ASYMPTOTICS OF THE PARTITION FUNCTION FOR RANDOM MATRICES VIA RIEMANN-HILBERT TECHNIQUES, AND APPLICATIONS TO GRAPHICAL ENUMERATION

N. M. ERCOLANI AND K. D. T-R MCLAUGHLIN

Abstract. We study the partition function from random matrix theory using a well known connection to orthogonal polynomials, and a recently developed Riemann-Hilbert approach to the computation of detailed asymptotics for these orthogonal polynomials. We obtain the first proof of a complete large N expansion for the partition function, for a general class of probability measures on matrices, originally conjectured by Bessis, Itzykson, and Zuber. We prove that the coefficients in the asymptotic expansion are analytic functions of parameters in the original probability measure, and that they are generating functions for the enumeration of labelled maps according to genus and valence. Central to the analysis is a large N expansion for the mean density of eigenvalues, uniformly valid on the entire real axis.

Contents

K. D. T-R McLaughlin was supported in part by NSF grant no. 9970328 and no. 0200749, and N. M. Ercolani was supported in part by NSF grant no. 0073087. The authors gratefully acknowledge the faculty and staff of the Ecole Normale Superieur, Paris, France, for their kind hospitality.

1. Motivation and Background

In this paper we will consider asymptotics for the following family of integrals:

$$Z_N(t_1, t_2, \dots, t_\nu) = \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \exp\left\{ -N^2 \left[\frac{1}{N} \sum_{j=1}^{N} V(\lambda_j;\ t_1, \dots, t_\nu) - \frac{1}{N^2} \sum_{j \neq \ell} \log|\lambda_j - \lambda_\ell| \right] \right\} d^N\lambda, \tag{1.1}$$

$$V(\lambda;\ t_1, \dots, t_\nu) = V_{\mathbf{t}}(\lambda) = V(\lambda) = \frac{1}{2}\lambda^2 + \sum_{k=1}^{\nu} t_k \lambda^k. \tag{1.2}$$

where the parameters $\{t_1, \dots, t_\nu\}$ are assumed to be such that the integral converges. For example, one may suppose that ν is even, and $t_\nu > 0$. We will use the following set for allowable $\mathbf{t} = (t_1, \dots, t_\nu)$. For any given $T > 0$ and $\gamma > 0$, define

$$\mathbb{T}(T, \gamma) = \{\mathbf{t} \in \mathbb{R}^\nu :\ |\mathbf{t}| \leq T,\ t_\nu > \gamma \sum_{j=1}^{\nu-1} |t_j|\}.$$

There are three main results in this paper, Theorems 1.1, 1.2, and 1.3. The first of these is the following theorem.

Theorem 1.1. *There is $T > 0$ and $\gamma > 0$ so that for $\mathbf{t} \in \mathbb{T}(\mathbf{T}, \gamma)$, one has the $N \to \infty$ asymptotic expansion*

$$\log\left(\frac{Z_N(\mathbf{t})}{Z_N(\mathbf{0})}\right) = N^2 e_0(\mathbf{t}) + e_1(\mathbf{t}) + \frac{1}{N^2} e_2(\mathbf{t}) + \cdots. \tag{1.3}$$

The meaning of this expansion is: if you keep terms up to order N^{-2k}, the error term is bounded by CN^{-2k-2}, where the constant C is independent of $\mathbf{t}$ for all $\mathbf{t} \in \mathbb{T}(\mathbf{T}, \gamma)$. For each j, the function $e_j(\mathbf{t})$ is an analytic function of the (complex) vector $\mathbf{t}$, in a neighborhood of $\mathbf{0}$. Moreover, the asymptotic expansion of derivatives of $\log(Z_N)$ may be calculated via term-by-term differentiation of the above series.

Remark: In the statement of the Theorem, $\mathbf{t} \in \mathbb{T}(T, \gamma)$. This is not the largest domain where the asymptotic expansion holds true. What is really required is the existence of a path Γ in $\mathbb{R}^\nu$ connecting $\mathbf{t}$ to $\mathbf{0}$ such that for all $\mathbf{t} \in \Gamma$, the associated equilibrium measure (see Section 3) is supported on a single interval, with strict variational inequality off the support, strict positivity on the interval of the support, and square-root vanishing at the endpoints. A global characterization of a maximal domain where the expansion holds true would lead us too far from the main focus of this paper, and we will not pursue this here.

