HAL open science

Random multi-geodesics on hyperbolic surfaces

Mingkun Liu

▶ To cite this version:

Mingkun Liu. Random multi-geodesics on hyperbolic surfaces. Geometric Topology [math.GT]. Université Paris Cité, 2022. English. NNT : 2022UNIP7320 . tel-04521297

HAL Id: tel-04521297

https://theses.hal.science/tel-04521297v1

Submitted on 26 Mar 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université Paris Cité

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

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 ansymptotique 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-géodesiques aléatoire.

Mots-clés

multi-géodésiques aléatoires sur les 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

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

Contents

1 Introduction **13**

1.1 Random hyperbolic multi-geodesics 13

1.2 Random square-tiled surfaces 17

1.3 Random permutations 19

1.4 Large genus asymptotic random geometry 21

1.5 Main results 23

2 Introduction (en français) **27**

2.1 Multi-géodésiques hyperboliques aléatoires 27

2.2 Surface à petits carreaux aléatoires 31

2.3 Permutations aléatoires 33

2.4 Géométrie aléatoire asymptotique en grand genre 35

2.5 Résultats principaux 38

3 Background **43**

3.1 Hyperbolic geometry of surfaces 43

3.1.1 Deformation spaces 44

3.1.2 Curves 44

3.1.3 Fenchel–Nielsen coordinates 45

3.1.4 Weil–Petersson volumes 45

3.2 Flat geometry of surfaces 46

3.2.1 (Half-)Translation surfaces 46

3.2.2 Period coordinates . . . 47
3.2.3 Masur–Veech volumes . . . 48
3.2.4 Stable graphs . . . 49
3.2.5 Graph polynomials . . . 50
3.3 Intersection numbers . . . 51
3.4 Measured laminations . . . 54
3.5 Thurston distance . . . 55
3.6 Earthquake . . . 55
3.7 GEM distribution and Poisson–Dirichlet distribution . . . 56
3.8 Analytic combinatorics . . . 57

4 Random multi-geodesics 61
4.1 Random multi-geodesics model . . . 61
4.2 Masur–Veech volumes . . . 63
4.3 Large genus asymptotics for intersection numbers . . . 65
4.4 Topology of large genus random multi-geodesics . . . 65

5 Length partition of random multi-geodesics 67
5.1 Mirzakhani's covering spaces . . . 68
5.2 Horospheres . . . 70
5.2.1 Horospherical measures . . . 71
5.2.2 Total mass . . . 71
5.2.3 Horospherical measures on the unit sphere bundle . . . 72
5.3 Equidistribution . . . 73
5.3.1 Escape to infinity? . . . 74
5.3.2 Absolute continuity . . . 76
5.4 Counting . . . 77
5.5 Statistics . . . 80

6 Large genus asymptotics 83
6.1 Introduction . . . 83

6.1.1 Main result 83
6.1.2 Square-tiled surface reformulation 84
6.1.3 Organization of the chapter 85
6.2 Size-biased sampling and GEM distribution 85
6.3 Length partition of a random multicurve 89
6.4 Density of the length distribution 91
6.5 Reduction in the asymptotic regime 93
6.6 Proof of the main theorem 94
6.6.1 Method of moments 95
6.6.2 Asymptotic expression of a related sum 99
6.6.3 Truncation error estimate 102
6.6.4 Proof of Theorem 151 105

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, especially when the genus of the surface is 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 *component* of γ ($k \leq 3g-3$ on a closed surface of genus g). 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 if this thesis is the following result that Mirzakhani proved in her thesis:

Theorem 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*

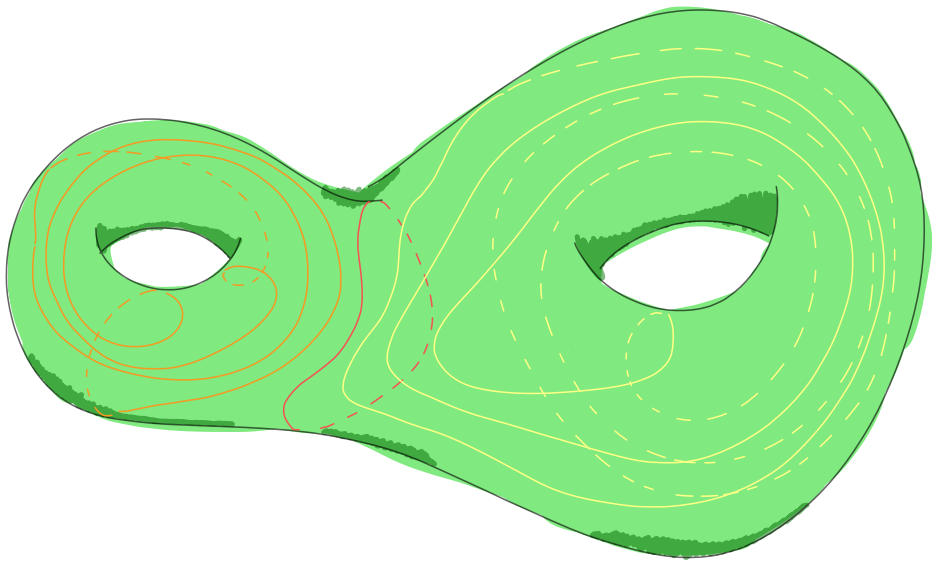

Figure 1.1: A multi-geodesic

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

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

as $R \to \infty$, where $c(\gamma)$ is a positive rational number, $B(X)$ denotes the Thurston volume of the unit ball (with respect to the length function) in the space of geodesic laminations on X, and b_g is the integral of the function B over the moduli space $\mathcal{M}_g$ of closed hyperbolic surfaces of genus g with respect to the Weil–Petersson measure, and is equal to the sum of all $c(\gamma)$, where $[\gamma]$ runs over all topological types of multicurves on Σ_g.

Remark 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], [Mir16] (filling curves), [ES16], [ES19] (non-simple multicurves), [ES20] (orbifolds), [EMM21], [AH21a] (effective counting), [EPS20] (general metric and length), [RS19] (geodesic current). On this topic, we strongly recommend the expository survey [AH22] and the forthcoming book [ES22].

Theorem 1 provides a model of *random* multi-geodesic.

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.

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. Endow $s_X(R)$ with the uniform probability measure to make it a probability space. Now, every question that we can ask to a particular multi-geodesic can be addressed to a random multi-geodesic, and the answer that one obtains is no longer deterministic, but "quantum". For example, the number of (connected) components of a random multi-geodesic (of length at most R) on a hyperbolic surface X of genus g varies between 1 and $3g-3$, and each number can be obtained with certain probability. Let $\mathrm{K}_{X,R}$ be the random variable associating a multi-geodesic to the number of its (connected) components.

Question 3. What is the limiting distribution of $\mathrm{K}_{X,R}$ as $R \to \infty$?

Another natural question to ask is

Question 4. What is the probability that a random multi-geodesic is separating? (We say that a multi-geodesic $m_1\gamma_1 + \cdots + m_k\gamma_k$ is non-separating if $X \smallsetminus (\gamma_1 \cup \cdots \cup \gamma_k)$ is connected.)

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

Question 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 5: the probability that a random multi-geodesic is of the topological type $[\gamma]$ is $c(\gamma)/b_g$, 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 questions above boil down to the "frequencies" $c(\gamma)$. For example, by calculating these numbers, Mirzakhani managed to prove in the same paper

Proposition 6. *On a closed hyperbolic surface of genus* 2*, the probability that a random primitive multi-geodesic is separating equals to* 1/49*.*

Question 3 is also partially answered: the number of components $\mathrm{K}_{X,R}$ converges to a random variable K_g depending only on g, and then the following question naturally arises:

Question 7. What is the limiting distribution of K_g as $g \to \infty$? And in particular, what is the average number of components of a random multi-geodesics on a large genus hyperbolic surface?

One may also wish to understand the *geometry* of a random multi-geodesic, for example, its *length partition*. The decomposition of a multi-geodesic $\gamma = m_1\gamma_1 + \cdots + m_k\gamma_k$ into connected components $m_i\gamma_i$, $1 \le i \le k$, 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 $\Delta_1^\infty := \{(x_1, x_2, \dots) \in [0,1]^{\mathbb{Z}_{\ge 1}} : x_1 + x_2 + \cdots = 1\}$,

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

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 $1, \dots, k$.

Now let us sample a random hyperbolic surface $X \in \mathcal{M}_g$, with respect to the Weil–Petersson probability measure on $\mathcal{M}_g$ for example, then pick a random multi-geodesic on X. By considering its length partition, we obtain a random variable $\hat{\ell}_g^{\downarrow}$.

Question 8. 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 $\Delta_1^{k-1} := \{(x_1, \dots, x_k) \in \mathbb{R}_{\ge 0}^k : x_1 + \cdots + x_k = 1\}$ 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 9. 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

Theorem 10 ([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 $1, \dots, 1$, 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^{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^{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 8, the procedure of choosing a random surface is redundant since the every hyperbolic surface has exactly the same statistics.

Remark 11. 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], 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], [DGZZ20c]. And in particular from spectral aspect; see e.g. [GLMST21], [Mon20], [MS20], [WX22], [LW21], [Hid21], [Rud22].

1.2 Random square-tiled surfaces

A *square-tiled* surface (or an *origami*) is a surface obtained by gluing squares. Here, every square is isometric to the standard Euclidean square $[0,1]^2 \subset \mathbb{R}^2$; the four sides are marked by "east", "south", "west", and "north" in anticlockwise; the two parallel sides "east" and "west" are said to be *vertical*, and the two sides "south" and "north" are said to be *horizontal*. See Figure 1.2. In gluing procedure, we assume that (so that the resulting square-tiled surface lies

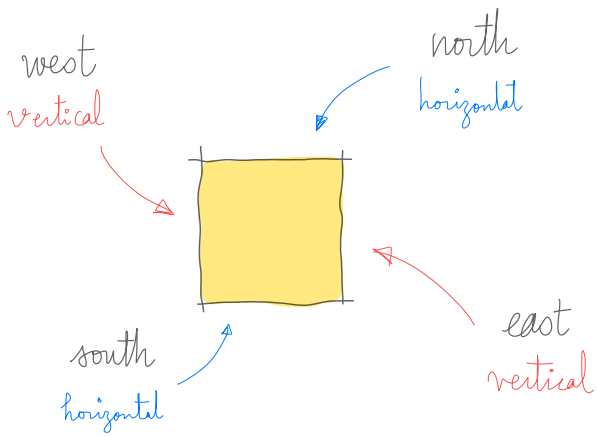

Figure 1.2: A square

in the principle stratum of holomorphic quadratic differentials)

1. the resulting surface is connected and orientable;
2. horizontal (resp. vertical) sides are glued together two-by-two;
3. when we turn around any point on the resulting surface (in the same direction), we end up back at the starting point after an angular displacement 2π or 3π (and nothing else).

Those points around which one has to travel 3π to get back are called *singularities* of the square-tiled surface. Figure 1.3 (and Figure 1.4) is an example, which is a square-tiled surface in $\mathcal{Q}(1^4)$ of genus 2 with 4 singularities and 3 cylinders. See e.g., [DGZZ21, Section 1.2] for a more detailed discussion.

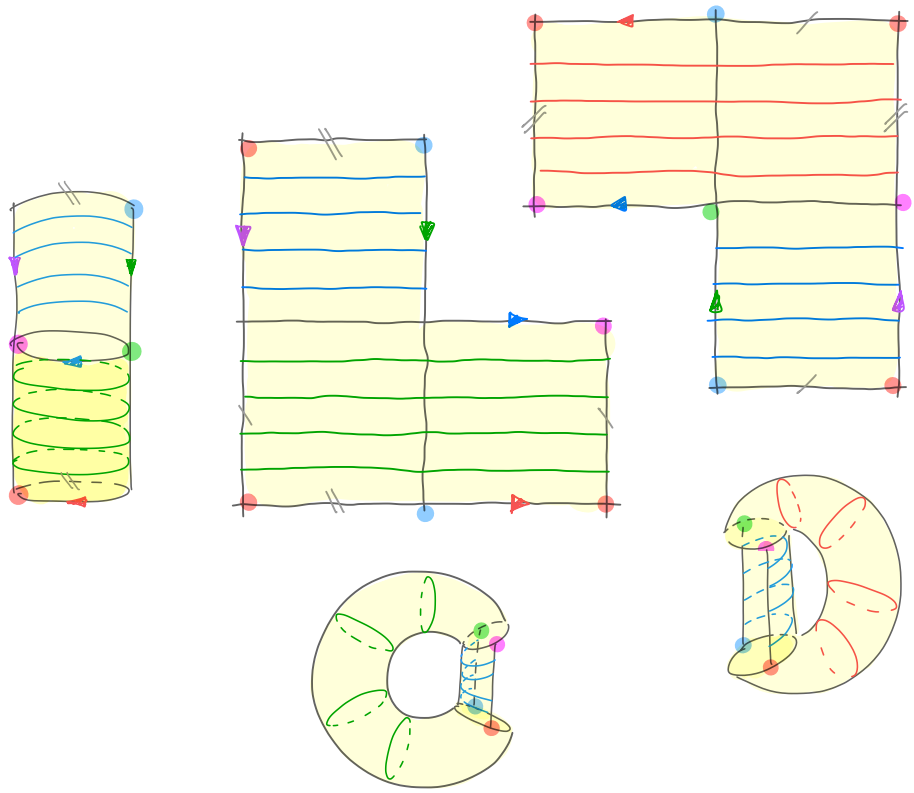

Figure 1.3: A square-tiled surface

Remark 12. Why square-tiled surfaces? There are many reasons but a particularly important one is that they correspond to integral points of the principal stratum of the moduli space of holomorphic quadratic differentials $\mathcal{Q}(1^{4g-4})$; see Section 3.2 for more details. (The definition of a square-tiled surface can be given in a more general way.)

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 *weighted stable graph* associated to a square-tiled surface is a graph that has a vertex for each singular leaf of the horizontal foliation, and for each horizontal cylinder has an edge that connects the two vertices (not necessarily distinct) corresponding to the two singular leaves that bound this cylinder; each vertex is decorated by the *ribbon graph* obtained by considering a tubular neighborhood of its corresponding singular leaf, and each edge is weighted by the height of its corresponding cylinder. See Figure 1.5 for the stable graph (on the left) of the square-tiled surface given by

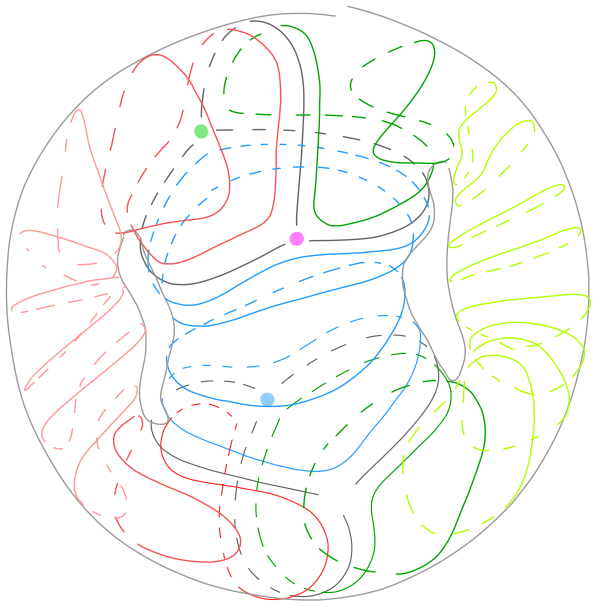

Figure 1.4: The square-tiled surface in Figure 1.3

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 3.2.4 for a precise definition of a weighted stable graph. We say that a square-tiled surface has type Γ if its associated weighted stable graph is Γ.

There is only a finite number of square-tiled surfaces of genus g tiled with N squares. Denote the set of all such surfaces by $\mathcal{ST}_g(N)$, and equip it with the uniform probability measure.

Question 13. What is the probability that a random square-tiled surface of genus g has k horizontal cylinders?

Question 14. What is the probability that a random square-tiled has a fixed given type?

The horizontal cylinder decomposition of a square-tiled surface enable us to write its total (flat) area N as the sum of the areas of each cylinder. Let us denote by $\hat{\mathrm{a}}^{\downarrow}_{g,N} \colon \mathcal{ST}_g(N) \to \Delta_1^\infty$ the random variable that associates a square-tiled surface to the vector whose i-th entry is the area of its i-th largest horizontal cylinder, divided by the total area N.

Question 15. What is the limiting distribution of $\hat{\mathrm{a}}^{\downarrow}_{g,N}$ as $N \to \infty$? And then as $g \to \infty$?

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 is to write down the map in the following form

$$\tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 1 & 7 & 2 & 4 & 6 & 5 & 3 \end{pmatrix} \tag{1.2}$$

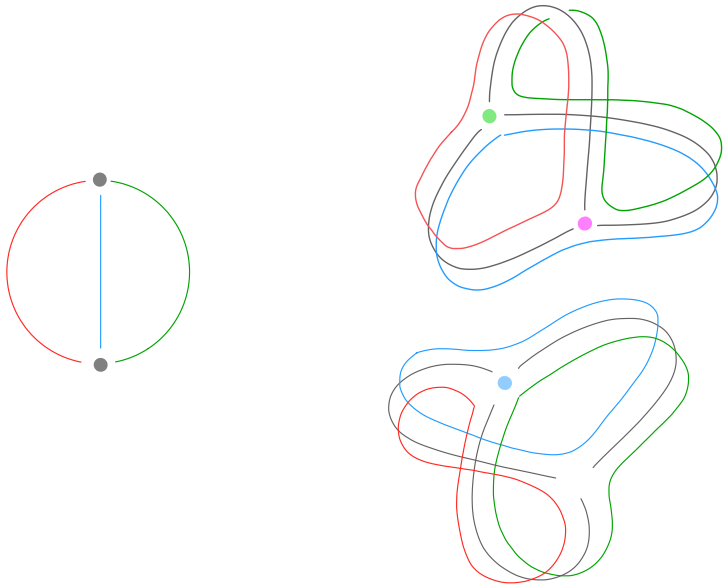

Figure 1.5: Associated stable graph and ribbon graphs

where each element in the first line is mapped to the element just below it; for example, $\tau(1) = 1$, $\tau(2) = 7$, etc. A permutation can also be written as a product of disjoint cycles. For example,

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

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

Given a permutation σ of n letters. We write $\mathrm{K}_n(\sigma)$ for the number of cycles in the cycle decomposition of σ, and denote by $\hat{\mathrm{c}}^{\downarrow}_{\theta,n}(\sigma)$ the n-vector whose i-th component is the length of the i-th longest cycle in σ divided by n. For the permutation defined by (1.2) or (1.3), we have $\mathrm{K}_n(\tau) = 4$ and $\hat{\mathrm{c}}^{\downarrow}_{\theta,n}(\tau) = (3/7, 2/7, 1/7, 1/7)$.

