Leading order asymptotics of a multi-matrix partition function for colored triangulations

by

Enrique Acosta Jaramillo

A Dissertation Submitted to the Faculty of the

Department of Mathematics

In Partial Fulfillment of the Requirements
For the Degree of

Doctor of Philosophy

In the Graduate College

The University of Arizona

2 0 1 3

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Enrique Acosta Jaramillo entitled
Leading order asymptotics of a multi-matrix partition function for colored triangulations
and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

______________________________________ Date: April 11, 2013
Kenneth D. T-R McLaughlin

______________________________________ Date: April 11, 2013
Nicholas M. Ercolani

______________________________________ Date: April 11, 2013
Hermann Flaschka

______________________________________ Date: April 11, 2013
Thomas G. Kennedy

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

______________________________________ Date: April 11, 2013
Kenneth D. T-R McLaughlin

Statement by Author

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

____ Date: April 11, 2013

Enrique Acosta Jaramillo

Dedication

A mis papás, y a Verónica.

Acknowledgments

First and foremost, I would like to thank my advisor, Ken McLaughlin, for his guidance, support, advice, and overwhelming enthusiasm. Working on this project would have been very different without his precise and concrete explanations on background material, his guidance, and his advice on the way to approach the relevant literature. I feel incredibly fortunate of having been able to work with him these past couple of years.

I would also like to thank Nick Ercolani for very helpful discussions throughout this project that always helped me see things from a broad perspective. Nick always surprised me by relating distant areas of Mathematics to whatever we were discussing in an incredibly fruitful way. Thank you also to Tom Kennedy for all his advice, both as a dissertation committee member and as Head for the Graduate Program; and to Hermann Flaschka for those wonderful meetings, from which I would always leave having learned something new, and with at least one more item on my mathematical reading list.

I thank my beautiful wife Verónica for all her support during these years, and more importantly, for bringing balance to my life. I would not have made it without her by my side.

Thank you to my parents for their unconditional support.

Finally, thank you to all of my peers in the Graduate Program for their company. In particular, thank you to Chol, Danny, Joe, Martin, Michael, Michael, Patrick, Richard, Shane and Yaron. It has been great to share this experience with all of you!

Table of Contents

List of Figures **8**

List of Tables **10**

Abstract **11**

Chapter 1. Introduction **12**
- 1.1. The Gaussian Unitary Ensemble (GUE) 12
- 1.2. Random matrix theory and enumeration of maps 13
- 1.3. The partition function for colored triangulations 16
- 1.4. This dissertation 17

Chapter 2. Combinatorial interpretation of $\widehat{Z}_N(t)$ **20**
- 2.1. The genus expansion 20
- 2.2. Combinatorial interpretation of $\widehat{Z}_N(t)$ 23
- 2.3. Some examples 27

Chapter 3. Overview of the Physics Literature **29**
- 3.1. Background on saddle point approach for the leading term. 29
- 3.2. Physics literature on the partition function $\widehat{Z}_N(t)$. 35
- 3.3. Comments regarding the physics literature. 41

Chapter 4. Leading order asymptotics **44**
- 4.1. Setup and heuristics. 44
- 4.2. Regularity assumptions. 46
- 4.3. Statement of the theorems. 47
- 4.4. Existence of extremal measures. 50
- 4.5. The finite dimensional kernel for $Z^{V,H}_{(N)}$. 53
- 4.6. Introduction of the diagonal term. 59
- 4.7. End of proof of theorem 4.3.2 63
- 4.8. Large deviation estimates and proof of theorem 4.3.5 66

Chapter 5. Variational Inequalities **73**
- 5.1. Extremal measures for positive (semi-)definite kernels. 73
- 5.2. Variational inequalities for the kernels $K_{V,H}$ 78
- 5.3. Positive definiteness of $K_{V,H}$ with $H(x,y) = \sqrt{1+t^2(x+y)^2}$ 85
- 5.4. Some results regarding the support of the extremal density for the case $H(x,y) = \sqrt{1+t^2(x+y)^2}$ 89

Table of Contents—*Continued*

Chapter 6. Asymptotics for the partition function for colored triangulations **98**
6.1. Reduction to an N dimensional integral. 99
6.2. Overview of the strategy 102
6.3. The associated Riemann-Hilbert problem 104
6.4. Definition of ρ_t 109
6.5. Proof of proposition 6.3.6 113
6.6. Analyticity of ρ_t in t. 123
6.7. Proof of (6.2.3) 126
6.8. Expansion around $t = 0$ 128
6.9. The family of elliptic curves degenerating to a nodal curve as $t \to 0$. 129

Appendix A. Combinatorial interpretation expectation of powers of traces **131**
A.1. The case of $\mathrm{Tr} M^4$ 131
A.2. The general case 139
A.3. The relationship with maps 143

Appendix B. The dual of a Diagram. **148**

Appendix C. The genus expansion for $\widehat{Z}_N(t)$ **155**
C.1. The partition function 155
C.2. Relation to labeled colored diagrams. 156
C.3. The genus expansion in terms of colored tri-valent maps 161
C.4. Colored Triangulations. 166

References **169**

List of Figures

Figure 2.1. A star of type $q = M_{i_1} M_{i_2} \dots M_{i_k}$ . . . 20
Figure 2.2. A star of type $M_1^2 M_3 M_2$ . . . 21
Figure 2.3. The two types of stars. . . . 23
Figure 2.4. A map with one vertex of type q_1 and one vertex of type q_2 . . 24
Figure 2.5. All colored tri-valent maps with two vertices. . . . 26
Figure 2.6. Colored triangulation of a sphere with two triangles. . . . 27
Figure 2.7. Construction of a sphere with four triangles. . . . 28
Figure 2.8. Canonical position for case (a). . . . 28

Figure 3.1. The Schwarz-Christoffel map SC and its inverse Γ . . . 38

Figure 6.1. Definition of SC and Γ . . . 106
Figure 6.2. Image under SC of a rectangular grid in the upper half plane. . 107
Figure 6.3. Graphs of $\rho_t(x)$ for $x \geq 0$ and $t = 1$, $t = 0.5$, $t = 0.1$, $t = 0.05$. . 113
Figure 6.4. Pre-images of the vertices of the polygon. . . . 115
Figure 6.5. Graph of $T(b)$ for $b > 0$. . . . 120

Figure A.1. n four-valent vertices. . . . 135
Figure A.2. The connection of the half-edges. . . . 136
Figure A.3. The indices in the connection of two half edges. . . . 136
Figure A.4. The three four-valent diagrams with one vertex. . . . 137
Figure A.5. The second vertex of valence 4. . . . 140
Figure A.6. A diagram with two vertices of valence two and two vertices of valence three. . . . 141
Figure A.7. Collapsing a vertex of valence 4 into a star of type $q_4(M) = M^4$. 143
Figure A.8. The diagrams embedded in their corresponding surfaces. . . . 145
Figure A.9. The bijection between diagrams and maps. . . . 147

Figure B.1. Labeling of the edges of the dual of face F. . . . 148
Figure B.2. Edge in the original diagram separating faces F and F'. . . . 149
Figure B.3. Definition of the pairing in the dual diagram. . . . 150
Figure B.4. Construction of the dual edges . . . 151
Figure B.5. The dual of a vertex. . . . 151
Figure B.6. A diagram with one vertex of valence four. . . . 152
Figure B.7. The dual of the vertices of the diagram in figure B.6 . . . 152
Figure B.8. The dual diagram of the diagram in figure B.6 . . . 152
Figure B.9. Construction of the dual . . . 154

Figure C.1. The labeled vertices of a tri-valent diagram. . . . 156
Figure C.2. Representation of a tri-valent diagram. . . . 157
Figure C.3. The two types of colored tri-valent vertices. . . . 158

List of Figures—*Continued*

Figure C.4. The four colored tri-valent diagrams with two vertices. 159
Figure C.5. From a colored tri-valent map to a colored tri-valent diagram. . 162
Figure C.6. Triangulation of a sphere. 167
Figure C.7. A colored triangulation of a sphere with two triangles. 167

List of Tables

Table 6.1. Values of $\beta(t)$ 122

Abstract

We study the leading order asymptotics of a Random Matrix theory partition function related to colored triangulations. This partition function comes from a three Hermitian matrix model that has been introduced in the physics literature. We provide a detailed and precise description of the combinatorial objects that the partition function counts that has not appeared previously in the literature. We also provide a general framework for studying the leading order asymptotics of an N dimensional integral that one encounters studying the partition function of colored triangulations. The results are obtained by generalizing well know results for integrals coming from Hermitian matrix models with only one matrix that give the leading order asympti-otics in terms of a finite dimensional variational problem. We apply these results to the partition function for colored triangulations to show that the minimizing density of the variational problem is unique, and agrees with the one proposed in the physics literature. This provides the first complete mathematically rigorous description of the leading order asymptotics of this matrix model for colored triangulations.

Chapter 1
Introduction

1.1 The Gaussian Unitary Ensemble (GUE)

The space of $N \times N$ Hermitian matrices is in natural bijection with $\mathbb{R}^{N^2}$, since a Hermitian matrix $M = (m_{ij})$ is uniquely determined by its N real entries on the diagonal and the real and imaginary parts of its entries above the diagonal. This space, endowed with the probability measure

$$d\widetilde{\mu}_N(M) := \frac{1}{\widetilde{Z}_N^{GUE}} \exp\left\{-\frac{1}{2}\mathrm{Tr} M^2\right\} dM, \tag{1.1.1}$$

is called the **Gaussian Unitary Ensemble (GUE)**, where dM is Lebesgue measure in the independent entries of the matrix

$$dM := \prod_i dm_{ii} \prod_{i<j} d\left(\mathrm{Re}\, m_{ij}\right)\, d\left(\mathrm{Im}\, m_{ij}\right),$$

and $\widetilde{Z}_N^{GUE}$ is the normalizing constant that makes $\widetilde{\mu}_N$ a probability measure.

By explicitly writing $\mathrm{Tr} M^2$ in terms of the variables in dM, one can see that the probability measure $\widetilde{\mu}_N$ is simply the joint probability distribution of the independent and normally distributed N^2 variables showing up in dM, where all variables have mean zero, and variance equal to either 1 or 1/2 (depending on whether they are on or off the diagonal). Using this, one can explicitly compute the normalizing constant by computing the gaussian integral

$$\widetilde{Z}_N^{GUE} = \int_{\mathbb{R}^{N^2}} \exp\left\{-\frac{1}{2}\mathrm{Tr} M^2\right\} = 2^{N/2}\pi^{N^2/2}. \tag{1.1.2}$$

1.1.3 Rescaled GUE. It is know that the real eigenvalues of the matrices in GUE are statistically confined to the interval $[-2\sqrt{N}, 2\sqrt{N}]$, and so it is natural to rescale the matrices in GUE by $1/\sqrt{N}$. The corresponding space of rescaled matrices has probability measure given by

$$d\mu_N(M) := \frac{1}{Z_N^{GUE}} \exp\left\{-\frac{N}{2}\mathrm{Tr}M^2\right\} dM. \tag{1.1.4}$$

We call this space the **rescaled Gaussian Unitary Ensemble**, and we call the measure μ_N **rescaled GUE measure**. Here dM is again Lebesgue measure in the independent entries of the matrix

$$dM := \prod_i dm_{ii} \prod_{i<j} d\left(\mathrm{Re}\, m_{ij}\right)\, d\left(\mathrm{Im}\, m_{ij}\right),$$

and Z_N^{GUE} is the normalizing constant that makes μ_N a probability measure

$$Z_N^{GUE} := \int \exp\left\{-\frac{N}{2}\mathrm{Tr}M^2\right\} dM = \sqrt{2^N \pi^{N^2}/N^{N^2}}.$$

1.1.5 Induced measure on the space of eigenvalues. The measures $\widetilde{\mu}_N$ and μ_N induce measures on the space of eigenvalues of the corresponding matrices. The space of *unordered* eigenvalues $\lambda_1, \ldots, \lambda_N$, which is naturally in bijection with $\mathbb{R}^N$, has induced measures given by (see for example chapter 5 in [7])

$$d\widetilde{\mu}_N^{\mathrm{ev}}(\lambda) = \frac{1}{\widetilde{Z}_N^{\mathrm{ev}GUE}} e^{-\sum_{i=1}^N \lambda_i^2/2} \prod_{i<j} (\lambda_i - \lambda_j)^2 d\lambda_1 \ldots d\lambda_N,$$

in the GUE case, and

$$d\mu_N^{\mathrm{ev}}(\lambda) = \frac{1}{Z_N^{\mathrm{ev}GUE}} e^{-N\sum_{i=1}^N \lambda_i^2/2} \prod_{i<j} (\lambda_i - \lambda_j)^2 d\lambda_1 \ldots d\lambda_N,$$

in the rescaled GUE case. Here $\widetilde{Z}_N^{\mathrm{ev}GUE}$ and $Z_N^{\mathrm{ev}GUE}$ are the appropriate normalizing constants that make $\widetilde{\mu}_N^{\mathrm{ev}}$ and μ_N^{ev} probability measures.

1.2 Random matrix theory and enumeration of maps

1.2.1 The genus expansion. In their paper [5] from 1978, the physicists Brézin, Parisi, Itzykson and Zuber described a very concrete formal relation between integrals with respect to rescaled GUE and enumeration of maps on orientable surfaces, which are graphs embedded in surfaces in such a way that their complement in the surface is a union of simply connected sets. Their ideas resulted from specializations of ideas of 't Hooft [27], and have subsequently been analyzed, made mathematically precise, and exploited by both physicists and mathematicians.

This formal relation from [5] is now referred to as **the genus expansion**, and in its (generalized) modern form, as stated for example in [15], states that one has the formal expansion

$$\frac{1}{N^2}\log\int\exp\left\{N\,\mathrm{Tr}\left(\sum_{i=1}^{s}t_i\,q_i(M_1,\ldots,M_m)\right)\right\}d\mu_N(M_1)\ldots d\mu_N(M_m)$$

$$\text{“}=\text{”}\sum_{g\geq 0}\frac{1}{N^{2g}}e_g(t_1,\ldots,t_s), \tag{1.2.2}$$

where μ_N is rescaled GUE measure (1.1.4), the q_i are monomials in the Hermitian matrices $M_1,\ldots,M_m$, the t_i are complex numbers assumed to make the integral exist, and where the “ = ” symbol means that (formal) partial derivatives in the t's on both sides agree when setting $t_1=\ldots=t_n=0$. The e_g are the formal generating functions for counts of maps of genus g defined by

$$e_g(t_1,\ldots,t_s)\text{“}=\text{”}\sum_{n_1,\ldots,n_s\geq 0}\mathcal{M}_g[(q_1,n_1),\ldots,(q_s,n_s)]\frac{t_1^{n_1}\ldots t_s^{n_s}}{n_1!\ldots n_s!}, \tag{1.2.3}$$

where the coefficients $\mathcal{M}_g[(q_1,n_1),\ldots,(q_s,n_s)]\in\mathbb{N}$ are related to the number of maps on an orientable surface of genus g with n_i vertices, of type q_i. We will make this more precise in chapter 2 when we discuss the combinatorial interpretation of the integral we will be studying in this document.

1.2.4 The special case $m = 1$. The case the case $m=1$ with only one matrix $M_1=M$ was the original setting in which the genus expansion was introduced in [5]. In this setting, the term $\sum_{i=1}^{s}t_i\,q_i(M)$ is just a polynomial $P(M)$ with the t's as coefficients, which up to relabeling of the t's (and introduction of more t's if necessary), may be written as

$$P(M) := \sum_{i=1}^{s} t_i M^i,$$

so that the genus expansion (1.2.2) takes the from

$$\frac{1}{N^2} \log \int \exp\left\{ N \operatorname{Tr}\left(\sum_{i=1}^{s} t_i\, M^i \right) \right\} d\mu_N(M) \text{ `` = '' } \sum_{g\geq 0} \frac{1}{N^{2g}} e_g(t_1, \ldots, t_s). \tag{1.2.5}$$

One can reduce the integral in (1.2.5), as Brézin, Parisi, Itzykson and Zuber did in [5], to an integral over the eigenvalues $\lambda_1, \ldots, \lambda_N$ of M, obtaining (see for example [7])

$$\int_{\mathbb{R}^{N^2}} \exp\left\{ N \operatorname{Tr}\left(P(M) \right) \right\} d\mu_N(M) = \int_{\mathbb{R}^N} \exp\left\{ N \sum_{j=1}^{N} P(\lambda_j) \right\} d\mu_N^{\mathrm{ev}}(\lambda). \tag{1.2.6}$$

This representation allowed the authors of [5] to study the asymptotics of the integral as $N \to \infty$, which by (1.2.5) seem to be intimately related to combinatorial counts, and we discuss their ideas in more depth in chapter 3.

1.2.7 Mathematically precise interpretation of the genus expansion. We remark that the deduction of (1.2.2) involves several formal manipulations involving limits and reordering of series, and (1.2.2) it is not expected to be an actual equality in general. Among other things, there are various convergence issues on both sides of the identity.

Only until recently, Ercolani and McLaughlin in their paper [12] from 2003, provided an interpretation of the genus expansion that is not just formal for the case $m = 1$ with only one matrix (1.2.5). They proved that if the polynomial $P(M)$ has leading term of even degree, then for t's lying in a particular subset Ω of $\mathbb{R}^N$ guaranteeing, among other things, that the integral converges, one can interpret (1.2.5) as an asymptotic expansion in N as $N \to \infty$, meaning that if one truncates the right hand side at some $g = g_0$, then the difference between the two sides is $O(1/N^{2g_0+2})$

uniformly in Ω. They also proved that the generating functions e_g are in fact analytic in the t's in a neighborhood of $t = 0$, so that (1.2.3) is an actual Taylor series expansion. Their methods involved Riemann Hilbert methods and orthogonal polynomials, which heavily relied on the representation (1.2.6).

More recently, using a different collection of ideas, Guionnet and Maurel-Segala in [15, 16] have shown that the interpretation of the genus expansion as an asymptotic expansion also holds for arbitrary m as in (1.2.2) for at least for the first two terms e_0 and e_1 (meaning they are the coefficients of the first two terms of an asymptotic expansion), under specific hypothesis on the term $\sum_{i=1}^{s} t_i \, q_i(M_1, \ldots, M_m)$ and the t's (for example if $\sum_{i=1}^{s} t_i \, q_i(M_1, \ldots, M_m)$ is strictly convex and self-adjoint, and the t's are small enough).

1.2.8 The leading order asymptotics give planar counts. Note that the results described in 1.2.7 in particular imply that the limit as $N \to \infty$ of the integral on the left of (1.2.2) is in fact the generating function for maps on a sphere, as long as the appropriate assumptions on $\sum_{i=1}^{s} t_i \, q_i(M_1, \ldots, M_m)$ are satisfied.

1.3 The partition function for colored triangulations

In this dissertation we will be studying the asymptotics of the partition function

$$\widehat{Z}_N(t) := \iiint \exp\left\{it\, N \operatorname{Tr}\left(ABC + ACB\right)\right\} d\mu_N(A) d\mu_N(B) d\mu_N(C), \tag{1.3.1}$$

where $d\mu_N$ is rescaled GUE measure (1.1.4). We will show that the integral on the right exists for real t in chapter 6, and will mostly assume that $t \in \mathbb{R}$.

1.3.2 Combinatorial interpretation. This particular partition function (1.3.1) has been studied in the physics literature in [6, 18, 20, 21, 13] because of its relation to counts of colored triangulations, which are described in these works as triangulations of genus g surfaces where the each edge is colored with one of three available colors

$\mathcal{A}, \mathcal{B}, \mathcal{C}$, in such a way that all three colors show up on the edges of each triangle. This interpretation is based on a form of the genus expansion (1.2.2), but as is common in the physics literature, the precise nature of the counts, which can depend on the symmetries of the map in a complicated way, is not made precise.

1.3.3 Leading order asymptotics in the physics literature. The leading order asymptotics of the partition function $N^{-2} \log \widehat{Z}_N(t)$ was studied in [6, 18, 20, 21, 13] using ideas from [5] that we will outline in chapter 3. The applicability of ideas from [5] relies strongly on the fact that $\widehat{Z}_N(t)$ can be expressed as an N-dimensional integral over the eigenvalues of just one of the matrices

$$\widehat{Z}_N(t) = \int_{\mathbb{R}^N} \prod_{1 \le i,j \le N} \frac{1}{\sqrt{1 + t^2(\lambda_i + \lambda_j)^2}} d\mu_N^{\mathrm{ev}}(\lambda), \tag{1.3.4}$$

which in a sense connects this $\widehat{Z}_N(t)$ to the case $m = 1$ in (1.2.2), but is manifestly not of the from of the integral on the right (1.2.6).

The contents of [6, 18, 20, 21, 13] contain claims that the limit $\lim_{N\to\infty} N^{-2} \log \widehat{Z}_N(t)$ exists and is analytic at $t = 0$, and give a Taylor expansion of it around $t = 0$ (which by the genus expansion is expected to correspond to counts for colored triangulations on a sphere). This expansion is obtained following saddle-point heursitics from [5], which we will discuss in detail in chapter 3.

1.4 This dissertation

One of the main goals of this dissertation is to provide a mathematical framework which describes the leading order asymptotics of integrals of the form

$$Z_N^{V,H} := \int_{\mathbb{R}^N} e^{-N \sum V(x_i)} \prod_{i<j} (x_i - x_j)^2 \prod_{1 \le i,j \le N} \frac{1}{H(x_i, x_j)} d^N x,$$

and to use this to give complete mathematical proofs of the leading order asymptotics conjectured in the physics literature for $\widehat{Z}_N(t)$.

In chapter 4 we prove the existence of the limit $\lim_{N\to\infty} N^{-2}\log Z_N^{V,H}$ under general hypothesis on V and H by generalizing results from [19], and prove the equality

$$\lim_{N\to\infty}\frac{1}{N^2}\log Z_N^{V,H} = \inf_{\mu\in\mathcal{M}^1(\mathbb{R})}\iint\left[\log\frac{H(x,y)}{|x-y|}+\frac{1}{2}V(x)+\frac{1}{2}V(y)\right]d\mu(x)d\mu(y),$$

where the infimum is taken over all probability measures in $\mathbb{R}$. We will use this to prove the existence of the limit $\lim_{N\to\infty} N^{-2}\log\widehat{Z}_N(t)$, and the equality

$$\lim_{N\to\infty}\frac{1}{N^2}\log\widehat{Z}_N(t) = \inf_{\mu\in\mathcal{M}^1(\mathbb{R})} I_0[\mu] - \inf_{\mu\in\mathcal{M}^1(\mathbb{R})} I_t[\mu],$$

where

$$I_t[\mu] := \iint\left[\log\frac{\sqrt{1+t^2(x+y)^2}}{|x-y|}+\frac{1}{4}x^2+\frac{1}{4}y^2\right]d\mu(x)d\mu(y).$$

In chapter 5 we show that the kernel

$$\log\frac{\sqrt{1+t^2(x+y)^2}}{|x-y|}$$

showing up in I_t is *positive definite*, a notion we introduce there, and show that this guarantees the the minimizer of the functional I_t is unique. We provide criteria to characterize this minimizer by generalizing results from the theory of logarithmic potentials with external fields, which we use in chapter 6 to show that the density ρ_t proposed in the physics literature does indeed give the unique minimizer of the functional I_t. We also prove that under the assumption that V is a polynomial and that the unique minimizer

$$I_{t,V}[\mu] := \iint\left[\log\frac{\sqrt{1+t^2(x+y)^2}}{|x-y|}+\frac{1}{2}V(x)+\frac{1}{2}V(y)\right]d\mu(x)d\mu(y),$$

has a continuous density $\rho_{t,V}(x)$ satisfying some technical properties, then $\rho_{t,V}(x)$ will be supported on finitely many intervals, and will be analytic (in x) at any point in the interior of its support.

In chapter 6 we provide a complete construction of the density ρ_t using the constructions in the physics literature, and in the process prove analyticity in t of various quantities of interest for to the asymptotics of $\widehat{Z}_N(t)$. In particular, we prove that the endpoints of the support of the destiny ρ_t are analytic in t, and prove analyticity of $\rho_t(x)$ in t for all x in the interior of its support. The analytic dependence of ρ_t on the parameter t will be used to show that $\lim_{N\to\infty} N^{-2}\log \widehat{Z}_N(t)$ is analytic in t around $t = 0$. All of these results are analogous to results that are available for one-matrix models ($m = 1$), but that were not known in this situation.

Finally, as mentioned in 1.3.2, the precise nature of the counts of colored triangulations given by the partition function $\widehat{Z}_N(t)$ is not described precisely in the physics literature, and we devote chapter 2 to careful description of the combinatorial interpretation of $\log \widehat{Z}_N(t)$. In particular, we will give a simple labeling scheme for these colored triangulations which trivializes their "automorphism group", so that one does not need to worry about the symmetries of the triangulations.

Summarizing, in chapter 2 we will discuss the genus expansion and the combinatorial interpretation of $\log \widehat{Z}_N(t)$. In chapter 3 we provide a detailed overview of the physics literature relevant to $\widehat{Z}_N(t)$. In chapters 4 and 5 we discuss the asymptotics of integrals of the form $Z_N^{V,H}$, and study its associated variational problem. In chapter 6 we use the contents of chapters 4 and 5 and the constructions in the physics literature to give precise statements regarding the leading order asymptotics of $\log \widehat{Z}_N(t)$. In the appendices we describe in detail the connection between matrix integrals and combinatorics of maps, and give a complete proof of the genus expansion (as a formal identity) for $\log \widehat{Z}_N(t)$.

Chapter 2
Combinatorial interpretation of $\widehat{Z}_N(t)$

We will use the contents of chapter 7 in [15] to precisely describe the combinatorial interpretation of the partition function $\widehat{Z}_N(t)$. We start by giving a brief summary of the relevant results and definitions from [15].

2.1 The genus expansion

2.1.1 Stars. We let $\mathbb{C}\langle M_1, \ldots, M_m\rangle$ be the the ring of polynomials with non-commuting variables $M_1, \ldots M_m$. Given a monomial $q(M_1, \ldots, M_m) = M_{i_1} M_{i_2} \ldots M_{i_k}$ in the M's, we associate to it **a star of type** q, which by definition is a vertex with k labeled half edges with labels $M_{i_1}, M_{i_2}, \ldots, M_{i_k}$, an orientation, and one special marked half edge M_{i_1} corresponding to the first term of q. We represent these stars graphically as in figure 2.1, where the special marked half edge is the one that has a hollow dot, and the arrow indicates the orientation of the vertex. It is useful to think of the M's as colors.

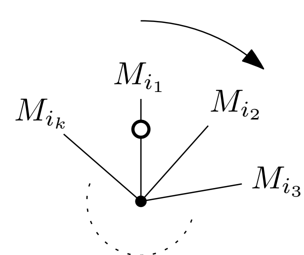

Figure 2.1. A star of type $q = M_{i_1} M_{i_2} \ldots M_{i_k}$

This association between monomials and stars is bijective by the fact that the special marked edge indicates what the first term of the corresponding monomial is, and the orientation specifies the ordering of the monomial. For example, the star in figure 2.2 is a star of type $M_1^2 M_3 M_2$.

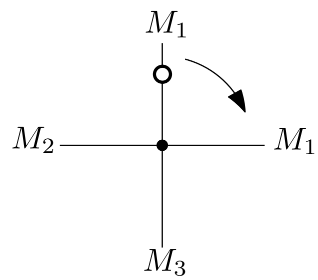

FIGURE 2.2. A star of type $M_1^2 M_3 M_2$

2.1.2 Definition. Given distinct monomials $q_1, q_2, \ldots, q_s$ in the variables $M_1, \ldots, M_m$, a **map of genus g with n_i stars of type q_i** for $i = 1, \ldots, s$ is a connected (multi)graph G embedded in an oriented surface of genus g with $\sum n_i$ vertices in such a way that:

- Each vertex of G has the structure of a star q_i for some $i = 1, \ldots, s$, where the orientation of the surface coincides with the orientation of each star.
- There are exactly n_i vertices in the graph G with the structure of a star of type q_i, and these vertices are labeled with the labels $v_1^{(i)}, v_2^{(i)}, \ldots, v_{n_i}^{(i)}$.
- The edges of G connect half edges of the stars with the same M-labels. In other words, of one thinks of the M's as colors, then the edges connect half edges with the same color, so one can think of the edges themselves having a color.
- The mutigraph G is the 1-skeleton of a CW-complex structure of the surface. In other words, the complement of G in the surface is the disjoint union of simply connected sets, which we call *faces*. The number of these faces is related to the genus g and the valences of the vertices by the Euler genus formula.

2.1.3 Equivalence of maps. We say that two maps of genus g with n_i stars of type q_i are **equivalent** if there is an orientation preserving homeomorphism of the surface that sends one map to the other that is compatible with the labels and the star types (in particular with the marking of the special half edge), and we define the number

$$\mathcal{M}_g[(q_1, n_1), \ldots, (q_s, n_s)] := \# \left\{ \begin{array}{c} \text{maps of genus } g \text{ with } n_i \text{ stars} \\ \text{of type } q_i \text{ up to equivalence} \end{array} \right\}.$$

We remark that this notion of equivalence does not agree with the notion *isotopy* (which identifies two maps that are smoothly deformable from one another) for $g \geq 1$ because of the existence of the so called *Dehn Twists*, which are orientation preserving homeomorphisms that result from the procedure of cutting a handle, making a complete twist on one of the sides, and gluing the handle back together. See [22, p.30] and references therein.

2.1.4 The genus expansion. The genus expansion, originally described by Brézin, Itzykson, Parisi and Zuber for $m = 1$ in [5], takes the following form in this setting.

2.1.5 Proposition (The Genus Expansion, [15] Lemma 7.12)**.** *One has the formal expansion*

$$\frac{1}{N^2} \log \int \exp \left\{ N \mathrm{Tr} \left(\sum_{i=1}^{s} t_i q_i(M_1, \ldots, M_m) \right) \right\} d\mu_N(M_1) \ldots d\mu_N(M_m)$$
$$\text{“} = \text{”} \sum_{g \geq 0} \frac{1}{N^{2g}} e_g(t_1, \ldots, t_s),$$

where μ_N is rescaled GUE measure as in (1.1.4), e_g is the formal generating function for maps of genus g defined by

$$e_g(t_1, \ldots, t_s) \text{“} = \text{”} \sum_{n_1, \ldots, n_s \geq 0} \mathcal{M}_g[(q_1, n_1), \ldots, (q_s, n_s)] \frac{t_1^{n_1} \ldots t_s^{n_s}}{n_1! \ldots n_s!},$$

and where the “ = ” symbol means that (formal) partial derivatives in the t's on both sides agree when setting $t_1 = \ldots = t_n = 0$.

2.1.6 Comments on the presentation in [15]. We remark that the description of the labels of the maps in [15] is not as precise as we have presented it 2.1.2. The labeling scheme we have described is compatible with the statements and proofs found there. We also remark that Guionnet in [15] incorrectly claims that equivalence of maps should be taken up to homeomorphism of the surface (at least for general term $\sum_{i=1}^{s} t_i q_i(M_1, \ldots, M_m)$). It is of fundamental importance that the homeomorphism be orientation preserving as we have stated in 2.1.3.

We present all the details of the genus expansion with a careful description of the labels in the particular case of the integral (1.3.1) we will be studying in appendix C. In appendix A we discuss the details of the connection between matrix integrals and combinatorial objects related to maps in a particularly simple situation.

2.2 Combinatorial interpretation of $\widehat{Z}_N(t)$

Using the form of the genus expansion given in 2.1.5, we are now ready to provide the combinatorial interpretation of the partition function $\widehat{Z}_N(t)$. We start by defining, the non-commuting monomials $q_1(A, B, C) = ABC$, $q_2(A, B, C) = ACB$ corresponding to the tri-valent stars shown in figure 2.3.

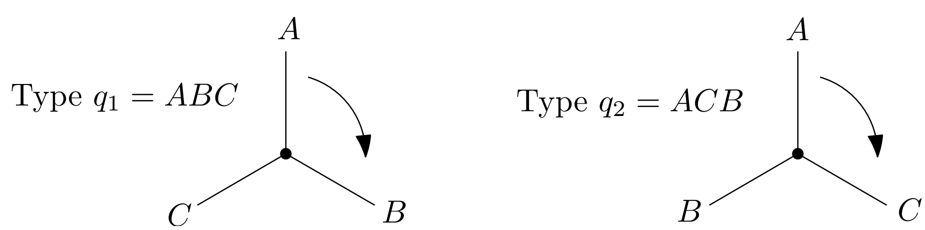

FIGURE 2.3. The two types of stars.

2.2.1 Removal of the special marking. We remark that we are removing the marking of the special half edge A from the stars described in 2.1.1, because it is unnecessary in this situation since these stars only have one half edge of type A.

2.2.2. By the genus expansion one has the formal identity

(2.2.3)
$$\frac{1}{N^2}\log\int \exp\left\{N\mathrm{Tr}\left(t_1q_1+t_2q_2\right)\right\}d\mu_N(A)d\mu_N(B)d\mu_N(C)\text{“}=\text{”}\sum_{g\geq 0}\frac{1}{N^{2g}}e_g(t_1,t_2)$$

where $d\mu_N$ is rescaled GUE as in (1.1.4) and

$$e_g(t_1,t_2):=\sum_{n_1,n_2\geq 0}\mathcal{M}_g[(q_1,n_1),(q_2,n_2)]\frac{t_1^{n_1}t_2^{n_2}}{n_1!n_2!},$$

is a formal generating function for maps with these two types of stars. For example, figure 2.4 shows the only map of genus 0 with one star of type q_1 and and one star of type q_2 up to equivalence. The orientation of the surface, which determines the types of the stars, is specified by the arrow. In particular, the coefficient of t_1t_2 in the expansion of $e_0(t_1,t_2)$ is 1.

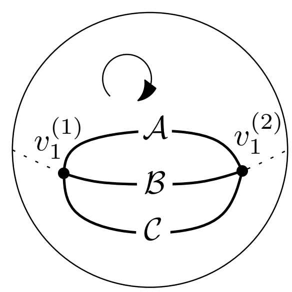

FIGURE 2.4. A map with one vertex of type q_1 and one vertex of type q_2

Note that the coefficients $\mathcal{M}_g[(q_1,n_1),(q_2,n_2)]$ are zero unless n_1+n_2 is even because all stars have valence three. If we now set $t_1=t_2=it$, then on the right side of (2.2.3) we get the partition function (1.3.1) and so (2.2.3) takes the form

$$\frac{1}{N^2}\log\widehat{Z}_N(t)\text{ “ = ” }\sum_{g\geq 0}\frac{1}{N^{2g}}e_g(t),$$

where

$$e_g(t):=\sum_{n\geq 0}\left(\sum_{n_1+n_2=2n}\binom{2n}{n_1}\mathcal{M}_g[(q_1,n_1),(q_2,n_2)]\right)\frac{(-1)^n}{(2n)!}t^{2n}.$$

2.2.4 Colored tri-valent maps. One can interpret the sum inside the parenthesis above as the number of maps of genus g with $2n$ vertices of type q_1 and q_2 with no restriction on the number of each type, and where labels for the vertices are taken from the set $\{1, \ldots, 2n\}$ (instead of the $v_i^{(j)}$), since the binomial coefficient is the number of ways to chose n_1 labels from the set $\{1, \ldots, 2n\}$ for the starts of type q_1.

We call these maps, with this particular labeling scheme for their vertices, *colored tri-valent maps*. Explicitly, a **colored tri-valent map with $2n$ vertices** is a graph embedded on a surface with $2n$ vertices labeled $1, \ldots, 2n$ where each vertex has the structure of a star of type either q_1 or q_2, where the complement of the graph in the surface is a union of simply connected sets, and where each edge has a color that agrees with the colors specified by the stars it connects.

In such a way, we obtain the following interpretation for $e_g(t)$:

$$e_g(t) = \sum_{n \geq 0} \left(\begin{array}{c} \text{number of colored tri-valent maps} \\ \text{of genus } g \text{ with } 2n \text{ stars of} \\ \text{types } q_1 \text{ or } q_2 \text{ up to equivalence} \end{array} \right) \frac{(-1)^n}{(2n)!} t^{2n}.$$

2.2.5. For example, there are two colored trivalent maps with genus zero, given by the two ways of labeling the vertices of the map in figure 2.4 with the labels 1 and 2. They are depicted in figure 2.5. Note that swapping the labels of the vertices gives a different colored tri-valent map because it changes their coloring scheme, corresponding to the type of star. This shows that the coefficient of $-t^2/2!$ in the expansion of $e_0(t)$ is 2.

Note that for colored tri-valent maps, changing the orientation of the surface has the effect of swapping the types of all the stars in the map.

2.2.6 Colored triangulations. Taking the dual of these colored tri-valent maps in the surface they are embedded in, as described briefly in [15], gives the interpretation for $e_g(t)$ as a formal generating function for colored triangulations. More precisely,

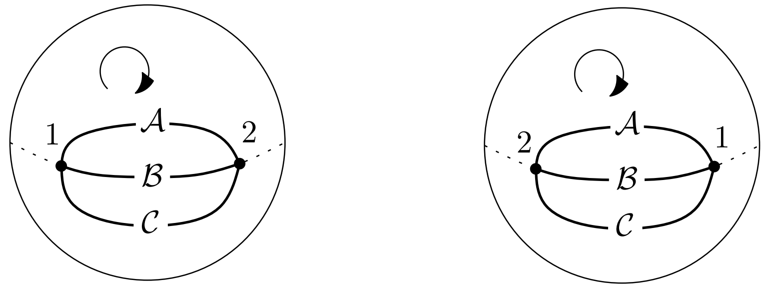

FIGURE 2.5. All colored tri-valent maps with two vertices.

we have

$$e_g(t) = \sum_{n\geq 0} \frac{(-1)^n}{(2n)!} \begin{pmatrix} \text{number of colored triangulations} \\ \text{with } 2n \text{ triangles on an orientable} \\ \text{surface of genus } g \text{ up to equivalence} \end{pmatrix} t^{2n}, \tag{2.2.7}$$

where by a **triangulation** we mean a multigraph (a graph that allows multiple edges between vertices) that is embedded inside a surface in such a way that the complement of the graph in the surface is a disjoint union of simply connected sets, and so that each of these sets has three distinct edges on its boundary. By a **colored triangulation** with $2n$ triangles we mean a triangulation of an oriented surface with $2n$ labeled triangles $1, 2, \ldots, 2n$ together with a coloring of each edge with one of the three colors $\mathcal{A}, \mathcal{B}, \mathcal{C}$ in such a way that each triangle has one edge of each of the three colors. We stress the fact that a colored triangulation consists not only of the multigraph with labels and colors, but also of a fixed orientation of the surface.

An example of a colored triangulation of a sphere using two triangles is shown in the figure 2.6, where the orientation on the surface is specified by the arrow. Note that in figure 2.6 the coloring scheme of the edges of triangle 1 following the orientation is $\mathcal{ABC}$, while for triangle 2 it is $\mathcal{ACB}$. These two schemes are the only two ways that the edges of a triangle can be colored using three distinct colors (up to cycling), and the coupling of the three matrices in the partition function $\widehat{Z}_N(t)$ has this form precisely to take into account for these two ways to color the edges.

We say that two triangulations are **equivalent** if there is an orientation preserving homeomorphism of the surface that takes vertices to vertices and edges to edges that

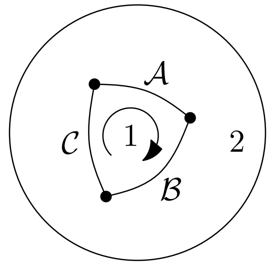

FIGURE 2.6. Colored triangulation of a sphere with two triangles.

is compatible with the colors and the labels.

2.2.8. We remark that the description of the counts in e_g do not involve unknown automorphism groups of the embedded graphs or triangulations. This precise description of the counts has not appeared previously in the literature

2.3 Some examples

2.3.1 The case $g = 0$ and $n = 1$ (two triangles). As follows by the discussion in 2.2.5 after taking the duals, there are only two colored triangulations on a sphere up to equivalence. One is depicted in figure 2.6, and the other one results from swapping the labels of the triangles.

2.3.2 The case $g = 0$ and $n = 2$ (four triangles). Ignoring colors and labels, there are two ways to make a sphere by glueing triangles together, where no two edges of the same triangle can be glued together (a condition that is necessary if one expects to be able to color their edges with three different colors). These two ways are depicted in figure 2.7, where the dots show the positions of the vertices of the triangles. Now, if we fix the orientation on these surfaces, then the orientation preserving homeomorphisms will correspond to rotations of these polyhedra in space.

In case (a), we can use these rotations to make sure that triangle 1 is the frontal face, and its edge with color $\mathcal{A}$ is the bottom one as depicted in figure 2.8 (the

FIGURE 2.7. Construction of a sphere with four triangles.

orientation is determined by the bent arrow in triangle 1). With triangle 1 and edge with color $\mathcal{A}$ in their fixed positions, there are six distinct ways to label the remaining triangles with labels $2, 3, 4$. Moreover, there are two ways to complete the coloring of the edges of triangle 1 (by choosing if it is a type I or II triangle), and each way can be seen to fix the coloring of all the remaining edges by the condition that each triangle needs to have the three colors in its edges. Thus, there are in total 12 colored triangulations of type (a).

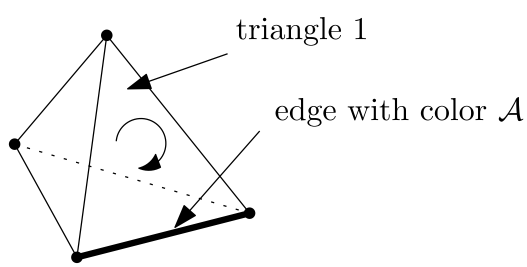

FIGURE 2.8. Canonical position for case (a).

Similar considerations show that there are 72 colored triangulations with four triangles of type (b), and so we see there are in total 84 colored triangulations with four triangles.

Chapter 3
Overview of the Physics Literature

3.1 Background on saddle point approach for the leading term.

The genus expansion, which we discussed in sections 1.2 and 2.1 establishes a surprising connection between random matrix theory and maps on surfaces. This connection was described in its modern from for the first time in the highly influential paper [5] by physicists Brézin, Itzykson, Parisi, and Zuber. In [5], the authors not only describe the genus expansion, but also describe a heuristc approach to obtain the leading order asymptotics for the integrals as the size of the matrices grows, related to counts of maps on a sphere.

In this section we present a brief summary of their ideas regarding the asymptotics of the integrals, since the heuristic arguments that physicists have for the partition function of colored triangulations, which we discuss in section 3.2, are entirely based in their heuristics.

We present their ideas, as the authors did themselves in [5] for the sake of concreteness, for the particular case of 4-valent planar maps.