The integral $Z_N(t_1, \dots, t_\nu) = Z_N(\mathbf{t})$ appears in a number of areas of mathematics and mathematical physics. We will explain some of the motivations for studying the asymptotics of $Z_N(\mathbf{t})$ in the following subsections. In what follows it will be useful to define the ratio:

$$\hat{Z}_N(\mathbf{t}) = Z_N(\mathbf{t})/Z_N(\mathbf{0}). \tag{1.4}$$

1.1. Motivation: statistical mechanics of a log-gas.

The integral (1.1) has the natural interpretation as the partition of function for a statistical mechanical system of particles on the line, with logarithmic interaction potential, in the presence of an external field whose potential is $V_{\mathbf{t}}(\lambda)$. The asymptotic behavior for $N \to \infty$ may then be interpreted as the limiting behavior of the statistical mechanical system in the low temperature and many particle asymptotic limit.

With N fixed, the integral $Z_N(t_1, \dots, t_\nu)$, and indeed all relevant statistical observables, may be expressed in terms of families (parameterized by $\{t_1, \dots, t_\nu\}$, and N) of associated orthogonal polynomials (see Section 1.4 below). By itself, this observation is not terribly useful. However, it turns out that these families of orthogonal polynomials are in a certain sense completely integrable. Indeed, all limiting asymptotic questions about the orthogonal polynomials (and hence concerning the statistical mechanical system) are in principle *explicitly computable* (see Section 3 below).

1.2. **Motivation: random matrix theory.** The integral (1.1) is also of fundamental importance in the theory of random matrices. As is well known (see, for example, the review text [21]), in the theory of random Hermitian matrices from the so-called "Unitary ensemble", one considers the measure on $N \times N$ Hermitian matrices given by

$$d\mu_{\mathbf{t}} = \frac{1}{Z_N} \exp\{-N \text{ Tr } [V_{\mathbf{t}}(M)]\} dM, \tag{1.5}$$

where dM is Lebesgue measure on the matrix entries, i.e.

$$dM = \prod_{j<k} dM_{jk}^{\mathrm{R}} dM_{jk}^{\mathrm{I}} \prod_{j=1}^{N} dM_{jj},$$

where M_{jk}^{R} denotes the real part of the matrix entry M_{jk}, and M_{jk}^{I} denotes the imaginary component of the matrix entry M_{jk}. It is a basic fact that the measure (1.5) induces a probability measure on the eigenvalues, with density

$$\frac{1}{Z_N} \exp\left\{-N^2\left[\frac{1}{N}\sum_{j=1}^{N} V_{\mathbf{t}}(\lambda_j) + \frac{1}{N^2}\sum_{j\neq\ell} \log|\lambda_j - \lambda_\ell|\right]\right\} d^N\lambda, \tag{1.6}$$

where $Z_N = Z_N(\mathbf{t})$ appearing in (1.6) is precisely the partition function defined in (1.1). Thus asymptotics such as those contained in (1.3) yield asymptotic information concerning the statistics of the eigenvalues of these random matrices. For example, by differentiating $\log Z_N$, one obtains

$$\frac{\partial}{\partial t_\ell} \log Z_N = -N \mathbb{E}\left(\text{ Tr } M^\ell \right), \tag{1.7}$$

where $\mathbb{E}$ denotes the expectation with respect to the probability measure $d\mu_{\mathbf{t}}$. Now if one has established (1.3), then in conjunction with (1.7), one learns the following:

$$\lim_{N\to\infty} \mathbb{E}\left(\frac{1}{N} \text{ Tr } M^\ell\right) = \frac{\partial}{\partial t_\ell} e_0(t_1, \ldots, t_\nu). \tag{1.8}$$

Similar calculations show that (1.3) implies

$$\begin{aligned} &\lim_{N\to\infty} \{\mathbb{E}\left(\text{ Tr } (M^n) \cdot \text{ Tr } (M^m)\right) - \mathbb{E}\left(\text{ Tr } M^m\right) \cdot \mathbb{E}\left(\text{ Tr } M^n\right)\} \\ &\quad = \frac{\partial^2}{\partial t_m \partial t_n} e_0(t_1, \ldots, t_\nu). \end{aligned} \tag{1.9}$$

Observe that (1.8) and (1.9) together imply that the fundamental random variables $\left\{\frac{1}{N}\text{Tr } M^\ell\right\}_{\ell=1}^{\infty}$ are asymptotically uncorrelated.

1.3. **Motivation: Graphical Enumeration.**

1.3.1. *The Viewpoint of Gaussian Expectations.* One way to try to think about the partition function $\hat{Z}_N$ in (1.5) is as a Gaussian expectation of the *interaction term*: $\exp\left\{-N \text{ Tr } \left[V_{\mathbf{t}}(M) - \frac{1}{2}M^2\right]\right\}$. This is the viewpoint taken in much of the physical literature on random matrix theory, for example [3].

