March 2023

Matrix Models of 2D Critical Phenomena

Nathan Hayford
University of South Florida

Matrix Models of 2D Critical Phenomena

by

Nathan Hayford

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Mathematics & Statistics
College of Arts and Sciences
University of South Florida

Co-Major Professor: Razvan Teodorescu, Ph.D.
Co-Major Professor: Seung-Yeop Lee, Ph.D.
David Rabson, Ph.D.
Dmitry Khavinson, Ph.D.
Evguenii Rakhmanov, Ph.D.
Sherwin Kouchekian, Ph.D.

Date of Approval:
March 29^{th}, 2023

Keywords: Random Matrix Theory, Orthogonal Polynomials, Statistical Mechanics,
Riemann-Hilbert, 2D Quantum Gravity, The Ising Model

DEDICATION

This dissertation is dedicated to the entire USF math department. This department has been practically family to me over the past few years, and has always been supportive of my mathematical endeavors. Thanks to all of you for everything, I'll miss you guys.

ACKNOWLEDGMENTS

There are an endless number of people I am indebted to for the completion of this work and my PhD; to properly thank everyone would take a full chapter on its own. I will try to keep this contained to a page or two, but know that I appreciate everything this department has done for me. I'd like to start by both of my advisors, Razvan Teodorescu and Seung-Yeop Lee. Razvan has always been supportive of my mathematical endeavors, and I am indebted to him for any breadth I can claim in my knowledge of mathematical physics. His creativity and approach to mathematics and physics have had a lasting impact on how I think. I am extremely grateful for Seung-Yeop, not only for proposing the project this thesis is based on, but also for his patience and attention to detail throughout the process this work. His unwavering focus is something I have come to wholly appreciate, and I try to emulate this same focus now in all of my work.

I also am grateful for the rest of my defense committee for spending the time and effort reading through this work; I'd like to say a few words about how each of them has personally impacted me. I would not be where I am now without the help of Dmitry Khavinson. I have to thank him for teaching me a good amount of the mathematics that I know, and for always taking the time to read my work. Dima has had unwavering support for my academic career since the beginning, and I have benefited from his guidance and wisdom, in math, and in life. I must thank Evguenii Rakhmanov for spending countless hours discussing mathematics with me. Zhenya has listened to my mathematical thoughts regardless of how far they stray from his area of expertise, and is always sure to make sure I "start from the beginning". I strive to have the same openness to new ideas in my later career, and am always surprised (although by now perhaps I shouldn't be) by how insightful he can be. Additionally, I am grateful for David Rabson, for advising me as an undergraduate, and for spending an enormous amount of time helping me develop as a researcher. Most importantly (at least to me), Dr. Rabson was my "math methods in physics" professor, and I will forever be indebted to him for being the first to show me that mathematical physics could be fun, and a serious career option. Finally, I'd like to thank Sherwin Kouchekian. Sherwin was my first introduction to hard analysis, and his beautiful presentation of and passion for the subject fully convinced me to that analysis truly was a alluring subject.

I must also thank a few more members of the math department who have had a profound influence on me. In particular, I am grateful to have met Catherine Bénéteau, whose thoughtful advice and mentorship has always been appreciated. I am also thankful for Boris Shekhtman, whose perspective on how mathematics should be done has shaped my own perspective. I thank Giacomo Micheli for his friendship and support during my time at USF: although our mathematical careers will probably never intersect, I hope to see him around in the future. Finally, I am indebted to Arthur Danielyan, Natăsa Jonoska, Brendan Nagle, Masahico Saito, and Dmytro Savchuk, for their boundless support and kindness, and for teaching me what it means to be a good mathematician and a good person. Outside of the department, I also have to thank Maurice Duits for letting help finish his and Seung-Yeop's project, for his kind mentorship, and for giving me a home at KTH for the next three years.

It goes without saying that the front office staff has been invaluable in my mathematical career; I truly would not be where I am without them (all forms I have to turn in would have certainly been 5+ years late). I sincerely thank Denise Marks, Anne-Marie Redfield, Matthew Harris, Anthony Inman, and Daniel Moreno for putting up with me and being the true champions of the department during my time here. You guys deserve more credit than you are often given.

I have many friends to thank that have have an influence on not only my mathematical career, but my life in general. I have to thank my good friend and collaborator Fudong Wang, who I grew up as a mathematician with, and has always been willing to discuss ideas with me. Thank you also to Lina Fajardo Gomez, not only for helping me make a number of the figures that appear in this thesis, but also for her advice and friendship. I must also thank Austin Dukes, for always being there for me, and for giving that talk one time in the Graduate Seminar. I am also grateful for Myrto Manolaki, Annamaria Iezzi, Mark Steele, Hunter Kantelis, Devon Conant, Alyssa Hayford, Eion Mulrenin, Boyoon Lee, Haowei (Alice) Chen, Mali Waugh, Marus Maley, Kaylie Dimisa-Nunez, Lauren Maley, and Nic Fontana for their friendship and support throughout my career.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The 2D Ising model has played an important role in the theory of phase transitions, as one of only a handful of exactly solvable models in statistical mechanics. The original model, introduced in the 1920s, has a rich mathematical structure. It thus came as a pleasant surprise when physicists studying matrix models of 2D gravity found that, coupled to quantum gravity, the planar Ising model still had an elegant solution. The methods used by V. Kazakov and his collaborators involved the method of orthogonal polynomials. However, these methods were formal, and no direct analytic derivation of the phase transition has been described in the literature since the original paper of V. Kazakov in 1986. In this work, we present a rigorous proof of Kazakov's results, using steepest descent analysis for biorthogonal polynomials. We are able to calculate the genus 0 partition function, and we also find that the phase transition is described by the string equation of a 3rd order reduction of the KP hierarchy, in agreement with the predictions of G. Moore, M. Douglas, and their collaborators. This is part of a forthcoming paper with Maurice Duits and Seung-Yeop Lee.

CHAPTER 1
INTRODUCTION

Начать с самого начала.

Start from the beginning.

– Evguenii Rakhmanov

1.1 The History of Phase Transitions.

The theory of phase transitions is one of the cornerstones of statistical mechanics. The second order phase transitions, or *continuous* phase transitions, are of particular interest, due in part to their wide range of applications. Examples of such transitions include the spontaneous magnetization of ferromagnetic materials below the Curie temperature, the percolation phase transition in polymer physics and geology, and the superconducting phase transition in condensed matter theory. The structure of these transitions have been intensively studied for almost a century, starting with Paul Eherenfest's classification scheme in the 30's [41]. Soon after, Lev Landau developed his phenomenological theory of continuous phase transitions [76], which characterizes the transition in terms of its critical exponents. The next major breakthrough in the theory of general continuous phase transitions came in the 60's, when Leo Kadanoff applied quantum field theoretic techniques (the renormalization group, or RG) to describe the Ising phase transition [65]; these ideas were developed further by Kenneth Wilson [105, 106]. The renormalization group approach to phase transitions describes the transition as a non-trivial fixed point of RG. One of the features of this approach was the ability to explain the phenomenon of *universality*: seemingly very different physical systems end up having identical critical exponents. Fixed points of the RG flow exhibit scale invariance; furthermore, many of the statistical systems of interest also enjoy translation invariance. These observations led the physicists A.A. Belavin, A.M. Polyakov, and A.B. Zamolodchikov (1984) to postulate that there should be manifest *conformal invariance* at these critical points [8]. In two dimensions, the group of local conformal transformations has a rich structure not present in other dimensions, which allowed for the *exact* solution of the field theories introduced in [8], and thus the exact prediction of the critical exponents of a whole series of universality classes. These are the celebrated *minimal models* of conformal field theory. The minimal models program was quite successful in its characterization of 2-dimensional continuous phase transitions, but left many open questions.

1.2 The History of the 2-Dimensional Ising Model.

The history of phase transitions is intimately linked with the history of the Ising model. Introduced by Ernst Ising at the suggestion of his doctoral advisor Wilhelm Lenz [60] in 1925, it was meant as a simplistic model of a ferromagnet. Ising himself only studied the 1-dimensional version of the model, which does not exhibit the spontaneous magnetization phase transition, and he incorrectly conjectured that the model does not admit a phase transition in any dimension. It was subsequently noted by Rudolph Peierls that higher-dimensional versions of the model likely *did* exhibit a phase transition [87] (the spontaneous magnetization transition). It took over 20 years for an exact solution to the 2D version of the model (the Ising model on $\mathbb{Z}^2$) to be announced by Lars Onsager [86] (1944), with credit due in part to his postdoc Bruria Kaufmann. However, Onsager's solution was not completely mathematically sound, and it took an additional decade for a fully rigorous solution to be provided by Chen-Ning Yang [108] in 1952. Since, the model has been celebrated as one of the most important models of a phase transition, and the model's surprisingly rich mathematical structure has been picked apart. The universality of the critical exponents appearing in the 2-dimensional model were verified for various lattices (cf. R.J. Baxter's book [7], Chapter 11, and references therein), but these considerations were markedly limited by the fact that computations could be performed explicitly only for choices of fairly regular lattices (e.g. square lattice, triangular lattice, etc.). The breakthrough of Leo Kadanoff [65] (1966) in applying renormalization group techniques to the Ising model shed further light on the phenomenon of universality in the Ising model. The advent of conformal field theory (CFT) techniques to describe 2-dimensional critical phenomena by A.A. Belavin, A.M. Polyakov, and A.B. Zamolodchikov [8] in 1984 continued the interest in universality in the 2D Ising model, and prompted Vladimir Kazakov in 1986 [70] to consider the Ising model coupled to 2-dimensional gravity; that is, the Ising model on a random lattice.

1.3 The History of 2-Dimensional Quantum Gravity.

A working theory of quantized gravity is one of the most fundamental open problems in theoretical physics, and has eluded physicists for the past several decades. One place where progress was successfully made was in the theory of 2-dimensional quantum gravity, as degrees of freedom present in the higher dimensional theories are absent in the 2-dimensional one. Part of the hope of the program was that that the exact solution to a theory of 2D gravity would shed light on how higher-dimensional versions might work; another major motivating factor was the first superstring revolution, in which the techniques of 2D gravity played an integral role. One of the first major works on the subject was Alexander Polyakov's 1981 paper on bosonic string theory [91], which demonstrated that the so-called Liouville (also called continuum) approach to 2D quantum gravity could be exactly solved. The drawback of the theory was that performing exact calculations proved difficult, and hindered the theory for the next decade or so. Indeed, the seminal paper of A.A. Belavin, A.M. Polyakov, and A.B. Zamolodchikov [8] in 1984 was described by Polyakov as "an unsuccessful attempt to solve the Liouville theory" [89]. A major development came in the works of Douglas and Shenker [35], Brézin and Kazakov [20], and Gross and Migdal [58, 59]. The ideas implicit in these works was to replace the integral over geometries in the continuum theory with a sum over discretized surfaces; in this way, many calculations that were previously inaccessible were now in sight. This technique was made possible by the advances made by the Saclay school of theoretical physics (Itzykson, Zuber, Parisi, Brézin, Kazakov, to name a few) in random matrix theory. It was realized in [19, 62] that random matrices could be used to perform a sum over discretized surfaces. Subsequent works by Vladimir Kazakov and collaborators [17, 70] showed that it might be possible to exactly solve a coupled theory of the minimal models of conformal field theory and gravity, and calculated the first result in this direction: the shift of the critical exponents of the $(4, 3)$ minimal model (corresponding to the critical 2D Ising model) when coupled to gravity. This work was later complemented by a corresponding continuum result, due to Kniznik, Polyakov, and Zamolodchikov [72]: their result became known as the *KPZ formula*. These results are what motivated the Douglas-Shenker, Brézin-Kazakov, and Gross-Migdal program. Their goal in part was to construct a unified theory of all $2D$ critical phenomena, with renormalization group flows between all of the critical points. As it turned out, the 2-matrix model was precisely the setting required to achieve this; the theory ended up being completely integrable, and the renormalization group flows between critical points were given by flows of the KP hierarchy.

1.4 The Subject of This Thesis.

From the standpoint of theoretical physics, the results described above describe one of the most outstanding achievements of the last few decades: an exact model of all 2D critical phenomena. However, from the mathematical standpoint, these works left much to the imagination. Many of the techniques used in the works of the 1980s and 1990s on random matrices related to 2D critical phenomena were non-rigorous, from the use of non-convergent matrix integrals to the ill-defined renormalization group flow. It thus became the task of a group of mathematical physicists to try and put these works on solid mathematical footing. The first few works in this direction were due to A. Fokas, A. Its, and A. Kitaev [47, 48], who were able to make a step in the right direction by providing a description of the first nontrivial critical point of the 1-matrix model (corresponding to a "pure" theory of gravity) in terms of orthogonal polynomials and Painlevé type equations. The next decade saw the birth of the method of steepest descent of Deift and Zhou [31], which opened the doors to the rigorous analysis of the more general families of orthogonal polynomials appearing in the 1-matrix model. These works built on pre-existing works of a number of mathematicians working in approximation theory (see [78, 93, 94, 102], among many others, as well as [13, 29, 30] for applications of the method of steepest descent to orthogonal polynomials), and finally culminated in the rigorous description of the critical point investigated by Fokas, Its and Kitaev by Maurice Duits and Arno Kuijlaars in [38] (2006), and by Pavel Bleher and Alfredo Deaño in [10]. Nowadays, the critical phenomenon in the 1-matrix model are widely accepted to be well-understood. However, the same cannot be said about the 2-matrix model, which has a much richer mathematical structure. Part of the reason for this deficit in knowledge was a lack of tools: the 1-matrix model could be studied using techniques of orthogonal polynomials, whereas the 2-matrix model was characterized by the more mysterious biorthogonal polynomials. The steepest descent analysis available for orthogonal polynomials did not apply directly to biorthogonal polynomials, which meant a fully rigorous analysis of their asymptotics (and thus a description of the Ising critical point, which only appears in the 2-matrix model) remained out of reach. This changed with a publication of Arno Kuijlaars and Ken McLaughlin [75] (2005), which characterized biorthogonal polynomials in terms of *multiple orthogonal polynomials*. Since there was already an extensive literature in approximation theory for the analysis of multiple orthogonal polynomials [1, 4, 85], including a Riemann-Hilbert formulation [5] amenable to steepest descent analysis, this allowed finally for the possibility of analysis of critical phenomenon in the 2- matrix model. Kuijlaars, Duits, and their collaborators began a program of analyzing critical phenomena in the 2-matrix model [36, 37, 39, 40], which took place in the early 2010's. However, the problem of providing a rigorous analysis of the Ising phase transition coupled to gravity remained open.

Finally, we arrive at the present day, and the subject of this thesis. In this thesis, we present two main results: first, we calculate rigorously the genus zero partition function of the 2D Ising model on a random 4-regular graph. We also for the first time give a rigorous description of the Ising critical point when coupled to gravity; this amounts to expressing the partition function in terms of a special solution of a higher-order Painlevé type equation (the "string equation" for the KdV-3 hierarchy). These results are based on steepest descent analysis for biorthogonal polynomials. This is the first rigorous work on the analysis of physically relevant critical points of the 2-matrix model, and we hope that it is the beginning of a program that will shed more light on universality in 2D critical phenomena.

1.5 Organization of the Thesis and Notations.

The rest of this thesis is organized as follows. In Chapter 2, we define the Ising model on a general graph, and discuss some of the general properties that Ising-type models enjoy. We then provide an exact solution to the 2D Ising model on $\mathbb{Z}^2$. In Chapter 3, we introduce random matrices, and define the 1 (respectively, 2)-matrix models. We show how these models relate to orthogonal (respectively, biorthogonal) polynomials, and show how biorthogonality may be reduced to multiple orthogonality. Finally, we discuss the combinatorial aspects of both matrix models, and discuss the connections to quantum gravity in 2 dimensions. In Chapter 4, we discuss some of the general techniques involved in steepest descent analysis for Riemann-Hilbert problems. We then demonstrate in the case of orthogonal polynomials how the machinery works. In this chapter, we also introduce the Riemann-Hilbert problem for multiple orthogonal polynomials, and show how it applies to the special case of biorthogonality. In Chapter 5, we prove the main results of the thesis: an explicit formula for the genus zero partition function of the quartic 2-matrix model, as well as a rigorous analysis of the multicritical point originally studied by Kazakov [70]. Finally, we conclude in Chapter 6 with some calculations related to the spectral curve of the cubic 2-matrix model. We also present some further open problems of interest, which will be the subject of future work.

We will frequently use a number of notations without comment; we list these here.

- We use the following asymptotic notations:
  - $f(x) = \mathcal{O}(g(x))$ as $x \to x_0$ if $\left|\frac{f(x)}{g(x)}\right| \leq M$ for $|x - x_0|$ sufficiently small (if $x_0 = \infty$, then this is interpreted as $|x|$ sufficiently large).
  - $f(x) = o(g(x))$ as $x \to x_0$ if $\left|\frac{f(x)}{g(x)}\right| \to 0$ as $x \to x_0$.
  - $f(x) \sim g(x)$ as $x \to x_0$ if $\frac{f(x)}{g(x)} \to 1$ as $x \to x_0$.
- Throughout, $\omega := e^{\frac{2\pi i}{3}} = -\frac{1}{2} + i\frac{\sqrt{3}}{2}$ is the principal third root of unity,
- We denote the $n \times n$ matrix with a 1 in the $(i,j)^{th}$ entry and zeros elsewhere by E_{ij},
- We denote the diagonal matrix D with entries $d_1, d_2, ..., d_n$ by $D = \text{diag}\ (d_1, d_2, ..., d_n)$. In other words, the below have the same meaning:

$$\text{diag}\ (d_1, d_2, ..., d_n) = \begin{pmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_n \end{pmatrix}.$$

- The third Pauli matrix $\sigma_3 := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. We will often write expressions such as $z^{C\sigma_3}$, or $e^{f(z)\sigma_3}$. These expressions are defined to be

$$z^{C\sigma_3} := \begin{pmatrix} z^C & 0 \\ 0 & z^{-C} \end{pmatrix}, \qquad e^{f(z)\sigma_3} := \begin{pmatrix} e^{f(z)} & 0 \\ 0 & e^{-f(z)} \end{pmatrix}.$$

- The matrix $\hat{\sigma}_{ij}$ is defined to be the 4×4 matrix which permutes the i^{th} and j^{th} row/column. For example, the matrix $\hat{\sigma}_{24}$ would be

$$\hat{\sigma}_{24} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

The matrix $\hat{\sigma}_{ij}$ permutes the i^{th} and j^{th} row and column of a given matrix A by conjugation; again using our example of $\hat{\sigma}_{24}$,

$$\hat{\sigma}_{24} A \hat{\sigma}_{24} = \hat{\sigma}_{24} \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix} \hat{\sigma}_{24} = \begin{pmatrix} a_{11} & a_{14} & a_{13} & a_{12} \\ a_{41} & a_{44} & a_{43} & a_{42} \\ a_{31} & a_{34} & a_{33} & a_{32} \\ a_{21} & a_{24} & a_{23} & a_{22} \end{pmatrix}.$$

- For readability purposes, blocks of zeros in matrices will be denoted simply by zero, where there is no cause for ambiguity. For example, the if A is a 3×3 matrix, then the expressions

$$\begin{pmatrix} 1 & 0 \; 0 \; 0 \\ 0 & \\ 0 & A \\ 0 & \end{pmatrix} = \begin{pmatrix} 1 & 0_{3\times 1} \\ 0_{1\times 3} & A \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & A \end{pmatrix}$$

all have identical meaning.

- If A is an $n \times n$ matrix and B is an $m \times m$ matrix, we define the matrix $A \oplus B$ to be the block diagonal matrix

$$A \oplus B := \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}.$$

- If $X(z)$ is the solution to a Riemann Hilbert problem defined on the contour γ, we write $J_X(z) : \gamma \to \mathbb{C}$ as its jump matrix: i.e.,

$$X_+(z) = X_-(z) J_X(z), \qquad z \in \gamma.$$

We will also sometimes write γ_X or Γ_X to denote the contour γ corresponding to the Riemann-Hilbert problem for $X(z)$.

CHAPTER 2
THE ISING MODEL

In this section, we discuss probably the first lattice model to be intensively studied: the Ising model. Historically, the model was suggested to E. Ising by his advisor, W. Lenz, as a simplistic model of ferromagnetism. As part of his thesis, Ising solved the 1-dimensional version of the model [60] (what we will call "the Ising model on $\mathbb{Z}$"), and incorrectly predicted that the model did not exhibit a finite-temperature phase transition. However, it was subsequently noted that higher-dimensional versions of the model likely *did* exhibit a phase transition [87] (the spontaneous magnetization transition). It was not until over 20 years later that an exact solution to the 2-D version of the model (the Ising model on $\mathbb{Z}^2$) was announced by L. Onsager [86], with credit due in part to his postdoc B. Kaufmann. However, Onsager's solution was not completely mathematically sound, and it took an additional decade for a fully rigorous solution to be provided by C. Yang [108]. Since, the model has been celebrated as one of the most important models of a phase transition, and the model's surprisingly rich mathematical structure has been picked apart.

In what follows, we will first define the Ising model on generic graphs, discuss the Kramers-Wannier duality, and then study the solution of the 2-dimensional Ising model first introduced by M. Kac and J.C. Ward [92]. We indicate the universal nature of the phase transition. Essentially all of what is stated in this chapter is now part of the classical literature on statistical mechanics. For further details, one should consult [7], for example.

2.1 Definition of the Model and Basic Properties.

Most of the graph-theoretic notions we introduce are fairly standard; we refer to [15] for further details on the relevant graph theory topics.

Definition 2.1. Let $G = (V, E)$ be a finite graph[1], and consider the collection of maps

$$\Sigma := \{\sigma : V \to \{\pm 1\}\}. \tag{2.1}$$

[1]Our notion of graph is actually a *multigraph*; in other words, we allow multiple edges between vertices and loops (an edge between a vertex and itself) in our definition. We will continue to refer to these as simply graphs, for ease of exposition.

We define a probability measure $\mathbb{P}$ on Σ (with the power set 2^{Σ} as a σ-algebra), dependent on two parameters, $\beta > 0$, the *inverse temperature*, and $h \in \mathbb{R}$, the *external field strength*. We call this measure the *Boltzmann distribution*. The measure is defined to be

$$\mathbb{P}(\sigma; \beta, h) := \frac{1}{Z} e^{-\beta H(\sigma)}, \tag{2.2}$$

where

$$H(\sigma; h) = -\sum_{(x,y)\in E} \sigma(x)\sigma(y) - h \sum_{x \in V} \sigma(x) \tag{2.3}$$

is the *Ising Hamiltonian*, and

$$Z = Z(\beta, h) := \sum_{\sigma \in \Sigma} e^{-\beta H(\sigma)} \tag{2.4}$$

is a normalization constant, called the *partition function*. We will sometimes simply write $\mathbb{P}(\sigma)$, $H(\sigma)$ when the dependence on the other variables is clear. We will also sometimes write $Z = Z_G$, to denote the dependence of the partition function on the particular graph we are working on.

We refer to the elements $\sigma \in \Sigma$ as *configurations*, and the values of $\sigma(x)$ as *spins*. We define expected values of quantities like $X(\sigma) : \Sigma \to \mathbb{R}$ as

$$\langle X \rangle := \frac{1}{Z} \sum_{\sigma \in \Sigma} X(\sigma) e^{-\beta H(\sigma)} \tag{2.5}$$

For most of our considerations, we will set $h = 0$; unless otherwise stated, we will assume from here on that $h = 0$. Let us make a few elementary remarks about this model. First, we see that the model favors configurations where all of the spins are aligned, i.e., configurations in which all the values of $\sigma(x)$ are the same are more probable. This is because such configurations have a smaller (more negative) energy, and thus their probability is larger, according to the definition (2.2). Another important property of the model is its $\mathbb{Z}_2$ symmetry: if $h = 0$, then

$$\langle \sigma \rangle = 0. \tag{2.6}$$

This follows from the fact that, for every configuration $\sigma \in \Sigma$, there exists a configuration σ' such that $\sigma' = -\sigma$; furthermore, $H(\sigma) = H(-\sigma)$. So, the collective contribution of these two configurations in the sum (2.5) is

$$\sigma e^{-\beta H(\sigma)} + \sigma' e^{-\beta H(\sigma')} = \sigma e^{-\beta H(\sigma)} - \sigma e^{-\beta H(\sigma)} = 0.$$

Since we can pair up the entire set of configurations Σ in the same manner, we find that $\langle \sigma \rangle = 0$.

The main object of interest in what follows will be the *free energy*, which is defined as

$$F(\beta) = -\frac{1}{\beta}\log Z(\beta). \tag{2.7}$$

If the graph $G = (V, E)$ is infinite in size, i.e., $|V| = +\infty$, then the partition function $Z(\beta)$ may diverge, and thus $F(\beta)$ is ill-defined as well. In such a situation, we will be interested in the *free energy per unit site*:

$$f(\beta) := -\lim_{|V|\to\infty} \frac{1}{\beta|V|}\log Z_{G'}(\beta), \tag{2.8}$$

where the limit is taken over an appropriate finite-size cutoff of the graph of interest. For example, one situation we shall meet is $G = \mathbb{Z}^n$, where the points of $\mathbb{Z}^n$ are treated as vertices, and two elements of $\mathbb{Z}^n$ are connected by an edge if their coordinates differ by ± 1 in only one place. A common regularization procedure is to take the sequence of graphs $G_N := \mathbb{Z}^n \cap ([-N, N]^n)$, the cutoff of $\mathbb{Z}^n$ inside the cube of size $2N + 1$. Then, one defines the free energy per unit site as

$$f_{\mathbb{Z}^n}(\beta) := -\lim_{N\to\infty} \frac{1}{\beta|G_N|}\log Z_{G_N}(\beta) = -\lim_{N\to\infty} \frac{1}{\beta(2N+1)^n}\log Z_{G_N}(\beta). \tag{2.9}$$

Let us briefly explain why the free energy is of such fundamental importance. We begin by re-expressing the free energy as

$$F(\beta) = U - \frac{1}{\beta}S, \tag{2.10}$$

where U is the average internal energy of the system, defined to be $U := \langle H \rangle$, and S is the *entropy* of the system, defined as

$$S := -\sum_{\sigma\in\Sigma} \mathbb{P}(\sigma)\log \mathbb{P}(\sigma) = -\langle \log \mathbb{P} \rangle. \tag{2.11}$$

We state this as a proposition:

Proposition 2.2. *The two expressions* (2.7), (2.10) *are equivalent.*

Proof. Let us start with the second expression, (2.10). We have that

$$\begin{aligned} F(\beta) = U - \frac{1}{\beta}S &= \langle H(\sigma) + \frac{1}{\beta}\log \mathbb{P}(\sigma)\rangle \\ &= \sum_{\sigma}\left[H(\sigma) + \frac{1}{\beta}\log\left(\frac{e^{-\beta H(\sigma)}}{Z}\right)\right]\frac{e^{-\beta H(\sigma)}}{Z} \\ &= \sum_{\sigma}\left[H(\sigma) - H(\sigma) - \frac{1}{\beta}\log Z\right]\frac{e^{-\beta H(\sigma)}}{Z} \\ &= \sum_{\sigma}\left(-\frac{1}{\beta}\log Z\right)\frac{e^{-\beta H(\sigma)}}{Z} = -\frac{1}{\beta}\log Z. \end{aligned}$$

□

The reason we consider the free energy is this: when the system is in thermal equilibrium with a heat bath at temperature $T = \beta^{-1}$, the free energy is minimized. In other words, it is *not* the energy of the system itself that is minimized, but rather the energy of the system *available to do work*: some of the energy is tangled up in the entropy, and is inaccessible to the system.

We make this idea a little more precise as follows. Let $\mathbb{Q}(\sigma)$ be an *arbitrary* probability distribution on the space of configurations Σ (note that, since $|\Sigma| < \infty$, the space of all such measures is a finite-dimensional submanifold of $\mathbb{R}^{|\Sigma|}$), and denote the space of all such distributions as $\mathcal{P}(\Sigma)$. For such distributions $\mathbb{Q} \in \mathcal{P}(\Sigma)$, we can define a notion of free energy of $\mathbb{Q}$ as

$$F(\beta;\mathbb{Q}) := \langle H\rangle_{\mathbb{Q}} - \frac{1}{\beta}S_{\mathbb{Q}}, \tag{2.12}$$

where $\langle H\rangle_{\mathbb{Q}}$ denotes the mathematical expectation of $H(\sigma)$ with respect to the measure $\mathbb{Q}$, and $S_{\mathbb{Q}}$ is the entropy of the distribution $\mathbb{Q}$, that is,

$$S_{\mathbb{Q}} = -\sum_{\sigma\in\Sigma}\mathbb{Q}(\sigma)\log\mathbb{Q}(\sigma). \tag{2.13}$$

Then, we have the following proposition:

Proposition 2.3. *For any fixed $\beta > 0$,*

$$F(\beta) = F(\beta;\mathbb{P}) = \min_{\mathbb{Q}\in\mathcal{P}(\Sigma)} F(\beta;\mathbb{Q}), \tag{2.14}$$

where $\mathbb{P}(\sigma) = \mathbb{P}(\sigma,\beta)$ is the measure defined in (2.2). In other words, among all probability distributions on Σ, the one which minimizes the free energy is the Boltzmann distribution (2.2).

Proof. Let $\mathbb{Q}^*$ be an extremum of the functional (2.12); we will show that Q^* is the Boltzmann distribution (2.2). Let $t > 0$ be sufficiently small, and let $\delta\mathbb{Q}$ be a *signed* measure on Σ, such that the measure $\mathbb{Q}_t(\sigma) := \mathbb{Q}^* + t\delta\mathbb{Q}(\sigma)$ belongs to the class $\mathcal{P}(\Sigma)$. Since $\mathbb{Q}^*$ was extremal, we must have that the first variation $\frac{dF(\beta;\mathbb{Q}_t)}{dt}\Big|_{t=0}$ must vanish on the space $\mathcal{P}(\Sigma)$. We have that:

$$F(\beta;\mathbb{Q}_t) - F(\beta;\mathbb{Q}^*) = t\sum_{\sigma}\left[H(\sigma) + \frac{1}{\beta}(\log\mathbb{Q}^*(\sigma) + 1)\right]\delta\mathbb{Q}(\sigma) + \mathcal{O}(t^2).$$

Since $Q^*(\sigma)$ is extremal, we must have that the first variation vanishes identically. Considered along with a Lagrange multiplier which fixes $\mathbb{Q}_t$ as a probability measure, we find that, for each $\sigma \in \Sigma$,

$$H(\sigma) + \frac{1}{\beta}(\log\mathbb{Q}^*(\sigma) + 1) = \lambda,$$

for some constant λ. Solving for $\mathbb{Q}^*(\sigma)$, we obtain that

$$\mathbb{Q}^*(\sigma) = \textit{const.} \times e^{-\beta H(\sigma)}. \tag{2.15}$$

The requirement that $\mathbb{Q}^*$ be a probability measure uniquely determines the constant of proportionality to be $Z(\beta)$; thus, $\mathbb{Q}^*$ is the Boltzmann distribution. □

This proposition justifies why the free energy is fundamental to equilibrium thermodynamics. There are also several more practical reasons the free energy is important to us. Many other thermodynamic quantities of interest, such as the average energy $U = \langle H \rangle$, and the average magnetization (at non-zero h, of course) $M(\beta) := \langle\sigma\rangle$, can be computed as derived quantities from the free energy. In the two examples we have given, it is easy to check from Equation (2.7) that (and here we assume $h \neq 0$)

$$U = \langle H \rangle = \frac{\partial}{\partial\beta}\left[\beta F(\beta)\right], \tag{2.16}$$

$$M = \langle\sigma\rangle = -\frac{\partial}{\partial h}\left[F(\beta)\right]. \tag{2.17}$$

Let us now define what we mean by *phase transition.*

Definition 2.4. Let X be a thermodynamic system (such as the above), with free energy $F(\beta, x_1, ..., x_N)$, dependent on temperature, and possibly some other macroscopic variables $x_1, ..., x_N$. X is said to have an n^{th} order phase transition if one of its n^{th} order derivatives with respect to $\beta, x_1, ..., x_N$ has a discontinuity for some fixed $0 < \beta < \infty$.

If the system in question is infinite in size, the above definition is modified to a discontinuity in the *free energy per unit site*. In our situation, the Ising free energy on a fixed graph depends in general on two parameters, the inverse temperature β and external field strength h. There is a very simple argument, which dates back to Peierls, that demonstrates that the Ising model on a finite graph *never* exhibits a phase transition:

Proposition 2.5. *Let $G = (V, E)$ be a finite graph, and $F(\beta, h)$ the free energy of the Ising model in external field h on G. Then, the Ising model on this graph does not exhibit a phase transition.*

Proof. Note that this is equivalent to showing that the *partition function* is an analytic function of β, h. Indeed, each Boltzmann factor $e^{\beta H}$ is an analytic function of β, h, since H is a linear function of these parameters. Since G is finite, the sum over Boltzmann factors is finite as well, and so the partition function is a finite sum of analytic functions. It follows that Z is analytic. □

Thus, if we are interested in describing phase transitions, we need to look to infinite-size graphs. What we shall see in the following sections is that the Ising model on $\mathbb{Z}^2$ (and, more generally, on the triangular and hexagonal lattices) exhibits a second-order phase transition in the magnetization parameter h at finite temperature $0 < \beta_c < \infty$.

2.2 High and Low Temperature Expansions: Kramers-Wannier Duality

We shall see that there is a duality between the high and low temperature expansions of the Ising model; these expansions will in fact allow us to predict the phase transition for the square-lattice Ising model. Formally speaking, these expansions are essentially obtained by "integrating out" the spin degrees of freedom in two different ways, to obtain an expression which depends only on the properties of the underlying graph. These results will show that the high temperature expansion of $Z_G(\beta)$ for the graph G is the same (up to a multiplicative factor) as the low-temperature expansion of $Z_{G^*}(\beta)$, where G^* is the *dual graph* to G. This relation is known as *Kramers-Wannier duality*, named after its discoverers Hendrik Kramers and Gregory Wannier [73].

Let us begin by searching for a high temperature ($\beta \to 0$) expansion of the partition function. For infinite graphs, this expansion should be interpreted as an asymptotic one; however, for finite graphs, we are lucky enough to be able to derive an exact expression in a small parameter $t := \tanh \beta$ for the partition function. First, we need a few graph-theoretic definitions.

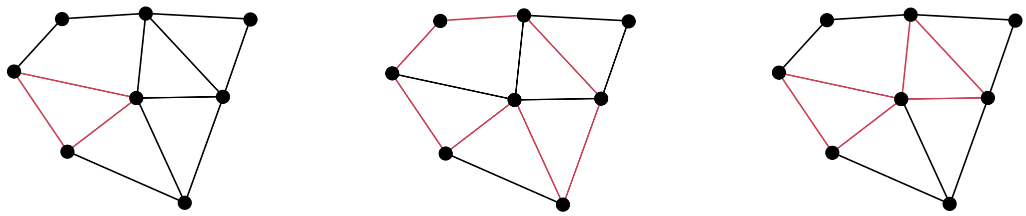


Figure 2.1. Three examples of closed curves on a given graph G, shown in red. The leftmost closed curve contributes a factor of t^3 to the partition function, and the central closed curve contributes a factor of t^7. Finally, the rightmost closed curve contributes a factor of t^6.

Definition 2.6. Let $G = (V, E)$ be a finite graph. A subcollection of edges $A \subset E$ in which every vertex in the graph (V, A) has even degree is called a *closed curve*[2] in G. The set of closed curves is written as $C(G)$. If $c \in C(G)$, then $|c|$ denotes the cardinality of the set of edges corresponding to c.

Such subgraphs are sometimes called *even subgraphs* in graph theory; however, we prefer the terminology closed curve, as it is more intuition-friendly. Some examples of closed curves on a graph G are shown in Figure (2.1); the naming "closed curve" is apparent from here.

Theorem 2.7. *Let $G = (V, E)$ be a finite graph, and $Z(\beta)$ be the partition function for the Ising model on G. Set $t = \tanh\beta$. Then*

$$Z(\beta) = 2^{|V|} [\cosh\beta]^{|E|} \sum_{c \in C(G)} t^{|c|} = 2^{|V|} [\cosh\beta]^{|E|} \sum_{k=0}^{|E|} c_k t^k, \tag{2.18}$$

where c_k is the number of closed curves of size k in G (the number of closed curves of size 0 *is defined to be* 1*).*

The above is called the *high temperature expansion* of the partition function, since β small is equivalent to t small, which follows from the asymptotic $x \sim \tanh x$.

Proof. Note that, since a product of spins can only take on one of two values, $\sigma(x)\sigma(y) = \pm 1$, we can rewrite the expression

$$e^{\beta\sigma(x)\sigma(y)} = \cosh\beta + \sinh\beta\sigma(x)\sigma(y) = \cosh\beta\,[1 + t\sigma(x)\sigma(y)]\,,$$

[2]Colloquially, such subsets are often referred to as *loops* in the physics literature. However, this notation is rather vague, as there are many other things that are often referred to as loops in physics; even within graph theory, there is already another object going by the name "loop". For this reason, we avoid this terminology, and prefer to use *closed curve*.

where we have introduced a new parameter $t = \tanh \beta$. Then, the partition function can be rewritten as

$$\begin{aligned} Z &= \sum_{\sigma\in\Sigma} e^{\beta \sum_{(x,y)\in E} \sigma(x)\sigma(y)} \\ &= \sum_{\sigma\in\Sigma} \prod_{(x,y)\in E} e^{\beta\sigma(x)\sigma(y)} \\ &= [\cosh \beta]^{|E|} \sum_{\sigma\in\Sigma} \prod_{(x,y)\in E} [1 + t\sigma(x)\sigma(y)] . \end{aligned}$$

We may further expand the product inside the sum:

$$\begin{aligned} \sum_{\sigma\in\Sigma} \prod_{(x,y)\in E} [1 + t\sigma(x)\sigma(y)] &= \sum_{\sigma\in\Sigma} \sum_{E'\subset E} t^{|E'|} \prod_{(x,y)\in E'} \sigma(x)\sigma(y) \\ &= \sum_{E'\subset E} t^{|E'|} \sum_{\sigma\in\Sigma} \prod_{(x,y)\in E'} \sigma(x)\sigma(y) \end{aligned}$$

In other words, we are summing over all possible subgraphs of G. We can actually explicitly evaluate the contribution from each subgraph; note that if any vertex does not have even degree, when we sum over all configurations, this particular subgraph will contribute nothing. To see this, suppose the vertex $x_0 \in V$ has odd degree in (V, E'); without loss of generality, suppose it appears exactly once, say in the edge $e_0 := (x_0, z)$. Partition Σ as $\Sigma := \Sigma_+ \cup \Sigma_-$, where $\Sigma_+ := \{\sigma : V \to \{\pm 1\} \mid \sigma(x_0) = 1\}$, and $\Sigma_- := \{\sigma : V \to \{\pm 1\} \mid \sigma(x_0) = -1\}$. Finally, denote Σ' to be the set of maps from $V \setminus \{x_0\}$ to $\{\pm 1\}$. Then,

$$\begin{aligned} \sum_{\sigma\in\Sigma} \prod_{(x,y)\in E'} \sigma(x)\sigma(y) &= \sum_{\sigma\in\Sigma} \left(\sigma(x_0)\sigma(z) \prod_{(x,y)\in E'\setminus e_0} \sigma(x)\sigma(y) \right) \\ &= \sum_{\sigma\in\Sigma_+} \left(\sigma(x_0)\sigma(z) \prod_{(x,y)\in E'\setminus e_0} \sigma(x)\sigma(y) \right) + \sum_{\sigma\in\Sigma_-} \left(\sigma(x_0)\sigma(z) \prod_{(x,y)\in E'\setminus e_0} \sigma(x)\sigma(y) \right) \\ &= \sum_{\sigma\in\Sigma'} \left(\sigma(z) \prod_{(x,y)\in E'\setminus e_0} \sigma(x)\sigma(y) \right) - \sum_{\sigma\in\Sigma'} \left(\sigma(z) \prod_{(x,y)\in E'\setminus e_0} \sigma(x)\sigma(y) \right) \\ &= 0. \end{aligned}$$

On the other hand, if every vertex has even degree, since $\sigma(x)^2 \equiv 1$ for any $x \in V$, we get that

$$\sum_{\sigma\in\Sigma} \prod_{(x,y)\in E'} \sigma(x)\sigma(y) = \sum_{\sigma\in\Sigma} 1 = 2^{|V|}.$$

The set of subgraphs with this property is precisely $C(G)$; a few examples of closed curves and their corresponding contributions to the partition function are shown in Figure (2.1). This concludes the proof. □

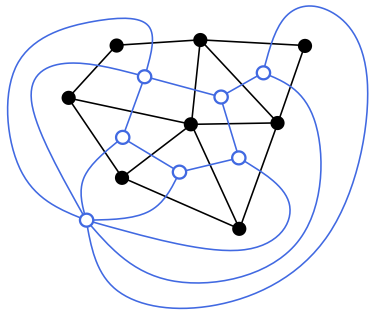
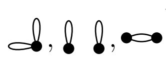
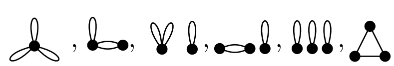


Figure 2.2. A graph G, shown in black, and its dual, G^*, shown in blue. The vertices of the dual graph G^* are the faces of the graph G; any two vertices of G^* are connected by an edge if their corresponding faces share an edge of G.

What we have proven is that the partition function for the Ising model on a graph G acts as a generating function for closed curves on G. We can express the first few terms of this expansion in graphical form:

$C\cdot Z(\beta) = 1+\#\Big\{$  $\Big\}\cdot t+\#\Big\{$  $\Big\}\cdot t^2+\#\Big\{$  $\Big\}\cdot t^3+\mathcal{O}(t^4),$ (2.19)

where $C := 2^{-|V|}\,[\cosh\beta]^{-|E|}$, and the notation $\#\{\cdot\}$ is interpreted as *the number of subgraphs in G isomorphic to one of the graphs* '$\cdot$'. We remark that disconnected graphs also contribute to the above sum.

We can similarly obtain a low temperature ($\beta \to \infty$) expansion. In this case, we notice the following: if β is very large, $e^{-\beta H}$ is exponentially small, and so the factors contributing the most to the sum are the ones with the smallest energy H. At this point, we specialize to *planar* graphs, for which there is a simple graphical interpretation of this expansion. A *planar graph* is a graph which admits an embedding into the plane without self-intersection. Planarity is needed for the following reason: there is a clear notion of a *face* of a planar graph: it is a region bounded by some collection of edges. Since none of the edges are overlapping, this region is well-defined. In order to calculate the low-temperature expansion, we again need another graph-theoretic notion: the *dual* of a planar graph.

Definition 2.8. Let $G = (V, E)$ be a planar graph. The *dual graph* of G, denoted G^*, is the graph formed by the following procedure: for each face of G, we place a vertex of G^*. Any two vertices of G^* are connected by an edge if their corresponding faces share an edge of G.

A depiction of a graph and its dual are shown in Figure (2.2). The use of the word *dual* is indeed justified; the dual of the dual graph is isomorphic to the original graph, $G^{**} = G$. Another readily apparent fact is that the number of edges in the dual graph, $|E^*|$, is the same as the number of edges in the original graph, $|E|$. With these points in place, we can proceed to construct the low temperature expansion of the Ising model.

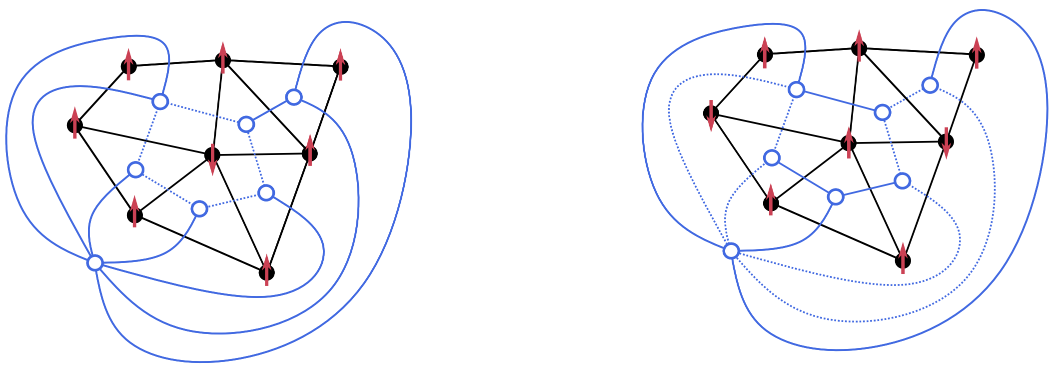


Figure 2.3. Two example configurations with one and two spins misaligned on the graph G; the dual graph G^* is shown in blue. Note that the dashed edges on the dual graph correspond to the bonds (edges) which are frustrated, i.e. contribute a higher energy to $H(\sigma)$. These edges form a closed curve c in G^*. Since $|E^*|=|E|$, we have that the contribution from each of these configurations is $s^{|c|}$. For the leftmost configuration, this evaluates to s^5; for the rightmost configuration, this evaluates to s^7.

Theorem 2.9. *Let $G=(V,E)$ be a finite graph, and $Z_G(\beta)$ be the partition function for the Ising model on G. Set $s=e^{-2\beta}$. Then,*

$$Z_G(\beta) = 2e^{\beta|E|}\sum_{c\in C(G^*)} s^{|c|} = 2e^{\beta|E|}\sum_{k=0}^{|E^*|}\tilde{c}_k s^k, \tag{2.20}$$

where $\tilde{c}_k$ is the number of closed curves of size k in G^* (the number of closed curves of size 0 is defined to be 1).

Proof. We again note that, if β is very large, then $e^{-\beta H}$ is exponentially small, and so the factors contributing the most to the sum are the ones with the smallest energy H. Obviously, the largest contribution to the sum comes from those configurations σ which minimize $H(\sigma)$; there are two of these, $\sigma(v)\equiv 1$ and $\sigma(v)\equiv -1$. These contribute a factor of $e^{\beta|E|}$. What about the next-lowest contribution? Consider a configuration with one (two) spin(s) misaligned on the graph G, as shown in Figure 2.3. Note that the dashed edges on the dual graph correspond to the bonds (edges) which have frustration, i.e. contribute a higher energy to $H(\sigma)$. These edges form a closed curve c in G^*. Now, since $|E^*|=|E|$, we have that the contribution from this configuration is $e^{\beta(|E|-2|c|)}$. For each closed curve in the dual graph, there are precisely two configurations σ which correspond to this curve, one of which is minus the other. We therefore see that we can write the partition function as a sum over closed curves in the dual graph:

$$Z_G(\beta) = \sum_{c\in C(G^*)} 2e^{\beta(|E|-2|c|)} = 2e^{\beta|E|}\sum_{c\in C(G^*)} s^{|c|}, \tag{2.21}$$

where we have set $s=e^{-2\beta}$. This concludes the proof. ∎

Thus, we have obtained two expansions of the Ising partition function, at high and low temperature. Both expressions are in terms of a sum over closed curves, either in G or G^*; we see that, identifying parameters

$$\tanh \beta_{high} = t = s = e^{-2\beta_{low}},$$

the high temperature expansion on G is equivalent to the low temperature expansion on the dual graph G^*. This is the essence of the *Kramers-Wannier duality*:

Theorem 2.10. *Let $G = (V, E)$ be a planar graph, and G^* its dual graph. Let $Z_G(\beta)$, $Z_{G^*}(\beta)$ denote the partition function for the Ising model on G and G^*, respectively. Then,*

$$Z_{G^*}(\beta) = 2e^{\tilde{\beta}|E|}2^{|V^*|}\left[\cosh(\beta)\right]^{|E|} Z_G(\tilde{\beta}), \tag{2.22}$$

where $\tilde{\beta} = -\frac{1}{2}\log\left[\tanh \beta\right]$.

Proof. On one hand, we can express $Z_{G^*}(\beta)$ using the high-temperature expansion:

$$2^{-|V^*|}\left[\cosh(\beta)\right]^{-|E|} Z_{G^*}(\beta) = \sum_{k=0}^{|E|} \tilde{c}_k t^k,$$

where $t = \tanh \beta$, and $\tilde{c}_k$ is the number of closed curves of size k in G^*. On the other hand, we can write the low-temperature expansion of the Ising model on G as

$$2e^{-\tilde{\beta}|E|} Z_G(\tilde{\beta}) = \sum_{k=0}^{|E|} \tilde{c}_k s^k, \tag{2.23}$$

where $s = e^{-2\tilde{\beta}}$, and $\tilde{c}_k$ again counts the number of closed curves of size k in G^*. Thus, if we set $\tilde{\beta} = -\frac{1}{2}\log\left[\tanh \beta\right]$ in the above expression, the right hand sides of both expressions we have derived become identical, and so we find that

$$2^{-|V^*|}\left[\cosh \beta\right]^{-|E|} Z_{G^*}(\beta) = 2e^{-\tilde{\beta}|E|} Z_G(\tilde{\beta}),$$

when $\tilde{\beta} = -\frac{1}{2}\log\left[\tanh \beta\right]$. Rearranging yields the desired result. □

Remark 2.11. This duality allows us to predict the temperature at which the phase transition occurs in the case of the square lattice Ising model. We shall later *prove* that this is indeed the case, but it is still instructive to note the argument made originally by Kramers and Wannier [73]. As we have previously mentioned, since $\mathbb{Z}^2$ is of infinite size, we are interested in the free energy per unit site, $f(\beta)$. It is also apparent that, since the graph is indeed of infinite size, it is possible for a phase transition to occur. Assuming that there are indeed only two phases, a high-temperature and low-temperature one, we see that the two phases must meet at some point of non-analyticity of $f(\beta)$, say $\beta = \beta_c$. This means that the high-temperature expansion of $f(\beta)$ (respectively, low temperature expansion) should converge all the way up to this critical temperature. Moreover, since $\mathbb{Z}^2$ is self-dual, the high-temperature and low-temperature expansions should be proportional, and we must have the equality

$$\beta_c = -\frac{1}{2}\log\left[\tanh\beta_c\right]. \tag{2.24}$$

Solving this equation for β_c, we obtain that

$$\beta_c = \frac{1}{2}\log(1+\sqrt{2}), \tag{2.25}$$

which is indeed the critical temperature of the Ising model on $\mathbb{Z}^2$.

2.3 Exact Solution of the 2D Ising Model: Combinatorial Method

We now discuss the combinatorial approach to the solution to the $2D$ Ising model; this technique was first introduced by Kac and Ward in [64], and later refined by Potts and Ward in [92]. See also the expository article [24], which revisits some of these methods with more detail.

Recall that the partition function for the Ising model on a graph $G = (V, E)$ admits the (high-temperature) expansion

$$Z(\beta) = 2^{|V|}[\cosh\beta]^{|E|}\sum_{c\in C(G)}[\tanh\beta]^{|c|}; \tag{2.26}$$

here, $C(G)$ denotes the set of all closed curves (possibly disconnected) in G. We make use of the combinatorial fact that the *logarithm* of a generating function counting the number of objects of a given kind produces the generating function for the number of *connected* objects of the same kind (cf. R.P. Stanley's book [98], Chapter 5). Furthermore, if the graph in question is *vertex transitive* [3], this sum reduces to a sum over loops based at a fixed vertex v_0:

$$\beta\cdot F(\beta) := -\log Z(\beta) = -|V|\log 2 - |E|\log(\cosh\beta) - |V|\sum_{c\in C_0(G)}[\tanh\beta]^{|c|}; \tag{2.27}$$

here, $C_0(G)$ denotes the set of all loops in G based at $v_0 \in V$.

[3] Intuitively, one can think a vertex transitive graph is one which 'looks the same' from the point of view of every vertex v. The proper definition is as follows (cf. [15]): G is vertex-transitive if any two vertices can be mapped into each other by an automorphism of the graph.

Let us now further specialize to the case of a specific graph: the square lattice Ising model (i.e., the Ising model on $\mathbb{Z}^2$). In this case, if we take a finite cutoff, then take the limit as $N \to \infty$, dividing by $|E|$, we obtain that (note that the handshaking lemma gives us that $4|V|= 2|E|$):

$$\begin{aligned}\beta f(\beta) &:= \lim_{|E|\to\infty} \frac{\beta}{|E|} \log Z(\beta) = -\frac{1}{2}\log 2 - \log(\cosh\beta) - \frac{1}{2}\sum_{c\in C_0(\mathbb{Z}^2)} [\tanh\beta]^{|c|} \\ &= -\frac{1}{2}\log 2 - \log(\cosh\beta) - \frac{1}{2}\sum_{k=0}^{\infty} R_k t^k, \end{aligned} \tag{2.28}$$

where R_k is the number of loops in $\mathbb{Z}^2$ based at the origin, and we have again abbreviated $t = \tanh\beta$. Thus, the problem of computing the free energy has been reduced to a combinatorial one: we have only to count the number of loops (closed curves) based at the origin appearing in the lattice. We remark that the same technique applies to any vertex transitive graph, and so in particular this method also applies to the triangular and hexagonal lattices. For now, we shall stick with $G = \mathbb{Z}^2$. With this technique in mind, we state the following theorem:

Theorem 2.12. *Let $f(\beta)$ denote the free energy per unit site for the Ising model on $\mathbb{Z}^2$. Then,*

$$\beta f(\beta) = -\frac{1}{2}\log 2 + \frac{1}{2}\int_{\mathbb{T}^2} \log\Delta(\beta;\theta_1,\theta_2)\frac{d\theta_1 d\theta_2}{(2\pi)^2}, \tag{2.29}$$

where the function $\Delta(\beta;\theta_1,\theta_2)$ is given by

$$\Delta(\beta;\theta_1,\theta_2) := \cosh^2(2\beta) - \sinh\left[2\beta(\cos(\theta_1) + \cos(\theta_2)\right]. \tag{2.30}$$

Proof. We give a sketch of the proof found in [64, 92]; for further details, one should consult these works. The main idea is to calculate the generating function

$$\sum_{k=0}^{\infty} R_k t^k, \tag{2.31}$$

where R_k is the number of closed curves based at the origin in $\mathbb{Z}^2$, and t is a parameter. It is clear that, if we can accomplish this, we are done. For each path γ in $\mathbb{Z}^2$ starting at the origin (closed or not), we introduce a *weight* $\rho(\gamma)$, or, as we shall sometimes refer to it, an *amplitude*, thought of as a weight on the vertices. The identity path γ_0 (where we remain at the origin) is assigned weight $\rho(\gamma_0) = 1$, and the weights of the other paths are defined iteratively. If the path γ_n continues into the next vertex in a straight line, as shown in the leftmost picture of Figure (2.4), then the weight of the path γ_{n+1} is given as $\rho(\gamma_{n+1}) = t\rho(\gamma_n)$. If the path γ_n approaches the next vertex from the left (respectively, right), then the weight of γ_{n+1} is set to be

$\rho(\gamma_{n+1}) = \alpha t \rho(\gamma_n)$ (respectively, $\rho(\gamma_{n+1}) = \bar{\alpha} t \rho(\gamma_n)$), where $\alpha = e^{\pi i/4}$, and accounts for the turning of the path. We again refer to Figure (2.4). Obviously, any path γ of length n satisfies $|\rho(\gamma)|= t^n$; if the path γ returns to the origin, it is also clear that the weight $\rho(\gamma)$ is real-valued.

For each $(x, y) \in \mathbb{Z}^2$, and every $n \geq 0$, we define the function $U_n(x, y)$, $D_n(x, y)$, $L_n(x, y)$ and $R_n(x, y)$ as the amplitudes that a path γ reaches the position (x, y) in exactly n steps from above, below, the left, and the right, respectively. In other words, $U_n(x, y)$ is, for example,

$$U_n(x, y) = \sum_{\gamma:0\to(x,y)} \rho(\gamma),$$

where the sum is taken over all paths γ of length n whose terminal vertex is (x, y), and reach this vertex by approaching at the final step (x, y) from $(x, y + 1)$. All other amplitudes are defined in a similar manner.

Since paths cannot double back on the place they came from, we obtain the following recursive formulae for the amplitudes $U_n(x, y)$, $D_n(x, y)$, $L_n(x, y)$ and $R_n(x, y)$:

$$\begin{aligned}
U_n(x, y) &= tU_{n-1}(x, y - 1) && +0 && +t\alpha L_{n-1}(x, y - 1) && +t\bar{\alpha} R_{n-1}(x, y - 1),\\
D_n(x, y) &= 0 && +tD_{n-1}(x, y + 1) && +t\bar{\alpha} L_{n-1}(x, y + 1) && +t\alpha R_{n-1}(x, y + 1),\\
L_n(x, y) &= t\bar{\alpha} U_{n-1}(x + 1, y) && +t\alpha D_{n-1}(x + 1, y) && +tL_{n-1}(x + 1, y) && +0,\\
R_n(x, y) &= t\alpha U_{n-1}(x - 1, y) && +t\bar{\alpha} D_{n-1}(x - 1, y) && +0 && +tR_{n-1}(x - 1, y).
\end{aligned}$$

Now, suppose $F_n \in \{U_n, D_n, L_n, R_n\}$, and define the function

$$\hat{F}_n(u, v) := \sum_{x,y\in\mathbb{Z}^2} F_n(x, y)u^x v^y,$$

for $|u|= |v|= 1$. It follows by Fourier inversion that

$$F_n(x, y) = \int_{|u|=1} \int_{|v|=1} \hat{F}_n(u, v)u^{-x} v^{-y} \frac{du}{2\pi iu} \frac{dv}{2\pi iv}.$$

The recursions on the amplitudes $U_n(x, y)$, $D_n(x, y)$, $L_n(x, y)$ and $R_n(x, y)$ in turn imply recursions for the functions $\hat{U}_n(u, v), \hat{D}_n(u, v), \hat{L}_n(u, v), \hat{R}_n(u, v)$:

$$\begin{aligned}
\hat{U}_n(u, v) &= tv\hat{U}_{n-1}(u, v) && +0 && +t\alpha v\hat{L}_{n-1}(u, v) && +t\bar{\alpha} v\hat{R}_{n-1}(u, v),\\
\hat{D}_n(u, v) &= 0 && +t\bar{v}\hat{D}_{n-1}(u, v) && +t\bar{\alpha}\bar{v}\hat{L}_{n-1}(u, v) && +t\alpha\bar{v}\hat{R}_{n-1}(u, v),\\
\hat{L}_n(u, v) &= t\bar{\alpha}\bar{u}\hat{U}_{n-1}(u, v) && +t\alpha\bar{u}\hat{D}_{n-1}(u, v) && +t\bar{u}\hat{L}_{n-1}(u, v) && +0,\\
\hat{R}_n(u, v) &= t\alpha u\hat{U}_{n-1}(u, v) && +t\bar{\alpha} u\hat{D}_{n-1}(u, v) && +0 && +tu\hat{R}_{n-1}(u, v).
\end{aligned}$$

Alternatively, defining the column vector $\hat{\psi}_n(u,v) := \langle \hat{U}_n(u,v), \hat{D}_n(u,v), \hat{L}_n(u,v), \hat{R}_n(u,v)\rangle^T$, one can write the above system of equations compactly as the matrix equation

$$\hat{\psi}_n(u,v) = t\mathbb{K}(u,v;\alpha)\hat{\psi}_{n-1}(u,v),$$

where

$$\mathbb{K}(u,v;\alpha) := \begin{pmatrix} v & 0 & \alpha v & \bar{\alpha} v \\ 0 & \bar{v} & \bar{\alpha}\bar{v} & \alpha\bar{v} \\ \bar{\alpha}\bar{u} & \alpha\bar{u} & \bar{u} & 0 \\ \alpha u & \bar{\alpha} u & 0 & u \end{pmatrix}.$$

It follows that

$$\hat{\psi}_n(u,v) = t^n\mathbb{K}^n(u,v;\alpha)\hat{\psi}_0(u,v),$$

and, by the Fourier inversion formula, we see that the total number of oriented closed curves of size n (unbased, but passing through zero) is given by

$$A_n t^n = -\frac{1}{2}\int_{|u|=1}\int_{|v|=1} \operatorname{tr}\mathbb{K}^n(u,v;\alpha)\frac{du}{2\pi i u}\frac{dv}{2\pi i v}.$$

If we only want to count the closed curves based at 0, we have only to divide by the combinatorial factor of n:

$$R_n t^n = -\frac{1}{2}\int_{|u|=1}\int_{|v|=1} \frac{1}{n}\operatorname{tr}\mathbb{K}^n(u,v;\alpha)\frac{du}{2\pi i u}\frac{dv}{2\pi i v}.$$

Thus, the generating function we were looking for is

$$\begin{aligned}\sum_{n=0}^{\infty} R_n t^n &= -\frac{1}{2}\sum_{n=0}^{\infty}\int_{|u|=1}\int_{|v|=1} \frac{1}{n}\operatorname{tr}\mathbb{K}^n(u,v;\alpha)\frac{du}{2\pi i u}\frac{dv}{2\pi i v} \\ &= -\frac{1}{2}\int_{|u|=1}\int_{|v|=1} \log\det\left[\mathbb{I} - t\mathbb{K}(u,v;\alpha)\right]\frac{du}{2\pi i u}\frac{dv}{2\pi i v}.\end{aligned}$$

Evaluation of this determinant yields the desired result. □

Remark 2.13. This is not the only solution method for calculating the exact free energy of the two-dimensional Ising model, and in fact is not even the original method employed by Onsager to derive the free energy. Onsager and Kaufman used the so-called *transfer matrix method* to calculate the free energy, correlation functions, and spontaneous magnetization [68, 69, 86]. This method makes a very important aspect of this model much more transparent: its *quantum group symmetry*, and connection to quantum

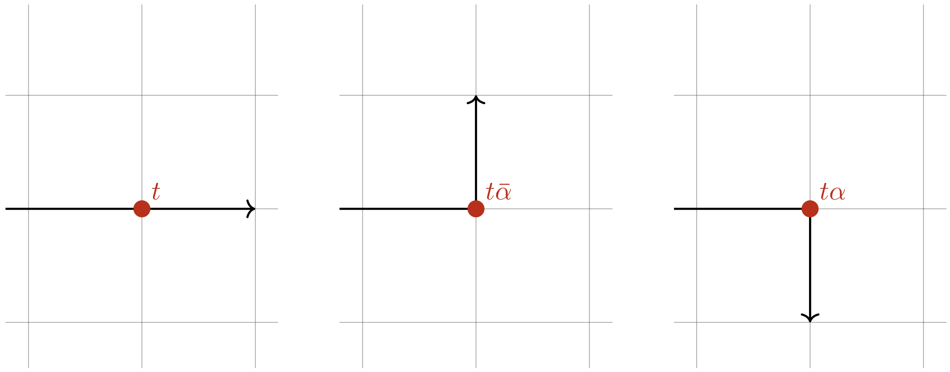


Figure 2.4. The loop 'building blocks'. Each vertex is assigned a weight t, $t\alpha$, $t\bar{\alpha}$, which depends on the direction the path turns in $\mathbb{Z}^2$.

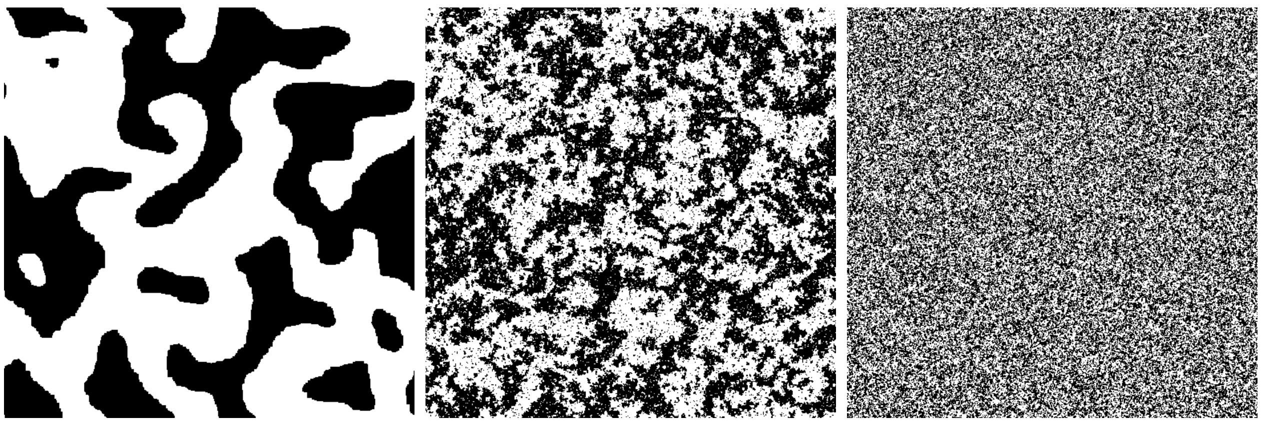


Figure 2.5. *(Left).* An average configuration of the Ising model below the critical temperature ($\beta > \beta_c$). Note the appearance of large-scale features, and the absence of most "noise". *(Middle).* The critical Ising model ($\beta = \beta_c$). Here, the correlation length diverges, and we see features of every scale. *(Right).* The high temperature Ising model ($\beta < \beta_c$). There is an absence of any large-scale features, the model is mostly noise.

integrable systems. Confer with [7] for a general introduction to the transfer matrix method, and [63] for details on the quantum integrability of the 2D Ising model and related lattice models.

From this result, Onsager was able to infer the formula for the spontaneous magnetization of the Ising model, as well as obtain a formula for the correlation length. We state the following propositions without proof:

Proposition 2.14. *(C.N. Yang, [108]). The spontaneous magnetization of the Ising model, defined as*

$$M := \lim_{h\to 0} M(\beta, h) = \lim_{h\to 0} \frac{\partial f(\beta, h)}{\partial h}, \tag{2.32}$$

is given by the explicit expression

$$M(\beta) = \begin{cases} [1 - \sinh^{-4}(2\beta)]^{1/8}, & \beta > \beta_c, \\ 0 & \beta < \beta_c. \end{cases} \tag{2.33}$$

Proposition 2.15. *Let $\langle \sigma_{00}\sigma_{NN} \rangle$ denote the diagonal correlation of the spins at sites $(0,0)$ and (N,N), on the square $[0,N] \times [0,N]$. Then,*

$$\langle \sigma_{00}\sigma_{NN} \rangle = \det_{1 \le n < m \le N} \varphi_{n-m}, \tag{2.34}$$

where $\varphi_k = \frac{1}{2\pi}\int_0^{2\pi} \Phi(\theta) e^{-ik\theta} d\theta$ is the k^{th} Fourier coefficient of the function

$$\Phi(\theta) = \left(\frac{\sinh^2(2\beta) - e^{-i\theta}}{\sinh^2(2\beta) - e^{i\theta}} \right)^{1/2}. \tag{2.35}$$

Furthermore, as $N \to \infty$, putting $t := [\sinh(2\beta)]^{-2}$,

$$\langle \sigma_{00}\sigma_{NN} \rangle = \begin{cases} (1-t)^{1/4} \left[1 + \mathcal{O}(N^{-2})\right], & \beta > \beta_c, \\ \frac{C}{N^{1/4}} \left[1 + \mathcal{O}(N^{-2})\right], & \beta = \beta_c, \\ \frac{t^{N/2}}{(\pi N)^{1/2}(1-t)^{1/4}} \left[1 + \mathcal{O}(N^{-2})\right] & \beta < \beta_c, \end{cases} \tag{2.36}$$

where the constant $C = 2^{1/12} e^{3\zeta'(1)}$, and $\zeta(s)$ is Riemann's zeta function.

One often writes t^N as $\exp\{N\xi\}$, where $\xi := \log t$ is called the *correlation length.* It is apparent that, at low temperatures, the Ising model has residual magnetization, and the correlation length is finite. This is apparent even in simulations of the Ising model, as Figure (2.5) shows: the leftmost panel of this figure depicts the typical configuration of the Ising model at some fixed temperature $\beta > \beta_c$ (i.e., below the critical temperature). The large-scale features indicate that the correlation length is finite and positive: approximately speaking, the average size of the magnetic domains (the large patches of white or black) should be on the order of the correlation length. At the critical temperature $\beta = \beta_c$, the correlation length diverges logarithmically; this is visible in the appearance of magnetic domains of arbitrary scales.

Remark 2.16. We additionally remark on the fact that the correlations of the Ising model can be represented as a Toeplitz-type determinant. We do not give a full account of the theory of Toeplitz determinants here, as it will take us too far from the topic of the current work. However, we feel it is important to acknowledge this connection, as most of the modern theory of Toeplitz determinants was developed to study the correlations of the Ising and other exactly solvable lattice models. Most of the modern interest in the Ising model comes from the rigorous calculation of its correlation functions, and is intimately tied to the Riemann-Hilbert analysis and Painlevé equations we shall meet in the subsequent chapters of this thesis.

The connection of Toeplitz matrices and correlations in the Ising model was first noted in [107], and would rely on Szëgo's famous limit formula for Toeplitz determinants [99] to compute the large N limit. However, Szëgo's formula required the symbol $\Phi(\theta)$ to be strictly positive, whereas the symbol coming from the correlations of the Ising model has an apparent algebraic singularity (2.35). This led M. E. Fisher and R. E. Hartwig to formulate a conjecture for the form of the limiting determinant of a Toeplitz matrix with algebraic singularities [46], which allowed one to write down the formulas given in the proposition above. This became the famous *Fisher–Hartwig conjecture*. Some partial results were obtained by various authors in [6, 16, 42, 104]. The full resolution to the conjecture came in the work of P. Deift, A. Its, and I. Krasovsky [28], who used Riemann-Hilbert techniques for orthogonal polynomials on the unit circle to verify the conjecture.

CHAPTER 3

MATRIX MODELS & 2D QUANTUM GRAVITY.

In this chapter, we introduce matrix models, and explain their connection to $2D$ quantum gravity. We begin introducing some basic matrix models, and discuss the connection to orthogonal and (for the 2-matrix model) biorthogonal polynomials. We discuss the combinatorial interpretation of matrix integrals, and show how certain matrix integrals can be interpreted as models of $D < 1$ matter theories coupled to topological quantum gravity.

3.1 The 1-Matrix Model: Invariant Ensembles.

Let $n \geq 1$ be a fixed integer, and X an $n \times n$ Hermitian matrix, i.e. $X^\dagger = X$, where "$\dagger$" denotes the conjugate transpose. We let $\mathcal{H}_n$ denote the space of all such matrices. We denote the entries of this matrix in terms of its real and imaginary components: $X_{ij} = x_{ij} + iy_{ij}$. Note that the Hermitian constraint implies that the diagonal entries X_{ii} are purely real. The space of Hermitian matrices forms an n^2-dimensional vector space over $\mathbb{R}$; as such, it is naturally equipped with an invariant (Haar) measure:

$$dX = \prod_{i=1}^{n} dx_{ii} \prod_{1 \leq i < j \leq n} dx_{ij} dy_{ij}. \tag{3.1}$$

This is, of course, just the product measure on all of the entries of the matrix. Now, let $V(x)$ be a monic polynomial of even degree. We define the following probability measure on $\mathcal{H}_n$:

$$d\mathbb{P}(X) = d\mathbb{P}_V(X) := Z^{-1} \exp\{-n \operatorname{tr} V(X)\} \, dX. \tag{3.2}$$

Here, Z is a constant which makes $d\mathbb{P}$ a probability measure: $Z = \int \exp\{-n \operatorname{tr} V(X)\} \, dX$. We often call Z the *partition function.*

Definition 3.1. A *Hermitian 1-matrix model* is a probability measure of the form (3.2) equipped on the space of Hermitian matrices $\mathcal{H}_n$.

Remark 3.2. Note that we took V to be a monic polynomial of even degree so as to guarantee the convergence of the partition function Z. However, we will often allow for an extension of this definition to *arbitrary* polynomials, with the interpretation that we are taking the integral not over the space of Hermitian matrices any longer, but rather over some suitably chosen contours in the complex plane. We think of the resulting object as being an "analytic continuation" of what appeared in the case where the integral was well-defined.

If $f : \mathcal{H}_n \to \mathbb{C}$, we denote its expected value, if defined, by

$$\langle f(X) \rangle_n := \int_{\mathcal{H}_n} f(X) d\mathbb{P}(X) = \frac{1}{Z} \int_{\mathcal{H}_n} f(X) \exp\left\{-n \operatorname{tr} V(X)\right\} dX. \tag{3.3}$$

Of particular interest are expected values of *class functions*, i.e., functions such that, for any fixed $X \in \mathcal{H}_n$, and any unitary transformation $U \in U(n)$,

$$f(U^\dagger X U) = f(X). \tag{3.4}$$

Consequentially, we can think of f as a function of only the eigenvalues of the matrix X.

There are many celebrated examples of matrix models. One of the oldest is the *Gaussian Unitary ensemble* (GUE):

Example 3.3. *The Gaussian Unitary ensemble (GUE).* This is the ensemble of $n \times n$ matrices equipped with the "Gaussian" probability measure

$$d\mathbb{P}(X) = Z^{-1} \exp\left\{-\frac{N}{2} \operatorname{tr} X^2\right\} dX. \tag{3.5}$$

(Here $N > 0$ is a parameter). Indeed, the descriptor "Gaussian" is correct: if we write out this measure in terms of the entries of the matrix X, we find that

$$d\mathbb{P}(X) = Z^{-1} \exp\left\{-\frac{N}{2} \sum_{i=1}^{n} x_{ii}^2 - N \sum_{1 \le i < j \le n} \left(x_{ij}^2 + y_{ij}^2\right)\right\} \prod_{i=1}^{n} dx_{ii} \prod_{1 \le i < j \le n} dx_{ij} dy_{ij}, \tag{3.6}$$

i.e., the matrix X is constructed out of Gaussian random variables:

$$X_{ij} \sim \mathcal{N}(0, N) + i\mathcal{N}(0, N), \qquad i < j, \qquad\qquad X_{ii} \sim \mathcal{N}(0, \frac{1}{2} N). \tag{3.7}$$

Additionally, Equation (3.6) allows us to compute the normalization constant explicitly, since we know how to integrate Gaussians:

$$Z = Z_{GUE} = 2^{n/2}\pi^{n^2/2}N^{-n^2}. \tag{3.8}$$

We shall discuss the GUE in more detail later on.

An important observation is that the density of the probability measures above are *invariant* with respect to unitary conjugation: this follows immediately from the fact that $\operatorname{tr} V(X)$ is a class function:

$$\operatorname{tr} V(U^\dagger X U) = \operatorname{tr} V(X), \tag{3.9}$$

for any $U \in U(n)$. Since we are typically only interested in expected values of class functions, one may wonder if the measure dX admits a decomposition $dX = h(\lambda_1, ..., \lambda_n) \prod_n d\lambda_i g(U) dU$, where dU is the Haar measure on the unitary group, and $\prod_n d\lambda_i$ is the product measure on the eigenvalues of X. If this were the case, then, for any class function f,

$$\begin{aligned}
\langle f(X)\rangle_n &= \frac{1}{Z}\int_{\mathcal{H}_n} f(X)\exp\left\{-n\operatorname{tr} V(X)\right\} dX \\
&= \frac{1}{Z}\int_{\mathbb{R}^n} f(\lambda_1, ..., \lambda_n)\exp\left\{-n\sum_{k=1}^{n} V(\lambda_i)\right\} h(\lambda_1, ..., \lambda_n)\prod_{k=1}^{n} d\lambda_k \int_{U(n)} g(U)dU \\
&= \frac{1}{Z'}\int_{\mathbb{R}^n} f(\lambda_1, ..., \lambda_n)\exp\left\{-n\sum_{k=1}^{n} V(\lambda_i)\right\} h(\lambda_1, ..., \lambda_n)\prod_{k=1}^{n} d\lambda_k,
\end{aligned}$$

where we have absorbed the integral over the unitary group (the volume of the unitary group, in fact) into the partition function. Thus, we would be able to reduce the integral over n^2 variables to an integral over n variables. This is indeed possible to do:

Proposition 3.4. *Write $X \in \mathcal{H}_n$ as $X = U^\dagger D U$, where U is a unitary matrix, and $D := diag(\lambda_1, ..., \lambda_n)$. Then, the Haar measure on $\mathcal{H}_n$ can be decomposed as*

$$dX = [\Delta(\lambda_1, ..., \lambda_n)]^2 \prod_n d\lambda_i g(U) dU, \tag{3.10}$$

where $g(U)$ is a function only of the 'angular' variables of X, and

$$\Delta(\lambda_1, ..., \lambda_n) := \prod_{1\le i<j\le n} (\lambda_i - \lambda_j) \tag{3.11}$$

is called the Vandermonde determinant.

Proof. The factor (3.11) appears as the Jacobian of the transformation from the matrix variables $\{x_{ij}, y_{ij}\}$ to the eigenvalue-angular variables $\{\lambda_i, u_{ij}\}$. This is most easily seen by considering the induced Riemannian metric on the space of Hermitian matrices, canonically given by

$$g = \mathrm{tr}(\delta X^\dagger \delta X) = \sum_{i=1}^{n} dx_{ii}^2 + 2 \sum_{1 \le i < j \le n} \left(dx_{ij}^2 + dy_{ij}^2 \right), \tag{3.12}$$

where here δX denotes the Hermitian matrix of covectors

$$(\delta X)_{ij} = \begin{cases} dx_{ij} + i dy_{ij}, & i < j, \\ dx_{ii}, & i = j. \end{cases}$$

Consider new coordinates on the space of matrices, induced by the diagonalization $X = U^\dagger D U$, where $U \in U(n)$ is defined up to multiplication on the left by a diagonal unitary matrix, and $D = \mathrm{diag}\,(\lambda_1, ..., \lambda_n)$. Since U is a unitary matrix, we have that

$$0 = \delta \mathbb{I} = \delta \left[U U^\dagger \right] = U \delta U^\dagger + \delta U U^\dagger,$$

so the matrix of covectors $\delta u := \delta U U^\dagger$ is anti-Hermitian, and independent of the representative we chose for $U \in U(n)$. We can rewrite the matrix of covectors δX in these new coordinates:

$$\delta X = U^\dagger \left(\delta u D + \delta D + D \delta u^\dagger \right) U = U^\dagger \left(\delta D - [D, \delta u] \right) U,$$

so that

$$g = \mathrm{tr}(\delta X^\dagger \delta X) = \mathrm{tr}(\delta D)^2 - 2\, \mathrm{tr}(\delta D [D, \delta u]) + \mathrm{tr}([D, \delta u]^2).$$

By cyclicity of trace, we have that $\mathrm{tr}(\delta D [D, \delta u]) = 0$; by definition of δD, $\mathrm{tr}(\delta D)^2 = \sum_{i=1}^n d\lambda_i^2$. It remains to compute $\mathrm{tr}([D, \delta u]^2)$. Using the fact that $D_{ij} = \lambda_i \delta_{ij}$, we have that

$$\begin{aligned} ([D, \delta u])_{ij} &= \sum_{k=1}^{n} \left(D_{ik} \delta u_{kj} - \delta u_{ik} D_{kj} \right) \\ &= \sum_{k=1}^{n} \left(\lambda_i \delta_{ik} \delta u_{kj} - \delta u_{ik} \lambda_k \delta_{kj} \right) \\ &= (\lambda_i - \lambda_j) \delta u_{ij}. \end{aligned}$$

It follows that

$$\mathrm{tr}([D, \delta u]^2) = 2\sum_{i<j}(\lambda_i - \lambda_j)^2|\delta u_{ij}|^2,$$

so that

$$g = \mathrm{tr}(\delta X^\dagger \delta X) = \sum_{i=1}^{n} d\lambda_i^2 + 2\sum_{i<j}(\lambda_i - \lambda_j)^2|\delta u_{ij}|^2.$$

Now, the volume form corresponding to this metric is invariant under translations in the space of Hermitian matrices, since the metric is; thus, the volume form here *is* the Haar measure, up to some irrelevant normalization constant. Since the Riemannian volume element is given by (cf. [22]):

$$\mathrm{vol}_g := dX = \sqrt{\det g}\prod_{i=1}^{n} d\lambda_i \prod_{i<j} du_{ij},$$

we obtain that

$$dX = 2^{n(n-1)/2}\prod_{i<j}(\lambda_i - \lambda_j)^2 \prod_{i=1}^{n} d\lambda_i \prod_{i<j} du_{ij},$$

which is precisely the form of (3.10). □

The fact that the Jacobian factors into a product of eigenvalue and angular components indicates that the eigenvalues of the above matrix ensembles, considered as random variables, are independent from the *eigenvectors*, which are represented by the angular component of the integration. Since we are eventually only interested in spectral properties of these matrices (i.e., expected values of functions of the eigenvalues), we can 'integrate out' angular variables and work with the remaining measure, as we had hoped for. Indeed, we have the following corollary:

Corollary 3.5. *If $f : \mathcal{H}_n \to \mathbb{C}$ is a class function, then*

$$\begin{aligned}\langle f(X)\rangle_n &:= \frac{1}{Z}\int_{\mathcal{H}_n} f(X)\exp\{-nV(X)\}\,dX \\ &= \frac{1}{Z'}\int_{\mathbb{R}^n} f(\lambda_1, ..., \lambda_n)\prod_{i<j}(\lambda_i - \lambda_j)^2 \prod_{i=1}^{n} e^{n\sum_{i=1}^{n} V(\lambda_i)} d\lambda_i,\end{aligned} \tag{3.13}$$

where the constants Z, Z' are related by a constant independent of $V(X)$:

$$Z = Z'\,vol[U(n)/U_{diag}(n)]\,2^{n(n-1)/2}. \tag{3.14}$$

In fact, one can determine the partition function Z' even more explicitly. One of the most useful relations in random matrix theory is the connection to the theory of orthogonal polynomials. For this reason, we will briefly digress into the theory of orthogonal polynomials.

