

The Ising Model Coupled to 2D Gravity: Genus Zero Partition Function

Maurice DUITTS ^a, Nathan HAYFORD ^a and Seung-Yeop LEE ^b

^{a)} Department of Mathematics, Royal Institute of Technology (KTH), Stockholm, Sweden

E-mail: duits@kth.se, nhayford@kth.se

^{b)} Department of Mathematics and Statistics, University of South Florida, Tampa, FL, USA

E-mail: lees3@usf.edu

Received January 31, 2025, in final form September 03, 2025; Published online September 24, 2025

<https://doi.org/10.3842/SIGMA.2025.079>

Abstract. We compute the genus 0 free energy for the 2-matrix model with quartic interactions, which acts as a generating function for the Ising model's partition function on a random, 4-regular, planar graph. This is consistent with the predictions of Kazakov and Boulatov on this model, as well as subsequent confirmation of this formula using combinatorial methods. We also provide a new parametric formula for the free energy and give a characterization of the phase space. Our analysis is based on a steepest descent Riemann–Hilbert analysis of the associated biorthogonal polynomials and the corresponding isomonodromic τ -function. A key ingredient in the analysis is a parametrization of the spectral curve. This analysis lays the groundwork for the subsequent study of the multicritical point, which we will study in a forthcoming work.

Key words: 2-matrix model; Riemann–Hilbert analysis; asymptotic analysis; graphical enumeration; Ising model

2020 Mathematics Subject Classification: 30E15; 30E10; 30E25; 41A60; 05C30; 05C10

Contents

1	Introduction	2
1.1	The Ising model coupled to gravity and the 2-matrix model	3
1.2	A family of biorthogonal polynomials	6
1.3	Analytic continuation	7
1.4	Riemann–Hilbert problem	8
1.5	Spectral curve	9
2	Statement of results	10
2.1	Asymptotic behavior of the partition function	10
2.2	Structure of the phase space	12
2.2.1	$H = 0$	12
2.2.2	The general case	13
2.3	Spectral curve	15
2.4	Phase transitions and special points on the spectral curve	16
2.5	Strategy of the proof and an associated Riemann–Hilbert problem	18
3	Definition and analysis of the spectral curve	20
3.1	Definition of the spectral curve	20
3.2	Sheet structure of the spectral curve	22
3.3	The functions Ω	23

3.4	Definition of the contours Γ , Γ_1 , and Γ_2	26
3.5	Lensing inequalities	26
3.6	The generic (noncritical) case: $0 < b < 1$, $1 < a < b^{-1}$, $0 < c < b$	32
4	The first and second transformations $\mathbf{Y} \mapsto \mathbf{X} \mapsto \mathbf{U}$	34
4.1	Idea of the transformation	34
4.2	Riemann–Hilbert problem and asymptotics for the Pearcey-type integrals $w_j(z)$	34
4.3	The transformation $\mathbf{Y} \mapsto \mathbf{X}$	39
4.4	The transformation $\mathbf{X} \mapsto \mathbf{U}$	41
5	The third and fourth transformations $\mathbf{U} \mapsto \mathbf{T} \mapsto \mathbf{S}$	43
5.1	The transformation $\mathbf{U} \mapsto \mathbf{T}$	43
5.2	The transformation $\mathbf{T} \mapsto \mathbf{S}$	45
6	Construction of parametrices and the transformation $\mathbf{S} \mapsto \mathbf{R}$	46
6.1	Global parametrix	47
6.2	Calculation of the local parametrices	53
7	Proof of main theorem and concluding remarks	57
7.1	Calculation of the free energy	57
7.2	Concluding remarks	59
A	Expansion of the uniformizing coordinate near the branch points	60
A.1	Expansion of the uniformizing coordinate in the generic and critical cases	61
A.2	Expansions of the uniformizing coordinate at the multicritical point and γ_b	62
B	The isomonodromic tau function	63
C	Explicit formulae for the spectral curve and the critical surface	68
C.1	Implicit formula for the spectral curve	68
C.2	Implicit formula for the critical surface	69
D	Parametrization of the phase space	69
D.1	The critical surfaces	69
D.2	Proof of Proposition 2.5	71
E	Calculations from physics: the genus 0 partition function	72
E.1	Wick expansion of the partition function	72
E.2	Formal calculation of the genus 0 partition function	76
E.3	Comparison with explicit calculation	79
E.4	Large V asymptotics of $\sigma(\tau, t, H)$	82
	References	87

1 Introduction

The 2-matrix model is the probability measure on the space of pairs of $n \times n$ Hermitian matrices X, Y defined by

$$\frac{1}{Z_n} \exp N \operatorname{tr}(\tau XY - V_1(X) - V_2(Y)) dX dY, \quad (1.1)$$

where dX , dY are the Haar measures on the space of $n \times n$ Hermitian matrices, $N > 0$ is a large parameter (which will eventually be set to n the size of the matrix), $\tau > 0$ is constant called the coupling constant, and V_j are two polynomials of even degree and with positive leading coefficient. The normalizing constant Z_n , which depends on n , N , τ , and any parameters present in V_1 , V_2 , is called the *partition function*, and will essentially be the main object of study in this work.

An important motivation for studying the partition function of the 2-matrix model comes from its connection to 2-dimensional quantum gravity [21, 34, 47, 48]. It is conjectured in the physics literature that, by tuning the parameters in the polynomials V_1 and V_2 , the 2-matrix model contains all multi-critical models of type (q, p) with arbitrary q , whereas the one matrix models ($\tau = 0$) only contains critical models with $q = 2$. Perhaps the most important first example is the case of two quartic potentials, which is related to the Ising model coupled to 2-dimensional gravity and the type $(3, 4)$ critical model. The richer structure also makes the 2-matrix model harder to analyze when compared to the well-understood one matrix model and a rigorous treatment of the asymptotic behavior as $N \rightarrow \infty$ of the general 2-matrix model is an important open problem. To date, the only rigorous results in the literature are in [35, 39, 40], but these results do not cover the case of quantum gravity coupled to the Ising model. The current writing is the first in a series of three papers where we discuss this multi-critical point. More precisely, we will study the limiting behavior of its partition function Z_n as $n = N \rightarrow \infty$ in case of the quartic potentials:

$$V_1(x) = V(x; e^H t) := \frac{1}{2}x^2 + \frac{e^H t}{4}x^4 \quad \text{and} \quad V_2(y) = V(y; e^{-H} t) := \frac{1}{2}y^2 + \frac{e^{-H} t}{4}y^4,$$

where $H \in \mathbb{R}$ and t are parameters. In this case, we write the partition function as

$$Z_n(\tau, t, H; N) := \iint \exp N \operatorname{tr}(\tau XY - V(X; e^H t) - V(Y; e^{-H} t)) dX dY,$$

where the integration is carried out over pairs of $n \times n$ Hermitian matrices. Note that to ensure that the measure in (1.1) is finite we need that $t > 0$. However, for fixed N the partition function $Z_N(\tau, t, H; N)$ depends analytically on t and can be analytically extended to complex t with a branch point at $t = 0$. In fact, we are mainly interested in $t < 0$ and discuss this continuation in detail below. Our main results are for the *genus zero free energy* defined as

$$F(\tau, t, H) := \lim_{N \rightarrow \infty} \frac{1}{N^2} \log \frac{Z_N(\tau, t, H; N)}{Z_N(\tau, 0, 0; N)}. \quad (1.2)$$

In this paper, we will show that, up to some critical value, the analytic continuation for $t < 0$ the genus zero free energy exists, and we derive an analytic expression for it, confirming predictions from the physics literature [14, 57]. We will also see that at the multi-critical point

$$\tau = \frac{1}{4}, \quad t = -\frac{5}{72}, \quad H = 0, \quad (1.3)$$

the free energy undergoes a phase transition. In two forthcoming works [37, 49], we will address this phase transition by means of a multi-scaling limit and show that it gives rise to the $(3, 4)$ minimal model coupled to gravity, as conjectured. Before we come to the statement of our main results and strategy of the proof, we start with a more detailed historical discussion on this model.

1.1 The Ising model coupled to gravity and the 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 2nd order phase transition at finite temperature.

The model describes a ferromagnet with only nearest-neighbor interactions, and can be defined on any graph¹ $G := (V, E)$ with vertices V and edges E as follows. The Hamiltonian for the ferromagnetic Ising model is a functional on maps $\psi: V \rightarrow \{\pm 1\}$, and is defined as

$$\mathcal{H}(\psi; h) = - \sum_{(x,y) \in E} \psi(x)\psi(y) - h \sum_{x \in V} \psi(x).$$

The parameter $h \in \mathbb{R}$ is called the *magnetic field*. The partition function for this model is defined to be

$$Z_G(\beta; h) := \sum_{\psi} e^{-\beta \mathcal{H}(\psi; h)},$$

where $\beta > 0$ is a parameter called the *inverse temperature*, and the sum is taken over all $2^{|V|}$ maps $\psi: V \rightarrow \{\pm 1\}$. In other words, we are considering the Boltzmann distribution on the system at temperature β^{-1} . Typically, one is interested in calculating the *free energy* of this model, defined as

$$F_G(\beta; h) = -\frac{1}{\beta} \log Z_G(\beta; h).$$

If the graph is of infinite size (i.e., the number of vertices is infinite), one is instead interested in the free energy *per unit site*; that is, to say

$$f(\beta; h) := \lim_{|V| \rightarrow \infty} \frac{1}{|V|} F_G(\beta; h),$$

where $|V|$ denotes the number of vertices in G , and the limit is taken in an appropriate sense.

The model was introduced by Ising [50], although he only studied the 1-dimensional model, and incorrectly conjectured that the model in general did *not* exhibit a phase transition. It was not until over 20 years later that Onsager [73] announced that the 2-dimensional model indeed exhibited a phase transition.² The universality of the critical exponents appearing in the 2-dimensional model were verified for various lattices (cf. [68], or [6, 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 next breakthrough in the study of critical phenomena came in the 1960s, when Kadanoff applied quantum field-theoretic techniques (the renormalization group, or RG) to describe the Ising phase transition [53]; these ideas were subsequently further developed by Wilson [83, 84]. One of the features of the RG 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 Belavin, Polyakov and Zamolodchikov [7] in 1984 to postulate that *conformal invariance* should manifest at these fixed points. Their work resulted in the characterization of 2D critical phenomena by the celebrated *minimal models* of conformal field theory (CFT). The Ising critical point itself is the so-called (3, 4) minimal model.

Happening concurrently with these developments in statistical mechanics were the first developments in the study of 2-dimensional quantum gravity. Part of the hope of this program was that an exact solution of a 2D theory of gravity could shed light on how higher-dimensional

¹Throughout the present work, by graph we actually mean *multigraph*, i.e., we allow for loops and multiple edges between vertices.

²In 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 Yang [85].

versions might function; another motivating factor was a number of recent advances in string theory, in which the techniques of 2D gravity played an integral role. One of the first major works on this subject was Polyakov’s work on bosonic string theory [76], which demonstrated that the Liouville (also called continuum) approach to 2D gravity could be exactly solved. The drawback of this theory was that explicit calculations proved difficult, and hindered the theory for the next decade or so. Indeed, the aforementioned work of Belavin, Polyakov, and Zamolodchikov [7] was later described by Polyakov as “an unsuccessful attempt to solve the Liouville theory” [75]. An alternative approach to this problem was put forth by the Saclay school of theoretical physics: to replace the functional integral over geometries by a sum over discretized surfaces, who began to develop this approach in the late 70s [20, 51]. This sum could be calculated with the help of matrix integrals [15, 28, 56, 58]. This method was quite successful in the case of a pure theory of gravity, as the model was indeed completely integrable, and was closely linked to the theory of orthogonal polynomials, and of Painlevé equations [43, 44]. The next natural step in this program was to try and compute how matter interacted with a nontrivial gravitational background.

These developments together prompted 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

$$\mathcal{Z}_n(\beta; h) = \sum_{\substack{G: |V|=n, \\ G \text{ planar}}} Z_G(\beta; h), \quad (1.4)$$

where the sum is taken over all 4-regular, planar graphs with n vertices. In [57], Kazakov considered the formal generating function

$$\mathcal{Z}(\tau, t; H) = \sum_{n \in \mathbb{N}} \left(\frac{-t\tau}{4(1-\tau^2)^2} \right)^n \mathcal{Z}_n(\beta; h), \quad (1.5)$$

where we identify

$$\tau := e^{-2\beta}, \quad H := \beta h,$$

and t is a parameter. Kazakov demonstrated that the generating function (1.5) is equivalent to the planar ($n \rightarrow \infty$) limit of the free energy of the 2-matrix model (1.2):

$$F(\tau, t, H) = \mathcal{Z}(\tau, t, H).$$

In [14, 57], a formula was derived for this quantity, which subsequently allowed for the prediction of the shift of the critical exponents of the Ising model when coupled to gravity. Kazakov’s description of the critical point turned out to be in direct agreement with the newly predicted results of Kniznik, Polyakov and Zamolodchikov [59] arising from coupling certain CFTs to matter; this is the so-called KPZ formula. Subsequent analysis for this model and closely related ones was performed in [14, 19] (see also the work [87], which describes a generalization of the Ising/massless case to massive Majorana fermions). The culmination of this work in the physics literature were the papers of Douglas and Shenker [34], Brézin and Kazakov [21], and Gross and Migdal [47, 48]. These works identified the critical points of the 2-matrix model (more generally, the matrix-chain model) with the CFT minimal models coupled to gravity. These critical points were characterized in terms of so-called string equations, which come from additional symmetries

of reductions of the KP hierarchy. This work completed the picture of pure gravity provided by the 1-matrix model (which described the $(2, 2g + 1)$ critical points) to describe all (p, q) minimal models, for (p, q) coprime. For an overview of this story from the physics perspective, see the excellent review [32] and references therein.

In the mathematics literature, the “pure gravity” situation, in which no matter is present, stemmed from the works [34, 47, 61]. This was subsequently described rigorously by Fokas, Its and Kitaev in [43, 44], and more recently (in line with the language of this work) by Duits and Kuijlaars [38] (the quartic model), and Bleher and Deaño [12] (the cubic model). Similarly, the other critical points of the 1-matrix model are now widely accepted to be well-understood (see, for example, [26]). However, the same cannot be said for the 2-matrix model, due in part to the lack of orthogonal polynomial techniques that had been core to the study of the 1-matrix model. This changed with the work of Kuijlaars and McLaughlin [61], in which they found a characterization of this model in terms of multiple orthogonal polynomials. Since there was already a wide mathematical literature on the asymptotics of such polynomials [5, 72, 81], this opened the doors for the possibility of the analysis of critical phenomena in the 2-matrix model, the first physically interesting critical point being the one studied originally by Kazakov [57] corresponding to the Ising model. Some progress in this direction was made by Kuijlaars, Duits, and their collaborators [35, 36, 39, 40], but no direct analytic derivation of the Ising critical point has been written down. In this series of works (see also the forthcoming [37, 49]), we provide a rigorous analysis of the critical point appearing in Kazakov’s work, using steepest descent analysis.

As was mentioned before, the partition function (1.5) is a generating function for the Ising model on 4-regular planar maps. This generating function (along with its analog for 3-regular planar maps) has been studied using purely combinatorial techniques, bypassing the use of matrix integrals. The first rigorous derivation of the results of Kazakov and Boulatov for both the 3 and 4-regular generating functions have been obtained by Bousquet-Mélou and Schaeffer in [16]. Alternative bijective approaches were found by Bouttier, Di Francesco and Guitter in [17, 18], and later by Bernardi and Bousquet-Mélou in [8] using functional equations. More recently still, the groups of Albenque, Ménard and Schaeffer [1, 2] and the group of Chen and Turunen [23, 24, 80] constructed local limits of various families of random maps coupled to the Ising model. These local limits correspond to infinite random maps and correspond to a *thermodynamic limit*. These groups were able to rigorously establish the existence of the Ising phase transition coupled to a random background, and calculate some of the critical exponents, in agreement with the predictions of Boulatov and Kazakov. In the present work, we study the (analytic continuation of) the associated matrix integral originally put forth by Kazakov, and study it using analytical techniques, i.e., Riemann–Hilbert analysis of biorthogonal polynomials. These techniques will allow us in subsequent works [37, 49] to study the multi-scaling limit of the Ising critical point, and prove that the partition function of this model converges under the appropriate scaling to a τ -function for the so-called $(3, 4)$ string equation, thus rigorously proving the results of [19, 27].

1.2 A family of biorthogonal polynomials

The 2-matrix model considered above, with partition function (1.1), can be studied using orthogonal polynomial-type methods. For convenience, we will first introduce the parameter

$$q := e^H. \tag{1.6}$$

Using the Itzykson–Zuber integral over the unitary group (cf., for example, [51, 69, 88]), 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 reads³

$$Z_n(\tau, t, q; N) = \frac{C_{n,N}}{\tau^{\frac{n(n-1)}{2}}} \times \iint \Delta(x)\Delta(y) \exp \left\{ N \sum_{i=1}^n (\tau x_i y_i - V(x_i; qt) - V(y_i; q^{-1}t)) \right\} \prod_{i=1}^n dx_i dy_i,$$

where

$$V(z; t) = \frac{1}{2}z^2 + \frac{t}{4}z^4,$$

$\Delta(x)$, $\Delta(y)$ 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 :

$$C_{n,N} = \frac{(2\pi)^{n(n-1)}}{(\prod_{p=1}^n p!)^2} \left(\prod_{p=1}^{n-1} p! \right) N^{-\frac{n(n-1)}{2}} = \frac{1}{(n!)^2} \frac{(2\pi)^{n(n-1)}}{\prod_{p=1}^{n-1} p!} N^{-\frac{n(n-1)}{2}}.$$

By performing elementary row/column transformations on the Vandermonde determinants $\Delta(x)$, $\Delta(y)$, and applying Andréief's identity, we can rewrite the partition function in terms of the following family of biorthogonal polynomials [69]

$$\int_{\Gamma} \int_{\Gamma} p_k(z) q_j(w) \exp[N(\tau zw - V(z; qt) - V(w; q^{-1}t))] dz dw = h_j \delta_{kj}, \quad (1.7)$$

where $\Gamma := \mathbb{R}$ is taken to be the real line. In this language, the partition function can be expressed in terms of the orthogonality coefficients:

$$Z_n(\tau, t, q; N) = n! \frac{C_{n,N}}{\tau^{\frac{n(n-1)}{2}}} \prod_{j=0}^{n-1} h_j(\tau, t, q; N). \quad (1.8)$$

The study of the large- n asymptotic behavior of the partition function thus reduces to the study of the large n asymptotic behavior of the associated biorthogonal polynomials.

1.3 Analytic continuation

The critical point considered by Kazakov (cf. [19], after formula (1)) corresponds equation (1.3) in the present work. This leads to an immediate issue: the partition function (1.8) does not converge for $t < 0$. We therefore must interpret the partition function as an analytic continuation of the $t > 0$ partition function. For finite n , we make this continuation as follows. Note that since all relevant quantities (the biorthogonal polynomials, the partition function, etc.) can be written in terms of moments of the biorthogonality measure, it is sufficient to consider analytic continuation of the moments. For $k, j \geq 0$, we define

$$m_{jk}(\tau, t, q; N) := \int_{\Gamma_t} \int_{\Gamma_t} z^j w^k \exp \left\{ N g \left[\tau zw - \frac{1}{2}z^2 - \frac{1}{2}w^2 - \frac{tqz^4}{4} - \frac{tq^{-1}w^4}{4} \right] \right\} dz dw,$$

where the contour $\Gamma_t := \{x e^{i\theta} \mid x \in \mathbb{R}, \theta = \arg(t^{-1/4})\}$, and is oriented in the direction of increasing x . This definition coincides with the standard definition of the moments when $t > 0$, and is analytic in t , and so acts as a suitable analytic continuation of the functions $m_{jk}(\tau, t, q; N)$.

³Note that we have changed from H in (1.2) to q , by a slight abuse of notation. We hope that this will not cause confusion further in this article.

Evidently, the moments have a branch point 0, and so there are two possible analytic continuations we can make: one where we analytically continue t through the upper half plane, and one where we continue through the lower half plane. This would in principle require two separate analyses. However, due to the Schwarz symmetry (for real τ , q , N , and $|\arg t| < \pi$),

$$\overline{m_{jk}(\tau, \bar{t}, q; N)} = m_{jk}(\tau, t, q; N),$$

we can obtain one analytic continuation from the other by reflection. Thus, it is sufficient to consider an analytic continuation through the *upper half plane*. For $t < 0$, 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 assume that the parameters of the model all have fixed sign:

$$\tau > 0, \quad t < 0, \quad q > 0, \quad H \in \mathbb{R}.$$

1.4 Riemann–Hilbert problem

As observed in [61], the biorthogonal polynomials defined by (1.7) admit a Riemann–Hilbert formulation, given as follows. Consider the following Riemann–Hilbert problem (RHP) on Γ (recall that the contour Γ is defined as starting from $e^{3\pi i/4} \cdot \infty$, and ending at $e^{-\pi i/4} \cdot \infty$):

$$\mathbf{Y}_+(z) = \mathbf{Y}_-(z) \left[\mathbb{I} + e^{-NV(z;qt)} \begin{pmatrix} 0 & f(z) & \frac{f'(z)}{N\tau} & \frac{f''(z)}{(N\tau)^2} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \right], \quad z \in \Gamma,$$

where

$$f(z) = \int_{\Gamma} \exp[N(\tau zw - V(w; q^{-1}t))] dw.$$

The asymptotics of $\mathbf{Y}(z)$ are chosen to be

$$\mathbf{Y}(z) = \left[\mathbb{I} + \mathcal{O}\left(\frac{1}{z}\right) \right] \begin{pmatrix} z^n & 0 & 0 & 0 \\ 0 & z^{-n/3} & 0 & 0 \\ 0 & 0 & z^{-n/3} & 0 \\ 0 & 0 & 0 & z^{-n/3} \end{pmatrix}, \quad |z| \rightarrow \infty,$$

where n is a multiple of 3. *This restriction on n is taken for simplicity of exposition; the assumption is not essential* (see also [39, 40]). This RHP admits a unique exact solution, with the 1-1 entry of $\mathbf{Y}(z)$ being the degree n monic biorthogonal polynomial⁴ defined by the relations (1.7):

$$[\mathbf{Y}]_{11}(z) = p_n(z).$$

⁴Note that, in general, the relation (1.7) in fact defines *two* sets of polynomials $\{p_n(z)\}$, $\{q_n(w)\}$, which in general do not coincide, except in the special case that the potentials $V_1(z) = V_2(z)$ are the same. However, because the potentials in question in this work are related by $V_1(z; qt) = V_2(z, q^{-1}t)$, the corresponding polynomials are related by $q_n(z; \tau, t, q; N) = p_n(z; \tau, t, q^{-1}; N)$.

The partition function of the matrix model defined by (1.1) (more precisely, its analytic continuation) can be related to an isomonodromic τ -function, given in terms of $\mathbf{Y}(z)$ [9]. The τ -function is defined to be

$$d \log \tau_n := \operatorname{Res}_{z=\infty} \operatorname{tr} [\mathbf{Y}^{-1}(z) \mathbf{Y}'(z) d\Psi_0(z) \Psi_0^{-1}(z)],$$

where Ψ_0 is an explicit matrix function, and the differential is in the variables of deformation t, τ, q . Its definition will be given more precisely in the succeeding text. The differential of the partition function considered by Kazakov is proportional to this isomonodromic- τ function:

$$d \log \frac{Z_n(\tau, t, q; N)}{\tau^{\frac{-n(n-1)}{2}} C_{n,N}} = d \log \left[\left(\frac{\tau}{t^2} \right)^{\frac{n}{2}(\frac{n}{3}-1)} \tau_n \right].$$

Furthermore, one may calculate $Z_n(\tau, 0, 0; N)$ explicitly (cf. [70, Appendix A.49]):

$$Z_n(\tau, 0, 0; N) = \tau^{\frac{-n(n-1)}{2}} C_{n,N} \times N^{-\frac{n(n+1)}{2}} \frac{(2\pi)^n \tau^{n(n-1)/2}}{(1-\tau^2)^{n^2/2}} \prod_{p=1}^{n-1} p!.$$

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 particular, we have that

$$d \log \frac{Z_n(\tau, t, q; N)}{Z_n(\tau, 0, 0; N)} = d \log \left[\frac{(1-\tau^2)^{n^2/2}}{\tau^{n^2/3} t^{n(\frac{n}{3}-1)}} \tau_n \right].$$

The remainder of this work is devoted to the calculation of this τ -function.

1.5 Spectral curve

In recent decades, the Riemann–Hilbert approach has emerged as a powerful tool for computing the asymptotic behavior of polynomials satisfying orthogonality relations. Its breakthrough occurred with the seminal work [30], where the authors successfully applied the Deift–Zhou steepest descent method [31] to the Riemann–Hilbert problem for orthogonal polynomials on the real line with a varying weight $e^{-nV(x)}dx$. An important contribution to this analysis came from potential theory: the limiting zero distribution of the polynomials is given by the equilibrium measure μ_V for the Coulomb gas in the external field V :

$$\mu_V := \arg \min_{\nu \in \mathcal{M}_1(\mathbb{R})} \iint \log \frac{1}{|x-y|} d\nu(x) d\nu(y) + \int V(x) d\nu(x),$$

where $\mathcal{M}_1(\mathbb{R})$ is the set of all unit Borel measures supported on \mathbb{R} . By introducing the equilibrium measure into the Riemann–Hilbert problem (via a transformation involving the g -function constructed from the equilibrium measure) and employing the Deift–Zhou steepest descent method, one can obtain a full asymptotic description of the polynomials and their properties. When V is a polynomial, this equilibrium measure can be characterized in terms of hyperelliptic curve. Specifically, there exists a polynomial Q such that

$$y(x) = V'(x) - \int \frac{1}{t-x} d\mu_V(t) \tag{1.9}$$

satisfies the algebraic equation

$$y^2 - V'(x)y + Q(x) = 0.$$

This defines a curve known as the spectral curve, which plays a crucial role in the analysis. Indeed, if $Q(x)$ is known explicitly, then the spectral curve can be used to perform the steepest descent analysis without knowledge of the equilibrium measure. The defining feature relevant for steepest descent analysis is the so-called *S-property* of the harmonic function $\phi(x) := \operatorname{Re} \int y(x) dx$ [62]:

$$\frac{\partial}{\partial n_+} \phi(x) = \frac{\partial}{\partial n_-} \phi(x), \quad x \in \operatorname{supp} \mu,$$

where $\frac{\partial}{\partial n_{\pm}}$ denote the normal derivatives on either side of $\operatorname{supp} \mu$. It is precisely this property which allows for the opening of lenses, one of the standard transformations in the Deift–Zhou analysis. The polynomial Q is determined in terms of the first few moments of the equilibrium measure. In general, without prior knowledge of this measure, these moments are difficult to compute. Without the aid of potential theory, one would need to prove that such an *S-property* holds by other means.

For special potentials V_1 and V_2 , a Deift–Zhou steepest descent for the Riemann–Hilbert problem characterizing biorthogonal polynomials has been studied in [35, 39, 40]. These works demonstrated that the limiting zero distribution for the polynomials can be characterized by a vector equilibrium problem (see also [36]). This problem involves three measures that simultaneously minimize an energy functional. Since its full description is rather complicated, we omit it here to maintain the flow of presentation. Similar to the case of orthogonal polynomials on the real line described above, the minimizing vector of measures can be characterized by an algebraic curve. This aligns with predictions made in [41, 74] based on loop equations. It is important to note that the vector equilibrium problem requires the potentials to be such that the matrix integrals converge as integrals over the real line. It does not trivially extend to the case $t < 0$, which is of interest to us. Nevertheless, from the loop equation approach one still expects that, in the problem at hand, the limiting zero distribution is given by an algebraic curve of the form

$$\tau q X^4 + \tau q^{-1} Y^4 - t X^3 Y^3 - q X^3 Y - q^{-1} Y^3 X + t \tau^{-1} X^2 Y^2 + a X^2 + b Y^2 + c X Y + d = 0,$$

for some constants a, b, c , and d . Here $Y(X)$, $X(Y)$ are the analogs of (1.9):

$$\begin{aligned} \tau Y &= X + t q X^3 + \int \frac{d\mu(\zeta)}{\zeta - X}, \\ \tau X &= Y + t q^{-1} Y^3 + \int \frac{d\tilde{\mu}(\zeta)}{\zeta - Y}, \end{aligned}$$

and $\mu, \tilde{\mu}$ represent the limiting zero distributions of p_n and q_n , respectively.

However, finding the constants a, b, c and d explicitly is challenging. One important aspect of our analysis is the computation of the spectral curve for certain parameter values with $t < 0$, and we are able to prove that the function $Y(X)$ satisfies an *S-property* of its own (cf. Lemma 3.4, and the proof of Proposition 3.6). We will use this as input in the Deift–Zhou steepest descent analysis.

2 Statement of results

2.1 Asymptotic behavior of the partition function

We will now state our main result: an exact expression for $F(\tau, t, H)$ as given in (1.2). To this end, we first consider the equation

$$0 = \mathfrak{I}(\sigma; \tau, t, q) := -t - \frac{1}{9} \tau^2 \sigma (\sigma^2 - 3) - \frac{1}{3} \frac{\sigma}{(1 + \sigma)^2}$$

$$+ \frac{2}{3} \left(\frac{\sigma}{1 - \sigma^2} \right)^2 \left[\underbrace{\frac{1}{2}(q + q^{-1})}_{\cosh H} - 1 \right]. \quad (2.1)$$

Note that for fixed $q = e^H$ and $0 < \tau < 1$, this is an algebraic curve in the (t, σ) plane. By the implicit function theorem, since

$$\frac{\partial \mathfrak{J}}{\partial \sigma}(\sigma; \tau, t, q)|_{t=0, \sigma=0} = \frac{1}{3}(\tau^2 - 1),$$

for $t = 0$ there is a unique solution $\sigma(\tau, t, H)$ of (2.1) which is analytic in a neighborhood of $t = 0$, and such that $\sigma(\tau, 0, H) = 0$, for any $H \in \mathbb{R}$, $0 < \tau < 1$. The solution can be analytically continued along the negative part of the real- t axis. Either one hits a branch point at some critical value t_{cr} such that $\sigma(\tau, t_{\text{cr}}, H)$ is a double root of (2.1), or $\sigma(\tau, t, H)$ is well defined for all $t < 0$. In the latter case, we set $t_{\text{cr}} = -\infty$ (this case in practice does not happen for the values of τ, H we consider here).

Definition 2.1. For fixed H and $0 < \tau < 1$, we denote with $t_{\text{cr}}(\tau, H)$ the largest negative value such that the solution $\sigma(\tau, t, H)$, characterized by $\sigma(\tau, 0, H) = 0$, is real analytic for $t \in (t_{\text{cr}}, 0]$. Moreover, we define

$$\mathcal{D} = \{(\tau, t, H) \mid 0 < \tau < 1, H \in \mathbb{R}, t_{\text{cr}}(\tau, H) < t < 0\},$$

and refer to \mathcal{D} as the phase space.

Note that \mathcal{D} is somewhat implicitly defined, as we have not indicated how t_{cr} can be found in terms of τ and H . As we will show below, for $H = 0$ it is possible to give t_{cr} as an explicit function of τ . For $H \neq 0$, we will see that \mathcal{D} has a very convenient parametrization. That parametrization is still somewhat complicated and we have not been able to derive an explicit expression for t_{cr} as a function of both τ and H . We will postpone this discussion for now and first present the main result of the paper.

Theorem 2.2. *Let (τ, t, H) belong to the region \mathcal{D} . Then, as $n \rightarrow \infty$,*

$$\begin{aligned} F(\tau, t, H) &:= \lim_{n \rightarrow \infty} \frac{1}{n^2} \log \frac{Z_n(\tau, t, H; n)}{Z_n(\tau, 0, 0; n)} \\ &= \frac{3}{4} + \frac{1}{2} \log \frac{(1 - \tau^2)\sigma(\tau, t, H)}{-3t} - \int_0^{\sigma(\tau, t, H)} \left(\lambda(u) - \frac{1}{2}\lambda(u)^2 \right) \frac{du}{u}, \end{aligned} \quad (2.2)$$

where $\lambda(u)$ is the rational function

$$\lambda(u) = -\frac{1}{t} \left[\frac{1}{9} \tau^2 u(u^2 - 3) + \frac{1}{3} \frac{u}{(u+1)^2} - \frac{2}{3} \left(\frac{u}{u^2 - 1} \right)^2 [\cosh H - 1] \right],$$

and $\sigma(\tau, t, H)$ from Definition 2.1.

This result was first formally computed in [14, 57], based partially on the work [69]. Our results are consistent with this formula. We remark that an alternative form of this result was derived using combinatorial methods, cf. [18], or [42] and references therein.

Remark 2.3. One should compare (2.1) to formula (17) in [14], with the translation of notations

ours	τ	t	q	σ
theirs	c	$g = g(z)$	$B = [q + q^{-1} - 2]$	$-z$

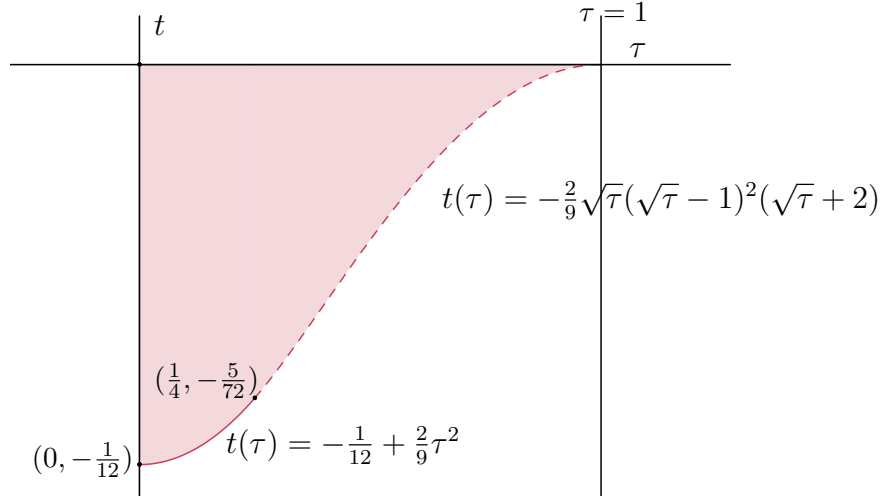


Figure 1. The phase portrait for the 2-matrix model with quartic interactions in the plane $H = 0$. The low-temperature critical curve is represented by the solid red line; the high-temperature critical curve is represented by the dashed red line.

2.2 Structure of the phase space

We will now discuss the structure of the phase space \mathcal{D} in greater detail. We will start by discussing the $H = 0$ first, as it is somewhat simpler, and then discuss the general situation.

2.2.1 $H = 0$

We begin with a description of the phase space when $H = 0$, as this region can be presented in a more transparent manner. All of the proofs of statements in this section follow from the more general results of the generic case ($H \neq 0$). For $H = 0$, $0 < \tau < 1$ we have the following values for $t_{\text{cr}}(\tau, 0)$:

$$t_{\text{cr}}(\tau, 0) = \begin{cases} -\frac{1}{12} + \frac{2}{9}\tau^2, & 0 < \tau \leq \frac{1}{4}, \\ -\frac{2}{9}\sqrt{\tau}(\sqrt{\tau}+1)^2(\sqrt{\tau}+2), & \frac{1}{4} < \tau < 1. \end{cases}$$

See also Figure 1. These curves can be obtained as the most negative real solution to the system of equations $\Im(\sigma; \tau, t, 0) = 0$, $\frac{\partial \Im}{\partial \sigma}(\sigma; \tau, t, 0) = 0$. The intersection of \mathcal{D} and the $H = 0$ plane is the region enclosed by $t = 0$, $\tau = 0$ and a curve connecting $(0, -\frac{1}{12}, 0)$ to $(1, 0, 0)$. This curve consist of two parts: one part connecting $(0, -\frac{1}{12}, 0)$ to the multicritical point $(\frac{1}{4}, -\frac{5}{72}, 0)$, and the other connects $(\frac{1}{4}, -\frac{5}{72}, 0)$ to $(1, 0, 0)$. We will refer to the first curve as the low-temperature critical curve and denote it by $\gamma_{\text{low},0}$. The second, denoted by $\gamma_{\text{high},0}$, will be referred to as the high-temperature critical curve. Note that these are precisely the critical curves that appeared in Kazakov's original work [57]. Note that, when $\tau \rightarrow 0$, the partition function $Z_N(\tau, t, 0; n)$ decouples into two independent partition functions for the 1-matrix model. The point $(0, -\frac{1}{12}, 0)$ is the critical point for the 1-matrix model that was studied in [38, 43, 44]. In this case, the phase transition can be described in terms of special solutions to the Painlevé I equation. At the multicritical point, $\sigma(\tau, t, 0)$ is a triple root and the description of this phase transition is the topic of a forthcoming paper.

We now proceed with general case.

2.2.2 The general case

Due to the aforementioned symmetry in $H < 0$ and $H > 0$, it will be convenient to use $q = e^H$ as in (1.6) and only consider $0 < q \leq 1$. We will then set

$$D = \{(\tau, t, q) \mid t < 0, \tau > 0, 0 < q \leq 1, (t, \tau, \log q) \in \mathcal{D}\}.$$

One of the important ingredients in our analysis is a particular parametrization of the phase space \mathcal{D} . This parametrization has not been used before in the literature, to the best of our knowledge.

We start by defining the set R of all possible values for our parameters

$$R := \{(a, b, c) \mid 1 \leq a \leq b^{-1}, 0 < c \leq b\}. \quad (2.3)$$

Then on R we define

$$\Pi: R \longrightarrow \mathbb{R}^3, \quad (a, b, c) \longmapsto (\tau, t, q),$$

where τ , t and q are defined as

$$\begin{aligned} \tau &= \tau(a, b, c) = \frac{1}{\sqrt{(a^4c^2 + a^2b^2c^2 + a^2 + c^2)(a^4b^2 + a^2b^2c^2 + a^2 + b^2)}}, \\ t &= t(a, b, c) = -\frac{a^2bc(a^4b^2c^2 + 3a^4 + 3a^2b^2 + 3a^2c^2 + 3b^2c^2 - 3)}{9(a^4c^2 + a^2b^2c^2 + a^2 + c^2)(a^4b^2 + a^2b^2c^2 + a^2 + b^2)}, \\ q &= q(a, b, c) = \frac{c(a^4b^2 + a^2b^2c^2 + a^2 + b^2)}{b(a^4c^2 + a^2b^2c^2 + a^2 + c^2)}. \end{aligned}$$

The Jacobian of this map is given by

$$\frac{\partial(\tau, t, q)}{\partial(a, b, c)} = -\frac{4ac(a^2 - 1)(a^2 + 1)(a^2b^2 - 1)(a^2c^2 - 1)(b^2c^2 - 1)(a^2 - b^2)(a^2 - c^2)}{3b(a^2b^2c^2 + a^4b^2 + a^2 + b^2)^{3/2}(a^2b^2c^2 + a^4c^2 + a^2 + c^2)^{7/2}}, \quad (2.4)$$

and this does not vanish in the interior of R . It does vanish on the surfaces defined by $a = 1$ and $a = b^{-1}$, and this is the reason that they are of particular importance to us. We define these components below.

Definition 2.4.

- (1) The low-temperature critical surface S_{low} , is the image under Π of $a = b^{-1}$, with $0 < b < 1$, $0 < c < b$.
- (2) The high-temperature critical surface S_{high} , is given by setting $a = 1$, with $0 < b < 1$, $0 < c < b$.
- (3) The $H \neq 0$ τ -critical curve γ_b , is given by setting $a = b = 1$, with $0 < c < 1$.