Write S_n for the set of all permutation of $\{1, \ldots, n\}$. The simplest (and the most fundamental) measure on S_n is no doubt the uniform probability measure, with respect to which $\mathbb{P}(\sigma) = 1/n!$ for all $\sigma \in S_n$. In this thesis, we shall consider a more general probability measure $\mathbb{P}_\theta$, where $\theta > 0$, on S_n introduced by Ewens in the context of population genetics [Ewe72] and defined by

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

Note that when $\theta = 1$, $\mathbb{P}_{\theta,n}$ is none other than the uniform probability on S_n, so $(\mathbb{P}_{\theta,n})_{\theta \in \mathbb{R}_{>0}}$ is a one-parameter deformation of the uniform probability measure on S_n. This model of random permutations has been widely studied. For a detailed treatment, we refer the reader to [ABT03, Example 2.19, Chapter 4] and references therein.

Remark 16. Consider the projection $\pi_n \colon S_n \to S_{n-1}$ that erases the last letter "n" in the cycle

decomposition. For example for the example defined by (1.3), $\pi_7(\tau) = (1)(23)(4)(56)$. This projection is the unique map that commutes with the two-side action of S_{n-1} for $n \geq 5$. The Ewens probability measure $\mathbb{P}_{\theta,n}$ enjoys the property that the push-forward measure of $\mathbb{P}_{\theta,n}$ by π_n coincides with $\mathbb{P}_{\theta,n-1}$. See [Ols11] for more details.

As the reader may have already guessed: we want to know

Question 17. What is the probability that a random permutation has k cycles?

Question 18. What is the distribution of $\hat{\mathrm{c}}^{\downarrow}_{\theta,n}$ as $n \to \infty$?

Remark 19. 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 $\mathrm{K}_{n,i}(\sigma)$ denotes the number of cycles of length i in $\sigma \in S_n$. Define

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

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 Question 17 and Question 18 since the Riemann zeta function $\zeta(x)$ goes to 1 very rapidly as $x \to +\infty$.

1.4 Large genus 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 $6g - 6$ letters,

are similar as $g \to \infty$.

Square-tiled surface defined in the beginning of Section 1.2 can be seen as integral points in the moduli space $\mathcal{Q}_g$ of *holomorphic quadratic differentials* of genus g (more precisely, its principal stratum), and the *Masur–Veech volume* of this space $\mathrm{vol}_{\mathrm{MV}}(\mathcal{Q}_g)$ can be obtained by counting square-tiled surfaces. See, e.g., [DGZZ16, Appendix A] for a detailed general discussion, [DGZZ20b] for a discussion in genus zero, and [Zor02] for a detailed discussion for Abelian differentials.

Remark 20. 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], etc. The approach for the calculation of Masur–Veech volumes (Abelian and quadratic) by counting square-tiled surfaces was proposed independently by Eskin–Masur, and Kontsevich–Zorich [Zor02], and these volumes has been intensively extensively studied in recent

years from many 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.

In the paper [Mir08a] where Mirzakhani constructs a conjugacy between the Teichmüller horocycle flow over $\mathcal{Q}_g$ and the earthquake flow on $\mathcal{M}_g$, and establishes the ergodicity of the latter, as a byproduct, she finds the following surprising identity:

$$b_g = C_g \cdot \mathrm{vol}_{\mathrm{MV}}(\mathcal{Q}_g) \tag{1.4}$$

for all $g \geq 2$, where b_g is same constant that appears in Theorem 1, and C_g is some constant depending only on g, which was later determined independently by Arana-Herrera [AH20b], and Monin–Telpukhovskiy [MT19].

It is proved in [DGZZ21] that the Masur–Veech volume of $\mathcal{Q}_g$ can be written as a sum over contributions from all (weighted) stable graphs, each of which corresponds to a cylinder decomposition. Moreover, each of these contributions can be expressed explicitly in term of intersection numbers between so called *psi-classes* on the Deligne–Mumford compactification $\overline{\mathcal{M}}_{g,n}$ of the moduli space of smooth complex curves.

By comparing with the formula for frequencies of multicurves $c(\gamma)$ given by Mirzakhani [Mir08b, Theorem 5.3], the authors of [DGZZ21] found that the equation (1.4) holds even after specializing to any particular multicurve and its corresponding weighted stable graph. Therefore, the random geometry of multi-geodesics and that of square-tiled surfaces are closely related; and the key to the understand both of them is hidden in the intersection numbers of psi-classes.

Informally, 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. 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 proved the year after by Kontsevich [Kon92], and later several different proofs have been found in, e.g., [OP09], [KL07], [Mir07c], and [ABC+20]. However, 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 A. Aggarwal [Agg21].

Based on A. Aggarwal's deep work, Delecroix–Goujard–Zograf–Zorich proved a series of results in [DGZZ20c] about large genus random square-tiled surfaces.

Theorem 21 ([DGZZ20c]). *Informally speaking, 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* $\log(6g-6)/2$.

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

Theorem 22 ([DGZZ20c])**.** *Informally speaking, 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* $\log(6g-6)/2$.

On the other hand, it is also known that, roughly speaking, the distribution of the number of cycles of 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 $6g-6$ elements.

1.5 Main results

The first step is to prove the following generation of Theorem 10 to multi-geodesics of arbitrary topological type.

Theorem 23 ([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* (3.3).

Remark 24. The density function $\bar{F}_\gamma$ can be expressed explicitly in term 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 V. Erlandsson, H. Parlier, and J. Souto [EPS20] suggests that the same statistics probably holds even for a large class of metrics.

Remark 25. 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 the knowledge of its topological type), and study its large genus asymptotics. In a joint work with Vincent Delecroix, we proved

Theorem 26 ([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 27. The same convergence still holds for random primitive multi-geodesics.

The random square-tiled surfaces counterpart of the preceding theorem is

Theorem 28 ([DL22])**.** *The random variable $\hat{\mathrm{a}}^{\downarrow}_{g,N}$ 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 26 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 29. It is known that the random variable $\hat{\mathrm{c}}^{\downarrow}_{\theta,n}$ representing the length partition 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 30. Another interesting example of a Poisson–Dirichlet distribution arises in integer factorization. Every integer $k \geq 2$ can be uniquely decomposed into a product of prime numbers. 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 a random integer between 2 and n picked uniform at random, and define the random variable by

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

Billingsley proved in [Bil72] that this random variable converges in distribution to the Poisson–Dirichlet law of parameter 1.

Remark 31. 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 $\mathbb{R}_{>0}$)

$$\frac{e^{-x}}{x} \sum_{n=1}^{\infty} (\cosh(x/n) - 1) = \frac{e^{-x}}{x} \sum_{n=1}^{\infty} \frac{\zeta(2n)}{(2n)!} x^{2n}. \tag{1.5}$$

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

$$\int_a^b \frac{e^{-x}}{x} \sum_{n=1}^{\infty} (\cosh(x/n) - 1) \tag{1.6}$$

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 (1.6). 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 (1.5); and the presence of the higher-order terms is due to the fact that our multi-geodesics are not necessarily primitive.

Remark 32. 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 $\hat{\ell}_{X,\gamma,R}$ as $R \to \infty$, 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 33. 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 [Yak] 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.

We obtain therefore a picture of a typical multi-geodesic on a large genus hyperbolic surface: it is non-separating, has about $\log(g)/2$ components, and we can barely see other

components than the three largest, which altogether take about 98% of the total length of the multi-geodesic.

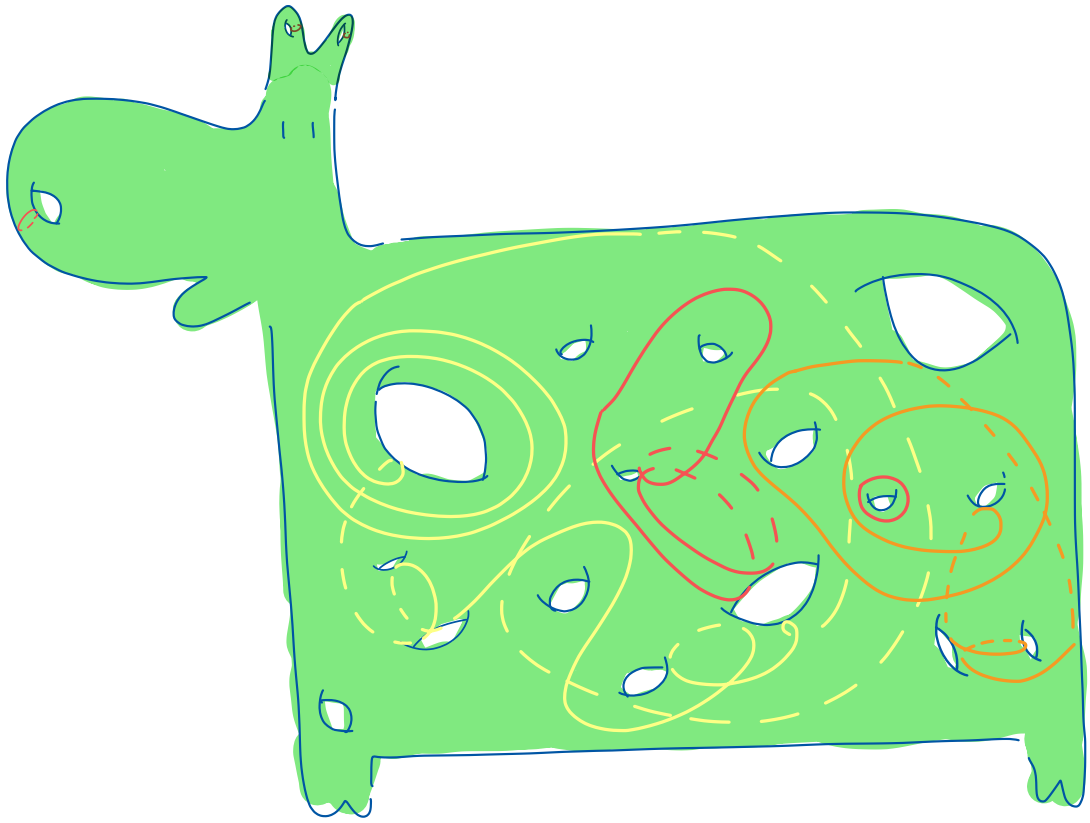

Figure 1.6: A typical multi-geodesic on a large genus hyperbolic surface

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 ne s'intersectent pas, 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 que 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

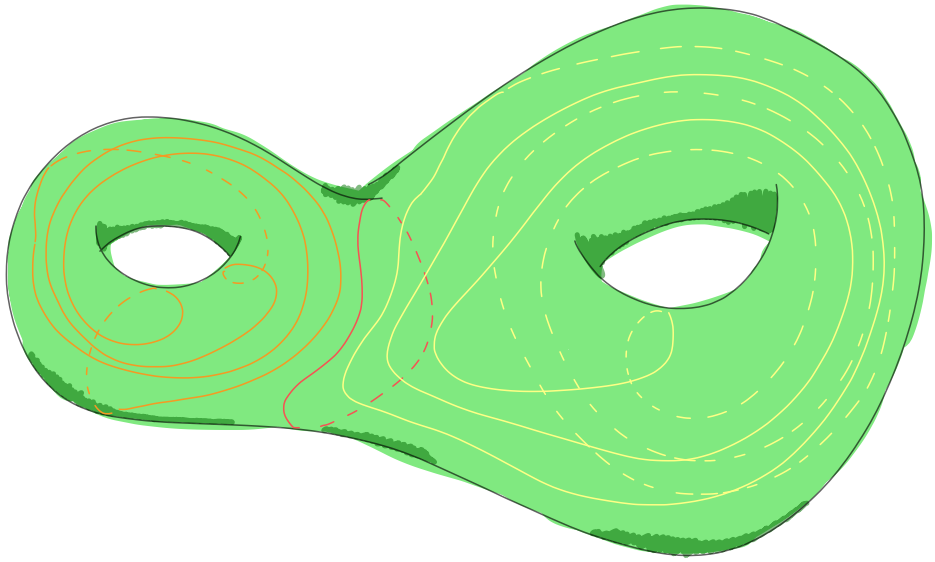

Figure 2.1: Une multi-géodésique

Théorème 34 ([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 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 $c(\gamma)$ est un nombre rationnel positif, $B(X)$ signifie le volume de Thurston de la boule unité (par rapport à la fonction de longueur) dans l'espace des laminations géodésiques sur X, et b_g est l'intégrale de la fonction B sur l'espace de modules $\mathcal{M}_g$ des surfaces hyperboliques closes de genre g par rapport à la mesure de Weil–Petersson, et est égale à la somme de tous les $c(\gamma)$, où $[\gamma]$ parcourt l'ensemble des types topologiques des multi-courbes sur Σ_g.

Remarque 35. 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 toujours un champs de recherche très actif. Pour des développements plus récents, voir e.g. [AH20a], [Mir16] (courbes remplissantes), [ES16], [ES19] (multi-courbes non simples), [ES20] (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 à paraître [ES22].

Théorème 34 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.

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 $s_X(R)$ de la probabilité uniforme, qui en fait un espace probabilisé. Maintenant, n'importe quelle question que l'on pose à une multi-géodésique particulière peut également s'adresser à une multi-géodésique aléatoire, et la réponse que l'on attend n'est plus déterministe, mais « quantique ». Par exemple, le nombre de composantes (connexes) d'une multi-géodésique aléatoire sur une surface hyperbolique de genre g varie entre 1 et $3g-3$, et chaque nombre peut être obtenu avec une certaine probabilité. Soit $\mathrm{K}_{X,R}$ la variable aléatoire qui associe une multi-géodésique au nombre de ses composantes (connexes).

Question 36. Quelle est la loi limite de $\mathrm{K}_{X,R}$ lorsque $R \to \infty$?

Une autre question que l'on se pose naturellement est la suivante:

Question 37. Quelle est la probabilité qu'une multi-géodésique aléatoire soit séparante (on dit qu'une multi-géodésique $m_1\gamma_1 + \cdots + m_k\gamma_k$ est non séparante si $X \smallsetminus (\gamma_1 \cup \cdots \cup \gamma_k)$ est connexe) ?

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

Question 38. Quelle est la probabilité que 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 38 : 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, en calculant ces nombres, Mirzakhani a pu démontrer dans le même papier

Proposition 39. *Sur une surface hyperbolique de genre* 2*, la probabilité qu'une multi-géodésique primitive aléatoire soit séparante est égale à* 1/49.

Question 36 reçoit également une réponse partielle : le nombre de composantes $\mathrm{K}_{X,R}$ converge vers une variable aléatoire K_g qui ne dépend que de g, et la question suivante se pose alors naturellement:

Question 40. Quelle est la loi limite de K_g quand $g \to \infty$? Et en particulier, quel est la moyenne du nombre de composantes d'une multi-géodésique aléatoire sur une surface hyperbolique aléatoire de grand genre ?

On souhaiterait également comprendre la *géométrie* d'une multi-géodésique aléatoire, par exemple, sa *partition des longueurs*. 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 $1, \dots, k$.

Maintenant, échantillonnons une surface hyperbolique aléatoire $X \in \mathcal{M}_g$, par rapport à la mesure de probabilité de Weil–Petersson sur $\mathcal{M}_g$ par exemple, puis choisissons une multi-géodésique aléatoire sur X. En considérant sa partition des longueurs, on obtient une variable aléatoire $\hat{\ell}_g^{\downarrow}$.

Question 41. 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 42. 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 démontre

Théorème 43 ([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 $1, \dots, 1$, 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 41, la procédure de choix d'une surface aléatoire est redondante car toutes les surfaces hyperboliques ont exactement les même statistiques.

Remarque 44. 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 ruban*, *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], 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], [DGZZ20c]. Et surtout d'un point de vue spectral ; voir e.g. [GLMST21], [Mon20], [MS20], [WX22], [LW21], [Hid21], [Rud22].

2.2 Surface à petits carreaux aléatoires

Une *surfaces à petits carreaux* (ou un *origami*) est une surface obtenue en collant les carrés. Ici, chaque carré est isométrique au carré euclidien standard $[0,1]^2 \subset \mathbb{R}^2$, les quatre côtés sont marqués par « esst», « south », « west », et « north » dans le sens inverse des aiguilles d;une montre ; les deux côtés parallèles « east » and « west » sont dits « verticaux » et les deux côtés « south » and « north » sont dits « horizontaux » ; voir Figure 2.2. Dans la procédure de collage, on suppose que (de sorte que la surface à petit carreaux obtenue se trouve dans la strate principale des différentielles quadratiques holomorphes)

1. la surface résultante est connexe et orientable;

2. les arêtes horizontales (resp. verticales) sont collés ensemble deux-par-deux ;

3. quand on tourne (dans la même direction) autour de n'important quel point de la surface obtenue, on revient au point de départ après un déplacement angulaire de 2π ou de 3π (et rien d'autre).

Les points autour desquels il faut parcourir 3π pour revenir au point de départ sont appelés les *singularités*. Figure 2.3 (and 2.4) est un exemple, qui est une surfaces à petits carreaux dans $\mathcal{Q}(1^4)$ de genre 2 avec 4 singularités et 3 cylindres. Voir e.g., [DGZZ21, Section 1.2] pour une discussion plus en détail.