3.1.1 The genus expansion for the case of one matrix and pure valence 4. If we define

$$\widehat{Z}_N^{[4]}(t) := \int_{\mathbb{R}^{N^2}} \exp\left\{-N \operatorname{Tr}\left[tM^4\right]\right\} d\mu_N(M),$$

then by the genus expansion $\widehat{Z}_N^{[4]}(t)$ is the formal generating function counting maps with stars of type M^4 (in other words, vertices of valence 4 with a special marking on one of their half edges)

$$\frac{1}{N^2}\log \widehat{Z}_N^{[4]}(t) \text{``}=\text{''} e_0^{[4]}(t) + \frac{e_1^{[4]}(t)}{N^2} + \frac{e_2^{[4]}(t)}{N^4} + \dots \tag{3.1.2}$$

where $e_g(t)$ is defined by

$$e_g^{[4]}(t) := \sum_{n\geq 1} \frac{(-1)^{n+1}}{n!} \begin{pmatrix} \text{number of maps with } n \\ \text{vertices of valence 4 on a} \\ \text{surface of genus } g \text{ up} \\ \text{to equivalence} \end{pmatrix} t^n. \tag{3.1.3}$$

3.1.4. As in (1.2.6), one can reduce to an integral over the eigenvalues of the matrix

$$\widehat{Z}_N^{[4]}(t) = \int_{\mathbb{R}^N} \exp\left\{-N \sum_{j=1}^N t\lambda_j^4\right\} d\mu_N^{\text{ev}}(\lambda),$$

which we write as

$$\widehat{Z}_N^{[4]}(t) = \frac{Z_N^{[4]}(t)}{Z_N^{[4]}(0)},$$

where

$$Z_N^{[4]}(t) := \int \exp\left\{-N \sum_i V_t(x_i)\right\} \prod_{i<j} (x_i - x_j)^2 d^N x,$$

and

$$V_t(y) := \frac{y^2}{2} + ty^4.$$

3.1.5 Heuristics for leading order asymptotics. By writing

$$Z_N^{[4]}(t) = \int \exp\left[-N^2 R_{N,t}(x)\right] d^N x,$$

where

$$R_{N,t}(x) := \frac{1}{N} \sum_i V_t^{[4]}(x_i) + \frac{1}{N^2} \sum_{i\neq j} \log \frac{1}{|x_i - x_j|},$$

the authors of [5] claim that the main contribution to $N^{-2} \log Z_N^{[4]}(t)$ for large N should be given by

$$-\frac{1}{N^2} \log Z_N^{[4]}(t) \underset{N\to\infty}{\simeq} R_{N,t}(x^*(t)) \tag{3.1.6}$$

where $x^*(t) = (x_1^*(t), x_2^*(t), \ldots, x_N^*(t))$ minimize $R_{N,t}(x)$, and so should satisfy the equations

$$V_t'(x_i^*(t)) = \frac{1}{N} \sum_{j=1}^{N}{}^{(i)} \frac{1}{x_i^*(t) - x_j^*(t)}, \tag{3.1.7}$$

for $i = 1, \ldots, n$, where the (i) on the sum means it avoids the value $j = i$.

Combining this with (3.1.2), the authors of [5] conclude that the generating function for genus zero maps should be given by

$$e_0^{[4]}(t) \text{``}=\text{''} \lim_{N\to\infty} -R_{N,t}(x^*(t)).$$

3.1.8 The saddle-point equation. The authors of [5] then consider the associated continuous problem, where they assume that the coordinates of $x^*(t)$ have a limiting density $\rho_t^{[4]}(x)$ as $N \to \infty$. More explicitly, they assume that the measure

$$\frac{1}{N}\sum_{i=1}^{N} \delta_{x_i^*(t)}$$

converges (in some unspecified sense) to the measure with continuous density $\rho_t^{[4]}(x)$.

Under these assumptions, (3.1.6) can be rewritten as

$$(3.1.9)\quad -\frac{1}{N^2}\log Z_N^{[4]}(t) \underset{N\to\infty}{\simeq} \int V_t(x)\rho_t^{[4]}(x)dx + \iint \left(\log\frac{1}{|x-y|}\right)\rho_t^{[4]}(x)\rho_t^{[4]}(y)dxdy$$

while (3.1.7) takes the form

$$(3.1.10)\qquad V_t'(x) = p.v.\int_{\mathbb{R}} \frac{\rho_t^{[4]}(y)}{x-y}dy$$

for $x \in \operatorname{supp} \rho_t^{[4]}$. This last equation they call the **saddle-point equation** for $\rho_t^{[4]}$.

3.1.11 The Sokhotski-Plemelj formulas. Under the assumption that the density is sufficiently well behaved (e.g., if it is Hölder continuous), one can relate the principal value integral in (3.1.10) to the unknown density $\rho_t^{[4]}$ explicitly through the Sokhotski-Plemelj formulas (see for example [14]), which state that the function

$$W_t^{[4]}(z) := \int_{\mathbb{R}} \frac{\rho_t^{[4]}(y)}{z-y}dy,$$

defined for z in the complement of the support of $\rho_t^{[4]}$ satisfies

$$(3.1.12)\qquad W_t^{[4]}(x \pm i0) = p.v.\int_{\mathbb{R}} \frac{\rho_t^{[4]}(y)}{x-y}dy \mp i\pi\rho_t^{[4]}(x), \quad x \in \operatorname{supp} \rho_t^{[4]},$$

where we have used the notation

$$f(x \pm i0) := \lim_{\varepsilon \downarrow 0} f(x \pm i\varepsilon)$$

for the boundary values of a function.

3.1.13 The function $W_t^{[4]}$. The authors then assume that $\rho_t^{[4]}$ is continuous, is sufficiently well behaved so that the Sokhotski-Plemelj formulas are applicable, and has compact support $[-\beta_t, \beta_t]$ where β_t depends on some unknown way on t. It then follows that the complex valued function $W_t^{[4]}$ satisfies the following properties:

1. $W_t^{[4]}(z)$ is analytic outside $[-\beta_t, \beta_t]$.
2. $W_t^{[4]}$ has the expansion
$$W_t^{[4]}(z) = \frac{1}{z} + O\left(\frac{1}{z^2}\right)$$
around $z = \infty$, since $\int \rho_t^{[4]} = 1$, as it is a probability density.
3. The jump discontinuity along $[-\beta_t, \beta_t]$ satisfies (see (3.1.10) and (3.1.12))
$$W_t^{[4]}(x - i0) + W_t^{[4]}(x + i0) = 2V_t'(x) = 2x + 8tx^3, \quad x \in [-\beta_t, \beta_t], \tag{3.1.14}$$
(here they are implicitly assuming the boundary values are finite).

The authors then exhibit a function that satisfies the above conditions, given by (see for example section 6.7 in [7])

$$W_t^{[4]}(z) = \frac{1}{2}V_t'(z) - \sqrt{z^2 - \beta(t)^2}\left(\frac{1}{2} + 2tz^2 + t\beta(t)^2\right),$$

where the square root is positive for $z > \beta(t)$, and

$$\beta_t = \beta(t),$$

is the analytic branch of β as a function of t defined by the equation

$$3t\beta^4 + \beta^2 - 4 = 0,$$

that is positive for $t \geq 0$ (this last condition comes from enforcing coefficient $1/z$ in expansion of $W_t^{[4]}$ around $z = \infty$ to be 1).

3.1.15 The density $\rho_t^{[4]}$ and explicit description of $e_0^{[4]}(t)$. Once they have this candidate function $W_t^{[4]}$, the authors of [5] use (3.1.12) once more to obtain

$$\rho_t^{[4]}(x) = \frac{1}{\pi}\sqrt{\beta(t)^2 - x^2}\left(\frac{1}{2} + 2tx^2 + t\beta(t)\right), \tag{3.1.16}$$

which gives them a density that is expected to make the asymptotics (3.1.9) hold.

Using this, they found

$$\begin{aligned} e_0^{[4]}(t) &= \frac{1}{384}\left(\beta(t)^2 - 4\right)\left(36 - \beta(t)^2\right) - \log\left(\frac{\beta(t)}{2}\right) \\ &= \sum_{n=1}(-1)^{n+1}\frac{12^n(2n-1)!}{n!(n+2)!}t^n \\ &= 2t - 18t^2 + 288t^3 + \dots, \end{aligned} \tag{3.1.17}$$

and then verified that the first terms of the expansion do in fact agree with counts of genus zero maps (3.1.3). Subsequent analysis by various authors then showed that (3.1.17) does indeed give the generating function for genus zero maps (3.1.3). We will elaborate on this below.

3.1.18 Comments. We remark that the above heuristic arguments relied not only on the assumption that the asymptotics in (3.1.9) hold, but also on the fact that the genus expansion (3.1.2) is more that just formal, in the sense that

$$\lim_{N\to\infty}\frac{1}{N^2}\log\widehat{Z}_N^{[4]}(t) = e_0^{[4]}(t)$$

is an actual equality.

As expected from the correctness of the first terms of the expansion of $e_0^{[4]}$ in (3.1.17), many of the results and methods described in [5] have subsequently been justified (some proofs were supplied by the authors of [5] themselves). In particular, there are now mathematical proofs of the following facts:

- The limit $\lim_{N\to\infty} -N^{-2}\log Z_N^{[4]}(t)$ does exist for $t > 0$.

- The limit is the minimizer of the functional

$$I_t^{[4]}[\sigma] := \int V_t^{[4]}(x)d\sigma(x) + \iint \left(\log\frac{1}{|x-y|}\right) d\sigma(x)d\sigma(y)$$

  over the space of all Borel probability measures on $\mathbb{R}$.

- This minimizer is unique for each t, and is given by $\rho_t^{[4]}(x)dx$ with $\rho_t^{[4]}$ as in (3.1.16) as correctly claimed in [5].

- Any weak limit of the measures $N^{-1}\sum \delta_{x_i^*(t)}$, where the $x^*(t)$ satisfy (3.1.7) converges weakly to $\rho_t^{[4]}(x)dx$, so $\rho_t^{[4]}(x)$ is in fact the continuous version of the minimizers $x^*(t)$.

- The functions $e_g(t)$ are analytic in a neighborhood of zero, and the formal identity

$$\frac{1}{N^2}\log \widehat{Z}_N^{[4]}(t) \text{``}=\text{''} e_0^{[4]}(t) + \frac{1}{N^2}e_1^{[4]}(t) + \frac{1}{N^4}e_2^{[4]}(t) + \dots$$

  can be interpreted as an asymptotic expansion in N in the sense that there exists $t_0 > 0$ and an $N_0 > 0$ such that for all $G \geq 0$ there exists a constant C_G such that

$$\left|\frac{1}{N^2}\log \widehat{Z}_N^{[4]}(t) - \left(e_0^{[4]}(t) + \frac{e_1^{[4]}(t)}{N^2} + \dots + \frac{e_G^{[4]}(t)}{N^{2G}}\right)\right| < \frac{C_G}{N^{2G+2}}$$

  for all $t \in [0, t_0]$ and $N \geq N_0$, and similar bounds hold for derivatives in t. This in particular implies that

$$\lim_{N\to\infty} \frac{1}{N^2}\log \widehat{Z}_N^{[4]}(t) = e_0^{[4]}(t)$$

  is really the generating function for 4-valent genus zero diagrams.

- All of the above can be appropriately generalized to more general V of the form

$$V(x; t_1, \dots, t_{2n}) = \frac{1}{2}x^2 + t_1 x + t_2 x^2 + \dots t_{2n}x^{2n}$$

giving counts for arbitrary number of vertices of valences up to $2n$, and under further modifications, to polynomials with leading terms with odd exponents.

Most of the proofs of these statements can be found in the wonderful book [7] by Percy Deift. The statement regarding the interpretation of the genus expansion as an asymptotic expansion was proved in 2003 by Ercolani and McLaughlin in [12].

3.1.19 Comment regarding the saddle-point equation. We remark that for general V, the analogue condition (3.1.10) is necessary, but not sufficient to characterize the minimizer. This will be discussed in more detail in Chapter 5.

3.2 Physics literature on the partition function $\widehat{Z}_N(t)$.

The large N asymptotics of the partition function $N^{-2}\log \widehat{Z}_N(t)$ was studied in the physics literature in [6, 18, 20, 21, 13] because of its relation to colored triangulations, even though no precise description of the "automorphism group" of these triangulations is given. We present here a short summary of the heuristic arguments and results from [6, 18, 20, 21], which we will subsequently expand in chapter 6. We will refer to the authors of these papers in this section as *the authors*, but we remark that the contents of each of these papers differs from what is presented here, and this a summary of their ideas regarding the partition function $\widehat{Z}_N(t)$, instead of the contents of their papers. We will discuss the differences between the contents of their papers and what is presented here in section 3.3.

The key fact about $\widehat{Z}_N(t)$ that has been exploited in the physics literature is that one can write $\widehat{Z}_N(t)$ as an integral over the eigenvalues of one of the matrices, given by

$$\widehat{Z}_N(t) = \int_{\mathbb{R}^N} \prod_{1\le i,j\le N} \frac{1}{\sqrt{1+t^2(\lambda_i+\lambda_j)^2}} d\mu_N^{\text{ev}}(\lambda),$$

as we stated in the introduction (1.3.4). We supply the details of this computation in chapter 6, which as a by-product shows that $\widehat{Z}_N(t)$ exists for $t \in \mathbb{R}$. By defining

$$Y_N(t) := \int_{\mathbb{R}^N} e^{-N\sum \lambda_i^2/2} \prod_{i<j} (\lambda_i - \lambda_j)^2 \prod_{i,j} \frac{1}{\sqrt{1+t^2(\lambda_i+\lambda_j)^2}} d^N\lambda.$$

one can further write

$$\widehat{Z}_N(t) = \frac{Y_N(t)}{Y_N(0)}.$$

In this way, the counting problem for colored triangulations is recast into the framework of matrix integrals with a single matrix, similar to the one discussed in section 3.1. Even though $Y_N(t)$ is not of the form of the integrals discussed there because of the extra product $\prod_{i,j}$ in the integrand, the fact that $Y_N(0) = Z_N^{[4]}(0)$ is in both cases Z_N^{GUE} as defined in 1.1.3, suggests that the same sort of approach may be applicable, and this is what the authors exploit.

3.2.1 Assumption that $t \in \mathbb{R}$ and $t > 0$. From now on we will assume that $t \in \mathbb{R}$ (which is not always the case in the papers [6, 18, 20, 21, 13]), which guarantees that the quantities above are defined and the equalities hold. When needed below, we will further assume that $t > 0$.

3.2.2. Following the same collection of ideas in [5] that we summarized in section 3.1, the authors assume that

$$\frac{1}{N^2}\log Y_N(t) \underset{N\to\infty}{\simeq} \int \frac{1}{2}x^2\rho_t(x)dx + \iint \left(\log \frac{\sqrt{1+t^2(x+y)^2}}{|x-y|}\right)\rho_t(x)\rho_t(y)dxdy,$$

where the probability density $\rho_t(x)$ satisfies the *saddle-point equation*

$$x = p.v. \int \frac{2\rho_t(y)}{x-y}dy - \int \left(\frac{t}{t(x+y)+i} + \frac{t}{t(x+y)-i}\right)\rho_t(y)dy \tag{3.2.3}$$

for $x \in \operatorname{supp}\rho_t$, which is the analogue of (3.1.10).

They further assume that ρ_t is continuous, even, and is supported on a single interval $[-\beta_t, \beta_t]$ where β_t depends on some unknown way on t, and the write (3.2.3) in terms of

$$W_t(z) := \int_{\mathbb{R}} \frac{\rho_t(y)}{z-y}dy,$$

as (using that W_t is odd since ρ_t is even)

$$x = W_t(x+i0) + W_t(x-i0) - W_t\left(x - \frac{i}{t}\right) - W_t\left(x + \frac{i}{t}\right), \tag{3.2.4}$$

for $x \in [-\beta_t, \beta_t]$, which is the analogue of (3.1.14) above. Again, $W_t(z)$ is assumed to have finite boundary values on $[-\beta_t, \beta_t]$, and the fact that ρ_t is a probability density implies that

$$W_t(z) = \frac{1}{z} + \frac{\mathfrak{m}_1(t)}{z^2} + \frac{\mathfrak{m}_2(t)}{z^3} + \dots$$

around $z = \infty$, where

$$\mathfrak{m}_i(t) := \int x^i \rho_t(x) dx.$$

The main goal is again to try to identify the function W_t from its properties, and in this way recover the unknown density ρ_t. This is completely analogous to what was described in section 3.1, the only difference being that the equations are somewhat more involved. In particular, equation (3.2.4) is degenerate for $t = 0$, and from now on we assume that $t > 0$.

The authors then define the function

$$\zeta_t(z) := z^2 + \frac{2i}{t}\left(W_t\left(z + \frac{i}{2t}\right) - W_t\left(z - \frac{i}{2t}\right)\right),$$

which by the assumptions on W_t is analytic outside the two cuts $\pm i/2t + [-\beta_t, \beta_t]$ and has finite extensions to the cuts. The motivation behind the definition of ζ_t is that the saddle-point equation (3.2.4) is equivalent to

$$\zeta_t\left(x + \frac{i}{2t} \pm i0\right) = \zeta_t\left(x - \frac{i}{2t} \mp i0\right), \quad x \in [-\beta_t, \beta_t], \tag{3.2.5}$$

and since W_t is odd (again, since ρ_t is assumed to be even), one can further check that

$$\begin{aligned} \zeta_t(-z) &= \zeta_t(z), \\ \zeta_t(\overline{z}) &= \overline{\zeta_t(z)}. \end{aligned} \tag{3.2.6}$$

These symmetries imply that ζ_t is real valued on $\mathbb{R}$ and $i\mathbb{R}$, and together with (3.2.5) imply that boundary values of ζ_t along the two cuts $\pm i/2t+[-\beta_t,\beta_t]$ are also real. This suggests looking for functions with these properties, and the authors assume that ζ_t must be come from some conformal map that maps the complement in the first quadrant of the segment $i/2t+[0,\beta_t]$ onto the upper half plane, and is then extended to the complement in the whole complex plane of the two cuts $\pm i/2t+[-\beta_t,\beta_t]$ by enforcing the symmetries (3.2.6).

Explicitly, the authors first consider a Schwarz-Christoffel map $SC(s;t,\beta)$ that maps the upper half plane to the complement in the first quadrant of the segment $i/2t+[0,\beta]$ (see figure 3.1), were they allow β to be independent of t. They then let $\Gamma(z;\beta,t)$ be the inverse of SC, and extend $\Gamma(z;\beta,t)$ analytically to the complement in the whole complex plane of the two cuts $\pm i/2t+[-\beta,\beta]$ by defining

$$\begin{aligned}\Gamma(-z;\beta,t) &= \Gamma(z;\beta,t),\\ \Gamma(\overline{z};\beta,t) &= \overline{\Gamma(z;\beta,t)}.\end{aligned}$$

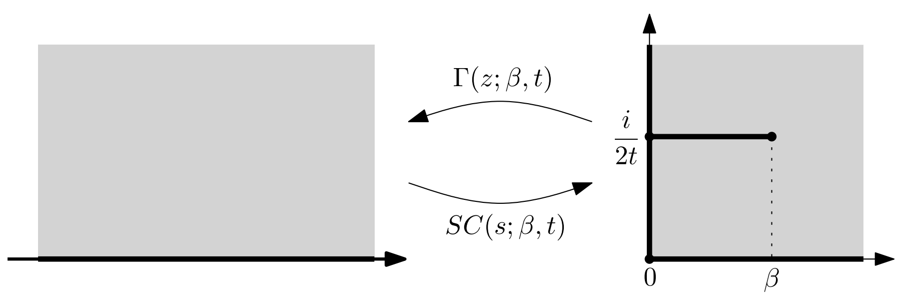

FIGURE 3.1. The Schwarz-Christoffel map SC and its inverse Γ

Now, the expansion of ζ_t at ∞ can be computed to be given by

$$\zeta_t(z)=z^2+\left(\frac{2}{t^2}\right)\frac{1}{z^2}+\left(\frac{12t^2\mathfrak{m}_2(t)-1}{2t^4}\right)\frac{1}{z^4}+\ldots, \tag{3.2.7}$$

and the authors then find conditions on real constants a_1 and a_2 so that $a_1\Gamma+a_2$ has an expansion of the form $z^2+(2/t^2)z^{-2}+O(1/z^4)$ at infinity. These constants a_1,a_2 are given as polynomials in parameters b_1, b_2, b_3 that come from choosing the pre-images of the vertices in the Schwarz-Christoffel map, and so depend in a complicated way on t and β (they are related by equations involving elliptic integrals, which we discuss in detail in chapter 6). The expansion $a_1\Gamma + a_2 = z^2 + (2/t^2)z^{-2} + O(1/z^4)$ at infinity not only makes $a_1 = a_1(\beta, t)$ and $a_2 = a_2(\beta, t)$ depend on β and t in a transcendental way, but also imposes an algebraic relation between b_1, b_2, b_3 and t.

The authors then assume that there is a solution to these relations, and assume that ζ_t agrees with the corresponding map $a_1\Gamma + a_2$. Using this assumption and (3.2.7) allows them to extract the second moment $\mathfrak{m}_2(t)$ of the density ρ_t in terms of b_1, b_2, b_3 and t by comparing the coefficients of $1/z^4$ in the expansion at infinity.

This second moment is in a sense all they are looking for to obtain the genus zero generating function $e_0(t)$, because defining

$$d\rho_{N,t}(\lambda) := \frac{1}{Y_N(t)} e^{-N\sum \lambda_i^2/2} \prod_{i<j} (\lambda_i - \lambda_j)^2 \prod_{i,j} \frac{1}{\sqrt{1 + t^2(\lambda_i + \lambda_j)^2}} d^N\lambda,$$

they claim that the relation

$$t\frac{d}{dt}\left[\frac{1}{N^2}\log \widehat{Y}_N(t)\right] = -1 + \int_{\mathbb{R}^N} \left(\frac{1}{N}\sum \lambda_i^2\right) d\rho_{N,t}(\lambda), \tag{3.2.8}$$

(which one obtains by making a change of variables scaling the λ by t), becomes in the limit $N \to \infty$

$$t\frac{d}{dt}\left[\lim_{N\to\infty} \frac{1}{N^2}\log \widehat{Y}_N(t)\right] = -1 + \mathfrak{m}_2(t), \tag{3.2.9}$$

which by the genus expansion then would give

$$te_0'(t) = -1 + \mathfrak{m}_2(t).$$

Using the equations they have relating all the parameters, they then compute the first coefficients of $\mathfrak{m}_2(t)$

$$\mathfrak{m}_2(t) = 1 - 2t^2 + 14t^4 - 138t^6 + \ldots$$

which then gives them the first terms in the expansion for $e_0(t)$

$$e_0(t) = -\frac{2}{2!}t^2 + \frac{84}{4!}t^4 - \frac{16560}{6!}t^6 + \dots,$$

which agrees with counts of colored triangulations that one can make for low values of n.

3.2.10. We remark that the above heuristics rely on the following assumptions:

- The existence of the limit $\lim_{N\to\infty} N^{-2} \log \widehat{Z}_N(t)$.
- The fact that there exists a probability density for which ρ_t for which

$$\lim_{N\to\infty} \frac{1}{N^2} \log Y_N(t) = \int \frac{1}{2}x^2 \rho_t(x)dx + \iint \log\left(\frac{\sqrt{1+t^2(x+y)^2}}{|x-y|}\right) \rho_t(x)\rho_t(y)dxdy.$$

- The fact that ρ_t is continuous, even, compactly supported on a single interval, and sufficiently well behaved so that the Sokhotski-Plemelj formulas (3.1.12) are applicable.
- The equality of ζ_t with the conformal map $a_1\Gamma + a_2$.
- The fact that (3.2.8) does become (3.2.9) as $N \to \infty$, which not only involves commuting the $N \to \infty$ limit and the derivative in t, but also as relies heavily on the fact that the minimizer the functional

$$I_t[\mu] := \int \frac{1}{2}x^2 d\mu(x) + \iint \log\left(\frac{\sqrt{1+t^2(x+y)^2}}{|x-y|}\right) d\mu(x)d\mu(y),$$

  is unique (a fact we will prove in chapter 5).
- The equality $\lim_{N\to\infty} N^{-2} \log \widehat{Y}_N(t) = e_0(t)$ (a weak version of the asymptotic expansion discussed in 3.1.18 in this setting).

3.2.11. In the following chapters we will provide proofs for all of the bullets in 3.2.10 but the last one, which we regrettably have been unable to prove for the moment.

We remark that because of the extra product $\prod_{i,j}$ in Y_N, the mathematical literature that deals with the asymptotics of these sorts of integrals is not directly applicable. We will develop the relevant theory in chapter 4.

We will also use the formulas relating the parameters b_1, b_2, b_3 with β and t, and the algebraic relation between b_1, b_2, b_3 and t to show that this system of equations defines β as a real analytic function of t with an analytic extension to $t = 0$, in analogy with the results discussed in section 3.1. We will also show that b_1, b_2, b_3, and also a_1, a_1 are analytic in t for $t > 0$ and analyze their behaviour around $t = 0$. We will use this to prove that $\rho_t(x)$ with x in the interior of its support is analytic in t for $t > 0$.

3.3 Comments regarding the physics literature.

We discuss here the relation between the contents of the papers [6, 18, 20, 21, 13], and what we discussed in section 3.2.

3.3.1. Cicuta in [6] mentions the combinatorial model $\widehat{Z}_N(t)$ for colored triangulations, and presents the reduction to the eigenvalues of just one of the matrices $\widehat{Y}_N(t)$, but then proceeds to simplify the model, and so does not study $\widehat{Y}_N(t)$ directly.

3.3.2. Hoppe in [18] was studying a related partition function

$$\tilde{Y}_N(t) := \frac{1}{t^{N^2}} \int e^{-N\sum V_t(\lambda_i)} \prod_{i<j} (\lambda_i - \lambda_j)^2 \prod_{i,j} \frac{1}{\sqrt{1 + (\lambda_i - \lambda_j)^2}} d^N\lambda$$

(note the negative sign in the term inside the square root), where

$$V_t(x) := \frac{1}{2t^2} x^2.$$

This integral comes from the two matrix model with interaction $[A, B]^2$ where $[A, B]$ is the commutator of the two matrices.

Hoppe provides the heuristic arguments for the leading order asymptotics of $\tilde{Y}_N$, introduces the map ζ_t, and has the idea of identifying it with the inverse of a Schwarz-Christoffel map, but does not complete the computations to find the analogue of $e_0(t)$.

3.3.3. Kazakov, Kostov and Nekrasov in [20] complete the computations for $\tilde{Y}(t)$ while changing the approach with the parameters of the Schwarz-Christoffel map and the associated elliptic integrals, and in 3.2 and chapter 6 we have followed and completed Hoppe's original approach since the authors of [20] start with an overly conditioned system of equations for the parameters of the conformal map Γ that can be justified only once one has proved it existence. Kazakov, Kostov and Nekrasov then claim that the large N asymtotics do not change in the large N limit if one replaces $(\lambda_i + \lambda_j)$ by $(\lambda_i - \lambda_j)$ in $\tilde{Y}_N(t)$, giving a rescaled version of $Y_N(t)$ (we will explain what we mean by this in 3.3.4), and so claim their solution applies to colored triangulations. We will prove that their claim is correct, once we show that the minimizing ρ_t is in fact and even function.

3.3.4. Both papers [18, 20] we have mentioned perform their constructions with the eigenvalues rescaled by t, and this rescaling results in a generalized density of eigenvalues $\tilde{\rho}_t$ (we are using this term loosely to refer to the analogue of ρ_t) that has vanishing support as $t \to 0$, i.e., it is converging to a point mass at the origin. This is expected since before the scaling, for $t = 0$ the limiting density is the Wigner semicircle law $\rho_0(x) = (2\pi)^{-1}\sqrt{4 - x^2}$, and $\tilde{\rho}_t$ can then be seen to be given by

$$\tilde{\rho}_t(x) = \frac{1}{t}\rho_t\left(\frac{x}{t}\right).$$

3.3.5. Kostov in [21] does directly discuss $Y_N(t)$, but does so for the case when t is purely imaginary where we see issues with the convergence of the integral Y_N, and the density ρ_t being real valued. We will not be discussing this situation careful, and will be assuming throughout that t is real, except when discussing the analyticity of various quantities as functions of t.

3.3.6. Finally, Eynard and Kristjansen in [13] take an alternate approach to analyze the saddle-point equation which leads them to an iterative (and formal) procedure to obtain the coefficients of e_0, but does not give a closed form expression like the one we described in section 3.2.

Chapter 4
Leading order asymptotics

In this chapter we will study the leading order asymptotics of $N^{-2}\log Z_N^{V,H}$ as $N\to\infty$ where

$$Z_N^{V,H} := \int e^{-N\sum V(x_i)}\prod_{i<j}(x_i-x_j)^2\prod_{1\le i,j\le N}\frac{1}{H(x_i,x_j)}d^Nx.$$

In particular, we show that under reasonable hypothesis on V and H, the limit

$$\lim_{N\to\infty}\frac{1}{N^2}\log Z_N^{V,H},$$

exists.

4.1 Setup and heuristics.

Following the heuristics we discussed in section 3.1, writing the integrand in $Z_N^{V,H}$ as an exponential we obtain

$$\exp\left[-N^2\left(\frac{1}{N}\sum_i\left(V(x_i)+\frac{1}{N}\log H(x_i,x_i)\right)+\frac{1}{N^2}\sum_{i\neq j}\log\frac{H(x_i,x_j)}{|x_i-x_j|}\right)\right],$$

and as $N\to\infty$, we expect the main contributions to the integral to come from the minimums of the expression being multiplied by $-N^2$. Assuming the existence of a limiting density $\rho_{V,H}$ for the measure $\frac{1}{N}\sum_{i=1}^N\delta_{x_i^*}$ where $x=(x_1^*,\dots,x_N^*)$ minimizes the expression being multiplied by $-N^2$, and assuming that the $N^{-1}\log H(x,x)$ term is negligible in the large N limit, one expects that $\lim_{N\to\infty}-N^{-2}\log Z_N^{V,H}$, if it exists, should be equal to

$$\int V(x)d\rho_{V,H}(x)dx+\iint\log\frac{H(x,y)}{|x-y|}d\rho_{V,H}(x)d\rho_{V,H}(y)dxdy,$$

which one can also write as

$$\iint\left[\log\frac{H(x,y)}{|x-y|}+\frac{1}{2}V(x)+\frac{1}{2}V(y)\right]d\rho_{V,H}(x)d\rho_{V,H}(y)dxdy,$$

since $\rho_{V,H}$ is a probability density.

Thus, one is led to analyze the functional

$$\iint \left[\log \frac{H(x,y)}{|x-y|} + \frac{1}{2}V(x) + \frac{1}{2}V(y)\right] d\mu(x)d\mu(y)$$

over the set $\mathcal{M}^1(\mathbb{R})$ of all positive Borel probability measures on $\mathbb{R}$ and to relate this to the asymptotics of $N^{-2}\log Z_N^{V,H}$.

We will show that under reasonable assumptions, the above functional does have minimizers, and that the asymptotics claimed above do hold. This will be done by adapting the arguments for the case $H \equiv 1$ from Johansson's paper [19] and Deift's book [7].

4.1.1 Notation. Throughout this chapter we will use the notation

$$I_{V,H}[\sigma] := \iint K_{V,H}\, d^2\sigma = \iint K_{V,H}(x,y)\, d\sigma(x)d\sigma(y),$$

where σ is a Borel probability measure, and where the kernel $K_{V,H}$ is given by

$$K_{V,H}(x,y) := \log \frac{H(x,y)}{|x-y|} + \frac{1}{2}V(x) + \frac{1}{2}V(y).$$

Note that one has

$$\begin{aligned} I_{V,H}[\sigma] &= \int V d\sigma + \iint \log \frac{H(x,y)}{|x-y|} d\sigma(x)d\sigma(y) \\ &= \int V d\sigma + \iint \log H(x,y)\, d^2\sigma + \iint \log \frac{1}{|x-y|} d\sigma(x)d\sigma(y) \end{aligned}$$

whenever the integrals exist and are finite.

We denote the infimum of $I_{V,H}[\sigma]$ over the set $\mathcal{M}^1(\mathbb{R})$ of all positive Borel probability measures on $\mathbb{R}$ by

$$\inf I_{V,H} := \inf_{\sigma \in \mathcal{M}^1(\mathbb{R})} I_{V,H}[\sigma].$$

4.1.2. We remark that $I_{V,H}[\sigma]$ may not exist for general $K_{V,H}$, in which case the definition of $\inf I_{V,H}$ may lack meaning, but below we will assume conditions on $K_{V,H}$ that guarantee that these quantities are well defined.

4.2 Regularity assumptions.

Our arguments will apply to functions H, V satisfying the following **regularity conditions**:

1. V is continuous.

2. H is continuous and $H \geq 1$.

3. The function
$$\psi_V(x) := V(x) - \log(x^2+1)$$
tends to ∞ as $|x| \to \infty$.

These regularity conditions imply the following facts that will be essential in the arguments.

4.2.1 Lower bound for $\boldsymbol{\psi_V}$. The growth condition on ψ_V implies that

$$\psi_V(x) \geq \min \psi_V > -\infty.$$

4.2.2 Bounds and growth of $\boldsymbol{K_{V,H}}$. Note that for all x, y in $\mathbb{R}$ we have that $|x-y| \leq \sqrt{1+x^2}\sqrt{1+y^2}$ and so $\log|x-y|^{-1} \geq -\frac{1}{2}\log[(x^2+1)(y^2+1)]$. This implies that

$$\begin{aligned} K_{V,H}(x,y) &\geq \log H(x,y) + \frac{1}{2}\big(V(x) - \log(x^2+1)\big) + \frac{1}{2}\big(V(y) - \log(y^2+1)\big) \\ &\geq \frac{1}{2}\psi_V(x) + \frac{1}{2}\psi_V(y), \end{aligned}$$

since $\log H \geq 0$, and so we see that $K_{V,H}$ is bounded from below by $\min \psi_V$, and that $K_{V,H} \to \infty$ as $|x|, |y| \to \infty$.

4.2.3 $\boldsymbol{I_{V,H}[\sigma]}$ is well defined. Since $K_{V,H}$ is bounded from below, then $I_{V,H}[\sigma] = \iint K_{V,H}\, d^2\sigma$ is a well defined element of $(-\infty, \infty]$ for all $\sigma \in \mathcal{M}^1(\mathbb{R})$.

4.2.4 Finiteness of inf $I_{V,H}$. To see that $\inf I_{V,H} < \infty$, just take any measure with compact support and finite logarithmic energy. For example, take $d\sigma(x) = \chi_{[-1,1]}(x)dx$ where $\chi_{[-1,1]}(x)$ is the characteristic function of the interval $[-1,1]$. Then one has $\int\int \log|x-y|^{-1}d\sigma(x)d\sigma(y) < \infty$, and so $I_{V,H}[\sigma] < \infty$ since by continuity, $\log H(x,y)$ and $V(x)$ are bounded in $[0,1]\times[0,1]$ and $[0,1]$ respectively. The fact that $-\infty < \inf I_{V,H}$ follows from the fact that K is bounded from below by the regularity hypothesis.

4.2.5 Finiteness of $\int_{\mathbb{R}} e^{-V(x)}dx$. The growth condition on ψ_V implies that for $|x|$ large enough we have $e^{-V(x)} \leq (1+x^2)^{-1}$, so that

$$\int_{\mathbb{R}} e^{-V(x)}dx < \infty.$$

4.3 Statement of the theorems.

4.3.1 Definition of $Z_{(N)}^{V,H}$. We will denote the integral without the diagonal term in the double product $\prod_{i,j} H(x_i,x_j)^{-1}$ by

$$Z_{(N)}^{V,H} := \int e^{-N\sum V(x_i)} \prod_{i<j}(x_i-x_j)^2 \prod_{1\leq i\neq j\leq N} \frac{1}{H(x_i,x_j)} d^N x.$$

This integral will play a very important role in the arguments that follow since it will be easier to adapt the arguments in the literature to this integral, because the delicate arguments occur along the diagonal where the logarithmic part $\log|x-y|$ of the kernel is singular.

4.3.2 Theorem. *If H and V satisfy the regularity conditions, then the limits*

$$\lim_{N\to\infty} \frac{1}{N^2}\log Z_N^{V,H} \quad \textit{and} \quad \lim_{N\to\infty} \frac{1}{N^2}\log Z_{(N)}^{V,H}$$

exist and are equal. The value of the common limit is given by

$$\lim_{N\to\infty} \frac{1}{N^2}\log Z_N^{V,H} = -\inf I_{V,H}$$

Moreover, there exists a measure μ^* *attaining the infimum*

$$I_{V,H}[\mu^*] = \inf I_{V,H} = \inf_{\sigma \in \mathcal{M}^1(\mathbb{R})} I_{V,H}[\sigma]$$

and any such measure has compact support, finite logarithmic energy and no point masses.

We will prove theorem 4.3.2 in section 4.7, after having developed the relevant results regarding the functional $I_{V,H}$.

4.3.3 Extremal and equilibrium measures. We will refer to any measure μ^* attaining the infimum

$$I[\mu^*] = \inf_{\sigma \in \mathcal{M}^1(\mathbb{R})} I[\sigma],$$

as an **extremal measure**. If μ^* is the *unique* measure satisfying this property, we will call it, following the terminology for the case when $H \equiv 1$, the **equilibrium measure**.

4.3.4. We remark that the requirement that the extremal measure be unique is a common assumption in all the literature regarding these types of limits since this is true case when $H \equiv 1$. In our adaptations of the arguments of [19], we show that this assumption is not necessary for theorem 4.3.2 to hold. Nonetheless, if one does have uniqueness of the extremal measure, one can prove stronger statements such as:

4.3.5 Theorem. *If* H *and* V *satisfy the regularity conditions and the extremal measure* $\mu_{V,H}$ *of* $I_{V,H}$ *is unique, then for any continuous bounded function* $\phi : \mathbb{R}^k \to \mathbb{R}$ *we have*

$$\lim_{N\to\infty} \frac{1}{N} E_N \left[\exp\left\{ \frac{1}{N^{k-1}} \sum_{i_1,\ldots,i_k=1}^{N} \phi(x_{i_1},\ldots,x_{i_k}) \right\} \right]$$
$$= \lim_{N\to\infty} E_N \left[\frac{1}{N^k} \sum_{i_1,\ldots,i_k=1}^{N} \phi(x_{i_1},\ldots,x_{i_k}) \right]$$
$$= \int_{\mathbb{R}^k} \phi(x_1,\ldots,x_k)\, d\mu_{V,H}(x_1)\ldots d\mu_{V,H}(x_k),$$

where $E_N[\cdot]$ is expectation with respect to the probability measure with density

$$\rho_N^{V,H}(x_1,\ldots,x_N) := \frac{1}{Z_N^{V,H}} \prod_{1\le i,j\le N} \frac{1}{H(x_i,x_j)} \prod_{i<j}(x_i-x_j)^2 e^{-N\sum_{i=1}^N V(x_i)}.$$

Furthermore, if $u_N^{V,H}(x_1,\ldots,x_k)$ is the k-point function of $\rho_N^{V,H}$, given by

$$u_N^{V,H}(x_1,\ldots,x_k) := \int_{\mathbb{R}^{N-k}} \rho_N^{V,H}(x_1,\ldots,x_N)dx_{k+1}dx_{k+2}\ldots dx_N,$$

then for every continuous bounded function $\phi:\mathbb{R}^k\to\mathbb{R}$ we have

$$\begin{aligned}\lim_{N\to\infty}\int_{\mathbb{R}^k} &\phi(x_1,\ldots,x_k)u_N^{V,H}(x_1,\ldots,x_k)dx_1\ldots dx_k\\ =\int_{\mathbb{R}^k} &\phi(x_1,\ldots,x_k)\, d\mu_{V,H}(x_1)\ldots d\mu_{V,H}(x_k).\end{aligned}$$

We will prove theorem 4.3.5 in section 4.8.

4.3.6. We remark that the condition that ϕ be bounded in the above theorem can be relaxed by imposing extra growth conditions on V. For example, if V grows faster than a logarithm, then the functions ϕ in the theorem are allowed to grow like polynomials. We will discuss these matters in section 4.8.

4.3.7 Generalization to allow H to be bounded by an arbitrary constant. We remark that the condition $H\ge 1$ in the regularity hypothesis 4.2 can be relaxed to

$$H \ge c_H > 0,$$

for some constant c_H and the statements of the theorems above will still hold. What one really needs in many of the arguments is to be able to separate

$$\log\frac{H(x,y)}{|x-y|} = \log H(x,y) + \log\frac{1}{|x-y|}$$

without running into issues. The arguments become a little more cumbersome under these more general hypothesis, and the generalizations once we have proved the theorems are usually simple, which is why we decided the keep the condition $H \geq 1$ throughout.

For example, to prove that

$$\lim_{N\to\infty} \frac{1}{N^2} \log Z_N^{V,H} = - \inf_{\sigma \in \mathcal{M}^1(\mathbb{R})} \iint \left[\log \frac{H(x,y)}{|x-y|} + \frac{1}{2}V(x) + \frac{1}{2}V(y) \right] d\sigma(x) d\sigma(y)$$

with the relaxed hypothesis $H \geq c_H > 0$, note that $Z_N^{V,H} = c_H^{-N^2} Z_N^{\widetilde{H},V}$ where $\widetilde{H} = H/c \geq 1$, and so we have

$$-\frac{1}{N^2} \log Z_N^{V,H} = \log c_H - \frac{1}{N^2} \log Z_N^{V,\widetilde{H}}.$$

One can then use the theorem on $Z_N^{V,\widetilde{H}}$ to obtain

$$\lim_{N\to\infty} -\frac{1}{N^2} \log Z_N^{V,H} = \log c_H + \inf I_{\widetilde{H},V},$$

but we have

$$\log c_H + \inf I_{\widetilde{H},V} \quad = \quad \inf_{\sigma \in \mathcal{M}^1(\mathbb{R})} \iint \left[\log \frac{H(x,y)}{|x-y|} + \frac{1}{2}V(x) + \frac{1}{2}V(y) \right] d\sigma(x) d\sigma(y).$$

by taking the $\log c_H$ inside the integral for $I_{\widetilde{H},V}$. This is precisely the same statement for $Z_N^{V,H}$ as for $Z_N^{V,\widetilde{H}}$.

4.4 Existence of extremal measures.

We start the proof of theorem 4.3.2 by proving the existence of an extremal measure and some of its properties.

4.4.1 Proposition. *If V and H satisfy the regularity hypothesis, then there exists an extremal measure μ^* (i.e., $I_{V,H}[\mu^*] = \inf I_{V,H}$), and any extremal measure has compact support, finite logarithmic energy $\iint \log |x-y|^{-1} d\mu^*(x) d\mu^*(y) < \infty$, and no point masses. In particular, for any extremal measure μ^* one has*

$$\inf I_{V,H} = I_{V,H}[\mu^*] = \int V d\mu^* + \iint \log H \, d^2\mu^* + \iint \log \frac{1}{|x-y|} d\mu^*(x) d\mu^*(y),$$

since all integrals are finite.

The rest of this section is devoted to the proof of the above proposition.

4.4.2. The fact that any extremal measure has compact support follows from the following lemma, which proves the even stronger statement that the measures that are close to attaining the infimum $\inf I_{V,H}$ have uniformly bounded support. The proof is adapted from Chapter 1 in [26], and will considerably simplify the arguments in the proof of proposition 4.4.1.