Moreover, the intersection of these critical surface with $H = 0$ can be obtained by further setting $b = c$ and we thus define:

- (4) The low-temperature $H = 0$ critical curve $\gamma_{\text{low},0}$ is given by setting $a = b^{-1}$, $c = b$.
- (5) The high-temperature $H = 0$ critical curve $\gamma_{\text{high},0}$ is given by setting $a = 1$, $c = b$.
- (6) The multicritical point is given by setting $a = b = c = 1$.

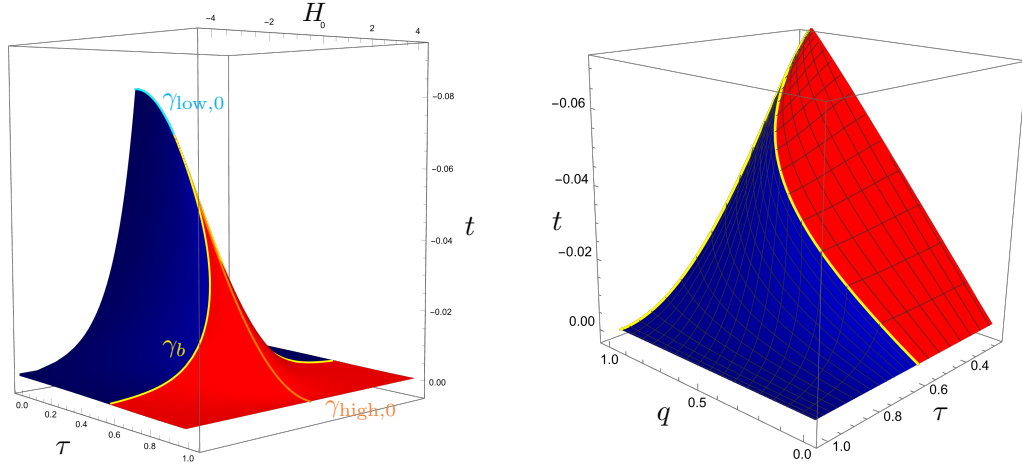


Figure 2. The left picture illustrates the phase portrait for the 2-matrix model with quartic interactions in the (τ, t, H) plane. The region where the spectral curve is of genus zero analyzed in this work lies in the half space where $\tau > 0$, $t < 0$, and is bounded by the red and blue surfaces. The genus zero partition function $F(\tau, t, H)$ is continuous on this surface, and in the genus zero region; on the surface, F is analytic away from the low-temperature critical curve, shown here in light blue. The multicritical point $(\frac{1}{4}, -\frac{5}{72}, 0)$ is the meeting point of the high-temperature critical curve (orange), the two phase boundaries joining the low and high temperature phases (yellow), and the low-temperature critical curve. The picture on the right represents the phase diagram in the coordinates τ, t, q and only the part $0 < q \leq 1$.

The critical curve γ_b is the interface between the low-temperature and high-temperature critical surfaces. A graphical representation of these critical surfaces is given in Figure 2. From the parametrization of the critical surfaces, it is not hard to compute the normal to these surfaces. It is interesting to note that the normal is continuous when passing from the low-temperature to the high-temperature critical surface.

Before we comment further on these critical surface, we first mention the necessary technical result that our map Π indeed parametrizes the phase space D and that the critical surfaces indeed are the sets defined by $t = t_{\text{cr}}$.

Proposition 2.5. *The set $\Pi(R)$ is the region enclosed by $\tau = 0$, $\tau = 1$, $q = 0$, $q = 1$, $t = 0$, and the critical surfaces $S_{\text{low}} \cup S_{\text{high}} \cup \gamma_b$. Moreover, $D = \Pi(R)$ and $\sigma(t, \tau, H) = a^2bc$.*

The proof of this proposition is given in Appendix D.

The critical surfaces S_{low} , S_{high} are implicit in the subsequent works of Boulatov and Kazakov [14], although they are never explicitly derived. Our parametrizations based on (3.4)–(3.6) of these surfaces is new and turn out to be very useful in our analysis of the Riemann–Hilbert problem.

Remark 2.6. When (a, b, c) tend to a point on the critical surface, the Jacobian $\frac{\partial(\tau, t, q)}{\partial(a, b, c)} \rightarrow 0$. In terms of the equation $0 = \mathfrak{I}(\sigma; \tau, t, q)$, this may be interpreted as the root $\sigma = \sigma(\tau, t, q)$ becoming a double root, i.e.,

$$\mathfrak{I}(\sigma; \tau, t, q) = \frac{\partial \mathfrak{I}}{\partial \sigma}(\sigma; \tau, t, q) = 0,$$

for (a, b, c) on the critical surface and away from the multicritical point. At the multicritical point $(\tau, t, q) = (\frac{1}{4}, -\frac{5}{72}, 1)$, the root σ becomes a *triple root* of the equation $0 = \mathfrak{I}(\sigma; \tau, t, q)$. One can thus characterize the critical surface as a subset of the zero locus of the discriminant of equation (2.1). In Appendix C, we write an explicit formula for this discriminant, and thus furnish an implicit formula for the critical surface.

2.3 Spectral curve

An important ingredient in our analysis is the spectral curve for the limiting zero distribution of the polynomials, which we will now introduce. First observe that, for $(a, b, c) \in R$, the function σ is bounded:

$$0 \leq \sigma \leq 1,$$

with $\sigma = 1$ only if $a = b^{-1}$, $c = b$ (this is precisely the low-temperature $H = 0$ critical curve). Then we define the rational functions

$$X(v) = \sqrt{-\frac{\tau\sigma}{3t}} \left(v + \frac{\sigma q^{-1} - 1}{\tau(\sigma^2 - 1)} v^{-1} - \frac{\sigma}{3q} v^{-3} \right), \quad (2.5)$$

$$Y(v) = \sqrt{-\frac{\tau\sigma}{3t}} \left(v^{-1} + \frac{\sigma q - 1}{\tau(\sigma^2 - 1)} v - \frac{\sigma q}{3} v^3 \right). \quad (2.6)$$

One can eliminate the parameter v from the pair of functions $(X(v), Y(v))$ to obtain an implicit formula for the Riemann surface these functions parametrize

$$\mathfrak{S}(X(v), Y(v)) = 0.$$

Here $\mathfrak{S}(X, Y)$ is a degree 6 polynomial in X and Y , with rational coefficients in the variables τ , t , q , and σ (recall that σ is defined as a special solution to the algebraic equation (2.1)). Explicitly, this polynomial is

$$\begin{aligned} \mathfrak{S}(X, Y) = & \tau q X^4 + \tau q^{-1} Y^4 - t X^3 Y^3 - q X^3 Y - q^{-1} Y^3 X + t \tau^{-1} X^2 Y^2 \\ & + \mathfrak{s}_2(\sigma; \tau, t, q) X^2 + \mathfrak{s}_2(\sigma; \tau, t, q^{-1}) Y^2 + \mathfrak{s}_1(\sigma; \tau, t, q) X Y + \mathfrak{s}_0(\sigma; \tau, t, q). \end{aligned}$$

where $\mathfrak{s}_i(\sigma; \tau, t, q)$ are given by

$$\begin{aligned} \mathfrak{s}_2(\sigma; \tau, t, q) = & \frac{1 - 6\sigma - 3\sigma^2}{27\tau t(\sigma + 1)^3} - \frac{1}{27\tau t} (\sigma^3 \tau^2 + 3\sigma^2 \tau^2 - 9\sigma \tau^2 - 27\tau^2 + 1) \\ & - \frac{(q - 1)\sigma}{27\tau t(\sigma^2 - 1)^3} [(9\tau^2 - 9) + 9(q + 1)\sigma + (4q^{-1} - (28\tau^2 + 6))\sigma^2 \\ & \quad - 6(q + 1)\sigma^3 + (30\tau^2 + 3)\sigma^4 + (q + 1)\sigma^5 - 12\sigma^6 \tau^2 + \sigma^8 \tau^2], \\ \mathfrak{s}_1(\sigma; \tau, t, q) = & -\frac{8(5\sigma + 3)}{81(\sigma + 1)^2 t} - \frac{\sigma^6 \tau^2 - 15\sigma^4 \tau^2 + 27\sigma^2 \tau^2 - 3\sigma^2 + 243\tau^2 + 24\sigma + 171}{243t} \\ & - \frac{4\sigma^3(\sigma^2 + 3)}{81t(\sigma^2 - 1)^2} [q + q^{-1} - 2], \\ \mathfrak{s}_0(\sigma; \tau, t, q) = & -\frac{\sigma}{19683t^2 \tau q^2 (\sigma^2 - 1)^4} \\ & \times \left[15q\sigma^6 \tau^2 + 9q\sigma^5 t + (12 - 111\tau^2)q\sigma^4 + 15\left(q^2 - \frac{6}{5}tq + 1\right)\sigma^3 \right. \\ & \quad \left. + (177\tau^2 - 15)q\sigma^2 - 54\left(q^2 - \frac{1}{6}tq + 1\right)\sigma + 81(1 - \tau^2)q \right]^2. \end{aligned}$$

These constants are given for completeness and we will not use their exact expression in this paper. Indeed, it is the uniformization that we will be working with. Note also, that since the curve has uniformization in terms of rational functions, it has genus zero.

The spectral curve provides us with a 4-sheeted Riemann surface. In the interior of the phase space, it is easy to see from the uniformization that there are four branch points (given by $X(v)$

with v a solution to $X'(v) = 0$ denoted by $X = \pm\alpha, \pm\beta$ with $0 < \alpha < \beta$. The sheet structure is represented in Figure 5. The first sheet has a branch cut at $[-\alpha, \alpha]$ at which it is glued to the second sheet in a crosswise manner. The second sheet has additional cuts at $(-\infty, -\beta]$ and $[\beta, \infty)$ at which it is glued in a crosswise manner to the third and fourth sheet respectively. On each sheet $j = 1, 2, 3, 4$, we define the uniformization coordinate $v_j(X)$ as the map such that $X(v_j(X))$ is the identity map on the j^{th} sheet (these maps are uniquely determined by their large X behavior; the expansions of rescaled versions of the functions $v_j(X)$ for $X \rightarrow \infty$ are given in Appendix A. See also Remark 3.2 for the exact relation between the coordinates u and v). Then, we can define the function $Y(X)$ on our Riemann surface by setting its value on the j^{th} sheet, $Y_j(X)$, to be $Y_j(X) = Y(v_j(X))$.

Although we do not prove this explicitly, it can be shown from our analysis that zeros of the polynomials $p_n(x)$ accumulate on the interval $(-\alpha, \alpha)$ and have limiting distribution, which we denote by μ . We can then recover the measure μ from the boundary behavior of Y_1 on the cut,

$$d\mu(s) = \frac{\tau}{2\pi i} [Y_{1,+}(s) - Y_{1,-}(s)] \mathbf{1}_{[-\alpha, \alpha]}(s) ds. \quad (2.7)$$

There is also a second measure that plays a role for us. Its defined by taking boundary values on the other cuts,

$$d\nu(s) = -\frac{\tau}{2\pi i} [Y_{3,+}(s) - Y_{3,-}(s)] \mathbf{1}_{(-\infty, -\beta]}(s) ds - \frac{\tau}{2\pi i} [Y_{4,+}(s) - Y_{4,-}(s)] \mathbf{1}_{[\beta, \infty)}(s) ds. \quad (2.8)$$

The interpretation of ν for the asymptotic behavior of the polynomials is not as obvious as it is for μ and it serves mostly as an auxiliary measure in our analysis. Note that from the representations (2.7) and (2.8) it is far from obvious that these measures are positive and we will put significant effort in proving this. The positivity is important for our lensing inequalities in the Riemann–Hilbert analysis.

Both measures μ and ν are absolutely continuous and have an analytic density. Away from the critical surfaces, these densities vanish as a square root near the endpoints. On the critical surfaces, one or both of these measures will have different behaviors at their endpoints and this will break down the limiting behavior of the partition function.

Remark 2.7. When changing roles of X and Y , the Y coordinate has branch points at $Y = \pm\tilde{\alpha}, \pm\tilde{\beta}$ with $0 < \tilde{\alpha} \leq \tilde{\beta}$. The zeros of the polynomials $q_n(x)$ accumulate on an interval $(-\tilde{\alpha}, \tilde{\alpha})$ and have limiting distribution, which we denote by $\tilde{\mu}$. Moreover, (X, Y) given by

$$\tau X = Y + tq^{-1}Y^3 + \int \frac{d\tilde{\mu}(\zeta)}{\zeta - Y}$$

are on the spectral curve and defines X as a function on Y on the first sheet with an analytic continuation to the other sheets of the surface.

2.4 Phase transitions and special points on the spectral curve

We will comment now on the phase transitions on the critical surfaces of the parameter space.

When the parameters approach the critical surface, the measures will observe different behavior of μ and ν at their endpoints:⁵

1. *The generic (non-critical) case.* The measures μ, ν vanish as square roots at their finite endpoints:

$$\frac{d\mu}{ds} = \rho |X \mp \alpha|^{1/2} [1 + \mathcal{O}(|X \mp \alpha|)], \quad X \rightarrow \pm\alpha,$$

⁵Throughout, $\rho, \tilde{\rho} > 0$ represent irrelevant constant factors. All limits are taken from inside the support of the measure, i.e., $X \nearrow \alpha$, $X \searrow -\alpha$, $X \nearrow -\beta$, and $X \searrow \beta$, respectively.

$$\frac{d\nu}{ds} = \tilde{\rho}|X \mp \beta|^{1/2}[1 + \mathcal{O}(|X \mp \beta|)], \quad X \rightarrow \pm\beta.$$

2. *The low-temperature critical surface S_{low} .* An extra zero at the endpoint of ν :

$$\begin{aligned} \frac{d\mu}{ds} &= \rho|X \mp \alpha|^{1/2}[1 + \mathcal{O}(|X \mp \alpha|)], & X \rightarrow \pm\alpha, \\ \frac{d\nu}{ds} &= \tilde{\rho}|X \mp \beta|^{3/2}[1 + \mathcal{O}(|X \mp \beta|)], & X \rightarrow \pm\beta. \end{aligned}$$

3. *The high-temperature critical surface and curve $S_{\text{high}} \cup \gamma_{\text{high},0}$.* An extra zero at the endpoint of μ :

$$\begin{aligned} \frac{d\mu}{ds} &= \rho|X \mp \alpha|^{3/2}[1 + \mathcal{O}(|X \mp \alpha|)], & X \rightarrow \pm\alpha, \\ \frac{d\nu}{ds} &= \tilde{\rho}|X \mp \beta|^{1/2}[1 + \mathcal{O}(|X \mp \beta|)], & X \rightarrow \pm\beta. \end{aligned}$$

4. *The low-temperature $H = 0$ critical curve.* Extra zeros at the endpoints of both measures:

$$\begin{aligned} \frac{d\mu}{ds} &= \rho|X \mp \alpha|^{3/2}[1 + \mathcal{O}(|X \mp \alpha|)], & X \rightarrow \pm\alpha, \\ \frac{d\nu}{ds} &= \tilde{\rho}|X \mp \beta|^{3/2}[1 + \mathcal{O}(|X \mp \beta|)], & X \rightarrow \pm\beta. \end{aligned}$$

5. *The $H \neq 0$ τ -critical curve.* The supports of μ and ν touch, and vanish locally as a cube root:

$$\begin{aligned} \frac{d\mu}{ds} &= \rho|X \mp \alpha|^{1/3}[1 + \mathcal{O}(|X \mp \alpha|^{1/3})], & X \rightarrow \pm\alpha, \\ \frac{d\nu}{ds} &= \tilde{\rho}|X \mp \alpha|^{1/3}[1 + \mathcal{O}(|X \mp \beta|^{1/3})], & X \rightarrow \pm\alpha. \end{aligned}$$

6. *The multicritical point.* The supports of μ and ν touch, and an extra zero at touching point:

$$\begin{aligned} \frac{d\mu}{ds} &= \rho|X \mp \alpha|^{4/3}[1 + \mathcal{O}(|X \mp \alpha|^{1/3})], & X \rightarrow \pm\alpha, \\ \frac{d\nu}{ds} &= \tilde{\rho}|X \mp \alpha|^{4/3}[1 + \mathcal{O}(|X \mp \beta|^{1/3})], & X \rightarrow \pm\alpha. \end{aligned}$$

The behaviors of the above measures are depicted in Figure 3. All these critical behaviors break our analysis in a crucial way and will lead to different scaling limits for the partition function.

The situation on the S_{high} and S_{low} is similar to the 1-matrix model critical point. Indeed also in that case, there is a merging of critical point to a branch point. This means that in a double scaling limit, the partition function is described by a special solution to the Painlevé I equation. For the Riemann–Hilbert analysis, it means that one should be able to treat the local parametrices in terms of Painlevé I parametrices, as has been performed in previous works [12, 38] in the 1-matrix model.

The situation on the critical curve γ_b is intriguing. The situation is different from the rest for the critical surface. Nevertheless, we still expect that the phase transition is described by the same Painlevé I equation as on the critical surfaces. We believe that a 3×3 Painlevé I parametrix (similar to the one arising from the Lax pair in [52]) should be used. We hope to address this in a future work.

Finally, the multi-critical point is the most important point of the critical surface and we shall treat this in the sequel to this work [37, 49].

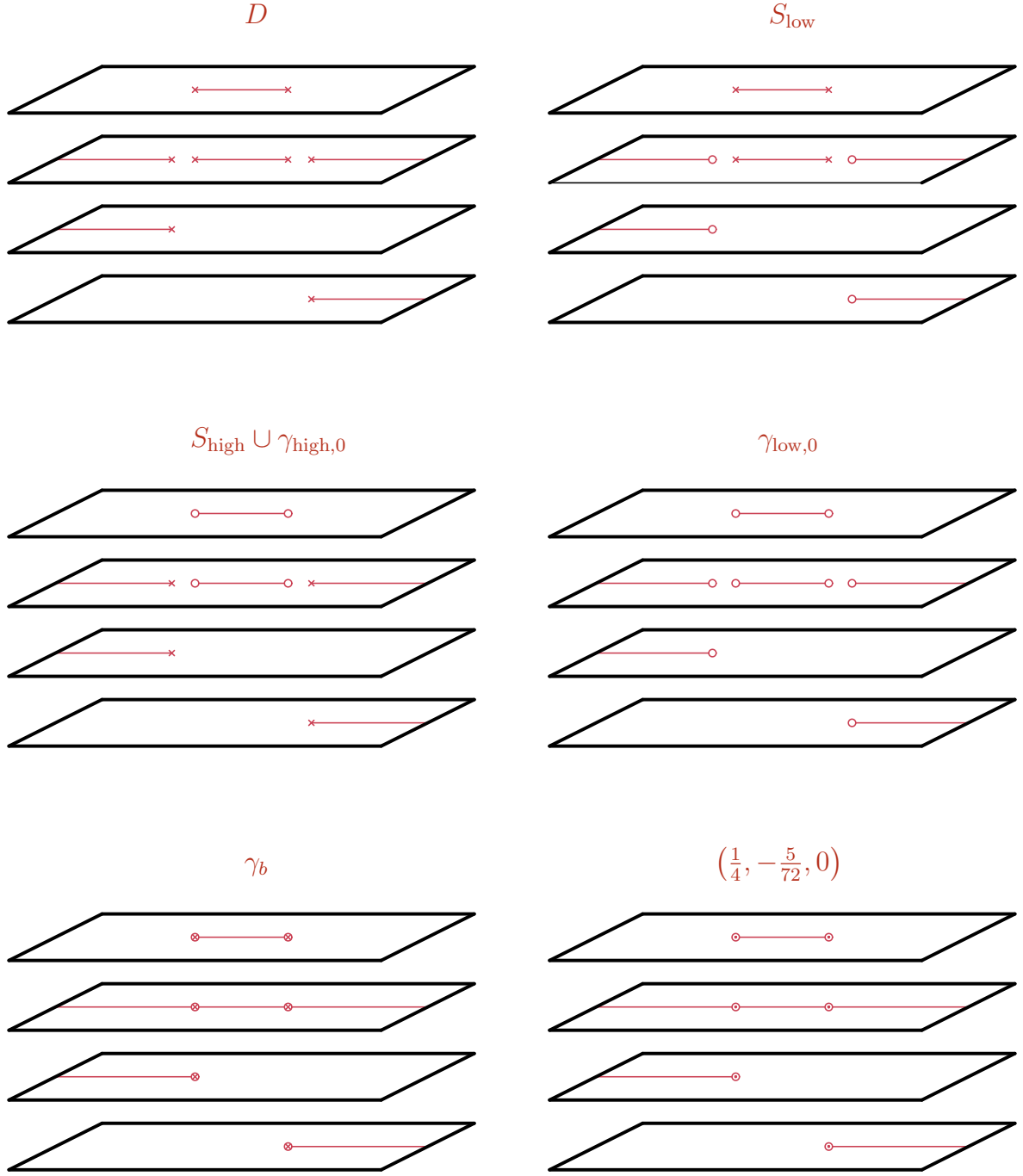


Figure 3. Possible endpoint behaviors of the measures μ and ν in various regions of the phase space. Each region is characterized by the behavior of μ, ν at their endpoints. The symbols $\times, \circ, \otimes, \odot$ correspond to the following vanishing exponents of μ, ν near their endpoints: $\times \leftrightarrow \frac{1}{2}$, $\circ \leftrightarrow \frac{3}{2}$, $\otimes \leftrightarrow \frac{1}{3}$, $\odot \leftrightarrow \frac{4}{3}$.

2.5 Strategy of the proof and an associated Riemann–Hilbert problem

The strategy for the proof of Theorem 2.2 is a Deift–Zhou steepest descent analysis [31] of \mathbf{Y} , i.e., a sequence of explicit invertible transformations

$$\mathbf{Y} \mapsto \mathbf{X} \mapsto \mathbf{U} \mapsto \mathbf{T} \mapsto \mathbf{S} \mapsto \mathbf{R},$$

with the final Riemann–Hilbert problem for \mathbf{R} having jumps that tend to the identity matrix as $n \rightarrow \infty$, uniformly and in L^2 , with normalized behavior at infinity. Standard argument then allow us to obtain an asymptotic expansion for \mathbf{R} , and thus, utilizing the invertibility of the transformations, an asymptotic expansion of \mathbf{Y} . This can be used to write an expression for the partition function (1.4).

In Section 3, we analyze the spectral curve of this model. Section 4 is devoted to the first and second transformations $\mathbf{Y} \mapsto \mathbf{X} \mapsto \mathbf{U}$. The transformations $\mathbf{U} \mapsto \mathbf{T} \mapsto \mathbf{S}$ which correspond to the opening of lenses are completed in Section 5. In Section 6, we construct the parametrices, and perform the final transformation $\mathbf{S} \mapsto \mathbf{R}$. We conclude with Section 7, in which we furnish a proof of Theorem 2.2.

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 $(i, j)^{\text{th}}$ entry and zeros elsewhere by E_{ij} ,
- The Pauli matrices

$$\sigma_1 := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 := \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

In particular, for the third Pauli matrix σ_3 , we will often write expressions such as $z^{C\sigma_3}$, or $e^{f(z)\sigma_3}$. These expressions are defined to be

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

- 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 $J_X(z): \gamma \rightarrow \mathbb{C}$ as its jump matrix: i.e.,

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

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

3 Definition and analysis of the spectral curve

In this section, we define the spectral curve and an associated set of functions, which we will later use in the second transformation. These functions are typically referred to as “ g -functions” (or collectively as *the* g -function) in the Riemann–Hilbert community, see [30]. This function is used to transform the original Riemann–Hilbert problem, which contains exponentially growing terms, into a Riemann–Hilbert problem with oscillatory terms. It is required that this special function satisfy certain inequalities, which we refer to as *lensing inequalities*, so that we can eventually transform the resulting Riemann–Hilbert problem with oscillations into one which is exponentially close to a RHP with constant jumps; this transform is often referred to as the *opening of lenses*. We also prove a number of inequalities necessary for the later “lensing” transformations. In what follows, the role of the g -function will be played by a collection of functions $\Omega_j(z)$.

We begin by constructing the spectral curve, and give a basic analysis of the spectral curve for all values of the parameters (τ, t, q) in the region D bounded by $\tau = 0$, $t = 0$, and $q = 1$ planes, the *infinite temperature plane* $\tau = 1$, and a “critical surface”, which comprises the remaining boundary components of the region D .

3.1 Definition of the spectral curve

The analysis of the spectral curve 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

$$\begin{aligned} Z &= \iint \Delta(x)\Delta(y) \exp \left\{ -n \sum_{i=1}^n \left(\frac{1}{2}x_i^2 + \frac{1}{2}y_i^2 + \frac{tq}{4}x_i^4 + \frac{tq^{-1}}{4}y_i^4 - \tau x_i y_i \right) \right\} \prod_{i=1}^n dx_i dy_i \\ &= \iint \exp - \left\{ \sum_{i < j} \log \frac{1}{|x_i - x_j|} + \sum_{i < j} \log \frac{1}{|y_i - y_j|} \right. \\ &\quad \left. + n \sum_{i=1}^n \left(\frac{1}{2}x_i^2 + \frac{1}{2}y_i^2 + \frac{tq}{4}x_i^4 + \frac{tq^{-1}}{4}y_i^4 - \tau x_i y_i \right) \right\} \prod_{i=1}^n dx_i dy_i. \end{aligned}$$

In the large n limit, under the appropriate scaling, we expect the x_i , y_i to accumulate to a limiting distributions that together minimize the term in the exponent; these distributions, μ and $\tilde{\mu}$ should be subject to the stationarity conditions that are given by

$$X + tqX^3 + \int \frac{d\mu(\zeta)}{\zeta - X} - \tau Y = 0,$$

$$Y + tq^{-1}Y^3 + \int \frac{d\tilde{\mu}(\zeta)}{\zeta - Y} - \tau X = 0.$$

Expanding at infinity, we find that

$$X + tqX^3 - \frac{1}{X} - \tau Y = \mathcal{O}(X^{-2}), \quad (3.1)$$

$$Y + tq^{-1}Y^3 - \frac{1}{Y} - \tau X = \mathcal{O}(Y^{-2}). \quad (3.2)$$

We will construct solutions to these equations from an algebraic curve $\mathfrak{S}(X, Y) = 0$, by making the ansatz that it has genus 0, i.e., that it is rationally parameterized.

Proposition 3.1. *A 3-parameter family of solutions to the stationarity equations (3.1) and (3.2) is given by the rational functions*

$$X(u; a, b, c) = A \int^u \frac{(u^2 - a^2)(u^2 - b^2)}{u^4} du = A \left(u + \frac{a^2 + b^2}{u} - \frac{a^2 b^2}{3u^3} \right), \quad (3.3)$$

$$Y(u; a, b, c) = X(u^{-1}; a, c, b) = B \left(\frac{1}{u} + (a^2 + c^2)u - \frac{a^2 c^2}{3}u^3 \right),$$

with τ, t, q, A , and B defined parametrically in terms of a, b , and c as

$$\tau = \tau(a, b, c) = \frac{1}{\sqrt{(a^4 c^2 + a^2 b^2 c^2 + a^2 + c^2)(a^4 b^2 + a^2 b^2 c^2 + a^2 + b^2)}}, \quad (3.4)$$

$$t = t(a, b, c) = -\frac{a^2 b c (a^4 b^2 c^2 + 3a^4 + 3a^2 b^2 + 3a^2 c^2 + 3b^2 c^2 - 3)}{9(a^4 c^2 + a^2 b^2 c^2 + a^2 + c^2)(a^4 b^2 + a^2 b^2 c^2 + a^2 + b^2)}, \quad (3.5)$$

$$q = q(a, b, c) = \frac{c(a^4 b^2 + a^2 b^2 c^2 + a^2 + b^2)}{b(a^4 c^2 + a^2 b^2 c^2 + a^2 + c^2)}, \quad (3.6)$$

$$A = A(a, b, c) = \sqrt{\frac{3(a^4 c^2 + a^2 b^2 c^2 + a^2 + c^2)}{a^4 b^2 c^2 + 3a^4 + 3a^2 b^2 + 3a^2 c^2 + 3b^2 c^2 - 3}}, \quad (3.7)$$

$$B = B(a, b, c) = A(a, c, b).$$

Remark 3.2. Before we give the proof, we first make a slight abuse of notation. Note that we have defined $X(v)$ and $Y(v)$ already in (2.5) and (2.6). It would be more prudent to use notations $\hat{X}(u; a, b, c)$ and $\hat{Y}(u; a, b, c)$. However, this is simply a scaling of the variables: $\hat{X}(u) = X(v\sqrt{A/B})$ and $\hat{Y}(u) = Y(v\sqrt{A/B})$. Thus the u and v variable give different parametrizations of the same objects. To avoid cumbersome notation, we will simply write $X(u; a, b, c)$ and $Y(u; a, b, c)$. A convenient property of the abc -parametrization is that the branch points of $X(u)$ and $Y(u)$ are at $u = a, b$, and $u = b, c$, respectively.

Proof. We can generically set

$$\begin{aligned} X(u) &= A[u + \alpha_0 + \alpha_1 u^{-1} + \alpha_2 u^{-2} + \alpha_3 u^{-3}], \\ Y(u) &= B[u^{-1} + \beta_0 + \beta_1 u + \beta_2 u^2 + \beta_3 u^3]. \end{aligned}$$

This fixes the positions of the infinities of the X - and Y -coordinates in the uniformizing plane as $u = \infty$, $u = 0$, respectively. One then sees that 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. For example, as $u \rightarrow \infty$, $tX(u)^3 = \mathcal{O}(u^3)$, 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 $A, B, \{\alpha_k\}$, and $\{\beta_k\}$ must then be chosen so that the asymptotics

$$\begin{aligned} X(u) + tqX(u)^3 - \frac{1}{X(u)} - \tau Y(u) &= \mathcal{O}(u^{-2}), & u \rightarrow \infty, \\ Y(u) + tq^{-1}Y(u)^3 - \frac{1}{Y(u)} - \tau X(u) &= \mathcal{O}(u^2), & u \rightarrow 0, \end{aligned}$$

hold. Since the stationarity equations are odd functions of X, Y , we can assume (without loss of generality) that

$$X(u) = -X(-u), \quad Y(u) = -Y(-u),$$

this implies that the branch points of X (resp. Y) will be symmetric about $u = 0$. This fixes the form of $X(u), Y(u)$ as

$$X(u) = A \left(u + \frac{a^2 + b^2}{u} - \frac{a^2 b^2}{3u^3} \right), \quad Y(u) = B \left(\frac{1}{u} + (c^2 + d^2)u - \frac{c^2 d^2}{3}u^3 \right),$$

for some parameters a, b, c, d (note that the branch points here are at $u = \pm a, b, u = \pm c, d$, respectively). We also have a degree of freedom available arising from reparametrization: $u \rightarrow \lambda u$. We use this freedom to fix $d = a$. Inserting these expressions for $X(u), Y(u)$ into the stationarity equation (3.1), we obtain that

$$\begin{aligned} & \left(tA^3 + \frac{1}{3}\tau B a^2 c^2 \right) u^3 + (A + 3qtA^3(a^2 + b^2) - \tau B(a^2 + c^2))u \\ & + \left(A(a^2 + b^2) + tA^3(-a^2 b^2 + (a^2 + b^2)^2 + (a^2 + b^2)(2a^2 + 2b^2)) - \frac{1}{A} - \tau B \right) u^{-1} \\ & = \mathcal{O}(u^{-2}). \end{aligned}$$

A similar equation at $u = 0$ holds upon interchanging $u \rightarrow u^{-1}, b \leftrightarrow c, A \leftrightarrow B, q \leftrightarrow q^{-1}$. The requirement that the above vanishes to order $\mathcal{O}(u^{-2})$ (and the equivalent condition at $u = 0$) yield a total of 6 equations on the coefficients; one of these equations is redundant, and so we can solve the remaining equations for A, B, t, τ, q in terms of the parameters a, b, c . The unique solution (up to a sign, and possibly a relabelling $b \leftrightarrow c$) is given by equations (3.4)–(3.7). This completes the proof. \blacksquare

We have now obtained a 3-parameter family of solutions to the stationarity equations up to terms of order $\mathcal{O}(X^{-2})$. Each fixed triple (a, b, c) parametrizes a Riemann surface. However, if one considers all possible values (a, b, c) , not all of them are relevant or suited for steepest descent analysis. Indeed, it will also be important that the measures μ and ν as given in (2.7) and (2.8) are positive when we are to open lenses. These constraints will be satisfied for all the parameters in our phase space R as defined in (2.3).

3.2 Sheet structure of the spectral curve

The associated Riemann surface given by the parametrization $(X(u), Y(u))$ 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 ∞ . This discussion is repeated from the introduction, for sake of clarity, and because of our slight abuse of notation on the relation between the coordinates u, v .

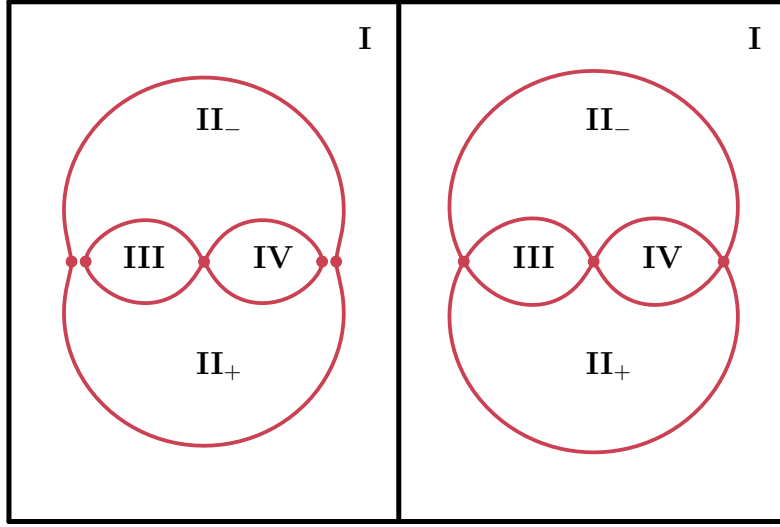


Figure 4. The branch cuts in the uniformizing plane $u = x + iy$ away from the multicritical point (left) and at the multicritical point (right). The noncritical case has $a = 1.0184$, $b = c = 0.9100$. The images of each sheet under the uniformizing map are labelled I, ..., IV. Note that, at criticality, the second sheet is split into two connected components; this is consistent with the picture in the physical plane, depicted in Figures 5 and 6.

Away from the multicritical point and curve γ_b , the spectral curve is 4-sheeted; this family of spectral curves have generically the same structure, and are shown in Figure 5. 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 $a = b = c = 1$, and on the curve γ_b , the curve further degenerates, and the branch points $\pm\beta \rightarrow \pm\alpha$. This family of spectral curves is shown in Figure 6. 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 $[\alpha, \infty)$.

In the uniformization plane, the spectral curve is shown at and away from the multicritical point $(1, 1, 1)$ in Figure 4.

3.3 The functions Ω

Of particular importance is the following (multi-valued) function:

$$\tau\Omega(X) = \tau \int Y dX.$$

This function will play the role of what is often referred to as the g -function in our Deift/Zhou steepest descent analysis of the Riemann–Hilbert problem. The function $\tau\Omega(X)$ is in general a multivalued function in \mathbb{C} ; its real part, however, can be defined in a single-valued manner on the spectral curve. The function $\Omega(u) := \int Y(u)X'(u)du$ in the uniformization coordinate is explicit:

$$\begin{aligned} \tau\Omega(u) &= \tau \int Y dX = \tau \int Y(u)X'(u)du \\ &= \frac{1}{a^2b^4c^2 + 3a^2b^2 + 3a^2c^2 + 3b^4 + 3b^2c^2 - 3} \left(-\frac{1}{4}b^2c^2u^4 - \frac{3}{4}a^2b^2u^{-4} \right) \end{aligned} \quad (3.8)$$

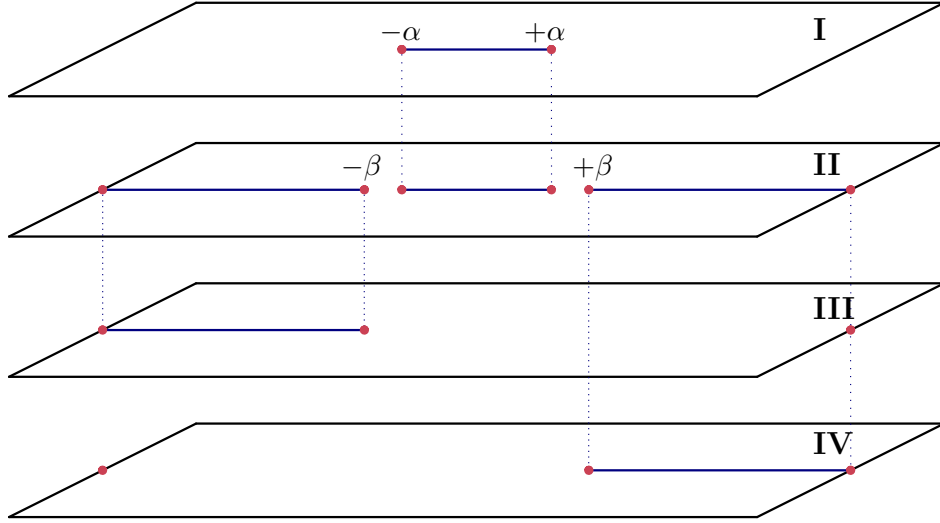


Figure 5. A representative example of a the spectral curve in the physical plane, when the associated parameters $(\tau, t, H) \notin \gamma_b \cup \{(\frac{1}{4}, -\frac{5}{72}, 0)\}$; the sheets are labelled I, \dots , IV.

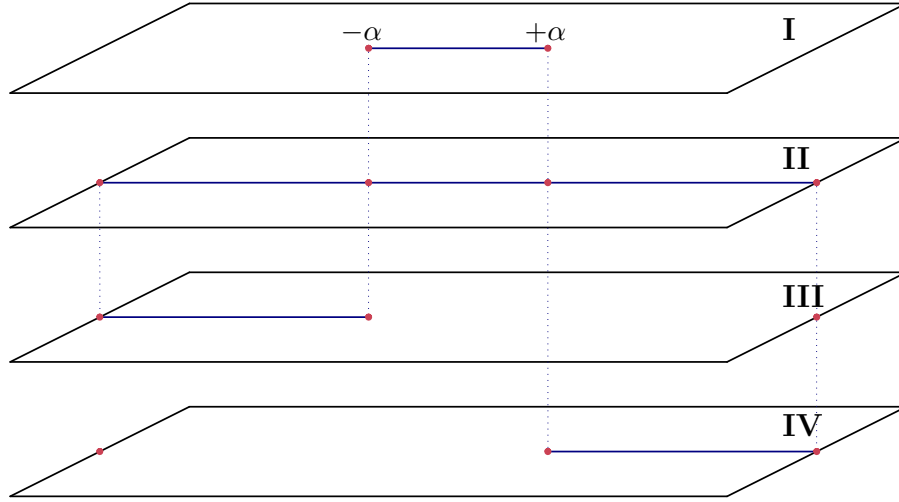


Figure 6. The spectral curve, when the associated parameters $(\tau, t, H) \in \gamma_b \cup \{(\frac{1}{4}, -\frac{5}{72}, 0)\}$. Note that the branch points at $z = \alpha$ and $z = \beta$ have merged here.

$$+\frac{1}{2}(a^2b^2c^2 + b^4c^2 + 3b^2 + 3c^2)u^2 - \frac{3}{2}(a^2b^4 + a^2b^2c^2 - a^2 - b^2)u^{-2} \Big) - \log(u).$$

From here on, we shall relabel $X(u)$ as

$$X(u) =: z(u),$$

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 expres-

sion for the expansion of $\Omega_j(z) := \Omega(u_j(z))$ at infinity that holds for all $(a, b, c) \in R$. We have the following proposition.