Remarque 45. Pourquoi des surfaces à petits carreaux ? Il y a de nombreuses raisons mais l'une d'entre elles, particulièrement importante, est qu'elles correspondent à des points d'entier de la strate principale de l'espace de modules des différentielles quadratiques holomorphes $\mathcal{Q}(1^{4g-4})$; voir Section 3.2 pour plus de détails. (La définition d'une surface à carreaux peut être donnée de manière plus générale.)

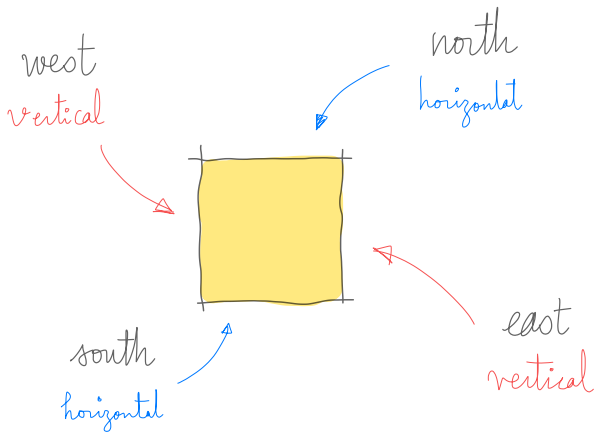

Figure 2.2: Un carré

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. Le *graphe stable pondéré* associé à une surfaces à petits carreaux est un graphe qui possède un sommet pour chaque feuille singulière de la foliation horizontale, et pour chaque cylindre horizontal, une arête qui relie les deux sommets (pas nécessairement distincts) correspondant aux deux feuilles singulières qui délimitent ce cylindre ; chaque somment est décoré par le *graphe en ruban* obtenu en considérant un voisinage tubulaire de sa feuille singulière correspondant. 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 ruban (à droite) qui correspondent aux deux sommets du graphe stable. Voir Section 3.2.4 pour une définition précise d'un graphe stable pondéré. On dit que une surfaces à petits carreaux est de type Γ, où Γ est un graphe stable pondéré, si son graphe stable pondéré est Γ.

Il n'y a qu'un nombre fini de surfaces à petits carreaux de genre g carrelées de N carrés. On note $\mathcal{ST}_g(N)$ l'ensemble de toutes ces surfaces, et le munit de la mesure de probabilité uniforme.

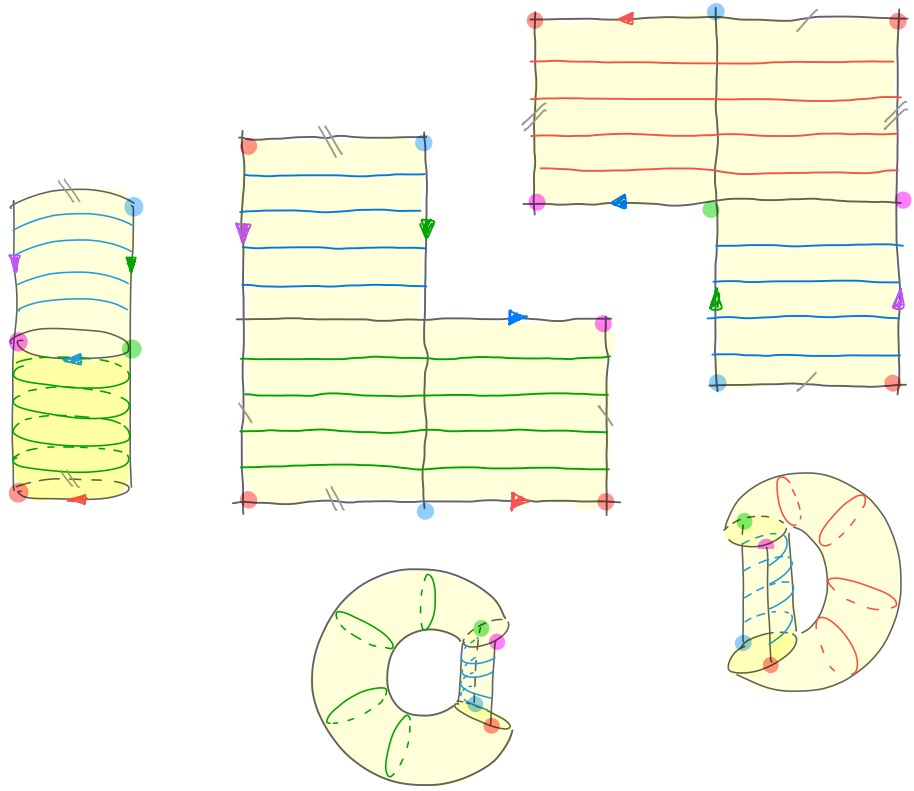

Figure 2.3: Une surfaces à petits carreaux dans $\mathcal{Q}(1^4)$ avec trois cylindres horizontaux

Question 46. Quelle est la probabilité qu'une surface à petits carreaux aléatoire de genre g ait k cylindres horizontaux ?

Question 47. Quelle est la probabilité qu'une surfaces à petits carreaux ait un type fixé donnée ?

La décomposition en cylindres horizontaux d'une surfaces à petits carreaux nous permet d'écrire son aire totale N comme la somme des aires de chaque cylindre. On désigne par $\hat{\mathrm{a}}^{\downarrow}_{g,N} \colon \mathcal{ST}_g(N) \to \Delta_1^\infty$ la variable aléatoire qui associe une surfaces à petits carreaux au vecteur dont la i-ième entrée est l'aire de son i-ième plus grand cylindre horizontal, divisée par l'aire totale N.

Question 48. What is the limiting distribution of $\hat{\mathrm{a}}^{\downarrow}_{g,N}$ as $N \to \infty$? And then as $g \to \infty$? Quelle est la loi limite de $\hat{\mathrm{a}}^{\downarrow}_{g,N}$ lorsque $N \to \infty$? Et puis quand $g \to \infty$?

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 sous la forme suivante

$$\tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 1 & 7 & 2 & 4 & 6 & 5 & 3 \end{pmatrix} \tag{2.2}$$

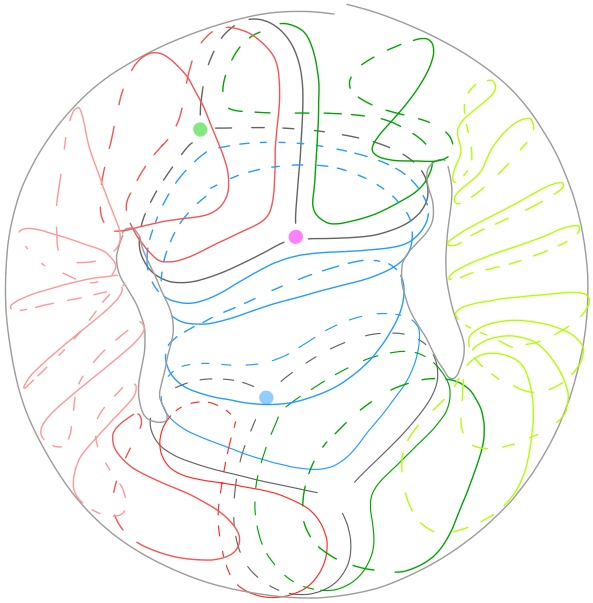

Figure 2.4: Surface obtenue

où chaque élément de la première ligne envoyé à l'élément juste en dessous ; par exemple $\tau(1) = 1$, $\tau(2) = 7$, etc. 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.

Étant donné une permutation σ de n lettres. On écrit $\mathrm{K}_n(\sigma)$ pour le nombre de cycles dans la décomposition en cycles de σ, et on désigne par $\hat{\mathrm{c}}^{\downarrow}_{\theta,n}(\sigma)$ le n-vecteur dont la i-ième composante est la longueur de la i-ième plus long cycle dans σ divisé par n. Pour la permutation définie par (1.2) ou (1.3), on a $\mathrm{K}_n(\tau) = 4$ et $\hat{\mathrm{c}}^{\downarrow}_{\theta,n}(\tau) = (3/7, 2/7, 1/7, 1/7)$.

On écrit S_n l'ensemble de toutes les permutations de $\{1, \dots, n\}$. La mesure la plus simple (et la plus fondamentale) sur S_n est sans doute la mesure de probabilité uniforme, pour laquelle $\mathbb{P}(\sigma) = 1/n!$ pour tout $\sigma \in S_n$. Dans cette thèse, on considéra une mesure plus générale $\mathbb{P}_\theta$, où $\theta > 0$, sur S_n introduite par Ewens dans le contexte de la génétique des populations [Ewe72] et définie par

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

Remarquons que lorsque $\theta = 1$, $\mathbb{P}_{\theta,n}$ n'est rien d'autre que la probabilité uniforme sur S_n, donc $(\mathbb{P}_{\theta,n})_{\theta \in \mathbb{R}_{>0}}$ est une déformation à un paramètre de la mesure de probabilité uniforme sur S_n. Ce modèle de permutations aléatoire a été largement étudié. Pour un traiement détaillé, Nous renvoyons le lecteur à [ABT03, Example 2.19, Chapter 4] et aux références qui s'y trouvent.

Remarque 49. Considérons la projection $\pi_n \colon S_n \to S_{n-1}$ qui efface la dernière lettre dans la décomposition en cycles. Par exemple, au case de (1.3), $\pi_7(\tau) = (1)(23)(4)(56)$. Cette projection

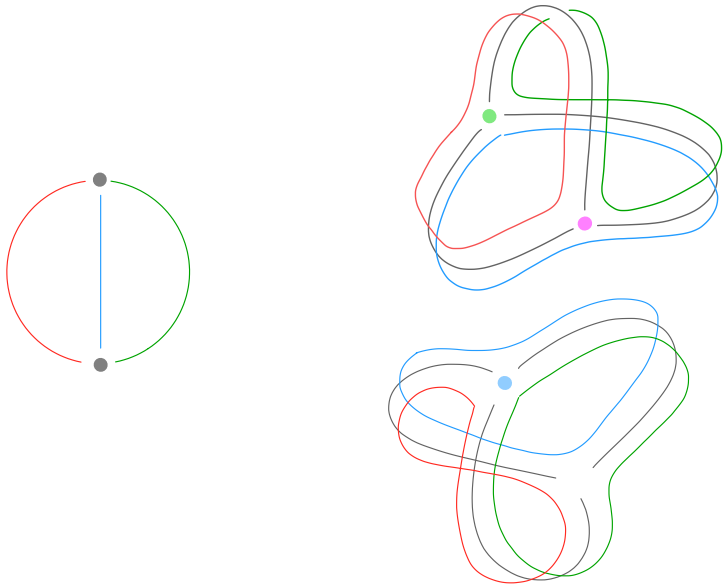

Figure 2.5: Graphe stable et graphes en ruban associés

est l'unique application qui commute avec action bilatérale de S_{n-1} pour $n \geq 5$. La mesure de probabilité d'Ewens $\mathbb{P}_{\theta,n}$ possède la propriété que la poussé-en-avant de $\mathbb{P}_{\theta,n}$ par π_n coïncide avec $\mathbb{P}_{\theta,n-1}$. Voir [Ols11] pour plus de détails.

Comme le lecteur l'a peut-être déjà deviné : nous voulons savoir

Question 50. Quelle est la probabilité qu'une permutation ait k cycles ?

Question 51. Quelle est la loi limite de $\hat{\mathrm{c}}^{\downarrow}_{\theta,n}$ lorsque $n \to \infty$?

Remarque 52. 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 $\mathrm{K}_{n,i}(\sigma)$ le nombre de cycles de longueur i dans $\sigma \in S_n$. On définit

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

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 Question 50 et Question 51 car la fonction zêta de Riemann $\zeta(x)$ converge vers 1 très rapidement lorsque $x \to +\infty$.

2.4 Géométrie aléatoire asymptotique en grand genre

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 $6g-6$ lettres,

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

Les surfaces à petits carreaux définis au début de Section 2.2 peuvent être vus comme des points d'entier dans l'espace de modules $\mathcal{Q}_g$ des *différentielles quadratiques holomorphes* de genre g (plus précisément, sa strate principale), et le *Masur–Veech volume* $\mathrm{vol}_{\mathrm{MV}}(\mathcal{Q}_g)$ de cet espace s'obtient en comptant les surfaces à petits carreaux. Voir, e.g., [DGZZ16, Appendix A] pour une discussion générale détaillée, [DGZZ20b] pour une discussion au cas de genre zéro, et [Zor02] discussion détaillée pour les différentielles abéliennes.

Remarque 53. 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.

Dans l'article [Mir08a] où Mirzakhani construit une conjugaison entre le flot horocyclique de Teichmüller sur $\mathcal{Q}_g$ et le flot de tremblement de terre sur $\mathcal{M}_g$, et établit l'ergodicité de ce dernier, comme un sous-produit, elle a trouvé l'identité surprenante suivante :

$$b_g = C_g \cdot \mathrm{vol}_{\mathrm{MV}}(\mathcal{Q}_g) \tag{2.4}$$

pour tout $g \geq 2$, où b_g est la même constante qui apparaît dans Théorème 34, et C_g est une constante qui dépend seulement de g, qui a été déterminée plus tard indépendamment par Arana-Herrera [AH20b], et Monin–Telpukhovskiy [MT19].

Il est prouvé dans [DGZZ21] que le volume de Masur–Veech de $\mathcal{Q}_g$ s'écrit comme la somme sur les contributions de tous les graphes stable (pondérés), dont chacune correspond à une décomposition en cylindres. De plus, chacune de ces contributions s'exprime explicitement en terme de nombres d'intersection entre les *classes psi* sur la compactification de Deligne–Mumford $\overline{\mathcal{M}}_{g,n}$ de l'espace de modules des courbes complexes lises. En comparant avec la formule pour les fréquences des multi-courbes $c(\gamma)$ donnée par Mirzakhani [Mir08b, Theorem 5.3], les auteurs de [DGZZ21] ont trouvé que l'équation (2.4) vaut même après s'être spécialisé sur une multi-courbe particulière et son graphe stable pondéré correspondant. Ainsi, la géométrie aléatoire des multi-géodésiques et celle des surfaces à petits carreaux sont étroitement liée ; et la clé pour les comprendre est cachée dans les nombres d'intersection des classes psi.

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. 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] prétend 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 par la suite plusieurs preuves différentes ont été trouvées dans, 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 [DGZZ20c] sur les surfaces à petits carreaux aléatoires de grand genre.

Théorème 54 ([DGZZ20c])**.** *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 $\log(6g-6)/2$.*

Le pont 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 55 ([DGZZ20c])**.** *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 $\log(6g-6)/2$.*

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 $6g-6$ éléments.

2.5 Résultats principaux

La première étape consiste à montrer la généralisation suivante de Théorème 43 aux multi-géodésiques de type topologique arbitraire.

Théorème 56 ([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 (3.3).

Remarque 57. La fonction de densité $\bar{F}_\gamma$ s'exprime explicitement en terme 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 V. Erlandsson, H. Parlier, et J. Souto [EPS20] suggèrent que la même statistique est probablement valable pour une plus grande classe de métriques.

Remarque 58. 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 59 ([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 60. La même convergence est valable pour les multi-géodésiques primitive aléatoires.

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

Théorème 61 ([DL22]). *La variable aléatoire $\hat{\mathrm{a}}^{\downarrow}_{g,N}$ 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 26 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 62. 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 63. Un autre exemple intéressant de loi de 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 définie par

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

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

Remarque 64. 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 $\mathbb{R}_{>0}$)

$$\frac{e^{-x}}{x} \sum_{n=1}^{\infty} (\cosh(x/n) - 1) = \frac{e^{-x}}{x} \sum_{n=1}^{\infty} \frac{\zeta(2n)}{(2n)!} x^{2n}. \tag{2.5}$$

En particulier, pour tout $a, b \in \mathbb{R}_{\geq 0}$ avec $a < b$, le nombre de composantes de longueur (normalisée) dans $[a/(6g-6), b/(6g-6)]$ converge vers une loi de Poisson de paramètres

$$\int_a^b \frac{e^{-x}}{x} \sum_{n=1}^{\infty} (\cosh(x/n) - 1) \tag{2.6}$$

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 (1.6). 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 (1.5) ; et la présence des termes d'ordre supérieur est due au fait que nos multi-géodésiques ne sont pas nécessairement primitives.

Remarque 65. 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 $\hat{\ell}_{X,\gamma,R}$ lorsque $R \to \infty$ 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 66. 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 [Yak] 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.

On obtient donc une image d'une multi-géodésique typique sur une surface hyperbolique de grand genre : elle n'est pas séparante, a environ $\log(g)/2$ composantes, et on peut à peine voir d'autres composantes que les trois plus grandes, qui prennent ensemble environ 98% de la longueur totale de la multi-géodésique.

