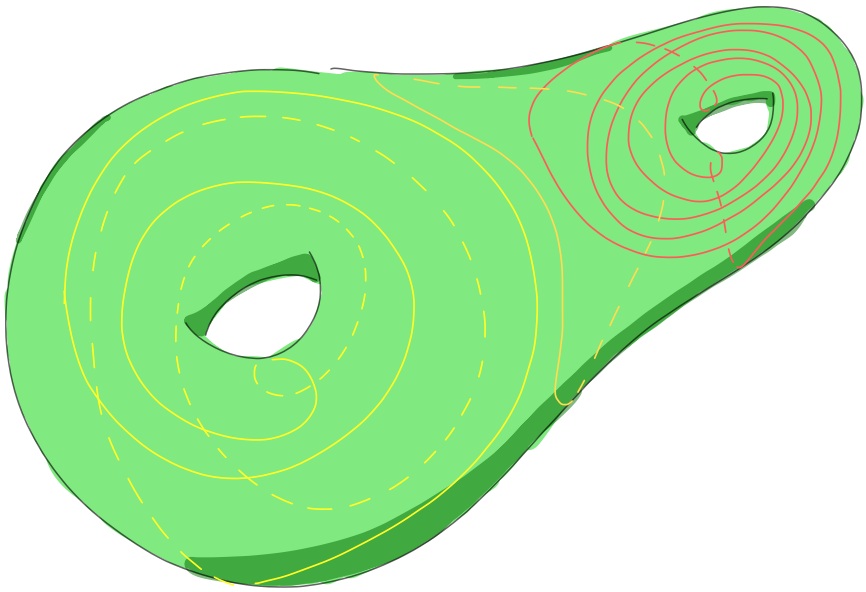

Figure 2.1: Une multi-géodésique

longueurs au plus R avec le même type topologique que celui de γ (deux courbes sont dites de la même type topologique si elles se trouvent dans le même orbite du groupe modulaire). Alors on a

$$|s_X(\gamma, R)| \sim c(\gamma) \cdot \frac{B(X)}{b_g} \cdot R^{6g-6} \tag{2.1}$$

lorsque $R \to \infty$, où $c(\gamma)$, $B(X)$ et b_g sont des constantes explicites qui ne dépendent que de γ, X, et g respectivement.

Remarque 2.1.2. Une géodésique fermée est rarement simple. D'après un théorème connu dû à Huber [Hub59] et Selberg, le nombre de géodésique fermée primitive (orientée) de longueur au maximum R est asymptotiquement équivalent à $\exp(R)/R$ lorsque $R \to \infty$. Un résultat similaire vaut pour les variétés hyperboliques, et est généralisé aux variétés compactes négativement courbées par Margulis dans sa thèse [Mar04], dans une approche complètement différente (systèmes dynamiques). Avant le travail révolutionnaire de Mirzakhani, des progrès sur le comptage des géodésiques fermées simples avaient été réalisés dans [BS85], [MR95a], [MR95b], et [Riv01]. Le comptage des géodésiques fermées sur une les surfaces hyperboliques fait encore aujourd'hui l'objet de recherches très actives. Pour des développements plus récents, voir e.g. [AH20a] (courbes remplissantes), [Mir16], [ES16] (multi-courbes non simples), [ES20] (sur les orbifolds), [EMM21], [AH21a] (comptage effectif), [EPS20] (métrique et longueur générales), [RS19] (courants géodésiques). Sur ce sujet, on recommande vivement l'article d'enquête [AH22] et le livre [ES22].

Théorème 2.1.1 prévoit un modèle de multi-géodésiques aléatoires.

De manière générale, un modèle d'une structure aléatoire est grosso modo une mesure de probabilité sur l'espace de modules M des objets considérés. Si l'on croit que « toutes les multi-géodésiques naissent égales », alors la mesure de probabilité uniforme semble être le choix le plus naturel. Toutefois, l'ensemble des multi-géodésiques sur une surface hyperbolique est un

ensemble infini dénombrable, et il n'existe pas de mesure de probabilité uniforme sur de tels ensembles. Essentiellement pour la même raison, il est impossible de choisir un entier positif « uniformément au hasard ». Néanmoins, un ordinateur n'a aucune difficulté à générer un entier compris entre un et, disons, un milliard. L'idée générale est de considérer une certaine *fonction de complexité*, notée $h\colon M \to R$, telle que pour tout $x \in M$, $h(x) < \infty$, et pour tout nombre n donné (qui pourrait être arbitrairement grand), $|\{x \in M : h(x) \leq n\}| < \infty$. On obtient ainsi une famille de mesures de probabilité uniformes, indexées par n, supportées sur un sous-ensemble de M de plus en plus grand, et on s'intéresse aux comportements asymptotiques des variables aléatoires (fonctions mesurables) sur M, lorsque n tend vers l'infini.

Il paraît que la fonction de *longueur* est un choix parfait. Soit X une surface hyperbolique, et soit R un nombre réel positif. L'ensemble $s_X(R)$ des multi-géodésiques sur X de longueur inférieure ou égale à R, est un ensemble fini. On munit désormais $s_X(R)$ de la probabilité uniforme, qui en fait un espace probabilisé. Une multi-géodésique $m_1\gamma_1 + \cdots + m_k\gamma_k$ sur X est dite *non-séparante* si $X \smallsetminus (\gamma_1 \cup \cdots \cup \gamma_k)$. Notons $p_{X,R}$ la probabilité qu'une multi-géodésique choisie uniformément au hasard soit non-séparante. Voici une question naturelle à poser :

Question 2.1.3. Quelle est la limite (si existe) de $p_{X,R}$ lorsque $R \to \infty$? Autrement dit, quelle est la probabilité qu'une multi-géodésique aléatoire sur X soit non-séparante ?

Et voici une autre :

Question 2.1.4. Combien de composantes une géodésique aléatoire sur X a-t-elle ?

Les deux questions ci-dessus ne concernent que la topologie d'une multi-géodésique, et on se ramène à la question suivante :

Question 2.1.5. Quelle est la probabilité qu'une multi-géodésique ait un type topologique donné ?

La structure de produit de la constante avant R^{6g-6} dans la formule (2.1) conduit immédiatement au fait surprenant suivant : $\lim_{R\to\infty} |s_X(\gamma, R)|/|s_X(R)|$ existe, et est égale à $c(\gamma)/b_g$. Cela réponde à Question 2.1.5 : la probabilité qu'une multi-géodésique aléatoire est de type topologique $[\gamma]$ est $c(\gamma)/b_g$, et cette probabilité ne dépend pas de la métrique hyperbolique X. On peut donc parler des « multi-courbes aléatoires » au lieu de multi-géodésiques aléatoires. Donc, toutes les questions ci-dessus se réduisent à ces *fréquences* $c(\gamma)$. Par exemple, Mirzakhani a pu calculer ses nombres et démontrer

Proposition 2.1.6. *Sur une surface hyperbolique de genre* 2*, la probabilité qu'une géodésique fermée simple aléatoire soit séparante est égale à* $1/49$.

En général, pour tout $g \geq 2$, il existe $p_g \in]0, 1[$ et une loi de probabilité K_g sur $\{1, \ldots, 3g-3\}$ telles qu'une multi-géodésique aléatoire ait p_g de chances d'être non-séparante et que le nombre de ses composantes soit distribué selon la loi K_g.

Néanmoins, le calcul de p_g et de K_g devient rapidement presque ingérable lorsque g croît. Cependant, bien que les choses soient super compliquées quand g est grand, il s'avère que quand g est super grand, les choses deviennent beaucoup plus simples. En particulier, les gens ont trouvé des réponses assez jolies (Théorème 2.4.4) aux questions suivantes :

Question 2.1.7. Quelle est la probabilité qu'une multi-géodésique aléatoire sur une surface hyperbolique de grand genre soit séparante ? Plus précisément, que peut-on dire de p_g lorsque $g \to \infty$?

Question 2.1.8. Combien de composantes une multi-géodésique aléatoire sur une surface hyperbolique de grand genre a-t-elle ? Plus précisément, que peut-on dire de K_g lorsque $g \to \infty$?

Et la *géométrie* ? La décomposition d'une multi-géodésique $\gamma = m_1\gamma_1 + \cdots + m_k\gamma_k$ en composantes connexes $m_i\gamma_i$, $1 \leq i \leq k$, nous permet d'écrire la longueur totale $\ell_X(\gamma)$ de γ comme la somme des longueurs de ses composantes. Ainsi, γ définit un vecteur, sur le simplexe standard de dimension infinie $\Delta_1^\infty := \{(x_1, x_2, \dots) \in [0,1]^{\mathbb{Z}_{\geq 1}} : x_1 + x_2 + \cdots = 1\}$

$$\hat{\ell}_X^{\downarrow}(\gamma) := \frac{1}{\ell_X(\gamma)} (m_1\ell_X(\gamma_1), \dots, m_k\ell_X(\gamma_k))^{\downarrow} \in \Delta_1^\infty$$

où $(x_1, x_2, \dots)^{\downarrow}$ signifie l'arrangement de $(x_1, x_2, \dots)$ dans l'ordre décroissant ; par exemple, $(1, 0, 2, 4)^{\downarrow} = (4, 2, 1, 0)$. On remarque $\hat{\ell}_X^{\downarrow}(\gamma)$ ne dépend pas de l'étiquetage des composantes de γ par $1, \dots, k$.

Question 2.1.9. Quelle est la loi limite (si existe) de $\hat{\ell}_X^{\downarrow}(\gamma)$ lorsque $R \to \infty$?

La loi limite dépende a priori de X. Maintenant, échantillonnons une surface hyperbolique aléatoire X de genre g (de certaine manière), puis choisissons une multi-géodésique aléatoire sur X. Ainsi, on obtient une variable aléatoire $\hat{\ell}_g^{\downarrow}$ qui représente la forme d'une multi-géodésique sur une surface hyperbolique aléatoire.

Question 2.1.10. Quelle est la loi de $\hat{\ell}_g^{\downarrow}$? Que se passe-t-il quand $g \to \infty$?

La variable aléatoire $\hat{\ell}_g^{\downarrow}$ est délicate car la topologie des multi-géodésiques devient intriquée quand le genre devient grand (le nombre de types topologiques des multi-géodésiques sur Σ_g croît de façon super-exponentielle en fonction de g). Commençons par une question plus simple : que se passe-t-il si on connaît déjà le type topologique de la multi-géodésique aléatoire ? Plus précisément, étant donné une multi-géodésique ordonnée $(m_1\gamma_1, \dots, m_k\gamma_k)$, son *vecteur des longueurs normalisé* est défini par

$$\hat{\ell}_X(m_1\gamma_1, \dots, m_k\gamma_k) = \frac{1}{\ell_X(\gamma)} \cdot (m_1\ell_X(\gamma_1), \dots, m_k\ell_X(\gamma_k)) \in \Delta_1^{k-1}$$

où $\Delta_1^{k-1} := \{(x_1, \dots, x_k) \in \mathbb{R}_{\geq 0}^k : x_1 + \cdots + x_k = 1\}$ est le simplex standard de dimension $k-1$.

Notons $s_X(\gamma, R)$ l'ensemble des multi-géodésiques de type topologique $[\gamma]$ sur X de longueur au plus R, et on le munit la probabilité uniforme. La variable aléatoire $\hat{\ell}_X$ sur $s_X(\gamma, R)$ dépend de γ et de R, bien que l'application sous-jacente n'en dépend pas. On le note $\hat{\ell}_{X,\gamma,R}$ afin d'insister ces dépendances.

Question 2.1.11. Quelle est la loi limite de $\hat{\ell}_{X,\gamma,R}$ lorsque $R \to \infty$?

L'étude de la loi limite de $\hat{\ell}_{X,\gamma,R}$ a été initialisée par Mirzakhani dans [Mir16], où elle a démontré le résultat suivant :

Théorème 2.1.12 ([Mir16, Theorem 1.2]). *Si $\gamma = (\gamma_1, \dots, \gamma_{3g-3})$ donne une décomposition en pantalons de X, alors $\hat{\ell}_{X,\gamma,R}$ converge en loi vers la loi de Dirichlet d'ordre $3g-3$ avec paramètres $(2, \dots, 2)$, i.e., la loi limite de $\hat{\ell}_{X,\gamma,R}$ admet pour densité $(6g-7)! \cdot x_1 \cdots x_{3g-3}$ par rapport à la mesure de Lebesgue sur le simplexe standard Δ_1^{3g-4}. Autrement dit, pour toute partie ouverte U de Δ^{3g-4},*

$$\lim_{R\to\infty} \mathbb{P}(\hat{\ell}_{X,\gamma,R} \in U) = (6g-7)! \int_U x_1 \cdots x_{3g-3}\, \lambda(dx).$$

où λ désigne la mesure de Lebesgue sur Δ^{3g-4}.

Donc la loi limite de $\hat{\ell}_{X,\gamma,R}$ ne dépend pas de la métrique hyperbolique lorsque γ est une décomposition en pantalons, et au fait, cela n'arrive jamais ; voir Section 2.5. Par conséquent, dans Question 2.1.10, la procédure de choix d'une surface aléatoire est redondante car toutes les surfaces hyperboliques ont exactement les même statistiques.

Remarque 2.1.13. L'étude des surface aléatoires a une longue histoire en physique et en mathématiques. Le modèle combinatoire, qui porte les noms comme *carte*, *graphes en rubans*, *graphes plongés dans* (ou *dessinés sur*) une surface, a été largement étudié au cours de ces dernières années, surtout dans le cas planaire ; voir, e.g., [LG19], et les références qui s'y trouvent. On peut construire une surface hyperbolique aléatoire en recolant des triangles idéaux [BM04], [Pet17], [BCP21], [SW22], ou fixant une surface et regardant ses revêtements [MP20], [MNP22], [HM21]. Lancée par [GPY11] [Mir13], l'étude des surfaces hyperboliques aléatoires échantillonnées par rapport à la mesure de Weil–Petersson a reçu une attention croissante au cours des dernières années. De nombreux efforts ont été déployés pour étudier le régime en grand genre ; voir par exemple [MP19], [NWX20], [PWX21], [DGZZ20b]. Et surtout d'un point de vue spectral ; voir e.g. [GLMST21], [Mon20], [MS20], [WX22], [LW21], [Hid21], [Rud22]. On recommande vivement au lecteur intéressé de lire [Mon21, Introduction] consacrée à ce sujet.

2.2 Surfaces à petits carreaux aléatoires

Une *surfaces à petits carreaux* (ou un *origami*) est une surface obtenue en collant des carrés. Ici, chaque carré est isométrique au carré euclidien standard $[0,1]^2 \subset \mathbb{R}^2$; ses quatre côtés sont étiquetés « haut », « gauche », « bas », et « droite » dans le sens inverse des aiguilles d'une montre ; les deux côtés parallèles « gauche » et « droite » sont dits « verticaux » et les deux côtés « haut » et « bas » sont dits « horizontaux » ; voir Figure 2.2. Quand on colle des carrés, les

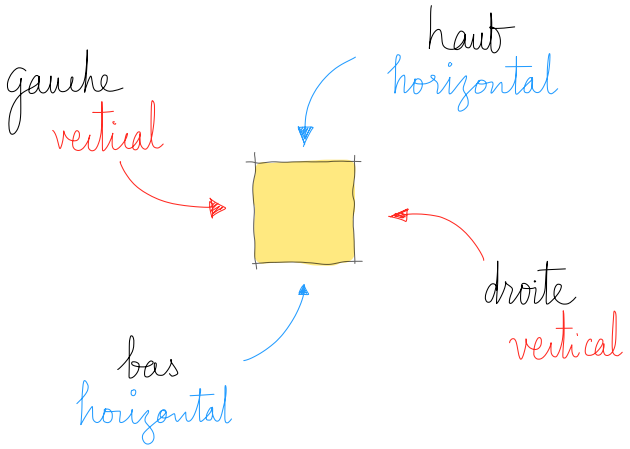

Figure 2.2: Un carré

côtés horizontaux (resp. verticaux) sont collés ensemble deux par deux, et on assume qu'il y a 4 ou 6 carrés autour de chaque sommet sur la surface obtenue. Ses sommets entourés par 6 carrés sont appelés *singularités*. Figure 2.3 (et 2.4) est un exemple d'une surfaces à petits carreaux

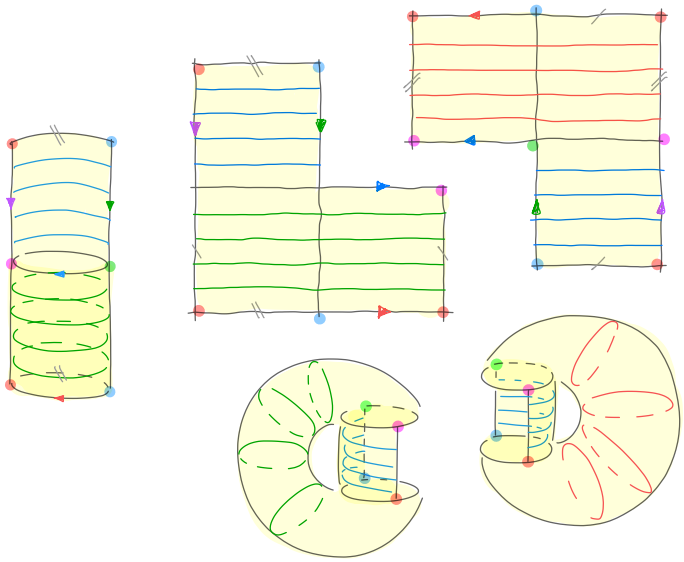

Figure 2.3: Une surfaces à petits carreaux de genre 2 avec 4 singularités et 3 cylindres

de genre 2 avec 4 singularités. Une surface à petits carreaux est feuilletée par des géodésiques plates horizontales. On dit qu'une géodésique plate horizontale est *singulière* si elle passe par au moins une singularité, et *régulière* si elle ne passe aucune singularité. Une surface à petits carreaux se décompose en feuilles singulières et *cylindres horizontaux maximaux* constitués de géodésiques plates horizontale régulières parallèles. La décomposition en cylindre d'une surface q à petits carreaux peut être encodée dans un *graphe stable pondéré* (qui est grosso modo un graphe dont les arrêtes sont pondérées), et chaque feuille singulière de q définit un *graphe en rubans* Voir Figure 2.5 pour le graphe stable (à gauche) associé à la surface à petits carreaux donnée par Figure 2.3 et Figure 2.4, et les deux graphes en rubans (à droite) qui correspondent aux deux sommets du graphe stable. Voir Section 3.6.1 pour une définition précise de cette construction. On dit que deux surfaces à petits carreaux ont le même *type combinatoire* si elles ont le même graphe stable pondéré associé.

Il n'y a qu'un nombre fini de surfaces à petits carreaux de genre g carrelées d'au plus N carrés. Notons $\mathcal{ST}_g(N)$ l'ensemble de toutes ces surfaces. Pour une surfaces à petits carreaux tirée uniformément au hasard de $\mathcal{ST}_g(N)$, on a envie de poser les questions suivantes :

Question 2.2.1. Quelle est la probabilité que q n'ait qu'une feuille singulière.

Question 2.2.2. Combien de cylindres q a-t-elle ?

Question 2.2.3. Quelle est l'aire du cylindre le plus grand de q ?

On peut reformuler la dernière question de la façon suivante. Considérons la variable aléatoire $\hat{a}^{\downarrow}_{g,N} \colon \mathcal{ST}_g(N) \to \Delta_1^\infty$ qui envoie chaque surface à la suite dont la i-ième entrée est l'aire de son i-ième plus grand cylindre horizontal, divisée par l'aire totale N de la surface.

Question 2.2.4. Quelle est la loi limite (si existe) de $\hat{a}^{\downarrow}_{g,N}$ lorsque $N \to \infty$? Et que se passe-t-il quand $g \to \infty$?

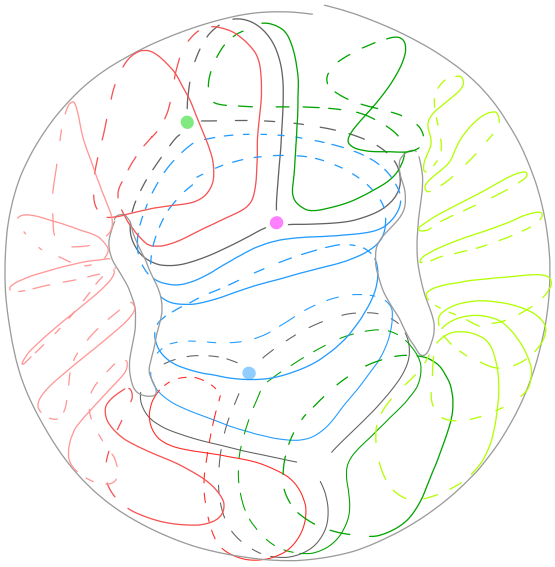

Figure 2.4: Surface obtenue

2.3 Permutations aléatoires

Une permutation aléatoire de n lettres est une bijection de $\{1, \ldots, n\}$ sur lui-même. La manière la plus directe de représenter une permutation est d'écrire cette application dans un tableau comme celui-ci :

i	1	2	3	4	5	6	7
$\tau(i)$	1	7	2	4	6	5	3

(2.2)

Une permutation s'écrit également comme un produit de *cycles* disjoints. Par exemple,

$$\sigma = (1)\,(2\,7\,3)\,(4)\,(5\,6) \tag{2.3}$$

où $1 \mapsto 1$, $2 \mapsto 7 \mapsto 3 \mapsto 2$, etc.

Notons S_n l'ensemble de toutes les permutations de $\{1, \ldots, n\}$, et considérons les permutations aléatoires. Le modèle plus simple (et la plus fondamentale) de permutations aléatoires est la mesure de probabilité uniforme S_n, pour laquelle $\mathbb{P}(\sigma) = 1/n!$ pour tout $\sigma \in S_n$. Nous considérerons une famille de mesures plus générale $\mathbb{P}_{\theta,n}$ paramétrée par $\theta > 0$, définie par

$$\mathbb{P}_{\theta,n}(\sigma) := \frac{\theta^{\mathrm{K}_n(\sigma)}}{Z_{\theta,n}}, \qquad \text{où} \quad Z_{\theta,n} := \sum_{\sigma \in S_n} \theta^{\mathrm{K}_n(\sigma)} = \theta(\theta+1)\cdots(\theta+n-1).$$

où $K_n(\sigma)$ désigne le nombre de cycles dans σ. Ce modèle, appelé le *modèle d'Ewens*, a été introduite par Ewens dans le contexte de la génétique des populations [Ewe72]. Remarquons que lorsque $\theta = 1$, $\mathbb{P}_{\theta,n}$ n'est rien d'autre que la probabilité uniforme sur S_n, et dans une telle permutation aléatoire de paramètre θ, on espère trouver beaucoup de (petits) cycles si $\theta > 1$ est grand, et peu de (grands) cycles si $\theta < 1$ est proche de 0. Voici deux questions naturelle à poser :

Question 2.3.1. Lorsque $n \to \infty$, combien de cycles une permutation aléatoire d'Ewens a-t-elle ?

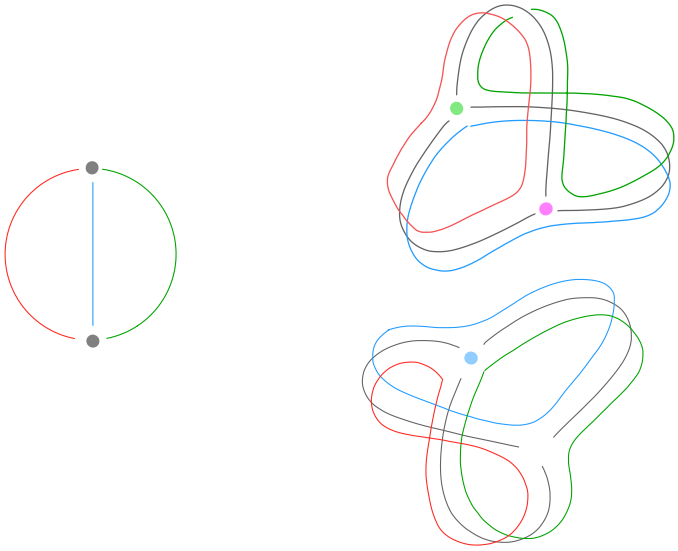

Figure 2.5: Graphe stable et graphes en rubans associés

Question 2.3.2. Quelle est la taille de son plus grand cycle ?

La deuxième question admet un raffinement suivant. Étant donné une permutation $\sigma \in S_n$, on note $\hat{c}^{\downarrow}_{\theta,n}(\sigma)$ le n-vecteur dont la i-ième composante est la taille de la i-ième plus long cycle dans σ divisé par n. Pour la permutation définie par (2.2) ou (1.3), on a $\hat{c}^{\downarrow}_{\theta,n}(\tau) = (3/7, 2/7, 1/7, 1/7)$.

Question 2.3.3. Quelle est la loi limite (si existe) de $\hat{c}^{\downarrow}_{\theta,n}$ lorsque $n \to \infty$?

Remarque 2.3.4. En fait, nous souhaitons considérer une mesure légèrement plus générale définie comme suit. Soit $\theta = (\theta_i)_{i=1}^{\infty}$ une suite de nombres réels non négatifs, et soit $K_{n,i}(\sigma)$ le nombre de cycles de longueur i dans $\sigma \in S_n$. On définit

$$\mathbb{P}_{\theta,n}(\sigma) = \frac{\theta_1^{K_{n,1}(\sigma)} \cdots \theta_n^{K_{n,n}(\sigma)}}{Z_{\theta,n}}, \qquad Z_{\theta,n} := \sum_{\sigma \in S_n} \theta_1^{K_{n,1}(\sigma)} \cdots \theta_n^{K_{n,n}(\sigma)}.$$

Le cas le plus intéressant de la suite θ est $(\zeta(2i)/2)_{i\geq 1}^{\infty}$. Cependant, $(\zeta(2i)/2)_{i=1}^{\infty}$ et $(1/2)_{i=1}^{\infty}$ donnent la même réponse à toutes les questions qu'on a posées dans cette section car grosso modo la fonction zêta de Riemann $\zeta(x)$ converge vers 1 très rapidement lorsque $x \to +\infty$.

2.4 Géométrie aléatoire asymptotique

Dans cette section, on verrons que

1. une multi-géodésique aléatoire sur une hyperbolique surface de genre g,
2. une surfaces à petits carreaux aléatoire de genre g,
3. une Evens(1/2) permutation aléatoire de g lettres,

se comportent de manière très similaire lorsque $g \to \infty$.

Le premier point est le suivant : les multi-géodésiques aléatoires sur une surface hyperbolique et les surfaces carrées aléatoires sont pratiquement la même chose, comme nous l'expliquons maintenant.

Les surfaces à petits carreaux peuvent être considérées comme des points d'entier dans la strate principale $\mathcal{Q}(1^{4g-4})$ de l'espace de modules des *différentielles quadratiques holomorphes* de genre g, et le *Masur–Veech volume* $\mathrm{vol}_{\mathrm{MV}}(\mathcal{Q}_g)$ de cette strate est étroitement liée au comptage asymptotique des surfaces à petits carreaux. Voir Section 3.3. Plus précisément, Delecroix–Goujard–Zograf–Zorich a démontré dans [DGZZ21] que $\mathrm{vol}_{\mathrm{MV}}(\mathcal{Q}(1^{4g-4}))$ s'écrit sous la forme d'une somme sur tous les graphes stables pondérés, et la contribution d'un graphe particulier s'interpréter comme la probabilité qu'une surface à petits carreaux aléatoire ait le type combinatoire donné par ce graphe (modulo une constante de normalisation). D'autre part, le même graphe représente aussi le type topologique d'une multi-géodésique γ, et d'après les travaux de Mirzakhani [Mir08a], de Monin–Telpukhovskiy [MT19], d'Arana-Herrera [AH20b], d'Erlandsson–Souto [ES22], et de Delecroix–Goujard–Zograf–Zorich [DGZZ21], la contribution de ce graphe dans le volume de Masur–Veech is égale à la fréquence $c(\gamma)$ modulo une constante de normalisation qui ne dépend que du genre de la surface ; voir Chapter 4. Donc, la probabilité qu'une multi-géodésique aléatoire ait le type topologique déterminé par un graphe stable pondéré est égale à la probabilité qu'une surface à petits carreaux ait le type combinatoire donné par le même graphe.

multi-géodésique $m_1\gamma_1 + \cdots + m_k\gamma_k$	surface à petits carreaux
composante $m_i\gamma_i$ de la multi-géodésique	cylindre C_i
multiplicité m_i	hauteur de C_i
longueur $\ell_X(\gamma_i)$	périmètre de C_i
longueur $m_i\ell_X(\gamma_i)$	aire de C_i
une composante connexe de $X \smallsetminus \{\gamma_1, \ldots, \gamma_k\}$	une feuille singulière
non-séparante	toutes les singularités se trouvent sur la même feuille

Table 2.1: Multi-géodésiques versus surfaces à petits carreaux

Remarque 2.4.1. Le volume de Masur–Veech joue un rôle important dans l'étude de la dynamique et la géométrie des billards dans les polygones rationnels, des échanges d'intervalles, etc. En particulier, elle est étroitement liée aux constantes de Siegel–Veech, et aux exposants de Lyapunov du fibré de Hodge le long le flot géodésique de Teichmüller. Voir, [EMZ03], [MZ08], [EKZ14], [Gou15], [AEZ16], [DZ20], etc. L'approche pour le calcul des volumes de Masur–Veech (abeliens et quadratiques) par comptage des surfaces à petits carreaux a été proposée indépendamment par Eskin–Masur, et Kontsevich–Zorich [Zor02], et ces volumes ont été intensivement étudiés ces dernières années sous de nombreux points de vue différents ; voir [EO01], [EO06], [EOP08], [GM20] (théorie de Hurwitz, quasi-modularité), [Sau18], [Sau21], [CMSZ20], [CMS+19] (théorie d'intersection), [AEZ16], [Agg20], [Agg21], [DGZZ20a], [DGZZ21] (combinatoire) [ABC+19] (récurrence topologique), etc.

Maintenant le problème est : comment pouvons-nous calculer ces nombres ($c(\gamma)$ ou volumes

Masur–Veech) ? De manière assez surprenante, il s'avère que ces nombres-là s'expriment en termes de *nombres d'intersection entre les classes de psi* qui font partie des objets classiques dans la géométrie algébrique.

Remarque 2.4.2. Se basant sur ses travaux sur le volume de Weil–Petersson de l'espace de modules des surfaces hyperboliques à bord totalement géodésique, Mirzakhani a donné dans [Mir07c] une formule explicite de $c(\gamma)$ qui fait intervenir des nombres d'intersection des classes de psi. Dans [DGZZ21], les auteurs ont calculé le volume de Masur–Veech de l'espace de modules des différentielles quadratiques contribué par un graphe stable, et a obtenu presque la même expression ; les résultats du comptage des graphes en rubans métriques dûs à Kontsevich [Kon92] et à Norbury [Nor13] jouent un rôle crucial.

De manière informelle, le i-ième *fibré en droites tautologique* $\mathcal{L}_i \to \overline{\mathcal{M}}_{g,n}$ sur $\overline{\mathcal{M}}_{g,n}$ au point $(C; x_1, \ldots, x_n) \in \overline{\mathcal{M}}_{g,n}$ est l'espace cotangent de C au i-ième point marqué x_i, où $\overline{\mathcal{M}}_{g,n}$ désigne la compactification de Deligne–Mumford de l'espace de modules des courbes complexes lisses de genre g avec n points marqués. La i-ième *class de psi*, notée par ψ_i, est la première classe de Chern $c_1(\mathcal{L}_i) \in H^2(\overline{\mathcal{M}}_{g,n}, \mathbb{Q})$ de $\mathcal{L}_i$. Motivé par l'équivalence entre les deux modèles de gravité quantique en dimension 2, Witten a formulé une conjecture frappante dans [Wit91] prétendant qu'une fonction génératrice de ces nombres d'intersection obéit à l'*équation de KdV* (nommée d'après Korteweg et de Vries, l'équation de KdV est un modèle mathématique des vagues en eau peu profonde), et par conséquent, tous ces nombres d'intersection sont complètement déterminés de manière récursive. La conjecture de Witten a été prouvée l'année d'après par Kontsevich [Kon92], et plusieurs preuves différentes ont été trouvées par la suite, e.g., [OP09], [KL07], [Mir07c], et [ABC+20]. Toutefois, ces effrayantes relations récursives sont délicates à analyser.

L'espoir se lève à l'infini : dans [ADG+20] les auteurs ont conjecturé que lorsque le genre tend vers l'infini, ces nombres obscurs admettent une expression asymptotique simple et fermée. Quelques mois plus tard, cette conjecture a été transformée en théorème par A. Aggarwal [Agg21].

En s'appuyant sur les travaux profonds d'Aggarwal's, Delecroix–Goujard–Zograf–Zorich a prouvé une série de résultats dans [DGZZ20b] sur les surfaces à petits carreaux aléatoires de grand genre :

Théorème 2.4.3 ([DGZZ20b]). *De manière informelle, lorsque $g \to \infty$,*

1. *la probabilité que chaque cylindre horizontal maximal d'une surface à petits carreaux soit primitif, est égale à $1/\sqrt{2}$.*

2. *la probabilité que toute les singularités coniques d'une surfaces à petits carreaux se trouvent sur la même feuille de la foliation horizontale et la même feuille de la foliation verticale, tend vers 1.*

3. *La distribution du nombre de cylindres d'une surface à petits carreaux est très bien approximée par la loi de Poisson de paramètre $\log(g)/2$.*

La dictionnaire entre les fréquences des multi-courbes et les volumes de Masur–Veech leur permet de traduire les résultats ci-dessus dans la description équivalente suivante des multi-géodésiques aléatoires :

Théorème 2.4.4 ([DGZZ20b]). *De manière informelle, lorsque $g \to \infty$,*

1. *la probabilité qu'une multi-géodésique aléatoire soit primitive, tend vers $1/\sqrt{2}$.*

2. *la probabilité qu'une multi-géodésique aléatoire soit séparante, tend vers* 0.

3. *la distribution du nombre de composantes d'une multi-géodésique aléatoire est très bien approximée par la loi de Poisson de paramètre* $\log(g)/2$.

D'autre part, il est également connu que, grosso modo, la distribution du nombre de cycles dans une permutation aléatoire choisie selon la mesure de probabilité d'Ewens de paramètre θ est très bien approximée par la loi de Poisson de paramètre θn.

Cette thèse fournit quelques preuves supplémentaires qu'une multi-géodésique aléatoire sur une surface hyperbolique aléatoire de genre g et une surfaces à petits carreaux aléatoire aléatoire de genre g, se comportent comme une Ewens 1/2 permutation aléatoire de g éléments.

2.5 Résultats principaux

La première étape consiste à montrer la généralisation suivante de Théorème 2.1.12 aux multi-géodésiques de type topologique arbitraire.

Théorème 2.5.1 ([Liu19]). *Soit* $\gamma = (m_1\gamma_1, \dots, m_k\gamma_k)$ *une multi-géodésique ordonnée sur* X. *La variable aléatoire* $\hat{\ell}_{X,\gamma,R}$ *converge en loi vers une variable aléatoire qui admet pour densité polynomiale par rapport à la mesure de Lebesgue sur* Δ^{k-1} *donnée par, une constante de normalisation près,*

$$(x_1, \dots, x_k) \mapsto \bar{F}_\gamma(x_1/m_k, \dots, x_k/m_k)$$

où $\bar{F}_\gamma$ *est la partie homogène de degré supérieure du polynôme* F_γ *associé à* γ *défini par* (3.7).

Remarque 2.5.2. La densité $\bar{F}_\gamma$ s'exprime explicitement en termes de nombres d'intersection de classes psi sur $\overline{\mathcal{M}}_{g,n}$, et en particulier elle ne dépend pas de la métrique hyperbolique X. En outre, les travaux de Erlandsson, Parlier, et Souto [EPS20] suggèrent que la même statistique est probablement valable pour une plus grande classe de métriques.

Remarque 2.5.3. Ce résultat est motivé par la version équivalente pour les surfaces à petits carreaux donnée dans [DGZZ21]. Des résultats très poches des nôtres ont été prouvées indépendamment par Arana-Herrera [AH21b, AH20a].

L'étape suivante consiste à étendre ce résultat à une multi-géodésique aléatoire générale (sans connaissance de son type topologique), et à étudier son asymptotique en grand genre. Dans un travail en commun avec Vincent Delecroix, nous avons prouvé que

Théorème 2.5.4 ([DL22]). *La variable aléatoire* $\hat{\ell}^{\downarrow}_{X,R}$ *converge en loi vers une variable aléatoire* L_g *lorsque* $R \to \infty$, *et* L_g *converge en loi vers la loi de Poisson–Dirichlet distribution de paramètre* 1/2 *lorsque* $g \to \infty$.

Remarque 2.5.5. La même convergence reste valable pour les multi-géodésiques primitive aléatoires.