To better explain this we briefly review some facts about Gaussian measures:
A measure μ on $\mathbb{R}^n$ is called *Gaussian* if its characteristic function, $\phi(\mathbf{k}) := \int_{\mathbb{R}^n} e^{i(\mathbf{k},\mathbf{x})} d\mu$, has the form

$$\phi(\mathbf{k}) = \exp\{i(\rho, \mathbf{k}) - \frac{1}{2}(\mathbf{Q}\mathbf{k}, \mathbf{k})\}, \tag{1.10}$$

where ρ is the mean and $\mathbf{Q}$ is called the covariance of μ. For mean zero and nonsingular covariance,

$$d\mu = \exp\{-\frac{1}{2}(\mathbf{A}\mathbf{x}, \mathbf{x})\}\mathbf{dx} \tag{1.11}$$

where $\mathbf{A} = \mathbf{Q}^{-1}$. We consider $d\mu$ of the form (1.11).

The salient feature of Gaussian measures, and the one which is fundamental for applications to graphical enumeration, is that expectations of general polynomial functions can be reduced to quadratic expectations.

The explicit recipe for this, generally referred to as the Wick formula, states that for linear functions ℓ_i on $\mathbb{R}^n$,

$$\langle \ell_1 \ell_2 \ldots \ell_{2k} \rangle = \sum \langle \ell_{r_1} \ell_{s_1} \rangle \langle \ell_{r_2} \ell_{s_2} \rangle \ldots \langle \ell_{r_k} \ell_{s_k} \rangle \tag{1.12}$$

where $\langle \cdot \rangle$ denotes expectation with respect to $d\mu$ and the sum is taken over the $(2k-1)!!$ $(= (2k-1)(2k-3) \ldots 1)$ Wick couplings of $1, 2, \ldots, 2k$. A *Wick coupling* is a partition of $1, 2, \ldots, 2k$ into couples (r_i, s_i) such that $r_1 < r_2 \cdots < r_k$ and $s_i > r_i$. The quadratic expectations are completely determined by $\langle x_i y_j \rangle = q_{ij}$ where q_{ij} is the ij^{th} entry of the covariance matrix Q. Expectations of odd polynomials vanish.

The proof of the Wick formula follows from a comparison of the Taylor coefficients of the characteristic function $\phi(\mathbf{k})$ and its logarithm. We refer the reader to [23, P. 9] for details.

The random matrix measure defining the Gaussian Unitary Ensemble (GUE) and given by

$$d\mu = 2^{-\frac{N}{2}} \pi^{-\frac{N^2}{2}} \exp\left\{ -\frac{1}{2} \mathrm{Tr} M^2 \right\} dM \tag{1.13}$$

is manifestly a Gaussian measure. It is straightforward to work out the covariance matrix from which one may conclude that

$$\langle m_{ij} m_{ji} \rangle = 1, \text{ and } \langle m_{ij} m_{kl} \rangle = 0 \text{ for } (i,j) \neq (k,l). \tag{1.14}$$

1.3.2. *Enumerating Maps.* There are by now a number of striking examples which illustrate the power of random matrix methods for calculating explicit solutions of combinatorial problems related to graphical enumeration. One of these is the problem of enumerating maps.

A *map* is a graph which is embedded into a Riemann surface so that

(1) the (images of the) edges do not intersect;
(2) dissecting the surface along the edges decomposes it into a union of open cells; these cells are called the *faces* of the map.

Given this definition, one can give a precise combinatorial description of a class of maps in terms of edge identifications between a collection of faces. For instance one can pose the question of how many maps can be constructed from a single face, having $2k$ edges around its boundary, by identifying its edges in pairs. It is straightforward to see that this number is $(2k-1)!!$. A more subtle question is to ask how many of these maps lie on a surface of genus g. It turns out that this counting problem has a direct and natural reinterpretation in terms of the combinatorics of Wick couplings. Let $\varepsilon_g(k)$ denote the number of one face maps with $2k$ edges on a Riemann surface of genus g. A generating function for these numbers can be directly expressed in terms of a random matrix moment [18]:

$$\langle \mathrm{Tr} M^{2k} \rangle = N^{k+1} \sum_{g=0}^{[k/2]} \varepsilon_g(k) \left(\frac{1}{N^2} \right)^g \tag{1.15}$$

There are extensions of these kinds of calculations to more general classes of maps along with remarkable applications to the calculation of geometric invariants of moduli spaces of Riemann surfaces. We refer the reader to [17] for a good general description of these results.

1.3.3. *Diagrammatic Expansions.* We return now to considering the evaluation of the partition function for the *deformed unitary ensemble* $\hat{Z}_N(\mathbf{t})$. Here we will summarize some beautiful work relating the asymptotic expansion (1.3) to some problems in enumerative geometry. As mentioned above, in [3] the authors asserted the existence of the expansion (1.3), and presented a very elegant consistency argument, which is described in subsection 1.5. The components of $\mathbf{t}$ are viewed as deformation parameters; when $\mathbf{t} = 0$ one has the partition function of the original Gaussian Unitary Ensemble.