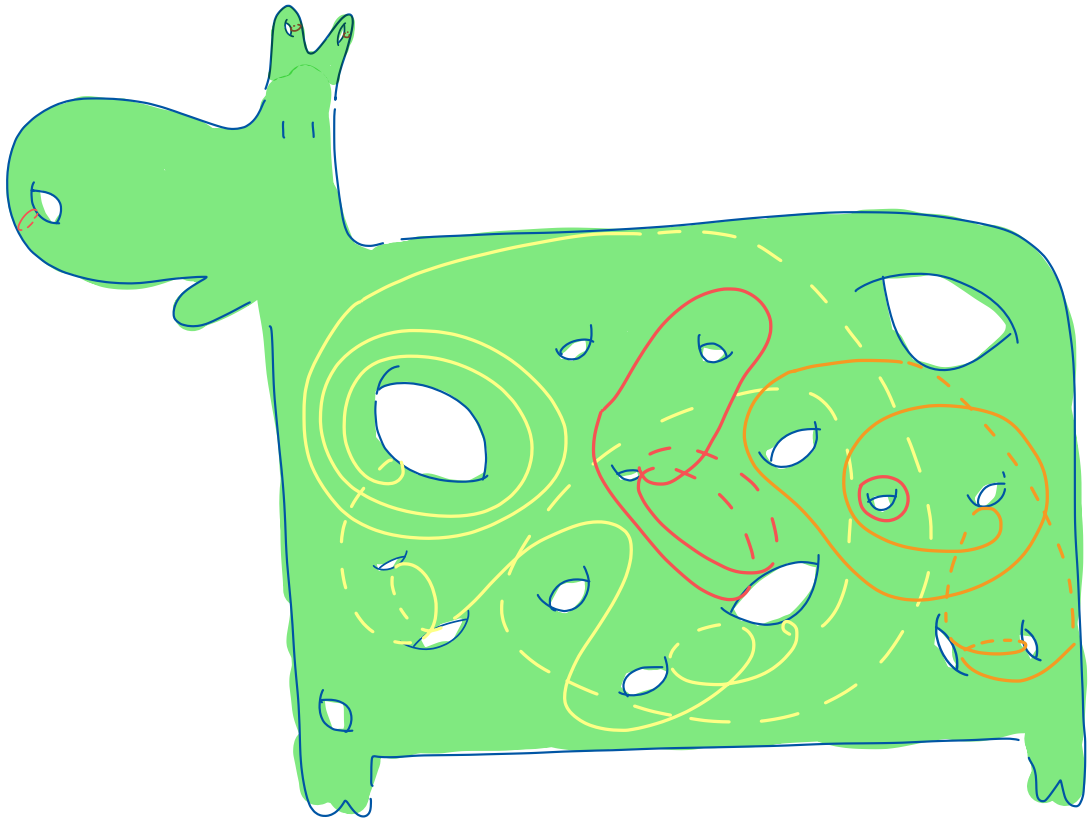

Figure 2.6: Une multi-géodesique typique sur une surface hyperbolique de grand genre

Chapter 3

Background

Throughout this text we shall use the symbol Σ_g to denote a connected, closed, oriented, topological surface of genus $g \geq 2$, and denote by $\Sigma_{g,n}$ a connected closed oriented topological 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.

3.1 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 67. 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*

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

Non-trivial examples of a hyperbolic surface can be obtained by the quotient of $\mathbb{H}^2$ by the action of a discrete subgroup (a *Fuchsian group*). The *modular group* $\mathrm{PSL}(2, \mathbb{Z})$, and the congruence subgroups

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

where $N \geq 2$, are typical examples of Fuchsian groups. For a detailed discussion of Fuchsian groups, see [Kat92].

Example 68. For any triple $(L_1, L_2, L_3) \in \mathbb{R}^3_{\geq 0}$, 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. However, the hyperbolic tori with a cusp are in 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 a topological surface Σ_g form a *moduli space* $\mathcal{M}_g$.

3.1.1 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 *equivalent* if $\phi_2 \circ \phi_1^{-1}$ is isotropic to an isometry. The *Teichmüller space*, denoted by $\mathcal{T}(\Sigma_g)$ or simply $\mathcal{T}_g$, is the set of such equivalence classes.

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 $\mathrm{Mod}_g := \mathrm{Homeo}^+(\Sigma_g)/\mathrm{Homeo}_0(\Sigma_g)$.

The group $\mathrm{Homeo}^+(\Sigma_g)$ acts (properly and discontinuously) 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 $\mathcal{T}_{g,n}(L_1, \ldots, L_n)$ and moduli space $\mathcal{M}_{g,n}(L_1, \ldots, L_n)$ of oriented complete hyperbolic surfaces of genus g with n (labeled) totally geodesic boundary components of lengths $L_1, \ldots, L_n \geq 0$ respectively can be defined in a similar manner.

3.1.2 Curves

In the introduction, the theorems are stated in terms of geodesics. Nevertheless, it is more convenient to work with (the free homotopy classes of) the topological curves, and they are actually equivalent for our purposes. A curve in a topological space X is (the image of) a continuous application $\mathbb{S}^1 \to X$. In this paper, we are interested in curves up to free homotopy. A closed curve is said to be *simple* if it does not intersect itself. A *multicurve* is a finite multiset of disjoint simple curves, and a multicurve is *ordered* (or *labeled*) if its underlying set is labeled. We will often write an ordered multicurve γ as an ordered list $(m_1\gamma_1, \ldots, m_k\gamma_k)$, and its unlabeled counterpart as a formal sum $\overline{\gamma} = m_1\gamma_1 + \cdots + m_k\gamma_k$ where $m_i \in \mathbb{Z}_{\geq 1}$, $1 \leq i \leq k$.

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 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 lie in the same mapping class group orbit. The three following subgroup, associated to γ, of the

mapping class group Mod_g will be useful later in the paper: $\mathrm{Stab}(\overline{\gamma})$ 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 $1 \leq i \leq k$, and $\mathrm{Stab}^+(\gamma)$ which fixes every γ_i and its orientation for all $1 \leq i \leq k$.

Let $X \in \mathcal{T}_g$. If a closed curve α on X is not homotopic to a point, then α is freely homotopic to a unique geodesic with the minimum length over all curves in the free homotopy class of α, and we write $\ell_X(\alpha)$ for the length of this geodesic.

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

3.1.3 Fenchel–Nielsen coordinates

A *pair of pants* is a surface that is homeomorphic to $\Sigma_{0,3}$, in other words, a sphere with three holes. A *pants decomposition* of $\Sigma_{g,n}$ is a set of disjoint simple closed curves $\{\alpha_1, \ldots, \alpha_{3g-3+n}\}$ on $\Sigma_{g,n}$ such that $\Sigma_{g,n} \smallsetminus \{\alpha_1, \ldots, \alpha_{3g-3+n}\}$ is a disjoint union of pairs of pants.

Fix an ordered pants decomposition $(\alpha_1, \ldots, \alpha_{3g-3+n})$ of $\Sigma_{g,n}$. Given $X \in \mathcal{T}_{g,n}(L_1, \ldots, L_n)$ (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 the length one turns before gluing two pairs of pants along α_i. See [Bus92, Section 1.7] for a precise definition). These $6g - 6 + 2n$ parameters are called *Fenchel–Nielsen coordinates*. The application $\mathcal{T}_{g,n}(L_1, \ldots, L_n) \to (\mathbb{R}_{>0} \times \mathbb{R})^{3g-3+n}$ given by $X \mapsto (\ell_{\alpha_i}(X), \tau_{\alpha_i}(X))_{i=1}^k$ is a bijection (see [Bus92, Chapter 6]).

3.1.4 Weil–Petersson volumes

Theorem 69 ([Wol83])**.** *Given a pants decomposition $\{\alpha_1, \ldots, \alpha_{3g-3+n}\}$, the formula*

$$\sum_{i=1}^{3g-3+n} d\ell_{\alpha_i} \wedge d\tau_{\alpha_i}$$

defines a symplectic form which has the same expression in any other Fenchel–Nielsen coordinates. In particular, it is invariant under the action of the mapping class group.

The symplectic form thus defined is the so-called *Weil–Petersson symplectic form*, and we shall denote it by ω. See [Wol83] for a more intrinsic definition.

Every symplectic form defines a volume form. The *Weil–Petersson volume* of the moduli space $\mathcal{M}_{g,n}(L_1, \ldots, L_n)$ is defined by

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

The following fundamental result is due to Mirzakhani.

Theorem 70 ([Mir07c])**.** *The Weil–Petersson volume $V_{g,n}(L_1, \ldots, L_n)$ is a symmetric polyno-*

mial 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* $\kappa_1 = [\omega]/2\pi^2 \in H^2(\overline{\mathcal{M}}_{g,n}, \mathbb{Q})$ *is the first Mumford class.*

See Section 3.3 for (slightly) more precise definitions of ψ_i and κ_1.

3.2 Flat geometry of surfaces

3.2.1 (Half-)Translation surfaces

There are (at least) three equivalent definitions of a *translation surfaces* of genus g.

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 [Wri15, Section 1] for a detailed discussion.

Example 71. If the collection of polygons in the third definition of a (half-)translation surface is is indeed a finite collection of squares isomorphic to $[0,1]^2 \subset \mathbb{R}^2 \simeq \mathbb{C}$, then the resulting surface is called a *square-tiled surface.*

3.2.2 Period coordinates

Let $\mathcal{H}_g$ denote the moduli space of (the set of) Abelian differentials of genus g, and let $\mathcal{Q}_g$ denote the moduli 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 each stratum corresponds to a partition of $2g-2$ (resp. $4g-4$) representing 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 piece-wise linear integral complex orbiford of dimension $4g-3$. Let $(C,\alpha) \in \mathcal{H}(1^{2g-2})$ be an Abelian differnetial, $Z(\alpha)$ be the set of its zeros, and let $H^1(C, Z(\alpha); \mathbb{C})$ denote the cohomology group of C relative to $Z(\alpha)$. It can be proved that there exists a open neighbourhood U of (C,α) in $\mathcal{H}(1^{2g-2})$ such that for all $(C',\alpha') \in U$, $H^1(C', Z(\alpha'); \mathbb{C})$ can be identified with $H^1(C, Z(\alpha); \mathbb{C})$. Thus, the principal stratum $\mathcal{H}(1^{2g-2})$ is locally modelled on $H^1(C, Z(\alpha); \mathbb{C})$ via the *period map*

$$\beta \mapsto \left(\beta \mapsto \int_\gamma \beta\right).$$

By choosing a basis $\{\gamma_i\}_{i=1}^{2g+2g-2-1}$ for the relative homology group $H_1(C, Z(\alpha); \mathbb{Z})$, the map

$$\beta \mapsto \left(\int_{\gamma_1} \beta, \dots, \int_{\gamma_{4g-3}} \beta\right) \in \mathbb{C}^{4g-3}$$

defines a local period coordinate system of $\mathcal{H}(1^{2g-2})$ near (C,α), and the integral points correspond to square-tiled (translation) surface via

$$C \to \mathbb{T}^2, \qquad z \mapsto \int_\bullet^z \beta \mod \mathbb{Z} \oplus i\mathbb{Z}$$

where $\bullet$ can be taken as, for example, the first zero of β.

A piece-wise linear integral 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 any stratum) of the moduli space of quadratic differentials $\mathcal{Q}(1^{4g-4})$ that we are going to discuss can be defined in a similar (but more technical) way. For a detailed discussion, see for instance, [DGZZ16, Appendix A], [FM14, Section 2.3].

Given a quadratic differential $(C, q) \in \mathcal{Q}_g(1^{4g-4})$, there exists a unique ramified double covering $\pi\colon \hat{C} \to C$ branched over the zeros $Z(q)$ of q, such that q lifts to the square of an Abelian differential on $\hat{C}$, denoted by $\sqrt{q}$, with only simple zeros. ($\hat{C}$ is of genus $4g-3$ by Riemann–Hurwitz formula.) In other words, there exists $(\hat{C}, \sqrt{q}) \in \mathcal{H}(1^{4g-4})$ such $\pi^* q = \sqrt{q} \otimes \sqrt{q}$. Consider the involution $\iota\colon \hat{C} \to \hat{C}$ that exchanges the two sheets of the ramified double covering. The cohomology group $H^1(\hat{C}, \hat{Z}(q); \mathbb{C})$ of $\hat{C}$ relative to $\hat{Z}(q) := \pi^{-1}(Z(q))$ splits into

$$H^1_-(\hat{C}, \hat{Z}; \mathbb{C}) \oplus H^1_+(\hat{C}, \hat{Z}; \mathbb{C})$$

where $\iota^*\alpha = -\alpha$ for all $\alpha \in H^1_-(\hat{C}, \hat{Z}; \mathbb{C})$ and $\iota^*\alpha = \alpha$ for all $\alpha \in H^1_+(\hat{C}, \hat{Z}; \mathbb{C})$. (Note that $[\sqrt{q}] \in H^1(\hat{C}, \hat{Z}; \mathbb{C})^-$.) Near (C, q), the principal stratum $\mathcal{Q}(1^{4g-4})$ is locally modelled on $H^1_-(\hat{C}, \hat{Z}(q); \mathbb{C})^-$ via the period map

$$p \mapsto \left(\gamma \mapsto \int_\gamma \sqrt{p}\right) \in \mathbb{C}^{6g-6}.$$

By choosing a basis $\{\gamma_1, \dots, \gamma_{6g-6}\}$ of $H_1(\hat{C}, \hat{Z}(q); \mathbb{Z})$, we obtain a system of period coordinates

$$p \mapsto \left(\int_{\gamma_1} \sqrt{p}, \cdots, \int_{\gamma_{6g-6}} \sqrt{p}\right)$$

on $\mathcal{Q}(1^{4g-4})$ near q, and the lattice points can be seen as square-tiled (half-translation) surface by considering

$$C \to \mathbb{T}^2/\pm \simeq \mathbb{C}\mathrm{P}^1, \qquad z \mapsto \int_\bullet^z \sqrt{p} \mod \mathbb{Z} \oplus i\mathbb{Z}.$$

Remark 72. The moduli space $\mathcal{Q}_g$ of holomorphic quadratic differential on Σ_g can be identified with the total space of the cotangent bundle $T^*\mathcal{M}_g$ over $\mathcal{M}_g$, thus $\mathcal{Q}_g$ carries a tautological symplectic form ω_g, and therefore a volume form. It turns out that this volume element differs from the one defined by periodic coordinates only by a constant depending only on g. See [DGZZ21, Section 2.1] for more details.

3.2.3 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

$$\mathrm{area}(C, q) := \int_C |q|$$

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

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

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

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

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

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 asymptotic square-tiled surfaces counting

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

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.

3.2.4 Stable graphs

Given a mutlicurve $m_1\gamma_1 + \cdots + m_k\gamma_k$ on Σ_g, one can associate with it with a *stable graph* in the following way. 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. See Figure 3.1 for an example. Note that the resulting graph does not depend on $m_1, \ldots, m_k$, but its edges can be naturally weighted the m_i's.

Given a square-tiled surface q, a (weighted) stable graph can also be constructed as follows. Consider the horizontal flat geodesic foliation of q. Then q can be decomposed into singular leaves and maximal horizontal cylinders consisting of regular leaves. The stable graph 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.

More formally, a stable graph consists of the following data

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

satisfying the following properties:

1. The pair (V, E) defines a connected graph, with the vertex set V and edge set E. The set H is the set of *half-edges*.

2. The map v assigns each half-edge to its adjacent vertex.

3. 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.

4. 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.

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

1. $\psi \circ \iota_1 = \iota_2 \circ \psi$. In other words, ψ maps an edge of Γ_1 to en edge of Γ_2.

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.

3. $g_1 = g_2 \circ \phi$.

Note that ψ determines ϕ but it is convenient to record automorphism as a pair (ϕ, ψ). We denote by $\mathrm{Aut}(\Gamma)$ the set of automorphisms of Γ.

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

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, we have

$$|\mathrm{Aut}(\Gamma)| = [\mathrm{Stab}(\gamma_1 + \cdots + \gamma_k) : \mathrm{Stab}^+(\gamma_1, \ldots, \gamma_k)], \tag{3.1}$$

and

$$|\mathrm{Aut}(\Gamma, \mathrm{m})| = [\mathrm{Stab}(m_1\gamma_1 + \cdots + m_k\gamma_k) : \mathrm{Stab}^+(\gamma_1, \ldots, \gamma_k)]. \tag{3.2}$$

3.2.5 Graph polynomials

Given a stable graph associated to the multicurve $\gamma = m_1\gamma_1 + \cdots + m_k\gamma_k$, we associate to each edge e a variable x_e, and define the associated *graph polynomial* by the formula

$$F_\gamma(x_e : e \in E) = \prod_e x_e \cdot \prod_v V_{g(v),n(v)}(x_{e(h)} : h \in H,\ v(h) = v). \tag{3.3}$$

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 P_γ, and $\bar{V}_{g,n}$ for that of $V_{g,n}$.

Example 73. If $\{\gamma_1, \dots, \gamma_{3g-3}\}$ is a pants decomposition, then $V_{g(v),n(v)} = 1$ for all $v \in V$, and $F_\gamma(x_1, \dots, x_{3g-3}) = \bar{F}_\gamma(x_1, \dots, x_{3g-3}) = x_1 \cdots x_{3g-3}$.

Example 74. 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,

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

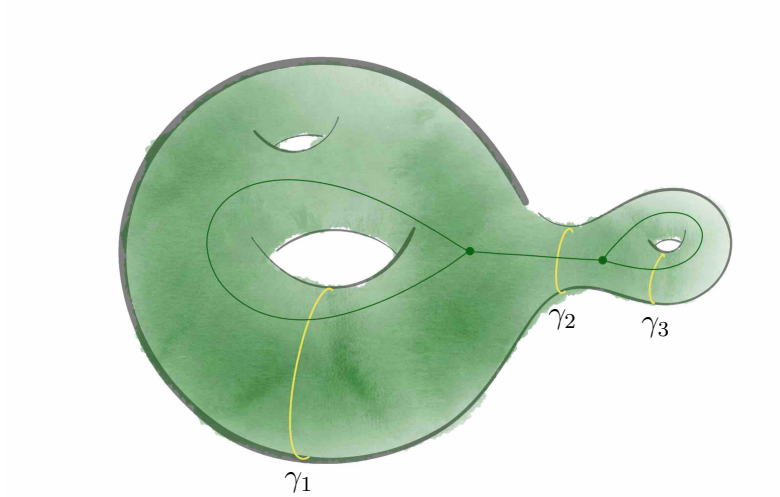

Figure 3.1: Example 74

3.3 Intersection numbers

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

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

The Deligne–Mumford moduli space of stable complex curves $\mathcal{M}_{g,n}$ 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 is the Chern class of $c_1(\mathcal{L}_i)\in H^2(\overline{\mathcal{M}}_{g,n},\mathbb{Q})$. 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 $\kappa_1 := \pi_*(\phi_{n+1}^2)\in H^2(\overline{\mathcal{M}}_{g,n},\mathbb{Q})$. From now now, 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 intersection numbers satisfy some recursive relations. The following two equations can be proved using standard algebro-geometric method.

Proposition 75 (String equation)**.**

$$\langle \tau_{d_1}\cdots\tau_{d_n}\tau_0\rangle_g = \sum_{i=1}^{n}\langle \tau_{d_1}\cdots\tau_{d_i-1}\cdots\tau_{d_n}\rangle_g$$

Proposition 76 (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 $\langle\tau_0^3\rangle=1$, and $\langle\tau_1\rangle_1=1/24$. 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 77. *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 78. *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 can deduce all intersection numbers from these two equations (even for $g = 2$). Motivated by the equivalence between different models for 2-dimensional quantum gravity, Witten proposes 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 number for all $g, n \geq 0$ with $2g - 2 + n > 0$.