La contre partie des surfaces à petits carreaux aléatoire du théorème précédent est

Théorème 2.5.6 ([DL22]). *La variable aléatoire* $\hat{a}^{\downarrow}_{g,N}$ *converge en loi vers* L_g *lorsque* $N \to \infty$, *et* L_g *converge en loi vers la loi de Poisson–Dirichlet de paramètre* 1/2 *lorsque* $g \to \infty$.

La loi de Poisson–Dirichlet est une loi de probabilité bien connue qui apparaît naturellement dans une variété de contextes notamment l'arithmétique, la combinatoire, les statistiques bayésiennes, et la génétique des populations ; voir [ABT03] et [PY97] et les références qui s'y trouvent. En particulier, si une suite aléatoire $V = (V_1, V_2, \dots) \in \Delta_1^\infty$ suit la loi de Poisson–Dirichlet de paramètre θ, alors le n-ième moment de V_i est donnée explicitement par la formule suivante [SL66] :

$$\mathbb{E}(V_i^n) = \frac{\Gamma(\theta+1)}{\Gamma(\theta+n)} \int_0^\infty \frac{\theta E_1(x)^{i-1}}{(i-1)!} x^{n-1} e^{-x-\theta E_1(x)}\, dx$$

où

$$E_1(x) := \int_x^\infty \frac{e^{-y}}{y}\, dy.$$

Donc, Théorème 1.5.4 implique que, grosso modo, les longueurs moyennes des trois premières plus longues composantes d'une multi-géodésique aléatoire sur une surface hyperbolique de grand genre sont approximativement, $75,8\%$, $17,1\%$, et $4,9\%$, respectivement, et la variance de la longueur de la plus grande composante est environ $0,0037$.

Remarque 2.5.7. Il est connu que la variable aléatoire $\hat{c}^{\downarrow}_{\theta,n}$ représentant la partition des longueurs des cycles d'une Ewens(θ) permutation aléatoire définie dans Section 2.3 converge en loi vers la loi de Poisson–Dirichlet de paramètre θ lorsque $n \to \infty$; voir [ABT03, Section 5.4]. Le cas où $\theta = 1$ est dû à Kingman [Kin77] et à Vershik–Shmidt [VS77].

Remarque 2.5.8. Le même phénomène Poisson–Dirichlet paraît dans la factorisation des nombres entiers. Chaque entier $k \geq 2$ se décompose de manière unique en un produit de nombres premiers. On écrit $p_i(k)$ pour le i-ième plus grand facteur premier dans k. Par exemple,

$$744 = 31 \times 3 \times 2^3, \qquad p_1(744) = 31,\ p_2(744) = 3,\ p_3(744) = p_4(744) = p_5(744) = 2$$

et $p_6(744) = p_7(744) = \cdots = 1$ par convention. Maintenant, considérons un entier aléatoire entre 2 et n choisi uniformément au hasard, et définissons la variable aléatoire par

$$\hat{\ell}^{\downarrow}_n : \{2, 3, \dots, n\} \to \Delta_1^\infty, \qquad k \mapsto \frac{1}{\log(k)} \left(\log p_1(k), \log p_2(k), \dots\right).$$

Billingsley démontre dans [Bil72] que cette variable aléatoire converge en loi vers la loi de Poisson–Dirichlet de paramètre 1.

Remarque 2.5.9. La convergence vers une loi de Poisson–Dirichlet ne dit rien sur les petites composantes (de longueur o(1)). Dans un travail en cours avec Vincent Delecroix, nous essayons de montrer que, le nombre de petites composantes (d'ordre g^{-1}) dans une multi-géodésique aléatoire sur une surface hyperbolique de genre g, lorsque $g \to \infty$, converge vers un processus de Poisson d'intensité (sur $\mathbb{R}_{>0}$)

$$\frac{e^{-x}}{x} \sum_{n=1}^\infty (\cosh(x/n) - 1) = \frac{e^{-x}}{x} \sum_{n=1}^\infty \frac{\zeta(2n)}{(2n)!} x^{2n}. \tag{2.4}$$

En particulier, pour tout $a, b \in \mathbb{R}_{\geq 0}$ avec $a < b$, le nombre de composantes de longueur (normal-

isée) dans $[a/(6g-6), b/(6g-6)]$ converge vers une loi de Poisson de paramètres

$$\int_a^b \frac{e^{-x}}{x} \sum_{n=1}^{\infty} (\cosh(x/n) - 1) \tag{2.5}$$

quand $g \to \infty$. Remarquons que pour une permutation aléatoire échantillonnée selon la mesure d'Ewens de paramètre $\theta = (\zeta(2k)/2)_{k=1}^{\infty}$, le nombre de cycles de longueur comprise entre a et b (longueur normalisée dans $[a/n, b/n]$) suit la loi de Poisson de paramètre

$$\sum_{k=a}^{b} \frac{\theta_k}{k} = \sum_{k=a}^{b} \frac{\zeta(2k)}{2k},$$

qui est très proche de (1.5). Finalement, il faut observer que l'intensité du processus de Poisson qui apparaît dans [MP19] (comptage de géodésiques primitives courtes sur une surface hyperbolique aléatoire à la Weil–Petersson de grand genre) et dans [JL21] (comptage de boucles primitives courtes dans une carte unicellulaire aléatoire de grand genre) en quelque sorte ressemblent au premier terme dans (1.4) ; et la présence des termes d'ordre supérieur est due au fait que nos multi-géodésiques ne sont pas nécessairement primitives.

Remarque 2.5.10. Notre modèle de multi-géodésiques aléatoires ne voit pas de multi-géodésique courtes. Par exemple, si on enlève un nombre fini de multi-géodésiques, alors par exemple la loi limite $\hat{\ell}_{g,\gamma}$ de $\hat{\ell}_{X,\gamma,R}$ ne sera pas perturbée, et cela explique en partie pourquoi $\hat{\ell}_{g,\gamma}$ ne dépend pas de la métrique hyperbolique X. Il serait intéressant de regarder d'autres modèles de multi-géodésiques aléatoires (sur une surface aléatoire), par exemple, celui où la probabilité de choisir γ est proportionnelle à $\exp(-\ell_X(\gamma) \cdot t)$, où t est un paramètre.

Remarque 2.5.11. Notre modèle de surfaces à petits carreaux aléatoires peut être modifié de façon directe pour étudier les surfaces à petits carreaux aléatoires dans d'autres strates (quadratiques et abéliennes). Très récemment, des progrès ont été réalisés dans la strate minimale $\mathcal{H}(2g-2)$ des différentielles abéliennes [Yak22] qui généralise un résultat dans [Sau18]. La démonstration est de nature combinatoire où la bijection entre les cartes unicellulaires et les arbres planaires avec des sommets distingués établie dans [Cha11] joue un rôle crucial. Pourtant, on ne sait pas si les contreparties des surfaces à petits carreaux dans toute strate antre que $\mathcal{Q}(1^{4g-4})$ existent en géométrie hyperbolique. Un autre modèle de surfaces à petits carreaux a été étudié par Shrestha dans [Shr22] en utilisant la théorie des représentations du groupe symétrique S_n.

Donc, une multi-géodésique typique sur une surface hyperbolique de grand genre ressemble à la suivante : elle est non-séparante, a environ $\log(g)/2$ composantes, mais nous ne verrons qu'environ trois composantes si nous ne regardons pas attentivement, puisque elles prennent déjà environ 98% de la longueur totale de la multi-géodésique.

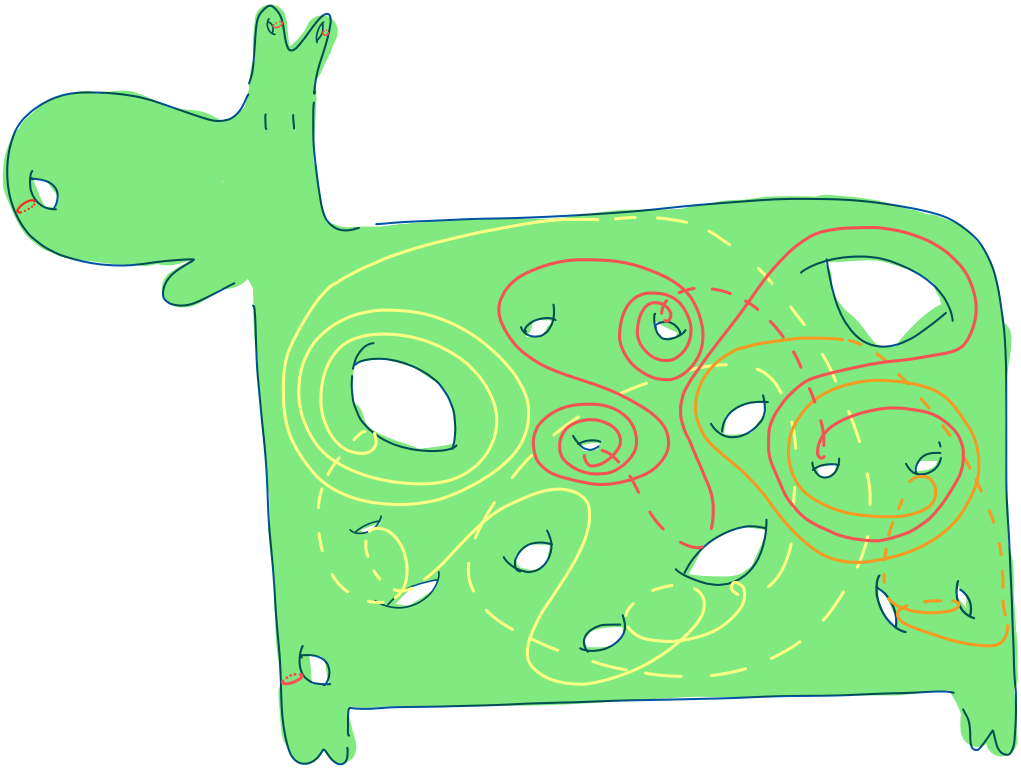

Figure 2.6: Une multi-géodésique typique sur une surface hyperbolique de grand genre

Chapter 3

Background

3.1 Topology of surfaces

Topological surfaces are smoothable and two smooth surfaces are diffeomorphic if and only if they are homeomorphic. Any connected, closed, orientable surface is homeomorphic to the connected sum of a sphere with a finite number of tori, and any connected, compact, orientable surface is homeomorphic to a closed surface removing a finite number of open disks with disjoint closures.

Throughout this thesis we shall use the symbol Σ_g to denote a connected, closed, oriented, smooth surface of genus $g \geq 2$, and denote by $\Sigma_{g,n}$ a connected closed oriented smooth surface of genus g with n boundary circles labeled by $\{1, \ldots, n\}$ with $2g - 2 + n > 0$.

Later we shall talk about hyperbolic surfaces homeomorphic to $\Sigma_{g,n}$ with totally geodesic boundary components of given lengths. If a boundary component has length equal to zero (and is therefore a *cusp*) then the corresponding circle boundary of $\Sigma_{g,n}$ should be though of as a puncture.

3.2 Hyperbolic geometry of surfaces

A *hyperbolic surface* is a 2-dimensional Riemannian manifold whose metric is complete and has constant sectional curvature equal to -1. There is a unique (up to isometry) simply connected hyperbolic surface, called the *hyperbolic plane* $\mathbb{H}^2$. So equivalently, a hyperbolic surface can be defined to be a surface locally modelled on $\mathbb{H}^2$. The hyperbolic plane has many models. For example, the *Poincaré half-plane model*, which is the upper-half plane $\{z = x + yi \in \mathbb{C} : x \in \mathbb{R},\ y \in \mathbb{R}_{>0}\}$ equipped with the metric

$$\frac{dx^2 + dy^2}{y^2}.$$

Example 3.2.1 (Fuchsian groups)**.** A trivial example of a hyperbolic surface is the hyperbolic plane $\mathbb{H}^2$. The group of orientation-preserving isometries of $\mathbb{H}^2$ is isomorphic to $\mathrm{PSL}(2, \mathbb{R})$ which acts on $\mathbb{H}^2$ by *Möbius transformations*

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z := \frac{az + b}{cz + d}.$$

A *Fuchsian group* is a discrete subgroup of $\mathrm{PSL}(2,\mathbb{R})$. A subgroup of $\mathrm{PSL}(2,\mathbb{R})$ is Fuchsian if and only if it acts properly discontinuously on $\mathbb{H}^2$. Non-trivial examples can be obtained as quotients of $\mathbb{H}^2$ by actions of Fuchsian groups. The *modular group* $\Gamma(1) \coloneqq \mathrm{PSL}(2,\mathbb{Z})$, and

$$\Gamma(k) \coloneqq \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{PSL}(2,\mathbb{Z}) : \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \bmod k \right\}$$

where $k \geq 2$, are typical examples of Fuchsian groups. Note that $X(k) \coloneqq \mathbb{H}^2/\Gamma(k)$ is non-compact for any $k \geq 1$. Examples of cocompact Fuchsian group can be obtained by some arithmetic constructions (Fuchsian groups *derived from a division quaternion algebra*). The *Klein quartic* (a compact Riemann surface of genus 3 with automorphism group isomorphic to $\mathrm{PSL}(2,\mathbb{F}_7)$, which is the lowest genus Hurwitz surface) is a thereby-obtained surface. For a detailed discussion of Fuchsian groups, see [Kat92].

Example 3.2.2 (Pairs of pants)**.** Unlike in Euclidean geometry, on $\mathbb{H}^2$ there exist geodesic hexagons whose all interior angles are right angles. Further, given $a, b, c \in \mathbb{R}_{\geq 0}$, there exists a unique (up to isometry) right-angled geodesic hexagon H with consecutive sides A, A', B, B', C, C' such that the lengths of A, B, C are a, b, c respectively (see [Bus92, Theorem 2.4.2]). If we take two copies of H and glue them along the A', B', C' sides using isometries (reflections), we get a hyperbolic surface which is topologically a sphere with three holes, called a *pair of pants*. It turns out that for any given triple $(L_1, L_2, L_3) \in \mathbb{R}^3_{\geq 0}$, up to isometry, there exists a unique surface hyperbolic surface of genus 0 with 3 labeled totally geodesic boundary components of lengths L_1, L_2, and L_3 respectively (see [Bus92, Theorem 3.1.7]).

Example 3.2.3 (Tori with one cusp)**.** The hyperbolic tori with a cusp are in natural bijection with the modular curve $\mathbb{H}^2/\mathrm{PSL}(2,\mathbb{Z})$. Unlike higher-dimensional hyperbolic manifolds, hyperbolic metric on a surface can be deformed, and the hyperbolic metrics on Σ_g form a *moduli space*.

3.2.1 Deformation spaces

Consider the set of orientation-preserving homeomorphisms $\phi\colon \Sigma_g \to X$ where X is an oriented complete hyperbolic surface of genus g. Two such homeomorphisms $\phi_1\colon \Sigma_g \to X_1$ and $\phi_2\colon \Sigma_g \to X_2$ are said to be *equivalent* if $\phi_2 \circ \phi_1^{-1}$ is isotropic to an isometry. The *Teichmüller space* (of hyperbolic surface of genus g), denoted by $\mathcal{T}(\Sigma_g)$ or simply $\mathcal{T}_g$, is the set of such equivalence classes.

Remark 3.2.4. We can define the *Teichmüller space of Riemann surfaces* in a similar way. It follows from the uniformization theorem and the Killing–Hopf theorem that these two spaces are in natural bijection. The Teichmüller space can therefore be studied via complex analysis. See [Hub06] and [IT92] for a detailed discussion.

Remark 3.2.5. The Teichmüller space can also be defined as a representation space for surface group $\pi_1(\Sigma_g)$ as follows. Consider the space of faithful representations $\pi_1(\Sigma_g) \to \mathrm{Isom}^+(\mathbb{H}^2) \simeq \mathrm{PSL}(2,\mathbb{R})$ with discrete image. The group $\mathrm{PSL}(2,\mathbb{R})$ acts on this space by conjugation, and the quotient, which is a connected component of the representation variety $\mathrm{Hom}(\pi_1(\Sigma_g), G)/G$ where $G = \mathrm{PSL}(2,\mathbb{R})$, is in natural bijection with $\mathcal{T}(\Sigma_g)$ (see [FM12, Proposition 10.2]). For a more general discussion, see [Lab13, Chapeter 5].

Let $\mathrm{Homeo}^+(\Sigma_g)$ denote the group of self-homeomorphisms of Σ_g that preserve the orientation, and let $\mathrm{Homeo}_0(\Sigma_g)$ denote its subgroup of homeomorphisms isotropic to the identity. The *mapping class group*, denoted by $\mathrm{Mod}(\Sigma_g)$ or simply Mod_g, is the quotient group $\mathrm{Mod}_g \coloneqq \mathrm{Homeo}^+(\Sigma_g)/\mathrm{Homeo}_0(\Sigma_g)$.

The group $\mathrm{Homeo}^+(\Sigma_g)$ acts (properly discontinuously; see [FM12, Theorem 12.2]) from the right on $\mathcal{T}_g$ by precomposition, and $\mathrm{Homeo}_0(\Sigma_g)$ acts trivially. The *moduli space*, denoted by $\mathcal{M}(\Sigma_g)$ or $\mathcal{M}_g$, is the quotient $\mathcal{T}_g/\mathrm{Mod}_g$.

The Teichmüller space $\mathcal{T}_{g,n}(L_1,\dots,L_n)$ and moduli space $\mathcal{M}_{g,n}(L_1,\dots,L_n)$ of oriented complete hyperbolic surfaces of genus g with n (labeled) totally geodesic boundary components of lengths $L_1,\dots,L_n \geq 0$ respectively can be defined in a similar manner.

3.2.2 Curves

A closed curve is said to be *simple* if it does not intersect itself. A *multicurve* is a finite multiset of disjoint simple closed curves. A multicurve is *ordered*, or *labeled*, if its underlying set is. We will often write an ordered multicurve γ as an list $(m_1\gamma_1,\dots,m_k\gamma_k)$, and its unlabeled counterpart as a formal sum $\overline{\gamma} = m_1\gamma_1+\cdots+m_k\gamma_k$ where $m_i \in \mathbb{Z}_{\geq 1}$ is the *multiplicity* of γ_i for all $1 \leq i \leq k$.

On a hyperbolic surface X, in the free homotopy class of each closed curve which is not freely homotopic to a point or a loop around a cusp, there exists a unique geodesic representative (see [FM12, Proposition 1.3]). Given a closed curve α on Σ_g (and therefore on X), we write $\ell_X(\alpha)$ for the length of the unique geodesic in its free homotopic class.

The group $\mathrm{Homeo}^+(\Sigma_g)$ acts on the set of closed curves on Σ_g by postcomposition, and the action of the subgroup $\mathrm{Homeo}_0(\Sigma_g)$ stabilizes sets of curves in the same free homotopy class. Thus the mapping class group Mod_g acts on the set of free homotopy classes of closed curves on Σ_g. We say that two closed curves α and β have the same *topological type* if they (more precisely, their free homotopy classes) lie in the same Mod_g-orbit.

Given an ordered multicurve $\gamma = (m_1\gamma_1,\dots,m_k\gamma_k)$, the three following subgroups of the mapping class group Mod_g will be useful later:

- $\mathrm{Stab}(\overline{\gamma})$ which fixes the multicurve $\overline{\gamma} = m_1\gamma_1+\cdots+m_k\gamma_k$ (but the γ_i's can be permutated),
- $\mathrm{Stab}(\gamma)$ which fixes every γ_i for all $1 \leq i \leq k$,
- $\mathrm{Stab}^+(\gamma)$ which fixes every γ_i and its orientation for all $1 \leq i \leq k$.

The notions of length and topological type extend naturally to multicurves.

3.2.3 Fenchel–Nielsen coordinates

Recall that a pair of pants is a surface that is homeomorphic to $\Sigma_{0,3}$, in other words, a sphere with three holes. Hyperbolic pairs of pants are building blocks of hyperbolic surface. Roughly speaking, any hyperbolic surface, compact or with totally geodesic boundary components, can be constructed by gluing hyperbolic pairs of pants. A little more precisely, if the two boundary components of two pairs of pants (might be the same) have the same length, then we can glue them together get a new surface, and when we glue two boundary components, there is a choice to make, namely, the position of the one boundary component relative to the other. Indeed, every hyperbolic surface can be obtained in this way.

A *pants decomposition* of $\Sigma_{g,n}$ is a set of disjoint simple closed curves $\{\alpha_1, \dots, \alpha_{3g-3+n}\}$ on $\Sigma_{g,n}$ such that $\Sigma_{g,n} \smallsetminus \{\alpha_1, \dots, \alpha_{3g-3+n}\}$ is a disjoint union of pairs of pants. Fix an ordered pants decomposition $(\alpha_1, \dots, \alpha_{3g-3+n})$ of $\Sigma_{g,n}$. Given $X \in \mathcal{T}_{g,n}(L_1, \dots, L_n)$ (or $\mathcal{T}_g$), we can associate for each α_i two parameters: the length of $\ell_{\alpha_i}(X) \in \mathbb{R}_{>0}$, and the *twist parameter* $\tau_{\alpha_i}(X) \in \mathbb{R}$ corresponding to how much one turns before gluing two pairs of pants along α_i (we actually need something more than an ordered pants decomposition to determine these parameters. For example, an initial configuration. Or a set of *seams*; see [FM12, Section 10.6.1] for more details). These $6g - 6 + 2n$ parameters are called *Fenchel–Nielsen coordinates*. The map

$$\mathcal{T}_{g,n}(L_1, \dots, L_n) \to (\mathbb{R}_{>0} \times \mathbb{R})^{3g-3+n}, \qquad X \mapsto (\ell_{\alpha_i}(X), \tau_{\alpha_i}(X))_{i=1}^{3g-3+n}$$

is a homeomorphism (see, [FM12, Section 10.6] or [Hub06, Theorem 7.6.3]. It is actually a real analytic diffeomorphism; see [Bus92, Theorem 6.3.2] for a proof in the case $n = 0$).

3.2.4 Weil–Petersson volumes

The Teichmüller space of closed Riemann surfaces of genus at least 2 can be endowed with a complex structure via the *Bers embedding*, and the its cotangent space at X is naturally isomorphic to the space $\mathcal{Q}(X)$ of holomorphic quadratic differential on X; see [Hub06, Chapter 6]. The Hermitian inner product on $\mathcal{Q}(X)$ defined by

$$\langle q_1, q_2 \rangle := \int_X \frac{\overline{q_1} q_2}{\rho^2}, \qquad \forall\, q_1, q_2 \in \mathcal{Q}(X),$$

where ρ is the hyperbolic metric on X, gives rise a metric on the Teichmüller space, called the *Weil–Petersson metric*. Taking the imaginary part of this metric, we obtain a 2-form $\tilde{\omega}$, which turns out to be closed (in other words, the Weil–Petersson metric is Kähler); see [Hub06, Theorem 7.7.2]. The 2-form $\tilde{\omega}$ and its rescaling $\omega := \tilde{\omega}/2$ are so-called *Weil–Petersson symplectic form*, or simply *Weil–Petersson form*. [Gol84]

Theorem 3.2.6 ([Wol85])**.** *Let $\{\alpha_1, \dots, \alpha_{3g-3}\}$ be a pants decomposition of Σ_g. The Weil–Petersson form ω has the following expression*

$$\sum_{i=1}^{3g-3} d\ell_{\alpha_i} \wedge d\tau_{\alpha_i}. \tag{3.1}$$

In particular, it is invariant under the action of the mapping class group.

This construction does not work well for Teichmüller space of hyperbolic surfaces with totally geodesic boundary components with prescribed lengths, however, the symplectic form on $\mathcal{T}_{g,n}(L_1, \dots, L_n)$ defined by (3.1) is still $\mathrm{Mod}(\Sigma_{g,n})$-invariant; see [Wol13, Lecture 4].

Every symplectic form induces a volume form, and hence a measure (the *Weil–Petersson measure*). The *Weil–Petersson volume* of the moduli space $\mathcal{M}_{g,n}(L_1, \dots, L_n)$ is defined to be

$$V_{g,n}(L_1, \dots, L_n) := \frac{1}{(3g-3+n)!} \int_{\mathcal{M}_{g,n}(L_1,\dots,L_n)} \omega^{\wedge(3g-3+n)}$$

This volume turns out to be finite. In fact, it is a polynomial in $L_1^2, \dots, L_n^2$, by the following fundamental result is due to Mirzakhani.

Theorem 3.2.7 ([Mir07c]). *The Weil–Petersson volume $V_{g,n}(L_1,\dots,L_n)$ is a symmetric polynomial in $L_1^2,\dots,L_n^2$ of degree $3g-3+2n$. More precisely,*

$$V_{g,n}(L_1,\dots,L_n) = \sum_{\substack{(d_0,d_1,\dots,d_n)\in\mathbb{Z}_{\geq 0}\\ d_0+d_1\cdots+d_n=3g-3+n}} \frac{(2\pi^2)^{d_0}}{2^{d_1+\cdots+d_n}d_0!d_1!\cdots d_n!}\left(\int_{\overline{\mathcal{M}}_{g,n}} \kappa_1^{d_0}\psi_1^{d_1}\cdots\psi_n^{d_n}\right) L_1^{2d_1}\cdots L_n^{2d_n}$$

where $\overline{\mathcal{M}}_{g,n}$ is the Deligne–Mumford compactification, $\psi_i \in H^2(\overline{\mathcal{M}}_{g,n},\mathbb{Q})$ is the i-th psi-class, and $\kappa_1 \in H^2(\overline{\mathcal{M}}_{g,n},\mathbb{Q})$ is the first Mumford class.

See Section 3.4 for (slightly) more precise definitions of ψ_i and κ_1.

3.2.5 Measured laminations

A *geodesic lamination*, or simply a *lamination*, on a closed hyperbolic surface X is a closed subset of X foliated by simple geodesics (not necessarily closed). A *measured geodesic lamination*, or simply a *measured lamination*, is a lamination together with a transverse measure of full support on it (roughly speaking, a transverse measure on a lamination is an assignment of weights to each leaf of the lamination). These notations was initially introduced by Thurston to complete the space of simple closed geodesics on X.

We denote by $\mathcal{ML}(X)$ the space of measured laminations on $X \in \mathcal{T}_g$. A topology can be assigned to $\mathcal{ML}(X)$ in several equivalent ways, for example, the topology generated by the family of semi-distances $d_\gamma(\lambda_1,\lambda_2) := |\lambda_1(\gamma)-\lambda_2(\gamma)|$ between $\lambda_1,\lambda_2 \in \mathcal{ML}(X)$ where γ runs over all simple closed curves on X, or via *geometric currents* (see [Bon88]). Multi-geodesics on X are dense in $\mathcal{ML}(X)$.

Using *train tracks*, $\mathcal{ML}(X)$ can be given an atlas consisting of charts to $\mathbb{R}^{6g-6}$, and the transition functions are piecewise integral linear. In other words, $\mathcal{ML}(X)$ is an piecewise integral linear real manifold of dimension $6g-6$. Since $\mathcal{ML}(X_1)$ and $\mathcal{ML}(X_2)$ are in one-to-one correspondence (via a homeomorphism between $\partial\widetilde{X}_1$ and $\partial\widetilde{X}_2$), sometimes we use the notation $\mathcal{ML}_g$ to denote the space of measured laminations without reference to any particular hyperbolic metric on Σ_g.

Further, the integral points in the charts of $\mathcal{ML}(X)$ are in natural bijection with the space $\mathcal{ML}_g(\mathbb{Z})$ of (free homotopy classes of) integral multicurves on Σ_g.

The length function defined on the set of multicurves can be extended to a continuous function $\ell_X\colon \mathcal{ML}(X) \to \mathbb{R}_{>0}$. As a result, we have $\ell_X(r\cdot\lambda) = r\cdot\ell_X(\lambda)$ and $\ell_{Xh}(h^{-1}\lambda) = \ell_X(\lambda)$ for all $r \in \mathbb{R}_{>0}$, $\lambda \in \mathcal{ML}(X)$, and $h \in \mathrm{Mod}(X)$.

Write δ_x for the Dirac measure centered at x. The following limit

$$\mu_{\mathrm{Th}} := \lim_{R\to\infty} \frac{1}{R^{6g-6}} \sum_{\gamma\in\mathcal{ML}_g} \delta_{\gamma/R}$$

exists, and is called the *Thurston measure*. Thurston measure is mapping class group invariant, and homogeneous i.e. $\mu_{\mathrm{Th}}(r\cdot U) = r^{6g-6}\cdot\mu_{\mathrm{Th}}(U)$ for any $r \in \mathbb{R}_{>0}$ and $U \subset \mathcal{ML}_g$ measurable.

For more details on this subject, the reader should consult the survey [Bon01]; [Thu02, Chapter 8, 9], [Kap01, Chapter 11], [PT07, Chapter 3], [PH92] are also recommended.

Measured laminations allow us to define a new distance on $\mathcal{T}_g$ as follows. For any $X_1, X_2 \in \mathcal{T}_g$,

set

$$d(X_1, X_2) := \sup_{\lambda \in \mathcal{ML}_g} \log \frac{\ell_{X_1}(\lambda)}{\ell_{X_2}(\lambda)}.$$

The *Thurston distance* between X_1 and X_2 is defined by

$$d_{\mathrm{Th}}(X_1, X_2) := \max\{d(X_1, X_2), d(X_1, X_2)\}.$$

The *Thurston distance ball* centered at $X \in \mathcal{T}_g$ of radius ϵ is defined to be

$$\mathbb{B}_X(\epsilon) := \{Y \in \mathcal{T}_g : d_{\mathrm{Th}}(X, Y) \leq \epsilon/2\}.$$

The reason for this choice of radius is that, for small ϵ, for example, $0 < \epsilon < 1$, we have

$$e^{\epsilon} < 1 + 2x \qquad \text{and} \qquad e^{-\epsilon} > 1 - 2x.$$

We have therefore, for any $\lambda \in \mathcal{ML}_g$ and any $Y \in \mathbb{B}_X(\epsilon)$,

$$(1 - \epsilon) \cdot \ell_X(\lambda) \leq \ell_Y(\lambda) \leq (1 + \epsilon) \cdot \ell_X(\lambda). \tag{3.2}$$

Thurston distance balls are well-defined on $\mathcal{M}_g^{\gamma}$, and on $\mathcal{M}_g$, since the Thurston distance is Mod_g-invariant.

3.2.6 Earthquake

Let γ be a simple closed geodesic on $X \in \mathcal{T}_g$. Roughly speaking, we can deform the hyperbolic metric by cutting X along γ, turning counterclockwise by distance t, and regluing. A precise definition can be found in e.g. [Bus92, Section 3.3]. Thus each simple closed curve γ on Σ_g defines a *twist flow* $\mathrm{tw}_\gamma \colon \mathcal{T}_g \times \mathbb{R} \to \mathcal{T}_g$ on the Teichmüller space. For $t \in \mathbb{R}$, we write $\mathrm{tw}_\gamma^t(\bullet)$ for $\mathrm{tw}_\gamma(\bullet, t)$.

The twist flow can be extended to any measured lamination $\lambda \in \mathcal{ML}_g$, and we have $(\mathrm{tw}_\lambda^t(X)) \cdot h = \mathrm{tw}_{h^{-1}\lambda}^t(Xh)$ and $\mathrm{tw}_{r\lambda}^t(X) = \mathrm{tw}_\lambda^{rt}$ for all $t \in \mathbb{R}$, $h \in \mathrm{Mod}_g$, and $r \in \mathbb{R}_{>0}$.

Let $\mathcal{PT}_g := \mathcal{T}_g \times \mathcal{ML}_g$ be the bundle of measured laminations over the Teichmüller space, and let $\mathcal{P}^1\mathcal{T}_g := \{(X, \lambda) \in \mathcal{PT}_g : \ell_X(\lambda) = 1\}$ be the unit sphere bundle of $\mathcal{PT}_g$ with respect to the length function.

The mapping class group acts on $\mathcal{PT}_g$ from the right via $(X, \lambda) \cdot h := (X \cdot h, h^{-1} \cdot \lambda)$. This action is well-defined on $\mathcal{P}^1\mathcal{T}_g$ since it preserves the length function $\ell(X, \lambda) := \ell_X(\lambda)$. Write $\mathcal{PM}_g := \mathcal{PT}_g/\mathrm{Mod}_g$ and $\mathcal{P}^1\mathcal{M}_g := \mathcal{P}^1\mathcal{T}_g/\mathrm{Mod}_g$.

The *earthquake flow* tw^t on $\mathcal{PT}_n$ is defined by

$$\mathrm{tw}^t(X, \lambda) := (\mathrm{tw}_\lambda^t(X), \lambda).$$

The earthquake flow commutes with the action of the mapping class group, and therefore descends to $\mathcal{PM}_g$, and to $\mathcal{P}^1\mathcal{M}_g$ (since the earthquake preserves the length function).

The Thurston measure on $\mathcal{ML}_g$ induces a measure on $\{\lambda \in \mathcal{ML}_g : \ell_X(\lambda) = 1\}$ in the following way: let $U \subset \{\lambda \in \mathcal{ML}_g : \ell_X(\lambda) = 1\}$ be an open subset. The *Thurston measure* of U is defined to be

$$\mu_{\mathrm{Th}}\{s \cdot \lambda \in \mathcal{ML}_g : \lambda \in U,\ s \in [0, 1]\}.$$

The measure ν_g on $\mathcal{P}^1\mathcal{T}_g$ defined by

$$\nu_g(U) := \int_{\mathcal{T}_g} \mu_{\mathrm{Th}}\{s\lambda \in \mathcal{ML}_g : (X,\lambda) \in U,\ s \in [0,1]\}\, dX$$

for any open subset $U \subset \mathcal{P}^1\mathcal{T}_g$, is invariant both under the earthquake flow (since μ_{WP} is) and under the action of the mapping class group (since μ_{Th} and μ_{WP} are), and hence descends to a measure on $\mathcal{P}^1\mathcal{M}_g$ that (by abuse of notation) we shall also denote by ν_g. The total mass of ν_g

$$b_g = \int_{\mathcal{M}_g} B(X)\, dX$$

where $B(X) := \mu_{\mathrm{Th}}\{\lambda \in \mathcal{ML}_g : \ell_X(\lambda) \leq 1\}$ is finite [Mir08a, Theorem 3.3].

The following result is fundamental.

Theorem 3.2.8 ([Mir08a])**.** *The earthquake flow on $\mathcal{P}^1\mathcal{M}_g$ is ergodic with respect to ν_g.*

We recommend [Wri18] for an expository survey on this topic.

3.3 Flat geometry of surfaces

We can also consider metrics on Σ_g with vanishing curvature, namely flat metrics. It follows from the Gauss–Bonnet theorem that smooth flat metrics exists on tori. However, flat metrics exist (and there are lots of them) if singularities are allowed. In this thesis, we consider only flat surfaces with the simplest types of singularities, namely *translation surfaces* and *half-translation surfaces*. Theses surfaces have been extensively studied over the past two decades from various aspects. The interested reader may care to consult e.g. [Zor06], [Wri15], [Zvo19], and [Mas22] for surveys on this topic.

3.3.1 (Half-)Translation surfaces

There are (at least) three equivalent definitions of a *translation surface*:

1. A pair (C,α) where C is a complex curve of genus g and α is a non-zero *Abelian differential* (a holomorphic one-form) on C;

2. A flat metric on Σ_g with *conical singularities* of cone angles with values in $2\pi\mathbb{Z}_{\geq 2}$;

3. A surface homeomorphic to Σ_g obtained from identifying two-by-two the edges of a finite collection of polygons in $\mathbb{C}$ via maps of the form $z \mapsto z + a$ for some $a \in \mathbb{C}$.

Similarly, there are (at least) three equivalent definitions of a *half-translation surface*:

1. A pair (C,q) where C is a complex curve of genus g and q is a non-zero *holomorphic quadratic differential* on C, i.e., a holomorphic section of the line bundle $K_C \otimes K_C$ where K_C is the canonical line bundle (the cotangent bundle in this case) of C;

2. A flat metric on Σ_g with conical singularities of cone angles with values in $\pi\mathbb{Z}_{\geq 3}$;

3. A surface homeomorphic to Σ_g obtained from identifying two-by-two the edges of a finite collection of polygons in $\mathbb{C}$ via maps of the form $z \mapsto \pm z + a$ for some $a \in \mathbb{C}$.

The equivalences between the above three definitions are not immediate. We encourage the reader to consult [DVH, Introduction] and [Wri15, Section 1] for a detailed discussion.

Example 3.3.1. If the collection of polygons in the third definition of a (half-)translation surface consists only squares isomorphic to $[0,1]^2 \subset \mathbb{R}^2 \simeq \mathbb{C}$, then the resulting surface is called a *square-tiled surface*.

3.3.2 Period coordinates

Let $\mathcal{H}_g$ denote the space of Abelian differentials of genus g, and let $\mathcal{Q}_g$ denote the space of holomorphic quadratic differentials of genus g. It follows from the Riemann–Roch theorem that any Abelian differential (resp. holomorphic quadratic differential) of genus g has, counted with multiplicity, $2g-2$ zeros (resp. $4g-4$ zeros). Thus $\mathcal{H}_g$ (resp. $\mathcal{Q}_g$) is naturally stratified, and the strata are indexed by partitions of $2g-2$ (resp. $4g-4$) corresponding to the orders of the zeros.