3.2 Orthogonal Polynomials.

Many canonical objects in random matrix theory can be computed explicitly in terms of quantities related to orthogonal polynomials. For the invariant ensembles, it will suffice to consider the family of (monic) orthogonal polynomials defined by the relation

$$\langle p_k, p_j \rangle := \int_{\mathbb{R}} p_k(\lambda) p_j(\lambda) e^{-V(\lambda)} d\lambda = h_j \delta_{jk}, \tag{3.15}$$

where $V(\lambda)$ is again a monic polynomial of even degree. Here, we will review some of the properties of such orthogonal polynomials. The first fact that we shall recall is that the above family of polynomials always satisfies a 3-term recursion relation.

Proposition 3.6. *Suppose $\{p_k(\lambda)\}$ satisfy the orthogonality relation (3.15). Then, there exist constants a_n, b_n such that*

$$p_{n+1}(\lambda) = (\lambda + a_n) p_n(\lambda) + b_n p_{n-1}(\lambda), \tag{3.16}$$

for each $n \geq 1$.

Proof. Fix $n \geq 1$, and consider the polynomial $\lambda p_n(\lambda) = \lambda^{n+1} + \ldots$. As a polynomial of degree $n+1$, we can express $\lambda p_n(\lambda)$ as a linear combination of the first $n+1$ monic orthogonal polynomials:

$$\lambda p_n(\lambda) = p_{n+1}(\lambda) + \sum_{k=0}^{n} \alpha_k^{(n)} p_k(\lambda).$$

Note that we have taken $p_{n+1}(\lambda)$ with coefficient 1, so that $\lambda p_n(\lambda)$ is indeed monic. Now, for any $0 \leq k \leq n$, we have that

$$\langle \lambda p_n, p_k \rangle = \langle p_n, \lambda p_k \rangle = \langle p_n, p_{k+1} + \ldots \rangle,$$

which can only be nonzero if $k = n$, or $k = n-1$. We thus have that

$$\lambda p_n(\lambda) = p_{n+1}(\lambda) + \alpha_n^{(n)} p_n(\lambda) + \alpha_{n-1}^{(n)} p_{n-1}(\lambda);$$

taking $a_n = -\alpha_n^{(n)}$, $b_n = -\alpha_{n-1}^{(n)}$ yields the desired result. We can also use the orthogonality relation to explicitly evaluate the coefficients a_n, b_n; taking inner products of (3.16) with $p_{n-1}(\lambda)$, we have that

$$\begin{aligned}\langle p_{n+1}, p_{n-1}\rangle &= \langle \lambda p_n, p_{n-1}\rangle + a_n\langle p_n, p_{n-1}\rangle + b_n\langle p_{n-1}, p_{n-1}\rangle \\ \Leftrightarrow 0 &= \langle p_n, \lambda p_{n-1}\rangle + b_n h_{n-1} \\ \Leftrightarrow 0 &= \langle p_n, p_n + ...\rangle + b_n h_{n-1} = h_n + b_n h_{n-1},\end{aligned}$$

so that $b_n = -\frac{h_n}{h_{n-1}}$. □

Of course, when $n = 0$, the same argumentation as above holds, and we have the alternate expression

$$\lambda p_0(\lambda) = \lambda = p_1(\lambda) + a_0 p_0(\lambda), \tag{3.17}$$

for some constant a_0.

The 3-term recursion relation leads to another useful identity for the projector onto the first n orthogonal polynomials, called the *Christoffel-Darboux kernel*:

$$K_n(\lambda, \mu) := \sum_{k=0}^{n-1} \frac{p_k(\lambda)p_k(\mu)}{h_k}. \tag{3.18}$$

This kernel has the following reproducing property; for any polynomial $q(\lambda)$ of degree at most $n-1$,

$$q(\lambda) = \int_{\mathbb{R}} K_n(\lambda, \mu)q(\mu)e^{-V(\mu)}d\mu. \tag{3.19}$$

In particular, the kernel itself is a degree $n-1$ polynomial in λ, and so we have that

$$K_n(\lambda, \mu) = \int_{\mathbb{R}} K_n(\lambda, \nu)K_n(\nu, \mu)e^{-V(\nu)}d\nu. \tag{3.20}$$

The 3-term recursion relation (3.16) allows us to provide a simpler expression for the kernel $K_n(\lambda, \mu)$.

Proposition 3.7. *For any $n \geq 1$,*

$$K_n(\lambda, \mu) = \frac{1}{h_{n-1}} \frac{p_n(\lambda)p_{n-1}(\mu) - p_{n-1}(\lambda)p_n(\mu)}{\lambda - \mu}. \tag{3.21}$$

Proof. Using the recursion relation (3.16), we have that

$$\begin{aligned}
(\lambda-\mu)K_n(\lambda,\mu) &= \sum_{k=0}^{n-1}\frac{\lambda p_k(\lambda)p_k(\mu)}{h_k} - \sum_{k=0}^{n-1}\frac{p_k(\lambda)\mu p_k(\mu)}{h_k} \\
&= \sum_{k=0}^{n-1}\frac{1}{h_k}p_k(\mu)(p_{k+1}(\lambda)-a_kp_k(\lambda)-b_kp_{k-1}(\lambda)) \\
&- \sum_{k=0}^{n-1}\frac{1}{h_k}p_k(\lambda)(p_{k+1}(\mu)-a_kp_k(\mu)-b_kp_{k-1}(\mu)) \\
&= \sum_{k=0}^{n-1}\frac{1}{h_k}p_k(\mu)(p_{k+1}(\lambda)-b_kp_{k-1}(\lambda)) - \sum_{k=0}^{n-1}\frac{1}{h_k}p_k(\lambda)(p_{k+1}(\mu)-b_kp_{k-1}(\mu)).
\end{aligned}$$

Now, using the fact that $b_k = -\frac{h_k}{h_{k-1}}$, and that b_0 is vacuously 0 we find that the series above telescopes to

$$(\lambda-\mu)K_n(\lambda,\mu) = \frac{1}{h_{n-1}}\left[p_n(\lambda)p_{n-1}(\mu) - p_{n-1}(\lambda)p_n(\mu)\right].$$

This concludes the proof. □

We now have a number of useful lemmas involving determinants.

Lemma 3.8. *Suppose $f(\lambda,\mu)$ is a function satisfying the reproducing property*

$$f(\lambda,\mu) = \int_{\mathbb{R}} f(\lambda,\nu)f(\nu,\mu)e^{-V(\nu)}d\nu, \tag{3.22}$$

and consider the $n\times n$ matrix $\mathbf{M}_n$ with entries $(\mathbf{M}_n)_{i,j} = f(\lambda_i,\lambda_j)$. Then,

$$\int_{\mathbb{R}} \det(\mathbf{M}_n)e^{-V(\lambda_n)}d\lambda_n = (C-n+1)\det(\mathbf{M}_{n-1}), \tag{3.23}$$

where $C = \int f(\lambda,\lambda)e^{-V(\lambda)}d\lambda$.

Proof. Denote by S_n the set of permutations of $\{1,2,\ldots,n\}$ and by $S_{n,-j}$ to be the set of bijective maps $\sigma' : \{1,2,\ldots,n\} \mapsto \{1,2,\ldots,n\}$ with $\sigma'(j) = n$. By Leibniz formula for determinants, we have that

$$\begin{aligned}
\int \det(\mathbf{M}_n) e^{-V(\lambda_n)} d\lambda_n &= \sum_{\sigma \in S_n} (-1)^{\sigma} \prod_{i=1}^{n} \int f(\lambda_{\sigma(i)}, \lambda_i) e^{-V(\lambda_n)} d\lambda_n \\
&= \sum_{j=1}^{n} \sum_{\sigma' \in S_{n,-j}} (-1)^{\sigma'+1} \prod_{i=1}^{n} \int f(\lambda_{\sigma'(i)}, \lambda_i) e^{-V(\lambda_n)} d\lambda_n \\
&= \sum_{j=1}^{n-1} \sum_{\sigma' \in S_{n,-j}} (-1)^{\sigma'+1} \prod_{i \neq j, i=1}^{n-1} f(\lambda_{\sigma'(i)}, \lambda_i) \int f(\lambda_{\sigma'(n)}, \lambda_n) f(\lambda_n, \lambda_j) e^{-V(\lambda_n)} d\lambda_n \\
&+ \sum_{\sigma \in S_{n-1}} (-1)^{\sigma} \prod_{i=1}^{n-1} f(\lambda_{\sigma(i)}, \lambda_i) \int f(\lambda_n, \lambda_n) e^{-V(\lambda_n)} d\lambda_n
\end{aligned}$$

By the reproducing property (3.22) for f, and the definition of C, we find that

$$\begin{aligned}
\int \det(\mathbf{M}_n) e^{-V(\lambda_n)} d\lambda_n &= \sum_{j=1}^{n-1} \sum_{\sigma' \in S_{n,-j}} (-1)^{\sigma'+1} \prod_{i \neq j, i=1}^{n-1} f(\lambda_{\sigma'(i)}, \lambda_i) f(\lambda_{\sigma'(n)}, \lambda_j) \\
&+ C \sum_{\sigma' \in S_{n,-n}} (-1)^{\sigma'} \prod_{i=1}^{n-1} f(\lambda_{\sigma'(i)}, \lambda_i) \\
&= -\sum_{j=1}^{n-1} \sum_{\sigma' \in S_{n,-j}} (-1)^{\sigma'} \prod_{i=1}^{n-1} f(\lambda_{\sigma'(n)}, \lambda_j) + C \det(\mathbf{M}_{n-1}) \\
&= -\sum_{j=1}^{n-1} \sum_{\sigma \in S_{n-1}} (-1)^{\sigma} \prod_{i=1}^{n-1} f(\lambda_{\sigma(n)}, \lambda_j) + C \det(\mathbf{M}_{n-1}) \\
&= (C - n + 1) \det(\mathbf{M}_{n-1}).
\end{aligned}$$

□

Proposition 3.9. *Let $\{p_k(\lambda)\}$ be the orthogonal polynomials defined by (3.15), and consider the matrix*

$$\mathbf{P}_n := \begin{pmatrix} p_0(\lambda_1) & p_1(\lambda_1) & \ldots & p_{n-1}(\lambda_1) \\ p_0(\lambda_2) & p_1(\lambda_2) & \ldots & p_{n-1}(\lambda_2) \\ \vdots & \vdots & \ddots & \vdots \\ p_0(\lambda_n) & p_1(\lambda_n) & \ldots & p_{n-1}(\lambda_n) \end{pmatrix}. \tag{3.24}$$

Then,

$$\int_{\mathbb{R}^n} [\det(\mathbf{P}_n)]^2 e^{-\sum_{i=1}^{n} V(\lambda_i)} \prod_{i=1}^{n} d\lambda_i = n! \prod_{i=0}^{n-1} h_i. \tag{3.25}$$

Proof. Define $\mathbf{H}_n := \mathrm{diag}(h_0, ..., h_{n-1})$. Then, since $[\det(\mathbf{P}_n)]^2 = \det(\mathbf{H}_n) \cdot \det[\mathbf{P}_n \mathbf{H}_n^{-1} \mathbf{P}_n^T]$, we have that

$$[\det(\mathbf{P}_n)]^2 = \det(\mathbf{H}_n) \cdot \det[\mathbf{P}_n \mathbf{H}_n^{-1} \mathbf{P}_n^T] = \det[K_n(\lambda_i, \lambda_j)],$$

where $K_n(\lambda, \mu)$ is the reproducing kernel (3.18). This kernel satisfies the hypotheses of Lemma (3.8), with norming constant $\int_{\mathbb{R}} K_n(\lambda, \lambda) d\lambda = n$. Thus, we find that

$$\begin{aligned}\int_{\mathbb{R}^n} [\det(\mathbf{P}_n)]^2 e^{-\sum_{i=1}^n V(\lambda_i)} \prod_{i=1}^n d\lambda_i &= \det(\mathbf{H}_n) \int_{\mathbb{R}^{n-1}} \int_{\mathbb{R}} \det_{n\times n}[K_n(\lambda_i, \lambda_j)] d\lambda_n \prod_{i=1}^{n-1} d\lambda_i \\ &= \det(\mathbf{H}_n) \int_{\mathbb{R}^{n-1}} \det_{(n-1)\times(n-1)}[K_n(\lambda_i, \lambda_j)] \prod_{i=1}^{n-1} d\lambda_i;\end{aligned}$$

proceeding inductively using Lemma (3.8), we obtain finally that

$$\int_{\mathbb{R}^n} [\det(\mathbf{P}_n)]^2 e^{-\sum_{i=1}^n V(\lambda_i)} \prod_{i=1}^n d\lambda_i = n! \det(\mathbf{H}_n).$$

Since $\det(\mathbf{H}_n) = \prod_{i=0}^{n-1} h_i$, we are done. □

Consequentially, we see that

Corollary 3.10. *The partition function Z' in (3.13) is*

$$Z' = n! \prod_{i=0}^{n-1} h_i, \tag{3.26}$$

where h_i are the norming constants for the monic orthogonal polynomials with respect to the weight $e^{nV(\lambda)}$.

Proof. This follows immediately from the fact that the determinant of the matrix (called the *Vandermonde matrix*)

$$\det \begin{pmatrix} 1 & \lambda_1 & \dots & \lambda_1^{n-1} \\ 1 & \lambda_2 & \dots & \lambda_2^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \lambda_n & \dots & \lambda_n^{n-1} \end{pmatrix} = \prod_{i<j} (\lambda_i - \lambda_j) = \Delta(\lambda_1, ..., \lambda_n), \tag{3.27}$$

and from the fact that the matrix $\mathbf{P}_n$ can be obtained from

$$\begin{pmatrix} 1 & \lambda_1 & \dots & \lambda_1^{n-1} \\ 1 & \lambda_2 & \dots & \lambda_2^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \lambda_n & \dots & \lambda_n^{n-1} \end{pmatrix}$$

by adding scalar multiples of the columns to the left of any given one, an operation which does not change the determinant of the matrix. □

The partition function is one of the most fundamental objects in random matrix theory; the above shows that this function can be computed explicitly in terms of quantities related to the orthogonal polynomials. The next theorem shows that many other quantities of interest can be computed in terms of the orthogonal polynomials.

Proposition 3.11. *(Heine Formula). Fix $n \geq 1$, and consider the invariant ensemble defined in (3.2). Then,*

$$p_n(\lambda) = \langle \det(\lambda \mathbb{I} - X) \rangle_n, \tag{3.28}$$

$$K_n(\lambda, \mu) = \langle \det(\lambda \mathbb{I} - X) \det(\mu \mathbb{I} - X) \rangle_{n-1}. \tag{3.29}$$

Proof. Let us begin with the first identity; obviously, the right hand side $\langle \det(\lambda \mathbb{I} - X) \rangle_n$ is a degree n polynomial in λ, and so it suffices to prove that

$$\int_{\mathbb{R}} \langle \det(\lambda \mathbb{I} - X) \rangle_n \lambda^k e^{nV(\lambda)} d\lambda = 0, \qquad k = 0, ..., n-1. \tag{3.30}$$

By integrating out angular variables, we have that

$$\begin{aligned} \langle \det(\lambda \mathbb{I} - X) \rangle_n &= \frac{1}{Z'} \int_{\mathbb{R}^n} \prod_{i=1}^{n} (\lambda - \lambda_i) \Delta(\lambda_1, ..., \lambda_n)^2 e^{-n \sum_{i=1}^{n} V(\lambda_i)} \prod_{i=1}^{n} d\lambda_i \\ &= \frac{1}{Z'} \int_{\mathbb{R}^n} \Delta(\lambda_1, ..., \lambda_n) \Delta(\lambda_1, ..., \lambda_n; \lambda) e^{-n \sum_{i=1}^{n} V(\lambda_i)} \prod_{i=1}^{n} d\lambda_i, \end{aligned}$$

where $\Delta(\lambda_1, ..., \lambda_n; \lambda)$ denotes the Vandermonde determinant in the variables $\lambda_1, ..., \lambda_n$ and λ. Thus, for any fixed $0 \leq k \leq n-1$, labelling $\lambda \equiv \lambda_{n+1}$:

$$\begin{aligned}\int_{\mathbb{R}} \langle \det(\lambda \mathbb{I} - X) \rangle_n \lambda^k e^{-nV(\lambda)} d\lambda &= \frac{1}{Z'} \int_{\mathbb{R}^n} \int_{\mathbb{R}} \Delta(\lambda_1, ..., \lambda_n; \lambda) \lambda^k \Delta(\lambda_1, ..., \lambda_n) e^{-n\sum_{i=1}^n V(\lambda_i)} d\lambda \prod_{i=1}^n d\lambda_i \\ &= \frac{1}{n+1} \sum_{\ell=1}^{n+1} \int_{\mathbb{R}^{n+1}} \Delta(\lambda_1, ..., \lambda_n, \lambda_{n+1}) \Delta(\lambda_1, ..., \lambda_n, \lambda_\ell^k) e^{-n\sum_{i=1}^n V(\lambda_i)} \prod_{k=1}^{n+1} d\lambda_k \\ &= \frac{1}{n+1} \int_{\mathbb{R}^{n+1}} \Delta(\lambda_1, ..., \lambda_n, \lambda_{n+1}) \det P(\lambda_1, ..., \lambda_{n+1}) e^{-n\sum_{i=1}^n V(\lambda_i)} \prod_{k=1}^{n+1} d\lambda_k,\end{aligned}$$

where P is the matrix

$$P(\lambda_1, ..., \lambda_{n+1}) = \begin{pmatrix} 1 & \lambda_1 & \dots & \lambda_1^{n-1} & \lambda_1^k \\ 1 & \lambda_2 & \dots & \lambda_2^{n-1} & \lambda_2^k \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & \lambda_{n+1} & \dots & \lambda_{n+1}^{n-1} & \lambda_{n+1}^k \end{pmatrix},$$

which vanishes identically when $0 \leq k \leq n-1$. This uniquely fixes $\langle \det(\lambda \mathbb{I} - X) \rangle_n$ as the monic orthogonal polynomial with respect to the weight $e^{-nV(\lambda)}$. A similar calculation yields that $K_n(\lambda, \mu) = \langle \det(\lambda \mathbb{I} - X) \det(\mu \mathbb{I} - X) \rangle_{n-1}$. $\square$

3.3 The 2-Matrix Model and Other Generalizations

In this section, we consider the 2-matrix model, and its connection to the theory of biorthogonal polynomials. Let $n \geq 1$, and suppose $V(X), W(Y)$ are monic polynomials of even degree. The 2-matrix model is defined by the following probability measure on $\mathcal{H}_n \times \mathcal{H}_n$:

$$d\mathbb{P}(X, Y) = Z^{-1} \exp\{n \operatorname{tr}[\tau XY - V(X) - W(Y)]\} dX dY, \tag{3.31}$$

where $0 < \tau < 1$, and dX, dY denote the Haar measures on the space of $n \times n$ Hermitian matrices. The normalization constant Z is again called the partition function, and is given by the integral

$$Z = \iint \exp\{n \operatorname{tr}[\tau XY - V(X) - W(Y)]\} dX dY. \tag{3.32}$$

Expected values in this matrix ensemble of functions $f(X, Y) : \mathcal{H}_n \times \mathcal{H}_n \to \mathbb{C}$ are defined by

$$\langle f(X, Y) \rangle_n := \frac{1}{Z} \iint f(X, Y) \exp\{n \operatorname{tr}[\tau XY - V(X) - W(Y)]\} dX dY \tag{3.33}$$

We will again be interested in expected values of *class functions*, i.e., functions which are invariant under unitary conjugations:

$$f(U_1^\dagger X U_1, U_2^\dagger Y U_2) = f(X, Y), \tag{3.34}$$

for any $U_1, U_2 \in U(n)$. Consequentially, the function f depends only on the eigenvalues of X and Y, which we will denote by $\{x_i\}$, $\{y_i\}$, respectively. We will use the same sort of abuse of notation as in the previous section, and write $f(X,Y) = f(x_1, ..., x_n, y_1, ..., y_n)$ for class functions. From our previous discussion on the 1-matrix model, it is clear that the measures dX, dY factorize into eigenvalue and angular components:

$$dXdY = \prod_{i<j}(x_i - x_j)^2 \prod_{i<j}(y_i - y_j)^2 \prod_{i=1}^{n} dx_i dy_i dU_1 dU_2, \tag{3.35}$$

where U_1, U_2 are invariant measures on independent copies of the unitary group. However, it is now far less obvious that the measure on the eigenvalues is still independent of the measure on the eigenvectors, due to the presence of the term $\mathrm{tr}(XY)$ in the exponent. Let us try to write the expected value of a class function $f(X,Y)$ as an integral only on eigenvalue variables. Set $X = U_1^\dagger D_X U_1$, $Y = U_2^\dagger D_Y U_2$, where $D_X = \text{diag}\,(x_1, ..., x_n)$, and $D_Y = \text{diag}\,(y_1, ..., y_n)$. We abbreviate the Vandermonde factors $\Delta(x_1, ..., x_n)$, $\Delta(y_1, ..., y_n)$ as Δ_X and Δ_Y, respectively. We then have that

$$\langle f(X,Y)\rangle_n = \frac{1}{Z}\iint f \cdot \exp\{-n\sum_{i=1}^{n} V(x_i) + W(y_i)\}\Delta_X^2\Delta_Y^2 \exp\{n\tau\,\mathrm{tr}\,XY\}\prod_{i=1}^{n} dx_i dy_i dU_1 dU_2, \tag{3.36}$$

where dU_i here are invariant measures on the unitary group $U(n)/U_{diag}(n)$. Now, we can rewrite

$$\exp\{n\tau\,\mathrm{tr}\,XY\} = \exp\{n\tau\,\mathrm{tr}(U_1^\dagger D_X U_1)(U_2^\dagger D_Y U_2)\} = \exp\{n\tau\,\mathrm{tr}(U_2 U_1^\dagger D_X U_1 U_2^\dagger D_Y)\} \tag{3.37}$$

Making the change of variables (in U_1, say) to $\tilde{U} = U_1 U_2^\dagger$, and using the invariance of the Haar measure (i.e., $dU_1 = d\tilde{U}$), we have that

$$\begin{aligned}\langle f(X,Y)\rangle_n &= \frac{1}{Z}\iint f \cdot e^{-n\sum_{i=1}^{n}[V(x_i)+W(y_i)]}\Delta_X^2\Delta_Y^2 e^{n\tau\,\mathrm{tr}(\tilde{U}^\dagger D_X \tilde{U} D_Y)} d\tilde{U} dU_2 \prod_{i=1}^{n} dx_i dy_i \\ &= \frac{1}{Z'}\iint_{\mathbb{R}^n\times\mathbb{R}^n} f \cdot e^{-n\sum_{i=1}^{n}[V(x_i)+W(y_i)]}\Delta_X^2\Delta_Y^2 \int e^{n\tau\,\mathrm{tr}(\tilde{U}^\dagger D_X \tilde{U} D_Y)} d\tilde{U} \prod_{i=1}^{n} dx_i dy_i.\end{aligned} \tag{3.38}$$

From here, we see that the only obstruction to computing expected values in eigenvalue variables is the integral over the unitary group

$$\int e^{n\tau\,\mathrm{tr}(\tilde{U}^\dagger D_X \tilde{U} D_Y)} d\tilde{U}. \tag{3.39}$$

In its current form, the dependence on the eigenvalue coordinates $\{x_i\}$, $\{y_i\}$ is unclear. However, we are in luck: this integral is well-studied in the random matrix community, and has been explicitly evaluated for us by a number of sources [62, 80, 112]. It is called the *Itzykson-Zuber* or *Harish-Chandra* integral; we reproduce its evaluation in the following proposition:

Proposition 3.12. *(Itzykson-Zuber integral over the unitary group). Set*

$$I(D_X, D_Y; \tau) := \int e^{n\tau\, \mathrm{tr}(\tilde{U}^\dagger D_X \tilde{U} D_Y)} d\tilde{U}. \tag{3.40}$$

Then,

$$I(D_X, D_Y; \tau) = (n\tau)^{-n(n-1)/2} \frac{\det(e^{n\tau x_i y_i})}{\Delta_X \Delta_Y}. \tag{3.41}$$

Proof. Consider the heat equation on the space of $n \times n$ Hermitian matrices $\mathcal{H}_n$:

$$\left(\frac{\partial}{\partial t} - \frac{1}{2}\left[\sum_{i=1}^{n} \frac{\partial^2}{\partial X_{ii}^2} + \frac{1}{2}\sum_{i<j}\left(\frac{\partial^2}{\partial \mathrm{Re}(X_{ij})^2} + \frac{\partial^2}{\partial \mathrm{Im}(X_{ij})^2}\right)\right]\right) f(X;t) = 0.$$

Since $\mathcal{H}_n \cong \mathbb{R}^{n^2}$, we can easily write down the heat kernel for the above equation:

$$K(X,Y;t) = \left(\frac{n}{2\pi t}\right)^{n^2/2} \exp\left[-\frac{n}{2t}\,\mathrm{tr}(X-Y)^2\right].$$

In other words, $K(X,Y;t)$ satisfies the PDE (3.3), subject to the boundary condition

$$K(X,Y;t) \to \delta(X-Y) \tag{3.42}$$

Let us consider a solution to (3.3) with class-function initial data $f(X) = f(U^\dagger X U)$, for any $U \in U(n)$. By the convolution theorem, $u(X,t) = (f * u)(X,t)$, where $*$ here denotes convolution. Evaluating this convolution, and switching over to eigenvalue variables, $X = U_1^\dagger D_X U_1$, $Y^\dagger = U_2^\dagger D_Y U_2$, we obtain that

$$\begin{aligned} u(X,t) =& \frac{1}{(2\pi t)^{n^2/2}} \int_{\mathbb{R}^n} \Delta_Y^2 f(Y) \exp\left(-\frac{n}{2t}\mathrm{tr}(X^2+Y^2)\right) \prod_{i=1}^{n} dy_i \\ &\times \int_{U(n)/U_{diag}(n)} dW \exp\left(\frac{1}{t}\mathrm{tr} XWYW^\dagger\right), \end{aligned} \tag{3.43}$$

where $W = U_1 U_2^\dagger$. Note that the inner integral is the object we want to compute. Furthermore, the above formula implies that if the initial data satisfies $f(X) = f(U^\dagger X U)$, for any $U \in U(n)$, then $u(X) = u(U^\dagger X U, t)$ for all $t > 0$ as well. Let us see examine the above equation in eigenvalue coordinates. Now, a classical formula from Riemannian Geometry (cf. [22], for example) tells us that, on a Riemannian manifold M with metric g, the Laplace-Beltrami operator is given in local coordinates $(z^1, ..., z^m)$ by the formula

$$\hat{D}_{LB} f = \frac{1}{\sqrt{\det g}} \sum_{i=1}^{m} \frac{\partial}{\partial z_i} \left(\sqrt{\det g} \sum_{j=1}^{m} g^{ij} \frac{\partial f}{\partial z_j} \right)$$

With the help of this formula, and our formula for the metric on $\mathcal{H}_n$ in eigenvalue coordinates (3.12), we can derive a formula satisfied by $u(X,t)$ in the eigenvalue coordinates:

$$\begin{aligned} 0 &= \left(\frac{\partial}{\partial t} - \frac{1}{2} \hat{D}_{\mathcal{H}_n} \right) u(X,t) = \left(\frac{\partial}{\partial t} - \frac{1}{2} \hat{D}_{eigenvalues} \right) u(X,t) \\ &= \frac{\partial u}{\partial t} - \frac{1}{2} \frac{1}{\Delta_X^2} \sum_{i=1}^{n} \frac{\partial}{\partial x_i} \left(\Delta_X^2 \frac{\partial u}{\partial x_i} \right) \\ &= \frac{\partial u}{\partial t} - \frac{1}{2} \frac{1}{\Delta_X^2} \sum_{i=1}^{n} \left(2\Delta_X \frac{\partial \Delta_X}{\partial x_i} \frac{\partial u}{\partial x_i} + \Delta_X^2 \frac{\partial^2 u}{\partial x_i^2} \right); \end{aligned}$$

Multiplying the above equation by the Vandermonde determinant Δ_X, and using the fact that $\sum_{i=1}^{n} \frac{\partial^2 \Delta_X}{\partial x_i^2} = 0$, the above equation becomes

$$\begin{aligned} 0 &= \frac{\partial \Delta_X u}{\partial t} - \frac{1}{2} \sum_{i=1}^{n} \left(2 \frac{\partial \Delta_X}{\partial x_i} \frac{\partial u}{\partial x_i} + \Delta_X \frac{\partial^2 u}{\partial x_i^2} \right) \\ &= \frac{\partial \Delta_X u}{\partial t} - \frac{1}{2} \sum_{i=1}^{n} \left(\underbrace{u \frac{\partial^2 \Delta_X}{\partial x_i^2}}_{=0} + 2 \frac{\partial \Delta_X}{\partial x_i} \frac{\partial u}{\partial x_i} + \Delta_X \frac{\partial^2 u}{\partial x_i^2} \right) \\ &= \left(\frac{\partial}{\partial t} - \frac{1}{2} \sum_{i=1}^{n} \frac{\partial^2}{\partial x_i^2} \right) \Delta_X u(X,t), \end{aligned}$$

and so $\Delta_X u(X,t)$ satisfies a heat equation of a different form. The generic solution for initial data $g(x_1, ..., x_n)$ to the above heat equation is

$$v(x_1, ..., x_n, t) = \int_{\mathbb{R}^n} \frac{1}{(2\pi t)^{n/2}} \exp\left(-\frac{1}{2t} \sum_{i=1}^{n} (x_i - y_i)^2 \right) g(y_1, ..., y_n) \prod_{i=1}^{n} dy_i;$$

In particular, for initial data $g(X) = \Delta_X f(X)$, which depends only on the eigenvalues of X, we obtain the formula

$$\Delta_X u(X,t) = \int_{\mathbb{R}^n} \frac{1}{(2\pi t)^{n/2}} \exp\left(-\frac{1}{2t} \sum_{i=1}^{n} (x_i - y_i)^2 \right) \Delta_Y f(Y) \prod_{i=1}^{n} dy_i \tag{3.44}$$

multiplying the previous equation by Δ_X, we see that we have two expressions for the same function $\Delta_X u(X,t)$, given by equations (3.43) and (3.44). Equating these expressions yields

$$\Delta_X \frac{1}{(2\pi t)^{n^2/2}} \int_{\mathbb{R}^n} dY \Delta_Y^2 f(Y) \exp\left(-\frac{1}{2t}\mathrm{tr}(X^2+Y^2)\right) \int_{U(n)/U_{diag}(n)} dW \exp\left(\frac{1}{t}\mathrm{tr}XWYW^\dagger\right)$$
$$= \int_{\mathbb{R}^n} \frac{1}{(2\pi t)^{n/2}} \exp\left(-\frac{1}{2t}\sum_{i=1}^n (x_i-y_i)^2\right) \Delta_Y f(Y) \prod_{i=1}^n dy_i$$

Setting $f(Y) = \prod_{i=1}^n \delta(Y-Y_0)$, we obtain that

$$\Delta_X \Delta_{Y_0}^2 \frac{1}{(2\pi t)^{n^2/2}} \exp\left(-\frac{1}{2t}\mathrm{tr}(X^2+Y_0^2)\right) \int_{U(n)/U_{diag}(n)} dW \exp\left(\frac{1}{t}\mathrm{tr}XWY_0W^\dagger\right)$$
$$= \frac{1}{(2\pi t)^{n/2}} \exp\left(-\frac{1}{2t}\mathrm{tr}(X-Y_0)^2\right) \Delta_{Y_0}.$$

Rearranging, and relabelling $Y_0 = Y$, we get

$$\int_{U(n)/U_{diag}(n)} dW \exp\left(\frac{1}{t}\mathrm{tr}XWYW^\dagger\right) = (2\pi t)^{-n(n-1)/2} \frac{\exp(\frac{1}{t}\mathrm{tr}XY)}{\Delta_Y \Delta_X};$$

finally,

$$\int_{U(n)/U_{diag}(n)} dW \exp\left(\frac{1}{t}\mathrm{tr}\ XWYW^\dagger\right) = (2\pi t)^{-n(n-1)/2} \frac{\det e^{\frac{1}{t}x_i y_j}}{\Delta_X \Delta_Y},$$

as desired. □

With the explicit evaluation of this integral completed, we see that the expected value of the class function $f(X,Y)$ can now be expressed explicitly in eigenvalue coordinates:

Corollary 3.13. *For any class function $f : \mathcal{H}_n \times \mathcal{H}_n \to \mathbb{C}$,*

$$\langle f(X,Y)\rangle_n = \frac{1}{Z''} \iint f \cdot \Delta_X \Delta_Y e^{n\sum_{i=1}^n [\tau x_i y_i - V(x_i) - W(y_i)]} \prod_{i=1}^n dx_i dy_i, \tag{3.45}$$

where Z'' is related to Z by the equation

$$Z = Z''\mathit{vol}\ (U(n)/U_{diag}(n)) \cdot (n\tau/2)^{-n(n-1)/2} \tag{3.46}$$

Just as the 1-matrix model was intimately connected with orthogonal polynomials, we shall see that the 2-matrix model is intimately connected with *biorthogonal polynomials*.

3.4 Biorthogonal Polynomials: A Special Case of Multiple Orthogonality

Let V, W be monic polynomials of even degree, $0 < \tau < 1$. The monic biorthogonal polynomials $\{p_k(x)\}$, $\{q_k(y)\}$ with respect to the weights V, W are defined by the relation

$$\int_{\mathbb{R}} \int_{\mathbb{R}} p_k(x) q_j(y) e^{\tau xy - V(x) - W(y)} dxdy = h_j \delta_{kj}. \tag{3.47}$$

Biorthogonal polynomials enjoy many of the properties that classical orthogonal polynomials satisfy, such as finite-term recursion relations, possession of a closed-form expression for the Christoffel-Darboux kernel, moment-determinant formulas for the partition function, and so on (see [44] and references therein, for example). However, the tools available to study the asymptotics of orthogonal polynomials, in particular, a Riemann-Hilbert formulation for the polynomials (which we shall discuss in Chapter 4), is not immediately present. Suppose $\deg V = v$, $\deg W = q$. Then we have the following theorem realizing biorthogonal polynomials as a special case of multiple orthogonality [75]:

Theorem 3.14. *(A.B.J. Kuijlaars and K.T-R. McLaughlin) Set* $D := \deg W - 1 = q - 1$. *Let* $k \geq 1$*, and consider the monic biorthogonal polynomial* $p_k(x)$ *of degree* k. *Let* $n_k = \left[\frac{k+j}{D}\right]$, $j = 0, ..., D$, *where* $[\cdot]$ *denotes the integral part. Then, we have that*

$$\int_{\mathbb{R}} p_k(x) x^i w_j(x) dx = 0, \qquad \textit{for } i = 0, ..., n_j - 1,\ j = 0, ..., D, \tag{3.48}$$

where the weights $w_j(x)$ *are defined to be*

$$w_j(x) := \int_{\mathbb{R}} y^j e^{\tau xy - V(x) - W(y)} dy. \tag{3.49}$$

Moreover, these conditions uniquely determine $p_k(x)$ *among all monic polynomials of degree* k.

There is of course an equivalent proposition characterizing the polynomials $q_k(y)$. This theorem is the key that allows us to analyze the asymptotics of biorthogonal polynomials, as there is a well-established literature on the asymptotics of multiple orthogonal polynomials [1, 4, 85], as well as a Riemann-Hilbert formulation of such polynomials [5], which we will discuss in Chapter 4. We now proceed to the proof of the theorem.

Proof. The proof essentially follows from an integration by parts argument. Since $\deg W = q$, we have that, for any fixed $j = 0, ..., q-1$, setting $D := q - 1$,

$$\frac{d^i}{dy^i}\left[y^j e^{-W(y)}\right] = \pi_{Di+j}(y)e^{-W(y)}, \tag{3.50}$$

where $\pi_{Di+j}(y)$ is a polynomial of degree exactly $Di + j$. Now, for any function $g(x)$, integrating by parts k times (and using the fact that boundary terms cancel), we find:

$$\begin{aligned}
&\iint g(x)\pi_{Di+j}(y)e^{\tau xy - V(x) - W(y)}dxdy \\
&\quad = \int g(x)e^{-V(x)} \int \frac{d^i}{dy^i}\left[y^j e^{-W(y)}\right] e^{\tau xy}dydx \\
&\quad = (-1)^i \int g(x)e^{-V(x)} \int y^j e^{-W(y)} \frac{d^i}{dy^i}\left[e^{\tau xy}\right] dydx \\
&\quad = (-\tau)^i \int g(x)x^i e^{-V(x)} \int y^j e^{-W(y)+\tau xy}dydx \\
&\quad = (-\tau)^i \int g(x)x^i w_j(x)dx.
\end{aligned}$$

Now, taking g to be the monic biorthogonal polynomial $p_k(x)$, we see that the left hand side of the above equation is zero provided $Di + j < k$. This implies the integral is zero for $i < n_j - 1$, for $j = 0, ..., q-1$, where n_j are as defined in the theorem; thus, this implies that the biorthogonal polynomial $p_k(x)$ satisfies

$$0 = \int p_k(x)x^i w_j(x)dx, \qquad \textit{for } i = 0, ..., n_j - 1,\ j = 0, ..., q-1.$$

Conversely, suppose that g is a monic polynomial of degree k, satisfying the relations (3.48). Then, the left hand side of the equation for g vanishes for $i < n_j - 1$, $j = 0, ..., q-1$. On the other hand, the polynomials π_{Di+j} for $i < n_j - 1$, $j = 0, ..., q-1$ form a basis for the space of polynomials of degree $\leq k-1$; it follows that g must be the monic biorthogonal polynomial $p_k(x)$. ☐

3.5 Matrix Models and Graph Combinatorics.

We now discuss the elegant connection between the Hermitian matrix models and the combinatorics of planar graphs. Most of what we will say here is not new; we refer to the excellent surveys [49, 113] for further details about these connections. We begin with the more widely known connection of the 1-matrix model to the combinatorics of planar (more generally, genus g) graphs. We will then proceed to discuss the 2-matrix model, and its relation to counting the number of 2-colored graphs. This connection allows us to link the discussion of the 2-matrix model to the Ising model on *random* graphs.

3.5.1 Combinatorics of the 1-Matrix Model.

Our starting point will be the Gaussian Unitary ensemble: we will compute a generating function for its moments.

Proposition 3.15. *Let $J \in \mathcal{H}_n$ be a fixed Hermitian matrix, and consider the quantity*

$$f(J) := \langle \exp(\operatorname{tr} JX) \rangle_n. \tag{3.51}$$

Then,

$$f(J) = \exp\left[\frac{1}{2N} \operatorname{tr} J^2\right]. \tag{3.52}$$

Proof. The proof of this proposition is a direct computation, which follows by "completing the square", and utilizing the cyclicity of trace:

$$\begin{aligned}
\langle \exp(\operatorname{tr} JX) \rangle_n &= \frac{1}{Z} \int_{\mathcal{H}_n} \exp\left[-\frac{N}{2} \operatorname{tr}\left(X^2 - \frac{2}{N} JX\right)\right] dX \\
&= \frac{\exp(\frac{1}{2N} \operatorname{tr} J^2)}{Z} \int_{\mathcal{H}_n} \exp\left[-\frac{N}{2} \operatorname{tr}\left(X - \frac{1}{N} J\right)^2\right] dX \\
&= \frac{\exp(\frac{1}{2N} \operatorname{tr} J^2)}{Z} \int_{\mathcal{H}_n} \exp\left[-\frac{N}{2} \operatorname{tr}\left(X - \frac{1}{N} J\right)^2\right] d(X - \frac{1}{N} J) \\
&= \exp\left[\frac{1}{2N} \operatorname{tr} J^2\right],
\end{aligned}$$

where in the last line we have used the translation invariance of the Haar measure on $\mathcal{H}_n$. □

It follows at once that we can compute any moment of the GUE, as $f(J)$ acts as a generating function. This is made precise by the following corollary:

Corollary 3.16. *For any $1 \le i, j, k, \ell \le n$, if X_{ij}, $X_{k\ell}$ are entries of a GUE matrix, then*

$$\langle X_{ij} \rangle_n = 0, \qquad \langle X_{ij} X_{k\ell} \rangle_n = \frac{1}{N} \delta_{i\ell} \delta_{jk}. \tag{3.53}$$

Proof. Notice that

$$\left.\frac{\partial f}{\partial J_{ji}}\right|_{J=0} = \langle X_{ij} \rangle_n, \qquad \left.\frac{\partial^2 f}{\partial J_{ji} \partial J_{\ell k}}\right|_{J=0} = \langle X_{ij} X_{k\ell} \rangle.$$

Calculating the derivatives explicitly yields the result. Note that the factor of two is accounted for because the matrix J is Hermitian. □

We could have already guessed the formulae (3.53) from the fact that the GUE measure (3.6) is made up of Gaussians. Indeed, this is essentially the trick that we have used: we have calculated the moment generating function of a collection of Gaussian variables. In fact, since the entries are Gaussian, we actually have enough information here to calculate any higher order moment as a polynomial in the moments (3.53). This result is known as Wick's theorem [1]:

Theorem 3.17. *Let $\phi_1, ..., \phi_N$ be a collection of centered normal random variables. Then,*

$$\left\langle \prod_{i=1}^{N} \phi_i \right\rangle = \sum_{\pi \in \Pi_N} \prod_{(a,b)\in\pi} \langle \phi_a \phi_b \rangle, \tag{3.54}$$

where Π_N denotes the set of all pairings of the set $1, ..., 2N$.

Some remarks are in order.

Remark 3.18. It is worth developing some intuition for the set of all pairings, Π_N. If $N = 4$, for example, then we have that

$$\Pi_4 = \Big\{ \{\{1,2\},\{3,4\}\}, \{\{1,4\},\{2,3\}\}, \{\{1,3\},\{2,4\}\} \Big\}. \tag{3.55}$$

So, for example, if $\phi_1, \phi_2, \phi_3, \phi_4$ are centered normal random variables, then

$$\langle \phi_1\phi_2\phi_3\phi_4 \rangle = \langle \phi_1\phi_2 \rangle\langle \phi_3\phi_4 \rangle + \langle \phi_1\phi_4 \rangle\langle \phi_2\phi_3 \rangle + \langle \phi_1\phi_3 \rangle\langle \phi_2\phi_4 \rangle. \tag{3.56}$$

In fact, the theorem is still valid if we consider repetitions of the random variables ϕ_k: in this case, we treat the set of indices as a multiset, and apply Wick's theorem. As an example of this, we have that

$$\begin{aligned} \langle \phi_1^2\phi_2\phi_3 \rangle &= \langle \phi_1^2 \rangle\langle \phi_2\phi_3 \rangle + \langle \phi_1\phi_3 \rangle\langle \phi_1\phi_2 \rangle + \langle \phi_1\phi_2 \rangle\langle \phi_1\phi_3 \rangle \\ &= \langle \phi_1^2 \rangle\langle \phi_2\phi_3 \rangle + 2\langle \phi_1\phi_2 \rangle\langle \phi_1\phi_3 \rangle. \end{aligned} \tag{3.57}$$

Furthermore, if $N = 2M+1$ is odd, then it is apparent that Π_N is empty, and the expected value $\left\langle \prod_{i=1}^{N} \phi_i \right\rangle = 0$. In general, there are $\frac{(2N)!}{2^N (N)!}$ possible pairings of the set $\{1, ..., N\}$.

The above theorem is useful in that it shows us that, in order to compute *any* moment of a collection of Gaussian variables, $\{\phi_a\}$, it is enough to know their covariances $\langle \phi_a \phi_b \rangle$. In our situation, the collection of Gaussian variables of interest are the matrix variables $\{X_{ij}\}$, which, by our previous observations, satisfy the hypotheses of Wick's theorem. We can therefore compute any expected value of moments of interest.

[1] The theorem is named after its popularizer, the Italian physicist Gian Carlo Wick, who rediscovered and popularized the theorem in theoretical physics [103]. The theorem in fact dates back to a 1918 paper of the Russian probabilist Leon Isserlis [61].

We will sometimes call the process of rewriting of an expected value of products of centered Gaussians in terms of pairings *Wick expansion.*

Our main focus will be on expected values of products of traces of the matrix variables, for example $\langle \operatorname{tr} X^4 \rangle_n$, or $\langle (\operatorname{tr} X^3)^2 \rangle_n$. From here on, we will take the parameter N in the GUE to be $N := n$, the size of the matrix.

As a first example let us compute $\langle \operatorname{tr} X^4 \rangle_n$ using Wick's theorem, and Equations (3.53).

Example 3.19. $\langle \operatorname{tr} X^4 \rangle_n$. First, we have that

$$\langle \operatorname{tr} X^4 \rangle_n = \sum_{i_1,i_2,i_3,i_4=1}^{n} \langle X_{i_1 i_2} X_{i_2 i_3} X_{i_3 i_4} X_{i_4 i_1} \rangle;$$

using Wick's theorem on the index set $\{i_1, i_2, i_3, i_4\}$, we find that

$$\begin{aligned}\langle \operatorname{tr} X^4 \rangle_n &= \sum_{i_1,i_2,i_3,i_4=1}^{n} \langle X_{i_1 i_2} X_{i_2 i_3} X_{i_3 i_4} X_{i_4 i_1} \rangle \\ &= \sum_{i_1,i_2,i_3,i_4=1}^{n} \Big(\langle X_{i_1 i_2} X_{i_2 i_3} \rangle \langle X_{i_3 i_4} X_{i_4 i_1} \rangle + \langle X_{i_1 i_2} X_{i_4 i_1} \rangle \langle X_{i_2 i_3} X_{i_3 i_4} \rangle + \langle X_{i_1 i_2} X_{i_3 i_4} \rangle \langle X_{i_2 i_3} X_{i_4 i_1} \rangle \Big).\end{aligned}$$

Using Equation (3.53), we obtain that

$$\begin{aligned}\langle \operatorname{tr} X^4 \rangle_n &= \sum_{i_1,i_2,i_3,i_4=1}^{n} \frac{1}{n^2} \delta_{i_1 i_3} + \sum_{i_1,i_2,i_3,i_4=1}^{n} \frac{1}{n^2} \delta_{i_2 i_4} + \sum_{i_1,i_2,i_3,i_4=1}^{n} \frac{1}{n^2} \delta_{i_2 i_3} \delta_{i_1 i_4} \delta_{i_3 i_4} \delta_{i_1 i_2} \\ &= n + n + \frac{1}{n} = 2n + \frac{1}{n}.\end{aligned}$$

So, two of the pairings from the Wick formula contributed a factor of n, whereas the remaining term contributed a factor of $\frac{1}{n}$. Before trying to calculate things in general, let's consider another example: $\langle (\operatorname{tr} X^3)^2 \rangle_n$.

Example 3.20. $\langle (\operatorname{tr} X^3)^2 \rangle_n$. In this case, we can write

$$\begin{aligned}\langle (\operatorname{tr} X^3)^2 \rangle_n &= \left\langle \left(\sum_{i_1,i_2,i_3=1}^{n} X_{i_1 i_2} X_{i_2 i_3} X_{i_3 i_1} \right)^2 \right\rangle_n \\ &= \sum_{i_1,i_2,i_3=1}^{n} \sum_{j_1,j_2,j_3=1}^{n} \langle X_{i_1 i_2} X_{i_2 i_3} X_{i_3 i_1} X_{j_1 j_2} X_{j_2 j_3} X_{j_3 j_1} \rangle_n.\end{aligned}$$

Wick's theorem tells us that there are $\frac{6!}{2^3 3!} = 15$ contributing pairings to the expected value; let us compute a few of these pairings. For example, the pairing

$$\pi := \Big\{ \{(i_1, i_2), (i_2, i_3)\}, \{(i_3, i_1), (j_1, j_2)\}, \{(j_2, j_3), (j_3, j_1)\} \Big\}$$

contributes a factor of

$$\begin{aligned}
&\sum_{i_1,i_2,i_3=1}^{n} \sum_{j_1,j_2,j_3=1}^{n} \langle X_{i_1 i_2} X_{i_2 i_3} \rangle_n \langle X_{i_3 i_1} X_{j_1 j_2} \rangle_n \langle X_{j_2 j_3} X_{j_3 j_1} \rangle_n \\
&= \sum_{i_1,i_2,i_3=1}^{n} \sum_{j_1,j_2,j_3=1}^{n} \left(\frac{1}{n} \delta_{i_1 i_3} \right) \left(\frac{1}{n} \delta_{i_3 j_2} \delta_{i_1 j_1} \right) \left(\frac{1}{n} \delta_{j_2 j_1} \right) = 1
\end{aligned}$$

to the sum; on the other hand the pairing

$$\pi' := \Big\{ \{(i_1, i_2), (j_1, j_2)\}, \{(i_2, i_3), (j_2, j_3)\}, \{(i_3, i_1), (j_3, j_1)\} \Big\}$$

contributes a factor of

$$\begin{aligned}
&\sum_{i_1,i_2,i_3=1}^{n} \sum_{j_1,j_2,j_3=1}^{n} \langle X_{i_1, i_2} X_{j_1, j_2} \rangle_n \langle X_{i_2, i_3} X_{j_2, j_3} \rangle_n \langle X_{i_3, i_1} X_{j_3, j_1} \rangle_n \\
&= \sum_{i_1,i_2,i_3=1}^{n} \sum_{j_1,j_2,j_3=1}^{n} \left(\frac{1}{n} \delta_{i_1 j_2} \delta_{i_2 j_1} \right) \left(\frac{1}{n} \delta_{i_2 j_3} \delta_{i_3 j_2} \right) \left(\frac{1}{n} \delta_{i_3 j_1} \delta_{i_1 j_3} \right) = \frac{1}{n^2}.
\end{aligned}$$

Carefully calculating the contributions from the remaining pairings, we find that

$$\langle (\operatorname{tr} X^3)^2 \rangle_n = 12 + \frac{3}{n^2}. \tag{3.58}$$

With the above examples worked out, a pattern starts to emerge. We now develop a diagrammatic approach to calculating expected values of the form

$$\left\langle \prod_{p=1}^{V} \operatorname{tr} X^{i_p} \right\rangle_n \tag{3.59}$$

By Wick's theorem, we can decompose the above expected value into a sum over $D := \sum_{p=1}^{V} i_p$ pairings of the indices of the trace, with each pairing π contributing a factor of

$$\sum_{I} \prod_{\{(j_a, j_b), (j_c, j_d)\} \in \pi} \langle X_{j_a j_b} X_{j_c j_d} \rangle_n, \tag{3.60}$$

where the summation I runs over all of the index sets of the traces $\operatorname{tr} X^{i_p}$, $p = 1, ..., V$, and $\vec{i} := (i_1, ..., i_V)$ is a V-tuple of positive integers. To each such pairing π, we associate the following diagram R_π:

1. Draw V vertices, each with degree $i_1, ..., i_V$, with half-edges labelled by the indices of the trace of the corresponding vertex.

2. Pair the edges according to the pairing prescribed by π.

The resulting diagram is called a *ribbon graph* or *fat graph*. To each such ribbon graph R_π, we can associate a quantity known as the *genus* of the graph. The *genus* of a ribbon graph R is the smallest genus g such that the ribbon graph R can be embedded onto this surface without self-intersection. If the ribbon graph is disjoint, its genus is the sum of the genera of its connected components.

The ribbon graph associated to the pairing π will by construction have V vertices, and $\frac{D}{2}$ edges, by the handshaking lemma (note that, if D is not even, then the full expected value is automatically zero). Each face in the resulting ribbon graph corresponds to a collection of indices which will be identified by Wick's theorem, since $\langle X_{ab} X_{cd} \rangle_n = \frac{1}{n} \delta_{ad} \delta_{bc}$. Thus, the number of independent summation indices I' left over after accounting for the identified indices is the same as the number of faces, which we shall denote by F. The result is that

$$\sum_I \prod_{\{(j_a, j_b), (j_c, j_d)\} \in \pi} \langle X_{j_a j_b} X_{j_c j_d} \rangle_n = \sum_I \frac{1}{n^{D/2}} \prod_{\{(j_a, j_b), (j_c, j_d)\} \in \pi} \delta_{j_a j_d} \delta_{j_b j_c} = \sum_{I'} \frac{1}{n^{D/2}} = n^{F - D/2}. \tag{3.61}$$

Now, since the number of edges is $E := D/2$, and the total number of vertices is V, we see that the contribution of the pairing π (equivalently, the ribbon graph R_π) is

$$n^{\chi(R_\pi) - V}, \tag{3.62}$$

where $\chi(R_\pi) := V - E + F$ is the *Euler characteristic* of the surface on which R_π embeds. If we now sum over all such ribbon graphs, we obtain the following theorem:

Theorem 3.21. *Let X be an $n \times n$ GUE matrix, and let $V \geq 1$, $\vec{i} := (i_1, ..., i_V)$ a V-tuple of positive integers. Then,*

$$n^V \left\langle \prod_{p=1}^{V} \operatorname{tr} X^{i_p} \right\rangle_n = \sum_{g \geq 0} \frac{e_g(\vec{i})}{n^{2g-2}}, \tag{3.63}$$

where $e_g(\vec{i})$ counts the number of genus g ribbon graphs on V vertices, with degrees $(i_1, ..., i_V)$.

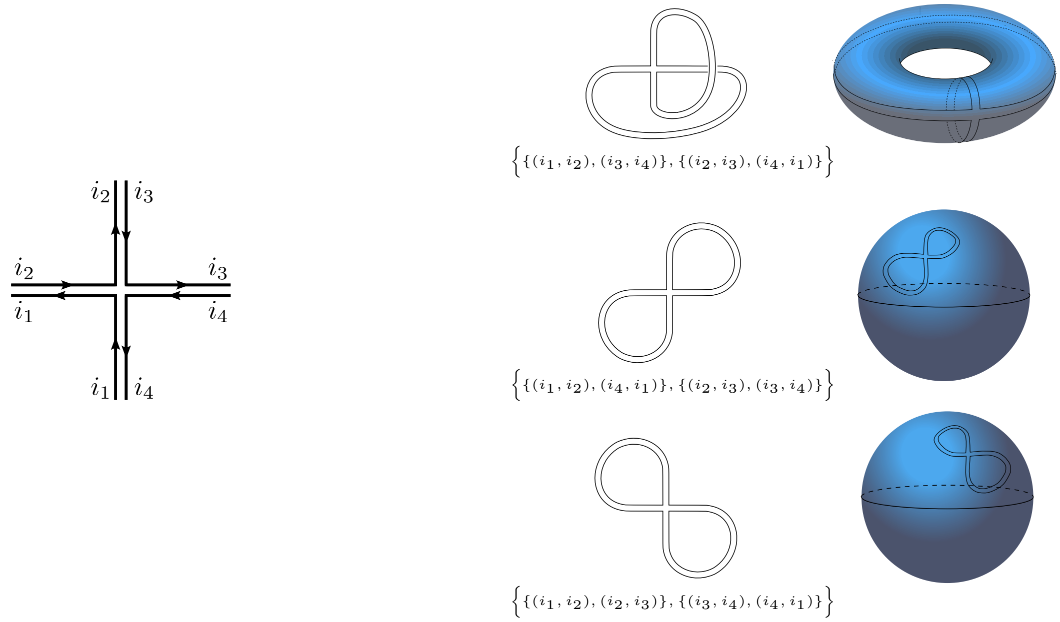


Figure 3.1. *(Left).* The "bare" diagram contributing to $\langle \operatorname{tr} X^4 \rangle_n$; there are four outgoing half-edges from the vertex, corresponding to the four matrix indices of $\operatorname{tr} X^4$ that need to be paired. *(Right).* The three possible pairings of the half-edges, and the corresponding surfaces they embed onto. There are two genus 0 contributions, and one genus 1 contribution.

Proof. As before, let $D := \sum_p i_p$; without loss of generality, assume D is even. The diagrammatic interpretation of Wick's theorem yields that the left hand side of Equation (3.63) is

$$n^V \left\langle \prod_{p=1}^{V} \operatorname{tr} X^{i_p} \right\rangle_n = n^V \sum_{R_\pi} n^{\chi(R_\pi) - V} = \sum_{R_\pi} n^{\chi(R_\pi)},$$

where the sum runs over all $\frac{D!}{2^{D/2}(D/2)!}$ possible ribbon graphs (pairings) of the half edges. The result then follows immediately, due to the relation $\chi = 2 - 2g$. □

Corollary 3.22. *With the notations of the previous theorem,*

$$\sum_{g \geq 0} e_g(\vec{i}) = \frac{D!}{2^{D/2}(D/2)!}. \tag{3.64}$$

One may go back and re-perform the calculations in Examples 3.19 and 3.20, using the diagrammatic techniques explained above. The "bare" diagram for the calculation of $\langle \operatorname{tr} X^4 \rangle_n$, as well as the set of pairings (diagrams), are shown in Figure (3.1). The "bare" diagram for the calculation of $\langle (\operatorname{tr} X^3)^2 \rangle_n$ is shown in Figure (3.2); the 15 contributing diagrams are shown in Table (1).

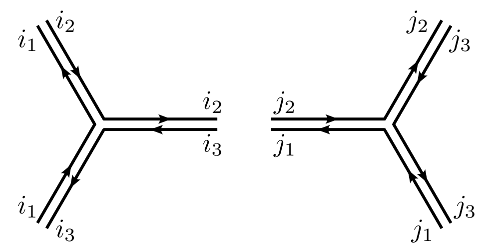


Figure 3.2. The "bare" diagram contributing to $\langle (\operatorname{tr} X^3)^2 \rangle_n$. There are two cubic vertices, one for each copy of $\operatorname{tr} X^3$. A pairing of the indices π results in a corresponding pairing of the edges. The 15 possible pairings of the edges are shown in Table (1).

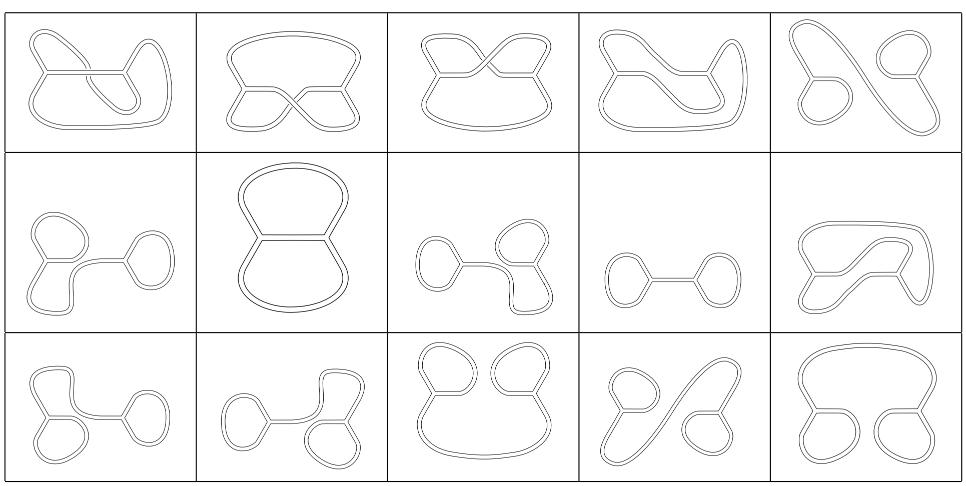


Table 1. The 15 contributing diagrams for the expected value $\langle (\operatorname{tr} X^3)^2 \rangle_n = 12 + \frac{3}{n^2}$. Note that all diagrams are planar, except for the first three: these contribute a factor of $\frac{1}{n^2}$ each, instead of a factor of 1 (as in the planar case).

The above expansion is often called the *genus expansion* or *topological expansion*, for obvious reasons. Often, one is interested in the genus expansion for a particular class of graphs, which admit some "regularity". We will consider the following generating integral:

$$\langle \exp\left[-Nt \operatorname{tr} X^4\right] \rangle_n = \frac{1}{Z_{GUE}} \int_{\mathcal{H}_n} \exp\left\{-N \operatorname{tr}\left[\frac{1}{2}X^2 + tX^4\right]\right\} dX, \tag{3.65}$$

where $t > 0$ is some parameter. Since Z_{GUE} is some explicit constant which we have already calculated, we will drop it from further considerations and consider the integral in (3.65) alone. Historically, this integral is what led to many of the intriguing connections between random matrices and the combinatorics of ribbon graphs [19, 62]. The integral acts as a generating function for 4-regular graphs, in the following sense. Making the change of variables $\tilde{X} = N^{1/2}X$, and relabelling $\tilde{X}$ to X, the integral (3.65) becomes (up to an overall factor of N^{n^2}, which we absorb into Z_{GUE}):

$$Z(t; N) := \int_{\mathcal{H}_n} \exp\left\{-\operatorname{tr}\left[\frac{1}{2}X^2 + \frac{t}{N}X^4\right]\right\} dX \tag{3.66}$$

If we expand $\exp\left[-\frac{t}{N} \operatorname{tr} X^4\right]$ as a perturbation series, we obtain

$$\exp\left[-\frac{t}{N} \operatorname{tr} X^4\right] = \sum_{m=0}^{\infty} \frac{1}{m!}\left(-\frac{t}{N}\right)^m \left(\operatorname{tr} X^4\right)^m. \tag{3.67}$$

Upon taking the expected value, we obtain

$$\begin{aligned} Z(t; N) &= \sum_{m=0}^{\infty} \frac{1}{m!}\left(-\frac{t}{N}\right)^m \langle \left(\operatorname{tr} X^4\right)^m \rangle_n \\ (\textit{Theorem } (3.21)) &= \sum_{m=0}^{\infty} \frac{1}{m!}\left(-\frac{t}{N}\right)^m \sum N^{m+\chi} \\ &= \sum_{m=0}^{\infty} \frac{1}{m!}(-t)^m \sum N^{\chi}, \end{aligned}$$

where the internal sum runs over all ribbon graphs (pairings) contributing to the term $\langle (\operatorname{tr} X^4)^m \rangle_n$, i.e. all 4-regular ribbon graphs with n vertices. The above object is *almost* what we want, but there is one remaining calculation which must be performed. Note that there was no assumption on the connectedness of the diagrams in question. Indeed, in general, the diagrams contributing to $\langle (\operatorname{tr} X^4)^m \rangle_n$ need not be connected; this is first apparent at $m = 2$. In principle, we would like to consider only connected diagrams. We are in luck: it is a commonly known fact in combinatorics that taking logarithms achieves this goal. The general principle is this: if we have an exponential generating function which counts the number of labelled objects, then its logarithm counts the number of connected objects of the same kind. We again refer to R.P. Stanley's book [98], Chapter 5, for further details on this fact. Using this, the result we are interested in is the following:

$$\log Z(t;N) = \sum_{m=1}^{\infty} \frac{(-t)^m}{m!} \sum{}' N^{\chi}, \tag{3.68}$$

where the inner sum $\sum'$ now runs over all *connected* ribbons graphs contributing to the term $\langle (\operatorname{tr} X^4)^m \rangle_n$. Now, the largest Euler characteristic that can appear is $\chi = 2$, coming from the *planar* or *spherical* diagrams; these diagrams contribute a factor of N^2 each. If we divide Equation (3.68) by N^2, then the planar diagrams contribute a factor of 1 each, and all other diagrams contribute a factor subleading in N. Interchanging the order of summation, what we have shown is the following:

Theorem 3.23. *Consider the function*

$$F(t;N) := \frac{1}{N^2} \log Z(t;N) = \frac{1}{N^2} \log \left\langle \exp\left[\frac{t}{N} \operatorname{tr} X^4\right] \right\rangle_n, \tag{3.69}$$

where the average $\langle \cdot \rangle_n$ is taken in the $n \times n$ GUE with parameter N set to 1. Then,

$$F(t;N) = \sum_{g=0}^{\infty} f_g(t), \tag{3.70}$$

where the function $f_g(t)$ is given by

$$f_g(t) = \sum_{m=0}^{\infty} \frac{(-t)^m}{m!} \mathcal{N}_g(m), \tag{3.71}$$

where $\mathcal{N}_g(m)$ is the number of 4-regular, connected ribbon graphs of genus g on m vertices.

The above should be interpreted as an asymptotic expansion, not an actual convergent series. Theorem (3.23) first appeared (without mathematical proof, albeit) in the mid-1970s work of the Saclay school of theoretical physics [19, 62]. Since this time, there has been much work in the mathematical physics community making the above result fully rigorous (cf. [11, 12, 44, 45], and references therein), as well as related results, such as for the "cubic" model, involving powers of $\operatorname{tr} X^3$, which very apparently needs some care in its definition. The tools used to make such results rigorous are tools we have already met: the method of orthogonal polynomials. For instance, for the above model, one would consider the family of polynomials orthogonal with respect to the weight $w(z;t) := \exp\left[\frac{1}{2}z^2 + tz^4\right]$:

$$\int p_n(z;t) p_m(z;t) \exp\left[\frac{1}{2}z^2 + tz^4\right] dz = \delta_{nm}. \tag{3.72}$$

We have been intentionally vague in defining the contour of integration; in principle, it should be taken to be the real line. However, if $t < 0$, the integral is divergent for any choice of polynomials anyways, and so one must make sense of the above by defining an appropriate *analytic continuation* of the polynomials in the variable t. This amounts to redefining the contour of orthogonality. The analysis of these polynomials has been the subject of extensive research [10, 12, 38, 47, 48]. This analysis in turn informs us about the combinatorics of the associated ribbon diagrams. Here, we list some selected results regarding these enumerative problems; the original ideas can likely be attributed to [19]; rigorous proofs of these results can be found in the aforementioned mathematical physics literature.

Theorem 3.24. *1. Define the function $\xi(t)$ as*

$$\xi = \xi(t) := \frac{-1 + \sqrt{1 + 48t}}{24t}. \tag{3.73}$$

For $g = 0, 1, 2$, with the notations of Theorem (3.23),

$$f_0(t) = \frac{1}{24}(\xi - 1)(9 - \xi) - \frac{1}{2}\log \xi, \tag{3.74}$$

$$f_1(t) = \frac{1}{12}\log(2 - \xi), \tag{3.75}$$

$$f_2(t) = \frac{1}{6!}(82 + 21\xi - 3\xi^2)\frac{(1 - \xi)^3}{(2 - \xi)^5} \tag{3.76}$$

2. The numbers $\mathcal{N}_g(m)$ have the following asymptotics

$$\mathcal{N}_g(m) = K_g 48^m m!\, m^{\frac{5g-7}{2}}[1 + O(m^{-1/2})], \qquad m \to \infty, \tag{3.77}$$

For some explicit constants K_g (cf. [12]).

3.5.2 Combinatorics of the 2-Matrix Model

Just as the 1-matrix model is connected to the combinatorics of genus g graphs, the 2-matrix model is connected to the combinatorics of *colored* or *labelled* graphs. This connection is what allows us to make a link between the 2-matrix model and the Ising model on random graphs. This idea was introduced in [70]. Subsequently, these ideas were expanded upon in the theoretical physics literature [17, 18, 25, 26, 34, 56], but no real mathematical treatment of this model has appeared in the literature to date. In this subsection, we discuss some of the basic combinatorial interpretations of the 2-matrix model, in analogy to the previous section. With the right parameter identification, we show that the logarithm of the partition function of the 2-matrix model acts as a generating function for the Ising model on random graphs of a certain kind.

Similarly to before, we begin by considering the *Gaussian* 2-matrix model, which is defined as the following probability measure on $\mathcal{H}_n \times \mathcal{H}_n$:

$$d\mathbb{P}(X,Y) := Z_n^{-1} \exp\left\{N \operatorname{tr}\left[\tau XY - \frac{1}{2}X^2 - \frac{1}{2}Y^2\right]\right\} dXdY, \tag{3.78}$$

where $Z_n := Z_n(\tau, N)$ is a normalization constant, chosen so that $d\mathbb{P}(X,Y)$ is indeed a probability measure. By writing out the above measure in the matrix variables, we again see that all of the entries Re X_{ij}, Im X_{ij}, Re Y_{ij}, Im Y_{ij} are centered Gaussian random variables. What makes the 2-matrix model more interesting structurally than the 1-matrix model is the fact that the X and Y entries have nontrivial covariance. We summarize this idea in the next proposition, in which we compute the moment generating function for the measure (3.78).

Proposition 3.25. *Let $J, K \in \mathcal{H}_n$ be a fixed Hermitian matrices, and consider the quantity*

$$f(J,K) := \langle \exp(\operatorname{tr} JX + KY)\rangle_n, \tag{3.79}$$

where the expected value is taken with respect to the measure (3.78). Then,

$$f(J,K) = \exp\left\{\frac{1}{N(1-\tau^2)} \operatorname{tr}\left[\frac{1}{2}J^2 + \frac{1}{2}K^2 - \tau JK\right]\right\}. \tag{3.80}$$

Proof. The proof is similar in technique to Proposition (3.15), so we only sketch the proof here, and omit further details. The essence of the proof is to again "complete the square", except that now we deal with a 'vector' of Hermitian matrices, $\vec{X} := (X,Y)^T$, and write the measure (3.78) as

$$d\mathbb{P}(X,Y) = Z_n^{-1} \exp\left\{\frac{N}{2} \operatorname{tr}\langle \vec{X}, Q\vec{X}\rangle\right\} dXdY,$$

where the matrix $Q := Q(\tau)$ is

$$Q = Q(\tau) := \begin{pmatrix} 1 & \tau \\ \tau & 1 \end{pmatrix}.$$

The problem of calculating $f(J,K)$ is thus reduced to making an appropriate change of variables; since the measures dX, dY are translation invariant, this is an achievable task. The solution is

$$f(J,K) = \exp\left\{\frac{1}{2N} \operatorname{tr}\langle \vec{J}, Q^{-1}\vec{J}\rangle\right\},$$

where $\vec{J} := (J,K)^T$. Rewriting the inner product in terms of its components yields the result. □

Corollary 3.26. *For any* $1 \le i, j, k, \ell \le n$,

$$\begin{aligned} \langle X_{ij}\rangle_n &= 0, & \langle Y_{ij}\rangle_n &= 0, \\ \langle X_{ij}X_{k\ell}\rangle_n = \langle Y_{ij}Y_{k\ell}\rangle_n &= \tfrac{1}{N(1-\tau^2)}\delta_{i\ell}\delta_{jk}, & \langle X_{ij}Y_{k\ell}\rangle_n = \langle Y_{ij}X_{k\ell}\rangle_n &= \tfrac{\tau}{N(1-\tau^2)}\delta_{i\ell}\delta_{jk}. \end{aligned} \tag{3.81}$$

Proof. Notice that

$$\left.\frac{\partial f}{\partial J_{ji}}\right|_{J,K=0} = \langle X_{ij}\rangle_n, \qquad \left.\frac{\partial f}{\partial K_{ji}}\right|_{J,K=0} = \langle Y_{ij}\rangle_n,$$

and that

$$\left.\frac{\partial^2 f}{\partial J_{ji}\partial J_{\ell k}}\right|_{J,K=0} = \langle X_{ij}X_{k\ell}\rangle_n, \qquad \left.\frac{\partial^2 f}{\partial J_{ji}\partial K_{\ell k}}\right|_{J,K=0} = \langle X_{ij}Y_{k\ell}\rangle_n,$$

and similarly for all other expected values of interest. Calculating the derivatives explicitly yields the result. Note that the factor of two is again accounted for because the matrices J, K are Hermitian. □

The key observation to make here is that there we can now distinguish between the covariances of "alike" matrix entries (X-X and Y-Y type covariances) and "unalike" covariances (X-Y covariances). Thus, we expect the underlying combinatorics should reflect this difference. This is indeed the case; as before, it will be useful to first consider a few examples; as before, we take the parameter $N := n$, the size of the matrix.

Example 3.27. $\langle \operatorname{tr} X^4\rangle_n$. This example is calculated in an identical manner as before (cf. Example (3.19)), the only difference being an overall multiplicative factor:

$$\langle \operatorname{tr} X^4\rangle_n = \left(\frac{1}{1-\tau^2}\right)^2 \left(2n + \frac{1}{n}\right). \tag{3.82}$$

We remark that an identical formula for $\langle \operatorname{tr} Y^4\rangle_n$ holds.

Example 3.28. $\langle (\operatorname{tr} X^3)^2\rangle_n$. This example is again essentially identical to its 1-matrix model analog, with the exception being an overall multiplicative factor: factor:

$$\langle (\operatorname{tr} X^3)^2\rangle_n = \left(\frac{1}{1-\tau^2}\right)^3 \left(12 + \frac{3}{n^2}\right). \tag{3.83}$$

An identical formula for $\langle (\operatorname{tr} Y^3)^2\rangle_n$ holds. However, this example is slightly more instructive, as it admits a counterpart which does *not* appear in the 1-matrix model, as illustrated in the next example.

Example 3.29. $\langle \operatorname{tr} X^3 \operatorname{tr} Y^3\rangle_n$. Here, we have the possibility of both types of covariances appearing in the Wick expansion of the expected value. We have that:

$$\langle \operatorname{tr} X^3 \operatorname{tr} Y^3\rangle_n = \sum_{i_1,i_2,i_3=1}^{n} \sum_{j_1,j_2,j_3=1}^{n} X_{i_1i_2}X_{i_2i_3}X_{i_3i_1}Y_{j_1j_2}Y_{j_2j_3}Y_{j_3j_1},$$

Which, by Wick's theorem, we can write as a sum over all pairings of the indices. Let us consider the following contributing pairings:

$$\pi_a := \Big\{\{(i_1,i_2),(i_2,i_3)\},\{(i_3,i_1),(j_1,j_2)\},\{(j_2,j_3),(j_3,j_1)\}\Big\},$$
$$\pi_b := \Big\{\{(i_1,i_2),(j_1,j_2)\},\{(i_2,i_3),(j_3,j_1)\},\{(i_3,i_1),(j_2,j_3)\}\Big\},$$
$$\pi_c := \Big\{\{(i_1,i_2),(j_1,j_2)\},\{(i_2,i_3),(j_2,j_3)\},\{(i_3,i_1),(j_3,j_1)\}\Big\}.$$

(Note that these three pairings correspond to the three nonisomorphic types of graphs contributing to the Wick expansion in the 1-matrix model; the first two are planar, and the last is non-planar). The pairing π_a contributes a factor of

$$\langle X_{i_1i_2}X_{i_2i_3}\rangle_n\langle X_{i_3i_1}Y_{j_1j_2}\rangle_n\langle Y_{j_2j_3}Y_{j_3j_1}\rangle_n = \left(\frac{1}{1-\tau^2}\right)^3\tau;$$

a similar calculation shows that the pairing π_b contributes a factor of

$$\langle X_{i_1i_2}X_{j_1j_2}\rangle_n\langle X_{i_2i_3}Y_{j_3j_1}\rangle_n\langle X_{i_3i_1}Y_{j_2j_3}\rangle_n = \left(\frac{1}{1-\tau^2}\right)^3\tau^3.$$

Finally, the pairing π_c contributes a factor of

$$\langle X_{i_1i_2}X_{j_2j_2}\rangle_n\langle X_{i_2i_3}Y_{j_2j_3}\rangle_n\langle X_{i_3i_1}Y_{j_3j_1}\rangle_n = \left(\frac{1}{1-\tau^2}\right)^3\tau^3\cdot\frac{1}{n^2}.$$

All other pairings come from ribbon graphs which are isomorphic to the ones represented by one of the above three; thus, summing up, we find that

$$\langle \operatorname{tr} X^3 \operatorname{tr} Y^3\rangle_n = \left(\frac{1}{1-\tau^2}\right)^3\left(9\tau+3\tau^3+\frac{3\tau^3}{n^2}\right).$$

The last two examples show that the 2-matrix model can "detect" something the 1-matrix model could not: *colorings* (or *labellings*, as we shall sometimes say) of the vertices [2]. The factor of τ in the last calculation accounted for the number of edges connecting an X vertex to a Y vertex. We now develop a diagrammatic

[2] In fact, the 2 matrix model is able to distinguish edge colors, if we consider expected values of the form $\langle \operatorname{tr} XYXY\rangle_n$. However, for simplicity (and because it is the situation we are interested in), we confine ourselves to consider only products of traces of X^{i_p},Y^{j_q}.

approach to calculating expected values of the form

$$\left\langle \prod_{p=1}^{k} \operatorname{tr} X^{i_p} \prod_{q=1}^{\ell} \operatorname{tr} Y^{j_q} \right\rangle_n , \tag{3.84}$$

where $(i_1, ..., i_p)$ and $(j_1, ..., j_p)$ are sequences of positive integers. Let $K = \sum_{p=1}^{k} i_p$, $L = \sum_{q=1}^{\ell} j_q$, and $E = K + L$ (this will be the total number of edges in the associated ribbon graph R_π). As was the case before, we can Wick expand the above into a sum over all possible pairings of the indices, with each pairing π contributing a factor of

$$\sum_{I} \prod_{\{(r_a,r_b),(r_c,r_d)\}\in\pi} \langle Z_{r_a r_b} Z_{r_c r_d} \rangle_n, \tag{3.85}$$

where the sum I runs over all of the indices of the traces $\operatorname{tr} X^{i_p}$, $p = 1, ..., k$, and $\operatorname{tr} Y^{j_q}$, $q = 1, ..., j$, and

$$Z_{r_a r_b} := \begin{cases} X_{r_a r_b}, & (r_a, r_b) \text{ belongs to the index set of one of the traces } \operatorname{tr} X^{i_p},\ p = 1, ..., k, \\ Y_{r_a r_b}, & (r_a, r_b) \text{ belongs to the index set of one of the traces } \operatorname{tr} Y^{i_q},\ q = 1, ..., j. \end{cases} \tag{3.86}$$

To each such pairing π, we associate a (colored) ribbon graph R_π as follows:

1. Draw k X-colored vertices, each with degree i_p, $p = 1, ..., k$, with half-edges labelled by the indices of the trace of the corresponding vertex.