To state this conjecture, we need some preparation. Let us define the following generating function

$$F(t_0, t_1, \dots) := \left\langle \exp\left(\sum_{i \geq 0} t_i \tau_i\right) \right\rangle \tag{3.4}$$

which can also be written as

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

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

$$F(t_0, t_1, \dots) = \frac{1}{24} t_1 + \frac{1}{6} t_0^3 + \frac{1}{48} t_1^2 + \frac{1}{24} t_0 t_2 + \frac{1}{6} t_0^3 t_1 + \frac{1}{1152} t_4 + \frac{1}{72} t_1^3 + \frac{1}{12} t_0 t_1 t_2 + \frac{1}{48} t_0^2 t_3 + \cdots$$

The Witten conjectures is the following

Theorem 79 (Witten–Kontsevich)**.** *The second derivative $U := \partial^2 F / \partial t_0^2$ 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 Kontsevich [Kon92], and several proofs have been found since then [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

$$b_0 := \frac{1}{8} + \sum_{i \geq 0} a_i a_{-i}, \qquad b_k := \frac{1}{2} \sum_{\substack{(i,j) \in \mathbb{Z}^2 \\ i+j=k}} a_i a_j, \quad \text{if } k \neq 0.$$

We say that F satisfies the *Virasoro constraints* if

$$(a_{-(2k+3)} - b_{-2k}) \exp(G) = 0 \tag{3.5}$$

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 $k \geq 1$, (3.5) is equivalent to

$$\begin{aligned}(2k+3)!!\langle \tau_{d_1}\cdots\tau_{d_n}\tau_{k+1}\rangle_g = &\sum_{i=1}^{k}\frac{(2d_i+2k+1)!!}{(2d_i-1)!!}\langle \tau_{d_1}\cdots\tau_{d_i+k}\cdots\tau_{d_n}\rangle_g \\ &+\frac{1}{2}\sum_{\substack{(i,j)\in\mathbb{Z}^2_{\geq 0}\\ i+j=k}}(2i+1)!!\,(2j+1)!!\,\langle \tau_{d_1}\cdots\tau_{d_n}\tau_i\tau_j\rangle_{g-1} \\ &+\frac{1}{2}\sum_{\substack{(i,j)\in\mathbb{Z}^2_{\geq 0}\\ i+j=k-1}}(2i+1)!!\,(2j+1)!!\sum_{\substack{g_1,g_2,I_1,I_2\\ I\sqcup J=\{1,\ldots,n\}}}\langle \tau_{d_I}\tau_i\rangle\langle \tau_{d_J}\tau_j\rangle,\end{aligned}$$

3.4 Measured laminations

Multicurves can be regarded as "lattice points" in the space of *measured laminations* $\mathcal{ML}_g$. We will only need the following properties of this space. See, e.g., [Kap01, Chapter 11] for more details.

1. $\mathcal{ML}_g$ is a $6g-6$-dimensional real manifold equipped with a natural piece-wise integral linear structure, i.e., $\mathcal{ML}_g$ has an natural atlas whose transition functions are piece-wise in $\mathrm{GL}(6g-6,\mathbb{Z})$.

2. The integral points in the coordinate charts of $\mathcal{ML}_g$, denoted by $\mathcal{ML}_g(\mathbb{Z})$, are in natural bijection with the (free homotopy classes of) integral multicurves on Σ_g.

3. The action of the mapping class group on the set of multicurves extends to $\mathcal{ML}_g$.

4. $(\mathbb{R}_{>0},\times)$ acts on $\mathcal{ML}_g$, and for any multicurve $\overline{\gamma}=m_1\gamma_1+\cdots+m_k\gamma_k$, and any $r\in\mathbb{R}_{>0}$, $r\cdot(m_1\gamma_1+\cdots+m_k\gamma_k)=r\,m_1\gamma_1+\cdots+r\,m_k\gamma_k$. We denote the quotient by $\mathbb{P}(\mathcal{ML}_g)$.

5. Given $X\in\mathcal{T}_g$, the length function ℓ_X defined on the set of multicurves extends to $\mathcal{ML}_g$. Moreover, for any $\lambda\in\mathcal{ML}_g$, $\ell_{X\cdot h}(h^{-1}\cdot\lambda)=\ell_X(\lambda)$ for any $h\in\mathrm{Mod}_g$, and $\ell_X(r\cdot\lambda)=r\cdot\ell_X(\lambda)$ for any $r\in\mathbb{R}_{>0}$.

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

7. $\mathcal{ML}_g$ carries a natural mapping class group invariant measure μ_{Th} defined by asymptotic counting of integral points, called the *Thurston measure*. For any open subset $U\subset\mathcal{ML}_g$, $\mu_{\mathrm{Th}}(t\cdot U)=t^{6g-6}\mu_{\mathrm{Th}}(U)$.

3.5 Thurston distance

Let $X_1, X_2 \in \mathcal{T}_g$. Set

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

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, e.g. $0 < \epsilon < 1$,

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

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

$$(1-\epsilon) \cdot \ell_X(\lambda) \leq \ell_Y(\lambda) \leq (1+\epsilon) \cdot \ell_X(\lambda) \tag{3.6}$$

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

3.6 Earthquake

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

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

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 both invariant 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 80 ([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.

3.7 GEM distribution and Poisson–Dirichlet distribution

The *beta distribution* $\mathrm{Beta}(a,b)$ with parameters (a,b) is a probability distribution whose density function is

$$\mathbb{1}_{]0,1[}(x) \cdot \frac{x^{a-1}(1-x)^{b-1}}{\mathrm{B}(a,b)} \qquad \text{where} \qquad \mathrm{B}(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.$$

One of the simplest ways to define a *GEM distribution* works as follows through a so-called *stick breaking process.*

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. Call V_2 the length of the piece on the right, 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

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

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

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

The *Poisson–Dirichlet distribution* of parameter θ, is defined as the distribution of the descending order statistics $V^\downarrow = (V_1^\downarrow, V_2^\downarrow, \dots)$ of V.

The Poisson–Dirichlet distribution is first introduced by Kingman [Kin75] as follows: Let $X = \{X(t) : t > 0\}$ be the Gamma process with parameter θ, that is, a Lévy process on $(0, \infty)$ with intensity $\theta\, x^{-1}e^{-x}$. The sample paths of X is increasing and purely discontinuous. Denote by $J^{\downarrow}(1) \geq J^{\downarrow}(2) \geq \cdots \geq 0$ be the jumps of X listed in descending order. Then the distribution of the random sequence $(J^{\downarrow}(1)/X(\theta), J^{\downarrow}(2)/X(\theta), \dots)$ is called the Poisson–Dirichlet distribution with parameter θ. Or equivalently: Let $N = \{N(t) : t > 0\}$ be the Poisson point process on $(0, \infty)$ with intensity $\theta x^{-1}e^{-x}$ and let $X(1) > X(2) > \cdots > 0$ be its points. The sum $S = X(1) + X(2) + \cdots$ is almost surely finite, and the sequence of random variables $(X(1)/S, X(2)/S, \dots)$ follows the Poisson–Dirichlet distribution of parameter θ.

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

Proposition 81 ([DJ89])**.** *Let $\theta \in \mathbb{R}_{>0}$. If $X = (X_1, X_2, \dots)$ follows the* GEM(θ) *distribution, then finite finite-dimensional distribution of $(X_1, \dots, X_r)$ has a density function given by the formula*

$$\frac{\theta^r (1 - x_1 - \cdots - x_r)^{\theta - 1}}{(1 - x_1)(1 - x_1 - x_2) \cdots (1 - x_1 - \cdots - x_{r-1})}.$$

3.8 Analytic combinatorics

When one is dealing with an enumeration problem, it is often helpful to encode the information in a *generating function*. For example, if we denote by C_n the number of full binary trees with $n + 1$ leaves, then C_n is the famous (n-th) *Catalan number*

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

The generating function for these numbers is defined by

$$c(x) := \sum_{n=0}^{\infty} C_n x^n = \frac{1 - \sqrt{1 - 4x}}{2x}.$$

A generating function $f(x) = \sum_{n=0}^{\infty} a_n x^n$ is only a formal power series. However, its analytic property tell us a lot about the asymptotic behavior of its coefficients a_n. More precisely, the asymptotic behavior of a_n as $n \to \infty$ can be read from the behavior of $f(x)$ around its singularities.

The following tools, first developed in [FO90], are standard in singularity analyses of generating functions.

Theorem 82 ([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 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 83 ([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 84 ([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

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

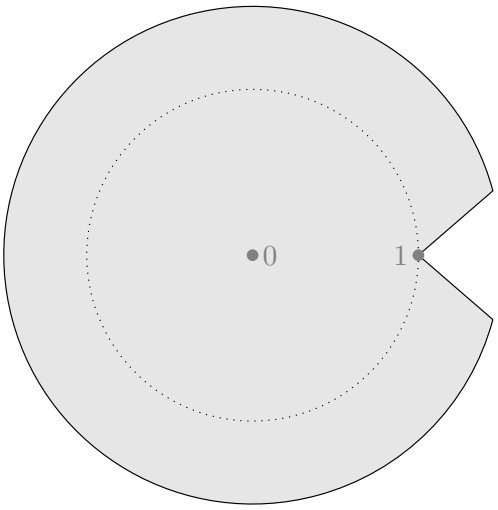

Figure 3.2: A Δ-domain

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 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 Random multi-geodesics model

Let $X \in \mathcal{M}_g$ be a hyperbolic surface. Let s_X denote that set of all multi-geodesic (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. 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.

Each question that one can ask to a particular multi-geodesic, for instance, "Is it separating?", "Does it has the same topological type than γ?", "What is its length partition?", can also be addressed to a random multi-geodesic. But the answers are now "quantum", for example, a random multi-geodesic can be separating with certain probability p, and non-separating with probability $1 - p$.

By the nature of the model we choose, it is necessary to understand the asymptotic behavior of $|s_X(R)|$ as $R \to \infty$, or more general, for fixed multicurve γ, the growth of cardinality

of

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

as $R \to \infty$.

G. McShane and I. Rivin [MR95a] [MR95b] showed that in the case X is a punctured torus, $|s_X(R)| \sim c_X L^2$ as $R \to \infty$, where c_X is a constant depending on X. Later, I. 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 generalized by M. Mirzakhani in her thesis:

Theorem 85 (Mirzakhani)**.** *For any $X \in \mathcal{M}_{g,n}$ and any multicurve γ on $\Sigma_{g,n}$, there exists a positive rational constant $c(\gamma)$ such that*

$$|s_X(\gamma, R)| \sim c(\gamma) \cdot \frac{B(X)}{b_{g,n}} \cdot R^{6g-6+2n}$$

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

$$b_{g,n} = \sum_{[\gamma]} c(\gamma) = \int_{\mathcal{M}_{g,n}} B(X)\, dX$$

where $[\gamma]$ runs over all topological types of multicurves on $\Sigma_{g,n}$ and dX is the Weil–Petersson measure on $\mathcal{M}_{g,n}$.

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

Corollary 86. *We have*

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

Remark 87. In [Mir08b] Mirzakhani gives an explicit formula for $c(\gamma)$.

A remarkable fact is that this ratio $c(\gamma)/b_{g,n}$, 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 γ (but not on the hyperbolic metric X), and one can therefore simply talk about the frequency of multicurves.

Example 88. 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 $c(\gamma) = 1/189$.

Example 89. Let $\alpha = (\alpha_1, \alpha_2)$ be a 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 $\alpha = 1/45$.

Example 90. The probability that a random primitive multi-geodesic on $X \in \mathcal{M}_2$ is separating is $1/49$.

4.2 Masur–Veech volumes

Mirzakhani constructs in [Mir08a] a conjugacy between the earthquake flow over $\mathcal{M}_g$ and the Teichmüller horocycle flow over $\mathcal{Q}_g$, and establishes the ergodicity of the earthquake flow, as a byproduct, she finds the following surprising identity.

Theorem 91 ([Mir08a])**.** *For any $g \geq 2$, there exists a constant C_g such that*

$$b_g = C_g \cdot \mathrm{vol}_{\mathrm{MV}}(\mathcal{Q}_g)$$

This constant was later determined via different geometric methods independently by F. Arana-Herrera and L. Monin–V. Telpukhovskiy.

Theorem 92 ([AH20b], [MT19])**.** *For any g, n with $2g-2+n>0$,*

$$b_{g,n} = \frac{\mathrm{vol}_{\mathrm{MV}}(\mathcal{Q}(1^{4g-4+n}, -1^n))}{2^{4g-1+n}(3g-3+n)(4g-4+n)!}.$$

In [DGZZ21], the authors showed that the Masur–Veech volume of the principal stratum of the moduli space of quadratic differentials $\mathcal{Q}(1^{4g-4+n}, -1^n)$ 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 93 ([DGZZ21, Theorem 1.21])**.** *Let $\gamma = m_1\gamma_1 + \cdots + m_k\gamma_k$ be a multicurve on $\Sigma_{g,n}$. Then we have*

$$\frac{c(\gamma)}{[\mathrm{Stab}(\gamma_1+\cdots+\gamma_k) : \mathrm{Stab}^+(\gamma_1,\ldots,\gamma_k)]} = \frac{1}{2^{4g-1+n}(3g-3+n)(4g-4+n)!}\frac{\mathrm{vol}_{\mathrm{MV}}(\Gamma(\gamma),(m_1,\ldots,m_k))}{|\mathrm{Aut}(\Gamma,(m_1,\ldots,m_k))|},$$

or equivalently,

$$\frac{c(\gamma)}{[\mathrm{Stab}(\gamma_1+\cdots+\gamma_k) : \mathrm{Stab}(\gamma)]} = \frac{\mathrm{vol}_{\mathrm{MV}}(\Gamma(\gamma),(m_1,\ldots,m_k))}{2^{4g-1+n}(3g-3+n)(4g-4+n)!}.$$

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

Let I be a finite set. Give a function $\mathrm{m}\colon I \to \mathbb{Z}_{\geq 1}$, we define a linear operator $\mathcal{Y}_{\mathrm{m}}$ 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

$$\mathcal{Y}_{\mathrm{m}} : \prod_{i\in I} x_i^{n_i} \mapsto \prod_{i\in I} \frac{n_i!}{\mathrm{m}(i)^{n_i+1}}. \tag{4.2}$$

Remark 94. 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

$$\mathcal{Y}_{\mathrm{m}}(\bullet) = (n+k-1)! \int_{\Delta_1^{k-1}} \bullet\, \lambda(dx)$$

where λ is the Lebesgue measure on Δ_1^{k-1}.

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 (4.2) can be rewritten as

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

where $m_i := \mathrm{m}(i)$. Given $m \in \mathbb{Z}_{\geq 1} \cup \{+\infty\}$, let us collect different $\mathcal{Y}_{\mathrm{m}}$ with the same index set together and define

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

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

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

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

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

Remark 95. 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 term of intersection numbers of ψ-classes by the work of Kontsevich [Kon92] and Norbury [Nor13].

Remark 96. Let $\gamma = (m_1\gamma_1, \ldots, m_k\gamma_k)$ be an ordered multicurve on Σ_g and Γ be its associated stable graph. The polynomial $\bar{F}_\gamma$ appears in Theorem 104 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 97 ([DGZZ21])**.** *The volume contribution of* $\mathrm{vol}_{\mathrm{MV}}(\mathcal{Q}_{g,n})$ *from* (Γ, m) *is*

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

We can also write

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

Note that although the P_Γ and $\bar{F}_\gamma$ arise for different reasons, both of them can be expressed in term of intersection numbers between psi-classes on $\overline{\mathcal{M}}_{g,n}$ (by [Kon92], [Nor13], and [Mir07c]). 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 (3.5) satisfied by the generating function of intersection numbers of psi-classes defined by (3.4) 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. Nevertheless, Aggarwal–Delecroix–Goujard–Zograf–Zorich conjectured in [ADG+20] that these numbers have a simple asymptotic formula as the genus goes to infinite. A lower bound [DGZZ20d] was proved later, and the whole conjecture was turned into a theorem by A. Aggarwal (in a stronger form) [ADG+20] using ingenious and deeply insightful probabilistic methods.

Theorem 98 ([Agg21])**.** *For* $g, n \in \mathbb{Z}_{\geq 0}$ *with* $2g - 2 + n > 0$ *and* $(d_1, \ldots, d_n) \in \mathbb{Z}^n_{\geq 0}$ *with* $d_1 + \cdots + d_n = 3g - 3 + n$*, let* $\epsilon(d_1, \ldots, d_n)$ *be defined by*

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

We have

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

This result is the key ingredient of [DGZZ20c] and [DL22] (last chapter of this thesis).

4.4 Topology of large genus random multi-geodesics

Based on the work of A. Aggarwal [Agg20] about large genus asymptotics for intersection numbers between psi-classes and for the Masur–Veech volume, Delecroix–Goujard–Zograf–Zorich proved in their recent paper give a beautiful answer to Question 7 in the introduction.

Theorem 99 ([DGZZ20c, Theorem 1.12]). *For all $z \in \mathbb{C}$ such that $|z| < 8/7$, the following asymptotic relation is valid as $g \to \infty$*

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

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

The preceding theorem implies that the distribution of 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 100 ([DGZZ20c, Theorem 1.13]). *For any $\kappa \in]1, 1.236]$ such that $\kappa \log(6g-6)/2$ is an integer, we have*

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

A more telling consequence of Theorem 99 might be the following, which also answers Question 7 in the introduction.