4.4.3 Lemma. *If V and H satisfy the regularity hypothesis, then there exists a constant $T > 0$ so that for every $\sigma \in \mathcal{M}^1(\mathbb{R})$ with $I_{V,H}[\sigma] < \inf I_{V,H} + 1$ and* $\operatorname{supp} \sigma \nsubseteq [-T, T]$*, there is a $\tilde{\sigma} \in \mathcal{M}^1([-T, T])$ with $I_{V,H}[\tilde{\sigma}] < I_{V,H}[\sigma]$. In particular, any extremal measure has support contained in $[-T, T]$.*

Proof. By the regularity hypothesis we know that (see 4.2.2) $K_{V,H} \geq \frac{1}{2}\psi_V(x) + \frac{1}{2}\psi_V(y)$, and so there exists a $T > 0$ so that

$$K_{V,H}(x, y) > \inf I_{V,H} + 1 \quad \text{for } (x, y) \notin [-T, T]^2.$$

Now note that if $I_{V,H}[\sigma] < \inf I_{V,H} + 1$, then $\sigma([-T, T]) > 0$ since otherwise, by letting $A = [-T, T]^2$ and B be its complement in $\mathbb{R}^2$, then the choice of T implies that

$$\inf I_{V,H} + 1 > I_{V,H}[\sigma] = \left(\iint_A + \iint_B \right) K_{V,H}\, d^2\sigma = \iint_B K_{V,H}\, d^2\sigma > \inf I_{V,H} + 1$$

which is clearly impossible. Define now

$$\tilde{\sigma} = \frac{1}{\sigma([-T, T])} \cdot \sigma\,|_{[-T,T]},$$

and write the integral in $I_{V,H}[\sigma]$ as the sum of the integral over $A = [-T, T]^2$ and its complement B. The choice of T implies that

$$I_{V,H}[\sigma] > \sigma([-T, T])^2 I_{V,H}[\tilde{\sigma}] + (1 - \sigma([-T, T])^2)(\inf I_{V,H} + 1),$$

(one would have equality if $\sigma([-T,T]) = 1$, but we are excluding that case by assuming $\operatorname{supp}\sigma \nsubseteq [-T,T]$). Since by hypothesis $I_{V,H}[\sigma] < \inf I_{V,H} + 1$, this implies

$$I_{V,H}[\sigma] > \sigma([-T,T])^2 I_{V,H}[\tilde{\sigma}] + (1 - \sigma([-T,T])^2) I_{V,H}[\sigma],$$

and this is equivalent to $I_{V,H}[\tilde{\sigma}] < I_{V,H}[\sigma]$. □

4.4.4 Tightness. We recall that a sequence of measures $\{\sigma_n\} \subseteq \mathcal{M}^1(\mathbb{R})$ is **tight** if for all $\varepsilon > 0$ there is an M such that $\sigma_n(\{|x| \geq M\}) < \varepsilon$ for all n. The important fact we will use is that a tight sequence $\{\sigma_n\} \subseteq \mathcal{M}^1(\mathbb{R})$ has a weakly convergent subsequence in $\mathcal{M}^1(\mathbb{R})$. We remark that tightness is important to guarantee that weak limits of measures do not loose mass. That is, if $\{\sigma_n\}$ converges weakly to σ and is tight, then $\sigma(\mathbb{R}) = 1$. This may not be true without the hypothesis of tightness. See [7, p. 135] for a discussion of these matters.

4.4.5 Existence of an extremal measure. Under the assumption that V and H satisfy the regularity hypothesis, and by the above lemma, we can chose a sequence $\sigma_n \in \mathcal{M}^1([-T,T])$ that gives the infimum

$$\inf I_{V,H} = \lim_{n\to\infty} I_{V,H}[\sigma_n],$$

and since all σ_n have support contained in $[-T,T]$, the sequence is tight. Therefore, we can further assume that the sequence is weakly convergent with limit measure $\mu^* \in \mathcal{M}^1([-T,T])$ by what was discussed in 4.4.4. We will show that $\inf I_{V,H} = I_{V,H}[\mu^*]$.

Let

$$f_m(x,y) := \min(m, K_{V,H}(x,y)),$$

be a sequence of continuous functions that monotonically increase to $K_{V,H}(x,y)$ as $m \to \infty$. By the simple inequality $K_{V,H} \geq f_m$ we have

$$I_{V,H}[\sigma_n] = \iint K_{V,H}\, d^2\sigma_n \geq \iint f_m\, d^2\sigma_n$$

and so, letting $n \to \infty$, we get

$$\inf I_{V,H} \geq \lim_{n\to\infty} \iint f_m \, d^2\sigma_n = \iint f_m \, d^2\mu^*,$$

since the f_m are continuous and bounded and $\sigma_n \otimes \sigma_n$ converge weakly to $\mu^* \otimes \mu^*$ in $\mathcal{M}^1([-T,T]^2)$. Letting then $m \to \infty$ and using the monotone convergence theorem ($K_{V,H}$ is bounded from below) we finally obtain $\inf I_{V,H} \geq I[\mu^*]$, which proves that $\inf I_{V,H} = I_{V,H}[\mu^*]$.

4.4.6. To finish the proof of proposition 4.4.1 it only remains to prove that any extremal measure μ^* has finite logarithmic energy and has no point masses, but these two facts follow easily from what has been proven. Explicitly, since since $K_{V,H}(x,x) = \infty$ by the $\log|x-y|^{-1}$ term, $d\mu^*$ can't have any point masses since $\inf I_{V,H} < \infty$. For the finiteness of the logarithmic energy, note that V and $\log H$ are bounded in the compact support of μ^* and so

$$\inf I_{V,H} = I_{V,H}[\mu^*] = \iint \left[\log \frac{1}{|x-y|} + \log H(x,y) + \frac{1}{2}V(x) + \frac{1}{2}V(y)\right] d\mu^*(x) d\mu^*(y)$$

can be finite only if $\iint \log|x-y| d\mu^*(x) d\mu^*(y) < \infty$.

4.5 The finite dimensional kernel for $Z_{(N)}^{V,H}$.

4.5.1 Comment about the $(\cdot)$ notation. In what follows we will systematically index quantities using parenthesis $(\cdot)$ if we are dealing with the integral $Z_{(N)}^{V,H}$ defined in 4.3.1 where we ignore the diagonal term. Indices without parentheses will be reserved for when we consider the integral $Z_N^{V,H}$ involving the diagonal term.

4.5.2 Definition of $S_{(N)}$. We define the operator $S_{(N)}$ on functions of two variables by

$$(S_{(N)}f)(x_1,\ldots,x_N) := \sum_{1\leq i\neq j\leq N} f(x_i,x_j)$$

where $f(x_1, x_2)$ is a function of two variables.

4.5.3 Definitions. For $x = (x_1, \ldots, x_N) \in \mathbb{R}^N$ define

$$\begin{aligned} S_{(N)}K_{V,H}(x) &:= (S_{(N)}K_{V,H})(x) = \sum_{1 \le i \ne j \le N} K_{V,H}(x_i, x_j), \\ d_{(N)} &:= \frac{1}{N(N-1)} \inf_{x \in \mathbb{R}^N} S_{(N)}K_{V,H}(x), \end{aligned}$$

where we note that

$$S_{(N)}K_{V,H}(x) = \sum_{i \ne j} \left(\log \frac{H(x_i, x_j)}{|x_i - x_j|} \right) + (N-1) \sum_{i=1}^{N} V(x_i),$$

so that

$$\exp\left(-S_{(N)}K_{V,H}(x)\right) = e^{-(N-1)\sum V(x_i)} \prod_{i<j} (x_i - x_j)^2 \prod_{i \ne j} \frac{1}{H(x_i, x_j)}$$

which is very similar to the integrand of $Z_{(N)}^{V,H}$ defined in 4.3.1, the difference being the factor of $N-1$ instead of N in the sum of the $V(x_i)$.

4.5.4. We remark that the point in defining the above quantities is that by the above equations we have

$$Z_{(N)}^{V,H} = \int \exp\left[-S_{(N)}K_{V,H}(x) - \sum_i V(x_i) \right] d^N x,$$

and for large $N \to \infty$ we expect

$$\begin{aligned} \exp\left[-S_{(N)}K_{V,H}(x) - \sum_i V(x_i) \right] &= \exp\left[-N^2 \left(\frac{1}{N^2} S_{(N)}K_{V,H}(x) + \frac{1}{N^2} \sum_i V(x_i) \right) \right] \\ &\simeq \exp\left[-N^2 \cdot \frac{1}{N^2} S_{(N)}K_{V,H}(x) \right] \\ &\simeq \exp\left[-N^2 \cdot \frac{1}{N(N-1)} S_{(N)}K_{V,H}(x) \right]. \end{aligned}$$

The minimizers of $S_{(N)}K_{V,H}(x)$, are the ones that will be playing the role of the minimizers discussed in the heuristics of section 4.1 in the arguments that follow.

4.5.5. We also remark that the expression

$$\frac{1}{N(N-1)} S_{(N)} K_{V,H}(x)$$

showing up in the definition of $d_{(N)}$ and in the heuristics above can be interpreted as a double integral of $K_{V,H}$ with respect to a uniform discrete measure supported at the x_i, where the diagonal terms are excluded, as they have to be since $K_{V,H}(t,t) = \infty$. Thus, $d_{(N)}$ is in some sense a finite version of $\inf I_{V,H}$, and in this section we will prove that $\lim_{N\to\infty} d_{(N)} = \inf I_{V,H}$, which will be a first step in proving that $\lim N^{-2} \log Z_{(N)}^{V,H} = -\inf I_{V,H}$.

4.5.6 The sequence $\{d_{(N)}\}$ is increasing. We remark that one can adapt the proof in [7, p. 148], originally due to Fekete, to show that $\{d_{(N)}\}$ is in fact an increasing sequence. We do not supply the details here.

4.5.7 Proposition. *If V and H satisfy the regularity hypothesis, then*

$$\lim_{N\to\infty} d_{(N)} = \inf I_{V,H}.$$

Moreover, there exist $x^{(*)} = (x_1^{(*)}, \ldots, x_N^{(*)}) \in \mathbb{R}^N$ that attain the infimum defining $d_{(N)}$, and if

$$\nu_{(N)} := \frac{1}{N} \sum_i \delta(x_i^{(*)}),$$

is the uniform discrete measure supported at one of these $x^{(*)}$, then $\{\nu_{(N)}\}_N$ is tight, and any weakly convergent subsequence of $\{\nu_{(N)}\}$ converges weakly to an extremal measure.

The proof of proposition 4.5.7 consists of paragraphs 4.5.8 to 4.5.13. After the proof we discuss further generalizations.

4.5.8 Proof that $d_{(N)} \leq \inf I_{V,H}$. For any $x \in \mathbb{R}^N$ we have

$$\frac{1}{N(N-1)} S_{(N)} K_{V,H}(x) \geq d_{(N)},$$

and integrating on both sides with respect to $d\mu^*(x_1)\dots d\mu^*(x_N)$ where μ^* is any extremal measure giving the infimum $I[\mu^*] = \inf I_{V,H}$, we obtain $\inf I_{V,H} \geq d_{(N)}$ since the integral of each $K_{V,H}(x_i, x_j)$ is $\inf I_{V,H}$.

4.5.9 Existence of $x^{(*)} \in \mathbb{R}^N$ attaining the infimum $d_{(N)}$. We have

$$\begin{aligned} \exp\left(-S_{(N)}K_{V,H}(x)\right) &= e^{-(N-1)\sum V(x_i)}\prod_{i<j}(x_i-x_j)^2\prod_{i\neq j}\frac{1}{H(x_i,x_j)} \\ &\leq e^{-(N-1)\sum V(x_i)}\prod_{i<j}(x_i-x_j)^2 \end{aligned}$$

since $H \geq 1$, which reduces us to the case $H \equiv 1$. Deift in [7, p. 131] shows this last quantity vanishes at infinity as follows: The inequality $|x-y| \leq \sqrt{1+x^2}\sqrt{1+y^2}$ implies that $\prod_{i<j}(x_i-x_j)^2 \leq \prod_i(1+x_i^2)^{N-1}$, so that

$$e^{-(N-1)\sum V(x_i)}\prod_{i<j}(x_i-x_j)^2 \leq \left(\prod_i e^{-V(x_i)}(1+x_i^2)\right)^{N-1}.$$

Now, $e^{-V(x)}(1+x^2) = \exp(-\psi_V(x))$, and since $-\infty < \min\psi_V \leq \psi_V \to \infty$ as $|x| \to \infty$ by the regularity hypothesis, we get that $e^{-V(x)}(1+x^2) < c$ for some constant c. Thus, for any $k = 1, \dots, N$ we have

$$e^{-(N-1)\sum V(x_i)}\prod_{i<j}(x_i-x_j)^2 \leq c^{(N-1)(N-1)}(1+x_k^2)e^{-V(x_k)} \to 0$$

as $|x_k| \to \infty$, which proves the claim.

Since $\exp(-S_{(N)}K_{V,H}(x))$ is continuous, the above argument proves that there are $x^{(*)}$ which attain the suppremum $\sup_{x\in\mathbb{R}^N}\left[\exp(-S_{(N)}K_{V,H}(x))\right]$, so that

$$d_{(N)} = \frac{1}{N(N-1)}S_{(N)}K_{V,H}(x^{(*)}).$$

4.5.10 Tightness of $\{\nu_{(N)}\}_{N\in\mathbb{N}}$. Define now

$$\nu_{(N)} = \frac{1}{N}\sum_{i=1}^{N}\delta_{x_i^*},$$

with x_i^* as in 4.5.9. Then the inequality $K_{V,H}(x,y) \geq \frac{1}{2}\psi_V(x) + \frac{1}{2}\psi_V(y)$ (from the regularity hypothesis 4.2) implies that

$$
\begin{aligned}
(4.5.11) \qquad d_{(N)} &= \frac{1}{N(N-1)} S_{(N)} K_{V,H}\left(x^{(*)}\right), \\
&= \frac{1}{N(N-1)} \sum_{1 \leq i \neq j \leq N} K_{V,H}\left(x_i^{(*)}, x_j^{(*)}\right), \\
&\geq \frac{1}{N(N-1)} \sum_{1 \leq i \neq j \leq N} \frac{1}{2}\psi_V\left(x_i^{(*)}\right) + \frac{1}{2}\psi_V\left(x_j^{(*)}\right), \\
&= \frac{1}{N(N-1)} \sum_{i=1}^{N} \psi_V\left(x_i^{(*)}\right)(N-1), \\
&= \int \psi_V \, d\nu_{(N)}.
\end{aligned}
$$

Now, since $\psi_V \to \infty$ as $|x| \to \infty$, for any L there exists and M_L so that $\psi_V(x) \geq L$ for $|x| > M_L$. Then, denoting by $A_L = [-M_L, M_L]$, and B_L its complement in $\mathbb{R}$, the above inequality together with $\inf I_{V,H} \geq d_{(N)}$ (from 4.5.8) give

$$
\inf I_{V,H} \geq \int \psi_V \, d\nu_{(N)} = \left(\int_{A_L} + \int_{B_L}\right) \psi_V \, d\nu_{(N)} \geq -|\min \psi_V| + L \int_{B_L} \psi_V \, d\nu_{(N)}.
$$

Letting $L \to \infty$ we see that we must have $\int_{B_L} \psi_V \, d\nu_{(N)} \to 0$ so that the sequence $\{\nu_{(N)}\}$ is tight since $M_L \to \infty$ as $L \to \infty$.

4.5.12 Definition of $\boldsymbol{\nu^{(*)}}$. By tightness of $\{\nu_{(N)}\}$, there is a subsequence with a weak limit $\nu^{(*)} \in \mathcal{M}^1(\mathbb{R})$. We will show below that $\nu^{(*)}$ is extremal, thus showing that any weak limit of the $\nu_{(N)}$ is extremal.

4.5.13 Proof that $\boldsymbol{\lim_{N\to\infty} d_{(N)} = \inf I_{V,H}}$ and $\boldsymbol{\nu^{(*)}}$ is an extremal measure. For any real s we can integrate the function $\min(s, K_{V,H}(x,y))$ with respect to $\nu_{(N)} \otimes$

$\nu_{(N)}$ since it is continuous and bounded from above and below, and we have

$$\begin{aligned} S_{(N)}K_{V,H}(x^{(*)}) &= \sum_{1\le i\ne j\le N} K_{V,H}\left(x_i^{(*)}, x_j^{(*)}\right), \\ &\ge \sum_{1\le i\ne j\le N} \min\left[s, K_{V,H}\left(x_i^{(*)}, x_j^{(*)}\right)\right], \\ &= \sum_{1\le i,j\le N} \min\left[s, K_{V,H}\left(x_i^{(*)}, x_j^{(*)}\right)\right] - Ns, \\ &= N^2 \iint \min\left(s, K_{V,H}(x,y)\right) d\nu_{(N)}(x) d\nu_{(N)}(y) - Ns, \end{aligned}$$

so that

$$\begin{aligned} d_{(N)} &= \frac{1}{N(N-1)} S_{(N)}K_{V,H}(x^{(*)}), \\ &\ge \frac{N^2}{N(N-1)} \iint \min\left(s, K_{V,H}(x,y)\right) d\nu_{(N)}(x) d\nu_{(N)}(y) - \frac{s}{N-1}. \end{aligned}$$

Let now $\{N_k\}_{k\in\mathbb{N}}$ be the subsequence of the N's that makes $\{\nu_{(N_k)}\}_{k\in\mathbb{N}}$ converge weakly to $\nu^{(*)}$ as $k\to\infty$. Then, taking $N\to\infty$ in the above inequality along this subsequence we get

$$\liminf_{k\to\infty} d_{(N_k)} \ge \iint \min\left(s, K_{V,H}(x,y)\right) d\nu^{(*)}(x) d\nu^{(*)}(y),$$

and then using monotone convergence letting $s\to\infty$ we obtain

$$\liminf_{k\to\infty} d_{(N_k)} \ge I[\nu^{(*)}] \ge \inf I_{V,H}.$$

Since we already knew that $\inf I_{V,H} \ge d_{(N)}$ from 4.5.8, we also have $\inf I_{V,H} \ge \limsup_{k\to\infty} d_{(N_k)}$, and this shows that $\lim_{k\to\infty} d_{(N_k)} = \inf I_{V,H}$, and that $I[\nu^{(*)}] = \inf I_{V,H}$ so that $\nu^{(*)}$ is extremal as claimed.

It remains to prove that the whole sequence $\{d_{(N)}\}$ converges to $\inf I_{V,H}$. We remark that this would follow from $\inf I_{V,H} \ge \limsup d_{(N)}$ and $\lim_{k\to\infty} d_{(N_k)} = \inf I_{V,H}$ if we knew the sequence was increasing (see 4.5.6), but instead we give here a direct argument. First note that (4.5.11) implies that

$$\infty > \inf I_{V,H} \ge d_{(N)} \ge \int \psi_V \, d\nu_{(N)} \ge \min \psi_V > -\infty$$

so that the sequence $\{d_{(N)}\}$ is bounded. Now, assume that $\{d_{(N'_k)}\}_{k\in\mathbb{N}}$ is any convergent subsequence of $\{d_{(N)}\}$ and let

$$d' = \lim_{k\to\infty} d_{(N'_k)}$$

be its limit. Since $\{\nu_{(N'_k)}\}$ is tight, we can further refine $\{N'_k\}$ to $\{N''_k\}$ so that $\nu_{(N''_k)}$ converges weakly to some ν'', say. If we repeat the above argument with the sequence $\{N''_k\}$ instead of with $\{N_k\}$ we obtain

$$\inf I_{V,H} \geq \lim_{k\to\infty} d_{(N''_k)} \geq I[\nu''] \geq \inf I_{V,H}$$

and so $d' = \inf I_{V,H}$, i.e., any convergent subsequence of $\{d_{(N)}\}$ converges to $\inf I_{V,H}$, and so $\lim_{N\to\infty} d_{(N)} = \inf I_{V,H}$ as claimed. □

4.5.14. This concludes the proof of Proposition 4.5.7. With this at hand, we could proceed to prove the part of Theorem 4.3.2 that pertains $Z^{V,H}_{(N)}$, but we will instead first discuss in section 4.6 how what we have proved is applicable to the integral $Z^{V,H}_N$ involving the diagonal part of the double product involving the H's .

4.5.15. We remark that if the extremal measure happens to be unique, then one can adapt the arguments in [7, p. 147] to give the following generalization of proposition 4.5.7. We will not be needing this, so will not supply the details.

4.5.16 Proposition. *If V and H satisfy the regularity hypothesis and the extremal measure μ^* is unique, then the whole sequence $\{\nu_{(N)}\}$ converges weakly to μ^*.*

4.6 Introduction of the diagonal term.

4.6.1. Note that one can move the diagonal term $\prod_i H(x_i, x_i)$ to the "V"-term in $Z^{V,H}_N$ by writing

$$Z^{V,H}_N = \int_{\mathbb{R}^N} e^{-N\sum V_N(x_i)} \prod_{i<j}(x_i - x_j)^2 \prod_{1\leq i\neq j\leq N} \frac{1}{H(x_i, x_j)} d^N x$$

where

$$V_N(x) := V(x) + \frac{1}{N}\log H(x,x).$$

This is what was done in the heuristics of section 4.1 where we claimed that the $(1/N)\log H(x,x)$ term should negligible. Note moreover, that we can write

$$Z_N^{V,H} = \int_{\mathbb{R}^N} \exp\left(-S_{(N)}K_N(x) - \sum_i V_N(x_i)\right) d^N x$$

where

$$\begin{aligned} K_N(x,y) &:= K_{V_N,H}(x,y) \\ &= \log\frac{H(x,y)}{|x-y|} + \frac{1}{2}V_N(x) + \frac{1}{2}V_N(y), \end{aligned}$$

and as usual,

$$S_{(N)}K_N(x) = \sum_{1\le i\ne j\le N} K_N(x_i,x_j).$$

Thus, by writing equations this way, we see that this setting is entirely analogous to the one we described for $Z_{(N)}^{V,H}$ in section 4.5 (compare the formulas for $Z_{(N)}^{V,H}$ and $Z_N^{V,H}$ involving $S_{(N)}K$ and $S_{(N)}K_N$ respectively), but now we have a kernel K_N that depends on N through V_N.

Let d_N be the analogue of $d_{(N)}$, i.e.,

$$d_N := \frac{1}{N(N-1)} \inf_{x\in\mathbb{R}^N} S_{(N)}K_N(x).$$

4.6.2 Proposition. *If H and V satisfy the regularity hypothesis, then*

$$\lim_{N\to\infty} d_N = \lim_{N\to\infty} d_{(N)} = \inf I_{V,H}.$$

We prove the proposition 4.6.2 below in 4.6.5. We start by proving facts that we will need about the N dependent variational functional I_N we define below.

4.6.3 N-dependent variational theory. Note that since $\log H \ge 0$ we have that $V_N \ge V$, and this implies in particular that

$$\psi_N(x) := V_N(x) - \log(x^2+1) \ge \psi_V(x),$$

so that V_N and H satisfy the regularity hypothesis 4.2 if V and H do. Applying the variational theory from 4.4 to the N-dependent variational problem

$$I_N[\sigma] := \iint K_N \, d^2\sigma,$$

where $N = 1, 2, \ldots$ and

$$F_N := \inf_{\sigma \in \mathcal{M}^1(\mathbb{R})} I_N[\sigma],$$

we see that F_N is finite, and there an extremal measure μ_N^*

$$F_N = I_N[\mu_N^*],$$

with compact support, no point masses, and finite logarithmic energy.

4.6.4 Proof that $\lim_{N\to\infty} F_N = \inf I_{V,H}$. Since $\log H \geq 0$ we have $V_N \geq V_{N+1} \geq V$ so that $K_N \geq K_{N+1} \geq K_{V,H}$. Thus,

$$F_1 \geq F_2 \geq \ldots \geq \inf I_{V,H},$$

which shows that the sequence $\{F_N\}$ converges since $\inf I_{V,H}$ is finite.

If we let μ^* be an extremal measure for $I_{V,H}$, so that $\inf I_{V,H} = I_{V,H}[\mu^*]$. Then we have

$$\begin{aligned} I_N[\mu^*] &= I_{V,H}[\mu^*] + \frac{1}{N}\int \log H(x,x) d\mu^*(x) \\ &= \inf I_{V,H} + \frac{1}{N}\int \log H(x,x) d\mu^*(x), \end{aligned}$$

where $\int \log H(x,x) d\mu^*(x) < \infty$ since $\log H$ is continuous and μ^* has compact support. Moreover, since $I_N[\mu^*] \geq F_N$ by the definition of F_N, we obtain

$$\inf I_{V,H} + \frac{1}{N}\int \log H(x,x) d\mu^*(x) = I_N[\mu^*] \geq F_N$$

and letting $N \to \infty$ this gives $\inf I_{V,H} \geq \lim_{N\to\infty} F_N$ so that

$$\lim_{N\to\infty} F_N = \inf I_{V,H},$$

as claimed.

4.6.5 Proof of proposition 4.6.2. Note that since $K_N \geq K_{V,H}$, we also have

$$d_N \geq d_{(N)}. \tag{4.6.6}$$

Then, just as in the proof that $\inf I_{V,H} \geq d_{(N)}$ in 4.5.8, for any x we have

$$\frac{1}{N(N-1)} S_{(N)} K_N(x) \geq d_N,$$

and integrating with respect to $d\mu_N^*(x_1) \ldots d\mu_N^*(x_N)$ where μ_N^* is any extremal measure giving the infimum $I[\mu_N^*] = F_N$, we obtain $F_N \geq d_N$. Combining this with inequality (4.6.6) gives

$$F_N \geq d_N \geq d_{(N)},$$

and letting $N \to \infty$ finally shows that

$$\lim_{N\to\infty} d_N = \inf I_{V,H},$$

since $\lim_{N\to\infty} d_{(N)} = \lim_{N\to\infty} F_N = \inf I_{V,H}$. □

4.6.7. We remark that with what we now have, one can follow the same arguments as with the finite dimensional kernel for $Z_{(N)}^{V,H}$ to prove the following proposition.

4.6.8 Proposition. *There is an $x^* \in \mathbb{R}^N$ which attains the infimum defining d_N*

$$d_N = \frac{1}{N(N-1)} S_{(N)} K_N(x^*),$$

and if we define

$$\nu_N = \sum \delta(x_i^*)$$

then the sequence $\{\nu_N\}$ is tight, and any weakly convergent subsequence of $\{\nu_N\}$ converges weakly to an extremal measure for $I_{V,H}$. Moreover, if the extremal measure μ^* is unique, the whole sequence $\{\nu_N\}$ converges to μ^* .

For the proof, one follows the above arguments with V_N instead of V, and one uses the contents of 4.6.3, and the inequalities $V_N \geq V$ and $K_N \geq K_{V,H}$ to reduce to arguments we have already presented. We do not include the details here.

4.7 End of proof of theorem 4.3.2

In this section we will prove that if V and H satisfy the regularity hypothesis, then

$$\liminf_{N\to\infty} -\frac{1}{N^2}\log Z_{(N)}^{V,H} \geq \inf I_{V,H}, \tag{4.7.1}$$

and that

$$\limsup_{N\to\infty} -\frac{1}{N^2}\log Z_{N}^{V,H} \leq \inf I_{V,H}. \tag{4.7.2}$$

From these two inequalities the remaining unproven statement of theorem 4.3.2 (equation (4.7.3) below) will follow since since $H \geq 1$ implies that $Z_N^{V,H} \leq Z_{(N)}^{V,H}$ and so

$$-\frac{1}{N^2}\log Z_{(N)}^{V,H} \leq -\frac{1}{N^2}\log Z_{N}^{V,H},$$

and one can take separately lim inf and lim sup on this inequality, which by (4.7.1) and (4.7.2) immediately implies that

$$\lim_{N\to\infty} -\frac{1}{N^2}\log Z_{(N)}^{V,H} = \lim_{N\to\infty} -\frac{1}{N^2}\log Z_{N}^{V,H} = \inf I_{V,H}. \tag{4.7.3}$$

Thus, to conclude the proof of theorem 4.3.2 it only remains to prove (4.7.1) and (4.7.2), which we now do.

4.7.4 Proof of (4.7.1). Note that

$$\begin{aligned}
Z_{(N)}^{V,H} &= \int_{\mathbb{R}^N} \exp\left[-S_{(N)}K_{V,H}(x) - \sum_i V(x_i)\right] d^N x,\\
&\leq \int_{\mathbb{R}^N} \exp\left[-N(N-1)d_{(N)} - \sum_i V(x_i)\right] d^N x,\\
&= e^{-N(N-1)d_{(N)}}\left(\int_{\mathbb{R}} e^{-V(x)}dx\right)^N,
\end{aligned}$$

since $d_{(N)} := (N(N-1))^{-1}\inf_{x\in\mathbb{R}^N} S_{(N)}K_{V,H}(x)$. This shows that

$$-\frac{1}{N^2}\log Z_{(N)}^{V,H} \geq \frac{N(N-1)}{N^2}d_{(N)} - \frac{1}{N}\log\int_{\mathbb{R}} e^{-V(x)}dx,$$

and letting $N \to \infty$ (recall that $\int_{\mathbb{R}} e^{-V(x)}dx < \infty$ by the growth conditions of ψ_V), we obtain

$$\liminf_{N\to\infty} -\frac{1}{N^2}\log Z_{(N)}^{V,H} \geq \lim_{N\to\infty} d_{(N)} = \inf I_{V,H}.$$

□

4.7.5 Lemma. *Given $\varepsilon > 0$, there is a measure with continuous compactly supported density $\phi_\varepsilon(x)$ so that*

$$I_{V,H}[\phi_\epsilon(x)dx] \leq \inf I_{V,H} + \varepsilon/2.$$

Proof. Let

$$\psi_\delta(x) = \frac{1}{2\delta}\int_{x-\delta}^{x+\delta} d\mu^*(t)$$

where μ^* is any extremal measure so that $I_{V,H}[\mu^*] = \inf I_{V,H}$. Since μ^* has compact support and that it has no point masses, it follows that $\psi_\delta(x)$ is continuous and has compact support. Moreover, one can check that $\psi_\delta(x)dx$ is a probability measure, and that $\psi_\delta(x)dx$ converges weakly to μ^* as $\delta \to 0$ (both facts follow by interchanging orders of integration). We show that $\lim_{\delta\to 0} I_{V,H}[\psi_\delta(x)dx] = I_{V,H}[\mu^*] = \inf I_{V,H}$ which will prove the claim.

First recall that

$$I_{V,H}[\sigma] = \iint \left[\log\frac{1}{|x-y|} + \log H(x,y) + \frac{1}{2}V(x) + \frac{1}{2}V(y)\right] d\sigma(x)d\sigma(y)$$

where $H \geq 1$ and V are continuous. It follows that by the weak convergence and the fact that the ψ_δ have compact support that

$$\iint \left[\log H(x,y) + \frac{1}{2}V(x) + \frac{1}{2}V(y)\right] d\psi_\delta(x)d\psi_\delta(y)dy$$

converges as $\delta \to 0$ to

$$\iint \left[\log H(x,y) + \frac{1}{2}V(x) + \frac{1}{2}V(y)\right] d\mu^*(x)d\mu^*(y)$$

so we only have to worry about the $\log|x-y|^{-1}$ term. This reduces us to the case $H \equiv 1$, and the details are worked out in [7, p.150]. □

4.7.6 Proof of (4.7.2). Let $\varepsilon > 0$, and set $E = \{x \in \mathbb{R} \mid \phi_\varepsilon(x) > 0\}$ where ϕ_ε is as given by the lemma. Then we have

$$\begin{aligned} Z_N^{V,H} &= \int_{\mathbb{R}^N} \exp\left[-S_{(N)}K_N(x) - \sum_i V_N(x_i)\right] d^N x \\ &\geq \int_{E^N} \exp\left[-S_{(N)}K_N(x) - \sum_i V_N(x_i)\right] d^N x \\ &= \int_{E^N} \exp\left[-S_{(N)}K_N(x) - \sum_i V_N(x_i) - \sum_i \log(\phi_\epsilon(x_i))\right] \prod_i \phi_\epsilon(x_i) d^N x, \end{aligned}$$

and if we then use Jensen's inequality

$$\int e^{f(x)} d\sigma(x) \geq e^{\int f(x) d\sigma(x)},$$

with the probability measure $d\sigma(x) = \prod_i \phi_\epsilon(x_i) d^N x$, one obtains

$$\begin{aligned} \log Z_N^{V,H} &\geq -\int_{E^N} S_{(N)}K_N(x) \prod_i \phi_\epsilon(x_i) d^N x - \int_{E^N} \sum_i V_N(x_i) \prod_i \phi_\epsilon(x_i) d^N x \\ &\quad -\int_{E^N} \sum_i \log(\phi_\epsilon(x_i)) \prod_i \phi_\epsilon(x_i) d^N x \\ &= -N(N-1) I_N[\phi_\varepsilon] - N \int_E V_N(x)\phi_\varepsilon(x) dx - N \int_E \log \phi_\varepsilon(x)\phi_\varepsilon(x) dx. \end{aligned}$$

Note that since ϕ_ε has compact support then $\int_E \log \phi_\varepsilon(x)\phi_\varepsilon(x)dx$ is finite, and $V_N(x)$ is bounded in E. Thus, if we denote by $V_\varepsilon = \max_{x\in E} V(x)$ and $C_\varepsilon = \max_{x\in E} \log H(x,x)$ we have for $x \in E$

$$\int_E V_N(x)\phi_\varepsilon(x) dx \leq V_\varepsilon + \frac{1}{N} C_\varepsilon$$

so that

$$\frac{1}{N^2} \log Z_N^{V,H} \geq -\frac{N(N-1)}{N^2} I_N[\phi_\varepsilon] + \frac{D_\varepsilon}{N} - \frac{C_\varepsilon}{N^3},$$

for some constant D_ε that depends on ε. This implies that

$$\liminf_{N\to\infty} \frac{1}{N^2} \log Z_N^{V,H} \geq -I_{V,H}[\phi_\varepsilon]$$

since

$$\begin{aligned}\lim_{N\to\infty} I_N[\phi_\varepsilon] &= \lim_{N\to\infty} I_{V,H}[\phi_\varepsilon] + \frac{1}{N}\int \log H(x,x)\phi_\varepsilon(x)dx \\ &= I_{V,H}[\phi_\varepsilon],\end{aligned}$$

as $\int \log H(x,x)\phi_\varepsilon(x)dx$ is finite by the compact support of $\phi_\varepsilon(x)dx$. By the choice of ϕ_ε, we have

$$\liminf_{N\to\infty} \frac{1}{N^2}\log Z_N^{V,H} \geq -(\inf I_{V,H} + \varepsilon/2), \tag{4.7.7}$$

and this implies (4.7.2) since ε was arbitrary. □

4.8 Large deviation estimates and proof of theorem 4.3.5

In this section we discuss some of the details of the proof of theorem 4.3.5, as well as its strengthenings under stronger hypothesis on the growth of V. We start by showing that the large deviation estimate, which is one of the key ingredients in the proof from [7] for the case $H \equiv 1$ continues to hold.

We let

$$\rho_N^{V,H}(x) := \frac{1}{Z_N^{V,H}} e^{-N\sum V(x_i)} \prod_{i<j}(x_i - x_j)^2 \prod_{i,j} \frac{1}{H(x_i,x_j)},$$

be the probability density in $\mathbb{R}^N$ associated to $Z_N^{V,H}$, and given $c > 0$ we define the (compact) set

$$A_{N,c} = \left\{ x \in \mathbb{R}^N : \frac{1}{N^2} S_{(N)} K_N(x) \leq \inf I_{V,H} + c \right\}.$$

4.8.1 Lemma (Large deviation estimate)**.** *If V and H satisfy the regularity hypothesis 4.2, then for any $c > 0$ there exists an N_c so that for $N \geq N_c$ and any constant $a \geq 0$ one has*

$$\mu_N^{V,H}(\mathbb{R}^N \setminus A_{N,c+a}) \leq e^{-aN^2},$$

where $\mu_N^{V,H}$ is the probability measure with density $\rho_N^{V,H}$.

Proof. Since $\lim_{N\to\infty} N^{-2}\log Z_N^{V,H} = -\inf I_{V,H}$, it follows that given $\varepsilon > 0$ there is an N_ε such that for $N \geq N_\varepsilon$ we have

$$\frac{1}{N^2}\log Z_N^{V,H} \geq -(\inf I_{V,H} + \varepsilon).$$

Using this bound one can remove the $Z_N^{V,H}$ term in computations of probabilities with respect to $\rho_N^{V,H}(x)dx$ since for $N \geq N_\varepsilon$ we will have $Z_N^{V,H} \geq \exp(-N^2(\inf I_{V,H} + \varepsilon))$, and so, for any borel set A we have (recall 4.6.1)

$$\begin{aligned}
\mu_N^{V,H}(A) &= \frac{1}{Z_N^{V,H}}\int_A e^{-N\sum V_N(x_i)}\prod_{i<j}(x_i - x_j)^2\prod_{i\neq j}\frac{1}{H(x_i,x_j)}d^N x \\
&\leq e^{N^2(\inf I_{V,H}+\varepsilon)}\int_A e^{-N\sum V_N(x_i)}\prod_{i<j}(x_i - x_j)^2\prod_{i\neq j}\frac{1}{H(x_i,x_j)}d^N x \\
&= e^{\varepsilon N^2}\int_A \exp\left[-S_{(N)}K_N(x) - \sum V_N(x_i) + N^2\inf I_{V,H}\right]dx,
\end{aligned}$$

as long as $N \geq N_\varepsilon$.

In particular, taking $A = \mathbb{R}^N \setminus A_{N,c+a}$ we have by the definition of $A_{N,c+a}$ that

$$-S_{(N)}K_N(x) + N^2\inf I_{V,H} < -N^2(c+a),$$

for $x \in A = \mathbb{R}^N \setminus A_{N,c+a}$, and so

$$\begin{aligned}
\mu_N^{V,H}(\mathbb{R}\setminus A_{N,c+a}) &\leq e^{\varepsilon N^2}\int_A \exp\left[-\sum V_N(x_i) - N^2(c+a)\right]d^N x \\
&= e^{-N^2(c+a-\varepsilon)}\int_A \exp\left[-\sum V_N(x_i)\right]d^N x \\
&\leq e^{-N^2(c+a-\varepsilon)}\left(\int_{\mathbb{R}} e^{-V_N(x)}dx\right)^N \\
&\leq e^{-N^2(c+a-\varepsilon)}\left(\int_{\mathbb{R}} e^{-V(x)}dx\right)^N
\end{aligned}$$

since $V_N \geq V$ (recall that $\int e^{-V} < \infty$ by the regularity hypothesis 4.2). Now taking $\varepsilon = \varepsilon(c) < c/2$ (note that ε was arbitrary until now) the above estimates give that for $N \geq N_\varepsilon$

$$\mu_N^{V,H}(\mathbb{R}\setminus A_{N,c+a}) \leq e^{-N^2 a}e^{-N^2 c/2}\,(\text{const.})^N$$

and then we can find $N_c \geq N_\varepsilon$ so that for $N \geq N_c$ we have

$$\mu_N^{V,H}(\mathbb{R} \setminus A_{N,c+a}) \leq e^{-N^2 a},$$

and this concludes the proof of the Lemma. □

4.8.2. As mentioned above, the large deviation estimate from lemma 4.8.1 is the essential ingredient in the proof of theorem 4.3.5. The proof is considerably long, but with what has been worked out here, it is essentially no different than the proof in [7] for the case $H \equiv 1$. We consider it illuminating to show how the large deviation estimate and the uniqueness of the extremal measure are used in the proof, and so we include here a proof of the special case that if $\phi : \mathbb{R} \to \mathbb{R}$ is continuous and bounded, then

$$\lim_{N\to\infty} \frac{1}{Z_N^{V,H}} \int_{\mathbb{R}^N} \left(\frac{1}{N} \sum_{i=1}^{N} \phi(x_i) \right) \rho_N^{V,H}(x) d^N x = \int \phi(s) d\mu_N^{V,H}(s). \tag{4.8.3}$$

We prove (4.8.3) in paragraph 4.8.5. For the proof we will need the following lemma:

4.8.4 Lemma. *If V and H satisfy the regularity hypothesis, $\eta > 0$ is arbitrary, $a = (a_1^{(\eta)}, \ldots, a_N^{(\eta)}) \in A_{N,2\eta}$, and we define*

$$\nu_{\eta,N} := \frac{1}{N} \sum \delta_{a_i^{(\eta)}},$$

then $\{\nu_{\eta,N}\}_N$ is tight, and any weak limit ν_η of $\{\nu_{\eta,N}\}_N$ satisfies

$$I_{V,H}[\nu_\eta] \leq \inf I_{V,H} + 2\eta.$$

Moreover, if one takes $\eta = 1/n$ and uses the above result to obtain weak limits $\nu_{1/n}$ for $n = 1, 2, \ldots$, then the sequence $\{\nu_{1/n}\}_n$ is tight, and any weak limit converges to an extremal measure of $I_{V,H}$.

Proof. By definition of $A_{N,2\eta}$ we have

$$\inf I_{V,H} + \eta \geq \frac{1}{N^2} S_N K_N(a) \geq \frac{N-1}{N^2} \sum \psi_V(a_i) = \frac{N-1}{N} \int \psi_V(x) d\nu_{\eta,N}(x),$$

(since $K_N(x,y) \geq \psi_N(x)/2 + \psi_N(y)/2 \geq \psi_V(x)/2 + \psi_V(y)/2$). Thus

$$\int \psi_V(x) d\nu_{\eta,N}(x) \leq const.$$

which shows that $\{\nu_{\eta,N}\}_N$ is tight as $\psi_V(x) \to \infty$ as $|x| \to \infty$. Let ν_η be any weak limit (say for $N = N_j$ with $j \to \infty$), then we have

$$\begin{aligned} \iint \min(l, K_N(x,y)) d\nu_{\eta,N}(x) d\nu_{\eta N}(y) &\leq S_N K_N(a) + \text{diagonal terms} \\ &\leq \frac{l}{N} + \inf I_{V,H} + 2\eta \end{aligned}$$

by the definition of $A_{N,2\eta}$. Now take $N_j \to \infty$, and then take $l \to \infty$ which shows that $I_{V,H}[\nu_\eta] \leq \inf I_{V,H} + 2\eta$ and concludes the first part of the statement.

For the second part, take $\eta = 1/n$ and and obtain the weak weak limits $\nu_{1/n}$ for $n = 1, 2, \ldots$. Now note that by the inequality $\int \psi_V(s) d\rho_{\eta,N}(s) \leq const.$ from above, we have that $\int (\psi_V(s) - \min \psi_V) \, d\nu_{\eta,N}(s) \leq const.$ too, where the integrand is positive. Now, for any $b > 0$ we have by weak convergence that

$$\begin{aligned} \int_{-b}^{b} (\psi_V(s) - \min \psi_V) \, d\nu_\eta(s) &= \lim_{j\to\infty} \int_{-b}^{b} (\psi_V(s) - \min \psi_V) \, d\nu_{\eta,N_j}(s) \\ &\leq \lim_{j\to\infty} \int_{-\infty}^{\infty} (\psi_V(s) - \min \psi_V) \, d\nu_{\eta,N_j}(s) \\ &\leq \lim_{j\to\infty} const. = const. \end{aligned}$$

Thus, for any $b > 0$ we have $\int_{-b}^{b} (\psi_V(s) - \min \psi_V) \, d\nu_\eta(s) \leq const.$, and taking $b \to \infty$ using monotone convergence we get

$$\int (\psi_V(s) - \min \psi_V) \, d\nu_\eta(s) \leq const.$$

which implies that $\{\nu_{1/n}\}_n$ is tight.