2. Draw ℓ Y-colored vertices, each with degree j_q, $q = 1, ..., \ell$, with half-edges labelled by the indices of the trace of the corresponding vertex.

3. Pair the edges according to the pairing prescribed by π.

The resulting diagram is called a 2-colored ribbon graph. We will sometimes denote a 2-colored ribbon graph (associated to a pairing π) on $V := k + \ell$ vertices, with k vertices of the first color, and $j := V - k$ of the second color, by $R_\pi(k, j)$. The *size* of such a ribbon graph is the number of vertices, denoted $|R_\pi(k, j)|= V$. We call such a coloring a *coloring of type* (k, j). The contribution of such a graph to the sum we will denote by $w[R_\pi(k, j)]$, or by $w[R_\pi]$, when the coloring type is clear. Given a pairing π, we see that all of the factor accounted for in the first theorem appear again (i.e., accounting for the number of edges, vertices, faces, etc.), with one additional factor here: the contribution will be multiplied by a power of τ^D, where D counts the number of edges connecting X-colored vertices to Y-colored vertices. Thus, we see that $w[R_\pi]$ is (upon identifying $N := n$):

$$w[R_\pi] = \frac{1}{n^E} \left(\frac{1}{1-\tau^2} \right)^E n^F \tau^D = \left(\frac{1}{1-\tau^2} \right)^E n^{\chi(R_\pi)-V} \tau^D. \tag{3.87}$$

Let U denote the number of edges in R_π connecting X-colored vertices to X-colored vertices and Y-colored vertices to Y-colored vertices. Then, we have the equality $D + U = E$, the total number of edges in R_π. If we define $S := D - U$, then $D = \frac{1}{2}(E + S)$, and we can express the weight $w[R_\pi]$ finally as

$$w[R_\pi] = \left(\frac{\tau^{1/2}}{1-\tau^2}\right)^E n^{\chi(R_\pi)-V} \tau^{\frac{1}{2}S} \tag{3.88}$$

(note that, E is always even, as the graph comes from a pairing, so there is no problem defining the square root). What we have proven is the following theorem:

Theorem 3.30. *Let X,Y be Gaussian random matrices, distributed according to the measure (3.78), with parameter $N := n$. Let $(i_1, ..., i_k)$, $(j_1, ..., j_\ell)$ be a k- (respectively, ℓ)-tuple of positive integers, and set $V := k + l$. Furthermore, put $E := \frac{1}{2}\left(\sum_p i_p + \sum_q j_q\right)$. Then,*

$$n^V \left\langle \prod_{p=1}^{k} \operatorname{tr} X^{i_p} \prod_{q=1}^{\ell} \operatorname{tr} Y^{j_q} \right\rangle_n = \left(\frac{\tau^{1/2}}{1-\tau^2}\right)^E \sum_{g\geq 0} \frac{e_g(\tau)}{n^{2g-2}}, \tag{3.89}$$

where $e_g(\tau)$ is a polynomial in τ, defined by

$$e_g(\tau) = \sum_{R_\pi} \tau^{\frac{1}{2}S(R_\pi)} \tag{3.90}$$

where the sum runs over all genus g unordered (k, j)-colored ribbon graphs, and $S(R_\pi)$ is the number of edges in R_π between like-colored vertices minus the number of edges between unalike vertices.

Note that, in the case when all of the vertices are of one color (say, X-colored), the internal sum $e_g(\tau)$ reduces to its analog in Theorem (3.21).

We now consider the generating function

$$Z_n(t, \tau; N) := \int_{\mathcal{H}_n} \int_{\mathcal{H}_n} \exp\left\{\operatorname{tr}\left[\tau XY - \frac{1}{2}X^2 - \frac{1}{2}Y^2 - \frac{t}{N}X^4 - \frac{t}{N}Y^4\right]\right\} dXdY, \tag{3.91}$$

which is the 2-matrix analog of the generating function (3.65), after a suitable rescaling. Expanding $\exp\left[-\frac{t}{N}\operatorname{tr} X^4\right]$, $\exp\left[-\frac{t}{N}\operatorname{tr} Y^4\right]$ as series in t, we obtain that

$$\exp\left[-\frac{t}{N}\operatorname{tr} X^4\right]\exp\left[-\frac{t}{N}\operatorname{tr} Y^4\right] = \sum_{M=0}^{\infty} \frac{(-t/N)^M}{M!} \sum_{k+j=M} \frac{M!}{k!\,j!}(\operatorname{tr} X^4)^k (\operatorname{tr} Y^4)^j. \tag{3.92}$$

If we insert this expression into Equation (3.91), and divide by the "Gaussian" partition function $Z_n(0,\tau;N)$, we obtain that

$$\frac{Z_n(t,\tau;N)}{Z_n(0,\tau;N)} = \sum_{M=0}^{\infty} \frac{(-t/N)^M}{M!} \sum_{k+j=M} \frac{M!}{k!\,j!} \langle (\operatorname{tr} X^4)^k (\operatorname{tr} Y^4)^j \rangle_n. \tag{3.93}$$

It is important to notice that, provided $k+j=M$, $\frac{M!}{k!j!}$ counts the number of 2-colorings of M objects, with k of the first color and j of the second. Expanding the expected value $\langle (\operatorname{tr} X^4)^k (\operatorname{tr} Y^4)^j \rangle_n$ using Wick's theorem and the diagrammatic rules we established, we obtain that

$$\frac{Z_n(t,\tau;N)}{Z_n(0,\tau;N)} = \sum_{M=0}^{\infty} \frac{(-t/N)^M}{M!} \sum_{k+j=M} \frac{M!}{k!\,j!} \sum_{|R_\pi(k,j)|=M} w[R_\pi(k,j)] \tag{3.94}$$

where the innermost sum runs over all 2-colored ribbon graphs on M vertices with k edges of the first color and j of the second, $k+j=M$. The colorings in this case are *unlabelled*, in the sense that any graph of a fixed type with coloring (k,j) are considered to be the same. However, we can count *labelled* colorings by noticing that:

1. If we color any labelled ribbon graph R_π in two different ways with precisely k vertices of the first color and j of the second, the resulting weights these diagrams contribute to the sum are identical,
2. If $k+j=M$, there are precisely $\frac{M!}{k!j!}$ possible colorings of type (k,j) on a given graph on M vertices.

Thus, we see that inner sum over k,j can be interpreted as a sum over the distinct possible colorings of the vertices, and we have that

$$\frac{Z_n(t,\tau;N)}{Z_n(0,\tau;N)} = \sum_{M=0}^{\infty} \frac{(-t/N)^M}{M!} \sum_{|R_\pi|=M} w[R_\pi], \tag{3.95}$$

where the internal sum now runs over all *labelled, 2-colored* ribbon graphs on N-vertices, with *any* coloring scheme. For our purposes, it is useful switch the order of summation over colorings and graphs, i.e., to write the above as

$$\frac{Z_n(t,\tau;N)}{Z_n(0,\tau;N)} = \sum_{M=0}^{\infty} \frac{(-t/N)^M}{M!} \sum_{\substack{\text{ribbon}\\\text{graphs}\\R_\pi}} \sum_{\substack{\text{colorings of}\\R_\pi}} w[R_\pi], \tag{3.96}$$

where the sum

$$\sum_{\substack{\text{ribbon}\\\text{graphs}\\R_\pi}}$$

is a sum over all *uncolored, labelled* ribbon graphs on M vertices, and

$$\sum_{\substack{\text{colorings of}\\ R_\pi}}$$

runs over all possible 2-colorings of R_π. Now, using the formula for the weights $w[R_\pi]$ of colored ribbon graphs we derived earlier (cf. Equation (3.88)), and using the fact that $E = 2M$ by the handshaking lemma, we have that (again putting $N := n$):

$$\frac{Z_n(t,\tau;n)}{Z_n(0,\tau;n)} = \sum_{M=0}^{\infty} \frac{1}{M!} \left(\frac{-t\tau}{(1-\tau^2)^2} \right)^M \sum_{\substack{\text{ribbon}\\ \text{graphs}\\ R_\pi}} n^{\chi(R_\pi)} \sum_{\substack{\text{colorings of}\\ R_\pi}} \tau^{\frac{1}{2}S(R_\pi)}. \tag{3.97}$$

For a fixed graph R_π, this sum is nothing but the partition function for the Ising model on this graph, with the parameter identification $\tau = e^{-2\beta}$. Thus, we have shown that

$$\frac{Z_n(t,\tau;n)}{Z_n(0,\tau;n)} = \sum_{M=0}^{\infty} \frac{1}{M!} \left(\frac{-t\tau}{(1-\tau^2)^2} \right)^M \sum_{\substack{\text{ribbon}\\ \text{graphs}\\ R_\pi}} n^{\chi(R_\pi)} Z_{R_\pi} \left(-\frac{1}{2} \log \tau \right). \tag{3.98}$$

As was the case in the previous section, taking logarithms of the above expression yields instead a sum over *connected* ribbon graphs:

$$\log \frac{Z_n(t,\tau;n)}{Z_n(0,\tau;n)} = \sum_{M=1}^{\infty} \frac{1}{M!} \left(\frac{-t\tau}{(1-\tau^2)^2} \right)^M {\sum_{\substack{\text{ribbon}\\ \text{graphs}\\ R_\pi}}}' n^{\chi(R_\pi)} Z_{R_\pi} \left(-\frac{1}{2} \log \tau \right), \tag{3.99}$$

where $'$ denotes the sum only over connected diagrams. If we divide through by n^2, and take the limit as $n \to \infty$, we obtain the following theorem.

Theorem 3.31. *Define the function*

$$F_n(t,\tau,N) := \frac{1}{n^2} \log \frac{Z_n(t,\tau;n)}{Z_n(0,\tau;n)}, \tag{3.100}$$

where $Z_n(t,\tau;N)$ is as in (3.91). Set

$$f(t,\tau) := \lim_{n\to\infty} \frac{1}{n^2} F_n(t,\tau,n). \tag{3.101}$$

Then,

$$f(t,\tau) = \sum_{M=0}^{\infty} \left(\frac{-t\tau}{(1-\tau^2)^2} \right)^M \mathcal{Z}_M(\tau), \tag{3.102}$$

where

$$\mathcal{Z}_M(\tau) = \sum_{|R_\pi|=M} Z_{R_\pi}\left[-\frac{1}{2}\log\tau \right], \tag{3.103}$$

and the sum runs over all 4*-regular, connected, planar ribbon graphs on* M *vertices, and* $Z_{R_\pi}(\beta)$ *is the partition function for the Ising model on* R_π*.*

It is the result of this thesis, and a forthcoming work with Maurice Duits and Seung-Yeop Lee, to give a rigorous analysis of the above generating function, and prove a theorem analogous to Theorem (3.23). One such result, which has been known for quite some time (and actually predates the interpretation of $f(t,\tau)$ as a generating function for the Ising model on random graphs) [17, 70, 80] gives an exact formula for $f(t,\tau)$:

Theorem 3.32.

$$f(t,\tau) = \frac{1}{2}\log\frac{\tau z(\tau,t)}{2t} - \int_0^{z(\tau,t)} \frac{d\zeta}{\zeta}\left[k(\zeta;\tau,t) - \frac{1}{2}k^2(\zeta;\tau,t) \right] - \log\frac{\tau}{2(1-\tau^2)} - 1, \tag{3.104}$$

where $k(\zeta;\tau,t)$ *is defined as*

$$k(\zeta;\tau,t) := \frac{\zeta}{t}\left[\frac{1}{(1-3\zeta)^2} - \tau^2 + 3\tau^2\zeta^2 \right], \tag{3.105}$$

and $z = z(\tau,t)$ *is implicitly determined as the unique solution of the fifth-order equation*

$$t = z\left[\frac{1}{(1-3z)^2} - \tau^2 + 3\tau^2 z^2 \right] \tag{3.106}$$

which satisfies

$$\lim_{t\to 0} \frac{z(\tau,t)}{t} = \frac{1}{(1-\tau^2)}. \tag{3.107}$$

In Chapter 5, we provide the first fully rigorous proof of this theorem.

3.6 Matrix Models and Quantum Gravity.

In this section, we summarize some of the progress made using matrix models in the theory of quantum gravity coupled to conformal matter. This is a subject that was extensively studied in the 1980s through the 1990s by various authors [17, 18, 25, 26, 34, 47, 48, 56, 58, 70, 72, 91], with an especially well-studied treatment of the "pure" gravity case (see in particular references [47, 48], and further results related to the associated Painlevé transcendent [10, 38, 66, 67]). We do not attempt to give an overarching survey of the aspects of the theory of 2D gravity. Here, our goal is to give a flavor of the approach some of the theoretical physicists working in this area took during this time period. We also try to demonstrate how matrix models became a central tool for answering a number of questions in this theory, and point out some of the results that are known rigorously, as well as some interesting conjectures that were made that were never rigorously proven. Most of what we write here should be taken with a grain of salt: everything is conjectural and non-rigorous unless otherwise stated. We mainly follow some review articles on the subject, put out by P. Ginsparg, G. Moore, and their collaborators [50, 55, 57].

Let us begin by attempting to describe a "pure" theory of quantum gravity, that is, with no external matter fields present. In 2 dimensions, classical (relativistic) gravity on a fixed compact 2-manifold M without boundary, with cosmological constant Λ, is described by the extrema of the *Einstein-Hilbert action*:

$$S[g] := \int_M \left[\frac{\mu}{2\pi} R(g) - \Lambda\right] \sqrt{g}dxdy, \tag{3.108}$$

where g is a metric on M, $R(g)$ is the scalar curvature (= twice the Gaussian curvature) of g, and $\sqrt{g}dxdy$ is the area element on M. By the Gauss-Bonnet Theorem [21], the above can be rewritten as

$$S[g] = \mu\chi(M) - \Lambda A(M), \tag{3.109}$$

where $\chi(M)$ is the Euler characteristic of M, and $A(M)$ is the Riemannian area of M with respect to the metric g. The above formula tells us that any classical theory of gravity is essentially trivial, because the essential part of the action, the part involving the scalar curvature, is a topological invariant. However, in a quantum theory of gravity, we are interested in the interpretation of the following formal generating function, called the *Feynman integral* or *partition function*:

$$\mathcal{Z}(\mu, \Lambda) := \sum_h \int e^{S[g]}\mathcal{D}g = \sum_h \int e^{\mu\chi(M)-\Lambda A(M)}\mathcal{D}g, \tag{3.110}$$

where the sum is taken over all possible genera of surfaces, and the integration is carried out over all (conformal equivalence classes) of metrics on M. This sum over genera, or "sum over topologies", as it is commonly referred to, is what makes *quantum* gravity in 2 dimensions interesting. To quote Di Francesco, Ginsparg, and Zinn-Justin [50],

> In the quantum case, however, even two dimensional gravity is non-trivial because large quantum fluctuations may change the genus of the surface and the partition function hence involves a sum

over surfaces of all genus. In addition, on higher genus surfaces there are non-trivial topological sectors.

In parallel to statistical physics, in quantum field theory (in our situation, quantum gravity), the main object of interest is the partition function (3.110). However, we immediately arrive at an issue: the partition function (3.110) is ill-defined, as the integral over the metric $\mathcal{D}g$ is not properly defined. Thus, we must try to give an appropriate interpretation of $\mathcal{Z}(\mu, \Lambda)$, and hope that any way we regularize the integral, the results will agree. A possible approach that garnered attention in the theoretical physics community was the so-called Liouville approach, introduced by Polyakov [91]. The method involves quantum field-theoretic techniques, which places it on shaky mathematical footing; however, this by no means discounts the results obtained in this direction.

Another possible approach to interpreting the partition function $\mathcal{Z}(\mu, \Lambda)$ is to discretize the surfaces we are summing over first, and then perform a continuum limit. In this case, the true partition function is approximated by

$$\mathcal{Z}_{approx}(\mu, \Lambda) = \sum_h \sum_T e^{\mu\chi(T) - \Lambda A(T)}, \tag{3.111}$$

where the internal sum is taken over all *triangulations*, or perhaps *quadrangulations* of a surface of genus g, depending on the context. If we assert that all triangles in our triangulation have equal area, by possible rescaling of Λ, we can take $A(T)$ to be the number of triangles in the triangulation. This puts us in a situation very close to the one of the previous two sections: a sum over all possible graphs of a given size, with weights given by a constant to the Euler characteristic of that graph. Since triangulations of a surface are dual to 3-regular graphs, and quadrangulations dual to 4-valent graphs (cf. Figure (3.3)), we see that $\mathcal{Z}_{approx}(\mu, \Lambda)$ and the generating function

$$F(t; N) = \frac{1}{N^2} \log Z(t; N) = \frac{1}{N^2} \log \left\langle \exp\left[\frac{t}{N} \operatorname{tr} X^4\right] \right\rangle_n$$

defined in Theorem (3.23) are closely linked. In $F(t; N)$, graphs of all sizes participate. To match a theory of quantum gravity, we want to take a "double scaling limit" in which we tune the coupling parameter t in the right way so that surfaces of infinite size (i.e., quadrangulations with an infinite number of vertices dominate $F(t, N)$. In other words, we want to consider the large M coefficient of the series $F(t; N)$ (see (3.68)), as this is precisely the generating function we are interested in. Calculation of the radius of convergence of the series $F(t, N)$ (as per standard complex variables, this is equivalent to calculating the nearest singularity of $F(t; N)$ to the origin in the t plane) determines the critical value of the coupling constant t at which we must consider $F(t, N)$. As it turns out, $F(t, N)$ is an analytic function of t, and there is a critical value of t, $t_c = -\frac{1}{48}$,

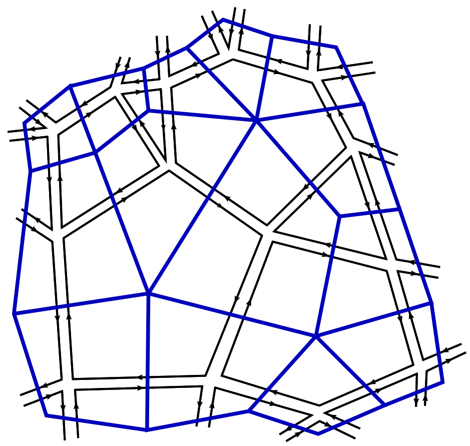


Figure 3.3. The quartic graphs generated by $\langle \exp(-t \operatorname{tr} X^4) \rangle_n$ are dual to "quadrangulations", that is, graphs in which every face is a quadrilateral. Thus, one can think of this generating function as a generating function of quadrangulations.

in our current normalization, where the function $F(t, N)$ diverges. In the language of statistical mechanics, this means that $F(t, N)$ has a phase transition occurring at $t = t_c$. This value was first calculated in [19], and the characterization of the transition was subsequently given in [34, 47, 48, 58]. The relevant theorem is the following (and this is indeed a theorem, in the honest mathematical sense). The statement is given for critical point of the 1-matrix model with cubic interactions $\operatorname{tr} X^3$, which is located at $t_c^2 = 1/(108\sqrt{3})$ [19].

Theorem 3.33. *(P. Bleher, A. Deaño) Given $\varepsilon > 0$ and $\delta > 0$, consider the double scaling regime $N^{4/5}(t - t_c) = c_1 \lambda$, $c_1 = 2^{-12/5} 3^{-7/4}$, and fix a neighborhood in the complex plane $D_R = \{\lambda \in \mathbb{C} : |\lambda| < R\}$. Let $\{\lambda_{\alpha,j}\}_{j=1}^{J}$ be the set of poles of $y(\lambda)$ (a special solution to the Painlevé I equation, cf. [10]) in D_R. The partition function $Z_N(t)$ can be written in the following way:*

$$Z_N(t) = Z_N^{\text{reg}}(t) Z_N^{\text{sing}}(\lambda) \left(1 + \mathcal{O}(N^{-\varepsilon})\right), \tag{3.112}$$

for $\lambda \in D_R \setminus \cup_j D(\lambda_j, \delta)$. Here the regular part is

$$Z_N^{\text{reg}}(t) = e^{N^2[A + B(t - t_c) + C(t - t_c)^2] + D}, \tag{3.113}$$

where the constants A, B, C and D are explicit in terms of the genus 0 partition function. The singular part of the partition function is

$$Z_N^{\text{sing}}(\lambda) = e^{-Y(\lambda)}, \tag{3.114}$$

where $Y(\lambda)$ solves the differential equation

$$Y''(\lambda) = y(\lambda), \tag{3.115}$$

with boundary condition

$$Y(\lambda) = \frac{2\sqrt{6}}{45}(-\lambda)^{5/2} - \frac{1}{48}\log(-\lambda) + \mathcal{O}((-\lambda)^{-5/2}), \qquad (-\lambda) \to \infty. \tag{3.116}$$

A similar result is implicit in the earlier work of [38] for the quartic matrix model, although they make no direct statement about the partition function, only the recursion coefficients.

It is also of great interest to calculate how different matter fields $\phi(z)$ interact with the pure theory of gravity. One class of quantum field theories relevant to statistical physics are the minimal models of conformal field theory.

In short, a *conformal field theory* (CFT) can be thought of mathematically as follows. We will be extraordinarily vague, as the full definition of a conformal field theory (more precisely, a vertex operator algebra) will take us too far astray from our original goal. For a full treatment of conformal field theory, one should consult [51, 54], for the relevant physical context, and [52] for the relevant mathematical treatment of the theory. A CFT is a (possibly infinite) collection of representations of the local conformal algebra (in 2 dimensions, this is the celebrated Virasoro algebra), along with a rule for how such representations interact, called the *operator product expansion.* Normally, the product of two such representations is a third, unrelated representation, and, upon taking an any number of products, we obtain any number of new representations.

The minimal models are distinguished among in that their operator product content is *finitely generated.* To the layman, this means that minimal models can be characterized by a finite set of parameters: the *central charge* c, which is indexed by a pair of coprime integers (p, q):

$$c = c(p, q) := 1 - 6\frac{(p-q)^2}{pq}, \tag{3.117}$$

and the *conformal dimensions* of its primary fields:

$$\Delta_{r,s} := \frac{(pr - qs)^2 - (p-q)^2}{4pq}, \tag{3.118}$$

for $1 \leq r \leq q-1$, $1 \leq s \leq p-1$. These models were introduced originally in [8] as some of the very first examples of exactly solvable quantum field theories; their introduction was for the primary purpose of describing universality classes of phase transitions in 2 dimensions. The general theory was this: each minimal model corresponded to the universality class of some 2D phase transition. The first few correspondences were:

1. The $(4,3)$ minimal model describes the critical Ising universality class,
2. The $(5,2)$ minimal model describes the Yang-Lee edge singularity (cf. [51, 109, 110])
3. The $(6,5)$ minimal model describes the critical 3-Potts model universality class (cf. [7, 51]),

and so on. The word "describe" here is intentionally vague; with the techniques allowed by CFT, physicists were only able to compute the critical exponents and a few very basic correlation functions of these theories. This being said, one should not discount their achievements: it took a great deal of effort and ingenuity to calculate this, and it was already apparent to them that much heavier machinery would be needed to compute more.

One direction of interest which would add more significance to the idea of universality was to see how CFTs (in particular, the minimal models) reacted when coupled to 2D gravity. Let us describe what is meant by *coupling to gravity* in this context. Suppose we are given a classical (scalar) field theory on a *fixed* Riemannian manifold (M,g), described by the action

$$S_{matter}[\phi] := \int \mathcal{L}(\phi, \partial_\mu \phi)\sqrt{g}d^2x, \tag{3.119}$$

where $\mathcal{L}(\phi, \partial_\mu\phi)$ is some coordinate-invariant density on M (note that this means in general that $\mathcal{L}$ has an implicit dependence on the metric g). As a bare quantum field theory, one is then interested in the calculation of

$$Z_{matter}(\hbar) = \int e^{-\frac{1}{\hbar}S_{matter}[\phi]}\mathcal{D}\phi. \tag{3.120}$$

By *coupling to quantum gravity*, we mean that we now consider the quantity

$$Z_{coupled}(\mu, \lambda, \hbar) = \sum_h \int e^{S[g]-\frac{1}{\hbar}S_{matter}[\phi]}\mathcal{D}\phi\mathcal{D}g. \tag{3.121}$$

The above (formal) integral does not in general split into a product of two integrals, as $S_{matter}[\phi]$ often depends on the metric in a non-trivial way. The problem of describing how a given field theory "reacts" when coupled to gravity was one of the most important problems in the original theory of 2D gravity, as free theories (i.e., theories of gravity in the absence of any matter) could only provide a limited amount of information. One of the first results in this direction was the result of Vladimir Kazakov [70], who predicted how the critical exponents of the Ising model should react when coupled to gravity. His prediction led was based purely on matrix model techniques: one could couple the Ising field to gravity by appropriately tuning the parameters of the generating function $F_n(t, \tau; N)$, defined in (3.100), so that infinite-size graphs again dominate, as in the 1-matrix case. Kazakov's result led to progress in the continuum theory of quantum gravity: after some preliminary work of Polyakov [90], A. Knizhnik, A. Polyakov, and A. Zamolodchikov (KPZ) derived a general formula for how the critical exponents of the (p, q) minimal model change when coupled to 2d gravity [72]. Their result was as follows: the coupled primary field dimensions $\Delta_{r,s}$ are related to the gravitationally-coupled primary field dimensions $\hat{\Delta}_{r,s}$ by the relation

$$\hat{\Delta}_{r,s} - \Delta_{r,s} = \frac{\hat{\Delta}_{r,s}(1 - \hat{\Delta}_{r,s})}{c+2}, \tag{3.122}$$

where c is the central charge (3.117) of the theory. This formula, along with the groundbreaking work of Kazakov and company, motivated the Douglas-Shenker [35], Brézin-Kazakov [20], and Gross-Migdal [58, 59] program. Their goal in part was to construct a unified theory of all $2D$ critical phenomena, with renormalization group flows between all of the critical points. The key insight that allowed these groups to perform computations previously inaccessible in the continuum theory was the observation that the 2-matrix model contained all of the (p, q)-minimal models coupled to gravity as critical points. This was an completion of the 1-matrix model, which contained only the series of minimal models indexed by $(2, 2p+1)$, for $p \geq 1$. In the works [50, 55, 57], the general program is outlined, and many explicit computations are performed, but these computations are still lacking in that they rely on non- convergent matrix integrals to describe phase transitions in the 2-matrix model. In Chapter 5, we shall prove a result about the partition function of the 2-matrix model at the Ising critical point, analogous to Theorem (3.33). In the language of this section, this gives a rigorous description of the $(4, 3)$-minimal model (corresponding to the critical Ising model) coupled to 2D gravity.

CHAPTER 4

STEEPEST DESCENT ANALYSIS: AN OVERVIEW

In this chapter, we outline the basic ideas of steepest descent analysis for Riemann-Hilbert problems, focusing in particular on the application to orthogonal polynomial ensembles. By now, these ideas are well-established in the literature (cf. [29, 30], for example). Nothing we present here is new; the purpose is this chapter is to introduce the main methods of steepest descent, and to provide a "simplest" example of the problem, so that the ideas in the next chapter (where the application of this method is less straightforward) is more palatable.

4.1 The Riemann-Hilbert Problem: Existence and Uniqueness Theorems

4.1.1 Preliminaries

In this section, we establish some notations, and the relevant background information about Cauchy-type integrals. This material is by now well-established in the literature (see for example [53, 84]). In the interest of brevity, we only state the relevant results, and omit the proofs. The interested reader should consult the aforementioned references for full details.

We call a family of contours $\Gamma \subset \mathbb{C}$ *complete* if it divides $\mathbb{C}$ into two complementary regions, say, D_+ and D_-. Examples of such systems of contours are shown in Figure (4.1). We will always assume that the

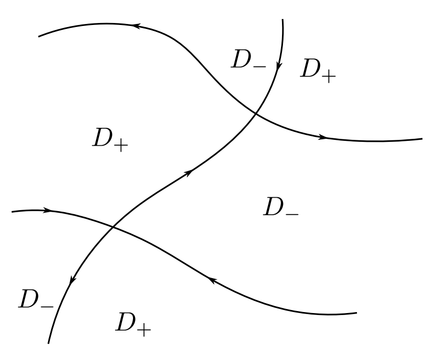


Figure 4.1. A complete system of contours, dividing the plane into regions D_+ and D_-.

contours Γ are piecewise smooth; this is overkill, but in what follows we will not meet any situation where this is not the case.

Then, if $f(z)$ is any function defined in $\mathbb{C} \setminus \Gamma$, and $z \in \Gamma$, we set

$$f_+(z) := \lim_{\substack{\zeta \to z \\ \zeta \in D_+}} f(\zeta), \tag{4.1}$$

$$f_-(z) := \lim_{\substack{\zeta \to z \\ \zeta \in D_-}} f(\zeta). \tag{4.2}$$

Since we are often interested in limits of functions defined in terms of Cauchy-type integrals, we assert that the above limits are actually *non-tangential* limits, i.e., ζ approaches z from within an angle from within D_+ (respectively, D_-). By a slight abuse of notation (which is mainly made for ease of notation), we will often just write "lim", and the prefix *non-tangential* will be implied. This being said, we shall continue through this section to emphasize that the limits taken are nontangential.

Suppose $\varphi : \Gamma \to \mathbb{C}$ is Hölder-α continuous on Γ. We define a function $C[\varphi](z)$, holomorphic in $\mathbb{C} \setminus \Gamma$, via the Cauchy-type integral

$$C[\varphi](z) := \frac{1}{2\pi i} \int_\Gamma \frac{\varphi(\zeta)}{\zeta - z} d\zeta. \tag{4.3}$$

A well-known theorem in complex analysis which is attributed to mathematicians J. Sokhotski and J. Plemelj, asserts that $C[\varphi](z)$ has well-defined (non-tangential!) boundary values, which are given by, for $z \in \Gamma$,

$$C[\varphi]_+(z) = \frac{1}{2}\varphi(z) + \frac{1}{2\pi i} \fint_\Gamma \frac{\varphi(\zeta)}{\zeta - z} d\zeta, \tag{4.4}$$

$$C[\varphi]_-(z) = -\frac{1}{2}\varphi(z) + \frac{1}{2\pi i} \fint_\Gamma \frac{\varphi(\zeta)}{\zeta - z} d\zeta. \tag{4.5}$$

An important observation is that we can rearrange (4.4), (4.5) to obtain

$$C[\varphi]_+(z) - C[\varphi]_-(z) = \varphi(z), \tag{4.6}$$

$$C[\varphi]_+(z) + C[\varphi]_-(z) = \frac{1}{\pi i} \fint_\Gamma \frac{\varphi(\zeta)}{\zeta - z}. \tag{4.7}$$

An important generalization of the above theorems is to the case of square-integrable functions, i.e., the case when $\varphi \in L^2(\Gamma)$:

$$L^2(\Gamma) := \left\{ f : \Gamma \to \mathbb{C} \,\middle|\, ||f||_{L^2(\Gamma)} := \left(\int_\Gamma |f(z)|^2 |dz| \right)^{1/2} < \infty \right\}. \tag{4.8}$$

In this situation, it happens that the boundary values of $C[\varphi](z)$ exist pointwise almost everywhere (a.e.), and moreover belong to $L^2(\Gamma)$ themselves[1]. Consequentially, the operators

$$\mathbf{C}_+[\varphi](z) := \lim_{\substack{\zeta\to z\\ \zeta\in D_+}} C[\varphi](\zeta), \tag{4.9}$$

$$\mathbf{C}_-[\varphi](z) := \lim_{\substack{\zeta\to z\\ \zeta\in D_-}} C[\varphi](\zeta) \tag{4.10}$$

are well-defined bounded operators from $L^2(\Gamma)$ to $L^2(\Gamma)$. These operators satisfy several important relations.

Proposition 4.1. *The operators $\mathbf{C}_\pm : L^2(\Gamma) \to L^2(\Gamma)$ have the following properties:*

1. $\mathbf{C}_+ - \mathbf{C}_- = \mathbf{1}$*, where* $\mathbf{1}$ *is the identity operator on* $L^(\Gamma)$*,*
2. $\mathbf{C}_+ \circ \mathbf{C}_- = \mathbf{C}_- \circ \mathbf{C}_+ = 0$*, and finally*
3. $\mathbf{C}_+ \circ \mathbf{C}_+ = \mathbf{C}_+$*, and similarly* $(-\mathbf{C}_-) \circ (-\mathbf{C}_-) = -\mathbf{C}_-$*.*

This proposition follows essentially from Cauchy's theorem, combined with the Plemelj formulae. Indeed, 1. is just a rewriting of (4.6). The above proposition establishes that the operators $\pm\mathbf{C}_\pm$ are orthogonal projections on $L^2(\Gamma)$, whose sum is the identity operator.

4.1.2 Riemann-Hilbert Problems and "Small-Norm" Theory

We will now apply the technology of the previous section to an important class of boundary value problems known as *Riemann-Hilbert problems* (RHPs). The standard version of a RHP we shall meet is formulated as follows:

Problem 4.2. *Suppose we are given a complete system of contours Γ, and an $n \times n$ matrix-valued function $J : \Gamma \to \mathbb{C}$ with entries $J_{ij}(z) \in L^2(\Gamma)$, $i, j = 1, ..., n$, with $\det J(z) \equiv 1$. We must construct a matrix-valued function $\mathbf{X}(z)$, with entries holomorphic $\mathbf{X}_{ij}(z)$ in $\mathbb{C} \setminus \Gamma$, such that*

$$\begin{cases} \mathbf{X}_+(z) = \mathbf{X}_-(z)J(z), & z \in \Gamma, \\ \mathbf{X}(z) = \mathbb{I} + \mathcal{O}(z^{-1}), & z \to \infty. \end{cases} \tag{4.11}$$

Here, the interpretation of $+/-$ is that the boundary values are taken entrywise. If the system of contours contains an arc which goes off to infinity, the limit in the second equation meant to be taken off of this arc.

[1] This result is non-trivial; the fact that the Hölder-α functions remain Hölder-α can be traced back to Plemelj and Privalov. However, the corresponding result for $L^2(\Gamma)$ functions did not appear for some time in the literature; the result is nowadays often attributed to Coifman, McIntosh, and Meyer [23].

We call $J(z)$ the *jump matrix* for $\mathbf{X}(z)$; consequentially, the first condition is called the *jump condition*. Condition 2 regarding the asymptotics of $\mathbf{X}(z)$ is called the *normalization condition*.

RHPs such as (4.2) appear commonly in mathematics and mathematical physics, and it is of vital importance to establish the existence and uniqueness of solutions to such problems. As it turns out, uniqueness is almost a triviality, as the following proposition will show.

Proposition 4.3. *If Problem* (4.2) *has a solution, it is unique.*

Proof. First, suppose $\mathbf{X}(z)$ is a solution to the Riemann-Hilbert problem (4.2), and set $f(z) := \det \mathbf{X}(z)$. Then, the $f(z)$ is a holomorphic function in $\mathbb{C} \setminus \Gamma$, with boundary values

$$f_+(z) = \det \mathbf{X}_+(z) = \det [\mathbf{X}_-(z)J(z)] = \det \mathbf{X}_-(z) = f_-(z), \qquad z \in \Gamma,$$

by virtue of the fact that $\det J(z) \equiv 1$. Thus, $f(z)$ extends to a continuous function in $\mathbb{C}$, which is holomorphic off of Γ. By Morera's theorem, it follows that $f(z)$ is entire. Furthermore, the normalization condition $\mathbf{X}(z) = \mathbb{I} + \mathcal{O}(z^{-1})$ implies that

$$f(z) = 1 + \mathcal{O}(z^{-1}), \qquad z \to \infty.$$

Thus, by Liouville's theorem, $f(z) \equiv 1$. It follows that solutions to Problem (4.2) are invertible.

Now, suppose $\mathbf{X}(z)$, $\tilde{\mathbf{X}}(z)$ are solutions to Problem (4.2), and set

$$\mathbf{E}(z) := \tilde{\mathbf{X}}(z)\mathbf{X}^{-1}(z).$$

The entries of $\mathbf{E}(z)$ are holomorphic in $\mathbb{C} \setminus \Gamma$, since $\mathbf{X}(z)$ is invertible. Furthermore, $\mathbf{E}(z)$ is continuous across Γ, since, for $z \in \Gamma$,

$$\mathbf{E}_+(z) = \tilde{\mathbf{X}}_+(z)\mathbf{X}_+^{-1}(z) = \tilde{\mathbf{X}}_-(z)J(z)J^{-1}(z)\mathbf{X}_-^{-1}(z) = \tilde{\mathbf{X}}_-(z)\mathbf{X}_-^{-1}(z) = \mathbf{E}_-(z).$$

Thus, the entries of $\mathbf{E}(z)$ extend to entire functions. The normalization condition $\mathbf{X}(z) = \mathbb{I} + \mathcal{O}(z^{-1})$ can then be used entrywise to determine the entries $\mathbf{E}_{ij}(z)$. On the diagonal, we have that, for $i = 1, ..., n$,

$$\mathbf{E}_{ii}(z) = 1 + \mathcal{O}(z^{-1}), \qquad z \to \infty,$$

and so Liouville's theorem implies that $\mathbf{E}_{ii}(z) \equiv 1$. Similarly, for $i \neq j$, $i, j = 1, ..., n$,

$$\mathbf{E}_{ij}(z) = \mathcal{O}(z^{-1}), \qquad z \to \infty,$$

and so Liouville's theorem implies that $\mathbf{E}_{ij}(z) \equiv 0$. Thus, we have shown that $\mathbf{E}(z) \equiv \mathbb{I}$; equivalently,

$$\tilde{\mathbf{X}}(z) \equiv \mathbf{X}(z).$$

□

This sort of argumentation is standard in Riemann-Hilbert analysis, and we shall meet it frequently in practice. Now, the previous proposition indeed establishes the uniqueness for the RHPs we will be interested in. However, we need additionally to show that a solution actually exists. As is typical in existence-type results, this is not an easy task. We will find that this is possible to do, provided that the jump matrix is "close" (in a sense that we will establish below) to the identity matrix $\mathbb{I}$. Formally speaking, the general principle is as follows: if $||J(z) - \mathbb{I}||$ is small (in some appropriately chosen sense), then a solution to Problem $\mathbf{X}(z)$ exists, and satisfies $||\mathbf{X}(z) - \mathbb{I}||$ is small (again, in some appropriately chosen sense). Such a Riemann-Hilbert problem will be called a *small norm Riemann-Hilbert problem.* We now prove existence of small norm RHPs; this is done by establishing an equivalence of the RHP (4.2) and a certain singular integral equation.

In the interest of generality, suppose the jump matrix $J(z)$ admits a factorization into a pair of invertible matrices:

$$J(z) = v_-^{-1}(z)v_+(z), \tag{4.12}$$

where $v_\pm(z)$ are defined on Γ. Of course, we always have the trivial factorization $v_+(z) := J(z)$, $v_-(z) = \mathbb{I}$; however, in practice, other factorizations can be more convenient. Additionally, this factorization makes the presentation of results more symmetric. Set

$$w_+(z) := v_+(z) - \mathbb{I}, \qquad w_-(z) := \mathbb{I} - v_-(z). \tag{4.13}$$

We define the singular integral operator

$$\mathbf{C}_w[\mathbf{X}](z) := \mathbf{C}_+[\mathbf{X}w_-](z) + \mathbf{C}_-[\mathbf{X}w_+](z), \tag{4.14}$$

where here $\mathbf{C}_\pm$ are defined as the Cauchy operators (4.9),(4.10), applied element-wise to the quantity in square brackets (we express these operators in an identical manner, by a slight abuse of notation). Then, we have the following proposition.

Proposition 4.4. *Consider the singular integral equation*

$$\mathbf{R}(z) - \mathbf{C}_w[\mathbf{R}](z) = \mathbb{I}, \qquad z \in \Gamma \tag{4.15}$$

Then, both problem the Riemann-Hilbert problem (4.2) *and Equation* (4.15) *simultaneously have solutions, or not.*

Proof. Suppose $\mathbf{R}(z)$ is the solution to the singular integral equation (4.15). We claim that

$$\mathbf{X}(z) := \mathbb{I} + C[\mathbf{R}w_+](z) + C[\mathbf{R}w_-](z). \tag{4.16}$$

is the solution to the Riemann-Hilbert problem (4.2). To see that this is the case, we must check that the jump condition and normalization condition are satisfied; the fact that the solution is holomorphic off of Γ is obvious, by properties of the Cauchy integral. Let us calculate the jump condition first. On one hand, we have that

$$\begin{aligned}
\mathbf{X}_+(z) &= \mathbb{I} + \mathbf{C}_+[\mathbf{R}w_+](z) + \mathbf{C}_+[\mathbf{R}w_-](z) \\
&= \mathbf{R}(z) + \mathbf{C}_+[\mathbf{R}w_+](z) - \mathbf{C}_-[\mathbf{R}w_+](z) \\
&= \mathbf{R}(z) + (\mathbf{C}_+ - \mathbf{C}_-)\,[\mathbf{R}w_+](z) \\
&= \mathbf{R}(z)(\mathbb{I} + w_+(z)) = \mathbf{R}(z)v_+(z),
\end{aligned}$$

where the second line comes directly from rearrangement of the singular integral equation, and we have used the fact that $\mathbf{C}_+ - \mathbf{C}_- = \mathbf{1}$. On the other hand, we have that

$$\begin{aligned}
\mathbf{X}_-(z) &= \mathbb{I} + \mathbf{C}_-[\mathbf{R}w_+](z) + \mathbf{C}_-[\mathbf{R}w_-](z) \\
&= \mathbf{R}(z) - \mathbf{C}_+[\mathbf{R}w_-](z) + \mathbf{C}_-[\mathbf{R}w_-](z) \\
&= \mathbf{R}(z) - (\mathbf{C}_+ - \mathbf{C}_-)[\mathbf{R}w_-](z) \\
&= \mathbf{R}(z)(\mathbb{I} - w_-(z)) = \mathbf{R}(z)v_-(z).
\end{aligned}$$

Thus, since $J(z) = v_-^{-1}(z)v_+(z)$, we see that

$$\mathbf{X}_+(z) = \mathbf{R}(z)v_+(z) = \mathbf{R}(z)v_-(z)[v_-^{-1}(z)v_+(z)] = \mathbf{X}_-(z)J(z),$$

and so $\mathbf{X}(z)$ satisfies the jump condition of (4.2). The normalization condition is much easier to check, since, as $z \to \infty$, the Cauchy transform of an integrable function behaves like $\mathcal{O}(z^{-1})$. In other words,

$$C[\mathbf{R}w_+](z) = \mathcal{O}(z^{-1}), z \to \infty, \qquad C[\mathbf{R}w_-](z) = \mathcal{O}(z^{-1}), z \to \infty.$$

Thus, we have that

$$\mathbf{X}(z) = \mathbb{I} + \mathcal{O}(z^{-1}), \qquad z \to \infty.$$

We have therefore shown that solvability of the singular integral equation (4.15) implies solvability of the RHP (4.2).

Conversely, suppose the Riemann-Hilbert problem (4.2) has a solution. Without loss of generality, let us take $v_+(z) = J(z)$, $v_-(z) = \mathbb{I}$. Then, since (4.2) has a solution, we can write

$$\mathbf{X}_+(z) = \mathbf{X}_-(z)J(z) = \mathbf{X}_-(z)(J(z) - \mathbb{I}) + \mathbf{X}_-(z),$$

or, rearranging,

$$\mathbf{X}_+(z) - \mathbf{X}_-(z) = \mathbf{X}_-(z)(J(z) - \mathbb{I});$$

since $\mathbf{X}_-(z)$ exists, the above additive Riemann-Hilbert problem can be resolved using the Plemelj formulae:

$$\mathbf{X}(z) = \mathbb{I} + C[\mathbf{X}_-(J - \mathbb{I})](z);$$

note that the addition of the identity matrix is due to the identity normalization of the Riemann-Hilbert problem for $\mathbf{X}(z)$. If we take the limit as $z \to \Gamma$ from the $-$ side of the contour, we find that $\mathbf{X}_-(z)$ solves the singular integral equation

$$\mathbf{X}_-(z) = \mathbb{I} + C[\mathbf{X}_-(J - \mathbb{I})](z).$$

Comparison with (4.15) in the case where $v_+(z) = J(z)$, $v_-(z) = \mathbb{I}$ shows that the singular integral equation has a solution; namely, $\mathbf{X}_-(z)$. □

We have now reduced the problem of existence of a solution to Problem (4.2) to the existence of the solution to the singular integral equation (4.15). Note that this equation is of the form

$$(\mathbf{1} - \mathbf{C}_w)\mathbf{R}(z) = \mathbb{I}; \tag{4.17}$$

thus, we see that if the norm of the operator $\mathbf{C}_w$ is sufficiently small (less than one), the operator $(\mathbf{1} - \mathbf{C}_w)$ will be invertible, with the inverse given explicitly by the Neumann series

$$(\mathbf{1} - \mathbf{C}_w)^{-1} = \mathbf{1} + \mathbf{C}_w + \mathbf{C}_w \circ \mathbf{C}_w + ... = \mathbf{1} + \sum_{k=1}^{\infty} (\mathbf{C}_w)^k,$$

where $(\mathbf{C}_w)^k$ is interpreted as $\mathbf{C}_w$ composed with itself k times. Thus, in order to guarantee the existence of the original RHP, it is enough to guarantee that the operator $\mathbf{C}_w$ indeed has sufficiently small norm. Indeed, since the Cauchy operators $\mathbf{C}_{\pm}$ were bounded as operators on $L^2(\Gamma)$, we have that

$$||\mathbf{C}_w||_{L^2(\Gamma)} \leq M_\Gamma \left(||w_+||_\infty + ||w_-||_\infty \right), \tag{4.18}$$

where here $||\cdot||_{L^2(\Gamma)}$ is the operator norm for operators from $L^2(\Gamma)$ to $L^2(\Gamma)$, and $||\cdot||_\infty$ is the supremum norm, taken entrywise. In practice, we have very little control over the constant M_Γ, but often we *do* have control over the supremum norms of the matrix functions $w_\pm(z)$. In particular, if there exists $\epsilon > 0$ sufficiently small such that $||w_+||_\infty, ||w_+||_\infty < \frac{\epsilon}{2M}$, then the singular integral equation (4.15), and thus the Riemann-Hilbert problem (4.2) has a unique solution. In the situation we will be interested in, we will have a *family* of Riemann-Hilbert problems of the form (4.2), where the jump matrix $J(z) := J(z;\epsilon)$ satisfies

$$||J(z;\epsilon) - \mathbb{I}||_\infty < C\epsilon, \tag{4.19}$$

for some positive constant $C > 0$. In this case, we have the following proposition:

Proposition 4.5. *Consider a family of Riemann-Hilbert problems of the form* (4.2)*, depending on* ϵ*, whose jump matrix satisfies the condition* (4.19)*. Then, this family has a solution for all* $\epsilon > 0$*, sufficiently small. Furthermore, the solution may be expanded as a series in powers of* ϵ*:*

$$\mathbf{X}(z) = \mathbb{I} + \mathbf{X_1}\epsilon + \mathbf{X_2}\epsilon^2 + \mathcal{O}(\epsilon^3), \tag{4.20}$$

for some matrix-valued functions $\mathbf{X_1}(z)$*,* $\mathbf{X_2}(z)$*, etc., independent of* ϵ*.*

Proof. By what we have already established above, it is enough to show that the matrices $w_+(z)$, $w_-(z)$ have sufficiently small supremum norm. Indeed, if we take the factorization of $J(z,\epsilon) = v_-^{-1}(z)v_+(z)$, with

$v_+(z) = J(z,\epsilon)$, $v_-(z) = \mathbb{I}$, we have that

$$||w_-(z)||_\infty = ||\mathbb{I} - v_-(z)||_\infty = 0,$$
$$||w_+(z)||_\infty = ||v_+(z) - \mathbb{I}||_\infty = ||J(z;\epsilon) - \mathbb{I}||_\infty < C\epsilon.$$

So, for sufficiently small $\epsilon > 0$, we can arrange for the operator norm of $\mathbf{C}_w$ to be less than 1, and so the singular integral equation has a solution, and so the Riemann-Hilbert problem (4.2) has a solution. The Neumann series expansion of the inverse operator $(\mathbf{1} - \mathbf{C}_w)^{-1}$ yields that the solution to the singular integral equation has an expansion in powers of ϵ; since the solution of the Riemann-Hilbert problem is explicit in terms of the solution of the integral equation, it has such an expansion as well. This proves (4.20). □

Remark 4.6. If $J(z;\epsilon)$ furthermore has an expansion of the form

$$J(z;\epsilon) = \mathbb{I} + \epsilon J_1(z) + \epsilon^2 J_2(z) + \mathcal{O}(\epsilon^3),$$

where $J_1(z)$, $J_2(z), \cdots$ are some uniformly bounded functions on Γ, then we can actually explicitly express the corrections to the ϵ-expansion of $R(z)$. By writing

$$R(z) = \mathbb{I} + \epsilon R_1(z) + \epsilon^2 R_2(z) + \mathcal{O}(\epsilon^3),$$

and inserting into the small-norm Riemann Hilbert problem, we obtain *additive* RHPs for the $R_j(z)$'s, which can be immediately solved via the Plemelj formula. For example, the first few equations determining the $R_j(z)$'s are

$$R_{1,+}(z) = R_{1,-}(z) + J_1(z),$$
$$R_{2,+}(z) = R_{2,-}(z) + J_2(z) + R_{1,-}(z)J_1(z),$$

and so on. All $R_j(z) = \mathcal{O}(z^{-1})$ as $z \to \infty$, so that $R(z)$ is identity normalized. Thus, we have that

$$R_1(z) = C[J_1](z),$$

so that, to first order in ϵ, $R(z) = \mathbb{I} + \epsilon C[J_1](z) + \mathcal{O}(\epsilon^2)$. This solution can then be inserted into the second equation determining $R_2(z)$, and so forth. In principle, one can write down a system of Riemann-Hilbert problems that can be solved iteratively as we have shown above; however, in practice, one typically only

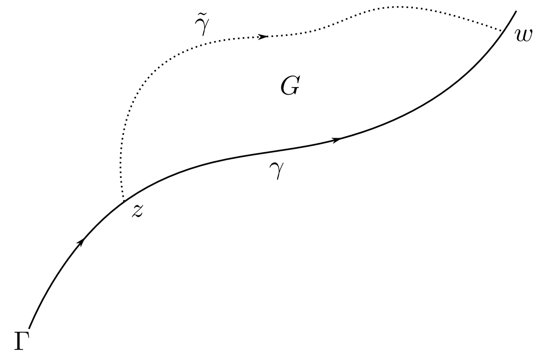


Figure 4.2. The contour Γ, and 'deformed' contour $\tilde{\Gamma} := \Gamma \cup \tilde{\gamma}$. The solid curve connecting the points z, w is γ, the dotted contour connecting these points is $\tilde{\gamma}$.

needs the first few terms in this series expansion. What we have demonstrated above is sufficient for our purposes.

4.1.3 Standard Riemann-Hilbert Techniques.

We will meet Riemann-Hilbert problems in later sections which characterize certain quantities of interest, such as orthogonal polynomials or biorthogonal polynomials. However, these RHPs will not be in the form of a small norm-type problem. The method of steepest descent is based on the following idea: we can make a series of explicit and invertible transformations on the given Riemann-Hilbert problem, to obtain a small-norm Riemann-Hilbert problem. By Proposition (4.19), we find that a solution exists whenever a certain parameter is small enough; by inverting the sequence of transformations we made, we thus have also proved existence for the original RHP. We will illustrate this idea in more detail in the forthcoming sections. However, it will be useful to first introduce some of the transformations we will meet.

Let us suppose Γ is a single smooth contour. Suppose $\mathbf{X}(z)$ is analytic in $\mathbb{C} \setminus \Gamma$, and satisfies the jump condition

$$\mathbf{X}_+(z) = \mathbf{X}_-(z)J(z), \tag{4.21}$$

for some invertible matrix-valued function $J : \Gamma \to \mathbb{C}$. The first kind of transformation we will come across is the analog of *deformation of contour* in Cauchy's integral theorem.

Proposition 4.7. *(Deformation of Contour). Suppose the jump matrix $J(z)$ extends analytically to the closure of a bounded domain G, as shown in Figure (4.2). Set $\gamma := \partial G \cap \Gamma$, $\tilde{\gamma} := \partial G \setminus \Gamma$, and $\tilde{\Gamma} := (\Gamma \setminus \gamma) \cup \tilde{\gamma}$,* *with orientations as in the Figure. Finally, define*

$$\tilde{\mathbf{X}}(z) := \begin{cases} \mathbf{X}(z)J^{-1}(z), & z \in G, \\ \mathbf{X}(z), & \textit{otherwise}. \end{cases} \tag{4.22}$$

Then, $\tilde{\mathbf{X}}(z)$ *is analytic in* $\mathbb{C} \setminus \tilde{\Gamma}$*, and is the unique solution to the following Riemann-Hilbert problem:*

$$\begin{cases} \tilde{\mathbf{X}}_+(z) = \tilde{\mathbf{X}}_-(z)J(z), & z \in \tilde{\Gamma}, \\ \tilde{\mathbf{X}}(z) = \mathbb{I} + \mathcal{O}(z^{-1}), & z \to \infty. \end{cases} \tag{4.23}$$

Proof. The fact that $\tilde{\mathbf{X}}(z) \equiv \mathbf{X}(z)$ for z sufficiently large (since G is a bounded domain) yields that $\tilde{\mathbf{X}}(z)$ is also identity-normalized. It remains to check the jump condition. For $z \in \tilde{\Gamma} \setminus \tilde{\gamma}$, $\tilde{\mathbf{X}}(z)$ and $\mathbf{X}(z)$ have the same boundary values, and so

$$\tilde{\mathbf{X}}_+(z) = \mathbf{X}_+(z) = \mathbf{X}_-(z)J(z) = \tilde{\mathbf{X}}_-(z)J(z).$$

On the other hand, for $z \in \tilde{\gamma}$, the function $\mathbf{X}(z)$ is analytic, and so

$$\tilde{\mathbf{X}}_+(z) = \mathbf{X}(z) = \mathbf{X}(z)J^{-1}(z)J(z) = \tilde{\mathbf{X}}_-(z)J(z),$$

by the definition of $\tilde{\mathbf{X}}(z)$ in the region G. Finally, we must check that the jump of $\tilde{\mathbf{X}}(z)$ on γ is eliminated. Indeed, we have that, for $z \in \gamma$,

$$\tilde{\mathbf{X}}_+(z) = \mathbf{X}_+(z)J^{-1}(z) = \mathbf{X}_-(z)J(z)J^{-1}(z) = \mathbf{X}_-(z) = \tilde{\mathbf{X}}_-(z).$$

It follows that $\tilde{X}(z)$ extends to an analytic function in a neighborhood of γ, by Morera's theorem. □

Remark 4.8. Note that, if the region G were unbounded, the same proof as the above applies; the only detail that changes is that there is possibly different behavior of $\tilde{\mathbf{X}}(z)$ if we approach infinity in the region G. However, if $J(z) = \mathbb{I} + \mathcal{O}(z^{-1})$, $z \to \infty$ in the region G, them the same statement as the above Proposition holds.

The next proposition is an extremely useful technique called *lens opening*. Often, we will meet jump matrices which are highly oscillatory on the jump contour, however, certain components become exponentially

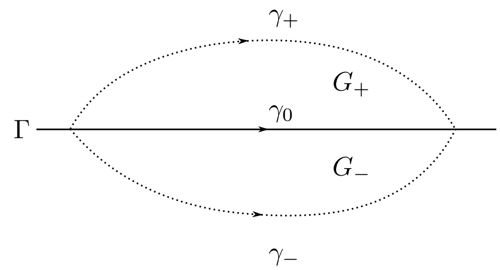


Figure 4.3. The contour structure for the 'opening of lenses' transformation.

small if we move above or below the contour just a little. The following lensing proposition tells us how to properly "break apart" the jump so that oscillations can be removed.

Proposition 4.9. *(Lens opening). Consider the family of contours and regions depicted in Figure* (4.3)*. Suppose the jump matrix $J(z)$ admits a factorization*

$$J(z) = v_-(z)D(z)v_+(z), \tag{4.24}$$

where $v_-(z)$ admits an analytic continuation into the closure of the region G_-, and $v_+(z)$ admits an analytic continuation into the closure of the region G_+. Then, set $\gamma_0 := G_+ \cap \Gamma (= G_- \cap \Gamma)$, $\gamma_+ := \partial G_+ \setminus \Gamma$, $\gamma_- := \partial G_- \setminus \Gamma$, with orientations as given in the Figure, and define

$$\tilde{\mathbf{X}}(z) := \begin{cases} \mathbf{X}(z)v_+^{-1}(z), & z \in G_+, \\ \mathbf{X}(z)v_-(z), & z \in G_-, \\ \mathbf{X}(z), & \textit{otherwise}. \end{cases} \tag{4.25}$$

Then, $\tilde{\mathbf{X}}(z)$ is the unique solution to the following Riemann-Hilbert problem:

$$\tilde{\mathbf{X}}_+(z) = \begin{cases} v_+(z), & z \in \gamma_+, \\ v_-(z), & z \in \gamma_-, \\ D(z), & z \in \gamma_0, \\ J(z), & \textit{otherwise}. \end{cases} \tag{4.26}$$

$$\tilde{\mathbf{X}}(z) = \mathbb{I} + \mathcal{O}(z^{-1}), \qquad z \to \infty. \tag{4.27}$$

Proof. The proof is again by direct calculation: if $z \in \Gamma \setminus \gamma_0$, then the boundary values of $\mathbf{X}(z)$ and $\tilde{\mathbf{X}}(z)$ coincide, and so

$$\tilde{\mathbf{X}}_+(z) = \mathbf{X}_+(z) = \mathbf{X}_-(z)J(z) = \tilde{\mathbf{X}}_-(z)J(z).$$

For $z \in \gamma_+$, we have that

$$\tilde{\mathbf{X}}_+(z) = \mathbf{X}(z) = \mathbf{X}(z)v_+^{-1}(z)v_+(z) = \tilde{\mathbf{X}}_-(z)v_+(z),$$

by definition of $\tilde{\mathbf{X}}(z)$. A similar calculation shows that $\tilde{\mathbf{X}}_+(z) = \tilde{\mathbf{X}}_-(z)v_-(z)$ on γ_-. Finally, for $z \in \gamma_0$,

$$\tilde{\mathbf{X}}_+(z) = \mathbf{X}_+(z)v_+^{-1}(z) = \mathbf{X}_-(z)J(z)v_+^{-1}(z) = \mathbf{X}_-(z)v_-(z)D(z) = \tilde{\mathbf{X}}_-(z)D(z).$$

This completes the proof. □

Remark 4.10. The region $G_+ \cup G_-$ is often called the *lens*. We also remark that the lens opening proposition is actually a generalization of Proposition (4.7), by taking $D(z) = v_-(z) = \mathbb{I}$, and $v_+(z) = J(z)$.

Example 4.11. There is a particularly important example of lens opening which we shall meet quite frequently in steepest descent analysis. In this case, we specialize to the 2×2 situation. Suppose that the jump matrix $J(z)$ takes the form

$$J(z) := \begin{pmatrix} e^{-in\theta_+(z)} & 1 \\ 0 & e^{in\theta_-(z)} \end{pmatrix}, \tag{4.28}$$

where $\theta_+(z)$ extends analytically to a region above the jump contour, $\theta_-(z)$ extends analytically to a region below the contour, and $\theta_+(z) - \theta_-(z) = 0$ for z on the contour. Then, $J(z)$ admits the factorization

$$J(z) = \begin{pmatrix} 1 & 0 \\ e^{in\theta_-(z)} & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ e^{-in\theta_+(z)} & 1 \end{pmatrix}. \tag{4.29}$$

In this situation, we would take

$$v_\pm(z) = \begin{pmatrix} 1 & 0 \\ e^{\mp in\theta_\pm(z)} & 1 \end{pmatrix}, \qquad D(z) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{4.30}$$

Another common technique that appears in Riemann-Hilbert analysis is *conjugation*. We first provide a simple example of this technique, then prove a more general version of the conjugation formula.

Proposition 4.12. *(The Effect of Conjugation, Version 1). Suppose* $\mathbf{X}(z)$ *satisfies the Riemann-Hilbert problem* (4.2)*, with jump matrix* $J(z)$*, and let* $m(z)$ *be an* $n \times n$ *matrix-valued invertible holomorphic function. Define*

$$\mathbf{M}(z) := m^{-1}(z)\mathbf{X}(z)m(z). \tag{4.31}$$

Then, $\mathbf{M}(z)$ *is the unique solution to the following Riemann-Hilbert problem:*

$$\begin{cases} \mathbf{M}_+(z) = \mathbf{M}_-(z)J_{\mathbf{M}}(z), & z \in \Gamma, \\ \mathbf{M}(z) = \mathbb{I} + \mathcal{O}(z^{-1}), & z \to \infty, \end{cases} \tag{4.32}$$

Here, $J_{\mathbf{M}}(z) = m^{-1}(z)J(z)m(z)$*.*

Proof. The proof is a direct computation: we have that

$$\begin{aligned} \mathbf{M}_+(z) &= m^{-1}(z)\mathbf{X}_+(z)m(z) = m^{-1}(z)\mathbf{X}_-(z)J(z)m(z) \\ &= m^{-1}(z)\mathbf{X}_-(z)m(z)m^{-1}(z)J(z)m(z) = \mathbf{M}_-(z)m^{-1}(z)J(z)m(z). \end{aligned}$$

The asymptotic condition follows similarly, since, as $z \to \infty$,

$$\mathbf{M}(z) = m^{-1}(z)[\mathbb{I} + \mathcal{O}(z^{-1})]m(z) = \mathbb{I} + \mathcal{O}(z^{-1}),$$

since $m(z)$ is invertible and holomorphic. □

Remark 4.13. Note that, if we had instead only multiplied $\mathbf{X}(z)$ on the *right* by $m(z)$, the jump matrix $J(z)$ would still be conjugated by $m(z)$. However, the Riemann-Hilbert problem satisfied by $\mathbf{X}(z)m(z)$ would no longer be identity-normalized: one would have the asymptotics $\mathbf{X}(z)m(z) = [\mathbb{I} + \mathcal{O}(z^{-1})]m(z)$, as $z \to \infty$.

The above proposition is enough for most purposes. However, it is sometimes the case that the matrix $m(z)$ we are conjugating by has jumps of its own, possibly on a completely different set of contours. We address this more general situation in the following proposition:

Proposition 4.14. *(The Effect of Conjugation, Version 2). Suppose* $\mathbf{X}(z)$ *satisfies the Riemann-Hilbert problem* (4.2) *with jump matrix* $J(z)$*, on the system of contours* Γ*. Suppose further that* $m(z)$ *is the solution to the following Riemann-Hilbert problem:* $m(z)$ *is analytic in* $\mathbb{C} \setminus \tilde{\Gamma}$*, and*

$$\begin{cases} m_+(z) = m_-(z)K(z), & z \in \tilde{\Gamma}, \\ m(z) = \mathbb{I} + \mathcal{O}(z^{-1}), & z \to \infty. \end{cases} \tag{4.33}$$

Then, the product $\mathbf{M}(z) := \mathbf{X}(z)m(z)$ *is the unique solution to the following Riemann-Hilbert problem:* $\mathbf{M}(z)$ *is analytic in* $\mathbb{C} \setminus (\Gamma \cup \tilde{\Gamma})$, *and*

$$\mathbf{M}_+(z) = \mathbf{M}_-(z) \begin{cases} m^{-1}(z)J(z)m(z), & z \in \Gamma \setminus \tilde{\Gamma}, \\ K(z), & z \in \tilde{\Gamma} \setminus \Gamma, \\ m_-^{-1}(z)J(z)m_+(z), & z \in \Gamma \cap \tilde{\Gamma}. \end{cases} \tag{4.34}$$

$$\mathbf{M}(z) = \mathbb{I} + \mathcal{O}(z^{-1}), \qquad z \to \infty. \tag{4.35}$$

Proof. Although the statement of the proposition might look daunting, the proof here is just as straightforward as it was in Proposition (4.12). First, note that the normalization condition satisfied by $\mathbf{M}(z)$ is obvious. Therefore, we have only to check that $\mathbf{M}(z)$ has the required jumps. Let $z \in \Gamma \setminus \tilde{\Gamma}$; then, as $\zeta \to z$ from the $+/-$ sides of Γ, only $\mathbf{X}(z)$ has a discontinuity, since $m(z)$ is analytic there. Therefore,

$$\mathbf{M}_+(z) = \mathbf{X}_+(z)m(z) = \mathbf{X}_-(z)J(z)m(z) = \mathbf{M}_-(z)[m^{-1}(z)J(z)m(z)],$$

by definition of $\mathbf{M}(z)$. Now, suppose $z \in \tilde{\Gamma} \setminus \Gamma$; only $m(z)$ has a discontinuity there, and $\mathbf{X}(z)$ is analytic. Thus,

$$\mathbf{M}_+(z) = \mathbf{X}(z)m_+(z) = \mathbf{X}(z)m_-(z)K(z) = \mathbf{M}_-(z)K(z).$$

Finally, suppose $z \in \Gamma \cap \tilde{\Gamma}$. In this case, both $\mathbf{X}(z)$ and $\mathbf{m}(z)$ have discontinuities, and so

$$\begin{aligned} \mathbf{M}_+(z) &= \mathbf{X}_+(z)m_+(z) = \mathbf{X}_-(z)J(z)m_+(z) \\ &= \mathbf{X}_-(z)m_-(z)m_-^{-1}(z)J(z)m_+(z) = \mathbf{M}_-(z)[m_-^{-1}(z)J(z)m_+(z)]. \end{aligned}$$

□

Remark 4.15. Note that the last jump can alternatively be written as

$$m_-^{-1}(z)J(z)m_+(z) = m_-^{-1}(z)J(z)m_-(z)K(z); \tag{4.36}$$

this is sometimes a more useful presentation of this jump.

4.2 A Riemann-Hilbert Problem For Orthogonal Polynomials

In this section, we define a Riemann-Hilbert problem related to orthogonal polynomials on the real line. This is intended to be an illustrative first example of the method of steepest descent, from which we can eventually approach more complicated examples (in particular, biorthogonal polynomials and the 2-matrix model). This being said, we do not introduce any new ideas here; what follows has appeared in the literature many times before [29, 30], and now appears in books such as P. Deift's [27]. Although we will not follow Deift's book very closely, most of what we say here is contained in his text.

Let us first establish the kinds of orthogonal polynomials that are relevant to random matrix theory. Let $V(z) = z^{2p} + ...$ be a monic of even degree $2p$, and consider the family of (monic) orthogonal polynomials defined by the relation

$$\int_{\mathbb{R}} \pi_n(z)\pi_m(z)e^{-NV(z)}dz = h_n\delta_{nm}. \tag{4.37}$$

Note that the requirement that the polynomials $\pi_n(z)$ are monic means that we cannot additionally require that the polynomials are *orthonormal*. This will not be an issue for us. Note that the polynomials also depend on the parameter $N > 0$; when we look for the asymptotics of these polynomials, we will eventually take $N = n$. Now consider the following Riemann-Hilbert problem:

Problem 4.16. *Construct a* 2×2 *matrix-valued function* $\mathbf{Y}(z)$*, analytic in* $\mathbb{C}\setminus\mathbb{R}$*, such that*

$$\begin{cases} \mathbf{Y}_+(z) = \mathbf{Y}_-(z)\left(\begin{smallmatrix} 1 & e^{-NV(z)} \\ 0 & 1 \end{smallmatrix}\right), & z\in\mathbb{R}, \\ \mathbf{Y}(z) = [\mathbb{I}+\mathcal{O}(z^{-1})]z^{n\sigma_3}, & z\to\infty. \end{cases} \tag{4.38}$$

Here, by $z^{n\sigma_3}$, we mean the matrix

$$z^{n\sigma_3} := \left(\begin{smallmatrix} z^n & 0 \\ 0 & z^{-n} \end{smallmatrix}\right). \tag{4.39}$$

When relevant, we shall denote the dependence of the above RHP on N, n as $\mathbf{Y}(z) = \mathbf{Y}_n(z;N)$. Note that, if N is large, this problem *does* have jumps that are close to the identity; however, this problem is not identity-normalized, and so it is not a Riemann-Hilbert problem of small-norm type. Despite this, it is clear that solutions to (4.16) are unique, provided that they exist. This follows from the fact that $\det\mathbf{Y}(z) = 1+\mathcal{O}(z^{-1})$ at infinity; the rest follows from the Liouville argument from the proof of Proposition (4.3).

The first surprising result, first noted by Fokas, Its, and Kitaev in [47], is that this Riemann-Hilbert problem is related to the orthogonal polynomials defined by the relation (4.37).

Proposition 4.17. *The solution to the Riemann-Hilbert problem* (4.16) *is given by*

$$\mathbf{Y}(z) = \begin{pmatrix} \pi_n(z) & C[\pi_n e^{-NV}](z) \\ -\frac{2\pi i}{h_{n-1}}\pi_{n-1}(z) & -\frac{2\pi i}{h_{n-1}}C[\pi_{n-1}e^{-NV}](z) \end{pmatrix}, \tag{4.40}$$

where $\pi_n(z)$ *are uniquely defined by the relation* (4.37), *and* $C[f](z) := \frac{1}{2\pi i}\int_{\mathbb{R}} \frac{f(\zeta)}{\zeta - z} d\zeta$.

Proof. Let us consider the first column of $\mathbf{Y}$. The jump condition implies that the first column in fact is continuous across $\mathbb{R}$:

$$\mathbf{Y}_{11,+}(z) = \mathbf{Y}_{11,-}(z), \qquad \mathbf{Y}_{12,+}(z) = \mathbf{Y}_{12,-}(z), \qquad z \in \mathbb{R},$$

and so $\mathbf{Y}_{11}$, $\mathbf{Y}_{12}$ extend to entire functions. The normalization condition implies that

$$\mathbf{Y}_{11}(z) = z^n + \mathcal{O}(z^{n-1}), \qquad \mathbf{Y}_{12}(z) = \mathcal{O}(z^{n-1}).$$

Therefore, by Liouville's theorem, $\mathbf{Y}_{11}(z)$ is a monic polynomial of degree n, and $\mathbf{Y}_{12}(z)$ is a polynomial of degree at most $n-1$. We now consider the second column. The 2-1 entry of $\mathbf{Y}$ satisfies the jump condition

$$\mathbf{Y}_{21,+}(z) = \mathbf{Y}_{21,-}(z) + \mathbf{Y}_{11}(z)e^{-NV(z)}, \qquad z \in \mathbb{R};$$

By the Plemelj formulae, the solution to the above RHP is

$$\mathbf{Y}_{21}(z) = C[\mathbf{Y}_{11}e^{-NV}](z) = \frac{1}{2\pi i}\int_{\mathbb{R}} \frac{\mathbf{Y}_{11}(\zeta)e^{-NV(\zeta)}}{\zeta - z} d\zeta.$$

Now, writing $\frac{1}{\zeta - z} = -\sum_{k=0}^{\infty} \frac{\zeta^k}{z^{k+1}}$, we see that

$$\mathbf{Y}_{21}(z) = \frac{\mathbf{Y}_{11}(\zeta)e^{-NV(\zeta)}}{\zeta - z} d\zeta = -\frac{1}{2\pi i}\sum_{k=0}^{\infty} \frac{1}{z^{k+1}} \int_{\mathbb{R}} \zeta^k \mathbf{Y}_{11}(\zeta)e^{-NV(\zeta)} d\zeta.$$

But, the asymptotic condition on $\mathbf{Y}_{21}(z)$ requires that

$$\mathbf{Y}_{21}(z) = \mathcal{O}(z^{-n-1}),$$

i.e., that the first n terms in the summation above vanish. This amounts to the vanishing of the integrals

$$\int_{\mathbb{R}} \zeta^k \mathbf{Y}_{11}(\zeta)e^{-NV(\zeta)} d\zeta = 0, k = 0, 1, ..., n-1.$$

Since $\mathbf{Y}_{11}(z)$ was a monic polynomial of degree n, these uniquely determines $\mathbf{Y}_{11}(z)$ as the n^{th} monic *orthogonal* polynomial. The same analysis applied to the second row yields the desired 2-2 entry; the constant $-\frac{2\pi i}{h_{n-1}}$ appears so that $\mathbf{Y}_{22}(z) = z^{-n} + \mathcal{O}(z^{-n-1})$, as opposed to $\mathbf{Y}_{22}(z) = cz^{-n} + \mathcal{O}(z^{-n-1})$, for some constant c. □

In fact, we can derive many classical identities satisfied by such orthogonal polynomials with relative ease using the Riemann-Hilbert formulation.

Proposition 4.18. *For fixed N, $V(z)$, and given $n \geq 0$, there exists a matrix polynomial $A_n(z) := A_1 z + A_0$ such that*

$$\mathbf{Y}_{n+1}(z) = A_n(z)\mathbf{Y}_n(z).$$

Proof. We have already noted that $\mathbf{Y}_n(z)$ is globally invertible. Next, consider the quantity

$$A_n(z) := \mathbf{Y}_{n+1}(z)\mathbf{Y}_n^{-1}(z).$$

$A_n(z)$ is holomorphic everywhere, except possibly the real axis. Since the jumps of $\mathbf{Y}_{n+1}(z)$, $\mathbf{Y}_n(z)$ across $\mathbb{R}$ are the same, and given by an explicit invertible matrix $J(z)$, we have that

$$A_{n,+}(z) = \mathbf{Y}_{n+1,+}(z)\mathbf{Y}_{n,+}^{-1}(z) = \mathbf{Y}_{n+1,-}(z)J(z)J^{-1}(z)\mathbf{Y}_{n,-}^{-1}(z) = \mathbf{Y}_{n+1,-}(z)\mathbf{Y}_{n,-}^{-1}(z) = A_{n,-}(z).$$

Thus, by Morera's theorem, the $A_n(z)$ extends to an entire matrix-valued function. Since, as $z \to \infty$,

$$A_n(z) = \mathbf{Y}_{n+1}(z)\mathbf{Y}_n^{-1}(z) = \left[\mathbb{I} + \mathcal{O}(z^{-1})\right] z^{(n+1)\sigma_3} z^{-n\sigma_3} \left[\mathbb{I} + \mathcal{O}(z^{-1})\right] = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right) z + \mathcal{O}(1).$$

Thus, by Liouville's theorem, $A_n(z)$ *is* a degree 1 polynomial of the form

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} z + A_0,$$

for some constant matrix A_0. In fact, if one is more careful with the asymptotic calculations, one can determine the matrix A_0 explicitly:

$$A_0 = \begin{pmatrix} a_{n+1} - a_n & \frac{h_n}{2\pi i} \\ -\frac{2\pi i}{h_n} & 0 \end{pmatrix},$$

where a_n is defined to be the subleading coefficient of $\pi_n(z)$, $\pi_n(z) = z^n + a_n z^{n-1} + ...$, and h_n are the orthogonality coefficients of the polynomials. □

Comparing the 1-1 entry of the equation $\mathbf{Y}_{n+1}(z) = A_n(z)\mathbf{Y}_n(z)$ yields the classical 3-term recurrence relation we derived for classical orthogonal polynomials in Chapter 3. This is not the only classical formula that can be derived from the Riemann-Hilbert formulation. In fact, one can also recover the Christoffel-Darboux kernel, as the next proposition shows.

Proposition 4.19. *The Christoffel-Darboux kernel, as defined in Equation* (3.18)*, admits the expression*

$$K_n(\lambda, \mu) = -\frac{1}{2\pi i}\frac{\left[\mathbf{Y}_n^{-1}(\lambda)\mathbf{Y}_n(\mu)\right]_{21}}{\lambda - \mu}. \tag{4.41}$$

Proof. Let us calculate the $2-1$ entry of the matrix $\left[\mathbf{Y}_n^{-1}(\lambda)\mathbf{Y}_n(\mu)\right]_{21}$. We have that

$$\mathbf{Y}_n^{-1}(\lambda)\mathbf{Y}_n(\mu) = \begin{pmatrix} * & * \\ \frac{2\pi i}{h_{n-1}}\pi_{n-1}(\lambda) & \pi_n(\lambda) \end{pmatrix}\begin{pmatrix} \pi_n(\mu) & * \\ -\frac{2\pi i}{h_{n-1}}\pi_{n-1}(\mu) & * \end{pmatrix} = \begin{pmatrix} * & * \\ \frac{2\pi i}{h_{n-1}}[\pi_n(\mu)\pi_{n-1}(\lambda) - \pi_n(\lambda)\pi_{n-1}(\mu)] & * \end{pmatrix}.$$

Dividing through by $2\pi i(\mu - \lambda)$, and comparing with the result of Proposition (3.7), yields the result. ☐

However, despite all of the useful identities we can derive directly from the Riemann-Hilbert formulation, we cannot say much about the asymptotics of the polynomials. For this, we shall need to employ the *method of steepest descent.*

4.3 The Method of Steepest Descent Applied to Orthogonal Polynomials

In order to study the asymptotics of the polynomials, we must employ the *method of steepest descent*, a technique first introduced in [31] to study a similar asymptotic problem in the theory of integrable systems. The idea here is the following: we will consider $\frac{1}{n} =: \epsilon$ as a small parameter, and try to convert the Riemann-Hilbert problem (4.16) to a small-norm problem. This will be achieved by applying a sequence of explicit and invertible transformations

$$\mathbf{Y} \longmapsto \mathbf{T} \longmapsto \mathbf{S} \longmapsto \mathbf{R},$$

where $\mathbf{R}$ is a small-norm RHP. Each of the above transformations accomplishes a particular goal. Let us formally outline them here. The first task is to remove 'growing part' of the asymptotics, i.e., to remove $z^{n\sigma_3}$. This is the essence of the first transformation $\mathbf{Y} \longmapsto \mathbf{T}$, and is often called the "g-function" transformation, due to the appearance of a function common to all such transformations in the steepest descent method. Upon removing the 'growing part' of the asymptotics, we will find that the first transformation has left some highly oscillatory jumps on the real line. In order to get rid of these oscillations, we will need to "push" the jump off the real line and onto "lenses"; this constitutes the second transformation $\mathbf{T} \longmapsto \mathbf{S}$. After performing this transformation, we will be left with only exponentially small, or constant jumps. From here on, we must try our best to find a solution to the remaining constant jump Riemann-Hilbert problem, this "parametrix" solution must match the jumps of $\mathbf{S}$, while staying identity-normalized at infinity. If we are successful in finding such a parametrix, then "dividing out" this parametrix from $\mathbf{S}$ will yield a small-norm Riemann-Hilbert problem, for which we have established existence and uniqueness. Tracing back the steps above yields a large-n asymptotic formula for the polynomials $\pi_n(z)$.