Theorem 101 ([DGZZ20c]). *As $g \to \infty$, we have*

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

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

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

On the other hand, in the same article, the authors also proved

Theorem 102 ([DGZZ20c, Theorem 1.1]). *The probability that a random multicurve is separating tends to zero as the genus of the underlying surface goes to infinity.*

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

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 M. Mirzakhani in [Mir16], where she proves

Theorem 103 ([Mir16, Theorem 1.2])**.** *If $\{\gamma_1, \dots, \gamma_{3g-3}\}$ is a pants decomposition of X, then $\hat{\ell}_{X,\gamma,R}$ converges in law to the Dirichlet distribution of order $3g-3$ with parameters $1, \dots, 1$, 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 $\Delta^{3g-4} := \{(x_1, \dots, x_{3g-3}) \in \mathbb{R}_{\geq 0} : x_1 + \cdots + x_{3g-3} = 1\}$ of dimension $3g-4$. In other words, for any open subset U of Δ^{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^{3g-4}.

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

Theorem 104. *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 (3.3).

Remark 105. 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 70). 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 104 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 article of F. Arana-Herrera [AH21b] appeared on the arXiv. The paper [AH20a] and the current paper are devoted to similar circle of problems and use similar circle of ideas, though they were written in parallel and completely independently. In particular, paper [AH21b] proves a much more general version of our Theorem 114 ([AH21b, Theorem 1.3]), which is one of the key ingredients allowing to attack the counting problem and the length statistics. We learned from the paper [AH21b] that this kind of statistics was initially conjectured by S. Wolpert. Papers [AH21b] and [AH20a] established results closely related to Theorem 104.

Proposition 117 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 121.

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 106. *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_γ/Mod_g. The resulting map $P_\gamma/\mathrm{Mod}_g \to \mathcal{M}_g^\gamma$ is a local isomorphism of symplectic orbifolds. All that remains now is to show that the map $P_\gamma/\mathrm{Mod}_g \to \mathcal{M}_g^\gamma$ 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 $s \in \mathrm{Stab}(\gamma)$ 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 107. 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 108. *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

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

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 $g \geq 3$, 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 109 (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 110. 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 $\mathcal{S}_{R,\gamma}^A \subset \mathcal{M}_g^\gamma$ can be written as $\ell_\alpha^{-1}(R \cdot A)$ where $R \cdot A$ is defined by $\{(x_1, \ldots, x_k) \in \mathbb{R}_{>0}^k : (x_1, \ldots, x_k)/R \in A\}$.

5.2.1 Horospherical measures

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

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

The horospherical measure $\mu_{R,\gamma}^A$ 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 $\mathcal{M}_g$, both of which by abuse of notation we shall also denote by $\mu_{R,\gamma}^A$. Note that $\mathcal{M}_g^\gamma \to \mathcal{M}_g$ is (ramified) covering map of infinite degree. However, its restriction on $\mathcal{S}_{R,\gamma}^A$ is of finite degree. Thus $\mu_{R,\gamma}^A$ on $\mathcal{M}_g$ is the push-forward measure of $\mu_{R,\gamma}^A$ by $\mathcal{M}_g^\gamma \to \mathcal{M}_g$. So for any open subset U of $\mathcal{M}_g$,

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

In particular, the total mass of $\mu_{R,\gamma}^A$ on $\mathcal{M}_g^\gamma$ and on $\mathcal{M}_g$ differ only 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 $\mathcal{M}_g^\gamma$ and $\mathcal{M}_g$.

Proposition 111. *The total mass of $\mu_{R,\gamma}^A$ on $\mathcal{M}_g$ is*

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

where $R \cdot A := \{(x_1, \ldots, x_k) \in \mathbb{R}_{\geq 0}^k : (x_1, \ldots, x_k)/R \in A\}$, λ is the Lebesgue measure on $\Delta_R^{k-1} := \{(x_1, \ldots, x_k) \in \mathbb{R}_{\geq 0} : x_1 + \cdots + x_k = R\}$, and P_γ is defined by the formula (3.3).

Proof. In the light of Remark 110, by taking f in Theorem 109 to be the following indicator function

$$\mathbb{1}\left\{(x_1, \ldots, x_k) \in \mathbb{R}_{>0}^k : \begin{array}{l} R \leq m_1 x_1 + \cdots + m_k x_k \leq R + \epsilon, \\ (m_1 x_1, \ldots, m_k x_k)/(m_1 x_1 + \cdots + m_k x_k) \in A \end{array}\right\}$$

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 112. *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 (3.3).

Remark 113. It results from Theorem 70 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 114. *We have the 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 115. *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 114. □

The proof of Theorem 114 rests on the following series of propositions.

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

Proposition 116. *The measure* ν *is invariant under the earthquake flow.*

Proposition 117. *The measure* ν *is absolutely continuous with respect to* ν_g.

Proposition 118. *The measure* ν *is a probability measure.*

Proof of Theorem 114. Proposition 116, Proposition 117, and Theorem 114 imply that ν and ν_g differ by a multiplicative constant, and it follows from Proposition 118 that this constant is 1. □

Proposition 116 is immediate.

Proof of Proposition 116. 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 117 and Proposition 118, which are more technical.

5.3.1 Escape to infinity?

In this subsection, we prove Proposition 118. The key ingredient is the following non-divergence result for earthquake flow due to Y. Minsky and B. Weiss.

Theorem 119 ([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 118. 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 119 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 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 109 to be the indicator function
$$\mathbb{1}\{(x_1, \ldots, x_k) \in \mathbb{R}^k_{\geq 0} : R \leq x_1 + \cdots + x_k \leq R + h,\ x_i < \epsilon\},$$

and we obtain

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

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

We now suppose that α is disjoint from $\gamma_1,\ldots,\gamma_k$. Denote by (γ,α) the ordered multicurve $(\gamma_1,\ldots,\gamma_k,\alpha)$. Again, by applying Theorem 109, f being the indicator function

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

we obtain the corresponding horospherical volume

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

where $\Delta^k_{[R,R+h]} := \{(x_1,\ldots,x_k) \in \mathbb{R}^k_{\geq 0} : R \leq x_1 + \cdots + x_k \leq R+h\}$. 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 112 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 $(\tau_{\gamma_1},\ldots,\tau_{\gamma_k})$-coordinates plane.) By Theorem 119,

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

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

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

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 117.

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 $f \le Cg$, and we write $f = \Theta_K(g)$ if there exists C, depending only on K, such that $(1/C)g \le f \le Cg$.

The key to the proof is the following estimates.

Theorem 120 ([Mir07a], [Mir16], [AH21b])**.** *Let $\epsilon \in (0,1)$, and let K be a compact subset of $\mathcal{M}_g$. We have*

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 121. 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 in a more general form without proof [Mir16, Proposition 2.1.b]. The proof of [Mir07a, Theorem 5.5.b] is concise and hard to follow. See also the footnote on p. 390 in [Wri20]. A much stronger estimate is obtained by F. Arana-Herrera in a different approach [AH21b, Proposition 1.5].

The rest of the proof of Proposition 117 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 122. *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 $U_x := \{\lambda \in \mathcal{ML}_g : \ell_x(\lambda) \le 1,\ [\lambda] \in U\}$.

Proof. It is enough to prove this for $A = \Delta^{k-1}$. By (3.6), for any $y \in \mathbb{B}_x(\epsilon)$, we have

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

and so

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

Hence Theorem 120.2 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 120.1 and Corollary 112. □

We need one further technical lemma.

Lemma 123. *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 120.1, 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 117. 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 118, we may assume that N is contained in a compact set $K \subset \mathcal{P}^1\mathcal{M}_g$. The preceding lemma 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 122 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

Theorem 124. *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 108, 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

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

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

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

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

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

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 125. *Let* $X \in \mathcal{M}_g$, $\gamma = (m_1\gamma_1, \dots, m_k\gamma_k)$ *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 have

Proof of Theorem 124. This follows at once from Theorem 125 and Lemma 108. □

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

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

To prove Theorem 125, 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 126. *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 (3.6) 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 $y \in \mathcal{B}_{(1+\epsilon)R}^+$. The 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 125. 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 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 $\mathcal{B}_R \subset \mathcal{M}_g^\gamma$, and 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 set of all $Y \in \mathcal{M}_g^\gamma$ that projects 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 126 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 115 and Corollary 112,

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

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,

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

and similarly,

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

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 104. Theorem 124 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 112. □

Example 127. 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 104 reduces to Theorem 103.

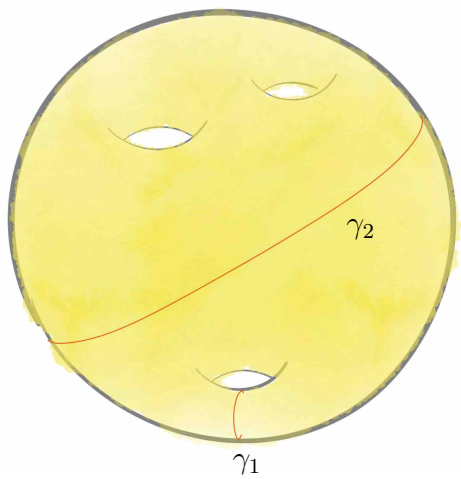

Figure 5.1: Example 127

Example 128. If $\gamma = (\gamma_1, \gamma_2)$, 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. 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 129. Let (γ_1, γ_2) be an order multicurve such that, for $i = 1, 2$, γ_i is separating, and γ_i bounds two surfaces of types $(g_i, 1)$ (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 $g_1 < g_2$.

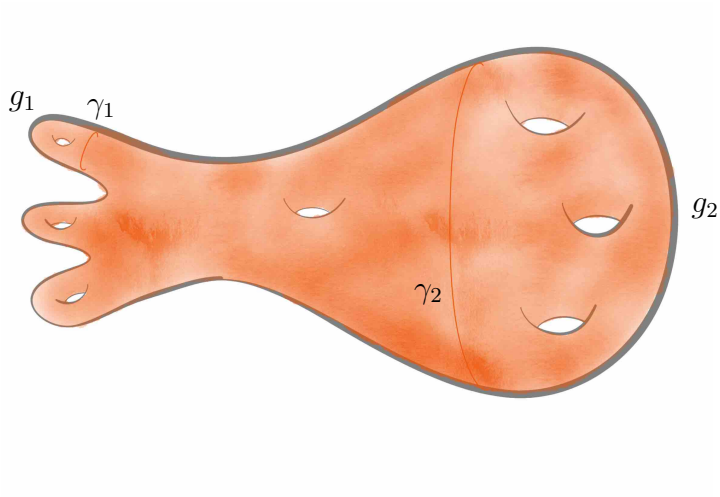

Figure 5.2: Example 128

Chapter 6

Large genus asymptotics

The essential content of this chapter is adapted from [DL22], 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 104 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, and so their lengths can not be putted into a vector in any reasonable 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

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

where $v^{\downarrow}$ 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 104.

Theorem 130. *The random vector $\hat{\ell}^{\downarrow}_{X,m,R}$ converges in distribution to a random vector $L^{\downarrow}_{g,m}$ which depends only on g and m. In other words, there exists a probability measure $\mu^{\downarrow}_{g,m}$ such that*

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

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 142).

The main result of this chapter is the following.

Theorem 131. *For any $m\in\mathbb{Z}_{\geq 1}\cup\{+\infty\}$, the sequence $(L^{\downarrow}_{g,m})_{g\geq 2}$ 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 $L^{\downarrow}_{g,1}$ and $L^{\downarrow}_{g,+\infty}$ 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 131 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 $\hat{\mathrm{a}}^{\downarrow}(q)\in\Delta_1^\infty$ 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 132 ([DGZZ21])**.** *Let $g\geq 2$, $m\in\mathbb{Z}_{\geq 1}\cup\{+\infty\}$, and let $L^{\downarrow}_{g,m}$ be the same random variable as in Theorem 130. Then as $N\to\infty$ we have the following convergence in distribution*

$$\frac{1}{|\mathcal{ST}(m,N)|}\sum_{M\in\mathcal{ST}(m,N)}\delta_{\hat{\mathrm{a}}^{\downarrow}(M)}\to L^{\downarrow}_{g,m}.$$

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 130 and Theorem 132 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 132 that Theorem 131 admits the following reformulation.

Corollary 133. *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 130; see Theorem 142 in Section 6.3. The formula follows from the work of M. Mirzakhani on random pants decompositions [Mir16] and the result of F. Arana-Herrera [AH21b] and M. 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 150 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 [DGZZ20c].

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 151.

6.2 Size-biased sampling and GEM distribution

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

Let us begin by introducing some notation. Let s be a positive real number, and define

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

Note that the canonical embeddings $\Delta^{i}_{\leq s}\hookrightarrow\Delta^{j}_{\leq s}$ given by $(x_1,\dots,x_k)\mapsto(x_1,\dots,x_k,0,\dots,0)$ for $i\leq j$ defines a direct system, and $\Delta^{\infty}_{\leq s}$ is its direct limit (the same for Δ^{∞}_{s}). We equip

$\Delta^\infty_{\le s}$ with the topology induced from the product topology $[0,s]^{\mathbb{Z}_{\ge 1}}$, and Δ^∞_s with the induced topology given by the canonical embedding $\Delta^\infty_s \hookrightarrow \Delta^\infty_{\le s}$. Note that the closure of Δ^∞_s in $\Delta^\infty_{\le s}$ is $\Delta^\infty_{\le s}$.

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 $\lambda_1 \ge \lambda_2 \ge \cdots \ge 0$, i.e., a point in Δ^∞_s, and a random partition is a probability measure on this set. However, the following seemingly awkward idea works perfectly 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. Given a partition of s represented as a point $x = (x_1, x_2, \dots) \in \Delta^\infty_s$ (this representation is not unique), let S_∞ denote the set of bijections from $\mathbb{Z}_{\ge 1}$ to itself. By Kolmogorov extension theorem, x defines a probability measure $\mathbb{P}_x$ on S_∞ in the following way.

1. If $s = 0$, i.e. $x_i = 0$ for all $i \ge 1$, then $\mathbb{P}_x(\sigma) = 1$ if σ is the identity, and zero otherwise.

2. If $s > 0$, set $\mathbb{P}_x(\sigma(1) = i_1) = x_{i_1}/s$, and for any $n \ge 1$, set

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

   to be

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

   ii. 1 if $s = x_{i_1} + \cdots + x_{i_n}$ and $i_{n+1} = n+1$.

   iii. 0 otherwise.

Note that if $s - x_1 - \cdots - x_n > 0$, then

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

if $i_1, \dots i_n$ are all distinct and $s - x_{i_1} - \cdots - x_{i_{n-1}} > 0$.

The *size-biased permutation* of x is a sequence of random variables $X = (X_1, X_2, \dots)$ 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 $\mathrm{SBP}\colon \mathcal{B}(\Delta^\infty_{\le 1}) \times \Delta^\infty_{\le 1} \to [0,1]$, where $\mathcal{B}(\Delta^\infty_{\le 1})$ stands for the Borel

σ-algebra of $\Delta^\infty_{\leq 1}$, 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 $\Delta^\infty_{\leq 1}$ and $x \in \Delta^\infty_{\leq 1}$. We are therefore able to construct the size-biased permutation $\mu\,\mathrm{SBP}$ (or μ^*) of any probability measure on μ on $\Delta^\infty_{\leq s}$ by setting

$$(\mu\,\mathrm{SBP})(B) := \int_{\Delta^\infty_{\leq 1}} \mu(dx)\,\mathrm{SBP}(B,x).$$

Let $X = (X_1, X_2, \dots) \in \Delta^\infty_s$ 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 134. 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 135. *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 136. *Let $X = (X_1, \dots, X_k) \in \Delta_s^{k-1}$ 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

$$\Delta^{k-r-1}_{s-x_1-\cdots-x_r} := \{(x_{r+1},\dots,x_k) \in \mathbb{R}^{k-r}_{\geq 0} : x_{r+1}+\cdots+x_k = 1-x_1-\cdots-x_r\}$$

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 135. We shall show that if the statement holds for $2 \leq r \leq k-1$, 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 137 ([DJ89, Theorem 3])**.** *Let $(\mu_n)_{n\geq 1}$ be a sequence of probability measures such that as $n\to\infty$, $(\mu_n)_{n\geq 1}$ 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 138. *Let $X^{(n)} \in \Delta_1^\infty$ be a sequence of random sequences, and let $\theta \in \mathbb{R}_{>0}$. Then the sequence of sorted sequence $X^{(n)\downarrow}$ converges in distribution to $\mathrm{PD}(\theta)$ if and only if the sequence of size-biased permutations $X^{(n)*}$ converges in distribution to $\mathrm{GEM}(\theta)$.*

The size-biased permutation and descending ordered statistics commute also with conditioning.

Lemma 139. *Let $(\mu_n)_{n\geq 1}$ be a sequence of probability measures, and $(a_n)_{n\geq 1}$ be a sequence of non-negative real numbers such that $a_1 + a_2 + \cdots = 1$. Then*

$$\sum_{n=1}^\infty a_n \mu_n^* = \left(\sum_{n=1}^\infty a_n \mu_n\right)^*, \qquad \sum_{n=1}^\infty a_n \mu_n^\downarrow = \left(\sum_{n=1}^\infty a_n \mu_n\right)^\downarrow.$$

6.3 Length partition of a random multicurve

The aim of this section is to state and prove a refinement of Theorem 130 that provides an explicit description of the random variable $L_{g,m}^\downarrow$.

Remark 140. The sample space (domain of definition) of the limit $L_{g,\gamma}$ of $\hat{\ell}_{X,\gamma,R}$ as $R \to \infty$ is not well-defined (the sequence of uniform probability measures on $s_X(\gamma, R) \hookrightarrow s_X(\gamma, \infty)$ converges to the zero measure). When we talk about the convergence of $\hat{\ell}_{X,\gamma,R}$, we talk about the convergence of the push-forward measures on Δ_1^{k-1} (or on Δ_1^∞), and 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 $\lambda\colon E \to \{1, \ldots, |E|\}$ 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 104 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 labelling λ, so from now on we shall write simply $\ell_{X,\gamma,R}^\downarrow$, $\ell_{X,\gamma,R}^*$, $L_{g,\gamma}^\downarrow$, $L_{g,\gamma}^*$.