Proposition 3.3 (Expansion of $\Omega_j(z)$ at $z = \infty$). *For any fixed $(a, b, c) \in R$, with $t(a, b, c)$, $\tau(a, b, c)$, $q(a, b, c)$ defined as in (3.4)–(3.6), we have the following expansions of the functions $\Omega_j(z)$:*

$$\tau\Omega_1(z) = \frac{tq}{4}z^4 + \frac{1}{2}z^2 - \log z + \ell_0 + \frac{C_0}{z^2} + \mathcal{O}(z^{-4}), \quad (3.9)$$

$$\tau\Omega_2(z) = \begin{cases} -\frac{3\omega^2}{4} \frac{\tau^{4/3}}{(-tq^{-1})^{1/3}} z^{4/3} - \frac{\omega}{2} \frac{\tau^{2/3}}{(-tq^{-1})^{2/3}} z^{2/3} \\ + \frac{1}{3} \log z + \ell_1 + \frac{\omega^2 C_1}{z^{2/3}} + \mathcal{O}\left(\frac{1}{z^{4/3}}\right), & \text{Im } z > 0, \\ -\frac{3\omega}{4} \frac{\tau^{4/3}}{(-tq^{-1})^{1/3}} z^{4/3} - \frac{\omega^2}{2} \frac{\tau^{2/3}}{(-tq^{-1})^{2/3}} z^{2/3} \\ + \frac{1}{3} \log z + \ell_1 + \frac{\omega C_1}{z^{2/3}} + \mathcal{O}\left(\frac{1}{z^{4/3}}\right), & \text{Im } z < 0, \end{cases} \quad (3.10)$$

$$\tau\Omega_3(z) = -\frac{3}{4} \frac{\tau^{4/3}}{(-tq^{-1})^{1/3}} z^{4/3} - \frac{1}{2} \frac{\tau^{2/3}}{(-tq^{-1})^{2/3}} z^{2/3} + \frac{1}{3} \log z + \ell_1 + \frac{C_1}{z^{2/3}} + \mathcal{O}\left(\frac{1}{z^{4/3}}\right), \quad (3.11)$$

$$\tau\Omega_4(z) = \begin{cases} -\frac{3\omega}{4} \frac{\tau^{4/3}}{(-tq^{-1})^{1/3}} z^{4/3} - \frac{\omega^2}{2} \frac{\tau^{2/3}}{(-tq^{-1})^{2/3}} z^{2/3} \\ + \frac{1}{3} \log z + \ell_1 + \frac{\omega C_1}{z^{2/3}} + \mathcal{O}\left(\frac{1}{z^{4/3}}\right), & \text{Im } z > 0, \\ -\frac{3\omega^2}{4} \frac{\tau^{4/3}}{(-tq^{-1})^{1/3}} z^{4/3} - \frac{\omega}{2} \frac{\tau^{2/3}}{(-tq^{-1})^{2/3}} z^{2/3} \\ + \frac{1}{3} \log z + \ell_1 + \frac{\omega^2 C_1}{z^{2/3}} + \mathcal{O}\left(\frac{1}{z^{4/3}}\right), & \text{Im } z < 0. \end{cases} \quad (3.12)$$

Here the constants $\ell_0 := \ell_0(a, b, c)$, $\ell_1 := \ell_1(a, b, c)$ are defined as

$$\begin{aligned} \ell_0(a, b, c) &= \frac{9a^6c^2 + 20a^4b^2c^2 + 9a^2b^4c^2 + 18a^4 + 18a^2b^2 + 18a^2c^2 + 18b^2c^2}{6(a^4b^2c^2 + 3a^4 + 3a^2b^2 + 3a^2c^2 + 3b^2c^2 - 3)} \\ &\quad - \log[A(a, b, c)], \\ \ell_1(a, b, c) &= -\frac{3(2a^6b^2 + 2a^4b^4 + 2a^4b^2c^2 + 2a^2b^4c^2 + a^4 + 4a^2b^2 + b^4)}{2a^2b^2(a^4b^2c^2 + 3a^4 + 3a^2b^2 + 3a^2c^2 + 3b^2c^2 - 3)} \\ &\quad - \frac{1}{3} \log\left[\frac{1}{3}a^2b^2A(a, b, c)\right]. \end{aligned}$$

and the constant $C_1 := C_1(a, b, c)$ is defined as

$$\begin{aligned} C_1(a, b, c) &= \frac{c^{3/2}\tau^{7/3}}{18b^{3/2}(-t)^{4/3}q^{1/6}} (3a^8b^4c^2 + 3a^6b^6c^2 + 3a^8b^2 + 3a^6b^4 + 3a^6b^2c^2 \\ &\quad + 3a^4b^6 + 3a^4b^4c^2 + 3a^2b^6c^2 - a^6 - 9a^4b^2 - 9a^2b^4 - b^6). \end{aligned} \quad (3.13)$$

The constant C_0 has an explicit expression in terms of a, b, c , but it is irrelevant in further calculations.

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 (3.8)). The expansions of the uniformization coordinate on each sheet near the branch points are given in Appendix A. ■

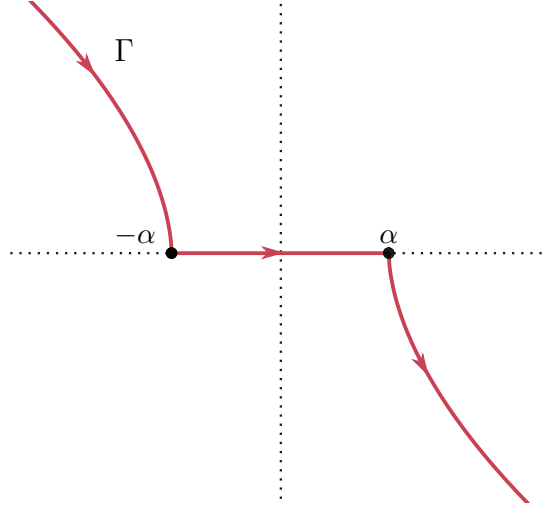


Figure 7. The contour Γ , which starts at $e^{3\pi i/4} \cdot \infty$, passes through $-\alpha$, $+\alpha$ on the real axis, and then goes off to $e^{-\pi i/4} \cdot \infty$.

3.4 Definition of the contours Γ , Γ_1 , and Γ_2

Our next step is 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 7.

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 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 8. We label the regions above Γ_1 , below Γ_2 , and bounded between Γ_1 and Γ_2 by Ω_u , Ω_ℓ , and Ω_c , respectively.

3.5 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. We will first need some setup in the uniformizing plane. Define the polar functions

$$r_+(\theta; a, b) = \sqrt{\frac{1}{2}(a^2 + b^2) + \sqrt{\frac{4}{3}a^2b^2 \sin^2 \theta + \frac{1}{4}(a^2 - b^2)^2}}, \quad \theta \in (-\pi, 0) \cup (0, \pi),$$

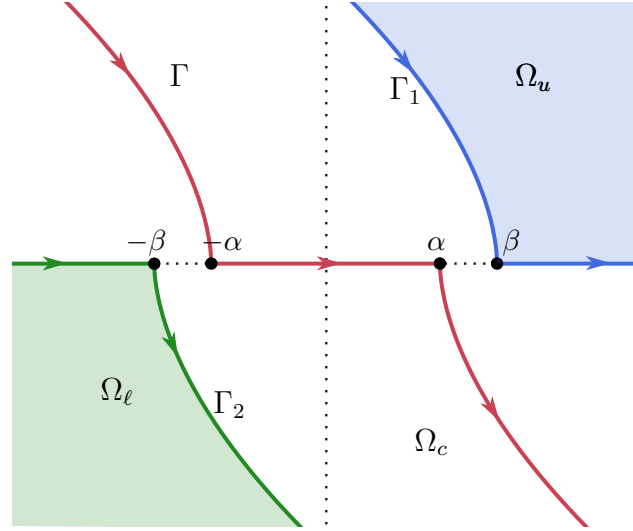


Figure 8. The contours Γ , Γ_1 , and Γ_2 . Also depicted are the regions Ω_ℓ and Ω_u (shaded), as well as Ω_c (unshaded).

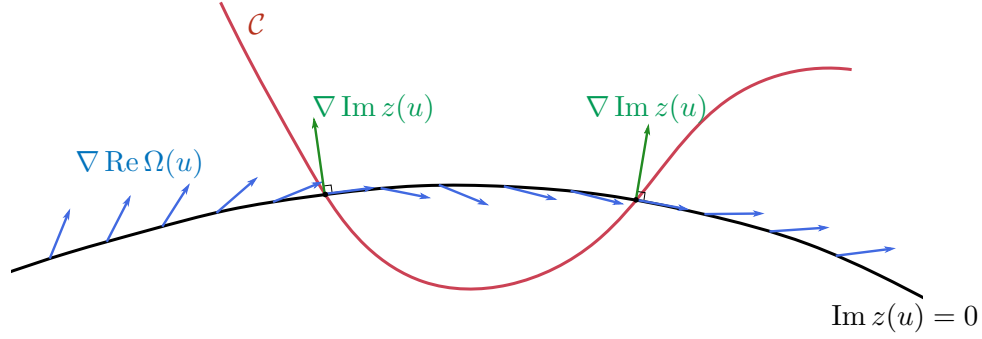


Figure 9. The curve $\text{Im } z = 0$, and the curve \mathcal{C} , which characterizes the places that $\nabla \text{Re } \Omega(u)$ and $\nabla \text{Im } z(u)$ are perpendicular. At the places where \mathcal{C} and $\text{Im } z = 0$ intersect, the normal vectors to these curves are perpendicular, implying that the direction of steepest descent may change sign, as shown above.

$$r_-(\theta; a, b) = \sqrt{\frac{1}{2}(a^2 + b^2) - \sqrt{\frac{4}{3}a^2b^2 \sin^2 \theta + \frac{1}{4}(a^2 - b^2)^2}},$$

$$\theta \in \left(-\frac{\pi}{3}, 0\right) \cup \left(0, \frac{\pi}{3}\right) \cup \left(\frac{2\pi}{3}, \pi\right) \cup \left(\pi, \frac{4\pi}{3}\right).$$

These functions define the preimages of the branch cuts (without the branch points) in the uniformizing plane in polar form: $r_+(\theta; a, b)$ defines the preimage of the cut connecting sheets 1 and 2, and $r_-(\theta; a, b)$ defines the preimages of the cuts between sheets 2 and 4 (for $\theta \in (-\frac{\pi}{3}, \frac{\pi}{3}) \setminus \{0\}$) and sheets 2 and 3 (for $\theta \in (\frac{2\pi}{3}, \frac{4\pi}{3}) \setminus \{\pi\}$). One can obtain these expressions by solving the equation $\text{Im } z(re^{i\theta}) = 0$ for r (see equation (3.3)). Note that $r_\pm(\theta; a, b) = r_\pm(\theta; b, a)$. There are six connected components of these curves, corresponding to the \pm sides of each branch cut. We now can state the main lemma.

Lemma 3.4. *Define the vector field*

$$\hat{\mathbf{n}} := \frac{\nabla \text{Im } z(u)}{\|\nabla \text{Im } z(u)\|}$$

(note that this is the normal vector to the preimages of the branch cuts in the uniformizing plane under the mapping $z(u)$). For any $(a, b, c) \in R = \{(a, b, c) \mid 0 < b \leq 1, 1 \leq a \leq b^{-1}, 0 < c \leq b\}$, the function

$$\nabla \operatorname{Re} \Omega(u) \cdot \hat{\mathbf{n}} = \frac{\partial}{\partial n} \operatorname{Re} \Omega(u)$$

is of constant sign on each connected component of the preimages of the branch cuts.

Proof. Let \mathcal{C} denote the curve where $\frac{\partial}{\partial n} \operatorname{Re} \Omega(u) = 0$. 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 $\operatorname{Im} z(u) = 0$ and \mathcal{C} do not intersect.⁶

First, since $\Omega(u)$ and $z(u)$ are analytic functions, the Cauchy–Riemann equations yield that $\partial[\operatorname{Re} \Omega(u)] = \frac{1}{2} \partial \Omega(u)$, and similarly $\partial[\operatorname{Im} z(u)] = \frac{1}{2i} \partial z(u)$ (here ∂ denotes the holomorphic derivative in u). It follows that the $\nabla \operatorname{Re} \Omega(u)$ is perpendicular to $\nabla \operatorname{Im} z(u)$ if and only if the quotient of these two expressions is purely imaginary,

$$\frac{\partial \Omega(u)}{i \partial z(u)} \in i\mathbb{R}.$$

It follows that the curve \mathcal{C} is characterized by the condition

$$\operatorname{Im} \frac{\partial \Omega(u)}{\partial z(u)} = 0,$$

which is equivalent to the condition $\operatorname{Im} \frac{\partial \Omega}{\partial z}(u) = \operatorname{Im} Y(u) = 0$ (cf. equation (3.8)). We thus must prove that the components of the branch cuts (a subset of $\operatorname{Im} z(u) = 0$ as defined before) and the corresponding components of $\operatorname{Im} Y(u) = 0$ do not intersect. We have already parameterized the branch cuts in the uniformizing plane in polar form; we can similarly parameterize the corresponding components of $\operatorname{Im} Y(u) = 0$. Since $X(u; a, b) = Y(u^{-1}, a, b)$, we have that

$$\operatorname{Im} Y(re^{i\theta}) = 0 \implies r = \frac{1}{r_{\pm}(\theta; a, c)}.$$

Therefore, we must prove that the following equations have no solution, for any value of $(a, b, c) \in R$, and any θ in the corresponding parameter ranges for r_+ , r_- :

- (1) $r_+(\theta; a, b) = \frac{1}{r_+(\theta; a, c)},$
- (2) $r_-(\theta; a, b) = \frac{1}{r_-(\theta; a, c)},$
- (3) $r_+(\theta; a, b) = \frac{1}{r_-(\theta; a, c)},$
- (4) $r_-(\theta; a, b) = \frac{1}{r_+(\theta; a, c)}.$

To show this, we state a sequence of inequalities we shall make use of. Provided $0 < c \leq b \leq 1$, for any θ in the parameter ranges for r_{\pm} , and any $1 \leq a \leq b^{-1}$, we claim that the following inequalities hold:

- (i) $r_+(\theta; a, b) \geq 1$, with equality only if $a = 1$,
- (ii) $r_-(\theta; a, b) \leq 1$, with equality only if $a = b = 1$,
- (iii) $r_+(\theta; a, b) \geq r_+(\theta; a, c)$, with equality only if $b = c$,
- (iv) $r_-(\theta; a, b) \geq r_-(\theta; a, c)$ with equality only if $b = c$.

Let us momentarily assume that (i)–(iv) hold, and see how they imply (1)–(4). For the equation (1), we have that

$$(1) \iff r_+(\theta; a, b)r_+(\theta; a, c) = 1,$$

⁶Here it is important to note that we have omitted the branch points from the definition of these curves. Indeed, $\operatorname{Im} z(u) = 0$ and \mathcal{C} may intersect at the branch points: this occurs in the case of criticality. However, it does not affect any of the arguments that follow.

and since

$$r_+(\theta; a, b)r_+(\theta; a, c) \stackrel{(iii)}{\geq} r_+(\theta; a, c)^2 \stackrel{(i)}{\geq} 1,$$

we have only to check that the equation $r_+(\theta; a, b)^2 = 1$ has no solutions, for $0 < b \leq 1$, $1 \leq a \leq b^{-1}$. Similarly, for equation (2),

$$(2) \iff r_-(\theta; a, b)r_-(\theta; a, c) = 1,$$

and so

$$r_-(\theta; a, b)r_-(\theta; a, c) \stackrel{(iv)}{\leq} r_-(\theta; a, b)^2 \stackrel{(ii)}{\leq} 1.$$

Thus, we have only to check that the equation $r_-(\theta; a, b)^2 = 1$ has no solution, for $0 < b \leq 1$, $1 \leq a \leq b^{-1}$. For equations (3) and (4), we obtain that

$$\begin{aligned} r_+(\theta; a, b)r_-(\theta; a, c) &\stackrel{(iv)}{\leq} r_+(\theta; a, b)r_-(\theta; a, b), \\ r_-(\theta; a, b)r_+(\theta; a, c) &\stackrel{(iii)}{\leq} r_+(\theta; a, b)r_-(\theta; a, b), \end{aligned}$$

and so we have only to show that $r_+(\theta; a, b)r_-(\theta; a, b) \leq 1$, for $0 < b \leq 1$, $1 \leq a \leq b^{-1}$. In summary, our original problem involving equations (1)–(4) has been reduced to showing that the following equations:

$$r_+(\theta; a, b)^2 = 1, \quad r_-(\theta; a, b)^2 = 1, \quad r_+(\theta; a, b)r_-(\theta; a, b) = 1, \quad (3.14)$$

have no solutions, provided $0 < b \leq 1$, $1 \leq a \leq b^{-1}$. Now, let us check that the first two equations (3.14) have no solutions. We have that

$$\begin{aligned} 1 = r_{\pm}(\theta)^2 &\iff 1 - \frac{1}{2}(a^2 + b^2) = \sqrt{\frac{4}{3}a^2b^2 \sin^2 \theta + \frac{1}{4}(a^2 - b^2)^2} \\ &\iff a^2b^2 - a^2 - b^2 + 1 = \frac{4}{3}a^2b^2 \sin^2 \theta \\ &\iff \frac{3}{4} \left(1 - \frac{1}{b^2} - \frac{1}{a^2} + \frac{1}{a^2b^2} \right) = \sin^2 \theta. \end{aligned}$$

Since $\theta = 0, \pi$ are excluded from the range of r_{\pm} , we need that the left hand side of the above equation to satisfy one of the inequalities

$$\frac{3}{4} \left(1 - \frac{1}{b^2} - \frac{1}{a^2} + \frac{1}{a^2b^2} \right) \leq 0 \quad \text{or} \quad \frac{3}{4} \left(1 - \frac{1}{b^2} - \frac{1}{a^2} + \frac{1}{a^2b^2} \right) > 1.$$

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

$$1 > r_+(\theta)r_-(\theta) \iff 1 > \frac{1}{4}(a^2 + b^2)^2 - \frac{4}{3}a^2b^2 \sin^2 \theta - \frac{1}{4}(a^2 - b^2)^2$$

$$\begin{aligned} &\iff a^2b^2 - 1 < \frac{4}{3}a^2b^2 \sin^2 \theta \\ &\iff \frac{3}{4} \left(1 - \frac{1}{a^2b^2} \right) < \sin^2 \theta, \end{aligned}$$

since $1 \leq a \leq b^{-1}$, $\frac{3}{4} \left(1 - \frac{1}{a^2b^2} \right) < 0$, and the desired inequality holds.

What remains to be shown is that the inequalities (i)–(iv) hold. Indeed, we have that

$$\begin{aligned} r_+(\theta; a, b) &= \sqrt{\frac{1}{2}(a^2 + b^2) + \sqrt{\frac{4}{3}a^2b^2 \sin^2 \theta + \frac{1}{4}(a^2 - b^2)^2}} \\ &\geq \sqrt{\frac{1}{2}(a^2 + b^2) + \sqrt{0 + \frac{1}{4}(a^2 - b^2)^2}} \\ &= \sqrt{\frac{1}{2}(a^2 + b^2) + \frac{1}{2}(a^2 - b^2)} = a \geq 1, \end{aligned}$$

which proves (i). Similarly,

$$\begin{aligned} r_-(\theta; a, b) &= \sqrt{\frac{1}{2}(a^2 + b^2) - \sqrt{\frac{4}{3}a^2b^2 \sin^2 \theta + \frac{1}{4}(a^2 - b^2)^2}} \\ &\leq \sqrt{\frac{1}{2}(a^2 + b^2) - \sqrt{0 + \frac{1}{4}(a^2 - b^2)^2}} \\ &= \sqrt{\frac{1}{2}(a^2 + b^2) - \frac{1}{2}(a^2 - b^2)} = b \leq 1, \end{aligned}$$

proving (ii).

Let us now prove the inequalities (iii), (iv). Set

$$f(x) := \frac{1}{2}(a^2 + x^2), \quad g(x) = \frac{4}{3}a^2x^2 \sin^2 \theta + \frac{1}{4}(a^2 - x^2)^2.$$

Then,

$$r_{\pm}(\theta; a, x) = \sqrt{f(x) \pm \sqrt{g(x)}}.$$

Note that (iii), (iv), are equivalent to showing that $r_{\pm}(\theta; a, x)$ is monotone in x , for any fixed $0 < a < x$, and any θ in the corresponding parameter ranges for r_{\pm} . We have that

$$\frac{d}{dx} \sqrt{f(x) \pm \sqrt{g(x)}} = \frac{f'(x) \pm \frac{g'(x)}{2\sqrt{g(x)}}}{2\sqrt{f(x) \pm \sqrt{g(x)}}}.$$

We must therefore show that

$$x \pm \frac{x(\frac{8}{3}a^2 \sin^2 \theta - (a^2 - x^2))}{2\sqrt{\frac{4}{3}a^2x^2 \sin^2 \theta + \frac{1}{4}(a^2 - x^2)^2}} > 0,$$

for $\sin^2 \theta < 1$ in the “+” case, and $\sin^2 \theta < \frac{3}{4}$ in the “−” case. Indeed, in the “+” case, since $0 < x < a$, algebraic manipulation yields that

$$\begin{aligned} &2\sqrt{\frac{4}{3}a^2x^2 \sin^2 \theta + \frac{1}{4}(a^2 - x^2)^2} + \frac{8}{3}a^2 \sin^2 \theta - (a^2 - x^2) \\ &> 2\sqrt{\frac{4}{3}a^2x^2 \sin^2 \theta + \frac{1}{4}(a^2 - x^2)^2} + \frac{8}{3}a^2 \sin^2 \theta > 0, \end{aligned}$$

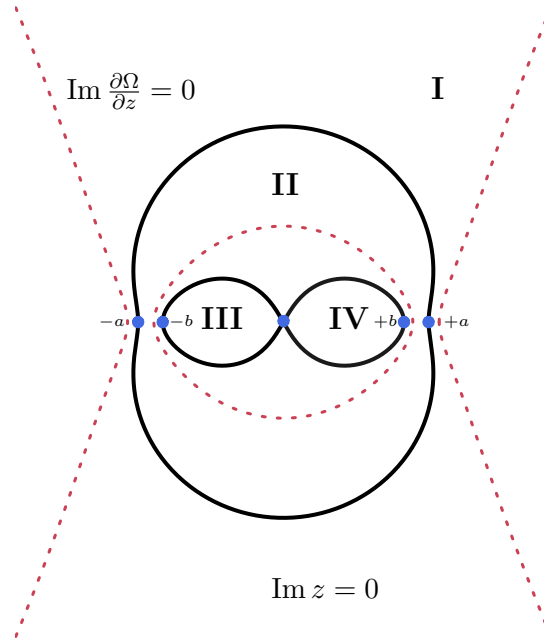


Figure 10. Images of the branch cuts $\text{Im } z(u) = 0$ and $\text{Im } \frac{\partial \Omega}{\partial z} = 0$ in the uniformizing plane, for $a = 1.059$, $b = 0.880$, $c = 0.880$. The curves indeed do not intersect, and so the direction of steepest descent is constant along each connected component of the branch cuts.

since all quantities above are positive. In the “ $-$ ” case, algebraic manipulation yields the following equivalent inequality:

$$\begin{aligned}
 2\sqrt{\frac{4}{3}a^2x^2\sin^2\theta + \frac{1}{4}(a^2 - x^2)^2} &> \frac{8}{3}a^2\sin^2\theta - (a^2 - x^2) \\
 \iff 4\left(\frac{4}{3}a^2x^2\sin^2\theta + \frac{1}{4}(a^2 - x^2)^2\right) &> \left(\frac{8}{3}a^2\sin^2\theta - (a^2 - x^2)\right)^2 \\
 \iff \frac{64}{9}a^4\sin^2\theta\left(\frac{3}{4} - \sin^2\theta\right) &> 0.
 \end{aligned}$$

Since $\sin^2\theta < \frac{3}{4}$ (recall the range of θ for r_-), we find that the above is positive, and so

$$x - \frac{x\left(\frac{8}{3}a^2\sin^2\theta - (a^2 - x^2)\right)}{2\sqrt{\frac{4}{3}a^2x^2\sin^2\theta + \frac{1}{4}(a^2 - x^2)^2}} > 0.$$

It follows that the inequalities (iii) and (iv) hold.

Thus, we have proven that, provided $(a, b, c) \in R = \{(a, b, c) \mid 0 < b \leq 1, 1 \leq a \leq b^{-1}, 0 < c \leq b\}$, 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 10 for a particular choice of the parameters (a, b, c) ; one can see explicitly that these curves do not intersect. ■

This lemma is fundamental in proving the required lensing inequalities; we now proceed to the proofs of these.

3.6 The generic (noncritical) case: $0 < b < 1$, $1 < a < b^{-1}$, $0 < c < b$

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 3.5. *Local expansions of $\Omega_j(z)$ around the branch points.*

(1) Around $z = \pm\alpha := z(\pm a)$

$$\begin{aligned}\tau\Omega_1(z) &= \eta_\alpha(z) + q_1(z - \alpha)^{3/2} + \mathcal{O}((z - \alpha)^2), \\ \tau\Omega_2(z) &= \eta_\alpha(z) - q_1(z - \alpha)^{3/2} + \mathcal{O}((z - \alpha)^2),\end{aligned}$$

and

$$\begin{aligned}\tau\Omega_1(z) &= \begin{cases} \eta_{-\alpha}(z) + iq_1(z + \alpha)^{3/2} + \mathcal{O}((z + \alpha)^2), & \text{Im } z > 0, \\ \eta_{-\alpha}(z) - iq_1(z + \alpha)^{3/2} + \mathcal{O}((z + \alpha)^2), & \text{Im } z < 0, \end{cases} \\ \tau\Omega_2(z) &= \begin{cases} \eta_{-\alpha}(z) - iq_1(z + \alpha)^{3/2} + \mathcal{O}((z + \alpha)^2), & \text{Im } z > 0, \\ \eta_{-\alpha}(z) + iq_1(z + \alpha)^{3/2} + \mathcal{O}((z + \alpha)^2), & \text{Im } z < 0, \end{cases} \\ \eta_{\pm\alpha}(z) &= \Omega(\pm\alpha) \mp \alpha \frac{a^2b^6 - 3a^2b^2 - 3b^4 - 3}{2a^2 + 6b^2}(z \mp \alpha).\end{aligned}$$

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 $z = \pm\beta := z(\pm b)$

$$\begin{aligned}\tau\Omega_2(z) &= \begin{cases} \eta_\beta(z) + i\tilde{q}_1(z - \beta)^{3/2} + \mathcal{O}((z - \beta)^2), & \text{Im } z > 0, \\ \eta_\beta(z) - i\tilde{q}_1(z - \beta)^{3/2} + \mathcal{O}((z - \beta)^2), & \text{Im } z < 0, \end{cases} \\ \tau\Omega_4(z) &= \begin{cases} \eta_\beta(z) - i\tilde{q}_1(z - \beta)^{3/2} + \mathcal{O}((z - \beta)^2), & \text{Im } z > 0, \\ \eta_\beta(z) + i\tilde{q}_1(z - \beta)^{3/2} + \mathcal{O}((z - \beta)^2), & \text{Im } z < 0, \end{cases}\end{aligned}$$

and

$$\begin{aligned}\tau\Omega_2(z) &= \eta_{-\beta}(z) + \tilde{q}_1(z + \beta)^{3/2} + \mathcal{O}((z + \beta)^2), \\ \tau\Omega_3(z) &= \eta_{-\beta}(z) - \tilde{q}_1(z + \beta)^{3/2} + \mathcal{O}((z + \beta)^2), \\ \eta_{\pm\beta}(z) &= \Omega(\pm\beta) \mp \beta \frac{a^2b^6 - 3a^2b^2 - 3b^4 - 3}{2a^2 + 6b^2}(z \mp \beta).\end{aligned}$$

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 3.6. *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 $\phi_4(z) - \phi_2(z) = \text{Re}(\Omega_4 - \Omega_2)(z) \sim \text{Re}[-2iq(z - \beta)^{3/2}] > 0$ 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 $\zeta \rightarrow z$ from the upper and lower half planes, respectively. Thus, since $\Omega(u) = \Omega(\bar{u})$, 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

$$\frac{\partial \text{Re } \Omega_2}{\partial n_+} = -\frac{\partial \text{Re } \Omega_4}{\partial n_-}, \quad \frac{\partial \text{Re } \Omega_2}{\partial n_-} = -\frac{\partial \text{Re } \Omega_4}{\partial n_+},$$

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 3.4 differ only by an overall positive factor $|z'(u)| > 0$). On the other hand, by our observation that $\Omega(u) = \Omega(\bar{u})$, we obtain the equalities

$$\frac{\partial \text{Re } \Omega_2}{\partial n_+} = \frac{\partial \text{Re } \Omega_2}{\partial n_-}, \quad \frac{\partial \text{Re } \Omega_4}{\partial n_+} = \frac{\partial \text{Re } \Omega_4}{\partial n_-}.$$

We thus compute that

$$\frac{\partial}{\partial n_{\pm}}[\text{Re}(\Omega_4 - \Omega_2)(z)] = 2\frac{\partial}{\partial n_{\pm}} \text{Re}(\Omega_4)(z).$$

Since this quantity is positive locally near $z = \beta$, Lemma 3.4 allows us to conclude that $\frac{\partial}{\partial n_{\pm}}[\text{Re}(\Omega_4 - \Omega_2)(z)] > 0$ 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 3.5, 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 3.4 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 3.7. *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 $z \in \Gamma_1 \cap \{\text{Im } z > 0\}$,
- (2) $\phi_2(z) - \phi_3(z) > 0$ for $z \in \Gamma_2 \cap \{\text{Im } z < 0\}$,
- (3) $\phi_1(z) - \phi_2(z) > 0$ for $z \in \Gamma \setminus \{\text{Im } z = 0\}$.

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 (3.5), 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 (3.10) and (3.12), 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 (3.6), 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 $z \in \Gamma_1 \cap \{\operatorname{Im} z > 0\}$. ■

Remark 3.8 (inequalities on the critical surface). In fact, all of the inequalities necessary for lensing hold on each component of the critical surface. However, this requires a separate analysis in each case, as one must perform a local expansion of the functions $\Omega_j(z)$ around each of the branch points, and the characteristic property of each of the components of the critical surface, as we shall see in the next remark, is that the functions $\Omega_j(z)$ have *different* expansions in these regions in general. Since we are only interested in the genus 0 partition function in this work, we omit the proof of these inequalities: this being said, the astute reader should have no trouble replicating the calculations performed above in each of the critical cases. We will demonstrate the proof of the relevant inequalities at the multicritical point explicitly in Part III of this work [37].

4 The first and second transformations $\mathbf{Y} \mapsto \mathbf{X} \mapsto \mathbf{U}$

4.1 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:

$$\frac{t}{N^2 \tau^2 q} f'''(z) + f'(z) - N \tau^2 z f(z) = 0, \quad (4.1)$$

whose solutions are the Pearcey-type integrals

$$w_j(z) := \int_{\gamma_j} \exp \left[N \left(\tau z w - \underbrace{\frac{1}{2} w^2 - \frac{t}{4q} w^4}_{-V(q^{-1}w)} \right) \right] dw.$$

It is also useful to notice that any solution $w(z) = w(z; \tau, t, q, N)$ to (4.1) satisfies the partial differential equations

$$\begin{aligned} \frac{\partial w}{\partial t} &= -\frac{1}{4N^3 \tau^4} \frac{\partial^4 w}{\partial z^4} = \frac{q}{4N \tau^2 t} \frac{\partial^2 w}{\partial z^2} - \frac{qz}{4t} \frac{\partial w}{\partial z} - \frac{q}{4t} w, \\ \tau \frac{\partial w}{\partial \tau} &= z \frac{\partial w}{\partial z}. \end{aligned}$$

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 $\mathbf{Y} \mapsto \mathbf{X}$.

4.2 Riemann–Hilbert problem and asymptotics for the Pearcey-type integrals $w_j(z)$

We begin by utilizing classical steepest descent analysis to determine the asymptotics of each of the integrals

$$w_j(z) = \int_{\gamma_j} \exp[N(\tau z w - V(w))] dw, \quad (4.2)$$

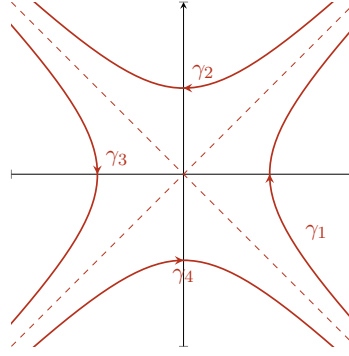


Figure 11. The contours γ_j , $j = 1, \dots, 4$.

as $z \rightarrow \infty$, where γ_j , $j = 1, \dots, 4$ are one of the contours in Figure 11. Before studying the asymptotics of the integrals (4.2), we will study the solutions to the so-called saddle point equation. The saddle point equation is defined to be

$$0 = \frac{d}{dw} \left[\tau zw - \frac{1}{2} w^2 - \frac{t}{4q} w^4 \right] \iff w^3 + \frac{q}{t} w - \frac{\tau q}{t} z = 0.$$

The large z asymptotics of the integrals (4.1) are determined in terms of certain solutions to the saddle point equation, in a sense we shall make explicit later on. We define $s_k(z)$, $k = 1, 2, 3$ to be particular solutions to the saddle point equation, defined as follows. First, we define the Riemann surface of this equation as \mathcal{S} , consisting of three sheets \mathcal{S}_k , $k = 1, 2, 3$,

$$\begin{aligned} \mathcal{S}_1 &:= \mathbb{C} \setminus ((-\infty, -z^*] \cup [z^*, \infty)), \\ \mathcal{S}_2 &:= \mathbb{C} \setminus [z^*, \infty), \\ \mathcal{S}_3 &:= \mathbb{C} \setminus (-\infty, -z^*]. \end{aligned}$$

Here $z^* := \pm \frac{2}{3\tau\sqrt{3}} \sqrt{\frac{q}{-t}} > 0$ are the branch points of the saddle point equation. Let $s_2(z)$ be the unique solution of the saddle point equation which is real for $z \in \mathbb{R} \setminus \{\pm z^*\}$. This solution admits an analytic continuation to all of \mathcal{S}_2 . We then define $s_1(z)$, $s_3(z)$ as the analytic continuations of $s_2(z)$ to sheets \mathcal{S}_1 and \mathcal{S}_3 , respectively. We have the following lemma, regarding the properties of the $s_k(z)$.

Lemma 4.1.

(1) *Boundary values of $s_k(z)$. The following relations hold:*

$$s_{3,\pm}(z) = s_{1,\mp}(z), \quad z \in (-\infty, -z^*], \quad s_{1,\pm}(z) = s_{2,\mp}(z), \quad z \in [z^*, \infty).$$

(2) *Asymptotics of $s_k(z)$. As $z \rightarrow \infty$,*

$$\begin{aligned} s_1(z) &= \begin{cases} -\omega^2 \left(\frac{\tau q}{-t} \right)^{1/3} z^{1/3} [1 + \mathcal{O}(z^{-1/3})], & \text{Im } z > 0, \\ -\omega \left(\frac{\tau q}{-t} \right)^{1/3} z^{1/3} [1 + \mathcal{O}(z^{-1/3})], & \text{Im } z < 0, \end{cases} \\ s_2(z) &= -\left(\frac{\tau q}{-t} \right)^{1/3} z^{1/3} [1 + \mathcal{O}(z^{-1/3})], \\ s_3(z) &= \begin{cases} -\omega \left(\frac{\tau q}{-t} \right)^{1/3} z^{1/3} [1 + \mathcal{O}(z^{-1/3})], & \text{Im } z > 0, \\ -\omega^2 \left(\frac{\tau q}{-t} \right)^{1/3} z^{1/3} [1 + \mathcal{O}(z^{-1/3})], & \text{Im } z < 0. \end{cases} \end{aligned}$$

- (3) *Symmetries of $s_k(z)$.* For each $k = 1, 2, 3$, we have that $s_k(-z) = -s_k(z)$, $s_k(\bar{z}) = \overline{s_k(z)}$. Furthermore, if $\operatorname{Re} z < 0$, then $\operatorname{Re} s_2(z) > 0$.

These properties are straightforward to verify; the proof is so similar to that of [40, Lemmas 2.1 and 2.3] that we omit it. For further details, we refer to this work. Consequentially, we have the following corollary.

Corollary 4.2. *For $s_k(z)$ defined as above, $k = 1, 2, 3$, put*

$$\theta_k(z) := \tau s_k(z) - V(s_k(z)), \quad k = 1, 2, 3.$$

(Note that $\theta'_k(z) = \tau s_k(z)$, so that $\theta_k(z)$ are, up to a factor of τ , the primitives of the functions $s_k(z)$). Then, the functions $\theta_k(z)$ have the large- z asymptotics

$$\begin{aligned} \theta_1(z) &= \begin{cases} -\frac{3\omega^2}{4} \frac{\tau^{4/3}}{(-tq^{-1})^{1/3}} z^{4/3} - \frac{\omega}{2} \frac{\tau^{2/3}}{(-tq^{-1})^{2/3}} z^{2/3} \\ + \frac{q}{6t} - \frac{\omega^2}{54\tau^{2/3}(-tq^{-1})^{4/3} z^{2/3}} + \mathcal{O}\left(\frac{1}{|z|^{4/3}}\right), & \operatorname{Im} z > 0, \\ -\frac{3\omega^2}{4} \frac{\tau^{4/3}}{(-tq^{-1})^{1/3}} z^{4/3} - \frac{\omega}{2} \frac{\tau^{2/3}}{(-tq^{-1})^{2/3}} z^{2/3} \\ + \frac{q}{6t} - \frac{\omega^2}{54\tau^{2/3}(-tq^{-1})^{4/3} z^{2/3}} + \mathcal{O}\left(\frac{1}{|z|^{4/3}}\right), & \operatorname{Im} z < 0, \end{cases} \\ \theta_2(z) &= -\frac{3}{4} \frac{\tau^{4/3}}{(-tq^{-1})^{1/3}} z^{4/3} - \frac{1}{2} \frac{\tau^{2/3}}{(-tq^{-1})^{2/3}} z^{2/3} + \frac{q}{6t} - \frac{1}{54\tau^{2/3}(-tq^{-1})^{4/3} z^{2/3}} + \mathcal{O}\left(\frac{1}{|z|^{4/3}}\right), \\ \theta_3(z) &= \begin{cases} -\frac{3\omega}{4} \frac{\tau^{4/3}}{(-tq^{-1})^{1/3}} z^{4/3} - \frac{\omega^2}{2} \frac{\tau^{2/3}}{(-tq^{-1})^{2/3}} z^{2/3} \\ + \frac{q}{6t} - \frac{\omega}{54\tau^{2/3}(-tq^{-1})^{4/3} z^{2/3}} + \mathcal{O}\left(\frac{1}{|z|^{4/3}}\right), & \operatorname{Im} z > 0, \\ -\frac{3\omega^2}{4} \frac{\tau^{4/3}}{(-tq^{-1})^{1/3}} z^{4/3} - \frac{\omega}{2} \frac{\tau^{2/3}}{(-tq^{-1})^{2/3}} z^{2/3} \\ + \frac{q}{6t} - \frac{\omega^2}{54\tau^{2/3}(-tq^{-1})^{4/3} z^{2/3}} + \mathcal{O}\left(\frac{1}{|z|^{4/3}}\right), & \operatorname{Im} z < 0. \end{cases} \end{aligned} \quad (4.3)$$

With this information about the saddle points $s_k(z)$ in place, we are ready to state the following proposition.