Note that the jump matrix for $\mathbf{Y}$ depends on a parameter N. From here on, we shall take this parameter $N = n$, for reasons that shall soon become clear.

4.3.1 The First Transformation

The goal of the first transformation is to bring the asymptotics of the RHP (4.16) to an identity-normalized form. Seemingly, if we just multiply $\mathbf{Y}$ by $z^{-n\sigma_3}$ on the right, we have solved the problem. However, this transformation causes problems of its own; a pole is then introduced at the origin, of order n. We therefore need some alternative to this. One possibility is to (as it is put by P. Miller [83]) "smear out" the pole into a branch cut, which Riemann-Hilbert analysis is better equipped to deal with. We put

$$g(z) := \int_{\mathbb{R}} \log(z - x)\rho(x)dx, \tag{4.42}$$

for some unit Borel measure $\rho(x)dx$, with compact support. Then, since $g(z) = \log z + \mathcal{O}(z^{-1})$, $z \to \infty$, the function

$$e^{-ng(z)\sigma_3} = z^{-n\sigma_3}[\mathbb{I} + \mathcal{O}(z^{-1})], z \to \infty, \tag{4.43}$$

and so multiplication by this function will properly normalize the Riemann-Hilbert problem for $\mathbf{Y}$. Notice further that $\det e^{-ng(z)\sigma_3} \equiv 1$, and so we can always invert this transformation. We therefore set

$$\mathbf{U}(z) := \mathbf{Y}(z)e^{-ng(z)\sigma_3}. \tag{4.44}$$

This is sort of a preliminary transformation; the full first transformation will come later. We then have that $\mathbf{U}(z)$ satisfies a Riemann-Hilbert problem of its own, which is identity-normalized:

Problem 4.20. *The function* $\mathbf{U}(z)$ *is holomorphic in* $\mathbb{C}\setminus\mathbb{R}$*, and is the solution to the following Riemann-Hilbert problem:*

$$\begin{cases} \mathbf{U}_+(z) = \mathbf{U}_-(z)\begin{pmatrix} e^{-n[g_+(z)-g_-(z)]} & e^{-n[V(z)-g_+(z)-g_-(z)]} \\ 0 & e^{n[g_+(z)-g_-(z)]} \end{pmatrix}, & z \in \mathbb{R}, \\ \mathbf{U}(z) = \mathbb{I} + \mathcal{O}(z^{-1}), & z \to \infty. \end{cases} \tag{4.45}$$

Here, $g_\pm(z)$ *denote the boundary values from above and below the real axis.*

Proof. As we have already observed, the asymptotic condition for $\mathbf{U}(z)$ is met by how we defined the transformation. As for the jump condition, we have that

$$\begin{aligned} \mathbf{U}_+(z) &= \mathbf{Y}_+(z)e^{-ng_+(z)\sigma_3} = \mathbf{Y}_-(z)\left(\begin{smallmatrix} 1 & e^{-nV(z)} \\ 0 & 1 \end{smallmatrix}\right)e^{-ng_+(z)\sigma_3} \\ &= \mathbf{Y}_-(z)e^{-ng_-(z)\sigma_3}e^{ng_-(z)\sigma_3}\left(\begin{smallmatrix} 1 & e^{-nV(z)} \\ 0 & 1 \end{smallmatrix}\right)e^{-ng_+(z)\sigma_3} \\ &= \mathbf{U}_-(z)e^{ng_-(z)\sigma_3}\left(\begin{smallmatrix} 1 & e^{-nV(z)} \\ 0 & 1 \end{smallmatrix}\right)e^{-ng_+(z)\sigma_3}; \end{aligned}$$

a straightforward calculation completes the result. ∎

We must choose $\rho(z)$ in such a way that we can eventually transform our RHP into a small-norm problem. In order to find the "right" choice of $\rho(z)$, let us express the boundary values $g_\pm(z)$ more explicitly. Since

$$[\log(z-x)]_\pm = \begin{cases} \log|z-x|, & z > x, \\ \log|z-x| \pm i\pi, & z < x, \end{cases} \tag{4.46}$$

we have that

$$g_\pm(z) = \int_{\mathbb{R}} \log|z-x|\rho(x)dx \pm i\pi \int_z^\infty \rho(x)dx. \tag{4.47}$$

In particular, we have that

$$g_+(z) - g_-(z) = 2\pi i \int_z^\infty \rho(x)dx. \tag{4.48}$$

From here, we see that, when n is large, the diagonal of the jump matrix is rapidly oscillating on the support $\rho(z)$. On the other hand, the off-diagonal part of the jump is characterized by the function

$$V(z) - g_+(z) - g_-(z) = 2\int_{\mathbb{R}} \log\frac{1}{|z-x|}\rho(x)dx + V(z). \tag{4.49}$$

The key observation here is that Equation (4.49) is precisely the variational derivative of a certain energy functional, namely,

$$E(\rho) := \iint_{\mathbb{R}\times\mathbb{R}} \log \frac{1}{|z-x|}\rho(x)\rho(z)dzdx + \int_{\mathbb{R}} V(z)\rho(z)dz. \tag{4.50}$$

In physical terms, this energy functional describes a distribution of electrons interacting through the 2-dimensional Coulomb potential (in 2D, this potential scales with distance r as $\log \frac{1}{r}$), confined to the real line, which experience an external potential $V(z)$ [2]. We develop some of the relevant potential theory in Appendix A; for a more detailed description of the general potential theory techniques involved here, we refer to [97]. For now, we appeal to physical intuition to proceed. Existence of the minimizer to the functional (4.50) is guaranteed on physical grounds – the charges must go *somewhere*, and the fact that we have chosen $V(z)$ to grow at infinity sufficiently fast implies that the charges must sit on some compactly supported set. For simplicity, from here on we will assume that $V(z)$ is additionally *convex*, so that the minimizer is unique, and supported on a single interval $[\alpha, \beta]$. In equilibrium, the charges must be stationary, within the confining potential, and the effective potential must increase off of the support of the charges. This amounts to the following variational equations:

$$2\int_{\mathbb{R}} \log \frac{1}{|z-x|}\rho(x)dx + V(z) = \ell_0, \qquad z \in \text{supp } \rho, \tag{4.51}$$

$$2\int_{\mathbb{R}} \log \frac{1}{|z-x|}\rho(x)dx + V(z) > \ell_0, \qquad z \notin \text{supp } \rho. \tag{4.52}$$

These variational conditions are proven in Appendix A; let us see how the variational conditions apply to the situation at hand. If we take $\rho(z)$ to be the equilibrium measure (i.e., the measure such that the above variational conditions are met), then $g(z)$ is just the analytic completion of the potential of this measure. The jump matrix for $\mathbf{U}(z)$ then reads, for $z \in \text{supp } \rho$,

$$\begin{pmatrix} e^{-n[g_+(z)-g_-(z)]} & e^{-n\ell_0} \\ 0 & e^{n[g_+(z)-g_-(z)]} \end{pmatrix} \tag{4.53}$$

So, the diagonal part of the jump is rapidly oscillating, and term in the upper diagonal is constant; this is almost exactly the situation we met in Example (4.11). A slight modification of the above RHP puts the jump into the desired form; we set

$$\mathbf{T}(z) := e^{\frac{n}{2}\ell_0\sigma_3}\mathbf{U}(z)e^{-\frac{n}{2}\ell_0\sigma_3} = e^{\frac{n}{2}\ell_0\sigma_3}\mathbf{Y}(z)e^{-n[\frac{1}{2}\ell_0+g(z)]\sigma_3}; \tag{4.54}$$

[2] The appearance of logarithmic potentials in the asymptotic theory of orthogonal polynomials can be traced back to E.A. Rakhmanov [93], although it is interesting to note that the earlier work of the Saclay school of theoretical physics [19] makes implicit use of the same techniques, albeit far less rigorously, and with no reference to orthogonal polynomials.

this is the first transformation in the steepest descent method. Since $\det e^{\frac{n}{2}\ell_0\sigma_3} = 1$, we again have that the transformation is invertible. The new function $\mathbf{T}(z)$ satisfies the following RHP:

Problem 4.21. *The function* $\mathbf{T}(z)$ *is analytic in* $\mathbb{C}\backslash\mathbb{R}$*, and is the unique solution to the following Riemann-Hilbert problem:*

$$\begin{cases} \mathbf{T}_+(z) = \mathbf{T}_-(z) \begin{pmatrix} e^{-n[g_+(z)-g_-(z)]} & e^{-n[V(z)-g_+(z)-g_-(z)-\ell_0]} \\ 0 & e^{n[g_+(z)-g_-(z)]} \end{pmatrix}, & z \in \mathbb{R}, \\ \mathbf{T}(z) = \mathbb{I} + \mathcal{O}(z^{-1}), & z \to \infty. \end{cases} \tag{4.55}$$

Proof. The conjugation of $\mathbf{U}(z)$ by $e^{\frac{n}{2}\ell_0\sigma_3}$ does not have any impact on the asymptotics. The jump matrix of $\mathbf{T}(z)$ can be calculated in a similar manner as before: one finds that the jump of $\mathbf{T}(z)$ is just the jump of $\mathbf{U}(z)$ conjugated by $e^{\frac{n}{2}\ell_0\sigma_3}$. □

There are two key observations. The first is that the transformation $\mathbf{Y} \mapsto \mathbf{T}$ yields an identity-normalized RHP. Secondly, let us observe the effect of this transformation on the jump matrix. If $z \in \mathbb{R} \setminus \operatorname{supp} \rho$, then $g_+(z) = g_-(z)$, as $g(z)$ is continuous away from the support of ρ. Thus, the jump matrix takes the form

$$\begin{pmatrix} 1 & e^{-n[V(z)-g_+(z)-g_-(z)-\ell_0]} \\ 0 & 1 \end{pmatrix}, \qquad z \in \mathbb{R} \setminus \operatorname{supp} \rho. \tag{4.56}$$

The variational inequality (4.52) tells us that $V(z) - g_+(z) - g_-(z) - \ell_0 > 0$, and so for n sufficiently large, the jump matrix is exponentially close to the identity matrix, away from the support of ρ. On the other hand, if $z \in \operatorname{supp} \rho$, the jump matrix of $\mathbf{T}(z)$ takes the form

$$\begin{pmatrix} e^{-n[g_+(z)-g_-(z)]} & 1 \\ 0 & e^{n[g_+(z)-g_-(z)]} \end{pmatrix}, \qquad z \in \operatorname{supp} \rho, \tag{4.57}$$

which follows from the variational equality (4.51). This is precisely the form of the jump matrix of (4.11).

4.3.2 The Second Transformation

In the second transformation, we open a lens around the support of the equilibrium measure ρ, as in Example (4.11). Define $\theta(z) := \frac{1}{2\pi i}[g_+(z) - g_-(z)]$. Then, for $z \in \text{supp } \rho$, the jump matrix of $\mathbf{T}$ takes the form

$$J_{\mathbf{T}}(z) = \begin{pmatrix} e^{-2\pi i n\theta(z)} & 1 \\ 0 & e^{2\pi i n\theta(z)} \end{pmatrix}, \qquad z \in \text{supp } \rho. \tag{4.58}$$

Since $\rho(z) > 0$ the function $\theta(z)$ is decreasing along the real axis, i.e. $\frac{\partial\theta}{\partial x} < 0$, for any point in supp ρ. Write $\theta_\pm(z) := u(z) + iv_\pm(z)$ to be the analytic continuation of $\theta(z)$ to regions just above and below the support of $\rho(z)$, respectively (note that $v_\pm(z) \to 0$ as $z \to$ a point in the support). By the Cauchy-Riemann equations, we have that

$$\pm\frac{\partial v_\pm}{\partial y} = \frac{\partial u}{\partial x} < 0, \tag{4.59}$$

this implies that, just above supp ρ, the function $e^{-2\pi i n\theta_+(z)}$ is exponentially decreasing in n, and similarly, just below supp ρ, the function $e^{2\pi i n\theta_-(z)}$ is exponentially decreasing in n. Letting G_+, G_- be lens-shaped regions above and below the support of ρ, as depicted in Figure (4.3), we define the matrix $\mathbf{S}$ by

$$\mathbf{S}(z) = \begin{cases} \mathbf{T}(z)\begin{pmatrix} 1 & 0 \\ e^{2\pi i n\theta_+(z)} & 1 \end{pmatrix}, & z \in G_+, \\ \mathbf{T}(z)\begin{pmatrix} 1 & 0 \\ e^{2\pi i n\theta_-(z)} & 1 \end{pmatrix}, & z \in G_-, \\ \mathbf{T}(z), & \textit{otherwise.} \end{cases} \tag{4.60}$$

It then follows that $\mathbf{S}(z)$ is analytic in $\mathbb{C}\backslash\Gamma$, where Γ consists of the real axis unioned with the lens boundaries, and is the unique solution to the following Riemann-Hilbert problem:

Problem 4.22. $\mathbf{S}(z)$ *is analytic in* $\mathbb{C} \setminus \Gamma$*, where* Γ *consists of the real axis unioned with the lens boundaries, and satisfies*

$$\mathbf{S}_+(z) = \mathbf{S}_-(z) \begin{cases} \begin{pmatrix} 1 & 0 \\ e^{-2\pi i n\theta_+(z)} & 1 \end{pmatrix}, & z \in \gamma_+, \\ \begin{pmatrix} 1 & 0 \\ e^{2\pi i n\theta_-(z)} & 1 \end{pmatrix}, & z \in \gamma_-, \\ \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, & z \in supp\ \rho, \\ \begin{pmatrix} 1 & e^{-n[V(z)-g_+(z)-g_-(z)-\ell_0]} \\ 0 & 1 \end{pmatrix} & z \in \mathbb{R} \setminus supp\ \rho. \end{cases} \tag{4.61}$$

$$\mathbf{S}(z) = \mathbb{I} + \mathcal{O}(z^{-1}), \qquad z \to \infty. \tag{4.62}$$

Proof. This follows immediately from Proposition (4.9) and Example(4.11). □

From here, it seems as though we are almost done; all of the jumps of $\mathbf{S}(z)$ are exponentially close to the identity matrix; the only thing left to correct is the constant jump on supp ρ. We stress that what remains is a Riemann-Hilbert problem with *constant* jumps, and thus no longer depends on $V(z)$ or n. Our next task is to ignore the exponentially small jumps of $\mathbf{S}(z)$, and try to approximate $\mathbf{S}(z)$ by the solution to the remaining constant jump RHP. This is the essence of the final transformation.

4.3.3 The Final Transformation

We again point out that we have made the ansatz that supp $\rho = [\alpha, \beta]$, i.e. is connected. This is really the first place where this assertion becomes relevant. If we ignore the exponentially small jumps of $\mathbf{S}(z)$, we are left with the following Riemann-Hilbert problem for a 2×2 matrix valued function $M(z)$:

Problem 4.23. *Global Parametrix RHP. Find a* 2×2 *matrix-valued function* $M(z)$*, analytic in* $\mathbb{C} \setminus [\alpha, \beta]$*, such that*

$$\begin{cases} M_+(z) = M_-(z) \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, & z \in [\alpha, \beta], \\ M(z) = \mathbb{I} + \mathcal{O}(z^{-1}), & z \to \infty. \end{cases} \tag{4.63}$$

This Riemann-Hilbert problem is often called the *global parametrix*. Our goal is to try and solve this Riemann-Hilbert problem as explicitly as possible. This is indeed possible to do; the key observation here is that the jump matrix is diagonalizable:

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = U^\dagger \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} U, \tag{4.64}$$

where $U := \frac{1}{\sqrt{2}}\left(\begin{smallmatrix} i & 1 \\ -i & 1 \end{smallmatrix}\right)$. Then, using the conjugation technique we learned in Proposition (4.12), the matrix $F(z) := UM(z)U^\dagger$ is the solution to the Riemann-Hilbert problem

$$\begin{cases} F_+(z) = F_-(z) \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, & z \in [\alpha, \beta], \\ F(z) = \mathbb{I} + \mathcal{O}(z^{-1}), & z \to \infty. \end{cases} \tag{4.65}$$

The diagonal form of the jump matrix implies that the matrix equations "decouple" into four scalar Riemann-Hilbert problems for the entries of $F(z)$. The solution can readily be found by means of the Plemelj formulae:

$$F(z) = \left(\frac{z-\beta}{z-\alpha}\right)^{\frac{1}{4}\sigma_3}; \tag{4.66}$$

This can be checked by direct verification. Thus, the solution to the global parametrix is

$$M(z) = U^\dagger \left(\frac{z-\beta}{z-\alpha}\right)^{\frac{1}{4}\sigma_3} U. \tag{4.67}$$

We are almost done! If we multiply $\mathbf{S}(z)$ on the right by the matrix $M^{-1}(z)$, the constant jump on the interval $[\alpha, \beta]$ cancels, and we have a good approximate solution to the Riemann-Hilbert problem, as all remaining jumps are exponentially small. Unfortunately, this is only partially true: the global parametrix matches $\mathbf{S}(z)$ well only away from the turning points $z = \alpha, \beta$. However, it is a bad approximation near these points, as it has a $\frac{1}{4}$-root singularity there. Therefore, we need to search for a better "local" model solution near the turning points; with this idea in mind, we define small discs D_α, D_β around $z = \alpha, \beta$, respectively. These local models will be called *local parametrices*. The local parametrices must have the following properties:

- The local parametrices must match the jumps of $\mathbf{S}(z)$ *exactly* inside the discs D_α, D_β,

- The local parametrices must agree with the global parametrix M(z) on the boundaries of the discs D_α, D_β.

First, we will need to know the local behavior of the function $\theta(z)$ near the endpoints of the support. This is a standard calculation in potential theory, and is performed in Appendix A. One finds that, generically near the turning points,

$$2\pi\theta(z) = c_0(z-\beta)^{3/2}[1+\mathcal{O}(z-\beta)], \qquad z\to\beta, \tag{4.68}$$

and similarly near $z=\alpha$ with β replaced with α in the above formula. For now, we focus on the disc at $z=\beta$; the computations at $z=\alpha$ are similar. Then, the function

$$\xi(z) := -(3/4)^{-2/3} n^{2/3} (2\pi\theta(z))^{2/3} \tag{4.69}$$

defines a conformal mapping of the disc D_β to a neighborhood of the origin in the ξ-plane. Moreover, all of the jumps of $\mathbf{S}(z)$ incident to $z=\beta$ can be rewritten in terms of the local coordinate ξ. We have that, in the ξ-plane:

$$\mathbf{S}_+ = \mathbf{S}_- \begin{cases} \begin{pmatrix} 1 & e^{-\frac{4}{3}\xi^{-3/2}} \\ 0 & 1 \end{pmatrix}, & \xi > 0, \\ \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, & \xi < 0, \\ \begin{pmatrix} 1 & 0 \\ e^{\frac{4}{3}\xi^{3/2}} & 1 \end{pmatrix}, & \xi \in \xi(\gamma_\pm). \end{cases} \tag{4.70}$$

Because of the freedom we have in choosing the lens boundaries, we can choose $\gamma_\pm$ so that their images in the ξ-plane are the rays $\arg\xi = \pm\frac{2\pi}{3}$, respectively. On the other hand, the global parametrix $M(z)$ can also be written locally in terms of the coordinate ξ:

$$M(z) = h(z)\xi^{\frac{1}{4}\sigma_3}U, \tag{4.71}$$

for some non-vanishing matrix-valued analytic function $h(z)$ in a neighborhood of $z=\beta$ (note: the function $h(z)$ depends on n!). Of course, left multiplication by an analytic function will not change the jumps of any Riemann-Hilbert problem with right jumps. We therefore pose the following Riemann-Hilbert problem:

Problem 4.24. *Construct a 2×2 matrix-valued analytic function $P_\beta(\xi)$, such that*

$$P_{\beta,+}(\xi) = P_{\beta,-}(\xi) \begin{cases} \begin{pmatrix} 1 & e^{-\frac{4}{3}\xi^{-3/2}} \\ 0 & 1 \end{pmatrix}, & \xi > 0, \\ \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, & \xi < 0, \\ \begin{pmatrix} 1 & 0 \\ e^{\frac{4}{3}\xi^{3/2}} & 1 \end{pmatrix}, & \arg\xi = \pm\frac{2\pi}{3}, \end{cases} \tag{4.72}$$

$$P_\beta(\xi) = [\mathbb{I} + \mathcal{O}(\xi^{-1})]\xi^{\frac{1}{4}\sigma_3}U, \qquad \xi \to \infty. \tag{4.73}$$

If we can construct a solution to the Riemann-Hilbert problem for $P_\beta(\xi)$, then the function $\Psi_\beta(\xi) := h(z)P_\beta(\xi)$ matches the jumps of $\mathbf{S}(z)$ exactly near $z = \beta$, and asymptotically behaves like the global parametrix $M(z)$, as $n \to \infty$ (equivalently, by how we defined ξ, as $\xi \to \infty$). As it turns out, a simple transformation takes the above RHP into a more familiar form: If we define $\hat{P}_\beta(\xi) = P_\beta(\xi)e^{-\frac{2}{3}\xi^{3/2}\sigma_3}$, we reach the following constant-jump Riemann-Hilbert problem:

$$\hat{P}_{\beta,+}(\xi) = \hat{P}_{\beta,-}(\xi) \begin{cases} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, & \xi > 0, \\ \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, & \xi < 0, \\ \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, & \arg\xi = \pm\frac{2\pi}{3}, \end{cases} \tag{4.74}$$

$$\hat{P}_\beta(\xi) = [\mathbb{I} + \mathcal{O}(z^{-1})]\xi^{\frac{1}{4}\sigma_3}Ue^{-\frac{2}{3}\xi^{3/2}\sigma_3}, \qquad \xi \to \infty. \tag{4.75}$$

The asymptotic expression for $\hat{P}_\beta(\xi)$ is familiar: it looks a lot like the asymptotic expansion of the *Airy function* $\mathrm{Ai}(\xi)$:

$$\mathrm{Ai}(\xi) = \frac{\xi^{-1/4}}{2\sqrt{\pi}}e^{-\frac{2}{3}\xi^{3/2}}[1 + \mathcal{O}(\xi^{-3/2})], \qquad \xi \to \infty. \tag{4.76}$$

And so, we can hope to write the solution to the local parametrix in terms of the Airy function $\mathrm{Ai}(\xi)$. If we were not so clever in recognizing this expansion, it would still have been possible to come to the same conclusion a different way: notice that the jumps of $\hat{P}_\beta$ and $\frac{d\hat{P}_\beta}{d\xi}$ agree, and so the function

$$B(\xi) := \frac{d\hat{P}_\beta}{d\xi}\hat{P}_\beta^{-1} \tag{4.77}$$

extends to an entire function in the ξ plane. The usual Liouville argument then gives us that $B(\xi) = B_1\xi + B_0$, for some matrices B_1 and B_0. This system reduces to a first order differential equation for one of the entries, which eventually reads

$$y''(\xi) = \xi y(\xi), \tag{4.78}$$

so we would have eventually come to this conclusion anyways (albeit by a longer computation). With this in mind, we set

$$y_0(\xi) = \mathrm{Ai}(\xi), \qquad y_1(\xi) = \omega\,\mathrm{Ai}(\omega\xi), \qquad y_2(\xi) = \omega^2\,\mathrm{Ai}(\omega^2\xi), \tag{4.79}$$

where $\mathrm{Ai}(\xi)$ is the Airy function. It is well known that $y_j''(\xi) + y_j(\xi) = 0$, $j = 1, 2, 3$, and that $y_0(\xi) + y_1(\xi) = y_2(\xi) = 0$. We define a 2×2 matrix-valued function $\Psi^{Airy}(\xi)$ by

$$\Psi^{Airy}(\xi) = \begin{cases} \begin{pmatrix} y_0(\xi) & -y_2(\xi) \\ y_0'(\xi) & -y_2'(\xi) \end{pmatrix}, & 0 < \arg\xi < \frac{2\pi}{3}, \\ \begin{pmatrix} -y_1(\xi) & -y_2(\xi) \\ -y_1'(\xi) & -y_2'(\xi) \end{pmatrix}, & \frac{2\pi}{3} < \arg\xi < \pi, \\ \begin{pmatrix} -y_2(\xi) & y_1(\xi) \\ -y_2'(\xi) & y_1'(\xi) \end{pmatrix}, & -\pi < \arg\xi < -\frac{2\pi}{3}, \\ \begin{pmatrix} y_0(\xi) & y_1(\xi) \\ y_0'(\xi) & y_1(\xi) \end{pmatrix}, & -\frac{2\pi}{3} < \arg\xi < 0. \end{cases} \tag{4.80}$$

The function $\Psi^{Airy}(\xi)$ is then analytic everywhere except the rays $\arg\xi = \pm\frac{2\pi}{3}$ and $\mathbb{R}$, and is in fact the unique solution to the following Riemann Hilbert problem:

$$\begin{cases} \Psi^{Airy}(\xi) \text{ is analytic off of the rays } \arg\xi = \pm\dfrac{2\pi}{3} \text{ and } \mathbb{R}, \\ \Psi^{Airy}_+(\xi) = \Psi^{Airy}_-(\xi)\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right), & \xi > 0, \\ \Psi^{Airy}_+(\xi) = \Psi^{Airy}_-(\xi)\left(\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right), & \arg\xi = \pm\frac{2\pi}{3}, \\ \Psi^{Airy}_+(\xi) = \Psi^{Airy}_-(\xi)\left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right), & \xi \in (-\infty, 0), \\ \Psi^{Airy}(\xi) = \frac{i}{2\sqrt{\pi}}\xi^{-\frac{1}{4}\sigma_3}U\left[\mathbb{I} + \mathcal{O}(\xi^{-2/3})\right]e^{-\frac{2}{3}\xi^{3/2}\sigma_3}, & \xi \to \infty. \end{cases} \tag{4.81}$$

The contours/jumps of the Airy parametrix are shown in Figure (4.4).

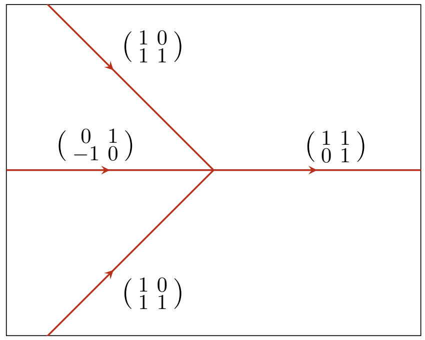


Figure 4.4. The jumps and contours of the Airy parametrix.

We see from here that $\hat{P}_\beta(\xi) = -2i\sqrt{\pi}\xi^{\frac{1}{2}\sigma_3}\Psi^{Airy}(\xi)$, and therefore that

$$P_\beta(\xi) = -2i\sqrt{\pi}\xi^{\frac{1}{2}\sigma_3}\Psi^{Airy}(\xi)e^{\frac{2}{3}\xi^{3/2}\sigma_3}, \tag{4.82}$$

and therefore that

$$\Psi_\beta(z) = -2i\sqrt{\pi}h(z)[\xi(z)]^{\frac{1}{2}\sigma_3}\Psi^{Airy}(\xi(z))e^{\frac{2}{3}[\xi(z)]^{3/2}\sigma_3}. \tag{4.83}$$

Nearly identical computations yield the parametrix at $z = \alpha$, we do not repeat the calculations here, but call the parametrix there $\Psi_\alpha(\xi)$. Finally, we can make the last transformation. Set

$$\mathbf{R}(\xi) = \begin{cases} \mathbf{S}(z)M^{-1}(z), & z \in \mathbb{C} \setminus (D_\alpha \cup D_\beta), \\ \mathbf{S}(z)\Psi_\beta^{-1}(z), & z \in D_\beta, \\ \mathbf{S}(z)\Psi_\alpha^{-1}(z), & z \in D_\alpha. \end{cases} \tag{4.84}$$

Note that, by construction, the matrices $M(z)$, $\Psi_\alpha(z)$, and $\Psi_\beta(z)$ all have determinant identically equal to 1, and so this transformation is invertible. We claim that $\mathbf{R}(z)$ is the solution to a small-norm Riemann-Hilbert problem:

Proposition 4.25. *The function* $\mathbf{R}(z)$ *is the solution to the following Riemann-Hilbert problem:*

$$\mathbf{R}_+(z) = \mathbf{R}_-(z) \begin{cases} \mathbb{I} + \mathcal{O}(e^{-n}), z \notin D_\alpha, D_\beta, \\ \mathbb{I} + \mathcal{O}(n^{-1}), z \in D_\alpha \cup D_\beta \end{cases} \tag{4.85}$$

$$\mathbf{R}(z) = \mathbb{I} + \mathcal{O}(z^{-1}), \qquad z \to \infty. \tag{4.86}$$

Proof. The jumps of $\Psi_\alpha(z)$, $\Psi_\beta(z)$ were chosen so as to exactly match the jumps of $\mathbf{S}(z)$; thus, inside the disc, there are no jumps, and $\mathbf{R}(z)$ extends to an analytic function there. Similarly, on the segment $[\alpha, \beta] \setminus (D_\alpha \cup D_\beta)$, the global parametrix matches the jump of $\mathbf{S}(z)$ exactly, and so the jump there *is* the identity matrix. The jumps of $\mathbf{S}(z)$ on the lens boundaries $\gamma_\pm$ and on the rest of the real line are not matched, but are exponentially close to the identity, so that

$$\mathbf{R}_+(z) = \mathbf{R}_-(z)[\mathbb{I} + \mathcal{O}(e^{-n})], \qquad z \notin D_\alpha, D_\beta.$$

All that remains to check is the jump on the boundaries of the discs D_α, D_β. This is tedious, but follows from how we constructed the local parametrices: one finds that

$$\mathbf{R}_+(z) = \mathbf{R}_-(z)[\mathbb{I} + \mathcal{O}(n^{-1})], \qquad n \to \infty, z \in \partial D_\alpha \cup D_\beta. \tag{4.87}$$

□

Thus, we have reached a small-norm RHP; by small norm theory, the solution to the Riemann-Hilbert problem for $\mathbf{R}(z)$ exists, and has the expansion

$$\mathbf{R}(z) = \mathbb{I} + \mathcal{O}(n^{-1}), \qquad n \to \infty. \tag{4.88}$$

Since the sequence of transformations we found were all invertible, we can run them backwards, and obtain an exact asymptotic expression for the original Riemann-Hilbert problem for $\mathbf{Y}(z)$.

4.3.4 Universality in the 1-Matrix Model.

Let us briefly discuss some of the consequences of the above analysis. In particular, we can derive a *universality* result about the local laws of the 1-matrix model. We shall not prove any of the statements we make below, but emphasize that they can be demonstrated by explicit use of the asymptotic formulas derived from steepest descent analysis; we indicate how one should go about calculating these quantities.

Proposition 4.26. *The limiting density of eigenvalues of the* 1*-matrix model in external field* $V(x)$ *is given by the measure* $\rho(x)dx$*, where* $\rho(x)dx$ *is the unique minimizer among all unit Borel measures supported on* $\mathbb{R}$ *of the functional*

$$E(\rho) := \iint_{\mathbb{R}\times\mathbb{R}} \log\frac{1}{|z-x|}\rho(x)\rho(z)dzdx + \int_{\mathbb{R}} V(z)\rho(z)dz. \tag{4.89}$$

Proof. As stated before, we only indicate how to prove such a statement from the asymptotic formulae we derived. Note that the density of eigenvalues can be computed at finite N using the Christoffel-Darboux kernel:

$$\rho_N(x) = \frac{1}{N} K_N(x,x);$$

since the Christoffel-Darboux kernel is explicit in terms of $\mathbf{Y}(z)$, we can calculate the large N limit of $\rho_N(x)$ as

$$\rho(x) = \lim_{N\to\infty} \rho_N(x) = \lim_{N\to\infty} \frac{1}{N} K_N(x,x) = \lim_{N\to\infty} -\lim_{y\to x} \frac{\left[\mathbf{Y}_N^{-1}(y)\mathbf{Y}_N(x)\right]_{21}}{2\pi i N(y-x)}.$$

□

For the remainder of this section, we let $\rho(x)$ denote the density of eigenvalues of the 1-matrix model in external field $V(x)$. We are now able to state a universality result for the 1-matrix model:

Proposition 4.27. *1. (Universality in the Bulk.) Let x be a point on the interior of the support of $\rho(x)dx$, and ξ, ζ lying in a compact subset of $\mathbb{R}$. Then, the limit*

$$K(\xi,\zeta) := \lim_{N\to\infty} \frac{K_N\left(x + \frac{\rho(x)}{N}\xi, x + \frac{\rho(x)}{N}\zeta\right)}{K_N(x,x)} \tag{4.90}$$

exists, and is independent of the choice of potential $V(x)$ (as long as the matrix integral is convergent). Moreover, $K(\xi,\zeta)$ is given by

$$K(\xi,\zeta) = \frac{\sin(\xi-\zeta)}{\xi-\zeta}. \tag{4.91}$$

2. (Universality of the Edge.) Let x be a point on the boundary of the support of $\rho(x)dx$, and assume $\rho(x)$ is not critical (i.e., $\rho(z) \sim c_0(z-x)^{1/2}[1+o(1)], z \to x$). Let ξ, ζ lie in a compact subset of $\mathbb{R}$. Then, the limit

$$K(\xi,\zeta) := \lim_{N\to\infty} \frac{K_N\left(x + \frac{\xi\rho(x)}{N}, x + \frac{\zeta\rho(x)}{N}\right)}{K_N(x,x)} \tag{4.92}$$

exists, and is given by $K(\xi,\zeta)$ is given by

$$K(\xi,\zeta) = \frac{\operatorname{Ai}(\xi)\operatorname{Ai}'(\zeta) - \operatorname{Ai}(\zeta)\operatorname{Ai}'(\xi)}{\xi-\zeta}, \tag{4.93}$$

where $\operatorname{Ai}(\xi)$ is the Airy function.

Proof. Again, since the Christoffel-Darboux kernel is given explicitly in terms of the solution to the Riemann-Hilbert problem $\mathbf{Y}(z)$, and we have the asymptotics of $\mathbf{Y}(z)$ explicitly from the steepest descent analysis, all one must do is apply the formula we derived for the asymptotics for $\mathbf{Y}(z)$. We remark that the edge universality result is indeed *universal*, since (as we can see in Appendix A, the behavior of the support of the equilibrium measure at an endpoint of the support is generically $\rho(z) \sim c_0(z-x)^{1/2}[1+o(1)], z \to x$, as stated in the hypothesis of the proposition. □

4.4 A Riemann-Hilbert Problem For Biorthogonal Polynomials.

Here, we formulate a Riemann-Hilbert problem for biorthogonal polynomials. We first must establish a Riemann-Hilbert problem for Type-II multiple orthogonal polynomials. As a definition (cf. [1]), we say a polynomial $p_{\vec{n}}(x)$ is a *multiple orthogonal polynomial of type II* with respect to the weights $w_1(x), w_2(x), ..., w_N(x) > 0$ (supported on $\mathbb{R}$, for example), of index $\vec{n} := (n_1, ..., n_N)$, if the following conditions are satisfied:

$$\deg p_{\vec{n}} \leq |\vec{n}| = \sum_{k=1}^{N} n_k \tag{4.94}$$

$$\int p_{\vec{n}}(x) x^j w_k(x) dx = 0, \qquad j = 0, ..., n_k, \qquad k = 1, ..., N. \tag{4.95}$$

Of course, when $N = 1$, we reduce back to the case of classical orthogonal polynomials with respect to the weight $w_1(x)$. It is known that multiple orthogonal polynomials satisfy a Riemann-Hilbert problem of their own.

Problem 4.28. *Riemann-Hilbert problem for type II multiple orthogonal polynomials. Find an* $(N+1) \times (N+1)$ *matrix-valued function* $Y : \mathbb{C} \to M_{N+1}(\mathbb{C})$, *analytic in* $\mathbb{C} \setminus \mathbb{R}$, *which satisfies the conditions*

$$\begin{cases} Y_+(z) = Y_-(z) \begin{pmatrix} 1 & w_1(z) & w_2(z) & \dots & w_N(z) \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}, & z \in \mathbb{R}, \\ Y(z) = \left[\mathbb{I} + \mathcal{O}(z^{-1})\right] \textit{diag}\ (z^{|\vec{n}|}, z^{-n_1}, z^{-n_2}, ..., z^{-n_N}), & z \to \infty \end{cases} \tag{4.96}$$

The solution to Problem (4.28) is given in terms of multiple orthogonal polynomials, as the next Proposition shows.

Proposition 4.29. *The solution to Problem (4.28) is*

$$Y(z) = \begin{pmatrix} P_{\vec{n}}(z) & \vec{R}_{\vec{n}}(z) \\ d_1 P_{\vec{n}-\vec{e}_1}(z) & d_1 \vec{R}_{\vec{n}-\vec{e}_1}(z) \\ \vdots & \vdots \\ d_N P_{\vec{n}-\vec{e}_N}(z) & d_N \vec{R}_{\vec{n}-\vec{e}_N}(z) \end{pmatrix}, \tag{4.97}$$

where $P_{\vec{n}}(x)$ is the monic multiple orthogonal polynomial with respect to the weights $w_1(x), w_2(x), ..., w_N(x) > 0$, of index $\vec{n} := (n_1, ..., n_N)$, $\vec{e}_k$ is the unit vector with a 1 in the k^{th} component and 0's otherwise, and $\vec{R}_{\vec{m}}(z) = (R_{\vec{m},1}, R_{\vec{m},2}, ..., R_{\vec{m},N})$ is the column vector function with components

$$R_{\vec{m},j}(z) = \int P_{\vec{m}}(x) w_j(x) \frac{dx}{z-x}, \qquad j = 1, ..., N, \tag{4.98}$$

and the constants d_j are defined as

$$\frac{1}{d_j} = \int x^{n_j - 1} P_{\vec{n}-\vec{e}_j}(x) w_j(x) \frac{dx}{z-x}. \tag{4.99}$$

We do not prove this proposition, as it is almost identical in style to the proof of the equivalent proposition for orthogonal polynomials, from Problem (4.16). Now by our observations from Chapter 3, biorthogonal polynomials have reformulation as multiple orthogonal polynomials (cf. Theorem (3.14), Chapter 3). Thus, we also have a Riemann-Hilbert problem for biorthogonal polynomials; namely, the problem (4.28). This RHP is our main tool in the next chapter, where we shall apply it to the 2-matrix model with quartic interactions.

CHAPTER 5
THE ISING MODEL COUPLED TO 2D GRAVITY.

5.1 Introduction.

This chapter is where we shall prove the main results of this thesis. In this section, we summarize what we have learned about the Ising model and random matrices, and provide an overview of what we are going to prove in this chapter.

5.1.1 The Ising Model Coupled to Gravity and a 2-Matrix Model.

The 2-dimensional Ising model has long been a source of interest in statistical physics, as it is an exactly solvable lattice model which exhibits a 2^{nd} order phase transition at finite temperature. The model describes a ferromagnet with only nearest-neighbor interactions, and can be defined on any graph $G := (V, E)$ with vertices V and edges E as follows. The Hamiltonian for the ferromagnetic Ising model is a functional on maps $\sigma : V \to \{\pm 1\}$, and is defined as

$$H(\sigma) = -\sum_{(x,y)\in E} \sigma(x)\sigma(y). \tag{5.1}$$

The partition function for this model is defined to be

$$Z_G(\beta) := \sum_{\sigma} e^{-\beta H(\sigma)}, \tag{5.2}$$

where $\beta > 0$ is a parameter called the *inverse temperature*, and the sum is taken over all maps $\sigma : V \to \{\pm 1\}$, so that there are $2^{|V|}$ terms in the summation. In other words, we are considering the Boltzmann distribution on the system at temperature β^{-1}. To say we have *exactly solved* the model in this context is to say that we have found an analytic expression for the free energy of the model, that is, an explicit expression for the quantity

$$F_G(\beta) = -\frac{1}{\beta} \log Z_G(\beta). \tag{5.3}$$

If the graph is of infinite size (i.e., the number of vertices is countably infinite), one is instead interested in the free energy *per unit site*; that is to say,

$$f_G(\beta) := \lim_{|G|\to\infty} \frac{1}{|G|} F_G(\beta), \tag{5.4}$$

where the limit is taken in an appropriate sense.

The model was introduced by E. Ising [60], although he only studied the 1-dimensional model, and incorrectly conjectured that the model did *not* exhibit a phase transition. It was not until over 20 years later that L. Onsager [86] announced that the 2-dimensional model indeed exhibited a phase transition[1]. The universality of the critical exponents appearing in the 2-dimensional model were verified for various lattices (cf. R.J. Baxter's book [7], Chapter 11, and references therein), but these considerations were markedly limited by the fact that computations could be performed explicitly only for choices of fairly regular lattices (e.g. square lattice, triangular lattice, etc.). The advent of conformal field theory (CFT) techniques to describe 2-dimensional critical phenomena by A.A. Belavin, A.M. Polyakov, and A.B. Zamolodchikov [8] in 1984 revived interest in the 2D Ising model, and prompted V. Kazakov to consider the Ising model coupled to 2-dimensional gravity; that is, the Ising model on a random lattice. More precisely, he found that the 2D Ising model on a random 4-regular planar graph could be described by the large N-limit of a 2-matrix model. Here, 4-*regular* means that each vertex is connected to four edges. The partition function for the Ising model on a random 4-regular planar graph with n-vertices is defined as

$$\mathcal{Z}_n(\beta) = \sum_{\substack{G:|G|=n,\\ G\ planar}} Z_G(\beta), \tag{5.5}$$

where the sum is taken over all 4-regular graphs with n vertices. In [70], V. Kazakov considered the formal generating function

$$\mathcal{Z}(\tau, t) = \sum_{n\in\mathbb{N}} \left(\frac{-4t\tau}{(1-\tau^2)^2}\right)^n \mathcal{Z}_n(\beta), \tag{5.6}$$

where $\tau := e^{-2\beta}$, and t is a parameter. Kazakov then considered the partition function of the 2-matrix model with quartic interactions:

$$Z_{matrix}(\tau, t; n, N) := \iint dXdY \exp\left\{N \operatorname{tr}[\tau XY - \frac{1}{2}X^2 - \frac{1}{2}Y^2 - \frac{t}{4}X^4 - \frac{t}{4}Y^4]\right\}, \tag{5.7}$$

[1] In fact, Onsager himself did not provide a proof, instead only furnishing the expression for the free energy. It took until 1952 for a fully rigorous proof to be published by C. Yang [108].

where X, Y are an $n \times n$ matrices, $N > 0$ is a parameter, and dX, dY represent the Haar measure on the space of $n \times n$ Hermitian matrices. The *free energy* of this matrix model is defined to be

$$F(\tau, t; n, N) := \frac{1}{N^2} \log Z_{matrix}(\tau, t; n, N). \tag{5.8}$$

Kazakov then demonstrated that the generating function (5.6) is equivalent to the planar ($n \to \infty$) limit of the free energy of the 2-matrix model (5.8), up to an additive constant independent of τ, t:

$$\lim_{n \to \infty} F(\tau, t; n, n) = \mathcal{Z}(\tau, t). \tag{5.9}$$

Kazakov's description of the critical point turned out to be in direct agreement with the newly predicted results of V.G. Kniznik, Polyakov and Zamolodchikov [25, 72] arising from coupling certain CFTs (for the Ising model, the (4,3) minimal model) to matter; this is the so-called KPZ formula. Subsequent analysis for this model and closely related ones was performed in [17, 18]. It was later shown [18, 25, 34, 56] that the phase transition was locally well described by a particular solution to the 'string equation' for KdV_3. These developments happened concurrently with developments in the theory of 2D quantum gravity, in particular descriptions of matrix models of gravity. The 'pure gravity' situation, in which no matter is present, was the subject of [19], and was subsequently described rigorously in [47, 48], and more recently (in line with the language of this work) [10, 38].

Kazakov and his collaborators from theoretical physics used formal matrix integral techniques to obtain their results. No direct analytic derivation of this description of the Ising critical point has yet been written down in the literature, although some closely related models [36, 37, 39, 40] have been described rigorously. In this work, we provide a rigorous analysis of the critical point appearing in Kazakov's work, using steepest descent analysis for biorthogonal polynomials.

5.1.2 A Family of Biorthogonal Polynomials and Associated RHP

The 2-matrix model considered above, with partition function (5.7), can be studied using orthogonal polynomial-type methods. Using the Itzykson-Zuber integral over the unitary group (cf. [62, 80, 112], for example), one can convert the integral over the spaces of formal hermitian matrices to an integral over eigenvalue coordinates; the partition function in the eigenvalue coordinates reads

$$Z_{matrix}(\tau, t; n, N) = \tau^{\frac{-n(n-1)}{2}} C_{n,N} \iint \Delta(x)\Delta(y) \exp\left\{ N \sum_{i=1}^{n} (\tau x_i y_i - \frac{1}{2}x_i^2 - \frac{1}{2}y_i^2 - \frac{t}{4}x_i^4 - \frac{t}{4}y_i^4) \right\} \prod_{i=1}^{n} dx_i dy_i, \tag{5.10}$$

where $\Delta(x), \Delta(y)$ denote the usual Vandermonde determinants in the variables $\{x_i\}$, $\{y_i\}$, respectively, and $C_{n,N}$ is a constant independent of the parameters τ, t:

$$C_{n,N} = \frac{(2\pi)^{n(n-1)}}{\left(\prod_{p=1}^{n} p!\right)^2} \left(\prod_{p=1}^{n-1} p!\right) N^{-\frac{n(n-1)}{2}} = \frac{1}{(n!)^2} \frac{(2\pi)^{n(n-1)}}{\prod_{p=1}^{n-1} p!} N^{-\frac{n(n-1)}{2}}. \tag{5.11}$$

We are thus naturally led to consider the biorthogonal polynomials

$$\int_\Gamma \int_\Gamma p_k(z) q_j(w) \exp\left[n(\tau z w - V(z) - Q(w))\right] dz dw = h_j \delta_{kj}, \tag{5.12}$$

where

$$V(z) = Q(z) = \frac{1}{2} z^2 + \frac{t}{4} z^4. \tag{5.13}$$

In this language, the partition function (that is, the normalizing constant that makes (5.10) a probability measure) can be expressed in terms of the orthogonality coefficients:

$$Z_{matrix}(\tau, t; n, N) = n!\, \tau^{\frac{-n(n-1)}{2}} C_{n,N} \prod_{j=0}^{n-1} h_j(\tau, t) \tag{5.14}$$

The critical point considered by Kazakov [18, 70] corresponds to

$$\tau = \frac{1}{4}, \qquad t = -\frac{5}{72}. \tag{5.15}$$

This leads to an immediate issue: the partition function (5.14) does not converge for $t < 0$. Thus, we must consider an analytic continuation of the partition function in t. We analytically continue each of the $h_j(\tau, t)$ through the upper half plane,

$$h_j(\tau, t) \longrightarrow h_j(\tau, t e^{\pi i}) \tag{5.16}$$

This can be equivalently achieved by deforming the integration contour Γ, starting where t is positive (and so the expression is for the partition function is a convergent series, and thus analytic), and continuing by a rotation in the z-plane of the contours so that t is allowed to take negative values. The contour Γ is defined as starting from $e^{3\pi i/4} \cdot \infty$, and ending at $e^{-\pi i/4} \cdot \infty$. We remark that we have the freedom to later redefine the contour Γ locally, as long as we retain its asymptotic properties. We shall indeed redefine Γ in a more precise manner in the next section, in order to guarantee that certain inequalities are satisfied on it.

From here on, we thus consider the contour Γ as chosen before, and $t < 0$, $\tau > 0$.

As observed in [75], the biorthogonal polynomials defined by (5.12) admit a Riemann-Hilbert formulation, given as follows.

Consider the following Riemann-Hilbert problem (RHP) on Γ:

$$\mathbf{Y}_+(z) = \mathbf{Y}_-(z)\left[\mathbb{I} + e^{-NV(z)}\begin{pmatrix} 0 & f(z) & \frac{f'(z)}{N\tau} & \frac{f''(z)}{(N\tau)^2} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}\right], \qquad z \in \Gamma, \tag{5.17}$$

where

$$f(z) = \int_\Gamma \exp\left[N(\tau z w - V(w))\right] dw \tag{5.18}$$

The asymptotics of $\mathbf{Y}_n(z)$ are chosen to be

$$\mathbf{Y}(z) = \left[\mathbb{I} + \mathcal{O}\left(\frac{1}{z}\right)\right]\begin{pmatrix} z^n & 0 & 0 & 0 \\ 0 & z^{-n/3} & 0 & 0 \\ 0 & 0 & z^{-n/3} & 0 \\ 0 & 0 & 0 & z^{-n/3} \end{pmatrix}, \qquad |z| \to \infty, \tag{5.19}$$

where n is a multiple of 3. This restriction is taken for simplicity of exposition; the assumption is not essential. This RHP admits a unique exact solution, with the 1-1 entry of $\mathbf{Y}(z)$ being the degree n monic biorthogonal polynomial[2] defined by the relations (5.12).

The analytic continuation of the partition function of the matrix model (5.7) can be related to an isomonodromic $\boldsymbol{\tau}$-function, given in terms of $\mathbf{Y}(z)$ [9]. The $\boldsymbol{\tau}$-function is defined to be

$$d \log \boldsymbol{\tau}_n := \operatorname*{Res}_{z=\infty} \operatorname{tr}\left[\mathbf{Y}^{-1}(z)\, \mathbf{Y}'(z) d\Psi_0(z) \Psi_0^{-1}(z)\right], \tag{5.20}$$

where Ψ_0 is an explicit matrix function, and the differential is in the variables of deformation t, τ. The differential of the partition function considered by Kazakov is proportional to this isomonodromic-$\boldsymbol{\tau}$ function:

$$d \log Z_{matrix}(\tau, t; n, N) = d \log\left[\left(\frac{\tau}{t^2}\right)^{\frac{n}{2}\left(\frac{n}{3}-1\right)} \boldsymbol{\tau}_n\right]. \tag{5.21}$$

[2]Note that, in general, the relation (5.12) in fact defines *two* sets of polynomials $\{p_n(z)\}$, $\{q_n(w)\}$, which in general do not coincide. However, in the special case that the potentials $V(z) = Q(z)$ are the same, the sequences of polynomials are identical, i.e. $p_n(z) = q_n(z)$.

Thus, by computing the isomonodromic τ-function, we can compute all quantities of interest related to the original partition function of the 2-matrix model. In [70, 80], it was formally shown that

$$\lim_{n\to\infty} \frac{1}{n^2} \log \frac{Z_{matrix}(\tau,t;n,n)}{Z_{matrix}(\tau,0;n,n)} = \frac{1}{2}\log\frac{\tau z(\tau,t)}{2t} - \int_0^{z(\tau,t)} \frac{d\zeta}{\zeta}\left[k(\zeta;\tau,t) - \frac{1}{2}k^2(\zeta;\tau,t)\right]$$
$$- \log\frac{\tau}{2(1-\tau^2)} + 1, \tag{5.22}$$

where $k(\zeta;\tau,t)$ is defined as

$$k(\zeta;\tau,t) := \frac{\zeta}{t}\left[\frac{1}{(1-3\zeta)^2} - \tau^2 + 3\tau^2\zeta^2\right], \tag{5.23}$$

and $z = z(\tau,t)$ is implicitly determined as the unique solution of the fifth-order equation

$$t = z\left[\frac{1}{(1-3z)^2} - \tau^2 + 3\tau^2 z^2\right] \tag{5.24}$$

which satisfies

$$\lim_{t\to 0}\frac{z(\tau,t)}{t} = \frac{1}{(1-\tau^2)}. \tag{5.25}$$

The expression (5.22) may be further integrated explicitly; we prefer to leave it in this form.

One may calculate $Z_{matrix}(\tau,0;n,n)$ explicitly (cf. [81], Appendix A.49):

$$Z_{matrix}(\tau,0;n,N) = \tau^{\frac{-n(n-1)}{2}} C_{n,N} N^{-\frac{n(n-1)}{2}} \frac{(2\pi)^n}{\prod_{p=1}^{n-1} p!} \frac{\tau^{n(n-1)/2}}{(1-\tau^2)^{n^2/2}} \tag{5.26}$$

Using steepest descent analysis, we can compute the isomonodromic $\boldsymbol{\tau}$-function defined by (5.20), and consequentially, due to Equations (5.21), (5.26), we can prove rigorously the formula derived in [70, 80], using the below formula:

$$\lim_{n\to\infty}\frac{1}{n^2} d\log\frac{Z_{matrix}(\tau,t;n,n)}{Z_{matrix}(\tau,0;n,n)} = -\frac{dt}{3t} + \frac{d\tau}{6\tau} - \frac{1+\tau^2}{\tau(1-\tau^2)}d\tau + \lim_{n\to\infty}\frac{1}{n^2} d\log\boldsymbol{\tau}_n. \tag{5.27}$$

5.1.3 Notations and Overview of the Remainder of the Paper.

The rest of the paper is devoted to a Deift-Zhou steepest descent analysis [31] of $\mathbf{Y}$, i.e., a sequence of explicit invertible transformations

$$\mathbf{Y} \longmapsto \mathbf{X} \longmapsto \mathbf{U} \longmapsto \mathbf{T} \longmapsto \mathbf{S} \longmapsto \mathbf{R},$$

with the final Riemann-Hilbert problem for $\mathbf{R}$ has jumps that tend to the identity matrix as $n \to \infty$, uniformly and in L^2, with normalized behavior at infinity. We thus obtain an asymptotic expansion for $\mathbf{R}$, and thus, utilizing the invertiblity of the transformations, an asymptotic expansion of $\mathbf{Y}$. This can be used to write an expression for the partition function (5.5).

The remainder of the paper is organized as follows. Section §2 is devoted to the first transformation $\mathbf{Y} \mapsto \mathbf{X}$, the second transformation $\mathbf{X} \mapsto \mathbf{U}$ is performed in §3. The transformations $\mathbf{U} \mapsto \mathbf{T} \mapsto \mathbf{S}$ are completed in §4. In the transformation $\mathbf{U} \mapsto \mathbf{T}$, we open lenses off of the cut between the branch points; in the transformation $\mathbf{T} \mapsto \mathbf{S}$, we open lenses around the central branch cut. In §5, we construct the parametrices, and perform the final transformation $\mathbf{S} \mapsto \mathbf{R}$.

We will make use of some notations frequently; we establish them here for the convenience of the reader.

- Throughout, $\omega := e^{\frac{2\pi i}{3}} = -\frac{1}{2} + i\frac{\sqrt{3}}{2}$ is the principal third root of unity,
- We denote the 4×4 matrix with a 1 in the $(i,j)^{th}$ entry and zeros elsewhere by E_{ij},
- The third Pauli matrix $\sigma_3 := \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$. We will often write expressions such as $z^{C\sigma_3}$, or $e^{f(z)\sigma_3}$. These expressions are defined to be

$$z^{C\sigma_3} := \begin{pmatrix} z^C & 0 \\ 0 & z^{-C} \end{pmatrix}, \qquad e^{f(z)\sigma_3} := \begin{pmatrix} e^{f(z)} & 0 \\ 0 & e^{-f(z)} \end{pmatrix}.$$

- The matrix $\hat{\sigma}_{ij}$ is defined to be the 4×4 matrix which permutes the i^{th} and j^{th} row/column. For example, the matrix $\hat{\sigma}_{24}$ would be

$$\hat{\sigma}_{24} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

  The matrix $\hat{\sigma}_{ij}$ permutes the i^{th} and j^{th} row and column of a given matrix A by conjugation; again using our example of $\hat{\sigma}_{24}$,

$$\hat{\sigma}_{24} A \hat{\sigma}_{24} = \hat{\sigma}_{24} \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix} \hat{\sigma}_{24} = \begin{pmatrix} a_{11} & a_{14} & a_{13} & a_{12} \\ a_{41} & a_{44} & a_{43} & a_{42} \\ a_{31} & a_{34} & a_{33} & a_{32} \\ a_{21} & a_{24} & a_{23} & a_{22} \end{pmatrix}.$$

- For readability purposes, blocks of zeros in matrices will be denoted simply by zero, where there is no cause for ambiguity. For example, the if A is a 3×3 matrix, then the expressions

$$\begin{pmatrix} 1 & 0 \; 0 \; 0 \\ 0 & \\ 0 & A \\ 0 & \end{pmatrix} = \begin{pmatrix} 1 & 0_{3\times 1} \\ 0_{1\times 3} & A \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & A \end{pmatrix}$$

  all have identical meaning.

- If A is an $n \times n$ matrix and B is an $m \times m$ matrix, we define the matrix $A \oplus B$ to be the block diagonal matrix

$$A \oplus B := \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}.$$

- If $X(z)$ is the solution to a Riemann Hilbert problem defined on the contour γ, we write $J_X(z) : \gamma \to \mathbb{C}$ as its jump matrix: i.e.,

$$X_+(z) = X_-(z) J_X(z), \qquad z \in \gamma.$$

  We will also sometimes write γ_X or Γ_X to denote the contour γ corresponding to the Riemann-Hilbert problem for $X(z)$.

5.2 Definition and Analysis of the Spectral Curve.

In this section, we define the spectral curve and g-functions, which we will later use in the second transformation. We also prove a number of inequalities necessary for the later "lensing" transformations.

We begin by constructing the spectral curve, and give a basic analysis of the spectral curve for all values of the parameters (τ, t) in the region D bounded by τ/t axes, the *infinite temperature line* $\tau = 1$, and the critical curve

$$t(\tau) = \begin{cases} -\frac{1}{12} + \frac{2}{9}\tau^2, & 0 < \tau < \frac{1}{4}, \\ -\frac{2}{9}\sqrt{\tau}(\sqrt{\tau} - 1)^2(\sqrt{\tau} + 2), & \frac{1}{4} < \tau < 1. \end{cases} \tag{5.28}$$

We call the components of this curve the *low-temperature critical curve*, for $0 < \tau < \frac{1}{4}$, and the *high temperature critical curve*, for $\frac{1}{4} < \tau < 1$. This curve is depicted in Figure (5.1). These are precisely the critical curves that appeared in Kazakov's original work [70]. As it will turn out, we will be able to explicitly parameterize the spectral curve for every value of the parameters on these phase transition lines, as well as the subcritical region depicted in the figure.

We define four different regions of the parameter space as follows:

1. The generic (non-critical) case, i.e $(\tau, t) \in \text{int}(D)$,
2. The low-temperature critical curve, when $t = -\frac{1}{12} + \frac{2}{9}\tau^2$, $0 < \tau < \frac{1}{4}$,
3. The high-temperature critical curve, when $t = -\frac{2}{9}\sqrt{\tau}(\sqrt{\tau} - 1)^2(\sqrt{\tau} + 2)$, $\frac{1}{4} < \tau < 1$,
4. The multi-critical case, when $\tau = \frac{1}{4}, t = -\frac{5}{72}$.

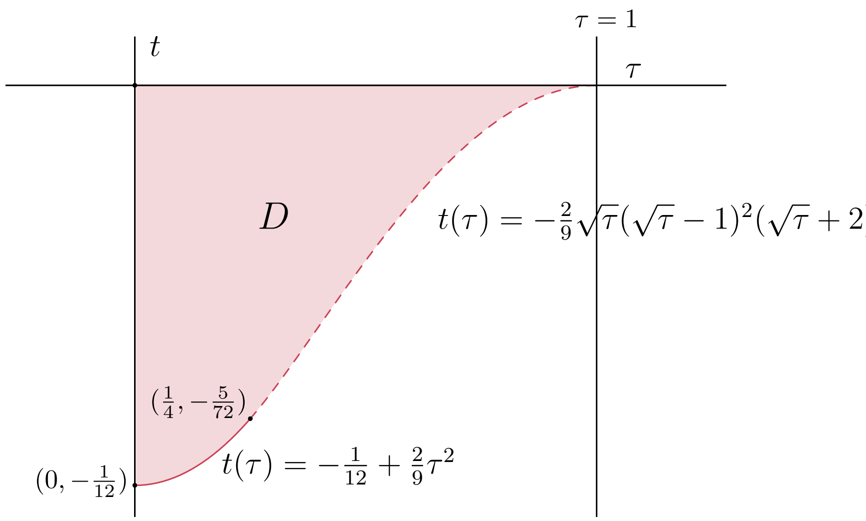
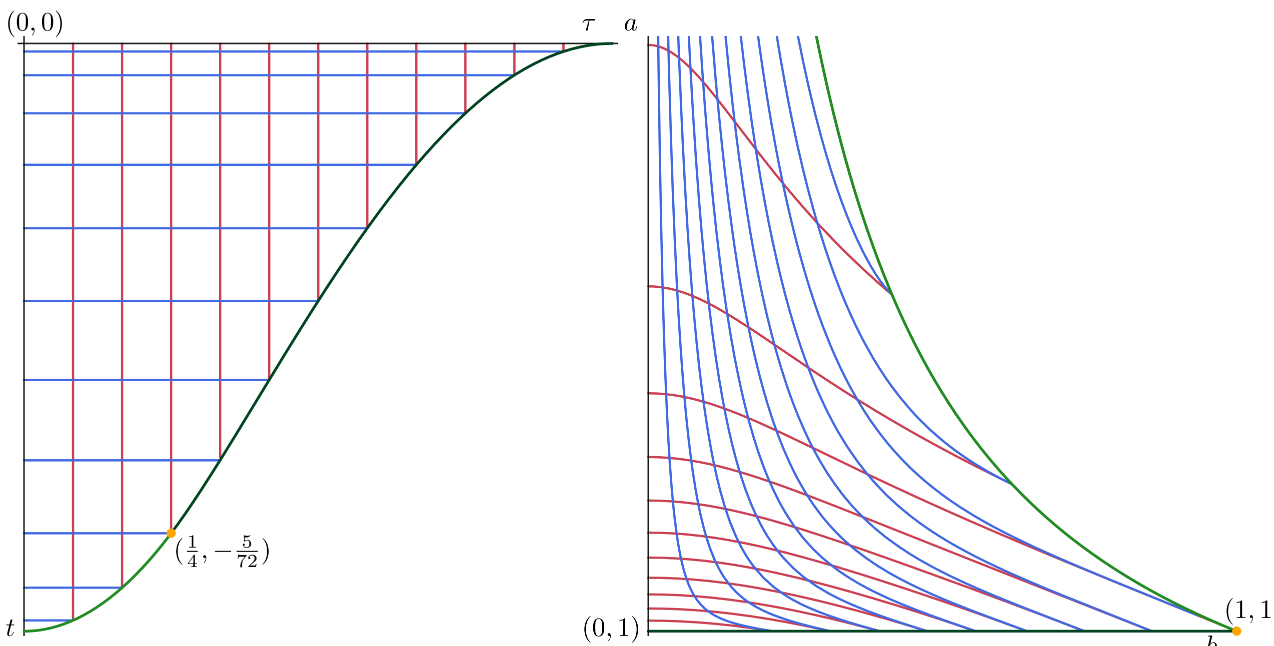


Figure 5.1. The phase portrait for the 2-matrix model with quartic interactions in the (τ, t) plane. In this paper, we study the shaded region D, along with the critical curve and multicritical point. The low-temperature critical curve is represented by the solid red line; the high-temperature critical curve is represented by the dashed red line. Note that the critical curve intersects the $\tau = 0$ axis precisely at the critical point of the 1-matrix model; this is in agreement with the intuition that the 2-matrix model decouples at 0 temperature.



(a) The (τ, t) phase plane. **(b)** The (a, b) phase plane.

Figure 5.2. The region D in the (τ, t) plane, and the region R in the (a, b) plane. Note that the axis $\tau = 0$ is mapped to infinity.

5.2.1 Definition of the Spectral Curve.

The analysis of the spectral curve and g-function is based on the ansatz that the curve is of genus zero, and thus rationally parameterized. Let us momentarily discuss the situation formally; once we have formulated a workable object, we shall prove that it is the correct one. Consider the partition function

$$\begin{aligned} Z &= \iint \Delta(x)\Delta(y) \exp\left\{-n\sum_{i=1}^{n}(\frac{1}{2}x_i^2 + \frac{1}{2}y_i^2 + \frac{t}{4}x_i^4 + \frac{t}{4}y_i^4 - \tau x_i y_i)\right\}\prod_{i=1}^{n} dx_i dy_i \\ &= \iint \exp\left\{\sum_{i<j}\log\frac{1}{|x_i - x_j|} + \sum_{i<j}\log\frac{1}{|y_i - y_j|} - n\sum_{i=1}^{n}(\frac{1}{2}x_i^2 + \frac{1}{2}y_i^2 + \frac{t}{4}x_i^4 + \frac{t}{4}y_i^4 - \tau x_i y_i)\right\}\prod_{i=1}^{n} dx_i dy_i. \end{aligned}$$

In the large n limit, under the appropriate scaling, we expect the eigenvalues x_i, y_i should accumulate to a continuous density; this density should be subject to the stationarity conditions

$$\begin{aligned} X + tX^3 + \int \frac{d\mu(\zeta)}{\zeta - X} - \tau Y = 0, \\ Y + tY^3 + \int \frac{d\mu(\zeta)}{\zeta - Y} - \tau X = 0. \end{aligned}$$

(Note that we have used the same density μ here to represent the density of the X and Y eigenvalues; this is admissible since the potentials are the same). These stationarity conditions are the analog of the ones for the 1-matrix model. Expanding at infinity, we find that

$$X + tX^3 - \frac{1}{X} - \tau Y = \mathcal{O}(X^{-2}), \tag{5.29}$$

$$Y + tY^3 - \frac{1}{Y} - \tau X = \mathcal{O}(Y^{-2}). \tag{5.30}$$

We expect that there should be a polynomial P in two variables such that $P(X, Y) = 0$; this Riemann surface is the spectral curve. The effective potential, which is constant on the support of the eigenvalues, and gives the dominant contribution to the partition function (i.e., it is the g-function we are seeking) can be recovered either as $\tau\Omega(X) = \tau\int Y dX$.

We now make the ansatz that the spectral curve is genus 0, i.e. that it is rationally parameterized. We have the following proposition:

Proposition 5.1. *A 2-parameter family of solutions to the stationarity equations* (5.29),(5.30) *is given by the rational functions*

$$X(u) = A\int^u \frac{(u^2-a^2)(u^2-b^2)}{u^4}du = A\left(u + \frac{a^2+b^2}{u} - \frac{a^2b^2}{3u^3}\right), \tag{5.31}$$

$$Y(u) = X(u^{-1}) = A\left(\frac{1}{u} + (a^2+b^2)u - \frac{a^2b^2}{3}u^3\right), \tag{5.32}$$

with A, t, τ *defined parametrically in terms of* a *and* b *as:*

$$t = t(a,b) = -\frac{a^2b^2(a^4b^4+3a^4+6a^2b^2+3b^4-3)}{9(a^2+b^2)^2(a^2b^2+1)^2}, \qquad \tau = \tau(a,b) = \frac{1}{(a^2+b^2)(a^2b^2+1)}, \tag{5.33}$$

$$A = A(a,b) = ab\sqrt{-\frac{\tau}{3t}} = \sqrt{\frac{3(a^2+b^2)(a^2b^2+1)}{a^4b^4+3a^4+6a^2b^2+3b^4-3}}. \tag{5.34}$$

Proof. By symmetry of the equations (5.29), (5.30) upon interchange of X and Y, we have (without loss of generality) that $Y(u) = X(u^{-1})$; this fixes the positions of the infinities of the X and Y-coordinates in the uniformizing plane as $u = \infty$, $u = 0$, respectively. This also makes the second stationarity equation (5.30) redundant; if we can find a solution to (5.29) with the constraint $Y(u) = X(u^{-1})$ imposed, we have automatically also found a solution to (5.30). We can generically set

$$X(u) = A\left[u + \alpha_0 + \alpha_1 u^{-1} + \alpha_2 u^{-2} + \alpha_3 u^{-3}\right];$$

this is the most general form of a rational function with poles only at $u = 0, \infty$ which can possibly be a solution to the stationarity equations; this is because the leading terms on the right hand side have equal order: $tX(u)^3 = \mathcal{O}(u^3)$, and $Y(u) = X(u^{-1}) = \mathcal{O}(u^3)$, with all other terms of order $\mathcal{O}(u^2)$ or lower. The free parameters $A, \{\alpha_k\}$ can then be chosen appropriately so that the rational function

$$\mathcal{R}(u) := X(u) + tX(u)^3 - \frac{1}{X(u)} - \tau Y(u) = \mathcal{O}(u^{-2}), \qquad u \to \infty.$$

The quadratic term and constant term of the expansion of $\mathcal{R}(u)$ at infinity require that the coefficients $\alpha_2, \alpha_0 = 0$. This now fixes the form of $X(u)$ as

$$X(u) = A\left[u + \alpha_1 u^{-1} + \alpha_3 u^{-3}\right].$$

The function $X(u)$ has four (finite) critical points $X'(u_k^*) = 0$, $k = 1,2,3,4$. The above form of $X(u)$ fixes the critical points to be symmetric: $X'(\pm u_1^*) = 0$, $X'(\pm u_2^*) = 0$. We take these branch points as a new set of parameters, and write $X(u)$ generically as

$$X(u) = A \int^u \frac{(u^2-a^2)(u^2-b^2)}{u^4} du = A \left(u + \frac{a^2+b^2}{u} - \frac{a^2b^2}{3u^3} \right),$$
$$Y(u) = X(u^{-1}) = A \left(\frac{1}{u} + (a^2+b^2)u - \frac{a^2b^2}{3} u^3 \right).$$

Inserting this expression for $X(u)$ into the stationarity equation (5.29), we obtain that

$$\left(tA^3 + \frac{1}{3}\tau A a^2 b^2 \right) u^3 + \left(A + 3tA^3(a^2+b^2) - \tau A(a^2+b^2) \right) u$$
$$+ \left(A(a^2+b^2) + tA^3(-a^2b^2 + (a^2+b^2)^2 + (a^2+b^2)(2a^2+2b^2)) - \frac{1}{A} - \tau A \right) u^{-1} = \mathcal{O}(u^{-2});$$

the requirement that the stationarity equation be satisfied up to order $\mathcal{O}(X^{-2})$ is equivalent that the first three coefficients in the above expansion vanish identically. This gives us a system of 3 equations, which we can solve for t, τ, and A; the unique solution (up to a determination of the sign of the square root) is

$$t = t(a,b) = -\frac{a^2b^2(a^4b^4 + 3a^4 + 6a^2b^2 + 3b^4 - 3)}{9(a^2+b^2)^2(a^2b^2+1)^2}, \qquad \tau = \tau(a,b) = \frac{1}{(a^2+b^2)(a^2b^2+1)},$$
$$A = A(a,b) = ab\sqrt{-\frac{\tau}{3t}} = \sqrt{\frac{3(a^2+b^2)(a^2b^2+1)}{a^4b^4+3a^4+6a^2b^2+3b^4-3}}.$$

This completes the proof. □

We have now obtained a 2-parameter family of solutions to the stationarity equations up to terms of order $\mathcal{O}(X^{-2})$. Each fixed pair (a,b) parameterizes a Riemann surface. We claim that we can parameterize the region D of the phase plane in terms of this family of Riemann surfaces.

Proposition 5.2. *Let*

$$R := \{(a,b) \mid 0 < b \leq 1, 1 \leq a \leq b^{-1}\}. \tag{5.35}$$

Then, there is a bijection between the region R and the region D of the phase plane, induced by the mapping $(a,b) \mapsto (\tau(a,b), t(a,b))$:

$$D = \{(\tau(a,b), t(a,b)) \mid (a,b) \in R\}. \tag{5.36}$$

Furthermore, we have the following identifications:

1. *The low-temperature critical curve given by the boundary component $a = b^{-1}$,*

2. *The high-temperature critical curve given by the boundary component $a = 1$,*

3. *The multicritical point given by $a = b = 1$.*

Proof. The proofs that $(\tau(b^{-1},b),t(b^{-1},b))$ for $0<b\leq 1$ parameterizes the low-temperature critical curve $t=-\frac{1}{12}+\frac{2}{9}\tau^2$, for $0<\tau<\frac{1}{4}$, and that $(\tau(1,b),t(1,b))$ for $0<b<1$ parameterizes the high-temperature critical curve $t=-\frac{2}{9}\sqrt{\tau}(\sqrt{\tau}-1)^2(\sqrt{\tau}+2)$, for $\frac{1}{4}<\tau<1$, are straightforward and left to the reader. □

We point out that when $(a,b)=(1,1)$, we obtain the multi-critical point $(\tau,t)=(\frac{1}{4},-\frac{5}{72})$. We remark that $t(a,b),\tau(a,b),A(a,b)$ defined in equations (5.33), (5.34) all have constant sign for the values of $(a,b)\in R$:

$$t(a,b)<0,\ \tau(a,b)>0,\ \textit{and}\ A(a,b)>0.$$

A comparison of the phase planes are depicted in Figure (5.2). One cannot in principle approach the $\tau=0$ axis, since the objects we are computing are essentially one collection of eigenvalues averaged over another. The correct limit to take is to set

$$a=\frac{\lambda}{b},\qquad b\to 0. \tag{5.37}$$

In this limit, one can approach the $\tau=0$ axis in the space of parameters. However, all of the relevant objects $(X,Y,\Omega,$ etc.) are singular in this limit.

Remark 5.3. We have actually found more than just the spectral curve for the region D; if we take $a\in i\mathbb{R}$ to be purely imaginary and b to be real, we obtain the correct asymptotics for the full range $0<t<\infty$, $0<\tau<1$. However, the integration contours for the Riemann-Hilbert problem and the weights are different in this case, as is the basic structure of the Riemann surface. We therefore postpone analysis of this case to a later work.

The associated Riemann surface given by the parameterization $(X(u),Y(u))$ is called the *spectral curve*. Let us study the structure of the spectral curve; we shall treat the parametric curve (X,Y) as a branched covering of the sphere over the X-coordinate; by construction, the X-coordinate has branch points ($X'(u)=0$) at $u=\pm a$, $\pm b$, and ∞.

Away from the multicritical point, the spectral curve is 4-sheeted; this family of spectral curves have generically the same structure, and are shown in Figure (5.4). There are 4 branch points, all of which lie on the real axis: at $\pm\alpha:=X(\pm a)$, and $\pm\beta:=X(\pm b)$. We have the inequalities $0<\alpha<\beta<\infty$. The structure of the curve is as follows: Sheets 1 and 2 are glued along $[-\alpha,\alpha]$, sheet 2 is glued to sheet 3 along the interval $(-\infty,-\beta]$, and finally sheets 2 and 4 are glued along $[\beta,\infty)$. At the multicritical point $a=b=1$, the curve further degenerates, and the branch points $\pm\beta\to\pm\alpha$. The multicritical spectral curve is shown in Figure (5.5). In this case, sheets 1 and 2 are glued along the interval $[-\alpha,\alpha]$, sheet 2 is glued to sheet 3 along the interval $(-\infty,-\alpha]$, and finally sheet 2 is glued to sheet 4 along the interval $[\alpha,\infty)$.

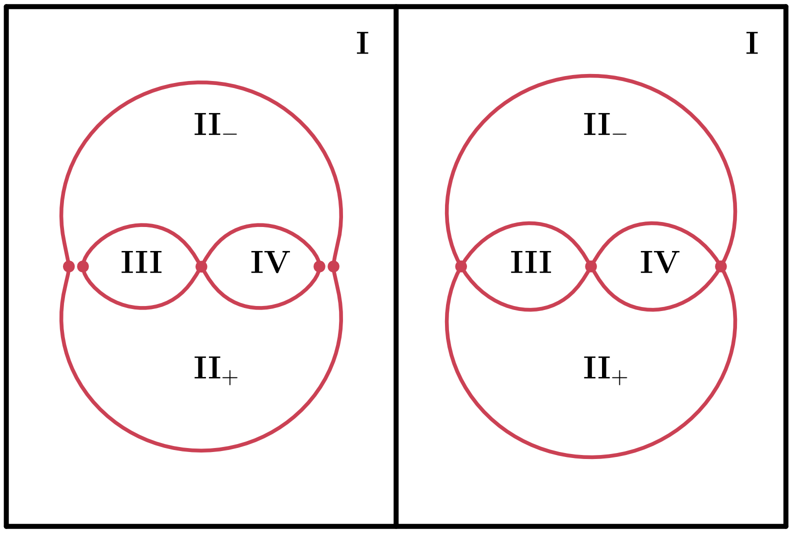


Figure 5.3. The branch cuts in the uniformizing plane $u = x + iy$ away from the multicritical point (left) and at the multicritical point. The noncritical case has $a = 1.0184, b = 0.9100$. The images of each sheet under the uniformizing map are labelled I,...,IV. Note that, at criticality, the second sheet is split into two connected components; this is consistent with the picture in the physical plane, depicted in Figures 5.4, 5.5.

In the uniformization plane, the spectral curve is shown at and away from the multicritical point $(1, 1)$ in Figure (5.3).

The candidate spectral curves must satisfy a number of additional properties. The curve must carry a "g-function", whose restrictions to each sheet must match the asymptotics of $\hat{\Theta}(z)$ at ∞. The g-function must also satisfy a certain collection of inequalities which will be necessary for the later lens-opening transformations.

We introduce the candidate g-function as the "effective potential" which we formally discussed earlier:

$$g(X) := \tau\Omega(X) = \tau \int Y dX. \tag{5.38}$$

We will not use the notation $g(X)$ for the g-function, as we prefer to stay with the notation $\tau\Omega(X)$. The function $\tau\Omega(X)$ is in general a multivalued function in $\mathbb{C}$. The function $\Omega(u) := \int Y(u)X'(u)du$ in the uniformization coordinate is explicit:

$$\begin{aligned}\tau\Omega(u) &= \tau \int Y dX = \tau \int Y(u)X'(u)du \\ &= \frac{1}{a^4b^4 + 3a^4 + 6a^2b^2 + 3b^4 - 3}\bigg(- \frac{1}{4}a^2b^2u^4 + \frac{1}{2}(a^2+b^2)(a^2b^2+3)u^2 - \log u \\ &- \frac{3}{2}(a^2+b^2)(a^2b^2-1)u^{-2} - \frac{3}{4}a^2b^2u^{-4}\bigg).\end{aligned} \tag{5.39}$$

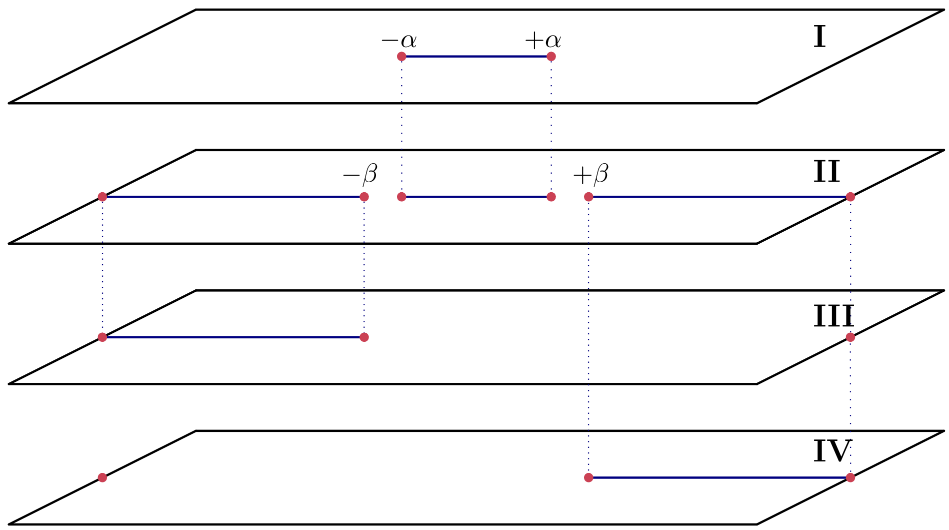


Figure 5.4. A representative example of a critical/generic surface in the physical plane; the sheets are labelled I,...,IV. If the curve is low-temperature critical, an extra zero of $\Omega(z)$ accumulates at $z = \pm\alpha$ and $z = \pm\beta$; if the curve is high-temperature critical, an extra zero accumulates at $z = \pm\alpha$.