Let ν be a weak limit of the sequence, then

$$\iint \min(l, K_N) d^2\nu_{1/n} \leq I_{V,H}[\nu_{1/n}] \leq \inf I_{V,H} + \frac{2}{n}$$

and taking $n \to \infty$ we get $\iint \min(l, K_N) d^2\nu \leq \inf I_{V,H}$. Finally, taking $l \to \infty$ shows that $I_{V,H}[\nu] \leq \inf I_{V,H}$ and so ν is extremal. □

4.8.5 Proof of (4.8.3). Say $|\phi| \le C$, then we have

$$\begin{aligned}
\int_{\mathbb{R}^N} \left(\frac{1}{N} \sum_{i=1}^{N} \phi(x_i) \right) \rho_N^{V,H}(x) dx &= \left(\int_{A_{N,2\eta}} + \int_{\mathbb{R}^N \setminus A_{N,2\eta}} \right) \left[\frac{1}{N} \sum_{i=1}^{N} \phi(x_i) \right] \rho_N^{V,H}(x) d^N x \\
&\le C\, \mu_N^{V,H}(\mathbb{R}^N \setminus A_{N,2\eta}) \\
&\quad + \int_{A_{N,2\eta}} \left(\frac{1}{N} \sum_{i=1}^{N} \phi(x_i) \right) \rho_N^{V,H}(x) d^N x
\end{aligned}$$

and so by the large deviation estimates 4.8.1 we have

$$\limsup_{N\to\infty} \int \left(\frac{1}{N} \sum_{i=1}^{N} \phi(x_i) \right) \rho_N^{V,H}(x) dx = \limsup_{N\to\infty} \int_{A_{N,2\eta}} \left(\frac{1}{N} \sum_{i=1}^{N} \phi(x_i) \right) \rho_N^{V,H}(x) dx.$$

Since $A_{N,2\eta}$ is compact, there is $x^* = (x_i^*)$ such that

$$\left(\frac{1}{N} \sum_{i=1}^{N} \phi(x_i^*) \right) = \sup_{x \in A_{N,2\eta}} \left(\frac{1}{N} \sum_{i=1}^{N} \phi(x_i) \right),$$

and using this we can write

$$\begin{aligned}
\limsup_{N\to\infty} \int \left(\frac{1}{N} \sum_{i=1}^{N} \phi(x_i) \right) \rho_N^{V,H}(x) dx &\le \limsup_{N\to\infty} \left(\frac{1}{N} \sum_{i=1}^{N} \phi(x_i^*) \right) \mu_N^{V,H}(A_{N,2\eta}) \\
&= \limsup_{N\to\infty} \left(\frac{1}{N} \sum_{i=1}^{N} \phi(x_i^*) \right) \\
&= \limsup_{N\to\infty} \int \phi(s) d\mu_{\eta,N}(s)
\end{aligned}$$

where

$$\mu_{\eta,N} = \frac{1}{N} \sum \delta_{x_i^*}$$

is the measure supported at these maximizers.

Now, chose a subsequence N_j of the N that makes the integrals on the right converge to the corresponding $\limsup$. By the lemma, the sequence $\{\mu_{\eta,N}\}_N$ is tight, and so, the from the subsequence $\{\mu_{\eta,N_j}\}_j$ we may chose a further subsequence (which we will contunue to denote by N_j) with weak limit μ_η so that

$$\begin{aligned}
\limsup_{N\to\infty} \int \phi(s) d\mu_{\eta,N}(s) &= \lim_{j\to\infty} \int \phi(s) d\mu_{\eta,N_j}(s) \\
&= \int \phi(s) d\mu_\eta(s)
\end{aligned}$$

since ϕ is continuous and bounded. Thus, we get

$$\limsup_{N\to\infty} \int \left(\frac{1}{N} \sum_{i=1}^{N} \phi(x_i) \right) \rho_N^{V,H}(x) d^N x \le \int \phi(s) d\mu_\eta(s),$$

for every $\eta > 0$.

Using the second statement in the lemma, let μ be a weak limit of $\{\mu_{1/n}\}_n$, so that μ is extremal. Then taking the limit as $n \to \infty$ on the right of the above equation we get

$$\limsup_{N\to\infty} \int \left(\frac{1}{N} \sum_{i=1}^{N} \phi(x_i) \right) \rho_N^{V,H}(x) d^N x \le \int \phi(s) d\mu(s),$$

again using the fact that ϕ is bounded and continuous.

One now repeats the whole argument with liminf taking as x^* the infimum of $N^{-1} \sum_{i=1}^{N} \phi(x_i^*)$ over $A_{N,\eta}$ (this is why we stated the lemma separately) to obtain

$$\int \phi(s) d\tilde{\mu}(s) \le \liminf_{N\to\infty} \int \left(\frac{1}{N} \sum_{i=1}^{N} \phi(x_i) \right) \rho_N^{V,H}(x) d^N x.$$

Now use uniqueness of the extremal measures to conclude that $\mu_{V,H} = \tilde{\mu} = \mu$, and so we finally conclude that

$$\lim_{N\to\infty} \int \left(\frac{1}{N} \sum_{i=1}^{N} \phi(x_i) \right) \rho_N^{V,H}(x) d^N x = \int \phi(s) \mu_{V,H}(s).$$

$\square$

4.8.6 Relaxation on the condition that ϕ be bounded. Under stronger assumptions on the growth on V, one can be generalize 4.3.5 to more general ϕ. For example, if V grows faster than a logarithm (e.g., like $(\log|x|)^{1+\varepsilon}$ for some $\varepsilon > 0$), then one can prove that (4.8.3) holds for polynomial ϕ. One way to do this is by using the bound on the one point function $u_N^{V,H}(t)$ given in lemma 4.8.7 below, which can be used to show that the integral outside a compact set is negligible in the limit $N \to \infty$.

4.8.7 Lemma. *If V and H satisfy the regularity assumptions, then there is a constant C so that*

$$u_N^{V,H}(t) \leq e^{CN}(1+t^2)^N e^{-NV(t)}.$$

Lemma 4.8.7 follows directly from the corresponding statement for the case $H \equiv 1$ given in lemma 4.4 in [19], under the assumption that $H \geq 1$. For the more general situation $H \geq c_H > 0$, one needs to adapt the proof given there, and use the large deviation estimate 4.8.1.

Chapter 5
Variational Inequalities

In this chapter we present a criterion which characterizes extremal measures analogous to the one described in [7] for logarithmic potentials with external fields. The criterion will apply to symmetric positive (semi-)definite kernels, which we define below. Using this, we will obtain a verifiable characterization of the extremal measure for kernels of the form $K_{V,H}$, and we then show that the kernel for colored triangulations satisfies the properties that guarantee the criterion is applicable.

5.1 Extremal measures for positive (semi-)definite kernels.

In this section we will assume that the kernel $K(x, y)$ is real valued and bounded from below on compact sets. To make equations less cluttered below, we will use the notation

$$\iint K d\sigma d\mu := \iint K(x, y) d\sigma(x) d\mu(y),$$
$$\iint K d^2\sigma := \iint K(x, y) d\sigma(x) d\sigma(y),$$

for general signed measures σ, μ on $\mathbb{R}$. We will reserve the notation

$$I_K[\sigma] := \iint K d^2\sigma,$$

for probability measures and will not use it for general signed measures.

5.1.1 Positive (semi-)definite kernels. We say that a kernel $K(x, y)$ that is bounded from below on compact sets is **positive semi-definite** if for any signed measure ν with mean zero and compact support one has

$$\iint K \left(d\nu^+ d\nu^+ + d\nu^- d\nu^-\right) \geq \iint K \left(d\nu^+ d\nu^- + d\nu^- d\nu^+\right), \tag{5.1.2}$$

where $\nu = \nu^+ - \nu^-$ is the decomposition of the measure into its positive and negative parts. Note that all integrals in the above inequality are well defined elements of $(-\infty, \infty]$ since $K(x,y)$ is bounded from below on the compact support of ν. We say that the kernel is **positive definite** if it is positive semi-definite and equality holds in 5.1.2 when both sides are finite only if $\nu = 0$.

In the case when all the integrals in the inequality are finite, then the inequality is equivalent to the more natural

$$\iint K d^2\nu \geq 0,$$

but this last integral may be undefined for general signed ν and K since we are not assuming K is bounded from above (the prototypical example being $K(x,y) = \log 1/|x-y|$).

5.1.3 Remark. There is some closely related literature on potential theory which is similar to the one we discuss below, for kernels for which $\iint K(x,y)d\nu(x)d\nu(y) \geq 0$ whenever ν is a measure for which the integral is defined (see for example [25]). These kernels are referred to as kernels of *positive type*. This definition assumes the finiteness of the integrals $\iint K\left(d\nu^+ d\nu^+ + d\nu^- d\nu^-\right)$ and $\iint K\left(d\nu^+ d\nu^- + d\nu^- d\nu^+\right)$ in the cases when the inequality needs to hold, and this is not sufficient for our purposes (see lemma 5.1.7 and 5.1.8 below). The analogues of positive definite kernels in this context are called kernels satisfying the *energy principle*. We will not be using this terminology here.

5.1.4 Symmetric kernels. We say that the kernel K is **symmetric** if $K(x,y) = K(y,x)$ for all x, y. Note that in the case when K is symmetric, the inequality in the definition of positive semi-definiteness can be written in the form

$$\iint K\left(d^2\nu^+ + d^2\nu^-\right) \geq 2 \iint K d\nu^+ d\nu^-, \tag{5.1.5}$$

since $\iint K d\nu^+ d\nu^- = \iint K d\nu^- d\nu^+ \in (-\infty, \infty]$.

5.1.6 Definition. We say that a measure $\sigma \in \mathcal{M}^1(\mathbb{R})$ is **decent** (with respect to K) if σ has compact support and $I_K[\sigma] < \infty$.

5.1.7 Lemma. *If K is a symmetric positive semi-definite kernel that is bounded from below on compact sets, and σ, μ are decent measures, then*

$$\iint K d(a_1\sigma + a_2\mu) d(b_1\sigma + b_2\mu) < \infty,$$

for any $a_i, b_i \in \mathbb{R}$. In particular, $\iint K d^2(\sigma - \mu)$ exists and we have

$$\iint K d^2(\sigma - \mu) \geq 0$$

since K is positive semi-definite.

Proof. Since K is symmetric and bounded from below on the compact support of $\sigma \otimes \mu$ and $\mu \otimes \sigma$, we know that $\iint K d\sigma d\mu = \iint K d\mu d\sigma$ are well defined as elements of $(-\infty, \infty]$. Moreover, since σ and μ are decent, then $\iint K d^2\sigma$ and $\iint K d^2\mu$ are finite, and so by bilinearity it suffices to prove $\iint K d\sigma d\mu = \iint K d\mu d\sigma$ is also finite. This follows by using positive semi-defininiteness with the compactly supported mean zero measure $\nu = \sigma - \mu$ and inequality (5.1.5). $\square$

5.1.8. We remark that the technical definition of positive (semi-)definiteness was made precisely so that the above lemma holds. Explicitly, we want the kernel itself and the decency of the measures to imply the existence of the integral in the statement of the lemma.

5.1.9. Let now K be a kernel that is bounded from below (not just on compact sets). Because of the lower bound of K, the integral $I_K[\sigma]$ exists for all $\sigma \in \mathcal{M}^1(\mathbb{R})$, and we can define

$$\inf I_K := \inf_{\sigma \in \mathcal{M}^1(\mathbb{R})} I_K[\sigma],$$

which we know is a well defined element of $(-\infty, \infty]$. As in the previous chapter, we refer to any measure μ^* attaining the infimum

$$\inf I_K = I_K[\mu^*],$$

as an **extremal** measure.

5.1.10 Theorem (Characterization of extremal measures for positive semi-definite kernels)**.** *Let K be a positive semi-definite symmetric kernel that is bounded from below, and assume that extremal measures exist and that any extremal measure is decent (so in particular $\inf I_K$ is finite). Then a decent measure $\mu^* \in \mathcal{M}^1(\mathbb{R})$ is an extremal measure if and only if*

$$\iint K d\mu^* d\sigma \geq \iint K d^2\mu^* \left(= I_K[\mu^*]\right),$$

for any decent measure σ.

Proof. ($\Rightarrow$) Let σ be a decent measure, and note that for any $t \in [0,1]$ we have

$$\mu^* + t(\sigma - \mu^*) = t\sigma + (1-t)\mu^* \in \mathcal{M}^1(\mathbb{R}),$$

so that we can write

$$I_K[\mu^* + t(\sigma - \mu^*)] = I_K[\mu^*] + 2t \iint K d\mu^* d(\sigma - \mu^*) + t^2 \iint K d^2(\sigma - \mu^*),$$

since all integrals involved are well defined and finite by lemma 5.1.7. Since μ^* is extremal, we have $I_K[\mu^*] \leq I_K[\mu^* + t(\sigma - \mu^*)]$, and so we conclude by the above equality that

$$0 \leq I_K[\mu^* + t(\sigma - \mu^*)] - I_K[\mu^*] = t\left(2 \iint K d\mu^* d(\sigma - \mu^*) + t \iint K d^2(\sigma - \mu^*)\right)$$

for all $t \in [0,1]$. Thus,

$$2 \iint K d\mu^* d(\sigma - \mu^*) + t \iint K d^2(\sigma - \mu^*) \geq 0,$$

for all $t \in [0,1]$, which for $t = 0$ is equivalent to the stated inequality.

($\Leftarrow$) Assume that μ^* is a decent measure that satisfies the stated properties, and let $\tilde{\mu}$ be an extremal measure (assumed to exist by hypothesis). Then, using the lemma again,

$$\begin{aligned} I_K[\tilde{\mu}] &= I_K[\mu^* + (\tilde{\mu} - \mu^*)] \\ &= I_K[\mu^*] + 2\iint K d\mu^* d(\tilde{\mu} - \mu^*) + \iint K d^2(\tilde{\mu} - \mu^*), \end{aligned} \tag{5.1.11}$$

and since $I_K[\mu^*] \geq I_K[\tilde{\mu}]$, we obtain

$$0 \geq 2\iint K d\mu^* d(\tilde{\mu} - \mu^*) + \iint K d^2(\tilde{\mu} - \mu^*). \tag{5.1.12}$$

However, the hypothesis on μ^* applied with $\sigma = \tilde{\mu}$ implies that

$$\iint K d\mu^* d(\tilde{\mu} - \mu^*) \geq 0, \tag{5.1.13}$$

and so we must have $\iint K d^2(\tilde{\mu} - \mu^*) \leq 0$ by (5.1.12), which by positive semi-definiteness of the kernel implies that

$$\iint K d^2(\tilde{\mu} - \mu^*) = 0.$$

Using this equality back in (5.1.12) together with (5.1.13) then implies that

$$\iint K d\mu^* d(\tilde{\mu} - \mu^*) = 0,$$

which by (5.1.11) then gives $I_K[\mu^*] = I_K[\tilde{\mu}]$, so that μ^* is extremal. $\square$

5.1.14 Positive definiteness implies the extremal measure is unique. The following lemma, adapted from [26], shows that positive definiteness is sufficient to guarantee the uniqueness of the extremal measure, if it exists. We remark that uniqueness of the extremal measure is not expected for general semi-definite kernels, and that the characterization of the extremal measure in theorem 5.1.10 does not require the uniqueness of the extremal measure.

5.1.15 Lemma. *If K is a positive definite symmetric kernel that is bounded from below for which extremal measures exist and are always decent (so in particular* $\inf I_K$ *is finite), then the extremal measure is unique.*

Proof. Assume that μ_1 and μ_2 are any two extremal measures. Then we can write

$$I_K\left[\frac{1}{2}\mu_1 + \frac{1}{2}\mu_2\right] + \iint K d^2\left(\frac{1}{2}\mu_1 - \frac{1}{2}\mu_2\right) = \frac{1}{2}I_K[\mu_1] + \frac{1}{2}I_K[\mu_2] = \inf I_K,$$

where all integrals exist and and finite because of lemma 5.1.7 since μ_1 and μ_2 are decent by hypothesis. The second term on the left is positive by positive semi-definiteness, but we also have $I_K[(\mu_1+\mu_2)/2] \geq \inf I_K$. Therefore, we must have $\iint K d^2(\mu_1 - \mu_2) = 0$, and since K is positive definite, it then follows that $\mu_1 = \mu_2$. □

5.2 Variational inequalities for the kernels $K_{V,H}$

We now apply the contents of the previous section to symmetric kernels of the form we discussed in the previous chapter, i.e., of the form

$$K_{V,H}(x,y) = L_H(x,y) + \frac{1}{2}V(x) + \frac{1}{2}V(y), \tag{5.2.1}$$

where

$$L_H(x,y) := \log\frac{H(x,y)}{|x-y|}.$$

Most of the arguments are adapted from the case $H \equiv 1$ as presented in [7].

5.2.2 Assumptions. We will assume that $K_{V,H}$ satisfies the following assumptions: The kernel $K_{V,H}$ is of the form 5.2.1 and

- H and V are continuous, $H \geq 1$, and $V(x) - \log(x^2-1) \to \infty$ as $|x| \to \infty$ (these are the regularity conditions 4.2).
- $K_{V,H}$ is symmetric (which amounts to H being symmetric).

- $K_{V,H}$ is positive semi-definite.

We recall that we proved that under the regularity assumptions on V and H, the kernel $K_{V,H}$ is bounded from below, $\inf I_{V,H} = \inf_{\sigma\in\mathcal{M}^1(\mathbb{R})} I_{K_{V,H}}[\sigma]$ is well defined and finite, extremal measures attaining the infimum exist, and any extremal measure is decent (in fact, we proved that the infimum defining $\inf I_{V,H}$ can be taken over decent measures).

This implies in particular that all the statements from the previous section are applicable for these kernels.

5.2.3 Positive semi-definiteness only depends on $\boldsymbol{L_H}$. We also recall that the growth condition on V and the lower bound on H are what guarantee that $K_{V,H}$ is bounded from below, so that the integral $I_{K_{V,H}}[\sigma] = \iint K_{V,H}\, d^2\sigma$ exists as an element of $(-\infty,\infty]$ for any positive probability measure σ. The function L_H, however, does not have to be bounded from below (the prototypical example $L(x,y) = \log 1/|x-y|$ is certainly not), but our hypothesis on H guarantee that L_H is bounded from below on compact sets.

In particular, for any mean zero compactly supported measure ν we have

$$\iint K_{V,H}\, d^2\nu = \iint L_H\, d^2\nu,$$

since $\iint V d^2\nu = 0$, where the integral on the right is well defined since L_H is bounded from below on the compact support of $\nu\otimes\nu$. This proves that the positive (semi-) definiteness of kernels of the form (5.2.1) with H and V satisfying the regularity conditions only depends on L_H.

5.2.4 Remark on bound of $\boldsymbol{H}$. We remark that we can replace the condition $H \geq 1$ by $H \geq c_H > 0$ for some constant c_H in assumptions 5.2.2 and all statements in this section will also hold.

5.2.5 Proposition. *If $K_{V,H}$ satisfies assumptions 5.2.2, a decent measure $\mu \in \mathcal{M}^1(\mathbb{R})$ is extremal if and only if there exists a constant ℓ such that*

$$Q_{V,H}(x) := 2 \int \log \frac{H(x,y)}{|x-y|} d\mu(y) + V(x) = \ell, \quad \mu\text{-almost everywhere,}$$

and the inequality

$$\int Q_{V,H}(x) d\sigma(x) \geq \ell$$

holds for all decent measures σ.

Proof. The proof of the proposition follows from theorem 5.1.10 and the following two lemmas. □

5.2.6 Lemma. *If $K_{H,V}$ satisfies assumptions 5.2.2, and μ and σ are decent, then*

$$\iint K_{V,H}\, d\mu d\sigma = \iint L_H\, d\mu d\sigma + \frac{1}{2} \int V d\mu + \frac{1}{2} \int V d\sigma$$

where all integrals exist and are finite.

Proof. The integral $\iint L_H\, d\mu d\sigma$ is well defined as an element of $(-\infty, \infty]$ because L_H is bounded from below on the compact support of $\mu \otimes \sigma$ (see 5.2.3), and the integral $\iint K_{V,H}\, d\mu d\sigma$ exists and is finite by lemma 5.1.7. The last two integrals on the right are finite since V is continuous and the measures have compact support. □

5.2.7 Lemma. *If $K_{V,H}$ satisfies assumptions 5.2.2 and μ is a decent measure, then $\iint K_{V,H}\, d\mu d\sigma \geq \iint K_{V,H}\, d^2\mu$ for all decent measures σ if and only if there exists a constant ℓ such that*

$$Q_{V,H}(x) := 2 \int \log \frac{H(x,y)}{|x-y|} d\mu(y) + V(x) = \ell$$

μ-almost everywhere, and the inequality

$$\int Q_{V,H}(x) d\sigma(x) \geq \ell$$

holds for all decent measures σ.

Proof. (⇐) Regrouping the equality in lemma 5.2.6 we obtain

$$\begin{aligned}\iint 2K_{V,H}\, d\mu d\sigma &= \int \left[\int 2L_H(x,y)d\mu(y) + V(x)\right] d\sigma(x) + \int V(y)d\mu(y)\\ &\geq \ell + \int V(y)d\mu(y)\\ &= \int \left[\int 2L_H(x,y)d\mu(y) + V(x)\right] d\mu(x) + \int V(y)d\mu(y)\\ &= \iint 2K_{V,H}\, d^2\mu\end{aligned}$$

where the second inequality follows by hypothesis, and the third equality follows since the integrand is equal to ℓ μ-a.e., again by hypothesis.

(⇒) Using the equality in lemma 5.2.6 on both sides of $\iint 2K_{V,H}\, d\mu d\sigma \geq \iint 2K_{V,H}\, d^2\mu$ and canceling $\int V d\mu$ in both sides we obtain

$$\iint 2L_H(x,y)d\mu(y)d\sigma(x) + \int V d\sigma(x) \geq \iint 2L_H d^2\mu + \int V d\mu,$$

or

$$\int Q_{V,H}(x)d\sigma(x) = \int \left[2\int \log\frac{H(x,y)}{|x-y|}d\mu(y) + V(x)\right] d\sigma(x) \geq \ell \tag{5.2.8}$$

where

$$\ell := \iint 2L_H d^2\mu + \int V d\mu.$$

Let now $B := \{x : Q_{V,H}(x) < \ell\}$ and assume that $\mu(B) > 0$. Define

$$\sigma = \frac{1}{\mu(B)}\mu\mid_B$$

and note that σ is decent since μ is. Then we have $\int Q_{V,H}(x)d\sigma(x) < \ell$ which contradicts the inequality (5.2.8) which we know holds for all decent σ. Thus, we must have $\mu(B) = 0$. Using (5.2.8) with $\sigma = \mu$ then shows that $Q_{V,H}(x) = \ell$, μ-almost everywhere. □

5.2.9 Remark about the constant ℓ. We note that if μ is extremal, then the constant ℓ in the statement of the proposition is given by

$$\ell = 2\inf I_{V,H} - \int V d\mu,$$

as follows from the proof of the second lemma. In particular, the constant depends on μ, and there is no reason to expect the constants for two different extremal measures to be equal.

We finally arrive at a characterization of extremal measures which are absolutely continuous with respect to Lebesgue measure where there is no mention of an auxiliary measure σ.

5.2.10 Theorem. *If $K_{V,H}$ satisfies assumptions 5.2.2, a decent measure $\mu \in \mathcal{M}^1(\mathbb{R})$ with $d\mu(x) = \rho(x)dx$, and ρ continuous is extremal if and only if there exists a constant ℓ such that*

$$\begin{aligned} Q_{V,H}(x) = 2\int \log \frac{H(x,y)}{|x-y|} d\mu(y) + V(x) \quad &\geq \quad \ell, \quad \forall x \in \mathbb{R} \\ &= \quad \ell, \quad \text{on } \{x : \rho(x) > 0\} \end{aligned}$$

Proof. ($\Leftarrow$) This is automatic by proposition 5.2.5.

($\Rightarrow$) We first note that if $d\mu(x) = \rho(x)dx$ for some continuous ρ with compact support, then $Q_{V,H}(x)$ is continuous in x (one only needs the continuity of

$$\int \log \frac{1}{|x-y|} \rho(y)dy$$

since H and V are continuous and $H \geq 1$, and this is proven in corollary 13.1 in [24]).

Now, by proposition 5.2.5 we know that $Q_{V,H}(x) = \ell$, μ-almost everywhere. Assume by contradiction that there exists an $a \in \{x : \rho(x) > 0\}$ for which $Q_{V,H}(a) \neq \ell$. Then since $Q_{V,H}$ is continuous, we know that there is a δ for which $Q_{V,H}(x) \neq \ell$ for all $x \in (a-\delta, a+\delta)$. Moreover, since ρ is continuous, we know that $\{x : \rho(x) > 0\}$ is open, so we can also choose δ so that $(a-\delta, a+\delta) \subset \{x : \rho(x) > 0\}$. But then the interval $(a-\delta, a+\delta)$ is a set where $f(x) \neq \ell$ and for which $\mu[(a-\delta, a+\delta)] \neq 0$ which gives a contradiction. Thus, $Q_{V,H} \equiv \ell$ on the set $\{x : \rho(x) > 0\}$.

For the inequality $Q_{V,H} \geq \ell$, note that by proposition 5.2.5 we know that

$$\int Q_{V,H}(x) d\sigma(x) \geq \ell$$

for any decent σ. However, if there was an a for which $Q_{V,H}(a) < \ell$, then again, by the continuity of $Q_{V,H}$, we would be able to find a δ for which $Q_{V,H}(x) < \ell$ for all $x \in (a-\delta, a+\delta)$, and then, if we define

$$\sigma = \frac{1}{2\delta}\chi_{(a-\delta,a+\delta)}dx$$

where $\chi_{(a-\delta,a+\delta)}$ is the characteristic function of the interval $(a-\delta, a+\delta)$, then σ will be decent since Q is continuous and so $I_{V,H}[\sigma] = (1/2)\int (Q_{V,H} + V)\, d\sigma < \infty$. However, we would then have $\int Q_{V,H}(x)d\sigma(x) < \ell$ which contradicts the above inequality since σ is decent. □

5.2.11 The saddle-point equation. We remark that if V, H and ρ are sufficiently well behaved, then the function

$$Q_{V,H}(x) = 2\int \log\frac{H(x,y)}{|x-y|}d\mu(y) + V(x),$$

showing up in theorem 5.2.10 will be differentiable, and moreover, the equality on the support of ρ from theorem 5.2.10 will be equivalent to

$$0 = Q'_{V,H}(x) = V'(x) + 2\int \frac{\partial_x H(x,y)}{H(x,y)}\rho(y)dy - 2\,p.v.\int \frac{\rho(y)}{x-y}dy, \quad x \in \operatorname{supp}\rho.$$

The benefit of this equation is that there no longer is any mention of the unknown constant ℓ for the equality on the support, which in principle could be hard to find. This last equation is what physicists would call the **saddle-point equation** for ρ (see 3.1.8). It is know that the inequality off the support does not follow from the equality, so if ρ satisfies the saddle-point equation, it is not necessarily an equilibrium measure.

For example, for the case $H \equiv 1$, $V(x) = x^2/2$ where one knows that the there is a unique extremal measure with density $\rho(x) = (2\pi)^{-1}\sqrt{4-x^2}$ (see [7] for example) supported on the interval $[-2,2]$, one may take instead for $a > 0$ the density

$$\rho_a(x) = \frac{\sqrt{a^2-x^2}}{2\pi} + \frac{4-a^2}{4\pi\sqrt{a^2-x^2}},$$

supported on the interval $[-a, a]$, which agrees with ρ when $a = 2$. Note that $\rho_a \geq 0$ and $\int_{-a}^{a} \rho_a(x)dx = 1$, so that ρ_a is in fact a probability density. One can verify that ρ_a satisfies the saddle point equation on its support $[-a, a]$, but as we know by uniqueness, the only one that will satisfy the variational inequalities off the support will be $\rho_2 = \rho$. This example comes from the following considerations: The saddle-point equation in this situation takes the form

$$2\, p.v. \int \frac{\psi(y)}{x - y} dy = x, \quad x \in \text{supp } \psi,$$

which one can write as

$$g(x + i0) + g(x - i0) = x, \quad x \in \text{supp } \psi,$$

where $g(z) := \int_{\mathbb{R}} \frac{\psi(y)}{z-y} dy$. General considerations (transforming it into a Riemann Hilbert problem) show that this is solved by

$$g(z) = \frac{1}{2}\left(z - \sqrt{z^2 - a^2}\right)$$

for some a, and one can recover the density through the formula $g(x+i0)-g(x-i0) = -2i\pi\rho(x)$. Now, one can tweak $g(z)$ by adding another term

$$\widetilde{g}(z) = \frac{1}{2}\left(z - \sqrt{z^2 - a^2}\right) + \frac{c}{\sqrt{z^2 - a^2}}$$

for some constant c, since this does not break the property

$$\widetilde{g}(x + i0) + \widetilde{g}(x - i0) = x, \quad x \in (-a, a).$$

The corresponding ψ is again given by $\widetilde{g}(x + i0) - \widetilde{g}(x - i0)$ which gives the ρ_a above, after one choses c so that it has mass 1.

5.2.12. If one further knows that the density is supported on a union of finite intervals $[a_1, b_1], \ldots, [a_n, b_n]$ where $b_i < a_{i+1}$, then the the conditions in the theorem can be replaced by

$$\begin{aligned}
Q'_{V,H}(x) &= 0 \quad \text{for } x \in \bigcup (a_i, b_i) \\
\int_x^{a_1} Q'_{V,H}(x)dx &\leq 0 \quad \text{for } x < a_1 \\
\int_{b_i}^{x} Q'_{V,H}(x)dx &\geq 0 \quad \text{for } b_i < x < a_{i+1} \\
\int_{b_n}^{x} Q'_{V,H}(x)dx &\geq 0 \quad \text{for } x > b_n \\
\int_{b_i}^{a_{i+1}} Q'_{V,H}(x)dx &= 0 \quad \text{for } i = 1 \ldots, n-1,
\end{aligned}$$

where one manages to remove all mention of the unknown constant ℓ from the statement of the theorem. We stress again that this all relies on V, H and ρ being well behaved, and we will not make this more precise at the moment. We will work out all the details for the case of colored triangulations in chapter 6.

5.3 Positive definiteness of $K_{V,H}$ with $H(x, y) = \sqrt{1 + t^2(x + y)^2}$

We now further specialize to analyze kernels $K_{V,t}$ of the form

$$K_{V,t}(x, y) := \log \frac{\sqrt{1 + t^2(x + y)^2}}{|x - y|} + \frac{1}{2}V(x) + \frac{1}{2}V(y) \tag{5.3.1}$$

where $t \in \mathbb{R}$, V is continuous, and $V(x) - \log(x^2 + 1) \to \infty$ as $|x| \to \infty$ (so that $H(x, y) = \sqrt{1 + t^2(x + y)^2}$ and V satisfy the regularity conditions 4.2). We will prove that these kernels are always positive definite, and so in particular they satisfy assumptions 5.2.2 which will make the criteria characterizing equilibrium measures described in the previous section applicable.

As discussed in 5.2.3, the positive (semi-)definiteness of the kernel $K_{V,t}$ only depends on the logarithmic piece

$$L_t(x, y) := \log \frac{\sqrt{1 + t^2(x + y)^2}}{|x - y|},$$

and so we only need to prove that for any signed measure ν with mean zero and compact support one has

$$\iint L_t \left(d\nu^+ d\nu^+ + d\nu^- d\nu^-\right) \geq 2 \iint L_t d\nu^+ d\nu^-, \tag{5.3.2}$$

where $\nu = \nu^+ - \nu^-$ is the decomposition of the measure into its positive and negative parts.

Regarding the part of L_t involving $\log 1/|x-y|$, Deift [7, p. 143] proves that

$$\iint \log \frac{1}{|x-y|} \left(d^2\nu^+ + d^2\nu^-\right) = 2 \iint \log \frac{1}{|x-y|} d\nu^+ d\nu^- + \frac{1}{2} \int \frac{|\mathcal{F}\nu|^2}{|k|} dk, \tag{5.3.3}$$

where

$$\mathcal{F}\nu(k) := \int e^{ixk} d\nu(x)$$

is the Fourier transform of the measure. This in particular implies that the kernel $\log 1/|x-y|$ is positive definite since the equality of the double integrals in (5.3.3) when they are finite (see the definition of positive definiteness) holds only if $\mathcal{F}\nu = 0$, or equivalently only if $\nu = 0$.

We will find a similar identity for $\log \sqrt{1 + t^2(x+y)^2}$ which we will combine with (5.3.3) to prove positive definiteness of L_t. Before going into the details, we describe a heuristic argument explaining where identity (5.3.3) and the analogous one for $\log \sqrt{1 + t^2(x+y)^2}$ come from.

5.3.4 Heuristics. The basic strategy is to write integrals of the form

$$\iint g(x \pm y) d\nu(x) d\nu(y)$$

in terms of the Fourier transforms of g and ν. We will ignore all issues involving whether the integrals are defined or not, and whether hypothesis needed to apply theorems hold or not.

If we write

$$\iint g(x-y) d\nu(x) d\nu(y) = \int (g * \nu)\, d\nu(x),$$

and then use Plancherel's theorem (which with this convention for the Fourier transform takes the form $\int f\overline{h} = (2\pi)^{-1}\int \mathcal{F}f\overline{\mathcal{F}h}$), we obtain

$$\iint g(x-y)d\nu(x)d\nu(y) = \frac{1}{2\pi}\int \mathcal{F}\left(g*\nu\right)\overline{\mathcal{F}\nu}\,dk = \frac{1}{2\pi}\int \mathcal{F}g\left|\mathcal{F}\nu\right|^2 dk.$$

This last equation applied to $g(x) = \log 1/|x|$ gives Deift's formula (5.3.3) above, since

$$\mathcal{F}\left(\log\frac{1}{|x|}\right) = \frac{\pi}{|k|} + 2\pi\gamma\delta(k),$$

where γ is the Euler constant and $\delta(k)$ is a point mass at the origin (recall that $\int d\nu = 0$).

For integrals involving the sum $x+y$ instead of $x-y$ (such as $\log\sqrt{1+t^2(x+y)^2}$), with the extra assumption that g is even and real valued we can write

$$\iint g(x+y)d\nu(x)d\nu(y) = \int\left(\int g(x-y)d\nu(y)\right)d\tilde{\nu}(x),$$

where $\tilde{\nu}(A) = \nu(-A)$, and equality holds because g is even. Then, the same arguments as above give

$$\iint g(x+y)d\nu(x)d\nu(y) = \frac{1}{2\pi}\int \mathcal{F}g\,\mathcal{F}\nu\,\overline{\mathcal{F}\tilde{\nu}}dk = \frac{1}{2\pi}\operatorname{Re}\int \mathcal{F}g\left(\mathcal{F}\nu\right)^2 dk,$$

where the second equality follows because $\mathcal{F}\tilde{\nu} = \overline{\mathcal{F}\nu}$ and the fact that we know that the integral on the left hand side is real.

For example, for $g(x) = \log\sqrt{1+t^2x^2}$ one expects to have

$$\iint \log\sqrt{1+t^2(x+y)^2}d\nu(x)d\nu(y) = -\frac{1}{2}\operatorname{Re}\int \frac{e^{-|k|/t}}{|k|}\left(\mathcal{F}\nu\right)^2 dk, \tag{5.3.5}$$

since

$$\mathcal{F}\left(\log\sqrt{1+t^2x^2}\right) = -\frac{\pi e^{-|k|/t}}{|k|}.$$

We will prove identity (5.3.5) in the proof of the following proposition. This concludes the heuristics.

5.3.6 Proposition. *The kernels* $K_{V,t}$ *from* (5.3.1) *with* $t\in\mathbb{R}$ *are positive definite.*

Proof. The case when $t = 0$ is proved in [7] and was outlined above, so from now on we assume that $t > 0$. We will use the following identity [7, p. 144]: For any $\varepsilon > 0$ one has

$$\log\left(s^2 + \varepsilon^2\right) = \log \varepsilon^2 - 2\,\mathrm{Re} \int_0^\infty e^{-\varepsilon k} \frac{e^{isk} - 1}{k} dk.$$

Writing $\log \sqrt{1 + t^2(x+y)^2} = \log t + 2^{-1} \log\left(t^{-2} + (x+y)^2\right)$ and using this identity with $\varepsilon = 1/t$ and $s = x + y$ we obtain

$$\log \sqrt{1 + t^2(x+y)^2} = -\mathrm{Re} \int_0^\infty e^{-k/t} \frac{e^{i(x+y)k} - 1}{k} dk.$$

Thus, for any compactly supported mean zero measure ν we have

$$\begin{aligned} \iint \log \sqrt{1 + t^2(x+y)^2}\, d^2\nu &= -\mathrm{Re} \int_0^\infty \frac{e^{-k/t}}{k} \iint \left[e^{i(x+y)k} - 1\right] d^2\nu dk \\ &= -\mathrm{Re} \int_0^\infty \frac{e^{-k/t}}{k} \left(\mathcal{F}\nu\right)^2 dk \\ &= -\frac{1}{2}\,\mathrm{Re} \int \frac{e^{-|k|/t}}{|k|} \left(\mathcal{F}\nu\right)^2 dk \end{aligned}$$

where the second equality follows because $\int d\nu = 0$, and the third because $\mathcal{F}\nu(-k) = \overline{\mathcal{F}\nu(k)}$ since ν is real. This is precisely the identity (5.3.5) we gave in the heuristics above. Note that all the integrals are finite because ν has compact support and $\mathcal{F}\nu(k)$ is analytic and vanishes at $k = 0$.

Combining (5.3.3) and (5.3.5), where only (5.3.3) can involve infinite quantities gives

$$\iint L_t \left(d^2\nu^+ + d^2\nu^-\right) = 2 \iint L_t d\nu^+ d\nu^- + \int \frac{|\mathcal{F}\nu|^2 - e^{-|k|/t}\,\mathrm{Re}\left[\left(\mathcal{F}\nu\right)^2\right]}{2|k|} dk, \tag{5.3.7}$$

and so, to prove the positive semi-definiteness of L_t we only need to prove the integral on the right is non-negative. However, this follows from the fact that the integrand itself on the integral on the right is non-negative, which one can see by writing

$$|\mathcal{F}\nu|^2 - e^{-|k|/t}\,\mathrm{Re}\left[\left(\mathcal{F}\nu\right)^2\right] = u(k)^2(1 - e^{-|k|/t}) + v(k)^2(1 + e^{-|k|/t}) \geq 0 \tag{5.3.8}$$

where

$$\mathcal{F}\nu(k) = u(k) + iv(k),$$

is the expression of the Fourier transform of ν in terms of its real and imaginary parts. For positive definiteness, note that if $\iint L\,(d^2\nu^+ + d^2\nu^-)$ and $2\iint L d\nu^+ d\nu^-$ are finite and equal (see definition of positive definiteness), then by (5.3.7) we must have

$$\int \frac{|\mathcal{F}\nu|^2 - e^{-|k|/t}\,\mathrm{Re}\left[(\mathcal{F}\nu)^2\right]}{2|k|} dk = 0$$

which by (5.3.8) is equivalent to $u \equiv v \equiv 0$. □

5.3.9 Corollary. *There is a unique extremal measure for the kernel $K_{V,t}$ for each $t \in \mathbb{R}$. A compactly supported measure with continuous density ρ_t^V is the extremal measure if and only if there exists a constant $\ell = \ell(t)$ such that*

$$\begin{aligned} Q_{V,t}(x) := 2\int \log \frac{\sqrt{1+t^2(x+y)^2}}{|x-y|}\rho_t^V(y)dy + V(x) \quad &\geq \quad \ell, \quad \forall x \in \mathbb{R} \\ &= \quad \ell, \quad \textit{on } \operatorname{supp}\rho_t^V. \end{aligned}$$

5.4 Some results regarding the support of the extremal density for the case $H(x,y) = \sqrt{1+t^2(x+y)^2}$

In this section we show that one can also generalize arguments from [8] regarding the case $H \equiv 1$, to prove that the support of the extremal measure of the kernel

$$K_{V,t}(x,y) = \log \frac{\sqrt{1+t^2(x+y)^2}}{|x-y|} + \frac{1}{2}V(x) + \frac{1}{2}V(y), \tag{5.4.1}$$

is a union of finitely many intervals, under the assumptions that V is polynomial of even degree with positive leading coefficient, and that the extremal measure has a continuous density ρ_t^V that satisfies two technical assumptions (which are expected to hold). More explicitly, we will prove the following proposition.

5.4.2 Proposition. *If V is a polynomial of even degree with positive leading term, $t \in \mathbb{R}$, and if the (unique) extremal measure $\mu_{V,t}$ for the kernel $K_{V,t}$ (5.4.1) has a*

Hölder continuous density ρ_t^V for which $\int (\log|x-y|)\, \rho_t^V(y)dy$ is differentiable in x, then $\mu_{V,t}$ is supported on finitely many intervals, and the density ρ_t^V is in fact analytic in the interior of its support.

We remark that it is expected that in this situation the extremal measure does have have a density ρ_t^V which is Hölder continuous and for which $\int (\log|x-y|)\, \rho_t^V(y)dy$ is differentiable in x (see [8]), but we do not have a proof of this fact at the moment. We provide the proof of proposition 5.4.2 in paragraph 5.4.15, after we develop the necessary ingredients for its proof.

We also remark that under the assumptions on the extremal measure from proposition 5.4.2, the above result will apply to the leading order asymptotics of the partition function

$$\widetilde{Z}_N(t;t_1,\dots,t_{2n}) := \iiint \exp\left\{N\,\mathrm{Tr}\left[it\,(ABC+ACB)-\sum_{j=1}^{2n} t_jA^j\right]\right\} d\mu_N(A)d\mu_N(B)d\mu_N(C),$$

since one can check that the computations that reduce $\widehat{Z}_N(t)$ (1.3.1) to an N dimensional integral over the eigenvalues of matrix A (which we review in section 6.1), also show that the asymptotics of $\widetilde{Z}_N(t;t_1,\dots,t_{2n})$ are related to the functional $I_{H,V}$ with $H(x,y)=\sqrt{1+t^2(x+y)^2}$ and

$$V_{t_1,\dots,t_{2n}}(x)=\frac{1}{2}x^2+t_1x+t_2x^2+\dots t_{2n}x^{2n}.$$

By the contents of chapter 2, this partition function is related to combinatorial counts of maps with tri-valent vertices with the three colors (of types q_1 and q_2), together with vertices of color $\mathcal{A}$ of various valences.