To fix ideas, will will restrict our attention to the case considered in [3]. So in this subsection, we will set all parameters $t_j = 0$, except for $t = t_4$. (While in [3] the authors considered more general deformations, the authors gave a more detailed discussion for this case.)

Consider the Taylor series expansion corresponding to all parameters $t_j = 0$ except for $t = t_4$,

$$\exp\left\{-\frac{t}{N}\mathrm{Tr}M^4\right\} = \sum_{n\geq 0}\frac{1}{n!}\left(\frac{-t}{N}\right)^n\left(TrM^4\right)^n. \tag{1.16}$$

This is clearly globally convergent for all t. The difficulty arises when one considers the Gaussian expectation (using (1.13)) of this exponential function. Commuting the expectation integrals with the sum on the right hand side, [3] produced the *formal* Taylor series expansion of the partition function around $t = 0$:

$$\begin{aligned}\hat{Z}_N(t) \quad &= \quad \left\langle \exp\left\{-\frac{t}{N}\mathrm{Tr}M^4\right\}\right\rangle = \frac{1}{Z_N(\mathbf{0})}Z_N(0,0,0,t_4 = t, 0,\ldots,0) \\ \text{“}=\text{”} \quad &\sum_{n\geq 0}\frac{1}{n!}\left(\frac{-t}{N}\right)^n\langle (TrM^4)^n\rangle.\end{aligned} \tag{1.17}$$

The quotation marks indicate where one has proceeded formally. Indeed it is manifest that this is not a convergent series expansion since the integral corresponding to $\langle\exp\{-(t/N)\mathrm{Tr}M^4\}\rangle$ converges only for t with positive real part. Properly speaking, this is really shorthand for a series of identities relating *derivatives* of $\hat{Z}_N$ evaluated at $t = 0$ to expectations of powers of Tr M^4.

Again, use of the Wick calculus allows one to replace the matrix moments appearing in the above formal series by a generating function for enumerating a certain class of *labelled* 4-valent maps (or disjoint unions of maps) which we will refer to as diagrams. A *four valent diagram* consists of

(1) n 4-valent vertices;
(2) a labelling of the vertices by the numbers 1,2,...,n;
(3) a labelling of the edges incident to vertex σ (for $\sigma = 1,\ldots,n$) by letters $i_\sigma, j_\sigma, k_\sigma, \ell_\sigma$ (This alphabetic order corresponds to the cyclic order of the edges around the vertex);
(4) a partitioning of the labelled edges into pairs.

(Once the number n is specified, the set of all n-vertex, four valent diagrams is in one-to-one correspondence with the set of all pairings.) Connecting the edges according to the pairing yields a graph, together with a cyclic ordering of the edges around each vertex. For definiteness, we will adhere to the rule that this cyclic ordering corresponds to a clockwise orientation of the edges around a vertex.

If the underlying graph is connected, the cyclic ordering determines a map associated to the diagram. In that case we will call the diagram a g-map if the map is on a Riemann surface of genus g. (Note that a *g-map* as defined here has more structure than a *map* as defined in the previous subsection since a g-map carries a labelling of its vertices and edges.) If the graph is not connected, then the edge labelling associates a map to each connected component, and the conglomerate of maps will be called a g-diagram.

The use of the Wick calculus yields the following different representation of (1.17), which clearly demonstrates that the partition function is a generating function for the enumeration of *diagrams*:

$$\hat{Z}_N(t)\text{“}=\text{”}\sum_{n\geq 0}\frac{1}{n!}\left(\frac{-t}{N}\right)^n\sum_g \#\{\text{4-valent, n-vertex, g-diagrams}\}N^{2-2g+n}. \tag{1.18}$$

Exploiting the relationship between the terms of a Taylor series and those of its logarithm one can write down an equally formal representation for the the **logarithm** of the partition function which can be regarded as a generating function for connected n-vertex diagrams:

$$\log\hat{Z}_N(t)\text{“}=\text{”}\sum_{n\geq 0}\frac{1}{n!}\left(\frac{-t}{N}\right)^n\sum_{g\geq 0}\#\{\text{4-valent, n-vertex g-maps}\}N^{2-2g+n}. \tag{1.19}$$

Finally one can make another leap beyond rigor and resum the terms of the previous formal series, as was done in [3], to formally order the series by the genus of the surface to which the diagram maps. This yields what is referred to in the physics literature as the genus expansion,

$$\log\hat{Z}_N(t)\text{“}=\text{”}\sum_g E_g(t)N^{2-2g}, \tag{1.20}$$

where $E_g(t) = \sum_{n\geq 1} \frac{1}{n!}(-t)^n \kappa_g(n)$ is a formal series (possibly convergent) in which each of the coefficients $\kappa_g(n)$ is the number of connected maps of genus g with n vertices (all 4-valent).