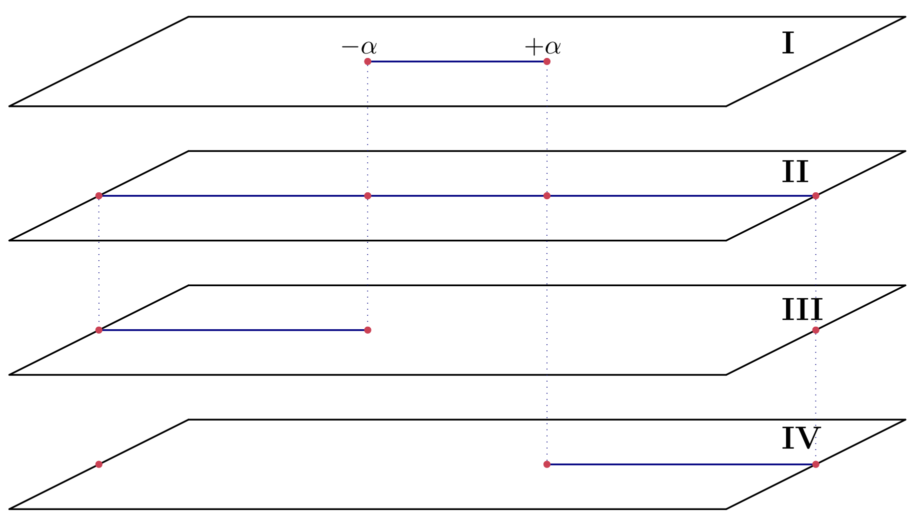


Figure 5.5. The (multi) critical spectral curve. Note that the branch points at $z = \alpha$ and $z = \beta$ have merged here.

From here on, we shall relabel $X(u)$ as

$$X(u) =: z(u), \tag{5.40}$$

as the notation is more convenient and consistent later on. We also define the uniformization mappings $u_j(z), j = 1, 2, 3, 4$, as the inverse functions to $z(u)$, such that $z(u_j(z))$ is the identity mapping in sheet j, away from the branch cuts.

We can expand Ω on each sheet of the spectral curve, by inverting the uniformization coordinate ($u_j(z)$, $j = 1, 2, 3, 4$), and inserting these expansions into the expression above for Ω. Since $z = \infty$ (corresponding to $u = 0$ on the lower three sheets, $u = \infty$ on the first sheet) is a common branch point for all curves in the family parameterized by R, we can write an expression for the expansion of $\Omega_j(z) := \Omega(u_j(z))$ at infinity that holds for all $(a, b) \in R$. We have the following proposition:

Proposition 5.4. *(Expansion of $\Omega_j(z)$ at $z = \infty$). For any fixed $(a, b) \in R$, with $t(a, b)$, $\tau(a, b)$ defined as in (5.33), we have the following expansions of the functions $\Omega_j(z)$:*

$$\tau\Omega_1(z) = \frac{t}{4}z^4 + \frac{1}{2}z^2 - \log z + \ell_0 + \frac{C_0}{z^2} + \mathcal{O}(z^{-4}), \tag{5.41}$$

$$\tau\Omega_2(z) = \begin{cases} -\frac{3\omega^2}{4}\frac{\tau^{4/3}}{(-t)^{1/3}}z^{4/3} - \frac{\omega}{2}\frac{\tau^{2/3}}{(-t)^{2/3}}z^{2/3} + \frac{1}{3}\log z + \ell_1 + \frac{\omega^2 C_1}{z^{2/3}} + \mathcal{O}\left(\frac{1}{z^{4/3}}\right), & Im\ z > 0, \\ -\frac{3\omega}{4}\frac{\tau^{4/3}}{(-t)^{1/3}}z^{4/3} - \frac{\omega^2}{2}\frac{\tau^{2/3}}{(-t)^{2/3}}z^{2/3} + \frac{1}{3}\log z + \ell_1 + \frac{\omega C_1}{z^{2/3}} + \mathcal{O}\left(\frac{1}{z^{4/3}}\right), & Im\ z < 0, \end{cases} \tag{5.42}$$

$$\tau\Omega_3(z) = -\frac{3}{4}\frac{\tau^{4/3}}{(-t)^{1/3}}z^{4/3} - \frac{1}{2}\frac{\tau^{2/3}}{(-t)^{2/3}}z^{2/3} + \frac{1}{3}\log z + \ell_1 + \frac{C_1}{z^{2/3}} + \mathcal{O}\left(\frac{1}{z^{4/3}}\right), \tag{5.43}$$

$$\tau\Omega_4(z) = \begin{cases} -\frac{3\omega}{4}\frac{\tau^{4/3}}{(-t)^{1/3}}z^{4/3} - \frac{\omega^2}{2}\frac{\tau^{2/3}}{(-t)^{2/3}}z^{2/3} + \frac{1}{3}\log z + \ell_1 + \frac{\omega C_1}{z^{2/3}} + \mathcal{O}\left(\frac{1}{z^{4/3}}\right), & Im\ z > 0, \\ -\frac{3\omega^2}{4}\frac{\tau^{4/3}}{(-t)^{1/3}}z^{4/3} - \frac{\omega}{2}\frac{\tau^{2/3}}{(-t)^{2/3}}z^{2/3} + \frac{1}{3}\log z + \ell_1 + \frac{\omega^2 C_1}{z^{2/3}} + \mathcal{O}\left(\frac{1}{z^{4/3}}\right), & Im\ z < 0. \end{cases} \tag{5.44}$$

Here, the constants $\ell_0 := \ell_0(a, b), \ell_1 := \ell_1(a, b)$ are defined as

$$\ell_0(a, b) = -\frac{9a^6b^2 + 20a^4b^4 + 9a^2b^6 + 18a^4 + 36a^2b^2 + 18b^4}{6(a^4b^4 + 3a^4 + 6a^2b^2 + 3b^4 - 3)} + \log A(a, b), \tag{5.45}$$

$$\ell_1(a, b) = -\frac{3(2a^6b^2 + 4a^4b^4 + 2a^2b^6 + a^4 + 4a^2b^2 + b^4)}{2a^2b^2(a^4b^4 + 3a^4 + 6a^2b^2 + 3b^4 - 3)} - \frac{1}{3}\log A(a, b) + \frac{1}{3}\log 3 - \frac{1}{3}\log a^2b^2. \tag{5.46}$$

and the constants $C_0 := C_0(a, b)$, $C_1 := C_1(a, b)$ are defined as

$$C_0(a, b) = -\frac{3}{2}\frac{(a^6b^2 + a^2b^6 - 3a^4 - 5a^2b^2 - 3b^4 + 3)(a^2 + b^2)^2(a^2b^2 + 1)}{(a^4b^4 + 3a^4 + 6a^2b^2 + 3b^4 - 3)^2}, \tag{5.47}$$

$$C_1(a, b) = \frac{(3a^6b^6 + 3a^6b^2 + 3a^4b^4 + 3a^2b^6 - a^4 - 8a^2b^2 - b^4)(a^2 + b^2)}{2b^4a^4(a^4b^4 + 3a^4 + 6a^2b^2 + 3b^4 - 3)}\left(\frac{abA^2(a, b)}{9}\right)^{1/3}. \tag{5.48}$$

Proof. The proof of this proposition is a straightforward calculation; as we have alluded to, one must first expand the uniformization coordinate at $z = \infty$ on each of the sheets, then insert this expansion into the expression for $\Omega(u)$ in the uniformization coordinate (see Equation (5.39)). The expansions of the uniformization coordinate on each sheet near the branch points are given in Appendix B. □

Thus, comparing the above expansions to the exponential part of the asymptotics of $\mathbf{X}(z)$ in Equation (5.121), we see that the $\Omega_j(z)$'s match exactly the terms appearing there, to order $z^{-2/3}$. Thus, the $\Omega_j(z)$'s have the correct asymptotics, and the g-function transformation will push through. However, we will need to ensure we can also open lenses; for this, we will need to prove that certain inequalities on the $\Omega_j(z)$'s hold. Here, we will need to break calculations up in to cases, as the structure of the contours and the $\Omega_j(z)$'s near them depends more intricately on the choices of (a, b). We will have 4 cases, which we have mentioned before, but list again here for the convenience of the reader (this time in (a, b) coordinates):

1. The generic case: $0 < b < 1$, $1 < a < b^{-1}$,
2. The low-temperature critical case: $0 < b < 1$, $a = b^{-1}$,
3. The high-temperature critical case: $0 < b < 1$, $a = 1$,
4. The (multi)-critical case: $a = b = 1$.

The next section is devoted to the proof of these inequalities.

5.2.2 Definition of the Contours Γ, Γ_1, and Γ_2.

Before we can proceed to the proof of the lensing inequalities, we first need to define a number of contour which our Riemann-Hilbert problem will rely on. These contours will be chosen so that the functions $\Omega_j(z)$ satisfy certain inequalities on them. The first such contour is Γ, on which the matrix-valued function $\mathbf{Y}(z)$ has jumps. We redefine Γ to be the contour starting at $e^{\frac{3\pi i}{4}} \cdot \infty$, passes through $z = -\alpha$, then continues along the real axis until it reaches $z = +\alpha$, then goes off again to infinity in the direction $e^{-\frac{\pi i}{4}}$. The modified contour Γ is depicted in Figure (5.6).

We also define two new contours, Γ_1 and Γ_2, which will appear in the first transformation. The contour Γ_2 is defined to start at $z = -\infty$, then travel along the real axis until it reaches $z = -\beta$; the contour then goes off to infinity in the sector $-\frac{\pi}{4} < \arg(z + \beta) < 0$. The exact direction of approach to infinity will be established in the next section, when we will require certain inequalities to hold on Γ_2.

Similarly, we define Γ_1 to be the contour starting at infinity in the sector $0 < \arg(z - \beta) < \frac{3\pi}{4}$, and approaches the point $z = \beta$; then, Γ_1 goes off to infinity again along the positive real axis. Again, the exact

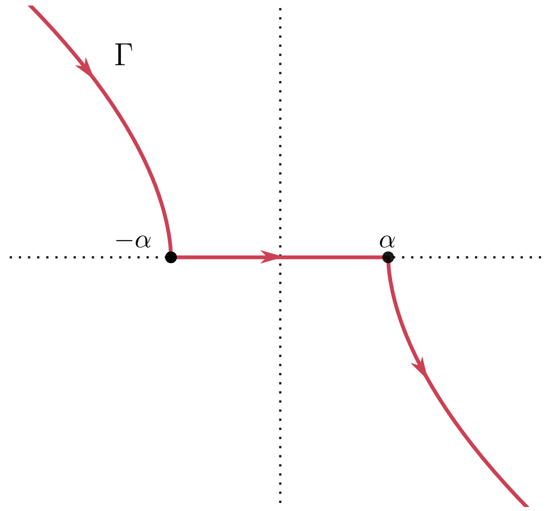


Figure 5.6. The contour Γ, which starts at $e^{3\pi i/4} \cdot \infty$, passes through $-\alpha$, $+\alpha$ on the real axis, and then goes off to $e^{-\pi i/4} \cdot \infty$.

specifications of Γ_1 will be established in the next section to guarantee that certain inequalities hold. The family of contours Γ, Γ_1, and Γ_2 are depicted in Figure (5.7). We label the regions above Γ_1, below Γ_2, and bounded between Γ_1 and Γ_2 by Ω_u, Ω_ℓ, an Ω_c, respectively.

5.2.3 Lensing Inequalities.

We now begin the work of proving that the function Ω we constructed indeed satisfies the inequalities necessary for lensing. The main idea of this section is contained in this subsection; it explains how we can extend local inequalities near the branch points to inequalities that hold on the full branch cuts. Most of the work that remains in the subsequent subsections involves expanding Ω near the branch points, and checking that the correct inequalities hold there. This check differs in the generic, critical, and multicritical cases, as the expansions of Ω near the branch points are different in each case. However, we stress that the proof is essentially the same, and relies only on the following lemma.

Lemma 5.5. *Consider the normal vector to the preimages of the branch cuts in the uniformizing plane under the mapping $z(u)$:*

$$\hat{\mathbf{n}} := \frac{\nabla Im\ z(u)}{||\nabla Im\ z(u)||}. \tag{5.49}$$

For any $(a, b) \in R = \{(a, b) \mid 0 < b \leq 1, 1 \leq a \leq b^{-1}\}$, the function

$$\nabla Re\ \Omega(u) \cdot \hat{\mathbf{n}} = \frac{\partial}{\partial n} Re\ \Omega(u) \tag{5.50}$$

is of constant sign on each connected component of the preimages of the branch cuts.

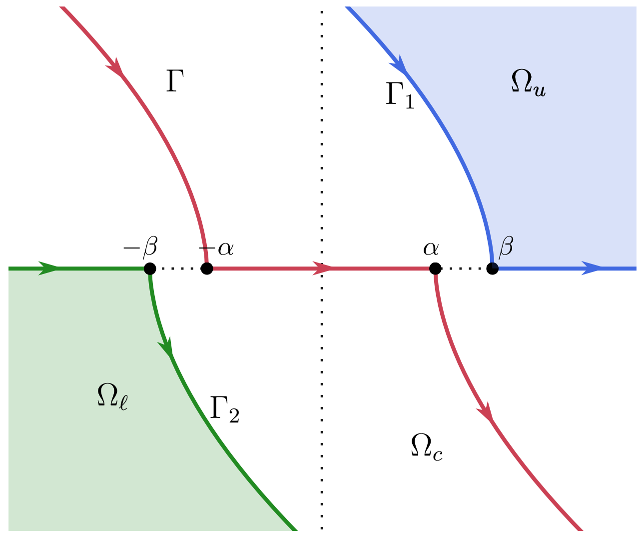


Figure 5.7. The contours Γ, Γ_1, and Γ_2. Also depicted are the regions Ω_ℓ and Ω_u (shaded), as well as Ω_c (unshaded).

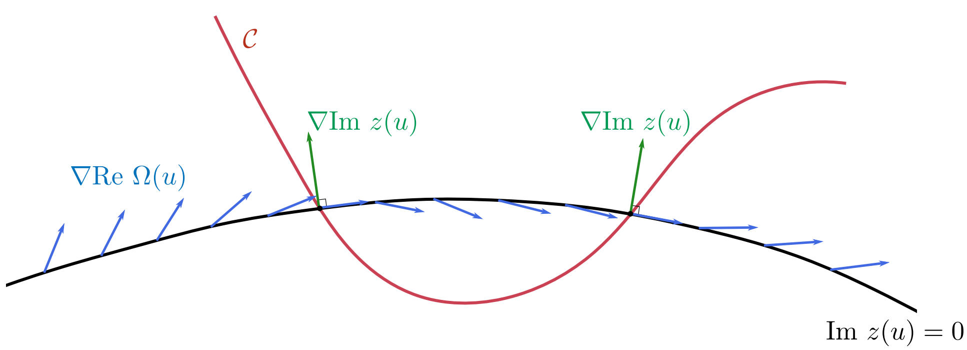


Figure 5.8. The curve $\text{Im } z = 0$, and the curve $\mathcal{C}$, which characterizes the places that $\nabla\text{Re } \Omega(u)$ and $\nabla\text{Im } z(u)$ are perpendicular. At the places where $\mathcal{C}$ and $\text{Im } z = 0$ intersect, the normal vectors to these curves are perpendicular, implying that the direction of steepest descent may change sign, as shown above.

Proof. Let $\mathcal{C}$ denote the curve where $\frac{\partial}{\partial n}\text{Re } \Omega(u) = 0$. In order to prove that the direction of steepest descent is constant on each component of the branch cuts, it is sufficient to check that the curves Im $z(u) = 0$ and $\mathcal{C}$ intersect at most at the points such that $z'(u) = 0$: namely, $\pm a, \pm b$.

First, since $\Omega(u)$ and $z(u)$ are analytic functions, the Cauchy-Riemann equations yield that $\partial[\text{Re } \Omega(u)] = \frac{1}{2}\partial\Omega(u)$, and similarly $\partial[\text{Im } z(u)] = \frac{i}{2}\partial z(u)$ (here ∂ denotes the holomorphic derivative in u). It follows that the $\nabla\text{Re } \Omega(u)$ is perpendicular to $\nabla\text{Im } z(u)$ if and only if the quotient of these two expressions is purely imaginary:

$$\frac{\Omega'(u)}{iz'(u)} \in i\mathbb{R}.$$

It follows that the curve $\mathcal{C}$ is characterized by the condition

$$\text{Im } \frac{\Omega'(u)}{z'(u)} = 0, \tag{5.51}$$

which is equivalent to the condition Im $\frac{\partial\Omega}{\partial z}(u) = \text{Im } Y(u) = 0$ (cf. Equation (5.39)). We thus must prove that Im $Y(u) = 0$ and Im $z(u) = 0$ intersect at most at the branch points. We can parameterize these curves in the uniformizing plane in polar form with relatively simple expressions. Defining

$$r_\pm(\theta) = \sqrt{\frac{1}{2}(a^2+b^2) \pm \sqrt{\frac{4}{3}a^2b^2\sin^2\theta + \frac{1}{4}(a^2-b^2)^2}}, \tag{5.52}$$

we have that

$$\text{Im } z(re^{i\theta}) = 0 \Longleftrightarrow r = r_\pm(\theta) \qquad \text{Im } Y(re^{i\theta}) = 0 \Longleftrightarrow r = \frac{1}{r_\pm(\theta)}. \tag{5.53}$$

This last identity follows from the fact that Im$X(u) = \overline{Y(1/\bar{u})}$. We must check that these curves do not intersect; the equations that determine intersection are

$$r_+(\theta) = \frac{1}{r_+(\theta)}, \qquad r_-(\theta) = \frac{1}{r_-(\theta)}, \qquad r_+(\theta) = \frac{1}{r_-(\theta)}. \tag{5.54}$$

Now, let us check that the first two equations have at most the branch points as solutions. We have that:

$$\begin{aligned} 1 = r_\pm(\theta)^2 \Longleftrightarrow & 1 - \frac{1}{2}(a^2+b^2) = \sqrt{\frac{4}{3}a^2b^2\sin^2\theta + \frac{1}{4}(a^2-b^2)^2} \\ \Longleftrightarrow & a^2b^2 - a^2 - b^2 + 1 = \frac{4}{3}a^2b^2\sin^2\theta \\ \Longleftrightarrow & \frac{3}{4}\left(1 - \frac{1}{b^2} - \frac{1}{a^2} + \frac{1}{a^2b^2}\right) = \sin^2\theta \end{aligned}$$

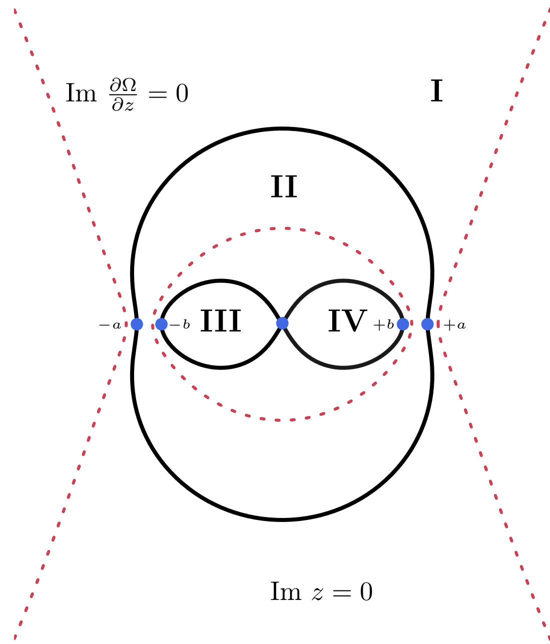


Figure 5.9. Images of the branch cuts Im $z(u) = 0$ and Im $\frac{\partial\Omega}{\partial z} = 0$ in the uniformizing plane, for $a = 1.059, b = 0.880$. The curves indeed do not intersect, and so the direction of steepest descent is constant along each connected component of the branch cuts.

Since the branch points occur at $\theta = 0, \pi$, we need that the left hand side of the above equation to satisfy one of the inequalities

$$\frac{3}{4}\left(1 - \frac{1}{b^2} - \frac{1}{a^2} + \frac{1}{a^2b^2}\right) \leq 0, \qquad or \qquad \frac{3}{4}\left(1 - \frac{1}{b^2} - \frac{1}{a^2} + \frac{1}{a^2b^2}\right) > 1. \tag{5.55}$$

Indeed, if $0 < b \leq 1$, then

$$\frac{1}{b^2} + \frac{1}{a^2} - \frac{1}{a^2b^2} \geq 1 + \frac{1}{a^2} - \frac{1}{a^2} = 1,$$

so that

$$\frac{3}{4}\left(1 - \frac{1}{b^2} - \frac{1}{a^2} + \frac{1}{a^2b^2}\right) \leq 0,$$

and the necessary inequality holds in the region R. On the other hand, for the other intersection equation, we have that

$$\begin{aligned}1 = r_+(\theta)r_-(\theta) &\iff 1 = \frac{1}{4}(a^2+b^2)^2 - \frac{4}{3}a^2b^2\sin^2\theta - \frac{1}{4}(a^2-b^2)^2 \\ &\iff a^2b^2 - 1 = \frac{4}{3}a^2b^2\sin^2\theta \\ &\iff \frac{3}{4}\left(1 - \frac{1}{a^2b^2}\right) = \sin^2\theta;\end{aligned}$$

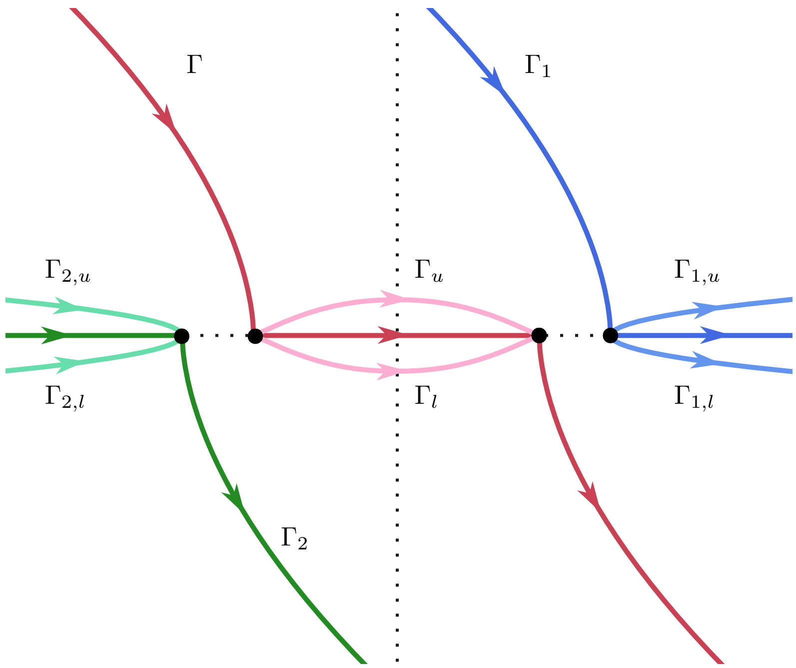


Figure 5.10. The opened lenses.

thus, we need that the left hand side of the above equation satisfies one of the following inequalities:

$$\frac{3}{4}\left(1-\frac{1}{a^2b^2}\right) > 1, \qquad or \qquad \frac{3}{4}\left(1-\frac{1}{a^2b^2}\right) \leq 0. \tag{5.56}$$

We see that, provided $1 \leq a \leq b^{-1}$, the second inequality holds.

Thus, we have proven that, provided $(a,b) \in R = \{(a,b) \mid 0 < b \leq 1, 1 \leq a \leq b^{-1}\}$, the only intersection points of the branch cuts and the places where the direction of steepest descent of $\Omega(u)$ changes sign are at (possibly) the branch points. The branch cuts $\text{Im } z = 0$ and the curves $\text{Im } \frac{\partial \Omega}{\partial z} = 0$ in the uniformizing plane are shown in Figure 5.9 for a particular choice of the parameters (a,b); one can see explicitly that these curves do not intersect. □

These differences are precisely the functions we will need in lensing; we now proceed to check that these differences indeed have the correct sign.

5.2.3.1 The Generic (Noncritical) Case: $0 < b < 1$, $1 < a < b^{-1}$. It remains to check that the $\Omega_j(z) := \Omega(u_j(z))$ have the correct sign around each of the lenses. This amounts to expanding $\Omega(u)$ around each of the branch points, and checking that locally, the signs of the quantities we will be interested are positive. The fact that these inequalities hold globally across each component of the support follows from the fact the intersection lemma we have just proven. We have the following proposition:

Proposition 5.6. *Local expansions of $\Omega_j(z)$ around the branch points .*

1. *Around $z = \pm\alpha := z(\pm a)$.*

$$\tau\Omega_1(z) = \eta_\alpha(z) + q_1(z-\alpha)^{3/2} + \mathcal{O}((z-\alpha)^2), \tag{5.57}$$

$$\tau\Omega_2(z) = \eta_\alpha(z) - q_1(z-\alpha)^{3/2} + \mathcal{O}((z-\alpha)^2), \tag{5.58}$$

and

$$\tau\Omega_1(z) = \begin{cases} \eta_{-\alpha}(z) + iq_1(z+\alpha)^{3/2} + \mathcal{O}((z+\alpha)^2), & Im\ z > 0, \\ \eta_{-\alpha}(z) - iq_1(z+\alpha)^{3/2} + \mathcal{O}((z+\alpha)^2), & Im\ z < 0, \end{cases} \tag{5.59}$$

$$\tau\Omega_2(z) = \begin{cases} \eta_{-\alpha}(z) - iq_1(z+\alpha)^{3/2} + \mathcal{O}((z+\alpha)^2), & Im\ z > 0, \\ \eta_{-\alpha}(z) + iq_1(z+\alpha)^{3/2} + \mathcal{O}((z+\alpha)^2), & Im\ z < 0, \end{cases}. \tag{5.60}$$

$\eta_{\pm\alpha}(z) = \Omega(\pm a) \mp \alpha \frac{a^2b^6 - 3a^2b^2 - 3b^4 - 3}{2a^2 + 6b^2}(z \mp \alpha)$. *Note that the constant* $q_1 := \frac{2(a^4-b^4)(a^4-1)(1-a^4b^4)}{a^{1/2}A^{3/2}(a^2-b^2)^{3/2}} > 0$.

2. *Around $z = \pm\beta := z(\pm b)$.*

$$\tau\Omega_2(z) = \begin{cases} \eta_\beta(z) + i\tilde{q}_1(z-\beta)^{3/2} + \mathcal{O}((z-\beta)^2), & Im\ z > 0, \\ \eta_\beta(z) - i\tilde{q}_1(z-\beta)^{3/2} + \mathcal{O}((z-\beta)^2), & Im\ z < 0, \end{cases} \tag{5.61}$$

$$\tau\Omega_4(z) = \begin{cases} \eta_\beta(z) - i\tilde{q}_1(z-\beta)^{3/2} + \mathcal{O}((z-\beta)^2), & Im\ z > 0, \\ \eta_\beta(z) + i\tilde{q}_1(z-\beta)^{3/2} + \mathcal{O}((z-\beta)^2), & Im\ z < 0, \end{cases} \tag{5.62}$$

and

$$\tau\Omega_2(z) = \eta_{-\beta}(z) + \tilde{q}_1(z+\beta)^{3/2} + \mathcal{O}((z+\beta)^2), \tag{5.63}$$

$$\tau\Omega_3(z) = \eta_{-\beta}(z) - \tilde{q}_1(z+\beta)^{3/2} + \mathcal{O}((z+\beta)^2). \tag{5.64}$$

Here, $\eta_{\pm\beta}(z) = \Omega(\pm\beta) \mp \beta\frac{a^2b^6-3a^2b^2-3b^4-3}{2a^2+6b^2}(z\mp\beta)$*. Note that the constant* $\tilde{q}_1 = \frac{2(a^4-b^4)(1-b^4)(1-a^4b^4)}{b^{1/2}A^{3/2}(a^2-b^2)^{3/2}} > 0$*.*

Proposition 5.7. *Lensing around the cuts. Let* $\Omega_j(z) = \phi_j(z) + i\psi_j(z)$*,* $j = 1, 2, 3, 4$*. Then, the following inequalities hold:*

1. $\phi_4(z) - \phi_2(z) > 0$ *in a lens around* $[\beta, \infty)$*,*

2. $\phi_3(z) - \phi_2(z) > 0$ *in a lens around* $(-\infty, -\beta]$*,*

3. $\phi_2(z) - \phi_1(z) > 0$ *in a lens around* $[-\alpha, \alpha]$*.*

Proof. We have only to expand the differences $\phi_j(z) - \phi_k(z)$ around the branch points; our previous lemmas guarantee that if the correct inequality holds locally, it holds globally as well. For z sufficiently close to β, the difference $\phi_4(z) - \phi_2(z) = \text{Re } (\Omega_4 - \Omega_2)(z) \sim \text{Re } \left[-2iq(z-\beta)^{3/2}\right] > 0$ in the sector $0 < \arg(z-\beta) < \frac{\pi}{3}$ for $|z-\beta|$ small enough. Now, for any $z \in (\beta, \infty)$, $u_{2,+}(z) = \overline{u_{4,-}(z)}$, where $u_{j,\pm}(z)$ denote the continuous limits of $u_j(\zeta)$ as $\zeta \to z$ from the upper and lower half planes, respectively. Thus, since $\overline{\Omega(u)} = \Omega(\bar{u})$, we have that

$$\overline{\Omega(u_{2,+}(z))} = \Omega(u_{4,-}(z)),$$

and so $\text{Re } (\Omega_4 - \Omega_2)(z) = 0$ for $z \in [\beta, \infty)$. Now, by definition of the sheets 2, 4, we have that

$$\frac{\partial \text{Re } \Omega_2}{\partial n_+} = -\frac{\partial \text{Re } \Omega_4}{\partial n_-}, \qquad \frac{\partial \text{Re } \Omega_2}{\partial n_-} = -\frac{\partial \text{Re } \Omega_4}{\partial n_+},$$

where $\frac{\partial}{\partial n_\pm}$ denote the normal derivatives in the upper/lower half planes in the z-coordinate, respectively (note that these normal derivatives and the one appearing in Lemma (5.5) differ only by an overall positive factor $|z'(u)| > 0$). On the other hand, by our observation that $\overline{\Omega(u)} = \Omega(\bar{u})$, we obtain the equalties

$$\frac{\partial \text{Re } \Omega_2}{\partial n_+} = \frac{\partial \text{Re } \Omega_2}{\partial n_-}, \qquad \frac{\partial \text{Re } \Omega_4}{\partial n_+} = \frac{\partial \text{Re } \Omega_4}{\partial n_-}.$$

We thus compute that

$$\frac{\partial}{\partial n_\pm}\left[\text{Re } (\Omega_4 - \Omega_2)(z)\right] = 2\frac{\partial}{\partial n_\pm}\text{Re } (\Omega_4)(z).$$

Since this quantity is positive locally near $z = \beta$, Lemma (5.5) allows us to conclude that $\frac{\partial}{\partial n_\pm}\left[\text{Re } (\Omega_4 - \Omega_2)(z)\right] > 0$ for all $z \in [\beta, \infty)$. Therefore, we can open a lens around $[\beta, \infty)$.

Similarly, near $z = -\beta$, again using the local expansions of Proposition (5.6), we have that $\phi_3(z) - \phi_2(z) = \text{Re } (\Omega_3 - \Omega_2)(z) > 0$ in the sector $\frac{2\pi}{3} < \arg(z+\beta) < \pi$ for $|z+\beta|$ sufficiently small. Thus, Lemma (5.5) guarantees that we can open a lens around $(-\infty, \beta]$.

Finally, near $z = +\alpha$ (respectively, $z = -\alpha$), the difference $\phi_2(z) - \phi_1(z) = \text{Re } (\Omega_2 - \Omega_1)(z) > 0$ for $\frac{2\pi}{3} < \arg z < \pi$ and $|z - \alpha|$ sufficiently small (respectively, $0 < \arg(z - \beta) < \frac{\pi}{3}$ and $|z + \alpha|$ sufficiently small). Thus, we can also open a lens around the central cut $[-\alpha, \alpha]$.

□

Proposition 5.8. *Inequalities off the real axis. Let* $\Omega_j(z) = \phi_j(z) + i\psi_j(z)$, $j = 1, 2, 3, 4$. *Then, the following inequalities hold:*

1. $\phi_2(z) - \phi_4(z) > 0$ *for* $z \in \Gamma_1 \cap \{Im\ z > 0\}$,
2. $\phi_2(z) - \phi_3(z) > 0$ *for* $z \in \Gamma_2 \cap \{Im\ z < 0\}$,
3. $\phi_1(z) - \phi_2(z) > 0$ *for* $z \in \Gamma \setminus \{Im\ z = 0\}$.

Proof. We prove $\phi_2(z) - \phi_4(z) > 0$ for $z \in \Gamma_1 \cap \{\text{Im } z > 0\}$; the proofs of the other inequalities follow from similar argumentation. Using (5.6), we have that $\phi_2(z) - \phi_4(z) > 0$ for $|z - \beta|$ sufficiently small in the sector $\frac{2\pi}{3} < |\arg(z - \beta)| < \pi$. Furthermore, at infinity, using Equations (5.42),(5.44), we see that $\phi_2(z) - \phi_4(z) > 0$ for $|z|$ sufficiently large in the sector $\frac{3\pi}{4} < |\arg z| < \pi$. Consider the domain

$$E := \{z \mid \phi_2(z) - \phi_4(z) > 0\};$$

by the lensing inequalities (5.7), the boundary of this domain is bounded away from the branch cuts, and $\phi_2(z) - \phi_4(z) = 0$ there. Since $\phi_2(z) - \phi_4(z)$ is not identically zero, the maximum principle tells us that the domain E is necessarily unbounded, and reaches infinity in the sector $\frac{3\pi}{4} < |\arg z| < \pi$. Thus, we may freely redefine Γ_1 so that $\phi_2(z) - \phi_4(z) > 0$ along Γ_1 for all $z \in \Gamma_1 \cap \{\text{Im } z > 0\}$. □

5.2.3.2 The Low-Temperature Critical Curve: $0 < b < 1, a = b^{-1}$. We repeat the calculations of the previous section for the family of curves with $a = b^{-1}$, i.e. the low-temperature critical curves. We have already proven that if an inequality holds locally around the branch cut, it holds globally. Thus, we have only to check that the relevant inequalities occur in the correct direction. To accomplish this, we must first expand the function $\Omega(u)$ around the branch points.

Proposition 5.9. *Local expansions of* $\Omega_j(z)$ *around the branch points .*

1. *Around* $z = \pm\alpha := z(\pm b^{-1})$.

$$\tau\Omega_1(z) = \eta_\alpha(z) - q_2(z - \alpha)^{5/2} + \mathcal{O}((z - \alpha)^3), \tag{5.65}$$

$$\tau\Omega_2(z) = \eta_\alpha(z) + q_2(z - \alpha)^{5/2} + \mathcal{O}((z - \alpha)^3), \tag{5.66}$$

$$\tau\Omega_1(z) = \begin{cases} \eta_{-\alpha}(z) + iq_2(z+\alpha)^{5/2} + \mathcal{O}((z+\alpha)^3), & Im\ z > 0, \\ \eta_{-\alpha}(z) - iq_2(z+\alpha)^{5/2} + \mathcal{O}((z+\alpha)^3), & Im\ z < 0, \end{cases} \tag{5.67}$$

$$\tau\Omega_2(z) = \begin{cases} \eta_{-\alpha}(z) - iq_2(z-\alpha)^{5/2} + \mathcal{O}((z-\alpha)^3), & Im\ z > 0, \\ \eta_{-\alpha}(z) + iq_2(z+\alpha)^{5/2} + \mathcal{O}((z+\alpha)^3), & Im\ z < 0, \end{cases} \tag{5.68}$$

$$\tag{5.69}$$

Here, $\eta_{\pm\alpha}(z) := \tau\Omega(b^{-1}) \pm \tau\frac{3b^4+1}{b^2(b^4+3)}\alpha(z\mp\alpha) - \frac{\tau}{2b^2}(z\mp\alpha)^2$. *Note that the constant* $q_2 := \frac{8}{5}\frac{b^{-5/2}A^{-5/2}}{(3b^8+4b^4+3)\sqrt{1-b^4}} > 0$.

2. *Around* $z = \pm\beta := z(\pm b)$.

$$\tau\Omega_2(z) = \begin{cases} \eta_{\beta}(z) + i\tilde{q}_2(z-\beta)^{5/2} + \mathcal{O}((z-\beta)^3), & Im\ z > 0, \\ \eta_{\beta}(z) - i\tilde{q}_2(z-\beta)^{5/2} + \mathcal{O}((z-\beta)^3), & Im\ z > 0, \end{cases}, \tag{5.70}$$

$$\tau\Omega_4(z) = \begin{cases} \eta_{\beta}(z) - i\tilde{q}_2(z-\beta)^{5/2} + \mathcal{O}((z-\beta)^3), & Im\ z > 0, \\ \eta_{\beta}(z) + i\tilde{q}_2(z-\beta)^{5/2} + \mathcal{O}((z-\beta)^3), & Im\ z > 0, \end{cases}, \tag{5.71}$$

and

$$\tau\Omega_2(z) = \eta_{-\beta}(z) + \tilde{q}_2(z+\beta)^{5/2} + \mathcal{O}((z+\beta)^3), \tag{5.72}$$

$$\tau\Omega_3(z) = \eta_{-\beta}(z) - \tilde{q}_2(z+\beta)^{5/2} + \mathcal{O}((z+\beta)^3). \tag{5.73}$$

Here, $\eta_{\pm\beta}(z) = \tau\Omega(\pm b) \pm \tau\frac{b^2(b^4+3)}{3b^4+1}\beta(z\mp\beta) - \frac{\tau b^2}{2}(z\mp\beta)^2$. *Note that the constant* $\tilde{q}_2 := \frac{8}{5}\frac{b^{5/2}A^{-5/2}}{(3b^8+4b^4+3)\sqrt{1-b^4}} > 0$.

As before, for the lens opening procedure to work, we need certain inequalities to hold on the cuts. The inequalities we need are summarized in the following proposition.

Proposition 5.10. *Lensing around the cuts.*

Let $\Omega_j(z) = \phi_j(z) + i\psi_j(z)$, $j = 1, 2, 3, 4$. *Then, the following inequalities hold:*

1. $\phi_4(z) - \phi_2(z) > 0$ *in a lens around* $[\beta, \infty)$,
2. $\phi_3(z) - \phi_2(z) > 0$ *in a lens around* $(-\infty, -\beta]$,
3. $\phi_2(z) - \phi_1(z) > 0$ *in a lens around* $[-\alpha, \alpha]$.

Proof. The proof technique here is identical to the proof of the lensing Proposition (5.7) in the previous subsection; we thus do not repeat the proof here. □

Proposition 5.11. *Inequalities off the real axis. Let* $\Omega_j(z) = \phi_j(z) + i\psi_j(z)$*,* $j = 1, 2, 3, 4$*. Then, the following inequalities hold:*

1. $\phi_2(z) - \phi_4(z) > 0$ *for* $z \in \Gamma_1 \cap \{Im\ z > 0\}$*,*

2. $\phi_2(z) - \phi_3(z) > 0$ *for* $z \in \Gamma_2 \cap \{Im\ z < 0\}$*,*

3. $\phi_1(z) - \phi_2(z) > 0$ *for* $z \in \Gamma \setminus \{Im\ z = 0\}$*.*

Proof. The proof here is so similar to the generic case that we omit it. □

5.2.3.3 The High-Temperature Critical Curve: $0 < b < 1, a = 1$.

The high-temperature critical curve is characterized by the condition $a = 1$. As in the previous sections, all that is left to do is to show that the correct inequalities hold on each of the cuts. This requires us to expand the $\Omega_j(z)$'s around each of the branch points.

Proposition 5.12. *Local expansions of* $\Omega_j(z)$ *around the branch points .*

1. *(around* $z = z(\pm 1) := \pm\alpha$*).*

$$\tau\Omega_1(z) = \eta_\alpha(z) - q_3(z-\alpha)^{5/2} + \mathcal{O}((z-\alpha)^3), \tag{5.74}$$

$$\tau\Omega_2(z) = \eta_\alpha(z) + q_3(z-\alpha)^{5/2} + \mathcal{O}((z-\alpha)^3), \tag{5.75}$$

and

$$\tau\Omega_1(z) = \begin{cases} \eta_{-\alpha}(z) + iq_3(z+\alpha)^{5/2} + \mathcal{O}((z+\alpha)^3), & Im\ z > 0, \\ \eta_{-\alpha}(z) - iq_3(z+\alpha)^{5/2} + \mathcal{O}((z+\alpha)^3), & Im\ z < 0, \end{cases} \tag{5.76}$$

$$\tau\Omega_2(z) = \begin{cases} \eta_{-\alpha}(z) - iq_3 + \mathcal{O}((z+\alpha)^3), & Im\ z > 0, \\ \eta_{-\alpha}(z) + iq_3(z+\alpha)^{5/2} + \mathcal{O}((z+\alpha)^3), & Im\ z < 0, \end{cases}. \tag{5.77}$$

Here, $\eta_{\pm\alpha}(z) = \tau\Omega(\pm\alpha) + \pm\tau\alpha(z \mp \alpha) + \frac{\tau}{2}(z \mp \alpha)^2$*. Note that the constant* $q_3 := \frac{8}{5}\frac{1}{(1-b^2)^{5/2}(2b^2+3)A^{5/2}} > 0$*.*

2. *(around $z = z(\pm b) := \pm\beta$).*

$$\tau\Omega_2(z) = \begin{cases} \eta_\beta(z) + i\tilde{q}_3(z-\beta)^{3/2} + \mathcal{O}((z-\beta)^2), & Im\ z > 0, \\ \eta_\beta(z) - i\tilde{q}_3(z-\beta)^{3/2} + \mathcal{O}((z-\beta)^2), & Im\ z < 0, \end{cases} \tag{5.78}$$

$$\tau\Omega_4(z) = \begin{cases} \eta_\beta(z) - i\tilde{q}_3(z-\beta)^{3/2} + \mathcal{O}((z-\beta)^2), & Im\ z > 0, \\ \eta_\beta(z) + i\tilde{q}_3(z-\beta)^{3/2} + \mathcal{O}((z-\beta)^2), & Im\ z < 0, \end{cases} \tag{5.79}$$

and

$$\tau\Omega_2(z) = \eta_{-\beta}(z) + \tilde{q}_3(z+\beta)^{3/2} + \mathcal{O}((z+\beta)^2), \tag{5.80}$$

$$\tau\Omega_3(z) = \eta_{-\beta}(z) - \tilde{q}_3(z+\beta)^{3/2} + \mathcal{O}((z+\beta)^2). \tag{5.81}$$

Here, $\eta_{\pm\beta}(z) = \tau\Omega(\pm\beta) \pm \tau\beta\frac{-b^6+3b^4+3b^2+3}{6b^2+2}(z \mp \beta)$. Note that the constant $\tilde{q}_3 := A^{-3/2}\frac{\tau(b^4-1)^2(b^2+1)}{b^{5/2}(2b^2+3)\sqrt{1-b^2}} > 0$.

In order to open lenses, we need certain inequalities to hold on the cuts. In particular, we will need the following proposition.

Proposition 5.13. *Lensing around the cuts. Let $\Omega_j(z) = \phi_j(z) + i\psi_j(z)$, $j = 1,2,3,4$. Then, the following inequalities hold:*

1. *$\phi_4(z) - \phi_2(z) > 0$ in a lens around $[\beta, \infty)$,*

2. *$\phi_3(z) - \phi_2(z) > 0$ in a lens around $(-\infty, -\beta]$,*

3. *$\phi_2(z) - \phi_1(z) > 0$ in a lens around $[-\alpha, \alpha]$.*

Proof. Again, the proof technique is identical to that appearing in the generic case of Proposition (5.7); thus, we omit it. □

Proposition 5.14. *Inequalities off the real axis. Let $\Omega_j(z) = \phi_j(z) + i\psi_j(z)$, $j = 1,2,3,4$. Then, the following inequalities hold:*

1. *$\phi_2(z) - \phi_4(z) > 0$ for $z \in \Gamma_1 \cap \{Im\ z > 0\}$,*

2. *$\phi_2(z) - \phi_3(z) > 0$ for $z \in \Gamma_2 \cap \{Im\ z < 0\}$,*

3. *$\phi_1(z) - \phi_2(z) > 0$ for $z \in \Gamma \setminus \{Im\ z = 0\}$.*

Proof. The proof here is again similar to the generic and high temperature cases; therefore, we omit it. □

5.2.3.4 The Multicritical Case: $a = b = 1$.

In the case where $(a,b) \to (1,1)$ (equivalently, $(\tau, t) \to (\frac{1}{4}, -\frac{5}{72})$), the branch points $\pm\beta$ merge with $\pm\alpha$. The local structure of Ω near these branch points is quite different from that of the critical and generic cases, in that *three* sheets meet at $\pm\alpha$ instead of *two*. We begin as in previous sections, with a proposition characterizing the local behavior of the $\Omega_j(z)$'s.

Proposition 5.15. *The functions $\Omega_j(z)$ have the following expansions about the branch points $z = \pm\alpha$: Near $z = +\alpha$:*

$$\Omega_1(z) = \eta_+(z) - \frac{48}{35}\left(\frac{2}{\alpha}\right)^{7/3}(z-\alpha)^{7/3} + \mathcal{O}\left((z-\alpha)^{8/3}\right), \tag{5.82}$$

$$\Omega_2(z) = \begin{cases} \eta_+(z) - \frac{48\omega^2}{35}\left(\frac{2}{\alpha}\right)^{7/3}(z-\alpha)^{7/3} + \mathcal{O}\left((z-\alpha)^{8/3}\right), & Im\ z > 0, \\ \eta_+(z) - \frac{48\omega}{35}\left(\frac{2}{\alpha}\right)^{7/3}(z-\alpha)^{7/3} + \mathcal{O}\left((z-\alpha)^{8/3}\right), & Im\ z < 0, \end{cases} \tag{5.83}$$

$$\Omega_4(z) = \begin{cases} \eta_+(z) - \frac{48\omega}{35}\left(\frac{2}{\alpha}\right)^{7/3}(z-\alpha)^{7/3} + \mathcal{O}\left((z-\alpha)^{8/3}\right), & Im\ z > 0, \\ \eta_+(z) - \frac{48\omega^2}{35}\left(\frac{2}{\alpha}\right)^{7/3}(z-\alpha)^{7/3} + \mathcal{O}\left((z-\alpha)^{8/3}\right), & Im\ z < 0, \end{cases} \tag{5.84}$$

and $\Omega_3(z)$ is regular in a neighborhood of $z = +\alpha$.

Near $z = -\alpha$:

$$\Omega_1(z) = \begin{cases} \eta_-(z) + \frac{48\omega}{35}\left(\frac{2}{\alpha}\right)^{7/3}(z+\alpha)^{7/3} + \mathcal{O}\left((z+\alpha)^{8/3}\right), & Im\ z > 0, \\ \eta_-(z) + \frac{48\omega^2}{35}\left(\frac{2}{\alpha}\right)^{7/3}(z+\alpha)^{7/3} + \mathcal{O}\left((z+\alpha)^{8/3}\right), & Im\ z < 0, \end{cases} \tag{5.85}$$

$$\Omega_2(z) = \begin{cases} \eta_-(z) + \frac{48\omega^2}{35}\left(\frac{2}{\alpha}\right)^{7/3}(z+\alpha)^{7/3} + \mathcal{O}\left((z+\alpha)^{8/3}\right), & Im\ z > 0, \\ \eta_-(z) + \frac{48\omega}{35}\left(\frac{2}{\alpha}\right)^{7/3}(z+\alpha)^{7/3} + \mathcal{O}\left((z+\alpha)^{8/3}\right), & Im\ z < 0, \end{cases} \tag{5.86}$$

$$\Omega_3(z) = \eta_-(z) + \frac{48}{35}\left(\frac{2}{\alpha}\right)^{7/3}(z+\alpha)^{7/3} + \mathcal{O}\left((z+\alpha)^{8/3}\right), \tag{5.87}$$

and $\Omega_4(z)$ is regular in a neighborhood of $z = -\alpha$. Here, $\eta_\pm(z) := \frac{6}{5} \pm \alpha(z \mp \alpha) - \frac{1}{2}(z \mp \alpha)^2$.

Proof. The proof of this proposition follows the same procedure as in the previous cases; however, we find it useful to present this case separately. At the multicritical point, the uniformizing coordinate takes the form $z(u) = \sqrt{\frac{2}{15}}\left(3u + \frac{6}{u} - \frac{1}{u^3}\right)$, and $\Omega(z(u)) = -\frac{1}{10}u^4 + \frac{8}{5}u^2 - \frac{3}{10}u^{-4} - 4\log u$. Expanding in the uniformizing coordinate around the images of the branch points $u = \pm 1$, we obtain that

$$\Omega(z(u)) - \eta_\pm(z(u)) = \mp\frac{48}{35}(u \mp 1)^7 + \mathcal{O}((u \mp 1)^8).$$

Expanding $z(u)$ in the uniformizing coordinate,

$$z(u) \mp \alpha = \frac{\alpha}{2}(u \mp 1)^3 + \mathcal{O}((u \pm 1)^4).$$

Thus, locally, $u \pm 1$ has the expansion (with to be determined choice of branch)

$$u \pm 1 \sim \left(\frac{2}{\alpha}\right)^{1/3} (z \pm \alpha)^{1/3}.$$

The choice of branch is determined by the argument of $u \pm 1$ in the uniformzing plane lying in the correct sheet, along with the specification that $(z \pm \alpha)^{1/3}$ is always taken with the principal branch cut. This yields the result. □

Proposition 5.16. *Let $\Omega_j(z) = \phi_j(z) + i\psi_j(z)$, $j = 1, 2, 3, 4$. Then, the following inequalities hold:*

1. $\phi_4(z) - \phi_2(z) > 0$ *for z in a lens around* $[\alpha, \infty)$,
2. $\phi_3(z) - \phi_2(z) > 0$ *for z in a lens around* $(-\infty, -\alpha]$,
3. $\phi_2(z) - \phi_1(z) > 0$ *for z in a lens around* $[-\alpha, \alpha]$.

Proof. Again, let $\Omega_j(z) = \phi_j(z) + i\psi_j(z)$.

Since, by the lemmas, the direction of steepest descent of Re $\Omega(u)$ is constant along each of the connected components of Im $z = 0$, we have only to check that the necessary inequalities hold near the branch points $z = \pm\alpha$.

Near $z = +\alpha$, using Proposition (5.15), we have that

$$\phi_4(z) - \phi_2(z) = \text{Re }\left[\Omega_4(z) - \Omega_2(z)\right] = \begin{cases} \text{Re }\left[\frac{48(\omega^2-\omega)}{35}\left(\frac{2}{\alpha}\right)^{7/3}(z-\alpha)^{7/3} + \mathcal{O}\left((z-\alpha)^{8/3}\right)\right], & \text{Im } z > 0, \\ \text{Re }\left[\frac{48(\omega-\omega^2)}{35}\left(\frac{2}{\alpha}\right)^{7/3}(z-\alpha)^{7/3} + \mathcal{O}\left((z-\alpha)^{8/3}\right)\right], & \text{Im } z < 0, \end{cases}$$

from which we can read off that $\phi_4(z) - \phi_2(z) > 0$ in the sector $|\arg(z-\alpha)| < \frac{3\pi}{7}$ for $|z - \alpha|$ sufficiently small; this proves 1.

Near $z = -\alpha$, again using Proposition (5.15), we have that

$$\phi_3(z) - \phi_2(z) = \text{Re }\left[\Omega_3(z) - \Omega_2(z)\right] = \begin{cases} \text{Re }\left[-\frac{48(\omega^2-1)}{35}\left(\frac{2}{\alpha}\right)^{7/3}(z+\alpha)^{7/3} + \mathcal{O}\left((z+\alpha)^{8/3}\right)\right], & \text{Im } z > 0, \\ \text{Re }\left[-\frac{48(\omega-1)}{35}\left(\frac{2}{\alpha}\right)^{7/3}(z+\alpha)^{7/3} + \mathcal{O}\left((z+\alpha)^{8/3}\right)\right], & \text{Im } z < 0, \end{cases}$$

from which we can read off that $\phi_3(z) - \phi_2(z) > 0$ in the sector $\frac{4\pi}{7} < |\arg(z+\alpha)| \leq \pi$ for $|z+\alpha|$ sufficiently small, so that 2. follows.

Finally, we determine the behavior of $\phi_2(z) - \phi_1(z)$ around either branch point; near $z = +\alpha$, say. We have that

$$\phi_2(z) - \phi_1(z) = \text{Re}\,[\Omega_2(z) - \Omega_1(z)] = \begin{cases} \text{Re}\,[\frac{48(1-\omega^2)}{35}\left(\frac{2}{\alpha}\right)^{7/3}(z-\alpha)^{7/3} + \mathcal{O}\left((z-\alpha)^{8/3}\right)], & \text{Im}\, z > 0, \\ \text{Re}\,[\frac{48(1-\omega)}{35}\left(\frac{2}{\alpha}\right)^{7/3}(z-\alpha)^{7/3} + \mathcal{O}\left((z-\alpha)^{8/3}\right)], & \text{Im}\, z < 0; \end{cases}$$

this expansion tells us that $\phi_2(z) - \phi_1(z) > 0$ in particular in the sector $\frac{4\pi}{7} < |\arg(z-\alpha)| \leq \pi$, for $|z-\alpha|$ sufficiently small. A similar analysis near $z = -\alpha$ shows that $\phi_2(z) - \phi_1(z) > 0$ in the sector $|\arg(z-\alpha)| < \frac{3\pi}{7}$ for $|z+\alpha|$ sufficiently small. Thus, we have established locally the inequalities 1. through 3.; as a consequence of Proposition (5.5), the inequalities hold in a neighborhood of each of the segments of the real line. □

Proposition 5.17. *Inequalities off the real axis. Let $\Omega_j(z) = \phi_j(z) + i\psi_j(z)$, $j = 1,2,3,4$. Then, the following inequalities hold:*

1. $\phi_2(z) - \phi_4(z) > 0$ *for* $z \in \Gamma_1 \cap \{Im\ z > 0\}$,

2. $\phi_2(z) - \phi_3(z) > 0$ *for* $z \in \Gamma_2 \cap \{Im\ z < 0\}$,

3. $\phi_1(z) - \phi_2(z) > 0$ *for* $z \in \Gamma \setminus \{Im\ z = 0\}$.

5.3 The First Transformation $\mathbf{Y} \rightarrow \mathbf{X}$.

5.3.1 Idea of the Transformation.

The main idea of the transformation $\mathbf{Y} \mapsto \mathbf{X}$ is illustrated in [39], who refer to an unpublished manuscript of Bertola, Harnad, and Its as the origin of the idea. The point is that the weights appearing in the jump matrix, $f(z)$ and its derivatives, satisfy a modified form of the Pearcey differential equation:

$$\frac{t}{N^2\tau^2} f'''(z) + f'(z) - N\tau^2 z f(z) = 0, \tag{5.88}$$

whose solutions are the Pearcey-type integrals

$$w_j(z) := \int_{\gamma_j} \exp\,[N(\tau z w \underbrace{- \frac{1}{2}w^2 - \frac{t}{4}w^4}_{-V(w)})]dw. \tag{5.89}$$

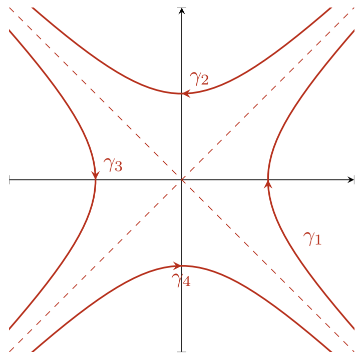


Figure 5.11. The contours γ_j, $j = 1, ..., 4$.

It is also useful to notice that any solution $w(z) = w(z; \tau, t, N)$ to (5.88) satisfies the partial differential equations

$$\frac{\partial w}{\partial t} = -\frac{1}{4N^3\tau^4}\frac{\partial^4 w}{\partial z^4} = \frac{1}{4N\tau^2 t}\frac{\partial^2 w}{\partial z^2} - \frac{z}{4t}\frac{\partial w}{\partial z} - \frac{1}{4t}w, \tag{5.90}$$

$$\tau\frac{\partial w}{\partial \tau} = z\frac{\partial w}{\partial z}. \tag{5.91}$$

In the following subsection, we analyze the asymptotics of these integrals, and consider an associated Riemann-Hilbert problem for the $w_j(z)$'s, which we shall make use of in the first transformation $\mathbf{Y} \mapsto \mathbf{X}$.

5.3.2 RHP and Asymptotics For the Pearcey-Type Integrals

We begin by utilizing classical steepest descent analysis to determine the asymptotics of each of the integrals

$$w_j(z) = \int_{\gamma_j} \exp\left[N(\tau z w - V(w))\right] dw, \tag{5.92}$$

as $z \to \infty$, where γ_j, $j = 1, ..., 4$ are one of the contours in Figure (5.11). Making the change of variables $w = z^{1/3}\zeta$, the functions $w_j(z)$ can be rewritten as

$$w_j(z) = z^{1/3}\int_{z^{1/3}\cdot\gamma_j} \exp\left[Nz^{4/3}\left\{\tau\zeta - \frac{t}{4}\zeta^4 - \frac{1}{2z^{2/3}}\zeta^2\right\}\right] d\zeta, \tag{5.93}$$

The large z saddle points of the integrand are determined by the roots of the polynomial

$$0 = \frac{d}{d\zeta}\left[\tau\zeta - \frac{t}{4}\zeta^4 - \frac{1}{2z^{2/3}}\zeta^2\right] = \tau - t\zeta^3 - \frac{1}{z^{2/3}}\zeta \tag{5.94}$$

We label the solutions of the above equation as $s_k(z)$, $k = 1,2,3$; these solutions have the asymptotic expansion

$$s_k(z) = -\frac{\tau^{1/3}\omega^{k-1}}{(-t)^{1/3}} - \frac{\omega^{1-k}}{3\tau^{1/3}(-t)^{2/3}} z^{-2/3} + \frac{\omega^{k-1}}{81\tau^{5/3}(-t)^{4/3}} z^{-2} + \frac{\omega^{1-k}}{243\tau^{11/3}} z^{-8/3} + \mathcal{O}(z^{-4}). \tag{5.95}$$

Here, we take the principal branch of the cube root. Since we only need the large z asymptotics of the saddle points, this definition of the $s_k(z)$'s is sufficient for our purposes. The exponent evaluated at the saddle points is then

$$\begin{aligned} N\theta_k(z) :&= Nz^{4/3}\left[\tau s_k(z) - \frac{t}{4}s_k(z)^4 - \frac{1}{2z^{2/3}}s_k(z)^2\right] \\ &= N\left[-\frac{3\omega^{k-1}}{4}\frac{\tau^{4/3}}{(-t)^{1/3}}z^{4/3} - \frac{\omega^{1-k}}{2}\frac{\tau^{2/3}}{(-t)^{2/3}}z^{2/3} + \frac{1}{6t} - \frac{\omega^{k-1}}{54\tau^{2/3}(-t)^{4/3}z^{2/3}} + \mathcal{O}\left(\frac{1}{|z|^{4/3}}\right)\right], \end{aligned} \tag{5.96}$$

as $z \to \infty$. We also need the quantities

$$\begin{aligned} &z^{1/3}\sqrt{\frac{-2\pi}{Nz^{4/3}\left[z^{-2/3} + 3ts_k(z)^2\right]}} \\ &= i\sqrt{\frac{2\pi}{3N}}\left(\frac{\omega^{1-k}}{(-t)^{1/6}\tau^{1/3}z^{1/3}} - \frac{1}{6(-t)^{1/2}\tau z} - \frac{\omega^{k-1}}{72}\frac{1}{(-t)^{5/6}\tau^{5/3}z^{5/3}} + \mathcal{O}\left(\frac{1}{|z|^{7/3}}\right)\right). \end{aligned} \tag{5.97}$$

We always choose the branch cuts along the negative real axis, and we have chosen principal branch of the square root above. Now, the saddle point approximation tells us that the functions $w_j(z)$ are approximated by

$$w_j(z) = \int_{\gamma_j} \exp\left[N(\tau zw - V(w))\right]dw = z^{1/3}\sqrt{\frac{-2\pi}{N\left[z^{2/3} + 3tz^{4/3}s_*(z)^2\right]}}e^{N\theta_*(z)}\left(1 + \mathcal{O}\left(\frac{1}{z^{4/3}}\right)\right), \tag{5.98}$$

$z \to \infty$, where $s_*(z)$ represents the nearest contributing saddle point to the contour γ_j, which gives the dominant contribution to the integral. For the k^{th} saddle point, we call the dominant contribution $S_k(z)$:

$$\begin{aligned} S_k(z) &:= z^{1/3}\sqrt{\frac{-2\pi}{N\left[z^{2/3} + 3tz^{4/3}s_k(z)^2\right]}}e^{N\theta_k(z)} \\ &= i\sqrt{\frac{2\pi}{3N}}\left(\frac{\omega^{1-k}}{(-t)^{1/6}\tau^{1/3}z^{1/3}} - \frac{1}{6(-t)^{1/2}\tau z} - \frac{\omega^{k-1}}{72}\frac{1}{(-t)^{5/6}\tau^{5/3}z^{5/3}} + \mathcal{O}\left(\frac{1}{|z|^{7/3}}\right)\right) \\ &\times \exp\left[N\left[-\frac{3\omega^{k-1}}{4}\frac{\tau^{4/3}}{(-t)^{1/3}}z^{4/3} - \frac{\omega^{1-k}}{2}\frac{\tau^{2/3}}{(-t)^{2/3}}z^{2/3} + \frac{1}{6t} - \frac{\omega^{k-1}}{54\tau^{2/3}(-t)^{4/3}z^{2/3}} + \mathcal{O}\left(\frac{1}{|z|^{4/3}}\right)\right]\right]. \end{aligned} \tag{5.99}$$

From this formula, we can write the asymptotics for each of the $w_j(z)$'s by determining the dominant saddle point; this saddle point depends on $\arg z$. We have the following large-z asymptotic formulae:

$$w_1(z) = \begin{cases} -S_2(z)[1+\mathcal{O}(z^{-4/3})], & 0 < \arg z < \pi, \\ -S_3(z)[1+\mathcal{O}(z^{-4/3})], & -\pi < \arg z < 0. \end{cases} \tag{5.100}$$

$$w_2(z) = \begin{cases} S_3(z)[1+\mathcal{O}(z^{-4/3})], & -\frac{\pi}{2} < \arg z < \pi, \\ S_1(z)[1+\mathcal{O}(z^{-4/3})], & -\pi < \arg z < -\frac{\pi}{2}. \end{cases} \tag{5.101}$$

$$w_3(z) = -S_1(z)[1+\mathcal{O}(z^{-4/3})], \qquad |\arg z| < \pi, \tag{5.102}$$

$$w_4(z) = \begin{cases} S_1(z)[1+\mathcal{O}(z^{-4/3})], & \frac{\pi}{2} < \arg z < \pi, \\ S_2(z)[1+\mathcal{O}(z^{-4/3})], & -\pi < \arg z < \frac{\pi}{2}. \end{cases} \tag{5.103}$$

The signs in front of the $S_k(z)$'s are chosen to be consistent with the orientation of the contours.

Define the row vectors

$$\vec{w}_j(z) := \left(w_j(z), \frac{w_j'(z)}{N\tau}, \frac{w_j''(z)}{(N\tau)^2} \right), \tag{5.104}$$

where $'$ here denotes the derivative with respect to z. We now define a 3×3 matrix $\mathbf{W}(z)$ as

$$\mathbf{W}(z) = \begin{cases} \begin{pmatrix} -\vec{w}_2(z) \\ \vec{w}_3(z) \\ \vec{w}_1(z) \end{pmatrix}, & z \in \Omega_u, \\ \begin{pmatrix} \vec{w}_3(z) + \vec{w}_4(z) \\ \vec{w}_3(z) \\ \vec{w}_1(z) \end{pmatrix}, & z \in \Omega_c, \\ \begin{pmatrix} \vec{w}_4(z) \\ \vec{w}_3(z) \\ \vec{w}_1(z) \end{pmatrix}, & z \in \Omega_\ell. \end{cases} \tag{5.105}$$

$\mathbf{W}(z)$ is defined so that it is the unique solution to the following Riemann-Hilbert problem:

Theorem 5.18. $\mathbf{W}(z)$ *is the unique solution to the Riemann-Hilbert problem*

$$\tilde{\mathbf{W}}_+(z) = \begin{cases} J_a \tilde{\mathbf{W}}_-(z), & z \in \Gamma_1, \\ J_b \tilde{\mathbf{W}}_-(z), & z \in \Gamma_2, \end{cases} \tag{5.106}$$

where the matrices J_a, J_b *are defined as*

$$J_a = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad J_b = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{5.107}$$

Note that $J_a J_b = J_b J_a = J$*. Furthermore,* $\mathbf{W}(z)$ *has asymptotics*

$$\mathbf{W}(z) = c_N \cdot e^{N\Lambda(z)} A(z) B(z) \hat{K} \left[\mathbb{I}_{3\times 3} + \mathcal{O}\left(\frac{1}{z}\right) \right], \qquad z \to \infty. \tag{5.108}$$

Here, the constant $c_N := i\sqrt{\frac{2\pi}{3N}}$*. The matrix* $\Lambda(z)$ *is defined to be*

$$\Lambda(z) = \begin{cases} diag(\hat{\theta}_3(z), \hat{\theta}_1(z), \hat{\theta}_2(z)), & Im\ z > 0, \\ diag(\hat{\theta}_2(z), \hat{\theta}_1(z), \hat{\theta}_3(z)), & Im\ z < 0, \end{cases} \tag{5.109}$$

where the functions $\hat{\theta}_j(z)$ *are defined by the exact formulas*

$$\hat{\theta}_j(z) = -\frac{3\omega^{k-1}}{4} \frac{\tau^{4/3}}{(-t)^{1/3}} z^{4/3} - \frac{\omega^{1-k}}{2} \frac{\tau^{2/3}}{(-t)^{2/3}} z^{2/3} + \frac{1}{6t}. \tag{5.110}$$

The matrices $A(z)$*,* $B(z)$*, and* $\hat{K}$ *are defined by*

$$A(z) = \begin{cases} \begin{pmatrix} -\omega & 1 & \omega^2 \\ -1 & 1 & 1 \\ -\omega^2 & 1 & \omega \end{pmatrix}, & Im\ z > 0, \\ \begin{pmatrix} \omega^2 & -1 & -\omega \\ -1 & 1 & 1 \\ -\omega & 1 & \omega^2 \end{pmatrix}, & Im\ z < 0. \end{cases} \tag{5.111}$$

$$B(z) = \begin{pmatrix} z^{-1/3} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & z^{1/3} \end{pmatrix}, \tag{5.112}$$

$$\hat{K} = \begin{pmatrix} (-t)^{-1/6}\tau^{1/3} & 0 & -\frac{n+27t}{54(-t)^{13/6}\tau^{1/3}} \\ 0 & (-t)^{-1/2} & 0 \\ 0 & 0 & -(-t)^{-5/6}\tau^{1/3} \end{pmatrix}. \tag{5.113}$$

Proof. The proof of this theorem is just a tedious check that the formulas we have previously derived for the asymptotics of the $w_j(z)$'s indeed guarantee that the $\mathbf{W}(z)$ solves the above Riemann-Hilbert problem. Uniqueness follows from standard arguments with Morera's and Liouville's theorem. We remark that the terms decaying in z as $z \to \infty$ in the functions $\theta_j(z)$ can be absorbed into the asympototic expansion $\mathbb{I} + \mathcal{O}(z^{-1})$, by passing this part of the expansion to the right of the matrices $A(z)B(z)$. More precisely, the expansion of the Laurent series is [3]

$$\mathbb{I}_{3\times 3} + \begin{pmatrix} 0 & \frac{-N^3-72N^2t-891Nt^2-810t^3}{5832Nt^3\tau} & 0 \\ \frac{-N+9t}{54\tau t} & 0 & \frac{N^3+36N^2t-81Nt^2-810t^3}{5832Nt^3\tau} \\ 0 & \frac{-N-9t}{54\tau t} & 0 \end{pmatrix} z^{-1} + \mathcal{O}(z^{-2}).$$

□

We furthermore have the following proposition, which will be useful later in the computation of the τ-function.

Proposition 5.19. $\mathbf{W}(z)$ *satisfies the following differential equations:*

$$\frac{\partial \mathbf{W}}{\partial z} = \mathbf{W} \cdot \mathcal{M}^z(z), \qquad \frac{\partial \mathbf{W}}{\partial t} = \mathbf{W} \cdot \mathcal{M}^t(z), \qquad \frac{\partial \mathbf{W}}{\partial \tau} = \mathbf{W} \cdot \mathcal{M}^\tau(z) \tag{5.114}$$

where the matrices $\mathcal{M}^z(z)$, $\mathcal{M}^t(z)$, *and* $\mathcal{M}^\tau(z)$ *are defined by*

$$\mathcal{M}^z(z) = N\tau \begin{pmatrix} 0 & 0 & \tau z/t \\ 1 & 0 & -1/t \\ 0 & 1 & 0 \end{pmatrix}, \tag{5.115}$$

$$\mathcal{M}^t(z) = \frac{N}{4t} \begin{pmatrix} -1/N & \tau z/t & -\tau^2 z^2/t + \frac{1}{Nt} \\ -\tau z & -(1/t + 2/N) & 2\tau z/t \\ 1 & -\tau z & -(1/t + 3/N) \end{pmatrix}, \tag{5.116}$$

$$\mathcal{M}^\tau(z) = Nz \begin{pmatrix} 0 & 0 & \tau z/t \\ 1 & 0 & -1/t \\ 0 & 1 & 0 \end{pmatrix}. \tag{5.117}$$

Proof. This is an immediate consequence of the differential equation (5.88) and the relations $\frac{\partial w}{\partial t} = -\frac{1}{4N^3\tau^4}\frac{\partial^4 w}{\partial z^4}$, and $\tau\frac{\partial w}{\partial \tau} = z\frac{\partial w}{\partial z}$. As a consistency check, one may verify that the compatability conditions between these equations (i.e., the zero-curvature equations) hold trivially. □

[3] The regular series expansion here either by (i) calculating subleading asymptotics of the saddle point expansion to high enough order, or (ii) using the fact that the asymptotic expansion itself should satisfy the differential equation(s) (5.114), and determining the coefficients of the subleading expansion recursively.

Remark 5.20. It is useful to notice that the function $A(z)B(z)$ is a solution to the following Riemann-Hilbert problem: $A(z)B(z)$ is analytic in $\mathbb{C}\setminus\mathbb{R}$, and satisfies the jump condition

$$A_+(z)B_+(z) = \begin{cases} \left(\begin{smallmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{smallmatrix}\right) A_-(z)B_-(z), & z>0, \\ \left(\begin{smallmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{smallmatrix}\right) A_-(z)B_-(z), & z<0. \end{cases} \tag{5.118}$$

5.3.3 The Transformation $\mathbf{Y} \to \mathbf{X}$

We now define the transformation $\mathbf{Y} \mapsto \mathbf{X}$. We set

$$\mathbf{X}(z) := \left(\begin{smallmatrix} 1 & 0 \\ 0 & c_N \hat{K} \end{smallmatrix}\right) \mathbf{Y}(z) \begin{pmatrix} e^{-NV(z)} & 0 \\ 0 & \mathbf{W}^{-1}(z) \end{pmatrix}. \tag{5.119}$$

By construction, $\mathbf{X}(z)$ is piecewise analytic on $\mathbb{C}\setminus(\Gamma_1\cup\Gamma_2\cup\Gamma)$. $\mathbf{X}(z)$ satisfies the following RHP:

Proposition 5.21. $\mathbf{X}(z)$ *is the unique solution to the following Riemann-Hilbert Problem:*

$$\mathbf{X}_+(z) = \mathbf{X}_-(z) \times \begin{cases} \left(\begin{smallmatrix} 1 & 0 \\ 0 & J_a^{-1} \end{smallmatrix}\right) & z\in\Gamma_1, \\ \left(\begin{smallmatrix} 1 & 0 \\ 0 & J_b^{-1} \end{smallmatrix}\right) & z\in\Gamma_2, \\ \left(\begin{smallmatrix} 1 & 1\,0\,0 \\ \vec{0}_{1\times 3} & \mathbb{I}_{3\times 3} \end{smallmatrix}\right) & z\in\Gamma, \end{cases} \tag{5.120}$$

Subject to the asymptotic condition

$$\mathbf{X}(z) = \left[\mathbb{I} + \mathcal{O}\left(\frac{1}{z}\right)\right] \left(\begin{smallmatrix} 1 & 0 \\ 0 & B^{-1}(z)A^{-1}(z) \end{smallmatrix}\right) \left(\begin{smallmatrix} z^n & 0 & 0 & 0 \\ 0 & z^{-n/3} & 0 & 0 \\ 0 & 0 & z^{-n/3} & 0 \\ 0 & 0 & 0 & z^{-n/3} \end{smallmatrix}\right) e^{-N\hat{\Theta}(z)}, \tag{5.121}$$

Where $\Theta(z)$ *is defined as the diagonal matrix*

$$\hat{\Theta}(z) = \begin{pmatrix} V(z) & 0 \\ 0 & \Lambda(z) \end{pmatrix}, \tag{5.122}$$

where $\Lambda(z)$ *is the diagonal* 3×3 *matrix defined by* (5.109).

Proof. The asymptotic condition (5.121) follows almost immediately from the definition of $\mathbf{X}(z)$; along with the definition of the matrix $\mathbf{W}(z)$. Indeed, this is obvious in the regions Ω_u and Ω_ℓ, and follows from definition of $\mathbf{W}(z)$. The only detail to check is that the asymptotics of $\mathbf{W}(z)$ in the region Ω_c are correct.

We see that $\mathbf{W}(z)$ in this region is obtained by adding the recessive solution $-\vec{w}_1(z)$ to the first row; since this solution is recessive in the region Ω_c, the asymptotics of $-\vec{w}_2(z)-\vec{w}_1(z)(=\vec{w}_3(z)+\vec{w}_4(z))$ are the same as the asymptotics of $-\vec{w}_2(z)$ there. Thus, the asymptotics of $\mathbf{X}(z)$ from the proposition hold.

Now, we address the jump conditions of $\mathbf{X}(z)$. Since the matrix function

$$\begin{pmatrix} e^{-NV(z)} & 0 \\ 0 & \mathbf{W}^{-1}(z) \end{pmatrix}$$

has jumps only on Γ_1, Γ_2 (arising from the jumps of $\mathbf{W}(z)$), and these contours do not intersect Γ, the first two jump conditions are clearly satisfied. It remains to check the jump of $\mathbf{X}(z)$ across Γ. The jump of $\mathbf{Y}(z)$ on Γ is

$$J_{\mathbf{Y}} = \mathbb{I} + e^{-NV(z)} \begin{pmatrix} 0 & f(z) & \frac{f'(z)}{N\tau} & \frac{f''(z)}{(N\tau)^2} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} =: \mathbb{I} + e^{-NV(z)} \begin{pmatrix} 0 & & \vec{f}(z) & \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

where $\vec{f}(z)$ is the row vector

$$\vec{f}(z) := \left(f(z), \frac{f'(z)}{N\tau}, \frac{f''(z)}{(N\tau)^2} \right).$$

Now, since Γ is homologically equivalent to $-\gamma_2 - \gamma_1$, we have that $\vec{L}(z) = -\vec{w}_2(z) - \vec{w}_1(z)$. Thus, the jump of $\mathbf{X}(z)$ across Γ is

$$\begin{aligned} &\begin{pmatrix} e^{NV(z)} & 0 \\ 0 & \mathbf{W}(z) \end{pmatrix} \left[\mathbb{I} + e^{-NV(z)} \begin{pmatrix} 0 & & \vec{f}(z) & \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \right] \begin{pmatrix} e^{-NV(z)} & 0 \\ 0 & \mathbf{W}^{-1}(z) \end{pmatrix} \\ &= \mathbb{I} + \begin{pmatrix} 0 & & \vec{f}(z)\,\mathbf{W}^{-1}(z) & \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = \mathbb{I} + \begin{pmatrix} 0 & & (-\vec{w}_2(z)-\vec{w}_1(z))\,\mathbf{W}^{-1}(z) & \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \end{aligned}$$

Here, we have used the relations

$$\begin{aligned} \vec{w}_1(z)\,\mathbf{W}^{-1}(z) &= (0,0,1), & \vec{w}_2(z)\,\mathbf{W}^{-1}(z) &= (-1,0,-1), \\ \vec{w}_3(z)\,\mathbf{W}^{-1}(z) &= (0,1,0), & \vec{w}_4(z)\,\mathbf{W}^{-1}(z) &= (1,-1,0), \end{aligned}$$

resulting from the fact that, in a neighborhood of Γ, the matrix $\mathbf{W}(z)$ admits the expression

$$\mathbf{W}(z) = \begin{pmatrix} \vec{w}_3(z) + \vec{w}_4(z) \\ \vec{w}_3(z) \\ \vec{w}_1(z) \end{pmatrix},$$

along with the identity $\mathbf{W}(z)\,\mathbf{W}^{-1}(z) = \mathbb{I}_{3\times 3}$.

□

Remark 5.22. Note that, if we had equivalently chosen $\gamma_3 + \gamma_4$ as the homological representative for Γ, the same resulting jump matrix is obtained.

5.4 The Second Transformation $\mathbf{X} \to \mathbf{U}$.

Define the matrix

$$\mathbf{G}(z) := \begin{pmatrix} \exp\left[n\tau\Omega_1(z)\right] & 0 & 0 & 0 \\ 0 & \exp\left[n\tau\Omega_2(z)\right] & 0 & 0 \\ 0 & 0 & \exp\left[n\tau\Omega_3(z)\right] & 0 \\ 0 & 0 & 0 & \exp\left[n\tau\Omega_4(z)\right] \end{pmatrix}. \tag{5.123}$$

We now are ready to perform the transformation $\mathbf{X} \mapsto \mathbf{U}$. Set

$$\mathbf{U}(z) := \left[\mathbb{I} - nC_1 \cdot E_{24}\right] e^{-nL}\, \mathbf{X}(z)\, \mathbf{G}(z), \tag{5.124}$$

where $\mathbf{G}(z)$ is defined as above, C_1 is the constant appearing in the $z^{-2/3}$ term of the asymptotics of the $\Omega_j(z)$'s (cf. Equation (5.48)), and L is the diagonal constant matrix

$$L := \operatorname{diag}\,(\ell_0, \ell_1 - \frac{1}{6t}, \ell_1 - \frac{1}{6t}, \ell_1 - \frac{1}{6t}). \tag{5.125}$$

Clearly, $\mathbf{U}(z)$ is analytic in $\mathbb{C} \setminus (\Gamma \cup \Gamma_1 \cup \Gamma_2)$; the goal of this subsection is to show that $\mathbf{U}(z)$ is the unique solution to its own Riemann-Hilbert problem.