For example, $\mathcal{H}_2$ has 5 strata: $\mathcal{H}(4)$ (one zero of order 4), $\mathcal{H}(3,1)$ (one triple zero, one simple zero), $\mathcal{H}(2,2)$ (two double zeros), $\mathcal{H}(2,1,1)$ (one double zero, two simple zeros), and $\mathcal{H}(1,1,1,1)$ (all zeros are simple). From now on we shall focus on the *principle stratum* $\mathcal{H}(1,\dots,1) = \mathcal{H}(1^{2g-2})$ (resp. $\mathcal{Q}(1,\dots,1) = \mathcal{Q}(1^{4g-4})$) consisting of Abelian differentials (resp. quadratic differentials) with only simple zeros. The principal stratum is open and dense in $\mathcal{H}_g$ (resp. $\mathcal{Q}_g$).

The principle stratum $\mathcal{H}(1^{2g-2})$ is a piecewise integral linear complex orbiford of dimension $4g-3$. Here we sketch a proof which can be found in [Wri15, Proposition 1.15]. Let S be a closed (topological) surface with $2g-2$ marked points $p_1,\dots,p_{2g-2}$. A *marked Abelian differentials* is an Abelian differential $(C,\alpha) \in \mathcal{H}(1^{2g-2})$ together with a *marking*, namely a homeomorphism $\varphi\colon S \to C$ such that the i-th marked point is sent to the i-th zero of α for any $1 \le i \le 2g-2$. Two marked Abelian differentials (C_1,α_1,φ_1) and (C_2,α_2,φ_2) are said to be equivalent if $\varphi_2 \circ \varphi_1^{-1}\colon C_1 \to C_2$ is isotropic to a biholomorphism f and $f^*\alpha_2 = \alpha_1$. Write $\widetilde{\mathcal{H}}(1^{4g-4})$ for the space of equivalent classes of marked Abelian differentials. The mapping class group $\mathrm{Mod}(S)$ acts on $\widetilde{\mathcal{H}}(1^{4g-4})$ by precomposition with the marking, and the quotient is nothing but $\mathcal{H}(1^{4g-4})$.

Let $H^1(S,\{p_1,\dots,p_{2g-2}\};\mathbb{C})$ denote the first cohomology group of S relative to the marked points $\{p_1,\dots,p_{2g-2}\}$. The *period map* defined by

$$\widetilde{\mathcal{H}}(1^{2g-2}) \longrightarrow H^1(S,\{p_1,\dots,p_{2g-2}\};\mathbb{C}), \qquad [(C,\alpha,\varphi)] \longmapsto \left(\gamma \mapsto \int_{\varphi(\gamma)} \alpha\right)$$

is a local injection whose image is open. Thus, apart from the locus of (C,α) where $\mathrm{Aut}(C)$ is not trivial, $\mathcal{H}(1^{2g-2})$ is locally modelled on $H^1(S,\{p_1,\dots,p_{2g-2}\};\mathbb{C})$. Choosing a symplectic basis $\{\gamma_i\}_{i=1}^{2g}$ for the absolute homology group $H_1(S,\mathbb{Z})$, and one curve γ_{2g+i} from p_i to p_{2g-2} for each $1 \le i \le 2g-3$, we obtain a basis $\{\gamma_i\}_{i=1}^{4g-3}$ for the relative homology group $H_1(S,\{p_1,\dots,p_{2g-2}\};\mathbb{Z})$. The map

$$\widetilde{\mathcal{H}}(1^{2g-2}) \longrightarrow \mathbb{C}^{4g-3}, \qquad [(C,\alpha,\varphi)] \longmapsto \left(\int_{\varphi(\gamma_1)} \alpha, \dots, \int_{\varphi(\gamma_{4g-3})} \alpha\right)$$

defines a local coordinate system of $\widetilde{\mathcal{H}}(1^{2g-2})$.

Integral points in period coordinates defines square-tiled (translation) surfaces by

$$C \longrightarrow \mathbb{T}^2, \qquad z \longmapsto \int_{z_0}^{z} \alpha \mod \mathbb{Z} \oplus i\mathbb{Z}$$

where z_0 is a reference point which can be taken as, for example, the first zero of α.

A piecewise integral linear structure comes with a family of a one-parameter family of Lebesgue measures (each two of which differ by a scaling constant in $\mathbb{R}_{>0}$), among them being the *Masur–Veech measure*, with respect to which the canonical integral lattice in $\mathcal{H}(1^{2g-2})$ has covolume one.

The period coordinates and the Masur–Veech measure of the principal stratum (in fact of any stratum) of the moduli space of quadratic differentials can be defined in a similar (but a little more technical) way. For a detailed discussion, see for instance, [DGZZ16, Appendix A], [FM14, Section 2.3].

3.3.3 Masur–Veech volumes

Each holomorphic quadratic differential can be locally written in the form $f(z)\ dz \otimes dz$ where f is holomorphic function. A non-zero quadratic differential $(C, q) \in \mathcal{Q}(1^{4g-4})$ defines a flat metric $|q|$ on C (if $q = f(z)\ dz \otimes dz$ locally, then $|q| = |f(z)|\,(dx^2 + dy^2)$ locally). The area of C with respect to this metric

$$\operatorname{area}(C, q) \coloneqq \int_C |q|$$

is finite.

The volume of $\mathcal{Q}(1^{4g-4})$ with respect to the Masur–Veech measure is infinite (basically because the Lebesgue measure of a vector space is infinite). Nevertheless, by the independent work of H. Masur [Mas82] and W. Veech [Vee82], despite its non-compactnes, the volume of

$$\{(C, q) \in \mathcal{Q}(1^{4g-4}) : \operatorname{area}(X, q) \leq r\}$$

is finite. This gives rise to a notion of volume for the moduli space of quadratic differentials $\mathcal{Q}(1^{4g-4})$, called the *Masur–Veech volume*, defined by

$$\operatorname{vol}_{\mathrm{MV}}(\mathcal{Q}(1^{4g-4})) = 2(6g-6) \cdot \mu_{\mathrm{MV}}\{(C, q) \in \mathcal{Q}(1^{4g-4}) : \operatorname{area}(C, q) \leq 1/2\}.$$

Since $\mathcal{Q}(1^{4g-4})$ is open dense in $\mathcal{Q}_g$, sometimes we simply write $\operatorname{vol}_{\mathrm{MV}}(\mathcal{Q}_g)$ for $\operatorname{vol}_{\mathrm{MV}}(\mathcal{Q}(1^{4g-4}))$.

Since square-tiled (half-translation) surfaces correspond to integral points in $\mathcal{Q}(1^{4g-4})$. The Masur–Veech volume of $\mathcal{Q}(1^{4g-4})$ can be evaluated by counting square-tiled surfaces

$$\operatorname{vol}_{\mathrm{MV}}(\mathcal{Q}(1^{4g-4})) = 2(6g-6) \cdot \lim_{N \to \infty} \frac{|\mathcal{ST}(\mathcal{Q}(1^{4g-4}), 2N)|}{N^{6g-6}}$$

where $\mathcal{ST}(\mathcal{Q}(1^{4g-4}), 2N)$ denotes the set of square-tiled surfaces in $\mathcal{Q}(1^{4g-4})$ tiled by $2N$ squares.

3.4 Intersection theory on $\overline{\mathcal{M}}_{g,n}$

In this section, we (very) briefly discuss the intersection theory on $\overline{\mathcal{M}}_{g,n}$, the Delgine–Mumford compactification of the moduli space of smooth complex curves of genus g with n marked points

where $2g-2+n>0$. The read may wish to consult for instance [Zvo12], [LZ04, Chapter 4], [Vak08], [HM98] for a more detailed introduction.

Write $\mathcal{M}_{g,n}$ for the set of isomorphism classes of compact complex algebraic curves of genus g with n marked points. It turns out that $\mathcal{M}_{g,n}$ admits a natural structure of a smooth complex orbifold of dimension $3g-3+n$.

In [DM69], Deligne and Mumford compactifies $\mathcal{M}_{g,n}$ by adding *stable* curves (simplest singular curves). The resulting space is called the Deligne–Mumford compactification $\overline{\mathcal{M}}_{g,n}$, which is a compact smooth complex orbiford of dimension $3g-3+n$ which contains $\mathcal{M}_{g,n}$ as an open dense sub-orbiford.

has n *tautological line bundle* $\mathcal{L}_1,\dots,\mathcal{L}_n \to \overline{\mathcal{M}}_{g,n}$ over $\overline{\mathcal{M}}_{g,n}$ such that the fiber of $\mathcal{L}_i$ at $(C;x_1,\dots,x_n)\in\overline{\mathcal{M}}_{g,n}$ is the cotangent space of C at the i-th marked point x_i. The i-th *psi-class* ψ_i is the first Chern class $c_1(\mathcal{L}_i)\in H^2(\overline{\mathcal{M}}_{g,n},\mathbb{Q})$ of $\mathcal{L}_i$. Let $\pi\colon \overline{\mathcal{M}}_{g,n+1}\to\overline{\mathcal{M}}_{g,n}$ be the forgetful map. The *first Mumford class* κ_1 is defined to be $\kappa_1 := \pi_*(\psi_{n+1}^2)\in H^2(\overline{\mathcal{M}}_{g,n},\mathbb{Q})$. From now on, we shall focus on the psi-classes. The following notation is convenient:

$$\langle \tau_{d_1}\cdots\tau_{d_n}\rangle_g := \int_{\overline{\mathcal{M}}_{g,n}} \psi_1^{d_1}\cdots\psi_n^{d_n}.$$

It turns out all these numbers satisfy some recursive relations. The following two equations can be proved using standard algebro-geometric method.

Proposition 3.4.1. *For $2g-2+n>0$, we have*

- *String equation:*
$$\langle \tau_{d_1}\cdots\tau_{d_n}\tau_0\rangle_g = \sum_{i=1}^{n}\langle \tau_{d_1}\cdots\tau_{d_i-1}\cdots\tau_{d_n}\rangle_g.$$
- *Dilaton equation:*
$$\langle \tau_{d_1}\cdots\tau_{d_n}\tau_1\rangle_g = (2g-2+n)\cdot\langle \tau_{d_1}\cdots\tau_{d_n}\rangle_g.$$

See [Wit91] for more details.

It follows from a direct computation that $\langle\tau_0^3\rangle_0=1$, and $\langle\tau_1\rangle_1=1/24$ (see [Zvo12, Proposition 2.26]). The string equation together with the initial condition $\langle\tau_0^3\rangle_0=1$ completely determine all intersection numbers between psi-class on $\mathcal{M}_{0,n}$, where $n\geq 3$ (since $d_1+\cdots+d_n+d_{n+1}=n-2$ implies that there is at least one $d_i=0$), and we have

Corollary 3.4.2. *For $d_1+\cdots+d_n=n-3$ we have*

$$\langle \tau_{d_1}\cdots\tau_{d_n}\rangle_0 = \binom{n-3}{d_1,\dots,d_n} = \frac{(n-3)!}{d_1!\cdots d_n!}.$$

Similarly, the dilaton equation together with the string equation and the initial condition $\langle\tau_1\rangle_1=1/24$ determine all intersection numbers between psi-classes on $\mathcal{M}_{1,n}$ where $n\geq 1$.

Corollary 3.4.3. *For $d_1+\cdots+d_n=n$, we have*

$$\langle \tau_{d_1}\cdots\tau_{d_n}\rangle_1 = \frac{1}{24}\binom{n}{d_1,\dots,d_n}\left(1-\sum_{k=2}^{n}\frac{(k-2)!(n-k)!}{n!}\,e_k(d_1,\dots,d_n)\right)$$

where e_k is the k-th elementary symmetric function defined by

$$e_k(d_1,\dots,d_n) := \sum_{i_1<\cdots<i_k} d_{i_1}\cdots d_{i_k}.$$

However, it is not surprising that we cannot deduce all intersection numbers from these two equations (even for $g=2$).

Motivated by the equivalence between different models for 2-dimensional quantum gravity, E. Witten proposed an extremely striking conjecture in [Wit91], which together with the string equation and the initial condition $\langle\tau_0^3\rangle_0 = 1$ determines completely all intersection numbers for all (g,n).

To state this conjecture, let us start by defining the following generating function

$$F(t_0,t_1,\dots) := \left\langle \exp\left(\sum_{i\geq 0} t_i\tau_i\right)\right\rangle \tag{3.3}$$

which can also be written as

$$F(t_0,t_1,\dots) = \sum_{n\geq 0}\frac{1}{n!}\sum_{(i_1,\dots,i_n)\in\mathbb{Z}^n_{\geq 1}}\langle\tau_{i_1}\cdots\tau_{i_n}\rangle\, t_{i_1}\cdots t_{i_n} = \sum_{\substack{(d_0,d_1,\dots)\\ d_0,d_1,\dots\geq 0}}\langle\tau_0^{d_0}\tau_1^{d_1}\cdots\rangle\,\frac{t_0^{d_0}t_1^{d_1}\cdots}{d_0!d_1!\cdots}$$

This frightening generating function starts with (see [KL07])

$$F(t_0,t_1,\dots) = \frac{1}{24}t_1 + \frac{1}{6}t_0^3 + \frac{1}{48}t_1^2 + \frac{1}{24}t_0t_2 + \frac{1}{6}t_0^3t_1 + \frac{1}{1152}t_4 + \frac{1}{72}t_1^3 + \frac{1}{12}t_0t_1t_2 + \frac{1}{48}t_0^2t_3 + \cdots$$

The Witten conjectures is the following.

Theorem 3.4.4 (Witten–Kontsevich)**.** *The second derivative $U := \partial^2F/\partial t_0^2$ of F satisfies the KdV equation*

$$\frac{\partial U}{\partial t_1} = U\frac{\partial U}{\partial t_0} + \frac{1}{12}\frac{\partial^3 U}{\partial t_0^3}.$$

Witten conjecture was first proved by M. Kontsevich [Kon92]. Since then, several proofs with different approaches have been found, [KL07], [OP09], [Mir07c], [ABC+20], etc.

Let $G(p_1,p_3,p_5,\dots)$ be the generating function obtained from $F(t_1,t_2,t_3,\dots)$ by substituting $t_k = (2k-1)!!\,p_{2k+1}$.

Let us define two family $(a_k)_{k\in\mathbb{Z}}$ and $(b_k)_{k\in\mathbb{Z}}$ of linear operators on $\mathbb{Q}[[p_1,p_2,\dots]]$. First, for $k=0$, set $\alpha_0=0$, and for $k\geq 1$, set

$$a_k : g(p_1,p_2,\dots)\mapsto p_k\cdot g(p_1,p_2,\dots),\qquad a_{-k} : g(p_1,p_2,\dots)\mapsto k\cdot\frac{\partial g}{\partial p_k},$$

Then define

$$b_0 := \frac{1}{8} + \sum_{i\geq 0} a_ia_{-i},\qquad b_k := \frac{1}{2}\sum_{i+j=k} a_ia_j,\quad \text{if } k\neq 0.$$

We say that F satisfies the *Virasoro constraints* if

$$(a_{-(2k+3)} - b_{-2k})\exp(G) = 0 \tag{3.4}$$

for all $k \geq -1$.

The Virasoro constrains when $k = -1$ and $k = 0$ are none other than the string equation and the dilaton equation respectively, and the Witten conjecture is equivalent to the Virasoro constrains [DVV91], [FKN91]. For $k \geq 1$, (3.4) is equivalent to

$$\begin{aligned}(2k+3)!!\langle \tau_{d_1} \cdots \tau_{d_n} \tau_{k+1} \rangle_g = \sum_{i=1}^{k} \frac{(2d_i+2k+1)!!}{(2d_i-1)!!} \langle \tau_{d_1} \cdots \tau_{d_i+k} \cdots \tau_{d_n} \rangle_g \\ + \frac{1}{2} \sum_{i+j=k-1} (2i+1)!!\,(2j+1)!!\,\langle \tau_{d_1} \cdots \tau_{d_n} \tau_i \tau_j \rangle_{g-1} \\ + \frac{1}{2} \sum_{\substack{i_1+i_2=k-1 \\ I_1 \sqcup I_2 = \{1,\ldots,n\} \\ g_1+g_2=g}} (2i_1+1)!!\,(2i_2+1)!!\langle \tau_{I_1} \tau_{i_1} \rangle_{g_1} \langle \tau_{I_2} \tau_{i_2} \rangle_{g_2}.\end{aligned}$$

3.5 Probability theory

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $X \colon \Omega \to E$, where E is a measurable space, be a random variable. The *probability distribution*, or simply the *distribution*, of X is the push-forward (probability) measure $X_*(\mathbb{P})$ on E. When we talk about a distribution on certain measurable space E without mentioning any random variable, we are referring to a probability measure on E. For example, the *beta distribution* with parameters $(a, b) \in \mathbb{R}^2_{>0}$, denoted by $\mathrm{Beta}(a, b)$, is a probability distribution on $[0, 1]$ that has density function (with respect to the Lebesgue measure)

$$\frac{\Gamma(a)\,\Gamma(b)}{\Gamma(a+b)} \cdot x^{a-1}(1-x)^{b-1}$$

where Γ denotes the usual gamma function. A multivariate generalization of the beta distribution is the *Dirichlet distribution*. The Dirichlet distribution of order k with parameter $(\alpha_1, \ldots, \alpha_k) \in \mathbb{R}^k_{>0}$, denoted by $\mathrm{Dir}(\alpha_1, \ldots, \alpha_k)$, is a probability distribution on the standard simplex of dimension $k-1$

$$\Delta_1^{k-1} \coloneqq \{(x_1, \ldots, x_k) \in \mathbb{R}^k_{\geq 0} : x_1 + \cdots + x_k = 1\}$$

that has a density function with respect to the Lebesgue measure on Δ_1^{k-1} given by the formula

$$\frac{1}{\mathrm{B}(\alpha_1, \ldots, \alpha_k)}\, x_1^{\alpha_1-1} \cdots x_k^{\alpha_k-1}$$

where $\mathrm{B}(\alpha_1, \ldots, \alpha_k) \coloneqq \Gamma(\alpha_1) \cdots \Gamma(\alpha_k)/\Gamma(\alpha_1 + \cdots + \alpha_k)$ is the multivariate beta function. Finally, the *gamma distribution* of parameter $\alpha > 0$ is a probability distribution on $x > 0$ with density $x^{\alpha-1}e^{-x}/\Gamma(\alpha)$.

For us a partition is not a partition of a positive integer, but a partition of a positive real number, say r ($r = 1$ in most cases). More precisely, a *partition* of r is a countable multiset λ of positive real numbers such that $\sum_{x \in \lambda} x = r$. In this thesis, we are particularly interested in distributions of random partitions. And when we study (large components of) random partitions, the *Poisson–Dirichlet distribution* pops up somewhere with high probability.

The Poisson–Dirichlet distribution is first introduced by Kingman in [Kin75] as the limit

of a sequence of Dirichlet distributions as follows. Consider a sequence of random variables $X_n = (X_{n,1}, \dots, X_{n,n})$ following $\mathrm{Dir}(\theta/n, \dots, \theta/n)$. Then the Poisson–Dirichlet distribution of parameter $\theta > 0$, denoted by $\mathrm{PD}(\theta)$, is the limiting distribution of its descending order statistics $X_n^{\downarrow}$ as $n \to \infty$. The Poisson–Dirichlet distribution can also be defined via a Poisson point process as follows. Let $N = \{N(t) : t > 0\}$ be the Poisson point process on $]0, \infty[$ with intensity $\theta x^{-1} e^{-x}$ and let $P_1 > P_2 > \cdots > 0$ be its points. The sum $S = P_1 + P_2 + \cdots$ is almost surely finite, and the sequence of random variables $(P_1/S, P_2/S, \dots)$ follows $\mathrm{PD}(\theta)$.

Another way to define the same distribution, and perhaps the simplest way, is to use the *stick breaking process* as we now explain. We start with a horizontal stick of length 1. Pick a random point on the stick and break the stick into two parts (left and right) at that point. Define V_1 as the length of the piece on the right and throw it away. Then, pick again a random point on the remaining stick and break it into two pieces. Write down the length of the piece on the right V_2, throw it away, and so on. More formally, let $U_1, U_2, \dots$ be a sequence of independent and identically distributed $\mathrm{Beta}(0, \theta)$ random variables. Consider the random sequence

$$V \coloneqq (V_1, V_2, V_3 \dots) \coloneqq (U_1, (1 - U_1)U_2, (1 - U_1)(1 - U_2)U_3, \dots).$$

Almost surely, this process does not terminates in finite number of steps. Nevertheless, almost surely, the series $V_1 + V_2 + \cdots$ converges to 1. Therefore, we obtain a random variable V whose distribution is defined on the standard infinite-dimensional simplex

$$\Delta_1^{\infty} \coloneqq \{(x_1, x_2, \dots,) \in [0, 1]^{\mathbb{Z}_{\geq 1}} : x_1 + x_2 + \cdots = 1\}.$$

The distribution of V, denoted by $\mathrm{GEM}(\theta)$, is called the *GEM distribution* (named after Griffiths, Engen, and McCloskey) of parameter θ, and its descending ordered statistics is nothing but $\mathrm{PD}(\theta)$.

The Poisson–Dirichlet distribution is "rather less than user-friendly". However, the marginal distribution of a GEM distribution has a simple description.

Proposition 3.5.1 ([DJ89])**.** *Let $\theta \in \mathbb{R}_{>0}$. If $X = (X_1, X_2, \dots)$ follows the $\mathrm{GEM}(\theta)$ distribution, then r-th marginal distribution $(X_1, \dots, X_r)$ of X has density function*

$$\mathbb{1}_{x_1 + \cdots + x_r \leq 1}(x_1, \dots, x_r) \frac{\theta^r (1 - x_1 - \cdots - x_r)^{\theta - 1}}{(1 - x_1)(1 - x_1 - x_2) \cdots (1 - x_1 - \cdots - x_{r-1})}$$

with respect to the Lebesgue measure.

See [ABT03, Section 5.7] for the basic properties of Poisson–Dirichlet and GEM distributions.

3.6 Combinatorics

3.6.1 Stable graphs

The topological structures of many geometric objects appearing in this thesis can be encoded into some combinatorial objects called *stable graphs*. Roughly speaking, a stable graph is the dual graph of a primitive multicurve on a surface. More formally, a stable graph Γ consists of the data

$$\Gamma = (V,\ E,\ H,\ g \colon V \to \mathbb{Z}_{\geq 0},\ \iota \colon H \to H)$$

satisfying the following properties:

- The pair (V, E) defines a connected graph, with vertex set V and edge set E. The set H is the set of *half-edges*.
- The map v assigns each half-edge to its adjacent vertex.
- The map ι is an involution, such that the 2-cycles of ι are in bijection with E, and the fixed points of ι are in bijection with L.
- The *genus function* g assigns each vertex x to its genus (the genus of the surface corresponding to x), such that the *stability condition*

$$2g(x) - 2 + n(x) > 0$$

  is satisfied, where $n(x)$ denotes the number of edges and legs adjacent to x.

A *weighted stable graph* is a pair (Γ, μ) where Γ is a stable graph and $\mu\colon E(\Gamma) \to \mathbb{Z}_{\geq 1}$ is a function assigning an (positive) integer weight to each edge of Γ.

Given a multi-geodesic (more a multicurve) $\gamma = m_1\gamma_1 + \cdots + m_k\gamma_k$, we associate a (weighted) stable graph with it as follows. Cut the surface along $\gamma_1, \ldots, \gamma_k$. To each connected component S of $\Sigma_g \smallsetminus \{\gamma_1, \ldots, \gamma_k\}$, we associate a vertex, and we decorate this vertex with the genus of S. For each component γ_i of γ, we draw an edge that connects the two vertices (which could be the same) corresponding to the two connected components of $\Sigma_g \smallsetminus \{\gamma_1, \ldots, \gamma_k\}$ bounded by γ_i. We now obtain a stable graph Γ_γ which does not depend on $m_1, \ldots, m_k$, and a weighted stable graph $(\Gamma_\gamma, \mu_\gamma)$ where μ_γ maps the edge corresponding to γ_i to m_i.

See Figure 3.1 for an example.

A weighted stable graph associated to a square-tiled half-translation surface q can also be constructed similarly. Using the horizontal flat geodesic foliation of q, q can be decomposed into singular leaves and maximal horizontal cylinders consisting of regular leaves. The stable graph Γ_q has a vertex for each singular leaf, and has an edge for each cylinder. Each edge connects the two vertices (can be the same) corresponding to the two singular leaves that bound the cylinder corresponding to the edge. Each edge can be decorated with the height of its corresponding cylinder, and each vertex carries a ribbon graph (or simply the genus of this ribbon graph) obtained by considering a tubular neighborhood of the singular leaf corresponding to the vertex. An example is given in Introduction.

The symmetries of a multi-geodesic or a square-tiled surface can be captured by its associated stable graph. An *isomorphism* between two stable graphs $\Gamma_1 = (V_1, E_1, H_1, g_1, \iota_1)$ and $\Gamma_2 = (V_2, E_2, H_2, g_2, \iota_2)$ is a pair (ϕ, ψ) of bijections $\phi\colon V_1 \to V_2$ and $\psi\colon H_1 \to H_2$ such that

- $\psi \circ \iota_1 = \iota_2 \circ \psi$. In other words, ψ maps an edge of Γ_1 to en edge of Γ_2.
- $\phi \circ v_1 = v_2 \circ \phi$ where $v_i\colon H_i \to V_i$, $i = 1, 2$ is the maps assigning a half-edge to its adjacent vertex.
- $g_1 = g_2 \circ \phi$.

Note that ψ determines ϕ but it is convenient to record automorphism as a pair (ϕ, ψ). An *isomorphism* between two weighted stable graphs (Γ_1, m_1) and (Γ_2, m_2) is an isomorphism (ϕ, ψ)

between Γ_1 and Γ_2 such that $\mathrm{m}_1 \circ e_1 = \mathrm{m}_2 \circ e_2 \circ \psi$ where $e_i \colon H_i \to E_i$, $i = 1, 2$, is the map which takes each half-edge in Γ_i to the edge containing it. We denote by $\mathrm{Aut}(\Gamma)$ the set of automorphisms of Γ, and by $\mathrm{Aut}(\Gamma, \mu)$ the set of automorphism of (Γ, μ).

Note that on Σ_g, the set of primitive multicurves and the set of stable graphs are in natural bijection, and the set of multicurves and the set of weighted stable graphs are in natural bijection. Further, let $\gamma = m_1\gamma_1 + \cdots + m_k\gamma_k$, then we have

$$|\mathrm{Aut}(\Gamma_\gamma)| = [\mathrm{Stab}(\gamma_1 + \cdots + \gamma_k) : \mathrm{Stab}^+(\gamma_1, \ldots, \gamma_k)], \tag{3.5}$$

and

$$|\mathrm{Aut}(\Gamma_\gamma, \mu_\gamma)| = [\mathrm{Stab}(m_1\gamma_1 + \cdots + m_k\gamma_k) : \mathrm{Stab}^+(\gamma_1, \ldots, \gamma_k)]. \tag{3.6}$$

3.6.2 Graph polynomials

Given a stable graph Γ, we associate to each edge $e \in E$ a variable x_e, and define the associated *graph polynomial* by the formula

$$F_\Gamma(x_e : e \in E) = \prod_e x_e \cdot \prod_v V_{g(v),n(v)}(x_{e(h)} : h \in H,\ v(h) = v). \tag{3.7}$$

where e runs through the edge set E, v runs through the vertex set V, $V_{g(v),n(v)}$ is the Weil–Petersson volume of $\mathcal{M}_{g(v),n(v)}$, $e(h)$ is the edge that contains the half-edge h, and $v(h)$ denotes the vertex incident to h. Note that F_Γ is of degree $2d - k$.

Let us write $\bar{F}_\Gamma$ for the top-degree homogeneous part of F_Γ, and $\bar{V}_{g,n}$ for that of $V_{g,n}$. Finally, we write F_γ and $\bar{F}_\gamma$ for F_{Γ_γ} and $\bar{F}_{\Gamma_\gamma}$ respectively.

Example 3.6.1. If $\{\gamma_1, \ldots, \gamma_{3g-3}\}$ is a pants decomposition, then $V_{g(v),n(v)} = 1$ for all $v \in V$, and $F_\gamma(x_1, \ldots, x_{3g-3}) = \bar{F}_\gamma(x_1, \ldots, x_{3g-3}) = x_1 \cdots x_{3g-3}$.

Example 3.6.2. Let $(\gamma_1, \gamma_2, \gamma_3)$ be an ordered multicurve on Σ_3 as in Firgure 3.1. The Weil–Petersson volume polynomial $V_{1,3}(x_1, x_2, x_3)$ is equal to (see [Do15])

$$\left(\frac{\mathrm{m}_{(3)}}{1152} + \frac{\mathrm{m}_{(2,1)}}{192} + \frac{\mathrm{m}_{(1,1,1)}}{96} + \frac{\pi^2\,\mathrm{m}_{(2)}}{24} + \frac{\pi^2\,\mathrm{m}_{(1,1)}}{8} + \frac{13\pi^4\,\mathrm{m}_{(1)}}{24}\right)(x_1^2, x_2^2, x_3^2) + \frac{14\pi^6}{9}$$

where m is the monomial symmetric polynomial. For example,

$$\mathrm{m}_{(2,1)}(x_1, x_2, x_3) = x_1^2x_2 + x_1x_3^2 + x_1x_2^2 + x_2^2x_3 + x_1x_3^2 + x_2x_3^2, \quad \mathrm{m}_{(1)}(x_1, x_2, x_3) = x_1 + x_2 + x_3.$$

So the top-degree part of $V_{1,3}$ is

$$\bar{V}_{1,3}(x_1, x_1, x_2) = \frac{2x_1^6 + x_2^6}{1152} + \frac{2x_1^6 + 2x_1^4x_2^2 + 2x_1^2x_2^4}{192} + \frac{x_1^4x_2^2}{96},$$

and therefore,

$$\bar{F}_\gamma(x_1, x_2, x_3) = \frac{2x_1^6x_2x_3 + x_1x_2^6x_3}{1152} + \frac{x_1^7x_2x_3 + x_1^5x_2^3x_3 + x_1^3x_2^5x_3 + x_1^5x_2^3x_3}{96}.$$

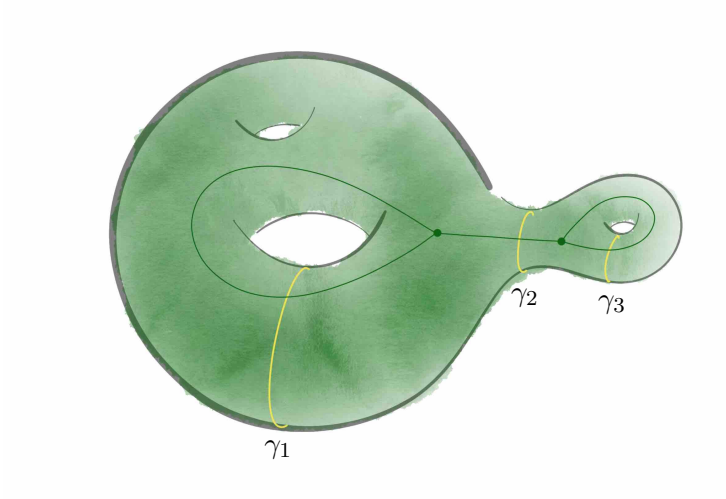

Figure 3.1: Example 3.6.2

3.6.3 Singularity analysis of generating functions

In Section 6.6, we use some standard tool in analytic combinatorics called *singularity analysis of generating functions* (see [FS09, Chapter VI]). When dealing with an enumeration problem, namely, counting the number of objects of a given size, it may be helpful to encode *these* numbers into a *generating function*. For example, let us count the number of full binary trees with n branching nodes (hence $n+1$ leaves). Denote this number by C_n, and the corresponding generating function is defined by

$$C(z) := \sum_{n=0}^{\infty} C_n z^n.$$

It follows from the observation "every full binary tree is of size 0 or it can be split into the left and the right subtrees by removing its root" that $C(x)$ satisfies the equation

$$C(z) = 1 + z\,C(z)^2.$$

Thus,

$$C(z) = \frac{1-\sqrt{1-4z}}{2z}, \tag{3.8}$$

and therefore, by the binomial theorem,

$$C_n = \frac{1}{n+1}\binom{2n}{n},$$

which is the well-known Catalan number. Apart from finding the explicit formula C_n, generating functions are even more useful when we are only interested in the *asymptotic behavior* of C_n as $n \to \infty$. In fact, up to now, the function $C(z)$ is only a formal power series, but we let $z \in \mathbb{C}$, then (3.8) defines a complex-valued function, and it turns out that the asymptotic behavior of C_n can be read from the singularities of $C(z)$ as a complex function. This idea is first developed explicitly in [FO90]. The following results can be found in Flajolet and Sedgewick's foundational book [FS09].

Theorem 3.6.3 ([FS09, Theorem VI.1]). *Let α be an arbitrary complex number in $\mathbb{C} \setminus \mathbb{Z}_{\leq 0}$. The coefficient of z^n in*

$$f(z) = \frac{1}{(1-z)^\alpha}$$

admits for large n a complete asymptotic expansion in descending power of n,

$$[z^n]\, f(z) \sim \frac{n^{\alpha-1}}{\Gamma(\alpha)} \left(1 + \frac{e_1}{n} + \frac{e_2}{n^2} + \cdots\right)$$

where $[z^n]\, f(z)$ denotes the coefficient of z^n in the formal power series $f(z) = \sum_{n\geq 0} a_n z^n$, and e_k is a polynomial in α of degree $2k$. In particular:

$$[z^n]\, f(z) \sim \frac{n^{\alpha-1}}{\Gamma(\alpha)} \left(1 + \frac{\alpha(\alpha-1)}{2n} + \frac{\alpha(\alpha-1)(\alpha-2)(3\alpha-1)}{24n^2} + \mathrm{O}\left(\frac{1}{n^3}\right)\right).$$

Theorem 3.6.4 ([FS09, Theorem VI.2]). *Let α be an arbitrary complex number in $\mathbb{C} \setminus \mathbb{Z}_{\leq 0}$. The coefficient of z^n in*

$$f(z) = \frac{1}{(1-z)^\alpha} \left(\frac{1}{z} \log \frac{1}{1-z}\right)^\beta$$

admits for large n a complete asymptotic expansion in descending power of n,

$$[z^n]\, f(z) \sim \frac{n^{\alpha-1}}{\Gamma(\alpha)} (\log n)^\beta \left(1 + \frac{C_1}{\log n} + \frac{C_2}{\log^2 n} + \cdots\right)$$

where

$$C_k = \binom{\beta}{k} \cdot \Gamma(\alpha) \cdot \frac{d^k}{ds^k} \frac{1}{\Gamma(s)}\bigg|_{s=\alpha}.$$

The following is [FS09, Definition VI.1].

Given $\phi \in \,]1, \pi/2[$ and $R \in \mathbb{R}_{>1}$, the open domain $\Delta(\phi, R)$ is defined as

$$\Delta(\phi, R) := \{z \in \mathbb{C} : |z| < R,\ z \neq 1,\ |\arg(z-1)| > \phi\}.$$

A domain is a Δ-*domain* at 1 if it is a $\Delta(\phi, R)$ for some R and ϕ; see Figure 3.2 for an example. For a complex number $\zeta \neq 0$, a Δ-domain at ζ is the image by the map $z \mapsto \zeta \cdot z$ of a Δ-domain at 1. A function is Δ-*analytic* if it is analytic in some Δ-domain.

Define

$$\mathcal{S} := \left\{(1-z)^{-\alpha} \cdot \lambda(z)^\beta : \alpha, \beta \in \mathbb{C}\right\}, \qquad \text{where} \qquad \lambda(z) := \frac{1}{z} \log \frac{1}{1-z}.$$

Theorem 3.6.5 ([FS09, Theorem VI.5]). *Let $f(z)$ be analytic in $|z| < \rho$ and have a finite number of singularities on the circle $|z| = \rho$ at points $\zeta_j = \rho\, e^{i\theta_j}$, for $j = 1, \ldots, r$. Assume that there exists a Δ-domain Δ_0 such that $f(z)$ is analytic in the indented disc $D := \zeta_1 \cdot \Delta_0 \cap \cdots \cap \zeta_j \cdot \Delta_0$ with $\zeta \cdot \Delta_0$ the image of Δ_0 by the mapping $z \mapsto \zeta \cdot z$.*

Assume that there exists r functions $g_1, \ldots, g_r$, each a linear combination of elements from $\mathcal{S}$, and a function $\tau \in S$ such that

$$f(z) = g(z/\zeta_j) + \mathrm{O}(\tau(z/\zeta_j))$$

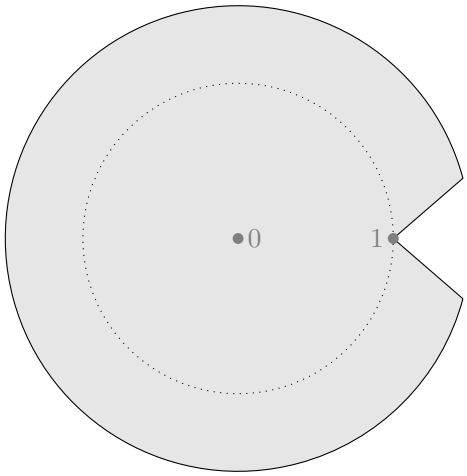

Figure 3.2: A Δ-domain

as $z \to \zeta_j$ *in* D*. Then the coefficients of* $f(z)$ *satisfy the asymptotic estimate*

$$[z^n]\, f(z) = \frac{[z^n]\, g_1(z)}{\zeta_1^n} + \cdots + \frac{[z^n]\, g_r(z)}{\zeta_r^n} + \mathrm{O}\left(\frac{\tau_n^*}{\rho^n}\right)$$

where each $[z^n]\, g_i(z)$ *has its coefficients determined by Theorem 3.6.4 and* $\tau_n^* = n^{a-1}(\log n)^b$ *if* $\tau(z) = (1-z)^{-\alpha}\lambda(z)^b$*.*

Chapter 4

Random multi-geodesics

In this chapter we define our model of random multi-geodesics on a closed hyperbolic surface, and briefly review some recent developments.