Remark 141. The random vector $\hat{\ell}_{X,\gamma,R}^\downarrow$ can also be defined directly by

$$\hat{\ell}_{X,\gamma,R}^\downarrow : s_X(\gamma, R) \to \Delta_1^\infty, \qquad \alpha \mapsto \hat{\ell}_X^\downarrow(\alpha)$$

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 $[\mathrm{Stab}(\underline{\alpha}) : \mathrm{Stab}(\alpha)]$-to-one correspondence and the function $\ell_{X,\gamma,A}^\downarrow$ 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

$$s_X(R, A) := \{\alpha : [\alpha] \in A,\ \ell_X(\alpha) \leq R\}$$

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 104 concerning random multicurves of a fixed topological type to random multicurves with topological types in any given set.

Theorem 142. *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

$$\mathbb{E}(\hat{\ell}^{\downarrow}_{X,A,R}(h)) = \sum_{[\alpha]\in A} \frac{|s_X(R, \alpha)|}{|s_X(A, R)|} \cdot \mathbb{E}(\hat{\ell}^{\downarrow}_{X,\alpha,R}(h)).$$

It follows from Theorem 85 that there exists $c(\gamma) \in \mathbb{Q}_{>0}$ such that

$$\lim_{R\to\infty} \frac{|s_X(R, \alpha)|}{|s_X(R, A)|} = \lim_{R\to\infty} \frac{|s_X(R, \alpha)|}{|s_X(R)|} \frac{|s_X(R)|}{|s_X(R, A)|} = \frac{c(\alpha)}{b_{g,A}}.$$

By Theorem 104 and Theorem 137,

$$\lim_{R\to\infty} \mathbb{E}(\hat{\ell}^{\downarrow}_{X,\alpha,R}(h)) = \mathbb{E}(L^{\downarrow}_{g,\alpha}(h)).$$

Now the result follows from Lemma 139 and the following elementary lemma. □

Lemma 143. *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 of the length distribution

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 144. *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 $x_i^{d_i}$, $1 \le i \le k$, 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 104.

Lemma 145. *Let $\gamma = (m_1\gamma_1, \ldots, m_k\gamma_k)$ 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

$$\frac{(6g-7)!}{\mathcal{Y}_{\underline{m}}(\bar{F}_\gamma)} \frac{\bar{F}_\gamma(x_1/m_1, \ldots, x_k/m_k)}{m_1 \cdots m_k}$$

where $\underline{m} = (m_1, \ldots, m_k)$. In other words, the normalizing constant in Theorem 104 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 straightforward computation using Lemma 144 and the definition of $\mathcal{Y}_{\underline{m}}$,

$$\int_{\Delta_1^{k-1}} \bar{F}_\gamma(x_1/m_1, \ldots, x_k/m_k)\, \lambda(dx) = \frac{m_1 \cdots m_k}{(6g-7)!} \cdot \mathcal{Y}_{\underline{m}}(\bar{F}_\gamma).$$

The rest of the assertion follows from Theorem 97. □

As we shall see in the next section, the following example turns out to be the most important case for our purposes.

Example 146. Let $\gamma = (m_1\gamma_1, \dots, m_k\gamma_k)$ be non-separating. Then its graph polynomial, denoted by $F_{g,k}(x_1, \dots, x_k)$, is equal to

$$\frac{1}{2^{3g-3-k}} \sum_{\substack{(d_{1-},d_{1+},\dots,d_{k-},d_{k+})\in\mathbb{Z}^{2k}_{\geq 0}\\ d_{i-}+d_{i+}+\cdots+d_{k-}+d_{k+}=3g-3}} \frac{x_1^{2d_1+1}\cdots x_k^{2d_i+1}}{d_{1-}!\,d_{1+}!\cdots d_{k-}!\,d_{k+}!} \int_{\overline{\mathcal{M}}_{g,2k}} \psi_{1-}^{d_{1-}}\psi_{1+}^{d_{1+}}\cdots\psi_{k-}^{d_{k-}}\psi_{k+}^{d_{k+}} \tag{6.1}$$

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

$$\frac{1}{2^{3g-3}(6g-6)k!\,c(\gamma)} \cdot \sum_{\substack{(d_{1-},d_{1+},\dots,d_{k-},d_{k+})\in\mathbb{Z}^{2k}_{\geq 0}\\ d_{i-}+d_{i+}+\cdots+d_{k-}+d_{k+}=3g-3}} \frac{x_1^{2d_1+1}\cdots x_k^{2d_i+1}}{m_1^{2d_1+2}\cdots m_k^{2d_k+2}} \frac{1}{d_{-}!\,d_{1+}!\cdots d_{k-}!\,d_{k+}!} \int_{\overline{\mathcal{M}}_{g,2k}} \psi_{1-}^{d_{1-}}\psi_{1+}^{d_{1+}}\cdots\psi_{k-}^{d_{k-}}\psi_{k+}^{d_{k+}}$$

where $d_i := d_{i-} + d_{i+}$ for all $1 \leq i \leq k$.

Corollary 147. *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*

$$\frac{(6g-7)!}{\mathcal{Z}_m(\bar{F}_{(\gamma_1,\dots,\gamma_k)})} \sum_{\substack{(m_1,\dots,m_k)\in\mathbb{Z}^k_{\geq 1}\\ m_1,\dots,m_k\leq m}} \frac{\bar{F}_{(m_1\gamma_1,\dots,m_k\gamma_k)}(x_1/m_1, \dots, x_k/m_k)}{m_1\cdots m_k}$$

Proof. This follows from Lemma 139, Lemma 145, and Theorem 97. □

6.5 Reduction in the asymptotic regime

The random variable $L^{\downarrow}_{g,m}$ appearing in Theorem 130 is delicate to study because it involves a huge number of terms. Using Theorem 98 from [Agg21] and [DGZZ20c] 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 $m_1, \ldots, m_k \leq m$, 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 148. 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 [DGZZ20c] in the following form (which is a corollary of Theorem 100).

Theorem 149. *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 150. *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 149 and the boundedness of h. □

6.6 Proof of the main theorem

The aim of this section is to prove the following result.

Theorem 151. *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 131 from Theorem 151.

Proof of Theorem 131. By Theorem 150, the random variables $L^*_{g,m}$ and $\widetilde{L}^*_{g,m,\kappa}$ have the same limit distribution as $g \to \infty$. Hence by Theorem 151, $L^*_{g,m}$ converges in distribution towards GEM(1/2) as $g \to \infty$. Finally Corollary 138 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 $X = (X_1, X_2, \dots) \in \Delta^\infty_{\le 1}$ be a random sequence. Given $r \in \mathbb{Z}_{\ge 1}$ $p = (p_1, \dots, p_r) \in \mathbb{Z}^r_{\ge 0}$ an r-tuple, we define

$$M_p(X) \coloneqq \mathbb{E}\left((1 - X_1) \cdots (1 - X_1 - \cdots - X_{r-1}) \cdot X_1^{p_1} \cdots X_r^{p_r}\right)$$

if $r \ge 2$, and $M_p(X) \coloneqq \mathbb{E}(X_1^{p_1})$ if $r = 1$.

The M_p-moments of a GEM distribution are particularly simple to compute.

Lemma 152. *Let $X = (X_1, X_2, \dots) \in \Delta^\infty_{\le 1}$ be a random series and $p = (p_1, \dots, p_r) \in \mathbb{Z}^r_{\ge 0}$ be an r-tuple. If $X \sim \mathrm{GEM}(\theta)$, 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 81,

$$M_p(X) = \int_{\Delta^r_{\le 1}} \theta^r x_1^{p_1} \cdots x_r^{p_r} (1 - x_1 - \cdots - x_r)^{\theta - 1}\, dx = \theta^r \int_{\Delta^r_1} x_1^{p_1} \cdots x_r^{p_r} x_{r+1}^{\theta-1}\, \lambda(dx)$$

Now the result follows from Lemma 144. □

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 153. *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 $r \ge 1$ and any $p = (p_1, \dots, p_r) \in \mathbb{Z}^r_{\ge 0}$, $\lim_{n\to\infty} M_p(X^{(n)}) = M_p(X^{(\infty)})$.*

Proof. The infinite-dimensional cube $[0,1]^{\mathbb{Z}_{\ge 1}}$ is compact with respect to the product topology by Tychonoff's theorem. The subset $\Delta^\infty_{\le 1}$ of $[0,1]^\infty$ is closed, and therefore compact. Let $C(\Delta^\infty_{\le 1}, \mathbb{R})$ denote the algebra of real-valued continuous functions on $\Delta^\infty_{\le 1}$, 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 $r \ge 2$ and $p_1, \dots, p_r \ge 0$. The set S is a separating subset of $C(\Delta^\infty_{\le 1}, \mathbb{R})$. Now it follows from Stone–Weierstrass theorem that the subalgebra generated by S is dense in $C(\Delta^\infty_{\le 1}, \mathbb{R})$ 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 154. *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 $p = (p_1, \dots, p_r) \in \mathbb{Z}^r_{\geq 0}$. The moment $M_p(\widetilde{L}^*_{g,m,\kappa})$ is asymptotically equivalent to*

$$\frac{\sqrt{\pi(m+1)/m}}{2(6g-6)^{p_1+\cdots+p_r+r-1/2}} \sum_{k=r}^{\kappa \log(6g-6)/2} \frac{1}{(k-r)!} \sum_{\substack{(j_1,\dots,j_k)\in\mathbb{Z}^k_{\geq 1} \\ j_1+\cdots+j_k=3g-3}} \prod_{i=1}^{k} \frac{\zeta_m(2j_i)}{2j_i} \prod_{i=1}^{r} \frac{(2j_i+p_i)!}{(2j_i-1)!}$$

where

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

is the partial Riemann zeta function.

Following [DGZZ20c, Equation (14)], we define

$$c_{g,k}(d_1,\dots,d_k) := \frac{g!\,(3g-3+2k)!}{(6g-5+4k)!} \frac{3^g}{2^{3g-6+5k}} (2d_1+2)!\cdots(2d_k+2)! \cdot \sum_{\substack{(d_{1-},d_{1+},\dots,d_{k-},d_{k+})\in\mathbb{Z}^{2k}_{\geq 0} \\ d_{i-}+d_{i+}+\cdots+d_{k-}+d_{k+}=3g-3}} \frac{1}{d_-!\,d_{1+}!\cdots d_{k-}!\,d_{k+}!} \int_{\overline{\mathcal{M}}_{g,2k}} \psi_{1-}^{d_{1-}}\psi_{1+}^{d_{1+}}\cdots\psi_{k-}^{d_{k-}}\psi_{k+}^{d_{k+}}$$

and $\tilde{c}_{g,k}(j_1,\dots,j_k) = c_{g-k,k}(j_1-1,\dots,j_k-1)$. The above coefficients were introduced in [DGZZ20c, Lemma 3.5], and we have

$$\lim_{g\to\infty} \sup_{\substack{k\in\mathbb{Z}_{\geq 1} \\ k^2\leq g/800}} |c_{g,k}(d_1,\dots,d_k)-1| = 0$$

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 $m_1,\dots,m_k \leq m$. Let us write $\widetilde{L}^*_{g,m,(k)}$ for $L^*_{g,\widetilde{T}_{g,m,(k)}}$ defined in Section 6.3.

Lemma 155. *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 147, 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 135, the density function of $\widetilde{L}^*_{g,m,(k)}$ is

$$\begin{aligned}&\frac{w_{g,k}\cdot k!}{\mathcal{Z}_m(\bar{F}_{g,k})\cdot(k-r)!}\frac{1}{(1-x_1)\cdots(1-x_1-\cdots-x_{r-1})}\\ &\quad\cdot \sum_{\substack{(j_1,\dots,j_k)\in\mathbb{Z}^k_{\geq 1}\\ j_1+\cdots+j_k=3g-3}} \tilde{c}_{g,k}(j_1,\dots,j_k) \prod_{i=1}^{k} \zeta_m(2j_i)\, \frac{x_i^{2j_i-1}}{(2j_i)!} \int_{\Delta^{k-r-1}_{1-x_1-\cdots-x_r}} x_{r+1}^{2j_{r+1}-1}\cdots x_k^{2j_k-1}\lambda(dx)\end{aligned} \tag{6.5}$$

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 144 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 $\widetilde{L}^*_{g,m,(k)}$

$$\begin{aligned}&\frac{w_{g,k}}{\mathcal{Z}_m(\bar{F}_{g,k})\cdot(k-r)!}\\ &\quad\cdot \sum_{\substack{(j_1,\dots,j_k)\in\mathbb{Z}^k_{\geq 1}\\ j_1+\cdots+j_k=3g-3}} \tilde{c}_{g,k}(j_1,\dots,j_k)\, \frac{(1-x_1-\cdots-x_r)^{2j_{r+1}+\cdots+2j_k-1}}{(1-x_1)\cdots(1-x_1-\cdots-x_{r-1})} \prod_{i=1}^{k}\frac{\zeta_m(2j_i)}{2j_i}\prod_{i=1}^{r}\frac{x_i^{2j_i-1}}{(2j_i-1)!}.\end{aligned}$$

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 144, we obtain

$$\int_{\Delta^r_{\leq 1}} x_1^{2j_1+p_1}\cdots x_r^{2j_r+p_r}\cdot(1-x_1-\cdots-x_r)^{2j_{r+1}+\cdots+2j_k-1}\,dx$$
$$= \int_{\Delta^r_1} x_1^{2j_1+p_1}\cdots x_r^{2j_r+p_r}\cdot x_{r+1}^{2j_{r+1}+\cdots+2j_k-1}\,\lambda(dx)$$
$$= \frac{(2j_1+p_1)!\cdots(2j_r+p_r)!\cdot(2j_{r+1}+\cdots+2j_k-1)!}{(6g-6+p_1+\cdots+p_r+r-1)!} \tag{6.7}$$

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 154. By Theorem 142,

$$M_p(\widetilde{L}^*_{g,m,\kappa}) = \sum_{k=1}^{\kappa\log(6g-6)/2}\frac{\widetilde{b}_{g,m,(k)}}{\widetilde{b}_{g,m,\kappa}}\cdot M_p\big(\widetilde{L}^*_{g,m,(k)}\big) = \frac{1}{\widetilde{b}_{g,m,\kappa}}\sum_{k=1}^{\kappa\log(6g-6)/2}\frac{\mathcal{Z}_m(\bar{F}_{g,k})}{(6g-6)!\,2^k\,k!}\cdot M_p\big(\widetilde{L}^*_{g,m,(k)}\big)$$

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 155, 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, [DGZZ20c, 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 149 and the fact that $\tilde{c}_{g,k}(j_1,\dots,j_k)\sim 1$ uniformly in $1\leq k\leq\kappa\log(6g-$

6)/2, and replacing (6.9) and (6.10) in (6.8), we obtain

$$M_p(\widetilde{L}^*_{g,m,\kappa}) \sim \frac{\sqrt{\pi(m+1)/m}}{2(6g-6)^{p_1+\cdots+p_r+r}} \sum_{k=1}^{\kappa\log(6g-6)/2} \frac{1}{(k-r)!} \sum_{\substack{(j_1,\ldots,j_k)\in\mathbb{Z}^k_{\geq 1}\\ j_1+\cdots+j_k=3g-3}} \prod_{i=1}^{k} \frac{\zeta_m(2j_i)}{2j_i} \prod_{i=1}^{r} \frac{(2j_i+p_i)!}{(2j_i-1)!}$$

the result desired. □

6.6.2 Asymptotic expression of a related sum

Let $\theta = (\theta_n)_{n\geq 1}$ be a sequence of non-negative real numbers and let $p = (p_1, \ldots, p_r) \in \mathbb{Z}^r_{\geq 0}$. This section is dedicated to the asymptotics in n of the numbers

$$S_{\theta,p,n} := \sum_{k=r}^{\infty} \frac{1}{(k-r)!} \sum_{\substack{(j_1,\ldots,j_k)\in\mathbb{Z}^k_{\geq 1}\\ j_1+\cdots+j_k=n}} \prod_{i=1}^{k} \frac{\theta_{j_i}}{2j_i} \prod_{i=1}^{r} \frac{(2j_i+p_i)!}{(2j_i-1)!} \tag{6.11}$$

which should be reminiscent of the formula from Theorem 154.

Let $\theta = (\theta_n)_{n\geq 1}$ 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 $(\theta_n)_{n\geq 1}$ 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 156. *Let $\theta = (\theta_i)_{i\geq 1}$ 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 [DGZZ20c, Lemma 3.8]. Since the proof is short, we include it for the sake of completeness.

Lemma 157. *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

$$\left|\log\left(1-\frac{z}{n^2}\right)\right| \le \frac{2}{n^2}\,|z|$$

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

$$-\sum_{n=2}^{m}\log\left(1-\frac{1}{n^2}\right) = \sum_{n=2}^{m}\left(2\log(n)-\log(n-1)-\log(n+1)\right) = \log\left(\frac{2m}{m+1}\right).$$

This completes the proof. □

Corollary 158. *Let $m \in \mathbb{Z}_{\ge 1}\cup\{+\infty\}$. For $\theta = (\zeta_m(2j))_{j\ge 1}$ 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 $n\in\mathbb{Z}_{\ge 0}$, 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 159. *Let $\theta=(\theta_i)_{i\ge 1}$ 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 $n\ge 0$,

$$[z^{2n}]\,G_{\theta,p}(z) = S_{\theta,p,n}$$

where $[z^n]\,f(z)$ denotes the operation of extracting the coefficient of z^n in the formal power series $f(z)=\sum_{n\ge 0}a_n z^n$.

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 160. *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 156. 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 160 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 $\mathbb{D}(0;\sqrt{R}) \smallsetminus \{-1,1\}$. 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 84, 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 154 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 154.

Theorem 161. *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 162 ([DGZZ20c, 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 163. *Let $h(z)$ be a holomorphic function on $\mathbb{D}(0;R) \setminus \{\,]-R,-1] \cup [1,R[\,\}$ 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 $\mathbb{D}(0;R) \setminus \{\,]-R,-1] \cup [1,R[\,\}$, we

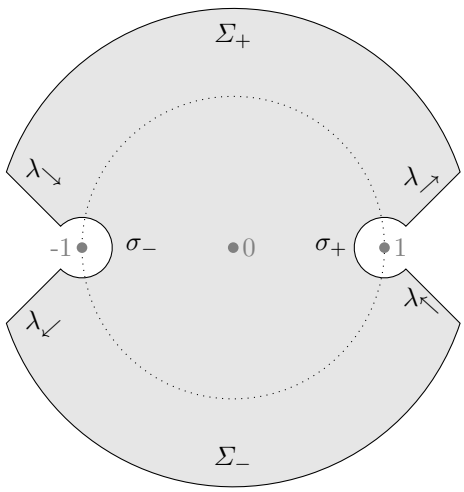

Figure 6.1: The contour γ.