Proposition 5.23. *The function* $\mathbf{U}(z)$ *is the unique solution to the following Riemann-Hilbert problem:*

$$
\mathbf{U}_+(z) = \mathbf{U}_-(z)\times
\begin{cases}
\mathbb{I} - E_{24}e^{-n\tau[\Omega_2(z)-\Omega_4(z)]}, & z \in \Gamma_1 \cap \{Im\ z > 0\}, \\
\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & e^{-n\tau[\Omega_{2,-}(z)-\Omega_{4,-}(z)]} & 0 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & e^{n\tau[\Omega_{2,-}(z)-\Omega_{4,-}(z)]} \end{pmatrix}, & z \in \Gamma_1 \cap \{Im\ z = 0\}, \\
\mathbb{I} - E_{23}e^{-n\tau[\Omega_2(z)-\Omega_3(z)]}, & z \in \Gamma_2 \cap \{Im\ z < 0\}, \\
\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & e^{-n\tau[\Omega_{2,-}(z)-\Omega_{3,-}(z)]} & -1 & 0 \\ 0 & 0 & e^{n\tau[\Omega_{2,-}(z)-\Omega_{3,-}(z)]} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, & z \in \Gamma_2 \cap \{Im\ z = 0\}, \\
\mathbb{I} + E_{12}e^{-n\tau[\Omega_1(z)-\Omega_2(z)]}, & z \in \Gamma \setminus \{Im\ z = 0\}, \\
\begin{pmatrix} e^{-n\tau[\Omega_{1,-}(z)-\Omega_{2,-}(z)]} & 1 & 0 & 0 \\ 0 & e^{n\tau[\Omega_{1,-}(z)-\Omega_{2,-}(z)]} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, & z \in \Gamma \cap \{Im\ z = 0\}.
\end{cases}
\tag{5.126}
$$

The asymptotics of $\mathbf{U}(z)$ *are given by*

$$
\mathbf{U}(z) = \left[\mathbb{I} + \mathcal{O}\left(\frac{1}{z^{1/3}}\right)\right]\begin{pmatrix} 1 & 0 \\ 0 & B^{-1}(z)A^{-1}(z) \end{pmatrix}, \qquad z \to \infty. \tag{5.127}
$$

Proof. The jump conditions satisfied by $\mathbf{U}(z)$ are readily verified from the definitions of $\mathbf{X}(z)$, $\mathbf{G}(z)$. Furthermore, the 'exponential asymptotics' of $\mathbf{X}(z)$ are removed by multiplication by $\mathbf{G}(z)$; comparison of formulas (5.41)-(5.44) with (5.121), along with the explicit expressions for the asymptotics of the functions $\theta_k(z)$ (see equation (5.96)) shows that this is indeed the case. Indeed, we have that, as $z \to \infty$,

$$
\begin{aligned}
\mathbf{U}(z) &= [\mathbb{I} - nC_1 \cdot E_{24}]\, e^{-nL}[\mathbb{I} + \mathcal{O}(z^{-1})]\begin{pmatrix} z^n & 0 \\ 0 & z^{-n/3}\mathbb{I}_{3\times 3} \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & B^{-1}(z)A^{-1}(z) \end{pmatrix}e^{-n\hat{\Theta}(z)}\,\mathbf{G}(z) \\
&= [\mathbb{I} - nC_1 \cdot E_{24}]\,[\mathbb{I} + \mathcal{O}(z^{-1})]\begin{pmatrix} 1 & 0 \\ 0 & B^{-1}(z)A^{-1}(z) \end{pmatrix}\begin{pmatrix} 1+\mathcal{O}(z^{-2}) & 0 \\ 0 & \mathbb{I}_{3\times 3}+n\hat{C}z^{-2/3}+\mathcal{O}(z^{-4/3}) \end{pmatrix},
\end{aligned}
$$

Here, $\hat{C}$ is the piecewise constant diagonal matrix

$$
\hat{C} = \begin{cases} \text{diag}\ (\omega^2 C_1, C_1, \omega C_1), & \text{Im}\ z > 0, \\ \text{diag}\ (\omega C_1, C_1, \omega^2 C_1), & \text{Im}\ z < 0. \end{cases}
$$

If we interchange the order of the last two matrices, we obtain that, as $z \to \infty$,

$$\begin{aligned}\mathbf{U}(z) &= [\mathbb{I} - nC_1 \cdot E_{24}]\,[\mathbb{I} + \mathcal{O}(z^{-1})]\left[\mathbb{I} + nC_1 \cdot E_{24} + \mathcal{O}(z^{-1/3})\right]\left(\begin{smallmatrix} 1 & 0 \\ 0 & B^{-1}(z)A^{-1}(z) \end{smallmatrix}\right) \\ &= \left[\mathbb{I} + \mathcal{O}(z^{-1/3})\right]\left(\begin{smallmatrix} 1 & 0 \\ 0 & B^{-1}(z)A^{-1}(z) \end{smallmatrix}\right).\end{aligned}$$

We will analyze the jumps of this new matrix after opening lenses; this is performed in the next section.

□

5.5 The Third and Fourth Transformations $\mathbf{U} \to \mathbf{T} \to \mathbf{S}$.

Here, we perform the lensing transformations. The first lensing transformation will open lenses around the unbounded branch cuts $(-\infty, -\beta] \cup [\beta, \infty)$; this will constitute the transformation $\mathbf{U} \mapsto \mathbf{T}$. The lens opening around the central cut $[-\alpha, \alpha]$ is performed next; this will constitute the transformation $\mathbf{T} \mapsto \mathbf{S}$. We remark here that the choice of (a, b) (i.e., whether the spectral curve is critical, generic, or multicritical) is irrelevant here, as all of the lensing propositions of Section 5.2.3 guarantee the same inequalities hold around the branch points. These cases will become distinguished when we later try to find a parametrix.

5.5.1 The Third Transformation $\mathbf{U} \to \mathbf{T}$

The opening of lenses here is based on the factorization of the jump matrix

$$\begin{pmatrix} e^{-ng_+(z)} & -1 \\ 0 & e^{-ng_-(z)} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -e^{-ng_-(z)} & 1 \end{pmatrix}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ -e^{-ng_+(z)} & 1 \end{pmatrix}, \tag{5.128}$$

where $g_+(z)$, $g_-(z)$ are the boundary values of one of the functions $\Omega_3(z) - \Omega_2(z)$, $\Omega_4(z) - \Omega_2(z)$ from above/below the contour.

By the lensing propositions of the previous section, there exist lens-shaped regions such as this depicted in Figure (5.10) around $(-\infty, -\beta]$ (respectively, $[\beta, \infty)$) such that the differences Re $[\Omega_3 - \Omega_2]$ (respectively, Re $[\Omega_4 - \Omega_2]$) are positive in this region. Define $\Gamma_{1,u}, \Gamma_{1,l}$ as the boundaries of the lensing region around $[\beta, \infty)$ in the upper and lower half planes, and similarly define $\Gamma_{2,u}, \Gamma_{2,l}$ as the boundaries of the lensing region around $(-\infty, -\beta]$. The sectors enclosed by these contours are labelled as follows:

- $\Sigma_{1,u}$ is the region enclosed by $\Gamma_{1,u}$ and $[\beta, \infty)$,
- $\Sigma_{1,l}$ is the region enclosed by $\Gamma_{1,l}$ and $[\beta, \infty)$,
- $\Sigma_{2,u}$ is the region enclosed by $\Gamma_{2,u}$ and $(-\infty, -\beta]$,

- $\Sigma_{2,l}$ is the region enclosed by $\Gamma_{2,l}$ and $(-\infty,-\beta]$.

These contours are depicted in Figure (5.10). Define matrices

$$V_1(z) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -e^{-n\tau[\Omega_4(z)-\Omega_2(z)]} & 0 & 1 \end{pmatrix}, \qquad V_2(z) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -e^{-n\tau[\Omega_3(z)-\Omega_2(z)]} & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{5.129}$$

We define the transformation $\mathbf{U} \mapsto \mathbf{T}$ by setting

$$\mathbf{T}(z) = \begin{cases} \mathbf{U}(z)V_1^{-1}(z), & z \in \Sigma_{1,u}, \\ \mathbf{U}(z)V_1(z), & z \in \Sigma_{1,l}, \\ \mathbf{U}(z)V_2^{-1}(z), & z \in \Sigma_{2,u}, \\ \mathbf{U}(z)V_2(z), & z \in \Sigma_{2,l}, \\ \mathbf{U}(z), & \textit{elsewhere}. \end{cases} \tag{5.130}$$

Clearly, $\mathbf{T}(z)$ is a piecewise analytic function off of the contours Γ_1,Γ_2, Γ, $\Gamma_{1,u}$, $\Gamma_{1,l}$, $\Gamma_{2,u}$, and $\Gamma_{2,l}$. In fact, $\mathbf{T}(z)$ is the unique solution to the following Riemann-Hilbert problem:

Proposition 5.24. *The function $\mathbf{T}(z)$ is the unique solution to the following RHP:*

$$\mathbf{T}_+(z) = \mathbf{T}_-(z) \begin{cases} \mathbb{I} - E_{42}e^{-n\tau[\Omega_4(z)-\Omega_2(z)]}, & z \in \Gamma_{1,u} \cup \Gamma_{1,l}, \\ \mathbb{I} - E_{32}e^{-n\tau[\Omega_3(z)-\Omega_2(z)]}, & z \in \Gamma_{2,u} \cup \Gamma_{2,l}, \\ \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, & z \in [\beta,\infty), \\ \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, & z \in (-\infty,-\beta], \\ \mathbb{I} - E_{24}e^{-n\tau[\Omega_2(z)-\Omega_4(z)]}, & z \in \Gamma_1 \cap \{Im\ z > 0\}, \\ \mathbb{I} - E_{23}e^{-n\tau[\Omega_2(z)-\Omega_3(z)]}, & z \in \Gamma_2 \cap \{Im\ z < 0\}, \\ \mathbb{I} + E_{12}e^{-n\tau[\Omega_1(z)-\Omega_2(z)]}, & z \in \Gamma \setminus [-\alpha,\alpha], \\ \begin{pmatrix} e^{-n\tau[\Omega_{1,-}(z)-\Omega_{2,-}(z)]} & 1 & 0 & 0 \\ 0 & e^{n\tau[\Omega_{1,-}(z)-\Omega_{2,-}(z)]} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, & z \in [-\alpha,\alpha]. \end{cases} \tag{5.131}$$

Furthermore, $\mathbf{T}(z)$ has asymptotics

$$\mathbf{T}(z) = \left[\mathbb{I} + \mathcal{O}(z^{-1/3})\right] \begin{pmatrix} 1 & 0 \\ 0 & B^{-1}(z)A^{-1}(z) \end{pmatrix}, \qquad z \to \infty. \tag{5.132}$$

Proof. The proof of this proposition follows immediately from the definition of $\mathbf{T}(z)$. $\square$

5.5.2 The Fourth Transformation $\mathbf{T} \to \mathbf{S}$

We now open the lens around the segment $[-\alpha, \alpha]$. This is based on the following factorization of the jump matrix:

$$\begin{pmatrix} e^{-n\tau[\Omega_{2,+}(z)-\Omega_{1,+}(z)]} & 1 & 0 & 0 \\ 0 & e^{-n\tau[\Omega_{2,-}(z)-\Omega_{1,-}(z)]} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ e^{-n\tau[\Omega_{2,-}(z)-\Omega_{1,-}(z)]} & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ e^{-n\tau[\Omega_{2,+}(z)-\Omega_{1,+}(z)]} & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{5.133}$$

By the lensing propositions of Section §3, there exists a lens-shaped region around $[-\alpha, \alpha]$ such that the difference Re $[\Omega_2 - \Omega_1](z) > 0$. Define contours Γ_u, Γ_l as the boundaries of this lens-shaped region in the upper and lower half planes, respectively. Further, put Σ_u, Σ_l to be the regions enclosed by $[-\alpha, \alpha]$ and Γ_u, Γ_l, respectively. Define the invertible matrix $V_0(z)$ in the lensed region $\Sigma_u \cup \Sigma_l$ by

$$V_0(z) := \begin{pmatrix} 1 & 0 & 0 & 0 \\ e^{-n\tau[\Omega_2(z)-\Omega_1(z)]} & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{5.134}$$

We define the piecewise analytic function $\mathbf{S}(z)$ by

$$\mathbf{S}(z) := \begin{cases} \mathbf{T}(z)V_0^{-1}(z), & z \in \Sigma_l, \\ \mathbf{T}(z)V_0(z), & z \in \Sigma_u, \\ \mathbf{T}(z), & \textit{otherwise}. \end{cases} \tag{5.135}$$

In this case, $\mathbf{S}(z)$ is the unique solution to the following RHP:

Proposition 5.25. *The function* $\mathbf{S}(z)$ *is the unique solution to the following Riemann-Hilbert problem:* $\mathbf{S}(z)$ *is piecewise analytic off the contours* Γ_1*,*Γ_2*,* Γ*,* $\Gamma_{1,u}$*,* $\Gamma_{1,l}$*,* $\Gamma_{2,u}$*,* $\Gamma_{2,l}$*,* Γ_l*, and* Γ_u*, satisfying the jump condition*

$$\mathbf{S}_+(z) = \mathbf{S}_-(z) \begin{cases} \mathbb{I} - E_{42} e^{-n\tau[\Omega_4(z)-\Omega_2(z)]}, & z \in \Gamma_{1,u} \cup \Gamma_{1,l}, \\ \mathbb{I} - E_{32} e^{-n\tau[\Omega_3(z)-\Omega_2(z)]}, & z \in \Gamma_{2,u} \cup \Gamma_{2,l}, \\ \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, & z \in [\beta, \infty), \\ \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, & z \in (-\infty, -\beta], \\ \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} & z \in [-\alpha, \alpha], \\ \mathbb{I} - E_{24} e^{-n\tau[\Omega_2(z)-\Omega_4(z)]}, & z \in \Gamma_1 \cap \{Im\ z > 0\}, \\ \mathbb{I} - E_{23} e^{-n\tau[\Omega_2(z)-\Omega_3(z)]}, & z \in \Gamma_2 \cap \{Im\ z < 0\}, \\ \mathbb{I} + E_{12} e^{-n\tau[\Omega_1(z)-\Omega_2(z)]}, & z \in \Gamma \setminus [-\alpha, \alpha], \\ \mathbb{I} + E_{21} e^{-n\tau[\Omega_2(z)-\Omega_1(z)]}, & z \in \Gamma_u \cup \Gamma_l. \end{cases} \tag{5.136}$$

Furthermore, $\mathbf{S}(z)$ *has asymptotics*

$$\mathbf{S}(z) = \left[\mathbb{I} + \mathcal{O}\left(z^{-1}\right)\right] \begin{pmatrix} 1 & 0 \\ 0 & B^{-1}(z)A^{-1}(z) \end{pmatrix}, \qquad z \to \infty. \tag{5.137}$$

Proof. Again, the proof of this proposition follows immediately from the definition of $\mathbf{S}(z)$. The fact that the stronger condition

$$\mathbf{S}(z) = \left[\mathbb{I} + \mathcal{O}\left(z^{-1}\right)\right] \begin{pmatrix} 1 & 0 \\ 0 & B^{-1}(z)A^{-1}(z) \end{pmatrix}, \qquad z \to \infty, \tag{5.138}$$

holds is due to the fact that the jumps of $A(z)B(z)$ match the jumps of $\mathbf{S}(z)$ at infinity, as per Remark (5.20). Thus, $\mathcal{O}(z^{-1/3})$ can be replaced with $\mathcal{O}(z^{-1})$ in the asymptotics of $\mathbf{S}(z)$. □

5.6 Construction of Parametrices and the Transformation $\mathbf{S} \to \mathbf{R}$.

All of the jumps of $\mathbf{S}$ are either constant, or exponentially small. Our next task is to try and eliminate these constant jumps. We will accomplish this task by searching for an approximate solution, called the *global parametrix* to the Riemann-Hilbert problem at hand; this approximate solution will match the constant jumps of $\mathbf{S}$ exactly, but the difference of jumps near the branch points will be 'bad', requiring us to find local approximations to the RHP (local parametrices) there. Aside from the proofs of the lensing inequalities, this

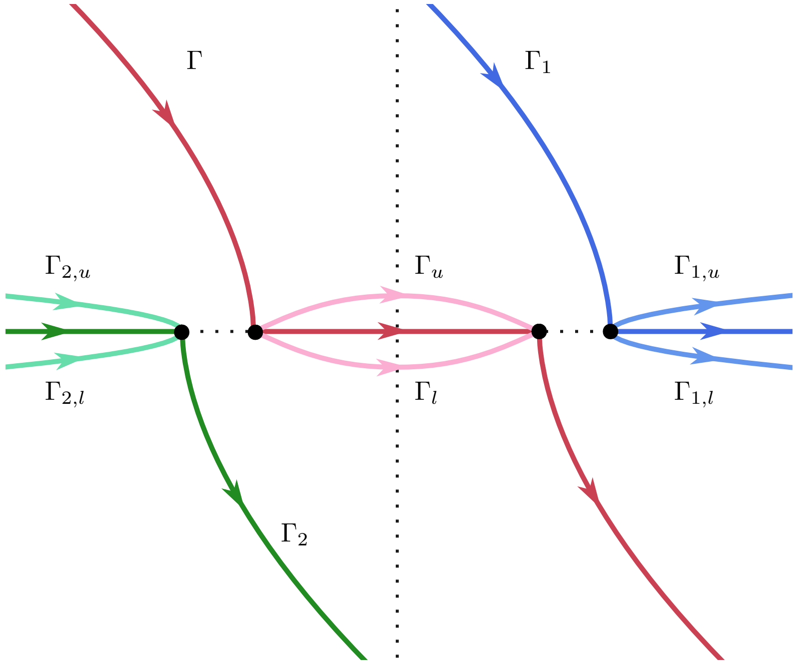


Figure 5.12. The opened lenses.

is really the first place we will see a difference between the multicritical, critical, and generic (non-critical) cases.

5.6.1 Global Parametrix.

If we ignore the exponentially small jumps of $\mathbf{S}(z)$, we obtain the following model RHP for a 4×4 matrix-valued function $M(z)$:

$$\begin{cases} M \text{ is analytic in } \mathbb{C} \setminus \mathbb{R}, & \\ M_+(z) = M_-(z) \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, & z \in (-\infty, -\beta], \\ M_+(z) = M_-(z) \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, & z \in [-\alpha, \alpha], \\ M_+(z) = M_-(z) \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, & z \in [\beta, \infty), \\ M(z) = \left[\mathbb{I} + \mathcal{O}(z^{-1})\right] \begin{pmatrix} 1 & 0 \\ 0 & B^{-1}(z)A^{-1}(z) \end{pmatrix}, & z \to \infty. \end{cases} \tag{5.139}$$

In general, solutions to (5.139) will not be unique. Uniqueness can be guaranteed by imposing additionally that

$$\begin{cases} M(z) = \mathcal{O}((z \mp \alpha)^{-1/4}), & z \to \pm\alpha, \\ M(z) = \mathcal{O}((z \mp \beta)^{-1/4}), & z \to \pm\beta. \end{cases} \tag{5.140}$$

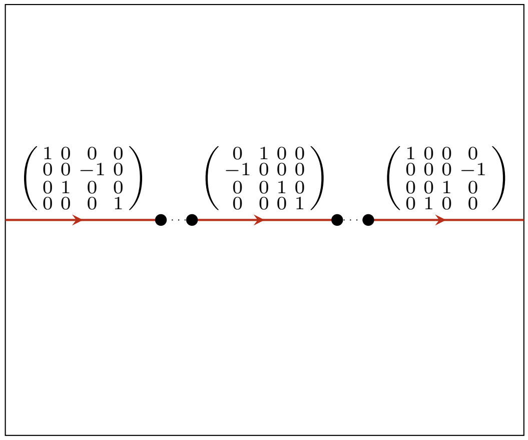


Figure 5.13. The jumps of the global parametrix $M(z)$ on $\mathbb{R}$. Note that, when $a, b \to 1$, branch points merge, and the cuts cover the whole real line.

Then, from the usual Liouville argument, it is clear that if a solution to (5.139) (with the constraint (5.140) imposed) exists, it is unique. We show that a solution exists by direct construction.

Proposition 5.26. *The solution to the global parametrix is given by*

$$M_{jk}(z) = \begin{cases} \psi_j(u_k(z)), & Im\ z > 0, \\ S_{jk}\psi_j(u_k(z)), & Im\ z < 0, \end{cases} \tag{5.141}$$

where $u_k(z)$ is the restriction of the uniformizing coordinate to the k^{th} sheet, and the functions $\psi_j(u)$ are given by

$$\begin{cases} \psi_1(u) = \frac{u^2}{\sqrt{(u^2-b^2)(u^2-a^2)}}, & \psi_2(u) = \left(-\frac{3\tau}{t}\right)^{1/6} \frac{2a^2b^2-3(a^2+b^2)u^2}{18u\sqrt{(u^2-b^2)(u^2-a^2)}} \\ \psi_3(u) = \frac{ab}{3\sqrt{(u^2-b^2)(u^2-a^2)}}, & \psi_4(u) = \left(-\frac{t}{27\tau}\right)^{1/6} \frac{u}{\sqrt{(u^2-b^2)(u^2-a^2)}}, \end{cases} \tag{5.142}$$

and $S = diag\ (1, -1, 1, 1)$.

Proof. Let us assume we are in either the generic (noncritical) or critical case, we have that $\alpha = \alpha(a, b) < \beta(a, b) = \beta$. Following [39], we will solve find it convenient to solve this problem in the uniformizing coordinate $z = z(u)$:

$$z(u) = A(a, b)\left(u + \frac{a^2 + b^2}{u} - \frac{a^2b^2}{3u^3}\right).$$

Consider the general row vector

$$\vec{\psi}(z) = \begin{cases} \left[f^u(u_1(z)), g^u(u_2(z)), h^u(u_3(z)), k^u(u_4(z))\right], & \text{Im } z > 0, \\ \left[f^l(u_1(z)), g^l(u_2(z)), h^l(u_3(z)), k^l(u_4(z))\right], & \text{Im } z < 0. \end{cases} \tag{5.143}$$

where f^u, g^u, h^u, k^u (respectively, f^l, g^l, h^l, k^l) are analytic functions away from the real axis, to be determined. Suppose $\vec{\psi}(z)$ satisfies the jumps of the global parametrix. Now, consider the analytic continuation of $\vec{\psi}$ through $(-\infty, -\beta]$: upon continuation, $u_2(z)$ and $u_3(z)$ are interchanged. On the other hand, the analytic continuation is determined by the jump condition of the global parametrix. This leads to the constraint

$$[f^u(u_1(z)), g^u(u_3(z)), h^u(u_2(z)), k^u(u_4(z))] = \left[f^l(u_1(z)), h^l(u_3(z)), -g^l(u_2(z)), k^l(u_4(z))\right], \tag{5.144}$$

for $z \in (-\infty, -\beta]$. In particular, this implies the equalities $f^u = f^l$, $k^u = k^l$, $g^u = h^l$, and $h^u = -g^l$. Similar analysis on the other cuts leads in particular to the further compatibility conditions

$$g^u = k^l, \qquad k^u = -g^l, \qquad f^u = -g^l, \qquad g^u = f^l. \tag{5.145}$$

Thus, the global parametrix depends on only on the unknown function: $f^u =: \psi$. Therefore,

$$\vec{\psi}(z) = \begin{cases} [\psi(u_1(z)), \psi(u_2(z)), \psi(u_3(z)), \psi(u_4(z))], & \text{Im } z > 0, \\ [\psi(u_1(z)), -\psi(u_2(z)), \psi(u_3(z)), \psi(u_4(z))], & \text{Im } z < 0. \end{cases} \tag{5.146}$$

We have thus shown that, in the upper half plane,

$$\vec{\psi}(z) = [\psi(u_1(z)), \psi(u_2(z)), \psi(u_3(z)), \psi(u_4(z))], \tag{5.147}$$

where $\vec{\psi}(z)$ defines a (possibly multivalued) analytic function in the uniformizing plane, which is defined by its components $\psi(u_j(z))$ in each of the images of the sheets $j = 1, 2, 3, 4$. Suppose $\vec{\psi}(z)$ satisfies the jump conditions of (5.139). Let us compute the monodromy of ψ around each of the branch points. In the uniformizing coordinate, the branch points are mapped on to $\pm a$ and $\pm b$ (Note: provided we are away from the multi-critical point, we have that $0 < b < 1 \leq a \leq b^{-1}$). By direct computation using the Riemann-Hilbert problem, we find that $\psi(u)$ has square-root singularities at each of the branch points $\pm a, \pm b$, and possibly a pole singularity at $u = 0$. The form of $\psi(u)$ is then fixed to be

$$\psi(u) = \frac{p_3(u)}{u\sqrt{(u^2 - a^2)(u^2 - b^2)}}, \tag{5.148}$$

where $p_3(u)$ is some polynomial of degree ≤ 3, and the positive branch cut of the square root is chosen, with branch cuts on the intervals $[-a, -b] \cup [b, a]$. The form of $\psi(u)$ is fixed by the following constraints. For any choice of $p_3(u)$, $\psi_k(u)$ is analytic in $\mathbb{C} \setminus \gamma$, where γ is the image of the cuts in the uniformizing plane, and satisfies the following conditions:

$$
\begin{cases}
\psi_+(u) = -\psi_-(u), & u \in \gamma, \\
\psi(u) = \mathcal{O}(1), & u \to \infty, \\
\psi(u) = \mathcal{O}(u^{-1}), & u \to 0, \\
\psi(u) = \mathcal{O}((u \mp a)^{-1/2}), & u \to \pm a, \\
\psi(u) = \mathcal{O}((u \mp b)^{-1/2}), & u \to \pm b.
\end{cases}
$$

These conditions guarantee that the entries of $M(z)$ are $\mathcal{O}(1)$ as $z \to \infty$ on the first sheet, and $\mathcal{O}(z^{1/3})$ on sheets 2–4. It also guarantees that $M(z)$ has the correct jumps. For a given row, we can use the normalization condition of the Riemann-Hilbert problem to determine the coefficients of $p_3(u)$.

Setting $p_3(u) = c_0 + c_1 u + c_2 u^2 + c_3 u^3$, then we have the following large z expansions of the functions $\psi(u_j(z))$:

$$
\psi(u_1(z)) = c_3 + \frac{c_2 A}{z} + \frac{((a^2+b^2)c_3 + 2c_1)A^2}{2z^2} + \frac{(3(a^2+b^2)c_2 + 2c_0)A^3}{2z^3} + \mathcal{O}(z^{-4}),
$$

$$
\psi(u_2(z)) = \begin{cases}
-\frac{3^{\frac{1}{3}} c_0}{(ab)^{\frac{5}{3}}} \frac{\omega^2 z^{\frac{1}{3}}}{A^{\frac{1}{3}}} + \frac{c_1}{ab} - \frac{2a^2b^2c_2 + 3c_0(a^2+b^2)}{3^{\frac{1}{3}}(ab)^{\frac{7}{3}}} \frac{\omega A^{\frac{1}{3}}}{z^{\frac{1}{3}}} + \frac{3^{\frac{1}{3}}(2a^2b^2c_3 - (a^2+b^2)c_1)}{6a^{\frac{5}{3}}b^{\frac{5}{3}}} \frac{\omega^2 A^{\frac{2}{3}}}{z^{\frac{2}{3}}} + \mathcal{O}(z^{-1}), & \operatorname{Im} z > 0, \\
-\frac{3^{\frac{1}{3}} c_0}{(ab)^{\frac{5}{3}}} \frac{\omega z^{\frac{1}{3}}}{A^{\frac{1}{3}}} + \frac{c_1}{ab} - \frac{2a^2b^2c_2 + 3c_0(a^2+b^2)}{3^{\frac{1}{3}}(ab)^{\frac{7}{3}}} \frac{\omega^2 A^{\frac{1}{3}}}{z^{\frac{1}{3}}} + \frac{3^{\frac{1}{3}}(2a^2b^2c_3 - (a^2+b^2)c_1)}{6a^{\frac{5}{3}}b^{\frac{5}{3}}} \frac{\omega A^{\frac{2}{3}}}{z^{\frac{2}{3}}} + \mathcal{O}(z^{-1}), & \operatorname{Im} z < 0,
\end{cases}
$$

$$
\psi(u_3(z)) = -\frac{3^{\frac{1}{3}} c_0}{(ab)^{\frac{5}{3}}} \frac{z^{\frac{1}{3}}}{A^{\frac{1}{3}}} + \frac{c_1}{ab} - \frac{2a^2b^2c_2 + 3c_0(a^2+b^2)}{3^{\frac{1}{3}}(ab)^{\frac{7}{3}}} \frac{A^{\frac{1}{3}}}{z^{\frac{1}{3}}} + \frac{3^{\frac{1}{3}}(2a^2b^2c_3 - (a^2+b^2)c_1)}{6a^{\frac{5}{3}}b^{\frac{5}{3}}} \frac{A^{\frac{2}{3}}}{z^{\frac{2}{3}}} + \mathcal{O}(z^{-1})
$$

$$
\psi(u_4(z)) = \begin{cases}
-\frac{3^{\frac{1}{3}} c_0}{(ab)^{\frac{5}{3}}} \frac{\omega z^{\frac{1}{3}}}{A^{\frac{1}{3}}} + \frac{c_1}{ab} - \frac{2a^2b^2c_2 + 3c_0(a^2+b^2)}{3^{\frac{1}{3}}(ab)^{\frac{7}{3}}} \frac{\omega^2 A^{\frac{1}{3}}}{z^{\frac{1}{3}}} + \frac{3^{\frac{1}{3}}(2a^2b^2c_3 - (a^2+b^2)c_1)}{6a^{\frac{5}{3}}b^{\frac{5}{3}}} \frac{\omega A^{\frac{2}{3}}}{z^{\frac{2}{3}}} + \mathcal{O}(z^{-1}), & \operatorname{Im} z > 0, \\
-\frac{3^{\frac{1}{3}} c_0}{(ab)^{\frac{5}{3}}} \frac{\omega^2 z^{\frac{1}{3}}}{A^{\frac{1}{3}}} + \frac{c_1}{ab} - \frac{2a^2b^2c_2 + 3c_0(a^2+b^2)}{3^{\frac{1}{3}}(ab)^{\frac{7}{3}}} \frac{\omega A^{\frac{1}{3}}}{z^{\frac{1}{3}}} + \frac{3^{\frac{1}{3}}(2a^2b^2c_3 - (a^2+b^2)c_1)}{6a^{\frac{5}{3}}b^{\frac{5}{3}}} \frac{\omega^2 A^{\frac{2}{3}}}{z^{\frac{2}{3}}} + \mathcal{O}(z^{-1}), & \operatorname{Im} z < 0,
\end{cases}
$$

Since $\left(\begin{smallmatrix} 1 & 0 \\ 0 & A(z)B(z) \end{smallmatrix}\right) M(z)$ has no jumps near infinity, it admits the expansion

$$
\left(\begin{smallmatrix} 1 & 0 \\ 0 & A(z)B(z) \end{smallmatrix}\right) M(z) = \mathbb{I} + \mathcal{O}(z^{-1}), z \to \infty. \tag{5.149}
$$

This fact allows us to determine the constants $\{c_i\}$ for each row; the rest of the proof follows from direct computation. A similar analysis in the lower half plane may be performed, and the same result for $\psi(z)$ is obtained. The expression in the lower half plane can also be obtained by analytic continuation of the solution in the upper half plane, in accordance with the jump conditions of the global parametrix. □

Remark 5.27. So far, we have ignored the multicritical case, when $a = b = 1$. In this case, the branch points in both the physical and uniformizing planes merge into a pair of branch points, say, $\pm\alpha$. If we follow the same procedure as before, we find that any row vector $\psi(u) = [\psi(u_1(z)), ..., \psi(u_4(z))]$ has no monodromy, and thus defines a meromorphic function on the spectral curve. In fact, if we complete the calculations, we find that the solution to the global parametrix in the multicritical case is just the degeneration of the general global parametrix as $a, b \to 1$. Direct inspection of the general global parametrix shows us that $(a, b) = (1, 1)$ is the only point in R where the rows of the global parametrix are single-valued in the uniformizing plane.

Remark 5.28. The final step in the steepest descent analysis is to calculate the local parametrices near the branch points, where the global parametrix is a bad approximation to $\mathbf{S}(z)$. In the next section, many of the jumps will involve differences of the $\Omega_j(z)$'s. For this reason, we introduce the notation

$$\delta\Omega_{ij}(z) := \Omega_i(z) - \Omega_j(z), \tag{5.150}$$

for $i, j \in \{1, ..., 4\}$.

We have now found an approximate solution to the Riemann-Hilbert problem for $\mathbf{S}(z)$. Indeed, if we consider the matrix

$$\mathbf{R}_{out}(z) := \mathbf{S}(z)M^{-1}(z), \tag{5.151}$$

we see that $\mathbf{R}_{out}(z) \to \mathbb{I}$ as $z \to \infty$. Furthermore, the jumps of $\mathbf{R}_{out}(z)$ are all exponentially small (in n), with one exception: near the branch points $\pm\alpha$, $\pm\beta$, the jumps are not close to the identity, as $n \to \infty$. Therefore, we must try to find a better approximation of $\mathbf{S}(z)$ near the branch points.

Remark 5.29. Here is the first place where we will see explicitly that the generic, critical, and multicritical cases differ. We are really only interested in the multicritical phase transition, as the other local parametrices are comparatively standard. We address the generic and critical cases here. Our description of these cases is brief, as they are closer to what appears in the established literature.

1. *The Generic Case: Airy Parametrices.*

   For (τ, t) off the critical curve, the situation is generic, and the behavior of the $\delta\Omega_{ij}(z)$'s near the branch points is

   $$\delta\Omega_{ij}(z) \sim (z \pm \alpha)^{3/2} \text{ (resp. } \sim (z \pm \beta)^{3/2} \text{).} \tag{5.152}$$

   (cf. Proposition (5.6)). By now, it is well-established in the literature that this behavior leads to Airy-type parametrices at each of the branch points. Since this computation is familiar, and since our interest lies mainly in the multicritical case, we omit the explicit calculation of the parametrices here.

2. *The Low-Temperature critical case: Painlevé I.*

For (τ, t) on the low-temperature critical curve (i.e., the curve defined by the equation $t = -\frac{1}{12} + \frac{2}{9}\tau^2$, $0 < \tau < \frac{1}{4}$), the behavior of the $\delta\Omega_{ij}(z)$'s near each of the four branch points $\pm\alpha$, $\pm\beta$, is

$$\delta\Omega_{ij}(z) \sim (z \pm \alpha)^{5/2} \text{ (resp. } \sim (z \pm \beta)^{5/2} \text{).} \tag{5.153}$$

(cf. Proposition (5.9)). It is also widely acknowledged in the literature (cf. [38], for example) that such local behavior leads to a Painlevé I-type Riemann-Hilbert problem, which has been the subject of intensive study [66, 67]. We expect that the same analysis applies to the situation at hand, and leads to a description of the partition function in terms of the same solution to Painlevé I that appears in the description of the critical 1-matrix model [38, 47, 48]. Again, since we are mainly interested in the behavior of the partition function at the multicritical point, we omit the explicit calculation of the parametrices here.

3. *The High-Temperature critical case: Painlevé I and Airy.*

For (τ, t) on the high-temperature critical curve (i.e., the curve defined by the equation $t = -\frac{2}{9}\sqrt{\tau}(\sqrt{\tau} - 1)^2(\sqrt{\tau} + 2)$, $\frac{1}{4} < \tau < 1$), the behavior of the $\delta\Omega_{ij}(z)$'s the branch points is

$$\delta\Omega_{ij}(z) \sim (z \pm \alpha)^{5/2}, \qquad \delta\Omega_{ij}(z) \sim (z \pm \beta)^{3/2}, \tag{5.154}$$

as $z \to \pm\alpha$, $z \to \pm\beta$, respectively (cf. Proposition (5.12)). This indicates that the appropriate local parametrices to use near $z = \pm\beta$ are Airy-type parametrices, by our previous commentary for the generic situation. Similarly, we expect that near the branch points $z = \pm\alpha$, one should use Painlevé I parametrices. Thus, the local parametrix structure on the high-temperature critical curve is a mix of the low-temperature critical and generic situations. On the support of the main cut, the density is still critical, and Painlevé I parametrices are needed; however, the other cuts cease to be "critical", as they were in the low-temperature regime.

We again omit the explicit calculation of the parametrices here, as our main interests lie elsewhere.

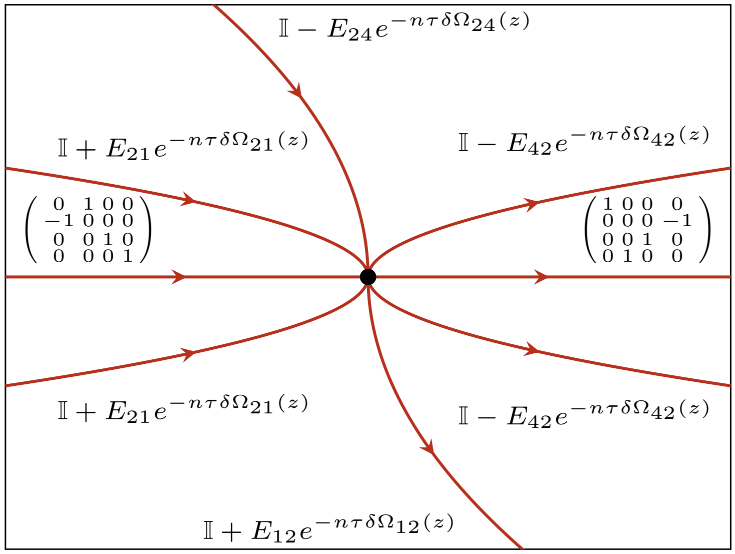


Figure 5.14. The jumps of the Riemann-Hilbert problem for $\mathbf{S}(z)$ in a small disc at $z = \alpha$. We must match these jumps exactly. Note that only the rows (columns) $1, 2$, and 4 participate nontrivially.

5.6.2 Local Parametrices: The Multicritical Case.

The multicritical case is distinguished from the other critical cases in the sense that:

- The local behavior of the functions $\delta\Omega_{jk}(z) = \textit{const.}\ (z-\alpha)^{7/3}[1 + \mathcal{O}((z-\alpha)^{8/3}))]$, instead of $\sim (z-\alpha)^{5/2}$,
- The critical phenomenon is characterized by the merging of two branch points, instead of a merging of a zero with a branch point.

The moving branch points make it difficult to analyze the behavior in a neighborhood of criticality. This type of problem has been addressed before in the literature [14, 32, 36]. The main idea is to define a *modified spectral curve*. This curve has the same asymptotic behavior as the actual spectral curve, as we approach multicriticality, but is characterized by the condition that the branch points are fixed at $\pm\alpha$ over the varying parameters. This gives locally the wrong behavior; however, since we are going to construct local parametrices anyways, this is admissible. We begin this subsection with a preliminary analysis of this modified spectral curve, and then proceed to construct local parametrices.

5.6.2.1 The Modified Spectral Curve. We actually seek to construct a *family* of modified spectral curves, which have the following properties:

1. At (multi)criticality, the modified spectral curve coincides with the multicritical spectral curve,
2. As $z \to \infty$ on any of the sheets, the modified spectral curve has the *exact* asymptotics of the true spectral curve, up to a certain order, i.e., equations (5.29), (5.30) hold,
3. When τ, t are away from their multicritical values, the branching structure of $\Omega_j(z)$ near the branch points is the same as that of the multicritical spectral curve (i.e., the branch points do not split upon deformation); furthermore, the terms of order $(z \pm \alpha)^{4/3}$ appearing in the expansion of the function $\Omega_j(z)$ at the critical points $z = \mp\alpha$ vanish, as they correspond to an irrelevant translation in the scaling limit.

We will only need to calculate the modified spectral curve in a small neighborhood of the multicritical curve, i.e., to first order in the deformation parameters. We move (τ, t) away from their multicritical values $(\tau_c, t_c) := (\frac{1}{4}, -\frac{5}{72})$ by

$$\begin{pmatrix} \tau \\ t \end{pmatrix} = \begin{pmatrix} \tau_c \\ t_c \end{pmatrix} + \delta_\perp \begin{pmatrix} -1 \\ 9 \end{pmatrix} + \delta_{||} \begin{pmatrix} 9 \\ 1 \end{pmatrix}. \tag{5.155}$$

Note that the vector $(-1, 9)^T$ is proportional to the normal vector of the critical curve at (τ_c, t_c), and the vector $(9, 1)^T$ is proportional to the tangent vector to the critical curve at this point. Thus, $\delta_{||}$ describes the displacement from multicriticality in the tangential direction, and $\delta_\perp$ describes the displacement from multicriticality in the normal direction.

As it turns out, the above conditions uniquely fix for us a choice of modified curve:

Lemma 5.30. *There exist functions*

$$A(\delta_{||}, \delta_\perp), B(\delta_{||}, \delta_\perp), H(\delta_{||}, \delta_\perp), K(\delta_{||}, \delta_\perp), M(\delta_{||}, \delta_\perp), N(\delta_{||}, \delta_\perp), \tag{5.156}$$

(real) analytic in a neighborhood of $(0, 0)$ *in the* $(\delta_{||}, \delta_\perp)$ *plane, such that, if we define*

$$z(u; \delta_{||}, \delta_\perp) := A(\delta_{||}, \delta_\perp) \left(u + \frac{2}{u} - \frac{1}{3u^3} \right), \tag{5.157}$$

$$\begin{aligned} Y(u; \delta_{||}, \delta_\perp) := B(\delta_{||}, \delta_\perp) &\left(\frac{1}{u} + H(\delta_{||}, \delta_\perp) u + K(\delta_{||}, \delta_\perp) u^3 \right) \\ &+ M(\delta_{||}, \delta_\perp) \left(\frac{1}{u-1} + \frac{1}{u+1} \right) + N(\delta_{||}, \delta_\perp) \left(\frac{1}{(u-1)^2} - \frac{1}{(u+1)^2} \right), \end{aligned} \tag{5.158}$$

then, the following Conditions hold in a neighborhood of $(0,0)$*, whenever*

$$\tau = \tau_c + 9\delta_{||} - \delta_{\perp}. \tag{5.159}$$

$$t = t_c + \delta_{||} + 9\delta_{\perp}. \tag{5.160}$$

C1. When $\delta_{||} = \delta_{\perp} = 0$*, the functions* $z(u;0,0)$*,* $Y(u;0,0)$ *coincide with the parameterization of the true multicritical spectral curve,*

C2. For $z = z(u;\delta_{||},\delta_{\perp})$*,* $Y = Y(u;\delta_{||},\delta_{\perp})$*, the expansions*

$$tz(u)^3 + z(u) - \frac{1}{z(u)} - \tau Y(u) = \mathcal{O}(u^{-2}), \qquad u \to \infty, \tag{5.161}$$

$$tY(u)^3 + Y(u) - \frac{1}{Y(u)} - \tau z(u) = \mathcal{O}(u^{2}), \qquad u \to 0, \tag{5.162}$$

hold, uniformly in a neighborhood of $(\delta_{||},\delta_{\perp}) = (0,0)$*,*

C3. Define $\Omega(u;\delta_{||},\delta_{\perp}) := \int Y(u)z'(u)du$*, uniformization coordinates* $u_j(z)$*,* $j = 1,...,4$ *as before, and* $\Omega_j(z) := \Omega(u_j(z);\delta_{||},\delta_{\perp})$*. Then, putting* $\alpha = \alpha(\delta_{||},\delta_{\perp}) := z(1;\delta_{||},\delta_{\perp})$*, the expansion of* $\Omega_j(z)$*,* $j = 1,2,4$ *about* $z = \alpha$ *has vanishing coefficient of the term* $(z-\alpha)^{4/3}$*. Similarly, the expansion of* $\Omega_j(z)$*,* $j = 1,2,3$ *about* $z = -\alpha$ *has vanishing coefficient of the term* $(z+\alpha)^{4/3}$*.*

Furthermore, to first order in $(\delta_{||},\delta_{\perp})$*,*

$$A(\delta_{||},\delta_{\perp}) = \sqrt{\frac{6}{5}}\left(1 + \frac{36}{5}\delta_{||} + \frac{324}{5}\delta_{\perp} + \mathcal{O}(|\delta|^2)\right) \tag{5.163}$$

$$B(\delta_{||},\delta_{\perp}) = \sqrt{\frac{6}{5}}\left(1 + \frac{81}{5}\delta_{||} + \frac{90847}{1440}\delta_{\perp} + \mathcal{O}(|\delta|^2)\right), \tag{5.164}$$

$$H(\delta_{||},\delta_{\perp}) = 2 - 72\delta_{||} + \frac{493}{36}\delta_{\perp} + \mathcal{O}(|\delta|^2), \tag{5.165}$$

$$K(\delta_{||},\delta_{\perp}) = -\frac{1}{3} + 24\delta_{||} + \frac{83}{108}\delta_{\perp} + \mathcal{O}(|\delta|^2), \tag{5.166}$$

$$M(\delta_{||},\delta_{\perp}) = \sqrt{\frac{6}{5}}\left(\frac{656}{3}\delta_{\perp} + \mathcal{O}(|\delta|^2)\right), \tag{5.167}$$

$$N(\delta_{||},\delta_{\perp}) = \sqrt{\frac{6}{5}}\left(\frac{328}{3}\delta_{\perp} + \mathcal{O}(|\delta|^2)\right). \tag{5.168}$$

(Here, $\mathcal{O}(|\delta|^2)$ *denotes terms of order* 2 *and higher in the Taylor expansion of the above functions in* $\delta_{||},\delta_{\perp}$*).*

Proof. Assume the form of $z(u;\delta_{||},\delta_{\perp}), Y(u;\delta_{||},\delta_{\perp})$ in terms of the functions $A(\delta_{||},\delta_{\perp})$ through $N(\delta_{||},\delta_{\perp})$. Additionally, suppose τ, t are as in Equations (5.159), (5.160). The requirement that Condition C2 (that the modified spectral curve has the *exact* asymptotics of the true spectral curve for $\tau = \tau(\delta_{||},\delta_{\perp})$, $t = t(\delta_{||},\delta_{\perp})$)

holds implies the following equations must hold:

$$
\begin{aligned}
0 &= tA^3 - \tau BK, \\
0 &= A + 6tA^3 - \tau BH, \\
0 &= 11tA^3 + 2A - \frac{1}{A} - \tau(B + 2M), \\
0 &= tB^3 + \frac{\tau}{3}A, \\
0 &= B + 3tB^3H - 3tB^2(4N - 2M) - 2\tau A, \\
0 &= BH - 2M + 4N + t[B\{2B(BK - 2M + 8N) + (BH - 2M + 4N)^2\} \\
&\quad + 2(BH - 2M + 4N)^2B + (BK - 2M + 8N)B^2] - \frac{1}{B} - \tau A.
\end{aligned}
$$

One of the above equations is redundant; the remaining 5 equations can be solved uniquely for for A, H, K, M, and N as rational functions of t, τ, and the function $B(\delta_{||}, \delta_\perp)$:

$$
\begin{aligned}
A &= -\frac{3tB^3}{\tau}, \qquad & H &= -\frac{3tB^2}{\tau^4}(\tau^2 + 54B^6t^3), \\
K &= -\frac{27t^4B^8}{\tau^4}, \qquad & M &= -\frac{891B^{12}t^5 + 18B^6t^2\tau^2 + 3B^4t\tau^4 - \tau^4}{6\tau^4tB^3}, \\
N &= -\frac{405B^{12}t^5 + 9B^6t^2\tau^2 + 9B^4t\tau^4 + B^2\tau^4 - \tau^4}{12\tau^4B^3t}.
\end{aligned}
$$

Now, Condition C3 requires that the expansion of $\Omega_j(z)$, $j = 1, 2, 4$ about $z = \alpha := z(1)$ has no term of the form $(z - \alpha)^{4/3}$, and similarly the expansion of $\Omega_j(z)$, $j = 1, 2, 3$ about $z = -\alpha := z(-1)$ should have no term of the form $(z + \alpha)^{4/3}$. Because of the symmetry of the coefficients of the poles in $Y(u)$, we only need to verify the first condition holds; the condition at $z = -\alpha$ will then hold automatically. The requirement that the coefficient of the term $(z - \alpha)^{4/3}$ in the expansion of the $\Omega_j(z)$'s implies the equation

$$
-\frac{(603B^4t + 7B^2 + 5)\tau^4 + 1575t^2\tau^2B^6 + 132111t^5B^{12}}{1536(-tB^3/\tau)^{1/3}\tau^4tB^3} = 0,
$$

which in particular will hold if the numerator vanishes identically. The theorem we are trying to prove has thus been reduced to proving the existence of an implicit function $B(\delta_{||}, \delta_\perp)$ of the equation

$$
\Phi(B, \delta_{||}, \delta_\perp) := (603B^4t + 7B^2 + 5)\tau^4 + 1575t^2\tau^2B^6 + 132111t^5B^{12} = 0, \tag{5.169}
$$

where $\tau = \tau_c + 9\delta_{||} - \delta_\perp$, $t = t_c + \delta_{||} + 9\delta_\perp$. When $\delta_{||} = \delta_\perp = 0$, $B = \sqrt{\frac{6}{5}}$ is a solution to the above equation. The implicit function theorem guarantees the existence of a function $B(\delta_{||}, \delta_\perp)$ in a neighborhood of $\delta_{||} = \delta_\perp = 0$ satisfying Equation (5.169), provided that

$$\left.\frac{\partial \Phi}{\partial B}\right|_{(0,0,\sqrt{6/5})} \neq 0.$$

Indeed, one can calculate that $\frac{\partial \Phi}{\partial B}|_{(0,0,\sqrt{6/5})} = -3\sqrt{\frac{6}{5}}$, and so an implicit function exists. The expansion of $B(\delta_{||}, \delta_\perp)$ to first order is

$$B(\delta_{||}, \delta_\perp) = \sqrt{\frac{6}{5}}\left(1 + \frac{81}{5}\delta_{||} + \frac{90847}{1440}\delta_\perp + \mathcal{O}(|\delta|^2)\right);$$

since A, H, K, M, and N all depend rationally on B, $\delta_{||}$, and $\delta_\perp$, we can obtain Taylor expansions of these functions as functions of $\delta_{||}$, $\delta_\perp$ as well. One may readily check that this procedure yields Equations (5.163)–(5.168). When $\delta_{||} = \delta_\perp = 0$, it is immediately apparent that $z(u; 0, 0)$, $Y(u; 0, 0)$ coincide with the parameterization of the true multicritical spectral curve, so that Condition C1 is satisfied. This completes the proof of the Lemma. □

Remark 5.31. We call the family of Riemann surfaces parameterized by $\left(z(u; \delta_{||}, \delta_\perp), Y(u; \delta_{||}, \delta_\perp)\right)$ the *modified spectral curve*, in light of our previous commentary. The previous lemma shows that this curve exists, provided that we are close enough to the multicritical point, i.e. for $\delta_{||}, \delta_\perp$ sufficiently small. Note that the branch points in the z-coordinate indeed do not split upon deformation: the branch points of the modified spectral curve above are at $z = \alpha = \pm z(1; \delta_{||}, \delta_\perp)$. When $\delta_{||} = \delta_\perp = 0$, we recover the multicritical spectral curve. We also remark that, the poles of Y only appear upon deformation in the normal direction. This is apparent in the above Lemma, as, to first order in $\delta_{||}, \delta_\perp$, the coefficients of the poles M and N depend only on $\delta_\perp$.

Crucially, the modified spectral curve has been constructed so that, outside a sufficiently small neighborhood of $z = \pm\alpha$, the lensing inequalities we proved earlier still hold. This seemingly introduces a new issue: the inequalities necessary to open lenses do not hold nearby the branch points. However, since we are going to introduce local lenses around the branch points anyways, we actually do not need these inequalities to hold here. This is the objective of introducing the modified curve: we now have a workable form for the local parametrices discs about the branch points (to be constructed later in this section), while retaining the required lensing inequalities outside of these discs. We now formally state a Lemma that guarantees the lensing inequalities still hold outside of local discs about $z = \pm\alpha$.

Lemma 5.32. *For any $\epsilon > 0$ sufficiently small, there exist discs $D_+ = \{z : |z-\alpha| < \epsilon\}$, $D_- = \{z : |z+\alpha| < \epsilon\}$, such that, setting $\Delta := D_+ \cup D_-$, and $\Omega_j(z) = \phi_j(z) + i\psi_j(z)$, $j = 1, 2, 3, 4$, the inequalities*

1. $\phi_4(z) - \phi_2(z) > 0$ *for* $z \in (\Gamma_{1,u} \cup \Gamma_{1,l}) \setminus \Delta$,

2. $\phi_3(z) - \phi_2(z) > 0$ *for* $z \in (\Gamma_{2,u} \cup \Gamma_{2,l}) \setminus \Delta$,

3. $\phi_2(z) - \phi_1(z) > 0$ *for* $z \in (\Gamma_u \cup \Gamma_l) \setminus \Delta$,

hold. Furthermore, with the same notations as above, the inequalities

1. $\phi_2(z) - \phi_4(z) > 0$ *for* $z \in \Gamma_1 \cap (\{Im\ z > 0\} \setminus \Delta)$,

2. $\phi_2(z) - \phi_3(z) > 0$ *for* $z \in \Gamma_2 \cap (\{Im\ z < 0\} \setminus \Delta)$,

3. $\phi_1(z) - \phi_2(z) > 0$ *for* $z \in (\Gamma \setminus \{Im\ z = 0\}) \setminus \Delta$,

hold.

Proof. We shall prove the first of the above 6 inequalities, i.e. that $\phi_4(z) - \phi_2(z) > 0$ for $z \in (\Gamma_{1,u} \cup \Gamma_{1,l}) \setminus \Delta$. The proof of the remaining inequalities follows from identical arguments.

Denote the minimum of the function $\phi_4(z) - \phi_2(z)$ on $(\Gamma_{1,u} \cup \Gamma_{1,l}) \setminus \Delta$ by $M = M(\delta_{||}, \delta_\perp)$. M is a continuous function of $\delta_{||}$, $\delta_\perp$; furthermore, since we have taken the minimum outside the discs D_+, D_-, there exists a constant $c > 0$ such that $M(0,0) > c > 0$, by Proposition (5.16). By continuity of M, it follows that $\phi_4(z) - \phi_2(z) > 0$ on $(\Gamma_{1,u} \cup \Gamma_{1,l}) \setminus \Delta$ for $\delta_{||}$, $\delta_\perp$ sufficiently small. □

We will also need the following proposition in the next section, when we construct the local parametrices.

Proposition 5.33. *There exist functions* $K_0(z)$, $K_1(z)$, $K_2(z)$, *analytic in a neighborhood of* $z = \alpha$, *such that*

$$\Omega_1(z) = c_0(z-\alpha)^{7/3} + K_0(z) + K_2(z)(z-\alpha)^{2/3} + K_1(z)(z-\alpha)^{1/3}, \tag{5.170}$$

$$\Omega_2(z) = \begin{cases} \omega^2 c_0(z-\alpha)^{7/3} + K_0(z) + \omega K_2(z)(z-\alpha)^{2/3} + \omega^2 K_1(z)(z-\alpha)^{1/3}, & Im\ z > 0, \\ \omega c_0(z-\alpha)^{7/3} + K_0(z) + \omega^2 K_2(z)(z-\alpha)^{2/3} + \omega K_1(z)(z-\alpha)^{1/3}, & Im\ z < 0, \end{cases} \tag{5.171}$$

$$\Omega_4(z) = \begin{cases} \omega c_0(z-\alpha)^{7/3} + K_0(z) + \omega^2 K_2(z)(z-\alpha)^{2/3} + \omega K_1(z)(z-\alpha)^{1/3}, & Im\ z > 0, \\ \omega^2 c_0(z-\alpha)^{7/3} + K_0(z) + \omega K_2(z)(z-\alpha)^{2/3} + \omega^2 K_1(z)(z-\alpha)^{1/3}, & Im\ z < 0. \end{cases} \tag{5.172}$$

Here, $c_0 = c_0(\delta_{||}, \delta_\perp)$ *has the expansion* $c_0 := -\frac{9}{28}(30)^{1/6} + \mathcal{O}(|\delta|)$, *and we define* $c_0^* := c_0(0,0)$.

Proof. The proof of this proposition is a direct computation. For simplicity, consider $\Omega_1(z)$. Near $z = \alpha$, we have that

$$\Omega_1(z) = \sum_{k=0}^{\infty} C_k (z-\alpha)^{k/3};$$

Thus, defining

$$K_0(z) := \sum_{k=0}^{\infty} C_{3k}(z-\alpha)^k, \qquad K_1(z) := \sum_{\substack{k=0\\k\neq 2}}^{\infty} C_{3k+1}(z-\alpha)^k, \qquad K_2(z) := \sum_{k=0}^{\infty} C_{3k+2}(z-\alpha)^k,$$

it follows immediately that $K_0(z), K_1(z)$, and $K_2(z)$ are holomorphic in a neighborhood of $z=\alpha$, and that

$$\Omega_1(z) = c_0(z-\alpha)^{7/3} + K_0(z) + K_2(z)(z-\alpha)^{2/3} + K_1(z)(z-\alpha)^{1/3}.$$

The same functions define $\Omega_2(z)$, $\Omega_4(z)$, but with different choices of branch of $(z-\alpha)^{1/3}$. This completes the proof of the proposition. □

Proposition 5.34. *Set*

$$\boldsymbol{\eta}(z;\delta_{||},\delta_{\perp}) := \frac{25}{54}(30)^{-5/6}\frac{K_2(z)-K_2(\alpha)}{z-\alpha}, \tag{5.173}$$

$$\boldsymbol{\nu}(z;\delta_{||},\delta_{\perp}) := \frac{5}{1312}(30)^{-1/6}K_1(z), \tag{5.174}$$

$$\boldsymbol{\mu}(\delta_{||},\delta_{\perp}) = K_2(\alpha). \tag{5.175}$$

Then,

$$\lim_{n\to\infty} n^{2/7}\boldsymbol{\eta}\left(\alpha+\frac{\xi}{n^{3/7}}; n^{-2/7}\eta, n^{-6/7}\nu\right) = \eta, \tag{5.176}$$

$$\lim_{n\to\infty} n^{6/7}\boldsymbol{\nu}\left(\alpha+\frac{\xi}{n^{3/7}}; n^{-2/7}\eta, n^{-6/7}\nu\right) = \nu, \tag{5.177}$$

$$\lim_{n\to\infty} n^{5/7}\boldsymbol{\mu}\left(n^{-2/7}\eta, n^{-6/7}\nu\right) = 0. \tag{5.178}$$

The convergence is uniform in any compact subset of the ξ plane.

Proof. Expanding $\boldsymbol{\eta}(z;\delta_{||},\delta_{\perp})$ first about $z=\alpha$, we have that

$$\frac{54}{25}(30)^{5/6}\boldsymbol{\eta}(z;\delta_{||},\delta_{\perp}) = C_5(\delta_{||},\delta_{\perp}) + C_8(\delta_{||},\delta_{\perp})(z-\alpha) + \mathcal{O}((z-\alpha)^2).$$

Substituting $z=\alpha+\frac{\xi}{n^{3/7}}$, we find that

$$\frac{54}{25}(30)^{5/6}\boldsymbol{\eta}(\alpha+\frac{\xi}{n^{3/7}};\delta_{||},\delta_{\perp}) = C_5(\delta_{||},\delta_{\perp}) + \mathcal{O}(n^{-3/7}).$$

Expanding in $\delta_{||}, \delta_{\perp}$,

$$\frac{54}{25}(30)^{5/6}\boldsymbol{\eta}(\alpha+\frac{\xi}{n^{3/7}};\delta_{||},\delta_{\perp}) = \frac{54}{25}(30)^{5/6}\delta_{||} - \frac{373}{4800}(30)^{5/6}\delta_{\perp} + \mathcal{O}(|\delta|^2) + \mathcal{O}(n^{-3/7}),$$

where $\mathcal{O}(|\delta|^2)$ denotes terms of order 2 in the parameters $(\delta_{||}, \delta_{\perp})$. Substituting $\delta_{||} = n^{-2/7}\eta$, $\delta_{\perp} = n^{-6/7}\nu$, we obtain

$$\frac{54}{25}(30)^{5/6}\boldsymbol{\eta}(\alpha+\frac{\xi}{n^{3/7}};n^{-2/7}\eta,n^{-6/7}\nu) = \frac{54}{25}(30)^{5/6}n^{-2/7}\eta + \mathcal{O}(n^{-3/7}).$$

Finally, multiplying through by $\frac{25}{54}(30)^{-5/6}n^{2/7}$,

$$n^{2/7}\boldsymbol{\eta}(\alpha+\frac{\xi}{n^{3/7}};n^{-2/7}\eta,n^{-6/7}\nu) = \eta + \mathcal{O}(n^{-1/7}).$$

Taking the limit as $n \to \infty$, we obtain (5.176).

Let us now show the equality (5.177) holds. Expanding $\boldsymbol{\nu}(z;\delta_{||},\delta_{\perp})$ about $z=\alpha$, since $C_4(\delta_{||},\delta_{\perp}) \equiv 0$ by the definition of the modified spectral curve (and the definition of $K_1(z)$):

$$\frac{1312}{5}(30)^{1/6}\boldsymbol{\nu}\left(z;\delta_{||},\delta_{\perp}\right) = C_1(\delta_{||},\delta_{\perp}) + \mathcal{O}((z-\alpha)^3).$$

Substituting $z = \alpha + \frac{\xi}{n^{3/7}}$, we obtain

$$\frac{1312}{5}(30)^{1/6}\boldsymbol{\nu}\left(\alpha+\frac{\xi}{n^{3/7}};\delta_{||},\delta_{\perp}\right) = C_1(\delta_{||},\delta_{\perp}) + \mathcal{O}(n^{-9/7}).$$

Expanding in $(\delta_{||}, \delta_{\perp})$,

$$\begin{aligned}\frac{1312}{5}(30)^{1/6}\boldsymbol{\nu}\left(\alpha+\frac{\xi}{n^{3/7}};\delta_{||},\delta_{\perp}\right) &= \frac{1312}{5}(30)^{1/6}\delta_{\perp} + \frac{3936}{25}(30)^{1/6}\delta_{||}\delta_{\perp} \\ &\quad - \frac{1252919}{225}(30)^{1/6}\delta_{\perp}^2 + \mathcal{O}(|\delta|^3) + \mathcal{O}(n^{-9/7}),\end{aligned}$$

where $\mathcal{O}(|\delta|^3)$ denotes terms of order 3 in the parameters $(\delta_{||}, \delta_{\perp})$. Substituting $\delta_{||} = n^{-2/7}\eta$, $\delta_{\perp} = n^{-6/7}\nu$, we obtain

$$\frac{1312}{5}(30)^{1/6}\boldsymbol{\nu}\left(\alpha+\frac{\xi}{n^{3/7}};n^{-2/7}\eta,n^{-6/7}\eta\right) = \frac{1312}{5}(30)^{1/6}n^{-6/7}\nu + \mathcal{O}(n^{-8/7}).$$

Finally, multiplying through by $\frac{5}{1312}(30)^{-1/6}n^{6/7}$,

$$n^{6/7}\boldsymbol{\nu}\left(\alpha+\frac{\xi}{n^{3/7}};n^{-2/7}\eta,n^{-6/7}\eta\right) = \nu + \mathcal{O}(n^{-2/7}).$$

Taking the limit as $n \to \infty$, we obtain (5.177).

Finally, expanding $\boldsymbol{\mu}(\delta_{||}, \delta_{\perp})$ in $\delta_{||}, \delta_{\perp}$, we find that

$$\boldsymbol{\mu}(\delta_{||}, \delta_{\perp}) = \frac{164}{5}(30)^{1/6}\delta_{\perp}\left[1 + \mathcal{O}(|\delta|)\right],$$

where $\mathcal{O}(|\delta|)$ denotes terms of order 1 or higher in $\delta_{||}, \delta_{\perp}$. Substituting $\delta_{||} = n^{-2/7}\eta$, $\delta_{\perp} = n^{-6/7}\nu$, we therefore find that

$$\boldsymbol{\mu}(n^{-2/7}\eta, n^{-6/7}\nu) = \mathcal{O}(n^{-6/7}).$$

Multiplying through by $n^{5/7}$, and taking the limit as $n \to \infty$, we obtain (5.178). $\square$

Remark 5.35. We remark that the constant $\hat{C}$ appears in the expansion of $K_2(\alpha)$ about $(\delta_{||}, \delta_{\perp}) = (0, 0)$: Expanding $K_2(\alpha)$ in $(\delta_{||}, \delta_{\perp})$, we find that

$$K_2(\alpha) = \frac{164}{5}(30)^{1/3}\delta_{\perp} + \mathcal{O}(|\delta|^2) = \hat{C}\delta_{\perp} + \mathcal{O}(|\delta|^2).$$

5.6.2.2 Construction of the Parametrices. We now construct the local parametrices; define discs $D_{\pm}$ of sufficiently small radii (to be determined) around $z = \alpha$, $z = -\alpha$. Our new parametrix $\hat{M}(z)$ will be defined as

$$\hat{M}(z) = \begin{cases} M(z), & z \in \mathbb{C} \setminus (D_{\pm}), \\ P_{+\alpha}(z), & z \in D_{+}, \\ P_{-\alpha}(z), & z \in D_{-}. \end{cases} \tag{5.179}$$

The functions $P_{\pm\alpha}(z)$ will be chosen to so that the following conditions are met:

1. $P_{\pm\alpha}(z)$ matches the jumps of $\mathbf{S}(z)$ *exactly* in the discs $D_{\pm}$,

2. $P_{\pm\alpha}(z) = M(z)[\mathbb{I} + \mathcal{O}(n^{-\delta})]$, as $z \to \partial D_{\pm}$, for some $\delta > 0$.

Before calculating the parametrices, it is useful to notice that, by the symmetry of the functions $\Omega_j(z)$, we have that

$$P_{-\alpha}(z) = \hat{\sigma}_{34} P_{+\alpha}(-z)\hat{\sigma}_{34},$$

and so it is sufficient to calculate the parametrix at $z = \alpha$ only, and use the above symmetry relation as the definition of $P_{-\alpha}(z)$. We therefore relabel $P_{\alpha}(z) =: P(z)$. We need two auxilliary functions for the construction of the parametrix, $F(\xi)$ and $\Psi(\xi; \eta, \mu, \nu)$; we define these functions here. The first such function, which we shall call $F(\xi)$, is chosen to match the jumps of the global parametrix in a neighborhood of $z = \alpha$

We will momentarily ignore the 3^{rd} row and column of global parametrix for this, as the jumps only involve the rows/columns 1,2, and 4. Let $\xi \in \mathbb{C}$ be an auxilliary variable, and set

$$F(\xi) = \mathrm{diag}(1, \xi^{1/3}, \xi^{-1/3}) \cdot C(\xi), \tag{5.180}$$

where $\xi^{\pm 1/3}$ are taken with the principal branch cut, and $C(\xi)$ is a piecewise constant matrix, to be determined:

$$C(\xi) := \begin{cases} C_+, & \mathrm{Im}\ \xi > 0, \\ C_-, & \mathrm{Im}\ \xi < 0. \end{cases} \tag{5.181}$$

Set $\Xi(\xi) := \mathrm{diag}\ (1, \xi^{1/3}, \xi^{-1/3})$, and note that

$$\Xi_+(\xi) = \Xi_-(\xi) J_\Xi, \qquad \xi < 0, \tag{5.182}$$

where $J_\Xi = \mathrm{diag}\ (1, \omega, \omega^2)$, $\omega = e^{\frac{2\pi i}{3}}$. The matrix $C(\xi)$ is to be determined by the jump conditions. We want $F(\xi)$ to satisfy the jump conditions

$$F_+(\xi) = F_-(\xi) \begin{cases} M_1 := \left(\begin{smallmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{smallmatrix}\right), & \xi < 0, \\ M_2 := \left(\begin{smallmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{smallmatrix}\right), & \xi > 0. \end{cases} \tag{5.183}$$

(Note that M_1, M_2 are just the jumps of the global parametrix to the left/right of $z = \alpha$ with the third row and column deleted, respectively). For $\xi > 0$, we have that

$$\Xi(\xi) C_+ = F_+(\xi) = F_-(\xi) = \Xi(\xi) C_- M_2, \tag{5.184}$$

and so we find the constraint

$$C_+ = C_- M_2. \tag{5.185}$$

This determines one of the matrices C_+, C_-. To determine the other matrix, we impose the jump condition for $\xi < 0$:

$$\Xi_-(\xi) J_\Xi C_+ = \Xi_+(\xi) C_+ = F_+(\xi) = F_-(\xi) M_1 = \Xi_-(\xi) C_- M_1, \tag{5.186}$$

which yields the additional constraint equation

$$J_\Xi C_+ = C_- M_1; \tag{5.187}$$

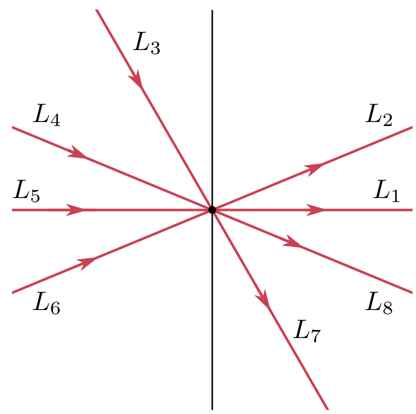


Figure 5.15. The jump contours of the Ψ-Riemann Hilbert problem in the ξ-plane.

This equation can either be expressed as a set of linear equations of C_+ or C_-:

$$J_\Xi C_+ = C_+ M_2^{-1} M_1 \qquad \text{or} \qquad J_\Xi C_- M_2 = C_- M_1 \tag{5.188}$$

For example, the general solution of the equation for $C_- := (c_{ij})$ is

$$C_- = \begin{pmatrix} c_{13} & -c_{13} & c_{13} \\ c_{21} & -\omega c_{21} & \omega^2 c_{21} \\ \omega^2 c_{33} & \omega c_{33} & c_{33} \end{pmatrix} \tag{5.189}$$

We choose the rest of the constants so for later convenience:

$$C_- = \begin{pmatrix} 1 & -1 & 1 \\ \omega^2 & -1 & \omega \\ \omega & -1 & \omega^2 \end{pmatrix} \tag{5.190}$$

C_+ can be constructed by multiplying C_- on the right by M_2:

$$C_+ = C_- M_2 = \begin{pmatrix} 1 & 1 & 1 \\ \omega^2 & \omega & 1 \\ \omega & \omega^2 & 1 \end{pmatrix}. \tag{5.191}$$

By construction, we have the following proposition:

Proposition 5.36. *The function $\hat{\sigma}_{34}[F(\xi) \oplus 1]\hat{\sigma}_{34}$ matches the jumps of the global parametrix in a neighborhood of $z = \alpha$.*

Define contours in the ξ plane $L_1, ..., L_8$ as the images of the following contours in the z-plane under the map $\xi(z) = n^{3/7}(z - \alpha)$:

- L_1 is the image of the contour $z > \alpha$,
- L_2 is the image of the contour $\Gamma_{1,u}$,

- L_3 is the image of the contour $\Gamma_1 \cap \{\text{Im } z > 0\}$,
- L_4 is the image of the contour Γ_u,
- L_5 is the image of the contour $z < \alpha$,
- L_6 is the image of the contour Γ_l,
- L_7 is the image of the contour $\Gamma \cap \{\text{Im } z < 0\}$,
- L_8 is the image of the contour $\Gamma_{1,l}$.

These contours are depicted in Figure (5.15). We now define the matrix valued function $\Psi(\xi;\eta,\mu,\nu)$ as the unique solution to the following 3×3 Riemann-Hilbert problem:

$$\Psi_+(\xi;\eta,\mu,\nu) = \Psi_-(\xi;\eta,\mu,\nu) \times \begin{cases} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, & \xi \in L_1, \\ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix}, & \xi \in L_2 \cup L_8, \\ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix}, & \xi \in L_3, \\ \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, & \xi \in L_4 \cup L_6, \\ \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, & \xi \in L_5, \\ \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, & \xi \in L_7. \end{cases} \tag{5.192}$$

satisfying the normalization condition

$$\Psi(\xi;\eta,\mu,\nu) = \left[\mathbb{I} + \mathcal{O}(\xi^{-1})\right] F(\xi) e^{-\vartheta(\hat{D}\xi^{1/3};\eta,\mu,\nu)}, \qquad |\xi| \to \infty. \tag{5.193}$$

where $\vartheta(\xi;\eta,\mu,\nu) := \frac{3}{7}\xi^7 + \eta\xi^5 + \mu\xi^2 + \nu\xi$, $F(\xi)$ is as defined above, and $\hat{D}$ is the matrix

$$\hat{D} := \begin{cases} \text{diag } (1, \omega^2, \omega), & \text{Im } \xi > 0, \\ \text{diag } (1, \omega, \omega^2), & \text{Im } \xi < 0. \end{cases} \tag{5.194}$$

Existence of the solution to this problem we postpone to a later work; for now, we will assume the existence, and describe how the function $\Psi(\xi;\eta,\mu,\nu)$ can be used to construct the parametrix. Define first the scaled coordinate $\xi(z)$ by

$$\xi(z) = n^{3/7}(z-\alpha). \tag{5.195}$$

Proposition 5.37. *Put $\mathcal{Q}(z) := \text{diag}\,(K_0(z), K_0(z), \Omega_3(z), K_0(z))$, where $K_0(z)$ is as defined in Proposition (5.33). Define the function*

$$P(z) := M(z)\hat{\sigma}_{34}\left[F(c_1\xi(z))^{-1}\Psi(c_1\xi(z); c_2 n^{2/7}\boldsymbol{\eta}(z), c_3 n^{5/7}\boldsymbol{\mu}, c_4 n^{6/7}\boldsymbol{\nu}(z)) \oplus 1\right]\hat{\sigma}_{34}e^{-n\mathcal{Q}(z)}\,\mathbf{G}(z), \quad (5.196)$$

where the constants c_1, c_2, and c_3 are defined as

$$c_1 := (7c_0/3)^{3/7}, \qquad c_2 := \frac{54}{25}(30)^{5/6}c_1^{-5/3}, \qquad c_3 := c_1^{-2/3}, \qquad c_4 := \frac{1312}{5}(30)^{1/6}c_1^{-1/3}, \quad (5.197)$$

$M(z)$ is the solution to the global parametrix problem (5.139), $F(z)$, $\Psi(z)$ are as defined above, and $\mathbf{G}(z)$ is the matrix appearing in the second transformation, defined by (5.123). Then, $P(z)$ satisfies the following conditions:

1. *$P(z)$ matches the jumps of $\mathbf{S}(z)$ exactly in the disc D_+,*
2. *$P(z) = M(z)[\mathbb{I} + \mathcal{O}(n^{-1/7})]$, for $z \in \partial D_+$, as $n \to \infty$.*

Proof. By construction, we see that the jumps of $P(z)$ indeed match exactly the jumps of $P(z)$ inside a sufficiently small disc D_+ (chosen small enough so that the series expansions defining the functions $K_j(z)$ converge). Indeed, if z belongs to one of the jump contours off the real axis, then $M(z)$, $F(c_1\xi(z))$, and $\mathbf{G}(z)$ are analytic, so that

$$\begin{aligned}
P_+(z) &= M\hat{\sigma}_{34}\left[F^{-1}\Psi_+ \oplus 1\right]\hat{\sigma}_{34}e^{-n\mathcal{Q}}\,\mathbf{G} \\
&= M\hat{\sigma}_{34}\left[F^{-1}\Psi_- J_\Psi \oplus 1\right]\hat{\sigma}_{34}e^{-n\mathcal{Q}}\,\mathbf{G} \\
&= M\hat{\sigma}_{34}\left[F^{-1}\Psi_- \oplus 1\right]\hat{\sigma}_{34}\left(\hat{\sigma}_{34}\left[J_\Psi \oplus 1\right]\hat{\sigma}_{34}\right)e^{-n\mathcal{Q}}\,\mathbf{G} \\
&= P_-(z)\left[\mathbf{G}^{-1}\left(\hat{\sigma}_{34}\left[J_\Psi \oplus 1\right]\hat{\sigma}_{34}\right)\mathbf{G}\right].
\end{aligned}$$

By the definition of the jumps J_Ψ, one finds that the above coincides exactly with the jumps of $\mathbf{S}(z)$ there. Now, if z belongs to one of the jump contours on the real axis, all of the functions M, Ψ, F, and $\mathbf{G}$ have jumps. In particular, by the definition of F, we have that

$$J_M = \hat{\sigma}_{34}\left[J_F \oplus 1\right]\hat{\sigma}_{34}. \quad (5.198)$$

Thus,

$$\begin{aligned}
P_+(z) &= M_+\hat{\sigma}_{34}\left[F_+^{-1}\Psi_+\oplus 1\right]\hat{\sigma}_{34}e^{-n\mathcal{Q}}\,\mathbf{G}_+ \\
&= M_-J_M\hat{\sigma}_{34}\left[J_F^{-1}F_-^{-1}\Psi_-J_\Psi\oplus 1\right]\hat{\sigma}_{34}e^{-n\mathcal{Q}}J_G\,\mathbf{G}_-\,J_G \\
(\mathit{Equation}\,(5.198)) &= M_-\hat{\sigma}_{34}\left[F_-^{-1}\Psi_-\oplus 1\right]\hat{\sigma}_{34}\left(\hat{\sigma}_{34}\left[J_\Psi\oplus 1\right]\hat{\sigma}_{34}\right)e^{-n\mathcal{Q}}J_G\,\mathbf{G}_-\,J_G \\
&= P_-(z)\left[\mathbf{G}_-^{-1}\left(\hat{\sigma}_{34}\left[J_\Psi\oplus 1\right]\hat{\sigma}_{34}\right)J_G\,\mathbf{G}_-\,J_G\right] \\
&= P_-(z)\left[\mathbf{G}_-^{-1}\left(\hat{\sigma}_{34}\left[J_\Psi\oplus 1\right]\hat{\sigma}_{34}\right)\mathbf{G}_+\right].
\end{aligned}$$

Again, by the definition of the jump matrices J_Ψ, one can calculate that the jumps above again coincide with those of $\mathbf{S}(z)$.