4.1 Model

Let X be a closed hyperbolic surface of genus $g \geq 2$. Let s_X denote that set of all multi-geodesics (or all free homotopy classes of multicurves) on X.

A model of random multi-geodesics on X is nothing but a probability measure on s_X. The uniform probability measure seems to be the most natural choice if we believe that all multi-geodesics are born equal and no one is more special than the others. However, such a probability measure does not exist since s_X is countable and infinite; essentially for the same reason it is impossible to "pick a positive integer uniformly at random". Nevertheless, a computer has no difficult generating a positive integer not exceeding, say, a billion. The general idea is to define a "complexity function" on the "moduli space" M of objects under consideration, say $H\colon M \to \mathbb{R}$, such that $H(x) < \infty$ for all $x \in M$, and $|\{x \in M : H(x) < N\}| < \infty$ for any $N > 0$. By doing so, we obtain a family of probability measures indexed by N supported on a larger and larger subset of M as N increases, and we are interested in the asymptotic behaviors of random variables on M as $N \to \infty$.

For our purposes, the length function $\ell_X\colon s_X \to \mathbb{R}$ seems to be a perfect choice for the complexity function. From now on, we shall consider $s_X(R)$, the set of multi-geodesics on X of length at most R, as a probability space equipped with the uniform probability measure.

To understand of the topology and the geometry of a random multi-geodesic sampled according to our model, it would be necessary to work out the asymptotics the growth of the cardinality of

$$s_X(\gamma, R) \coloneqq \{\alpha \in \mathrm{Mod}_g \cdot \gamma : \ell_X(\alpha) \leq R\} \tag{4.1}$$

as $R \to \infty$, for fixed multicurve γ.

McShane and Rivin [MR95a] [MR95b] showed that in the case of a punctured torus, $|s_X(R)| \sim c_X R^2$ as $R \to \infty$, where c_X is a constant depending on X. Later, Rivin [Riv01] proved that the polynomial growth of $|s_X(R)|$ holds for any genus. More precisely, for any $X \in \mathcal{M}_g$, there exists c_X such that

$$c_X^{-1} R^{6g-6} \leq |s_X(R)| \leq c_X R^{6g-6}.$$

Their results were elegantly extended by Mirzakhani in her thesis:

Theorem 4.1.1 (Mirzakhani)**.** *For any $X \in \mathcal{M}_g$ and any multicurve γ on Σ_g, there exists a positive rational constant $c(\gamma)$ such that*

$$|s_X(\gamma, R)| \sim c(\gamma) \cdot \frac{B(X)}{b_g} \cdot R^{6g-6}$$

as $R \to \infty$, where $B(X)$ is the Thurston volume of the unit ball in the space of measured laminations $\mathcal{ML}(X)$ with respect to the length function ℓ_X, and

$$b_g := \int_{\mathcal{M}_g} B(X)\, dX = \sum_{[\gamma]} c(\gamma)$$

where $[\gamma]$ runs over all topological types of multicurves on Σ_g and dX is the Weil–Petersson measure on $\mathcal{M}_g$.

The product structure of the asymptotic growth constant $c(\gamma)B(X)/b_g$ leads to the following immediate corollary.

Corollary 4.1.2. *We have*

$$\lim_{R\to\infty} \frac{|s_X(R,\gamma)|}{|s_X(R)|} = \frac{c(\gamma)}{b_g}.$$

Remark 4.1.3. In [Mir08b] Mirzakhani gives an explicit formula for $c(\gamma)$.

A remarkable fact is that this ratio $c(\gamma)/b_g$, which can be interpreted as "the probability that a random multi-geodesic on X has the same topological type than γ", depends only on the topological type of the curve (but not on the geometry of the surface), and we can therefore simply talk about random (topological) multicurves.

Example 4.1.4. Let γ be a simple closed curve on Σ_2. If γ is non-separating, then $c(\gamma) = 16/63$; if γ is separating, then $c(\gamma) = 1/189$. Therefore, the probability that a random simple closed geodesic on $X \in \mathcal{M}_2$ is separating is $1/49$.

Example 4.1.5. Let $\alpha = (\alpha_1, \alpha_2)$ be a primitive multicurve on Σ_2. If both of α_1 and α_2 are non-separating, then $c(\alpha) = 8/15$; if one is non-separating and the other is separating, then $c(\alpha) = 1/45$.

4.2 Masur–Veech volumes

The following result bridges the hyperbolic world and the flat world of surfaces:

Theorem 4.2.1 ([Mir08a], [AH20b], [MT19], [ES22])**.** *For any $g \geq 2$,*

$$b_g = \frac{1}{2^{4g-1}(3g-3)(4g-4)!}\, \mathrm{vol}_{\mathrm{MV}}(\Omega(1^{4g-4})).$$

This result was first proved by Mirzakhani (as a byproduct) in [Mir08a], where she constructs a conjugacy between the earthquake flow over the bundle of measured laminations over the moduli space and the Teichmüller horocycle flow over the bundle of holomorphic quadratic

differentials over the moduli space, and establishes the ergodicity of the earthquake flow. Since then, several proofs using different methods were found by Arana-Herrera, Monin–Telpukhovskiy, and Erlandsson–Souto.

In [DGZZ21], the authors showed that the Masur–Veech volume of the principal stratum of the moduli space of quadratic differentials $\mathcal{Q}(1^{4g-4})$ can be written as a sum over contributions from all (weighted) stable graphs. By a direct comparison with the formula of $c(\gamma)$ in [Mir08b, Theorem 5.3], they found that the same relation holds in the following form:

Theorem 4.2.2 ([DGZZ21, Theorem 1.21])**.** *Let $\gamma = m_1\gamma_1 + \cdots + m_k\gamma_k$ be a multicurve on Σ_g. We have*

$$\frac{c(\gamma)}{[\mathrm{Stab}(\gamma_1+\cdots+\gamma_k) : \mathrm{Stab}^+(\gamma_1,\ldots,\gamma_k)]} = \frac{1}{2^{4g-1}(3g-3)(4g-4)!}\frac{\mathrm{vol}_{\mathrm{MV}}(\Gamma_\gamma,\mu_\gamma)}{|\mathrm{Aut}(\Gamma_\gamma,\mu_\gamma)|},$$

where μ is the evident weight assignment. Or equivalently,

$$\frac{c(\gamma)}{[\mathrm{Stab}(\gamma_1+\cdots+\gamma_k) : \mathrm{Stab}(\gamma)]} = \frac{\mathrm{vol}_{\mathrm{MV}}(\Gamma_\gamma,\mu_\gamma)}{2^{4g-1}(3g-3)(4g-4)!}.$$

Let us briefly review the formula for Masur–Veech volumes given in [DGZZ21].

Let I be a finite set. Give a function $\mu\colon I \to \mathbb{Z}_{\geq 1}$, we define a linear operator $\mathcal{Y}_\mu$ on the rational polynomial algebra $\mathbb{Q}[x_i]_{i\in I}$ in variables indexed by I, given by (the linear extension of) the following formula

$$\mathcal{Y}_\mu\colon \prod_{i\in I} x_i^{n_i} \longmapsto \prod_{i\in I}\frac{n_i!}{\mu(i)^{n_i+1}}. \tag{4.2}$$

For us, the index set I is always the edge set E of a stable graph Γ. If the edges of Γ are labeled by $\{1,2,\ldots,k\}$, where $k = |E|$, then μ can be determined by a vector $(m_1,\ldots,m_k)$ where $m_i := \mu(i)$, and m_i(4.2) can be rewritten as

$$\mathcal{Y}_{(m_1,\ldots,m_k)}(x_1^{n_1}\cdots x_k^{n_k}) = \frac{n_1!\cdots n_k!}{m_1^{n_1+1}\cdots m_k^{n_k+1}}.$$

Remark 4.2.3. This operator is somewhat similar to the Laplace transform. Its restriction on $\mathbb{Q}_n[x_1,\ldots,x_k]$ (polynomials with k variables of degree n) can be written as

$$\mathcal{Y}_{(m_1,\ldots,m_k)}(\bullet) = (n+k-1)!\int_{\Delta^{k-1}_{1;m_1,\ldots,m_k}} \bullet\,\lambda(dx)$$

where $\Delta^{k-1}_{1;m_1,\ldots,m_k} := \{(x_1,\ldots,x_k)\in\mathbb{R}_{\geq 1} : m_1x_1+\cdots+m_kx_k = 1\}$ and λ is the Lebesgue measure on it.

Given $m \in \mathbb{Z}_{\geq 1}\cup\{+\infty\}$, let us collect different $\mathcal{Y}_\mu$ with the same index set together and define

$$\mathcal{Z}_m := \sum_{\substack{\mu\colon I\to\mathbb{Z}_{\geq 1}\\ \max_{i\in I}\mu(i)\leq m}} \mathcal{Y}_\mu = \sum_{\substack{(m_1,\ldots,m_k)\in\mathbb{Z}^k_{\geq 1}\\ m_i\leq m,\ 1\leq i\leq m}} \mathcal{Y}_{(m_1,\ldots,m_k)},$$

or equivalently,

$$\mathcal{Z}_m(x_1^{n_1}\cdots x_k^{n_k}) = n_1!\cdot\zeta_m(m_1+1)\cdots n_k!\cdot\zeta_m(m_k+1)$$

where

$$\zeta(s) := \sum_{n=1}^{m} \frac{1}{n^s}$$

is the partial Riemann zeta function.

The last ingredient is (another) graph polynomial P_Γ associated to a stable graph Γ, whose variables are indexed by the edges of Γ, defined by

$$\begin{aligned} &P_\Gamma(x_e : e \in E) \\ &\qquad := \frac{2^{6g-6}(4g-4)!}{(6g-7)!} \frac{1}{2^{|V|-1}} \frac{1}{|\mathrm{Aut}(\Gamma)|} \cdot \prod_e x_e \cdot \prod_v N_{g(v),n(v)}(x_{e(h)} : h \in H, e(h) = v), \end{aligned}$$

where e runs over the edge set E of Γ and v runs over the vertex set V of Γ, and the polynomial $N_{g(v),n(v)}$ is given by

$$N_{g,n}(x_1, \dots, x_n) = \sum_{\substack{(d_1,\dots,d_n)\in\mathbb{Z}^n_{\geq 0} \\ d_1+\cdots+d_n=3g-3+n}} \frac{1}{2^{5g-6+2n}} \frac{x_1^{2d_1} \cdots x_n^{2d_n}}{d_1! \cdots d_n!} \int_{\overline{\mathcal{M}}_{g,n}} \psi_1^{d_1} \cdots \psi_n^{d_n}.$$

Remark 4.2.4. The presence of the polynomial $N_{g,n}$ is related to the ribbon graph counting (of given genus with prescribed perimeters), and it can be written in terms of intersection numbers of ψ-classes by the work of Kontsevich [Kon92] and Norbury [Nor13].

Remark 4.2.5. Let $\gamma = (m_1\gamma_1, \dots, m_k\gamma_k)$ be an ordered multicurve on Σ_g and Γ be its associated stable graph. The polynomial $\bar{F}_\gamma$ appears in Theorem 5.0.2 and P_Γ are related by the simple relation

$$P_\Gamma = 2^{4g-2} \frac{(4g-4)!}{(6g-7)!} \frac{1}{|\mathrm{Aut}(\Gamma)|} \cdot \bar{F}_\gamma$$

and $N_{g,n} = 2^{-(2g-3+n)} \cdot \bar{V}_{g,n}$, where $\bar{V}_{g,n}$ is the top-degree part of the Weil–Petersson volume polynomial of the moduli space of hyperbolic surfaces that appears in $\bar{F}_\gamma$.

One of the main results of [DGZZ21] is the following:

Theorem 4.2.6 ([DGZZ21]). *The volume contribution of* $\mathrm{vol}_{\mathrm{MV}}(\mathcal{Q}_g(1^{4g-4}))$ *from* (Γ, μ) *is*

$$\mathrm{vol}_{\mathrm{MV}}(\Gamma, \mu) = \mathcal{Y}_\mu(P_\Gamma).$$

We can also write

$$c(\gamma) = [\mathrm{Aut}(\Gamma) : \mathrm{Aut}(\Gamma, \mu)] \cdot \mathcal{Y}_\mu(P_\Gamma) = [\mathrm{Stab}(\gamma_1 + \cdots + \gamma_k) : \mathrm{Stab}(\gamma)] \cdot \mathcal{Y}_\mu(P_\Gamma).$$

Note that although P_Γ and $\bar{F}_\gamma$ arise for different reasons, both of them can be expressed in terms of intersection numbers between psi-classes on $\overline{\mathcal{M}}_{g,n}$. Therefore, the key to understand random multi-geodesics and random square-tiled surfaces is the large genus asymptotics for these intersection numbers.

4.3 Large genus asymptotics for intersection numbers

Recall that the Virasoro constrains (3.4) satisfied by the generating function of intersection numbers of psi-classes defined by (3.3) completely determines all these numbers. However, these terrifying recursive relations are delicate to analyse, and one has no reason to expect that, except in a few exceptional cases like [Zog19], these intersection numbers admit simple closed expressions. However, Delecroix–Goujard–Zograf–Zorich conjectured in [DGZZ21] that these numbers have a simple asymptotic formula as the genus goes to infinite, and a lower bound [DGZZ20c] was proved later. The whole conjecture (in a even stronger form) was turned into a theorem by Aggarwal [Agg21] using deeply insightful probabilistic methods.

Theorem 4.3.1 ([Agg21]). *For $g, n \in \mathbb{Z}_{\geq 0}$ with $2g-2+n>0$ and $(d_1,\dots,d_n) \in \mathbb{Z}^n_{\geq 0}$ with $d_1+\cdots+d_n = 3g-3+n$, let $\epsilon(d_1,\dots,d_n)$ be defined by*

$$\int_{\overline{\mathcal{M}}_{g,n}} \psi_1^{d_1}\cdots\psi_n^{d_n} = \frac{(6g-5+2n)!!}{(2d_1+1)!!\cdots(2d_n+1)!!}\cdot\frac{1}{g!\cdot 24^g}\cdot(1+\epsilon(d_1,\dots,d_n)).$$

We have

$$\lim_{g\to\infty}\sup_{n<\sqrt{g/800}}\ \sup_{\substack{(d_1,\dots,d_n)\in\mathbb{Z}^n_{\geq 0}\\ d_1+\cdots+d_n=3g-3+n}} \epsilon(d_1,\dots,d_n) = 0.$$

This result is one of the key ingredients of [DGZZ20b] and [DL22] (last chapter of this thesis). Later, a weaker version was proved by Jindong Guo and Di Yang in [GY21].

4.4 Topology of large genus random multi-geodesics

Based on the work of Aggarwal [Agg20] about large genus asymptotics for intersection numbers between psi-classes and for the Masur–Veech volume, in their recent paper, Delecroix–Goujard–Zograf–Zorich gave beautiful answers to Question 1.1.7 and Question 1.1.8 in the introduction.

Theorem 4.4.1 ([DGZZ20b, Theorem 1.1]). *The probability that a random multicurve is separating tends to zero as the genus of the underlying surface goes to infinity.*

Theorem 4.4.2 ([DGZZ20b]). *As $g\to\infty$, we have*

$$\mathbb{E}(K_g) = \frac{\log(6g-6)}{2} + \frac{\gamma}{2} + \log(2) + \mathrm{o}(1)$$

$$\mathbb{V}\mathrm{ar}(K_g) = \frac{\log(6g-6)}{2} + \frac{\gamma}{2} + \log(2) - \frac{3\,\zeta(2)}{4} + \mathrm{o}(1)$$

where $\gamma \approx 0.57721$ is the Euler–Mascheroni constant.

In fact, they proved the following result, which is much stronger than Theorem 4.4.2.

Theorem 4.4.3 ([DGZZ20b, Theorem 1.12]). *For all $z\in\mathbb{C}$ such that $|z|<8/7$, the following asymptotic relation is valid as $g\to\infty$*

$$\mathbb{E}(z^{K_g}) = (6g-6)^{(z-1)/2}\cdot\frac{z\,\Gamma(3/2)}{\Gamma(z/2)}\cdot(1+\mathrm{o}(1)).$$

Moreover, for any compact set Q in the open disk $|z| < 8/7$, there exists $\delta(Q) > 0$, such that for all $z \in Q$ the error term has the form $\mathrm{O}(g^{-\delta(Q)})$.

The preceding theorem implies that the distribution of the number of components K_g can be approximated by the Poisson distribution of parameter $\log(6g-6)/2$ in a very strong sense called "mod-Poisson convergence" (see [KN10]). This convergence is stronger than a central limit theorem, and in particular implies (see e.g., [Hwa94]) the following large derivation result which shall be useful for us later.

Theorem 4.4.4 ([DGZZ20b, Theorem 1.13])**.** *For any $\kappa \in]1, 1.236]$ such that $\kappa \log(6g-6)/2$ is an integer, we have*

$$\mathbb{P}(K_g > \kappa \log(6g-6)/2 + 1) = \frac{(6g-6)^{-\frac{x\log(x)-x+1}{2}}}{\sqrt{2\pi\kappa\log(6g-6)/2}} \frac{\kappa}{\kappa-1} \left(\frac{\sqrt{\pi}}{2\Gamma(1+x/2)} + \mathrm{O}\left(\frac{1}{\log g}\right)\right)$$

To sum up, topologically speaking, a *typical* random multi-geodesic on a large genus (random or not) hyperbolic surface is non-separating, and has roughly $\log(g)/2$ components.

In the two next chapter, we investigate the geometry of a random multi-geodesic on a hyperbolic surface, and more precisely, we shall focus on the length partition.

Chapter 5

Length partition of random multi-geodesics

The content of this chapter is essentially taken from [Liu19].

The study of the length partition of random multi-geodesics is initialed by Mirzakhani in [Mir16], where she proves the following result

Theorem 5.0.1 ([Mir16, Theorem 1.2])**.** *If $\{\gamma_1, \dots, \gamma_{3g-3}\}$ is a pants decomposition of X, then $\hat{\ell}_{X,R,\gamma}$ converges in law to the Dirichlet distribution of order $3g-3$ with parameters $(2, \dots, 2)$, i.e., the limit distribution of $\hat{\ell}_{X,\gamma,R}$ admits density function $(6g-7)! \cdot x_1 \cdots x_{3g-3}$ with respect to the Lebesgue measure on the standard simplex $\Delta_1^{3g-4} := \{(x_1, \dots, x_{3g-3}) \in \mathbb{R}_{\geq 0} : x_1 + \cdots + x_{3g-3} = 1\}$ of dimension $3g-4$. In other words, for any open subset U of Δ_1^{3g-4},*

$$\lim_{R\to\infty} \mathbb{P}(\hat{\ell}_{X,R,\gamma} \in U) = (6g-7)! \int_U x_1 \cdots x_{3g-3}\, \lambda(dx).$$

where λ is the Lebesgue measure on Δ_1^{3g-4}.

Our main result of this chapter is the following generation of the preceding theorem to any arbitrary topological type.

Theorem 5.0.2. *Let $\gamma = (m_1\gamma_1, \dots, m_k\gamma_k)$ be an ordered multi-geodesic on X. The random variable $\hat{\ell}_{X,R,\gamma}$ converges in law to a random variable which admits a polynomial density with respect to the Lebesgue measure on Δ^{k-1} given by, up to a normalizing constant,*

$$(x_1, \dots, x_k) \mapsto \bar{F}_\gamma(x_1/m_k, \dots, x_k/m_k)$$

where $\bar{F}_\gamma$ is top-degree (homogeneous) part of the graph polynomial F_γ associated to γ defined by (3.7).

Remark 5.0.3. The function $\bar{F}_\gamma$ is a homogeneous polynomial of degree $6g-7$ whose coefficients can be expressed in terms of the psi-classes on the Deligne–Mumford compactification of the moduli space of smooth complex curves $\overline{\mathcal{M}}_{g,n}$ (see Theorem 3.2.7). In particular, it depends only upon the topological type of γ, but not upon the hyperbolic metric X.

Idea of the proof

The limit distribution that we are after boils down to the asymptotics of multicurves counting under constraints, which can be transformed to a problem of approximating to the number of

"lattice points" within a horoball in a covering space of the moduli space. By considering tiling of the covering space by translates of a fundamental domain for the action of the mapping class group, it would not be unreasonable to expect that this number might be proportional to the volume of the horoball divided by the volume of the moduli space, and finally this is not so far from the truth. We proceed using techniques that Margulis introduced in his thesis [Mar04], and the equidistribution theorem for large horospheres initially established by Mirzakhani [Mir07a]. Similar methods were also applied in, e.g., [EM93].

Theorem 5.0.2 can be generalized to hyperbolic surfaces with cusps if Mirzakhani's work on the ergodicity of the earthquake flow can be generalized to such surfaces, which seems to be the case (see [Mir16]).

Remark

While the author was finishing this paper, the paper by Arana-Herrera [AH21b] appeared on the arXiv. Both paper are devoted to a similar circle of problems and use a similar circle of ideas, though they were written in parallel and completely independently. In particular, Arana-Herrera proves a much more general version of our Theorem 5.3.1 ([AH21b, Theorem 1.3]), which is one of the key ingredients allowing to attack the counting problem and the length statistics. We learned from [AH21b] that this kind of statistics was initially conjectured by Wolpert. Papers [AH21b] and [AH20a] established results closely related to Theorem 5.0.2.

Proposition 5.3.4 below is based on a theorem stated by M. Mirzakhani but presented without a detailed proof. The paper [AH21b] contains a detailed proof of an even stronger estimate which implies, in particular, the statements of this theorem; see Remark 5.3.8.

5.1 Mirzakhani's covering spaces

Let $\gamma = (m_1\gamma_1, \dots, m_k\gamma_k)$ be an ordered multicurve on Σ_g. The quotient space

$$\mathcal{M}_g^\gamma := \mathcal{T}_g / \operatorname{Stab}(\gamma)$$

introduced by M. Mirzakhani in her thesis plays an important role in this paper.

Write $\pi^\gamma \colon \mathcal{T}_g \to \mathcal{M}_g^\gamma$ and $\pi_\gamma \colon \mathcal{M}_g^\gamma \to \mathcal{M}_g$ for the two natural projections ("raising and lowering the index"). Let us consider the product space $P_\gamma := \mathcal{T}_g \times \operatorname{Mod}_g \cdot (\gamma_1, \dots, \gamma_k)$ of the Teichmüller space and the mapping class group orbit of $(\gamma_1, \dots, \gamma_k)$. The mapping class group Mod_g acts on P (from the right) via $(X; \gamma_1, \dots, \gamma_k) \cdot h = (X \cdot h; h^{-1}\gamma_1, \dots, h^{-1}\gamma_k)$.

Lemma 5.1.1. *The quotient $P_\gamma/\operatorname{Mod}_g$ is isomorphic to $\mathcal{M}_g^\gamma$ as symplectic orbifolds.*

Proof. Consider the map $P_\gamma \to \mathcal{M}_g^\gamma$ defined by $(X, h\gamma) \mapsto \pi^\gamma(Xh)$. This map is surjective, and descends to the quotient $P_\gamma/\operatorname{Mod}_g$. The resulting map $P_\gamma/\operatorname{Mod}_g \to \mathcal{M}_g^\gamma$ is a local isomorphism of symplectic orbifolds. All that remains now is to show that the map $P_\gamma/\operatorname{Mod}_g \to \mathcal{M}_g^\gamma$ is injective. Let $(X_1, h_1\gamma)$, $(X_2, h_2\gamma) \in P_\gamma$ such that $\pi^\gamma(X_1h_1) = \pi^\gamma(X_2h_2)$. By definition, there exists $s \in \operatorname{Stab}(\gamma)$ such that $X_1h_1s = X_2h_2$. Therefore

$$(X_2, h_2\gamma) = (X_1h_1sh_2^{-1}, h_2\gamma) \sim (X_1h_1s, \gamma) \sim (X_1h_1, \gamma),$$

which proves the injectivity. □

Remark 5.1.2. The length $\ell_X(\alpha)$, where $X \in \mathcal{M}_g^\gamma$, is not well-defined in general. However, it is if α has the same topological type as that of γ.

The next lemma is a simple fact, but for our purposes it will be very important: it transforms the multicurves counting that we are after to a "lattice points" counting problem on $\mathcal{M}_g^\gamma$.

Lemma 5.1.3. *Let $\gamma = (m_1\gamma_1, \dots, m_k\gamma_k)$ be an ordered multicurve, $X \in \mathcal{T}_g$, $R \in \mathbb{R}_{>0}$, and $A \subset \Delta^{k-1}$ be an open subset. The set*

$$\{h\gamma : h \in \mathrm{Mod}_g,\ \ell_X(h\gamma) \le R,\ \hat{\ell}_X(h\gamma) \in A\}$$

and the set

$$\{[(X, h\gamma)] \in \mathcal{M}_g^\gamma : h \in \mathrm{Mod}_g,\ \ell_X(h\gamma) \le R,\ \hat{\ell}_X(h\gamma) \in A\}$$

where by $[(X, h\gamma)]$ we mean the image of $(X, h\gamma)$ under $P \to P/\mathrm{Mod}_g$, are in bijection given by $h\gamma \mapsto [(X, h\gamma)]$,

Proof. The given map is obviously surjective. Suppose that $[(X, h_1\gamma)] = [(X, h_2\gamma)]$, then $h_1^{-1}h_2 \in \mathrm{Stab}(\gamma)$, and therefore $h_1\gamma = h_2\gamma$. The injectivity follows. □

Next, let us review another covering space of $\mathcal{M}_g$ that Mirzakhani introduced. By considering the Fenchel–Nielsen coordinates associated to a pants decomposition that contains $\gamma_1, \dots, \gamma_k$, the Teichmüller space $\mathcal{T}_g$ can be written as

$$\{(\ell_e, \tau_e, X_v) : e \in E,\ v \in V,\ \ell_e \in \mathbb{R}_{>0},\ \tau_e \in \mathbb{R},\ X_v \in \mathcal{T}_{g(v),n(v)}(\ell_{e(h)} : h \in H,\ v(h) = v)\} \quad (5.1)$$

where V (resp. E; H) is the vertex (resp. edge; half-edge) set of the stable graph associated to γ. The group

$$G_\gamma := \prod_e \mathbb{Z} \times \prod_v \mathrm{Mod}_{g(v),n(v)}$$

acts naturally on $\mathcal{T}_g$ written in the form (5.1) (each copy of $\mathbb{Z}$ acts as the Dehn twist about a γ_i), and G_γ can be identified with $\mathrm{Stab}^+(\gamma)$. The quotient $C_\gamma := \mathcal{T}_g/G_\gamma$ is of the form

$$\{(\ell_e, \tau_e, X_v) : e \in E,\ v \in V,\ \ell_e \in \mathbb{R}_{>0},\ \tau_e \in \mathbb{R}/\ell_e\mathbb{Z},\ X_v \in \mathcal{M}_{g(v),n(v)}(\ell_{e(h)} : h \in H,\ v(h) = v)\}.$$

Since $G_\gamma \simeq \mathrm{Stab}^+(\gamma)$ is a subgroup of $\mathrm{Stab}(\gamma)$, $\mathcal{T}_g \to \mathcal{M}_g^\gamma$ factors through a (ramified) covering map $C_\gamma \to \mathcal{M}_g^\gamma$. The degree of this covering map is

$$\kappa_\gamma = 2^{M(\gamma)} \cdot [\mathrm{Stab}(\gamma) : \langle \mathrm{Stab}^+(\gamma), \mathrm{Stab}_0(\gamma)\rangle] \quad (5.2)$$

where $M(\gamma)$ is the number of i such that γ_i bounds a surface homeomorphic to $\Sigma_{1,1}$, and $\langle \mathrm{Stab}^+(\gamma), \mathrm{Stab}_0(\gamma)\rangle$ stands for the subgroup of $\mathrm{Stab}(\gamma)$ generated by $\mathrm{Stab}^+(\gamma)$ and the kernel $\mathrm{Stab}_0(\gamma)$ of the action of $\mathrm{Stab}(\gamma)$ on $\mathcal{T}_g$. Note that $\mathrm{Stab}_0(\gamma)$ is trivial when $g \ge 3$, and is isomorphic to $\mathbb{Z}/2\mathbb{Z}$ if $g = 2$ (generated by the hyperelliptic involution which fixes the free homotopy class of every simple closed curve on Σ_2). For more details, see the footnote on p. 369–370 in [Wri20].

Integrating functions over C_γ (and $\mathcal{M}_g^\gamma$) is far less delicate than integrating function over $\mathcal{M}_g$. Starting from this observation Mirzakhani was able to calculate the integrals of an important class of functions defined on $\mathcal{M}_g$, that she called "geometric functions".

Theorem 5.1.4 (Mirzakhani's integration formula)**.** *Let $\gamma = (\gamma_1, \dots, \gamma_k)$ be an ordered multicurve, and $f\colon \mathbb{R}^k_{>0} \to \mathbb{R}$ be a measurable function. Let $X \in \mathcal{M}_g$, and choose a $\widetilde{X} \in \pi^{-1}(X) \in \mathcal{T}_g$. We define $f_\gamma\colon \mathcal{M}_g \to \mathbb{R}$ by the formula*

$$f_\gamma(X) := \sum_{(\alpha_1,\dots,\alpha_k)\in \mathrm{Mod}_g\cdot(\gamma_1,\dots,\gamma_k)} f(\ell_{\widetilde{X}}(\alpha_1), \dots, \ell_{\widetilde{X}}(\alpha_k))$$

Note that $f_\gamma(X)$ does not depend on the choice of $\widetilde{X}$. We have

$$\begin{aligned}\int_{\mathcal{M}_g} f_\gamma(X)\, dX &= \int_{\mathcal{M}_g^\gamma} f(\ell_X(\gamma_1), \dots, \ell_X(\gamma_k))\, dX \\ &= \kappa_\gamma \int_{\mathbb{R}^k_{>0}} f(x_1, \dots, x_k) \cdot P_\gamma(x_1, \dots, x_k)\, dx_1 \cdots dx_k.\end{aligned}$$

5.2 Horospheres

Let $\gamma = (m_1\gamma_1, \dots, m_k\gamma_k)$ be an ordered multicurve, $\overline{\gamma} = m_1\gamma_1 + \cdots + m_k\gamma_k$ be its unlabeled counterpart, and A be an open subset of the standard simplex $\Delta^{k-1} := \{(x_1, \dots, x_k) \in \mathbb{R}_{\geq 0} : x_1 + \cdots + x_k = 1\}$ of dimension $k-1$.

The *horosphere* of radius R associated to γ and A on $\mathcal{T}_g$ is defined by

$$\tilde{\mathcal{S}}^A_{R,\gamma} := \{X \in \mathcal{T}_g : \ell_X(\gamma) = R,\ \hat{\ell}_X(\gamma) \in A\}.$$

Similar notions can be defined on $\mathcal{M}_g^\gamma$ and on $\mathcal{M}_g$ by

$$\mathcal{S}^A_{R,\gamma} := \pi^\gamma(\tilde{\mathcal{S}}_L) \subset \mathcal{M}_g^\gamma, \qquad \bar{\mathcal{S}}^A_{R,\gamma} := \pi(\tilde{\mathcal{S}}_R) \subset \mathcal{M}_g$$

where $\pi^\gamma\colon \mathcal{T}_g \to \mathcal{M}_g^\gamma$ and $\pi\colon \mathcal{T}_g \to \mathcal{M}_g$ are the natural projections.

Remark 5.2.1. Let $\alpha = (m_1\alpha_1, \dots, m_k\alpha_k)$ be an ordered multicurve. The length vector $\ell_\alpha\colon X \mapsto (m_i\ell_X(\alpha_i))_{i=1}^k$ is well-defined for $X \in \mathcal{M}_g^\gamma$ if $\alpha_1 + \cdots + \alpha_k$ has the same topological type as $\gamma_1 + \cdots + \gamma_k$. The horosphere $\mathcal{S}^A_{R,\gamma} \subset \mathcal{M}_g^\gamma$ can be written as $\ell_\alpha^{-1}(R \cdot A)$ where $R \cdot A$ is defined by $\{(x_1, \dots, x_k) \in \mathbb{R}^k_{>0} : (x_1, \dots, x_k)/R \in A\}$.

5.2.1 Horospherical measures

We can choose $d-k$ simple closed curves $\alpha_{k+1}, \dots, \alpha_d$ such that $\{\gamma_1, \dots, \gamma_k, \alpha_{k+1}, \dots, \alpha_d\}$ is a pants decomposition. In the associated Fenchel–Nielsen coordinates, the horosphere $\tilde{\mathcal{S}}^A_{R,\gamma}$ is an open subset of a simplex. Let μ_Δ denote the Weil–Petersson (Lebesgue) measure on this simplex. The *horospherical measure* $\mu^A_{R,\gamma}$ of an open subset $U \subset \mathcal{T}_g$ is defined to be

$$\mu^A_{R,\gamma}(U) := \mu_\Delta(U \cap \tilde{\mathcal{S}}^A_{R,\gamma}).$$

The horospherical measure $\mu^A_{R,\gamma}$ is invariant under the action of the mapping class group, and hence descends to a measure on $\mathcal{M}_g^\gamma$ and a measure on $\mathcal{M}_g$l both of which by abuse of notation we shall also denote by $\mu^A_{R,\gamma}$. Note that $\mathcal{M}_g^\gamma \to \mathcal{M}_g$ is (ramified) covering map of infinite degree. However, its restriction on $\mathcal{S}^A_{R,\gamma}$ is of finite degree. Thus $\mu^A_{R,\gamma}$ on $\mathcal{M}_g$ is the push-forward measure

of $\mu^A_{R,\gamma}$ by $\mathcal{M}^\gamma_g \to \mathcal{M}_g$. So for any open subset U of $\mathcal{M}_g$,

$$\mu^A_{R,\gamma}(\pi_\gamma^{-1}(U)) = [\mathrm{Stab}(\overline{\gamma}) : \mathrm{Stab}(\gamma)] \cdot \mu^A_{R,\gamma}(U).$$

In particular, the total masses of $\mu^A_{R,\gamma}$ on $\mathcal{M}^\gamma_g$ and on $\mathcal{M}_g$ differ only by a multiplicative constant depending only on γ.

5.2.2 Total mass

The horospherical measure on $\mathcal{T}_g$ has infinite total mass. Nevertheless, its total mass is finite on $\mathcal{M}^\gamma_g$ and $\mathcal{M}_g$.

Proposition 5.2.2. *The total mass of $\mu^A_{R,\gamma}$ on $\mathcal{M}_g$ is*

$$M^A_{R,\gamma} = \frac{\kappa_\gamma}{[\mathrm{Stab}(\overline{\gamma}) : \mathrm{Stab}(\gamma)]} \frac{1}{m_1 \cdots m_k} \int_{R\cdot A} F_\gamma(x_1/m_k, \ldots, x_k/m_k)\, \lambda(dx)$$

where $R \cdot A := \{(x_1, \ldots, x_k) \in \mathbb{R}^k_{\geq 0} : (x_1, \ldots, x_k)/R \in A\}$, λ is the Lebesgue measure on $\Delta^{k-1}_R := \{(x_1, \ldots, x_k) \in \mathbb{R}_{\geq 0} : x_1 + \cdots + x_k = R\}$, and P_γ is defined by the formula (3.7).