Proposition 4.3. *Define the function $S(w)$ by*

$$S(w) := \sqrt{\frac{2\pi}{N(1 + 3tq^{-1}w^2)}} e^{N(\tau zw - V(z))}.$$

Then, the functions $w_j(z)$, $j = 1, 2, 3, 4$, defined by (4.2), have large- z asymptotics given by

$$\begin{aligned} w_1(z) &= -S(s_3(z)) [1 + \mathcal{O}(z^{-4/3})], \\ w_2(z) &= \begin{cases} S(s_1(z)) [1 + \mathcal{O}(z^{-4/3})], & 0 < \arg z < \pi, \\ S(s_3(z)) [1 + \mathcal{O}(z^{-4/3})], & -\frac{\pi}{2} < \arg z < 0, \\ S(s_2(z)) [1 + \mathcal{O}(z^{-4/3})], & \pi < \arg z < -\frac{\pi}{2}, \end{cases} \\ w_3(z) &= -S(s_2(z)) [1 + \mathcal{O}(z^{-4/3})], \\ w_4(z) &= \begin{cases} S(s_2(z)) [1 + \mathcal{O}(z^{-4/3})], & \frac{\pi}{2} < \arg z < \pi, \\ S(s_3(z)) [1 + \mathcal{O}(z^{-4/3})], & 0 < \arg z < \frac{\pi}{2}, \\ S(s_1(z)) [1 + \mathcal{O}(z^{-4/3})], & -\pi < \arg z < 0. \end{cases} \end{aligned}$$

Proof. Classical steepest descent analysis (cf., for example, [71]) tells us that the asymptotics of the integrals $w_j(z)$ are given by

$$w_j(z) = \pm \sqrt{\frac{2\pi}{NV''(s_*(z))}} e^{N(\tau z s_*(z) - V(s_*(z)))} (1 + \mathcal{O}(z^{-2/3})),$$

where $*$ $\in \{1, 2, 3\}$, and $s_*(z)$ is the dominant contributing saddle point to the integral. The proposition then follows from determining the dominant saddle in each quadrant; the sign in front of the square root is determined by the orientation of the contour γ_j as it passes through the relevant saddle point. \blacksquare

We now define the row vectors

$$\vec{w}_j(z) := \left(w_j(z), \frac{w'_j(z)}{N\tau}, \frac{w''_j(z)}{(N\tau)^2} \right),$$

where $'$ here denotes the derivative with respect to z . We now define a 3×3 matrix $\mathbf{W}(z)$ as

$$\mathbf{W}(z) = \begin{cases} \begin{pmatrix} -\vec{w}_2(z) \\ \vec{w}_3(z) \\ \vec{w}_1(z) \end{pmatrix}, & z \in \Omega_u, \\ \begin{pmatrix} \vec{w}_3(z) + \vec{w}_4(z) \\ \vec{w}_3(z) \\ \vec{w}_1(z) \end{pmatrix}, & z \in \Omega_c, \\ \begin{pmatrix} \vec{w}_4(z) \\ \vec{w}_3(z) \\ \vec{w}_1(z) \end{pmatrix}, & z \in \Omega_\ell. \end{cases}$$

$\mathbf{W}(z)$ is defined so that it is the unique solution to the following Riemann–Hilbert problem.

Proposition 4.4. $\mathbf{W}(z)$ is the unique solution to the following Riemann–Hilbert problem:

- (1) $\mathbf{W}(z)$ is analytic in $\mathbb{C} \setminus (\Gamma_1 \cup \Gamma_2)$, with boundary values

$$\mathbf{W}_+(z) = \begin{cases} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \mathbf{W}_-(z) =: J_1 \mathbf{W}_-(z), & z \in \Gamma_1, \\ \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \mathbf{W}_-(z) =: J_2 \mathbf{W}_-(z), & z \in \Gamma_2 \end{cases}$$

(note that these matrices commute, and that their product is $J_1 J_2 = J$).

- (2) At infinity, $\mathbf{W}(z)$ is normalized as

$$\mathbf{W}(z) = c_N \cdot e^{N\Theta(z)} A(z) B(z) \hat{K} \left[\mathbb{I}_{3 \times 3} + \mathcal{O}\left(\frac{1}{z}\right) \right], \quad z \rightarrow \infty,$$

where the constant $c_N := i\sqrt{\frac{2\pi}{3N}} e^{N\frac{q}{6t}}$, the matrix $\Theta(z)$ is defined to be

$$\Theta(z) = \begin{cases} \text{diag}(\lambda_3(z), \lambda_1(z), \lambda_2(z)), & \text{Im } z > 0, \\ \text{diag}(\lambda_2(z), \lambda_1(z), \lambda_3(z)), & \text{Im } z < 0, \end{cases} \quad (4.4)$$

where the functions $\lambda_k(z)$ are defined by the exact formulas

$$\lambda_k(z) = -\frac{3\omega^{k-1}}{4} \frac{\tau^{4/3}}{(-tq^{-1})^{1/3}} z^{4/3} - \frac{\omega^{1-k}}{2} \frac{\tau^{2/3}}{(-tq^{-1})^{2/3}} z^{2/3}.$$

and finally, the matrices $A(z)$, $B(z)$, and \hat{K} are given by

$$A(z) = \begin{cases} \begin{pmatrix} -\omega & 1 & \omega^2 \\ -1 & 1 & 1 \\ -\omega^2 & 1 & \omega \end{pmatrix}, & \text{Im } z > 0, \\ \begin{pmatrix} \omega^2 & -1 & -\omega \\ -1 & 1 & 1 \\ -\omega & 1 & \omega^2 \end{pmatrix}, & \text{Im } z < 0, \end{cases} \quad B(z) = \begin{pmatrix} z^{-1/3} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & z^{1/3} \end{pmatrix},$$

$$\hat{K} = \begin{pmatrix} (-tq^{-1})^{-1/6} \tau^{-1/3} & 0 & -\frac{n+27tq^{-1}}{54(-tq^{-1})^{13/6} \tau^{1/3}} \\ 0 & (-tq^{-1})^{-1/2} & 0 \\ 0 & 0 & -(-tq^{-1})^{-5/6} \tau^{1/3} \end{pmatrix}.$$

Proof. The proof of this proposition 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 $z \rightarrow \infty$ in the functions $\theta_j(z)$ can be absorbed into the asymptotic 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 is⁷

$$\mathbb{I}_{3 \times 3} + \begin{pmatrix} 0 & \frac{-N^3 - 72N^2t/q - 891Nt^2/q^2 - 810t^3/q^3}{5832Nt^3/q^3\tau} & 0 \\ \frac{-N+9t/q}{54\tau t/q} & 0 & \frac{N^3+36N^2t/q-81Nt^2/q^2-810t^3/q^3}{5832Nt^3/q^3\tau} \\ 0 & \frac{-N-9t/q}{54\tau t/q} & 0 \end{pmatrix} z^{-1} + \mathcal{O}(z^{-2}). \quad \blacksquare$$

Remark 4.5. We have introduced functions $\lambda_k(z)$ in the above proposition. These functions have been chosen so that

$$\text{diag}(\theta_1(z), \theta_2(z), \theta_3(z)) = \Theta(z) + \mathcal{O}(1), \quad z \rightarrow \infty.$$

The reason for the introduction of these functions is to match the asymptotic theory of ordinary differential equations with rational coefficients; such an equation will have a large z expansion consistent with the above.

We furthermore have the following proposition, which will be useful later in the computation of the τ -function.

Proposition 4.6. $\mathbf{W}(z)$ satisfies the following differential equations:

$$\begin{aligned} \frac{\partial \mathbf{W}}{\partial z} &= \mathbf{W} \cdot \mathcal{M}^z(z), & \frac{\partial \mathbf{W}}{\partial t} &= \mathbf{W} \cdot \mathcal{M}^t(z), \\ \frac{\partial \mathbf{W}}{\partial \tau} &= \mathbf{W} \cdot \mathcal{M}^\tau(z), & \frac{\partial \mathbf{W}}{\partial H} &= \mathbf{W} \cdot \mathcal{M}^H(z), \end{aligned} \quad (4.5)$$

⁷The 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) (4.6), and determining the coefficients of the subleading expansion recursively.

where the matrices $\mathcal{M}^z(z)$, $\mathcal{M}^t(z)$, $\mathcal{M}^\tau(z)$, and $\mathcal{M}^H(z)$ are defined by

$$\begin{aligned}\mathcal{M}^z(z) &= N\tau \begin{pmatrix} 0 & 0 & \tau qz/t \\ 1 & 0 & -q/t \\ 0 & 1 & 0 \end{pmatrix}, \\ \mathcal{M}^t(z) &= \frac{N}{4t} \begin{pmatrix} -1/N & \tau qz/t & -\tau^2 qz^2/t + \frac{q}{Nt} \\ -\tau z & -(q/t + 2/N) & 2\tau qz/t \\ 1 & -\tau z & -(q/t + 3/N) \end{pmatrix}, \\ \mathcal{M}^\tau(z) &= Nz \begin{pmatrix} 0 & 0 & \tau qz/t \\ 1 & 0 & -q/t \\ 0 & 1 & 0 \end{pmatrix}, \\ \mathcal{M}^H(z) &= -\frac{N}{4} \begin{pmatrix} -1/N & \tau qz/t & -\tau^2 qz^2/t + \frac{q}{Nt} \\ -\tau z & -(q/t + 2/N) & 2\tau qz/t \\ 1 & -\tau z & -(q/t + 3/N) \end{pmatrix}.\end{aligned}$$

Proof. This can be inferred immediately from the differential equation (4.1) and the relations

$$\frac{\partial w}{\partial t} = -\frac{e^{-H}}{4N^3\tau^4} \frac{\partial^4 w}{\partial z^4}, \quad \tau \frac{\partial w}{\partial \tau} = z \frac{\partial w}{\partial z}, \quad \frac{\partial w}{\partial H} = -t \frac{\partial w}{\partial t}.$$

As a consistency check, one may verify that the compatibility conditions between these equations (i.e., the zero-curvature equations) hold trivially. \blacksquare

Remark 4.7. 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

$$A_+(z)B_+(z) = \begin{cases} \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{pmatrix} A_-(z)B_-(z), & z > 0, \\ \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} A_-(z)B_-(z), & z < 0. \end{cases}$$

This is consistent with the fact that, for $z \in \mathbb{R}$ sufficiently large, $\Theta(z)$ has jumps

$$\Theta_+(z) = \begin{cases} \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{pmatrix} \Theta_-(z), & z > R > 0, \\ \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \Theta_-(z), & z < -R < 0. \end{cases}$$

4.3 The transformation $\mathbf{Y} \mapsto \mathbf{X}$

We now define the transformation $\mathbf{Y} \mapsto \mathbf{X}$. We set

$$\mathbf{X}(z) := \begin{pmatrix} 1 & 0 \\ 0 & c_N \hat{K} \end{pmatrix} \mathbf{Y}(z) \begin{pmatrix} e^{-NV(z)} & 0 \\ 0 & \mathbf{W}^{-1}(z) \end{pmatrix}.$$

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 4.8. $\mathbf{X}(z)$ is the unique solution to the following Riemann–Hilbert Problem:

$$\mathbf{X}_+(z) = \mathbf{X}_-(z) \times \begin{cases} \begin{pmatrix} 1 & 0 \\ 0 & J_1^{-1} \end{pmatrix}, & z \in \Gamma_1, \\ \begin{pmatrix} 1 & 0 \\ 0 & J_2^{-1} \end{pmatrix}, & z \in \Gamma_2, \\ \begin{pmatrix} 1 & 1 & 0 & 0 \\ \vec{0}_{1 \times 3} & \mathbb{I}_{3 \times 3} \end{pmatrix}, & z \in \Gamma, \end{cases}$$

Subject to the normalization condition

$$\mathbf{X}(z) = \left[\mathbb{I} + \mathcal{O}\left(\frac{1}{z}\right) \right] \begin{pmatrix} 1 & 0 \\ 0 & B^{-1}(z)A^{-1}(z) \end{pmatrix} \begin{pmatrix} z^n & 0 & 0 & 0 \\ 0 & z^{-n/3} & 0 & 0 \\ 0 & 0 & z^{-n/3} & 0 \\ 0 & 0 & 0 & z^{-n/3} \end{pmatrix} e^{-N\Lambda(z)}, \quad (4.6)$$

where $\Lambda(z)$ is defined as the diagonal matrix

$$\Lambda(z) = \begin{pmatrix} V(z) & 0 \\ 0 & \Theta(z) \end{pmatrix},$$

where $\Theta(z)$ is the diagonal 3×3 matrix defined by (4.4).

Proof. The asymptotic condition (4.6) 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

$$J_{\mathbf{Y}} = \mathbb{I} + e^{-NV(z)} \begin{pmatrix} 0 & f(z) & \frac{f'(z)}{N\tau} & \frac{f''(z)}{(N\tau)^2} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} =: \mathbb{I} + e^{-NV(z)} \begin{pmatrix} 0 & \vec{f}(z) & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

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

$$\begin{pmatrix} e^{NV(z)} & 0 \\ 0 & \mathbf{W}(z) \end{pmatrix} \left[\mathbb{I} + e^{-NV(z)} \begin{pmatrix} 0 & \vec{f}(z) & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \right] \begin{pmatrix} e^{-NV(z)} & 0 \\ 0 & \mathbf{W}^{-1}(z) \end{pmatrix}$$

$$\begin{aligned}
&= \mathbb{I} + \begin{pmatrix} 0 & \vec{f}(z) \mathbf{W}^{-1}(z) & & \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = \mathbb{I} + \begin{pmatrix} 0 & (-\vec{w}_2(z) - \vec{w}_1(z)) \mathbf{W}^{-1}(z) & & \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \\
&= \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.
\end{aligned}$$

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 4.9. Note that, if we had equivalently chosen $\gamma_3 + \gamma_4$ as the homological representative for Γ , the same resulting jump matrix is obtained.

4.4 The transformation $\mathbf{X} \mapsto \mathbf{U}$

Define the matrix

$$\mathbf{G}(z) := \begin{pmatrix} \exp[n\tau\Omega_1(z)] & 0 & 0 & 0 \\ 0 & \exp[n\tau\Omega_2(z)] & 0 & 0 \\ 0 & 0 & \exp[n\tau\Omega_3(z)] & 0 \\ 0 & 0 & 0 & \exp[n\tau\Omega_4(z)] \end{pmatrix}.$$

We now are ready to perform the transformation $\mathbf{X} \mapsto \mathbf{U}$. Set

$$\mathbf{U}(z) := [\mathbb{I} - nC_1 \cdot E_{24}] e^{-nL} \mathbf{X}(z) \mathbf{G}(z),$$

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 (3.13)), and L is the diagonal constant (in z) matrix

$$L := \text{diag}(\ell_0, \ell_1, \ell_1, \ell_1).$$

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 4.10. *The function $\mathbf{U}(z)$ is the unique solution to the following Riemann–Hilbert problem:*

$$\mathbf{U}_+(z) = \mathbf{U}_-(z)$$

$$\begin{aligned}
& \begin{cases} \mathbb{I} - E_{24}e^{-n\tau[\Omega_2(z)-\Omega_4(z)]}, & z \in \Gamma_1 \cap \{\operatorname{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 \{\operatorname{Im} z = 0\}, \\ \mathbb{I} - E_{23}e^{-n\tau[\Omega_2(z)-\Omega_3(z)]}, & z \in \Gamma_2 \cap \{\operatorname{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 \{\operatorname{Im} z = 0\}, \\ \mathbb{I} + E_{12}e^{-n\tau[\Omega_1(z)-\Omega_2(z)]}, & z \in \Gamma \setminus \{\operatorname{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 \{\operatorname{Im} z = 0\}. \end{cases} \\
& \times \begin{cases} \mathbb{I} - E_{24}e^{-n\tau[\Omega_2(z)-\Omega_4(z)]}, & z \in \Gamma_1 \cap \{\operatorname{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 \{\operatorname{Im} z = 0\}, \\ \mathbb{I} - E_{23}e^{-n\tau[\Omega_2(z)-\Omega_3(z)]}, & z \in \Gamma_2 \cap \{\operatorname{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 \{\operatorname{Im} z = 0\}, \\ \mathbb{I} + E_{12}e^{-n\tau[\Omega_1(z)-\Omega_2(z)]}, & z \in \Gamma \setminus \{\operatorname{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 \{\operatorname{Im} z = 0\}. \end{cases}
\end{aligned}$$

The asymptotics of $\mathbf{U}(z)$ are given by

$$\mathbf{U}(z) = \left[\mathbb{I} + \mathcal{O}\left(\frac{1}{z^{1/3}}\right) \right] \begin{pmatrix} 1 & 0 \\ 0 & B^{-1}(z)A^{-1}(z) \end{pmatrix}, \quad z \rightarrow \infty.$$

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 (3.9)–(3.12) with (4.6), along with the explicit expressions for the asymptotics of the functions $\lambda_k(z)$ (see equation (4.3)) shows that this is indeed the case. Indeed, we have that, as $z \rightarrow \infty$,

$$\begin{aligned}
\mathbf{U}(z) &= [\mathbb{I} - nC_1 \cdot E_{24}]e^{-nL} [\mathbb{I} + \mathcal{O}(z^{-1})] \begin{pmatrix} z^n & 0 \\ 0 & z^{-n/3} \mathbb{I}_{3 \times 3} \end{pmatrix} \\
&\quad \times \begin{pmatrix} 1 & 0 \\ 0 & B^{-1}(z)A^{-1}(z) \end{pmatrix} e^{-n\hat{\Theta}(z)} \mathbf{G}(z) \\
&= [\mathbb{I} - nC_1 \cdot E_{24}] [\mathbb{I} + \mathcal{O}(z^{-1})] \begin{pmatrix} 1 & 0 \\ 0 & B^{-1}(z)A^{-1}(z) \end{pmatrix} \\
&\quad \times \begin{pmatrix} 1 + \mathcal{O}(z^{-2}) & 0 \\ 0 & \mathbb{I}_{3 \times 3} + n\hat{C}z^{-2/3} + \mathcal{O}(z^{-4/3}) \end{pmatrix}.
\end{aligned}$$

Here \hat{C} is the piecewise constant diagonal matrix

$$\hat{C} = \begin{cases} \operatorname{diag}(\omega^2 C_1, C_1, \omega C_1), & \operatorname{Im} z > 0, \\ \operatorname{diag}(\omega C_1, C_1, \omega^2 C_1), & \operatorname{Im} z < 0. \end{cases}$$

If we interchange the order of the last two matrices, we obtain that, as $z \rightarrow \infty$,

$$\begin{aligned}
\mathbf{U}(z) &= [\mathbb{I} - nC_1 \cdot E_{24}] [\mathbb{I} + \mathcal{O}(z^{-1})] [\mathbb{I} + nC_1 \cdot E_{24} + \mathcal{O}(z^{-1/3})] \begin{pmatrix} 1 & 0 \\ 0 & B^{-1}(z)A^{-1}(z) \end{pmatrix} \\
&= [\mathbb{I} + \mathcal{O}(z^{-1/3})] \begin{pmatrix} 1 & 0 \\ 0 & B^{-1}(z)A^{-1}(z) \end{pmatrix}.
\end{aligned}$$

We will analyze the jumps of this new matrix after opening lenses; this is performed in the next section. ■

5 The third and fourth transformations $\mathbf{U} \mapsto \mathbf{T} \mapsto \mathbf{S}$

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 3.5 guarantee the same inequalities hold around the branch points. These cases will become distinguished when we later try to find a parametrix.

5.1 The transformation $\mathbf{U} \mapsto \mathbf{T}$

The opening of lenses here is based on the factorization of the jump matrix

$$\begin{pmatrix} e^{-ng_+(z)} & -1 \\ 0 & e^{-ng_-(z)} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -e^{-ng_-(z)} & 1 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -e^{-ng_+(z)} & 1 \end{pmatrix},$$

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 those depicted in Figure 12 around $(-\infty, -\beta]$ (respectively, $[\beta, \infty)$) such that the differences $\text{Re}[\Omega_3 - \Omega_2]$ (respectively, $\text{Re}[\Omega_4 - \Omega_2]$) 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 12. Define matrices

$$V_1(z) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -e^{-n\tau[\Omega_4(z) - \Omega_2(z)]} & 0 & 1 \end{pmatrix},$$

$$V_2(z) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -e^{-n\tau[\Omega_3(z) - \Omega_2(z)]} & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

We define the transformation $\mathbf{U} \mapsto \mathbf{T}$ by setting

$$\mathbf{T}(z) = \begin{cases} \mathbf{U}(z)V_1^{-1}(z), & z \in \Sigma_{1,u}, \\ \mathbf{U}(z)V_1(z), & z \in \Sigma_{1,l}, \\ \mathbf{U}(z)V_2^{-1}(z), & z \in \Sigma_{2,u}, \\ \mathbf{U}(z)V_2(z), & z \in \Sigma_{2,l}, \\ \mathbf{U}(z), & \text{elsewhere.} \end{cases}$$

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.

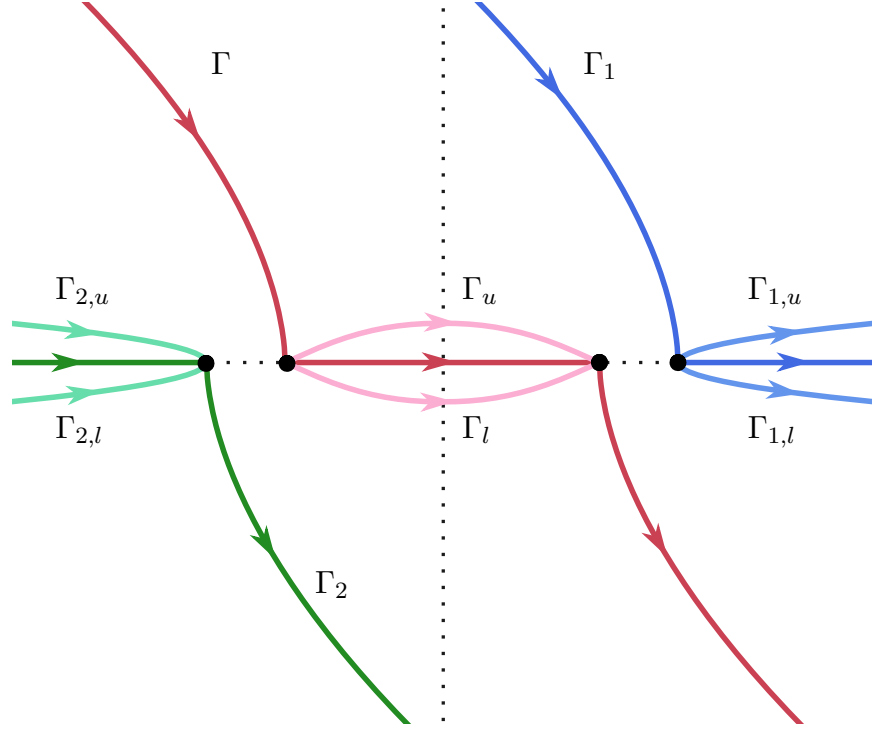


Figure 12. The opened lenses.

Proposition 5.1. *The function $\mathbf{T}(z)$ is the unique solution to the following RHP:*

$$\mathbf{T}_+(z) = \mathbf{T}_-(z) \begin{cases} \begin{pmatrix} \mathbb{I} - E_{42}e^{-n\tau[\Omega_4(z)-\Omega_2(z)]}, \\ \mathbb{I} - E_{32}e^{-n\tau[\Omega_3(z)-\Omega_2(z)]}, \\ \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, \\ \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \end{pmatrix} \begin{array}{l} z \in \Gamma_{1,u} \cup \Gamma_{1,l}, \\ z \in \Gamma_{2,u} \cup \Gamma_{2,l}, \\ \\ z \in [\beta, \infty), \\ \\ z \in (-\infty, -\beta], \end{array} \\ \begin{pmatrix} \mathbb{I} - E_{24}e^{-n\tau[\Omega_2(z)-\Omega_4(z)]}, \\ \mathbb{I} - E_{23}e^{-n\tau[\Omega_2(z)-\Omega_3(z)]}, \\ \mathbb{I} + E_{12}e^{-n\tau[\Omega_1(z)-\Omega_2(z)]}, \\ \begin{pmatrix} e^{-n\tau[\Omega_{1,-}(z)-\Omega_{2,-}(z)]} & 1 & 0 & 0 \\ 0 & e^{n\tau[\Omega_{1,-}(z)-\Omega_{2,-}(z)]} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \end{pmatrix} \begin{array}{l} z \in \Gamma_1 \cap \{\operatorname{Im} z > 0\}, \\ z \in \Gamma_2 \cap \{\operatorname{Im} z < 0\}, \\ z \in \Gamma \setminus [-\alpha, \alpha], \\ z \in [-\alpha, \alpha]. \end{array} \end{cases}$$

Furthermore, $\mathbf{T}(z)$ has asymptotics

$$\mathbf{T}(z) = [\mathbb{I} + \mathcal{O}(z^{-1/3})] \begin{pmatrix} 1 & 0 \\ 0 & B^{-1}(z)A^{-1}(z) \end{pmatrix}, \quad z \rightarrow \infty.$$

Proof. The proof of this proposition follows immediately from the definition of $\mathbf{T}(z)$. ■

5.2 The transformation $\mathbf{T} \mapsto \mathbf{S}$

We now open the lens around the segment $[-\alpha, \alpha]$. This is based on the following factorization of the jump matrix:

$$\begin{aligned} & \begin{pmatrix} e^{-n\tau[\Omega_{2,+}(z)-\Omega_{1,+}(z)]} & 1 & 0 & 0 \\ 0 & e^{-n\tau[\Omega_{2,-}(z)-\Omega_{1,-}(z)]} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ e^{-n\tau[\Omega_{2,-}(z)-\Omega_{1,-}(z)]} & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \\ &\quad \times \begin{pmatrix} 1 & 0 & 0 & 0 \\ e^{-n\tau[\Omega_{2,+}(z)-\Omega_{1,+}(z)]} & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \end{aligned}$$

By the lensing propositions of Section 3, there exists a lens-shaped region around $[-\alpha, \alpha]$ such that

$$\operatorname{Re}[\Omega_2 - \Omega_1](z) > 0.$$

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

$$V_0(z) := \begin{pmatrix} 1 & 0 & 0 & 0 \\ e^{-n\tau[\Omega_2(z)-\Omega_1(z)]} & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

We define the piecewise analytic function $\mathbf{S}(z)$ by

$$\mathbf{S}(z) := \begin{cases} \mathbf{T}(z)V_0^{-1}(z), & z \in \Sigma_l, \\ \mathbf{T}(z)V_0(z), & z \in \Sigma_u, \\ \mathbf{T}(z), & \text{otherwise.} \end{cases}$$

In this case, $\mathbf{S}(z)$ is the unique solution to the following RHP.

Proposition 5.2. *The function $\mathbf{S}(z)$ is the unique solution to the following Riemann–Hilbert problem: $\mathbf{S}(z)$ is piecewise analytic off the contours*

$$\Gamma_1, \Gamma_2, \Gamma, \Gamma_{1,u}, \Gamma_{1,l}, \Gamma_{2,u}, \Gamma_{2,l}, \Gamma_l, \Gamma_u,$$

satisfying the jump condition

$$\mathbf{S}_+(z) = \mathbf{S}_-(z) \begin{cases} \mathbb{I} - E_{42}e^{-n\tau[\Omega_4(z)-\Omega_2(z)]}, & z \in \Gamma_{1,u} \cup \Gamma_{1,l}, \\ \mathbb{I} - E_{32}e^{-n\tau[\Omega_3(z)-\Omega_2(z)]}, & z \in \Gamma_{2,u} \cup \Gamma_{2,l}, \\ \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, & z \in [\beta, \infty), \\ \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, & z \in (-\infty, -\beta], \\ \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, & z \in [-\alpha, \alpha], \\ \mathbb{I} - E_{24}e^{-n\tau[\Omega_2(z)-\Omega_4(z)]}, & z \in \Gamma_1 \cap \{\operatorname{Im} z > 0\}, \\ \mathbb{I} - E_{23}e^{-n\tau[\Omega_2(z)-\Omega_3(z)]}, & z \in \Gamma_2 \cap \{\operatorname{Im} z < 0\}, \\ \mathbb{I} + E_{12}e^{-n\tau[\Omega_1(z)-\Omega_2(z)]}, & z \in \Gamma \setminus [-\alpha, \alpha], \\ \mathbb{I} + E_{21}e^{-n\tau[\Omega_2(z)-\Omega_1(z)]}, & z \in \Gamma_u \cup \Gamma_l. \end{cases}$$

Furthermore, $\mathbf{S}(z)$ has asymptotics

$$\mathbf{S}(z) = \left[\mathbb{I} + \mathcal{O}(z^{-1}) \right] \begin{pmatrix} 1 & 0 \\ 0 & B^{-1}(z)A^{-1}(z) \end{pmatrix}, \quad z \rightarrow \infty.$$

Proof. Again, the proof of this proposition follows immediately from the definition of $\mathbf{S}(z)$. The fact that the stronger condition

$$\mathbf{S}(z) = \left[\mathbb{I} + \mathcal{O}(z^{-1}) \right] \begin{pmatrix} 1 & 0 \\ 0 & B^{-1}(z)A^{-1}(z) \end{pmatrix}, \quad z \rightarrow \infty,$$

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 4.7. Thus, $\mathcal{O}(z^{-1/3})$ can be replaced with $\mathcal{O}(z^{-1})$ in the asymptotics of $\mathbf{S}(z)$. ■

6 Construction of parametrices and the transformation $\mathbf{S} \mapsto \mathbf{R}$

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 is really the first place we will see a difference between the multicritical, critical, and generic (non-critical) cases.

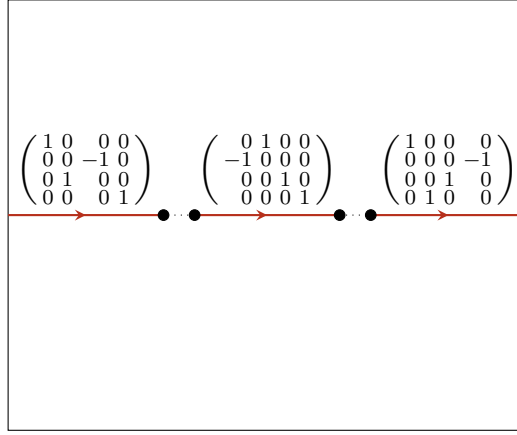


Figure 13. The jumps of the global parametrix $M(z)$ on \mathbb{R} . Note that, when $a, b \rightarrow 1$, branch points merge, and the cuts cover the whole real line.

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 $M(z)$:

$$\left\{ \begin{array}{ll} M \text{ is analytic in } \mathbb{C} \setminus \mathbb{R}, \\ M_+(z) = M_-(z) \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, & z \in (-\infty, -\beta], \\ M_+(z) = M_-(z) \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, & z \in [-\alpha, \alpha], \\ M_+(z) = M_-(z) \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, & z \in [\beta, \infty), \\ M(z) = [\mathbb{I} + \mathcal{O}(z^{-1})] \begin{pmatrix} 1 & 0 \\ 0 & B^{-1}(z)A^{-1}(z) \end{pmatrix}, & z \rightarrow \infty. \end{array} \right. \quad (6.1)$$

In general, solutions to (6.1) will not be unique. Uniqueness can be guaranteed by imposing additionally that

$$\begin{aligned} M(z) &= \mathcal{O}((z \mp \alpha)^{-1/4}), & z \rightarrow \pm\alpha, \\ M(z) &= \mathcal{O}((z \mp \beta)^{-1/4}), & z \rightarrow \pm\beta. \end{aligned} \quad (6.2)$$

Then, from the usual Liouville argument, it is clear that if a solution to (6.1) (with the constraint (6.2) imposed) exists, it is unique. We show that a solution exists by direct construction.

Proposition 6.1. *The solution to the global parametrix is given by*

$$M_{jk}(z) = \begin{cases} \psi_j(u_k(z)), & \text{Im } z > 0, \\ \psi_j(u_\ell(z))S_{\ell k}, & \text{Im } z < 0, \end{cases}$$

(here, summation over ℓ is implied) 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

$$\begin{aligned} \psi_1(u) &= \frac{u^2}{\sqrt{(u^2 - b^2)(u^2 - a^2)}}, & \psi_2(u) &= \frac{3^{2/3}A^{1/3}}{a^{1/3}b^{1/3}} \frac{2a^2b^2 - 3(a^2 + b^2)u^2}{18u\sqrt{(u^2 - b^2)(u^2 - a^2)}} \\ \psi_3(u) &= \frac{ab}{3\sqrt{(u^2 - b^2)(u^2 - a^2)}}, & \psi_4(u) &= \frac{a^{1/3}b^{1/3}}{3^{2/3}A^{1/3}} \frac{u}{\sqrt{(u^2 - b^2)(u^2 - a^2)}}, \end{aligned}$$

and $S = \text{diag}(1, -1, 1, 1) = [\sigma_3 \oplus \mathbb{I}_2]$.

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 $z = z(u)$:

$$z(u) = A(a, b) \left(u + \frac{a^2 + b^2}{u} - \frac{a^2 b^2}{3u^3} \right).$$

Consider the general row vector

$$\vec{\psi}(z) = \begin{cases} [f^u(u_1(z)), g^u(u_2(z)), h^u(u_3(z)), k^u(u_4(z))], & \text{Im } z > 0, \\ [f^l(u_1(z)), g^l(u_2(z)), h^l(u_3(z)), k^l(u_4(z))], & \text{Im } z < 0. \end{cases}$$

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

$$[f^u(u_1(z)), g^u(u_3(z)), h^u(u_2(z)), k^u(u_4(z))] = [f^l(u_1(z)), h^l(u_3(z)), -g^l(u_2(z)), k^l(u_4(z))],$$

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 yields the further compatibility conditions $g^u = k^l$, $k^u = -g^l$, $f^u = -g^l$, $g^u = f^l$. Thus, the global parametrix depends on only on the unknown function: $f^u =: \psi$. Therefore,

$$\vec{\psi}(z) = \begin{cases} [\psi(u_1(z)), \psi(u_2(z)), \psi(u_3(z)), \psi(u_4(z))], & \text{Im } z > 0, \\ [\psi(u_1(z)), -\psi(u_2(z)), \psi(u_3(z)), \psi(u_4(z))], & \text{Im } z < 0. \end{cases}$$

We have thus shown that, in the upper half plane,

$$\vec{\psi}(z) = [\psi(u_1(z)), \psi(u_2(z)), \psi(u_3(z)), \psi(u_4(z))],$$

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 (6.1). 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

$$\psi(u) = \frac{p_3(u)}{u\sqrt{(u^2 - a^2)(u^2 - b^2)}},$$

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:

$$\begin{aligned} \psi_+(u) &= -\psi_-(u), & u \in \gamma, \\ \psi(u) &= \mathcal{O}(1), & u \rightarrow \infty, \\ \psi(u) &= \mathcal{O}(u^{-1}), & u \rightarrow 0, \\ \psi(u) &= \mathcal{O}((u \mp a)^{-1/2}), & u \rightarrow \pm a, \\ \psi(u) &= \mathcal{O}((u \mp b)^{-1/2}), & u \rightarrow \pm b. \end{aligned}$$

These conditions guarantee that the entries of $M(z)$ are $\mathcal{O}(1)$ as $z \rightarrow \infty$ 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 $p_3(u) = c_0 + c_1u + c_2u^2 + c_3u^3$, then we have the following large z expansions of the functions $\psi(u_j(z))$:

$$\begin{aligned} \psi(u_1(z)) &= c_3 + \frac{c_2A}{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}}} \\ \quad + \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}), & \text{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}}} \\ \quad + \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}), & \text{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}}} \\ &\quad + \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}}} \\ \quad + \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}), & \text{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}}} \\ \quad + \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}), & \text{Im } z < 0, \end{cases} \end{aligned}$$

Since $\begin{pmatrix} 1 & 0 \\ 0 & A(z)B(z) \end{pmatrix} M(z)$ has no jumps near infinity, it admits the expansion

$$\begin{pmatrix} 1 & 0 \\ 0 & A(z)B(z) \end{pmatrix} M(z) = \mathbb{I} + \mathcal{O}(z^{-1}), z \rightarrow \infty.$$

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. \blacksquare

Remark 6.2. So far, we have ignored the multicritical case, when $a = b = c = 1$. In this case, the branch points in both the spectral 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 $\psi(u) = [\psi(u_1(z)), \dots, \psi(u_4(z))]$ 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 $a, b, c \rightarrow 1$. Direct inspection of the general global parametrix shows us that $(a, b, c) = (1, 1, 1)$ is the only point in R where the rows of the global parametrix are single-valued in the uniformizing plane. The same is true when we approach a point on the curves γ_b ; this is beyond the scope of the present work.

The next proposition will become useful when we study the local parametrix problem.

Proposition 6.3. *$M(z)$ satisfies the symmetry condition*

$$M(-z) = [\sigma_3 \oplus \sigma_3] M(z) [\sigma_3 \oplus \sigma_1]. \quad (6.3)$$

Proof. First, note that the previous proposition yields that

$$\psi_j(-u) = -\psi_j(u), \quad j \bmod 2 \equiv 0, \quad \psi_j(-u) = \psi_j(u), \quad j \bmod 2 \equiv 1.$$

Furthermore, the uniformizing coordinates $u_j(z)$ satisfy the following relations (cf. Appendix A for details):

$$\begin{aligned} u_1(-z) &= -u_1(z), & u_2(-z) &= -u_2(z), \\ u_3(-z) &= -u_4(z), & u_4(-z) &= -u_3(z). \end{aligned}$$

Thus, we can compute entrywise the transformation $M(-z)$. For example, if z belongs to the upper half plane, then $-z$ is in the lower half plane, and we obtain that

$$\begin{aligned} M_{11}(-z) &= \psi_1(u_1(-z)) = \psi_1(-u_1(z)) = \psi_1(u_1(z)), \\ M_{22}(-z) &= -\psi_2(u_2(-z)) = -\psi_2(-u_2(z)) = \psi_2(u_2(z)), \\ M_{33}(-z) &= \psi_3(u_3(-z)) = \psi_3(-u_4(z)) = \psi_4(u_4(z)). \end{aligned}$$

Note that $\text{Im } z > 0$ became relevant in the calculation for the 2-2 entry. Calculation of the remaining entries yields equation (6.3). \blacksquare

Remark 6.4. Note here that the right multiplication $[\sigma_3 \oplus \sigma_1] = [\sigma_3 \oplus \mathbb{I}_2][\mathbb{I}_2 \oplus \sigma_1]$ has two purposes: the block $[\mathbb{I}_2 \oplus \sigma_1] = \hat{\sigma}_{34}$ interchanges of columns 3 and 4, and the block $[\sigma_3 \oplus \mathbb{I}_2] = S$ accounts for the change in sign of the global parametrix upon moving from the upper to the lower half plane (and vice versa).

Remark 6.5. 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

$$\delta\Omega_{ij}(z) := \Omega_i(z) - \Omega_j(z),$$

for $i, j \in \{1, \dots, 4\}$.

We have now found an approximate solution to the Riemann–Hilbert problem for $\mathbf{S}(z)$. Indeed, if we consider the matrix

$$\mathbf{R}_{\text{out}}(z) := \mathbf{S}(z)M^{-1}(z),$$

we see that $\mathbf{R}_{\text{out}}(z) \rightarrow \mathbb{I}$ as $z \rightarrow \infty$. Furthermore, the jumps of $\mathbf{R}_{\text{out}}(z)$ 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 $n \rightarrow \infty$. Therefore, we must try to find a better approximation of $\mathbf{S}(z)$ near the branch points.

Remark 6.6. Here is the first place in the Riemann–Hilbert analysis where we will see explicitly that the generic, critical, and multicritical cases differ. In this work, we are interested only in the generic situation. In Part III, we will address the multicritical situation; the other critical cases we hope to study in a later work. We record here how the local parametrices for each of these cases should work out. One should compare this to the remarks in Section 2.4.