The authors of [3] fully appreciated that this intuitive "deduction" was completely formal and in fact suggested an approach for providing a rigorous derivation. This approach was never fully pursued (we briefly review its history in Section 1.5).

We have taken a different approach and our proof of Theorem 1.1 provides a rigorous validation of the form of the representation (1.20). When restricted to the four-valent case, our result may be summarized as follows.

A. There is a positive number T_4 such that for $0 \leq t \leq T_4$, the partition function $\hat{Z}_N(t)$ possesses the following asymptotic expansion.

$$N^{-2} \log \hat{Z}_N(t) = e_0(t) + N^{-2} e_1(t) + N^{-4} e_2(t) + \cdots . \tag{1.21}$$

The meaning of this expansion is: if you keep terms up to order N^{-2k}, the error term is bounded by CN^{-2k-2}. The coefficients $e_g(t)$ possess analytic continuations to a neighborhood of $t = 0$.

B. The coefficients $e_g(t)$ are related to the counting of 4-valent g-maps via the following formula:

$$(-1)^n \frac{\partial^n}{\partial t^n} e_g(0) = \#\{\text{4-valent, n-vertex, g-maps}\}.$$

For the case of more general deformations, and the full asymptotic expansion (1.3), we have the following theorem concerning the coefficients $e_g(t_1 \dots t_\nu)$.

Theorem 1.2. *The coefficients in the asymptotic expansion (1.3) satisfy the following relations. Let g be a nonnegative integer. Then*

$$e_g(t_1 \dots t_\nu) = \sum_{n_j \geq 1} \frac{1}{n_1! \dots n_\nu!} (-t_1)^{n_1} \dots (-t_\nu)^{n_\nu} \kappa_g(n_1, \dots, n_\nu) \tag{1.22}$$

in which each of the coefficients $\kappa_g(n_1, \dots, n_\nu)$ is the number of g-maps with n_j j-valent vertices for $j = 1, \dots, \nu$.

1.4. **Motivation: The Connection to Orthogonal Polynomials.** As mentioned above, the integral (1.1) is intimately connected to the theory of orthogonal polynomials. Consider the measure

$$w_N(x)dx := \exp\left[-NV_{\mathbf{t}}(x)\right]dx. \tag{1.23}$$

Let us define $\{p_j(x;\ N,\ \mathbf{t})\}_{j=0}^{\infty}$ to be the sequence of polynomials orthogonal with respect to the measure $w_N(x)dx$. That is, $\{p_j(x;\ N,\ \mathbf{t})\}_{j=0}^{\infty}$ satisfies

$$\int_{-\infty}^{\infty} p_j p_k w_N dx = \begin{cases} 0 & j \neq k \\ 1 & j = k \end{cases}, \tag{1.24}$$

and $p_j(x;\ N,\ \mathbf{t}) = \gamma_j^{(N)} x^j + \cdots$, $\gamma_k^{(N)} > 0$. (The leading coefficient $\gamma_k^{(n)}$ is of course dependent on the parameters $t_1, \dots, t_\nu$; however, we suppress this dependence for notational convenience.) The fact of the matter is that $Z_N(\mathbf{t})$ may also be defined via

$$Z_N(\mathbf{t}) = N! \prod_{\ell=0}^{N-1} \left(\gamma_\ell^{(N)}\right)^{-2}. \tag{1.25}$$

Z_N is also defined via

$$Z_N(t_1, \dots, t_\nu) = N! \begin{vmatrix} c_0 & c_1 & \cdots & c_{N-1} \\ c_1 & c_2 & \cdots & c_N \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ c_{N-1} & c_N & \cdots & c_{2N-2} \end{vmatrix}, \tag{1.26}$$

where $c_j = \int_{\mathbb{R}} x^j w_N(x)dx$ are the moments of the measure $w_N(x)dx$, and the determinant above is called a Hankel determinant (see, for example, Szegö's classic text [24]). The asymptotic expansion (1.3) constitutes a version of the strong Szegö limit theorem for Hankel determinants. The strong Szegö limit theorem concerns the asymptotic behavior of Toeplitz determinants associated to a given measure on the interval $(0, 2\pi)$. We refer the interested reader to [24] for more information.