To fix notation in this section, we define

$$H_t(x,y) := \sqrt{1+t^2(x+y)^2}. \tag{5.4.3}$$

5.4.4 A formula from the saddle-point equation. We first remark that if the density ρ_t^V is continuous, and $\int (\log|x-y|)\,\rho_t^V(y)dy$ is differentiable in x, then the saddle-point equation (see 5.2.11), which corresponds to the equality on the support of ρ_t^V in corollary 5.3.9, takes the form

$$0 = V'(x) + 2\int \frac{\partial_x H_t(x,y)}{H(x,y)}\rho_t^V(y)dy - 2\,p.v.\int \frac{\rho_t^V(y)}{x-y}dy, \quad x \in \operatorname{supp}\rho_t^V. \tag{5.4.5}$$

Now, defining

$$W_{V,t}(z) := \int \frac{\rho_t^V(y)}{z-y}dy, \quad z \notin \operatorname{supp}\rho_t^V,$$

then, under the assumption that $\rho_t^V(y)$ is Hölder continuous, the Sokhotski-Plemelj formulas

$$W_{V,t}(x \pm i0) = p.v.\int_{\mathbb{R}} \frac{\rho_t^V(y)}{x-y}dy \mp i\pi\rho_t^V(x), \quad x \in \operatorname{supp}\rho_t^V,$$

combined with equation (5.4.5) show that

$$W_{V,t}(x \pm i0) = \frac{1}{2}V'(x) + \int \frac{\partial_x H_t(x,y)}{H_t(x,y)}\rho_t^V(y)dy \mp i\pi\rho_t^V(x), \quad x \in \operatorname{supp}\rho_t^V.$$

Define

$$R_{V,t}(z) := W_{V,t}(z) - \left(\frac{1}{2}V'(z) + \int \frac{\partial_1 H_t(z,y)}{H_t(z,y)}\rho_t^V(y)dy\right), \tag{5.4.6}$$

(where the $\partial_1 H_t(z,y)$ term means $\partial_x(H_t(x,y))\,|_{x=z}$) for z outside $\operatorname{supp}\rho_t^V$ *and* outside the set of points where $\partial_1 H_t(z,y)/H_t(z,y)$ vanishes when $y \in \operatorname{supp}\rho_t^V$. This corresponds (under the assumption that t is real) to assuming that z does not lie on the support $\operatorname{supp}\rho_t^V$ or on the translates of the support of ρ_t^V given by $\pm i/t - \operatorname{supp}\rho_t^V$, since

$$\frac{\partial_1 H_t(x,y)}{H_t(x,y)} = \frac{t^2(x+y)}{1+t^2(x+y)^2}. \tag{5.4.7}$$

The above equations show that

$$\begin{aligned} R_{V,t}(x \pm i0) &= \mp i\pi\rho_t^V(x), \quad x \in \operatorname{supp}\rho_t^V \\ &\in \mathbb{R}, \qquad x \notin \operatorname{supp}\rho_t^V. \end{aligned} \tag{5.4.8}$$

We now show that one can express $(R_{V,t}(z))^2$ as a function that is analytic for $z \in \mathbb{R}$. We will present the arguments for general V and H.

5.4.9 The loop equations. We provide a precise derivation in the present situation of what are known in the physics literature as the *loop equations*. We will have to make assumptions on V and H for the arguments to hold, which we will make precise along the way.

Following ideas presented in [19, p. 161] for the case $H \equiv 1$, we make the substitution $x_i = y_i + \alpha\phi(y_i)$ in the integral

$$Z_N^{V,H} = \int_{\mathbb{R}^N} e^{-N\sum V(x_i)} \prod_{i<j}(x_i - x_j)^2 \prod_{1\le i,j\le N} \frac{1}{H(x_i, x_j)} d^N x,$$

where ϕ has continuous derivative with ϕ' bounded form below. This substitution is a diffeomorhism as long as $\alpha\phi'(y) \neq -1$ for all y, which can be guaranteed if $\alpha \geq 0$ is small enough. If one then differentiates with respect to α, and sets $\alpha = 0$, one obtains, after much manipulation, the identity

$$\begin{aligned} 0 \;=\; & N \int_{\mathbb{R}} \phi'(s) u_N^{V,H}(s) ds - N^2 \int_{\mathbb{R}} V'(s)\phi(s) u_N^{V,H}(s) ds \\ & + N(N-1) \iint_{\mathbb{R}^2} \frac{\phi(s) - \phi(r)}{s - r} u_N^{V,H}(s, r) ds dr \\ & - N^2 \iint_{\mathbb{R}^2} \frac{\phi(s)\partial_1 H(s,r) + \phi(r)\partial_2 H(s,r)}{H(s,r)} u_N^{V,H}(s,r) ds dr, \end{aligned} \tag{5.4.10}$$

where $u_N^{V,H}(x_1)$ and $u_N^{V,H}(x_1, x_2)$ are as in theorem 4.3.5, and $\partial_i H$ denotes partial derivative of H with respect to its i-th variable. The idea to differentiate with respect to a parameter after a change of variables is common in quantum field theory, and the equations that one obtains are referred to as *loop equations*.

Now, under the assumption that H is symmetric and differentiable (such as $H = H_t$ as in (5.4.3)), we have

$$\partial_2 H(s,r) = \partial_r(H(s,r)) = \partial_r(H(r,s)) = \partial_1 H(r,s),$$

and then using the fact that $u_N^{V,H}(s,r)$ is symmetric too (independent of the symmetry of H), we have

$$\begin{aligned}\iint_{\mathbb{R}^2} \phi(r)\frac{\partial_2 H(s,r)}{H(s,r)}u_N^{V,H}(s,r)dsdr &= \iint_{\mathbb{R}^2} \phi(r)\frac{\partial_1 H(r,s)}{H(r,s)}u_N^{V,H}(r,s)dsdr,\\ &= \iint_{\mathbb{R}^2} \phi(s)\frac{\partial_1 H(s,r)}{H(s,r)}u_N^{V,H}(s,r)dsdr,\end{aligned}$$

where in the last equality we changed the roles of r and s in the integral. Using this in the last integral in (5.4.10) then gives the identity

$$\begin{aligned}0 &= N\int \phi'(s)u_N^{V,H}(s)ds - N^2\int V'(s)\phi(s)u_N^{V,H}(s)ds\\ &\quad + N(N-1)\iint \frac{\phi(s)-\phi(r)}{s-r}u_N^{V,H}(s,r)dsdr\\ &\quad - 2N^2\iint \phi(s)\frac{\partial_1 H(s,r)}{H(s,r)}u_N^{V,H}(s,r)dsdr.\end{aligned}$$

Dividing by N^2, assuming that the extremal measure $\mu_{V,H}$ is unique, taking the limit $N\to\infty$, and assuming that theorem 4.3.5 is applicable for each term in the above equation, we obtain

$$\begin{aligned}0 &= -\int V'(s)\phi(s)d\mu_{V,H}(s) + \iint \frac{\phi(s)-\phi(r)}{s-r}d\mu_{V,H}(s)d\mu_{V,H}(r)\\ &\quad - 2\iint \frac{\partial_1 H(s,r)}{H(s,r)}\phi(s)d\mu_{V,H}(s)d\mu_{V,H}(r).\end{aligned} \tag{5.4.11}$$

Since all of the expressions in the above discussion from (5.4.10) to (5.4.11) are linear in ϕ, it can be separately applied to the real and imaginary parts of

$$\phi(y) := \frac{1}{z-y},$$

where $z\notin\mathbb{R}$, and then combined together. This gives

$$\begin{aligned}0 &= -\int \frac{V'(s)}{z-s}d\mu_{V,H}(s) + \iint \frac{1}{(z-s)(z-r)}d\mu_{V,H}(s)d\mu_{V,H}(r)\\ &\quad - 2\iint \frac{\partial_1 H(s,r)}{H(s,r)}\frac{1}{z-s}d\mu_{V,H}(s)d\mu_{V,H}(r),\end{aligned}$$

which one can rewrite in terms of

$$W_{V,H}(z) := \int \frac{1}{z-s} d\mu_{V,H}(s), \qquad z \notin \operatorname{supp} \mu_{V,H}, \tag{5.4.12}$$

as

$$\begin{aligned} 0 &= \int \frac{V'(z)-V'(s)}{z-s} d\mu_{V,H}(s)ds - V'(z)W_{V,H}(z) + (W_{V,H}(z))^2 \\ &\quad + 2\iint \left(\frac{\partial_1 H(z,r)}{H(z,r)} - \frac{\partial_1 H(s,r)}{H(s,r)} \right) \frac{1}{z-s} d\mu_{V,H}(s) d\mu_{V,H}(r) \\ &\quad -2\, W_{V,H}(z) \int \frac{\partial_1 H(z,r)}{H(z,r)} d\mu_{V,H}(r). \end{aligned}$$

Completing the square for $W_{V,H}(z)$ then gives

$$\begin{aligned} (R_{V,H}(z))^2 &= \left(\frac{1}{2} V'(z) + \int \frac{\partial_1 H(z,r)}{H(z,r)} d\mu_{V,H}(r) \right)^2 \\ &\quad - \int \frac{V'(z)-V'(s)}{z-s} d\mu_{V,H}(s)ds \\ &\quad -2 \iint \left(\frac{\partial_1 H(z,r)}{H(z,r)} - \frac{\partial_1 H(s,r)}{H(s,r)} \right) \frac{1}{z-s} d\mu_{V,H}(s) d\mu_{V,H}(r), \end{aligned} \tag{5.4.13}$$

where

$$R_{V,H}(z) := W_{V,t}(z) - \left(\frac{1}{2} V'(z) + \int \frac{\partial_1 H(z,r)}{H(z,r)} d\mu_{V,H}(r) \right). \tag{5.4.14}$$

We remark that the deduction of (5.4.13) depended on the fact that the extremal measure of $K_{V,H}$ is unique, that H is symmetric and differentiable with continuous partial derivatives, and that all the applications of theorem 4.3.5 were valid.

5.4.15 Proof of proposition 5.4.2. If we now go back to the case with $H_t(x,y) = \sqrt{1+t^2(x+y)^2}$, then all of the arguments in 5.4.9 are applicable since $\partial_1 H_t$ is continuous and bounded on $\mathbb{R}^2$, and for the limit of $\int V'(s)\phi(s)u_N^{V,H}(s)ds$ we can write

$$\int V'(s)\phi(s)u_N^{V,H}(s)ds = -\int \frac{V'(z)-V'(s)}{z-s} u_N^{V,H}(s)ds + V'(z) \int \frac{1}{z-s} u_N^{V,H}(s)ds$$

and use the fact that the first integrand on the left is a polynomial if V is a polynomial (see 4.8.6).

Thus, we find that (see (5.4.6), (5.4.13) and (5.4.14))

$$
\begin{aligned}
(R_{V,t}(z))^2 = &\left(\frac{1}{2}V'(z) + \int \frac{t^2(z+r)}{1+t^2(z+r)^2}\rho_t^V(r)dr\right)^2 - \int \frac{V'(z)-V'(s)}{z-s}\rho_t^V(s)ds \\
&- 2\iint \frac{t^2(1+t^2(z+r)(s+r))}{(1+t^2(z+r)^2)\,(1+t^2(s+r)^2)}\rho_t^V(s)\rho_t^V(r)dsdr,
\end{aligned}
$$

where we note that the last integrand is not singular at $z = s$ (compare to (5.4.13)). We denote by $F_{V,t}(z)$ the function of z on the right hand side of this equality.

Of fundamental importance is to note that since V is a polynomial, then $F_{V,t}(z)$ is analytic and real valued for $z \in \mathbb{R}$. Moreover, letting z approach the real axis to $x \in \mathbb{R}$ (say from above), and using (5.4.8), we obtain

$$
\begin{aligned}
F_{V,t}(x) &= -(2\pi\rho_t^V(x))^2, \quad x \in \operatorname{supp}\rho_t^V \\
&\geq 0, \qquad x \notin \operatorname{supp}\rho_t^V,
\end{aligned}
$$

which immediately implies the following equality of sets

$$\{x \in \mathbb{R} \mid \rho_t^V(x) > 0\} = \{x \in \mathbb{R} \mid F_{V,t}(x) < 0\}.$$

We know that the set on the left is bounded since the support of ρ_t^V is compact, and since $F_{V,t}(z)$ is analytic on $\mathbb{R}$, it can only change sign a finite number of times on this bounded set. This implies that $\operatorname{supp}\rho_t^V$ consists of finitely many intervals. The fact that ρ_t^V is analytic in the interior of its support follows from the fact that $F_{V,t}$ is analytic on $\mathbb{R}$. □

5.4.16 General formulas that do not assume the symmetry of H. We remark that even though the arguments in 5.4.9 were presented under the assumption that H is symmetric, one can obtain weaker formulas without assuming symmetry of H. Even though we will not be using these more general equations since the H for colored triangulations is symmetric, we supply the details for possible future applications.

By using the identity

$$\frac{\phi(s)\partial_1 H(s,r) + \phi(r)\partial_2 H(s,r)}{H(s,r)} = \phi(s)\frac{\partial_1 H(z,r)}{H(z,r)} + \phi(r)\frac{\partial_2 H(s,z)}{H(s,z)} + T[H,\phi,r,s,z],$$

where

$$T[H,\phi,r,s,z] := \phi(s)\left(\frac{\partial_1 H(s,r)}{H(s,r)} - \frac{\partial_1(z,r)}{H(z,r)}\right) + \phi(r)\left(\frac{\partial_2 H(s,r)}{H(s,r)} - \frac{\partial_2(s,z)}{H(s,z)}\right),$$

we can write equation (5.4.10) as

$$(5.4.17)\qquad \begin{aligned} 0 = {} & \frac{1}{N}\int \phi'(s)u_N^{V,H}(s)ds - \int (V'(z)-V'(s))\phi(s)u_N^{V,H}(s)ds \\ & - V'(z)\int \phi(s)u_N^{V,H}(s)ds + \frac{N(N-1)}{N^2}\iint \frac{\phi(s)-\phi(r)}{s-r}u_N^{V,H}(s,r)dsdr \\ & - \iint T[H,\phi,r,s,z]u_N^{V,H}(s,r)dsdr \\ & - \iint \phi(s)\left(\frac{\partial_1 H(z,r)}{H(z,r)} + \frac{\partial_2 H(r,z)}{H(r,z)}\right)u_N^{V,H}(s,r)dsdr, \end{aligned}$$

where in the last integral we have used the fact that $u_N^{V,H}(s,r) = u_N^{V,H}(r,s)$ to change the roles of s and r in the second term.

Under the assumption that the hypothesis for theorem 4.3.5 are satisfied for each term, then (5.4.10) becomes

$$\begin{aligned} 0 \;=\; & -\int V'(s)\phi(s)\mu_{V,H}(s) + \iint \frac{\phi(s)-\phi(r)}{s-r}d\mu_{V,H}(s)d\mu_{V,H}(r) \\ & - \iint \frac{\phi(s)\partial_s H(s,r) + \phi(r)\partial_r H(s,r)}{H(s,r)}d\mu_{V,H}(s)d\mu_{V,H}(r) \end{aligned}$$

while (5.4.17) becomes

$$\begin{aligned} 0 = {} & \int \frac{V'(z)-V'(s)}{z-s}d\mu_{V,H}(s) - V'(z)\int \frac{1}{z-s}d\mu_{V,H}(s) \\ & + \iint \frac{1}{(z-r)(z-s)}d\mu_{V,H}(s)d\mu_{V,H}(r) \\ & - \iint T[H,\phi,r,s,z]d\mu_{V,H}(s)d\mu_{V,H}(r) \\ & - \iint \phi(s)\left(\frac{\partial_1 H(z,r)}{H(z,r)} + \frac{\partial_2 H(r,z)}{H(r,z)}\right)d\mu_{V,H}(s)d\mu_{V,H}(r), \end{aligned}$$

or, in terms of $W_{V,H}(z)$ (see (5.4.12))

$$\begin{aligned} 0 = &\int \frac{V'(z) - V'(s)}{z - s} d\mu_{V,H}(s) - V'(z) W_{V,H}(z) + \ (W_{V,H}(z))^2 \\ - &\iint T[H, \phi, r, s, z] d\mu_{V,H}(s) d\mu_{V,H}(r) \\ - &\ W_{V,H}(z) \int \left(\frac{\partial_1 H(z,r)}{H(z,r)} + \frac{\partial_2 H(r,z)}{H(r,z)} \right) d\mu_{V,H}(r). \end{aligned}$$

Chapter 6
Asymptotics for the partition function for colored triangulations

In this chapter we go back to the study of the specific integral

$$\widehat{Z}_N(t) := \iiint \exp\left\{N \operatorname{Tr}\left[it(ABC + ACB)\right]\right\} d\mu_N(A)d\mu_N(B)d\mu_N(C)$$

related to colored triangulations.

In section 6.1 we provide the details of why

$$\widehat{Z}_N(t) = \frac{Y_N(t)}{Y_N(0)} =: \widehat{Y}_N(t), \qquad t \in \mathbb{R} \tag{6.0.1}$$

where

$$Y_N(t) := \int_{\mathbb{R}^N} e^{-N\sum \lambda_i^2/2} \prod_{i<j}(\lambda_i - \lambda_j)^2 \prod_{i,j} \frac{1}{\sqrt{1 + t^2(\lambda_i + \lambda_j)^2}} d^N\lambda,$$

with $d^N\lambda = d\lambda_1 \ldots d\lambda_N$. As far as we know, this formula first appeared in [6].

As we discussed in chapter 4, the leading order asymptotics of $Y_N(t)$ are related to the minimizer of the functional

$$I_t[\mu] := \iint \left[\log \frac{\sqrt{1 + t^2(x+y)^2}}{|x-y|} + \frac{1}{2}V(x) + \frac{1}{2}V(y)\right] d\mu(x)d\mu(y) = \iint K_t\, d^2\mu,$$

where

$$K_t(x, y) := \log \frac{\sqrt{1 + t^2(x+y)^2}}{|x-y|} + \frac{1}{2}V(x) + \frac{1}{2}V(y),$$

and

$$V(x) = \frac{1}{2}x^2.$$

In section 5.3 we proved that the kernels K_t are positive definite for any $t \in \mathbb{R}$. This implies that the extremal measure attaining the infimum

$$\inf I_t := \inf_{\mu \in \mathcal{M}^1(\mathbb{R})} I_t[\mu],$$

is unique for each $t \in \mathbb{R}$, and, as stated in corollary 5.3.9, that the criterion to characterize the density of an extremal measure (if such a density exists) described by theorem 5.2.10 is applicable.

In this chapter we return to the heuristics from the physics litarature we discussed in section 3.2, and show that the density proposed by the physicists is, in fact, the density of the extremal measure of I_t. The construction of the density ρ_t is considerably involved, and will make up a large portion of this chapter. Once ρ_t is constructed, we will discuss its analytic dependence on the parameter t, which will be used to show that $\lim_{N\to\infty} N^{-2} \log \widehat{Z}_N(t)$ is analytic in t around $t = 0$.

6.1 Reduction to an N dimensional integral.

In this section we prove the identity (6.0.1). We start by separating the integral over the matrix C, write the rescaled GUE measure $d\mu_N(C)$ explicitly in terms of dC as in (1.1.4), and complete the square to obtain

$$\begin{aligned} \widehat{Z}_N(t) &= \frac{1}{Z_N^{GUE}} \iint \Bigg(\exp\left\{-Nt^2 \mathrm{Tr}\left[ABAB + A^2B^2\right]\right\} \times \\ & \int \exp\left\{-\frac{N}{2}\mathrm{Tr}\left[(C - it(AB+BA))^2\right]\right\} dC \Bigg) d\mu_N(B) d\mu_N(A). \end{aligned}$$

Now, $\int \exp\{-\mathrm{Tr}\,(C - sM)^2\} dC = \int \exp\{-\mathrm{Tr}\, C^2\} dC$ for any Hermitian matrix M and real s by a simple substitution, and so

$$\int \exp\left\{-\frac{N}{2}\mathrm{Tr}\left[(C - it(AB+BA))^2\right]\right\} dC = \int \exp\left\{-\frac{N}{2}\mathrm{Tr} C^2\right\} dC,$$

because the left hand side is analytic as a function of $t \in \mathbb{C}$, and equal to the right hand side for $t \in i\mathbb{R}$, since $AB + BA$ is Hermitian. Given that their common value is Z_N^{GUE}, this gives

$$\widehat{Z}_N(t) = \iint \exp\left\{-Nt^2 \mathrm{Tr}\left[ABAB + A^2B^2\right]\right\} d\mu_N(B) d\mu_N(A),$$

which is a representation of $\widehat{Z}_N(t)$ that is interesting in its own right because it also has a combinatorial interpretation (4-valent stars with two colors $\mathcal{A}, \mathcal{B}$ of types $ABAB$ or A^2B^2).

Now, diagonalize the matrix $A = U_A \Lambda_A U_A^{-1}$ where Λ_A is the diagonal matrix $\Lambda_A := \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$, and U_A is unitary (see for example Chapter 5 in [7]). One can then move the conjugation of A by U_A to conjugation of B by U_A^{-1}, and use the invariance of $d\mu_N(B)$ under conjugation by unitary matrices to cancel the unitary part of the integral dU_A. This gives

$$\widehat{Z}_N(t) = \iint \exp\left\{-Nt^2 \operatorname{Tr}\left[\Lambda_A B \Lambda_A B + \Lambda_A^2 B^2\right]\right\} d\mu_N(B) d\mu_N^{\text{ev}}(\lambda),$$

where μ_N^{ev} is the induced probability measure on eigenvalues of A as defined in 1.1.5.

Writing $d\mu_N(B)$ explicitly gives

$$\widehat{Z}_N(t) = \frac{1}{Z_N^{GUE}} \iint \exp\left\{-N \operatorname{Tr}\left[\frac{1}{2}B^2 + t^2(\Lambda_A B \Lambda_A B + \Lambda_A^2 B^2)\right]\right\} dB d\mu_N^{\text{ev}}(\lambda).$$

The term in the exponential is a (diagonal) quadratic form in the real and imaginary entries of B, which one can further check is positive definite for all $\lambda \in \mathbb{R}^N$ only if $t \in \mathbb{R}$. We now perform the integral over the B matrix by using Gaussian integration as follows: We first rescale the matrix B by setting $\hat{B} = \sqrt{N} B$. This gives (see (1.1.1))

$$\widehat{Z}_N(t) = \frac{1}{\widetilde{Z}_N^{GUE}} \iint \exp\left\{-\operatorname{Tr}\left[\frac{1}{2}\hat{B}^2 + t^2(\Lambda_A \hat{B} \Lambda_A \hat{B} + \Lambda_A^2 \hat{B}^2)\right]\right\} d\hat{B} d\mu_N^{\text{ev}}(\lambda).$$

Now note that the term in the exponential is a quadratic form in the variables showing up in $d\hat{B}$, which we can write as

$$\begin{aligned} -\operatorname{Tr}\left[\frac{1}{2}\hat{B}^2 + t^2(\Lambda_A \hat{B} \Lambda_A \hat{B} + \Lambda_A^2 \hat{B}^2)\right] &= -\frac{1}{2}\operatorname{Tr}\left[\hat{B}^2 + 2t^2(\Lambda_A \hat{B} \Lambda_A \hat{B} + \Lambda_A^2 \hat{B}^2)\right] \\ &= -\frac{1}{2} b^T Q(t, \lambda) b \end{aligned}$$

where $\cdot^T$ stands for transpose and

$$b^T = (\hat{b}_{11}, \ldots, \hat{b}_{NN}, \operatorname{Re} \hat{b}_{12}, \ldots, \operatorname{Im} \hat{b}_{12})$$

are the N^2 variables showing up in $d\hat{B}$. Then, for values of t for which $Q(t,\lambda)$ is positive definite for all $\lambda \in \mathbb{R}^N$ (we will see below that this is the case if $t \in \mathbb{R}$), we will have have

$$\widehat{Z}_N(t) = \frac{1}{\widetilde{Z}_N^{GUE}} \int_{\mathbb{R}^N} \sqrt{\frac{(2\pi)^{N^2}}{\det Q(t,\lambda)}} d\mu_N^{\mathrm{ev}}(\lambda).$$

Using (1.1.2) we get

$$\widehat{Z}_N(t) = 2^{N(N-1)/2} \int_{\mathbb{R}^N} \frac{1}{\sqrt{\det Q(t,\lambda)}} d\mu_N^{\mathrm{ev}}(\lambda).$$

All that remains now is to find $\det Q(t,\lambda)$ and the t for which $Q(t,\lambda)$ is positive definite for all $\lambda \in \mathbb{R}^N$. By using $\Lambda_A = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$ and writing the traces $\mathrm{Tr}\hat{B}^2$, $\mathrm{Tr}[\Lambda_A \hat{B} \Lambda_A \hat{B}]$, and $\mathrm{Tr}[\Lambda_A^2 \hat{B}^2]$ in terms of the matrix entries separately, using the same indices on each sum, and the fact that $\hat{B}$ is hermitian so that $\hat{b}_{ij}\hat{b}_{ji} = |\hat{b}_{ij}|^2$, we obtain

$$\begin{aligned}
\mathrm{Tr}\left[\hat{B}^2 + 2t^2(\Lambda_A \hat{B} \Lambda_A \hat{B} + \Lambda_A^2 \hat{B}^2)\right] &= \sum_{i,j=1}^N |\hat{b}_{ij}|^2 (1 + 2t^2\lambda_i\lambda_j + 2t^2\lambda_i^2) \\
&= \sum_{i=1}^N \hat{b}_{ii}^2 (1 + 4t^2\lambda_i^2) + \sum_{i<j} 2|\hat{b}_{ij}|^2 (1 + 2t^2\lambda_i\lambda_j + t^2\lambda_i^2 + t^2\lambda_j^2) \\
&= \sum_{i=1}^N \hat{b}_{ii}^2 (1 + 4t^2\lambda_i^2) + \sum_{i<j} \left(\left(\mathrm{Re}\,\hat{b}_{ij}\right)^2 + \left(\mathrm{Im}\,\hat{b}_{ij}\right)^2 \right) 2\left(1 + t^2\left(\lambda_i + \lambda_j\right)^2\right)
\end{aligned}$$

so that $Q(t,\lambda)$ is diagonal, and

$$\det Q(t,\lambda) = 2^{N(N-1)} \prod_{i,j=1}^N \left(1 + t^2(\lambda_i + \lambda_j)^2\right).$$

Note that $Q(t,\lambda)$ is positive definite for all $\lambda \in \mathbb{R}^N$ only if $t \in \mathbb{R}$.

Going back to $\widehat{Z}_N(t)$, we see that the constant $2^{N(N-1)}$ will cancel, giving

$$\begin{aligned}
\widehat{Z}_N(t) &= \int_{\mathbb{R}^N} \prod_{i,j=1}^N \frac{1}{\sqrt{1 + t^2(\lambda_i + \lambda_j)^2}} d\mu_N^{\mathrm{ev}}(\lambda) \\
&= \frac{1}{Z_N^{\mathrm{evGUE}}} \int_{\mathbb{R}^N} e^{-\frac{N}{2}\sum \lambda_i^2} \prod_{i<j} (\lambda_i - \lambda_j)^2 \prod_{i,j=1}^N \frac{1}{\sqrt{1 + t^2(\lambda_i + \lambda_j)^2}} d^N\lambda \\
&= \frac{Y_N(t)}{Y_N(0)} = \widehat{Y}_N(t)
\end{aligned}$$

for $t \in \mathbb{R}$, by the definition of μ_N^{ev} (see 1.1.5), and the fact that $Z_N^{\mathrm{evGUE}} = Y_N(0)$. □

6.2 Overview of the strategy

We now recall the construction of a candidate density ρ_t for the minimizer of I_t in the physics literature that we discussed in section 3.2. We remark that the papers [18] and [20] contain a construction of a density ρ_t that satisfies the equality $Q_t(x) = \ell_t$ on the support of ρ_t, but do not verify the inequality off the support, and do not have the result from corollary 5.3.9 (which states that these two conditions are sufficient to know that ρ_t is, in fact, the minimizer of I_t). As discussed in 5.2.11, there may be densities that satisfy the equality on the support without being extremal.

6.2.1 Heuristics. If in the expression for $Y_N(t)$ one rescales the eigenvalues by t to move the t dependence to the potential $V(x) = x^2/2$, one then takes a t derivative, and then rescales the eigenvalues back, one obtains the identity

$$\frac{d}{dt}\left[\frac{1}{N^2}\log \widehat{Y}_N(t)\right] = -\frac{1}{t} + \frac{1}{t}\int \left(\frac{1}{N}\sum x_i^2\right)\rho_{N,t}(x_1,\dots,x_N)dx_1\dots dx_N, \tag{6.2.2}$$

where $\rho_{N,t}(x_1,\dots,x_N)$ is the probability density given by

$$\rho_{N,t}(x_1,\dots,x_N) := \frac{1}{Y_N(t)}\prod_{1\le i,j\le N}\frac{1}{\sqrt{1+t^2(x_i+x_j)^2}}\prod_{i<j}(x_i-x_j)^2 e^{-N\sum_{i=1}^N x_i^2/2}.$$

Letting $N\to\infty$ and using theorem 4.3.5 with the unbounded $\phi(x) = x^2$, one expects the integral in the right hand side of (6.2.2) to converge to $\int_{\mathbb{R}} x^2\rho_t(x)dx$, if ρ_t is, in fact, the unique minimizer of I_t. Assuming that the limit $N\to\infty$ and d/dt commute on the left side of (6.2.2), one expects that in the limit, identity (6.2.2) will take the from

$$\frac{d}{dt}\left[-I_t[\rho_t]\right] = -\frac{1}{t} + \frac{1}{t}\int_{\mathbb{R}} x^2\rho_t(x)dx. \tag{6.2.3}$$

This shows, heuristically, that to obtain the genus zero generating function e_0 one just needs to find a Taylor expansion for the second moment of the density ρ_t around $t = 0$. This second moment is what is found in [20] using a construction we review in the following section when carefully defining ρ_t.

6.2.4. We remark that the application of theorem 4.3.5 with the unbounded $\phi(x) = x^2$ can be justified by using the large deviation estimates, and using that in this situation $V(x)$ grows like a polynomial (as we discussed in 4.8.6). We also remark that the existence of $\log Z_N(t)$ for complex t close to $t = 0$ is a delicate matter that we will not analyze here.

6.2.5. We will give a direct proof for (6.2.3) in section 6.7, after we construct ρ_t and prove that it is analytic in t.

6.2.6 The saddle-point equation for ρ_t and definition of ζ_t. We recall that, under the assumption that the minimizer of I_t is absolutely continuous with continuous density ρ_t, the saddle-point equation for ρ_t as described in 5.2.11 takes the form

(6.2.7)
$$x = p.v. \int \frac{2\rho_t(y)}{x-y} dy - \int \left(\frac{t}{t(x+y)+i} + \frac{t}{t(x+y)-i} \right) \rho_t(y) dy, \quad x \in \operatorname{supp} \rho_t.$$

Assuming that ρ_t is even and supported on a single interval $[-\beta_t, \beta_t]$, one can write (6.2.7) in terms of the Cauchy transform

$$W_t(z) := \int_{\mathbb{R}} \frac{\rho_t(y)}{z-y} dy, \quad z \in \mathbb{C} \setminus [-\beta_t, \beta_t], \tag{6.2.8}$$

by using the Sokhotski–Plemelj formulas

$$W_t(x \pm i0) = p.v. \int_{\mathbb{R}} \frac{\rho_t(y)}{x-y} dy \mp i\pi\rho_t(x), \quad x \in [-\beta_t, \beta_t] \tag{6.2.9}$$

under the assumption that the boundary values of W_t are finite. Using these assumptions, and the fact that W_t is odd if ρ_t is even, equation (6.2.7) takes the form (assuming $t \neq 0$)

$$x = W_t(x+i0) + W_t(x-i0) - W_t\left(x - \frac{i}{t}\right) - W_t\left(x + \frac{i}{t}\right), \quad x \in [-\beta_t, \beta_t].$$

If one further defines

$$\zeta_t(z) := z^2 + \frac{2i}{t}\left(W_t\left(z + \frac{i}{2t}\right) - W_t\left(z - \frac{i}{2t}\right)\right), \tag{6.2.10}$$

then the saddle-point equation (6.2.7) translates into the following functional equation for ζ_t

$$\zeta_t\left(x+\frac{i}{2t}\pm i0\right)=\zeta_t\left(x-\frac{i}{2t}\mp i0\right),\quad x\in[-\beta_t,\beta_t]. \tag{6.2.11}$$

The fact that $W_t(z)$ is analytic in $\mathbb{C}\setminus[-\beta_t,\beta_t]$ with expansion at $z=\infty$ given by

$$W_t(z)=\frac{1}{z}+\frac{m_1(t)}{z^2}+\frac{m_2(t)}{z^3}+\ldots,$$

where

$$m_i(t):=\int x^i\rho_t(x)dx,$$

translates into the fact that ζ_t is analytic outside the two cuts $\pm i/2t+[-\beta_t,\beta_t]$, and has an expansion at infinity of the form

$$\begin{aligned}\zeta_t(z) \quad = \quad & z^2+\left(\frac{2}{t^2}\right)\frac{1}{z^2}+\left(\frac{12t^2m_2(t)-1}{2t^4}\right)\frac{1}{z^4}\\ & +\left(\frac{1-40t^2m_2(t)+80t^4m_4(t)}{8t^6}\right)\frac{1}{z^6}+\ldots.\end{aligned} \tag{6.2.12}$$

Finally, under the assumption that t is real, one can check that the function ζ_t must satisfy the symmetries

$$\zeta_t(-z)=\zeta_t(z)=\overline{\zeta_t(\overline{z})},\qquad t\in\mathbb{R}\setminus\{0\}.$$

One expects to be able to recover ρ_t from ζ_t by using the Plemelj formulas (6.2.9) once more, which give

$$\rho_t(x)=-\frac{t}{4\pi}\left[\zeta_t\left(x+\frac{i}{2t}+i0\right)-\zeta_t\left(x+\frac{i}{2t}-i0\right)\right].$$

In this way, the problem of finding ρ_t is recast into the problem of finding ζ_t.

6.3 The associated Riemann-Hilbert problem

The conditions satisfied by ζ_t that have no explicit mention of ρ_t make up the following problem that ζ_t solves. We call the function f_t in the problem differently in order to clearly differentiate between the hypothetical ζ_t related to the Cauchy transform of the unknown density ρ_t, and the function f_t which will be constructed.

6.3.1 Required properties for f_t. Assume that $t > 0$. The function f_t:

1. Is analytic on the complement of the two cuts $\pm i/2t + [-\beta, \beta]$, where β is an unknown positive real number, that possibly depends on t.

2. Satisfies the symmetries

$$f_t(-z) = f_t(z) = \overline{f_t(\overline{z})}.$$

3. Has finite boundary values when approaching the cuts from above and from below, which are related by

$$f_t\left(x + \frac{i}{2t} \pm i0\right) = f_t\left(x - \frac{i}{2t} \mp i0\right), \qquad x \in [\beta, \beta].$$

4. Has a pole of order two at infinity, and an expansion at infinity given by

$$f_t(z) = z^2 + \left(\frac{2}{t^2}\right)\frac{1}{z^2} + O\left(\frac{1}{z^4}\right).$$

We remark that conditions 2 and 3 together imply that, if f_t exists, then it is real valued on the real and imaginary axes, and its boundary values on the cuts from above and below are also real.

6.3.2. The clever construction introduced in [18] shows that it is not difficult to find an explicit function that satisfies $1 - 3$ and the weaker condition 4' given by

4'. f_t has a pole of order two at infinity, and an expansion at infinity given by

$$f_t(z) = z^2 + O\left(\frac{1}{z^2}\right).$$

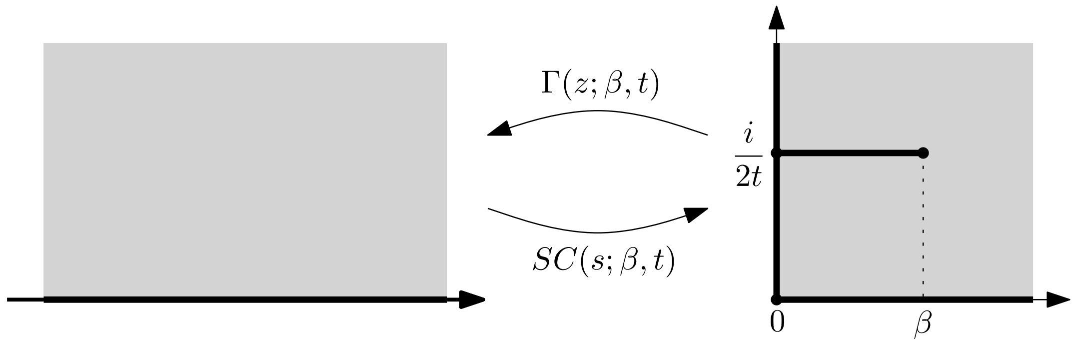

FIGURE 6.1. Definition of SC and Γ

The basic idea, as we discussed briefly in section 3.2, is the following: Let $\beta > 0$ be arbitrary, and let $\Gamma(z;\beta,t)$ be the inverse of the Schwarz–Christoffel map $SC(s;\beta,t)$ that maps the upper half plane to the complement of the segment $i/2t + [0,\beta]$ in the first quadrant $\operatorname{Re} z, \operatorname{Im} z > 0$, and that sends infinity to infinity (see figure 6.1).

Figure 6.2 shows the image under SC of an equally spaced rectangular grid on the s-plane created with the Schwarz-Christoffel toolbox for Matlab, by Tobin A. Driscoll.

Since $\Gamma(z;\beta,t)$ is real valued on the positive real and imaginary axes, one can analytically extend $\Gamma(z;\beta,t)$ to $\mathbb{C}\setminus\{\pm i/2t+[-\beta,\beta]\}$ by using the symmetry principle and defining $\Gamma(-z;\beta,t) = \Gamma(z;\beta,t)$ and $\Gamma(\overline{z};\beta,t) = \overline{\Gamma(z;\beta,t)}$. This extended map $\Gamma(z;\beta,t)$ satisfies all of the conditions required for f_t except for possibly the expansion at infinity. One can further check that $\Gamma(z;\beta,t)$ has a pole of order two at infinity, and so there is a unique linear combination $a_1\Gamma + a_2$, that has an expansion at infinity of the form $z^2 + O(1/z^2)$. Moreover, we will see that a_1 and a_2 are real, so that $a_1\Gamma + a_2$ will satisfy conditions $1-3, 4$', where 2 continues to hold since a_1 and a_2 are real. We will analyze this construction in much more detail in section 6.5.

6.3.3. Using general arguments from the theory of Riemann surfaces one can prove that this solution to $1-3, 4$' is unique among the functions with continuous extensions to the cuts. In fact, condition 2 is not required to prove uniqueness, and conditions

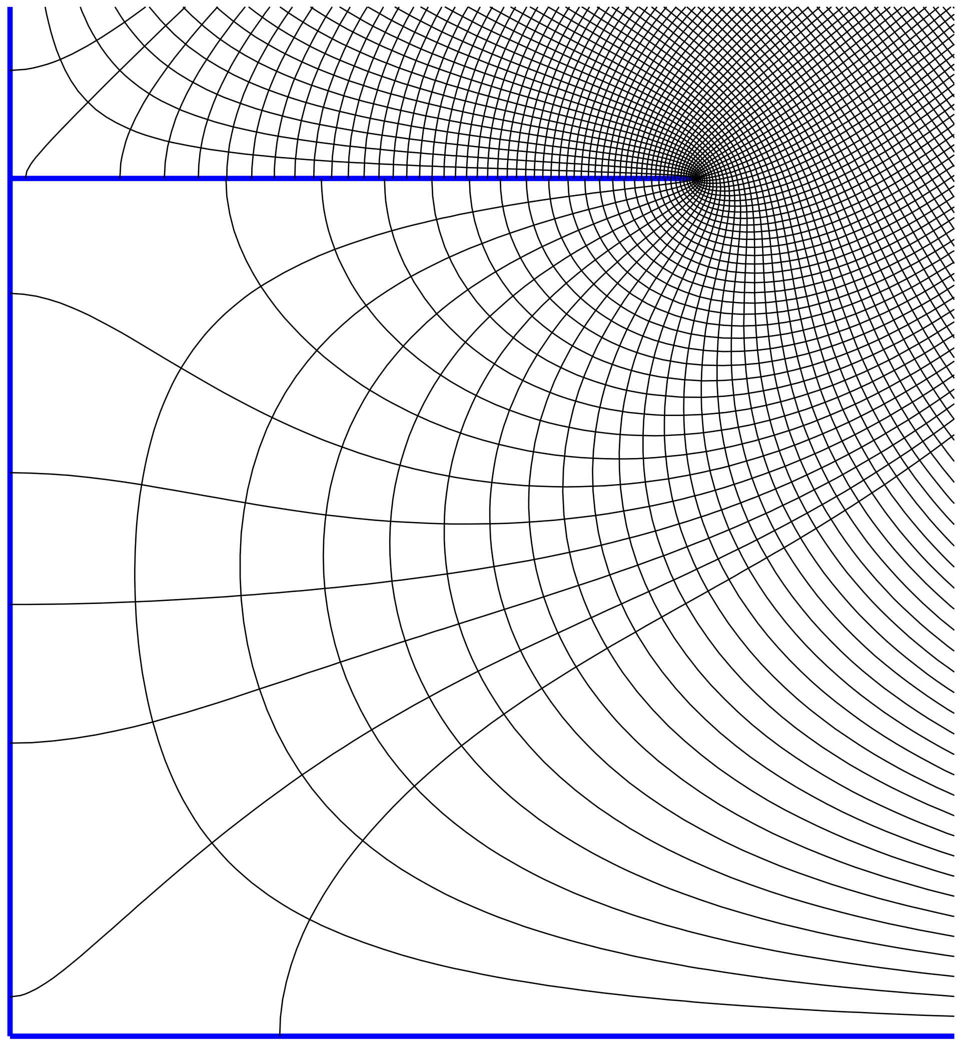

FIGURE 6.2. Image under SC of a rectangular grid in the upper half plane.

$1, 3$, and 4' can be cast in the form of a non-local Riemann-Hilbert problem with a unique solution.

6.3.4 Lemma. *For any $t, \beta > 0$, the function $a_1\Gamma(z; \beta, t) + a_2$ described above is the unique function with continuous extensions to the cuts that satisfies conditions $1, 3$ and 4'.*

Stated in different terms, for any $t, \beta > 0$, the function $a_1\Gamma(z; \beta, t) + a_2$ is the unique solution to the following non-local Riemann-Hilbert problem: Find a function

$f(z)$ with the following properties:

(RH1) f is analytic on the complement of the two cuts $\pm i/2t + [-\beta, \beta]$.

(RH2) f has continuous extensions to the cuts (including the endpoints) related by

$$f\left(x + \frac{i}{2t} \pm i0\right) = f\left(x - \frac{i}{2t} \mp i0\right), \qquad x \in [\beta, \beta].$$

(RH3) The expansion of f at $z = \infty$ given by

$$f(z) = z^2 + O\left(\frac{1}{z^2}\right).$$

Proof. By condition $(RH2)$ the function $f(z)$ defines a function on the elliptic curve constructed by opening up the two cuts $\pm i/2t + [-\beta, \beta]$ in the complex plane, and glueing them together forming a handle (note that one is using only one copy of the complex plane). Analyticity of this function (which by abuse of notation we will continue to call f) outside the point corresponding to $z = \infty$ shows that f must be an elliptic function of order two, with a single pole by $(RH3)$. By Riemann Roch, f must correspond to a (complex) linear combination of the Wierstrass $\wp$ function in the uniformized elliptic curve. Since the coefficients in the expansion at infinity give two conditions (because of the vanishing constant term), we conclude that f is unique. □

6.3.5. We will see that the the extra equation coming from the coefficient of z^{-2} in condition 4 will fix the value of β as an analytic function of t for $t > 0$ which admits an analytic extension to $t = 0$.