- (1) *The generic case: Airy parametrices.* For (τ, t, H) off the critical surface, the situation is generic, and the behavior of the $\delta\Omega_{ij}(z)$'s near the branch points is

$$\delta\Omega_{ij}(z) \sim (z \pm \alpha)^{3/2} \quad (\text{resp. } \sim (z \pm \beta)^{3/2}),$$

cf. Proposition 3.5. 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, we omit the explicit calculation of the all of the parametrices here. We present the calculation of the local parametrix at $z = -\beta$ in the next subsection.

- (2) $(\tau, t, H) \in \gamma_{\text{low},0}$: *Painlevé I.* For (τ, t, H) on the low-temperature critical curve (i.e., the curve defined by the equation $t = -\frac{1}{12} + \frac{2}{9}\tau^2$, $0 < \tau < \frac{1}{4}$, $H = 0$), the behavior of the $\delta\Omega_{ij}(z)$'s near each of the four branch points $\pm\alpha, \pm\beta$, is

$$\delta\Omega_{ij}(z) \sim (z \pm \alpha)^{5/2} \quad (\text{resp. } \sim (z \pm \beta)^{5/2}).$$

It is also widely acknowledged in the literature (cf., for example, [38]) that such local behavior leads to a Painlevé I-type Riemann–Hilbert problem, which has been the subject of intensive study [54, 55]. 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, 43, 44]. Since we are mainly interested in the behavior of the partition function in the generic case and at the multicritical point, we omit the explicit calculation of the parametrices here.

- (3) $(\tau, t, H) \in \gamma_{\text{high},0} \cup S_{\text{high}}$: *Painlevé I and Airy.* For (τ, t, H) on the high-temperature critical surface, the behavior of the $\delta\Omega_{ij}(z)$'s the branch points is

$$\delta\Omega_{ij}(z) \sim (z \pm \alpha)^{5/2}, \quad \delta\Omega_{ij}(z) \sim (z \pm \beta)^{3/2},$$

as $z \rightarrow \pm\alpha$, $z \rightarrow \pm\beta$, respectively. 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.

- (4) $(\tau, t, H) \in S_{\text{low}}$: *Airy and Painlevé I*. Similarly to the high-temperature surface, on the low-temperature surface both Airy and Painlevé I-type parametrices appear. In this case, the behavior of the $\delta\Omega_{ij}(z)$ ’s the branch points is

$$\delta\Omega_{ij}(z) \sim (z \pm \alpha)^{3/2}, \quad \delta\Omega_{ij}(z) \sim (z \pm \beta)^{5/2}.$$

Thus, one must construct Airy-type parametrices at $z = \pm\alpha$, and Painlevé I parametrices near $z = \pm\beta$.

- (5) $(\tau, t, H) \in \gamma_b$: *New Painlevé I parametrix*. On the critical curve γ_b (the interface of the surfaces S_{low} and S_{high}), the branch points at $z = +\alpha, +\beta$ (respectively, $z = -\alpha, -\beta$) merge, and one obtains a pair of cubic branch points $z = \pm\alpha$, at which the $\delta\Omega_{ij}(z)$ ’s behave as

$$\delta\Omega_{ij}(z) \sim (z \pm \alpha)^{5/3}.$$

To the best of our knowledge, an appropriate model Riemann–Hilbert problem for such a parametrix has not yet appeared in the literature. Some preliminary calculations (see also Part II [49]) show that the “right” parametrix is a 3×3 RHP for a function $\Psi(\xi; x)$, which has asymptotics

$$\Psi(\xi; x) = g(\xi) \left[\mathbb{I} + \frac{\Psi_1(x)}{\xi^{1/3}} + \mathcal{O}\left(\frac{1}{\xi^{2/3}}\right) \right] e^{\Theta(\xi; x)},$$

where $g(\xi) = \mathcal{O}(\xi^{1/3})$, $\Theta(\xi; x)$ is a diagonal matrix whose entries are analytic continuations of the function $\theta_1(\xi; x) = \frac{3}{5}\xi^{5/3} + x\xi^{1/3}$, and the jumps of Ψ appearing only from the Stokes phenomenon. Some calculation reveals that this model Riemann–Hilbert problem is related to the Painlevé I equation, in that $\frac{d}{dx}[\Psi_1(x)]_{11} = u(x)$, and $u(x)$ solves a scaled version of the Painlevé I equation. The associated Lax pair *has* appeared before in the literature [52]; one might hope to relate this parametrix to the “standard” 2×2 parametrix for Painlevé I via the techniques of Liechty and Wang [64], who studied a similar relation between two instances of Painlevé II parametrices. Both seem to be instances of the so-called $(p, q) \leftrightarrow (q, p)$ duality of string equations [45, 46] (for PII, $q = 2$, $p = 4$, and for PI, $q = 2$, $p = 3$). It would be interesting to investigate this correspondence in more detail. For now, we leave this problem open.

- (6) $(\tau, t, H) = (\frac{1}{4}, -\frac{5}{72}, 0)$: *A new critical phenomenon*. Finally, if we are at the multicritical point, we again find that the branch points at $z = +\alpha, +\beta$ (respectively, $z = -\alpha, -\beta$) merge, and one obtains a pair of cubic branch points $z = \pm\alpha$, at which the $\delta\Omega_{ij}(z)$ ’s behave as

$$\delta\Omega_{ij}(z) \sim (z \pm \alpha)^{7/3}.$$

This requires the construction of a new 3×3 parametrix, which we discuss fully in Part II [49]. This parametrix is connected to the $(3, 4)$ string equation [33], and is meant to describe the Ising phase transition coupled to gravity, cf. [21, 34, 47, 48]. We fully investigate how the multicritical partition function of the 2-matrix model studied in this work is connected to this equation in Part III of this work [37].

6.2 Calculation of the local parametrices

We now construct the local parametrices. Define four discs

$$D_{\pm\alpha} := \{z \in \mathbb{C} \mid |z \mp \alpha| < \epsilon\}, \quad D_{\pm\beta} := \{z \in \mathbb{C} \mid |z \mp \beta| < \epsilon\},$$

where $\epsilon > 0$ is a radius, to be determined. Our new parametrix $M^\infty(z)$ will be defined as

$$M^\infty(z) = \begin{cases} M(z), & z \in \mathbb{C} \setminus (D_{\pm\alpha} \cup D_{\pm\beta}), \\ P^{(\pm\alpha)}(z), & z \in D_{\pm\alpha}, \\ P^{(\pm\beta)}(z), & z \in D_{\pm\beta}. \end{cases}$$

The functions $P^{(\pm\alpha)}(z)$ will be chosen so that the following conditions are met:

- (1) $P^{(\pm\alpha)}(z)$ matches the jumps of $\mathbf{S}(z)$ *exactly* in the discs $D_{\pm\alpha}$,
- (2) $P^{(\pm\alpha)}(z) = [\mathbb{I} + \mathcal{O}(n^{-\delta})]M(z)$, for $z \in D_{\pm\alpha}$, $n \rightarrow \infty$ for some $\delta > 0$.

The functions $P^{(\pm\beta)}(z)$ should be chosen so that identical statements to the above hold, with α replaced with β . In the case of generic values of the parameters, one should expect that near the branch points $z = \pm\alpha, \pm\beta$, an Airy-type parametrix can be used to describe the local behavior of the Riemann–Hilbert problem at hand. This is a common calculation in the literature (cf., for example, [30]). Before calculating the parametrices, it is useful to notice that we only have to calculate two of the parametrices; the others can be calculated by symmetry. This statement is summarized in the following lemma.

Lemma 6.7. *Suppose $P^{(+\beta)}(z)$ satisfies conditions (1) and (2) above. Then, $P^{(-\beta)}(z)$ can be constructed as*

$$P^{(-\beta)}(z) = [\sigma_3 \oplus \sigma_3]P^{(+\beta)}(-z)[\sigma_3 \oplus \sigma_1].$$

Proof. Let us verify condition (2) first. Indeed, if z belongs to a sufficiently small neighborhood of $z = -\beta$, by Proposition 6.3, we have that

$$\begin{aligned} [\sigma_3 \oplus \sigma_3]P^{(+\beta)}(-z)[\sigma_3 \oplus \sigma_1] &= [\sigma_3 \oplus \sigma_3][\mathbb{I} + \mathcal{O}(n^{-\delta})]M(-z)[\sigma_3 \oplus \sigma_1] \\ &= [\mathbb{I} + \mathcal{O}(n^{-\delta})]M(z), \end{aligned}$$

where $M(z)$ above denotes the local expansion of $M(z)$ about $z = -\beta$. Now, let us check condition (1) By direct calculation,

$$\begin{aligned} P_+^{(-\beta)}(z) &= [\sigma_3 \oplus \sigma_3]P_+^{(+\beta)}(-z)[\sigma_3 \oplus \sigma_1] \\ &= [\sigma_3 \oplus \sigma_3]P_-^{(+\beta)}(-z)J_P(-z)[\sigma_3 \oplus \sigma_1] \\ &= P_-^{(-\beta)}(z)[\sigma_3 \oplus \sigma_1]J_P(-z)[\sigma_3 \oplus \sigma_1], \end{aligned}$$

where $J_P(z)$ denotes the jumps of $P^{(+\beta)}(z)$. We must verify that $[\sigma_3 \oplus \sigma_1]J_P(-z)[\sigma_3 \oplus \sigma_1]$ agrees with the jumps of $\mathbf{S}(z)$ in a neighborhood of $z = -\beta$; this is done by direct calculation. For example, for $z \in (-\infty, -\beta] \cap D_{-\alpha}$, $-z \in [\beta, \infty) \cap D_{+\alpha}$. Furthermore, we have that

$$[\sigma_3 \oplus \sigma_1]J_P(-z)[\sigma_3 \oplus \sigma_1] = [\sigma_3 \oplus \sigma_1] \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} [\sigma_3 \oplus \sigma_1] = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

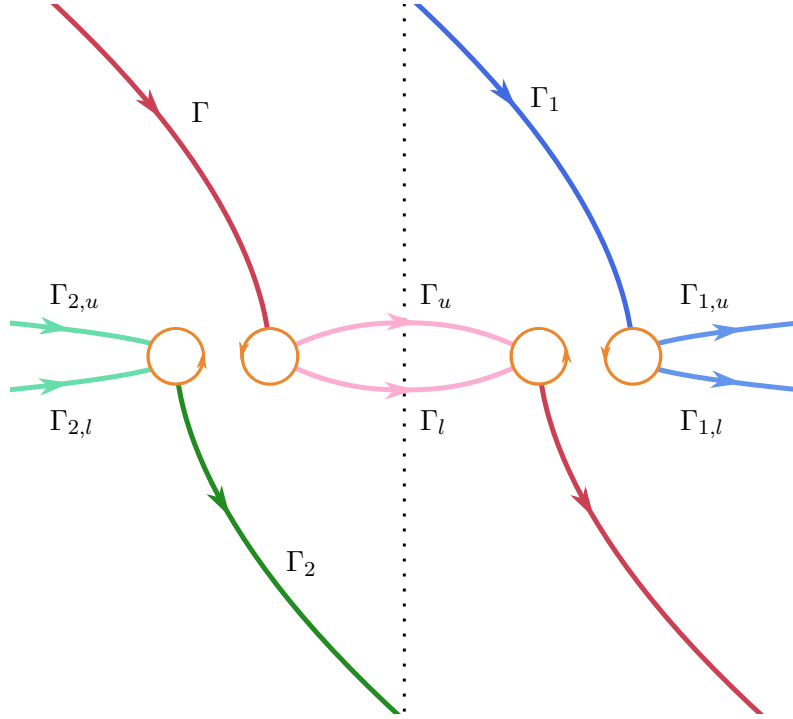


Figure 14. The opened lenses after the final transformation.

which indeed agrees with the jump of $\mathbf{S}(z)$ there, upon reorientation of the jump contour. Similarly, if $z \in \Gamma_2 \cap \{\operatorname{Im} z < 0\} \cap D_-$, then $-z \in \Gamma_1 \cap \{\operatorname{Im} z > 0\} \cap D_+$, and

$$\begin{aligned} [\sigma_3 \oplus \sigma_1] J_P(-z) [\sigma_3 \oplus \sigma_1] &= [\sigma_3 \oplus \sigma_1] [\mathbb{I} - E_{24} e^{-n\tau\delta\Omega_{24}(-z)}] [\sigma_3 \oplus \sigma_1] \\ &= \mathbb{I} + E_{23} e^{-n\tau\delta\Omega_{24}(-z)} = \mathbb{I} + E_{23} e^{-n\tau\delta\Omega_{23}(z)}, \end{aligned}$$

by the symmetry of the functions $\Omega_j(z)$ (and again reorientation of the jump contour). The rest of the calculations are similar, so we omit them. \blacksquare

We state without proof the same lemma for the parametrices at $z = \pm\alpha$.

Lemma 6.8. *Suppose $P^{(+\alpha)}(z)$ satisfies conditions (1) and (2) above. Then, $P^{(-\alpha)}(z)$ can be constructed as $P^{(-\alpha)}(z) = [\sigma_3 \oplus \sigma_3] P^{(+\alpha)}(-z) [\sigma_3 \oplus \sigma_1]$.*

Thus, it is sufficient to calculate the parametrix at $z = \alpha$, $z = \beta$ only, and use the above symmetry relations as the definition of $P^{(-\alpha)}(z)$, $P^{(-\beta)}(z)$.

For brevity, we only compute explicitly the parametrix at $z = -\beta$. The rest of the local parametrices are calculated in an identical manner.

Let $\mathcal{A}(\zeta)$ denote the unique solution to the following 2×2 Riemann–Hilbert problem:

$$\mathcal{A}_+(\zeta) = \mathcal{A}_-(\zeta) \times \begin{cases} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, & \arg \zeta = 0, \\ \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, & \arg \zeta = \pm 2\pi/3, \\ \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, & \arg \zeta = \pi. \end{cases}$$

$$\mathcal{A}(\zeta) = [\mathbb{I} + \mathcal{O}(\zeta^{-1})] \zeta^{-\frac{1}{4}\sigma_3} U e^{-\frac{2}{3}\zeta^{3/2}\sigma_3}, \quad \zeta \rightarrow \infty, \quad (6.4)$$

where here $U = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix}$. The solution to the RHP (6.4) in the sector $0 < \arg \zeta < 2\pi/3$ is given by

$$\mathcal{A}(\zeta) = \sqrt{2\pi} \begin{pmatrix} \text{Ai}(\zeta) & -\omega^2 \text{Ai}(\omega^2 \zeta) \\ -i \text{Ai}'(\zeta) & i\omega \text{Ai}'(\omega^2 \zeta) \end{pmatrix}.$$

Here $\text{Ai}(\zeta)$ denotes the usual Airy function; the solution to the RHP (6.4) in the other sectors can be obtained by analytic continuation, and application of the jump condition.

Define the conformal mapping

$$\zeta(z) := \left[\frac{3}{4} n \tau \delta \Omega_{23}(z) \right]^{2/3}.$$

Note that $\zeta(z)$ is univalent in a neighborhood of $z = -\beta$, since $\delta \Omega_{23}(z) = \text{const} \cdot (z + \beta)^{3/2} [1 + \mathcal{O}(z + \beta)]$. Finally, set

$$\hat{P}(z) := \mathcal{A}(\zeta(z)) e^{\frac{2}{3} \zeta(z)^{3/2} \sigma_3} \sigma_3.$$

We redefine the jump contours of $\mathcal{A}(\zeta)$ so that the jump contours of $\hat{P}(z)$ align with those of $\mathbf{S}(z)$, in a neighborhood of $z = -\beta$. The following lemma then holds.

Lemma 6.9. *The jumps of $1 \oplus \hat{P}(z) \oplus 1$ and $\mathbf{S}(z)$ agree, in a neighborhood of $z = -\beta$.*

Proof. This is an immediate consequence of the fact that multiplication of the matrix $\mathcal{A}(\zeta(z))$ by $\mathbf{p} := e^{\frac{2}{3} \zeta(z)^{3/2} \sigma_3} \sigma_3$ conjugates its jumps by this factor; in other words, $J_{\hat{P}}(z) = \mathbf{p}^{-1} J_{\mathcal{A}}(z) \mathbf{p}$. ■

Now, consider 2×2 matrix-valued function

$$\tilde{\mathbf{m}}(z) := (z + \beta)^{-\frac{1}{4} \sigma_3} U \sigma_3.$$

It follows immediately that the jump of $1 \oplus \tilde{\mathbf{m}} \oplus 1$ and the jump of the global parametrix $M(z)$ match in a neighborhood of $z = -\beta$. Therefore, one can locally represent the global parametrix as

$$M(z) = \mathfrak{H}(z) [1 \oplus \tilde{\mathbf{m}}(z) \oplus 1],$$

for some locally holomorphic matrix function $\mathfrak{H}(z)$. Furthermore, since $\zeta(z) = n^{2/3} (z + \beta) g(z)$, for some non-vanishing analytic function $g(z)$ in a neighborhood of $z = -\beta$, by a slight redefinition of $\mathfrak{H}(z)$, we can rewrite $M(z)$ as

$$M(z) = \tilde{\mathfrak{H}}(z) [1 \oplus n^{\frac{1}{6} \sigma_3} [\zeta(z)]^{-\frac{1}{4} \sigma_3} U \sigma_3 \oplus 1],$$

where $\tilde{\mathfrak{H}}(z)$ is some n -independent invertible function, analytic in a neighborhood of $z = -\beta$.

We now define the parametrix in a sufficiently small disc about $z = -\beta$ as

$$P^{(-\beta)}(z) := M(z) [1 \oplus \sigma_3 U^{-1} [\zeta(z)]^{\frac{1}{4} \sigma_3} \oplus 1] \hat{P}(z).$$

We remark that $P^{(-\beta)}(z)$ and $\hat{P}(z)$ have identical jumps inside the disc at $z = -\beta$, since

$$M(z) [1 \oplus \sigma_3 U^{-1} [\zeta(z)]^{\frac{1}{4} \sigma_3} \oplus 1]$$

has no jumps there, by our previous considerations. The parametrix at $z = +\beta$ can be constructed using our previous symmetry considerations, cf. Lemma 6.7. One can similarly construct parametrices at $P^{(\pm\alpha)}(z)$, by replacing in the definition of $\hat{P}(z)$: (1) the conformal map $\zeta(z)$ with $\zeta(z) := [\frac{3}{4} n \tau \delta \Omega_{24}(z)]^{2/3}$, (2) conjugating by $\hat{\sigma}_{34}$, and (3) readjusting the jump contours, if

needed. Such a construction is identical to that of $P_{-\beta}(z)$, and so we omit the explicit calculation.

Finally, set

$$\mathbf{R}(z) = \begin{cases} \mathbf{S}(z)M^{-1}(z), & z \text{ outside the discs at } z = \pm\alpha, z = \pm\beta, \\ \mathbf{S}(z)(P^{(\pm\alpha)})^{-1}(z), & z \text{ inside the discs at } z = \pm\alpha, \\ \mathbf{S}(z)(P^{(\pm\beta)})^{-1}(z), & z \text{ inside the discs at } z = \pm\beta. \end{cases}$$

We have the following proposition.

Proposition 6.10. *On the boundary of the discs at $z = \pm\alpha, \pm\beta$,*

$$\mathbf{R}_+(z) = \mathbf{R}_-(z)[\mathbb{I} + \mathcal{O}(n^{-1})], \quad n \rightarrow \infty.$$

Proof. We furnish a proof only on the boundary of the disc $z = -\beta$. Let us assume the boundary of the disc at $z = -\beta$ is oriented counterclockwise. Then,

$$\begin{aligned} \mathbf{R}_+(z) &= \mathbf{S}(z)(P^{(-\beta)})^{-1}(z) = \mathbf{S}(z)M^{-1}(z)M(z)(P^{(-\beta)})^{-1}(z) \\ &= \mathbf{R}_-(z)M(z)(P^{(-\beta)})^{-1}(z). \end{aligned}$$

Now, we have that

$$\begin{aligned} &M(z)(P^{(-\beta)})^{-1}(z) \\ &= M(z)[1 \oplus \hat{P}^{-1}(z) \oplus 1][1 \oplus [\zeta(z)]^{-\frac{1}{4}\sigma_3}U \oplus 1]M^{-1}(z) \\ &= M(z) \underbrace{[1 \oplus \sigma_3 e^{-\frac{2}{3}\zeta(z)^{3/2}}\mathcal{A}^{-1}(\zeta(z)) \oplus 1]}_{1 \oplus \hat{P}^{-1}(z) \oplus 1} [1 \oplus [\zeta(z)]^{-\frac{1}{4}\sigma_3}U \sigma_3 \oplus 1] \\ &\quad \times \underbrace{[1 \oplus \sigma_3 U^{-1}[\zeta(z)]^{\frac{1}{4}\sigma_3} n^{-\frac{1}{6}\sigma_3} \oplus 1]}_{M^{-1}(z)} \tilde{\mathfrak{H}}^{-1}(z) \\ &= \underbrace{\tilde{\mathfrak{H}}(z)[1 \oplus n^{\frac{1}{6}\sigma_3}[\zeta(z)]^{-\frac{1}{4}\sigma_3}U \sigma_3 \oplus 1]}_{M(z)} [1 \oplus \sigma_3 e^{-\frac{2}{3}\zeta(z)^{3/2}}\mathcal{A}^{-1}(\zeta(z)) \oplus 1] \\ &\quad \times [1 \oplus n^{-\frac{1}{6}\sigma_3} \oplus 1] \tilde{\mathfrak{H}}^{-1}(z) \\ &= \tilde{\mathfrak{H}}(z)[1 \oplus n^{\frac{1}{6}\sigma_3}[\zeta(z)]^{-\frac{1}{4}\sigma_3}U \sigma_3 \oplus 1][1 \oplus \sigma_3 U^{-1}[\zeta(z)]^{\frac{1}{4}\sigma_3}[\mathbb{I}_{2 \times 2} + \mathcal{O}(\zeta^{-1})] \oplus 1] \\ &\quad \times [1 \oplus n^{-\frac{1}{6}\sigma_3} \oplus 1] \tilde{\mathfrak{H}}^{-1}(z) \\ &= \tilde{\mathfrak{H}}(z)[1 \oplus n^{\frac{1}{6}\sigma_3} \oplus 1][1 \oplus [\mathbb{I}_{2 \times 2} + \mathcal{O}(\zeta^{-1})] \oplus 1][1 \oplus n^{-\frac{1}{6}\sigma_3} \oplus 1] \tilde{\mathfrak{H}}^{-1}(z). \end{aligned}$$

Since $\mathcal{O}(\zeta^{-1}) = \mathcal{O}(n^{-2/3})$ in the z -plane, by crude estimation we see that the above jump is at worst

$$[1 \oplus n^{\frac{1}{6}\sigma_3} \oplus 1][1 \oplus [\mathbb{I}_{2 \times 2} + \mathcal{O}(\zeta^{-1})] \oplus 1][1 \oplus n^{-\frac{1}{6}\sigma_3} \oplus 1] = \mathbb{I} + \mathcal{O}(n^{-1/3}), \quad n \rightarrow \infty$$

(conjugation by an n -independent holomorphic matrix does not change the error). In fact, as is typical for the Airy parametrix, a more detailed expansion of the above product reveals that cancellations occur, and one finds that indeed

$$\mathbf{R}_-(z)^{-1} \mathbf{R}_+(z) = M(z)(P^{(-\beta)})^{-1}(z) = \mathbb{I} + \mathcal{O}(n^{-1}), \quad n \rightarrow \infty.$$

Thus, we have proven what is desired. ■

Aside from the jumps on the local discs, all other jumps of $\mathbf{R}(z)$ are exponentially close to the identity matrix as $n \rightarrow \infty$, as we have already seen; furthermore, this bound improves as $z \rightarrow \infty$. If we let $\Gamma_{\mathbf{R}}$ denote the union of all of the jump contours of $\mathbf{R}(z)$, and $J_{\mathbf{R}}(z): \Gamma_{\mathbf{R}} \rightarrow \mathbb{C}$ denote the jumps of $\mathbf{R}(z)$, then we have that

$$J_{\mathbf{R}}(z) = \mathbb{I} + \mathcal{O}(e^{-cn(|z|+1)}), \quad n \rightarrow \infty,$$

uniformly for $z \in \Gamma_{\mathbf{R}} \setminus (D_{\pm\alpha} \cup D_{\pm\beta})$, for some positive constant $c > 0$. Using similar arguments as in [39, 40], we can conclude the following.

Proposition 6.11. *There exists a constant $C > 0$ such that, for every n , and uniformly for $z \in \mathbb{C} \setminus \Gamma_{\mathbf{R}}$,*

$$\|\mathbf{R}(z) - \mathbb{I}\| \leq \frac{C}{n(1 + |z|)}.$$

This allows us to conclude that $\mathbf{R}(z)$ admits a large n asymptotic expansion in powers of n^{-1} , cf. [30, Theorem 7.10]. Since all of the Riemann–Hilbert transformations we made were invertible, by tracing back these transformations we can obtain a large n asymptotic expansion of $\mathbf{Y}(z)$. This concludes the steepest descent analysis.

7 Proof of main theorem and concluding remarks

7.1 Calculation of the free energy

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 τ function, as per [9]. Explicitly, the τ -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 τ -differential is

$$\begin{aligned} d \log \tau_n(\tau, t) = & \left\langle \mathbf{Y}^{-1} \mathbf{Y}' \begin{pmatrix} 0 & 0 \\ 0 & \frac{\partial \mathbf{W}}{\partial \tau} \mathbf{W}^{-1} \end{pmatrix} \right\rangle d\tau + \left\langle \mathbf{Y}^{-1} \mathbf{Y}' \begin{pmatrix} 0 & 0 \\ 0 & \frac{\partial \mathbf{W}}{\partial T} \mathbf{W}^{-1} \end{pmatrix} \right\rangle dT \\ & + \left\langle \mathbf{Y}^{-1} \mathbf{Y}' \begin{pmatrix} 0 & 0 \\ 0 & \frac{\partial \mathbf{W}}{\partial \bar{T}} \mathbf{W}^{-1} \end{pmatrix} \right\rangle d\bar{T}, \end{aligned} \quad (7.1)$$

where $'$ denotes the derivative with respect to the spectral variable z , \mathbf{W} is the matrix which appears in the first transformation, $T := qt$, $\bar{T} = t/q$, and for a given matrix-valued 1-form $A(z)$, we have introduced the notation

$$\langle A(z) \rangle := \operatorname{Res}_{z=\infty} \operatorname{tr} [A(z)].$$

The partition function's differential is then

$$d \frac{\log Z_n(\tau, t, q; N)}{\tau^{\frac{-n(n-1)}{2}} C_{n,N}} = d \log \left[\left(\frac{\tau}{t^2} \right)^{\frac{n}{2}(\frac{n}{3}-1)} \tau_n \right],$$

and the free energy can be written as

$$d \log \frac{Z_n(\tau, t, q; N)}{Z_n(\tau, 0, 0; N)} = d \log \left[\frac{(1 - \tau^2)^{n^2/2}}{\tau^{n^2/3} t^{n(\frac{n}{3}-1)}} \tau_n \right].$$

The derivations of the above formulae for the τ -function and partition function are sketched in Appendix B. Combined with the asymptotic formulae of the previous sections, these formulae allow us to finally prove the main theorem. Let us recall the main theorem (this is identical to Theorem 2.2).

Theorem. *Let (τ, t, H) belong to the region \mathcal{D} . Then, as $n \rightarrow \infty$,*

$$\begin{aligned} F(\tau, t, H) &:= \lim_{n \rightarrow \infty} \frac{1}{n^2} \log \frac{Z_n(\tau, t, H; n)}{Z_n(\tau, 0, 0; n)} \\ &= \frac{3}{4} + \frac{1}{2} \log \frac{(1 - \tau^2)\sigma(\tau, t, H)}{-3t} - \int_0^{\sigma(\tau, t, H)} \left(\lambda(u) - \frac{1}{2} \lambda(u)^2 \right) \frac{du}{u}, \end{aligned}$$

where $\lambda(u)$ is the rational function

$$\lambda(u) = -\frac{1}{t} \left[\frac{1}{9} \tau^2 u(u^2 - 3) + \frac{1}{3} \frac{u}{(u+1)^2} - \frac{2}{3} \left(\frac{u}{u^2 - 1} \right)^2 [\cosh H - 1] \right],$$

and $\sigma(\tau, t, H)$ from Definition 2.1.

Proof. Let $\hat{\sigma} := \text{diag}(1, -1/3, -1/3, -1/3)$, and set

$$\mathbf{Y}^{(1)} := \lim_{z \rightarrow \infty} z [\mathbf{Y}(z) z^{-n\hat{\sigma}} - \mathbb{I}], \quad \mathbf{Y}^{(2)} := \lim_{z \rightarrow \infty} z^2 \left[\mathbf{Y}(z) z^{-n\hat{\sigma}} - \mathbb{I} - \frac{\mathbf{Y}^{(1)}}{z} \right].$$

Using the expression (7.1), and the fact that the matrices \mathcal{M}^\bullet , $\bullet \in \{\tau, t, q\}$, are all polynomials of degree at most 2 in z , we can derive an explicit expression for the tau function in terms of the entries of $\mathbf{Y}(z)$:

$$\begin{aligned} \varrho_\tau &= -\frac{\tau q n}{t} \left[\mathbf{Y}_{12}^{(1)} \mathbf{Y}_{41}^{(1)} + \mathbf{Y}_{22}^{(1)} \mathbf{Y}_{42}^{(1)} + \mathbf{Y}_{32}^{(1)} \mathbf{Y}_{43}^{(1)} + \mathbf{Y}_{42}^{(1)} \mathbf{Y}_{44}^{(1)} - 2 \mathbf{Y}_{42}^{(2)} + \frac{1}{\tau} \mathbf{Y}_{43}^{(1)} \right. \\ &\quad \left. - \frac{t}{\tau q} \mathbf{Y}_{34}^{(1)} - \frac{t}{\tau q} \mathbf{Y}_{23}^{(1)} \right], \\ \varrho_t &= -\frac{\tau}{4t} \varrho_\tau + \frac{n\tau q}{4t^2} \left[\mathbf{Y}_{43}^{(1)} + \mathbf{Y}_{32}^{(1)} - \frac{2n}{3\tau} - \frac{2t}{\tau q} \right], \quad \varrho_q = -\frac{t}{q} \varrho_t. \end{aligned}$$

The above three formulae tell us that we must compute the asymptotics of $\mathbf{Y}(z)$ to two sub-leading orders in z at infinity. In an appropriately chosen sector (chosen so that we are outside of the lenses) at infinity, the asymptotics of $\mathbf{Y}(z)$ we have calculated are given by

$$\mathbf{Y}(z) = \begin{pmatrix} 1 & 0 \\ 0 & c_N^{-1} \hat{K}^{-1} \end{pmatrix} e^{nL} [\mathbb{I} + nC_1 \cdot E_{24}] [\mathbb{I} + \mathcal{O}(n^{-2})] M(z) \mathbf{G}^{-1}(z) \begin{pmatrix} e^{nV(z)} & 0 \\ 0 & \mathbf{W}(z) \end{pmatrix}. \quad (7.2)$$

This implies that we must compute:

- the regular expansion of $M(z)B^{-1}(z)A^{-1}(z)$ at infinity to order z^{-2} ,
- the expansion of the effective potentials $\Omega_j(z)$ to second order, $j = 1, 2, 3, 4$; in particular, we write

$$\begin{aligned} \tau \Omega_3(z) &= -\frac{3}{4} \frac{\tau^{4/3}}{(-t/q)^{1/3}} z^{4/3} - \frac{1}{2} \frac{\tau^{2/3}}{(-t/q)^{2/3}} z^{2/3} + \frac{1}{3} \log z + \ell_1 \\ &\quad + \frac{C_1}{z^{2/3}} + \frac{C_2}{z^{4/3}} + \frac{C_3}{z^2} + \mathcal{O}(z^{-8/3}), \quad z \rightarrow \infty, \end{aligned}$$

with similar expressions for $\Omega_2(z)$, $\Omega_4(z)$,

- the regular part of the expansion of $\mathbf{W}(z)$, up to order z^{-2} .

The explicit expressions for C_1 , C_2 are given as

$$C_1(a, b, c) = \frac{c^{3/2} \tau^{7/3}}{18b^{3/2}(-t)^{4/3} q^{1/6}} (3a^8 b^4 c^2 + 3a^6 b^6 c^2 + 3a^8 b^2 + 3a^6 b^4 + 3a^6 b^2 c^2)$$

$$\begin{aligned}
& + 3a^4b^6 + 3a^4b^4c^2 + 3a^2b^6c^2 - a^6 - 9a^4b^2 - 9a^2b^4 - b^6), \\
C_2(a, b, c) = & -\frac{c^2\tau^{8/3}}{54b^2(-t)^{5/3}q^{1/3}} (6a^{10}b^4c^2 + 9a^8b^6c^2 + 6a^6b^8c^2 + 2a^{10}b^2 - 6a^8b^4 + 2a^8b^2c^2 \\
& - 6a^6b^6 - 6a^6b^4c^2 + 2a^4b^8 - 6a^4b^6c^2 + 2a^2b^8c^2 - a^8 \\
& - 10a^6b^2 - 12a^4b^4 - 10a^2b^6 - b^8).
\end{aligned}$$

With these calculations in hand, one may insert the asymptotic expansions derived above into (7.2), and furthermore insert these expressions into the expressions of ρ_τ , ρ_t , and ρ_q , to obtain

$$\begin{aligned}
\frac{1}{n^2}\varrho_\tau &= \frac{2}{3}\frac{C_1}{\tau^{1/3}}\left(-\frac{q}{t}\right)^{2/3} + 4C_2\left(-\frac{\tau q}{t}\right)^{1/3} + \mathcal{O}(n^{-2}), \\
\frac{1}{n^2}\varrho_t &= -\frac{\tau}{4t}\varrho_\tau + \frac{2}{\tau^{1/3}}C_1 - \frac{q^2}{27t^3} - \frac{2q}{3t^2} + \mathcal{O}(n^{-2}), \\
\frac{1}{n^2}\varrho_q &= -\frac{t}{q}\varrho_t + \mathcal{O}(n^{-2}).
\end{aligned}$$

Now, τ , t , q , C_1 , and C_2 all have explicit expressions in terms of a , b , c , and so we can compute the tau differential in these coordinates. We thus obtain an expression for dF . This expression is exact, and can thus be integrated. Since we know the value of this expression at $(\tau, 0, 0)$ (namely, $F(\tau, 0, 0) \equiv 0$), we obtain an exact expression for the function $F(\tau, t, H)$. The final result is yielded by comparing the result of the previously outlined calculation to the expression (2.2). ■

7.2 Concluding remarks

This work is the first in our series of papers on the Ising transition on the 2-matrix model: the study of the free energy of this model at the multicritical point will be the subject of [37, 49]. These works are perhaps one of the main reasons the 2-matrix model merits study, and will hopefully shed light on minimal matter coupled to gravity. Aside from this work, there are still many other interesting questions we did not pursue in this work which absolutely merit further study. We discuss some of these problems below.

Genus expansion of the free energy. One of the more pressing questions we did not address is the conjectured genus expansion of the free energy [32, 41]. It is conjectured that the following $N \rightarrow \infty$ asymptotic expansion holds, in the genus 0 region of phase space we considered in this work:

$$\frac{1}{N^2} \log \frac{Z_N(\tau, t, q; N)}{Z_N(\tau, 0, 0; N)} \sim F(\tau, t, q) + \sum_{g=1}^{\infty} \frac{F_g(\tau, t, q)}{N^{2g}},$$

where $F_g(\tau, t, q)$ is the analog of (1.5) (see also equation (1.4)), but with the sum taken over all connected, 4-regular, genus g maps. We fully expect that this is the case, and this result should follow from a more careful look at the Riemann–Hilbert analysis performed in this work. Concretely, there have been several works which give formulae for the genus 1 contribution to the free energy $F_1(\tau, t, H)$ [29, 41]. It is of therefore of interest to confirm if our results agree with the findings of these works.

Critical partition function and Painlevé I. One corollary of the famous Yang–Lee theorem [63, 86] is that the partition function for the Ising model should be an analytic function of all of its parameters, away from the segment $H = 0$, $[0, T_c)$, where T_c is the critical temperature. Indeed, this formally seems to be the case for the model studied here, even when coupled to gravity. However, the spectral curve seemingly undergoes a phase transition on the critical surface for $H \neq 0$, characterized by the critical curves γ_b depicted in Figure 2. This transition is

characterized by a merging of the measures μ_1 and μ_2 discussed in Remark 2.4. This does not necessarily imply that the partition function itself undergoes a transition, but certainly closer attention to this case is needed. Some preliminary analysis reveals the appearance of a special 3×3 parametrix, which involves Painlevé I. The Lax pair associated to this parametrix has appeared in the literature [52], but the parametrix itself requires a more careful analysis, especially if one wants to identify the particular solution appearing when one approaches γ_b . More generally, this parametrix, along with the $(3, 4)$ parametrix discussed in Part II of this work [49], belong to a hierarchy of string equations of type $(3, p)$, where $p = 3k + 1, 3k + 2$. A more comprehensive study of these equations, as well as the more general (q, p) string equations, is needed, if one is interested in completing the picture of minimal models coupled to gravity.

Yang–Lee zeros of the partition function. Another tantalizing feature of the model we studied comes from the works of [3, 4, 79]. It is well-known that the Ising model [63, 86] that the partition function for the Ising model, as a function of the complexified parameter e^H , has all of its zeros on the unit circle. This is often thought of as the mechanism through which the ferromagnetic phase transition occurs, and it comes with its own associated minimal CFT (the Yang–Lee edge, the so-called “(2, 5)” minimal model, cf. [22]). However, it is far from obvious that sums of such partition functions should also exhibit this feature. Nevertheless, the numerical results of [3, 4, 29, 79] seem to indicate that this is indeed the case. A proper study of this phenomenon would require the study of the partition function defined by (1.1) for *complex* values of the parameters, as one would need to study this model with imaginary external field ($H \rightarrow iH$). A drawback of our current approach is that the construction of the spectral curve relied on the ansatz that the associated “ S -curves” all lie on the real axis; it is not clear that this remains true when the parameters of the model are complexified (indeed, this is *not* the case for the 1-matrix model, cf., for example, [10, 13]). It would be interesting if one could characterize these S -curves in terms of some canonical quadratic differential, as has been performed for some closely related matrix models (see [62, 77] for this construction pertaining to the 1-matrix model with/without hard walls, and [65, 66, 67] for the construction related to the external source model).