Proof. In the light of Remark 5.2.1, by taking f in Theorem 5.1.4 to be the indicator function

$$\mathbb{1}\left\{(x_1, \ldots, x_k) \in \mathbb{R}^k_{>0} : \begin{array}{l} R \leq m_1x_1 + \cdots + m_kx_k \leq R + \epsilon, \\ (m_1x_1, \ldots, m_kx_k)/(m_1x_1 + \cdots + m_kx_k) \in A \end{array}\right\}$$

we obtain that $\mu^A_{R,\gamma}(\mathcal{M}_g) \cdot [\mathrm{Stab}(\overline{\gamma}) : \mathrm{Stab}(\gamma)]$ is equal to

$$\mu^A_{R,\gamma}(\mathcal{M}^\gamma_g) = \lim_{\epsilon \to 0} \frac{1}{\epsilon} \int_{\mathcal{M}_g} f_\gamma(X)\, dX = \frac{\kappa_\gamma}{m_1 \cdots m_k} \int_{\Delta_{R\cdot A}} F_\gamma(x_1/m_1, \ldots, x_k/m_k)\, \lambda(dx),$$

the result desired. □

Corollary 5.2.3. *The total mass $M^A_{R,\gamma}$ (of $\mu^A_{R,\gamma}$ on $\mathcal{M}_g$) is a polynomial in R of degree $2d-1 = 6g-7$. Write C^A_γ for its leading coefficient. We have*

$$M^A_{R,\gamma} \sim C^A_\gamma \cdot R^{2d-1} \tag{5.3}$$

as $R \to \infty$, and C^A_γ can be calculated by

$$C^A_\gamma = \frac{\kappa_\gamma}{[\mathrm{Stab}(\overline{\gamma}) : \mathrm{Stab}(\gamma)]} \frac{1}{m_1 \cdots m_k} \int_A \bar{F}_\gamma(x_1/m_1, \ldots, x_k/m_k)\, \lambda(dx),$$

where λ is the Lebesgue measure on Δ^{k-1} and $\bar{F}_\gamma$ is the top-degree homogeneous part of the graph polynomial F_γ defined by (3.7).

Remark 5.2.4. It results from Theorem 3.2.7 that the polynomial $\bar{F}_\gamma$ can be expressed in terms of intersections numbers of ψ-classes on the Deligne–Mumford compactification $\overline{\mathcal{M}}_{g,n}$.

5.2.3 Horospherical measures on the unit sphere bundle

Let ν_Δ denote the Lebesgue measure on

$$\mathcal{P}\tilde{\mathcal{S}}^A_{R,\gamma} := \{(X, \gamma/R) \in \mathcal{P}^1\mathcal{T}_g : \hat{\ell}_X(\gamma) \in A\}.$$

Note that $\mathcal{P}\tilde{\mathcal{S}}^A_{R,\gamma}$ projects via $\mathcal{P}^1\mathcal{T}_g \to \mathcal{T}_g$ to $\tilde{\mathcal{S}}^A_{R,\gamma}$, and is invariant under the earthquake flow. Let ν_Δ denote the Lebesgue measure on $\mathcal{P}\tilde{\mathcal{S}}^A_{R,\gamma}$. The *horospherical measure* $\nu^A_{R,\gamma}$ on $\mathcal{P}\mathcal{T}_g$ is defined by the formula

$$\nu^A_{R,\gamma}(U) := \nu_\Delta(U \cap \mathcal{P}\tilde{\mathcal{S}}^A_{R,\gamma})$$

where U is any open subset of $\mathcal{P}^1\mathcal{T}_g$. The measure $\nu^A_{R,\gamma}$ is Mod_g-invariant, and therefore descends to a measure on $\mathcal{P}^1\mathcal{M}_g$ which by abuse of notation we shall also denote by $\nu^A_{R,\gamma}$. Note that $\mu^A_{R,\gamma}$ is the push-forward of $\nu^A_{R,\gamma}$ via $\mathcal{P}^1\mathcal{M}_g \to \mathcal{M}_g$.

Notation

To simplify the notation, let us fix $X \in \mathcal{M}_g$, a multicurve $\gamma = (m_1\gamma_1, \ldots, m_k\gamma_k)$ on Σ_g, and an open subset A of the standard simplex $\Delta^{k-1} := \{(x_1, \ldots, x_k) \in \mathbb{R}_{\geq 0} : x_1 + \cdots + x_k = 1\}$. From now on we shall write μ_R for $\mu^A_{R,\gamma}$, ν_R for $\nu^A_{R,\gamma}$, and M_R for $M^A_{R,\gamma}$, unless otherwise stated.

5.3 Equidistribution

In this section, we establish the equidistribution of large horospheres.

Theorem 5.3.1. *We have weak convergence of probability measures on $\mathcal{P}^1\mathcal{M}_g$*

$$\frac{\nu_R}{M_R} \Rightarrow \frac{\nu_g}{b_g}$$

as $R \to \infty$.

The following immediate corollary is exceedingly useful late on:

Corollary 5.3.2. *We have weak convergence of probability measures on $\mathcal{M}_g$*

$$\frac{\mu_R}{M_R} \Rightarrow \frac{B(X)}{b_g}\,\mu_{\mathrm{WP}}$$

as $R \to \infty$.

Proof. This follows from the fact that μ_R is the push-forward of ν_R via $\mathcal{P}^1\mathcal{M}_g \to \mathcal{M}_g$ and Theorem 5.3.1. □

The proof of Theorem 5.3.1 rests on the following series of propositions.

Let ν be a weak limit of $(\nu_R/M_R)_{R>0}$.

Proposition 5.3.3. *The measure ν is invariant under the earthquake flow.*

Proposition 5.3.4. *The measure ν is absolutely continuous with respect to ν_g.*

Proposition 5.3.5. *The measure ν is a probability measure.*

Proof of Theorem 5.3.1. Proposition 5.3.3, Proposition 5.3.4, and Theorem 5.3.1 imply that ν and ν_g differ by a multiplicative constant, and it follows from Proposition 5.3.5 that this constant is 1. □

Proposition 5.3.3 is immediate.

Proof of Proposition 5.3.3. This follows from the fact that ν_R is invariant under the earthquake flow (since ν_g is). □

For the rest of this section we shall prove Proposition 5.3.4 and Proposition 5.3.5, which are more technical.

5.3.1 Escape to infinity?

In this subsection, we prove Proposition 5.3.5. The key ingredient is the following non-divergence result for the earthquake flow due to Minsky and Weiss.

Theorem 5.3.6 ([MW02, Theorem E2], [Mir07a, Corollary 5.12])**.** *For any $c > 0$, there exists $\epsilon > 0$, depending only on c, such that for any $x \in \mathcal{T}_g$ and any $\lambda \in \mathcal{ML}_g$, the following dichotomy holds:*

1. *There exists a simple closed curve α disjoint from λ, and $\ell_x(\alpha) < \epsilon$.*
2. *We have*
$$\liminf_{T\to\infty} \frac{|\{t \in [0,T] : \pi(\mathrm{tw}^t_\lambda(x)) \in \mathcal{M}_g^{\geq\epsilon}\}|}{T} > 1 - c$$
where $\pi\colon \mathcal{T}_g \to \mathcal{M}_g$ is the natural projection, and $\mathcal{M}_g^{\geq\epsilon}$ is the compact subset of $\mathcal{M}_g$ consisting of all surfaces whose shortest closed geodesic has length at least ϵ.

Proof of Proposition 5.3.5. It is enough to prove that for any $\delta > 0$, we can find a compact subset K_δ of $\mathcal{P}^1\mathcal{M}_g$ such that
$$\liminf_{R\to\infty} \frac{\nu_R(K_\delta)}{M_R} \geq 1 - \delta.$$
The strategy is to show that there exists $\epsilon > 0$, depending only on δ, such that the pre-image of $\mathcal{M}_g^{\geq\epsilon}$ under $\mathcal{P}^1\mathcal{M}_g \to \mathcal{M}_g$ possess the desired property. In other words,
$$\liminf_{R\to\infty} \frac{\mu_R(\mathcal{M}_g^{\geq\epsilon})}{M_R} \geq 1 - \delta.$$

Taking $c = \delta/2$, Theorem 5.3.6 allows us to write $\tilde{\mathcal{S}}_R \subset \mathcal{T}_g$ as the disjoint union of $\tilde{\mathcal{S}}_1$ and $\tilde{\mathcal{S}}_2$ corresponding to the two possibilities. For convenience, we shall adapt the convention that $\bar{\mathcal{S}}_*$ (resp. $\mathcal{S}_*$) denotes the image of $\tilde{\mathcal{S}}_*$ under $\mathcal{T}_g \to \mathcal{M}_g$ (resp. $\mathcal{T}_g \to \mathcal{M}_g^\gamma$), where $*$ is a certain index.

First, we show that $\mu_R(\bar{\mathcal{S}}_1) \leq \mu_R(\mathcal{S}_1) = \mathrm{o}(M_R)$ as $R \to \infty$ even when $A = \Delta^{k-1}$ (the subset of the simplex that we choose to define μ_R is the whole simplex). For any point in $\tilde{\mathcal{S}}_1$, at least one of the following holds:

1.1. α is freely homotopic to γ_i for some $1 \leq i \leq k$.

1.2. α is disjoint from $\gamma_1, \ldots, \gamma_k$.

Thus $\tilde{\mathcal{S}}_1$ can be written as the union of $\tilde{\mathcal{S}}_{1,1}$ and $\tilde{\mathcal{S}}_{1,2}$ corresponding to the two cases above. To simplify the notation, in the following estimates of $\mu_R(\mathcal{S}_{1,1})$ and $\mu_R(\mathcal{S}_{1,2})$ we assume that γ is primitive, i.e. $m_1 = \cdots = m_k = 1$ (the calculation differs from the general case only by a multiplicative constant).

For each i, the corresponding horospherical volume of $\mathcal{S}_{1,1}$ can be estimated by taking f in Theorem 5.1.4 to be the indicator function

$$\mathbb{1}\{(x_1, \dots, x_k) \in \mathbb{R}^k_{\geq 0} : R \leq x_1 + \cdots + x_k \leq R + h,\ x_i < \epsilon\},$$

and we obtain

$$\mu_R(\mathcal{S}_{1,1}) \leq \sum_{i=1}^{k} \lim_{h \to 0} \frac{\kappa_\gamma}{h} \int_0^\epsilon dx_i \int_{\Delta^{k-1}_{[R-x_i, R+h-x_i]}} dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_k\, P_\gamma(x_1, \dots, x_k),$$

where $\Delta^{k-1}_{[R-x_i, R+h-x_i]} := \{(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_k) \in \mathbb{R}^{k-1}_{\geq 0} : R \leq x_1 + \cdots + x_k \leq R + h\}$. Since P_γ is a polynomial of degree $2d-k$ and $x_1 \cdots x_k$ is a factor of P_γ, we have $\mu_R(\mathcal{S}_{1,1}) = \mathrm{O}(\epsilon^2 R^{2d-3})$.

We now suppose that α is disjoint from $\gamma_1, \dots, \gamma_k$. Denote by (γ, α) the ordered multicurve $(\gamma_1, \dots, \gamma_k, \alpha)$. Again, by applying Theorem 5.1.4, f being the indicator function

$$\mathbb{1}\{(x_1, \dots, x_k, y) \in \mathbb{R}^{k+1}_{>0} : R \leq x_1 + \cdots + x_k \leq R + h,\ y < \epsilon\},$$

we obtain the corresponding horospherical volume

$$\lim_{h \to 0} \frac{\kappa_\gamma}{h} \int_0^\epsilon dy \int_{\Delta^k_{[R, R+h]}} dx_1 \cdots dx_k\, P_{(\gamma,\alpha)}(x_1, \dots, x_k, y)$$

where $\Delta^k_{[R,R+h]} := \{(x_1, \dots, x_k) \in \mathbb{R}^k_{\geq 0} : R \leq x_1 + \cdots + x_k \leq R + h\}$. Since $P_{(\gamma,\alpha)}$ is a polynomial of degree $2d - k - 1$ of which y is a factor, and there are only finitely many topological types of (γ, α), we have $\mu_R(\mathcal{S}_{1,2}) = \mathrm{O}(\epsilon^2 R^{2d-3})$.

Using Corollary 5.2.3 we deduce

$$\frac{\mu_R(\bar{\mathcal{S}}_1)}{M_R} \leq \frac{\mu_R(\bar{\mathcal{S}}_{1,1}) + \mu_R(\bar{\mathcal{S}}_{1,2})}{M_R} \leq \frac{\mu_R(\mathcal{S}_{1,1}) + \mu_R(\mathcal{S}_{1,2})}{M_R} = \mathrm{O}(\epsilon^2 R^{-2}) = \mathrm{o}(1)$$

as $R \to \infty$.

Let us now consider $\bar{\mathcal{S}}_2$. One observes that every $p \in \bar{\mathcal{S}}_2$ lies in a unique 1-periodic earthquake flow orbit along the direction $\gamma_p := \gamma_1/\ell_p(\gamma_1) + \cdots + \gamma_k/\ell_p(\gamma_k)$, and $\bar{\mathcal{S}}_2$ can be written as the disjoint union of such orbits. (If one completes γ to a pants decomposition by adding $d - k$ simple curves, such orbits are parallel straight lines in the $(\tau_{\gamma_1}, \dots, \tau_{\gamma_k})$-coordinates plane.) By Theorem 5.3.6,

$$|\{t \in [0,1] : \pi(\mathrm{tw}^t_{\gamma_p}(p)) \in \mathcal{M}^{\geq \epsilon}_g\}| > 1 - \delta/2,$$

for all $p \in \mathcal{S}_2$. Thus,

$$\frac{\mu_R(\bar{\mathcal{S}}_2 \cap \mathcal{M}^{\geq \epsilon}_g)}{\mu_R(\bar{\mathcal{S}}_2)} \geq 1 - \delta/2.$$

Therefore,

$$\begin{aligned}\frac{\mu_R(\mathcal{M}_g^{\geq\epsilon})}{M_R} &= \frac{\mu_R(\mathcal{M}_g^{\geq\epsilon}\cap\bar{\mathcal{S}}_1)+\mu_R(\mathcal{M}_g^{\geq\epsilon}\cap\bar{\mathcal{S}}_2)}{M_R}\\ &= 0+\frac{\mu_R(\mathcal{M}_g^{\geq\epsilon}\cap\bar{\mathcal{S}}_2)}{\mu_R(\bar{\mathcal{S}}_2)}\frac{M_R-\mu_R(\bar{\mathcal{S}}_1)}{M_R}\geq(1-\delta/2)(1-\mathrm{o}(1))\end{aligned}$$

as $R\to\infty$. The $\mathrm{o}(1)$ term can be made, e.g. smaller than $\delta/2$, by increasing R. The proof is thus complete. □

5.3.2 Absolute continuity

In this subsection, we prove Proposition 5.3.4.

We use the following notation throughout the subsection. Let d denote $3g-3$. We write $f=\mathrm{O}_K(g)$ if there exists $C>0$, depending only on K, such that $f\leq Cg$, and we write $f=\Theta_K(g)$ if there exists C, depending only on K, such that $(1/C)g\leq f\leq Cg$.

The key to the proof are the following estimates.

Theorem 5.3.7 ([Mir07a], [Mir16], [AH21b])**.** *Let $\epsilon\in(0,1)$, and let K be a compact subset of $\mathcal{M}_g$. Then:*

1. *For any $x\in K$, $\mu_{\mathrm{WP}}(\mathbb{B}_x(\epsilon))=\Theta_K(\epsilon^{2d})$.*
2. *For any $x\in\pi^{-1}(K)\subset\mathcal{T}_g$, $\mu_R(\mathbb{B}_x(\epsilon))=\mathrm{O}_K(\epsilon^{2d-1}/R)$.*

Remark 5.3.8. The first part of the preceding theorem is [Mir07a, Theorem 5.5.a]. Mirzakhani proved the second part in the case when γ is a simple closed curve [Mir07a, Theorem 5.5.b], and claimed a more general version [Mir16, Proposition 2.1.b] without proof. The proof of [Mir07a, Theorem 5.5.b] is concise and not easy to follow. See also the footnote on p. 390 in [Wri20]. A much stronger estimate is obtained by Arana-Herrera in a different approach [AH21b, Proposition 1.5].

The rest of the proof of Proposition 5.3.4 can be adapted from Mirzakhani's original proof in the case when γ is simple. Let us sketch her arguments for the sake of self-containedness.

Corollary 5.3.9. *Let $U\subset\mathbb{P}(\mathcal{ML}_g)$ be open, $K\subset\mathcal{T}_g$ be compact, $x\in K$, and $p\colon\mathcal{T}_g\times\mathbb{P}(\mathcal{ML}_g)\to\mathcal{P}^1\mathcal{M}_g$ be the natural projection. For $\epsilon\in(0,1)$, we have*

$$\frac{\nu_R(p(\mathbb{B}_x(\epsilon)\times U))}{M_R}=\mathrm{O}_K(\nu_g(\mathbb{B}_x(\epsilon)\times U_x))$$

where $U_x\coloneqq\{\lambda\in\mathcal{ML}_g:\ell_x(\lambda)\leq1,\ [\lambda]\in U\}$.

Proof. It is enough to prove this for $A=\Delta^{k-1}$. By (3.2), for any $y\in\mathbb{B}_x(\epsilon)$, we have

$$(1-\epsilon)^{2d}\cdot\mu_{\mathrm{Th}}(U_x)\leq\mu_{\mathrm{Th}}(U_y)\leq(1+\epsilon)^{2d}\cdot\mu_{\mathrm{Th}}(U_x),\tag{5.4}$$

and so

$$\begin{aligned}&\#\{\alpha\in\mathrm{Mod}_g\cdot\gamma:[\alpha]\in U,\ \ell_y(\alpha)=R\text{ for some }y\in\mathbb{B}_x(\epsilon)\}\\ &\leq\#\{\alpha\in\mathrm{Mod}_g\cdot\gamma:[\alpha]\in U,\ (1-\epsilon)R\leq\ell_x(\alpha)\leq(1+\epsilon)R\}\\ &=\mathrm{O}_K(\epsilon R^{2d}\mu_{\mathrm{Th}}(U_x)).\end{aligned}$$

Hence Theorem 5.3.7.2 implies that $\nu_R(p(\mathbb{B}_x(\epsilon) \times U)) = \mathrm{O}_K(\epsilon^{2d} R^{2d-1} \mu_{\mathrm{Th}}(U_x))$. The result now follows from Theorem 5.3.7.1 and Corollary 5.2.3. □

We need one further technical lemma.

Lemma 5.3.10. *Let K be a compact subset of $\mathcal{P}^1\mathcal{T}_g$. For any $N \subset K$ with $\nu_g(N) = 0$, and any $\epsilon > 0$, there exists an open cover $\{\mathbb{B}_{X_i}(r_i) \times U_i : i \in \mathbb{Z}_{\geq 1}\}$ of N, where for all i, $X_i \in \mathcal{T}_g$, $r_i \in (0,1)$, and $U_i \subset \mathbb{P}(\mathcal{ML}_g)$ is open, such that*

$$\sum_{i \geq 1} \nu_g(\mathbb{B}_{X_i}(r_i) \times U_i) \leq \epsilon.$$

Proof. Fix a choice of Fenchel–Nielsen coordinates. There exists an open cover $\{B_{X_i}(r_i) \times U_i : i \in \mathbb{Z}_{\geq 1}\}$ of N, where $B_{X_i}(r_i)$ is the Euclidean ball of radius r_i centered at X_i, such that $\sum_{i \geq 1} \nu_g(B_{X_i}(r_i) \times U_i) \leq \epsilon$, and $\sup_{i \geq 1} r_i$ can be made as small as we please (since ν_g is a Lebesgue class measure). It follows from the compactness of $K \times [0,1]$ that there exists a constant s depending only on K such that $B_x(r) \subset \mathbb{B}_x(s \cdot r)$ for any $x \in K$ and any $r \in [0,1]$. By Theorem 5.3.7.1, there exists a constant s' depending only on K such that $\mu_{\mathrm{WP}}(\mathbb{B}_x(s \cdot r)) \leq s' \cdot \mu_{\mathrm{WP}}(B_x(r))$ for any $x \in K$, and any $r < 1/2s$. Therefore, by (5.4)

$$\sum_{i \geq 1} \nu_g(\mathbb{B}_{X_i}(s \cdot r_i) \times U_i) \leq 3^{2d} s' \sum_{i \geq 1} \nu_g(B_{X_i}(r_i) \times U_i) \leq 3^{2d} s' \epsilon = \mathrm{O}_K(\epsilon)$$

and the lemma follows. □

Proof of Proposition 5.3.4. It is sufficient, as before, to consider the case in which A is the whole simplex Δ^{k-1}. Let $N \subset \mathcal{P}^1\mathcal{M}_g$ with $\nu_g(N) = 0$. By Proposition 5.3.5, we may assume that N is contained in a compact set $K \subset \mathcal{P}^1\mathcal{M}_g$. Lemma 5.3.10 implies for any $\epsilon > 0$, there exists an open cover $\{\mathbb{B}_{X_i}(r_i) \times U_i : i \in \mathbb{Z}_{\geq 1}\}$ of N, such that

$$\sum_{i \geq 1} \nu_g(\mathbb{B}_{X_i}(r_i) \times U_i) \leq \epsilon.$$

Hence, it follows from Corollary 5.3.9 that

$$\sum_{i \geq 1} \frac{\nu_R(\mathbb{B}_{X_i}(r_i) \times U_i)}{M_R} = \sum_{i \geq 1} \mathrm{O}_K(\nu_g(\mathbb{B}_{X_i}(r_i) \times U_{X_i})) \leq \mathrm{O}_K(\epsilon).$$

The proof is thus complete. □

5.4 Counting

The main result of this section is the following theorem which is a refined version of [Mir08b, Theorem 1.1].

Theorem 5.4.1. *Let $X \in \mathcal{M}_g$, $\gamma = (m_1\gamma_1, \ldots, m_k\gamma_k)$ be an ordered multicurve, and $A \subset \Delta^{k-1}$ be open. We have*

$$\#\{\alpha \in \mathrm{Mod}_g \cdot \gamma : \ell_X(\alpha) \leq R,\ \hat{\ell}_X(\alpha) \in A\} \sim C_\gamma^A \frac{[\mathrm{Stab}(\overline{\gamma}) : \mathrm{Stab}(\gamma)]}{2d} \frac{B(X)}{b_g} R^{2d}.$$

as $R \to \infty$.

By virtue of Lemma 5.1.3, this multicurves counting problem can be transformed to a counting problem on $\mathcal{M}_g^\gamma$. Let us begin by introducing some definitions that we need to state our counting result on $\mathcal{M}_g^\gamma$.

The *horoball* (on $\mathcal{M}_g^\gamma$) is defined by

$$\mathcal{B}_R := \{X \in \mathcal{M}_g^\gamma : \ell_X(\gamma) \leq R,\ \hat{\ell}_X(\gamma) \in A\} = \bigcup_{0<r\leq R} \mathcal{S}_r$$

and its associated measure $\mu_{\leq R}$ is defined by the formula

$$\mu_{\leq R}(U) := \int_0^R \mu_r(U)\, dr = \mu_{\mathrm{WP}}(U \cap \mathcal{B}_R)$$

where U is any open subset of $\mathcal{M}_g^\gamma$. By abuse of notation, we shall also use $\mu_{\leq R}$ to denote the measure on $\mathcal{M}_g$ defined by the formula

$$\nu_{\leq R}(U) := \int_0^R \mu_r(U)\, dr = \mu_{\mathrm{WP}}(U \cap \pi_\gamma(\mathcal{B}_R)),$$

for any open subset U of $\mathcal{M}_g$. Let $X \in \mathcal{M}_g$ and let $N(R)$ denote the number of pre-images of X under $\pi_\gamma \colon \mathcal{M}_g^\gamma \to \mathcal{M}_g$ which lie within the horoball $\mathcal{B}_R \subset \mathcal{M}_g^\gamma$, i.e., $N(R) := \#\{\pi_\gamma^{-1}(X) \cap \mathcal{B}_R\}$.

We have the following counting result on $\mathcal{M}_g^\gamma$.

Theorem 5.4.2. *Let* $X \in \mathcal{M}_g$, $\gamma = (m_1\gamma_1, \ldots, m_k\gamma_k)$ *be an ordered multicurve, and* $A \subset \Delta^{k-1}$ *be open. Then we have*

$$N(R) \sim C_\gamma^A \frac{[\mathrm{Stab}(\overline{\gamma}) : \mathrm{Stab}(\gamma)]}{2d} \frac{B(X)}{b_g} R^{2d}.$$

as $R \to \infty$.

As an immediate corollary, we get the main result of the section:

Proof of Theorem 5.4.1. This follows at once from Theorem 5.4.2 and Lemma 5.1.3. □

We introduce a family of subsets $A_{a,b}$ of Δ^{k-1}, indexed by $a = (a_1, \ldots, a_{k-1}) \in [0,1]^{k-1}$ and $b = (b_1, \ldots, b_{k-1}) \in [0,1]^{k-1}$ such that $a_i < b_i$ for all $1 \leq i \leq k-1$, and defined by

$$A_{a,b} := \left\{(x_1, \ldots, x_k) \in \Delta^{k-1} : a_i \leq x_i \leq b_i,\ \forall 1 \leq i \leq k-1\right\}.$$

To prove Theorem 5.4.2, it is enough to check the case when $A = A_{a,b}$ for all a, b. In order to abbreviate our formulas, for the rest of this section we write

$$A := A_{a,b}, \qquad A_+ := A_{\frac{1-\epsilon}{1+\epsilon}a, \frac{1+\epsilon}{1-\epsilon}b}, \qquad A_- := A_{\frac{1+\epsilon}{1-\epsilon}a, \frac{1-\epsilon}{1+\epsilon}b},$$

where we adopt the convention that $\frac{1+\epsilon}{1-\epsilon}b_i = 1$ if $\frac{1+\epsilon}{1-\epsilon}b_i > 1$, and we write $\mathcal{B}_R^+ := \mathcal{B}_R^{A_+}$, $\mu_{\leq R}^+ := \mu_{\leq R}^+$, etc. The reason for the choice of A_+ and A_- is the following elementary lemma.

Lemma 5.4.3. *Choose* $\epsilon \in (0,1)$ *small enough to ensure that* A_- *and* A_+ *are well-defined, and let* $x, y \in \mathcal{M}_g^\gamma$ *with* $d_{\mathrm{Th}}(x,y) \leq \epsilon$. *We have*

(1). If $x \in \mathcal{B}^-_{(1-\epsilon)R}$, then $y \in \mathcal{B}_R$,

(2). If $x \in \mathcal{B}_R$, then $y \in \mathcal{B}^+_{(1+\epsilon)R}$.

Proof. Suppose that $x \in \mathcal{B}_R$. It follows from the inequality (3.2) that

$$\ell_y(\gamma) \leq (1+\epsilon)\ell_x(\gamma) \leq (1+\epsilon)R$$

and

$$\frac{1-\epsilon}{1+\epsilon}\, a_i \leq \frac{(1-\epsilon)\ell_x(m_i\gamma_i)}{(1+\epsilon)\ell_x(\gamma)} \leq \frac{\ell_y(m_i\gamma_i)}{\ell_y(\gamma)} \leq \frac{(1+\epsilon)\ell_x(m_i\gamma_i)}{(1-\epsilon)\ell_x(\gamma)} \leq \frac{1+\epsilon}{1-\epsilon}\, b_i,$$

which shows that $y \in \mathcal{B}^+_{(1+\epsilon)R}$. Part (1) of the lemma can be proved in a similar manner. □

Now we are ready to prove our main result of the section.

Proof of Theorem 5.4.2. We can choose $\epsilon \in (0,1)$ such that $\mathbb{B}_{Y_1}(\epsilon) \cap \mathbb{B}_{Y_2}(\epsilon) = \varnothing$ for any distinct pre-images Y_1, Y_2 of X under $\pi_\gamma \colon \mathcal{M}_g^\gamma \to \mathcal{M}_g$. Let us write

$$N_-(R) := \#\{Y \in \pi_\gamma^{-1}(X) \subset \mathcal{M}_g^\gamma : \mathbb{B}_Y(\epsilon) \subset \mathcal{B}_R\},$$

for the set of all $Y \in \mathcal{M}_g^\gamma$ such that Y projects to X and the Thurston distance ball of radius ϵ centered at Y is entirely included within the horoball $\mathcal{B}_R \subset \mathcal{M}_g^\gamma$. Furthermore, we write

$$N_+(R) := \#\{Y \in \pi_\gamma^{-1}(X) \subset \mathcal{M}_g^\gamma : \mathbb{B}_Y(\epsilon) \cap \mathcal{B}_R \neq \varnothing\}$$

for the set of all $Y \in \mathcal{M}_g^\gamma$ that project to X such that $\mathbb{B}_Y(\epsilon)$ intersects $\mathcal{B}_R$. By definition,

$$N_-(R) \leq N(R) \leq N_+(R).$$

It follows from Lemma 5.4.3 that

$$N_+(R) \cdot \mu_{\mathrm{WP}}(\mathbb{B}_X(\epsilon)) \leq \mu_{\mathrm{WP}}(\pi_\gamma^{-1}(\mathbb{B}_X(\epsilon)) \cap \mathcal{B}^+_{(1+\epsilon)R}) = \mu^+_{\leq(1+\epsilon)R}(\pi_\gamma^{-1}(\mathbb{B}_X(\epsilon))) \tag{5.5}$$

and

$$\mu^-_{(1-\epsilon)R}(\pi_\gamma^{-1}(\mathbb{B}_X(\epsilon))) = \mu_{\mathrm{WP}}(\pi_\gamma^{-1}(\mathbb{B}_X(\epsilon)) \cap \mathcal{B}^-_{(1-\epsilon)R}) \leq N_-(R) \cdot \mu_{\mathrm{WP}}(\mathbb{B}_X(\epsilon)). \tag{5.6}$$

For any open subset $U \subset \mathcal{M}_g$,

$$\mu_{\leq R}(\pi_\gamma^{-1}(U)) = [\mathrm{Stab}(\overline{\gamma}) : \mathrm{Stab}(\gamma)] \cdot \mu_{\leq R}(U), \tag{5.7}$$

We deduce from (5.5), (5.6), and (5.7) that

$$\mu^-_{\leq(1-\epsilon)R}(\mathbb{B}_X(\epsilon)) \leq \frac{N(R) \cdot \mu_{\mathrm{WP}}(\mathbb{B}_X(\epsilon))}{[\mathrm{Stab}(\overline{\gamma}) : \mathrm{Stab}(\gamma)]} \leq \mu^+_{\leq(1+\epsilon)R}(\mathbb{B}_X(\epsilon)).$$

where $\mathbb{B}_X(\epsilon) \subset \mathcal{M}_g$. Hence,

$$\lim_{R\to\infty} \frac{\mu^+_{(1+\epsilon)R}(\mathbb{B}_X(\epsilon))}{R^{2d}} = \lim_{R\to\infty} \frac{1}{R} \int_0^{(1+\epsilon)R} \frac{\mu^+_t(\mathbb{B}_X(\epsilon))}{R^{2d-1}}\, dt \tag{5.8}$$

$$= C^+ \cdot \lim_{R\to\infty} \frac{1}{R^{2d}} \int_0^{(1+\epsilon)L} t^{2d-1} \frac{\mu^+_t(\mathbb{B}_X(\epsilon))}{C^+ \cdot t^{2d-1}}\, dt \tag{5.9}$$

where $C^+ := C^{A+}_\gamma$ is given by (5.3). By Corollary 5.3.2 and 5.2.3,

$$\frac{\mu^+_t(\mathbb{B}_X(\epsilon))}{C^+ \cdot t^{2d-1}} = \frac{1}{b_g} \int_{\mathbb{B}_X(\epsilon)} B(Y)\, dY + \mathrm{o}(1)$$

as $t \to \infty$, where $\mathrm{o}(1)$ is bounded and the constant depends only on γ and A^+. Thus (5.9) is equal to

$$\frac{(1+\epsilon)^{2d}\, C^+}{2d\, b_g} \int_{\mathbb{B}_X(\epsilon)} B(Y)\, dY.$$

Therefore,

$$\frac{\mu_{\mathrm{WP}}(\mathbb{B}_X(\epsilon))}{[\mathrm{Stab}(\overline{\gamma}) : \mathrm{Stab}(\gamma)]} \limsup_{R\to\infty} \frac{N(R)}{R^{2d}} \le \frac{(1+\epsilon)^{2d}\, C^+}{2d\, b_g} \int_{\mathbb{B}_X(\epsilon)} B(Y)\, dY,$$

and similarly,

$$\frac{(1-\epsilon)^{2d}\, C^-}{2d\, b_g} \int_{\mathbb{B}_X(\epsilon)} B(Y)\, dY \le \frac{\mu_{\mathrm{WP}}(\mathbb{B}_X(\epsilon))}{[\mathrm{Stab}(\overline{\gamma}) : \mathrm{Stab}(\gamma)]} \liminf_{R\to\infty} \frac{N(R)}{R^{2d}}$$

where $C^- := C^{A-}_\gamma$. Taking $\epsilon \to 0$, we obtain

$$\lim_{R\to\infty} \frac{N(R)}{R^{2d}} = C^A_\gamma \frac{[\mathrm{Stab}(\overline{\gamma}) : \mathrm{Stab}(\gamma)]}{2d} \frac{B(X)}{b_g}.$$

This established the theorem. □

5.5 Statistics

Proof of Theorem 5.0.2. Theorem 5.4.1 implies

$$\lim_{R\to\infty} \mathbb{P}(\hat{\ell}_{X,\gamma,R} \in A) = \frac{C^A_\gamma}{C^{\Delta^{k-1}}_\gamma}.$$

The assertion now follows from Corollary 5.2.3. □

Example 5.5.1. If $\gamma = (\gamma_1, \ldots, \gamma_{3g-3})$ is a pants decomposition, then $g(v) = 0$, $n(v) = 3$, and $V_{g(v),n(v)} = 1$ for all $v \in V$. Thus $F_\gamma(x_1, \ldots, x_{3g-3}) = x_1 \cdots x_{3g-3}$, and Theorem 5.0.2 reduces to Theorem 5.0.1.

Example 5.5.2. Let $\gamma = (\gamma_1, \gamma_2)$, where γ_2 is separating and separates Σ_g into a torus with a hole and a surface of type $(g-1, 1)$, and γ_1 sits on the torus with a hole is non-separating as in

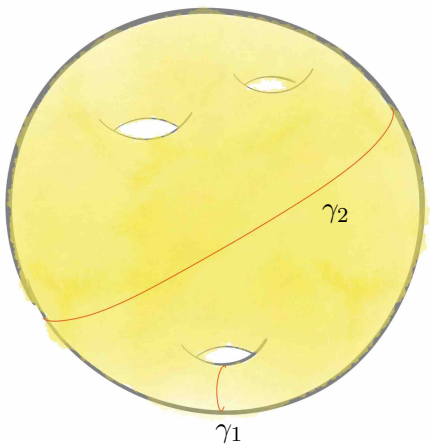

Figure 5.1: Example 5.5.1

Figure 5.1. Then its associated graph polynomial $\bar{F}_\gamma$ is equal to

$$x_1 x_2 \cdot \bar{V}_{0,3}(x_1, x_1, x_2) \cdot \bar{V}_{g-1,1}(x_2) = \text{constant} \cdot x_1 x_2^{6g-9}.$$

This implies that in a random multi-geodesic of topological type (γ_1, γ_2) on a hyperbolic surface of genus $g \gg 2$, the separating component is very likely to be much longer than the non-separating component.

Example 5.5.3. Let (γ_1, γ_2) be an ordered multicurve such that, for $i = 1, 2$, γ_i is separating, and γ_i bounds two surfaces of type $(g_i, 1)$ (genus g_i with 1 boundary component) and $(g - g_1 - g_2, 2)$ respectively, as shown in Firgure 5.2. Then $\bar{F}_\gamma$ is

$$x_1 x_2 \cdot \bar{V}_{g_1,1}(x_1)\, \bar{V}_{g_2,1}(x_2)\, \bar{V}_{g-g_1-g_2,2}(x_1, x_2) = \text{constant} \cdot x_1^{6g_1-3} x_2^{6g_2-3} \cdot \bar{V}_{g-g_1-g_2,2}(x_1, x_2)$$

where $\bar{V}_{g-g_1-g_2,2}$ is a symmetric polynomial. So in a typical multi-geodesic of type (γ_1, γ_2), the first component is shorter than the second if $g_1 < g_2$.

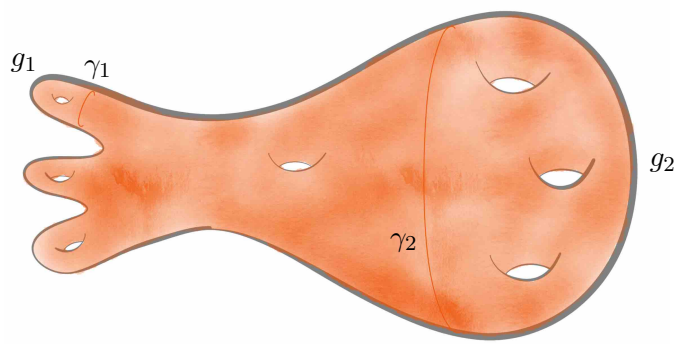

Figure 5.2: Example 5.5.2

Chapter 6

Large genus asymptotics

The essential content of this chapter is adapted from [DL22], which is a work joint with Vincent Delecroix.

6.1 Introduction

6.1.1 Main result

In the last chapter we have determined the distribution of the normalized length vector of a random multi-geodesic of given topological type. In what follows we use Theorem 5.0.2 to study the length partition of a random multi-geodesic without any topological constraints.