Let us now show that, as $n\to\infty$ on the boundary of the disc D_+, that $P(z) = M(z)[\mathbb{I}+\mathcal{O}(n^{-1/7})]$. Note that $\vartheta(c_1(\xi(z))^{1/3}, c_2n^{2/7}\boldsymbol{\eta}(z), c_3n^{5/7}\boldsymbol{\mu}, c_4n^{6/7}\boldsymbol{\nu}(z))$ is given by

$$\begin{aligned}
\vartheta(c_1[\xi(z)]^{1/3}, c_2n^{2/7}\boldsymbol{\eta}(z), c_3n^{6/7}\boldsymbol{\nu}(z)) &= nc_0(z-\alpha)^{7/3} + n\left(\frac{K_2(z)-K_2(\alpha)}{z-\alpha}\right)(z-\alpha)^{5/3} \\
&\quad + nK_2(\alpha)(z-\alpha)^{2/3} + nK_1(z)(z-\alpha)^{1/3} \\
&= n\Omega_1(z) - nK_0(z),
\end{aligned}$$

uniformly in a neighborhood of $z=\alpha$. Similar equalities can be verified for the other entries of the exponent $\vartheta(c_1\hat{D}[\xi(z)]^{1/3}, c_2n^{2/7}\boldsymbol{\eta}(z), c_3n^{5/7}\boldsymbol{\mu}, c_4n^{6/7}\boldsymbol{\nu}(z))$. Thus, we see that, as $n\to\infty$,

$$\begin{aligned}
&\hat{\sigma}_{34}\left[F(c_1\xi(z))^{-1}\Psi(c_1\xi(z); c_2n^{2/7}\boldsymbol{\eta}(z), c_3n^{5/7}\boldsymbol{\mu}, c_4n^{6/7}\boldsymbol{\nu}(z))\oplus 1\right]\hat{\sigma}_{34}e^{-n\mathcal{Q}(z)}\,\mathbf{G}(z) \\
&= \hat{\sigma}_{34}\left[F(c_1\xi(z))^{-1}\left[\mathbb{I}+\mathcal{O}(\xi(z)^{-1})\right]F(c_1\xi(z))\right]\hat{\sigma}_{34} \\
&= \mathbb{I}+\mathcal{O}(n^{-1/7}).
\end{aligned}$$

The last equality follows from the fact that conjugation of the $\mathcal{O}(z^{-1})$ term by $F(c_1\xi(z))$ yields terms of order $n^{-1/7}$. It thus follows that

$$P(z) = M(z)[\mathbb{I}+\mathcal{O}(n^{-1/7})].$$

□

Remark 5.38. The parameters η,μ,ν appearing in the model Riemann-Hilbert problem correspond to translations in temperature, magnetic field, and gravitation/matrix model parameter, respectively (cf. the lecture notes [50], page 57, for example). The magnetic parameter μ disappears completely from our con-

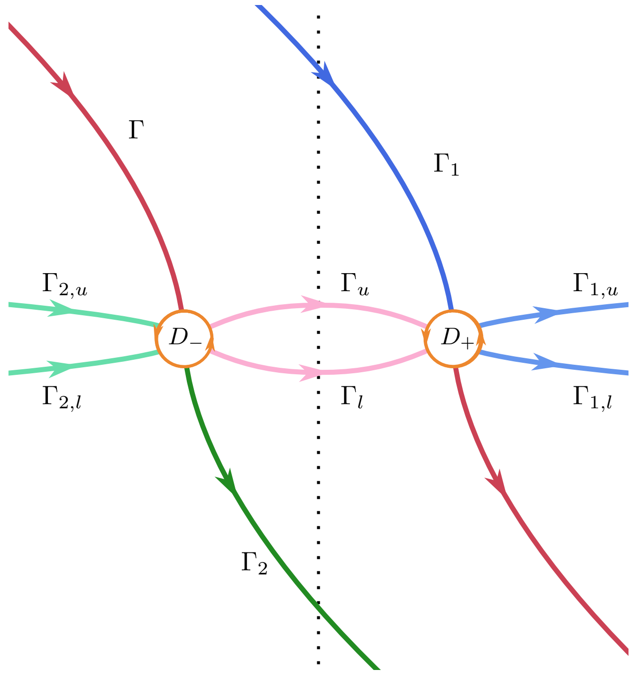


Figure 5.16. Opened lenses around the branch points (note: the spectral curve is at multicriticality here). The discs $D_\pm$ are bounded by the orange circles. The above depicts the system of contours Σ, on which $\mathbf{R}(z)$ has jumps.

siderations, since we are considering the 2-matrix model with symmetric potentials $V = W = \frac{1}{2}z^2 + \frac{t}{4}z^4$. If we had instead chosen the potentials as

$$V(X) = \frac{1}{2}X^2 + \frac{te^H}{4}X^4, \qquad W(Y) = \frac{1}{2}Y^2 + \frac{te^{-H}}{4}Y^4, \tag{5.199}$$

then we would have that a multicritical point appears at $(\tau, t, H) = (\frac{1}{4}, -\frac{5}{72}, 0)$, and taking small variations in μ would lead to the appearance of a nontrivial scaling of the μ term in the model Riemann-Hilbert problem for Ψ. This scaling will be the study of a future work; for now, we take $\mu = 0$ in the local parametrix, and write $\Psi(\xi; \eta, 0, \nu) := \Psi(\xi; \eta, \nu)$, by a slight abuse of notation.

5.6.3 The Final Transformation $\mathbf{S} \to \mathbf{R}$

Finally, we have obtained a "good" parametrix. We now set

$$\mathbf{R}(z) = \begin{cases} \mathbf{S}(z)M^{-1}(z), & z \in \mathbb{C} \setminus D_\pm, \\ \mathbf{S}(z)P^{-1}(z), & z \in D_+, \\ \mathbf{S}(z)\hat{\sigma}_{34}P^{-1}(-z)\hat{\sigma}_{34}, & z \in D_-. \end{cases} \tag{5.200}$$

We then have the following proposition:

Proposition 5.39. *The function* $\mathbf{R}(z)$ *is the unique solution to the following Riemann-Hilbert problem:* $\mathbf{R}(z)$ *is analytic in* $\mathbb{C} \setminus \Sigma$*, where* Σ *is the system of contours consisting of the lens boundaries outside of the discs* $D_\pm$*, and the contours* $\Gamma, \Gamma_1, \Gamma_2$ *outside of the discs* $D_\pm$*, and the boundaries of the discs* $\partial D_\pm$ *(see the Figure (5.16) for the jump contours of* $\mathbf{R}(z)$*). Furthermore,* $\mathbf{R}(z)$ *satisfies*

$$\mathbf{R}_+(z) = \mathbf{R}_-(z)\left[\mathbb{I} + \mathcal{O}(e^{-n})\right], n \to \infty, \tag{5.201}$$

for $z \in \Sigma \setminus \partial D_\pm$*, and*

$$\mathbf{R}_+(z) = \mathbf{R}_-(z)\left[\mathbb{I} + \mathcal{O}(n^{-1/7})\right], n \to \infty, \tag{5.202}$$

for $z \in \partial D_\pm$*.* $\mathbf{R}(z)$ *is also normalized as*

$$\mathbf{R}(z) = \mathbb{I} + \mathcal{O}(z^{-1/3}), \qquad z \to \infty. \tag{5.203}$$

Proof. By construction, the global parametrix matches the constant jumps/asymptotics of $\mathbf{S}(z)$ outside of the discs $D_\pm$ *exactly*, and so we are left only with exponentially small jumps of $\mathbf{S}(z)$ on the contours $\Sigma \setminus \partial D_\pm$. By the definition of the local parametrices, the jumps of $\mathbf{S}(z)$ are matched exactly in the discs $D_\pm$. On the boundary of these disc D_+, we have that

$$\mathbf{R}_+(z) = \mathbf{S}(z)P^{-1}(z) = \mathbf{S}(z)M^{-1}(z)M(z)P^{-1}(z) = \mathbf{R}_-(z)[M(z)P^{-1}(z)], \tag{5.204}$$

and a similar equality holds at $z = -\alpha$. Thus, we have only to show that $M(z)P^{-1}(z) = \mathbb{I} + \mathcal{O}(n^{-1/7})$, $n \to \infty$. Indeed, if we set $\hat{N} := \operatorname{diag}(1, c_1^{-1/3}n^{-1/7}, 1, c_1^{1/3}n^{1/7})$, we see that $\hat{N}F(\xi)$ is an n-independent function, so that

$$M(z)P^{-1}(z) = M(z)F(c_1\xi(z))^{-1}\hat{N}^{-1}\hat{N}\left[\mathbb{I} + \frac{A}{c_1\xi(z)} + \frac{B}{(c_1\xi(z))^2} + \mathcal{O}(\xi(z)^{-3})\right]\hat{N}^{-1}\hat{N}F(c_1\xi(z))M(z)^{-1},$$

as $n \to \infty$. By construction, the function $H(z) := M(z)F(c_1\xi(z))^{-1}\hat{N}^{-1}$ is an n-independent function, holomorphic in a neighborhood of $z = \alpha$. We also must take care to write the $c_1 := c_1(\eta, \nu) = \hat{c}^{3/7} + \mathcal{O}(n^{-2/7})$. The jump of $\mathbf{R}(z)$ across the boundary of the disc centered at $z = \alpha$ is then

$$\begin{aligned} J_{\mathbf{R}} &= H(z)\hat{N}[\mathbb{I} + A\xi^{-1} + B\xi^{-2} + \mathcal{O}(\xi^{-3})]\hat{N}^{-1}H(z)^{-1} \\ &= \mathbb{I} + \frac{a_{32}\hat{c}^{-1/7}n^{-1/7}}{z-\alpha}\left[H(z)E_{42}H(z)^{-1}\right] + \frac{\hat{c}^{-2/7}n^{-2/7}}{z-\alpha}\left[a_{12}H(z)E_{12}H(z)^{-1} + a_{31}H(z)E_{41}H(z)^{-1}\right] + \mathcal{O}(n^{-3/7}). \end{aligned}$$

The jump across the circle of sufficiently small radius $z = \alpha$ is thus of order $\mathbb{I} + \mathcal{O}(n^{-1/7})$. $\square$

Thus, by small norm theory, we have guaranteed a solution to the Riemann-Hilbert problem for $\mathbf{R}(z)$ exists for n sufficiently large. The remaining task is to develop the first few terms in the Neumann series for $\mathbf{R}(z)$, as these terms will contain the relevant recursion coefficients. Following [38, 74], we develop $\mathbf{R}(z)$ as a series in powers of $n^{-1/7}$, using the fact that the jump matrix $J_{\mathbf{R}}(z)$ has an expansion in powers of $n^{-1/7}$:

Lemma 5.40. *For n sufficiently large, the RHP for $\mathbf{R}(z)$ has a solution, which admits the asymptotic expansion*

$$\mathbf{R}(z) = \mathbb{I} + \sum_{k=1}^{\infty} R_k(z) n^{-k/7}, \tag{5.205}$$

as $n \to \infty$, uniformly for $z \in \mathbb{C} \setminus \Gamma_{\mathbf{R}}$. Furthermore, this expansion is valid uniformly near infinity, in the sense that, for any $K \geq 1$, there is a constant $C_K > 0$ such that

$$\left|\left|\mathbf{R}(z) - \mathbb{I} - \sum_{k=1}^{K-1} R_k(z) n^{-k/7}\right|\right| \leq C_K |z|^{-1} n^{-K/7}, \tag{5.206}$$

whenever z is sufficiently large. (Here, $||\cdot||$ denotes any matrix norm).

Proof. We have already seen that the jump matrix for $\mathbf{R}(z)$ is close to the identity matrix for n sufficiently large. It is also clear that there exist constants $C, \epsilon > 0$ such that

$$||J_{\mathbf{R}}(z) - \mathbb{I}|| \leq C e^{-\epsilon |z|^{4/3}}, \tag{5.207}$$

for z sufficiently large, in the unbounded components of $\Gamma_{\mathbf{R}}$. This follows directly from the form of the jump matrices there. The lemma then follows from standard arguments, cf. [29, 38, 74]. $\square$

We can now proceed to calculate the first few terms in the Neumann series for $\mathbf{R}(z)$.

5.6.3.1 Subleading asymptotics for $\mathbf{R}(z)$. Following [29, 38, 74], we can compute further terms in the Neumann series for $\mathbf{R}(z)$ by recursively solving the Riemann-Hilbert problems for the functions $R_k(z)$, which satisfy the following Riemann-Hilbert problems:

$$\begin{cases} R_k \text{ is analytic in } \mathbb{C} \setminus \partial D_{\pm}, \\ R_{k,+}(z) = R_{k,-}(z) + \sum_{\ell=1}^{k} R_{k-l,-}(z) J_{\ell}^{\pm}(z), & z \in \partial D_{\pm}, \\ R_k(z) = \mathcal{O}(z^{-1}), & z \to \infty, \end{cases} \tag{5.208}$$

with the initial condition $R_0(z) = \mathbb{I}$. Here, $J_\ell^\pm(z)$ are the large-n expansions of the jump matrix $J_{\mathbf{R}}(z)$ for $\mathbf{R}(z)$ on the boundaries of the discs $D_\pm$, respectively:

$$J_{\mathbf{R}}(z) = \mathbb{I} + \sum_{\ell=1}^{\infty} J_\ell^\pm(z) n^{-\ell/7} \qquad \text{on } \partial D_\pm.$$

For example, the first few $J_\ell^\pm(z)$ are given by

$$\begin{aligned}
J_1^+(z) &= \frac{a_{32}}{z-\alpha}\left[H(z)E_{42}H(z)^{-1}\right], \\
J_2^+(z) &= \frac{1}{z-\alpha}\left[a_{12}H(z)E_{12}H(z)^{-1} + a_{31}H(z)E_{41}H(z)^{-1}\right], \\
J_1^-(z) &= -\frac{a_{32}}{z+\alpha}\left[\sigma_{34}H(-z)E_{42}H(-z)^{-1}\sigma_{34}\right], \\
J_2^-(z) &= -\frac{1}{z+\alpha}\sigma_{34}H(-z)\left[a_{12}E_{12} + a_{31}E_{41}\right]H(-z)^{-1}\sigma_{34}
\end{aligned}$$

These RHPs are additive, and thus can be solved immediately by means of the Plemelj formulae. However, the fact that the jump matrices admit analytic continuations all the way up to the branch points allows us to compute the first few coefficients in a much more streamlined way. We have the following proposition.

Proposition 5.41. *The function $R_1(z)$ solving the Riemann-Hilbert problem above is given by*

$$R_1(z) = \begin{cases} \dfrac{1}{z-\alpha}\operatorname*{Res}_{z=\alpha} J_1^+(z) + \dfrac{1}{z+\alpha}\operatorname*{Res}_{z=-\alpha} J_1^-(z), & z \in \mathbb{C}\setminus D_\pm, \\ \dfrac{1}{z-\alpha}\operatorname*{Res}_{z=\alpha} J_1^+(z) + \dfrac{1}{z+\alpha}\operatorname*{Res}_{z=-\alpha} J_1^-(z) - J_1^+(z), & z \in D_+, \\ \dfrac{1}{z-\alpha}\operatorname*{Res}_{z=\alpha} J_1^+(z) + \dfrac{1}{z+\alpha}\operatorname*{Res}_{z=-\alpha} J_1^-(z) - J_1^-(z), & z \in D_-. \end{cases} \tag{5.209}$$

5.7 Calculation of the Partition Function.

We are now in a position where we are able to calculate the asymptotics of the partition function. As noted in the introduction, the partition function for the 2-matrix model can be written in terms of an isomonodromic $\boldsymbol{\tau}$ function, as per [9]. Explicitly, the $\boldsymbol{\tau}$ function is expressible in terms of the solution of the Riemann-Hilbert problem for $\mathbf{Y}(z)$. Since we have succeeded in finding the asymptotics of $\mathbf{Y}(z)$, we in turn can produce an asymptotic expression for the partition function. The expression for the $\boldsymbol{\tau}$-differential is

$$d\log\boldsymbol{\tau}_n(\tau,t) = \langle \mathbf{Y}^{-1}\mathbf{Y}'\left(\begin{smallmatrix} 0 & 0 \\ 0 & \frac{\partial \mathbf{W}}{\partial\tau}\mathbf{W}^{-1}\end{smallmatrix}\right)\rangle d\tau + 2\times\langle \mathbf{Y}^{-1}\mathbf{Y}_n'\left(\begin{smallmatrix} 0 & 0 \\ 0 & \frac{\partial \mathbf{W}}{\partial t}\mathbf{W}^{-1}\end{smallmatrix}\right)\rangle dt, \tag{5.210}$$

where $'$ denotes the derivative with respect to the spectral variable z, and $\mathbf{W}$ is the matrix which appears in the first transformation. The partition function's differential is then

$$d\log \mathcal{Z}_{\Updownarrow\dashv\sqcup\nabla\rangle\S} = d\log\left[\left(\frac{t}{\tau}\right)^{\frac{n}{2}\left(\frac{n}{3}-1\right)}\boldsymbol{\tau}_n\right]. \tag{5.211}$$

The derivations of the above formulae for the $\boldsymbol{\tau}$-function and partition function are sketched in Appendix D. Combined with the asymptotic formulae of the previous sections, these formulae allow us to state the following theorem:

Theorem 5.42. *Let (τ, t) belong to the interior of the region D. Then, as $n \to \infty$,*

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n^2}\log\frac{Z_{matrix}(\tau,t;n,n)}{Z_{matrix}(\tau,0;n,n)} = \frac{1}{2}\log\frac{\tau z(\tau,t)}{2t} - \int_0^{z(\tau,t)}\frac{d\zeta}{\zeta}\left[k(\zeta;\tau,t)-\frac{1}{2}k^2(\zeta;\tau,t)\right]\\ -\log\frac{\tau}{2(1-\tau^2)} - 1,\end{aligned} \tag{5.212}$$

where $k(\zeta;\tau,t)$ is defined as

$$k(\zeta;\tau,t) := \frac{\zeta}{t}\left[\frac{1}{(1-3\zeta)^2} - \tau^2 + 3\tau^2\zeta^2\right], \tag{5.213}$$

and $z = z(\tau,t)$ is implicitly determined as the unique solution of the fifth-order equation

$$t = z\left[\frac{1}{(1-3z)^2} - \tau^2 + 3\tau^2 z^2\right] \tag{5.214}$$

which satisfies

$$\lim_{t\to 0}\frac{z(\tau,t)}{t} = \frac{1}{(1-\tau^2)}. \tag{5.215}$$

Proof. The proof of this theorem follows from direct application of the above Riemann-Hilbert analysis to the formula for the partition function:

$$\lim_{n\to\infty}\frac{1}{n^2}d\log\frac{Z_{matrix}(\tau,t;n,n)}{Z_{matrix}(\tau,0;n,n)} = -\frac{dt}{3t} + \frac{d\tau}{6\tau} - \frac{1+\tau^2}{\tau(1-\tau^2)}d\tau + \lim_{n\to\infty}\frac{1}{n^2}d\log\boldsymbol{\tau}_n, \tag{5.216}$$

where $d\log\boldsymbol{\tau}_n$ is defined as in Appendix (D). We remark that the formulae in the Riemann-Hilbert analysis must be expanded to order z^{-2}, as the definition of the tau function contains coefficients of $\mathbf{Y}$ at infinity of this order. □

We are also able to prove the following theorem about the behavior of the partition function at the multicritical point:

Theorem 5.43. *Set* $\tau = \frac{1}{4} + n^{-2/7}\eta - \frac{1}{9}n^{-6/7}\nu$, *and* $t = -5/72 + \frac{1}{9}\eta n^{-2/7} + \frac{2}{9}\eta^2 n^{-4/7} + \nu n^{-6/7} - \frac{8}{9}\eta^3 n^{-6/7}$. *Then, as* $n \to \infty$,

$$\lim_{n\to\infty} \frac{1}{n^2} \log \frac{Z_{matrix}(\tau, t; n, n)}{Z_{matrix}(\tau, 0; n, n)} = \left(\int^{\nu} u(\sigma) d\sigma \right) d\nu + F(u, v) d\eta, \tag{5.217}$$

where u, v *are the functions appearing the string equation for* KdV_3.

CHAPTER 6

FURTHER WORK AND CONCLUSION.

In this chapter, we prevent further results that build upon the previous chapter. We also present some further open problems that are part of the larger program of this thesis, and will hopefully be the subject of future work.

6.1 The Spectral Curve For the Cubic 2-Matrix Model.

Much like in the case of the 1-matrix model, it is also of interest to study the *cubic* 2-matrix model, i.e., the formal generating function

$$Z_n(\tau,t,N) := \iint \exp \operatorname{tr}\left[\tau XY - \frac{1}{2}X^2 - \frac{1}{2}Y^2 - \frac{t}{3N^{1/2}}X^3 - \frac{t}{3N^{1/2}}Y^3\right] dXdY, \tag{6.1}$$

where the integral is taken over the space of $n \times n$ Hermitian matrices. The above integral does not converge for any value of t, and so care is needed to give an appropriate interpretation. As in the previous chapter, we interpret the above integral in terms of a family of biorthogonal polynomials, as an analytic continuation of their recursion coefficients. We consider the family of monic biorthogonal polynomials defined by the relation

$$\int_\Gamma \int_\Gamma p_k(x)q_j(y) \exp\left(N\left[\tau xy - \frac{1}{2}x^2 - \frac{1}{2}y^2 - \frac{t}{3}x^3 - \frac{t}{3}y^3\right]\right) dxdy = h_k\delta_{jk}, \tag{6.2}$$

where the contour Γ here is taken to start from $\infty \times e^{2\pi i/3}$ along the real axis, and then go back off to infinity along the positive real axis, as shown in Figure (6.1). These polynomials are well-defined, provided $t < 0$, $\tau > 0$.

As was the case for the biorthogonal polynomials of the previous chapter, the biorthogonal polynomials defined by the relation (6.2) satisfy a Riemann Hilbert problem. However, since the potentials $V = W = \frac{1}{2}z^2 + \frac{t}{3}z^3$ are *cubic*, the associated Riemann-Hilbert problem is 3×3. Define the function

$$f(z) = f(z;\tau,t,N) := \int_\Gamma \exp\left[N(\tau zw - V(w))\right]dw. \tag{6.3}$$

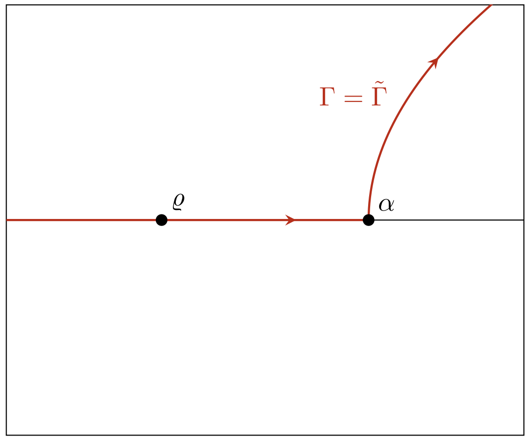


Figure 6.1. The contour Γ, marked along with some relevant points.

We then have the following RHP:

Problem 6.1. *Construct a* 3×3 *function, analytic in* $\mathbb{C} \setminus \Gamma$*, such that:*

$$\mathbf{Y}_+(z) = \mathbf{Y}_-(z) \left[\mathbb{I} + e^{-NV(z)} \begin{pmatrix} 0 & f(z) & \frac{f'(z)}{N\tau} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \right], \qquad z \in \Gamma, \tag{6.4}$$

subject to the normalization condition

$$\mathbf{Y}(z) = \left[\mathbb{I} + \mathcal{O}\left(\frac{1}{z}\right) \right] \begin{pmatrix} z^n & 0 & 0 \\ 0 & z^{-n/2} & 0 \\ 0 & 0 & z^{-n/2} \end{pmatrix}, \qquad |z| \to \infty. \tag{6.5}$$

We take n to be even here, for simplicity; this is not an essential assumption. The relation to the family of biorthogonal polynomials (6.2) is that the solution to the above problem has 1-1 entry $\mathbf{Y}_{11}(z) = p_n(z)$.

From here, one should proceed to the steepest descent analysis performed in Chapter 5. In the interest of brevity of this thesis, we shall not perform the full calculation here, instead opting to only discuss the spectral curve. This is by far the most difficult step of the analysis; in fact, once this calculation is in place, the rest of the steepest descent analysis is (more or less) identical to that of the previous chapter. The only substantial difference is that one must use the Airy function in the first transformation, instead of the Pearcey-type integrals. We shall see that the cubic 2-matrix model admits a multicritical phenomenon identical to that of the quartic 2-matrix model. We hope to complete the full analysis of the cubic 2-matrix model in a later work.

6.1.1 Analysis of the Spectral Curve.

We must search for a solution to the stationarity equations

$$X + tX^2 - \frac{1}{X} - \tau Y = \mathcal{O}(X^{-2}), \tag{6.6}$$

$$Y + tY^2 - \frac{1}{Y} - \tau X = \mathcal{O}(Y^{-2}). \tag{6.7}$$

in terms of rational functions (the genus 0 ansatz). We have the following proposition:

Proposition 6.2. *A 2-parameter family of solutions to the stationarity equations (6.6),(6.7) is given by the rational functions*

$$X(u) = A\int^u \frac{(u-a)(u-b)(u+a+b)}{u^3}du + K = A\left(u + \frac{a^2+ab+b^2}{u} - \frac{ab(a+b)}{2u^2}\right) + K, \tag{6.8}$$

$$Y(u) = X(u^{-1}) = A\left(\frac{1}{u} + (a^2+ab+b^2)u - \frac{ab(a+b)}{2}u^2\right) + K, \tag{6.9}$$

with A, t, τ, K defined parametrically in terms of a and b as:

$$t = t(a,b) = -\frac{ab(a+b)\left(\Delta - \frac{1}{2}(a^2+ab+b^2)\lambda - \sqrt{\Delta}\right)\sqrt{2\sigma(1+\sqrt{\Delta})}}{2(a^2+ab+b^2)^2\lambda^2}, \tag{6.10}$$

$$\tau = \tau(a,b) = \frac{\sqrt{\Delta}-1}{\lambda(a^2+ab+b^2)}, \tag{6.11}$$

$$A = A(a,b) = \sqrt{\frac{2(1+\Delta)}{\sigma}}, \tag{6.12}$$

$$K = K(a,b) = -\frac{(a^2+ab+b^2-\sqrt{\Delta}-1)\sqrt{2\sigma(1+\sqrt{\Delta})}}{\sigma(a+b)ab}, \tag{6.13}$$

where the parameters Δ, σ, λ are defined as

$$\Delta = 2a^6b^2 + 6a^5b^3 + 8a^4b^4 + 6a^3b^5 + 2a^2b^6 + a^4 + 2a^3b + 3a^2b^2 + 2ab^3 + b^4 - 2a^2 - 2ab - 2b^2 + 1, \tag{6.14}$$

$$\sigma = a^4b^2 + 2a^3b^3 + a^2b^4 + 2a^4 + 4a^3b + 6a^2b^2 + 4b^3a + 2b^4 - 2, \tag{6.15}$$

$$\lambda = 2a^4b^2 + 4a^3b^3 + 2a^2b^4 + a^2 + ab + b^2 - 2. \tag{6.16}$$

Proof. The proof of this proposition is identical to its analog in Chapter 5, so we do not present it in full here. We only remark that, after some tedious calculation, one can check that

$$t(a,b) < 0, \qquad \tau(a,b) > 0, \qquad A(a,b) > 0, \qquad K(a,b) < 0, \tag{6.17}$$

provided $0 < b < 1 < a < \frac{1}{b}$. For example, one can show $A > 0$ by first demonstrating that $\Delta, \sigma > 0$. This follows immediately from the direct application of the inequality $a > 1$ to Δ, σ:

$$\Delta > 2b^2 + 6b^5 + 9b^4 + 8b^3 + 3b^2 > 0,$$
$$\sigma > 3b^4 + 6b^3 + 7b^2 + 4b > 0.$$

Proofs of the remaining formulas and inequalities are straightforward, though tedious. □

We stress that the long formulas given above are not so important; the key point is that the spectral curve is still rationally parameterized, in a form very similar to the quartic case. Most importantly, the branching structure of the spectral curve is closely related, and is independent of the functions A, K, t, τ.

So far, what we have obtained is a 2-parameter family of solutions to the stationarity equations, up to terms of order $\mathcal{O}(X^{-2})$. Each fixed pair (a, b) parameterizes a Riemann surface. Similarly to the quartic case, we define some special curves in the (τ, t)-plane, as well as a multicritical point. We define the critical curves as follows:

1. *Low temperature critical curve.* The low temperature critical curve is defined to be the parametric curve $(\tau(s), t(s))$, $0 \le s \le 1$, where, setting $\psi(s) = \sqrt{3s^8 + 6s^6 + 10s^4 + 6s^2 + 3}$,

$$\tau(s) = \frac{(-s^2 + \psi(s))s^2}{3(s^4 + s^2 + 1)^2}, \tag{6.18}$$
$$t(s) = \frac{\sqrt{4s^4 + 2s^2 + 4}\sqrt{s^2 + \psi(s)}\left[s^8 + 2s^6 + \frac{11}{3}s^4 + 2s^2 + 1 - \frac{2}{3}s^2\psi(s)\right](s^2 + 1)^2}{12(s^4 + s^2 + 1)^4}. \tag{6.19}$$

2. *High temperature critical curve.* The high temperature critical curve is defined to be the parametric curve $(\tau(s), t(s))$, $0 \le s \le 1$, where, setting $\phi(s) = (s^2 + s)\sqrt{2s^2 + 2s + 3}$,

$$\tau(s) = \frac{\phi(s) - 1}{(s^2 + s + 1)^2(2s^2 + 2s - 1)}, \tag{6.20}$$
$$t(s) = \frac{\sqrt{6s^3 + 12s^2 + 14s + 8}\sqrt{1 + \phi(s)}\left[s^6 + 3s^5 + \frac{9}{2}s^4 + 4s^3 + \frac{3}{2}s^2 + \frac{1}{2} - \phi(s)\right](s + 1)s^{3/2}}{2(s^2 + s + 1)^4(2s^2 + 2s - 1)^2} \tag{6.21}$$

3. *The multicritical point.* Finally, we define the multicritical point as the $s \to 1$ limit of either of the above parametric curves, which turns out to be

$$\tau_c = \frac{2\sqrt{7} - 1}{27}, \qquad t_c = \frac{(4\sqrt{7} - 29)\sqrt{10 + 20\sqrt{7}}}{729}. \tag{6.22}$$

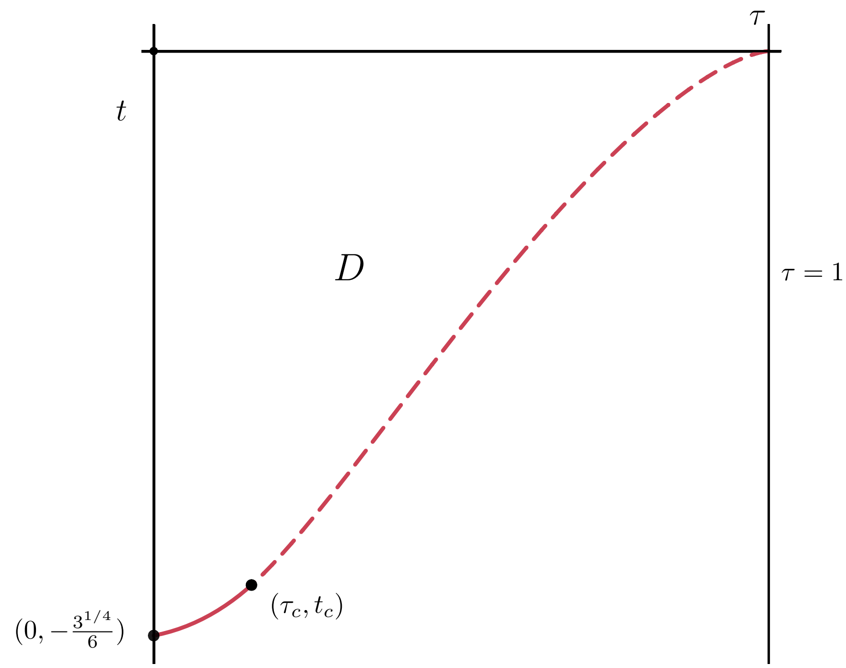


Figure 6.2. The phase portrait for the 2-matrix model with cubic interactions in the (τ, t) plane. The low-temperature critical curve is represented by the solid red line; the high-temperature critical curve is represented by the dashed red line. Note that the critical curve intersects the $\tau = 0$ axis precisely at the critical point of the cubic 1-matrix model; this is in agreement with the intuition that the 2-matrix model decouples at 0 temperature.

4. *The region* D. Finally, we define the region D as the region of the phase plane bounded by the $t = 0$ and $\tau = 0$ axes, and the critical curves.

The critical curves, multicritical point, and region D for the cubic 2-matrix model is are shown in Figure (6.2). Let us make a few comments about the above curves before proceeding to the next theorem. First, although the form of the above equations is rather complicated, in computations, the explicit form of these curves is never needed. We only present the curves in explicit form here for completeness [1]. Moreover, the critical curves have the following special properties:

- As $s \to 0$ on the low temperature critical curve, $\tau(s) \to 0$, and $t(s) \to -\frac{3^{1/4}}{6}$. This is to be compared with the critical point of the cubic 1-matrix model, cf. [19], Equation 52 (in their normalization, the critical point is 1/3 times this one), or [10, 11] (again with the normalization of [19]).
- The critical temperature at the multicritical point, τ_c, is first described in the seminal work of Boulatov and Kazakov [17] (see Equation 47). However, they do not explicitly express the critical value of t where the multicritical point occurs.

[1]We also present them here because their form seems to never have been written down in the literature; perhaps this is for the obvious reason that they are simply too long.

- One can also calculate the slope of critical curve at the multicritical point. As was the case with the quartic 2-matrix model, the expressions for the high-temperature and low-temperature critical curves meet at the multicritical point (τ_c, t_c), and agree up to 3^{rd} order in their Taylor expansions there. For later reference, we record the slope of the tangent line here:

$$\frac{t'_-(1)}{\tau'_-(1)} = -\frac{\sqrt{10+20\sqrt{7}}\left(-5+\sqrt{7}\right)}{90} = \frac{t'_+(1)}{\tau'_+(1)}, \tag{6.23}$$

  where $\pm$ denote the limits as $s \to 1$ along the high-temperature (respectively, low temperature) critical curves.

We now state a proposition, which characterizes the critical curves (and multicritical point) in terms of the positions of the branch points (a, b) in the uniformizing plane.

Proposition 6.3. *Let*

$$R := \{(a, b) \mid 0 < b \leq 1, 1 \leq a \leq b^{-1}\}. \tag{6.24}$$

Then, there is a bijection between the region R and the region D of the phase plane, induced by the mapping $(a, b) \mapsto (\tau(a, b), t(a, b))$:

$$D = \{(\tau(a, b), t(a, b)) \mid (a, b) \in R\}. \tag{6.25}$$

Furthermore, we have the following identifications:

1. *The low-temperature critical curve given by the boundary component $a = b^{-1}$,*
2. *The high-temperature critical curve given by the boundary component $a = 1$,*
3. *The multicritical point given by $a = b = 1$.*

Proof. We omit the proof of this proposition, for brevity. We only remark that the identifications of the low temperature and high temperature critical curves follows directly from their definition. □

The associated Riemann surface given by the parameterization $(X(u), Y(u))$ is called the *spectral curve.* Let us study the structure of the spectral curve; we shall treat the parametric curve (X, Y) as a branched covering of the sphere over the X-coordinate; by construction, the X-coordinate has branch points ($X'(u) = 0$) at $u = a$, b, $-(a + b)$, and ∞. For later convenience, we relabel $X(u) =: z(u)$, as it is more consistent with the Riemann-Hilbert problem at hand.

Away from the multicritical point, the spectral curve is 3-sheeted; this family of spectral curves have generically the same structure, and are shown in Figure (6.3). There are 3 finite branch points, all of which

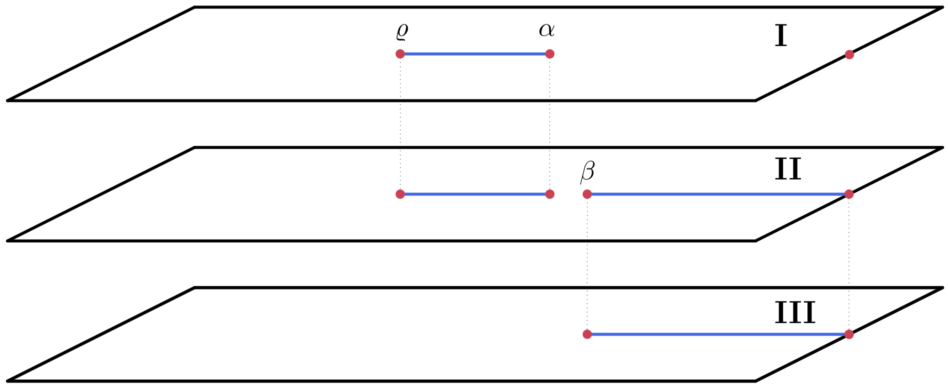


Figure 6.3. A representative example of a critical/generic surface in the physical plane; the sheets are labelled I,II,III. If the curve is low-temperature critical, an extra zero of $\Omega(z)$ accumulates at $z = \alpha$ and $z = \beta$; if the curve is high-temperature critical, an extra zero accumulates at $z = \alpha$.

lie on the real axis: at $\alpha := z(a)$, $\beta := z(b)$, and $\varrho := z(-(a+b))$. We have the inequalities

$$-\infty < \varrho < 0 < \alpha < \beta < \infty. \tag{6.26}$$

The structure of the curve is as follows: Sheets 1 and 2 are glued along the interval $[\varrho, \alpha]$, and sheets 2 and 3 are glued along the interval $[\beta, \infty)$. The spectral curve for generic values of the parameters is shown in Figure (6.3). At the multicritical point $a = b = 1$, the curve further degenerates, and the branch point $\beta \to \alpha$. The multicritical spectral curve is shown in Figure (6.4). In this case, sheets 1 and 2 are glued along the interval $[\varrho, \alpha]$, and sheet 2 is glued to sheet 3 along the interval $[\alpha, \infty)$. The spectral curve is alternatively shown at and away from the multicritical point in Figure (6.5). We define the uniformizing coordinates $u_j(z)$ as the map from the j^{th} sheet of the spectral curve in the z-coordinate, so that $z(u_j(z))$ is the identity map on the j^{th} sheet. We then define the function $\Omega(u)$:

$$\Omega(u) = \int Y(u)X'(u)du, \tag{6.27}$$

and set $\Omega_j(z) := \Omega(u_j(z))$, $j = 1, 2, 3$.

For the values of the parameters parametrizing the region D, we can prove that lensing is again possible. Along with some local calculations, this amounts to showing that the sign of the g-function doesn't change on the branch cuts; equivalently, that the curves $\operatorname{Im} z(u)$ and $\operatorname{Im} Y(u)$ intersect at most at the branch points, analogously to the quartic 2-matrix model. We have the following proposition.

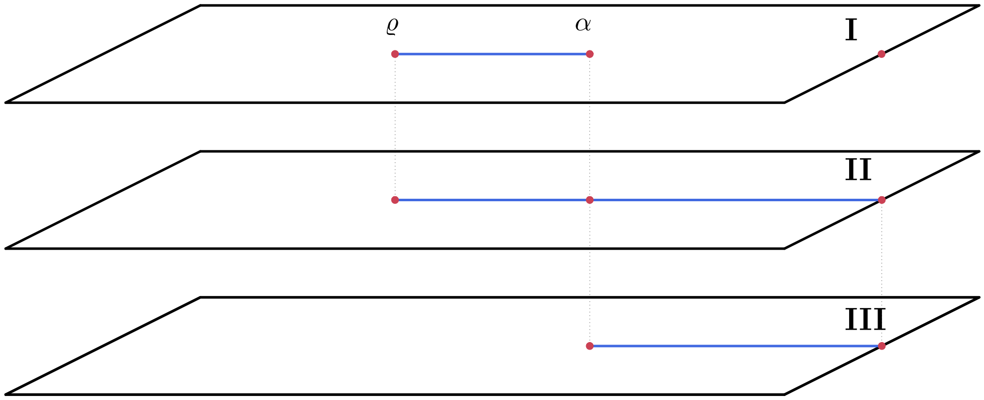


Figure 6.4. The critical spectral curve. Note that the branch points at $z = \alpha$, $z = \beta$ have merged.

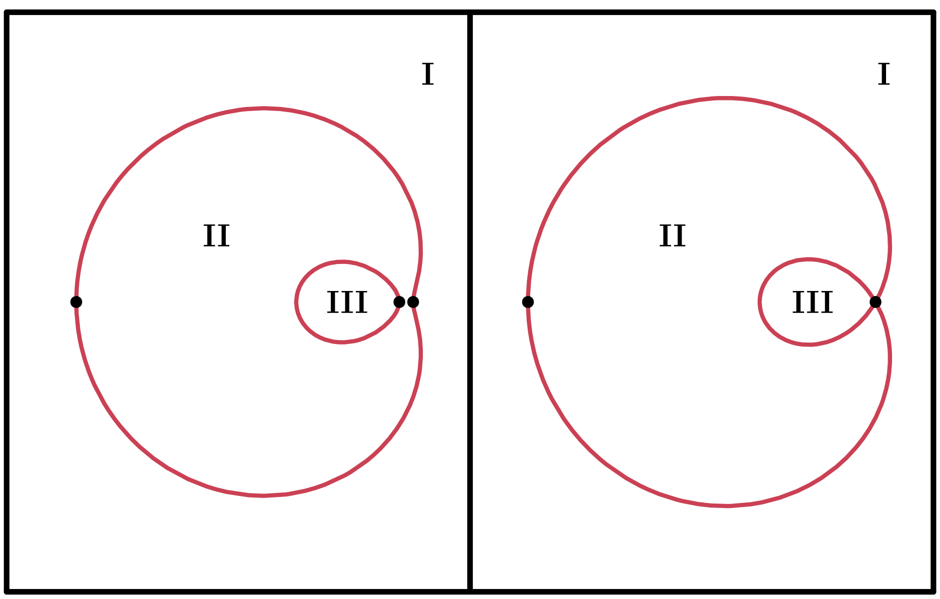


Figure 6.5. The branch cuts in the uniformizing plane $u = x + iy$ away from the multicritical point (left) and at the multicritical point. The noncritical case has $a = 1.009, b = 0.890$. The images of each sheet under the uniformizing map are labelled I,II,III. Note that, at criticality, the branch points at b, c merge; this is consistent with Figures (6.3), (6.4).

Lemma 6.4. *Consider the normal vector to the preimages of the branch cuts in the uniformizing plane under the mapping $z(u)$:*

$$\hat{\mathbf{n}} := \frac{\nabla \, Im \, z(u)}{||\nabla \, Im \, z(u)||}. \tag{6.28}$$

For any $(a, b) \in R = \{(a, b) \mid 0 < b \leq 1, 1 \leq a \leq b^{-1}\}$, the function

$$\nabla \, Re \, \Omega(u) \cdot \hat{\mathbf{n}} = \frac{\partial}{\partial n} \, Re \, \Omega(u) \tag{6.29}$$

is of constant sign on each connected component of the preimages of the branch cuts.

Proof. The proof of this lemma, as was explained in Chapter 5, Lemma (5.5), reduces to showing that the curves Im $z(u) = 0$, Im $Y(u) = 0$ intersect at most at the branch points. We sketch how the rest of the proof should go; we leave the full calculation for a later work. One can parameterize these curves in Cartesian coordinates $u = x + iy$ explicitly. Setting $\varpi := a^2 + ab + b^2$, we have that

$$\text{Im } z(u) = 0 \Leftrightarrow y = \pm\sqrt{\frac{ab(a+b)x^3 - \varpi x^2 + 1}{\varpi - ab(a+b)x}},$$
$$\text{Im } Y(u) = 0 \Leftrightarrow y = \pm\sqrt{\frac{1}{2}\varpi - x^2 \pm \frac{1}{2}\sqrt{\varpi^2 - 4ab(a+b)x}}.$$

One can show that the only solutions to the equations defining the intersections of Im $z(u) = 0$ and Im $Y(u) = 0$ are

$$x = \frac{\varpi - 1}{ab(a+b)}, \qquad \frac{\varpi + 1 \pm \sqrt{(\varpi+1)(\varpi-3)}}{2ab(a+b)}.$$

One can then show that the corresponding solutions for y are necessarily imaginary; thus, for real x, y, the curves do not intersect. □

With this lemma in place, all that is left is to check is that the g-function has the right sign where in lens-shaped regions about the branch cuts. This involves a straightforward calculation of the series expansions of the $\Omega_j(z)$ about the branch points, as was the case in Chapter 5. For brevity, we do not present these expansions here, but mention some of the properties one should expect upon performing these expansions.

1. Away from the critical curves ($0 < b < 1 < a < \frac{1}{b}$), the expansions of $\Omega_1(z)$, $\Omega_2(z)$ at $z = \varrho, \alpha$ are singular, and with the first singular term being $(z - \rho)^{3/2}$ (respectively, $(z - \alpha)^{3/2}$). This indicates that the local parametrix there will be of Airy type. At $z = \beta$, the expansions of $\Omega_2(z)$, $\Omega_3(z)$ are singular, with the first singular term being $(z - \beta)^{3/2}$. Thus, the parametrix there is also of Airy type.

2. On the low-temperature critical curve ($0 < b < 1$, $a = \frac{1}{b}$), the expansions of $\Omega_1(z)$, $\Omega_2(z)$ at $z = \varrho$ are singular, with the first singular term being $(z-\rho)^{3/2}$, indicating the local parametrix there is of Airy type. However, at $z = \alpha$, $\Omega_1(z)$, $\Omega_2(z)$ have an extra zero, and the first singular term appearing in the expansion of these functions there is $(z-\alpha)^{5/2}$. This is indicative of a Painlevé I-type parametrix. Similarly, at $z = \beta$, the expansions of $\Omega_2(z)$, $\Omega_3(z)$ are singular, and an extra zero accumulates there: the first singular term appearing in the expansion of these functions is $(z-\alpha)^{5/2}$, again indicating the appropriate local parametrix is of Painlevé I type.

3. On the high-temperature critical curve, ($0 < b < 1$, $a = 1$), the expansions of $\Omega_1(z)$, $\Omega_2(z)$ at $z = \varrho$ are singular, with the first singular term being $(z-\rho)^{3/2}$, indicating the local parametrix there is of Airy type. At the $z = \alpha$, however, $\Omega_1(z)$, $\Omega_2(z)$ have an extra zero, and the first singular term appearing in the expansion of these functions there is $(z-\alpha)^{5/2}$. This is indicative of a Painlevé I-type parametrix. At $z = \beta$, the expansions of $\Omega_2(z)$, $\Omega_3(z)$ are singular, with the first singular term being $(z-\beta)^{3/2}$. Thus, the parametrix there is also of Airy type.

4. At the multicritical point $a = b = 1$, the expansions of $\Omega_1(z)$, $\Omega_2(z)$ at $z = \varrho$ are singular, with the first singular term being $(z-\rho)^{3/2}$, indicating the local parametrix there is of Airy type. On the other hand, the branch points at $z = \alpha$, $z = \beta$ have merged, resulting in a branch point of order 2. At this point, the first singular term in the expansions of $\Omega_1(z)$, $\Omega_2(z)$, and $\Omega_3(z)$ is $(z-\alpha)^{7/3}$. This is an identical phenomenon to what occurred in the quartic 2-matrix model, and the same parametrix we used there (involving a special solution of the KdV_3 string equation) applies here.

The above remarks should be compared to the situation of the quartic 2-matrix model. The positions of the branch points in the uniformizing plane seem to describe the transitions from the generic situation to the different kinds of criticality in an identical manner. This suggests that these branch points might be good local coordinates when one attempts to investigate higher-order critical phenomenon. We reserve discussions of this, and complete proofs of all of the above remarks, for a later work.

6.2 Further Open Problems.

Here, we present a number of other problems which we have not fully addressed in this work, and plan to make part of a later research program.

- *Analysis of the Local Parametrix.* The model Riemann-Hilbert problem we introduced in Chapter 5 appearing in the multicritical parametrices is stated, but we provide no analysis of the existence of the solution, nor any asymptotic analysis of the solution. The first point is obviously important, and will be the subject of a future work. We hope to also investigate the second point (the asymptotic analysis of the Riemann-Hilbert problem), as this analysis would in turn inform us about the structure of the special solution to the KdV_3 string equation which appears in the expression for the multicritical partition function. This would be the 2-matrix analog of the works [66, 67].

- *Higher-genus criticality.* We have only found the partition function in part of the region of the phase plane where the spectral curve is of genus zero. In analogy to the 1-matrix model, the spectral curve accumulates extra zeros at the branch points at criticality, and if we were to move below the critical curves in the phase plane, these extra zeros would split off, and form their own branch points (and new branch cuts/supports of the measure). For the quartic 2-matrix model, directly below the low-temperature critical curve, the spectral curve immediately transitions to a curve of genus 4; directly below the high-temperature critical curve, the spectral curve immediately transitions to a curve of genus 2. This seems to imply the existence of another critical curve connecting these two regions of the phase plane. As far as calculations go, the quartic case seems rather inaccessible, because the genus is too large to say anything.

  However, the situation seems to be more manageable in the cubic case: directly below the low-temperature curve, the spectral curve transitions to a curve of genus 2, and directly below the high-temperature curve, the spectral curve transitions to a curve of genus 1. This is likely more tractable, and in fact, we can write the following ansatz for a solution to the stationarity equations (6.6), (6.7):

$$X(u) = A\wp(u;q) + B\zeta(u;q) - B\zeta\left(u - \frac{1+q}{2};q\right) + C, \tag{6.30}$$

$$Y(u) = X(1/u), \tag{6.31}$$

  for some undetermined constants A, B, C, q, and where $\wp(u;q)$ is the Weierstrass-$\wp$ function, with elliptic modulus q, and $\zeta(u;q)$ is the Weierstrass ζ-function (again with elliptic modulus q). This choice of parametrization is consistent with the expected branching structure of the spectral curve, and is the direct analog of the genus 0 ansatz given before, in the genus 1 case. We remark that V. Kazakov and A. Marshakov have investigated the physical interpretations of this higher genus criticality [71], but very little work in this direction for the 2-matrix model has been performed. This kind of problem is of interest in the approximation theory community [2, 79], as the problem of describing the asymptotics of multiple orthogonal polynomials in the situation where the spectral curve is of higher genus is still a very active area of research. It is also possible that uniformization is *not* the correct approach. Very similar models of multiple orthogonal polynomials on higher- genus Riemann surfaces

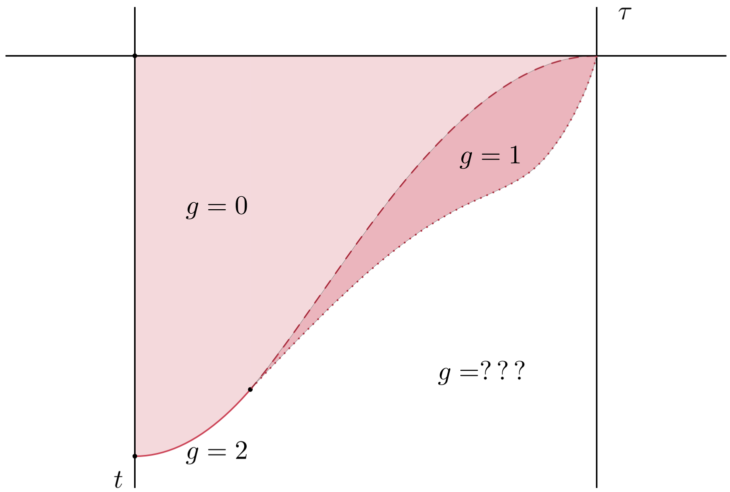


Figure 6.6. A conjecture of the full phase portrait for the cubic 2-matrix model should look like. There is a second high-temperature critical curve, coming from the transition of genus between the genus 1 and genus 2 regions of the phase plane.

have been investigated before in the literature [3, 79], using a parameterization of the spectral curve as a polynomial in X, Y:

$$P(X,Y) = X^3 - \Pi_2(Y)X^2 - \Pi_1(Y)X + \Pi_0(Y), \tag{6.32}$$

where $\Pi_i(Y)$ are polynomials in Y, $i = 0, 1, 2$. In fact, one can show that we have a similar parameterization of the spectral curve of the cubic 2-matrix model, with the $\Pi_i(Y)$ given by

$$\Pi_2(Y) = \frac{t}{\tau}Y^2 + \frac{1}{\tau}Y - \frac{1}{t}, \tag{6.33}$$

$$\Pi_1(Y) = \frac{1}{\tau}Y^2 + \frac{1}{t\tau^2}(\tau^3 - t^2 + \tau)Y + \alpha_0(t,\tau), \tag{6.34}$$

$$\Pi_0(Y) = Y^3 + \frac{1}{t}Y^2 + \alpha_0(t,\tau)Y + \alpha_1(t,\tau), \tag{6.35}$$

for parameters $\alpha_i(t,\tau)$ chosen so that the stationarity equations (6.6), (6.7) hold. In general, these functions have complicated dependence on t, τ. This could give an alternate approach to dealing with the cubic 2-matrix model.

- *The (p,q) Minimal Models Coupled to Gravity.* As was mentioned in the introduction, it is widely believed that all of the (p,q) minimal models coupled to gravity are incorporated by the 2-matrix model. In this sense, the 2-matrix model is a universal model for these kinds of phase transitions. This being said, very little mathematically rigorous work has been done on *any* critical phenomenon in the 2-matrix model, aside from the $(4,3)$ (Ising) minimal model studied in this work. This is one of the most interesting further directions of research. Some foundational work on the theoretical physics side has already been examined, and in particular, various physicists have identified precisely where the $(3,5)$, $(5,6)$, $(4,5)$, and $(3,8)$ minimal models coupled to gravity occur the 2-matrix model [100, 101]. More generally, J.-M. Daul, V.A. Kazakov, and I.K. Kostov claim to have identified all of the minimal models coupled to gravity within the 2-matrix model [26]. It is important to try and establish their claims rigorously.

- *The Genus Expansion.* It is well known that the free energy of the 1-matrix model admits a *genus expansion* in powers of $1/N^2$; the coefficients of this expansion count the number of connected ribbon graphs of a given genus [43, 44]. Similarly, as discussed in Chapter 3, the partition function for the 2-matrix model should admit an analogous genus expansion, with the coefficients counting the number of 2-colored, connected, genus g ribbon graphs. The genus expansion for the quartic 2-matrix model is in reach from the calculations of Chapter 5; we do not prove the existence of the expansion there, however, as the current work would *really* become too long. However, this expansion is still of interest, and is something that we would like to pursue in a future work.

- *Multiplicative Chaos, Log-Correlated Fields, and Liouville Gravity.* For some time now, there has been a large group of probabilists working on the same collection of conjectures related to 2D critical phenomena coupled to Liouville quantum gravity (cf. [82, 88, 95], for example). The problems they address are very closely related to many of the problems outlined in this thesis, such as the KPZ formula describing the change in critical exponents of the minimal models coupled to gravity. It would be of interest to see if the results found in this community have anything to say about critical phenomenon in the 2-matrix model, and vice-versa.

6.3 Conclusion.

To summarize, in this thesis, we have studied the first physically relevant critical phenomenon in the 2-matrix model, corresponding to the Ising model coupled to 2-dimensional gravity. We also calculated the genus-0 partition function of the quartic 2-matrix model, making the original results of Kazakov [70] fully rigorous. This problem is the first step in a larger program on critical phenomenon in the 2-matrix model. We hope to extend these results by studying the cubic 2-matrix model, and investigate the open problems described in Section (6.2).

REFERENCES

[1] A.I. Aptekarev. “Multiple Orthogonal Polynomials”. In: *J. Comput. Appl. Math.* 99.1-2 (1998), pp. 423–447.

[2] A.I. Aptekarev, A.B.J. Kuijlaars, and W. Van Assche. “Asymptotics of Hermite–Padé Rational Approximants for Two Analytic Functions with Separated Pairs of Branch Points (Case of Genus 0)”. In: *Int. Math. Res. Pap.* rpm007 (2008).

[3] A.I. Aptekarev, D.N. Toulyakov, and W. Van Assche. “Hyperelliptic uniformization of algebraic curves of the third order”. In: *J. Comp. Appl. Math.* 284 (2015), pp. 38–49.

[4] W. Van Assche and E. Coussement. “Some classical multiple orthogonal polynomials”. In: *J. Comput. Appl. Math.* 127 (2001), pp. 317–347.

[5] W. Van Assche, J.S. Geronimo, and A.B.J. Kuijlaars. “Riemann-Hilbert Problems for Multiple Orthogonal Polynomials”. In: *Special Functions 2000: Current Perspective and Future Directions*. Ed. by J. Bustoz, M.E.H. Ismail, and S.K. Suslov. Springer, 2001, pp. 33–88.

[6] E. Basor. “A localization theorem for Toeplitz determinants”. In: *Indiana Univ. Math. J.* 28.6 (1979), pp. 975–983.

[7] R.J. Baxter. *Exactly Solved Models in Statistical Mechanics*. 3rd ed. Academic Press Limited, 1989.

[8] A.A. Belavin, A.M. Polyakov, and A.B. Zamolodchikov. “Infinite conformal symmetry in two-dimensional quantum field theory”. In: *Nucl. Phys. B* 241.2 (1984), pp. 333–380.

[9] M. Bertola and O. Marchal. “The partition function of the two-matrix model as an isomonodromic τ function”. In: *J. Math. Phys.* 50.1 (2009), p. 013529.

[10] P. Bleher and A. Deaño. “Painlevé I Double Scaling Limit in the Cubic Random Matrix Model”. In: *Random Matrices Theory Appl.* 5.2 (2016).

[11] P. Bleher, A. Deaño, and M. Yattselev. “Topological Expansion in the Complex Cubic Log–Gas Model: One-Cut Case”. In: *J. Stat. Phys.* 166 (2017), pp. 784–827.

[12] P. Bleher, R. Gharakhloo, and K.T-R. McLaughlin. “Phase Diagram and Topological Expansion in the Complex Quartic Random Matrix Model”. In: *arXiv:2112.09412* (2022).

[13] P. Bleher and A. Its. “Semiclassical asymptotics of orthogonal polynomials, Riemann–Hilbert problem, and universality in the matrix model,” in: *Ann. Math.* 150 (1999), pp. 185–266.

[14] P. Bleher and A.B.J. Kuijlaars. “Large n limit of Gaussian random matrices with external source, part III: double scaling limit”. In: *Comm. Math. Phys.* 270 (2007), pp. 481–517.

[15] B. Bollobás. *Modern Graph Theory.* Graduate Texts in Mathematics 184. Springer, 1998.

[16] A. Böttcher and B. Silbermann. “Toeplitz matrices and determinants with Fisher-Hartwig symbols”. In: *J. Funct. Anal.* 63 (1985), pp. 178–214.

[17] D.V. Boulatov and V.A. Kazakov. “The Ising Model on a Random Planar Lattice: The Structure of the Phase Transition and the Exact Critical Exponents”. In: *Phys. Lett. B* 186.3 (1986), pp. 379–384.

[18] E. Brézin, M.R. Douglas, V.A. Kazakov, and S.H. Shenker. “The Ising model coupled to 2D gravity. A nonperturbative analysis”. In: *Phys. Lett. B* 237.1 (1990), pp. 43–46.

[19] E. Brézin, C. Itzykson, G. Parisi, and J.B. Zuber. “Planar diagrams”. In: *Comm. Math. Phys.* 59.1 (1978), pp. 35–51.

[20] E. Brézin and V.A. Kazakov. “Exactly solvable field theories of closed strings”. In: *Phys. Lett. B.* 236.2 (1990), pp. 144–150.

[21] M.P. do Carmo. *Differential Geometry of Curves and Surfaces.* Prentice-Hall, 1976.

[22] M.P. do Carmo. *Riemannian Geometry.* Birkhäuser, 1992.

[23] R.R. Coifman, A. McIntosh, and Y. Meyer. “L’intégrale de Cauchy définit un opérateur borné sur L^2 pur les courves lipschitziennes”. In: *Ann. Math.* 116 (1982), pp. 361–387.

[24] G.A.T.F. da Costa and A. L. Maciel. “Combinatorial Formulation of Ising Model Revisited”. In: *Rev. Bras. Ens. Fis.* 25.1 (2003), pp. 49–59.

[25] C. Crnković, P. Ginsparg, and G. Moore. “The Ising model, the Yang-Lee edge singularity, and 2D quantum gravity”. In: *Phys. Lett. B* 237.2 (1990), pp. 196–201.

[26] J.-M. Daul, V.A. Kazakov, and I.K. Kostov. “Rational Theories of 2D Gravity from the Two-Matrix Model”. In: *Nucl. Phys. B* 409 (1993), pp. 311–338.

[27] P. Deift. *Orthogonal Polynomials and Random Matrices: A Riemann-Hilbert Approach.* Amer. Math. Soc., 2000.

[28] P. Deift, A. Its, and I. Krasovsky. “Asymptotics of Toeplitz, Hankel, and Toeplitz+Hankel determinants with Fisher-Hartwig singularities”. In: *Ann. of Math.* 174 (2011), pp. 1243–1299.

[29] P. Deift, T. Kriecherbauer, K.T-R. McLaughlin, S. Venakides, and X. Zhou. "Strong asymptotics of orthogonal polynomials with respect to exponential weights". In: *Comm. Pure Appl. Math.* 52 (1999), pp. 1491–1552.

[30] P. Deift, T. Kriecherbauer, K.T-R. McLaughlin, S. Venakides, and X. Zhou. "Uniform asymptotics for polynomials orthogonal with respect to varying exponential weights and applications to universality questions in random matrix theory". In: *Comm. Pure Appl. Math.* 52 (1999), pp. 1335–1425.

[31] P. Deift and X. Zhou. "A steepest descent method for oscillatory Riemann-Hilbert problems. Asymptotics for the MKdV equation". In: *Ann. of Math.* 137.2 (1993), pp. 295–368.

[32] S. Delvaux, A.B.J. Kuijlaars, and L. Zhang. "Critical behavior of non-intersecting Brownian motions at a tacnode". In: *Comm. Pure and Appl. Math* 64 (2011), pp. 1305–1383.

[33] L.A. Dickey. *Soliton Equations and Hamiltonian Systems.* 2nd ed. Vol. 26. World Scientific, 2003.

[34] M.R. Douglas. "Strings in Less Than One Dimension and Generalized KdV Hierarchies". In: *Phys. Lett. B* 238.2–4 (1990), pp. 176–180.

[35] M.R. Douglas and S.H. Shenker. "Strings in less than one dimension". In: *Nucl. Phys. B.* 335.3 (1990), pp. 635–654.

[36] M. Duits and D. Geudens. "A critical phenomenon in the two-matrix model in the quartic/quadratic case". In: *Duke Math. J.* 162.8 (2013), pp. 1383–1462.

[37] M. Duits, D. Geudens, and A.B.J. Kuijlaars. "A vector equilibrium problem for the two-matrix model in the quartic/quadratic case". In: *Nonlinearity* 24 (2011), pp. 951–993.

[38] M. Duits and A.B.J. Kuijlaars. "Painlevé I asymptotics for orthogonal polynomials with respect to a varying quartic weight". In: *Nonlinearity* 19 (2006), pp. 2211–2245.

[39] M. Duits and A.B.J. Kuijlaars. "Universality in the Two-Matrix Model: A Riemann-Hilbert Steepest Descent Analysis". In: *Comm. Pure Appl. Math.* 62 (2009), pp. 1076–1153.

[40] M. Duits, A.B.J. Kuijlaars, and M.-Y. Mo. "The Hermitian two matrix model with an even quartic potential". In: *Memoirs of the American Mathematical Society* 217.1022 (2012).

[41] P. Ehrenfest. "Phasenumwandlungen im ueblichen und erweiterten Sinn, classifiziert nach den entsprechenden Singularitaeten des thermodynamischen Potentiales". In: *Proc. Royal Acad. Amsterdam* 36 (1933), pp. 153–157.

[42] T. Ehrhardt. "A status report on the asymptotic behavior of Toeplitz determinants with Fisher-Hartwig singularities". In: *Operator Theory: Adv. Appl.* 124 (2001), pp. 217–241.

[43] N.M. Ercolani and K.T-R. McLaughlin. “Asymptotics and integrable structures for biorthogonal polynomials associated to a random two-matrix model”. In: *Physica D* 152-153 (2001), pp. 232–268.

[44] N.M. Ercolani and K.T-R. McLaughlin. “Asymptotics of the partition function for random matrices via Riemann-Hilbert techniques and applications to graphical enumeration”. In: *Int. Math. Res. Not.* 14 (2003), pp. 755–820.

[45] N.M. Ercolani, K.T-R. McLaughlin, and V.U. Pierce. “Random Matrices, Graphical Enumeration and the Continuum Limit of Toda Lattices”. In: *Comm. Math. Phys.* 278 (2008), pp. 31–81.

[46] M. E. Fisher and R. E. Hartwig. “Toeplitz determinants, some applications, theorems and conjectures”. In: *Adv. Chem. Phys.* 15 (1968), pp. 333–353.

[47] A.S. Fokas, A.R. Its, and A.V. Kitaev. “Discrete Painlevé equations and their appearance in quantum gravity”. In: *Comm. Math. Phys.* 142 (1991), pp. 313–344.

[48] A.S. Fokas, A.R. Its, and A.V. Kitaev. “The isomonodromy approach to matrix models in 2D quantum gravity”. In: *Comm. Math. Phys.* 147 (1992), pp. 395–430.

[49] P. Di Francesco. “2D Quantum Gravity, Matrix Models, and Graph Combinatorics”. In: *Applications of Random Matrices in Physics*. Ed. by É. Brézin, V. Kazakov, D. Serban, P. Wiegmann, and A. Zabrodin. Springer, 2006, pp. 33–88.

[50] P. Di Francesco, P. Ginsparg, and J. Zinn-Justin. “2D Gravity and Random Matrices”. In: *arXiv:hep-th/9306153* (1994).

[51] P. Di Francesco, P. Mathieu, and D. Sénéchal. *Conformal Field Theory*. Graduate Texts in Contemporary Physics. Springer, 1997.

[52] E. Frenkel and D. Ben-Zvi. *Vertex Algebras and Algebraic Curves*. Mathematical Surveys and Monographs 88. AMS, 2004.

[53] F.D. Gakhov and I.N. Sneddon. *Boundary Value Problems*. Pergamon Press, 1966.

[54] P. Ginsparg. “Applied Conformal Field Theory”. In: *arXiv:hep-th/9108028* (1988).

[55] P. Ginsparg. “Matrix Models of 2D Gravity”. In: *arXiv:hep-th/9112013* (1991).

[56] P. Ginsparg, M. Goulian, M.R. Plesser, and J. Zinn-Justin. “(p, q) String Actions”. In: *Nucl. Phys. B* 342 (1990), pp. 539–563.

[57] P. Ginsparg and G. Moore. “Lectures on 2D gravity and 2D string theory (TASI 1992)”. In: *arXiv:hep-th/9304011* (1993).

[58] D.J. Gross and A.A. Migdal. "A Non-Perturbative Treatment of Two-Dimensional Quantum Gravity". In: *Nucl. Phys. B* 340.2-3 (1990), pp. 333–365.

[59] D.J. Gross and A.A. Migdal. "Nonperturbative two-dimensional quantum gravity". In: *Phys. Rev. Lett.* 64.2 (1990), pp. 127–130.

[60] E. Ising. "Contribution to the Theory of Ferromagnetism". In: *Z. Physik* 31 (1925), pp. 253–258.

[61] L. Isserlis. "On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables". In: *Biometrika* 12.1-2 (1918), pp. 134–139.

[62] C. Itzykson and J.B. Zuber. "The planar approximation II". In: *J. Math. Phys* 21 (1980), pp. 411–421.

[63] M. Jimbo and T. Miwa. *Algebraic Analysis of Solvable Lattice Models*. CBMS Regional Conference Series in Mathematics 85. AMS, 1995.

[64] M. Kac and J.C. Ward. "A Combinatorial Solution of the Two-Dimensional Ising Model". In: *Phys. Rev.* 88.6 (1952), pp. 1332–1337.

[65] L.P. Kadanoff. "Scaling laws for Ising models near T_c". In: *Physics Physique Fizika* 2.6 (1966), pp. 263–272.

[66] A.A. Kapaev. "Quasi-linear Stokes phenomenon for the Painlevé first equation". In: *J. Phys. A* 37 (2004), pp. 11149–11167.

[67] A.A. Kapaev and A.V. Kitaev. "Connection formulae for the first Painlevé transcendent in the complex domain". In: *Lett. Math. Phys.* 27 (1993), pp. 243–252.

[68] B. Kaufman. "Crystal Statistics. II. Partition Function Evaluated by Spinor Analysis". In: *Phys. Rev.* 76.8 (1949), pp. 1232–1243.

[69] B. Kaufman and L. Onsager. "Crystal Statistics. III. Short-Range Order in a Binary Ising Lattice". In: *Phys. Rev.* 76.8 (1949), pp. 1244–1252.

[70] V.A. Kazakov. "Ising Model on a Dynamical Planar Lattice: Exact Solution". In: *Phys. Lett. A* 119.3 (1986), pp. 140–144.

[71] V.A. Kazakov and A. Marshakov. "Complex curve of the two-matrix model and its tau-function". In: *J. Phys. A: Math. Gen.* 36 (2003), pp. 3107–3136.

[72] V.G. Knizhnik, A.M. Polyakov, and A.B. Zamolodchikov. "Fractal Structure of 2D-Quantum Gravity". In: *Mod. Phys. Lett. A* 3.8 (1988), pp. 819–826.

[73] H.A. Kramers and G.H. Wannier. "Statistics of the two-dimensional ferromagnet". In: *Phys. Rev.* 60.3 (1941), pp. 252–262.

[74] A.B.J. Kuijlaars, K.T-R McLaughlin, W. Van Assche, and M. Vanlessen. “The Riemann-Hilbert approach to strong asymptotics for orthogonal polynomials on $[-1, 1]$”. In: *Adv. Math.* 188 (2004), pp. 337–398.

[75] A.B.J. Kuijlaars and K.T-R. McLaughlin. “A Riemann-Hilbert problem for biorthogonal polynomials”. In: *J. Comp. Appl. Math.* 178 (2005), pp. 313–320.

[76] L. Landau. “On the Theory of Phase Transitions”. In: *Zh. Eksp. Teor. Fiz.* 7 (1937), pp. 19–32.

[77] N.S. Landkof. *Foundations of Modern Potential Theory.* Springer-Verlag, Berlin, 1972.

[78] D.S. Lubinsky and E.B. Saff. *Strong asymptotics for extremal polynomials associated with weights on R.* Lecture Notes in Mathematics 1305. Springer, Berlin–New York, 1988.

[79] A. Martínez-Finkelshtein and G.L.F. Silva. “Spectral Curves, Variational Problems and the Hermitian Matrix Model with External Source”. In: *Comm. Math. Phys.* 383 (2021).

[80] M.L. Mehta. “A Method of Integration over Matrix Variables”. In: *Comm. Math. Phys.* 79 (1981), pp. 327–340.

[81] M.L. Mehta. *Random Matrices.* 3rd. Academic Press, 2004.

[82] J. Miller and S. Sheffield. “Liouville quantum gravity and the Brownian map I: the QLE(8/3, 0) metric”. In: *Invent. Math.* 219.1 (2020), pp. 75–152.

[83] P. Miller. *Lectures on Random Matrix Theory for Course at SAMSI.* Lecture Notes. 2006.

[84] N.I. Muskhelishvili. *Singular Integral Equations.* Springer, 1958.

[85] E. Nikishin and V. Sorokin. *Rational Approximation and Orthogonality.* Translations of Mathematical Monographs 92, Amer. Math. Soc., 1991.

[86] L. Onsager. “Crystal statistics. I. A two-dimensional model with an order-disorder transition”. In: *Phys. Rev.* 65.3–4 (1944), pp. 117–149.

[87] R. Peierls. “On Ising’s Model of Ferromagnetism”. In: 32.3 (1936), pp. 477–481.

[88] J. Pfeffer. “Weak Liouville quantum gravity metrics with matter central charge $c \in (-\infty, 25)$”. In: *arXiv preprint arXiv:2104.04020* (2021).

[89] A.M. Polyakov. Lecture at Northeastern University. 1990.

[90] A.M. Polyakov. “Quantum Gravity in Two Dimensions”. In: *Mod. Phys. Lett. A* 2.11 (1987), pp. 893–898.

[91] A.M. Polyakov. “Quantum Theory of Bosonic Strings”. In: *Phys. Lett. B.* 103.3 (1981), pp. 207–210.

[92] R.B. Potts and J.C. Ward. "The Combinatrial [*sic*] Method and the Two-Dimensional Ising Model". In: *Prog. Theor. Phys.* 13.1 (1955), pp. 38–46.

[93] E.A. Rakhmanov. "On Asympototic Properties of Polynomials Orthogonal on the Real Axis". In: *Math. USSR Sb,* 47 (1984), pp. 155–193.

[94] E.A. Rakhmanov. "Strong asymptotics for orthogonal polynomials". In: *Methods of Approximation Theory in Complex Analysis and Mathematical Physics*. Ed. by A.A. Gonchar and E.B. Saff. Springer Berlin Heidelberg, 1993, pp. 71–97.

[95] R. Rhodes and V. Vargas. "KPZ formula for log-infinitely divisible multifractal random measures". In: *ESAIM: Probability and Statistics* 15 (2011), pp. 358–371.

[96] W. Rudin. *Functional Analysis*. McGraw-Hill, 1973.

[97] E.B. Saff and V. Totik. *Logarithmic Potentials with External Fields*. Springer-Verlag, 1997.

[98] R.P. Stanley. *Enumerative combinatorics. Vol. 2*. Cambridge Studies in Advanced Mathematics 62. Cambridge University Press, 1999.

[99] G. Szëgo. "Ein Grenzwertsatz über die Toeplitzschen Determinanten einer reellen positiven Funktion". In: *Math. Ann.* 76.4 (1915), pp. 490–503.

[100] T. Tada. "(q, p) critical point from two-matrix models". In: *Phys. Lett. B.* 259.4 (1991), pp. 442–447.

[101] T. Tada and M. Yamaguchi. "P and Q operator analysis for the two-matrix model". In: *Phys. Lett. B.* 250.1–2 (1990), pp. 38–43.

[102] V. Totik. *Weighted approximation with varying weight*. Lecture Notes in Mathematics 1569. Springer, Berlin, 1994.

[103] G.C. Wick. "The evaluation of the collision matrix". In: *Phys. Rev.* 80.2 (1950), pp. 268–272.

[104] H. Widom. "Toeplitz determinants with singular generating functions". In: *Amer. J. Math.* 95 (1973), pp. 333–383.

[105] K.G. Wilson. "Renormalization group and critical phenomena. I. Renormalization group and the Kadanoff scaling picture". In: *Phys. Rev. B* 4.9 (1971), pp. 3174–3183.

[106] K.G. Wilson. "Renormalization group and critical phenomena. II. Phase-space cell analysis of critical behavior". In: *Phys. Rev. B* 4.9 (1971), pp. 3184–3205.

[107] T.T Wu. "Theory of Toeplitz Determinants and the Spin Correlations of the Two-Dimensional Ising Model. I". In: *Phys. Rev.* 149.1 (1966), pp. 380–401.

[108] C.N. Yang. "The Spontaneous Magnetization of a Two-Dimensional Ising Model". In: *Phys. Rev.* 85.5 (1952), pp. 808–816.

[109] C.N. Yang and T.D. Lee. "Statistical Theory of Equations of State and Phase Transitions. I. Theory of Condensation". In: *Phys. Rev.* 87.3 (1952), pp. 404–409.

[110] C.N. Yang and T.D. Lee. "Statistical Theory of Equations of State and Phase Transitions. II. Lattice Gas and Ising Model". In: *Phys. Rev.* 87.3 (1952), pp. 410–419.

[111] A.B. Zamolodchikov. "Infinite additional symmetries in two-dimensional conformal quantum field theory". In: *Theor. Math. Phys.* 65 (1985), pp. 1205–1213.

[112] P. Zinn-Justin and J.B. Zuber. "On some integrals over the U(N) unitary group and their large N limit". In: *J. Phys. A: Math. Gen.* 36 (2003), pp. 3173–3193.

[113] A. Zvonkin. "Matrix integrals and map enumeration: An accessible introduction". In: *Math. Comput. Modelling* 26.8-10 (1997), pp. 281–304.

APPENDIX A
POTENTIAL THEORY.

Here, we establish some of the necessary tools and techniques we shall borrow from potential theory. We do not attempt to make a comprehensive review of all of potential theory; our aim here is only to explain the minimal amount of theory required to follow the main text (especially Chapter 4). None of the results we refer to in this section are new. One should consult [77, 97] and references therein for further details.

A.1 Harmonic, Superharmonic, and Subharmonic Functions.

Potential theory is intimately tied to the theory of harmonic functions. Recall that a C^2-function $u : \Omega \to \mathbb{R}$ defined on a domain $\Omega \subset \mathbb{C}$ is called *harmonic* if its *Laplacian*

$$\Delta u \equiv \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)u = 4\partial\bar{\partial}u = 0 \tag{A.1}$$

vanishes in Ω (Here, $\partial f, \bar{\partial} f$ are the holomorphic/ antiholomorphic derivatives of a function f, respectively). The following proposition states the relevant properties of harmonic functions:

Proposition A.1. *(Properties of harmonic functions). Let $u, v : \Omega \to \mathbb{R}$ be harmonic functions. Then:*

1. *(mean-value property) $(M_r u)(z_0) = u(z_0)$, for any disc $B_r(z_0) \subset \Omega$. Here, $(M_r u)$ represents the integral average of u along the boundary of the disc $B_r(z_0)$:*

$$(M_r u)(z_0) = \frac{1}{2\pi r}\int_0^{2\pi} u(z_0 + re^{i\theta})d\theta. \tag{A.2}$$

2. *(Identity principle) If $u = v$ on a nonempty open subset of Ω, then $u \equiv v$ everywhere in Ω.*

3. *(Maximum/minimum principle) If u attains its maximum/minimum at $z_0 \in \Omega$, then $u \equiv const$.*

4. *(Liouville theorem) If $\Omega = \mathbb{C}$, and u is bounded, then $u \equiv const$.*

We refer the reader to [77] for proofs of the above facts. We will also need the classes of *superharmonic* and *subharmonic* functions.

Definition A.2. (Superharmonic function). A function is called superharmonic if it is lower-semicontinuous ($\liminf_{\zeta \to z} v(\zeta) \geq v(z)$, for every $z \in \Omega$), and satisfies the super mean-value property: $(M_r v)(z) \leq v(z)$ for every $z \in \Omega$.

Likewise,

Definition A.3. A function v is called subharmonic if $-v$ is superharmonic, or, equivalently, it is upper-semicontinuous ($\limsup_{\zeta \to z} v(\zeta) \leq v(z)$, for every $z \in \Omega$) and satisfies the sub mean-value property: $(M_r v)(z) \geq v(z)$ for every $z \in \Omega$.

Obviously, we have that a function is harmonic if and only if it is both super- and sub-harmonic. Superharmonic functions enjoy the following properties:

Proposition A.4. *Let $u, v : \Omega \to \mathbb{C}$ be superharmonic functions. Then:*

1. $\min\{u, v\}$ *and* $\alpha u + \beta v$ *are superharmonic, for* $\alpha, \beta > 0$.
2. *(Minimum principle) If u attains its minimum at $z_0 \in \Omega$, then u is constant.*
3. *(Comparison to harmonic functions) If h is harmonic on Ω, and*

$$\liminf_{\zeta \to z} u(\zeta) \geq h(z)$$

for every $z \in \partial\Omega$, then $u(z) \geq h(z)$ in Ω.