A 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 15, 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 $(a, b, c) \in R = \{0 < b \leq 1, 0 < c \leq b, 1 \leq a \leq b^{-1}\}$, and so we list them first. Let $A = A(a, b, c)$. Then,

$$u_1(z) = \frac{z}{A} - \frac{(a^2 + b^2)A}{z} - \frac{\left(\frac{5}{3}a^2b^2 + a^4 + b^4\right)A^3}{z^3} - \frac{\left(\frac{14}{3}(a^4b^2 + a^2b^4) + 2a^6 + 2b^6\right)A^5}{z^5} + \mathcal{O}(z^{-7}),$$

$$u_2(z) = \begin{cases} -\left(\frac{a^2b^2}{3}\right)^{1/3} \frac{\omega A^{1/3}}{z^{1/3}} + \frac{(a^2 + b^2)A}{3z} - \left(\frac{3}{a^2b^2}\right)^{1/3} \frac{\omega^2(a^4 + a^2b^2 + b^4)A^{5/3}}{9z^{5/3}} + \mathcal{O}(z^{-7/3}), & \text{Im } z > 0, \\ -\left(\frac{a^2b^2}{3}\right)^{1/3} \frac{\omega^2 A^{1/3}}{z^{1/3}} + \frac{(a^2 + b^2)A}{3z} - \left(\frac{3}{a^2b^2}\right)^{1/3} \frac{\omega(a^4 + a^2b^2 + b^4)A^{5/3}}{9z^{5/3}} + \mathcal{O}(z^{-7/3}), & \text{Im } z < 0, \end{cases}$$

$$u_3(z) = -\left(\frac{a^2b^2}{3}\right)^{1/3} \frac{A^{1/3}}{z^{1/3}} + \frac{(a^2+b^2)A}{3z} - \left(\frac{3}{a^2b^2}\right)^{1/3} \frac{(a^4+a^2b^2+b^4)A^{5/3}}{9z^{5/3}} + \mathcal{O}(z^{-7/3}),$$

$$u_4(z) = \begin{cases} -\left(\frac{a^2b^2}{3}\right)^{1/3} \frac{\omega^2 A^{1/3}}{z^{1/3}} + \frac{(a^2+b^2)A}{3z} \\ -\left(\frac{3}{a^2b^2}\right)^{1/3} \frac{\omega(a^4+a^2b^2+b^4)A^{5/3}}{9z^{5/3}} + \mathcal{O}(z^{-7/3}), & \text{Im } z > 0, \\ -\left(\frac{a^2b^2}{3}\right)^{1/3} \frac{\omega A^{1/3}}{z^{1/3}} + \frac{(a^2+b^2)A}{3z} \\ -\left(\frac{3}{a^2b^2}\right)^{1/3} \frac{\omega^2(a^4+a^2b^2+b^4)A^{5/3}}{9z^{5/3}} + \mathcal{O}(z^{-7/3}), & \text{Im } z < 0, \end{cases}$$

On the other hand, the local expansions of the uniformizing coordinate differ in the non-critical/critical cases and multicritical cases. We indicate the behavior of the uniformization coordinate around these branch points here.

A.1 Expansion of the uniformizing coordinate in the generic and critical cases

All of the following expansions are valid for $0 < b < 1$, $1 \leq a \leq b$ (and also $0 < c \leq b$, but this parameter does not play a role). We also let $A := A(a, b, c) > 0$ be as in (3.7).

- *Expansion at $z = +\alpha$.* As $z \rightarrow \alpha$, letting $\zeta = z - \alpha$,

$$u_1(z) = a + \frac{a^{3/2}}{A^{1/2}(a^2-b^2)^{1/2}} \zeta^{1/2} + \frac{a^2(3a^2-7b^2)}{6A(a^2-b^2)^2} \zeta + \frac{a^{5/2}C_2}{72A^{3/2}(a^2-b^2)^{7/2}} \zeta^{3/2} + \mathcal{O}(\zeta^2),$$

$$u_2(z) = a - \frac{a^{3/2}}{A^{1/2}(a^2-b^2)^{1/2}} \zeta^{1/2} + \frac{a^2(3a^2-7b^2)}{6A(a^2-b^2)^2} \zeta - \frac{a^{5/2}C_2}{72A^{3/2}(a^2-b^2)^{7/2}} \zeta^{3/2} + \mathcal{O}(\zeta^2),$$

and $u_3(z)$, $u_4(z)$ have regular expansions. Here $C_2 := C_2(a, b, c) > 0$.

- *Expansion at $z = -\alpha$.* As $z \rightarrow -\alpha$, letting $\zeta = z + \alpha$,

$$u_1(z) = \begin{cases} -a + \frac{ia^{3/2}}{A^{1/2}(a^2-b^2)^{1/2}} \zeta^{1/2} + \frac{a^2(3a^2-7b^2)}{6A(a^2-b^2)^2} \zeta \\ + \frac{ia^{5/2}C_2}{72A^{3/2}(a^2-b^2)^{7/2}} \zeta^{3/2} + \mathcal{O}(\zeta^2), & \text{Im } \zeta > 0, \\ -a - \frac{ia^{3/2}}{A^{1/2}(a^2-b^2)^{1/2}} \zeta^{1/2} + \frac{a^2(3a^2-7b^2)}{6A(a^2-b^2)^2} \zeta \\ - \frac{ia^{5/2}C_2}{72A^{3/2}(a^2-b^2)^{7/2}} \zeta^{3/2} + \mathcal{O}(\zeta^2), & \text{Im } \zeta < 0, \end{cases}$$

$$u_2(z) = \begin{cases} -a - \frac{ia^{3/2}}{A^{1/2}(a^2-b^2)^{1/2}} \zeta^{1/2} + \frac{a^2(3a^2-7b^2)}{6A(a^2-b^2)^2} \zeta \\ - \frac{ia^{5/2}C_2}{72A^{3/2}(a^2-b^2)^{7/2}} \zeta^{3/2} + \mathcal{O}(\zeta^2), & \text{Im } \zeta > 0, \\ -a + \frac{ia^{3/2}}{A^{1/2}(a^2-b^2)^{1/2}} \zeta^{1/2} + \frac{a^2(3a^2-7b^2)}{6A(a^2-b^2)^2} \zeta \\ + \frac{ia^{5/2}C_2}{72A^{3/2}(a^2-b^2)^{7/2}} \zeta^{3/2} + \mathcal{O}(\zeta^2), & \text{Im } \zeta < 0, \end{cases}$$

and $u_3(z)$, $u_4(z)$ have regular expansions. Here $C_2 := C_2(a, b, c) > 0$.

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 $z \rightarrow \beta$, letting $\zeta = z - \beta$,

$$u_2(z) = \begin{cases} b + \frac{ib^{3/2}}{A^{\frac{1}{2}}(a^2 - b^2)^{\frac{1}{2}}} \zeta^{1/2} - \frac{(7a^2 - 3b^2)b^2}{6A(a^2 - b^2)^2} \zeta \\ \quad + \frac{ib^{5/2}\tilde{C}_2}{72A^{\frac{5}{2}}(a^2 - b^2)^{\frac{7}{2}}} \zeta^{\frac{3}{2}} + \mathcal{O}(\zeta^2), & \text{Im } \zeta > 0, \\ b - \frac{ib^{3/2}}{A^{\frac{1}{2}}(a^2 - b^2)^{\frac{1}{2}}} \zeta^{1/2} - \frac{(7a^2 - 3b^2)b^2}{6A(a^2 - b^2)^2} \zeta \\ \quad - \frac{ib^{5/2}\tilde{C}_2}{72A^{\frac{5}{2}}(a^2 - b^2)^{\frac{7}{2}}} \zeta^{\frac{3}{2}} + \mathcal{O}(\zeta^2), & \text{Im } \zeta < 0, \end{cases}$$

$$u_4(z) = \begin{cases} b - \frac{ib^{3/2}}{A^{\frac{1}{2}}(a^2 - b^2)^{\frac{1}{2}}} \zeta^{1/2} - \frac{(7a^2 - 3b^2)b^2}{6A(a^2 - b^2)^2} \zeta \\ \quad - \frac{ib^{5/2}\tilde{C}_2}{72A^{\frac{5}{2}}(a^2 - b^2)^{\frac{7}{2}}} \zeta^{\frac{3}{2}} + \mathcal{O}(\zeta^2), & \text{Im } \zeta > 0, \\ b + \frac{ib^{3/2}}{A^{\frac{1}{2}}(a^2 - b^2)^{\frac{1}{2}}} \zeta^{1/2} - \frac{(7a^2 - 3b^2)b^2}{6A(a^2 - b^2)^2} \zeta \\ \quad + \frac{ib^{5/2}\tilde{C}_2}{72A^{\frac{5}{2}}(a^2 - b^2)^{\frac{7}{2}}} \zeta^{\frac{3}{2}} + \mathcal{O}(\zeta^2), & \text{Im } \zeta < 0, \end{cases}$$

and $u_1(z), u_3(z)$ have regular expansions. Here $\tilde{C}_2 := \tilde{C}_2(a, b, c) > 0$.

- *Expansion at $z = -\beta$.* As $z \rightarrow -\beta$, letting $\zeta = z + \beta$,

$$u_2(z) = -b - \frac{b^{3/2}}{A^{\frac{1}{2}}(a^2 - b^2)^{\frac{1}{2}}} \zeta^{1/2} - \frac{(7a^2 - 3b^2)b^2}{6A(a^2 - b^2)^2} \zeta - \frac{b^{5/2}\tilde{C}_2}{72A^{\frac{5}{2}}(a^2 - b^2)^{\frac{7}{2}}} \zeta^{\frac{3}{2}} + \mathcal{O}(\zeta^2),$$

$$u_3(z) = -b + \frac{b^{3/2}}{A^{\frac{1}{2}}(a^2 - b^2)^{\frac{1}{2}}} \zeta^{1/2} - \frac{(7a^2 - 3b^2)b^2}{6A(a^2 - b^2)^2} \zeta + \frac{b^{5/2}\tilde{C}_2}{72A^{\frac{5}{2}}(a^2 - b^2)^{\frac{7}{2}}} \zeta^{\frac{3}{2}} + \mathcal{O}(\zeta^2),$$

and $u_1(z), u_4(z)$ have regular expansions. Here

$$\tilde{C}_2 := \tilde{C}_2(a, b) = 101a^4 - 30a^2b^2 + 9b^4 = C_2(b, a).$$

A.2 Expansions of the uniformizing coordinate at the multicritical point and γ_b

- *Expansion at $z = +\alpha$.* Let $\zeta := z - \alpha$, $A = A(1, 1, c) > 0$; as $\zeta \rightarrow 0$, we have the expansions:

$$u_1(z) = 1 + \left(\frac{3}{4A}\right)^{1/3} \zeta^{1/3} + \frac{3}{4} \left(\frac{3}{4A}\right)^{2/3} \zeta^{2/3} + \frac{21}{64A} \zeta + \frac{37}{192} \left(\frac{3}{4A}\right)^{4/3} \zeta^{4/3} + \mathcal{O}(\zeta^{5/3}),$$

$$u_2(z) = \begin{cases} 1 + \left(\frac{3}{4A}\right)^{1/3} \omega^2 \zeta^{1/3} + \frac{3}{4} \left(\frac{3}{4A}\right)^{2/3} \omega \zeta^{2/3} + \frac{21}{64A} \zeta \\ \quad + \frac{37}{192} \left(\frac{3}{4A}\right)^{4/3} \omega^2 \zeta^{4/3} + \mathcal{O}(\zeta^{5/3}), & \text{Im } \zeta > 0, \\ 1 + \left(\frac{3}{4A}\right)^{1/3} \omega \zeta^{1/3} + \frac{3}{4} \left(\frac{3}{4A}\right)^{2/3} \omega^2 \zeta^{2/3} + \frac{21}{64A} \zeta \\ \quad + \frac{37}{192} \left(\frac{3}{4A}\right)^{4/3} \omega \zeta^{4/3} + \mathcal{O}(\zeta^{5/3}), & \text{Im } \zeta < 0, \end{cases}$$

$$\begin{aligned}
u_3(z) &= -\frac{1}{3} + \frac{1}{64A}\zeta - \frac{27}{16384A^2}\zeta^2 + \frac{891}{4194304A^3}\zeta^3 + \mathcal{O}(\zeta^4), \\
u_4(z) &= \begin{cases} 1 + \left(\frac{3}{4A}\right)^{1/3} \omega \zeta^{1/3} + \frac{3}{4} \left(\frac{3}{4A}\right)^{2/3} \omega^2 \zeta^{2/3} + \frac{21}{64A}\zeta \\ \quad + \frac{37}{192} \left(\frac{3}{4A}\right)^{4/3} \omega \zeta^{4/3} + \mathcal{O}(\zeta^{5/3}), & \text{Im } \zeta > 0, \\ 1 + \left(\frac{3}{4A}\right)^{1/3} \omega^2 \zeta^{1/3} + \frac{3}{4} \left(\frac{3}{4A}\right)^{2/3} \omega \zeta^{2/3} + \frac{21}{64A}\zeta \\ \quad + \frac{37}{192} \left(\frac{3}{4A}\right)^{4/3} \omega^2 \zeta^{4/3} + \mathcal{O}(\zeta^{5/3}), & \text{Im } \zeta < 0. \end{cases}
\end{aligned}$$

- *Expansion at $z = -\alpha$.* Let $\zeta := z + \alpha$, $A = A(1, 1, c) > 0$; as $\zeta \rightarrow 0$, we have the expansions:

$$\begin{aligned}
u_1(z) &= \begin{cases} -1 + \left(\frac{3}{4A}\right)^{1/3} \omega \zeta^{1/3} - \frac{3}{4} \left(\frac{3}{4A}\right)^{2/3} \omega^2 \zeta^{2/3} + \frac{21}{64A}\zeta \\ \quad - \frac{37}{192} \left(\frac{3}{4A}\right)^{4/3} \omega \zeta^{4/3} + \mathcal{O}(\zeta^{5/3}), & \text{Im } \zeta > 0, \\ -1 + \left(\frac{3}{4A}\right)^{1/3} \omega^2 \zeta^{1/3} - \frac{3}{4} \left(\frac{3}{4A}\right)^{2/3} \omega \zeta^{2/3} + \frac{21}{64A}\zeta \\ \quad - \frac{37}{192} \left(\frac{3}{4A}\right)^{4/3} \omega^2 \zeta^{4/3} + \mathcal{O}(\zeta^{5/3}), & \text{Im } \zeta < 0, \end{cases} \\
u_2(z) &= \begin{cases} -1 + \left(\frac{3}{4A}\right)^{1/3} \omega^2 \zeta^{1/3} - \frac{3}{4} \left(\frac{3}{4A}\right)^{2/3} \omega \zeta^{2/3} + \frac{21}{64A}\zeta \\ \quad - \frac{37}{192} \left(\frac{3}{4A}\right)^{4/3} \omega^2 \zeta^{4/3} + \mathcal{O}(\zeta^{5/3}), & \text{Im } \zeta > 0, \\ -1 + \left(\frac{3}{4A}\right)^{1/3} \omega \zeta^{1/3} - \frac{3}{4} \left(\frac{3}{4A}\right)^{2/3} \omega^2 \zeta^{2/3} + \frac{21}{64A}\zeta \\ \quad - \frac{37}{192} \left(\frac{3}{4A}\right)^{4/3} \omega \zeta^{4/3} + \mathcal{O}(\zeta^{5/3}), & \text{Im } \zeta < 0, \end{cases} \\
u_3(z) &= -1 + \left(\frac{3}{4A}\right)^{1/3} \zeta^{1/3} - \frac{3}{4} \left(\frac{3}{4A}\right)^{2/3} \zeta^{2/3} + \frac{21}{64A}\zeta - \frac{37}{192} \left(\frac{3}{4A}\right)^{4/3} \zeta^{4/3} + \mathcal{O}(\zeta^{5/3}), \\
u_4(z) &= \frac{1}{3} + \frac{1}{64A}\zeta + \frac{27}{16384A^2}\zeta^2 + \frac{891}{4194304A^3}\zeta^3 + \mathcal{O}(\zeta^4).
\end{aligned}$$

We also note the following symmetry properties of these coordinates:

Proposition A.1. *The expansions $u_j(z)$ and $u_j(-z)$, $j = 1, 2, 3, 4$ are related by*

$$\begin{aligned}
u_1(-z) &= -u_1(z), & u_2(-z) &= -u_2(z), \\
u_3(-z) &= -u_4(z), & u_4(-z) &= -u_3(z).
\end{aligned}$$

Proof. This proposition can be observed directly from the expansions of $u_j(z)$ given above. ■

B The isomonodromic tau function

The partition function of the 2-matrix model was identified with an isomonodromic τ -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

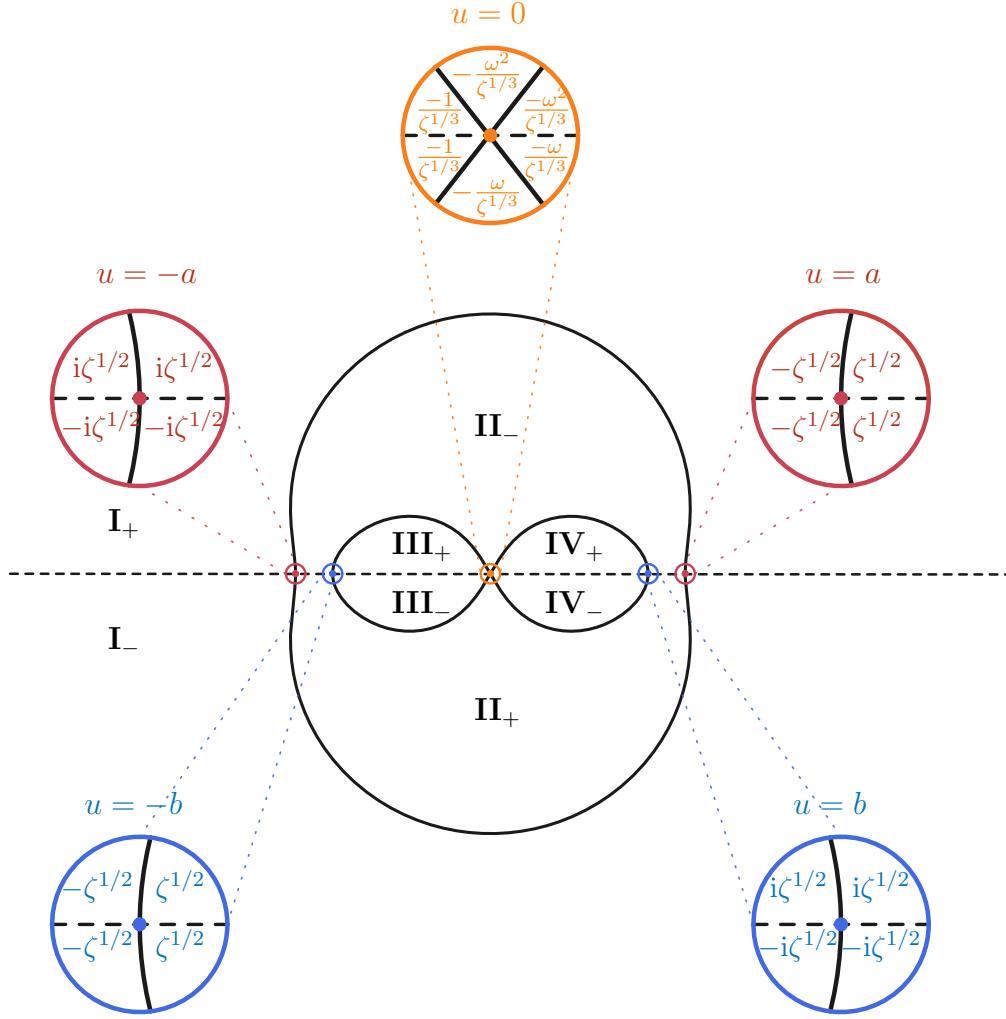


Figure 15. The uniformizing plane for generic values of the parameters (a, b, c) . The leading asymptotic behavior of $u_j(z)$ on each sheet is given in the appropriate local coordinate. For example, in a neighborhood of $u = a$, the local coordinate $\zeta = \text{const} \cdot (z - \beta)$, and $u_2(z) \sim -\zeta^{1/2}[1 + \mathcal{O}(\zeta^{1/2})]$, whereas $u_3(z) \sim \zeta^{1/2}[1 + \mathcal{O}(\zeta^{1/2})]$.

the 2-matrix model is an isomonodromic τ -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 $A(z)$,

$$\langle A(z) \rangle := \text{Res}_{z=\infty} \text{tr}[A(z)].$$

Recall that, in general, the n^{th} monic biorthogonal polynomial $p_n(z)$ with respect to the weight⁸

$$e^{NW(z,w)} := e^{N[\tau zw - V(z;T) - V(w;\bar{T})]}$$

(Here $V(z;T) := \frac{1}{2}z^2 + \frac{T}{4}z^4$, and we take T, \bar{T} as independent parameters in general) on contour(s) $(\Gamma, \bar{\Gamma})$ is given in terms of the solution to the following Riemann–Hilbert problem:

⁸Here we take $N > 0$ as a free parameter, in general different from the index of the biorthogonal polynomial. We will later set $N = n$.

- (1) $\mathbf{Y}_n(z)$ is a piecewise analytic function in $\mathbb{C} \setminus \Gamma$.
- (2) $\mathbf{Y}_n(z)$ has boundary values

$$\mathbf{Y}_{n,+}(z) = \mathbf{Y}_{n,-}(z)e^{-NV(z)} \left[\mathbb{I} + w(z)E_{12} + \frac{1}{N\tau}w'(z)E_{13} + \frac{1}{N^2\tau^2}w''(z)E_{14} \right], \quad z \in \Gamma,$$

where $w(z) := \int_{\Gamma} e^{N[\tau zw - V(w, \bar{T})]} dw$ (note that in the general situation, the integration contour Γ is allowed to differ from the jump contour Γ ; in this work, this distinction is irrelevant).

- (3) As $z \rightarrow \infty$,

$$\mathbf{Y}_n(z) = \left[\mathbb{I} + \frac{\mathbf{Y}_n^{(1)}}{z} + \frac{\mathbf{Y}_n^{(2)}}{z^2} + \mathcal{O}\left(\frac{1}{z^3}\right) \right] \begin{pmatrix} z^n & 0 & 0 \\ 0 & z^{-m_n-1}\mathbb{I}_{r_n} & 0 \\ 0 & 0 & z^{-m_n}\mathbb{I}_{3-r_n} \end{pmatrix},$$

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. If we put $h_n(\tau, T, \bar{T}; N)$ to be the norming constant for the monic n^{th} -order polynomials,

$$\int_{\Gamma} \int_{\Gamma} p_n(z) q_m(w) e^{NW(z,w)} dz dw = h_n(\tau, T, \bar{T}; N) \delta_{nm},$$

then the *partition function* of this model is defined to be

$$Z_n(\tau, T, \bar{T}; N) := \prod_{k=0}^{n-1} h_k(\tau, T, \bar{T}; N).$$

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,

$$\mathbf{Y}_n(z) = \begin{pmatrix} p_n(z) & C_{\Gamma}[p_n w](z) & \frac{1}{N\tau}C_{\Gamma}[p_n w'](z) & \frac{1}{N^2\tau^2}C_{\Gamma}[p_n w''](z) \\ Q_{n-1}(z) & C_{\Gamma}[Q_{n-1} w](z) & \frac{1}{N\tau}C_{\Gamma}[Q_{n-1} w'](z) & \frac{1}{N^2\tau^2}C_{\Gamma}[Q_{n-1} w''](z) \\ Q_{n-2}(z) & C_{\Gamma}[Q_{n-2} w](z) & \frac{1}{N\tau}C_{\Gamma}[Q_{n-2} w'](z) & \frac{1}{N^2\tau^2}C_{\Gamma}[Q_{n-2} w''](z) \\ Q_{n-3}(z) & C_{\Gamma}[Q_{n-3} w](z) & \frac{1}{N\tau}C_{\Gamma}[Q_{n-3} w'](z) & \frac{1}{N^2\tau^2}C_{\Gamma}[Q_{n-3} w''](z) \end{pmatrix},$$

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

$$C_{\Gamma}[f](z) := \frac{1}{2\pi i} \int_{\Gamma} \frac{f(x)}{x-z} dz$$

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*

$$\mathbf{Y}_{n+1}(z) = R_n(z) \mathbf{Y}_n(z),$$

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 $R_n^{(1)} = E_{11}$, whereas the matrix $(R_n^{(0)})_{jk}$ has entries as given in the following table:

	$k = \alpha_0$	$k = 1$	$k \neq 1, \alpha_0$
$j = \alpha_0$	$\frac{-(\mathbf{Y}_n^{(2)})_{\alpha_0,1} + \sum_{\ell \neq \alpha_0} (\mathbf{Y}_n^{(1)})_{\alpha_0,\ell} (\mathbf{Y}_n^{(1)})_{\ell,1}}{(\mathbf{Y}_n^{(1)})_{\alpha_0,1}}$	$-(\mathbf{Y}_n^{(1)})_{\alpha_0,1}$	$-(\mathbf{Y}_n^{(1)})_{\alpha_0,k}$
$j = 1$	$\frac{1}{(\mathbf{Y}_n^{(1)})_{\alpha_0,1}}$	0	0
$j \neq 1, \alpha_0$	$\frac{(\mathbf{Y}_n^{(1)})_{j,1}}{(\mathbf{Y}_n^{(1)})_{\alpha_0,1}}$	0	δ_{jk}

This implies that the matrix $R_n(z)$ is determined entirely in terms of $\mathbf{Y}_n(z)$.

The isomonodromic τ -differential corresponding to $\mathbf{Y} := \mathbf{Y}_n$ is defined to be

$$d \log \tau_n := \langle \mathbf{Y}_n^{-1} \mathbf{Y}_n' d\hat{\mathbf{W}} \hat{\mathbf{W}}^{-1} \rangle,$$

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}

$$d\hat{\mathbf{W}}\hat{\mathbf{W}}^{-1} = \begin{pmatrix} 0 & 0 \\ 0 & -\mathbf{W}^{-1} \frac{\partial \mathbf{W}}{\partial \tau} \end{pmatrix} d\tau + \begin{pmatrix} 0 & 0 \\ 0 & -\mathbf{W}^{-1} \frac{\partial \mathbf{W}}{\partial T} \end{pmatrix} dT + \begin{pmatrix} -Nz^4/4 & 0 \\ 0 & 0 \end{pmatrix} d\bar{T}.$$

In [9], the parameters of the isomonodromic τ -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

$$\mathbf{Y}_{n+1}(z) = R_n(z) \mathbf{Y}_n(z). \quad (\text{B.1})$$

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 B.1.

$$d \log \frac{\tau_{n+1}}{\tau_n} = -\langle R_n^{-1} R_n' d\mathbf{Y}_n \mathbf{Y}_n^{-1} \rangle.$$

Proof. By equation (B.1), 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 τ differentials is

$$\begin{aligned} d \log \frac{\tau_{n+1}}{\tau_n} &= \langle [\mathbf{Y}_n^{-1} R_n^{-1} R_n' \mathbf{Y}_n + \mathbf{Y}_n^{-1} \mathbf{Y}_n'] d\hat{\mathbf{W}} \hat{\mathbf{W}}^{-1} \rangle - \langle \mathbf{Y}_n^{-1} \mathbf{Y}_n' d\hat{\mathbf{W}} \hat{\mathbf{W}}^{-1} \rangle \\ &= \langle \mathbf{Y}_n^{-1} R_n^{-1} R_n' \mathbf{Y}_n d\hat{\mathbf{W}} \hat{\mathbf{W}}^{-1} \rangle. \end{aligned}$$

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

$$\mathbf{Y}_n d\hat{\mathbf{W}} \hat{\mathbf{W}}^{-1} \mathbf{Y}_n^{-1} = d\mathbf{Y}_n \mathbf{Y}_n^{-1} + \text{polynomial}.$$

Inserting the above into our expression for the τ -quotient, we obtain that

$$d \log \frac{\tau_{n+1}}{\tau_n} = \langle R_n^{-1} R'_n [\text{polynomial}] - R_n^{-1} R'_n d\mathbf{Y}_n \mathbf{Y}_n^{-1} \rangle;$$

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. \blacksquare

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 B.2. *The ratio of consecutive τ differentials, up to multiplication by a function independent of the isomonodromic times, is given by*

$$\frac{\tau_{n+1}}{\tau_n} = (\mathbf{Y}_n^{(1)})_{1, \alpha_0}.$$

Proof. This follows immediately from inspection of the previous proposition, and the explicit form of the matrices $R_n(z)$. For details, see [9]. \blacksquare

Furthermore, we can relate the coefficients $(\mathbf{Y}_n^{(1)})_{1, \alpha_0}$ to the biorthogonality coefficients h_n .

Proposition B.3. *The matrix coefficient $(\mathbf{Y}_n^{(1)})_{1, \alpha_0}$ is given in terms of the n^{th} normalizing constant of the biorthogonal polynomials:*

$$(\mathbf{Y}_n^{(1)})_{1, \alpha_0} = \left(\frac{T}{\tau} \right)^S h_n,$$

where $S \in \mathbb{N}$, $\alpha_0 \in \{0, 1, 2\}$ are such that $n = 3S + \alpha_0 - 1$.

Proposition B.4. *The isomonodromic τ -function τ_n is related to the partition function $Z_n(\tau, t, H; N)$ for n a multiple of 3 by the formula*

$$Z_n(\tau, T, \bar{T}; N) = \left(\frac{\tau}{T} \right)^{\frac{n}{2} \left(\frac{n}{3} - 1 \right)} \tau_n.$$

Remark B.5. We remark that there is a minor error in the statement of the above proposition in [9]. The power of (T/τ) should be inverse to what it reads in their Theorem 3.4 (see page 17). The proof is otherwise correct, and there are no essential changes to the results otherwise.

Thus, we have an explicit expression for the partition function in terms of quantities we can calculate, by making the change of variables $T = qt$, $\bar{T} = q^{-1}t$. However, the calculation of the differential in the variable \bar{T} is computationally difficult, as it would require us to calculate the expansion of \mathbf{Y} to 4 subleading terms. However, we can use the fact that the partition function is symmetric in T, \bar{T} ; in other words,

$$d \log Z_n(\tau, T, \bar{T}; N) = d \log Z_n(\tau, \bar{T}, T; N).$$

If we set

$$\varpi_\tau(\tau, T, \bar{T}, n; N) := \langle \mathbf{Y}_n^{-1} \mathbf{Y}'_n \hat{\mathbf{W}}_\tau \hat{\mathbf{W}}^{-1} \rangle, \quad \varpi_T(\tau, T, \bar{T}, n; N) := \langle \mathbf{Y}_n^{-1} \mathbf{Y}'_n \hat{\mathbf{W}}_T \hat{\mathbf{W}}^{-1} \rangle,$$

then we have the following proposition, which follows immediately from the symmetry

$$\log Z_n(T, \bar{T}) = \log Z_n(\bar{T}, T),$$

and our previous calculations.

Proposition B.6. *The partition function is given by the expression*

$$\begin{aligned} d \log Z_n = d \log \left(\frac{\tau}{T\bar{T}} \right)^{\frac{n}{2}(\frac{n}{3}-1)} + \varpi_\tau(\tau, T, \bar{T}, n; N) d\tau + \varpi_T(\tau, T, \bar{T}, n; N) dT \\ + \varpi_T(\tau, \bar{T}, T, n; N) d\bar{T}. \end{aligned}$$

C Explicit formulae for the spectral curve and the critical surface

Here we present explicit formulae for the spectral curve $\mathfrak{S}(X, Y) = 0$, and the critical surface $\Xi(\tau, t, q) = 0$. Both of these formulae are rather longer and not particularly enlightening. We have therefore relinquished their presentation in the main text and placed them here, for the sake of completeness.

C.1 Implicit formula for the spectral curve

One can eliminate the parameter u from the pair of functions $(X(u), Y(u))$ to obtain an implicit formula for the Riemann surface these functions parametrize

$$\mathfrak{S}(X(u), Y(u)) = 0.$$

$\mathfrak{S}(X, Y)$ is a degree 6 polynomial in X and Y , with rational coefficients in the variables τ, t, q , and σ (recall that σ is defined as a special solution to the algebraic equation (2.1)). Explicitly, this polynomial is

$$\begin{aligned} \mathfrak{S}(X, Y) = \tau q X^4 + \tau q^{-1} Y^4 - t X^3 Y^3 - q X^3 Y - q^{-1} Y^3 X + t \tau^{-1} X^2 Y^2 \\ + \mathfrak{s}_2(\sigma; \tau, t, q) X^2 + \mathfrak{s}_2(\sigma; \tau, t, q^{-1}) Y^2 + \mathfrak{s}_1(\sigma; \tau, t, q) X Y + \mathfrak{s}_0(\sigma; \tau, t, q). \end{aligned}$$

where $\mathfrak{s}_i(\sigma; \tau, t, q)$ are given by

$$\begin{aligned} \mathfrak{s}_2(\sigma; \tau, t, q) &= \frac{1 - 6\sigma - 3\sigma^2}{27\tau t(\sigma + 1)^3} - \frac{1}{27\tau t} (\sigma^3 \tau^2 + 3\sigma^2 \tau^2 - 9\sigma \tau^2 - 27\tau^2 + 1) \\ &\quad - \frac{(q - 1)\sigma}{27\tau t(\sigma^2 - 1)^3} [(9\tau^2 - 9) + 9(q + 1)\sigma + (4q^{-1} - (28\tau^2 + 6))\sigma^2 \\ &\quad - 6(q + 1)\sigma^3 + (30\tau^2 + 3)\sigma^4 + (q + 1)\sigma^5 - 12\sigma^6 \tau^2 + \sigma^8 \tau^2], \\ \mathfrak{s}_1(\sigma; \tau, t, q) &= -\frac{8(5\sigma + 3)}{81(\sigma + 1)^2 t} - \frac{\sigma^6 \tau^2 - 15\sigma^4 \tau^2 + 27\sigma^2 \tau^2 - 3\sigma^2 + 243\tau^2 + 24\sigma + 171}{243t} \\ &\quad - \frac{4\sigma^3(\sigma^2 + 3)}{81t(\sigma^2 - 1)^2} [q + q^{-1} - 2], \\ \mathfrak{s}_0(\sigma; \tau, t, q) &= -\frac{\sigma}{19683t^2 \tau q^2 (\sigma^2 - 1)^4} \left[15q\sigma^6 \tau^2 + 9q\sigma^5 t + (12 - 111\tau^2)q\sigma^4 \right. \end{aligned}$$

$$\begin{aligned}
& + 15 \left(q^2 - \frac{6}{5}tq + 1 \right) \sigma^3 + (177\tau^2 - 15)q\sigma^2 - 54 \left(q^2 - \frac{1}{6}tq + 1 \right) \sigma \\
& + 81(1 - \tau^2)q \Big]^2.
\end{aligned}$$

C.2 Implicit formula for the critical surface

Recall that the critical surface is characterized by the vanishing of the discriminant of the polynomial $\mathcal{J}(\sigma; \tau, t, q)$, defined in equation (2.1). In other words, let $\Xi(\tau, t, q)$ denote this discriminant. Then, we have that

$$(\tau, t, q) \in S_L \cup S_H \implies \Xi(\tau, t, q) = 0.$$

Explicitly, we can write this curve as ($C = \frac{1}{2}(q + q^{-1}) = \cosh(H)$)

$$\begin{aligned}
\mathcal{J}(\tau, t, C) := & 32400000t\tau^4C^5 + 1350000\tau^4(\tau^4 + 288t^2 - 2\tau^2 + 1)C^4 \\
& + 36000t\tau^2(3160\tau^6 + 28026t^2\tau^2 - 2745\tau^4 + 4374t^2 - 1215\tau^2)C^3 \\
& + (3840000\tau^{12} - 423014400t^2\tau^8 - 10560000\tau^{10} - 1700611200t^4\tau^4 \\
& - 235612800t^2\tau^6 + 7440000\tau^8 - 306110016t^6 + 2173003200t^4\tau^2 \\
& - 666646200t^2\tau^4 + 1440000\tau^6 + 25223400t^2\tau^2 - 2160000\tau^4)C^2 \\
& + (-10174464t\tau^{12} + 403107840t^3\tau^8 + 60549120t\tau^{10} - 3809369088t^5\tau^4 \\
& + 2205895680t^3\tau^6 - 197475840t\tau^8 - 3673320192t^7 + 8865853056t^5\tau^2 \\
& - 2979218880t^3\tau^4 + 147277440t\tau^6 - 25509168t^5 + 249930360t^3\tau^2 \\
& - 1451520t\tau^4 + 1275264t\tau^2)C \\
& - 65536\tau^{16} + 5308416t^2\tau^{12} + 1277952\tau^{14} - 161243136t^4\tau^8 - 29859840t^2\tau^{10} \\
& - 7827456\tau^{12} + 2176782336t^6\tau^4 - 362797056t^4\tau^6 - 21772800t^2\tau^8 \\
& + 15861760\tau^{10} - 11019960576t^8 + 8979227136t^6\tau^2 - 5048925696t^4\tau^4 \\
& + 806993280t^2\tau^6 - 12437760\tau^8 - 153055008t^6 + 583036704t^4\tau^2 \\
& + 42729120t^2\tau^4 + 2624256\tau^6 - 531441t^4 + 6601824t^2\tau^2 \\
& + 546048\tau^4 + 20736\tau^2 = 0.
\end{aligned}$$

One can readily check that $\vec{r}_{\text{low}}(b, c)$, $\vec{r}_{\text{high}}(b, c)$ (defined below), indeed both parametrize the above algebraic equation.

D Parametrization of the phase space

Here we discuss the parametrization of D_q by (3.4)–(3.6) in more detail and prove Proposition 2.5.

D.1 The critical surfaces

Let us start by studying the critical surfaces defined in Definition 2.4. Note that, by definition, the low and high temperature curves are given by the pair of parametric equations $\vec{r}_{\bullet}(b, c) = \langle \tau_{\bullet}(b, c), t_{\bullet}(b, c), q_{\bullet}(b, c) \rangle$, $\bullet \in \{\text{low}, \text{high}\}$, defined by

$$\vec{r}_{\text{low}}(b, c) := \left\langle \frac{b^3}{\sqrt{N_1(b, c)N_2(b, c)}}, -\frac{bc(b^6c^2 + \frac{4}{3}b^2c^2 + 1)}{3N_1(b, c)N_2(b, c)}, \frac{bcN_2(b, c)}{N_1(b, c)} \right\rangle,$$

$$\vec{r}_{\text{high}}(b, c) := \left\langle \frac{1}{\sqrt{M_1(b, c)M_2(b, c)}}, -\frac{bc(4b^2c^2 + 3b^2 + 3c^2)}{9M_1(b, c)M_2(b, c)}, \frac{cM_2(b, c)}{bM_1(b, c)} \right\rangle,$$

respectively, where

$$\begin{aligned} N_1(b, c) &= 2b^4c^2 + b^2 + c^2, & N_2(b, c) &= b^4 + b^2c^2 + 2, \\ M_1(b, c) &= b^2c^2 + 2c^2 + 1, & M_2(b, c) &= b^2c^2 + 2b^2 + 1. \end{aligned}$$

In both cases, the parameter range for (b, c) is $0 < b < 1$, $0 < c < b$. The critical curve γ_b , defined as the boundary connecting the surfaces S_{low} and S_{high} , is given parametrically as

$$\vec{r}(c) = \left\langle \frac{1}{\sqrt{(3c^2 + 1)(c^2 + 3)}}, -\frac{c(7c^2 + 3)}{9(c^2 + 3)(3c^2 + 1)}, \frac{c(c^2 + 3)}{3c^2 + 1} \right\rangle, \quad 0 < c < 1,$$

which results from the limit $b \rightarrow 1$ in either of the critical surfaces $\vec{r}_{\text{low}}, \vec{r}_{\text{high}}$.

Note that by changing variables from (τ, t, q) to (τ^2, t, q) the parametrizations become rational and thus the surface S_{low} and S_{high} are algebraic.

Lemma D.1. *The union $S_{\text{low}} \cup S_{\text{high}}$ of the low- and high-temperature critical surface is the graph of a function of $0 < \tau < 1$ and $0 < q < 1$.*

Proof. It suffices to show that the projection of $S_{\text{high}} \cup S_{\text{low}} \cup \gamma_b$ to the $t = 0$ plane is a bijection onto the square defined by $0 < \tau < 1$ and $0 < q < 1$.