The components of a multicurve do not come with any natural labels, can not be putted into a vector in any natural way. (In the last chapter we consider ordered mutlicurves with a fixed topological type, and more preciously, we fix an (artificially) ordered multicurve and study its mapping class group orbit.) This issue can be fixed by considering the evident order on $\mathbb{R}$: we do not know how to order the components in a multicurve but we do know how to order real numbers. More precisely, we consider the descending order statistics of the list of list of component lengths

$$\gamma = m_1\gamma_1 + \cdots + m_k\gamma_k \longmapsto \hat{\ell}_X^{\downarrow}(\gamma) := \frac{1}{\ell_X(\gamma)}\left(m_1\ell_X(\gamma_1), \ldots, m_k\ell_X(\gamma_k)\right)^{\downarrow} \in \Delta_1^{k-1}$$

where $v^{\downarrow} = (v_1^{\downarrow}, v_2^{\downarrow}, \ldots)$ denotes the sorted vector of $v = (v_1, v_2, \ldots)$ in descending order; for example, $(1, 9, 6, 8, 8, 4)^{\downarrow} = (9, 8, 8, 6, 4, 1)$.

Let $R \in \mathbb{R}_{>0}$, $m \in \mathbb{Z}_{\geq 1} \cup \{+\infty\}$. Let us denote by $s_X(m, R)$ the set of multi-geodesics $\gamma = m_1\gamma_1 + \cdots + m_k\gamma_k$ on $X \in \mathcal{M}_g$ (here k varies from 1 to $3g-3$) such that $\ell_X(\gamma) = m_1\ell_X(\gamma_1) + \cdots + m_k\ell_X(\gamma_k) \leq R$ and $m_i \leq m$ for all $1 \leq i \leq k$. Equip $s_X(m, R)$ with the uniform probability measure to make it a probability space, and write $\ell_{X,m,R}^{\downarrow}$ for the random variable on $s_X(m, R)$ whose underlying map is $\ell_X^{\downarrow}$.

The following result is a corollary of Theorem 5.0.2.

Theorem 6.1.1. *The random vector $\hat{\ell}_{X,m,R}^{\downarrow}$ converges in distribution to a random vector $L_{g,m}^{\downarrow}$ which depends only on g and m. In other words, there exists a probability measure $\mu_{g,m}$ such*

that

$$\frac{1}{|s_X(m,R)|}\sum_{\gamma\in s_X(m,R)}\delta_{\hat{\ell}_X^{\downarrow}(\gamma)}\Rightarrow\mu_{g,m}$$

as $R\to\infty$, where δ_x stands for the Dirac mass at x.

In fact we shall prove a stronger version of result above (Theorem 6.3.3).

The main result of this chapter is the following.

Theorem 6.1.2. *For any $m\in\mathbb{Z}_{\geq 1}\cup\{+\infty\}$, the sequence $(L_{g,m}^{\downarrow})_{g\geq 2}$ converges in law to the Poisson–Dirichlet distribution of parameter $1/2$ as $g\to\infty$.*

The most interesting cases are when $m=1$ (primitive multicurves) and when $m=+\infty$ (all multicurves). Let us emphasize that $L_{g,1}^{\downarrow}$ and $L_{g,+\infty}^{\downarrow}$ converge to the same limit as $g\to\infty$.

All marginals of the Poisson–Dirichlet law can be computed, see for example [ABT03, Section 4.11]. In particular if $V=(V_1,V_2,\dots)\sim\mathrm{PD}(\theta)$, then

$$\mathbb{E}((V_j)^n)=\frac{\Gamma(\theta+1)}{\Gamma(\theta+n)}\int_0^\infty\frac{(\theta E_1(x))^{j-1}}{(j-1)!}\,x^{n-1}e^{-x-\theta E_1(x)}\,dx$$

where

$$E_1(x):=\int_x^\infty\frac{e^{-y}}{y}\,dy.$$

The formulas can be turned into a computer program and values were tabulated in [Gri79, Gri88]. For $\theta=1/2$, we have

$$\mathbb{E}(V_1)\approx 0.758,\quad \mathbb{E}(V_2)\approx 0.171,\quad \text{and}\quad \mathbb{E}(V_3)\approx 0.049.$$

6.1.2 Square-tiled surface reformulation

Thanks to the correspondence between random multicurves and random square-tiled surfaces developed in [AH20b] and [DGZZ21], Theorem 6.1.2 can be reformulated in terms of square-tiled surfaces.

Let $q\in\mathcal{Q}(1^{4g-4})$ be a square-tiled surface tiled by N squares. Denote by $\hat{a}^{\downarrow}(q)\in\Delta_1^\infty$ the vector whose i-th entry is the area of the i-th largest cylinder of q devised by the total area N of q. The following is a particular case of [DGZZ21, Theorem 1.29].

Theorem 6.1.3 ([DGZZ21])**.** *Let $g\geq 2$, $m\in\mathbb{Z}_{\geq 1}\cup\{+\infty\}$, and let $\mu_{g,m}$ be the same probability measure on Δ_1^∞ as in Theorem 6.1.1. Then as $N\to\infty$ we have the following convergence of measures*

$$\frac{1}{|\mathcal{ST}(m,N)|}\sum_{M\in\mathcal{ST}(m,N)}\delta_{\hat{a}^{\downarrow}(M)}\Rightarrow\mu_{g,m}$$

where $\mathcal{ST}(m,N)$ denotes the set of square-tiled surfaces of genus g, of height at most m, and tiled by at most N squares.

An important difference to notice between Theorem 6.1.1 and Theorem 6.1.3 is that in the former the hyperbolic metric X is fixed and we sum over multi-geodesics on X while in the latter we sum square-tiled surfaces (flat metrics).

It follows from Theorem 6.1.3 that Theorem 6.1.2 admits the following reformulation.

Corollary 6.1.4. *The normalized (horizontal) area vector of a random square-tiled surface of genus g converges in distribution to* $\mathrm{PD}(1/2)$ *as g tends to ∞.*

6.1.3 Organization of the chapter

The first step of the proof consists in writing an explicit expression for the random variable $L^{\downarrow}_{g,m}$ that appears in Theorem 6.1.1; see Theorem 6.3.3 in Section 6.3. The formula follows from the work of Mirzakhani on random pants decompositions [Mir16] and the result of Arana-Herrera [AH21b] and Liu [Liu19] on random multicurves of fixed topological type. The expression of $L^{\downarrow}_{g,m}$ can be seen as a refinement of the formula for the Masur–Veech volume of the moduli space of quadratic differentials from [DGZZ21].

The formula for $L^{\downarrow}_{g,m}$ involves a super-exponential number of terms in g (one term for each topological type of multicurve on a surface of genus g). However, in the large genus limit only $\mathrm{O}(\log(g))$ terms contribute. This allows us to consider a simpler random variable $\widetilde{L}^{\downarrow}_{g,m,\kappa}$ which, asymptotically, coincides with $L^{\downarrow}_{g,m}$; see Theorem 6.5.3 in Section 6.5. This reduction is very similar to the one used for the large genus asymptotics of Masur–Veech volumes in [Agg21] and [DGZZ20b].

The core of our proof consists in proving the convergence of moments of the simpler variable $\widetilde{L}^{\downarrow}_{g,m,\kappa}$. We do not work directly with $\widetilde{L}^{\downarrow}_{g,m,\kappa}$ but its size-biased version $\widetilde{L}^{*}_{g,m,\kappa}$. The definition of size bias and the link with the Poisson–Dirichlet distribution is explained in Section 6.2. In Section 6.6, we show that the moments $\widetilde{L}^{*}_{g,m,\kappa}$ converge to the moments of $\mathrm{GEM}(1/2)$ which is the size-biased version of the Poisson–Dirichlet process $\mathrm{PD}(1/2)$; see Theorem 6.6.1.

6.2 Size-biased sampling

In this section we introduce the size-biased sampling of a (random) partition.

Let us begin by introducing some notation. Let s be a positive real number. We write

$$\begin{aligned}
\Delta^{k-1}_{s} &\coloneqq \{(x_1,\dots,x_k)\in[0,s]^k : x_1+\cdots+x_k=s\},\\
\Delta^{\infty}_{s} &\coloneqq \{(x_1,x_2,\dots)\in[0,s]^{\mathbb{Z}_{\geq 1}} : x_1+x_2+\cdots=s\},\\
\Delta^{k}_{\leq s} &\coloneqq \{(x_1,\dots,x_k)\in[0,s]^k : x_1+\cdots+x_k\leq s\},\\
\Delta^{\infty}_{\leq s} &\coloneqq \{(x_1,x_2,\dots)\in[0,s]^{\mathbb{Z}_{\geq 1}} : x_1+x_2+\cdots\leq s\}.
\end{aligned}$$

Note that the canonical embeddings $\Delta^{i}_{\leq s}\hookrightarrow\Delta^{j}_{\leq s}$ given by $(x_1,\dots,x_i)\mapsto(x_1,\dots,x_i,0,\dots,0)$ for $i\leq j$ defines a direct system, and $\Delta^{\infty}_{\leq s}$ is its direct limit (the same for Δ^{∞}_{s}). We equip $\Delta^{\infty}_{\leq s}$ with the topology induced from the product topology $[0,s]^{\mathbb{Z}_{\geq 1}}$, and Δ^{∞}_{s} with the induced topology given by the canonical embedding $\Delta^{\infty}_{s}\hookrightarrow\Delta^{\infty}_{\leq s}$. Note that Δ^{∞}_{s} is dense in $\Delta^{\infty}_{\leq s}$.

A partition of a positive real number s is a countable multiset λ of positive real numbers such that $\sum_{x\in\lambda}x=s$. To study random partitions, it would be convenient if the moduli space of partitions has a nice coordinate system. Of course, a partition can be uniquely represented as a decreasing sequence $\lambda_1\geq\lambda_2\geq\cdots\geq 0$, i.e., a point in Δ^{∞}_{s}, and a random partition is a probability measure on Δ^{∞}_{s}. However, the following seemingly awkward idea works surprisingly well: rather than arranging the parts of a partition in descending order, we put them in a random order. In other words, given a partition, we associate it with a probability measure on Δ^{∞}_{s} (rather than a point in Δ^{∞}_{s}).

This trick, called *size-biased sampling*, works as follows. For example, for the partition $\{1,2,3\}$, consider the random ordering $V=(V_1,V_2,V_3)$ of $\{1,2,3\}$ such that

$$\mathbb{P}(V_1=1)=\frac{1}{1+2+3},\qquad \mathbb{P}(V_1=2)=\frac{2}{1+2+3},\qquad \mathbb{P}(V_1=3)=\frac{3}{1+2+3},$$

in other words, the probability that the first entry is x proportional to its size (which is x). Then for the second entry V_2, we have

$$\mathbb{P}(V_2=2\mid V_1=1)=\frac{2}{2+3},\quad \mathbb{P}(V_2=3\mid V_1=1)=\frac{3}{2+3},\quad \mathbb{P}(V_2=1\mid V_1=2)=\frac{1}{1+3},\dots$$

in other word, the second element in the random ordering is chosen among all except the first element and the probability of picking an element is assumed to be always proportional to its size.

The formal definition is the following. Given a partition of s represented as a point $x=(x_1,x_2,\dots)\in\Delta_s^\infty$ (not necessarily sorted), let S_∞ denote the set of bijections from $\mathbb{Z}_{\geq 1}$ to itself. By Kolmogorov extension theorem, x defines a probability measure $\mathbb{P}_x$ on S_∞ in the following way.

- If $s=0$, i.e. $x_i=0$ for all $i\geq 1$, then $\mathbb{P}_x(\sigma)=1$ if σ is the identity, and zero otherwise.
- If $s>0$, set $\mathbb{P}_x(\sigma(1)=i_1)=x_{i_1}/s$, and for any $n\geq 1$, set

  $$\mathbb{P}_x(\sigma(n+1)=i_{n+1}\mid \sigma(1)=i_1,\dots,\sigma(n)=i_n)$$

  to be

  - $x_{i_{n+1}}/(1-x_1-\cdots-x_{i_n})$ if $i_1,\dots,i_{n+1}$ are distinct and $s-x_{i_1}-\cdots-x_{i_n}>0$,
  - 1 if $s=x_{i_1}+\cdots+x_{i_n}$ and $i_{n+1}=n+1$.
  - 0 otherwise.

Note that if $i_1,\dots,i_n$ are distinct and $s-x_{i_1}-\cdots-x_{i_n}>0$, then

$$\mathbb{P}_x(\sigma(1)=i_1,\dots,\sigma(n)=i_n)=\frac{x_{i_1}\cdots x_{i_n}}{s(s-x_{i_1})\cdots(s-x_{i_1}-\cdots-x_{i_{n-1}})}.$$

The *size-biased permutation* of x is a random sequence $X=(X_1,X_2,\dots)$ whose distribution is given by the formula

$$\mathbb{P}(X\in B)\coloneqq\mathbb{P}_x\{\sigma\in S_\infty:(x_{\sigma(1)},x_{\sigma(2)},\dots)\in B\}$$

where B is any Borel set of Δ_s^∞. Note that the resulting random reordering X of x depends only on the underlying partition that x represents.

From now on we assume that $s=1$. As x varies, the family of probability measures $\mathbb{P}_x$ defines a Markov kernel $\mathrm{SBP}\colon\mathcal{B}(\Delta_{\leq 1}^\infty)\times\Delta_{\leq 1}^\infty\to[0,1]$, where $\mathcal{B}(\Delta_{\leq 1}^\infty)$ stands for the Borel σ-algebra of $\Delta_{\leq 1}^\infty$, defined by

$$\mathrm{SBP}(B,x)=\mathbb{P}_x\{\sigma\in S_\infty:(x_{\sigma(1)},x_{\sigma(2)},\dots)\in B\}$$

where B is any Borel set of $\Delta_{\leq 1}^\infty$ and $x\in\Delta_{\leq 1}^\infty$. We are therefore able to construct the size-biased

permutation $\mu\,\mathrm{SBP}$ (or μ^*) of any probability measure on μ on $\Delta^\infty_{\leq 1}$ by setting

$$(\mu\,\mathrm{SBP})(B) \coloneqq \int_{\Delta^\infty_{\leq 1}} \mu(dx)\,\mathrm{SBP}(B,x).$$

Let $X = (X_1, X_2, \dots) \in \Delta^\infty_1$ be a random sequence with distribution $\mathbb{P}_X$. We write $X^* = (X^*_1, X^*_2, \dots)$ for the random variable whose distribution is the size-biased permutation $\mathbb{P}_X\,\mathrm{SBP}$ of $\mathbb{P}_X$, and $X^\downarrow = (X^\downarrow_1, X^\downarrow_2, \dots)$ for the random variable whose distribution is the push-forward of $\mathbb{P}_X$ under the sort operator $\Delta^\infty_1 \to \Delta^\infty_1$ that maps a sequence to its descending statistics.

Remark 6.2.1. The size-biased permutation X^*, of a random variable $X\colon \Omega \to \Delta^\infty_1$ defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, is not naturally defined as a random variable $\Omega \to \Delta^\infty_1$. Indeed, although we often talk about X^*, but it is its distribution, i.e. the push-forward measure $\mathbb{P}_X$ on Δ^∞_1 of $\mathbb{P}$ via $X\colon \Omega \to \Delta^\infty_1$, that matters to us.

The two following elementary lemmas will be useful later.

Lemma 6.2.2. *Let $X = (X_1, \dots, X_k) \in \Delta^{k-1}_s$ be a random vector. If the distribution of X admits a density function $f_X(x_1, \dots, x_k)$ (with respect to the Lebesgue measure λ on Δ^{k-1}_1), then the size-biased permutation of the distribution of X has density*

$$\frac{x_1 x_2 \cdots x_{k-1}}{s(s-x_1)\cdots(s-x_1-\cdots-x_{k-2})} \sum_{\sigma \in S_k} f(x_{\sigma(1)}, \dots, x_{\sigma(k)}).$$

Proof. Let μ denote the distribution of X. By definition, $(\mu\,\mathrm{SBP})(B)$ is equal to

$$\begin{aligned}
&= \int_{\Delta^{k-1}_s} \mathrm{SBP}(B,x)\,\mu(dx) \\
&= \int_{\Delta^{k-1}_s} \mathbb{P}_x\{\sigma \in S_k : (x_{\sigma(1)}, \dots, x_{\sigma(k)}) \in B\}\, f_X(x)\,\lambda(dx) \\
&= \int_{\Delta^{k-1}_s} \sum_{\sigma \in S_k} \frac{x_{\sigma(1)} x_{\sigma(2)} \cdots x_{\sigma(k-1)}}{s(s-x_{\sigma(1)})\cdots s(s-x_{\sigma(1)}-\cdots-x_{\sigma(k-2)})}\, \mathbb{1}_B(x_{\sigma(1)}, \dots, x_{\sigma(k)})\, f_X(x)\,\lambda(dx) \\
&= \int_{\Delta^{k-1}_s} \sum_{\sigma \in S_k} \frac{x_1 x_2 \cdots x_{k-1}}{s(s-x_1)\cdots(s-x_1-\cdots-x_{k-2})}\, \mathbb{1}_B(x_1, \dots, x_k)\, f_X(x_{\sigma(1)}, \dots, x_{\sigma(k)})\,\lambda(dx).
\end{aligned}$$

This proves the result. □

The following lemma can be interpreted as: the marginal distribution of the first r samples in a size-biased sampling is none other than the distribution of the r size-biased sampling.

Lemma 6.2.3. *Let $X = (X_1, \dots, X_k) \in \Delta^{k-1}_s$ be a random vector whose distribution μ admits density function f_X. Then for $r < k$, the marginal distribution of the first r coordinates of the size-biased permutation of μ has density function*

$$\frac{1}{(k-r)!} \frac{x_1 x_2 \cdots x_r}{s(s-x_1)\cdots(s-x_1-\cdots-x_{r-1})} \int_{\Delta^{k-r-1}_{s-x_1-\cdots-x_r}} \lambda(d(x_{r+1}, \dots, x_k)) \sum_{\sigma \in S_k} f_X(x_{\sigma(1)}, \dots, x_{\sigma(k)})$$

where

$$\Delta^{k-r-1}_{s-x_1-\cdots-x_r} \coloneqq \{(x_{r+1}, \dots, x_k) \in \mathbb{R}^{k-r}_{\geq 0} : x_{r+1} + \cdots + x_k = 1 - x_1 - \cdots - x_r\}$$

and λ is the Lebesgue measure on $\Delta^{k-r-1}_{s-x_1-\cdots-x_r}$.

Proof. The proof is by induction. The case $r = k-1$ is Lemma 6.2.2. We shall show that if the statement holds for $2 \le r \le k-1$, it holds for $r-1$. The marginal density function $f_{X^*}(x_1, \dots, x_{r-1})$ can be calculated by

$$
\begin{aligned}
&= \int_0^{s-x_1-\cdots-x_{r-1}} dx_r f_{X^*}(x_1, \dots, x_r) \\
&= \frac{1}{(k-r)!} \frac{x_1 \cdots x_{r-1}}{s(s-x_1)\cdots(s-x_1-\cdots-x_{r-1})} \\
&\quad \cdot \int_0^{s-x_1-\cdots-x_{r-1}} dx_r \frac{x_r}{s-x_1-\cdots-x_{r-1}} \int_{\Delta^{k-r-1}_{s-x_1-\cdots-x_r}} \lambda(d(x_{r+1}, \dots, x_k)) \sum_{\sigma \in S_k} f(x_{\sigma(1)}, \dots, x_{\sigma(k)})
\end{aligned}
$$

The integral after $x_1 \cdots x_{r-1}/((1-x_1)\cdots(1-x_1-\cdots-x_{r-1})$ is equal to

$$
\int_0^{1-x_1-\cdots-x_{r-1}} \frac{dx_r}{1-x_1-\cdots-x_{r-1}} \int_{\Delta^{k-r-1}_{s-x_1-\cdots-x_r}} \lambda(d(x_{r+1}, \dots, x_k))\, x_r \sum_{\sigma \in S_k} f(x_{\sigma(1)}, \dots, x_{\sigma(k)})
$$

which is, by the symmetry between $x_1, \dots, x_r$ in the last integral,

$$
\begin{aligned}
&\frac{1}{k-r+1} \int_0^{1-x_1-\cdots-x_{r-1}} dx_r \int_{\Delta^{k-r-1}_{s-x_1-\cdots-x_r}} \lambda(d(x_{r+1}, \dots, x_k)) \sum_{\sigma \in S_k} f(x_{\sigma(1)}, \dots, x_{\sigma(k)}) \\
&\qquad = \frac{1-x_1-\cdots-x_{r-1}}{k-r+1} \int_{\Delta^{k-r-2}_{1-x_1-\cdots-x_{r-1}}} \lambda(d(x_r, \dots, x_k)) \sum_{\sigma \in S_k} f(x_{\sigma(1)}, \dots, x_{\sigma(k)}),
\end{aligned}
$$

so the statement for $r-1$ follows. □

The following result shows "$*$" and "$\downarrow$" commute with limits.

Theorem 6.2.4 ([DJ89, Theorem 3])**.** *Let $(\mu_n)_{n \ge 1}$ be a sequence of probability measures such that as $n \to \infty$, $(\mu_n)_{n \ge 1}$ converges weakly to a probability measure μ on Δ_1^∞. Then as $n \to \infty$, μ_n^* converges weakly to μ^* and $\mu_n^\downarrow$ converges weakly to $\mu^\downarrow$.*

We will also use the above result in the following form.

Corollary 6.2.5. *Let $X^{(n)} = (X_1^{(n)}, X_2^{(n)}, \dots) \in \Delta_1^\infty$ be a sequence of random sequences and $\theta \in \mathbb{R}_{>0}$. The sequence of sorted sequences $X^{(n)\downarrow}$ of of $X^{(n)}$ converges in distribution to* $\mathrm{PD}(\theta)$ *if and only if the size-biased permutation $X^{(n)*}$ of $X^{(n)}$ converges in distribution to* $\mathrm{GEM}(\theta)$.

The size-biased permutation and sorting commute also with conditioning.

Lemma 6.2.6. *Let $(\mu_n)_{n \ge 1}$ be a sequence of probability measures, and $(p_n)_{n \ge 1}$ be a sequence of non-negative real numbers such that $p_1 + p_2 + \cdots = 1$. Then*

$$
\sum_{n=1}^\infty p_n \mu_n^* = \left(\sum_{n=1}^\infty p_n \mu_n \right)^*, \qquad \sum_{n=1}^\infty p_n \mu_n^\downarrow = \left(\sum_{n=1}^\infty p_n \mu_n \right)^\downarrow.
$$

6.3 Normalized length spectrum of a random multi-geodesic

The aim of this section is to state and prove a refinement of Theorem 6.1.1 that provides an explicit description of the random variable $L^{\downarrow}_{g,m}$.

Remark 6.3.1. The sample space of $\hat{\ell}_{X,\gamma,R}$ can be made to be s_X (the space of all multi-geodesics on X), nevertheless, although $\hat{\ell}_{X,\gamma,R}$ converges in distribution, the family of probability measures indexed by R supported on $s_{X,R}$ does not weakly converge to a σ-additive measure on s_X, and hence the sample space (domain of definition) of $L_{g,\gamma}$ is not s_X. However, since what we really care is the distributions of $\hat{\ell}_{X,\gamma,R}$ and $L_{g,\gamma}$ (which is a measure on Δ_1^{k-1}), the domain of definition of $L_{g,\gamma}$ has no importance.

Let γ be a multicurve. Let E the component set of γ and let $\lambda\colon E \to \{1,\dots,|E|\}$ be a bijection (a labeling). (Now we can write $\gamma = m_1\gamma_1 + \cdots + m_k\gamma_k$, where $k := |E|$.) Let us (temporarily) denote the labeled (or ordered) multicurve by the pair (γ,λ). Recall that Theorem 5.0.2 implies that a random variable $\hat{\ell}_{X,(\gamma,\lambda),R}$ converges to $L_{g,(\gamma,\lambda)}$ as $R \to \infty$. Note that the distribution of the descending statistics and size-biased permutation of $\ell_{X,(\gamma,\lambda),R}$ and of $L_{g,(\gamma,\lambda)}$ does not depend on the choice of the labelling λ, so from now on we shall write simply $\ell^{\downarrow}_{X,\gamma,R}$, $\ell^{*}_{X,\gamma,R}$, $L^{\downarrow}_{g,\gamma}$, $L^{*}_{g,\gamma}$.

Remark 6.3.2. The random vector $\hat{\ell}^{\downarrow}_{X,\gamma,R}$ can also be defined directly by

$$\hat{\ell}^{\downarrow}_{X,\gamma,R} : s_X(\gamma,R) \to \Delta_1^{\infty}, \qquad \alpha \mapsto \hat{\ell}^{\downarrow}_{X}(\alpha)$$

without choosing any labeling. Indeed, the two definitions are equivalent since the natural projection $s_X((\gamma,\lambda),R) \to s_X(\gamma,R)$ is a $[\mathrm{Stab}(m_1\gamma_1 + \cdots + m_k\gamma_k) : \mathrm{Stab}((m_1\gamma_1,\dots,m_k\gamma_k))]$-to-one correspondence and the function $\ell^{\downarrow}_{X,\gamma,A}$ is constant in each fiber. Similarly, $\hat{\ell}^{*}_{X,\gamma,R}$ can be defined as a random variable with distribution given by

$$\frac{1}{|s_X(\gamma,R)|}\sum_{\alpha\in s_X(\gamma,R)} \delta^{*}_{\hat{\ell}_X(\alpha)}.$$

Note that $\delta_{\hat{\ell}_X(\alpha)}$ is not well-defined since α is only a unordered multicurve, but its size-biased permutation $\delta^{*}_{\hat{\ell}_X(\alpha)}$ is.

Write $[\alpha]$ for the topological type of a multicurve α. Let A be a set of topological types of multicurves on Σ_g. Consider the set of multicurves

$$s_X(R,A) := \{\alpha : [\alpha] \in A,\ \ell_X(\alpha) \le R\}$$

and the random vectors

$$\hat{\ell}^{\downarrow}_{X,A,R} : s_X(R,A) \to \Delta_1^{\infty}, \qquad \alpha \to \hat{\ell}^{\downarrow}_{X}(\alpha).$$

Similarly, we define $\hat{\ell}^{*}_{X,A,R}$ to be a random variable with distribution

$$\frac{1}{|s_X(A,R)|}\sum_{\alpha\in s_X(A,R)} \delta^{*}_{\hat{\ell}_X(\alpha)}.$$

The following is a direct generalization of Theorem 5.0.2 concerning random multicurves of

a fixed topological type to random multicurves with topological types in any given set.

Theorem 6.3.3. *Given a set A of topological types of multicurves on Σ_g. Then as $R \to \infty$, $\hat{\ell}^{\downarrow}_{X,A,R}$ converges in distribution to a random variable $L^{\downarrow}_{g,A}$, $\hat{\ell}^{*}_{X,A,R}$ converges in distribution to a random variable $L^{*}_{g,A}$, and for any bounded continuous function $h\colon \Delta_1^\infty \to \mathbb{R}$, we have*

$$\mathbb{E}(L^{\downarrow}_{g,A}) = \sum_{[\alpha]\in A} \frac{c(\alpha)}{b_A} \cdot \mathbb{E}(L^{\downarrow}_{g,\alpha}(h)), \qquad \mathbb{E}(L^{*}_{g,A}) = \sum_{[\alpha]\in A} \frac{c(\alpha)}{b_A} \cdot \mathbb{E}(L^{*}_{g,\alpha}(h))$$

Proof. We prove the "$\downarrow$" part, the "$*$" part is completely analogous.

Let $h\colon \Delta_1^\infty \to \mathbb{R}$ be a bounded continuous function. Then

$$\mathbb{E}(\hat{\ell}^{\downarrow}_{X,A,R}(h)) = \sum_{[\alpha]\in A} \frac{|s_X(\alpha,R)|}{|s_X(A,R)|} \cdot \mathbb{E}(\hat{\ell}^{\downarrow}_{X,\alpha,R}(h)).$$

It follows from Theorem 4.1.1 that there exists $c(\gamma) \in \mathbb{Q}_{>0}$ such that

$$\lim_{R\to\infty} \frac{|s_X(\alpha,R)|}{|s_X(A,R)|} = \lim_{R\to\infty} \frac{|s_X(\alpha,R)|}{|s_X(R)|} \frac{|s_X(R)|}{|s_X(R,A)|} = \frac{c(\alpha)}{b_{g,A}}.$$

By Theorem 5.0.2 and Theorem 6.2.4,

$$\lim_{R\to\infty} \mathbb{E}(\hat{\ell}^{\downarrow}_{X,\alpha,R}(h)) = \mathbb{E}(L^{\downarrow}_{g,\alpha}(h)).$$

Now the result follows from Lemma 6.2.6 and the following elementary lemma. □

Lemma 6.3.4. *Let $(f_n)_{n\in\mathbb{Z}_{\geq 1}}$ and $(g_n)_{n\in\mathbb{Z}_{\geq 1}}$ be sequences of real-valued measurable functions on a measure space (X,Σ,μ) with point-wise limit f and g respectively, such that $|f_n(x)| \leq g_n(x)$ and g is integrable. Then f is integrable and*

$$\lim_{n\to\infty} \int_X f_n \, d\mu = \int_X f \, d\mu.$$

Proof. Since f_n converges point-wisely to f and f_n is measurable, f is measurable; f is integrable as it is bounded by a integrable function. Fatou's lemma implies

$$\int_X f \, d\mu \leq \liminf_{n\to\infty} \int_X f_n \, d\mu.$$

Applying again Fatou's lemma to $g - f$, we have

$$\int_X g \, d\mu - \int_X f \, d\mu = \int_X (g-f) \, d\mu \leq \int_X g \, d\mu - \limsup_{n\to\infty} \int_X f_n \, d\mu,$$

and therefore,

$$\limsup_{n\to\infty} \int_X f_n \, d\mu \leq \int_X f \, d\mu,$$

which completes the proof. □

Of course, the most interesting cases are when A consists of topological types of all multicurves or all primitive multicurves.

6.4 Density

Before we tackle the large genus asymptotics, it will be help to drive a more explicit expression for the density function of the limiting distribution distribution $L_{g,\gamma}$. The following elementary lemma is particularly useful in our calculations.

Lemma 6.4.1. *Let $d_1, \dots, d_k \in \mathbb{R}_{\geq 0}$. We have*

$$\int_{\Delta^k_{\leq r}} x_1^{d_1} \cdots x_k^{d_k}\, dx = \frac{d_1! \cdots d_k!}{(d_1 + \cdots d_k + k)!} \cdot r^{d_1 + \cdots + d_k + k}$$

and

$$\int_{\Delta^{k-1}_r} x_1^{d_1} \cdots x_k^{d_k}\, \lambda(dx) = \frac{d_1! \cdots d_k!}{(d_1 + \cdots + d_k + k - 1)!} \cdot r^{d_1 + \cdots + d_k + k - 1}$$

where λ is the Lebesgue measure on the simplex Δ^{k-1}_r, and by $a!$ we mean $\Gamma(a+1)$.

Proof. These identities can be deduced by induction using the identity

$$\int_0^x t^{a-1}(x-t)^{b-1}\, dt = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}\, x^{a+b-1}.$$

Here we give another proof of the second identity using Laplace transform (the first follows from the second). Let us define

$$I : \mathbb{R}_{>0} \to \mathbb{R}, \qquad r \mapsto \int_{\Delta^{k-1}_r} x_1^{d_1} \cdots x_k^{d_k}.$$

Note that $I(r)$ can be written as

$$I(r) = \int_{\mathbb{R}^k_{>0}} \delta(r - x_1 - \cdots - x_k) \cdot x_1^{d_1} \cdots x_k^{d_k}$$

where δ is the Dirac delta function. The Laplace transform of $I(r)$ is

$$\mathcal{L}\{I\}(z) = \int_0^\infty e^{-zr} I(r)\, dr = \int_{\mathbb{R}^k_{>0}} e^{-z(x_1 + \cdots + x_k)} \cdot x_1^{d_1} \cdots x_k^{d_k}\, dx_1 \cdots dx_k = \prod_{i=1}^k \int_0^\infty e^{-zx_i} x_i^{d_i}\, dx_i$$

which is the product of the Laplace transform of $x_i^{d_i}$, $1 \leq i \leq k$, and is equal to

$$\prod_{i=1}^k \mathcal{L}\{x_i^{d_i}\} = d_1! \cdots d_k! \cdot z^{d_1 + \cdots + d_k + k}.$$

Therefore,

$$I(r) = \mathcal{L}^{-1}(\mathcal{L}\{I\}(z))(r) = \mathcal{L}^{-1}\left(d_1! \cdots d_k! z^{d_1 + \cdots + d_k + k}\right) = \frac{d_1! \cdots d_k!}{(d_1 + \cdots + d_k - 1)!} \cdot r^{d_1 + \cdots + d_k + k - 1},$$

the result desired. □

The following lemma gives an explicit expression for the normalizing constant that appears in Theorem 5.0.2.

Lemma 6.4.2. *Let $\gamma = (m_1\gamma_1, \dots, m_k\gamma_k)$ be an ordered multicurve on Σ_g. The limiting distribution*

$$\lim_{R\to\infty} \frac{1}{|s_X(R,\gamma)|} \sum_{\alpha \in s_X(R,\gamma)} \delta_{\hat{\ell}_X(\alpha)}$$

exists and has density function

$$\frac{(6g-7)!}{\mathcal{Y}_{\underline{m}}(\bar{F}_\gamma)} \frac{\bar{F}_\gamma(x_1/m_1, \dots, x_k/m_k)}{m_1 \cdots m_k}$$

where $\underline{m} = (m_1, \dots, m_k)$. In other words, the normalizing constant in Theorem 5.0.2 is given by

$$\frac{(6g-7)!}{\mathcal{Y}_{\underline{m}}(F_\gamma)} \frac{1}{m_1 \cdots m_k} = \frac{2^{3g-3}(6g-6)\, k! \cdot c(\gamma)}{[\mathrm{Stab}(\gamma_1 + \cdots + \gamma_k) : \mathrm{Stab}(m_1\gamma_1 + \cdots + m_k\gamma_k)]} = \frac{k!\, \mathrm{vol}_{\mathrm{MV}}(\Gamma(\gamma), \underline{m})}{2^{g+1}(4g-4)!}.$$

Proof. By a straight-forward computation using Lemma 6.4.1 and the definition of $\mathcal{Y}_{\underline{m}}$,

$$\int_{\Delta_1^{k-1}} \bar{F}_\gamma(x_1/m_1, \dots, x_k/m_k)\, \lambda(dx) = \frac{m_1 \cdots m_k}{(6g-7)!} \cdot \mathcal{Y}_{\underline{m}}(\bar{F}_\gamma).$$

The rest of the assertion follows from Theorem 4.2.6. □

As we shall see in the next section, the following example turns out to be the most important case for our purposes.

Example 6.4.3. Let $\gamma = (m_1\gamma_1, \dots, m_k\gamma_k)$ be a non-separating ordered multicurve on Σ_g. All such multicurves have the same associated stable graph (which does not depend on the weights $m_1, \dots, m_k$), and the corresponding graph polynomial, denoted by $F_{g,k}(x_1, \dots, x_k)$, is equal to

$$\frac{1}{2^{3g-3-k}} \sum_{\substack{(d_{1-}, d_{1+}, \dots, d_{k-}, d_{k+}) \in \mathbb{Z}_{\geq 0}^{2k} \\ d_{i-} + d_{i+} + \cdots + d_{k-} + d_{k+} = 3g-3}} \frac{x_1^{2d_1+1} \cdots x_k^{2d_i+1}}{d_{1-}!\, d_{1+}! \cdots d_{k-}!\, d_{k+}!} \int_{\overline{\mathcal{M}}_{g,2k}} \psi_{1-}^{d_{1-}} \psi_{1+}^{d_{1+}} \cdots \psi_{k-}^{d_{k-}} \psi_{k+}^{d_{k+}} \tag{6.1}$$

where $d_i := d_{i-} + d_{i+}$ for all $1 \leq i \leq k$. Then the density function of the limiting distribution of $L_{g,\gamma}$ is

$$\frac{1}{2^{3g-3}(6g-6)k!\, c(\gamma)} \cdot \sum_{\substack{(d_{1-}, d_{1+}, \dots, d_{k-}, d_{k+}) \in \mathbb{Z}_{\geq 0}^{2k} \\ d_{i-} + d_{i+} + \cdots + d_{k-} + d_{k+} = 3g-3}} \frac{x_1^{2d_1+1} \cdots x_k^{2d_i+1}}{m_1^{2d_1+2} \cdots m_k^{2d_k+2}} \frac{1}{d_-!\, d_{1+}! \cdots d_{k-}!\, d_{k+}!} \int_{\overline{\mathcal{M}}_{g,2k}} \psi_{1-}^{d_{1-}} \psi_{1+}^{d_{1+}} \cdots \psi_{k-}^{d_{k-}} \psi_{k+}^{d_{k+}}$$

where $d_i := d_{i-} + d_{i+}$ for all $1 \leq i \leq k$.