It is well known in both the approximation theory literature and the random matrix theory literature (see, for example, [21],[22] and the references therein, or [5] and references) that $Z_N(\mathbf{t})$ satisfies the following leading order asymptotic behavior:

$$\lim_{N\to\infty} \frac{1}{N^2} \log Z_N(t_1, \dots t_\nu)$$
$$= \sup_{\mu\in\mathbb{A}} \left\{ -\int V_{\mathbf{t}}(\lambda)d\mu(\lambda) + \int\int \log|\lambda-\eta| d\mu(\lambda)d\mu(\eta) \right\},$$

where $\mathbb{A}$ is the set of all positive Borel measures on the real axis, with unit mass. The variational problem posed above is referred to in the approximation theory literature as the problem of determining the equilibrium measure for logarithmic potentials in the presence of an external field (see [22]). It is well known that the supremum is achieved at a unique measure μ^* (see, for example, [22]), and that for real analytic external fields, the equilibrium measure μ^* is supported on finitely many disjoint intervals, and on the interior of each interval, the equilibrium measure has analytic density [5].

1.5. **Brief discussion of history.** In [3], Bessis, Itzykson and Zuber considered asymptotics for the integral (1.1), but for the special case in which $t_j \equiv 0$ for $j \neq 4$. That is, they considered the integral

$$Z_N^{(\mathrm{BIZ})}(t) = \int\cdots\int \exp\left\{-N^2\left[\frac{1}{N}\sum_{j=1}^N\left(\frac{1}{2}\lambda_j^2 + t\lambda_j^4\right) - \frac{1}{N^2}\sum_{j\neq\ell}\log|\lambda_j-\lambda_\ell|\right]\right\}d^N\lambda. \tag{1.27}$$

They computed asymptotics for this integral by using (1.25), together with some reasonable assumptions on the asymptotics of ratios of the leading coefficients $\gamma_j^{(n)}$. The basic idea is to re-write (1.25) in yet another form, namely

$$\frac{1}{N^2}\log Z_N^{(\mathrm{BIZ})}(t) = \frac{1}{N^2}\log(N!) - \frac{2}{N}\log\left(\gamma_0^{(N)}\right) + \sum_{j=1}^{N-1}\frac{2(N-j)}{N^2}\log\left(b_{j-1}^{(N)}\right), \tag{1.28}$$

where $b_j^{(N)}$ denotes the recurrence coefficient associated to the orthogonal polynomial sequence, $xp_j(x) = b_j p_{j+1}(x) + a_j p_j(x) + b_{j-1}p_{j-1}(x)$. It easy easy to see that the recurrence coefficients are related directly to the leading coefficients $\gamma_j^{(N)}$, via $b_j = \gamma_j^{(N)}/\gamma_{j+1}^{(N)}$.

The authors then made the following asymptotic assumptions concerning the recurrence coefficients $b_j^{(N)}$:

$$\left(b_j^{(N)}\right)^2 = r_0\left(\frac{j}{N},t\right) + \frac{1}{N^2}r_2\left(\frac{j}{N},t\right) + \cdots + \frac{1}{N^{2k}}r_{2k}\left(\frac{j}{N},t\right) + \cdots, \tag{1.29}$$

where the functions $r_{2\ell}(x,t)$ are assumed to be infinitely differentiable functions of the variable x. With these assumptions in hand, Bessis, Itzykson, and Zuber then substitute (1.29) into (1.28), and apply the Euler-Maclaurin summation formula to replace each summation by a series of integrals.

Following that, there is an argument that *if an asymptotic description such as (1.29) holds, then one may deduce the form of the functions $r_{2\ell}$.* This is achieved by studying a nonlinear recursion relation satisfied by the recurrence coefficients $b_j^{(N)}$. This in turn leads to a prediction for the asymptotics of $Z_N^{(\mathrm{BIZ})}$. The authors compute explicitly the first three terms, and present a conjecture for all terms. Their conjecture is that for $j \geq 2$, $e_j^{(\mathrm{BIZ})}(t)$ is a rational function of t, with one pole in the finite plane, of order $5(j-1)$, located at $t = -1/48$.

Quite separately, a rigorous analysis of the recurrence coefficients for polynomials orthogonal with respect to weights of the form $e^{-x^{2m}}dx$ [20] was carried out by analyzing the very same nonlinear recursion relation. This yielded a proof of the leading order information contained in (1.29), but valid only if $j \to \infty$.

On the other hand, in [3], the authors use (1.29) as a uniform expansion, valid even for $j = 1$, as $N \to \infty$. It would seem plausible that such an expansion would hold for $j \to \infty$, $N \to \infty$, such that $j/N \to x > 0$. However, it is entirely possible that for j such that j/N approaches 0, there is some correction to (1.29) which may affect the asymptotic computation of $\log(Z_N^{(\mathrm{BIZ})})$ at some finite algebraic order in N.