The projection of the surface S_{high} to the $t = 0$ plane is obtained by setting $t = 0$ in its parameterization. The parameterization becomes singular near $b = c = 0$, but this can be resolved by setting $\tilde{c} = c/b$, leading to the map

$$(\tau, q) = \left(\frac{1}{\sqrt{(1 + 2\tilde{c}^2b^2 + \tilde{c}^2b^4)(1 + 2b^2 + \tilde{c}^2b^4)}}, \frac{\tilde{c}(1 + 2b^2 + \tilde{c}^2b^4)}{1 + 2\tilde{c}^2b^2 + \tilde{c}^2b^4} \right). \quad (\text{D.1})$$

Note that this defines a smooth map on \mathbb{R}^2 , but we are mainly interested in its restriction to the closure of the parameter space given by unit square $0 \leq b \leq 1$ and $0 \leq \tilde{c} < 1$. The map $(b, \tilde{c}) \mapsto (\tau, q)$ is differentiable and the Jacobian never vanishes in the open square. This implies that it is locally a diffeomorphism. The image of the boundary is given by the following observations:

- For $b = 0$, we find $\tau = 1$ and $0 \leq q \leq 1$.
- For $b = 1$, we obtain the projection of curve γ_b to the $t = 0$ plane.
- For $\tilde{c} = 0$, we find $q = 0$ and $1/\sqrt{3} \leq \tau \leq 1$.
- For $\tilde{c} = 1$, we find $q = 1$ and $1/4 \leq \tau \leq 1$.

It is also easy to verify that the map (D.1) provides a bijection from the boundary of the unit square to the boundary of this region.

Concluding, (D.1) is a continuous map from the closed unit square, that is locally diffeomorphic and maps the boundary bijective to the boundary of the image. Since ϕ is continuous and the unit square is compact, it is also a proper map (pre-image of compact sets are compact). By further invoking the Hadamard–Caccioppoli theorem [60], one find that the map defines a homeomorphism, from the closed unit disk onto its image. Moreover, the image of the boundary divides \mathbb{R}^2 into two components and the image equals the bounded component. In other words, the projection of S_{high} is the region enclosed by $\tau = 1$, $q = 0$, $q = 1$ and the projection of γ_b to the $t = 0$ plane.

In a similar fashion, one can show that the projection of S_{low} is the region enclosed by $\tau = 0$, $q = 0$, $q = 1$ and the projection of γ_b to the $t = 0$ plane. The projections together thus fill out the unit square bijectively and this proves the statement. \blacksquare

D.2 Proof of Proposition 2.5

Proof. We will characterize the image of the parametrization inspired by the following simple principle: if D is an open set with compact closure and $\phi: \overline{D} \rightarrow \mathbb{R}^n$ is a continuous function such that $\phi(D)$ open, then $\partial\phi(D) \subset \phi(\partial D)$. Since the Jacobian (2.4) does not vanish in the interior of R , the map Π maps R° to an open subset of \mathbb{R}^3 . However, R is not compact and Π is not continuous for $b \rightarrow 0$. Nevertheless, as we will show below by a limiting procedure, the boundary $\partial\Pi(R)$ can still be found from the behavior of Π near the boundary ∂R .

For $\varepsilon > 0$, define $R^\varepsilon = R \cap \{b > \varepsilon\}$. Then the closure of R^ε is compact and Π is continuous up to the boundary. The boundary of $\Pi(R^\varepsilon)$ consists of the images of the boundary part of R^ε given by respectively $a = b^{-1}$, $a = 1$, $c = b$, $c = 0$ and $b = \varepsilon$. The first two conditions, lead to the critical surfaces S_{low} and S_{high} (note that for $\varepsilon \downarrow 0$ this will only trace our part of these surface but when $\varepsilon \downarrow 0$ we retrieve the entire surfaces again). The condition $c = b$ gives us (part of) the plane $q = 1$. The condition $c = 0$ turns out to be obsolete as it leads to both $t = 0$ and $q = 0$ and this is the part where the critical surfaces meet the $t = 0$ plane. So it remains to investigate the behavior of the image of $b = \varepsilon$ when $\varepsilon \downarrow 0$. Note that the image of $b = \varepsilon$ under Π gives a smooth surface. We claim that for $\varepsilon \downarrow 0$ this surface will converge to parts of the planes $\tau = 0$ and $t = 0$.

To study the $\varepsilon \downarrow 0$ limit we start, as in the proof of Lemma D.1, by setting $c = \tilde{c}b$, and consider the parametrization by a , b and \tilde{c} . Now $0 \leq \tilde{c} \leq 1$ and the limit $b = \varepsilon \downarrow 0$ simply gives

$$\tau(a, b, c) \rightarrow 1/a^2, \quad t(a, b, c) \rightarrow 0, \quad q(a, b, c) \rightarrow \tilde{c}.$$

One can readily verify that this parameterizes the $t = 0$ wall of the phase space for $0 < q < 1$:

$$\{(\tau, t, q) \mid t = 0, 0 < \tau < 1, 0 < q < 1\}.$$

The $\tau = 0$ plane can be obtained in a similar way. To this end, set $a = \tilde{a}/b$, $\tilde{c} = c/b$, with $0 < \tilde{a} \leq 1$ and $0 \leq \tilde{c} \leq 1$. Then for $b = \varepsilon \downarrow 0$ we find

$$\tau(a, b, c) \rightarrow 0, \quad t(a, b, c) \rightarrow -\frac{\tilde{a}^2 \tilde{c}}{3(1 + \tilde{a}^2)(1 + \tilde{a}^2 \tilde{c}^2)}, \quad q(a, b, c) \rightarrow \frac{\tilde{c}(1 + \tilde{a}^2)}{1 + \tilde{a}^2 \tilde{c}^2}.$$

One can readily verify that the above parametrizes the following set:

$$\left\{(\tau, t, q) \mid \tau = 0, -\frac{1}{12}q < t < 0, 0 < q < 1\right\}. \quad (\text{D.2})$$

Equality $t = -q/12$ in this limiting procedure is obtained by setting $\tilde{a} = 1$ and thus $a = b^{-1}$. Therefore, (D.2) is the part of the plane $\tau = 0$ enclosed by $t = 0$, $q = 0$, $q = 1$ and the part of the boundary of S_{low} that is in the $\tau = 0$ plane.

Concluding, after taking the limit $\varepsilon \downarrow 0$, we find that Π maps R° to the open region in \mathbb{R}^3 for which the boundary consists of the critical surfaces $S_{\text{low}} \cup S_{\text{high}} \cup \gamma_b$ and the four planes defined by respectively $q = 0$, $q = 1$, $t = 0$ and $\tau = 0$.

Next we show that $\Pi(R) = D$. Note that, by Lemma D.1, we can write

$$\Pi(R) = \{(t, \tau, q) \mid 0 < \tau < 1, q \in \mathbb{R}, t_0(\tau, q) < t < 0\},$$

for some function t_0 . It remains to show that $t_0(\tau, q) = t_{\text{cr}}(\tau, q)$. We first note that it by a simple but tedious computation one can show that $\sigma = a^2 bc$ indeed solves (2.1), for τ , t , q as defined in (3.4)–(3.6). It is also clear that σ depends continuously on a , b , c in the interior of R and for $b \rightarrow 0$ we find $\sigma \rightarrow 0$. Moreover, since the Jacobian of Π is not vanishing in the interior of R , we see that σ is even differentiable as a function of t . We thus proved that $\sigma(t) = a^2 bc$ for $t_0 \leq t < 0$. Finally, at the boundary the Jacobian vanishes and $\sigma(t) = a^2 bc$ is no longer differentiable as a function of t and thus no longer analytic in t . By definition, we must have $t_{\text{cr}} = t_0$. This finishes the proof. \blacksquare

E Calculations from physics: the genus 0 partition function

In this appendix, we study the genus 0 free energy, which is defined in terms of the partition function

$$Z_n(\tau, t, H; N) := \int_{\mathcal{H}_n} \int_{\mathcal{H}_n} \exp \left\{ \text{tr} \left[\tau XY - \frac{1}{2} X^2 - \frac{1}{2} Y^2 - \frac{te^H}{4N} X^4 - \frac{te^{-H}}{4N} Y^4 \right] \right\} dX dY. \quad (\text{E.1})$$

The (genus 0) free energy is then defined as the ratio

$$F(\tau, t, H) = \lim_{N \rightarrow \infty} \frac{1}{N^2} \log \frac{Z_N(\tau, t, H; N)}{Z_N(\tau, 0, 0; N)}.$$

More precisely, we reproduce the following results from the physics literature:

- (1) $Z_n(\tau, t, H; N)$ is a generating function for partition functions of the Ising model on genus g 4-regular graphs, in a sense we shall make precise. This is the result of Kazakov (for $H = 0$) and Kazakov and Boulatov ($H \neq 0$) [14, 57]. This is presented in Appendix E.1.
- (2) The large N limit of the genus 0 partition function admits an exact expression in terms of the solution to a certain implicit equation. This is also the result of [14, 57]. This is presented in Appendix E.2.
- (3) We check that the results of these calculations indeed agree. We calculate the first few terms of this generating function by hand, and show that this result agrees with the exact formula from [14, 57]. This is presented in Appendix E.3.
- (4) The asymptotics of the Taylor coefficients of $\sigma(\tau, t, H)$ are derived, and explicitly so for the special case $H = 0$, which is again a result of [57].

We present these results in more detail here, as the calculations are somewhat involved and the methods from the physics literature are perhaps unfamiliar to the target audience of this work. This appendix is meant to hopefully clarify these results; we do not claim any originality here. We also stress that these methods are formal: they are best thought of as guiding principles.

E.1 Wick expansion of the partition function

In this appendix, we replicate in detail the Wick expansion of the free energy, and show that it is a formal generating function for the Ising model on random 4-regular genus g graphs. We begin by considering the Gaussian model:

$$d\mathbb{P}(X, Y) = \frac{1}{Z_n(\tau, 0, 0; N)} \exp \left\{ N \text{tr} \left[\tau XY - \frac{1}{2} X^2 - \frac{1}{2} Y^2 \right] \right\} dX dY. \quad (\text{E.2})$$

Let $\langle \cdot \rangle_n$ denote the expected value with respect to this measure. Then, one can directly compute the following.

Lemma E.1. *The matrix elements X_{ij} , Y_{ij} have the following means and covariances:*

$$\begin{aligned} \langle X_{ij} \rangle_n &= 0, & \langle Y_{ij} \rangle_n &= 0, \\ \langle X_{ij} X_{kl} \rangle_n &= \langle Y_{ij} Y_{kl} \rangle_n = \frac{1}{N(1 - \tau^2)} \delta_{il} \delta_{jk}, & \langle X_{ij} Y_{kl} \rangle_n &= \langle Y_{ij} X_{kl} \rangle_n = \frac{\tau}{N(1 - \tau^2)} \delta_{il} \delta_{jk}. \end{aligned} \quad (\text{E.3})$$

Proof. We only sketch the idea of the proof here. The main point is that one can explicitly compute the moment generating function of this model. Let $J, K \in \mathcal{H}_n$ be a fixed Hermitian matrices, and consider the quantity

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

where the expected value is taken with respect to the measure (E.2). By the usual trick of completing the square, one can compute that

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

Taking derivatives with respect to the appropriate matrix elements of J, K yields (E.3). \blacksquare

Now, since the entries of both X, Y are all centered Gaussian variables, we can apply Wick's theorem to calculate the expected value of any higher order moment in terms of the covariances (E.3). Let us now explain how to calculate expected values of the form

$$\left\langle \prod_{p=1}^k \operatorname{tr} X^{i_p} \prod_{q=1}^{\ell} \operatorname{tr} Y^{j_q} \right\rangle_n, \quad (\text{E.4})$$

where (i_1, \dots, i_p) and (j_1, \dots, j_p) are sequences of positive integers. The main takeaway will be that one can interpret this expected value digrammatically; we shall now explain this procedure. Let $K = \sum_{p=1}^k i_p$, $L = \sum_{q=1}^{\ell} j_q$, and $E = K + L$. We assume that E is even; otherwise, this expected value is automatically zero. We can apply Wick's theorem to (E.4), and convert this expected value into a sum over all possible pairings of the indices:

$$\left\langle \prod_{p=1}^k \operatorname{tr} X^{i_p} \prod_{q=1}^{\ell} \operatorname{tr} Y^{j_q} \right\rangle_n = \sum_I \sum_{\pi \in \Pi_E} \prod_{\{(r_a, r_b), (r_c, r_d)\} \in \pi} \langle Z_{r_a r_b} Z_{r_c r_d} \rangle_n,$$

where Π_E denotes the set of pairings of E elements, the sum I runs over all of the indices of the traces $\operatorname{tr} X^{i_p}$, $p = 1, \dots, k$, and $\operatorname{tr} Y^{j_q}$, $q = 1, \dots, \ell$, and

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

Interchanging the order of summation, we see that each pairing π contributing a factor of

$$w[R_\pi] := \sum_I \prod_{\{(r_a, r_b), (r_c, r_d)\} \in \pi} \langle Z_{r_a r_b} Z_{r_c r_d} \rangle_n$$

to the expected value. To each such pairing π , we associate a 2-colored ribbon graph R_π as follows:

- (1) Draw k X -colored vertices, each with degree i_p , $p = 1, \dots, k$, 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 , $q = 1, \dots, \ell$, with half-edges labelled by the indices of the trace of the corresponding vertex.
- (3) Pair the half-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 $\ell := V - k$ of the second color, by $R_\pi(k, \ell)$. The *size* of such a ribbon graph is the number of vertices, denoted $|R_\pi(k, \ell)| = V$. We call such a coloring a *coloring of type* (k, ℓ) . The contribution of such a graph to the Wick sum we will denote by $w[R_\pi(k, \ell)]$, or by $w[R_\pi]$, when the coloring type is clear.

Let us compute more explicitly the contribution to the sum of a particular pairing (ribbon graph) π (R_π). The ribbon graph associated to the pairing π will by construction have V vertices,

and E edges, by the handshaking lemma. Each face in the resulting ribbon graph corresponds to a collection of indices which will be identified by Wick's theorem, since $\langle Z_{ab}Z_{cd} \rangle_n = \frac{\star}{n(1-\tau^2)} \delta_{ad}\delta_{bc}$, where Z is either X , Y , and \star is either 1 or τ . 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 contribution will be multiplied by a factor 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 $N := n$):

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

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

$$w[R_\pi] = \left(\frac{\tau^{1/2}}{1-\tau^2} \right)^E n^{\chi(R_\pi)-V} \tau^{\frac{1}{2}S}. \quad (\text{E.5})$$

Note that, since 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 proposition:

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

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

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

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

where the sum runs over all genus g unordered (k, ℓ) -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 unlike vertices.

Now, let us apply this proposition to the matrix integral (E.1). Expanding $\exp\left[-\frac{t}{4N} \text{tr } X^4\right]$, $\exp\left[-\frac{t}{4N} \text{tr } Y^4\right]$ as series in t , we obtain that

$$\begin{aligned} & \exp\left[-\frac{t}{4N} \text{tr } X^4\right] \exp\left[-\frac{t}{4N} \text{tr } Y^4\right] \\ &= \sum_{V=0}^{\infty} \frac{(-t/4N)^V}{V!} \sum_{k+j=V} \frac{V!}{k!j!} (\text{tr } X^4)^k (\text{tr } Y^4)^j e^{(k-j)H}. \end{aligned}$$

If we insert this expression into equation (E.1), and divide by the Gaussian partition function $Z_n(\tau, 0, 0; N)$, we obtain that

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

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

$$\frac{Z_n(\tau, t, H; N)}{Z_n(\tau, t, H; N)} = \sum_{V=0}^{\infty} \frac{(-t/4N)^V}{V!} \sum_{k+j=V} \frac{V!}{k!j!} \sum_{|R_\pi(k,j)|=V} w[R_\pi(k,j)] e^{(k-j)H},$$

where the innermost sum runs over all 2-colored ribbon graphs on V vertices with k edges of the first color and j of the second, $k + j = V$. 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 $k + j = V$, there are precisely $\frac{V!}{k!j!}$ possible colorings of type (k, j) on a given graph on V 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

$$\frac{Z_n(\tau, t, H; N)}{Z_n(\tau, 0, 0; N)} = \sum_{V=0}^{\infty} \frac{(-t/4N)^V}{V!} \sum_{|R_\pi|=V} w_h[R_\pi],$$

where we have defined a modified weight function

$$w_h[R_\pi] := w[R_\pi] e^{(k-j)H},$$

and the internal sum now runs over all *labelled*, 2-colored ribbon graphs on V -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

$$\frac{Z_n(\tau, t, H; N)}{Z_n(\tau, 0, 0; N)} = \sum_{V=0}^{\infty} \frac{(-t/4N)^V}{V!} \sum_{\substack{\text{ribbon} \\ \text{graphs} \\ R_\pi}} \sum_{\text{colorings of } R_\pi} w_h[R_\pi] e^{(k-j)H},$$

where the sum

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

is a sum over all *uncolored*, *labelled* ribbon graphs on V 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 (E.5)), and using the fact that $E = 2V$ by the handshaking lemma, we have that (again putting $N := n$):

$$\frac{Z_n(\tau, t, H; n)}{Z_n(\tau, 0, 0; n)} = \sum_{V=0}^{\infty} \frac{1}{V!} \left(\frac{-t\tau}{4(1-\tau^2)^2} \right)^V \sum_{\substack{\text{ribbon} \\ \text{graphs} \\ R_\pi}} n^{\chi(R_\pi)} \sum_{\text{colorings of } R_\pi} \tau^{\frac{1}{2}S(R_\pi)} e^{(k-j)H}.$$

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 $\tau = e^{-2\beta}$, $H = \beta h$. Thus, we have shown that

$$\frac{Z_n(\tau, t, H; n)}{Z_n(\tau, 0, 0; n)} = \sum_{V=0}^{\infty} \frac{1}{V!} \left(\frac{-t\tau}{4(1-\tau^2)^2} \right)^V \sum_{\substack{\text{ribbon} \\ \text{graphs} \\ R_\pi}} n^{\chi(R_\pi)} Z_{R_\pi}(\beta, h).$$

A general principle in the theory of generating functions is the following. Suppose 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. For more details pertaining to this fact, one may consult, for example, the book [78, Section 5], instead a sum over *connected* ribbon graphs:

$$\log \frac{Z_n(\tau, t, H; n)}{Z_n(\tau, 0, 0; n)} = \sum_{V=1}^{\infty} \frac{1}{V!} \left(\frac{-t\tau}{4(1-\tau^2)^2} \right)^V \sum_{\substack{\text{ribbon} \\ \text{graphs} \\ R_\pi}} ' n^{\chi(R_\pi)} Z_{R_\pi}(\beta, h),$$

where $'$ denotes the sum only over connected diagrams, and β, h are related to τ, H via the formulae

$$\tau = e^{-2\beta}, \quad H = \beta h.$$

In particular, if we divide through by n^2 , and take a limit as $n \rightarrow \infty$, we obtain that

$$F(\tau, t, H) := \lim_{n \rightarrow \infty} \frac{1}{n^2} \log \frac{Z_n(\tau, t, H; n)}{Z_n(\tau, 0, 0; n)} = \sum_{V=1}^{\infty} \frac{1}{V!} \left(\frac{-t\tau}{4(1-\tau^2)^2} \right)^V \sum_{\substack{\text{ribbon} \\ \text{graphs} \\ R_\pi}} ' Z_{R_\pi}(\beta, h), \quad (\text{E.6})$$

where the internal sum runs over all connected, 4-valent planar ribbon graphs on V vertices.

E.2 Formal calculation of the genus 0 partition function

Here we calculate the genus 0 partition function in the limit as $N := n$ tends to infinity. Define the function $V(z; t) := \frac{1}{2}z^2 + \frac{t}{4}z^4$, and put

$$W(x, y) := \tau xy - V(x, te^H) - V(y, te^{-H}).$$

We consider the family of biorthogonal polynomials defined by the formula

$$h_k(\tau, t, H; N) \delta_{kj} = h_k \delta_{kj} = \iint P_k(x) Q_k(y) e^{NW(x, y)} dx dy,$$

where the integration is carried out over an appropriately chosen contour, so that the above integral makes sense. The genus zero generating function for the Ising model on random spherical quadrangulations is then

$$F(\tau, t, H) := \lim_{N \rightarrow \infty} \frac{1}{N^2} \log \prod_{k=0}^N \frac{h_k(\tau, t, H; N)}{h_k(\tau, 0, 0; N)}.$$

In order to calculate $F(\tau, t, H)$, we must first study some properties of the corresponding biorthogonal polynomials. We have the following lemma.

Lemma E.3. *The polynomials $\{P_k(x)\}$, $\{Q_k(y)\}$ satisfy the recursion relations*

$$xP_k(x) = P_{k+1}(x) + R_k P_{k-1}(x) + S_k P_{k-3}(x), \quad (\text{E.7})$$

$$yQ_k(y) = Q_{k+1}(y) + \tilde{R}_k Q_{k-1}(y) + \tilde{S}_k Q_{k-3}(y). \quad (\text{E.8})$$

Proof. We sketch the proof of (E.7). Obviously, $xP_k(x) = P_{k+1}(x) + \sum_{j=0}^k c_{jk} P_j(x)$. Multiplying this relation by $Q_\ell(y)$ and integrating against the measure $e^{NW(x,y)} dx dy$, one obtains

$$c_{k\ell} h_\ell = \iint x P_k(x) Q_\ell(y) e^{NW(x,y)} dx dy.$$

Using the identity $xe^{NW(x,y)} = \frac{1}{N\tau} \left(\frac{\partial}{\partial y} e^{NW(x,y)} \right) + \frac{1}{\tau} (y + te^{-H} y^3)$, one sees that

$$c_{k\ell} h_\ell = \iint P_k(x) Q_\ell(y) \left[\frac{1}{N\tau} \left(\frac{\partial}{\partial y} e^{NW(x,y)} \right) + \frac{1}{\tau} (y + te^{-H} y^3) \right] dx dy.$$

Since $\ell < k+1$, the first term vanishes identically, since integration by parts yields the integrand as $-P_k(x) Q'_\ell(y) e^{NW(x,y)}$, and $Q'_\ell(y)$ is a polynomial of degree less than k . Similarly, last two terms can only be nonzero if $\ell = k, k-1, k-3$; the k term is also zero, by the symmetry of the integration measure upon the change of variables $(x, y) \rightarrow (-x, -y)$. The result (E.8) is obtained in an identical manner. ■

We now show that the recursion coefficients satisfy additional relations.

Lemma E.4. *Define $f_k := h_k/h_{k-1}$. The coefficients $R_k, S_k, \tilde{R}_k, \tilde{S}_k$ satisfy the following relations:*

$$\tau S_k = te^{-H} f_k f_{k-1} f_{k-2}, \quad (\text{E.9})$$

$$\tau R_k = f_k [1 + te^{-H} (\tilde{R}_{k+1} + \tilde{R}_k + \tilde{R}_{k-1})], \quad (\text{E.10})$$

$$\frac{k}{N} = -\tau f_k + \tilde{R}_k + te^{-H} [\tilde{S}_{k+2} + \tilde{S}_{k+1} + \tilde{S}_k + \tilde{R}_k (\tilde{R}_{k+1} + \tilde{R}_k + \tilde{R}_{k-1})], \quad (\text{E.11})$$

$$\tau \tilde{S}_k = te^H f_k f_{k-1} f_{k-2}, \quad (\text{E.12})$$

$$\tau \tilde{R}_k = f_k [1 + te^H (R_{k+1} + R_k + R_{k-1})], \quad (\text{E.13})$$

$$\frac{k}{N} = -\tau f_k + R_k + te^H [S_{k+2} + S_{k+1} + S_k + R_k (R_{k+1} + R_k + R_{k-1})]. \quad (\text{E.14})$$

Proof. We only prove the first relation; the last three are proven identically. Multiplying equation (E.7) by $\tau Q_{k-3}(y)$, and integrating with respect to the measure $e^{NW(x,y)} dx dy$, we find that

$$\tau S_k h_{k-3} = \iint \tau x P_k(x) Q_{k-3}(y) e^{NW(x,y)} dx dy.$$

Using the identity $\tau x e^{NW(x,y)} = \frac{1}{N} \left(\frac{\partial}{\partial y} e^{NW(x,y)} \right) + (y + te^{-H} y^3)$, we see that

$$\tau S_k h_{k-3} = \iint \tau P_k(x) Q_{k-3}(y) \left[\frac{1}{N} \left(\frac{\partial}{\partial y} e^{NW(x,y)} \right) + (y + te^{-H} y^3) \right] dx dy.$$

The first term drops out upon integrating by parts; similarly, since $\deg[yQ_{k-3}] = k-2$, the second term also evaluates to 0. On the last term, we use iteratively the recursion formula (E.8):

$$\begin{aligned} \tau S_k h_{k-3} &= te^{-H} \iint P_k(x) [Q_k(y) + \text{lower degree polynomials}] e^{NW(x,y)} dx dy \\ &= te^{-H} h_k. \end{aligned} \quad \blacksquare$$

In planar limit, we assume the scaling $\frac{k}{N} \rightarrow \lambda$, $f_k \rightarrow f(\lambda)$, and so on. The above formulae (E.9)–(E.14) imply that, in the limit as $N \rightarrow \infty$,

$$\begin{aligned}\tau S(\lambda) &= te^{-H} f(\lambda)^3, \\ \tau R(\lambda) &= f(\lambda) [1 + 3te^{-H} \tilde{R}(\lambda)], \\ \lambda &= -\tau f(\lambda) + \tilde{R}(\lambda) + 3te^{-H} [\tilde{S}(\lambda) + \tilde{R}(\lambda)^2],\end{aligned}\tag{E.15}$$

$$\begin{aligned}\tau \tilde{S}(\lambda) &= te^H f(\lambda)^3, \\ \tau \tilde{R}(\lambda) &= f(\lambda) [1 + 3te^H R(\lambda)], \\ \lambda &= -\tau f(\lambda) + R(\lambda) + 3te^H [S(\lambda) + R(\lambda)^2],\end{aligned}\tag{E.16}$$

and these relations we can solve for $S(\lambda)$, $R(\lambda)$, $\tilde{R}(\lambda)$, and $\tilde{S}(\lambda)$ in terms of $f(\lambda)$, and the parameters τ , t , H :

$$\begin{aligned}S(\lambda) &= \frac{te^{-H}}{\tau} f(\lambda)^3, & \tilde{S}(\lambda) &= \frac{te^H}{\tau} f(\lambda)^3, \\ R(\lambda) &= \frac{\tau + 3te^{-H} f(\lambda)}{\tau^2 - 9t^2 f(\lambda)^2} f(\lambda), & \tilde{R}(\lambda) &= \frac{\tau + 3te^H f(\lambda)}{\tau^2 - 9t^2 f(\lambda)^2} f(\lambda).\end{aligned}$$

These formulae can then be subsequently substituted into either equations (E.15) or (E.16) (they yield the same equation in the end) to obtain an implicit formula for the function $f(\lambda) = f(\lambda; \tau, t, H)$:

$$\begin{aligned}\lambda &= -\tau f(\lambda) + \frac{\tau + 3te^{-H} f(\lambda)}{\tau^2 - 9t^2 f(\lambda)^2} f(\lambda) + \frac{3te^H (\tau + 3te^{-H} f(\lambda))^2}{(\tau^2 - 9t^2 f(\lambda)^2)^2} f(\lambda)^2 + \frac{3t^2 f(\lambda)^3}{\tau} \\ \iff \lambda &= -\tau f(\lambda) + \frac{3t^2}{\tau} f(\lambda)^3 + \frac{\tau f(\lambda)}{(\tau - 3t f(\lambda))^2} + \frac{6\tau^2 t f(\lambda)^2}{(\tau^2 - 9t^2 f(\lambda)^2)^2} [\cosh H - 1].\end{aligned}\tag{E.17}$$

In particular, when $t = H = 0$, we obtain the solution

$$f(\lambda; \tau, 0, 0) = \frac{\tau \lambda}{1 - \tau^2}.$$

Returning again to the calculation of the function $F(\tau, t, H)$, note that, by the definition of the variables f_k , $\prod_{k=0}^N h_k = h_0^N \prod_{k=1}^N f_k^{N-k}$, and so

$$\frac{1}{N^2} \log \prod_{k=0}^N \frac{h_k(\tau, t, H; N)}{h_k(\tau, 0, 0; N)} = \frac{1}{N} \log \frac{h_0(\tau, t, H; N)}{h_0(\tau, 0, 0; N)} + \sum_{k=1}^N \left(1 - \frac{k}{N}\right) \log \frac{f_k(\tau, t, H; N)}{f_k(\tau, 0, 0; N)} \frac{1}{N}.$$

Taking the limit as $N \rightarrow \infty$, the first term vanishes (the argument of the logarithm is bounded), and the second term approximates a Riemann integral; we thus obtain the formula

$$\begin{aligned}F(\tau, t, H) &= \lim_{N \rightarrow \infty} \frac{1}{N^2} \log \prod_{k=0}^N \frac{h_k(\tau, t, H; N)}{h_k(\tau, 0, 0; N)} = \int_0^1 (1 - \lambda) \log \frac{f(\lambda; \tau, t, H)}{f(\lambda; \tau, 0, 0)} d\lambda \\ &= \frac{3}{4} - \frac{1}{2} \log \frac{\tau}{1 - \tau^2} + \int_0^1 (1 - \lambda) \log f(\lambda; \tau, t, H) d\lambda.\end{aligned}\tag{E.18}$$

This, along with the implicit formula (E.17), are enough (in principle) to determine the function $F(\tau, t, H)$. Integrating the above formula by parts,

$$F(\tau, t, H) = \frac{3}{4} - \frac{1}{2} \log \frac{\tau}{1 - \tau^2} + \left[\left(\lambda - \frac{1}{2} \lambda^2 \right) \log f(\lambda; \tau, t, H) \right]_0^1$$

$$\begin{aligned}
& - \int_0^1 \left(\lambda - \frac{1}{2} \lambda^2 \right) \frac{f'(\lambda; \tau, t, H)}{f(\lambda; \tau, t, H)} d\lambda \\
& = \frac{3}{4} - \frac{1}{2} \log \frac{\tau}{1 - \tau^2} + \frac{1}{2} \log f(1; \tau, t, H) - \int_0^1 \left(\lambda - \frac{1}{2} \lambda^2 \right) \frac{f'(\lambda; \tau, t, H)}{f(\lambda; \tau, t, H)} d\lambda
\end{aligned}$$

(Here we have used the fact that $f(0; \tau, t, H) \equiv 0$). Making the change of variables $u = f(\lambda; \tau, t, H)$, $\frac{du}{u} = \frac{f'}{f} d\lambda$,

$$F(\tau, t, H) = \frac{3}{4} - \frac{1}{2} \log \frac{\tau}{1 - \tau^2} + \frac{1}{2} \log f(1; \tau, t, H) - \int_0^{f(1; \tau, t, H)} \left(\tilde{\lambda}(u) - \frac{1}{2} \tilde{\lambda}(u)^2 \right) \frac{du}{u},$$

where $\tilde{\lambda}(u)$ is defined by use of the implicit equation (E.17):

$$\tilde{\lambda}(u) = -\tau u + \frac{3t^2}{\tau} u^3 + \frac{\tau u}{(\tau - 3tu)^2} + \frac{6\tau^2 t u^2}{(\tau^2 - 9t^2 u^2)^2} [\cosh H - 1].$$

The integration can be carried out directly, since $\lambda(u)$ is a rational function. Finally, defining

$$\sigma(\tau, t, H) := -\frac{3t}{\tau} f(1; \tau, t, H),$$

and making an appropriate change of variables, we obtain the following Proposition:

Proposition E.5. *The genus 0 partition function admits the expression*

$$F(\tau, t, H) = \frac{3}{4} + \frac{1}{2} \log \frac{(1 - \tau^2) \sigma(\tau, t, H)}{-3t} - \int_0^{\sigma(\tau, t, H)} \left(\lambda(u) - \frac{1}{2} \lambda(u)^2 \right) \frac{du}{u},$$

where $\lambda(u)$ is the rational function

$$\lambda(u) = -\frac{1}{t} \left[\frac{1}{9} \tau^2 u (u^2 - 3) + \frac{1}{3} \frac{u}{(u + 1)^2} - \frac{2}{3} \left(\frac{u}{u^2 - 1} \right)^2 [\cosh H - 1] \right]$$

and $\sigma = \sigma(\tau, t, H)$ is defined implicitly by the equation

$$t = -\frac{1}{9} \tau^2 \sigma (\sigma^2 - 3) - \frac{1}{3} \frac{\sigma}{(1 + \sigma)^2} + \frac{2}{3} \left(\frac{\sigma}{1 - \sigma^2} \right)^2 [\cosh H - 1].$$

This is the result originally obtained by Kazakov ($H = 0$, cf. [57]) and Boulatov ($H \neq 0$, cf. [14]). If one expands the function $F(\tau, t, H)$ as a series in t , one obtains that

$$\begin{aligned}
F(\tau, t, H) &= 4\tau^{-1} \cosh H \left(\frac{-t\tau}{4(1 - \tau^2)^2} \right) + (8\tau^2 + 64 + 72\tau^{-2} \cosh(2H)) \frac{1}{2} \left(\frac{-t\tau}{4(1 - \tau^2)^2} \right)^2 \\
&+ \left(\frac{3456}{\tau^3} \cosh(H) (2 \cosh(2H) + \tau^4 + 2\tau^2 - 1) \right) \frac{1}{6} \left(\frac{-t\tau}{4(1 - \tau^2)^2} \right)^3 + \mathcal{O}(t^4). \quad (\text{E.19})
\end{aligned}$$

E.3 Comparison with explicit calculation

On the other hand, we can compute the first few contributions to the genus 0 free energy by hand. The following graphs (shown in Figure 16) contribute at genus zero:

$V = 1$. For one vertex, there is one type of graph contributing, shown in the figure. The partition function for the Ising model on this graph is, in the variables τ, H ,

$$Z_1^{(1)}(\tau, H) = 2\tau^{-1} \cosh H.$$

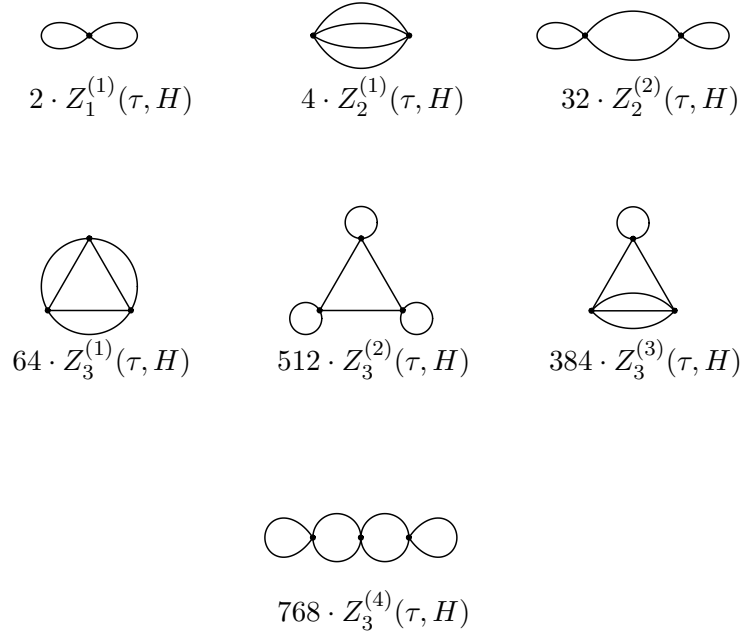


Figure 16. The isomorphism classes of graphs contributing at genus 0 to the free energy, up to $V = 3$ vertices. Each isomorphism class of graph contributes a factor of $N_i \cdot Z_V^{(i)}(\tau, H)$.

$V = 2$. For two vertices, there are two nonisomorphic types of graphs contributing, shown in the figure. Their corresponding partition functions are

$$Z_2^{(1)}(\tau, H) = 2\tau^{-2} \cosh(2H) + 2\tau^2, \quad Z_2^{(2)}(\tau, H) = 2\tau^{-2} \cosh(2H) + 2.$$

$V = 3$. There are four nonisomorphic types of graphs contributing, shown in the figure. Their partition functions are

$$\begin{aligned} Z_3^{(1)}(\tau, H) &= 2\tau^{-3} \cosh(3H) + 6\tau \cosh(H), \\ Z_3^{(2)}(\tau, H) &= 2\tau^{-3} \cosh(3H) + 2\tau^{-1} \cosh(H) + 4\tau \cosh(H), \\ Z_3^{(3)}(\tau, H) &= 2\tau^{-3} \cosh(3H) + 6\tau^{-1} \cosh(H), \\ Z_3^{(4)}(\tau, H) &= 2\tau^{-3} \cosh(3H) + 4\tau^{-1} \cosh(H) + 2\tau \cosh(H). \end{aligned}$$

Observe that we can write the genus zero free energy (E.6) equivalently as

$$F(\tau, t, H) = \sum_{V=0}^{\infty} \left(\sum_{|G|=V} N_G Z_G(\tau, H) \right) \frac{1}{V!} \left(\frac{-t\tau}{4(1-\tau^2)^2} \right)^V,$$

where the internal sum is taken over *isomorphism classes* of graphs G on V vertices, and N_G counts the number of labelled graphs isomorphic to G when one “forgets” the labelling. In this way, one can explicitly compute that

$$\begin{aligned} F(\tau, t, H) &= [2 \cdot Z_1^{(1)}(\tau, H)] \left(\frac{-t\tau}{4(1-\tau^2)^2} \right) \\ &\quad + [4 \cdot Z_2^{(1)}(\tau, H) + 32 \cdot Z_2^{(2)}(\tau, H)] \frac{1}{2} \left(\frac{-t\tau}{4(1-\tau^2)^2} \right)^2 \\ &\quad + [64 \cdot Z_3^{(1)}(\tau, H) + 512 \cdot Z_3^{(2)}(\tau, H) + 384 \cdot Z_3^{(3)}(\tau, H) + 768 \cdot Z_3^{(4)}(\tau, H)] \end{aligned}$$

$$\begin{aligned}
& \times \frac{1}{6} \left(\frac{-t\tau}{4(1-\tau^2)^2} \right)^3 + \mathcal{O}(t^4) \\
& = 4\tau^{-1} \cosh H \left(\frac{-t\tau}{4(1-\tau^2)^2} \right) + (8\tau^2 + 64 + 72\tau^{-2} \cosh(2H)) \frac{1}{2} \left(\frac{-t\tau}{4(1-\tau^2)^2} \right)^2 \\
& \quad + \left(\frac{3456}{\tau^3} \cosh(H) (2 \cosh(2H) + \tau^4 + 2\tau^2 - 1) \right) \frac{1}{6} \left(\frac{-t\tau}{4(1-\tau^2)^2} \right)^3 + \mathcal{O}(t^4),
\end{aligned}$$

which is in agreement with the formula (E.19) originally obtained by Kazakov.