Corollary 6.4.4. *Let $\gamma = (\gamma_1, \dots, \gamma_k)$ be an ordered primitive multicurve on Σ_g. Take $A_{\gamma,m}$ to be the collection of topological types of multicurves of the form $m_1\gamma_1 + \cdots + \gamma_k$ with $m_1, \dots, m_k \leq m$. Then the density function of size-biased permutation of $L_{g,A_{\gamma,m}}$ is the size-biased permutation of*

$$\frac{(6g-7)!}{\mathcal{Z}_m(\bar{F}_{(\gamma_1,\dots,\gamma_k)})} \sum_{\substack{(m_1,\dots,m_k) \in \mathbb{Z}_{\geq 1}^k \\ m_1,\dots,m_k \leq m}} \frac{\bar{F}_{(m_1\gamma_1,\dots,m_k\gamma_k)}(x_1/m_1, \dots, x_k/m_k)}{m_1 \cdots m_k}$$

Proof. This follows from Lemma 6.2.6, Lemma 6.4.2, and Theorem 4.2.6. □

6.5 Reduction in the asymptotic regime

The random variable $L^{\downarrow}_{g,m}$ appearing in Theorem 6.1.1 is delicate to study because it involves a huge number of terms. Using Theorem 4.3.1 from [Agg21] and [DGZZ20b] we show that we can restrict to a sum involving only $\mathrm{O}(\log(g))$ terms associated to non-separating multicurves.

We denote by $\Gamma_{g,k}$ the stable graph of genus g with a vertex of genus $g-k$ and k loops (which corresponds to non-separating multicurves on Σ_g with k components). To simplify the notation we fix a bijection between the edges of $\Gamma_{g,k}$ and $\{1,2,\dots,k\}$ so that its graph polynomial $F_{g,k} := F_{\Gamma_{g,k}}$ is a polynomial in $\mathbb{Q}[x_1,\dots,x_k]$. Note that because the edges in $\Gamma_{g,k}$ are not distinguishable, the polynomial $F_{g,k}$ is symmetric.

Let $m \in \mathbb{Z}_{\geq 1} \cup \{+\infty\}$. Let $T_{g,m}$ denote the set of topological types of multicurves on Σ_g of the form $m_1\gamma_1 + \cdots + m_k\gamma_k$ where $m_1,\dots,m_k \leq m$, and k varies from 1 to $3g-3$ (running through all possible values). Let us write $L^*_{g,m}$ for $L^*_{g,T_{g,m}}$ that we defined in Section 6.3, and define

$$b_{g,m} := \sum_{[\gamma]\in T_{g,m}} c(\gamma) \tag{6.2}$$

Note that $b_g = b_{g,+\infty}$. Now for given $\kappa \in \mathbb{R}_{>1}$, consider the set $\widetilde{T}_{g,m,\kappa}$ of topological types of non-separating multicurves on Σ_g of the form $m_1\gamma_1 + \cdots + m_k\gamma_k$ where $m_1,\dots,m_k \leq m$, and $k \leq \kappa\log(6g-6)/2$. Again, write $\widetilde{L}^*_{g,m,\kappa}$ for $L^*_{g,\widetilde{T}_{g,m,\kappa}}$, and define

$$\tilde{b}_{g,m,\kappa} := \sum_{[\gamma]\in \widetilde{T}_{g,m,\kappa}} c(\gamma) \tag{6.3}$$

Remark 6.5.1. We warn the reader that the constant $b_{g,m}$ in this article has nothing to do with the analogue of b_g in the context of surfaces of genus g with n boundaries which is denoted by $b_{g,n}$ in [Mir16] and [DGZZ21].

We will use the asymptotic results of [Agg21] and [DGZZ20b] in the following form (which is a corollary of Theorem 4.4.4).

Theorem 6.5.2. *For any $m \in \mathbb{Z}_{\geq 1} \cup \{+\infty\}$ and $\kappa \in \mathbb{R}_{>1}$, we have*

$$b_{g,m} \sim \widetilde{b}_{g,m,\kappa} \sim \frac{1}{\pi}\frac{1}{(6g-6)(4g-4)!}\sqrt{\frac{m}{m+1}}\left(\frac{4}{3}\right)^{4g-4}$$

as $g \to \infty$.

The following theorem allows to focus on a much smaller set of topological types.

Theorem 6.5.3. *Given $\kappa \in \mathbb{R}_{>1}$. For any bounded and continuous function $h\colon \Delta_1^\infty \to \mathbb{R}$, we have*

$$\mathbb{E}(L^*_{g,m}(h)) \sim \mathbb{E}(\widetilde{L}^*_{g,m,\kappa}(h))$$

as $g \to \infty$.

Proof. The expectation $\mathbb{E}(L^*_{g,m}(h))$ can be written as (via "the law of total expectation") a sum of expectations weighted by $c(\gamma)$ over all topological types $[\gamma]$. Now the result follows directly from Theorem 6.5.2 and the boundedness of h. □

6.6 Proof of the main theorem

The aim of this section is to prove the following result.

Theorem 6.6.1. *For $g \geq 2$ integral, $m \in \mathbb{Z}_{\geq 1} \cup \{+\infty\}$ and $\kappa \in \mathbb{R}_{>1}$. Then as g tends to infinity, the random variable $\widetilde{L}^*_{g,m,\kappa}$ defined in the last section converges in distribution to* GEM(1/2).

Let us first show how to derive our main result Theorem 6.1.2 from Theorem 6.6.1.

Proof of Theorem 6.1.2. By Theorem 6.5.3, the random variables $L^*_{g,m}$ and $\widetilde{L}^*_{g,m,\kappa}$ have the same limit distribution as $g \to \infty$. Hence by Theorem 6.6.1, $L^*_{g,m}$ converges in distribution towards GEM(1/2) as $g \to \infty$. Finally Corollary 6.2.5 shows that the convergence in distribution of $L^*_{g,m}$ to GEM(1/2) is equivalent to the convergence towards PD(1/2). This concludes the proof. □

6.6.1 Method of moments

In this section, we show that the convergence of a sequence of random variables $X^{(n)}$ is equivalent to the convergence of some special moments. This strategy called the *method of moments* is a standard tool in probability; see for example [Bil95, Section 30] for the case of real variables.

Let $X = (X_1, X_2, \dots) \in \Delta^\infty_{\leq 1}$ be a random sequence. Given $r \in \mathbb{Z}_{\geq 1}$ $p = (p_1, \dots, p_r) \in \mathbb{Z}^r_{\geq 0}$ an r-tuple, we define

$$M_p(X) := \mathbb{E}\left((1 - X_1) \cdots (1 - X_1 - \cdots - X_{r-1}) \cdot X_1^{p_1} \cdots X_r^{p_r}\right)$$

if $r \geq 2$, and $M_p(X) := \mathbb{E}(X_1^{p_1})$ if $r = 1$.

The M_p-moments of a GEM distribution are particularly simple to compute.

Lemma 6.6.2. *Let $X = (X_1, X_2, \dots) \in \Delta^\infty_{\leq 1}$ be a random series and $p = (p_1, \dots, p_r) \in \mathbb{Z}^r_{\geq 0}$ be an r-tuple. If $X \sim$ GEM(θ), then*

$$M_p(X) = \frac{\theta^r \cdot \Gamma(\theta) \cdot p_1! \cdots p_r!}{\Gamma(p_1 + \cdots + p_r + \theta + r + 1)}.$$

Proof. By Proposition 3.5.1,

$$M_p(X) = \int_{\Delta^r_{\leq 1}} \theta^r x_1^{p_1} \cdots x_r^{p_r} (1 - x_1 - \cdots - x_r)^{\theta - 1}\, dx = \theta^r \int_{\Delta^r_1} x_1^{p_1} \cdots x_r^{p_r} x_{r+1}^{\theta-1}\, \lambda(dx)$$

Now the result follows from Lemma 6.4.1. □

To prove that a sequence of random sequences converges in distribution to a GEM process, it is sufficient to check its M_p-moments.

Lemma 6.6.3. *A sequence of random sequences $X^{(n)} = (X_1^{(n)}, X_2^{(n)}, \dots) \in \Delta^\infty_1$ converges in distribution to a random sequence $X^{(\infty)}$ if and only if for any $r \geq 1$ and any $p = (p_1, \dots, p_r) \in \mathbb{Z}^r_{\geq 0}$, $\lim_{n\to\infty} M_p(X^{(n)}) = M_p(X^{(\infty)})$.*

Proof. The infinite-dimensional cube $[0,1]^{\mathbb{Z}_{\geq 1}}$ is compact with respect to the product topology by Tychonoff's theorem. The subset $\Delta^\infty_{\leq 1}$ of $[0,1]^\infty$ is closed, and therefore compact. Let $C(\Delta^\infty_{\leq 1}, \mathbb{R})$

denote the algebra of real-valued continuous functions on $\Delta^\infty_{\leq 1}$, and let S be the set of functions of the form $x_1^{p_1}$ or

$$(1-x_1)(1-x_1-x_2)\cdots(1-x_1-\cdots-x_{r-1})\cdot x_1^{p_1}\cdots x_r^{p_r}$$

with $r \geq 2$ and $p_1,\ldots,p_r \geq 0$. The set S is a separating subset of $C(\Delta^\infty_{\leq 1},\mathbb{R})$. Now it follows from Stone–Weierstrass theorem that the subalgebra generated by S is dense in $C(\Delta^\infty_{\leq 1},\mathbb{R})$ with respect to the uniform convergence topology. To complete the proof, note that S is closed under multiplication. □

We use the following asymptotic simplification of the moments.

Theorem 6.6.4. *Let $g \in \mathbb{Z}_{\geq 2}$, $m \in \mathbb{Z}_{\geq 1} \cup \{\infty\}$, $\kappa \in \mathbb{R}_{>1}$, $r \in \mathbb{Z}_{\geq 1}$, and $p = (p_1,\ldots,p_r) \in \mathbb{Z}^r_{\geq 0}$. The moment $M_p(\widetilde{L}^*_{g,m,\kappa})$ is asymptotically equivalent to*

$$\frac{\sqrt{\pi(m+1)/m}}{2(6g-6)^{p_1+\cdots+p_r+r-1/2}} \sum_{k=r}^{\kappa\log(6g-6)/2} \frac{1}{(k-r)!} \sum_{\substack{(j_1,\ldots,j_k)\in\mathbb{Z}^k_{\geq 1}\\ j_1+\cdots+j_k=3g-3}} \prod_{i=1}^k \frac{\zeta_m(2j_i)}{2j_i} \prod_{i=1}^r \frac{(2j_i+p_i)!}{(2j_i-1)!}$$

where

$$\zeta_m(s) := \sum_{n=1}^m \frac{1}{n^s}$$

is the partial Riemann zeta function.

Following [DGZZ20b, Equation (14)], we define

$$c_{g,k}(d_1,\ldots,d_k) := \frac{g!\,(3g-3+2k)!}{(6g-5+4k)!}\frac{3^g}{2^{3g-6+5k}}(2d_1+2)!\cdots(2d_k+2)!$$
$$\cdot \sum_{\substack{(d_{1-},d_{1+},\ldots,d_{k-},d_{k+})\in\mathbb{Z}^{2k}_{\geq 0}\\ d_{i-}+d_{i+}+\cdots+d_{k-}+d_{k+}=3g-3}} \frac{1}{d_-!\,d_{1+}!\cdots d_{k-}!\,d_{k+}!}\int_{\overline{\mathcal{M}}_{g,2k}} \psi_{1-}^{d_{1-}}\psi_{1+}^{d_{1+}}\cdots\psi_{k-}^{d_{k-}}\psi_{k+}^{d_{k+}}$$

and $\tilde{c}_{g,k}(j_1,\ldots,j_k) = c_{g-k,k}(j_1-1,\ldots,j_k-1)$. The above coefficients were introduced in [DGZZ20b, Lemma 3.5], and we have

$$\lim_{g\to\infty} \sup_{\substack{k\in\mathbb{Z}_{\geq 1}\\ k^2\leq g/800}} |c_{g,k}(d_1,\ldots,d_k)-1| = 0$$

This asymptotic result is a direct consequence of [Agg21, Theorem 9] that we stated in the introduction.

Given $m,k \in \mathbb{Z}_{\geq 1}$, let $\widetilde{T}_{g,m,(k)}$ denote the set of topological types of non-separating multicurves on Σ_g of the form $m_1\gamma_1+\cdots+m_k\gamma_k$ with $m_1,\ldots,m_k \leq m$. Let us write $\widetilde{L}^*_{g,m,(k)}$ for $L^*_{g,\widetilde{T}_{g,m,(k)}}$ defined in Section 6.3.

Lemma 6.6.5. *For any* $p = (p_1, \dots, p_r) \in \mathbb{Z}^r_{\geq 0}$,

$$M_p\big(\widetilde{L}^*_{g,m,(k)}\big) = \frac{w_{g,k} \cdot k!}{\mathcal{Z}_m(\bar{F}_{g,k})\,(k-r)!\,(6g-7+p_1+\cdots+p_r+r)!} \cdot \sum_{\substack{(j_1,\dots,j_k)\in\mathbb{Z}^k_{\geq 1}\\ j_1+\cdots+j_k=3g-3}} \tilde{c}_{g,k}(j_1,\dots,j_k) \prod_{i=1}^{k} \frac{\zeta_m(2j_i)}{2j_i} \prod_{i=1}^{r} \frac{(2j_i+p_i)!}{(2j_i-1)!}$$

where

$$w_{g,k} := \frac{(6g-5-2k)!\,(6g-7)!}{(g-k)!\,(3g-3-k)!}\,\frac{2^{3k-3}}{3^{g-k}}.$$

Proof. By Corollary 6.4.4, the density function of $\widetilde{L}^*_{g,m,(k)}$ is the size-biased permutation of the following density function

$$\frac{(6g-7)!}{\mathcal{Z}_m(\bar{F}_{g,k})} \sum_{\substack{(m_1,\dots,m_k)\in\mathbb{Z}^k_{\geq 1}\\ m_1,\dots,m_k\leq m}} \frac{\bar{F}_{g,k}(x_1/m_1,\dots,x_k/m_k)}{m_1\cdots m_k}. \tag{6.4}$$

The graph polynomial $\bar{F}_{g,k}$ (see (6.1)) is

$$\bar{F}_{g,k}(x_1,\dots,x_k) = \frac{w_{g,k}}{(6g-7)!} \sum_{\substack{(j_1,\dots,j_k)\in\mathbb{Z}^k_{\geq 1}\\ j_1+\cdots+j_k=3g-3}} \tilde{c}_{g,k}(j_1,\dots,j_k) \prod_{i=1}^{k} \frac{x_i^{2j_i-1}}{(2j_i)!}.$$

Hence, the density function (6.4) can be rewritten as

$$\frac{w_{g,k}}{\mathcal{Z}_m(\bar{F}_{g,k})} \sum_{\substack{(j_1,\dots,j_k)\in\mathbb{Z}^k_{\geq 1}\\ j_1+\cdots+j_k=3g-3}} \tilde{c}_{g,k}(j_1,\dots,j_k) \prod_{i=1}^{k} \zeta_m(2j_i)\,\frac{x_i^{2j_i-1}}{(2j_i)!}.$$

Therefore, it follows from Lemma 6.2.2, the density function of $\widetilde{L}^*_{g,m,(k)}$ is

$$\frac{w_{g,k}\cdot k!}{\mathcal{Z}_m(\bar{F}_{g,k})\cdot(k-r)!}\,\frac{1}{(1-x_1)\cdots(1-x_1-\cdots-x_{r-1})} \cdot \sum_{\substack{(j_1,\dots,j_k)\in\mathbb{Z}^k_{\geq 1}\\ j_1+\cdots+j_k=3g-3}} \tilde{c}_{g,k}(j_1,\dots,j_k) \prod_{i=1}^{k} \zeta_m(2j_i)\,\frac{x_i^{2j_i-1}}{(2j_i)!} \int_{\Delta^{k-r-1}_{1-x_1-\cdots-x_r}} x_{r+1}^{2j_{r+1}-1}\cdots x_k^{2j_k-1}\lambda(dx) \tag{6.5}$$

where $\Delta^{k-r-1}_{1-x_1-\cdots-x_r} := \{(x_{r+1},\cdots,x_k)\in\mathbb{R}^{k-r}_{\geq 0} : x_{r+1}+\cdots+x_k = 1\}$ and λ is the Lebesgue measure on $\Delta^{k-r-1}_{1-x_1-\cdots-x_r}$. In the above, we used the fact that the density (6.4) is a symmetric function. Hence the sum over all permutations of k elements only pops out a $k!$ coefficient. The value of the integral in the above sum follows from Lemma 6.4.1 and is equal to

$$\frac{(2j_{r+1}-1)!\cdots(2j_k-1)!}{(2j_{r+1}+\cdots+2j_k-1)!}\,(1-x_1-\cdots-x_r)^{2j_{r+1}+\cdots+2j_k-1}.$$

We end up with the following formula for the distribution of the r-marginal density of $\widetilde{L}^*_{g,m,(k)}$

$$\frac{w_{g,k}}{\mathcal{Z}_m(\bar{F}_{g,k})\cdot(k-r)!} \cdot \sum_{\substack{(j_1,\dots,j_k)\in\mathbb{Z}^k_{\geq 1}\\ j_1+\cdots+j_k=3g-3}} \tilde{c}_{g,k}(j_1,\dots,j_k)\,\frac{(1-x_1-\cdots-x_r)^{2j_{r+1}+\cdots+2j_k-1}}{(1-x_1)\cdots(1-x_1-\cdots-x_{r-1})}\prod_{i=1}^{k}\frac{\zeta_m(2j_i)}{2j_i}\prod_{i=1}^{r}\frac{x_i^{2j_i-1}}{(2j_i-1)!}.$$

From the above formula and the definition of the moment M_p, we have

$$M_p\big(\widetilde{L}^*_{g,m,(k)}\big) = \frac{w_{g,k}}{\mathcal{Z}_m(\bar{F}_{g,k})\cdot(k-r)!}\sum_{\substack{(j_1,\dots,j_k)\in\mathbb{Z}^k_{\geq 1}\\ j_1+\cdots+j_k=3g-3}} \tilde{c}_{g,k}(j_1,\dots,j_k)\prod_{i=1}^{k}\frac{\zeta_m(2j_i)}{2j_i}\prod_{i=1}^{r}\frac{1}{(2j_i-1)!} \cdot\int_{\Delta^r_{\leq 1}} x_1^{2j_1+p_1}\cdots x_r^{2j_r+p_r}\cdot(1-x_1-\cdots-x_r)^{2j_{r+1}+\cdots+2j_k-1}\,dx \tag{6.6}$$

Again, applying Lemma 6.4.1, we obtain

$$\begin{aligned}\int_{\Delta^r_{\leq 1}} x_1^{2j_1+p_1}\cdots x_r^{2j_r+p_r}\cdot(1-x_1-\cdots-x_r)^{2j_{r+1}+\cdots+2j_k-1}\,dx &= \int_{\Delta^r_1} x_1^{2j_1+p_1}\cdots x_r^{2j_r+p_r}\cdot x_{r+1}^{2j_{r+1}+\cdots+2j_k-1}\,\lambda(dx)\\ &= \frac{(2j_1+p_1)!\cdots(2j_r+p_r)!\cdot(2j_{r+1}+\cdots+2j_k-1)!}{(6g-6+p_1+\cdots+p_r+r-1)!}\end{aligned} \tag{6.7}$$

Substituting from (6.7) into (6.6), we obtain the desired result. □

Now we are ready for the proof of the main result of this section.

Proof of Theorem 6.6.4. By Theorem 6.3.3,

$$M_p(\widetilde{L}^*_{g,m,\kappa}) = \sum_{k=1}^{\kappa\log(6g-6)/2}\frac{\widetilde{b}_{g,m,(k)}}{\widetilde{b}_{g,m,\kappa}}\cdot M_p\big(\widetilde{L}^*_{g,m,(k)}\big) = \frac{1}{\widetilde{b}_{g,m,\kappa}}\sum_{k=1}^{\kappa\log(6g-6)/2}\frac{\mathcal{Z}_m(\bar{F}_{g,k})}{(6g-6)!\,2^k\,k!}\cdot M_p\big(\widetilde{L}^*_{g,m,(k)}\big)$$

where $\widetilde{b}_{g,m,\kappa}$ is defined in (6.3). Now substituting the formula for $M_p\big(\widetilde{L}^*_{g,m,(k)}\big)$ from Lemma 6.6.5, we have as $g\to\infty$ the asymptotic equivalence

$$M_p(\widetilde{L}^*_{g,m,\kappa}) \sim \frac{1}{\widetilde{b}_{g,m,\kappa}}\sum_{k=1}^{\kappa\log(6g-6)/2}\frac{w_{g,k}}{(6g-6)!\,2^k\,(k-r)!\cdot(6g-7+p_1+\cdots+p_r+r)!} \cdot\sum_{\substack{(j_1,\dots,j_k)\in\mathbb{Z}^k_{\geq 1}\\ j_1+\cdots+j_k=3g-3}}\tilde{c}_{g,k}(j_1,\dots,j_k)\prod_{i=1}^{k}\frac{\zeta_m(2j_i)}{2j_i}\prod_{i=1}^{r}\frac{(2j_i+p_i)!}{(2j_i-1)!}. \tag{6.8}$$

On the one hand, [DGZZ20b, Equation (3.13)] (in the proof of Theorem 3.4), we have

$$\frac{w_{g,k}}{(6g-6)!(6g-7+p_1+\cdots+p_r+r)!\,2^k} \sim \frac{1}{(4g-4)!}\frac{1}{2\sqrt{\pi}}\frac{1}{(6g-6)^{p_1+\cdots+p_r+r-3/2}}\left(\frac{4}{3}\right)^{4g-4} \tag{6.9}$$

as $g \to \infty$. On the other hand,

$$\frac{(6g-7)!}{(6g-7+p_1+\cdots+p_r+r)!} \sim \frac{1}{(6g-6)^{p_1+\cdots+p_r+r}}. \tag{6.10}$$

Now, using Theorem 6.5.2 and the fact that $\tilde{c}_{g,k}(j_1,\ldots,j_k) \sim 1$ uniformly in $1 \le k \le \kappa\log(6g-6)/2$, and replacing (6.9) and (6.10) in (6.8), we obtain

$$M_p(\widetilde{L}^*_{g,m,\kappa}) \sim \frac{\sqrt{\pi(m+1)/m}}{2(6g-6)^{p_1+\cdots+p_r+r}} \sum_{k=1}^{\kappa\log(6g-6)/2} \frac{1}{(k-r)!} \sum_{\substack{(j_1,\ldots,j_k)\in\mathbb{Z}^k_{\ge 1}\\ j_1+\cdots+j_k=3g-3}} \prod_{i=1}^{k}\frac{\zeta_m(2j_i)}{2j_i}\prod_{i=1}^{r}\frac{(2j_i+p_i)!}{(2j_i-1)!}$$

the result desired. □

6.6.2 Asymptotic expression of a related sum

Let $\theta = (\theta_n)_{n\ge 1}$ be a sequence of non-negative real numbers and let $p = (p_1,\ldots,p_r) \in \mathbb{Z}^r_{\ge 0}$. This section is dedicated to the asymptotics in n of the numbers

$$S_{\theta,p,n} := \sum_{k=r}^{\infty}\frac{1}{(k-r)!}\sum_{\substack{(j_1,\ldots,j_k)\in\mathbb{Z}^k_{\ge 1}\\ j_1+\cdots+j_k=n}}\prod_{i=1}^{k}\frac{\theta_{j_i}}{2j_i}\prod_{i=1}^{r}\frac{(2j_i+p_i)!}{(2j_i-1)!} \tag{6.11}$$

which should be reminiscent of the formula from Theorem 6.6.4.

Let $\theta = (\theta_n)_{n\ge 1}$ be a sequence of non-negative real numbers and let $g_\theta(z)$ be the formal series

$$g_\theta(z) := \sum_{n=1}^{\infty}\theta_n\frac{z_n}{n}. \tag{6.12}$$

We say that $(\theta_n)_{n\ge 1}$ is *admissible* if

1. g_θ converges in the open disk $\mathbb{D}(0;1) \subset \mathbb{C}$ centered at 0 of radius 1,
2. $g_\theta(z) + \log(1-z)$ extends to a holomorphic function on $\mathbb{D}(0;R)$ with $R > 1$.

Theorem 6.6.6. *Let $\theta = (\theta_i)_{i\ge 1}$ be admissible. We have*

$$S_{\theta,p,n} \sim \sqrt{\frac{e^\beta}{2}}\,\frac{p_1!\cdots p_r!}{2^{r-1}}\,\frac{n^{p_1+\cdots+p_r+r-1/2}}{\Gamma(p_1+\cdots+p_r+r+1/2)}$$

where β is the value of $g_\theta(z)+\log(1-z)$ at $z=1$.

The following result is essentially [DGZZ20b, Lemma 3.8]. Since the proof is short, we include it for the sake of completeness.

Lemma 6.6.7. *Given $m \in \mathbb{Z}_{\geq 1} \cup \{\infty\}$, define*

$$g_m(z) := \sum_{j\geq 1} \zeta_m(2j)\, \frac{z^j}{j}$$

Then $g_m(z)$ is summable for $z \in \mathbb{D}(0;1)$ and $g_m(z)+\log(1-z)$ extends to a holomorphic function on $\mathbb{D}(0;4)$. In particular, the sequence $(\zeta(2j))_{j\geq 1}$ is admissible. Moreover, the value of $g_m(z) + \log(1-z)$ at $z = 1$ is $\log((2m)/(m+1))$.

Proof. The series g_m converges in $\mathbb{D}(0;1)$ since $\zeta_m(2j)$ is bounded uniformly in j. Thus,

$$g_m(z) = \sum_{j=1}^{\infty}\sum_{n=1}^{m} \frac{1}{n^{2j}}\frac{z^j}{j} = \sum_{n=1}^{m}\sum_{j=1}^{\infty} \frac{1}{n^{2j}}\frac{z^j}{j} = -\sum_{n=1}^{m} \log\left(1-\frac{z}{n^2}\right)$$

and hence

$$g_m(z) + \log(1-z) = -\sum_{n=2}^{m} \log\left(1-\frac{z}{n^2}\right).$$

Each term $\log(1-z/n^2)$ in the sum is a holomorphic function in $\mathbb{D}(0;n^2)$. It follows from

$$\left|\log\left(1-\frac{z}{n^2}\right)\right| \leq \frac{2}{n^2}\,|z|$$

that the series $g_m(z)+\log(1-z)$ converges absolutely even for $m = +\infty$ for $|z| < 4$, and defines a holomorphic function in $\mathbb{D}(0;4)$. Finally, the value of $g_m(z) - \log(1-z)$ is

$$-\sum_{n=2}^{m} \log\left(1-\frac{1}{n^2}\right) = \sum_{n=2}^{m} (2\log(n) - \log(n-1) - \log(n+1)) = \log\left(\frac{2m}{m+1}\right).$$

This completes the proof. □

Corollary 6.6.8. *Let $m \in \mathbb{Z}_{\geq 1} \cup \{+\infty\}$. For $\theta = (\zeta_m(2j))_{j\geq 1}$ we have*

$$S_{\theta,p,n} \sim \sqrt{\frac{m}{m+1}}\,\frac{p_1!\cdots p_r!}{2^{r-1}}\,\frac{n^{p_1+\cdots p_r-1/2}}{\Gamma(p_1+\cdots+p_r+r+1/2)}$$

as $n \to \infty$.

Given $n \in \mathbb{Z}_{\geq 0}$, we define the differential operator D_n on $\mathbb{C}[[z]]$ by

$$(D_n f)(z) := z\frac{d^{n+1}}{dz^{n+1}}(z^n f(z)).$$

Let us start with some preliminary lemmas.

Lemma 6.6.9. *Let $\theta = (\theta_i)_{i\geq 1}$ and $g_\theta(z)$ be defined by (6.12), and $S_{\theta,p,n}$ be defined by (6.11). Define*

$$G_{\theta,p}(z) := \exp\left(\frac{1}{2}\,g_\theta(z^2)\right)\prod_{i=1}^{r} D_{p_i}\left(\frac{1}{2}\,g_\theta(z^2)\right).$$

We have, for any $n \geq 0$,

$$[z^{2n}]\,G_{\theta,p}(z) = S_{\theta,p,n}$$

where $[z^n]\, f(z)$ *denotes the operation of extracting the coefficient of* z^n *in the formal power series* $f(z) = \sum_{n\geq 0} a_n z^n$.

Proof. We observe first that

$$\begin{aligned}
S_{\theta,p,n} &= \sum_{k=1}^{\infty} \frac{1}{k!} \sum_{\substack{(j_1,\dots,j_{k+r})\in\mathbb{Z}_{\geq 1}^{k+r} \\ j_1+\cdots+j_{k+r}=n}} \left(\prod_{i=1}^{k} \frac{\theta_{j_i}}{2j_i}\right) \left(\prod_{i=1}^{r} \frac{(2j_{k+i}+p_{k+i})!}{(2j_{k+i}-1)!} \frac{\theta_{j_{k+i}}}{2j_{k+i}}\right) \\
&= \sum_{k=1}^{\infty} \sum_{\substack{(j_1,\dots,j_k,\tilde{j}_1,\dots\tilde{j}_r)\in\mathbb{Z}_{\geq 1}^{k+r} \\ j_1+\cdots+j_k+\tilde{j}_1+\cdots+\tilde{j}_r=n}} \left(\frac{1}{k!}\prod_{i=1}^{k} \frac{\theta_{j_i}}{2j_i}\right) \left(\prod_{i=1}^{r} \frac{(2\tilde{j}_i+p_i)!}{(2\tilde{j}_i-1)!} \frac{\theta_{\tilde{j}_i}}{2\tilde{j}_i}\right) \\
&= [z^{2n}] \left(\sum_{k=1}^{\infty} \frac{1}{k!} \left(\sum_{j=1}^{\infty} \frac{\theta_j}{2j} z^{2j}\right)^k\right) \prod_{i=1}^{r} \left(\sum_{j_i=1}^{\infty} \frac{(2j_i+p_i)!}{(2j_i-1)!} \frac{\theta_{j_i}}{2j_i} z^{2j_i}\right) \\
&= [z^{2n}] \exp\left(\frac{1}{2} g_\theta(z^2)\right) \prod_{i=1}^{r} \left(\sum_{j_i=1}^{\infty} \frac{(2j_i+p_i)!}{(2j_i-1)!} \frac{\theta_{j_i}}{2j_i} z^{2j_i}\right).
\end{aligned}$$

It follows from the fact

$$D_n(z^{2j}) = z\frac{d^{n+1}}{dz^{n+1}}(z^{2j+n}) = \frac{(2j+n)!}{(2j-1)!} z^{2j}$$

that

$$D_n\left(\frac{1}{2} g_\theta(z^2)\right) = \sum_{j=1}^{\infty} \frac{(2j+n)!}{(2j-1)!} \frac{\theta_j}{2j} z^{2j},$$

and the proof of the lemma is completed. □

Lemma 6.6.10. *For any* $n \in \mathbb{Z}_{\geq 0}$, *we have*

$$\frac{1}{n!} D_n(-\log(1\pm z)) = \frac{1}{(1\pm z)^{n+1}} - 1.$$

Proof. By Leibniz's rule,

$$\begin{aligned}
\frac{z}{n}\frac{d^{n+1}}{dz^{n+1}}\left(z^n \log\frac{1}{1-z}\right) &= \frac{z}{n!}\sum_{i=0}^{n} \binom{n+1}{i} n(n-1)\cdots(n-i+1) z^{n-i} \frac{(n-i)!}{(1-z)^{n+1-i}} \\
&= -1 + \sum_{i=0}^{n+1} \binom{n+1}{i} \left(\frac{z}{1-z}\right)^{n-i+1} \\
&= -1 + \left(1+\frac{z}{1-z}\right)^{n+1} \\
&= -1 + \frac{1}{(1-z)^{n+1}}.
\end{aligned}$$

The case $-\log(1+z)$ maybe be treated similarly. □

Proof of Theorem 6.6.6. By the assumption that θ is admissible, we can write $g_\theta(z^2) = -\log(1-z^2) + \beta + r_\theta(z)$ where $r_\theta(z)$ is holomorphic in $\mathbb{D}(0;\sqrt{R})$ and $r_\theta(1) = 0$. From Lemma 6.6.10 we

deduce that, for any $n \in \mathbb{Z}_{\geq 0}$,

$$D_n(g_\theta(z^2)) = \frac{n!}{(1-z)^{n+1}} + \frac{n!}{(1+z)^{n+1}} + r_{\theta,p}(z)$$

where $r_{\theta,p}$ is holomorphic in $\mathbb{D}(0;\sqrt{R})$. Therefore, $G_{\theta,p}$ can be extended to a holomorphic function in $\mathbb{D}(0;\sqrt{R}) \setminus \{-1,1\}$. Further, as $z \to 1$,

$$\begin{aligned} G_{\theta,p}(z) &= \exp\left(-\frac{1}{2}(\log(1-z)+\log(2)-\beta)+\mathrm{O}(1-z)\right)\prod_{i=1}^{r}\frac{p_i!}{2}\left(\frac{1}{(1-z)^{p_i+1}}+\mathrm{O}(1)\right) \\ &= \sqrt{\frac{e^\beta}{2}}\,\frac{p_1!\cdots p_r!}{2^r}\,\frac{1}{(1-z)^{p_1+\cdots+p_r+r+1/2}}\,(1+\mathrm{o}(1)), \end{aligned}$$

and similarly,

$$G_{\theta,p}(z) = \sqrt{\frac{e^\beta}{2}}\,\frac{p_1!\cdots p_r!}{2^r}\,\frac{1}{(1+z)^{p_1+\cdots+p_r+r+1/2}}\,(1+\mathrm{o}(1)),$$

as $z \to -1$. Now applying Theorem 3.6.5, we obtain

$$[z^{2n}]\,G_{\theta,p}(z) \sim 2\sqrt{\frac{e^\beta}{2}}\,\frac{p_1!\cdots p_r!}{2^r}\,\frac{(2n)^{p_1+\cdots+p_r+r-1/2}}{\Gamma(p_1+\cdots+p_r+r+1/2)}$$

as $n \to \infty$. The proof is thus completed. □

6.6.3 Truncation error estimate

Recall that Theorem 6.6.4 provides an expression for the moment $M_p(\widetilde{L}^*_{g,m,\kappa})$ which involves a sum which is a truncated version of $S_{\theta,p,n}$ from (6.11). In this section, we show that the difference between $S_{\theta,p,n}$ and its truncation is negligible compared to the asymptotics in Theorem 6.6.4.

Theorem 6.6.11. *Let $\theta = (\theta_n)_{n\geq 1}$ be admissible, $g_\theta(z)$ be define by (6.12), and $S_{\theta,p,n}$ be defined by (6.11). For any $\kappa \in \mathbb{R}_{>1}$, we have*

$$\sum_{k=r}^{\kappa\log(2n)/2}\frac{1}{(k-r)!}\sum_{\substack{(j_1,\ldots,j_k)\in\mathbb{Z}^k_{\geq 1}\\ j_1+\cdots+j_k=n}}\prod_{i=1}^{k}\frac{\theta_{j_i}}{2j_i}\prod_{i=1}^{r}\frac{(2j_i+p_i)!}{(2j_i-1)!} \sim S_{\theta,p,n} \tag{6.13}$$

as $n \to \infty$.

Bounding the coefficient in a Taylor expansion is a standard tool in asymptotic analysis as the "Big-O transfer". However, in our situation we need to bound the n-th Taylor coefficient of a function f_n which depends on n. To do so, we track down the dependences on the function inside the transfer theorem.

Lemma 6.6.12 ([DGZZ20b, Lemma 4.4])**.** *Let $\lambda, x \in \mathbb{R}_{>0}$. We have*

$$\sum_{k=\lceil x\lambda\rceil}^{\infty}\frac{\lambda^k}{k!} \leq \exp(-\lambda(x\log x - x)).$$

Lemma 6.6.13. *Let $h(z)$ be a holomorphic function on $\mathbb{D}(0;R) \smallsetminus \{\,]-R,-1] \cup [1,R[\,\}$ satisfying*

$$h(z) = -\frac{1}{2}\log(1-z^2) + \mathrm{O}(1).$$

as $z \to \pm 1$. Let $\kappa \in \mathbb{R}_{>0}$, $p,q \in \mathbb{Z}_{\geq 0}$. For each $n \in \mathbb{Z}_{\geq 1}$, define

$$f_n(z) := \frac{1}{(1-z)^p\,(1+z)^q} \sum_{k=\lfloor \kappa \log(n)/2 \rfloor}^{\infty} \frac{h(z)^k}{k!}.$$

Then,

$$[z^n]\, f_n(z) = \mathrm{O}\Big(n^{\max\{p,q\}-1-(\kappa\log\kappa-\kappa)/2}\Big)$$

as $n \to \infty$.