6.3.6 Proposition. *For each $t > 0$ there is a unique value $\beta = \beta(t) > 0$ for which there is a function satisfying conditions $1 - 4$ with continuous extensions to the cuts (by the lemma this function is unique). Moreover, $\beta(t)$ is analytic in t for $t > 0$, and admits an analytic extension to $t = 0$, with limiting value $\beta(0) = 2$.*

Furthermore, the solution f_t maps the complement of the segment $i/2t+[0,\beta(t)]$ in the first quadrant $\mathrm{Re}\, z, \mathrm{Im}\, z > 0$ conformally to the upper half plane, and the boundary values satisfy the inequality

$$f_t\left(x+\frac{i}{2t}+i0\right)-f_t\left(x+\frac{i}{2t}-i0\right)\le 0,$$

for all $x \in [0,\beta(t)]$.

The proof of the proposition is considerably involved, since f_t is constructed explicitly. This will be done in section 6.5. The definition of f_t is given in formula (6.5.21). The construction comes from [18], and was further analyzed in [20]. Our proposition adds a careful analysis of the analyticity in t of the quantities involved in the construction and of the endpoint of the support, which are essential to prove that the density ρ_t is analytic in t. .

Before going into the details of the construction of f_t we show how one can essentially backtrack the heuristic arguments described in section 6.2.6 to obtain the (unique) extremal measure of I_t from f_t.

6.4 Definition of ρ_t

6.4.1 Proposition. *If f_t is the solution to problem 6.3.1 given by proposition 6.3.6 (explicitly defined in (6.5.21) below), and one defines for $x \in \mathbb{R}$*

$$\rho_t(x) := -\frac{t}{4\pi}\left[f_t\left(x+\frac{i}{2t}+i0\right)-f_t\left(x+\frac{i}{2t}-i0\right)\right],$$

then ρ_t is non-negative, even, continuous, and locally the restriction of an analytic function in the interior of its support $[-\beta(t),\beta(t)]$.

Moreover, ρ_t is the unique extremal measure for I_t, and if one defines W_t and ζ_t as in equations (6.2.8) and (6.2.10) using this ρ_t, then $f_t = \zeta_t$ and all the formulas in 6.2.6 are valid.

Proof. Using the definition of ρ_t in the statement, define define W_t and ζ_t by formulas (6.2.8) and (6.2.10). We start by proving that $f_t = \zeta_t$. Using the definition of ρ_t, it is easy to see that for z outside the cuts one has

$$-\frac{2i}{t}W_t\left(z-\frac{i}{2t}\right)=\frac{1}{2\pi i}\int_{\gamma_1}\frac{f_t(s)}{s-z}ds,$$

where γ_1 is a loop around the top cut $i/2t + [-\beta(t), \beta(t)]$ in the clockwise direction that does not enclose z. Similarly, by using the fact that

$$\rho_t(x) = -\frac{t}{4\pi}\left[f_t\left(x-\frac{i}{2t}-i0\right)-f_t\left(x-\frac{i}{2t}+i0\right)\right],$$

which holds since f_t commutes with conjugation and its boundary values on the cuts are real, we see that

$$\frac{2i}{t}W_t\left(z+\frac{i}{2t}\right)=\frac{1}{2\pi i}\int_{\gamma_2}\frac{f_t(s)}{s-z}ds,$$

where γ_2 is a loop around the bottom cut $-i/2t + [-\beta(t), \beta(t)]$ in the clockwise direction that does not enclose z. By the definition of ζ_t, it follows that

$$\zeta_t(z) = z^2 + \frac{1}{2\pi i}\int_{\gamma_1+\gamma_2}\frac{f_t(s)}{s-z}ds.$$

This implies that $\zeta_t = f_t$ since the right hand side can be seen to be equal to f_t by Cauchy's theorem with f_t at z, and relating this to the residue of the integrand at infinity. In particular, this proves that ζ_t has finite boundary values on the cuts that satisfy (6.2.11).

The fact that ρ_t is non-negative follows from proposition 6.3.6. To prove that ρ_t is even recall that the boundary values of f_t along the cuts are real. Using the two symmetries $f_t(-z) = f_t(z) = \overline{f_t(\overline{z})}$ of f_t and the fact that $t \in \mathbb{R}$, it is easy to see that for $x \in [-\beta(t), \beta(t)]$ we have $\rho_t(-x) = \rho_t(x)$. To check that ρ_t is the restriction of an analytic function to the interval $(-\beta(t), \beta(t))$, note that for $x \in (-\beta(t), \beta(t))$ we can write

$$\rho_t(x) = -\frac{t}{4\pi}\left(f_t\left[x+\frac{i}{2t}+i0\right]-f_t\left[x-\frac{i}{2t}+i0\right]\right)$$

again by using the symmetries of f_t and the fact that the boundary values of f_t are real. This implies that $\rho_t(x) = h_t(x + i0)$ where

$$h_t(z) := -\frac{t}{4\pi}\left(f_t\left[z + \frac{i}{2t}\right] - f_t\left[z - \frac{i}{2t}\right]\right),$$

which is analytic outside the cuts $\pm i/t + [-\beta(t), \beta(t)]$ and the real interval $[\beta(t), \beta(t)]$ where it has a sign jump. Thus, ρ_t is the restriction to the real axis of the analytic function $\widetilde{h}_t(z)$ defined to be $h_t(z)$ above the real axis, and $-h_t(z)$ below the axis.

This concludes the proof that ρ_t is a continuous probability density. We now show that it satisfies the variational inequalities for I_t.

Since ρ_t is the restriction of an analytic function in the interior of its support, the real valued function $Q_t(x)$ from corollary 5.3.9 can be shown to be differentiable. Indeed, one only needs to show that

$$G_t(x) := \int_{-\beta(t)}^{\beta(t)} \log|x - y|\rho_t(y)dy$$

is differentiable. This is proved by defining

$$F_t(z) := \int_{-\beta(t)}^{\beta(t)} \log(z - y)\rho_t(y)dy, \quad z \in \mathbb{C} \setminus (-\infty, \beta(t)]$$

where $\log(z - y)$ is taken to be the principal branch of the complex logarithm in the z-plane with cut along $(-\infty, y]$, and using the fact that for $x \in (-\beta(t), \beta(t))$ one has $F_t(x + i0) = G_t(x) + i\pi \int_x^{\beta(t)} \rho_t(y)dy$. The claim that $G_t(x)$ is differentiable then follows from the fact that $F_t(x+i0)$ is differentiable because one can move the contour of integration below the real axis at x.

Thus, to prove that ρ_t satisfies the variational inequalities, it is sufficient to verify the following conditions

$$\begin{aligned} Q_t'(x) &= 0 \quad \text{for } x \in (-\beta(t), \beta(t)) \\ \int_{\beta(t)}^{x} Q_t'(x)dx &\geq 0 \quad \text{for } x \geq \beta(t) \\ \int_{-\beta(t)}^{x} Q_t'(x)dx &\geq 0 \quad \text{for } x \leq -\beta(t). \end{aligned} \tag{6.4.2}$$

Now, by the Plemelj formulas (6.2.9) which hold by what we showed above about ρ_t being the restriction of an analytic function, and the definition of ζ_t, we have

$$\frac{2i}{t}Q_t'(x) = \zeta_t\left(x + \frac{i}{2t} + i0\right) - \zeta_t\left(x - \frac{i}{2t} - i0\right),$$

and the equality in (6.4.2) follows from the fact that $\zeta_t = f_t$ and the properties of f_t. Regarding the inequalities, we have for $x \notin [-\beta(t), \beta(t)]$

$$f_t\left(x + \frac{i}{2t} + i0\right) - f_t\left(x - \frac{i}{2t} - i0\right) = 2i \operatorname{Im} f_t\left(x + \frac{i}{2t}\right)$$

since f_t commutes with conjugation. Then, since f_t maps the complement of the segment $i/2t + [0, \beta(t)]$ in the first quadrant $\operatorname{Re} z, \operatorname{Im} z > 0$ conformally to the upper half plane (see proposition 6.3.6), we conclude (again since $\zeta_t = f_t$) that

$$\begin{aligned} Q_t'(x) &\geq 0 \quad \text{if } x \geq \beta(t) \\ &\leq 0 \quad \text{if } x \leq -\beta(t) \end{aligned}$$

which gives the inequalities in (6.4.2) off the support. This completes the verification that ρ_t is indeed the (unique) extremal measure for I_t. □

6.4.3 The semicircle density. We remark that ρ_0 is the semicircle density $(2\pi)^{-1}\sqrt{4 - x^2}$ since $Y_N(0) = Z_N^{\text{GUE}}$, so that ρ_t is expected to converge to ρ_0, but the definition of ρ_t in terms of f_t given by proposition 6.4.1 does not allow one to see this directly. In fact, we will see that the function f_t itself blows-up as $t \searrow 0$, so the definition of ρ_t in proposition 6.4.1 is singular on more ways than it is apparent by just noticing the $i/2t$ terms. Nonetheless, in section 6.6 we prove that $\rho_t(x)$ depends analytically on t for $t > 0$ and x in the interior of its support.

The graphs in figure 6.3, created by numerically solving for $\beta(t)$ using equations we describe below, and using the Schwarz-Christoffel toolbox for Matlab created by Tobin A. Driscoll, illustrate the convergence of ρ_t to the semicircle density supported on $[-2, 2]$.

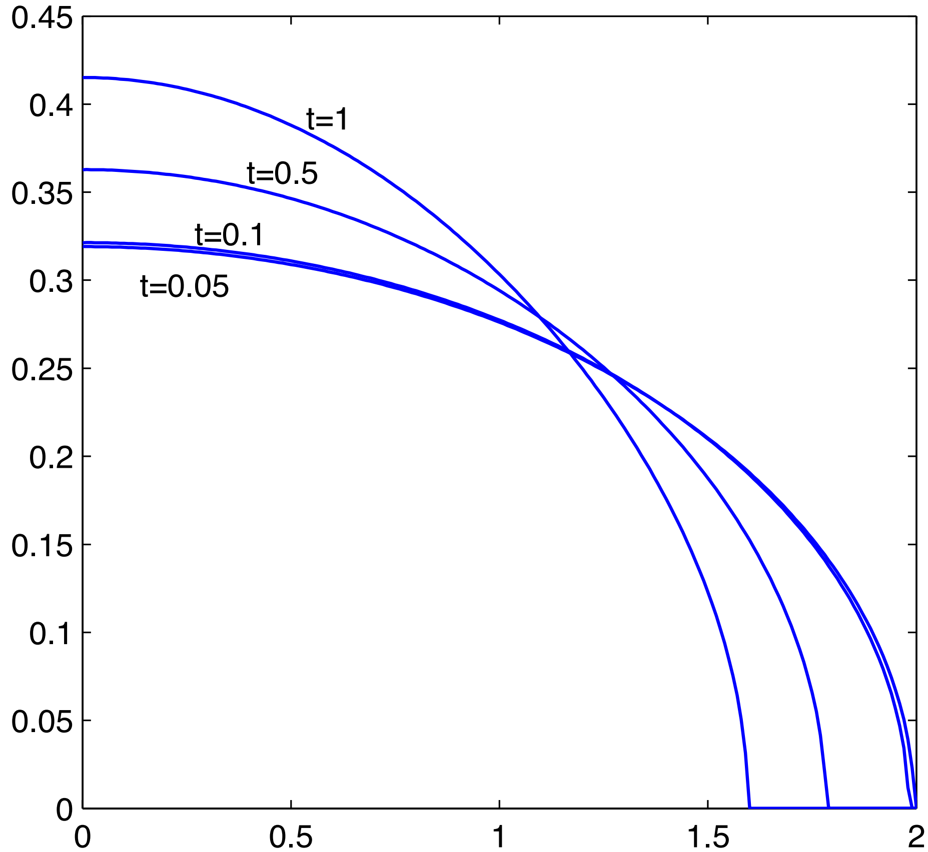

FIGURE 6.3. Graphs of $\rho_t(x)$ for $x \geq 0$ and $t = 1$, $t = 0.5$, $t = 0.1$, $t = 0.05$.

6.5 Proof of proposition 6.3.6

Let $\beta > 0$, and let $\Gamma(z; \beta, t)$ be the inverse of the Schwarz–Christoffel map $SC(s; \beta, t)$ that maps the upper half plane to the complement of the segment $i/2t + [0, \beta]$ in first quadrant $\text{Re}\, z, \text{Im}\, z > 0$ with the following pre-images for the vertices

$$\begin{array}{ccc} s & \leftrightarrow & z \\ \infty & & \infty \\ 0 & & 0 \\ -1 & & i/2t. \end{array}$$

The choice of pre-vertices fixes $SC(s;\beta,t)$ completely. It is given by

$$z = SC(s;\beta,t) = A\int_0^s \frac{s+c}{\sqrt{s(s+1)(s+b)}}ds$$

where the path of integration is taken inside the upper half s-plane $\text{Im}\, s > 0$, and for the square root we take a branch that is analytic in the upper half plane that gives the positive square root on the positive real axis as one approaches it from above. In most of what follows the choice of branch cuts will be irrelevant as long as they do not intersect the upper half plane. Here $A = A(\beta,t)$, $b = b(\beta,t)$ and $c = c(\beta,t)$ are uniquely determined by β and t by the theory of Schwarz–Christoffel maps (see [23, III, p. 323]), and the fact that

$$\begin{array}{ccc} s & \leftrightarrow & z \\ -c & & i/2t+\beta \\ -b & & i/2t \end{array}$$

so that in particular $-b < -c < -1$ (see figure 6.4).

More explicitly, A, b, and c are uniquely determined by β and t through the equations

$$\begin{aligned} \frac{i}{2t} &= A\int_0^{-1} \frac{s+c}{\sqrt{s(s+1)(s+b)}}ds, \\ \beta &= A\int_{-1}^{-c} \frac{s+c}{\sqrt{s(s+1)(s+b)}}ds, \\ -\beta &= A\int_{-c}^{-b} \frac{s+c}{\sqrt{s(s+1)(s+b)}}ds. \end{aligned}$$

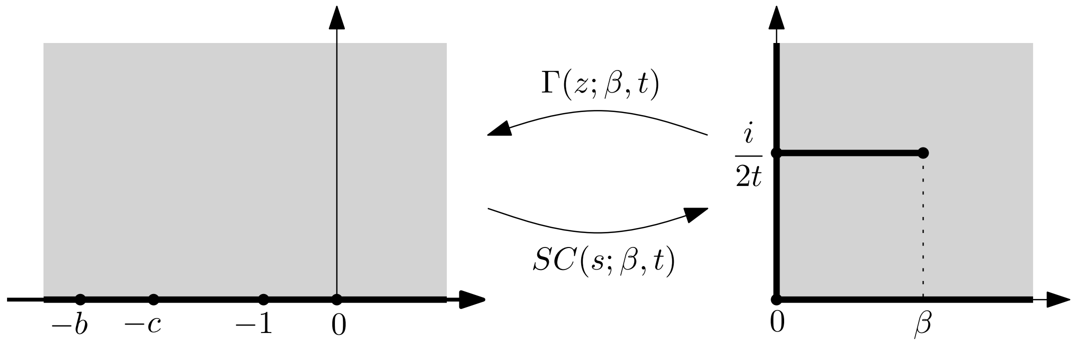

FIGURE 6.4. Pre-images of the vertices of the polygon.

Note that the last two equations imply that c is the value that makes the loop integral that goes around the interval $[-b,-1]$ vanish (here we take $[-b,-1]\cup[0,\infty)$ to be the branch cuts for the square root). I.e.,

$$0=\int_{-1}^{-b}\frac{z+c}{\sqrt{z(z+1)(z+b)}}dz=\frac{1}{2}\oint\frac{z+c}{\sqrt{z(z+1)(z+b)}}dz. \tag{6.5.1}$$

6.5.2. Now, the expansion of $\Gamma(z;\beta,t)$ at infinity is given by

$$\Gamma(z;\beta,t)=\frac{1}{4A^2}z^2+(2c-(1+b))+\frac{4A^2(2c-b+2bc-3c^2)}{3}\frac{1}{z^2}+O\left(\frac{1}{z^4}\right), \tag{6.5.3}$$

which can be seen as follows: For all z in the first quadrant outside the cut we have

$$z=A\int_0^{\Gamma(z;\beta,t)}\frac{s+c}{\sqrt{s(s+1)(s+b)}}ds,$$

so that for z close to infinity (which we know corresponds also to Γ close to infinity) we have

$$\begin{aligned}\frac{z}{A} &= \int_0^{\Gamma}\frac{s+c}{\sqrt{s(s+1)(s+b)}}ds\\ &= 2\sqrt{\Gamma}+\int_0^{\Gamma}\left[\frac{s+c}{\sqrt{s(s+1)(s+b)}}-\frac{1}{\sqrt{s}}\right]ds\\ &= 2\sqrt{\Gamma}+\int_0^{\infty}\left[\frac{s+c}{\sqrt{s(s+1)(s+b)}}-\frac{1}{\sqrt{s}}\right]ds\\ &\quad+\int_{\infty}^{\Gamma}\left[\frac{s+c}{\sqrt{s(s+1)(s+b)}}-\frac{1}{\sqrt{s}}\right]ds.\end{aligned}$$

Now, the first integral in the last expression is zero because of (6.5.1) by a contour deformation argument (taking $[0,\infty]$ as the branch of $\sqrt{z}$ as for $\sqrt{z(z+1)(z+b)}$). For the second integral we can taylor expand around $s=\infty$, and integrate term by term to obtain

$$\int_\infty^\Gamma \left[\frac{s+c}{\sqrt{s(s+1)(s+b)}} - \frac{1}{\sqrt{s}}\right] ds = \frac{1+b-2c}{\sqrt{\Gamma}} + \frac{4c+4bc-2b-3b^2-3}{12\Gamma^{3/2}} + \dots$$

This gives

$$z = 2A\sqrt{\Gamma} + \frac{A(1+b-2c)}{\sqrt{\Gamma}} + \frac{A\left(-3-3b^2-2b+4bc+4c\right)}{12\Gamma^{3/2}} + \dots$$

which one can invert by squaring

$$\left(\frac{z}{2A}\right)^2 = \Gamma + (1+b-2c) + \frac{b-2bc-2c+3c^2}{3\Gamma} + \dots$$

and using

$$\left(\frac{2A}{z}\right)^{2n} = \frac{1}{\Gamma^n} + O\left(\frac{1}{\Gamma^{n+1}}\right),$$

which gives (6.5.3).

6.5.4. Equation (6.5.3) shows that the linear combination $a_1\Gamma + a_2$ of Γ that solves the problem $1-3,4$' described in 6.3.2 is given by

$$4A^2\Gamma(z;\beta,t) + 4A^2(1+b-2c), \tag{6.5.5}$$

and it will satisfy condition 4 (see 6.3.1) only if the parameters satisfy the extra condition

$$8A^4t^2\left(2c-b+2bc-3c^2\right) = 3.$$

We show below that enforcing this last equality defines A, b, c and β as analytic functions of $t>0$, and that moreover, these functions admit an analytic or meromorphic extension to a complex neighborhood of $t=0$.

6.5.6 Lemma. *The system of equations*

$$\text{(I)} \qquad \frac{i}{2t} = A\int_0^{-1} \frac{z+c}{\sqrt{z(z+1)(z+b)}}dz$$

$$\text{(II)} \qquad \beta = A\int_{-1}^{-c} \frac{z+c}{\sqrt{z(z+1)(z+b)}}dz$$

$$\text{(III)} \qquad -\beta = A\int_{-c}^{-b} \frac{z+c}{\sqrt{z(z+1)(z+b)}}dz.$$

$$\text{(IV)} \qquad 0 = 3 - 8A^4t^2(2c - b + 2bc - 3c^2)$$

where

$$t, A, b, c, \beta > 0$$

$$b > c > 1$$

defines $(At), b, c, \beta$ *as (real) analytic functions of* t *for* $t > 0$*, where the function* $b(t)$ *is monotonically increasing in* t *for* $t > 0$*.*

Moreover, all of these these functions admit an analytic extension to complex t *in a neighborhood of* $t = 0$*. The case* $t = 0$ *is singular, since* A *grows without bound as* $t \searrow 0$*, and corresponds to* $b = c = 1$*,* $\beta = 2$*, and* $At \to 1/4$*. The Laurent expansions of these functions around* $t = 0$ *are given by*

$$\text{(6.5.7)} \qquad \begin{aligned} b(t) &= 1 + 16t + 128t^2 + 656t^3 + 2304t^4 + \dots \\ A(t) &= 1/4t - 1 + t + 3t^2 - 5t^3 - (55/2)t^4 + \dots \\ c(t) &= 1 + 8t + 48t^2 + 200t^3 + 576t^4 + 1116t^5 + \dots \\ \beta(t) &= 2 - 2t^2 + 15t^4 - 165t^6 + (8555/4)t^8 - (121599/4)t^{10} \\ &\quad +(3669003/8)t^{12} - (57721293/8)t^{14} + \dots \end{aligned}$$

6.5.8 Remark. Note that the coefficients of neither of these Taylor expansions seem to be converging to zero. This seems to indicate that all of these functions have a singularity in the complex disk $|t| \le 1$.

6.5.9 Remark. We remark that we have found no closed expression for these functions. As will be seen in the proof, one needs to locally invert an expression for t in terms of b.

6.5.10 Proof of lemma 6.5.6. For the proof it will be convenient to rescale A by t to remove its singular behavior. It will also be convenient to scale β, so we define the two new quantities

$$\begin{aligned} a &:= At \\ d &:= \beta t. \end{aligned}$$

As was mentioned before, adding (II) and (III) we obtain equation (6.5.1), and this shows that c is an analytic function of b. In the loop integral in (6.5.1) we are taking the cuts along $[-b,-1]\cup[0,\infty]$. One can also express c in terms b (which will be of use below) by using the complete elliptic integrals $E(k)$, $K(k)$ as

$$c(b) = b\frac{E(k)}{K(k)}, \tag{6.5.11}$$

where

$$E(k) = \int_0^{\pi/2} \sqrt{1-k^2\sin^2\alpha}\; d\alpha \tag{6.5.12}$$

$$K(k) = \int_0^{\pi/2} \frac{d\alpha}{\sqrt{1-k^2\sin^2\alpha}} \tag{6.5.13}$$

and

$$k^2 := 1-\frac{1}{b}. \tag{6.5.14}$$

Similarly, the integral in (I) is a loop integral which one may also write in terms of complete elliptic integrals giving an equation for a as a function of b

$$a(b) = \frac{K(k)}{2\pi\sqrt{b}}. \tag{6.5.15}$$

The reader verifying this might may find Legendre's relations [1, Ch 17] of use, which state that $E(k)K(k') + E(k')K(k) - K(k)K(k') = \pi/2$ where $k' = \sqrt{1-k^2}$. One can also rewrite (III) in terms of incomplete elliptic integrals, and get

$$d(b) = 2a\sqrt{b}\left(E(\phi, k) - \frac{c}{b}F(\phi, k)\right) \tag{6.5.16}$$

where

$$\phi(b) := \frac{1}{2}\arcsin\left(\frac{b+1-2c}{b-1}\right) + \frac{\pi}{4}. \tag{6.5.17}$$

These equations explicitly show that a, c and d are analytic functions of b for $b > 0$ (note that the expression inside the arcsine for ϕ is analytic at $b = 1$ because the numerator is analytic in b and vanishes at $b = 1$).

Writing (IV) in terms of a, b, c gives

$$t^2 = \frac{8}{3}a^4(2c - b + 2bc - 3c^2),$$

where the expression on the right is a function of b, which we denote by $T(b)$

$$T(b) := \frac{8}{3}a(b)^4(2c(b) - b + 2bc(b) - 3c(b)^2),$$

so that $t^2 = T(b)$.

The derivative of $T(b)$ can be expressed in the surprisingly compact form

$$T'(b) = \frac{-1}{2\pi^4 b(b-1)}E(k)K(k)\left(E(k) - K(k)\right)\left(bE(k) - K(k)\right),$$

which shows, in particular, that $T'(b) > 0$ for $b > 1$ (as suggested by the graph in figure 6.5). Together with the fact that $T(1) = 0$, this shows that for each $b > 1$ there are only two t's that make (IV) hold, given by $t = \pm\sqrt{T(b)}$ (real square root), and only one t that is positive. This shows that

$$t(b) := \sqrt{T(b)}, \qquad b > 1$$

is an analytic function of b for $b > 1$, with an analytic inverse. One can check that $t(b) \to \infty$ as $b \to \infty$.

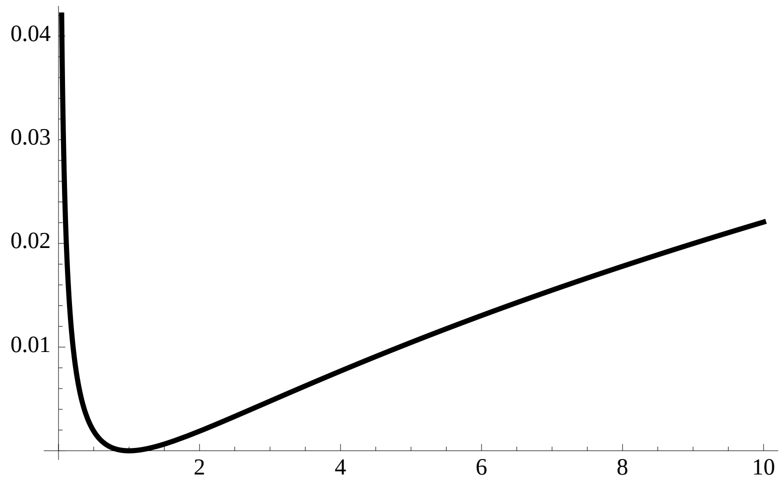

FIGURE 6.5. Graph of $T(b)$ for $b > 0$.

Regarding the singular case $t = 0$ corresponding to $b = 1$, we know that $t^2 = T(b)$ is analytic around $b = 1$, and so we have

$$t^2 = (b-1)^2 R(b)$$

for an analytic $R(b)$, where $R(1) \neq 0$ since $T'(1) = 0$ and $T''(1) \neq 0$ (see figure 6.5). Then, since $R(1) \neq 0$, we can define a branch of the square root of R close to $b = 1$ so that $(b-1)\sqrt{R(b)}$ agrees with $\sqrt{T(b)}$ (the positive real square root) for $b > 1$. We can then define

$$t(b) := (b-1)\sqrt{R(b)} \tag{6.5.18}$$

in a neighborhood of $b = 1$. This gives t as an analytic function of b close to $b = 1$ and agrees with $\sqrt{T(b)}$ for $b > 1$. For real b this amounts to defining

$$t(b) = \begin{cases} \sqrt{T(b)} & \text{for } b \geq 1 \\ -\sqrt{T(b)} & \text{for } 0 < b < 1. \end{cases}$$

From the fact that $t'(1) \neq 0$, we see that $t(b)$ is locally invertible around $b = 1$ with analytic inverse. Thus, we can define

$$b(t) := \text{inverse of } t(b)$$

which is an analytic function of t for $t \geq 0$ as claimed. This shows that a, b, c, d are all analytic functions of t for $t \geq 0$.

From the above, we see that the limiting value for b as $t \searrow 0$ is $b = 1$, which gives the values of the parameters in this singular case:

$$t = 0 \longleftrightarrow \begin{array}{rcl} a & = & 1/4 \\ b & = & 1 \\ c & = & 1 \\ d & = & 0 \end{array}$$

(in particular this shows that $A \to \infty$ as $t \to 0$).

Thus, the solutions of (I)-(V) are analytic functions of $t \in [0, \infty)$ (except for A which has a simple pole at $t = 0$), and given by

$$\begin{aligned} b(t) &= \text{inverse of } t(b) \\ A(t) &= \frac{K(k(t))}{2\pi t\sqrt{b(t)}} \\ c(t) &= b(t)\frac{E(k(t))}{K(k(t))} \\ \beta(t) &= 2t^2 A(t)\sqrt{b(t)}\left(E(\phi(b(t)), k(t)) - \frac{c(t)}{b(t)}F(\phi(b(t)), k(t))\right) \end{aligned}$$

where

$$\begin{aligned} k(t)^2 &= 1 - 1/b(t) \\ \phi(t) &= \phi(b(t)) := \frac{1}{2}\arcsin\left(\frac{b(t) + 1 - 2c(t)}{b(t) - 1}\right) + \frac{\pi}{4} \end{aligned}$$

The Taylor expansions in (6.5.7) were found using Mathematica. This concludes the proof of the lemma. □

6.5.19 Remark. Even though an explicit expression for the endpoint $\beta(t)$ of the density as a function of t is unlikely to be found with this approach since it requires an explicit expression for the inverse of $\sqrt{T(b)}$, we note that the above allows one to write down an explicit expression for β as a function of b. Concretely, for $b > 1$ we have

$$\beta = \frac{d(b)}{\sqrt{T(b)}}.$$

Calculating the behavior of β as $t \searrow 0$ is a tedious but straightforward task using equations (6.5.11)-(6.5.18), from which one can see that $\beta \nearrow 2$ as $t \searrow 0$. By numerically solving for b given a value of $t > 0$, one can find the approximate values of the endpoint β shown in table 6.1, which were used to create the graphs of ρ_t in page 113.

t	$\beta(t)$
1	1.59637
0.5	1.78324
0.1	1.98135
0.05	1.99509
0.01	1.9998

TABLE 6.1. Values of $\beta(t)$

6.5.20 Definition of f_t. The proposition shows that the problem 6.3.1 for f_t does indeed have a solution, given by (see (6.5.5))

$$f_t(z) := 4A(t)^2\Gamma(z;\beta(t),t) + 4A(t)^2(1 + b(t) - 2c(t)), \tag{6.5.21}$$

where A, b, c, β are as given by the proof of lemma 6.5.6. This in particular shows the inequality proposition 6.3.6 holds for $t > 0$ since $A(t), b(t), c(t)$ are all real valued.

For future use, we nota that in terms of A, b and c, the expansion of f_t at $z = \infty$ is given by (see (6.5.3))

$$\begin{aligned} f_t(z) &= 4A^2\Gamma_t(z) + 4A^2(1 + b - 2c) \\ &= z^2 + \frac{(2A)^4}{3}\left(2c - b + 2bc - 3c^2\right)\frac{1}{z^2} \\ &\quad + \frac{(2A)^6}{5}\left(-b - b^2 + 2c + 8bc + 2b^2c - 10c^2 - 10bc^2 + 10c^3\right)\frac{1}{z^4} \\ &\quad + \frac{(2A)^8}{63}\left(-9b - 20b^2 - 9b^3 + 18c + 134bc + 134b^2c + \dots\right)\frac{1}{z^6} + \dots \end{aligned} \tag{6.5.22}$$

This concludes the proof of proposition 6.3.6, and with it completes the construction of the density ρ_t.

6.6 Analyticity of ρ_t in t.

With the explicit construction of f_t and the analyticity of the coefficients $A(t), b(t), c(t)$ showing up un the definition of of f_t and $\Gamma(z; \beta(t), t)$, we are now ready to prove the following proposition.

6.6.1 Proposition. *The density $\rho_t(x)$ is analytic in t at $t = t_0 > 0$ for any $x_0 \in (-\beta(t_0), \beta(t_0))$.*

Proof. By the definition of ρ_t from proposition 6.4.1, it is sufficient to prove that $\Gamma\left(x_0 + i/2t \pm i0; \beta(t), t\right)$ are analytic in t at $t = t_0 > 0$ for any $x_0 \in (0, \beta(t_0))$. Fix $t_0 > 0$ and $x_0 \in (0, \beta(t_0))$, and define

$$s_0^{\pm} := \Gamma_{t_0}\left(x_0 + i/2t \pm i0; \beta(t_0), t_0\right),$$

where we know that s_0^+ and s_0^- are real and,

$$-b(t_0) < s_0^+ < -c(t_0) < s_0^- < -1.$$

For s in the upper half plane and $t > 0$ we define

$$SC_t(s) := SC(s; \beta(t), t),$$

where we recall that

$$SC_t(s) = A(t) \int_0^s \frac{s + c(t)}{\sqrt{s(s+1)(s+b(t))}} ds, \tag{6.6.2}$$

and where the path of integration is taken in the upper half-plane. We also let

$$\Gamma_t(z) := \Gamma(z; \beta(t), t),$$

be the inverse of $SC_t(s)$, defined in the complement of the segment $[0, \beta(t)] + i/2t$ in the first quadrant $\text{Re } z, \text{Im } z > 0$.

The proof consists of analytically extending the definition of $SC_t(s)$ in both s and t to complex neighborhoods of $s_0^{\pm}$ and t_0 by using (6.6.2), and then using complex inverse function theorem to obtain an integral representation of $\Gamma_t(z)$ that can be continued in z past the cut in both directions, and makes sense for complex t close to t_0. This relies on the analyticity of $A(t)$, $b(t)$ and $c(t)$ for complex t close to t_0, and must be done separately for each of the two points s_0^+ and s_0^-. We provide the details for s_0^+.

We start by extending the meaning of the square root in (6.6.2) to complex t close to t_0 by defining

$$R(s,t) := \sqrt{s}\sqrt{s+1}\sqrt{s+b(t)}$$

where $\sqrt{s}$ and $\sqrt{s+1}$ are principal branches, and for $\sqrt{s+b(t)}$ we take the branch cut in the s-plane to be $(-\infty, -b(t_0)] \cup L_t$ where L_t is the straight segment connecting $-b(t_0)$ and $-b(t)$. For real $t > 0$ this agrees with the square root in SC_t, and corresponds to taking the branch cuts $(-\infty, -b(t)] \cup [-1, 0]$, which do not affect the definition of $SC_t(s)$ for real $t > 0$.

Recall now that we showed that $tA(t), b(t), c(t)$ admit analytic extensions in the complex t-plane to a neighborhood of the real non-negative t-axis $t \geq 0$. Let $U_{s_0^+}$ be a complex disc around s_0^+ that does not contain $-b(t_0)$ or $-c(t_0)$, and let V_{t_0} be a complex neighborhood of t_0 where $A(t), b(t), c(t)$ are analytic, $A(t) \neq 0$ (which is possible since $A(t) \neq 0$ for any real $t > 0$), and that moreover satisfies

$$(-b[V_{t_0}] \cup -c[V_{t_0}]) \cap U_{s_0} = \emptyset,$$

where by $-b[V_{t_0}]$ and $-c[V_{t_0}]$ we mean the images of V_{t_0} under $t \mapsto -b(t)$ and $-c(t)$ respectively.

Note that $\sqrt{s+b(t)}$ is jointly analytic in the set $\mathbb{C} \setminus \{(-\infty, -b(t_0)] \cup -b[V_{t_0}]\} \times V_{t_0} \subseteq \mathbb{C} \times \mathbb{C}$ since $b(t)$ is analytic in V_{t_0}, and its values avoid the cuts in s of the square root. In particular, the function $R(s,t)$ defined above is jointly analytic in s and t on the set $\mathbb{C} \setminus \{(-\infty, -b(t_0)] \cup -b[V_{t_0}] \cup [-1, 0]\} \times V_{t_0} \subseteq \mathbb{C} \times \mathbb{C}$ and extends the square root function showing up in the definition $SC_t(s)$ to complex $t \in V_{t_0}$ for the s in their common domain (which excludes the $s \in -b[V_{t_0}]$).

For $(s,t) \in U_{s_0^+} \times V_{t_0}$ define

$$\widetilde{SC}_t^+(s) := A(t) \int_0^s \frac{s + c(t)}{R(s,t)} ds$$

where the path of integration is taken along the upper half plane avoiding the set $-b[V_{t_0}]$ up to a point in $U_{s_0^+}$ (say s_0^+), and then inside $U_{s_0^+}$ to the point s. By the above remarks on $R(s,t)$, if $(s,t) \in U_{s_0^+} \times V_{t_0}$ with s in the upper half s-plane and t is real, then $\widetilde{SC}_t^+(s) = SC_t(s)$, and so $\widetilde{SC}_t^+(s)$ is a local extension of $SC_t(s)$ both to complex t close to t_0, and to s past the real axis to a neighborhood of s_0^+. The important fact to note is that $\widetilde{SC}_t^+(s)$ is jointly analytic in s and t in $U_{s_0^+} \times V_{t_0}$ since the integrand is jointly analytic in s and t in $U_{s_0^+} \times V_{t_0}$, and the same holds for its derivative $\left(\widetilde{SC}_t^+\right)'(s)$ since it is the integrand itself.

For $t > 0$ and complex w with $\operatorname{Re} w > 0$, $\operatorname{Im} w > -1/2t$ and $w \notin (0, \beta(t))$ we define

$$G(w,t) := \Gamma_t \left(w + \frac{i}{2t} \right),$$

which for fixed t is the inverse of $s \mapsto SC_t(s) - i/2t$. Thus, $G(w,t)$ has a cut at $(0, \beta(t)]$ as a function of w, which corresponds to the cut of $\Gamma_t(z)$ at $i/2t + (0, \beta(t)]$. Note that we have

$$s_0^\pm = \Gamma_t \left(x_0 \pm i/2t + i0 \right) = G(x_0 \pm i0, t).$$

We want to show that $G(x_0 + i0, t)$ is (real) analytic at $t = t_0$. We will do this by showing that $G(w,t)$ admits an analytic extension in w to a neighborhood of x_0 (i.e., past the cut), and then show that this extension is jointly analytic in w and t at (x_0, t_0).

Since $s \neq -c(t)$ and $A(t) \neq 0$, for $(s,t) \in U_{s_0} \times V_{t_0}$, it follows that $\left(\widetilde{SC}_t^+\right)'(s) \neq 0$, and so $\widetilde{SC}_t^+(s)$ is locally invertible in s for all $t \in V_{t_0}$. If we let $G^+(w,t)$ be the inverse of $s \to \widetilde{SC}_t^+(s) - i/2t$, then we see that for real t close to t_0 this $G^+(w,t)$ is an analytic extension of $G(w,t)$ to a neighborhood of $w = x_0$ (i.e., an extension past the cut). In particular, for $t = t_0$ we obtain an extension of $G(w, t_0)$ to a neighborhood of x_0, so that we have

$$G(x_0 + i0, t_0) = G^+(x_0, t_0).$$

By the complex inverse function theorem we have the integral representation

$$G^+(x_0, t_0) = \int_{\gamma_0} \frac{s\left(\widetilde{SC}_{t_0}^+\right)'(s)}{\left(\widetilde{SC}_{t_0}^+(s) - i/2t_0\right) - x_0} ds,$$

where γ_0 is a sufficiently small loop contained in U_{s_0} enclosing s_0. Now, if t is complex and sufficiently close to t_0, then the above representation continues to hold in the sense that for all t in a complex neighborhood of t_0 we have

$$G^+(x_0, t) = \int_{\gamma_0} \frac{s\left(\widetilde{SC}_{t}^+\right)'(s)}{\left(\widetilde{SC}_{t}^+(s) - i/2t\right) - x_0} ds,$$

(we need to have $s_0^+(t) := G^+(x_0, t)$ inside γ_0 and the image of γ_0 under $\widetilde{SC}_t^+(s) - i/2t$ to be contained in the domain of definition of $G^+(w, t)$). Now, since the integrand above is analytic in t for all $s \in \gamma_0$ by construction, this shows that $G^+(x_0, t)$ is (real) analytic in t at $t = t_0$, and so

$$\Gamma_t \left[x_0 + \frac{i}{2t} + i0\right] = G^+(x_0 + i0, t) = G^+(x_0, t)$$

is (real) analytic in t at $t = t_0$ as claimed.

A similar argument works for the other boundary value s_0^-. □

6.7 Proof of (6.2.3)

We can now prove (6.2.3) using the the analyticity of ρ_t in t. The starting point is the identity

$$\begin{aligned}\lim_{N\to\infty} -\frac{1}{N^2} \log Z_N(t) &= I_t[\rho_t] \\ &= \iint \log \frac{\sqrt{1 + t^2(x+y)^2}}{|x-y|} \rho_t(x)\rho_t(y) dx dy + \int \frac{x^2}{2} \rho_t(x) dx\end{aligned}$$

Now make the substitution $u = tx$, $v = ty$, which gives

$$I_t[\rho_t] = \log t + \frac{1}{2t^2}\int u^2\psi_t(u)du + \iint \log \frac{\sqrt{1+(u+v)^2}}{|u-v|}\psi_t(u)\psi_t(v)dudv$$

where

$$\psi_t(u) := \frac{1}{t}\rho_t\left(\frac{u}{t}\right)$$

and the $\log t$ term comes from the change of variables in the double integral and the fact that ψ_t is also a probability density. Note that ψ_t converges to a point mass at the origin as $t \to 0$, but also that we will be dealing with $t > 0$ throughout. Now differentiate with respect to t to obtain

$$\begin{aligned} \frac{d}{dt}I_t[\rho_t] &= \frac{1}{t} - \frac{1}{t^3}\int u^2\psi_t(u)du + \frac{1}{2t^2}\int u^2\frac{\partial}{\partial t}\left[\psi_t(u)\right]du \\ &\quad + 2\iint \log \frac{\sqrt{1+(u+v)^2}}{|u-v|}\psi_t(v)\frac{\partial}{\partial t}\left[\psi_t(u)\right]dudv \end{aligned}$$

using symmetry in the double integral. Combining the integrals gives

$$\begin{aligned} \frac{d}{dt}I_t[\rho_t] &= \frac{1}{t} - \frac{1}{t^3}\int u^2\psi_t(u)du \\ &\quad + \int \left[\frac{1}{2t^2}u^2 + 2\int \log \frac{\sqrt{1+(u+v)^2}}{|u-v|}\psi_t(v)dv\right]\frac{\partial}{\partial t}\left[\psi_t(u)\right]du. \end{aligned}$$

The variational *equality* from corollary 5.3.9 then implies that the term inside parenthesis is equal to $\ell - \log t$ on the support of ψ_t, and so

$$\begin{aligned} \frac{d}{dt}I_t[\rho_t] &= \frac{1}{t} - \frac{1}{t^3}\int u^2\psi_t(u)du + (\ell - \log t)\int \frac{\partial}{\partial t}\left[\psi_t(u)\right]du \\ &= \frac{1}{t} - \frac{1}{t^3}\int u^2\psi_t(u)du, \end{aligned}$$

since $\int \frac{\partial}{\partial t}\left[\psi_t(u)\right]du = 0$ as ψ_t is a probability density. Going now back to the x variable in the last integral gives identity (6.2.3) for $t > 0$, as claimed.

6.8 Expansion around $t = 0$

Now that we know that (section 6.7)

$$\frac{d}{dt} I_t[\rho_t] = \frac{1}{t} - \frac{1}{t} m_2(t),$$

for $t > 0$, where

$$m_2(t) = \int x^2 \rho_t(x) dx,$$

we can find a Taylor expansion around $t = 0$ for

$$\lim_{N\to\infty} \log \widehat{Z}_N(t) = I_0[\rho_0] - I_t[\rho_t].$$

More precisely we will see that the right hand side admits and analytic extension to $t = 0$. One way to do this (this is what was done in the physics literature) is to notice that the expansion (6.2.12) of ζ_t at infinity contains $m_2(t)$ in the coefficient of z^{-4}, and since $\zeta_t = f_t$, one also has an expression the the coefficient of z^{-4} in terms of $A(t), b(t), c(t)$ by from (6.5.23). This gives the following equality

$$\frac{12t^2 m_2(t) - 1}{2t^4} = \frac{(2A)^6}{5} \left(-b - b^2 + 2c + 8bc + 2b^2c - 10c^2 - 10bc^2 + 10c^3\right),$$

from which one concludes that $m_2(t)$ admits a meromorphic extension to a neighborhood of $t = 0$ (recall that $A(t)$ has a pole at $t = 0$). By further using the expansions of A, b, c around $t = 0$ one sees that $m_2(t)$ is in fact analytic at $t = 0$, with expansion given by

$$\begin{aligned} m_2(t) \quad = \quad & 1 - 2t^2 + 14t^4 - 138t^6 + 1608t^8 - 20736t^{10} \\ & + 286452t^{12} - 4160274t^{14} + 62772488t^{16} \\ & - 976099152t^{18} + 15552756144t^{20} - 252856594128t^{22} \\ & + 4181199178176t^{24} - 70146006867072t^{26} + \dots \end{aligned}$$

This shows that $I_0[\rho_0] - I_t[\rho_t]$ admits an analytic extension to $t = 0$, with Taylor expansion given by

$$\begin{aligned} I_0[\rho_0] - I_t[\rho_t] &= -\frac{2}{2!}t^2 + \frac{84}{4!}t^4 - \frac{16560}{6!}t^6 + \frac{8104320}{8!}t^8 \\ &\quad -\frac{7524679680}{10!}t^{10} + \frac{11434247193600}{12!}t^{12} - \dots \end{aligned}$$

6.9 The family of elliptic curves degenerating to a nodal curve as $t \to 0$

Even though it was of no use in the above considerations, we would like to bring attention to the fact that the function f_t is a meromorphic function on an elliptic curve E_t, and that the isomorphism class of the elliptic curve is varying as t varies. This might give some insight to the situation in future considerations.