For this reason, it is of great interest to find some independent way to compute the asymptotics for $\log(Z_N(t_1, \ldots, t_\nu))$.

In a different direction, in [4], techniques from the theory of isomonodromic deformations, as well as techniques for the asymptotic analysis of Riemann-Hilbert problems were applied to the problem of determining asymptotics for the polynomials $p_n(z;\ N,\ t)$ orthogonal with respect to the 1-parameter family of measures $\exp[-N(x^2/2 + tx^4)]$. The results were used to prove that the associated random matrix model obeys the universality conjecture. That is, it was proved that the local statistics of the eigenvalues of random matrices with the probability measure $\exp[-N\ \mathrm{Tr}\ (M^2/2 + tM^4)]dM$ converge as $N \to \infty$ to a universal random point process, whose correlation functions are given in terms of the so-called sine kernel.

In addition, in [6]-[8], recent techniques for the asymptotic analysis of Riemann-Hilbert problems were applied to the problem of determining asymptotics for a wide class of orthogonal polynomials, which contain the orthogonal polynomials $p_n(z;\ N,\ \mathbf{t})$ as a subset. The results were used to establish this universality conjecture of random matrix theory, for a family of real analytic probability measures. As a by-product of the analysis in [8], asymptotics for the recurrence coefficients $b_{N-1}^{(N)}$ were obtained.

The upshot of this is that there is emerging a new way to compute asymptotics of $\log(Z_N(\mathbf{t}))$. In both [4] and [6]-[8], an important asymptotic technique is the Deift-Zhou steepest descent / stationary phase method for the asymptotic analysis of Riemann-Hilbert problems. This method was introduced in [11], and further developed in [12] and [10].

One way to proceed is to compute the asymptotics for the orthogonal polynomials, and then deduce the asymptotics for the recurrence coefficients $b_j^{(N)}$. However, this still requires an analysis of the recurrence coefficients which is *uniform* in j, i.e. for $j = 0$ through $j = N - 1$. This appears to be an onerous task, and so we shall not adopt this approach here. Instead, we will proceed, starting in the next subsection, by the direct analysis of yet another representation of $\log(Z_N)$ (formula (1.34)).

1.6. **A statistical mechanical formula, and orthogonal polynomials.** In this paper we will analyze the large N behavior of $Z_N(\mathbf{t})$ by direct analysis of the formula (1.33) for the logarithmic derivative of Z_N. The derivation of this formula proceeds as follows. One begins by computing the logarithmic derivative of Z_N:

$$\frac{\partial}{\partial t_\ell}\log(Z_N) = \frac{1}{Z_N}\int\cdots\int\left(-N\sum_{j=1}^{N}\lambda_j^\ell\right) \times \exp\left\{-N^2\left[\frac{1}{N}\sum_{j=1}^{N}V_{\mathbf{t}}(\lambda_j) - \frac{1}{N^2}\sum_{j\neq\ell}\log|\lambda_j-\lambda_\ell|\right]\right\}d^N\lambda. \tag{1.30}$$

Of course, this has the immediate interpretation as an expectation value, with respect to the probability measure (1.5):

$$\frac{\partial}{\partial t_\ell}\log(Z_N) = \mathbb{E}_N\left(-N\sum_{j=1}^{N}\lambda_j^\ell\right). \tag{1.31}$$

Now it is well known (see, for example, [21]) that expectation values of functions of the λ_js can be expressed in terms of the associated orthogonal polynomials. In particular, the mean density of the λ_j's is given by the so-called "one-point function",

$$\rho_N^{(1)}(\lambda) = \frac{1}{N}\exp[-NV_{\mathbf{t}}(\lambda)]\sum_{j=0}^{N-1}p_j(\lambda;\ N,\mathbf{t})^2. \tag{1.32}$$

Using this connection, one has the following remarkable formula for the partial derivatives of $\log(Z_N)$:

$$\frac{\partial}{\partial t_\ell}\log(Z_N) = -N^2\mathbb{E}\left(\frac{1}{N}\text{ Tr } M^\ell\right) = -N^2\int_{-\infty}^{\infty}\lambda^\ell \rho_N^{(1)}(\lambda)d\lambda \tag{1.33}$$

in which the first equality follows from (1.7). The fundamental theorem of calculus readily implies

$$Z_N(\mathbf{t}) = Z_N(\mathbf{0})\exp\left\{-N^2\int_0^{\mathbf{t}}\int_{\mathbb{R}}\rho_1^{(N)}(\lambda)\,\overrightarrow{\nabla_t}V d\lambda\cdot\vec{d\ell}\right\}. \tag{1.34}$$

Since $Z_N(\mathbf{t}=0)$ is explicitly known ($Z_N(\mathbf{0})=\cdots$), the formula (1.34) shows that if we have global asymptotics for $\rho_1^{(N)}$, we can deduce an asymptotic expansion for $Z_N(\mathbf{t})$. We will analyze the one-point function, globally on the real axis, and determine a uniform asymptotic representation for it. Then we will evaluate the integral appearing in the right hand side of (1.33), and from this obtain a uniform asymptotic expansion for (1.34). This is the procedure by which we shall compute the asymptotic expansion of $\log Z_N$.