Remark E.6. *High/low temperature limits of the free energy.* As a final remark, let us show that in the limit $\tau, H \rightarrow 0$, the free energy of the quartic-two matrix model reduces to twice the free energy of the quartic 1-matrix model. Intuitively, one should expect this result because τ acts as a coupling parameter between the matrices X and Y ; when $\tau, H \rightarrow 0$, we obtain two independent copies of the quartic 1-matrix model. Recall that the free energy for the quartic 1-matrix model is

$$\mathcal{F}(t) := \lim_{N \rightarrow \infty} \log \frac{Z_N^{(1)}(t; N)}{Z_N^{(1)}(0; N)},$$

where $Z_n^{(1)}(t; N)$ is given by the formula

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

Explicitly, one may evaluate $\mathcal{F}(t)$ to be (cf., for example, [11], or the more recent [13])

$$\mathcal{F}(t) = \int_0^1 (1-\lambda) \log \frac{-1 + \sqrt{1 + 12t\lambda}}{6t\lambda} d\lambda.$$

On the other hand, let us return to the formula for the partition function of the 2-matrix model (see the first expression in equation (E.18)):

$$\begin{aligned}
F(\tau, t, H) &= \int_0^1 (1-\lambda) \log \frac{f(\lambda; \tau, t, H)}{f(\lambda; \tau, 0, 0)} d\lambda \\
&= \int_0^1 (1-\lambda) \log \frac{(1-\tau^2)f(\lambda; \tau, t, H)}{\tau\lambda} d\lambda,
\end{aligned}$$

where $f(\lambda; \tau, t, H)$ is defined by the implicit equation

$$\lambda = -\tau f(\lambda) + \frac{\tau + 3te^{-H}f(\lambda)}{\tau^2 - 9t^2 f(\lambda)^2} f(\lambda) + \frac{3te^H(\tau + 3te^{-H}f(\lambda))^2}{(\tau^2 - 9t^2 f(\lambda)^2)^2} f(\lambda)^2 + \frac{3t^2 f(\lambda)^3}{\tau}. \quad (\text{E.20})$$

Define the new variable $\tilde{f}(\lambda; \tau, t, H) := \frac{(1-\tau^2)}{\tau} f(\lambda; \tau, t, H)$. The reason for this change of variables is that $\tilde{f}(\lambda; \tau, t, H) \rightarrow \lambda$ when $t \rightarrow 0$, i.e., the limit function is independent of τ . Under this change of variables, equation (E.20) becomes

$$0 = \lambda - \frac{(1 + 3te^{-H}\tilde{f})\tilde{f}}{1 - 9t^2\tilde{f}^2} - \frac{3te^H(1 + 3te^{-H}\tilde{f})^2}{(1 - 9t^2\tilde{f}^2)^2} + \mathcal{O}(\tau^2).$$

In the limit as $\tau, H \rightarrow 0$, one sees that

$$0 = \lambda - \frac{(1 + 3t\tilde{f})\tilde{f}}{1 - 9t^2\tilde{f}^2} - \frac{3t(1 + 3t\tilde{f})^2}{(1 - 9t^2\tilde{f}^2)^2}.$$

This equation can be solved for \tilde{f} :

$$\tilde{f}(t) = \frac{1 + 6t\lambda - \sqrt{1 + 12t\lambda}}{18t^2\lambda},$$

where we have taken the branch of the solution which tends to λ as $t \rightarrow 0$. Inserting this expression into the expression for the free energy of the 2-matrix model,

$$\begin{aligned} \lim_{\tau, H \rightarrow 0} F(\tau, t, H) &= \lim_{\tau, H \rightarrow 0} \int_0^1 (1 - \lambda) \log \frac{\tilde{f}(\tau, t, H)}{\lambda} d\lambda \\ &= \int_0^1 (1 - \lambda) \log \frac{1 + 6t\lambda - \sqrt{1 + 12t\lambda}}{18t^2\lambda^2} d\lambda \\ &= \int_0^1 (1 - \lambda) \log \left(\frac{-1 + \sqrt{1 + 12t\lambda}}{6t\lambda} \right)^2 d\lambda \\ &= 2 \int_0^1 (1 - \lambda) \log \frac{-1 + \sqrt{1 + 12t\lambda}}{6t\lambda} d\lambda \\ &= 2\mathcal{F}(t). \end{aligned}$$

More generally, we have the following limits:

$$\begin{aligned} \lim_{H \rightarrow 0} \lim_{\tau \rightarrow 0} F(\tau, (1 - \tau^2)t, H) &= \lim_{H \rightarrow 0} [\mathcal{F}(e^H t) + \mathcal{F}(e^{-H} t)] = 2\mathcal{F}(t), \\ \lim_{H \rightarrow 0} \lim_{\tau \rightarrow 1} F(\tau, (1 - \tau^2)t, H) &= \lim_{H \rightarrow 0} [\mathcal{F}(2e^H t) + \mathcal{F}(2e^{-H} t)] = 2\mathcal{F}(2t). \end{aligned}$$

The first limit follows from our previous considerations, without setting $H = 0$. The replacement of t by $(1 - \tau^2)t$ here is natural from the point of view of the graphical expansion of the 2-matrix model. The second limit follows from identical calculations. These limits correspond to the low and high temperature limits of the Ising model. As originally pointed out by Kazakov [57], the factor of 2 appearing in the $\tau \rightarrow 1$ limit comes from the fact that, for a graph G on V vertices, the partition function for the Ising model on that graph (recalling that $\tau = e^{-2\beta}$)

$$\lim_{\beta \rightarrow 0} Z_G(\beta, h) \rightarrow 2^V.$$

Thus, in the limit as $\tau \rightarrow 1$, each planar, 4-regular graph on V vertices should contribute the same factor of 2^V to the free energy.

E.4 Large V asymptotics of $\sigma(\tau, t, H)$

Here we calculate the large V asymptotics of the implicitly defined function $\sigma(\tau, t, H)$. The reason for these calculations is the following: in the works [14, 57], it is argued that the asymptotics of the Taylor coefficients of $F(\tau, t, H)$ are dominated by the nearest critical point to the origin of the equation $\mathfrak{I}(\sigma; \tau, t, H) = 0$. The authors make the assertion that the Taylor coefficients of F behave like $F_V \sim \text{const} \cdot V^\alpha [g(\tau, H)]^V$; it is eventually the expression $g(\tau, H)$ that they are interested in. Indeed, calculation of $g(\tau, H)$ is equivalent to finding the radius of convergence of F , considered as an analytic function of t . Since F is defined in terms of $\sigma(\tau, t, H)$, it is therefore plausible that this radius of convergence should coincide with the radius of convergence of $\sigma(\tau, t, H)$ as a function of t . However, we were unfortunately unable to show that this was the case, as it is not obvious that terms such as $\frac{1}{(\sigma+1)^3}$, $\log \frac{(1-\tau^2)\sigma}{-3t}$ do not alter the radius of convergence of F (the remaining terms in the expression for F are polynomial in σ , and thus have the same radius of convergence as that of σ). We therefore instead have opted to simply investigate the Taylor coefficients of $\sigma(\tau, t, H)$ as a function of t , and take the assumptions Kazakov makes about the implication of this function as fact.

Let us now set about calculating the Taylor coefficients of $\sigma(\tau, t, H)$ as a function of t . We do so explicitly in the case when $H = 0$, and also provide an expansion which holds as $H \rightarrow 0$. Recall that $\sigma = \sigma(\tau, t, H)$ is defined by the equation

$$\mathfrak{I}(\sigma; \tau, t, H) = -\frac{1}{9}\tau^2\sigma(\sigma^2 - 3) - \frac{1}{3}\frac{\sigma}{(1+\sigma)^2} + \frac{2}{3}\left(\frac{\sigma}{1-\sigma^2}\right)^2 [\cosh H - 1] - t = 0.$$

When $H = 0$, the last term vanishes identically. Note that, for any fixed τ, H , the above implicit equation is of special form: it is actually an equation defining an *inverse* function to $t(\sigma)$. We recall the following theorem, due to Lagrange (cf., for instance, [82, p. 149, Exercise 25]):

Theorem E.7 (Lagrange inversion formula). *Suppose $t = G(\sigma)$ is analytic in a neighborhood of $\sigma = 0$, and suppose $G(0) = 0$, $G'(0) \neq 0$. Then, an inverse function $\sigma = \sigma(t)$ exists in a neighborhood of $t = 0$, and is given by the power series*

$$\sigma(t) = \sum_{V=1}^{\infty} \sigma_V \frac{t^V}{V!},$$

where

$$\sigma_V = \lim_{\sigma \rightarrow 0} \frac{d^{V-1}}{d\sigma^{V-1}} \left(\frac{\sigma^V}{G(\sigma)^V} \right). \quad (\text{E.21})$$

The proof of this theorem follows almost immediately from the Cauchy integral formula. In our situation, for any fixed $0 < \tau < 1$, and $H \in \mathbb{R}$, we have that

$$\mathfrak{I}(0; \tau, t, H) = 0, \quad \mathfrak{I}_\sigma(0; \tau, t, H) = \frac{1}{3}(\tau^2 - 1) \neq 0,$$

and so we can develop a Taylor series expansion of $\sigma(\tau, t, H)$ about $t = 0$, by taking $G(\sigma)$ to be

$$G(\sigma) := \mathfrak{I}(\sigma; \tau, 0, H) = -\frac{1}{9}\tau^2\sigma(\sigma^2 - 3) - \frac{1}{3}\frac{\sigma}{(1+\sigma)^2} + \frac{2}{3}\left(\frac{\sigma}{1-\sigma^2}\right)^2 [\cosh H - 1].$$

By the Cauchy integral formula, along with formula (E.21), we can write

$$\sigma_V(\tau, H) = \frac{1}{2\pi i} \oint_{C_0} \frac{d\zeta}{\mathfrak{I}(\zeta; \tau, 0, H)^V} = \frac{1}{2\pi i} \oint_{C_0} e^{-V \log \mathfrak{I}(\zeta; \tau, 0, H)} d\zeta, \quad (\text{E.22})$$

where C_0 is a sufficiently small positively oriented circle enclosing the origin. We are interested in the large- V asymptotics of $\sigma_V(\tau, H)$. When $H = 0$, and $\tau \neq \frac{1}{4}$, these asymptotics can be calculated explicitly using classical steepest descent analysis (cf., for example, [71]). When $H = 0$, the saddle points of the integrand are

$$\zeta = 1, \quad -1 \pm \tau^{-1/2}, \quad -1 \pm i\tau^{-1/2}.$$

The dominant saddle point ζ^* is the one nearest to the origin. Thus, we see that

$$\zeta^*(\tau) = \begin{cases} 1, & 0 < \tau < \frac{1}{4}, \\ \tau^{-1/2} - 1, & \frac{1}{4} < \tau < 1. \end{cases}$$

(Note that this change in the dominant saddle point is ultimately the source of the phase transition). It follows that

$$\sigma_V(\tau, 0) = \frac{\pm}{2\pi i} \sqrt{\frac{2\pi}{V[-\log \mathfrak{I}(\zeta^*; \tau, 0, 0)]_{\zeta\zeta}}} \left[\frac{1}{\mathfrak{I}(\zeta^*; \tau, 0, 0)} \right]^V [1 + \mathcal{O}(V^{-1/2})],$$

where the sign \pm is chosen appropriately in accordance with the contour of steepest descent. This addresses the asymptotics of σ_V away from the critical point $\tau = \frac{1}{4}$. In fact, we can actually say more. We summarize the above result, as well as an extension of it to a neighborhood of the critical point, in the following proposition:

Proposition E.8. *Define $t_{\text{low}}(\tau) := -\frac{1}{12} + \frac{2}{9}\tau^2$, $t_{\text{high}}(\tau) = -\frac{2}{9}\sqrt{\tau}(\sqrt{\tau}-1)^2(\sqrt{\tau}+2)$. As $V \rightarrow \infty$, for any fixed $0 < \tau < 1$, $\tau \neq \frac{1}{4}$,*

$$\sigma_V(\tau, 0) = \begin{cases} \frac{1}{\sqrt{3\pi V}} \sqrt{\frac{8\tau^2 - 3}{16\tau^2 - 1}} t_{\text{low}}(\tau)^{-V} [1 + \mathcal{O}(V^{-1/2})], & 0 < \tau < \frac{1}{4}, \\ \frac{(\sqrt{\tau} - 1)}{2\sqrt{3\pi V}} \sqrt{\frac{2 + \sqrt{\tau}}{\tau(2\sqrt{\tau} - 1)}} t_{\text{high}}(\tau)^{-V} [1 + \mathcal{O}(V^{-1/2})], & \frac{1}{4} < \tau < 1. \end{cases} \quad (\text{E.23})$$

If τ belongs to a sufficiently small neighborhood of $\tau = \frac{1}{4}$, then we have the uniform in τ asymptotic

$$\sigma_V(\tau, 0) = [t_{\text{low}}(\tau)t_{\text{high}}(\tau)]^{-V/2} \left[\frac{a_0(\tau)}{V^{1/3}} \text{Ai}(V^{2/3}s(\tau)) + \frac{b_0(\tau)}{V^{2/3}} \text{Ai}'(V^{2/3}s(\tau)) \right] \times [1 + \mathcal{O}(V^{-1/3})],$$

Where $\text{Ai}(z)$ denotes the Airy function, $s(\tau) \geq 0$ is the continuous function

$$s(\tau) = \frac{3}{4} \left| \log \left[\frac{t_{\text{low}}(\tau)}{t_{\text{high}}(\tau)} \right] \right|^{2/3}, \quad (\text{E.24})$$

satisfying $s(\frac{1}{4}) = 0$, and $a_0(\tau)$, $b_0(\tau)$ are given by

$$a_0(\tau) = \frac{s(\tau)^{1/4}}{2} \left[\frac{2}{\sqrt{3}} \sqrt{\frac{3 - 8\tau^2}{1 - 16\tau^2}} + \sqrt{\frac{(1 - \sqrt{\tau})^2(2 + \sqrt{\tau})}{3\tau(1 - 2\sqrt{\tau})}} \right],$$

$$b_0(\tau) = \frac{s(\tau)^{-1/4}}{2} \left[\frac{2}{\sqrt{3}} \sqrt{\frac{3 - 8\tau^2}{1 - 16\tau^2}} - \sqrt{\frac{(1 - \sqrt{\tau})^2(2 + \sqrt{\tau})}{3\tau(1 - 2\sqrt{\tau})}} \right].$$

Proof. As discussed above, when $\tau \neq \frac{1}{4}$, all of the saddle points are simple, and classical steepest descent analysis yields the expansion (E.23). However, when $\tau \rightarrow \frac{1}{4}$, the saddle points $\zeta = 1$, $\zeta = -1 + \tau^{-1/2}$ coalesce, and the expansions in the low and high temperature regimes are no longer valid. Indeed, upon direct inspection of the expansions (E.23), one sees that the leading terms in both expressions diverge as $\tau \rightarrow \frac{1}{4}$. In principle, one can find an asymptotic expansion which holds precisely at $\tau = \frac{1}{4}$. However, we can find a more robust expansion, which holds uniformly for τ sufficiently close to $\frac{1}{4}$; this kind of asymptotic expansion was first demonstrated in [25]. The proof relies on the construction of a family of conformal maps $\zeta \mapsto u(\zeta; \tau)$, which satisfy the algebraic equation

$$-\log \mathfrak{I}(\zeta; \tau, 0, 0) = \frac{1}{3}u^3 - s(\tau)u + C(\tau). \quad (\text{E.25})$$

Here s , C are determined by the equations

$$-\log \mathfrak{I}(1; \tau, 0, 0) = -\frac{2}{3}s(\tau)^{3/2} + C(\tau),$$

$$-\log \mathfrak{I}(-1 + \tau^{-1/2}; \tau, 0, 0) = +\frac{2}{3}s(\tau)^{3/2} + C(\tau).$$

(note that these conditions guarantee that $\frac{du}{d\zeta} \neq 0, \infty$). One finds that

$$C(\tau) = -\frac{1}{2} \log[t_{\text{low}}(\tau)t_{\text{high}}(\tau)],$$

$$s(\tau) = \left((-1)^{\mathbf{1}_{\{\tau > \frac{1}{4}\}}} \frac{3}{4} \log \left[\frac{t_{\text{low}}(\tau)}{t_{\text{high}}(\tau)} \right] \right)^{2/3}.$$

(Note that $\frac{t_{\text{low}}}{t_{\text{high}}} > 1$ for $0 < \tau < \frac{1}{4}$, and is smaller than 1 for $\frac{1}{4} < \tau < 1$, and so the above expression for S is equivalent to (E.24)). Provided that τ is sufficiently close to $\frac{1}{4}$, $C(\tau)$, $s(\tau)$ are positive and real-valued. Following [25], we can then rewrite the integral (E.22) as

$$\sigma_V(\tau, 0) = \frac{e^{VC(\tau)}}{2\pi i} \oint_{\tilde{C}} e^{V[\frac{1}{3}u^3 - s(\tau)u]} \frac{d\zeta}{du} du,$$

where \tilde{C} is the image of the circle C under the conformal map $u(\zeta)$. Now, we can express $\frac{d\zeta}{du}$ as

$$\frac{d\zeta}{du} = \sum a_k(\tau)(u^2 - s(\tau))^k + \sum b_k(\tau)u(u^2 - s(\tau))^k;$$

the only terms we calculate explicitly are $a_0(\tau)$, $b_0(\tau)$, as they will be the dominant terms in the expansion. Differentiating (E.25), and expanding using L'Hôpital's rule, one can determine that

$$\left(\frac{d\zeta}{du} \Big|_{u=s^{1/2}} \right)^2 = \frac{4}{3} \frac{3-8\tau^2}{1-16\tau^2} s(\tau)^{1/2},$$

$$\left(\frac{d\zeta}{du} \Big|_{u=-s^{1/2}} \right)^2 = \frac{(1-\sqrt{\tau})^2(2+\sqrt{\tau})}{3\tau(1-2\sqrt{\tau})} s(\tau)^{1/2},$$

and so these derivatives are determined up to a sign, which can be determined for instance by requiring the asymptotic expansion we are developing match with classical steepest descent results precisely at $\tau = \frac{1}{4}$. This determines $a_0(\tau)$, $b_0(\tau)$ to be

$$a_0(\tau) = \frac{s(\tau)^{1/4}}{2} \left[\frac{2}{\sqrt{3}} \sqrt{\frac{3-8\tau^2}{1-16\tau^2}} + \sqrt{\frac{(1-\sqrt{\tau})^2(2+\sqrt{\tau})}{3\tau(1-2\sqrt{\tau})}} \right],$$

$$b_0(\tau) = \frac{s(\tau)^{-1/4}}{2} \left[\frac{2}{\sqrt{3}} \sqrt{\frac{3-8\tau^2}{1-16\tau^2}} - \sqrt{\frac{(1-\sqrt{\tau})^2(2+\sqrt{\tau})}{3\tau(1-2\sqrt{\tau})}} \right].$$

These functions are indeed continuous, as one can readily compute the left/right limits as $\tau \rightarrow \frac{1}{4}$. One finally obtains that

$$\sigma_V(\tau, 0) = e^{VC(\tau)} \left[\frac{a_0(\tau)}{V^{1/3}} \text{Ai}(V^{2/3}s(\tau)) + \frac{b_0(\tau)}{V^{2/3}} \text{Ai}'(V^{2/3}s(\tau)) \right] [1 + \mathcal{O}(V^{-1/3})]$$

$$= [t_{\text{low}}(\tau)t_{\text{high}}(\tau)]^{-V/2} \left[\frac{a_0(\tau)}{V^{1/3}} \text{Ai}(V^{2/3}s(\tau)) + \frac{b_0(\tau)}{V^{2/3}} \text{Ai}'(V^{2/3}s(\tau)) \right] [1 + \mathcal{O}(V^{-1/3})],$$

as desired. As a consistency check, for $\tau \neq \frac{1}{4}$, if one further replaces $\text{Ai}(V^{2/3}s(\tau))$ with its asymptotic expansion for large V (note that $s(\tau) \geq 0$, by construction), one can match the formulae (E.23). \blacksquare

Remark E.9. We can also extend some of the above calculations by continuity to the case of nonzero H , provided H is sufficiently small. However, in a neighborhood of the critical point $\tau = \frac{1}{4}$, $H = 0$, the above calculations do not apply, as we shall now explain.

Provided that $\tau \neq \frac{1}{4}$, one can continue the dominant saddle point ζ^* to nonzero H in an unambiguous manner, and the dominant saddle remains simple for sufficiently small H . For $\frac{1}{4} < \tau < 1$, recall that the dominant saddle point was $\zeta^* = -1 + \tau^{-1/2}$. In this case, ζ^* is an analytic function of H in a neighborhood of $H = 0$, and admits the expansion

$$\zeta^*(\tau, H) = -1 + \tau^{-1/2} - \frac{(1 - \sqrt{\tau})(1 + 2\sqrt{\tau} + 2\tau)}{2(1 - 2\sqrt{\tau})^4} H^2 + \mathcal{O}(H^4).$$

By continuity, this is the nearest saddle point to the origin when H is sufficiently small, for any fixed $\frac{1}{4} < \tau < 1$. The situation is slightly more complicated when $0 < \tau < \frac{1}{4}$. In this case, one sees that the saddle point at $\zeta^* = 1$ is actually the result of 4 simple saddle points for nonzero H . These saddle points do *not* depend analytically on H in a neighborhood of $H = 0$, and instead have a series expansion in powers of $|H|^{1/2}$. One finds that

$$\zeta = 1 \pm \frac{\sqrt{2}}{(1 - 16\tau^2)^{1/4}} |H|^{1/2} + \mathcal{O}(|H|), \quad 1 \pm \frac{i\sqrt{2}}{(1 - 16\tau^2)^{1/4}} |H|^{1/2} + \mathcal{O}(|H|).$$

These saddle points are continuous in H ; the one nearest to the origin for $|H| > 0$ is therefore

$$\zeta^* = 1 - \frac{\sqrt{2}}{(1 - 16\tau^2)^{1/4}} |H|^{1/2} + \mathcal{O}(|H|).$$

When $H \neq 0$, it is apparent that the saddle point at $\zeta = 1$ split into a quadruple of saddle points when $H \neq 0$, and for $0 < \tau < \frac{1}{4}$. This is also the case when $\frac{1}{4} < \tau < 1$, although the saddle point is irrelevant for the calculations at hand. However, when $\tau = \frac{1}{4}$, the saddle points at $\zeta = 1$, $\zeta = -1 + \tau^{-1/2}$ merge; when the external field H is “turned on” at this point, one finds that the saddle point at $\zeta = 1$ splits into *five* distinct saddle points, each which has the expansion

$$\zeta_j = 1 - 2^{1/5} \mathfrak{s}^j |H|^{2/5} + \mathcal{O}(|H|^{4/5}), \quad j = 0, \dots, 4,$$

where $\mathfrak{s} := e^{\frac{2\pi i}{5}}$ is a 5th root of unity. If one is interested in asymptotics precisely at $\tau = \frac{1}{4}$ (as one might be, if attempting to calculate the critical exponent δ from the Landau theory, cf. [14]), then one recognizes ζ_0 as the dominant saddle point, and can proceed to use classical steepest descent to calculate the asymptotics of $\sigma_V(\frac{1}{4}, H)$ there. A more interesting calculation, which is for the moment out of reach, would be to obtain *uniform* asymptotics in a neighborhood of the critical point $\tau = \frac{1}{4}$, $H = 0$, as we have done in the case when we ignored the parameter H . This would require an analysis similar to that of [25], but instead for a merging of 5 branch points simultaneously. Similarly to how the uniform asymptotics arising from the merging of *two* branch points was written in terms of solutions to a *second order* equation (the Airy equation), one should expect that the asymptotics arising from 5 merging branch points should be characterized in terms of a degree 5 equation, whose solutions are integrals of the form

$$P(x, a, b, c; \gamma) := \frac{1}{2\pi i} \int_{\gamma} e^{\frac{1}{6}\zeta^6 + a\zeta^4 + b\zeta^3 + c\zeta^2 - x\zeta} d\zeta,$$

where the parameters a, b, c, x are functions of τ, H , chosen so that all tend to 0 as $\tau \rightarrow \frac{1}{4}$, $H \rightarrow 0$. One must construct an associated conformal map, which is by no means a simple task. We hope to pursue this analysis in a later work.

Acknowledgements

MD was supported by the Swedish Research Council (VR), grant no. 2021-06015, and the European Research Council (ERC), Grant Agreement No. 101002013. NH was supported by

the European Research Council (ERC), Grant Agreement No. 101002013. Part of the work was completed while NH was at the University of South Florida. We would like to thank an anonymous referee, who pointed out a number of results in the literature which rigorously derive the genus zero free energy using combinatorial methods. We would also like to thank all of the referees, whose valuable comments helped improve the overall presentation of this manuscript.

References

- [1] Albenque M., Ménard L., Geometric properties of spin clusters in random triangulations coupled with an Ising model, [arXiv:2201.11922](#).
- [2] Albenque M., Ménard L., Schaeffer G., Local convergence of large random triangulations coupled with an Ising model, *Trans. Amer. Math. Soc.* **374** (2021), 175–217, [arXiv:1812.03140](#).
- [3] Ambjørn J., Anagnostopoulos K.N., Magnea U., Singularities of the partition function for the Ising model coupled to 2D quantum gravity, *Modern Phys. Lett. A* **12** (1997), 1605–1627, [arXiv:hep-lat/9705004](#).
- [4] Ambjørn J., Anagnostopoulos K.N., Magnea U., Complex zeros of the 2D Ising model on dynamical random lattices, *Nuclear Phys. B* **63** (1998), 751–753, [arXiv:hep-lat/9708014](#).
- [5] Aptekarev A.I., Multiple orthogonal polynomials, *J. Comput. Appl. Math.* **99** (1998), 423–447.
- [6] Baxter R.J., Exactly solved models in statistical mechanics, Academic Press, London, 1989.
- [7] Belavin A.A., Polyakov A.M., Zamolodchikov A.B., Infinite conformal symmetry in two-dimensional quantum field theory, *Nuclear Phys. B* **241** (1984), 333–380.
- [8] Bernardi O., Bousquet-Mélou M., Counting colored planar maps: algebraicity results, *J. Combin. Theory Ser. B* **101** (2011), 315–377, [arXiv:0909.1695](#).
- [9] Bertola M., Marchal O., The partition function of the two-matrix model as an isomonodromic τ function, *J. Math. Phys.* **50** (2009), 013529, 17 pages, [arXiv:0809.1598](#).
- [10] Bertola M., Tovbis A., Asymptotics of orthogonal polynomials with complex varying quartic weight: global structure, critical point behavior and the first Painlevé equation, *Constr. Approx.* **41** (2015), 529–587, [arXiv:1108.0321](#).
- [11] Bessis D., Itzykson C., Zuber J.-B., Quantum field theory techniques in graphical enumeration, *Adv. in Appl. Math.* **1** (1980), 109–157.
- [12] Bleher P., Deaño A., Painlevé I double scaling limit in the cubic random matrix model, *Random Matrices Theory Appl.* **5** (2016), 1650004, 58 pages, [arXiv:1310.3768](#).
- [13] Bleher P., Gharakhloo R., McLaughlin K.T.-R., Phase diagram and topological expansion in the complex quartic random matrix model, *Comm. Pure Appl. Math.* **77** (2024), 1405–1485, [arXiv:2112.09412](#).
- [14] Boulatov D.V., Kazakov V.A., The Ising model on a random planar lattice: the structure of the phase transition and the exact critical exponents, *Phys. Lett. B* **186** (1987), 379–384.
- [15] Boulatov D.V., Kazakov V.A., Kostov I.K., Migdal A.A., Analytical and numerical study of a model of dynamically triangulated random surfaces, *Nuclear Phys. B* **275** (1986), 641–686.
- [16] Bousquet-Mélou M., Schaeffer G., The degree distribution in bipartite planar maps: applications to the Ising model, [arXiv:math.CO/0211070](#).
- [17] Bouttier J., Di Francesco P., Guitter E., Combinatorics of bicubic maps with hard particles, *J. Phys. A* **38** (2005), 4529–4559, [arXiv:math.CO/0501344](#).
- [18] Bouttier J., Di Francesco P., Guitter E., Blocked edges on Eulerian maps and mobiles: application to spanning trees, hard particles and the Ising model, *J. Phys. A* **40** (2007), 7411–7440, [arXiv:math.CO/0702097](#).
- [19] Brézin É., Douglas M.R., Kazakov V., Shenker S.H., The Ising model coupled to 2D gravity. A nonperturbative analysis, *Phys. Lett. B* **237** (1990), 43–46.
- [20] Brézin É., Itzykson C., Parisi G., Zuber J.-B., Planar diagrams, *Comm. Math. Phys.* **59** (1978), 35–51.
- [21] Brézin É., Kazakov V.A., Exactly solvable field theories of closed strings, *Phys. Lett. B* **236** (1990), 144–150.
- [22] Cardy J.L., Conformal invariance and the Yang–Lee edge singularity in two dimensions, *Phys. Rev. Lett.* **54** (1985), 1354–1356.
- [23] Chen L., Turunen J., Critical Ising model on random triangulations of the disk: enumeration and local limits, *Comm. Math. Phys.* **374** (2020), 1577–1643, [arXiv:1806.06668](#).

- [24] Chen L., Turunen J., Ising model on random triangulations of the disk: phase transition, *Comm. Math. Phys.* **397** (2023), 793–873, [arXiv:2003.09343](#).
- [25] Chester C., Friedman B., Ursell F., An extension of the method of steepest descents, *Proc. Cambridge Philos. Soc.* **53** (1957), 599–611.
- [26] Claeys T., Vanlessen M., Universality of a double scaling limit near singular edge points in random matrix models, *Comm. Math. Phys.* **273** (2007), 499–532, [arXiv:math-ph/0607043](#).
- [27] Crnković v., Ginsparg P., Moore G., The Ising model, the Yang–Lee edge singularity, and 2D quantum gravity, *Phys. Lett. B* **237** (1990), 196–201.
- [28] David F., A model of random surfaces with nontrivial critical behaviour, *Nuclear Phys. B* **257** (1985), 543–576.
- [29] de Albuquerque L.C., Alves N.A., Dalmaz D., Yang–Lee zeros of the Ising model on random graphs of non planar topology, *Nuclear Phys. B* **580** (2000), 739–756, [arXiv:hep-th/9912270](#).
- [30] Deift P., Kriecherbauer T., McLaughlin K.T.-R., Venakides S., Zhou X., Strong asymptotics of orthogonal polynomials with respect to exponential weights, *Comm. Pure Appl. Math.* **52** (1999), 1491–1552.
- [31] Deift P., Zhou X., A steepest descent method for oscillatory Riemann–Hilbert problems. Asymptotics for the MKdV equation, *Ann. of Math.* **137** (1993), 295–368, [arXiv:math.AP/9201261](#).
- [32] Di Francesco P., Ginsparg P., Zinn-Justin J., 2D gravity and random matrices, *Phys. Rep.* **254** (1995), 133, [arXiv:hep-th/9306153](#).
- [33] Douglas M.R., Strings in less than one dimension and the generalized KdV hierarchies, *Phys. Lett. B* **238** (1990), 176–180.
- [34] Douglas M.R., Shenker S.H., Strings in less than one dimension, *Nuclear Phys. B* **335** (1990), 635–654.
- [35] Duits M., Geudens D., A critical phenomenon in the two-matrix model in the quartic/quadratic case, *Duke Math. J.* **162** (2013), 1383–1462, [arXiv:1111.2162](#).
- [36] Duits M., Geudens D., Kuijlaars A.B.J., A vector equilibrium problem for the two-matrix model in the quartic/quadratic case, *Nonlinearity* **24** (2011), 951–993, [arXiv:1007.3137](#).
- [37] Duits M., Hayford N., The Ising model coupled to 2D gravity: critical partition function, in preparation.
- [38] Duits M., Kuijlaars A.B.J., Painlevé I asymptotics for orthogonal polynomials with respect to a varying quartic weight, *Nonlinearity* **19** (2006), 2211–2245, [arXiv:math.CA/0605201](#).
- [39] Duits M., Kuijlaars A.B.J., Universality in the two-matrix model: a Riemann–Hilbert steepest-descent analysis, *Comm. Pure Appl. Math.* **62** (2009), 1076–1153, [arXiv:0807.4814](#).
- [40] Duits M., Kuijlaars A.B.J., Mo M.Y., The Hermitian two matrix model with an even quartic potential, *Mem. Amer. Math. Soc.* **217** (2012), v+105 pages, [arXiv:1010.4282](#).
- [41] Eynard B., Large- N expansion of the 2-matrix model, *J. High Energy Phys.* **2003** (2003), no. 1, 051, 38 pages, [arXiv:hep-th/0210047](#).
- [42] Eynard B., Counting surfaces. CRM Aisenstadt chair lectures, *Prog. Math. Phys.*, Vol. 70, [Birkhäuser](#), Cham, 2016.
- [43] Fokas A.S., Its A.R., Kitaev A.V., Discrete Painlevé equations and their appearance in quantum gravity, *Comm. Math. Phys.* **142** (1991), 313–344.
- [44] Fokas A.S., Its A.R., Kitaev A.V., The isomonodromy approach to matrix models in 2D quantum gravity, *Comm. Math. Phys.* **147** (1992), 395–430.
- [45] Fukuma M., Kawai H., Nakayama R., Explicit solution for p - q duality in two-dimensional quantum gravity, *Comm. Math. Phys.* **148** (1992), 101–116.
- [46] Ginsparg P., Goulian M., Plesser M.R., Zinn-Justin J., (p, q) string actions, *Nuclear Phys. B* **342** (1990), 539–563.
- [47] Gross D.J., Migdal A.A., A nonperturbative treatment of two-dimensional quantum gravity, *Nuclear Phys. B* **340** (1990), 333–365.
- [48] Gross D.J., Migdal A.A., Nonperturbative two-dimensional quantum gravity, *Phys. Rev. Lett.* **64** (1990), 127–130.
- [49] Hayford N., The Ising model coupled to gravity: higher-order Painlevé equations/the $(3, 4)$ string equation, [arXiv:2405.03260](#).
- [50] Ising E., Beitrag zur theorie des ferromagnetismus, *Z. Physik* **31** (1925), 253–258.

- [51] Itzykson C., Zuber J.-B., The planar approximation. II, *J. Math. Phys.* **21** (1980), 411–421.
- [52] Joshi N., Kitaev A.V., Treharne P.A., On the linearization of the first and second Painlevé equations, *J. Phys. A* **42** (2009), 055208, 18 pages, [arXiv:0806.0271](#).
- [53] Kadanoff L.P., Scaling laws for ising models near T_c , *Phys. Phys. Fiz.* **2** (1966), 263–272.
- [54] Kapaev A.A., Quasi-linear stokes phenomenon for the Painlevé first equation, *J. Phys. A* **37** (2004), 11149–11167, [arXiv:nlin.SI/0404026](#).
- [55] Kapaev A.A., Kitaev A.V., Connection formulae for the first Painlevé transcendent in the complex domain, *Lett. Math. Phys.* **27** (1993), 243–252.
- [56] Kazakov V.A., Bilocal regularization of models of random surfaces, *Phys. Lett. B* **150** (1985), 282–284.
- [57] Kazakov V.A., Ising model on a dynamical planar random lattice: exact solution, *Phys. Lett. A* **119** (1986), 140–144.
- [58] Kazakov V.A., Kostov I.K., Migdal A.A., Critical properties of randomly triangulated planar random surfaces, *Phys. Lett. B* **157** (1985), 295–300.
- [59] Knizhnik V.G., Polyakov A.M., Zamolodchikov A.B., Fractal structure of 2D-quantum gravity, *Modern Phys. Lett. A* **3** (1988), 819–826.
- [60] Krantz S.G., Parks H.R., The implicit function theorem. History, theory, and applications, *Mod. Birkhäuser Class.*, [Birkhäuser](#), New York, 2013.
- [61] Kuijlaars A.B.J., McLaughlin K.T.-R., A Riemann–Hilbert problem for biorthogonal polynomials, *J. Comput. Appl. Math.* **178** (2005), 313–320, [arXiv:math.CV/0310204](#).
- [62] Kuijlaars A.B.J., Silva G.L.F., S-curves in polynomial external fields, *J. Approx. Theory* **191** (2015), 1–37, [arXiv:1311.7026](#).
- [63] Lee T.D., Yang C.N., Statistical theory of equations of state and phase transitions. II. Lattice gas and Ising model, *Phys. Rev.* **87** (1952), 410–419.
- [64] Liechty K., Wang D., Two lax systems for the Painlevé II equation, and two related kernels in random matrix theory, *SIAM J. Math. Anal.* **48** (2016), 3618–3666, [arXiv:1601.01603](#).
- [65] Martínez-Finkelstein A., Silva G.L.F., Critical measures for vector energy: global structure of trajectories of quadratic differentials, *Adv. Math.* **302** (2016), 1137–1232, [arXiv:1509.06704](#).
- [66] Martínez-Finkelstein A., Silva G.L.F., Critical measures for vector energy: asymptotics of non-diagonal multiple orthogonal polynomials for a cubic weight, *Adv. Math.* **349** (2019), 246–315, [arXiv:1805.01748](#).
- [67] Martínez-Finkelstein A., Silva G.L.F., Spectral curves, variational problems and the Hermitian matrix model with external source, *Comm. Math. Phys.* **383** (2021), 2163–2242, [arXiv:1907.08108](#).
- [68] McCoy B.M., Wu T.T., The two-dimensional Ising model, [Harvard University Press](#), Cambridge, MA, 1973.
- [69] Mehta M.L., A method of integration over matrix variables, *Comm. Math. Phys.* **79** (1981), 327–340.
- [70] Mehta M.L., Random matrices, 3rd ed., *Pure Appl. Math. (Amsterdam)*, Vol. 142, Elsevier, Amsterdam, 2004.
- [71] Miller P.D., Applied asymptotic analysis, *Grad. Stud. Math.*, Vol. 75, [American Mathematical Society](#), Providence, RI, 2006.
- [72] Nikishin E.M., Sorokin V.N., Rational approximations and orthogonality, *Transl. Math. Monogr.*, Vol. 92, [American Mathematical Society](#), Providence, RI, 1991.
- [73] Onsager L., Crystal statistics. I. A two-dimensional model with an order-disorder transition, *Phys. Rev.* **65** (1944), 117–149.
- [74] Orantin N., Chain of matrices, loop equations, and topological recursion, in *The Oxford Handbook of Random Matrix Theory*, Oxford University Press, Oxford, 2011, 329–352, [arXiv:0911.5089](#).
- [75] Polyakov A.M., Lecture at Northeastern University (from the lecture notes of P. Ginsparg and G. Moore), [arXiv:hep-th/9304011](#).
- [76] Polyakov A.M., Quantum geometry of bosonic strings, *Phys. Lett. B* **103** (1981), 207–210.
- [77] Rakhmanov E.A., Orthogonal polynomials and S-curves, in *Recent Advances in Orthogonal Polynomials, Special Functions, and Their Applications*, *Contemp. Math.*, Vol. 578, [American Mathematical Society](#), Providence, RI, 2012, 195–239.
- [78] Stanley R.P., Enumerative combinatorics. Vol. 2, *Cambridge Stud. Adv. Math.*, Vol. 62, [Cambridge University Press](#), Cambridge, 1999.

-
- [79] Staudacher M., The Yang–Lee edge singularity on a dynamical planar random surface, *Nuclear Phys. B* **336** (1990), 349–362.
 - [80] Turunen J., Interfaces in the vertex-decorated Ising model on random triangulations of the disk [arXiv:2003.11012](#).
 - [81] Van Assche W., Coussement E., Some classical multiple orthogonal polynomials, *J. Comput. Appl. Math.* **127** (2001), 317–347, [arXiv:math.CA/0103131](#).
 - [82] Whittaker E.T., Watson G.N., A course of modern analysis, *Cambridge Math. Lib.*, [Cambridge University Press](#), Cambridge, 1996.
 - [83] Wilson K.G., Renormalization group and critical phenomena. I. Renormalization and the Kadanoff scaling picture, *Phys. Rev. B* **4** (1971), 3174–3183.
 - [84] Wilson K.G., Renormalization group and critical phenomena. II. Phase-space cell analysis of critical behavior, *Phys. Rev. B* **4** (1971), 3184–3205.
 - [85] Yang C.N., The spontaneous magnetization of a two-dimensional Ising model, *Phys. Rev.* **85** (1952), 808–816.
 - [86] Yang C.N., Lee T.D., Statistical theory of equations of state and phase transitions. I. Theory of condensation, *Phys. Rev.* **87** (1952), 404–409.
 - [87] Zamolodchikov A., Ishimoto Y., Massive Majorana fermion coupled to two-dimensional gravity and random lattice Ising model, *Theoret. and Math. Phys.* **147** (2006), 755–776.
 - [88] Zinn-Justin P., Zuber J.-B., On some integrals over the $U(N)$ unitary group and their large N limit, *J. Phys. A* **36** (2003), 3173–3193, [arXiv:math-ph/0209019](#).