More explicitly, using the equation

$$z = A(t) \int_0^{\Gamma_t(z)} \frac{s + c(t)}{\sqrt{s(s+1)(s+b(t))}} ds,$$

for z in the first quadrant outside the cut, one can find the differential equation that Γ_t satisfies, and conclude that the map

$$\begin{aligned} E_t &\to \mathbb{CP}^2, \\ p &\mapsto [\Gamma_t(p) : \Gamma_t'(p) : 1], \end{aligned}$$

gives a biholomorphic map onto the elliptic curve defined by

$$A(t)^2 y^2 (x + c(t))^2 = x(x+1)(x+b(t)),$$

which is isomorphic to the elliptic curve in Legendre normal form defined by

$$y^2 = x(x+1)(x+b(t)).$$

Since $b(t)$ is not constant for $t > 0$, we see that the isomorphism class of E_t is varying in t, and since $t \searrow 0$ corresponds to $b \searrow 1$, we see that the limiting curve E_0 is singular since the cubic gets a repeated factor. In other words, the family E_t degenerates to a nodal rational curve at $t = 0$.

The j invariant of E_t therefore only depends on $b(t)$, and it is given by

$$j(t) = 256\frac{(b(t)^2 - b(t) + 1)^3}{b(t)^2(b(t) - 1)^2}.$$

Under this identification the function Γ_t gets identified with the projection onto the x coordinate (in both algebraic curves).

Appendix A
Combinatorial interpretation expectation of powers of traces

In this appendix we show how the expectations of the form

$$\left\langle \left(\mathrm{Tr} M^1\right)^{n_1} \left(\mathrm{Tr} M^2\right)^{n_2} \dots \left(\mathrm{Tr} M^\nu\right)^{n_\nu} \right\rangle,$$

where $\langle f \rangle$ denotes the expectation of f with respect to GUE measure $\widetilde{\mu}_N$ (1.1.1)

$$\langle f \rangle := \int f d\widetilde{\mu}_N(M),$$

are related to counts of combinatorial objects. We start by discussing a particularly simple case.

A.1 The case of $\mathrm{Tr} M^4$

Note that in terms of the entries of the matrix we have

$$\mathrm{Tr} M^4 = \sum_{i,j,k,l=1}^{N} m_{ij} m_{jk} m_{kl} m_{li}$$

where one should notice the cycle present in the indices of the m's. Thus

$$\left(\mathrm{Tr} M^4\right)^n = \sum_{\substack{i_1,\dots,i_n \\ j_1,\dots,j_n \\ k_1,\dots,k_n \\ l_1,\dots,l_n}=1}^{N} \begin{array}{r} (m_{i_1 j_1} m_{j_1 k_1} m_{k_1 l_1} m_{l_1 i_1})(m_{i_2 j_2} m_{j_2 k_2} m_{k_2 l_2} m_{l_2 i_2}) \dots \\ \dots (m_{i_n j_n} m_{j_n k_n} m_{k_n l_n} m_{l_n i_n}) \end{array}$$

which we write compactly as

$$\left(\mathrm{Tr} M^4\right)^n = \sum_{\sigma} M_\sigma$$

where $\sigma = (i_1, i_2, \ldots, i_n, j_1, \ldots, j_n, k_1, \ldots k_n, l_1, \ldots, l_n)$ runs over the N^{4n} choices for each index from 1 to N and M_σ is defined as

$$M_\sigma = (m_{i_1 j_1} m_{j_1 k_1} m_{k_1 l_1} m_{l_1 i_1})(m_{i_2 j_2} m_{j_2 k_2} m_{k_2 l_2} m_{l_2 i_2}) \ldots (m_{i_n j_n} m_{j_n k_n} m_{k_n l_n} m_{l_n i_n}),$$

(the reason why we have kept the parenthesis will be apparent below).

Now, for the expectation of $(\mathrm{Tr} M^4)^n$ we have

$$\left\langle \left(\mathrm{Tr} M^4\right)^n \right\rangle = \sum_\sigma \langle M_\sigma \rangle \, ,$$

and to compute $\langle M_\sigma \rangle$ one may use *Wick's lemma.*

A.1.1 Lemma (Wick's Lemma)**.** *If $f_1, \ldots, f_{2m}$ are $2m$ linear functions of Gaussian mean zero independent random variables (meaning that each f_i is a linear combination of these random variables), then*

$$\langle f_1 \ldots f_{2m} \rangle = \sum_{\text{couplings}} \langle f_{i_1} f_{j_1} \rangle \langle f_{i_2} f_{j_2} \rangle \ldots \langle f_{i_m} f_{j_m} \rangle$$

where $\langle f \rangle$ denotes expectation with respect to the joint probability distribution of the Gaussian variables, and where a coupling of the set $\{f_1, f_2, \ldots, f_{2m}\}$ is a partition of the set into m sets of 2 elements

$$\{f_1, f_2, \ldots, f_n\} = \{f_{i_1}, f_{j_1}\} \sqcup \{f_{i_2}, f_{j_2}\} \sqcup \ldots \sqcup \{f_{i_m}, f_{j_m}\}$$

where the ordering is not important.

For a proof of Wick's lemma see lemma 7.6 in [15].

A.1.2. A way to visualize a coupling is to write the $2m$ terms next to each other

$$f_1 \quad f_2 \quad f_3 \quad \ldots f_{2m-1} \quad f_{2m}$$

and connect the two terms in the set $\{f_{i_*}, f_{j_*}\}$ with an arc for $* = 1, \ldots, m$. Regarding the number of couplings, note that there are $\binom{2m}{2}$ choices for the first pair, $\binom{2m-2}{2}$

choices for the second pair, and so on, and since the order of the m pairs is not important, we see that there are

$$\frac{\binom{2m}{2}\binom{2m-2}{2}\dots\binom{2}{2}}{m!} = \frac{(2m)!}{2^m m!} = 1\cdot 3\cdot 5\cdot \ldots \cdot (2m-1) =: (2m-1)!!$$

distinct couplings.

A.1.3. We remark that Wick's lemma is of particular use in this setting, since one can check that for products of two matrix entries one has

$$\langle m_{ij} m_{kl}\rangle = \delta_{il}\delta_{jk},$$

because if $i = l$ and $j = k$ then $m_{ij}m_{kl} = |m_{ij}| = (\text{Re } m_{ij})^2 + (\text{Im } m_{ij})^2$ and we get $\left\langle (\text{Re } m_{ij})^2 + (\text{Im } m_{ij})^2 \right\rangle = 2(1/2) = 1$, and if this is not the case, then $m_{ij}m_{kl}$ is a quadratic form of distinct and independent random variables, and so its expectation is equal to zero.

A.1.4. Going back to the expectation of $(\text{Tr}M^4)^n$, we can now write

$$\left\langle \left(\text{Tr}M^4\right)^n \right\rangle = \sum_{\sigma} \sum_{\substack{\text{couplings} \\ \text{of the } 4n \\ \text{terms in } M_\sigma}} \langle \mathcal{C}_1\rangle \dots \langle \mathcal{C}_{2n}\rangle$$

where $\mathcal{C}_i = m_{\alpha\beta}m_{\gamma\delta}$ if $\mathcal{C}_i$ is the couple corresponding to $\{m_{\alpha\beta}, m_{\gamma\delta}\}$. Note that this is jumbling-up the indices in a non-trivial way because of the cycles in the double indices.

Note that $\langle \mathcal{C}_1\rangle \dots \langle \mathcal{C}_{2n}\rangle$ is either 0 or 1, and it is 1 only when some conditions for the double indices in all of the couples in the coupling are satisfied. Specifically, if $\mathcal{C} = m_{\alpha\beta}m_{\gamma\delta}$, then

$$\langle \mathcal{C}\rangle = 1 \iff \alpha = \delta \text{ and } \beta = \gamma$$

which is independent of the actual values of $\alpha, \beta, \gamma, \delta$ as long as the equalities hold. This shows that the value of $\langle \mathcal{C}_1\rangle \dots \langle \mathcal{C}_{2n}\rangle$ only depends on the coupling, and not on the specific values of the indices. Because of this, we may change the order of summation above to obtain

$$\left\langle \left(\mathrm{Tr} M^4\right)^n \right\rangle = \sum_{\substack{\text{couplings} \\ \text{of the } 4n \text{ double} \\ \text{indices in } M_\sigma}} \sum_{\sigma} \langle \mathcal{C}_1 \rangle \ldots \langle \mathcal{C}_{2n} \rangle$$

where we are now considering the couplings of the generic $4n$ double indices

$$\{i_1 j_1, j_1 k_1, k_1 l_1, l_1 i_1, i_2 j_2, \ldots\ldots, i_n j_n, j_n k_n, k_n l_n, l_n i_n\}$$

and then assigning them specific values when we sum over σ. In this case $\mathcal{C} = m_{\alpha\beta} m_{\gamma\delta}$ if $\mathcal{C}$ is the couple corresponding to the indices $\{\alpha\beta, \gamma\delta\}$.

A.1.5 Example $n = 1$. We have

$$\mathrm{Tr} M^4 = \sum_{i,j,k,l=1}^{N} m_{ij} m_{jk} m_{kl} m_{li},$$

and if we compute $\langle \mathrm{Tr} M^4 \rangle$ using Wick's lemma we get

$$\left\langle \mathrm{Tr} M^4 \right\rangle = \sum_{\substack{\text{couplings} \\ \text{of the 4 double} \\ \text{indices in } M_\sigma}} \sum_{\sigma} \langle \mathcal{C}_1 \rangle \langle \mathcal{C}_2 \rangle$$

where in this case we need to consider the couplings of the four double indices

$$\{ij, jk, kl, li\}$$

(we can drop the subindex of the indices in this case since it is not necessary).

There are 3 such couplings, given by

A. $\{ij, jk\}, \{kl, li\}$

B. $\{ij, kl\}, \{jk, li\}$

C. $\{ij, li\}, \{jk, kl\}$

and so we have

$$\begin{aligned}\langle \mathrm{Tr} M^4 \rangle &= \sum_\sigma \langle m_{ij} m_{jk} \rangle \langle m_{kl} m_{li} \rangle + \langle m_{ij} m_{kl} \rangle \langle m_{jk} m_{li} \rangle + \langle m_{ij} m_{li} \rangle \langle m_{jk} m_{kl} \rangle \\ &= \sum_\sigma \delta_{ik} + \sum_\sigma \delta_{il} \delta_{jk} \delta_{ji} \delta_{kl} + \sum_\sigma \delta_{jl} \\ &= N^3 + N + N^3 \\ &= 2N^3 + N\end{aligned}$$

A.1.6 Four-valent diagrams. We now explain an interpretation for the exponents and the coefficients for N that we are obtaining.

To figure out which indices σ make $\langle \mathcal{C}_1 \rangle \ldots \langle \mathcal{C}_{2n} \rangle \neq 0$ for a given coupling we construct the following graph: we set up n vertices, corresponding to the n groups of indices (according to the parenthesis) in

$$M_\sigma = (m_{i_1 j_1} m_{j_1 k_1} m_{k_1 l_1} m_{l_1 i_1})(m_{i_2 j_2} m_{j_2 k_2} m_{k_2 l_2} m_{l_2 i_2}) \ldots (m_{i_n j_n} m_{j_n k_n} m_{k_n l_n} m_{l_n i_n})$$

and to each vertex we assign four double edges, as shown in figure A.1

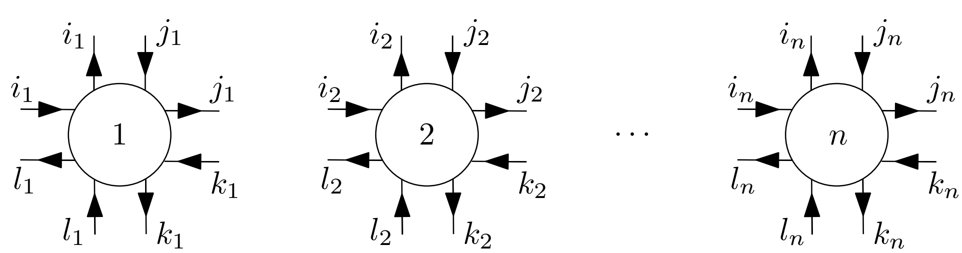

FIGURE A.1. n four-valent vertices.

Each double edge corresponds to one tuple of indices that appears in the group corresponding to that vertex. For example, the first vertex has a double edge corresponding to $i_1 j_1$ (pointing upwards in the picture). Now, if we are going to pair say $i_1 j_1$ with some other double index say $j_n k_n$ we can connect the double edges corresponding to them and if we do this for all edges we get a coupling.

We call the resulting directed labeled multi-graph a **diagram with four valent vertices**. One can clearly see that this construction is bijective, in the sense that there is a a bijection between the couplings and diagrams.

A.1.7 The reason why the double edges and arrows are useful. Note that to encode the coupling we really do not need double edges (we could use single edges with labels $(i_* j_*)$, for example). The real use of the double edges is to encode the information that makes $\langle \mathcal{C}_1 \rangle \ldots \langle \mathcal{C}_{2n} \rangle \neq 0$ as we now explain. Each edge in the double edge corresponds to one of the indices and we add orientations to the sides of the double edge by specifying that the first index corresponds to the edge with outgoing orientation, and the second index corresponds to the one with incoming orientation. Then when we join the edges and we make sure the orientations match, and for example, in the case of pairing $i_1 j_1$ with $j_n k_n$ this will look as shown in figure A.2.

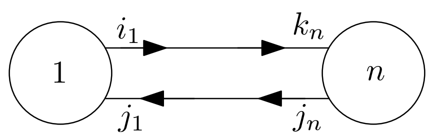

FIGURE A.2. The connection of the half-edges.

Note that making the orientations match is precisely matching the indices according to the rule which make $\langle m_{i_1 j_1} m_{j_n k_n} \rangle = 1$, i.e. $i_1 = k_n$ and $j_1 = j_n$.

In general coupling of the double edges matching the orientations is encoding the equalities of the indices that will make that couple to give $\langle m_c \rangle \neq 0$. If c is the couple corresponding to $\alpha\beta$ and $\gamma\delta$ then $\langle m_{\alpha\beta} m_{\gamma\delta} \rangle = 1$ if and only if $\alpha = \delta$ and $\beta = \gamma$, and this will be encoded in the graph as

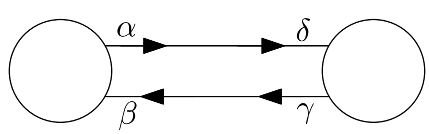

FIGURE A.3. The indices in the connection of two half edges.

A.1.8 The faces of a diagram. Thus, from the graph we construct from a given pairing we can clearly see which conditions on the $4n$ indices $\{i_\nu, j_\nu, k_\nu, l_\nu\}_{\nu=1}^n$ imply that for that particular coupling we obtain

$$\langle \mathcal{C}_1 \rangle \ldots \langle \mathcal{C}_{2n} \rangle = 1,$$

by following the labels of individual edges. These conditions are always given as equalities of particular indices (e.g. $i_1 = j_7 = i_3 = k_5$ and $j_3 = l_2$), and are always cycles, since each individual index appears both as an incoming and an incoming index, and so the rules eventually cycle back you where you started. We call these cycles of indices **faces**.

A.1.9 Example $n = 1$. As we discussed above, there are 3 couplings given by

A. $\{ij, jk\}, \{kl, li\}$

B. $\{ij, kl\}, \{jk, li\}$

C. $\{ij, li\}, \{jk, kl\}$

and the diagrams we obtain from the couplings are shown in figure A.4.

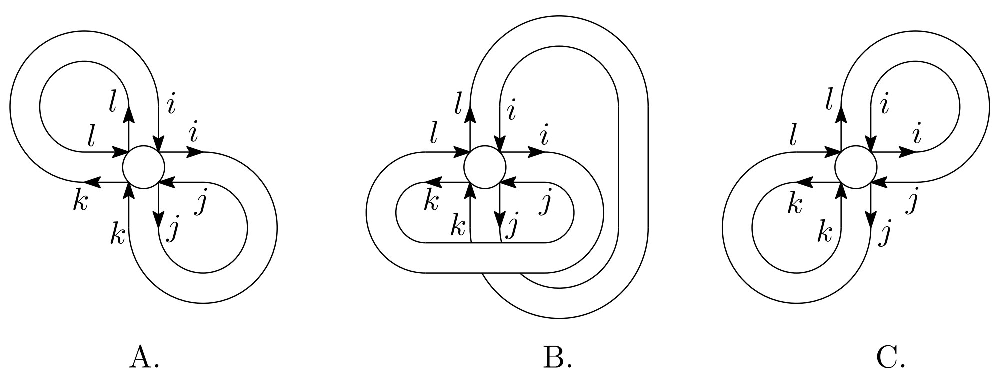

FIGURE A.4. The three four-valent diagrams with one vertex.

Note that following the arrows in the diagram corresponding to coupling A one obtains the cycles

$$\begin{array}{c} i \to k \to i \\ l \to l \\ j \to j \end{array}$$

from which we can read the conditions $i = k$, $l = l$, $j = j$ which are the conditions that make the term $\langle m_{ij} m_{jk}\rangle\langle m_{kl} m_{li}\rangle$ corresponding to the coupling be nonzero (and so equal to 1) as was discussed in A.1.5.

For coupling B we have the cycle

$$i \to l \to k \to j \to i$$

from which we see that for the term $\langle m_{ij} m_{kl}\rangle\langle m_{jk} m_{li}\rangle$ corresponding to the coupling to be nonzero (and so equal to 1) we must have $i = j = k = l$, so all indices must be equal. Note that in the computation

$$\begin{aligned} \left\langle \mathrm{Tr} M^4 \right\rangle &= \sum_\sigma \langle m_{ij} m_{jk}\rangle\langle m_{kl} m_{li}\rangle + \langle m_{ij} m_{kl}\rangle\langle m_{jk} m_{li}\rangle + \langle m_{ij} m_{li}\rangle\langle m_{jk} m_{kl}\rangle \\ &= \sum_\sigma \delta_{ik} + \sum_\sigma \delta_{il}\delta_{jk}\delta_{ji}\delta_{kl} + \sum_\sigma \delta_{jl} \\ &= N^3 + N + N^3 \\ &= 2N^3 + N \end{aligned}$$

the exponent of N that shows up is the number of faces of each coupling.

A.1.10 The count for an arbitrary power of TrM^4. For general n see that if the four-valent diagram of the coupling has F faces then

$$\sum_\sigma \quad \langle \mathcal{C}_1\rangle \ldots \langle \mathcal{C}_{2n}\rangle = N^F$$

since $\langle \mathcal{C}_1\rangle \ldots \langle \mathcal{C}_{2n}\rangle = 1$ only if all the indices in each cycle are equal, and there are N choices for the value of the index of each cycle (which we had already seen in the example $n = 1$ above).

Therefore,

$$\left\langle \left(\mathrm{Tr} M^4\right)^n \right\rangle = \sum_{\substack{\text{couplings of} \\ \text{the } 4n \text{ double} \\ \text{indices in } M_\sigma}} \sum_{\sigma} \langle \mathcal{C}_1 \rangle \dots \langle \mathcal{C}_{2n} \rangle = \sum_{F=1}^{\infty} A_F N^F$$

where A_F is the number of couplings with F faces (note this sum is finite since F can be at most $3n$).

A.2 The general case

There is nothing special about valence 4 in the above discussion, and one may consider expectations of the form

$$\left\langle \left(\mathrm{Tr} M^1\right)^{n_1} \left(\mathrm{Tr} M^2\right)^{n_2} \dots (\mathrm{Tr} M^\nu)^{n_\nu} \right\rangle .$$

Here, as we explain below, the Wick couplings will correspond to diagrams with n_1 vertices of valence 1, n_2 vertices of valence 2 and so on (we will see that the expectation is 0 if the sum of the valences of all the vertices is not even, so only cases where we can make couplings are of interest).

A.2.1 Definition. A **(labeled) diagram with n_j vertices of valence j** is an oriented labeled multi-graph with the following properties:

- There are $\sum n_j$ vertices which are numbered by tuples

  $$(a, b) = (\text{vertex } \#, \text{valence}),$$

  where b is half the number of edges incident to it (which we call the **valence** of the vertex) and $a = 1, \dots, n_b$.

- The $2b$ edges incident to vertex (a, b) are grouped in consecutive pairs while moving clockwise, with one oriented edge pointing outwards and one oriented inwards. The incoming edge of each double edge is always clockwise from the outgoing edge.

- The $2b$ edges incident to vertex (a, b) are labeled clockwise by

$$i_1^{(a,b)}, i_2^{(a,b)}, \ldots, i_b^{(a,b)},$$

where each index is the label of the incoming edge of a double edge, and also of the the outgoing edge of the next double edge (moving clockwise around the vertex). For example, figure A.5 shows the second vertex of valence 4.

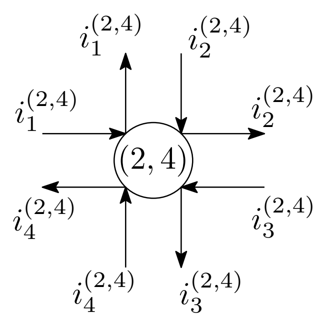

FIGURE A.5. The second vertex of valence 4.

- Each pair of oriented edges is paired to another pair of oriented edges (possibly from the same vertex) in such a way that the orientations match. We call the resulting double oriented edges the **edges** of the diagram. The intersections between edges are ignored.

For example, figure A.6 shows a diagram with two vertices of valence two, and two vertices of valence three.

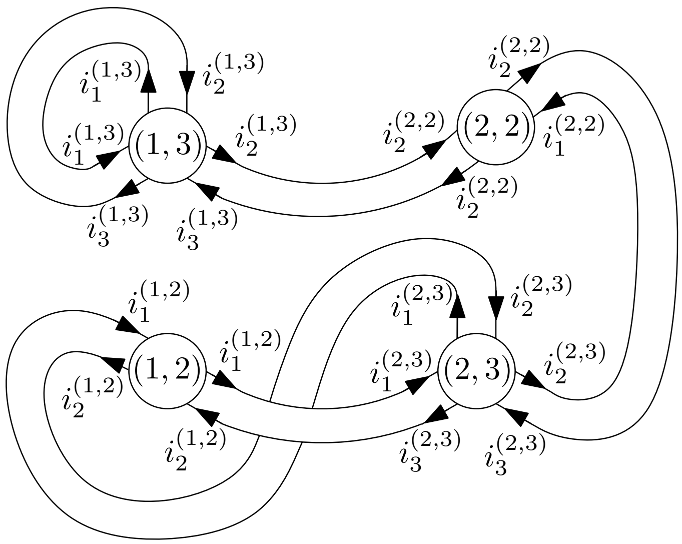

FIGURE A.6. A diagram with two vertices of valence two and two vertices of valence three.

A.2.2. We remark that the particular way in which the edges are connected is not important. What is important is the information of which edges are connected with each other. In particular, there are many ways to draw the diagram in figure A.6 (in particular the "intersection" of the edges can be avoided for this diagram).

A.2.3 Combinatorial interpretation of $(\mathrm{Tr}M^1)^{n_1}\,(\mathrm{Tr}M^2)^{n_2}\ldots(\mathrm{Tr}M^\nu)^{n_\nu}$. We start by expanding $(\mathrm{Tr}M^1)^{n_1}(\mathrm{Tr}M^2)^{n_2}\dots(\mathrm{Tr}M^\nu)^{n_\nu}$. We have

$$\begin{aligned}
(\mathrm{Tr}M^j)^{n_j} &= \left(\sum_{i_1,\dots,i_j=1}^{N} m_{i_1 i_2}\dots m_{i_j i_1}\right)^{n_j} \\
&= \left(\sum_{i_1^{(1,j)},\dots,i_j^{(1,j)}=1}^{N} m_{i_1^{(1,j)} i_2^{(1,j)}}\dots m_{i_j^{(1,j)} i_1^{(1,j)}}\right)\times \\
&\quad \left(\sum_{i_1^{(2,j)},\dots,i_j^{(2,j)}=1}^{N} m_{i_1^{(2,j)} i_2^{(2,j)}}\dots m_{i_j^{(2,j)} i_1^{(2,j)}}\right)\times\dots \\
&\quad \dots\times\left(\sum_{i_1^{(n_j,j)},\dots,i_j^{(n_j,j)}=1}^{N} m_{i_1^{(n_j,j)} i_2^{(n_j,j)}}\dots m_{i_j^{(n_j,j)} i_1^{(n_j,j)}}\right) \\
&= \sum_{\substack{1\le k\le n_j \\ 1\le i_1^{(k,j)},\dots,i_j^{(k,j)}\le N}} \prod_{s=1}^{n_j} m_{i_1^{(s,j)} i_2^{(s,j)}}\dots m_{i_j^{(s,j)} i_1^{(s,j)}},
\end{aligned}$$

one so one can write $(\mathrm{Tr}M^1)^{n_1}(\mathrm{Tr}M^2)^{n_2}\dots(\mathrm{Tr}M^\nu)^{n_\nu}$ as

$$\prod_{j=1}^{\nu}(\mathrm{Tr}M^j)^{n_j} = \sum_{\substack{1\le r\le\nu \\ 1\le k\le n_r \\ 1\le i_1^{(k,r)},\dots,i_j^{(k,r)}\le N}} \prod_{j=1}^{\nu}\prod_{s=1}^{n_j} m_{i_1^{(s,j)} i_2^{(s,j)}}\dots m_{i_j^{(s,j)} i_1^{(s,j)}}.$$

If we then use Wick's lemma on the expectation of the product on the right, then each resulting term from a coupling will correspond uniquely to one of the diagrams defined in A.2.1. Note that if $\sum_j n_j j$ (the sum of the valences of all the vertices, and the number of terms in the product above) is not even, then we cannot use Wick's lemma, but then $\left\langle\prod_{j=1}^{\nu}(\mathrm{Tr}M^j)^{n_j}\right\rangle = 0$ because in the last equation all monomials have an odd number of terms and making the substitution $M \leftrightarrow -M$ gives the result.

This gives a unique diagram for each coupling, and so gives a bijection between the couplings and the diagrams, and so one obtains

$$\left\langle \left(\mathrm{Tr}M^1\right)^{n_1} \left(\mathrm{Tr}M^2\right)^{n_2} \dots \left(\mathrm{Tr}M^\nu\right)^{n_\nu} \right\rangle = \sum_{F=1}^{\infty} A_{F,n_1,\dots n_\nu} N^F$$

where $A_{F,n_1,\dots n_\nu}$ is the number of diagrams with F faces with n_j vertices which are j-valent.

A.3 The relationship with maps

A.3.1 From vertices to stars. If one takes a diagram and collapses its double edges to single edges without orientation, one obtains a regular graph, and to be able to recover the diagram from this procedure one needs to retain some information. A way to do this is to specify the orientation of the collapsed vertices (using some convention), and leave a special marking on the edge that corresponded to the $i_1 i_2$ edge at each vertex (see figure A.5). An example of this "collapsing" procedure is shown in figure A.7, where the arrow around the collapsed vertex is there to specify the orientation of the vertex, so that when one "fattens" the edges back one knows which is the outgoing and which is the incoming arrow. In such a way, we obtain a star as discussed in 2.1.1, where we are ignoring the colors since we have discussed everything for the case of one matrix where the type of a star is determined by its valence. It is clear that this gives is a bijection between the vertices of diagrams and the stars 2.1.1.

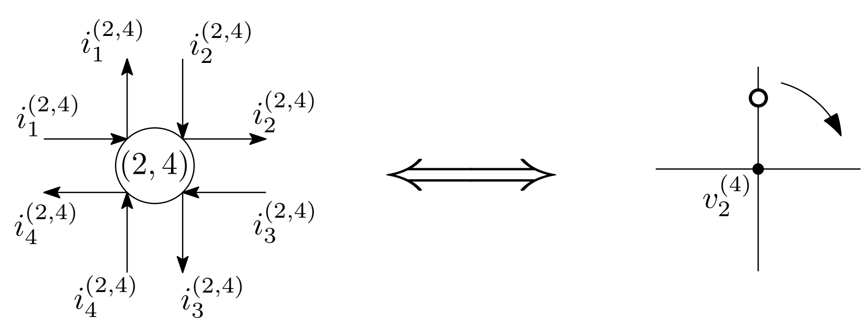

FIGURE A.7. Collapsing a vertex of valence 4 into a star of type $q_4(M) = M^4$.

A.3.2 The oriented surface defined by a connected diagram. Note that each face of the diagram (given by a cycle following the arrows in the diagram) defines a glueing map of a 2-cell (disc) onto the boundary of the diagram. If one takes a connected diagram, and glues these discs, one will end up with a connected surface. Since exactly two faces are glued to each edge of the diagram, the surface has no boundary, and since it is constructed using finitely many 2-cells, it will be compact. In such a way, each diagram can be interpreted as a cellular structure of a surface (see for example [17] for the theory of CW-complexes), and using cellular homology one can see that this surface is always orientable (one shows that $H_2 = \mathbb{Z}$ by studying the boundary map from faces to edges, and using the fact that every edge only shows up in only two faces and with opposite orientation if one uses the cycles in the diagrams to orient the faces). We do not supply the details here, but instead illustrate the construction for two of the diagrams we saw in the expansion of $\langle \mathrm{Tr}\ M^4 \rangle$ in figure A.8.

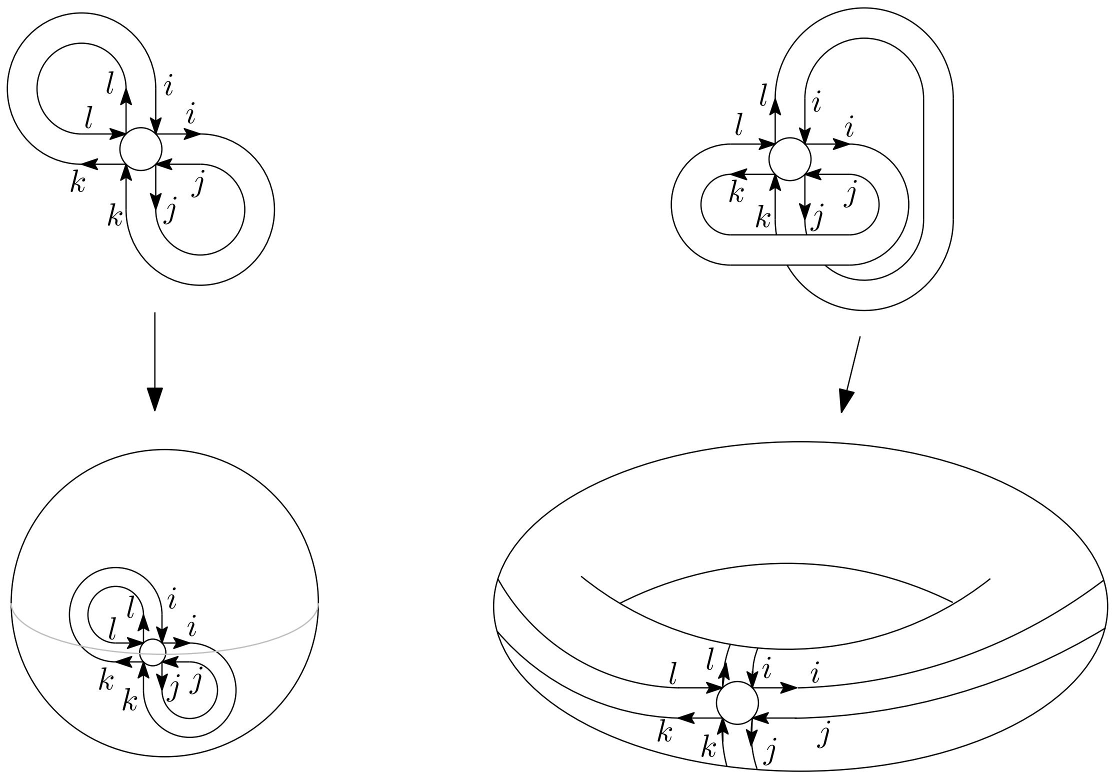

FIGURE A.8. The diagrams embedded in their corresponding surfaces.

A.3.3. We remark that because of the arrows on each vertex, the diagram not only defines the surface, but determines also determines an orientation on the surface. We call the corresponding oriented surface, together with its diagram as in figure A.8 the **embedding of the diagram**.

A.3.4 The genus of the surface. Using Euler's formula, one sees that the genus of the resulting surface is directly related to the number of faces of the diagram, which correspond to the number of discs used in the above construction. More precisely, if a *connected* diagram has n_j vertices of valence j, then the genus g and the number of faces are related by

$$\begin{aligned} 2-2g &= V-E+F \\ &= \sum_{j=1}^{\nu} n_j - \frac{1}{2}\sum_{j=1}^{\nu} j n_n + F \\ &= \frac{1}{2}\sum_{j=1}^{\nu}(j-2)n_j + F. \end{aligned}$$

We call g the **genus of the diagram**.

A.3.5 From diagrams to maps. Finally, if to an embedded diagram one preforms the collapsing construction described in A.3.1, one obtains a map as in chapter 2 (we described here for the case with only one matrix, where the types of the vertices - using the terminology in 2.1.1 - are only determined by their valence). This procedure is illustrated in figure A.9, where since all the vertices have type M^3, where we have simplified the labeling scheme of the diagram slightly since all vertices have the same type.

Note that the orientations at all the collapsed vertices agree, and correspond to a fixed orientation of the surface.

A.3.6 The notion of equivalence of maps. Since the integrals we have discussed have a precise interpretation in terms of counts of diagrams, to have a precise statement of the genus expansion in terms of maps (such as the one we gave in section 2.1), one needs to define equivalence of the maps to correspond to *equality* of the associated diagrams, where two diagrams are equal if the couplings that define them are the same. Viewing the diagram as the data of a CW-complex structure of the surface, together with a fixed orientation on this surface, it follows that the diagrams corresponding to two maps are equal if and only if there is an *orientation preserving homeomerophism* of the surface taking one map to the other that is compatible with all the labels and special markings on the maps. This is the definition of equivalence of maps we gave in 2.1.3.

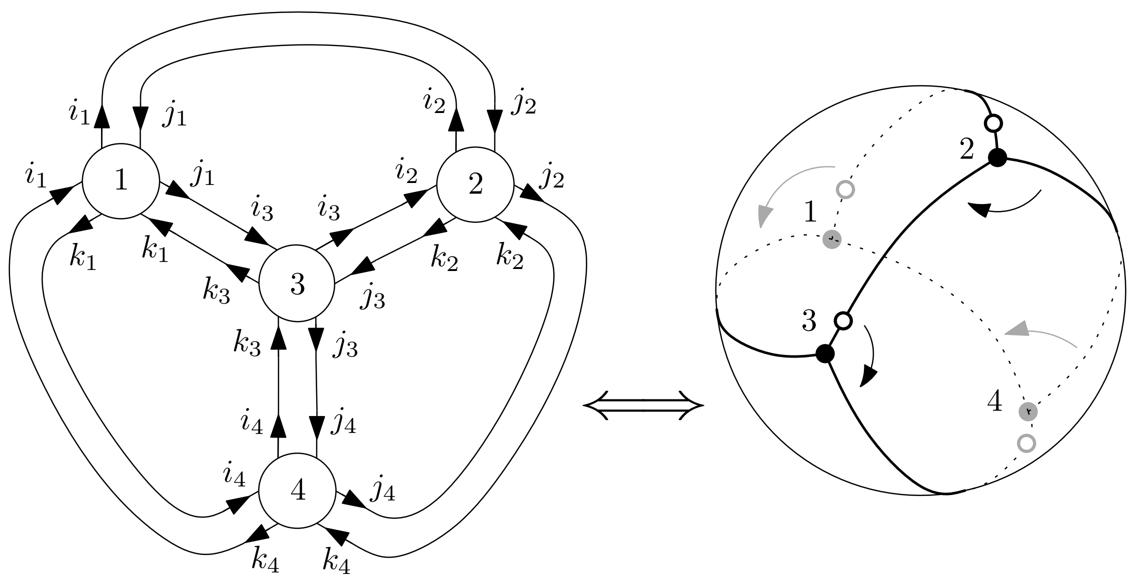

FIGURE A.9. The bijection between diagrams and maps.

Appendix B
The dual of a Diagram.

In this appendix we present a construction of the dual diagram of a diagram. The arguments will be given for the case in which there is only one matrix, where the type of a vertex (or star) is determined just by its valence. We assume the diagrams are labeled as in appendix A, where we recall that the faces of the diagram, which are the oriented cycles in the diagram, are not assumed to be labeled explicitly.

B.1 Definition of the dual diagram. We construct the dual diagram as follows:

1. Make a vertex for each face F of the diagram, and give each vertex as many (double) half-edges as the boundary of the face F has. If the original face had a label, use it as the label of this new vertex, otherwise leave the face unlabeled. Give the half-edges orientations with the incoming edge of each double edge clockwise from the outgoing edge (just as in the original diagram). Now label the edges counter-clockwise with the labels that the boundary of F has with its given orientation with each label appearing in two consecutive edges, each from a different double edge. For example, if F is the face

$$\sigma \to \tau \to \upsilon \to \dots,$$

then the corresponding (dual) vertex F has labeled edges as in figure B.1

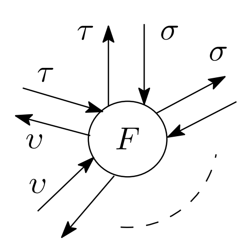

Figure B.1. Labeling of the edges of the dual of face F.

Note that the labeling scheme of the dual vertex is "increasing" in the opposite direction as the labels of a canonically labeled vertex (which increase clockwise). This will be discussed below.

2. Couple the double edges according to the following rule: If σ is any label in the original diagram, and the double edge with outgoing edge labeled with σ has incoming edge labeled with τ, then in the dual diagram we pair the outgoing edge τ with the incoming edge σ (note the reversal of the roles of incoming and outgoing indices).

   In the case when the diagram has canonical labeling, this corresponds to pairing outgoing $i_{\alpha}^{(a,b)}$ with incoming $i_{\alpha-1}^{(a,b)}$, where it is understood that if $\alpha = 1$, then $\alpha - 1$ is replaced by b.

3. As described below, the faces of this dual diagram are in one to one correspondence with the vertices of the original diagram. If the vertices of the original diagram had labels, use these labels to label the faces of the dual diagram.

We must check that the description of the pairing in step 2 is compatible when done on both ends of the double edge. This is the case because if in the original diagram we have an edge between faces F and F' as in the figure,

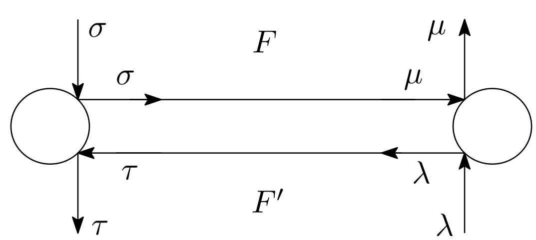

FIGURE B.2. Edge in the original diagram separating faces F and F'.

then in the dual diagram, the pairing of outgoing τ with incoming σ is the same as the pairing of outgoing μ with incoming λ as seen in the figure below.

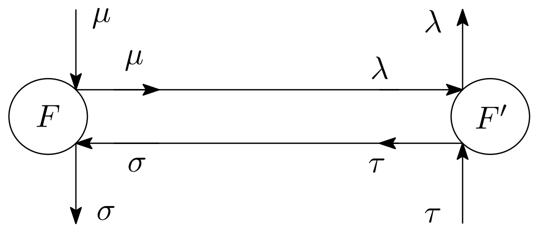

FIGURE B.3. Definition of the pairing in the dual diagram.

Regarding the faces of the dual diagram, note that they are in correspondence with the vertices of the original graph. For example, if the diagram has canonical labeling, the dual face to vertex (a, b) is the face corresponding to the cycle

$$i_b^{(a,b)} \to i_{b-1}^{(a,b)} \to \ldots \to i_1^{(a,b)} \to i_b^{(a,b)},$$

in the dual diagram (which is present in the dual diagram because of step 1). According to step 3, this dual face gets the label (a, b).

Thus, the dual diagram has as many (double) edges as the original diagram, it has as many faces as the number of vertices of the original diagram, and as many vertices as the number of faces of the original diagram.

B.2 Why the dual diagram is indeed a dual in a topological sense. If we assume the original diagram is connected, then we can view it as a map embedded in an oriented surface as described in A.3. Given a map, its dual is a graph embedded in the same surface which is constructed by defining a vertex inside each face, and then connecting these vertices by an edge if the corresponding faces share and edge (the faces may agree). When constructing this edge one must make sure that the only edge of the original diagram it intersects is the one that implies its existence.

This is precisely what we are accomplishing with the definition of the dual in step 2, and a nice way to see this is by drawing an edge and its dual next to each other as in the figure below for a canonically labeled diagram.

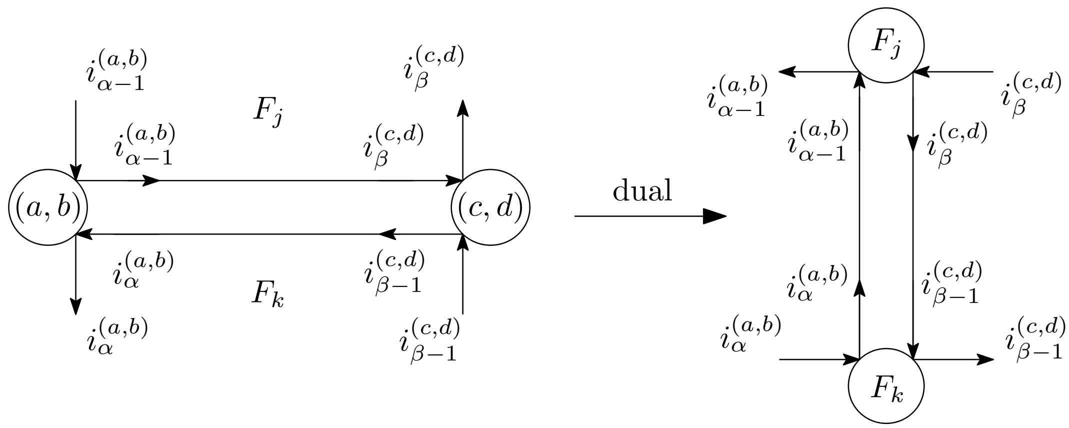

FIGURE B.4. Construction of the dual edges

Note that in both diagrams the local orientation of the boundaries of faces is the same around the vertices as can be seen in figure B.5.