We again refer to [77] for the proofs. We state one final proposition regarding superharmonic functions before studying potentials:

Proposition A.5. *Let (X, μ) be a finite measure space, $\Omega \subset \mathbb{C}$ a domain, and $v : \Omega \times X \to (-\infty, \infty]$ a measurable function satisfying*

1. *$v(z, \cdot)$ is superharmonic for any fixed value of the second argument;*
2. $\inf_{t \in X} v(z, t)$ *is locally bounded below on* Ω.

Then the function

$$u(z) = \int_X v(z, t)\, d\mu(t) \tag{A.3}$$

is superharmonic on Ω.

Proof. Since superharmonicity is a local property, it suffices to prove the theorem on discs contained in Ω. Fix such a precompact set $D \subset \Omega$. Property 2. implies that $\inf_t v(z, t)$ is bounded below on this set, and thus, without loss of generality, we may assume $v(z,t) \geq 0$ on $D \times X$. Now, if $z_n \to z$, then, by Fatou's lemma (and the lower semi-continuity of v), we have that

$$\begin{aligned}\liminf_{n\to\infty} u(z_n) &\geq \int_X \liminf_{n\to\infty} v(z_n,t)\, d\mu(t) \\ &\geq \int_X v(z,t) d\mu(t) = u(z)\end{aligned}$$

therefore, we conclude that u is lower semi-continuous. Furthermore, for any disc $B_r(z_0) \subset D$, by Fubini's theorem (and the super mean-value property of v), we have that

$$\begin{aligned}(M_r u)(z_0) &= \frac{1}{2\pi r}\int_0^{2\pi} u(z_0 + re^{i\theta})\, d\theta \\ &= \int_X \frac{1}{2\pi r}\int_0^{2\pi} v(z_0 + re^{i\theta}, t) d\theta\, d\mu(t) \\ &\leq \int_X v(z_0,t)\, d\mu(t) = u(z_0),\end{aligned}$$

i.e. u satisfies the super mean-value property. We conclude that u is superharmonic. □

Note the duality betweeen superharmonic and subharmonic functions: replacing 'superharmonic' with 'subharmonic', 'inf' with 'sup', and 'locally bounded below' with 'locally bounded above' yields an analogous theorem about subharmonic functions.

A.2 Logarithmic Potentials and Capacity.

We now discuss the theory of potentials. We begin with a few remarks about the class of measures we shall be considering from here on. The measure space in which we will usually be working is $\mathbb{C}$, with the usual σ-algebra of Borel sets, unless otherwise specified. Given a measure μ on this measure space, we define its *support* to be

$$\text{supp}\ \mu := \{z \in \mathbb{C} \mid \forall \epsilon > 0, \mu(B_\epsilon(z)) > 0\}. \tag{A.4}$$

One can think of supp μ as the set of points where the measure μ doesn't vanish identically. If $K \subset \mathbb{C}$, we let $\mathcal{M}_1(K)$ denote the set of positive, unit Borel measures compactly supported in K. We now define the *logarithmic potential* of a measure μ.

Definition A.6. Let μ be a finite Borel measure with compact support in $\mathbb{C}$. We define its potential, $U^\mu(z)$, to be

$$U^\mu(z) = \int \log \frac{1}{|z-w|} d\mu(w). \tag{A.5}$$

Potentials will be the primary objects of interest in this section. We now formulate some of the basic properties of potentials:

Proposition A.7. *$U^\mu(z)$ is a superharmonic function, and harmonic on $\mathbb{C}\setminus(\text{supp }\mu)$. Moreover, at infinity,*

$$U^\mu(z) = -\mu(\mathbb{C})\log|z| + O(1/|z|). \tag{A.6}$$

Proof. Using the fact that $\log\frac{1}{|z-w|}$ is a superharmonic function locally bounded above, by Theorem A.5 we have that $U^\mu(z)$ is superharmonic on $\mathbb{C}$. By the same theorem, applied now to $\mathbb{C}\setminus(\text{supp }\mu)$, $\log\frac{1}{|z-w|}$ is subharmonic here, and thus so is $U^\mu(z)$. Therefore U^μ is harmonic off of the support of μ. Finally, to see that equation (A.6) holds, write

$$U^\mu(z) = -\mu(\mathbb{C})\log|z| - \int \log\left|1-\frac{w}{z}\right| d\mu(z),$$

for $|z|$ sufficiently large, since μ is compactly supported, we see that the last term is $O(1/|z|)$. □

We now introduce the notion of *capacity* of a set, which will act as a measure of the ability of that set to hold charge.

Definition A.8. Let $K\subset\mathbb{C}$ be compact, and $\mu\in\mathcal{M}_1(K)$. The *logarithmic energy* of μ is defined as

$$I[\mu] := \int\int \log\frac{1}{|z-w|} d\mu(w)d\mu(z) = \int U^\mu(w)d\mu(w). \tag{A.7}$$

We define the *equilibrium energy* of K to be

$$I_0 := \inf_{\mu\in\mathcal{M}_1(K)} I[\mu]. \tag{A.8}$$

The *logarithmic capacity* of K is then defined to be

$$\text{cap}(K) := e^{-I_0}. \tag{A.9}$$

It is a well-known fact that (cf. [77] again), for compact subsets $K\subset\mathbb{C}$, there is always a measure $\mu_0\in\mathcal{M}_1(K)$ satisfying

$$I_0 = I[\mu_0] = \inf_{\mu\in\mathcal{M}_1(K)} I[\mu], \tag{A.10}$$

and so cap(K) is well-defined. We summarize some of the properties of capacity in the following proposition.

Proposition A.9. *1. Let $E \subset \mathbb{C}$, and $\alpha, \beta \in \mathbb{C}$. Then*

$$\operatorname{cap}(\alpha E + \beta) = \operatorname{cap}(E). \tag{A.11}$$

2. If $z_0 \in \mathbb{C}$, $r > 0$, then

$$\operatorname{cap}(\{|z - z_0| \leq r\}) = r. \tag{A.12}$$

3. If $z_0 \in \mathbb{C}$, $r > 0$, then

$$\operatorname{cap}(\{|z - z_0| = r\}) = r. \tag{A.13}$$

4. if $[a, b] \subset \mathbb{R}$ is an interval, then

$$\operatorname{cap}([a, b]) = \frac{1}{4}(b - a). \tag{A.14}$$

Proof. The first property is obvious from the definition of capacity. For sake of brevity, we omit the proofs of the remaining three properties of capacity, and again refer the reader to [77, 97] for the details. □

The definition of capacity can be extended to any measurable set E in the usual way, by setting

$$\operatorname{cap}(E) := \sup \{\operatorname{cap}(K) \mid K \subset E,\, K \text{ compact}\}. \tag{A.15}$$

Capacity is always a non-negative quantity; we call sets of capacity zero *polar*. A property is said to hold *quasi-everywhere* (q.e.) if it occurs everywhere except a polar set. We state without proving that every polar set is of measure zero, so that a property holding quasi-everywhere also occurs almost everywhere. Moreover, the countable union of polar sets is again polar.

A.3 The Weighted Energy Problem.

In our Riemann-Hilbert analysis of orthogonal polynomials, we wanted to minimize a functional of the form[1]

$$E[\mu] := \iint \log \frac{1}{|z - \zeta|} d\mu(z) d\mu(\zeta) + 2 \int V(z) d\mu(z), \tag{A.16}$$

with the minimum taken over all measures $\mu \in \mathcal{M}_1(K)$, for some closed set K (in the situation of Chapter 4, $K = \mathbb{R}$). We now pose conditions under which this sort of minimization problem makes sense, and the resulting variational conditions that arise for the extremizing measure.

[1]The factor of 2 in front of $V(z)$ here is inessential, and is taken for simplicity of notations of this appendix.

Definition A.10. Let $F \subset \mathbb{C}$ be a closed set, and $w : F \to \mathbb{R}_+$; we call w a *weight function* on F. A weight function $w(z)$ on F is said to be *a*dmissible if the following properties hold:

1. w is upper-semicontinuous,

2. The set of points $F_0 \subset F$ where $w(z) > 0$ has positive capacity, and finally,

3. If F is unbounded, then $|z|w(z) \to 0$ as $|z| \to \infty$ inside F.

The last condition is take to ensure $w(z)$ grows sufficiently fast at infinity. As an example (which, for us, is *the* example of interest), if we take $F = \mathbb{R}$, and $V(z)$ any monic polynomial of even degree $d \geq 2$, then the weight function

$$w(z) := \exp(-V(z)) \tag{A.17}$$

is admissible, since it is clearly upper-semicontinuous, positive on the whole real line (and thus condition 2. holds), and finally trivially satisfies the last growth condition. Motivated by this example, given an admissible weight, we define $V(z) = V(z; w) := \log|w(z)|$.

Given an admissible weight, we define the weighted energy functional

$$\begin{aligned} E_w[\mu] = E[\mu] &:= \iint \log \frac{1}{|z-\zeta| w(z) w(\zeta)} d\mu(z) d\mu(\zeta) \\ &= \iint \log \frac{1}{|z-\zeta|} d\mu(z) d\mu(\zeta) + 2 \int V(z) d\mu(z). \end{aligned} \tag{A.18}$$

Given an admissible weight function $w(z)$ on $F \subset \mathbb{C}$, we are interested in the following minimization problem:

$$\inf_{\mu \in \mathcal{M}_1(F)} E_w[\mu] =: V_0. \tag{A.19}$$

If λ is a minimizing measure for the above problem, define

$$\ell_0 := V_0 - \int V(z) d\lambda(z) = \iint \log \frac{1}{|z-\zeta|} d\lambda(z) d\lambda(\zeta) + \int V(z) d\lambda(z). \tag{A.20}$$

We have the following theorem:

Theorem A.11. *1. ℓ_0 is finite, and moreover, there is a unique measure $\lambda \in \mathcal{M}_1(F)$ such that the minimum is attained:*

$$\inf_{\mu \in \mathcal{M}_1(F)} E_w[\mu] = E_w[\lambda] = V_0. \tag{A.21}$$

2. The support of λ is a compact subset of F, and is of positive capacity,

3. *For quasi-every* $z \in F$,

$$U^{\lambda}(z) + V(z) \geq \ell_0, \tag{A.22}$$

and for quasi-every z *in the support of* λ,

$$U^{\lambda}(z) + V(z) = \ell_0. \tag{A.23}$$

Note that, in the case $w(z) := \exp(-V(z))$ with $V(z)$ a monic polynomial of even degree, conditions (A.22), (A.23) are precisely the variational conditions (4.51), (4.52) of Chapter 4. Theorem (A.11) is sometimes referred to as the *Frostman theorem* or *Gauss-Frostman theorem*; the conditions (A.22), (A.23) are called the *Frostman conditions*. On physical grounds, these principles are quite clear: if we have a collection of electrons in equilibrium with an external field $V(z)$, then on the support of the charges (i.e. supp λ), the effective potential $U^{\lambda}(z) + V(z)$ should be constant. Moreover, if the charges are truly in equilibrium, if we move away from the support of the charges, the effective potential should increase, otherwise the electrons would have moved to an area of lower potential energy. Before proving Theorem (A.11), we first need a few lemmas about weak-$*$ convergence. Recall that a sequence of Borel measures $\{\mu_k\}$ on F is said to converge *weakly* to μ if, for every $\varphi \in C(F)$, the sequence

$$\int_F \varphi d\mu_n \to \int_F \varphi d\mu, \tag{A.24}$$

as $n \to \infty$. We then write $\mu_n \xrightarrow{*} \mu$. This notion of convergence induces a topology on the space of measures, called the *weak-$*$ topology*. The set $\mathcal{M}_1(F)$ is of course a subset of the space of all Borel measures on F, and in fact, is compact in this topology, in the following sense:

Lemma A.12. *Let* $\{\mu_n\}$ *be a sequence of measures in* $\mathcal{M}_1(F)$. *Then* $\{\mu_n\}$ *has a convergent subsequence.*

Proof. The proof of this lemma is an elementary fact from real analysis; we refer the reader to [96], for example. □

With this in mind, we can prove the following lemma:

Lemma A.13. *(Fatou Lemma for the weighted energy) Suppose* $\mu_n \xrightarrow{*} \mu$ *in* $\mathcal{M}_1(F)$, *and suppose* $w : F \to \mathbb{R}_+$ *is an admissible weight function on* F. *Then,*

$$E_w[\mu] \leq \liminf_{n\to\infty} E_w[\mu_n]. \tag{A.25}$$

Proof. For simplicity, we assume that F is compact. The general case follows using the usual tools from real analysis. By assumption, we have that, for any $\varphi, \psi \in C(F)$,

$$\iint \varphi(z)\psi(\zeta)\, d\mu_n(\zeta)d\mu_n(z) \to \iint \varphi(z)\psi(\zeta)\, d\mu(\zeta)d\mu(z)$$

The Stone-Weierstrass theorem yields that, for any $f(z,\zeta) \in C(F \times F)$, there exist $\{\varphi_j(z), \psi_j(\zeta)\}_{j=1}^n$ such that $\sum \varphi_j(z)\psi_j(\zeta)$ is uniformly close to f. It follows that

$$\iint f(z,w)\, d\mu_n(\zeta)d\mu_n(z) \to \int\int f(z,\zeta)\, d\mu(\zeta)d\mu(z)$$

for any $f \in C(F \times F)$. Setting $f(z,\zeta) = \min\{\log \frac{1}{[|z-\zeta|w(z)w(\zeta)]}, m\} \in C(F \times F)$ for some fixed integer $m \geq 1$, we find that

$$\begin{aligned}
\liminf_{n\to\infty} E_w[\mu_n] &= \liminf_{n\to\infty} \iint \log \frac{1}{[|z-\zeta|w(z)w(\zeta)]} d\mu_n(\zeta)d\mu_n(z) \\
&\geq \iint \min\{\log \frac{1}{[|z-\zeta|w(z)w(\zeta)]}, m\} d\mu_n(\zeta)d\mu_n(z) \\
&= \iint \min\{\log \frac{1}{[|z-\zeta|w(z)w(\zeta)]}, m\} d\mu(\zeta)d\mu(z)
\end{aligned}$$

where the first equality follows from $\log \frac{1}{[|z-\zeta|w(z)w(\zeta)]} \geq \min\{\log \frac{1}{[|z-\zeta|w(z)w(\zeta)]}, m\}$, and the second from our previous observation. Since the functions $f_m = \min\{\log \frac{1}{[|z-\zeta|w(z)w(\zeta)]}, m\}$ are increasing, by the monotone convergence theorem, we have finally that

$$\liminf_{n\to\infty} E_w[\mu_n] \geq \iint \min\{\log \frac{1}{[|z-\zeta|w(z)w(\zeta)]}, m\} d\mu(\zeta)d\mu(z) \to E_w[\mu].$$

□

With these lemmas in place, we are now ready to prove Theorem (A.11):

Proof. (of Theorem (A.11)). Let us begin by showing that an extremizing measure indeed exists. Set $\ell_0 = \inf_{\mu\in\mathcal{M}_1(K)} E_w[\mu]$, and let $\{\mu_n\}$ be a sequence of measures from $\mathcal{M}_1(K)$ such that $E_w[\mu_n] \to \ell_0$. By Lemma (A.12), this sequence has a convergent subsequence; denote the limiting measure by $\mu_{n_k} \xrightarrow{*} \lambda$. By the lemma, we have that

$$E_w[\lambda] \leq \liminf_{k\to\infty} E_w[\mu_{n_k}] = \ell_0, \tag{A.26}$$

therefore λ extremizes $E_w[\mu]$. For the proof of finiteness of ℓ_0 and uniqueness of λ, we refer the reader to [97], Theorem 1.3. For brevity, we also omit the proof that the support of λ has positive capacity.

It remains to prove the equality and inequality (A.22), (A.23). Let

$$\Phi(z) := U^{\lambda}(z) + V(z) = \int \log \frac{1}{|z-w|} d\lambda(w) + V(z) \tag{A.27}$$

denote the "effective potential". We will prove the following assertion:

The set $E := \{z \in E \mid \Phi(z) < \ell_0\}$ is polar (of capacity 0).

From this assertion, we can immediately infer (A.22). We proceed by contradiction. Suppose, that E is of positive capacity; then, there is an integer $N \geq 1$ such that the set

$$E_1 := \left\{ z \in E \mid |z| \leq N, \Phi(z) < \ell_0 - \frac{1}{N} \right\} \tag{A.28}$$

is also of positive capacity; note that E_1 is also compact. On the other hand, since $\int \Phi(z) d\lambda(z) = \ell_0$, there must be a compact set $E_2 \subset \operatorname{supp} \lambda$ such that

$$\Phi(z) > \ell_0 - \frac{1}{2N}, \tag{A.29}$$

for every $z \in E_2$. Obviously, we must have that $m := \lambda(E_2) > 0$. Now, let $\sigma \in \mathcal{M}_1(E_1)$, such that $E_w[\sigma]$ is finite, and define the (signed) measure $\hat{\sigma}$ by

$$\hat{\sigma} = \begin{cases} m\sigma, & \text{on } E_1, \\ -\lambda, & \text{on } E_2, \\ 0, & \text{otherwise.} \end{cases} \tag{A.30}$$

We then have that, for any $t > 0$, sufficiently small, the measure $\lambda_t := \lambda + t\hat{\sigma} \in \mathcal{M}_1(F)$. Furthermore, if we compute the weighted energy of λ_t,

$$\begin{aligned} E_w[\lambda_t] &= V_0 + 2t \left[\iint \log \frac{1}{|z-\zeta|} d\lambda(z) d\hat{\sigma}(\zeta) + \int V(z) d\hat{\sigma}(z) \right] + t^2 \iint \log \frac{1}{|z-\zeta|} d\sigma(z) d\hat{\sigma}(\zeta) \\ &= V_0 + 2t \int \left[U^{\lambda}(z) + V(z) \right] d\hat{\sigma}(\zeta) + \mathcal{O}(t^2) \\ &< V_0 - \frac{2tm}{2N} + \mathcal{O}(t^2) < V_0, \end{aligned}$$

in contradiction with the fact λ was the minimizer to $E_w[\mu]$.

Now, we prove (A.23). Let $z_0 \in \text{supp } \lambda$, and suppose that

$$\Phi(z_0) > \ell_0. \tag{A.31}$$

By lower-semicontinuity of Φ, we have that there exists $\delta > 0$ such that $\Phi(z) > \ell_0 + \epsilon$ holds, for all $|z - z_0| < \delta$, and $\epsilon > 0$ sufficiently small (without loss of generality, $A := \overline{\{|z - z_0| < \epsilon\}} \subset \text{supp } \lambda$). Thus, we can write

$$\begin{aligned}\ell_0 = \int \Phi(z) d\lambda(z) &= \int_A \Phi(z) d\lambda(z) + \int_{(\text{supp } \lambda)\setminus A} \Phi(z) d\lambda(z) \\ &\geq \lambda(A)\,[\ell_0 + \epsilon] + [1 - \lambda(A)]\ell_0,\end{aligned}$$

which in turn implies that $\lambda(A) = 0$. But, this is a contradiction, since we assumed $A \subset \text{supp } \lambda$. The inequality

$$\Phi(z) := U^{\lambda}(z) + V(z) \leq \ell_0, \tag{A.32}$$

for $z \in \text{supp } \lambda$, follows. This inequality, along with the inequality (A.22), implies the equality (A.23). $\square$

The above theorem is extremely useful when it comes to calculation of equilibrium measures. In the next section, we show how the Gauss-Frostman theorem may be used to calculate the corresponding equilibrium measure, in the case when $V(z)$ is a monic polynomial of even degree.

A.4 Calculation of Equilibrium Measures.

Here, we show some examples of how the Gauss-Frostman theorem may be used to calculate (in a fairly explicit manner) the equilibrium measure in the external field $V(z)$, where $V(z)$ is a monic polynomial of even degree.

We have the following theorem:

Theorem A.14. *Let $V(z)$ be a monic polynomial of even degree $d \geq 2$, and consider the functional*

$$E_w[\mu] = \iint \log \frac{1}{|z - \zeta|} d\mu(z) d\mu(\zeta) + 2 \int V(z) d\mu(z). \tag{A.33}$$

Then the (unique) minimizing measure $\lambda \in \mathcal{M}_1(\mathbb{R})$ among all measures $\mu \in \mathcal{M}_1(\mathbb{R})$ is given by

$$d\lambda(z) = \frac{1}{2\pi} \sqrt{M(z)} dx, \tag{A.34}$$

for some polynomial $M(z)$ of degree $4d-2$. Furthermore, the support of λ consists of a union of at most $2d-1$ intervals.

Proof. Let us first establish that the support of λ consists of at most $2d-1$ intervals. We aim to calculate the *Cauchy transform* of λ:

$$C^{\lambda}(z) := \int \frac{d\lambda(\zeta)}{\zeta - z}. \tag{A.35}$$

The Plemelj formula then yields that we can recover λ from the boundary values of $C^{\lambda}(z)$:

$$\frac{d\lambda}{dz} = \frac{1}{2\pi i}\left[C^{\lambda}_{+}(z) - C^{\lambda}_{-}(z)\right], \qquad z \in \text{supp } \lambda, \tag{A.36}$$

where the $+/-$ subscripts denote the limits from above and below the real line. Differentiating the Gauss-Frostman condition (A.23) on the support of the measure λ, we find that

$$C^{\lambda}_{+}(z) + C^{\lambda}_{-}(z) + V'(z) = 0, \tag{A.37}$$

where we have used the fact that the derivative of the logarithmic potential on the support of the measure λ is the Cauchy principal value integral. Now, define the function

$$Q(z) := -[C^{\lambda}(z)]^2 - [C^{\lambda}(z)]V'(z). \tag{A.38}$$

The function $Q(z)$ is analytic everywhere except possibly supp λ, and at infinity, behaves as

$$Q(z) = 2dz^{2d-2} + \mathcal{O}(z^{2d-3}), \qquad z \to \infty.$$

as a consequence of the fact that $\deg V' = 2d-1$, and $C^{\lambda}(z) = -z^{-1} + \mathcal{O}(z^{-2})$, $z \to \infty$. Comparing the boundary values of $Q(z)$ across the support of λ, we find that

$$\begin{aligned} Q_{+}(z) - Q_{-}(z) &= -[C^{\lambda}_{+}(z)]^2 + [C^{\lambda}_{-}(z)]^2 - \left[C^{\lambda}_{+}(z) - C^{\lambda}_{-}(z)\right]V'(z) \\ &= -\left[C^{\lambda}_{+}(z) - C^{\lambda}_{-}(z)\right]\left[C^{\lambda}_{+}(z) + C^{\lambda}_{-}(z)\right] - \left[C^{\lambda}_{+}(z) - C^{\lambda}_{-}(z)\right]V'(z) \\ &= \left[C^{\lambda}_{+}(z) - C^{\lambda}_{-}(z)\right]V'(z) - \left[C^{\lambda}_{+}(z) - C^{\lambda}_{-}(z)\right]V'(z) = 0, \end{aligned}$$

where in the last line, we have used the "derivative" Frostman condition (A.37). Thus, $Q(z)$ is continuous across supp λ, and thus by Morera's theorem, extends to an entire function. Since $Q(z) = 2dz^{2d-2} + \mathcal{O}(z^{2d-3})$, $z \to \infty$, we can conclude that $Q(z)$ is a polynomial of degree $2d-2$, by Liouville's theorem. We thus have an explicit quadratic expression for the Cauchy transform of λ:

$$[C^{\lambda}(z)]^2 + V'(z)C^{\lambda}(z) + Q(z) = 0;$$

solving the above for $C^{\lambda}(z)$, we find that

$$C^{\lambda}(z) = \frac{1}{2}\left[V'(z) \pm \sqrt{(V'(z))^2 - 4Q(z)}\right]$$

The Plemelj formula then yields that

$$\frac{d\lambda}{dz} = \frac{1}{2\pi}\sqrt{4Q(z) - (V'(z))^2} \tag{A.39}$$

This establishes the form of the equilibrium measure λ. Since λ is supported on $\mathbb{R}$, the real roots of the polynomial $4Q(z) - (V'(z))^2$ determine the endpoints of the support; since there are at most $2(2d-1) = 4d-2$ real roots of this polynomial, there are at most $2d-1$ components of the support, with each component being an interval:

$$\text{supp }\lambda = \bigcup_{k=1}^{2d-1}[a_k, b_k].$$

□

Now that we have determined the (form of) the support of λ, we can more carefully determine the form of the equilibrium measure. We already have a generic form of the measure (see Equation (A.39)), but we can in fact be more precise. What follows is often colloquially referred to as the "square root trick". Let us suppose the equilibrium measure is supported on *exactly* $2d-1$ intervals [2], with endpoints $a_1 < b_1 < ... < a_{2d-1} < b_{2d-1}$. We seek a function $R(z)$ satisfying

1. $R^2(z) = \prod_{k=1}^{2d-1}(z-a_k)(z-b_k)$,
2. R is analytic in $\mathbb{C} \setminus (\bigcup_k [a_k, b_k])$.
3. $R(z) = z^{2d-1} + \mathcal{O}(z^{2d-2})$, $z \to \infty$.

There is an explicit formula for such a function:

$$R(z) = \prod_{k=1}^{2d-1}(z-a_k)^{1/2}(z-b_k)^{1/2}. \tag{A.40}$$

[2]In practice, this may not be the case, and one often finds the number of intervals by trial and error. If the "square root trick" doesn't work for $2d-1$ intervals, it may work for $2d-2$; since the equilibrium measure is unique, what we have found *is* the equilibrium measure. If the trick doesn't work for $2d-2$, then one must try $2d-3$ intervals, and so on. The key point here is that there are finitely many cases, and so the measure can be found via a finite process.

Furthermore, the function $R(z)$ has the following boundary values:

$$R_+(z) = \begin{cases} R_-(z), & z \in \mathbb{R} \setminus \left(\bigcup_{k=1}^{2d-1} [a_k, b_k] \right) \\ -R_-(z), & z \in \bigcup_{k=1}^{2d-1} [a_k, b_k]. \end{cases}$$

Then, the function $h(z) = \frac{C^\lambda(z)}{R(z)}$ is analytic in $\mathbb{C} \setminus \left(\bigcup_{k=1}^{2d-1} [a_k, b_k] \right)$, and satisfies the jump condition

$$h_+(z) - h_-(z) = \frac{V'(z)}{R_+(z)}, \qquad z \in \bigcup_{k=1}^{2d-1} [a_k, b_k].$$

Furthermore, the function $h(z)$ satisfies $h(z) = -z^{-2d+2} + \mathcal{O}(z^{-2d+3})$, as $z \to \infty$. Since, in particular, $h(z) = \mathcal{O}(z^{-1})$ as $z \to \infty$, we can explicitly find $h(z)$ by means of the Plemelj formula:

$$h(z) = \frac{1}{2\pi i} \sum_{k=1}^{2d-1} \int_{a_k}^{b_k} \frac{V'(\zeta) d\zeta}{R_+(\zeta)(\zeta - z)}; \tag{A.41}$$

the condition that the first $2d-1$ terms in the series expansion of $h(z)$ at infinity determine the positions of the endpoints of the support.

In particular, we obtain several results immediately from the following theorem. First, we note that, if $\deg V = 2q$, then the support of λ consists of (at most) q intervals. We also remark that, in a generic situation, the endpoints of the support of λ behave like $(x-a)^{1/2}$. This justifies most of the comments we made in the construction of local parametrices in Chapter (4).

APPENDIX B

EXPANSION OF THE UNIFORMIZING COORDINATE NEAR THE BRANCH POINTS.

Here, we list the relevant expansions of the uniformizing coordinate on each sheet of the spectral curve. We have included Figure (B.1), which depicts the leading order asymptotics of the uniformizing coordinate near each of the branch points, for the convenience of the reader. The expansions at infinity hold for all $(a,b) \in R = \{0 < b \leq 1, 1 \leq a \leq b^{-1}\}$, and so we list them first. Let $A = A(a,b) = ab\sqrt{-\tau(a,b)/3t(a,b)}$. Then,

$$u_1(z) = \frac{z}{A} - \frac{(a^2+b^2)A}{z} - \frac{(\frac{5}{3}a^2b^2 + a^4 + b^4)A^3}{z^3} - \frac{(\frac{14}{3}(a^4b^2 + a^2b^4) + 2a^6 + 2b^6)A^5}{z^5} + \mathcal{O}(z^{-7}), \tag{B.1}$$

$$u_2(z) = \begin{cases} -\left(\frac{a^2b^2}{3}\right)^{1/3} \frac{\omega A^{1/3}}{z^{1/3}} + \frac{(a^2+b^2)A}{3z} - \left(\frac{3}{a^2b^2}\right)^{1/3} \frac{\omega^2(a^4+a^2b^2+b^4)A^{5/3}}{9z^{5/3}} + \mathcal{O}(z^{-7/3}), & \text{Im } z > 0, \\ -\left(\frac{a^2b^2}{3}\right)^{1/3} \frac{\omega^2 A^{1/3}}{z^{1/3}} + \frac{(a^2+b^2)A}{3z} - \left(\frac{3}{a^2b^2}\right)^{1/3} \frac{\omega(a^4+a^2b^2+b^4)A^{5/3}}{9z^{5/3}} + \mathcal{O}(z^{-7/3}), & \text{Im } z < 0, \end{cases} \tag{B.2}$$

$$u_3(z) = -\left(\frac{a^2b^2}{3}\right)^{1/3} \frac{A^{1/3}}{z^{1/3}} + \frac{(a^2+b^2)A}{3z} - \left(\frac{3}{a^2b^2}\right)^{1/3} \frac{(a^4+a^2b^2+b^4)A^{5/3}}{9z^{5/3}} + \mathcal{O}(z^{-7/3}), \tag{B.3}$$

$$u_4(z) = \begin{cases} -\left(\frac{a^2b^2}{3}\right)^{1/3} \frac{\omega^2 A^{1/3}}{z^{1/3}} + \frac{(a^2+b^2)A}{3z} - \left(\frac{3}{a^2b^2}\right)^{1/3} \frac{\omega(a^4+a^2b^2+b^4)A^{5/3}}{9z^{5/3}} + \mathcal{O}(z^{-7/3}), & \text{Im } z > 0, \\ -\left(\frac{a^2b^2}{3}\right)^{1/3} \frac{\omega A^{1/3}}{z^{1/3}} + \frac{(a^2+b^2)A}{3z} - \left(\frac{3}{a^2b^2}\right)^{1/3} \frac{\omega^2(a^4+a^2b^2+b^4)A^{5/3}}{9z^{5/3}} + \mathcal{O}(z^{-7/3}), & \text{Im } z < 0, \end{cases} \tag{B.4}$$

On the other hand, the local expansions of the uniformizing coordinate differ in the noncritical/critical cases and multicritical cases. We indicate the behavior of the uniformization coordinate around these branch points here.

B.1 Expansion of the Uniformizing Coordinate in the Generic and Critical Cases.

All of the following expansions are valid for $0 < b < 1$, $1 \le a \le b$. We also let $A := A(a,b) > 0$ be as in (5.34).

- *Expansion at $z = +\alpha$.* As $z \to \alpha$, letting $\zeta = z - \alpha$,

$$u_1(z) = a + \frac{a^{3/2}}{A^{\frac{1}{2}}(a^2-b^2)^{\frac{1}{2}}}\zeta^{\frac{1}{2}} + \frac{a^2(3a^2-7b^2)}{6A(a^2-b^2)^2}\zeta + \frac{a^{5/2}C_2}{72A^{\frac{3}{2}}(a^2-b^2)^{\frac{7}{2}}}\zeta^{\frac{3}{2}} + \mathcal{O}(\zeta^2), \tag{B.5}$$

$$u_2(z) = a - \frac{a^{3/2}}{A^{\frac{1}{2}}(a^2-b^2)^{\frac{1}{2}}}\zeta^{\frac{1}{2}} + \frac{a^2(3a^2-7b^2)}{6A(a^2-b^2)^2}\zeta - \frac{a^{5/2}C_2}{72A^{\frac{3}{2}}(a^2-b^2)^{\frac{7}{2}}}\zeta^{\frac{3}{2}} + \mathcal{O}(\zeta^2), \tag{B.6}$$

and $u_3(z), u_4(z)$ have regular expansions. Here, $C_2 := C_2(a,b) = 9a^4 - 30a^2b^2 + 101b^4 > 0$.

- *Expansion at $z = -\alpha$.* As $z \to -\alpha$, letting $\zeta = z + \alpha$,

$$u_1(z) = \begin{cases} -a + \frac{ia^{3/2}}{A^{\frac{1}{2}}(a^2-b^2)^{\frac{1}{2}}}\zeta^{\frac{1}{2}} + \frac{a^2(3a^2-7b^2)}{6A(a^2-b^2)^2}\zeta + \frac{ia^{5/2}C_2}{72A^{\frac{3}{2}}(a^2-b^2)^{\frac{7}{2}}}\zeta^{\frac{3}{2}} + \mathcal{O}(\zeta^2), & \mathrm{Im}\,\zeta > 0, \\ -a - \frac{ia^{3/2}}{A^{\frac{1}{2}}(a^2-b^2)^{\frac{1}{2}}}\zeta^{\frac{1}{2}} + \frac{a^2(3a^2-7b^2)}{6A(a^2-b^2)^2}\zeta - \frac{ia^{5/2}C_2}{72A^{\frac{3}{2}}(a^2-b^2)^{\frac{7}{2}}}\zeta^{\frac{3}{2}} + \mathcal{O}(\zeta^2), & \mathrm{Im}\,\zeta < 0, \end{cases} \tag{B.7}$$

$$u_2(z) = \begin{cases} -a - \frac{ia^{3/2}}{A^{\frac{1}{2}}(a^2-b^2)^{\frac{1}{2}}}\zeta^{\frac{1}{2}} + \frac{a^2(3a^2-7b^2)}{6A(a^2-b^2)^2}\zeta - \frac{ia^{5/2}C_2}{72A^{\frac{3}{2}}(a^2-b^2)^{\frac{7}{2}}}\zeta^{\frac{3}{2}} + \mathcal{O}(\zeta^2), & \mathrm{Im}\,\zeta > 0, \\ -a + \frac{ia^{3/2}}{A^{\frac{1}{2}}(a^2-b^2)^{\frac{1}{2}}}\zeta^{\frac{1}{2}} + \frac{a^2(3a^2-7b^2)}{6A(a^2-b^2)^2}\zeta + \frac{ia^{5/2}C_2}{72A^{\frac{3}{2}}(a^2-b^2)^{\frac{7}{2}}}\zeta^{\frac{3}{2}} + \mathcal{O}(\zeta^2), & \mathrm{Im}\,\zeta < 0, \end{cases}, \tag{B.8}$$

and $u_3(z), u_4(z)$ have regular expansions. Here, $C_2 := C_2(a,b) = 9a^4 - 30a^2b^2 + 101b^4 > 0$.

The expansions at $z = \pm\beta$ can be obtained in a similar manner, by exchanging the roles of a, b in the expansions. We obtain that:

- *Expansion at $z = +\beta$.* As $z \to \beta$, letting $\zeta = z - \beta$,

$$u_2(z) = \begin{cases} b + \frac{ib^{3/2}}{A^{\frac{1}{2}}(a^2-b^2)^{\frac{1}{2}}}\zeta^{1/2} - \frac{(7a^2-3b^2)b^2}{6A(a^2-b^2)^2}\zeta + \frac{ib^{5/2}\tilde{C}_2}{72A^{\frac{5}{2}}(a^2-b^2)^{\frac{7}{2}}}\zeta^{\frac{3}{2}} + \mathcal{O}(\zeta^2), & \mathrm{Im}\,\zeta > 0, \\ b - \frac{ib^{3/2}}{A^{\frac{1}{2}}(a^2-b^2)^{\frac{1}{2}}}\zeta^{1/2} - \frac{(7a^2-3b^2)c^2}{6A(a^2-b^2)^2}\zeta - \frac{ib^{5/2}\tilde{C}_2}{72A^{\frac{5}{2}}(a^2-b^2)^{\frac{7}{2}}}\zeta^{\frac{3}{2}} + \mathcal{O}(\zeta^2), & \mathrm{Im}\,\zeta < 0, \end{cases} \tag{B.9}$$

$$u_4(z) = \begin{cases} b - \frac{ib^{3/2}}{A^{\frac{1}{2}}(a^2-b^2)^{\frac{1}{2}}}\zeta^{1/2} - \frac{(7a^2-3b^2)b^2}{6A(a^2-b^2)^2}\zeta - \frac{ib^{5/2}\tilde{C}_2}{72A^{\frac{5}{2}}(a^2-b^2)^{\frac{7}{2}}}\zeta^{\frac{3}{2}} + \mathcal{O}(\zeta^2), & \mathrm{Im}\,\zeta > 0, \\ b + \frac{ib^{3/2}}{A^{\frac{1}{2}}(a^2-b^2)^{\frac{1}{2}}}\zeta^{1/2} - \frac{(7a^2-3b^2)b^2}{6A(a^2-b^2)^2}\zeta + \frac{ib^{5/2}\tilde{C}_2}{72A^{\frac{5}{2}}(a^2-b^2)^{\frac{7}{2}}}\zeta^{\frac{3}{2}} + \mathcal{O}(\zeta^2), & \mathrm{Im}\,\zeta < 0, \end{cases} \tag{B.10}$$

and $u_1(z), u_3(z)$ have regular expansions. Here, $\tilde{C}_2 := \tilde{C}_2(a,b) = 101a^4 - 30a^2b^2 + 9b^4 = C_2(b,a)$.

- *Expansion at $z = -\beta$.* As $z \to -\beta$, letting $\zeta = z + \beta$,

$$u_2(z) = -b - \frac{b^{3/2}}{A^{\frac{1}{2}}(a^2-b^2)^{\frac{1}{2}}}\zeta^{1/2} - \frac{(7a^2-3b^2)b^2}{6A(a^2-b^2)^2}\zeta - \frac{b^{5/2}\tilde{C}_2}{72A^{\frac{5}{2}}(a^2-b^2)^{\frac{7}{2}}}\zeta^{\frac{3}{2}} + \mathcal{O}(\zeta^2), \tag{B.11}$$

$$u_3(z) = -b + \frac{b^{3/2}}{A^{\frac{1}{2}}(a^2-b^2)^{\frac{1}{2}}}\zeta^{1/2} - \frac{(7a^2-3b^2)c^2}{6A(a^2-b^2)^2}\zeta + \frac{b^{5/2}\tilde{C}_2}{72A^{\frac{5}{2}}(a^2-b^2)^{\frac{7}{2}}}\zeta^{\frac{3}{2}} + \mathcal{O}(\zeta^2), \tag{B.12}$$

and $u_1(z), u_4(z)$ have regular expansions. Here, $\tilde{C}_2 := \tilde{C}_2(a,b) = 101a^4 - 30a^2b^2 + 9b^4 = C_2(b,a)$.

B.2 Expansions of the Uniformizing Coordinate at the Multicritical Point.

- *Expansion at $z = +\alpha$.* Let $\zeta := z - \alpha$, $A = A(1,1) = \sqrt{6/5}$; as $\zeta \to 0$, we have the expansions:

$$u_1(z) = 1 + \left(\frac{3}{4A}\right)^{1/3}\zeta^{1/3} + \frac{3}{4}\left(\frac{3}{4A}\right)^{2/3}\zeta^{2/3} + \frac{21}{64A}\zeta + \frac{37}{192}\left(\frac{3}{4A}\right)^{4/3}\zeta^{4/3} + \mathcal{O}(\zeta^{5/3}) \tag{B.13}$$

$$u_2(z) = \begin{cases} 1 + \left(\frac{3}{4A}\right)^{1/3}\omega^2\zeta^{1/3} + \frac{3}{4}\left(\frac{3}{4A}\right)^{2/3}\omega\zeta^{2/3} + \frac{21}{64A}\zeta + \frac{37}{192}\left(\frac{3}{4A}\right)^{4/3}\omega^2\zeta^{4/3} + \mathcal{O}(\zeta^{5/3}), & \operatorname{Im}\zeta > 0, \\ 1 + \left(\frac{3}{4A}\right)^{1/3}\omega\zeta^{1/3} + \frac{3}{4}\left(\frac{3}{4A}\right)^{2/3}\omega^2\zeta^{2/3} + \frac{21}{64A}\zeta + \frac{37}{192}\left(\frac{3}{4A}\right)^{4/3}\omega\zeta^{4/3} + \mathcal{O}(\zeta^{5/3}), & \operatorname{Im}\zeta < 0 \end{cases}, \tag{B.14}$$

$$u_3(z) = -\frac{1}{3} + \frac{1}{64A}\zeta - \frac{27}{16384A^2}\zeta^2 + \frac{891}{4194304A^3}\zeta^3 + \mathcal{O}(\zeta^4), \tag{B.15}$$

$$u_4(z) = \begin{cases} 1 + \left(\frac{3}{4A}\right)^{1/3}\omega\zeta^{1/3} + \frac{3}{4}\left(\frac{3}{4A}\right)^{2/3}\omega^2\zeta^{2/3} + \frac{21}{64A}\zeta + \frac{37}{192}\left(\frac{3}{4A}\right)^{4/3}\omega\zeta^{4/3} + \mathcal{O}(\zeta^{5/3}), & \operatorname{Im}\zeta > 0, \\ 1 + \left(\frac{3}{4A}\right)^{1/3}\omega^2\zeta^{1/3} + \frac{3}{4}\left(\frac{3}{4A}\right)^{2/3}\omega\zeta^{2/3} + \frac{21}{64A}\zeta + \frac{37}{192}\left(\frac{3}{4A}\right)^{4/3}\omega^2\zeta^{4/3} + \mathcal{O}(\zeta^{5/3}), & \operatorname{Im}\zeta < 0. \end{cases} \tag{B.16}$$

- *Expansion at $z = -\alpha$.* Let $\zeta := z + \alpha$, $A = \sqrt{6/5}$; as $\zeta \to 0$, we have the expansions:

$$u_1(z) = \begin{cases} -1 + \left(\frac{3}{4A}\right)^{1/3}\omega\zeta^{1/3} - \frac{3}{4}\left(\frac{3}{4A}\right)^{2/3}\omega^2\zeta^{2/3} + \frac{21}{64A}\zeta - \frac{37}{192}\left(\frac{3}{4A}\right)^{4/3}\omega\zeta^{4/3} + \mathcal{O}(\zeta^{5/3}), & \operatorname{Im}\zeta > 0, \\ -1 + \left(\frac{3}{4A}\right)^{1/3}\omega^2\zeta^{1/3} - \frac{3}{4}\left(\frac{3}{4A}\right)^{2/3}\omega\zeta^{2/3} + \frac{21}{64A}\zeta - \frac{37}{192}\left(\frac{3}{4A}\right)^{4/3}\omega^2\zeta^{4/3} + \mathcal{O}(\zeta^{5/3}), & \operatorname{Im}\zeta < 0, \end{cases} \tag{B.17}$$

$$u_2(z) = \begin{cases} -1 + \left(\frac{3}{4A}\right)^{1/3}\omega^2\zeta^{1/3} - \frac{3}{4}\left(\frac{3}{4A}\right)^{2/3}\omega\zeta^{2/3} + \frac{21}{64A}\zeta - \frac{37}{192}\left(\frac{3}{4A}\right)^{4/3}\omega^2\zeta^{4/3} + \mathcal{O}(\zeta^{5/3}), & \operatorname{Im}\zeta > 0, \\ -1 + \left(\frac{3}{4A}\right)^{1/3}\omega\zeta^{1/3} - \frac{3}{4}\left(\frac{3}{4A}\right)^{2/3}\omega^2\zeta^{2/3} + \frac{21}{64A}\zeta - \frac{37}{192}\left(\frac{3}{4A}\right)^{4/3}\omega\zeta^{4/3} + \mathcal{O}(\zeta^{5/3}), & \operatorname{Im}\zeta < 0, \end{cases} \tag{B.18}$$

$$u_3(z) = -1 + \left(\frac{3}{4A}\right)^{1/3}\zeta^{1/3} - \frac{3}{4}\left(\frac{3}{4A}\right)^{2/3}\zeta^{2/3} + \frac{21}{64A}\zeta - \frac{37}{192}\left(\frac{3}{4A}\right)^{4/3}\zeta^{4/3} + \mathcal{O}(\zeta^{5/3}), \tag{B.19}$$

$$u_4(z) = \frac{1}{3} + \frac{1}{64A}\zeta + \frac{27}{16384A^2}\zeta^2 + \frac{891}{4194304A^3}\zeta^3 + \mathcal{O}(\zeta^4). \tag{B.20}$$

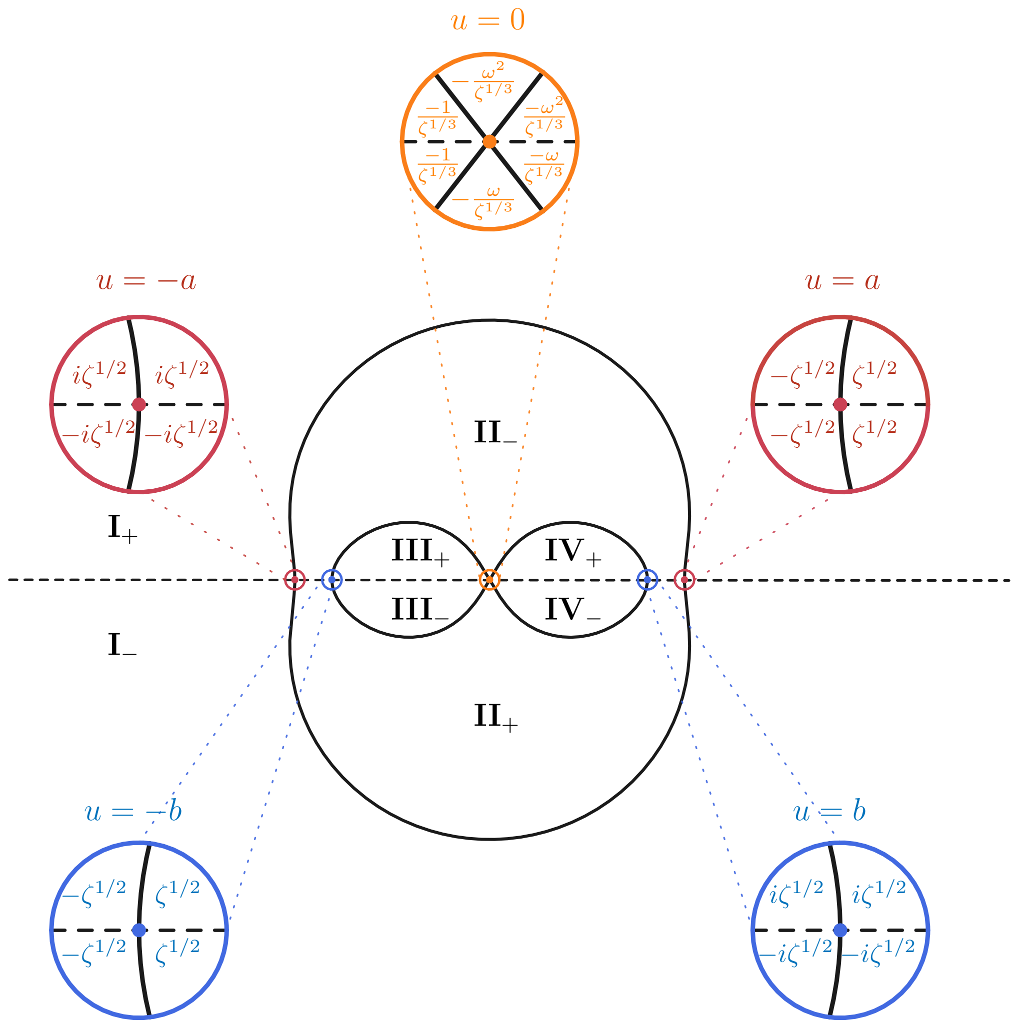


Figure B.1. The uniformizing plane for generic values of the parameters (a, b). The leading asymptotic behavior of $u_j(z)$ on each sheet is given in the appropriate local coordinate. For example, in a neighborhood of $u = a$, the local coordinate $\zeta = const \cdot (z - \beta)$, and $u_2(z) \sim -\zeta^{1/2}[1 + \mathcal{O}(\zeta^{1/2})]$, whereas $u_3(z) \sim \zeta^{1/2}[1 + \mathcal{O}(\zeta^{1/2})]$.

APPENDIX C

DERIVATION OF THE KdV_3 STRING EQUATION.

Consider the model Riemann-Hilbert problem for the function Ψ introduced in the local parametrices, defined by the function

$$F(\xi) = \text{diag}(1, \xi^{1/3}, \xi^{-1/3}) \cdot C(\xi), \tag{C.1}$$

where

$$C(\xi) = \begin{cases} \begin{pmatrix} 1 & 1 & 1 \\ \omega^2 & \omega & 1 \\ \omega & \omega^2 & 1 \end{pmatrix}, & \xi > 0, \\ \begin{pmatrix} 1 & -1 & 1 \\ \omega^2 & -1 & \omega \\ \omega & -1 & \omega^2 \end{pmatrix}, & \xi < 0. \end{cases} \tag{C.2}$$

The jumps of Ψ are defined as

$$\Psi_+(\xi; \eta, \mu, \nu) = \Psi_-(\xi; \eta, \mu, \nu) \times \begin{cases} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, & \xi \in L_1, \\ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix}, & \xi \in L_2 \cup L_8, \\ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix}, & \xi \in L_3, \\ \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, & \xi \in L_4 \cup L_6, \\ \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, & \xi \in L_5, \\ \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, & \xi \in L_7, \end{cases} \tag{C.3}$$

where the contours $L_1, ..., L_8$ are as in Figure (5.15), and satisfying the normalization condition

$$\Psi(\xi; \eta, \mu, \nu) = \left[\mathbb{I} + \mathcal{O}(\xi^{-1})\right] F(\xi) e^{-\vartheta(\hat{D}\xi^{1/3}; \eta, \mu, \nu)}, \qquad |\xi| \to \infty. \tag{C.4}$$

where $\vartheta(\xi; \eta, \mu, \nu) := \frac{3}{7}\xi^7 + \eta\xi^5 + \mu\xi^2 + \nu\xi$, $F(\xi)$ is as defined above, and $\hat{D}$ is the matrix

$$\hat{D} := \begin{cases} \text{diag } (1, \omega^2, \omega), & \text{Im } \xi > 0, \\ \text{diag } (1, \omega, \omega^2), & \text{Im } \xi < 0. \end{cases} \tag{C.5}$$

We immediately set $\mu = 0$, as it is irrelevant for our current considerations, and set $\Psi(\xi;\eta,\mu,\nu) := \Psi(\xi;\eta,\nu)$, by a slight abuse of notation. We tacitly assume the existence of the solution to this problem; we postpone the analysis of this to a later work. Here, we show how Kazakov's string equation may be derived from the string equation for the operators

$$\Xi(\xi;\eta,\nu) := \frac{\partial \Psi}{\partial \xi}\Psi^{-1}, \tag{C.6}$$

$$V(\xi;\eta,\nu) := \frac{\partial \Psi}{\partial \nu}\Psi^{-1}, \tag{C.7}$$

$$U(\xi;\eta,\nu) := \frac{\partial \Psi}{\partial \eta}\Psi^{-1} \tag{C.8}$$

Where $\Xi(\xi;\eta,\nu) = \Xi_2\xi^2 + \Xi_1\xi + \Xi_0$, $U(\xi;\eta,\nu) = U_1\xi + U_0$, and $V(\xi;\eta,\nu) = V_2\xi^2 + V_1\xi + V_0$. The matrices $\Xi_k(\eta,\nu)$, $V_k(\eta,\nu)$, and $U_k(\eta,\nu)$ are given by

$$\Xi_2 := \begin{pmatrix} 0&0&0\\ 0&0&1\\ 0&0&0 \end{pmatrix}, \qquad \Xi_1(\eta,\nu) := \begin{pmatrix} 0 & 1 & \frac{5}{3}\eta+a_{12} \\ \frac{5}{3}\eta-a_{31} & -a_{32} & a_{22}-a_{33} \\ 1 & 0 & a_{32} \end{pmatrix}, \tag{C.9}$$

and $\Xi_0(\eta,\nu)$ is given by the more complicated expression

$$\begin{aligned} \Xi_0(\eta,\nu) &:= \frac{5}{3}\eta\begin{pmatrix} a_{12}-a_{31} & -a_{32} & a_{11}-a_{33} \\ a_{22}-a_{11} & -a_{12} & a_{21}-a_{13} \\ a_{32} & 1 & a_{31} \end{pmatrix} + \\ &= \begin{pmatrix} a_{13}-a_{21}-a_{31}a_{12} & a_{11}-a_{22}-a_{12}a_{32} & b_{12}-a_{23}-a_{12}a_{33} \\ (a_{11}-a_{22}+a_{33})a_{31}+a_{32}a_{21}-b_{31}+a_{23} & a_{31}a_{12}+a_{33}a_{32}+a_{21}-b_{32} & a_{13}a_{31}-a_{22}a_{33}+a_{23}a_{32}+a_{33}^2+b_{22}-b_{33}+\frac{\nu}{3} \\ a_{33}-a_{11}-a_{32}a_{31} & a_{31}-a_{12}-a_{32}^2 & b_{32}-a_{13}-a_{32}a_{33} \end{pmatrix} \end{aligned} \tag{C.10}$$

$$V_1 := \begin{pmatrix} 0&0&0\\ 0&0&1\\ 0&0&0 \end{pmatrix}, \qquad V_0(\eta,\nu) := \begin{pmatrix} 0 & 1 & a_{12} \\ -a_{31} & -a_{32} & a_{22}-a_{33} \\ 1 & 0 & a_{32} \end{pmatrix}, \tag{C.11}$$

and, finally,

$$\begin{aligned} &U_2 := \begin{pmatrix} 0&0&1\\ 1&0&0\\ 0&0&0 \end{pmatrix}, \qquad U_1(\eta,\nu) := \begin{pmatrix} a_{12}-a_{31} & -a_{32} & a_{11}-a_{33} \\ a_{22}-a_{11} & -a_{12} & a_{21}-a_{13} \\ a_{32} & 1 & a_{31} \end{pmatrix}, \\ &U_0(\eta,\nu) := \\ &\begin{pmatrix} a_{33}a_{31}+a_{32}a_{21}-a_{11}a_{12}+b_{12}-b_{31} & a_{32}(a_{33}+a_{22}-a_{11})-a_{12}^2-b_{32}+a_{13} & a_{31}a_{13}-a_{13}a_{12}-a_{11}a_{33}+a_{32}a_{23}+a_{33}^2+b_{11}-b_{33} \\ a_{11}^2-a_{11}a_{22}+a_{12}a_{21}-a_{31}a_{21}+a_{13}a_{31}+b_{22}-b_{11} & a_{13}a_{32}-a_{21}a_{32}+a_{11}a_{12}+a_{23}-b_{12} & (a_{33}-a_{22}+a_{11})a_{31}+a_{12}a_{23}-a_{21}a_{33}-b_{13}+b_{21} \\ b_{32}-a_{11}a_{32}-a_{21}-a_{31}^2 & a_{33}-a_{22}-a_{32}a_{31}-a_{32}a_{12} & b_{31}-a_{13}a_{32}-a_{33}a_{31}-a_{23} \end{pmatrix}. \end{aligned} \tag{C.12}$$

The string equation itself is the compatibility condition of the operators $\partial_\xi - \Xi(\xi;\eta,\nu)$ and $\partial_\nu - V(\xi;\eta,\nu)$:

$$[\partial_\xi - \Xi(\xi;\eta,\nu), \partial_\nu - V(\xi;\eta,\nu)] = \partial_\nu \Xi - \partial_\xi V + [\Xi, V] = 0. \tag{C.13}$$

The other compatibility conditions give additional equations which further constrain the functions a_{ij}, b_{ij}.

Let us first briefly recount some facts about the generalized KdV hierarchies, in particular KdV_3 (also known as Gelfand-Dickey-3). For further details about KdV_3, one should consult [33]. Consider the differential operator

$$L = \partial^3 - \frac{3}{2}u\partial - \frac{3}{4}u' + \frac{3}{2}v, \tag{C.14}$$

where u, v are functions of the variable of differentiation ν ($\partial = \partial_\nu$), as well as an infinite collection of 'times' t_k. The operator L is chosen so that, under an infinitesmal change of coordinates $\nu \to \nu(\lambda)$, the functions u and v transform as

$$\tilde{u}(\lambda) = u(\nu(\lambda))\left(\frac{d\nu}{d\lambda}\right)^2 + 2\{\nu, \lambda\}, \tag{C.15}$$

$$\tilde{v}(\lambda) = v(\nu(\lambda))\left(\frac{d\nu}{d\lambda}\right)^3, \tag{C.16}$$

i.e. as an affine connection and as a rank 3 tensor, respectively (here, $\{\nu, \lambda\}$ denotes the Schwarzian derivative of ν with respect to λ) . The operator $\left(\frac{d\nu}{d\lambda}\right)^2 \circ L \circ \left(\frac{d\nu}{d\lambda}\right)$ then acts covariantly from the space of rank 1 tensors to rank 2 tensors. One can expand $L^{1/3}$ uniquely as a series in pseudodifferential operators:

$$L^{1/3} = \partial + \frac{1}{2}u\partial^{-1} + \mathcal{O}(\partial^{-2}), \tag{C.17}$$

with the coefficients of L being differential polynomials in the functions u, v. The equations of the KdV_3 hierarchy are generated by the equations

$$0 = [\partial_{t_k} - L_+^{k/3}, \lambda - L], \tag{C.18}$$

with $\partial_{t_k}\lambda = 0$, and k an integer which is not a multiple of 3. Here, the subscript '+' denotes the purely differential piece of the pseudodifferential operator $L_+^{k/3}$. The first equation in the hierarchy ($k = 2$) is the celebrated Boussinesq equation. The hierarchy has an infinite set of conserved flows, characterized by the equations

$$0 = [\partial_{t_k} - L_+^{k/3}, \partial_{t_j} - L_+^{j/3}]. \tag{C.19}$$

This hierarchy has infinitely many conserved quantities, as well as an associated bihamiltonian structure, cf. [33]. This hierarchy also has a number of additional symmetries. These symmetries do not commute amongst themselves, and are characterized by their explicit dependence on the 'times' $\{t_k\}$. The additional symmetries generate an infinite-dimensional W-algebra. In the case of the usual KdV ($= \mathrm{KdV}_2$) hierarchy, this algebra is the celebrated Virasoro algebra; in the case of KdV_3, one obtains the so-called Zamolodchikov

W_3 algebra (cf. [33, 111]). The first such symmetry is generated by the equation (setting $A := L_+^{4/3}$) $[L, A] = 1$. Explicitly, the operator A is the 4^{th} order differential operator

$$A = L_+^{4/3} = \partial^4 - \{u, \partial^2\} + \{v, \partial\} + \frac{1}{2}u^2 - \frac{1}{6}u'''. \tag{C.20}$$

(Here, $\{A, B\} := AB + BA$ denotes the anticommutator of operators). There is an additional non-trivial flow that we may add, coming from the operator $L_+^{2/3} = \partial^2 - u$; it is associated with the variable η, and yields a generalized string equation:

$$[L, \tilde{A}] = 1, \qquad \text{where} \qquad -\tilde{A} := L_+^{4/3} + \frac{5}{3}\eta L_+^{2/3}. \tag{C.21}$$

This is the so-called string equation. Equation (C.21) appears in the physics literature as the descriptor of the $(4, 3)$ minimal model coupled to gravity, under various normalizations [25, 34, 56]. The interpretation of the ν flow is as a descriptor of the transition to 'criticality', i.e. the transition from the non-critical to critical matrix model. The parameter η describes the Ising phase transition along the critical curve. It is the solutions to equation (C.21) that appear explicitly at the critical point of the matrix model in question. The string equation may be expressed explicitly as a pair of differential equations for the functions $u(\nu), v(\nu)$:

$$-1 = \partial\left[\frac{1}{12}u^{(4)} - \frac{3}{4}uu'' - \frac{3}{8}(u')^2 + \frac{1}{2}u^3 + \frac{3}{2}v^2 - \frac{5}{12}\eta(3u^2 - u'')\right], \tag{C.22}$$

$$0 = \partial\left[\frac{1}{2}v'' - \frac{3}{2}uv + \frac{5}{2}\eta v\right]. \tag{C.23}$$

Let us now return to the Lax operators arising from the local parametrices of the critical 2-matrix model. The string equation for the matrix model reads

$$0 = [\partial_\xi - \Xi(\xi; \eta, \nu), \partial_\nu - V(\xi; \eta, \nu)] = \partial_\nu \Xi - \partial_\xi V + [\Xi, V]. \tag{C.24}$$

We claim that the above equation can be reduced to a pair of equations for two functions, which we will show are equivalent to the KdV_3 string equation (C.21) (or, equivalently, (C.22) and (C.23)). This is an inevitably involved calculation; nevertheless, if one is careful enough, it is possible to recover the desired result. Here, we indicate the sequence of steps by which one may derive the string equation from (C.24). First, one should note that Equation (C.24) is already a first-order polynomial in ξ. Then, if one follows the following procedure:

- Solve the $\mathcal{O}(\xi)$ term in the 1-3 entry of the string equation for a_{11},

- Solve the $\mathcal{O}(\xi)$ term in the 2-1 entry of the string equation for a_{22},
- Solve the $\mathcal{O}(\xi)$ term in the 3-3 entry of the string equation for a_{31},
- Solve the $\mathcal{O}(\xi)$ term in the 2-3 entry of the string equation for b_{32} (from here on, the string equation is $\mathcal{O}(1)$, and so we refer only to the entry number),
- Solve the 1-1 entry of the string equation for b_{31},
- Solve the 1-2 entry of the string equation for a_{21},
- Solve the 1-3 entry of the string equation for b_{33}, and finally
- Solve the 2-2 entry of the string equation for b_{12}.

After this procedure, we are left with two independent equations for the unknown functions $a_{12}(\nu, \eta)$, $a_{32}(\nu, \eta)$ (some of the other functions, such as a_{33}, are eliminated automatically). Making the substitutions

$$a_{12}(\nu) = \frac{1}{2}\int^{\nu} v(t)dt + \frac{1}{4}u(\nu) - \frac{1}{8}\left(\int^{\nu} u(t)dt\right)^2, \tag{C.25}$$

$$a_{32}(\nu) = -\frac{1}{2}\int^{\nu} u(t)dt, \tag{C.26}$$

the remaining pair of equations are converted into (after an integration):

$$-1 = \frac{\partial}{\partial\nu}\left[\frac{1}{12}u^{(4)} - \frac{3}{4}uu'' - \frac{3}{8}(u')^2 + \frac{1}{2}u^3 + \frac{3}{2}v^2 - \frac{5}{12}\eta(3u^2 - u'')\right], \tag{C.27}$$

$$0 = \frac{\partial}{\partial\nu}\left[\frac{1}{2}v'' - \frac{3}{2}uv + \frac{5}{2}\eta v\right]; \tag{C.28}$$

these are precisely the string equation (C.21). One can work backwards from here to see that the matrices $\Xi(\xi; \eta, \nu)$ and $V(\xi; \eta, \nu)$ are completely determined in terms of the functions u and v; for example,

$$V(\xi; \eta, \nu) = E_{23}\xi + \begin{pmatrix} 0 & 1 & \frac{1}{2}\int v d\nu + \frac{1}{4}u - \frac{1}{8}(\int u d\nu)^2 \\ -\frac{1}{2}\int v d\nu + \frac{1}{4}u - \frac{1}{8}(\int u d\nu)^2 & -\frac{1}{2}\int u d\nu & v + \frac{1}{2}\int u d\nu \int v d\nu \\ 1 & 0 & \frac{1}{2}u d\nu \end{pmatrix} \tag{C.29}$$

One may also determine the dependence of the functions u, v on ν, using the zero-curvature equation between $\partial_\xi - \Xi$ and $\partial_\nu - U$; one finds that

$$\frac{\partial}{\partial\eta}u(\eta, \nu) = \frac{\partial}{\partial\nu}\left[\frac{1}{3}\left(\frac{5}{3}\eta - \frac{1}{2}u\right)u'' + \frac{1}{8}(u')^2 + \frac{1}{4}u^3 - \frac{5}{3}\eta u^2 - \frac{1}{2}v^2 + \frac{4}{3}\nu\right], \tag{C.30}$$

$$\frac{\partial}{\partial\eta}v(\eta, \nu) = \frac{1}{12}vu''' - \frac{1}{6}u''v' + \left(-\frac{25}{9}\eta^2 - \frac{5}{6}\eta u + \frac{1}{4}u^2\right)v' + \frac{5}{12}\left(\eta - \frac{3}{5}u\right)u'v. \tag{C.31}$$

APPENDIX D
THE ISOMONODROMIC TAU FUNCTION.

The partition function of the 2-matrix model was identified with an isomonodromic $\boldsymbol{\tau}$-function by Bertola and Marchal, cf. [9]. Their derivation is relatively straightforward, and applies almost directly to our situation. However, some of the details of the calculation are different enough that the proof merits discussion. We present the proof of the fact that the partition function for the 2-matrix model is an isomonodromic $\boldsymbol{\tau}$-function here. The proof mirrors almost directly that of Bertola and Marchal's; one should consult their work and references therein for further details and commentary. For sake of readability, let us introduce the notation, for a given matrix-valued 1-form $A(z)$,

$$\langle A(z)\rangle := \operatorname*{Res}_{z=\infty} \operatorname{tr}\left[A(z)\right].$$

Recall that, in general, the n^{th} monic biorthogonal polynomial $p_n(z)$ with respect to the weight [1]

$$e^{NW(z,w)} := e^{N\left[\tau z w-\frac{1}{2}z^2-\frac{t}{4}z^4-\frac{1}{2}w^2-\frac{\bar{t}}{4}w^4\right]} \tag{D.1}$$

(we take $\bar{t}$, t as independent parameters in general) on contour(s) (Γ,Γ) is given in terms of the solution to the following Riemann-Hilbert problem:

1. $\mathbf{Y}_n(z)$ is a piecewise analytic function in $\mathbb{C}\setminus\Gamma$,

2. $\mathbf{Y}_n(z)$ has boundary values

$$\mathbf{Y}_{n,+}(z) = \mathbf{Y}_{n,-}(z)e^{-NV(z)}\left[\mathbb{I}+w(z)E_{12}+\frac{1}{N\tau}w'(z)E_{13}+\frac{1}{N^2\tau^2}w''(z)E_{14}\right], z\in\Gamma, \tag{D.2}$$

where $w(z) := \int_\Gamma e^{N\left[\tau z w-\frac{1}{2}w^2-\frac{t}{4}w^4\right]}dw$,

[1]Here, we take $N>0$ as a free parameter, in general different from the index of the biorthogonal polynomial. We will later set $N=n$.

3. As $z \to \infty$,

$$\mathbf{Y}_n(z) = \left[\mathbb{I} + \frac{Y_{1,n}}{z} + \mathcal{O}\left(\frac{1}{z^2}\right)\right] \begin{pmatrix} z^n & 0 & 0 \\ 0 & z^{-m_n-1}\mathbb{I}_{r_n} & 0 \\ 0 & 0 & z^{-m_n}\mathbb{I}_{3-r_n} \end{pmatrix}, \tag{D.3}$$

where $m_n \in \mathbb{N}$, $r_n \in \{0,1,2\}$ are such that $n = 3m_n + r_n$.

Note that, when $n = 3k$ is a multiple of 3, the above coincides with the Riemann Hilbert problem analyzed in the present work.

We now summarize some of the basic results pertaining to this RHP which we will need in our analysis of the τ-differential.

The solution to this Riemann-Hilbert problem is given explicitly:

$$\mathbf{Y}_n(z) = \begin{pmatrix} p_n(z) & C_\Gamma[p_n w](z) & \frac{1}{N\tau}C_\Gamma[p_n w'](z) & \frac{1}{N^2\tau^2}C_\Gamma[p_n w''](z) \\ Q_{n-1}(z) & C_\Gamma[Q_{n-1}w](z) & \frac{1}{N\tau}C_\Gamma[Q_{n-1}w'](z) & \frac{1}{N^2\tau^2}C_\Gamma[Q_{n-1}w''](z) \\ Q_{n-2}(z) & C_\Gamma[Q_{n-2}w](z) & \frac{1}{N\tau}C_\Gamma[Q_{n-2}w'](z) & \frac{1}{N^2\tau^2}C_\Gamma[Q_{n-2}w''](z) \\ Q_{n-3}(z) & C_\Gamma[Q_{n-3}w](z) & \frac{1}{N\tau}C_\Gamma[Q_{n-3}w'](z) & \frac{1}{N^2\tau^2}C_\Gamma[Q_{n-3}w''](z), \end{pmatrix} \tag{D.4}$$

where Q_{n-1}, Q_{n-2}, and Q_{n-3} are some appropriately chosen polynomials of degrees $n-1$, $n-2$, and $n-3$, respectively, $p_n(z)$ is the n^{th} monic biorthogonal polynomial, and

$$C_\Gamma[f](z) := \frac{1}{2\pi i}\int_\Gamma \frac{f(x)}{x-z}dz \tag{D.5}$$

denotes the Cauchy transform with respect to the contour Γ. We can also relate the Riemann-Hilbert problem for $\mathbf{Y}_n(z)$ to the Riemann-Hilbert problem for $\mathbf{Y}_{n+1}(z)$ by means of a *raising operator*:

$$\mathbf{Y}_{n+1}(z) = R_n(z)\,\mathbf{Y}_n(z), \tag{D.6}$$

where $R_n(z) := R_n^{(1)}z + R_n^{(0)}$ is a degree 1 matrix-valued polynomial in z. The existence of $R_n(z)$ follows immediately from the fact that $\mathbf{Y}_{n+1}(z)\,\mathbf{Y}_n^{-1}(z)$ has no jumps, and thus extends to an entire function. The asymptotics of $\mathbf{Y}_{n+1}(z)\,\mathbf{Y}_n^{-1}(z)$ uniquely fix the form of $R_n(z)$. Setting $\alpha_0 := r_N + 1$, we have that

$R_n^{(1)} = E_{1,1}$, whereas the matrix $(R_n^0)_{jk}$ has entries as given in the following table:

$$
\begin{array}{c|ccc}
 & k=\alpha_0 & k=1 & k\neq 1,\alpha_0 \\
\hline
j=\alpha_0 & \frac{-(Y_{n,2})_{\alpha_0,1}+\sum_{\ell\neq\alpha_0}(Y_{N,1})_{\alpha_0,\ell}(Y_{N,1})_{\ell,1}}{(Y_{n,1})_{\alpha_0,1}} & -(Y_{n,1})_{\alpha_0,1} & -(Y_{n,1})_{\alpha_0,k} \\
j=1 & \frac{1}{(Y_{n,1})_{\alpha_0,1}} & 0 & 0 \\
j\neq 1,\alpha_0 & \frac{(Y_{n,1})_{j,1}}{(Y_{n,1})_{\alpha_0,1}} & 0 & \delta_{jk}
\end{array}
\tag{D.7}
$$

Note that the matrix $R_n(z)$ is determined entirely in terms of $\mathbf{Y}_n(z)$.

The isomonodromic $\boldsymbol{\tau}$-differential corresponding to $\mathbf{Y} := \mathbf{Y}_n$ is defined to be

$$d\log\boldsymbol{\tau}_n := \langle \mathbf{Y}_n^{-1}\,\mathbf{Y}_n'\,d\hat{\mathbf{W}}\hat{\mathbf{W}}^{-1}\rangle. \tag{D.8}$$

where $\hat{\mathbf{W}}(z)$ is essentially the augmented $\mathbf{W}$-matrix from the first transformation:

$$\hat{\mathbf{W}}(z) := \begin{pmatrix} e^{-NV(z)} & 0 \\ 0 & \mathbf{W}^{-1}(z) \end{pmatrix},$$

with the only difference here being that $V(z) = \frac{1}{2}z^2 + \frac{\bar{t}}{4}z^4$. For now, we treat the parameter N in the matrix $\hat{\mathbf{W}}$ as a *fixed* parameter independent of the index of the polynomial n; we shall later set $N = n$. Since multiplication by $\hat{\mathbf{W}}$ yields a constant jump RHP, we have the following formulae for the differential of $d\hat{\mathbf{W}}\hat{\mathbf{W}}^{-1}$: renders $\mathbf{Y}$

$$d\hat{\mathbf{W}}\hat{\mathbf{W}}^{-1} = \begin{pmatrix} 0 & 0 \\ 0 & \frac{\partial \mathbf{W}}{\partial\tau}\mathbf{W}^{-1} \end{pmatrix} d\tau + \begin{pmatrix} 0 & 0 \\ 0 & \frac{\partial \mathbf{W}}{\partial t}\mathbf{W}^{-1} \end{pmatrix} dt + \begin{pmatrix} z^4/4 & 0 \\ 0 & 0 \end{pmatrix} d\bar{t}.$$

When $\bar{t} = t$, we must have that the coefficients of dt, $d\bar{t}$ must be equal. This being the case, we may calculate the $\boldsymbol{\tau}$-differential by instead taking twice the second coefficient matrix:

$$d\log\boldsymbol{\tau}_n(\tau,t) = \langle \mathbf{Y}_n^{-1}\,\mathbf{Y}_n' \begin{pmatrix} 0 & 0 \\ 0 & \frac{\partial \mathbf{W}}{\partial\tau}\mathbf{W}^{-1} \end{pmatrix}\rangle d\tau + 2\times\langle \mathbf{Y}_n^{-1}\,\mathbf{Y}_n' \begin{pmatrix} 0 & 0 \\ 0 & \frac{\partial \mathbf{W}}{\partial t}\mathbf{W}^{-1} \end{pmatrix}\rangle dt. \tag{D.9}$$

In [9], the parameters of the isomonodromic $\boldsymbol{\tau}$-differential come from the coefficients of the potential (as opposed to our case, where one of the parameters is the coefficient of the interaction term XY; namely, τ), and the definition of $d\hat{\mathbf{W}}\hat{\mathbf{W}}^{-1}$ is slightly different. However, it is only important to the proof that $\hat{\mathbf{W}}$ renders the jumps of $\mathbf{Y}$ constant, which we have already seen (indeed, this was the point of the first transformation $\mathbf{Y} \mapsto \mathbf{X}$).

The biorthogonal polynomials double as a particular sequence of multiple orthogonal polynomials; Summarizing the arguments of [9], one can use the sequence of raising operators arising from the multiple orthogonality to produce the next *biorthogonal polynomial* in the sequence. Let us denote this raising operator generically by $R_n(z)$. $R_n(z)$ is defined so that

$$\mathbf{Y}_{n+1}(z) = R_n(z)\,\mathbf{Y}_n(z). \tag{D.10}$$

Generically, $R_n(z)$ is a degree 1 polynomial in z; its inverse is also a degree 1 polynomial in z. We have the following proposition:

Proposition D.1.

$$d\log\frac{\boldsymbol{\tau}_{n+1}}{\boldsymbol{\tau}_n} = -\langle R_n^{-1}R_n' d\,\mathbf{Y}_n\,\mathbf{Y}_n^{-1}\rangle \tag{D.11}$$

Proof. By equation (D.10), we have that

$$\mathbf{Y}_{n+1}^{-1}\,\mathbf{Y}_{n+1}' = \mathbf{Y}_n^{-1}\,R_n^{-1}R_n'\,\mathbf{Y}_n + \mathbf{Y}_n^{-1}\,\mathbf{Y}_n';$$

Thus, the quotient of $\boldsymbol{\tau}$ differentials is

$$\begin{aligned} d\log\frac{\boldsymbol{\tau}_{n+1}}{\boldsymbol{\tau}_n} &= \left\langle \left[\mathbf{Y}_n^{-1}\,R_n^{-1}R_n'\,\mathbf{Y}_n + \mathbf{Y}_n^{-1}\,\mathbf{Y}_n'\right] d\hat{\mathbf{W}}\hat{\mathbf{W}}^{-1}\right\rangle - \langle \mathbf{Y}_n^{-1}\,\mathbf{Y}_n'\,d\hat{\mathbf{W}}\hat{\mathbf{W}}^{-1}\rangle \\ &= \langle \mathbf{Y}_n^{-1}\,R_n^{-1}R_n'\,\mathbf{Y}_n\,d\hat{\mathbf{W}}\hat{\mathbf{W}}^{-1}\rangle. \end{aligned}$$

Now, recall that $\mathbf{Y}_n\,\mathbf{W}$ has constant jumps, and so the differential $d\,[\mathbf{Y}_n\,\mathbf{W}]\,\mathbf{W}^{-1}\,\mathbf{Y}_n^{-1}$ has coefficients which are polynomial in z, by the standard Liouville argument. This statement can be rewritten as

$$\mathbf{Y}_n\,d\hat{\mathbf{W}}\hat{\mathbf{W}}^{-1}\,\mathbf{Y}_n^{-1} = d\,\mathbf{Y}_n\,\mathbf{Y}_n^{-1} + \textit{polynomial}.$$

Inserting the above into our expression for the $\boldsymbol{\tau}$-quotient, we obtain that

$$d\log\frac{\boldsymbol{\tau}_{n+1}}{\boldsymbol{\tau}_n} = \left\langle R_n^{-1}R_n'[\textit{polynomial}] - R_n^{-1}R_n' d\,\mathbf{Y}_n\,\mathbf{Y}_n^{-1}\right\rangle; \tag{D.12}$$

since $R_n^{-1}R_n'$ is a polynomial in z, the first term is a polynomial in z, and thus has no residues at infinity. This completes the proof. $\square$

We can use the explicit form of the raising operators obtained before to get an expression for the τ-differential in terms of the coefficients of $\mathbf{Y}_n(z)$. The exact expression is summarized by the following proposition.

Proposition D.2. *The ratio of consecutive $\boldsymbol{\tau}$ differentials, up to multiplication by a function independent of the isomonodromic times, is given by*

$$\frac{\boldsymbol{\tau}_{n+1}}{\boldsymbol{\tau}_n} = (Y_{1,n})_{1,\alpha_0} \tag{D.13}$$

Proof. This follows immediately from inspection of the previous proposition, and the explicit form of the matrices $R_n(z)$. For details, see [9]. □

Furthermore, we can relate the coefficients $(Y_{1,n})_{1,\alpha_0}$ to the biorthogonality coefficients h_n:

Proposition D.3. *The matrix coefficient $(Y_{1,n})_{1,\alpha_0}$ is given in terms of the n^{th} normalizing constant of the biorthogonal polynomials:*

$$(Y_{1,n})_{1,\alpha_0} = \left(\frac{t}{\tau}\right)^S h_n, \tag{D.14}$$

where $S \in \mathbb{N}$, $\alpha_0 \in \{0, 1, 2\}$ are such that $n = 3S + \alpha_0 - 1$.

Proposition D.4. *The isomonodromic $\boldsymbol{\tau}$-function $\boldsymbol{\tau}_n$ is related to the partition function $\mathcal{Z}_n$ for n a multiple of 3 by the formula*

$$\mathcal{Z}_n = \left(\frac{\tau}{t^2}\right)^{\frac{n}{2}\left(\frac{n}{3}-1\right)} \boldsymbol{\tau}_n. \tag{D.15}$$

Remark D.5. We remark that the statement of the above Proposition in [9] is incorrect. The proof which comes before is correct, but the power of (t/τ) should be one over what it reads in their Theorem 3.4 (page 17); there are no essential changes to the results otherwise.