Proof. Let $0 < \eta < R-1$ and $0 < \phi < \pi/2$ and define the contour γ as the union $\sigma_+ \cup \sigma_- \cup \lambda_{\nearrow} \cup \lambda_{\nwarrow} \cup \lambda_{\searrow} \cup \lambda_{\swarrow} \cup \lambda_{\nearrow} \cup \Sigma_+ \cup \Sigma_-$ with

$$\begin{aligned}
\sigma_+ &= \{z : |z-1| = 1/n,\ |\arg(z-1)| \geq \phi\},\\
\sigma_- &= \{z : |z+1| = 1/n,\ |\arg(z-1)| \geq \phi\},\\
\lambda_{\nearrow} &= \{z : |z-1| \geq 1/n,\ |z| \leq 1+\eta, \arg(z-1) = \phi\},\\
\lambda_{\nwarrow} &= \{z : |z-1| \geq 1/n,\ |z| \leq 1+\eta, \arg(z-1) = -\phi\},\\
\lambda_{\searrow} &= \{z : |z+1| \geq 1/n,\ |z| \leq 1+\eta, \arg(z-1) = \pi-\phi\},\\
\lambda_{\swarrow} &= \{z : |z+1| \geq 1/n,\ |z| \leq 1+\eta, \arg(z-1) = -\pi+\phi\},\\
\Sigma_+ &= \{z : |z| = 1+\eta,\ \arg(z-1) \geq \phi,\ \arg(z+1) \leq \pi-\phi\},\\
\Sigma_- &= \{z : |z| = 1+\eta,\ \arg(z-1) \leq -\phi,\ \arg(z+1) \geq -\pi+\phi\}.
\end{aligned}$$

See Figure 6.1 for a picture of γ. Since f_n is holomorphic on $\mathbb{D}(0;R) \smallsetminus \{\,]-R,-1] \cup [1,R[\,\}$, we

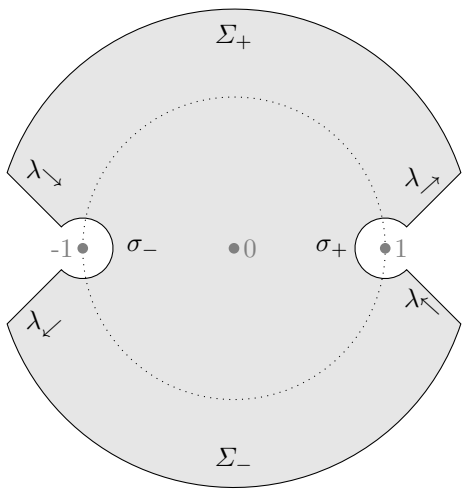

Figure 6.1: The contour γ.

have the Cauchy's residue theorem for its coefficients

$$[z^n]\, f_n(z) = \frac{1}{2\pi i}\int_\gamma \frac{f_n(z)}{z^{n+1}}\, dz. \tag{6.14}$$

Taking absolute values in (6.14) we obtain

$$|[z^n]\, f_n(z)| \le \frac{1}{2\pi}\int_\gamma \frac{|dz|}{|z|^{n+1}} \frac{1}{|1-z|^p|1+z|^q} \sum_{k=\lfloor \kappa \log(n)/2\rfloor}^{\infty} \frac{|h(z)|^k}{k!}. \tag{6.15}$$

The proof proceeds by analyzing the right-hand side in (6.15) for each piece of the contour γ.

Let us start with the small arc of the circle σ_+. The change of variables $z = 1 - e^{i\theta}/n$ yields

$$\left|\frac{1}{2\pi i}\int_{\sigma_+} \frac{f_n(z)dz}{z^{n+1}}\right| \le \frac{n^{p-1}}{2\pi\cdot(R+1)^q}\int_{-\pi+\phi}^{\pi-\phi} \frac{d\theta}{|1-e^{i\theta}/n|^{n+1}} \sum_{k=\lfloor \kappa \log(n)/2\rfloor}^{\infty} \frac{|h(1-e^{i\theta}/n)|^k}{k!}.$$

First $h(1-e^{i\theta}/n) = \log(n)/2 + \mathrm{O}(1)$ uniformly in θ. Hence, by Lemma 6.6.12, uniformly in θ as $n \to \infty$ we have

$$\sum_{k=\lceil \kappa \log(n)/2\rceil}^{\infty} \frac{|h(z)|^k}{k!} \le \exp\left(-(\kappa\log\kappa - \kappa)\cdot\frac{\log n + \mathrm{O}(1)}{2}\right) = \mathrm{O}\Big(n^{-\frac{1}{2}(\kappa\log\kappa-\kappa)}\Big).$$

Since $1/|1-e^{i\theta}/n|^{n+1}$ is uniformly bounded in n,

$$\left|\frac{1}{2\pi i}\int_{\sigma_+} \frac{f_n(z)\, dz}{z^{n+1}}\right| = \mathrm{O}\Big(n^{p-1-\frac{1}{2}(\kappa\log\kappa-\kappa)}\Big).$$

Similarly,

$$\left|\frac{1}{2\pi i}\int_{\sigma_-} \frac{f_n(z)\, dz}{z^{n+1}}\right| = \mathrm{O}\Big(n^{q-1-\frac{1}{2}(\kappa\log\kappa-\kappa)}\Big).$$

Let us now consider the case of $\lambda_\nearrow$. Let r be the positive solution of the equation $|1+re^{i\phi}| = 1+\eta$. Perform the change of variable $z = 1 + e^{i\phi}\cdot t/n$, we have

$$\begin{aligned}&\left|\frac{1}{2\pi i}\int_{\lambda_\nearrow} \frac{f(z)}{z^{n+1}}\, dz\right| \\ &\quad\le \frac{n^{p-1}}{2\pi\cdot(R+1)^q}\int_1^{nr} dt\cdot t^{-p}\left|1+e^{i\phi}t/n\right|^{-n-1} \sum_{k=\lfloor \kappa \log(n)/2\rfloor}^{\infty} \frac{|h(1+e^{i\phi}t/n)|^k}{k!}.\end{aligned}$$

For n large enough and uniformly in t, $|h(1+e^{i\phi}t/n)| = \log(n)/2 + \mathrm{O}(1)$. Lemma 6.6.12 gives

$$\sum_{k=\lfloor \kappa \log(n)/2\rfloor}^{\infty} \frac{|h(1+e^{i/\phi}t/n|^k}{k!} = \mathrm{O}\Big(n^{-\frac{1}{2}(\kappa\log\kappa-\kappa)}\Big).$$

From the boundedness of $|1+e^{i\phi}t/n|^{-n-1}$ it follows that

$$\int_1^{nr} t^{-p}|1+e^{i\phi}\cdot t/n|^{-n-1}\, dt = \mathrm{O}(1),$$

and therefore

$$\left|\frac{1}{2\pi i}\int_{\lambda_{\nearrow}}\frac{f(z)dz}{z^{n+1}}\right| = \mathrm{O}\Big(n^{-\frac{1}{2}(\kappa\log\kappa-\kappa)}\Big).$$

The same estimate is valid for the integral along the other three segments $\lambda_{\nwarrow}$, $\lambda_{\searrow}$, and $\lambda_{\swarrow}$.

For the two large demi-circles Σ_+ and Σ_-, we have

$$\left|\frac{1}{2\pi i}\int_{\Sigma_+}\frac{f(z)dz}{z^{n+1}}\right| \leq \frac{1}{2\pi}\cdot\eta^{p+1/2}\cdot(1+\eta)^{-n-1}\cdot 2\pi(1+\eta) = \frac{\eta^{p+1/2}}{(1+\eta)^{n+1}}$$

which decreases exponentially fast.

We conclude the proof by combining the above estimates. □

Proof of Theorem 6.6.11. Similarly to Lemma 6.6.9, if we write

$$G_{n,\theta,p}(z) := \sum_{k=\lfloor\log(n)/2\rfloor}^{\infty}\frac{1}{k!}\left(\frac{1}{2}g_\theta(z^2)\right)^k\cdot\prod_{i=1}^{r}D_{p_i}\left(\frac{1}{2}g_\theta(z^2)\right).$$

then $[z^{2n}]\,G_{n,\theta,p}(z)$ is the complement of the partial sum in the right hand side of (6.13). Following the proof of Theorem 6.6.6 we obtain as $z\to 1$

$$G_{n,\theta,p}(z) = \sum_{k=\lfloor\log(n)/2\rfloor}^{\infty}\frac{1}{k!}\left(\frac{1}{2}g_\theta(z^2)\right)^k\cdot\prod_{i=1}^{r}\frac{p!}{2}\left(\frac{1}{(1-z)^{p_i+1}}+\mathrm{O}(1)\right)$$

where the $\mathrm{O}(1)$ is uniform in n (it only depends on $g_\theta(z)$). Applying Lemma 6.6.13 we obtain

$$[z^{2n}]\,G_{n,\theta,p}(z) = \mathrm{O}\Big((2n)^{p_1+\cdots+p_r+r-1-\frac{1}{2}(\kappa\log\kappa-\kappa)}\Big).$$

For $\kappa>1$ we have $-1-\frac{1}{2}(\kappa\log\kappa-\kappa)<-1/2$ and the above sum is negligible compared to the asymptotics of the full sum $S_{\theta,p,n}$ from Theorem 6.6.6. □

6.6.4 Proof of Theorem 6.6.1

Now we are ready for the proof of the main result of the section.

Proof of Theorem 6.6.1. By Lemma 6.6.3, it suffices to prove the convergence of the moments $M_p(\widetilde{L}^*_{g,m,\kappa})$ for all $p=(p_1,\dots,p_r)$ towards the moments of the GEM(1/2) distribution that were computed in Lemma 6.6.2.

Now, Theorem 6.6.4 provides an asymptotic equivalence of $M_p(\tilde{L}^*_{g,m,\kappa})$ involving the sum

$$\sum_{k=r}^{\kappa\log(6g-6)/2}\frac{1}{(k-r)!}\sum_{\substack{(j_1,\dots,j_k)\in\mathbb{Z}^k_{\geq 1}\\ j_1+\cdots+j_k=3g-3}}\prod_{i=1}^{k}\frac{\zeta_m(2j_i)}{2j_i}\prod_{i=1}^{r}\frac{(2j_i+p_i)!}{(2j_i-1)!}.$$

The asymptotics of the above sum was then obtained from Corollary 6.6.8 and Theorem 6.6.11. Namely, the above is asymptotically equivalent to

$$\sqrt{\frac{m}{m+1}}\cdot\frac{p_1!\cdots p_r!}{2^{r-1}}\cdot\frac{(6g-6)^{p_1+\cdots+p_r+r-1/2}}{\Gamma(p_1+\cdots+p_r+r+1/2)}$$

as $g \to \infty$. Substituting this value in the formula of Theorem 6.6.4 we obtain as $g \to \infty$

$$M_p(\widetilde{L}^*_{g,m,\kappa}) \sim \frac{\sqrt{\pi}}{2^r} \cdot \frac{p_1! \cdots p_r!}{\Gamma(p_1 + \cdots + p_r + r + 1/2)}.$$

The above is the value of the M_p-moments of the distribution GEM(1/2) from Lemma 6.6.2 as $\theta = 1/2$ and $(\theta - 1)! = (-1/2)! = \Gamma(1/2) = \sqrt{\pi}$.

Since the convergence of $M_p(\widetilde{L}^*_{g,m,\kappa})$ holds for all $p = (p_1, \ldots, p_r)$, the sequence $\widetilde{L}^*_{g,m,\kappa}$ converges in distribution towards GEM(1/2). $\square$

Bibliography

[ABC+19] Jørgen Ellegaard Andersen, Gaëtan Borot, Séverin Charbonnier, Vincent Delecroix, Alessandro Giacchetto, Danilo Lewanski, and Campbell Wheeler. Topological recursion for Masur–Veech volumes, 2019.

[ABC+20] Jørgen Ellegaard Andersen, Gaëtan Borot, Séverin Charbonnier, Alessandro Giacchetto, Danilo Lewański, and Campbell Wheeler. On the Kontsevich geometry of the combinatorial Teichmüller space, 2020.

[ABO17] Jørgen Ellegaard Andersen, Gaëtan Borot, and Nicolas Orantin. Geometric recursion, 2017.

[ABT03] Richard Arratia, A. D. Barbour, and Simon Tavaré. *Logarithmic combinatorial structures: A probabilistic approach.* EMS Monogr. Math. Zürich: European Mathematical Society (EMS), 2003.

[ADG+20] Amol Aggarwal, Vincent Delecroix, Elise Goujard, Peter Zograf, and Anton Zorich. Conjectural large genus asymptotics of Masur–Veech volumes and of area Siegel–Veech constants of strata of quadratic differentials. *Arnold Math. J.*, 6(2):149–161, 2020.

[AEZ16] Jayadev S. Athreya, Alex Eskin, and Anton Zorich. Right-angled billiards and volumes of moduli spaces of quadratic differentials on $\mathbb{C}\mathrm{P}^1$. *Ann. Sci. Éc. Norm. Supér. (4)*, 49(6):1311–1386, 2016.

[Agg20] Amol Aggarwal. Large genus asymptotics for volumes of strata of Abelian differentials. *J. Am. Math. Soc.*, 33(4):941–989, 2020.

[Agg21] Amol Aggarwal. Large genus asymptotics for intersection numbers and principal strata volumes of quadratic differentials. *Invent. Math.*, 226(3):897–1010, 2021.

[AH20a] Francisco Arana-Herrera. Counting hyperbolic multi-geodesics with respect to the lengths of individual components, 2020.

[AH20b] Francisco Arana-Herrera. Counting square-tiled surfaces with prescribed real and imaginary foliations and connections to Mirzakhani's asymptotics for simple closed hyperbolic geodesics. *J. Mod. Dyn.*, 16:81–107, 2020.

[AH21a] Francisco Arana-Herrera. Effective mapping class group dynamics III: Counting filling closed curves on surfaces, 2021.

[AH21b] Francisco Arana-Herrera. Equidistribution of families of expanding horospheres on moduli spaces of hyperbolic surfaces. *Geom. Dedicata*, 210:65–102, 2021.

[AH22] Francisco Arana-Herrera. Counting problems from the viewpoint of ergodic theory: from primitive integer points to simple closed curves, 2022.

[BCP21] Thomas Budzinski, Nicolas Curien, and Bram Petri. The diameter of random Belyĭ surfaces. *Algebr. Geom. Topol.*, 21(6):2929–2957, 2021.

[Bil72] P. Billingsley. On the distribution of large prime divisors. *Period. Math. Hung.*, 2:283–289, 1972.

[Bil95] Patrick Billingsley. *Probability and measure.* Chichester: John Wiley & Sons Ltd., 3rd ed. edition, 1995.

[BM04] Robert Brooks and Eran Makover. Random construction of Riemann surfaces. *J. Differ. Geom.*, 68(1):121–157, 2004.

[Bon88] Francis Bonahon. The geometry of Teichmüller space via geodesic currents. *Invent. Math.*, 92(1):139–162, 1988.

[Bon01] Francis Bonahon. Geodesic laminations on surfaces. In *Laminations and foliations in dynamics, geometry and topology. Proceedings of the conference held at SUNY at Stony Brook, USA, May 18–24, 1998*, pages 1–37. Providence, RI: American Mathematical Society (AMS), 2001.

[BS85] Joan S. Birman and Caroline Series. Geodesics with bounded intersection number on surfaces are sparsely distributed. *Topology*, 24:217–225, 1985.

[Bus92] Peter Buser. *Geometry and spectra of compact Riemann surfaces*, volume 106 of *Prog. Math.* Boston, MA: Birkhäuser, 1992.

[Cha11] Guillaume Chapuy. A new combinatorial identity for unicellular maps, via a direct bijective approach. *Adv. Appl. Math.*, 47(4):874–893, 2011.

[CMS+19] D. Chen, M. Möller, A. Sauvaget, with an appendix by G. Borot, A. Giacchetto, and D. Lewanski. Masur–veech volumes and intersection theory: the principal strata of quadratic differentials, 2019.

[CMSZ20] Dawei Chen, Martin Möller, Adrien Sauvaget, and Don Zagier. Masur–Veech volumes and intersection theory on moduli spaces of Abelian differentials. *Invent. Math.*, 222(1):283–373, 2020.

[DGZZ16] Vincent Delecroix, Elise Goujard, Peter Zograf, and Anton Zorich. Square-tiled surfaces of fixed combinatorial type: equidistribution, counting, volumes of the ambient strata, 2016.

[DGZZ20a] Vincent Delecroix, Elise Goujard, Peter Zograf, and Anton Zorich. Contribution of one-cylinder square-tiled surfaces to Masur–Veech volumes. In *Some aspects of the theory of dynamical systems: a tribute to Jean-Christophe Yoccoz. Volume I*, pages 223–274. Paris: Société Mathématique de France (SMF), 2020.

[DGZZ20b] Vincent Delecroix, Elise Goujard, Peter Zograf, and Anton Zorich. Large genus asymptotic geometry of random square-tiled surfaces and of random multicurves, 2020.

[DGZZ20c] Vincent Delecroix, Elise Goujard, Peter Zograf, and Anton Zorich. Uniform lower bound for intersection numbers of ψ-classes. *SIGMA, Symmetry Integrability Geom. Methods Appl.*, 16:paper 086, 13, 2020.

[DGZZ21] Vincent Delecroix, Elise Goujard, Peter Zograf, and Anton Zorich. Masur–Veech volumes, frequencies of simple closed geodesics, and intersection numbers of moduli spaces of curves. *Duke Mathematical Journal*, 170(12):2633–2718, 2021.

[DJ89] Peter Donnelly and Paul Joyce. Continuity and weak convergence of ranked and size-biased permutations on the infinite simplex. *Stochastic Processes and their Applications*, 31(1):89–103, 1989.

[DL22] Vincent Delecroix and Mingkun Liu. Length partition of random multicurves on large genus hyperbolic surfaces, 2022.

[DM69] Pierre Deligne and David Mumford. The irreducibility of the space of curves of a given genus. *Publ. Math., Inst. Hautes Étud. Sci.*, 36:75–109, 1969.

[Do15] Norman Do. Moduli spaces of hyperbolic surfaces and their Weil–Petersson volumes. In *Handbook of moduli. Volume I*, pages 217–258. Somerville, MA: International Press; Beijing: Higher Education Press, 2015.

[DVH] Vincent Delecroix, Ferrán Valdez, and Pascal Hubert. *Infinite translation in the wild.*

[DVV91] Robbert Dijkgraaf, Herman Verlinde, and Erik Verlinde. Loop equations and virasoro constraints in non-perturbative two-dimensional quantum gravity. *Nuclear Physics B*, 348(3):435–456, 1991.

[DZ20] Vincent Delecroix and Anton Zorich. Cries and whispers in wind-tree forests. In *What's next? The mathematical legacy of William P. Thurston*, pages 83–115. Princeton, NJ: Princeton University Press, 2020.

[EKZ14] Alex Eskin, Maxim Kontsevich, and Anton Zorich. Sum of Lyapunov exponents of the Hodge bundle with respect to the Teichmüller geodesic flow. *Publ. Math., Inst. Hautes Étud. Sci.*, 120:207–333, 2014.

[EM93] Alex Eskin and Curt McMullen. Mixing, counting, and equidistribution in Lie groups. *Duke Mathematical Journal*, 71(1):181–209, 1993.

[EMM21] Alex Eskin, Maryam Mirzakhani, and Amir Mohammadi. Effective counting of simple closed geodesics on hyperbolic surfaces. *Journal of the European Mathematical Society*, 24(9):3059–3108, Oct 2021.

[EMZ03] Alex Eskin, Howard Masur, and Anton Zorich. Moduli spaces of Abelian differentials: the principal boundary, counting problems, and the Siegel–Veech constants. *Publ. Math., Inst. Hautes Étud. Sci.*, 97:61–179, 2003.

[EO01] Alex Eskin and Andrei Okounkov. Asymptotics of numbers of branched coverings of a torus and volumes of moduli spaces of holomorphic differentials. *Invent. Math.*, 145(1):59–103, 2001.

[EO06] Alex Eskin and Andrei Okounkov. Pillowcases and quasimodular forms. In *Algebraic geometry and number theory. In Honor of Vladimir Drinfeld's 50th birthday*, pages 1–25. Basel: Birkhäuser, 2006.

[EOP08] Alex Eskin, Andrei Okounkov, and Rahul Pandharipande. The theta characteristic of a branched covering. *Adv. Math.*, 217(3):873–888, 2008.

[EPS20] Viveka Erlandsson, Hugo Parlier, and Juan Souto. Counting curves, and the stable length of currents. *J. Eur. Math. Soc. (JEMS)*, 22(6):1675–1702, 2020.

[ES16] Viveka Erlandsson and Juan Souto. Counting curves in hyperbolic surfaces. *Geom. Funct. Anal.*, 26(3):729–777, 2016.

[ES20] Viveka Erlandsson and Juan Souto. Counting curves on orbifolds, 2020.

[ES22] Viveka Erlandsson and Juan Souto. *Mirzakhani's curve counting and geodesic currents*, volume 345 of *Prog. Math.* Birkhäuser Cham, 2022.

[Ewe72] W.J. Ewens. The sampling theory of selectively neutral alleles. *Theoretical Population Biology*, 3(1):87–112, 1972.

[FKN91] Masafumi Fukama, Hikaru Kawai, and Ryuichi Nakayama. Continuum schwinger-dyson equations and universal structures in two-dimensional quantum gravity. *International Journal of Modern Physics A*, 06(08):1385–1406, 1991.

[FM12] Benson Farb and Dan Margalit. *A Primer on Mapping Class Groups (PMS-49)*. Princeton University Press, 2012.

[FM14] Giovanni Forni and Carlos Matheus. Introduction to Teichmüller theory and its applications to dynamics of interval exchange transformations, flows on surfaces and billiards. *J. Mod. Dyn.*, 8(3-4):271–436, 2014.

[FO90] Philippe Flajolet and Andrew Odlyzko. Singularity analysis of generating functions. *SIAM J. Discrete Math.*, 3(2):216–240, 1990.

[FS09] Philippe Flajolet and Robert Sedgewick. *Analytic combinatorics*. Cambridge: Cambridge University Press, 2009.

[GLMST21] Clifford Gilmore, Etienne Le Masson, Tuomas Sahlsten, and Joe Thomas. Short geodesic loops and L^p norms of eigenfunctions on large genus random surfaces. *Geom. Funct. Anal.*, 31(1):62–110, 2021.

[GM20] Elise Goujard and Martin Möller. Counting Feynman-like graphs: quasimodularity and Siegel–Veech weight. *J. Eur. Math. Soc. (JEMS)*, 22(2):365–412, 2020.

[Gol84] William M Goldman. The symplectic nature of fundamental groups of surfaces. *Advances in Mathematics*, 54(2):200–225, 1984.

[Gou15] Elise Goujard. Siegel–Veech constants for strata of moduli spaces of quadratic differentials. *Geom. Funct. Anal.*, 25(5):1440–1492, 2015.

[GPY11] Larry Guth, Hugo Parlier, and Robert Young. Pants decompositions of random surfaces. *Geom. Funct. Anal.*, 21(5):1069–1090, 2011.

[Gri79] R. C. Griffiths. On the distribution of allele frequencies in a diffusion model. *Theor. Popul. Biol.*, 15:140–158, 1979.

[Gri88] R. C. Griffiths. On the distribution of points in a Poisson–Dirichlet process. *J. Appl. Probab.*, 25(2):336–345, 1988.

[GY21] Jindong Guo and Di Yang. On the large genus asymptotics of psi-class intersection numbers, 2021.

[Hid21] Will Hide. Spectral gap for weil–petersson random surfaces with cusps, 2021.

[HM98] Joe Harris and Ian Morrison. *Moduli of curves*, volume 187 of *Grad. Texts Math.* New York, NY: Springer, 1998.

[HM21] Will Hide and Michael Magee. Near optimal spectral gaps for hyperbolic surfaces, 2021.

[Hub59] Heinz Huber. Zur analytischen Theorie hyperbolischer Raumformen und Bewegungsgruppen. *Math. Ann.*, 138:1–26, 1959.

[Hub06] John Hamal Hubbard. *Teichmüller theory and applications to geometry, topology, and dynamics. Volume 1: Teichmüller theory. With contributions by Adrien Douady, William Dunbar, and Roland Roeder, Sylvain Bonnot, David Brown, Allen Hatcher, Chris Hruska, Sudeb Mitra.* Ithaca, NY: Matrix Editions, 2006.

[Hwa94] Hsien-Kuei Hwang. *Théorèmes limites pour les structures combinatoires et les fonctions arithmétiques.* PhD thesis, 1994. Thèse de doctorat dirigée par Jean-Marc Steyaert, Sciences appliquées Palaiseau, Ecole polytechnique 1994.

[IT92] Yoichi Imayoshi and Masahiko Taniguchi. *An introduction to Teichmüller spaces.* Tokyo: Springer-Verlag, 1992.

[JL21] Svante Janson and Baptiste Louf. Unicellular maps vs hyperbolic surfaces in large genus: simple closed curves, 2021.

[Kap01] Michael Kapovich. *Hyperbolic manifolds and discrete groups*, volume 183 of *Prog. Math.* Boston, MA: Birkhäuser, 2001.

[Kat92] Svetlana Katok. *Fuchsian groups.* Chicago: The University of Chicago Press, 1992.

[Kin75] J. F. C. Kingman. Random discrete distributions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 37(1):1–22, 1975.

[Kin77] J. F. C. Kingman. The population structure associated with the Ewens sampling formula. *Theor. Popul. Biol.*, 11:274–283, 1977.

[KL07] M. E. Kazarian and S. K. Lando. An algebro-geometric proof of Witten's conjecture. *J. Am. Math. Soc.*, 20(4):1079–1089, 2007.

[KN10] Emmanuel Kowalski and Ashkan Nikeghbali. Mod-Poisson convergence in probability and number theory. *Int. Math. Res. Not.*, 2010(18):3549–3587, 2010.

[Kon92] Maxim Kontsevich. Intersection theory on the moduli space of curves and the matrix Airy function. *Communications in Mathematical Physics*, 147(1):1–23, 1992.

[Lab13] François Labourie. *Lectures on representations of surface groups.* Zur. Lect. Adv. Math. Zürich: European Mathematical Society (EMS), 2013.

[LG19] Jean-François Le Gall. Brownian geometry. *Jpn. J. Math. (3)*, 14(2):135–174, 2019.

[Liu19] Mingkun Liu. Length statistics of random multicurves on closed hyperbolic surfaces, 2019.

[LW21] Michael Lipnowski and Alex Wright. Towards optimal spectral gaps in large genus, 2021.

[LZ04] S. K. Lando and A. K. Zvonkin. *Graphs on surfaces and their applications. Appendix by Don B. Zagier*, volume 141 of *Encycl. Math. Sci.* Berlin: Springer, 2004.

[Mar04] G. A. Margulis. *On some aspects of the theory of Anosov systems. With a survey by Richard Sharp: Periodic orbits of hyperbolic flows. Transl. from the Russian by S. V. Vladimirovna.* Springer Monogr. Math. Berlin: Springer, 2004.

[Mas82] Howard Masur. Interval exchange transformations and measured foliations. *Ann. Math. (2)*, 115:169–200, 1982.

[Mas22] Daniel Massart. A short introduction to translation surfaces, veech surfaces, and teichmüller dynamics. *Surveys in Geometry I*, page 343–388, 2022.

[Mir07a] Maryam Mirzakhani. Random hyperbolic surfaces and measured laminations. In *In the tradition of Ahlfors-Bers, IV. Proceedings of the 3rd Ahlfors-Bers colloquium, Ann Arbor, MI, USA, May 19–22, 2005*, pages 179–198. Providence, RI: American Mathematical Society (AMS), 2007.

[Mir07b] Maryam Mirzakhani. Simple geodesics and Weil–Petersson volumes of moduli spaces of bordered Riemann surfaces. *Invent. Math.*, 167(1):179–222, 2007.

[Mir07c] Maryam Mirzakhani. Weil–Petersson volumes and intersection theory on the moduli space of curves. *J. Am. Math. Soc.*, 20(1):1–23, 2007.

[Mir08a] Maryam Mirzakhani. Ergodic Theory of the Earthquake Flow. *International Mathematics Research Notices*, 2008, 01 2008. rnm116.

[Mir08b] Maryam Mirzakhani. Growth of the number of simple closed geodesies on hyperbolic surfaces. *Annals of Mathematics*, 168(1):97–125, 2008.

[Mir13] Maryam Mirzakhani. Growth of Weil–Petersson volumes and random hyperbolic surface of large genus. *J. Differ. Geom.*, 94(2):267–300, 2013.

[Mir16] Maryam Mirzakhani. Counting mapping class group orbits on hyperbolic surfaces, 2016.

[MNP22] Michael Magee, Frédéric Naud, and Doron Puder. A random cover of a compact hyperbolic surface has relative spectral gap $\frac{3}{16}-\varepsilon$. *Geometric and Functional Analysis*, May 2022.

[Mon20] Laura Monk. Benjamini–Schramm convergence and spectrum of random hyperbolic surfaces of high genus, 2020.

[Mon21] Laura Monk. *Geometry and Spectrum of typical hyperbolic surfaces.* PhD thesis, 2021.

[MP19] Mariam Mirzakhani and Bram Petri. Lengths of closed geodesics on random surfaces of large genus. *Comment. Math. Helv.*, 94(4):869–889, 2019.

[MP20] Michael Magee and Doron Puder. The asymptotic statistics of random covering surfaces, 2020.

[MR95a] Greg McShane and Igor Rivin. A norm on homology of surfaces and counting simple geodesics. *Int. Math. Res. Not.*, 1995(2):61–69, 1995.

[MR95b] Greg McShane and Igor Rivin. Simple curves on hyperbolic tori. *C. R. Acad. Sci., Paris, Sér. I*, 320(12):1523–1528, 1995.

[MS20] Etienne Le Masson and Tuomas Sahlsten. Quantum ergodicity for Eisenstein series on hyperbolic surfaces of large genus, 2020.

[MT19] Leonid Monin and Vanya Telpukhovskiy. On normalizations of Thurston measure on the space of measured laminations. *Topology Appl.*, 267:12, 2019. Id/No 106878.

[MW02] Yair Minsky and Barak Weiss. Nondivergence of horocyclic flows on moduli space. *J. Reine Angew. Math.*, 552:131–177, 2002.

[MZ08] Howard Masur and Anton Zorich. Multiple saddle connections on flat surfaces and the principal boundry of the moduli spaces of quadratic differentials. *Geom. Funct. Anal.*, 18(3):919–987, 2008.

[Nor13] Paul Norbury. String and dilaton equations for counting lattice points in the moduli space of curves. *Trans. Am. Math. Soc.*, 365(4):1687–1709, 2013.

[NWX20] Xin Nie, Yunhui Wu, and Yuhao Xue. Large genus asymptotics for lengths of separating closed geodesics on random surfaces, 2020.

[Ols11] Grigori Olshanski. Random permutations and related topics. In *The Oxford handbook of random matrix theory*, pages 510–533. Oxford: Oxford University Press, 2011.

[OP09] A. Okounkov and R. Pandharipande. Gromov–Witten theory, Hurwitz numbers, and matrix models. In *Algebraic geometry, Seattle 2005. Proceedings of the 2005 Summer Research Institute, Seattle, WA, USA, July 25–August 12, 2005*, pages 325–414. Providence, RI: American Mathematical Society (AMS), 2009.

[Pet17] Bram Petri. Random regular graphs and the systole of a random surface. *J. Topol.*, 10(1):211–267, 2017.

[PH92] R. C. Penner and J. L. Harer. *Combinatorics of train tracks*, volume 125 of *Ann. Math. Stud.* Princeton, NJ: Princeton University Press, 1992.

[PT07] Athanase Papadopoulos and Guillaume Théret. On Teichmüller's metric and Thurston's asymmetric metric on Teichmüller space. In *Handbook of Teichmüller theory. Volume I*, pages 111–204. Zürich: European Mathematical Society (EMS), 2007.

[PWX21] Hugo Parlier, Yunhui Wu, and Yuhao Xue. The simple separating systole for hyperbolic surfaces of large genus. *Journal of the Institute of Mathematics of Jussieu*, pages 1–10, May 2021.

[PY97] Jim Pitman and Marc Yor. The two-parameter Poisson–Dirichlet distribution derived from a stable subordinator. *Ann. Probab.*, 25(2):855–900, 1997.

[Riv01] Igor Rivin. Simple curves on surfaces. *Geom. Dedicata*, 87(1-3):345–360, 2001.

[RS19] Kasra Rafi and Juan Souto. Geodesic currents and counting problems. *Geom. Funct. Anal.*, 29(3):871–889, 2019.

[Rud22] Zeév Rudnick. GOE statistics on the moduli space of surfaces of large genus, 2022.

[Sau18] Adrien Sauvaget. Volumes and Siegel–Veech constants of $\mathcal{H}(2g-2)$ and Hodge integrals. *Geom. Funct. Anal.*, 28(6):1756–1779, 2018.

[Sau21] Adrien Sauvaget. The large genus asymptotic expansion of Masur–Veech volumes. *Int. Math. Res. Not.*, 2021(20):15894–15910, 2021.

[Shr22] Sunrose Shrestha. The topology and geometry of random square-tiled surfaces. *Geom. Dedicata*, 216(4):31, 2022. Id/No 38.

[SL66] L. A. Shepp and S. P. Lloyd. Ordered cycle lengths in a random permutation. *Trans. Am. Math. Soc.*, 121:340–357, 1966.

[SW22] Yang Shen and Yunhui Wu. The Cheeger constants of random belyi surfaces, 2022.

[Thu02] William Thurston. *Geometry and topology of three-manifolds.* 2002.

[Vak08] R. Vakil. The moduli space of curves and Gromov–Witten theory. In *Enumerative invariants in algebraic geometry and string theory. Lectures given at the C.I.M.E. summer school, Cetraro, Italy, June 6–11, 2005*, pages 143–198. Berlin: Springer, 2008.

[Vee82] William A. Veech. Gauss measures for transformations on the space of interval exchange maps. *Annals of Mathematics*, 115(2):201–242, 1982.

[VS77] A. M. Vershik and A. A. Shmidt. Limit measures arising in the asymptotic theory of symmetric groups. I. *Theory Probab. Appl.*, 22:70–85, 1977.

[Wit91] Edward Witten. Two-dimensional gravity and intersection theory on moduli space. In *Surveys in differential geometry. Vol. I: Proceedings of the conference on geometry and topology, held at Harvard University, Cambridge, MA, USA, April 27-29, 1990*, pages 243–310. Providence, RI: American Mathematical Society; Bethlehem, PA: Lehigh University, 1991.

[Wol85] Scott Wolpert. On the Weil–Petersson geometry of the moduli space of curves, 1985.

[Wol13] Scott A. Wolpert. Mirzakhani's volume recursion and approach for the Witten–Kontsevich theorem on moduli tautological intersection numbers. In *Moduli spaces of Riemann surfaces*, pages 221–266. Providence, RI: American Mathematical Society (AMS); Princeton, NJ: Institute for Advanced Study (IAS), 2013.

[Wri15] Alex Wright. Translation surfaces and their orbit closures: an introduction for a broad audience. *EMS Surv. Math. Sci.*, 2(1):63–108, 2015.

[Wri18] Alex Wright. Mirzakhani's work on earthquake flow, 2018.

[Wri20] Alex Wright. A tour through Mirzakhani's work on moduli spaces of Riemann surfaces. *Bull. Am. Math. Soc., New Ser.*, 57(3):359–408, 2020.

[WX22] Yunhui Wu and Yuhao Xue. Random hyperbolic surfaces of large genus have first eigenvalues greater than $\frac{3}{16} - \epsilon$. *Geom. Funct. Anal.*, 32(2):340–410, 2022.

[Yak22] Ivan Yakovlev. Contribution of n-cylinder square-tiled surfaces to masur-veech volume of $\mathcal{H}(2g-2)$, 2022.

[Zog19] P. G. Zograf. An explicit formula for Witten's 2-correlators. *J. Math. Sci., New York*, 240(5):535–538, 2019.

[Zor02] Anton Zorich. Square tiled surfaces and Teichmüller volumes of the moduli spaces of Abelian differentials. In *Rigidity in dynamics and geometry. Contributions from the programme Ergodic theory, geometric rigidity and number theory, Isaac Newton Institute for the Mathematical Sciences, Cambridge, UK, January 5–July 7, 2000*, pages 459–471. Berlin: Springer, 2002.

[Zor06] Anton Zorich. Flat surfaces. In *Frontiers in number theory, physics, and geometry I. On random matrices, zeta functions, and dynamical systems. Papers from the*

meeting, Les Houches, France, March 9–21, 2003, pages 437–583. Berlin: Springer, 2nd printing edition, 2006.

[Zvo12] Dimitri Zvonkine. An introduction to moduli spaces of curves and their intersection theory. In *Handbook of Teichmüller theory. Volume III*, pages 667–716. Zürich: European Mathematical Society (EMS), 2012.

[Zvo19] Dimitri Zvonkine. Translation surfaces and moduli spaces. *Lond. Math. Soc., Newsl.*, 2019(480):23–27, 2019.