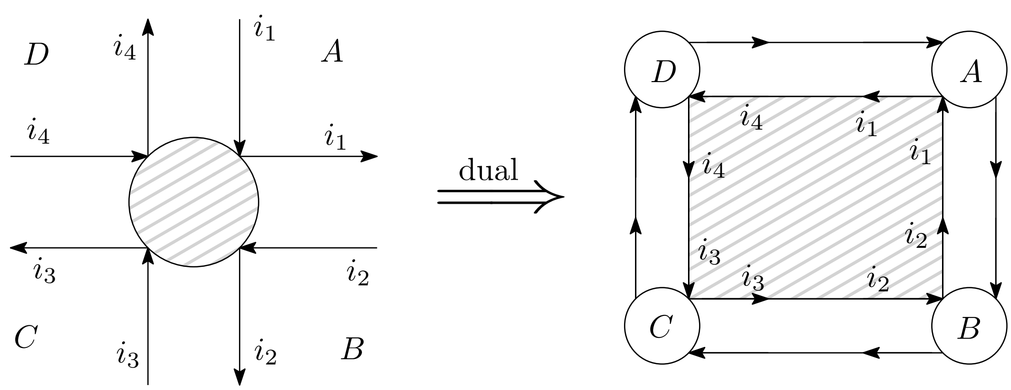

FIGURE B.5. The dual of a vertex.

Thus, this dualization is preserving the orientation of the surface, and this is the reason why we chose to make the construction with dual faces corresponding to decreasing sequences in the indices.

B.3 Note about non-connected diagrams. If the diagram is not connected, then one can see that the procedure constructs the dual to each diagram for each connected component with the same argument as above.

B.4 Example. Consider the diagram shown in figure B.6 with faces F_1, F_2, F_3.

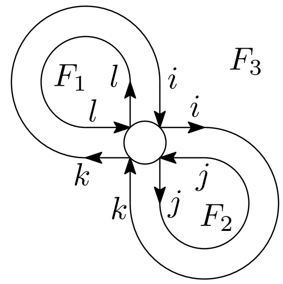

FIGURE B.6. A diagram with one vertex of valence four.

Then the dual consists of 3 vertices with edges labeled as shown in figure B.7, and

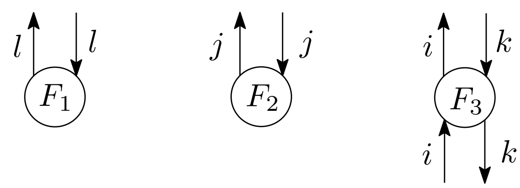

FIGURE B.7. The dual of the vertices of the diagram in figure B.6

for the pairings we join $l \to k$, $k \to j$, $j \to i$, $i \to l$, which gives the dual diagram presented in figure B.8 where the vertex of the original diagram corresponds to the only face

$$l \to k \to j \to i \to l.$$

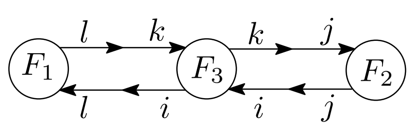

FIGURE B.8. The dual diagram of the diagram in figure B.6

B.5 How to recover the original diagram from its dual. The same procedure we described to construct the dual diagram from a diagram will give the original diagram out of its dual.

B.5.1 Proposition. *The dual of the dual of a diagram is the diagram itself.*

Proof. For simplicity we assume that the diagram is canonically labeled with vertices indexed by tuples $(a, b) = (\text{vertex } \#, \text{valence})$, where $b = 1, \ldots, \nu$ and $a = 1, \ldots, n_b$ (it is understood that if $n_j = 0$ then there are no vertices of valence j), and with the double edges around vertex (a, b) labeled clockwise by $i_1^{(a,b)}, i_2^{(a,b)}, \ldots, i_b^{(a,b)}$.

Let G be the original diagram, G' be its dual, and G'' be the dual of G'. As we described above, the face dual the vertex (a, b) in G is the cycle

$$i_b^{(a,b)} \to i_{b-1}^{(a,b)} \to \ldots \to i_1^{(a,b)} \to i_b^{(a,b)},$$

and was labeled as face (a, b) in the dual. We will temporarily denote this face with a prime as $(a, b)'$ to avoid confusion, but remark that its label is really (a, b). To construct the dual of G' step 1 tells us to construct one vertex for each face of G' and label it using the labels of the faces if present. Thus, G'' will have vertices corresponding to vertices of G, with the same labels. Again, for the sake of clarity, we refer to the vertex (a, b) of G'' as $(a, b)''$ to specify that it is a vertex of the double dual.

Now we define the double edges of vertices of G''. According to the procedure, we label the double edges around vertex (a, b) clockwise according to the reverse orientation of the face $(a, b)'$. Thus to vertex $(a, b)''$ in G'' we add b edges labeled clockwise in the order $i_1^{(a,b)} \to i_2^{(a,b)} \to \ldots \to i_b^{(a,b)}$. This is precisely the labeling and valence of the original vertex (a, b) of G.

Therefore, it only remains to be checked that step 2 defines the same pairings for G'' as the ones we have for G. Say we want to figure out what to pair the outgoing edge $i_\sigma^{(a,b)}$ in G'' as shown in the figure.

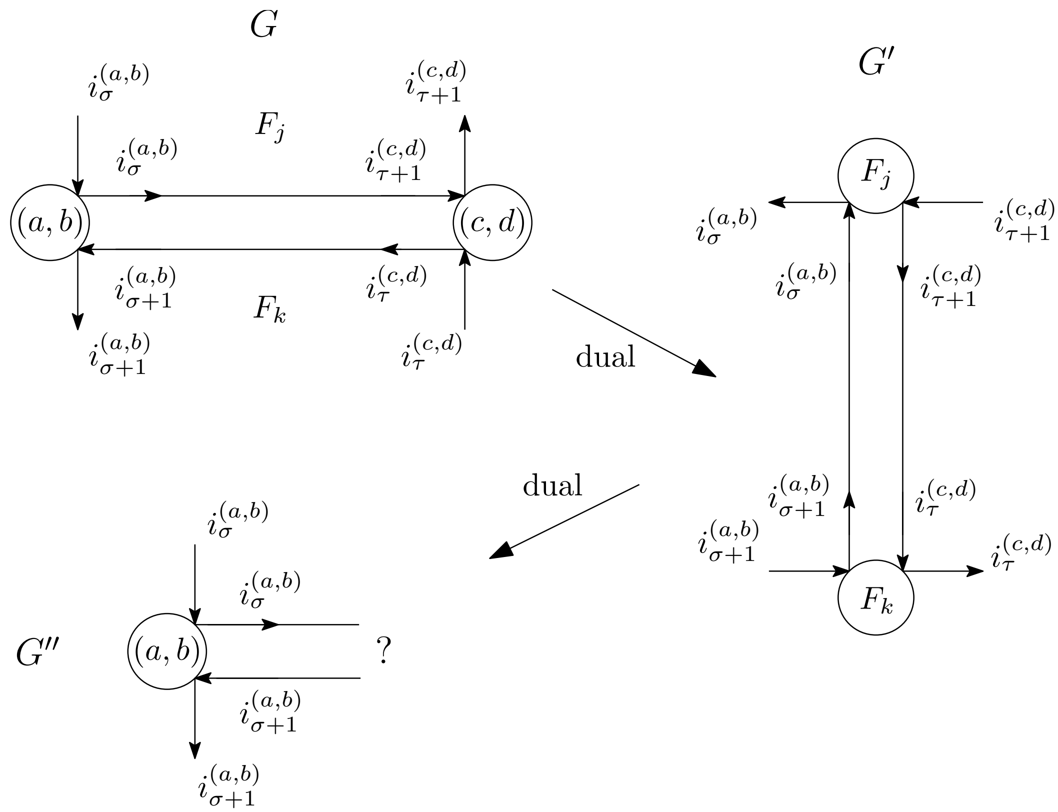

FIGURE B.9. Construction of the dual

According to step 2 we need to find $i_\sigma^{(a,b)}$ as an incoming edge in G' with outgoing edge say π, and then pair outgoing $i_\sigma^{(a,b)}$ with the incoming label π in G''. But by the way we constructed G' we see that π is no other than $i_{\tau+1}^{(c,d)}$ (see the figure), and so in G'' we pair outgoing $i_\sigma^{(a,b)}$ with incoming $i_{\tau+1}^{(c,d)}$. This is is precisely the same pairing as in G, and so this completes the proof that G'' is the same as G. $\square$

B.6 A note about the labels. If our diagrams happen to have extra labels for the vertices, edges or the faces, then in the construction of the dual we can transfer these labels. The same is true for colorings of vertices, edges of faces.

Appendix C
The genus expansion for $\widehat{Z}_N(t)$

In this appendix we provide a self-contained proof of the genus expansion

$$\frac{1}{N^2}\log \widehat{Z}_N(t) \text{ “ = ” } \sum_{g\geq 0}\frac{1}{N^{2g}}e_g(t),$$

as a formal identity (meaning $d^n/dt^n\,|_{t=0}$ agrees on both sides), where

$$\widehat{Z}_N(t) := \iiint \exp\left\{it\,N\,\mathrm{Tr}\,(ABC+ACB)\right\}d\mu_N(A)d\mu_N(B)d\mu_N(C), \tag{C.0.1}$$

and

$$e_g(t) = \sum_{n\geq 0}\frac{(-1)^n}{(2n)!}\begin{pmatrix}\text{number of colored triangulations}\\ \text{with } 2n \text{ triangles on an orientable}\\ \text{surface of genus } g \text{ up to equivalence}\end{pmatrix}t^{2n}, \tag{C.0.2}$$

as in 2.2.6.

C.1 The partition function

By defining

$$Z_N(t) := \iiint \exp\left\{-N\mathrm{Tr}\left[\frac{1}{2}\left(A^2+B^2+C^2\right)-it(ABC+ACB)\right]\right\}dAdBdC,$$

we can write partition function $\widehat{Z}_N(t)$ defined in (C.0.1) as

$$\widehat{Z}_N(t) = \frac{Z_N(t)}{Z_N(0)},$$

since $Z_N(0) = \left(Z_N^{GUE}\right)^3$ (see 1.1.3). Rescaling all three matrices in both numerator and denominator by $1/\sqrt{N}$ then gives integrals over *unscaled* GUE

$$\widehat{Z}_N(t) = \iiint \exp\left\{i\frac{t}{\sqrt{N}}\mathrm{Tr}\left[ABC+ACB\right]\right\}d\widetilde{\mu}_N(A)\widetilde{\mu}_N(B)\widetilde{\mu}_N(C), \tag{C.1.1}$$

where $\widetilde{\mu}_N$ is GUE measure as defined in 1.1.1.

C.1.2 The partition function as a formal generating function. By formally Taylor expanding the exponential in (C.1.1), and interchanging the series and integral symbols one sees that

$$\widehat{Z}_N(t) \text{``}=\text{''} \sum_{n\geq 0} i^n \frac{N^{-n/2}\langle(\mathrm{Tr}\,[ABC+ACB])^n\rangle}{n!} t^n \tag{C.1.3}$$

where

$$\langle f\rangle := \iiint f\, d\widetilde{\mu}_N(A)d\widetilde{\mu}_N(B)d\widetilde{\mu}_N(C)$$

is the the expectation with respect to the combined GUE measure, where the " = " symbol means that this equality is just formal and ignores all possible issues with convergence or interchanging of summation and integral signs.

C.2 Relation to labeled colored diagrams.

C.2.1. A **tri-valent diagram with n vertices** is diagram with n vertices of valence three as defined in A.2.1. Since all the vertices have the same valence, we simplify the labeling notation to remove mention of the vertex type. More specifically, we will label the n vertices as in figure C.1.

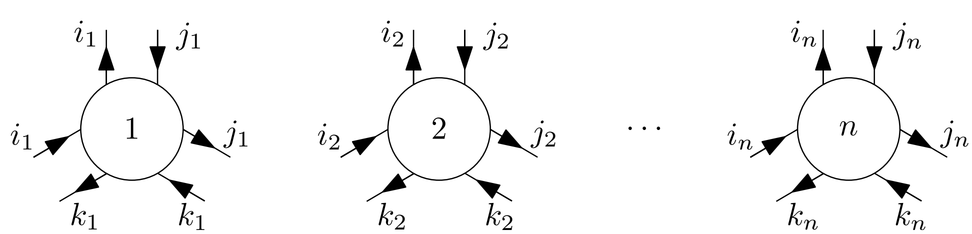

FIGURE C.1. The labeled vertices of a tri-valent diagram.

Note that since all vertices have valence three, the pairings force the number of edges to be even. Figure C.2 shows an example of a tri-valent diagram with four vertices. This diagram has two faces, given by

$$i_1 \to i_1$$

$$j_1 \to i_2 \to j_3 \to i_4 \to i_3 \to j_2 \to \ldots \to k_1 \to j_1,$$

and so defines a map in a surface of genus 1, as described in A.3.5.

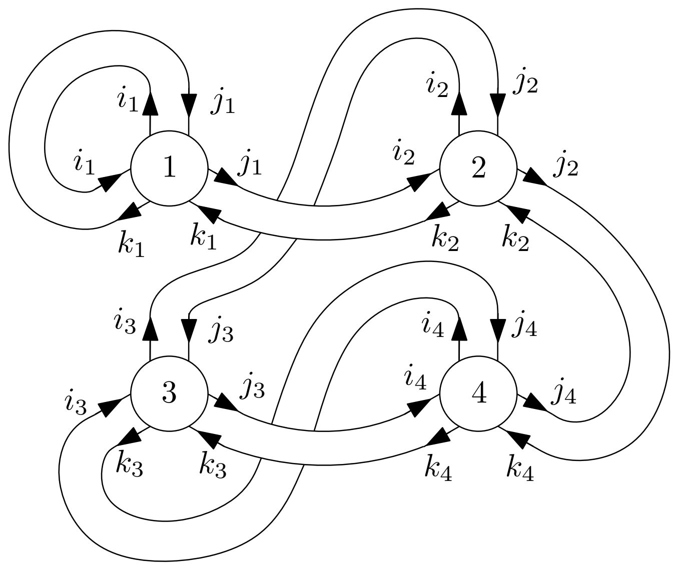

FIGURE C.2. Representation of a tri-valent diagram.

C.2.2. A **colored tri-valent diagram** is a tri-valent diagram together with an assignment of one of the colors $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ to each half-edge in such a way that the following conditions are satisfied:

- Half-edges ij of each vertex are assigned color $\mathcal{A}$.
- Paired half-edges of the diagram share the same color.
- Each vertex has one half-edge of each of the three colors.

We call the collection of vertices together with the coloring of their half-edges the **coloring scheme of the vertices** of the diagram.

C.2.3. We note that given these conditions, each vertex can be colored in only one of two ways (see figure C.3):

- The half-edge jk is colored with $\mathcal{B}$ and half-edge ki is colored with $\mathcal{C}$. We call a vertex colored in this way a **type I vertex.**
- The half-edge jk is colored with $\mathcal{C}$ and half-edge ki is colored with $\mathcal{B}$. We call these type of vertex a **type II vertex.**

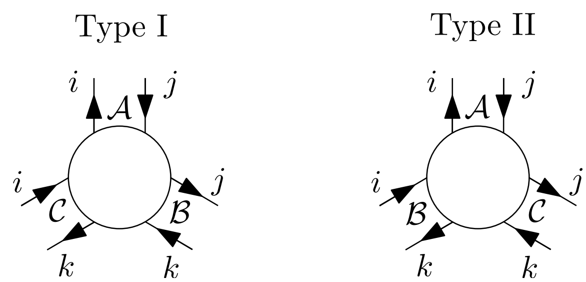

FIGURE C.3. The two types of colored tri-valent vertices.

C.2.4 Example. Figure C.4 shows all the colored tri-valent diagrams with two vertices. The top two give maps of genus zero, while the bottom two give maps of genus one.

C.2.5 Remark on the labels. The very rigid condition regarding half-edges ij always being colored with color $\mathcal{A}$ above is cooked up so that proposition C.2.6 below is true. In the end, this all boils down the fact that the integrals really count couplings via Wick's lemma A.1.1 and to how one expands $(\mathrm{Tr}\,[ABC + ACB])^n$ as a polynomial in the entries of the matrices. The fact that matrix A shows up first in both of the matrix monomials singles matrix A in a very particular way in the proof of C.2.6.

Of course, this is just a matter of convention, and if one prefers one may restate everything singling out a "special" color for the whole diagram and a special "half-edge" for each vertex which plays the role of the ij edge.

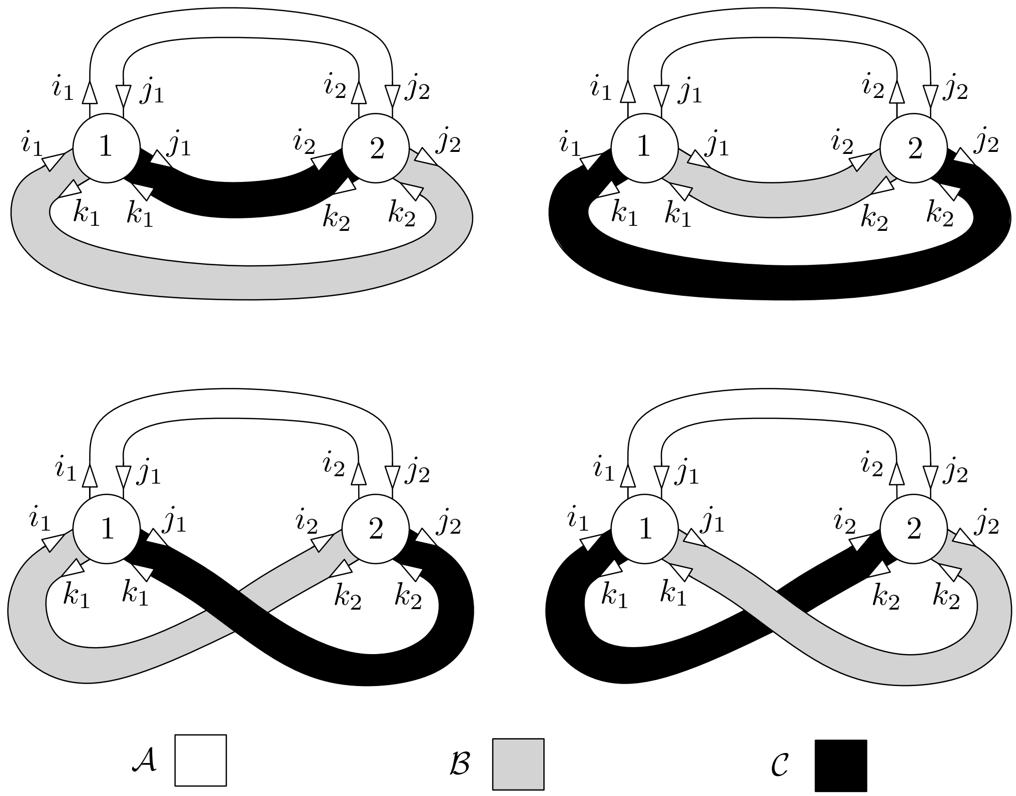

FIGURE C.4. The four colored tri-valent diagrams with two vertices.

C.2.6 Proposition. *If n is odd, then $\langle \mathrm{Tr}\,[ABC + ACB]^n \rangle = 0$. Otherwise, for $n > 0$ we have*

$$\langle \mathrm{Tr}\,[ABC + ACB]^{2n} \rangle = \sum_{F \geq 1} \begin{pmatrix} \text{number of colored} \\ \text{tri-valent diagrams} \\ \text{with } 2n \text{ vertices} \\ \text{and } F \text{ faces} \end{pmatrix} N^F.$$

Proof. By writting

$$\mathrm{Tr}\,[ABC + ACB] = \sum_{i,j,k=1}^{N} a_{ij} b_{jk} c_{ki} + a_{ij} c_{jk} b_{ki},$$

we have

$$(\mathrm{Tr}\,[ABC + ACB])^n = \sum_{\substack{i_1, \ldots, i_n \\ j_1, \ldots, j_n \\ k_1, \ldots, k_n}=1}^{N} \begin{array}{r} (a_{i_1 j_1} b_{j_1 k_1} c_{k_1 i_1} + a_{i_1 j_1} c_{j_1 k_1} b_{k_1 i_1})(a_{i_2 j_2} b_{j_2 k_2} \ldots \\ \ldots (a_{i_n j_n} b_{j_n k_n} c_{k_n i_n} + a_{i_n j_n} c_{j_n k_n} b_{k_n i_n}). \end{array}$$

We can shorten the notation by introducing the muti-index

$$\sigma := (i_1, i_2, \ldots, i_n, j_1, \ldots j_n, k_1, \ldots, k_n)$$

and defining for $p = 1, \ldots, n$

$$\begin{aligned} T_p^{\mathrm{I}}(\sigma) &:= a_{i_p j_p} b_{j_p k_p} c_{k_p i_p}, \\ T_p^{\mathrm{II}}(\sigma) &:= a_{i_p j_p} c_{j_p k_p} b_{k_p i_p}, \end{aligned}$$

to obtain

$$(\mathrm{Tr}\,[ABC + ACB])^n = \sum_\sigma (T_1^{\mathrm{I}}(\sigma) + T_1^{\mathrm{II}}(\sigma))(T_2^{\mathrm{I}}(\sigma) + T_2^{\mathrm{II}}(\sigma)) \ldots (T_n^{\mathrm{I}}(\sigma) + T_n^{\mathrm{II}}(\sigma))$$

where the sum runs over all indices in σ ranging from 1 to N, and the notation T^{I} and T^{II} was chosen to stand for *type* I *vertex* and *type* II *vertex* as defined in (C.2.3), which will be useful below. By expanding the product on the right we obtain

$$(\mathrm{Tr}\,[ABC + ACB])^n = \sum_{q_1, \ldots q_n \in \{\mathrm{I}, \mathrm{II}\}} \sum_\sigma T_1^{q_1}(\sigma) T_2^{q_2}(\sigma) \ldots T_n^{q_n}(\sigma),$$

and if n is odd, then $T_1^{q_1}(\sigma) T_2^{q_2}(\sigma) \ldots T_n^{q_n}(\sigma)$ is a monomial in the a's, b's and c's with an odd number of entries from each matrix, and so $\langle T_1^{q_1}(\sigma) T_2^{q_2}(\sigma) \ldots T_n^{q_n}(\sigma)\rangle = 0$ (by using the substitution $(a, b, c) \mapsto (-a, -b, -c)$ in the integral). This proves the first claim in the statement by the linearity of $\langle \cdot \rangle$.

Regarding even powers, again using linearity we write

$$\left\langle (\mathrm{Tr}\,[ABC + ACB])^{2n} \right\rangle = \sum_{q_1, \ldots q_{2n} \in \{\mathrm{I}, \mathrm{II}\}} \sum_\sigma \langle T_1^{q_1}(\sigma) T_2^{q_2}(\sigma) \ldots T_{2n}^{q_{2n}}(\sigma)\rangle \,,$$

and make the following identifications:

- We identify the term $\langle T_1^{q_1}(\sigma) T_2^{q_2}(\sigma) \ldots T_{2n}^{q_{2n}}(\sigma)\rangle$ with a collection of $2n$ tri-valent vertices as in figure C.1, which have no coloring type specified to them, and where we ignore the actual numerical values of the indices in σ coming from the sum over σ.

- Each vertex is colored according to the coloring scheme determined by the q's, so that vertex 1 is of type q_1, vertex 2 is of type of type q_2, and so on (see C.2.3), where we make the convention to always color the edge ij with color A (just as in the definition of T_p^{I} and T_p^{II}).

In such a way, the outside sum in the above equation corresponds to a *coloring scheme* of the $2n$ tri-valent vertices as defined in (C.2.2), with each coloring scheme appearing only once.

Then, when one applies Wick's lemma to $\langle T_1^{q_1}(\sigma)T_2^{q_2}(\sigma)\dots T_{2n}^{q_{2n}}(\sigma)\rangle$, non-zero contributions in the resulting sum will only come from couplings that pair the a's with the a's, the b's with the b's and the c's with the c's since the matrices are independent, i.e., the couplings that are compatible with the coloring schemes of the vertices. To each of these couplings we can associative a unique *colored tri-valent diagram* as defined above, and the coupling will contribute N^F to the sum where F is the number of faces as in appendix A. This gives the equality in the statement. □

C.2.7. By combining C.2.6 and C.1.3 we obtain the formal identity

$$\widehat{Z}_N(t) \text{``}=\text{''} 1 + \sum_{n\geq 1} \frac{(-1)^n}{(2n)!} \left(\sum_{F\geq 1} \begin{pmatrix} \text{number of colored} \\ \text{tri-valent diagrams} \\ \text{with } 2n \text{ vertices} \\ \text{and } F \text{ faces} \end{pmatrix} N^{F-n} \right) t^{2n}, \tag{C.2.8}$$

where we have separated the first term $1 = \langle \mathrm{Tr}\,[ABC + ACB]^0 \rangle$ from the others to avoid having to specify unnatural counts for diagrams with zero vertices.

C.3 The genus expansion in terms of colored tri-valent maps

C.3.1 The associated colored tri-valent maps. As discussed in appendix A, the connected diagrams have a particularly interesting geometric interpretation. By embedding a connected colored tri-valent diagram in the surface it defines, and then collapsing the edges as described in A.3, one obtains a colored tri-valent map as we defined in 2.2.4. Figure C.5 illustrates how the construction looks like locally, where the bent arrow is specifying the orientation of the surface.

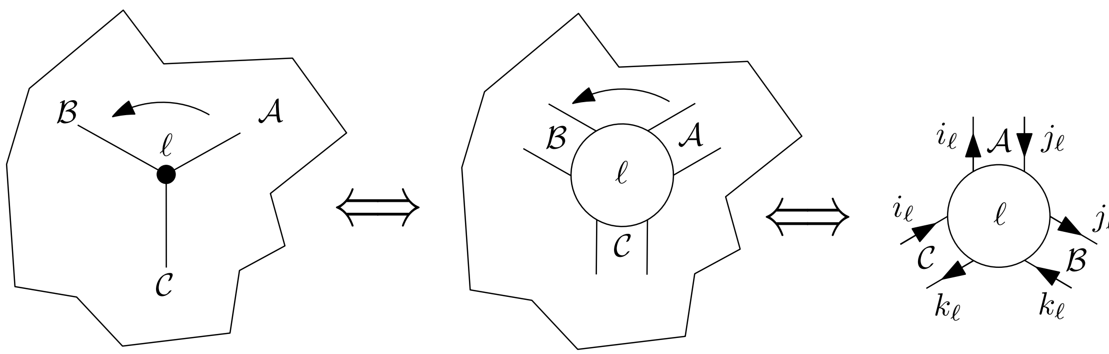

FIGURE C.5. From a colored tri-valent map to a colored tri-valent diagram.

C.3.2 The logarithm counts connected objects. It is well known in the combinatorics literature, and widely used in the applications of matrix integrals to combinatorics, that the logarithm of an exponential generating function for labeled objects is the formal generating function for the connected ones. In other words, one has the formal identity

$$\log\left[\sum_{n=0} \frac{b_n}{n!} t^n\right] \text{ “ = ” } \sum_{n=1} \frac{a_n}{n!} t^n,$$

where the a_n count only the connected objects while the b_n are the counts with any number of connected components. This statement is more like a principle than a general theorem, because what one means by labels, by connected, and how one "counts" can vary drastically from one application to another, and this fact makes it virtually is impossible to write a precise and general statement. In particular, the formal exponential series to which we want to take the logarithm on the right of C.2.8 has as coefficients Laurent polynomials in N with integer coefficients, i.e., our "numbers" b_n are Laurent polynomials.

For the principle to hold, one must work with an exponential generating series (with the $1/n!$ coefficient for the t^n term), and more importantly, the decomposition of whatever one is counting (e.g., into connected components) must be hereditary in the sense that each connected component of the decomposition is also an object of the same type.

C.3.3 Proposition. *One has the formal identity*

$$\log \widehat{Z}_N(t) \text{ “}=\text{” } \sum_{n\geq 1} \frac{(-1)^n}{(2n)!} \left(\sum_{F\geq 1} \left(\begin{array}{c} \text{number of colored} \\ \text{connected tri-valent} \\ \text{diagrams with } 2n \\ \text{vertices and } F \text{ faces} \end{array} \right) N^{F-n} \right) t^{2n},$$

which is obtained from the C.2.8 by taking the logarithm on both sides.

Proof. We just need to see that the logarithm of the formal power series on the left of C.2.8 singles out the connected diagrams. Equivalently, since $\log(1+z)$ and $\exp(z)-1$ are inverses to each other as formal power series (one has to avoid constant terms when composing formal power series, see [2]), we want to prove that if

$$Q_{2n}(N) := \sum_{F\geq 1} \left(\begin{array}{c} \text{number of colored tri-valent diagrams} \\ \text{with } 2n \text{ vertices and } F \text{ faces} \end{array} \right) N^{F-n},$$

and

$$P_{2n}(N) := \sum_{F\geq 1} \left(\begin{array}{c} \text{number of colored tri-valent connected} \\ \text{diagrams with } 2n \text{ vertices and } F \text{ faces} \end{array} \right) N^{F-n},$$

then

$$\exp\left[\sum_{n\geq 1} \frac{(-1)^n}{(2n)!} P_{2n}(N) t^{2n} \right] - 1 \text{ “}=\text{” } \sum_{n\geq 1} \frac{(-1)^n}{(2n)!} Q_{2n}(N) t^{2n} \tag{C.3.4}$$

as formal power series.

Let $p_{(2n,F)}$ be the coefficients of $P_{2n} = P_{2n}(N)$, so that

$$P_{2n} = \sum_{F\geq 0} p_{(2n,F)} N^{F-n}.$$

If we write

$$\begin{aligned} P_{2n}P_{2m} &= \left(\sum_{F_1\geq 1} p_{(2n,F_1)} N^{F_1-n} \right) \left(\sum_{F_2\geq 1} p_{(2m,F_2)} N^{F_2-m} \right) \\ &= \sum_{F\geq 1} \left(\sum_{F_1+F_2=F} p_{(2n,F_1)} p_{(2m,F_2)} \right) N^{F-(n+m)}, \end{aligned}$$

we can interpret the coefficient $\sum_{F_1+F_2=F} p_{(2n,F_1)}p_{(2m,F_2)}$ as the number of labeled colored tri-valent diagrams with F faces and precisely two 2 connected components, one with $2n$ vertices and another with $2m$ vertices, with the extra restriction, say, that we use $1, 2, \dots, 2n$ for the labels of the vertices of the first connected component, and we use $2n+1, 2n+2, \dots, 2(n+m)$ for the labels for the vertices in the second connected component (here we are using the minimal labels for the tri-valent diagrams discussed in C.3.1). We emphasize the fact that $\sum_{F_1+F_2=F} p_{(2n,F_1)}p_{(2m,F_2)}$ is not the number of diagrams with F faces and $2n+2m$ vertices and exactly two components, since we are not allowing the vertices of the components to have arbitrary labels from the set $\{1, \dots, 2n+2m\}$.

Similarly, for all $k \geq 1$ we have

$$P_{2n_1}P_{2n_2}\dots P_{2n_k} = \sum_{F\geq 0}\left(\sum_{F_1+F_2+\dots+F_k=F} p_{(2n_1,F_1)}\dots p_{(2n_k,F_k)}\right) N^{F-\sum n_i}$$

where the coefficient of $N^{F-\sum n_i}$ is the number of labeled diagrams with precisely k components, $2\sum n_i$ vertices, and F faces, with one component with $2n_1$ vertices with labels $1, \dots, 2n_1$, another component with $2n_2$ vertices and labels, $2n_1+1, \dots, 2(n_1+n_2)$, and so on.

To remove the condition on the labels for each component coming from a particular set, and also to remove the implicit ordering we are imposing on the components, we sum over the partitions of the labels for the vertices to get

$$\sum_{\substack{\{1,\dots,2n\}=S_1\sqcup S_2\dots\sqcup S_k \\ \text{with } |S_i|\neq 0 \text{ and even}}} P_{|S_1|}P_{|S_2|}\dots P_{|S_k|} = \sum_{F\geq 0} p^{(k)}_{(2n,F)}N^{F-n}$$

where $p^{(k)}_{(2n,F)}$ is the number of tri-valent diagrams with $2n$ vertices, F faces, and exactly k connected components, and where the sum is taken over all partitions of the set $\{1, \dots, 2n\}$ into k non-empty sets $S_1, \dots, S_k$, each with an even number of elements. The sets S_i specify the labels of the vertices that are used in the corresponding connected component.

Finally, summing over k we obtain the following relation between the Q's and the P's

$$\sum_{\substack{k\geq 1 \\ \{1,\ldots,2n\}=S_1\sqcup S_2\ldots\sqcup S_k \\ \text{with } |S_i|\neq 0 \text{ and even}}} P_{|S_1|}P_{|S_2|}\ldots P_{|S_k|} = Q_{2n}.$$

We now use the *exponential formula*, a formal identity for the exponential of an exponential generating function (see for example [2]), which states that for a_i in any commutative ring one has

$$\exp\left[\sum_{n=1} \frac{a_n}{n!}t^n\right] \text{ ``}=\text{'' } \sum_{n=0} \frac{b_n}{n!}t^n$$

where $b_0 = 1$ and for $n \geq 1$ the b's are given by

$$b_n = \sum_{\substack{k\geq 1 \\ \{1,\ldots,n\}=S_1\sqcup S_2\ldots\sqcup S_k \\ \text{with } S_i\neq\emptyset}} a_{|S_1|}a_{|S_2|}\ldots a_{|S_k|}$$

We apply this with $a_n = 0$ if n is odd, and $a_{2n} = (-1)^n P_{2n}$, which gives the identity C.3.4 and concludes the proof. $\square$

C.3.5. One can expand the first terms in the exponential formula above obtaining

$$\begin{aligned}
Q_2(N) &= P_2(N) \\
Q_4(N) &= P_4(N) + 3P_2(N)^2 \\
Q_6(N) &= P_6(N) + 15P_2(N)^3 + 15P_2(N)P_4(N) \\
&\vdots
\end{aligned}$$

One can also use an alternate and closed formula for the b_n given by

$$b_n = B_n(a_1, a_2, \ldots, a_n)$$

where B_n is the *n-th complete Bell polynomial.*

C.3.6 The genus expansion in terms of colored tri-valent diagrams. If we now use Euler's formula for connected tri-valent diagrams with $2n$ vertices we obtain

$$F - n = 2 - 2g,$$

and so we can write the formal identity for $\log \widehat{Z}_N(t)$ in proposition C.3.3 as

$$\log \widehat{Z}_N(t) \text{``}=\text{''} \sum_{n\geq 1} \frac{(-1)^n}{(2n)!} \left(\sum_{g\geq 0} \begin{pmatrix} \text{number of colored} \\ \text{connected tri-valent} \\ \text{diagrams with } 2n \\ \text{vertices and genus } g \end{pmatrix} N^{2-2g} \right) t^{2n}.$$

Finally, dividing by N^2 and formally interchanging the two summations we obtain the so called **genus expansion**

$$\frac{1}{N^2} \log \widehat{Z}_N(t) \text{``}=\text{''} \sum_{g\geq 0} \frac{1}{N^{2g}} e_g(t),$$

where

$$e_g(t) := \sum_{n\geq 1} \frac{(-1)^n}{(2n)!} \begin{pmatrix} \text{number of colored} \\ \text{connected tri-valent} \\ \text{diagrams with } 2n \\ \text{vertices and genus } g \end{pmatrix} t^{2n},$$

is the exponential generating function counting tri-valent colored diagrams of genus g.

C.4 Colored Triangulations.

C.4.1. By a **triangulation** of a surface we will mean a graph embedded in a surface in such a way that its complement is a union of simply connected sets (discs), and where all these discs have three edges in their boundary. These discs are the **triangles** of the triangulation. For example, figure C.6 shows a triangulation of the sphere with 4 triangles.

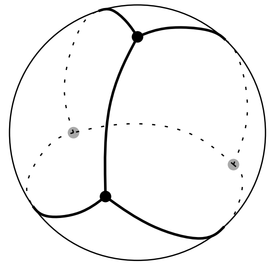

FIGURE C.6. Triangulation of a sphere.

A **colored triangulation with** $2n$ **triangles** on an orientable surface of genus g is a triangulation of the surface with $2n$ labeled triangles $1, 2, \ldots, 2n$ together with:

1. A coloring of each edge of the 1-skeleton with one of the three colors $\mathcal{A}, \mathcal{B}, \mathcal{C}$ in such a way that each triangle has one edge of one of the three colors.

2. A fixed orientation of the surface.

As we have mentioned before, the data of the orientation of the surface is important to gave a well defined connection with maps.

C.4.2 Example. Figure C.7 shows a colored triangulation of a sphere with two triangles. The orientation on the surface is specified by the arrow which determines what clockwise means in triangle 1.

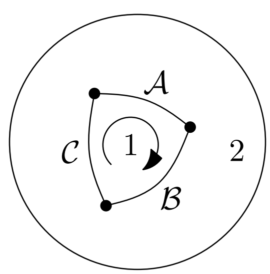

FIGURE C.7. A colored triangulation of a sphere with two triangles.

C.4.3 Type I and type II triangles. Note that the above conditions imply that the edges of each triangle can be colored in only one of two ways:

- The coloring order of the edges is $\mathcal{ABC}$ when one follows them in the clockwise direction (induced from the surface on each triangle). We call these **type I triangles.**

- The coloring order of the edges is $\mathcal{ACB}$ when one follows them in the clockwise direction. We call these **type II triangles.**

C.4.4 Example. On the example C.4.2 triangle 1 is a Type I triangle, and triangle 2 is a Type II triangle. If we reverse the orientationof the surface, then the types of the triangles get reversed too.

C.4.5 Duality. Using the construction of the dual of a diagram from appendix B, one can see that these colored triangulations are precisely the duals of colored trivalent diagrams (or maps) as we defined above. In particular, if we define equivalence of colored triangulations to agree with the equivalence of the dual maps, which in the end corresponds to equality of the diagrams, we obtain the following version of the genus expansion:

C.4.6 The genus expansion in terms of colored triangulations.

$$\frac{1}{N^2}\log \widehat{Z}_N(t) \text{``}=\text{''} \sum_{g\geq 0} \frac{1}{N^{2g}} e_g(t),$$

where

$$e_g(t) := \sum_{n\geq 0} \frac{(-1)^n}{(2n)!} \begin{pmatrix} \text{number of colored} \\ \text{triangulations with} \\ 2n \text{ triangles on a} \\ \text{surface of genus } g \\ \text{up to equivalence} \end{pmatrix} t^{2n},$$

where two triangulations are equivalent if there is an orientation preserving homeomorphism of the surface that takes vertices to vertices and edges to edges that is compatible with the colors and the labels.

References

[1] M. Abramowitz and I.A. Stegun. *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, volume 55. Dover publications, 1965.

[2] M. Aigner. *A Course in Enumeration*, volume 238. Springer, 2007.

[3] G. Anderson, A. Guionnet, and O. Zeitouni. *Introduction to random matrices (Studies in advanced mathematics, N 118).* Cambridge University Press, Cambridge, 2009.

[4] D. Bessis, C. Itzykson, and J.B. Zuber. Quantum field theory techniques in graphical enumeration. *Advances in Applied Mathematics*, 1(2):109–157, 1980.

[5] E. Brézin, C. Itzykson, G. Parisi, and J.B. Zuber. Planar diagrams. *Communications in Mathematical Physics*, 59(1):35–51, 1978.

[6] G.M. Cicuta, L. Molinari, and E. Montaldi. Matrix models and graph colouring. *Physics Letters B*, 306(3):245–251, 1993.

[7] P. Deift. *Orthogonal polynomials and random matrices: a Riemann-Hilbert approach*, volume 3. New York Univ Courant Inst, 1999.

[8] P. Deift, T. Kriecherbauer, and K.D.T-R McLaughlin. New results on the equilibrium measure for logarithmic potentials in the presence of an external field. *Journal of approximation theory*, 95(3):388–475, 1998.

[9] A. Dembo and O. Zeitouni. *Large deviations techniques and applications*, volume 38. Springer, 2009.

[10] P. Di Francesco. 2d quantum gravity, matrix models and graph combinatorics. *Applications of random matrices in physics*, pages 33–88, 2006.

[11] N.M. Ercolani and K.D.T-R. McLaughlin. Asymptotics and integrable structures for biorthogonal polynomials associated to a random two-matrix model. *Physica D: Nonlinear Phenomena*, 152:232–268, 2001.

[12] N.M. Ercolani and K.D.T-R. McLaughlin. Asymptotics of the partition function for random matrices via Riemann-Hilbert techniques and applications to graphical enumeration. *International Mathematics Research Notices*, 2003(14):755–820, 2003.

[13] B. Eynard and C. Kristjansen. An iterative solution of the three-colour problem on a random lattice. *Nuclear Physics B*, 516(3):529–542, 1998.

[14] F. D. Gakhov. *Boundary value problems.* Dover Publications, 1990.

[15] A. Guionnet. *Large Random Matrices: Lectures on Macroscopic Asymptotics: École d'Été de Probabilités de Saint-Flour XXXVI–2006*, volume 36. Springer, 2009.

[16] A. Guionnet and E. Maurel-Segala. Second order asymptotics for matrix models. *The Annals of Probability*, 35(6):2160–2212, 2007.

[17] A. Hatcher. *Algebraic Topology.* Cambridge University Press, 2001.

[18] J. Hoppe. Quantum theory of a massless relativistic surface and a two-dimensional bound state problem. *Soryushiron Kenkyu*, 80(3):145–202, 1989.

[19] K. Johansson. On fluctuations of eigenvalues of random Hermitian matrices. *Duke Mathematical Journal*, 91(1):151–204, 1998.

[20] V.A. Kazakov, I.K. Kostov, and N. Nekrasov. D-particles, matrix integrals and kp hierarchy. *Nuclear Physics B*, 557(3):413–442, 1999.

[21] I.K. Kostov. Exact solution of the three-color problem on a random lattice. *Physics Letters B*, 549(1):245–252, 2002.

[22] S.K. Lando and A.K. Zvonkin. *Graphs on surfaces and their applications*, volume 141. Springer, 2004.

[23] A.I. Markushevich and R.A. Silverman. *Theory of functions of a complex variable.* American Mathematical Society, 2005.

[24] Y. Mizuta. Continuity properties of potentials and beppo-levi-deny functions. *Hiroshima mathematical journal*, 23(1):79–153, 1993.

[25] N. Ninomiya. Etude sur la théorie du potentiel pris par rapport au noyau symétrique. *J. Inst. Polytech. Osaka City Univ*, 8:147–179, 1957.

[26] E.B. Saff and V. Totik. *Logarithmic potentials with external fields*, volume 316. Springer, 1997.

[27] G. 't Hooft. A planar diagram theory for strong interactions. *Nuclear Physics B*, 72(3):461–473, 1974.

[28] W.T. Tutte. A census of planar maps. *Canad. J. Math*, 15(2):249–271, 1963.

[29] A. Zvonkin. Matrix integrals and map enumeration: an accessible introduction. *Mathematical and Computer Modelling*, 26(8):281–304, 1997.