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 162, uniformly in θ as $n\to\infty$ we have

$$\sum_{k=\lceil \kappa\log(n)/2\rceil}^{\infty} \frac{|h(z)|^k}{k!} \le \exp\left(-(\kappa\log\kappa-\kappa)\cdot\frac{\log n + \mathrm{O}(1)}{2}\right) = \mathrm{O}\left(n^{-\frac{1}{2}(\kappa\log\kappa-\kappa)}\right).$$

Since $1/|1-e^{i\theta}/n|^{n+1}$ is uniformly bounded in n,

$$\left|\frac{1}{2\pi i}\int_{\sigma_+} \frac{f_n(z)\, dz}{z^{n+1}}\right| = \mathrm{O}\left(n^{p-1-\frac{1}{2}(\kappa\log\kappa-\kappa)}\right).$$

Similarly,

$$\left|\frac{1}{2\pi i}\int_{\sigma_-} \frac{f_n(z)\, dz}{z^{n+1}}\right| = \mathrm{O}\left(n^{q-1-\frac{1}{2}(\kappa\log\kappa-\kappa)}\right).$$

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 162 gives

$$\sum_{k=\lfloor \kappa\log(n)/2\rfloor}^{\infty} \frac{|h(1+e^{i/\phi}t/n|^k}{k!} = \mathrm{O}\left(n^{-\frac{1}{2}(\kappa\log\kappa-\kappa)}\right).$$

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

$$\left|\frac{1}{2\pi i}\int_{\Sigma_+}\frac{f(z)dz}{z^{n+1}}\right| \le \frac{1}{2\pi}\cdot\eta^{p+1/2}\cdot(1+\eta)^{-n-1}\cdot 2\pi(1+\eta) = \frac{\eta^{p+1/2}}{(1+\eta)^{n+1}}$$

which decreases exponentially fast.

We conclude the proof by combining the above estimates. □

Proof of Theorem 161. Similarly to Lemma 159, 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 156 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 163 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 156. □

6.6.4 Proof of Theorem 151

Now we are ready for the proof of the main result of the section.

Proof of Theorem 151. By Lemma 153, 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 152.

Now, Theorem 154 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 158 and Theorem 161. 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 154 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 152 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 $p = (p_1,\dots,p_r)$, the sequence $\widetilde{L}^*_{g,m,\kappa}$ converges in distribution towards GEM(1/2). □

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 (English).

[ADG+20] Amol Aggarwal, Vincent Delecroix, Élise 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** (2020), no. 2, 149–161 (English).

[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** (2016), no. 6, 1311–1386 (English).

[Agg20] Amol Aggarwal, *Large genus asymptotics for volumes of strata of Abelian differentials*, J. Am. Math. Soc. **33** (2020), no. 4, 941–989 (English).

[Agg21] ———, *Large genus asymptotics for intersection numbers and principal strata volumes of quadratic differentials*, Invent. Math. **226** (2021), no. 3, 897–1010 (English).

[AH20a] Francisco Arana-Herrera, *Counting hyperbolic multi-geodesics with respect to the lengths of individual components*, 2020.

[AH20b] ———, *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** (2020), 81–107 (English).

[AH21a] ———, *Effective mapping class group dynamics iii: Counting filling closed curves on surfaces*, 2021.

[AH21b] ———, *Equidistribution of families of expanding horospheres on moduli spaces of hyperbolic surfaces*, Geom. Dedicata **210** (2021), 65–102 (English).

[AH22] ———, *Counting problems from the viewpoint of ergodic theory: from primitive integer points to simple closed curves*, 2022.

[Bil72] P. Billingsley, *On the distribution of large prime divisors*, Period. Math. Hung. **2** (1972), 283–289 (English).

[Bil95] Patrick Billingsley, *Probability and measure.*, 3rd ed. ed., Chichester: John Wiley & Sons Ltd., 1995 (English).

[BM04] Robert Brooks and Eran Makover, *Random construction of Riemann surfaces*, J. Differ. Geom. **68** (2004), no. 1, 121–157 (English).

[BS85] Joan S. Birman and Caroline Series, *Geodesics with bounded intersection number on surfaces are sparsely distributed*, Topology **24** (1985), 217–225 (English).

[Bus92] Peter Buser, *Geometry and spectra of compact Riemann surfaces*, Prog. Math., vol. 106, Boston, MA: Birkhäuser, 1992 (English).

[Cha11] Guillaume Chapuy, *A new combinatorial identity for unicellular maps, via a direct bijective approach*, Adv. Appl. Math. **47** (2011), no. 4, 874–893 (English).

[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** (2020), no. 1, 283–373 (English).

[DGZZ16] Vincent Delecroix, Élise Goujard, Peter Zograf, and Anton Zorich, *Square-tiled surfaces of fixed combinatorial type: equidistribution, counting, volumes of the ambient strata*, 2016.

[DGZZ20a] ———, *Contribution of one-cylinder square-tiled surfaces to Masur–Veech volumes*, Some aspects of the theory of dynamical systems: a tribute to Jean-Christophe Yoccoz. Volume I, Paris: Société Mathématique de France (SMF), 2020, pp. 223–274 (English).

[DGZZ20b] ———, *Enumeration of meanders and Masur–Veech volumes*, Forum Math. Pi **8** (2020), 80 (English), Id/No e4.

[DGZZ20c] ———, *Large genus asymptotic geometry of random square-tiled surfaces and of random multicurves*, 2020.

[DGZZ20d] ———, *Uniform lower bound for intersection numbers of ψ-classes*, SIGMA, Symmetry Integrability Geom. Methods Appl. **16** (2020), paper 086, 13 (English).

[DGZZ21] ———, *Masur–Veech volumes, frequencies of simple closed geodesics, and intersection numbers of moduli spaces of curves*, Duke Mathematical Journal **170** (2021), no. 12, 2633 – 2718.

[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** (1989), no. 1, 89–103.

[DL22] Vincent Delecroix and Mingkun Liu, *Length partition of random multicurves on large genus hyperbolic surfaces*, 2022.

[Do15] Norman Do, *Moduli spaces of hyperbolic surfaces and their Weil–Petersson volumes*, Handbook of moduli. Volume I, Somerville, MA: International Press; Beijing: Higher Education Press, 2015, pp. 217–258 (English).

[DVV91] Robbert Dijkgraaf, Herman Verlinde, and Erik Verlinde, *Loop equations and virasoro constraints in non-perturbative two-dimensional quantum gravity*, Nuclear Physics B **348** (1991), no. 3, 435–456.

[DZ20] Vincent Delecroix and Anton Zorich, *Cries and whispers in wind-tree forests*, What's next? The mathematical legacy of William P. Thurston, Princeton, NJ: Princeton University Press, 2020, pp. 83–115 (English).

[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** (2014), 207–333 (English).

[EM93] Alex Eskin and Curt McMullen, *Mixing, counting, and equidistribution in Lie groups*, Duke Mathematical Journal **71** (1993), no. 1, 181 – 209.

[EMM21] Alex Eskin, Maryam Mirzakhani, and Amir Mohammadi, *Effective counting of simple closed geodesics on hyperbolic surfaces*, Journal of the European Mathematical Society **24** (2021), no. 9, 3059–3108.

[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** (2003), 61–179 (English).

[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** (2001), no. 1, 59–103 (English).

[EO06] ———, *Pillowcases and quasimodular forms*, Algebraic geometry and number theory. In Honor of Vladimir Drinfeld's 50th birthday, Basel: Birkhäuser, 2006, pp. 1–25 (English).

[EOP08] Alex Eskin, Andrei Okounkov, and Rahul Pandharipande, *The theta characteristic of a branched covering*, Adv. Math. **217** (2008), no. 3, 873–888 (English).

[EPS20] Viveka Erlandsson, Hugo Parlier, and Juan Souto, *Counting curves, and the stable length of currents*, J. Eur. Math. Soc. (JEMS) **22** (2020), no. 6, 1675–1702 (English).

[ES16] Viveka Erlandsson and Juan Souto, *Counting curves in hyperbolic surfaces*, Geom. Funct. Anal. **26** (2016), no. 3, 729–777 (English).

[ES19] ———, *Mirzakhani's curve counting*, 2019.

[ES20] ———, *Counting curves on orbifolds*, 2020.

[ES22] ———, *Mirzakhani's curve counting and geodesic currents*, Prog. Math., vol. 345, Birkhäuser Cham, 2022 (English).

[Ewe72] W.J. Ewens, *The sampling theory of selectively neutral alleles*, Theoretical Population Biology **3** (1972), no. 1, 87–112.

[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** (1991), no. 08, 1385–1406.

[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** (2014), no. 3-4, 271–436 (English).

[FO90] Philippe Flajolet and Andrew Odlyzko, *Singularity analysis of generating functions*, SIAM J. Discrete Math. **3** (1990), no. 2, 216–240 (English).

[FS09] Philippe Flajolet and Robert Sedgewick, *Analytic combinatorics*, Cambridge: Cambridge University Press, 2009 (English).

[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** (2021), no. 1, 62–110 (English).

[GM20] Elise Goujard and Martin Möller, *Counting Feynman-like graphs: quasimodularity and Siegel–Veech weight*, J. Eur. Math. Soc. (JEMS) **22** (2020), no. 2, 365–412 (English).

[Gou15] Elise Goujard, *Siegel–Veech constants for strata of moduli spaces of quadratic differentials*, Geom. Funct. Anal. **25** (2015), no. 5, 1440–1492 (English).

[GPY11] Larry Guth, Hugo Parlier, and Robert Young, *Pants decompositions of random surfaces*, Geom. Funct. Anal. **21** (2011), no. 5, 1069–1090 (English).

[Gri79] R. C. Griffiths, *On the distribution of allele frequencies in a diffusion model*, Theor. Popul. Biol. **15** (1979), 140–158 (English).

[Gri88] ———, *On the distribution of points in a Poisson–Dirichlet process*, J. Appl. Probab. **25** (1988), no. 2, 336–345 (English).

[Hid21] Will Hide, *Spectral gap for weil–petersson random surfaces with cusps*, 2021.

[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** (1959), 1–26 (German).

[Hwa94] Hsien-Kuei Hwang, *Théorèmes limites pour les structures combinatoires et les fonctions arithmétiques*, Ph.D. thesis, 1994, Thèse de doctorat dirigée par Jean-Marc Steyaert, Sciences appliquées Palaiseau, Ecole polytechnique 1994, p. 1 vol. (260 p.).

[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*, Prog. Math., vol. 183, Boston, MA: Birkhäuser, 2001 (English).

[Kat92] Svetlana Katok, *Fuchsian groups*, Chicago: The University of Chicago Press, 1992 (English).

[Kin75] J. F. C. Kingman, *Random discrete distributions*, Journal of the Royal Statistical Society. Series B (Methodological) **37** (1975), no. 1, 1–22.

[Kin77] ———, *The population structure associated with the Ewens sampling formula*, Theor. Popul. Biol. **11** (1977), 274–283 (English).

[KL07] M. E. Kazarian and S. K. Lando, *An algebro-geometric proof of Witten's conjecture*, J. Am. Math. Soc. **20** (2007), no. 4, 1079–1089 (English).

[KN10] Emmanuel Kowalski and Ashkan Nikeghbali, *Mod-Poisson convergence in probability and number theory*, Int. Math. Res. Not. **2010** (2010), no. 18, 3549–3587 (English).

[Kon92] Maxim Kontsevich, *Intersection theory on the moduli space of curves and the matrix Airy function*, Communications in Mathematical Physics **147** (1992), no. 1, 1 – 23.

[LG19] Jean-François Le Gall, *Brownian geometry*, Jpn. J. Math. (3) **14** (2019), no. 2, 135–174 (English).

[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*, Encycl. Math. Sci., vol. 141, Berlin: Springer, 2004 (English).

[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 (English).

[Mas82] Howard Masur, *Interval exchange transformations and measured foliations*, Ann. Math. (2) **115** (1982), 169–200 (English).

[Mir07a] Maryam Mirzakhani, *Random hyperbolic surfaces and measured laminations*, In the tradition of Ahlfors-Bers, IV. Proceedings of the 3rd Ahlfors-Bers colloquium, Ann Arbor, MI, USA, May 19–22, 2005, Providence, RI: American Mathematical Society (AMS), 2007, pp. 179–198 (English).

[Mir07b] ———, *Simple geodesics and Weil–Petersson volumes of moduli spaces of bordered Riemann surfaces*, Invent. Math. **167** (2007), no. 1, 179–222 (English).

[Mir07c] ———, *Weil–Petersson volumes and intersection theory on the moduli space of curves*, J. Am. Math. Soc. **20** (2007), no. 1, 1–23 (English).

[Mir08a] ———, *Ergodic Theory of the Earthquake Flow*, International Mathematics Research Notices **2008** (2008), rnm116.

[Mir08b] ———, *Growth of the number of simple closed geodesies on hyperbolic surfaces*, Annals of Mathematics **168** (2008), no. 1, 97–125.

[Mir13] ———, *Growth of Weil–Petersson volumes and random hyperbolic surface of large genus*, J. Differ. Geom. **94** (2013), no. 2, 267–300 (English).

[Mir16] ———, *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 (2022).

[Mon20] Laura Monk, *Benjamini-schramm convergence and spectrum of random hyperbolic surfaces of high genus*, 2020.

[MP19] Mariam Mirzakhani and Bram Petri, *Lengths of closed geodesics on random surfaces of large genus*, Comment. Math. Helv. **94** (2019), no. 4, 869–889 (English).

[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** (1995), no. 2, 61–69 (English).

[MR95b] _____, *Simple curves on hyperbolic tori*, C. R. Acad. Sci., Paris, Sér. I **320** (1995), no. 12, 1523–1528 (English).

[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** (2019), 12 (English), Id/No 106878.

[MW02] Yair Minsky and Barak Weiss, *Nondivergence of horocyclic flows on moduli space*, J. Reine Angew. Math. **552** (2002), 131–177 (English).

[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** (2008), no. 3, 919–987 (English).

[Nor13] Paul Norbury, *String and dilaton equations for counting lattice points in the moduli space of curves*, Trans. Am. Math. Soc. **365** (2013), no. 4, 1687–1709 (English).

[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*, The Oxford handbook of random matrix theory, Oxford: Oxford University Press, 2011, pp. 510–533 (English).

[OP09] A. Okounkov and R. Pandharipande, *Gromov-Witten theory, Hurwitz numbers, and matrix models*, Algebraic geometry, Seattle 2005. Proceedings of the 2005 Summer Research Institute, Seattle, WA, USA, July 25–August 12, 2005, Providence, RI: American Mathematical Society (AMS), 2009, pp. 325–414 (English).

[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 (2021), 1–10.

[PY97] Jim Pitman and Marc Yor, *The two-parameter Poisson–Dirichlet distribution derived from a stable subordinator*, Ann. Probab. **25** (1997), no. 2, 855–900 (English).

[Riv01] Igor Rivin, *Simple curves on surfaces*, Geom. Dedicata **87** (2001), no. 1-3, 345–360 (English).

[RS19] Kasra Rafi and Juan Souto, *Geodesic currents and counting problems*, Geom. Funct. Anal. **29** (2019), no. 3, 871–889 (English).

[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** (2018), no. 6, 1756–1779 (English).

[Sau21] _____, *The large genus asymptotic expansion of Masur–Veech volumes*, Int. Math. Res. Not. **2021** (2021), no. 20, 15894–15910 (English).

[SL66] L. A. Shepp and S. P. Lloyd, *Ordered cycle lengths in a random permutation*, Trans. Am. Math. Soc. **121** (1966), 340–357 (English).

[Vee82] William A. Veech, *Gauss measures for transformations on the space of interval exchange maps*, Annals of Mathematics **115** (1982), no. 2, 201–242.

[VS77] A. M. Vershik and A. A. Shmidt, *Limit measures arising in the asymptotic theory of symmetric groups. I*, Theory Probab. Appl. **22** (1977), 70–85 (English).

[Wit91] Edward Witten, *Two-dimensional gravity and intersection theory on moduli space*, 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, Providence, RI: American Mathematical Society; Bethlehem, PA: Lehigh University, 1991, pp. 243–310 (English).

[Wol83] Scott Wolpert, *On the symplectic geometry of deformations of a hyperbolic surface*, Ann. Math. (2) **117** (1983), 207–234 (English).

[Wri15] Alex Wright, *Translation surfaces and their orbit closures: an introduction for a broad audience*, EMS Surv. Math. Sci. **2** (2015), no. 1, 63–108 (English).

[Wri18] _____, *Mirzakhani's work on earthquake flow*, 2018.

[Wri20] _____, *A tour through Mirzakhani's work on moduli spaces of Riemann surfaces*, Bull. Am. Math. Soc., New Ser. **57** (2020), no. 3, 359–408 (English).

[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** (2022), no. 2, 340–410 (English).

[Yak] Ivan Yakovlev, *Contribution of n-cylinder square-tiled surfaces to Masur–Veech volume of $\mathcal{H}(2g-2)$*, in preparation.

[Zog19] P. G. Zograf, *An explicit formula for Witten's 2-correlators*, J. Math. Sci., New York **240** (2019), no. 5, 535–538 (English).

[Zor02] Anton Zorich, *Square tiled surfaces and Teichmüller volumes of the moduli spaces of Abelian differentials.*, 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, Berlin: Springer, 2002, pp. 459–471 (English).

[Zvo12] Dimitri Zvonkine, *An introduction to moduli spaces of curves and their intersection theory*, Handbook of Teichmüller theory. Volume III, Zürich: European Mathematical Society (EMS), 2012, pp. 667–716 (English).