The following theorem is a fundamental result which, following the above prescription, yields Theorem 1.1.

Theorem 1.3. *There is $T>0$ and $\gamma>0$ so that for all $\mathbf{t}\in\mathbb{T}(T,\gamma)$, the following expansion holds true:*

$$\int_{-\infty}^{\infty} f(\lambda)\rho_N^{(1)}(\lambda)d\lambda = f_0 + N^{-2}f_1 + N^{-4}f_2+\cdots, \tag{1.35}$$

provided the function $f(\lambda)$ is C^∞ smooth, and grows no faster than a polynomial for $\lambda\to\infty$. The coefficients f_j depend analytically on $\mathbf{t}$ for $\mathbf{t}\in\mathbb{T}(T,\gamma)$, and the asymptotic expansion may be differentiated term by term.

In the statement of the above Theorem, the meaning of the asymptotic expansion is this: if you stop after the term $N^{-2j}f_j$, the error term is $\mathcal{O}(N^{-(2j+2)})$.

Remark 1.4. We note that in some very special instances results of the above form may be deduced from known results in the literature. For instance, taking the Gaussian limit of (1.33) one has

$$\frac{\partial}{\partial t_{2k}}\log(Z_N)|_{\mathbf{t}=\mathbf{0}} = -N^2\int_{-\infty}^{\infty}\lambda^{2k}\rho_N^{(1)}(\lambda,\mathbf{t}=\mathbf{0})d\lambda = -N^2\langle\frac{1}{N}N^{-k}\text{ Tr } M^{2k}\rangle.$$

The last term, $-N^{-k+1}\langle\text{ Tr } M^{2k}\rangle$, which by formula (1.15) is equal to a *finite* series, $-N^2\sum_{g=0}^{[k/2]}\varepsilon_g(k)\left(\frac{1}{N^2}\right)^g$, is a generating function for counting one-face maps with k edges. This series is manifestly even in N. In the setting of (1.33) this series is actually a generating function for counting $2k$-valent one-vertex maps (see also (2.24)-(2.26)). There is no inconsistency in this since by duality these two countings are equivalent.

Remark 1.5. Additionally, in [2], Albeverio et al have recently considered asymptotic expansions of the form (1.35), for the special function $f_z(\lambda)=(z-\lambda)^{-1}$, under the assumption that the external field is even, but they allow for the equilibrium measure to be supported on one or two intervals. They assert an asymptotic expansion for such integrals which is in inverse powers of N; our analysis shows that there is a complete expansion, in inverse powers of N^2.

The paper is organized as follows. In Section 2 we present a self-contained description of the connection between random matrix integrals and the enumeration of maps on Riemann surfaces, according to genus and valence; we also prove Theorem 1.2, using Theorem 1.1. In Section 3 we summarize the procedure developed in [6]-[8] to establish rigorous uniform global asymptotics for the associated orthogonal polynomials. In Section 4 we use the results of Section 3 to derive exact formulae for $\rho_N^{(1)}(\lambda)$, the mean density of eigenvalues at finite N, in terms of the Riemann–Hilbert procedure. Explicit in the formulae derived is *a complete asymptotic expansion* for $\rho_N^{(1)}$. In Section 5, we use the formulae derived in Section 4, to establish Theorem 1.3.

2. Enumeration of g-maps

In this section we will explain the connection between the evaluation of the Gaussian moments $\langle\prod_{j=1}^{\nu}(TrM^j)^{n_j}\rangle$, and g-diagrams and g-maps (as defined in subsection 1.3.3). Following that we will prove Theorem 1.2 which states that the asymptotic expansion (1.3) is a generating function for g-maps, enumerating them according to the number of vertices with different valences and the genus.

2.1. **Gaussian matrix integrals and enumeration of diagrams.** For simplicity we first discuss in detail the case of the pure moment $(TrM^4)^n$; afterward we will indicate how this extends to general mixed moments.

The expression $(TrM^4)^n$ has the form

$$\sum m_{i_1,j_1} m_{j_1,k_1} m_{k_1,\ell_1} m_{\ell_1,i_1} m_{i_2,j_2} m_{j_2,k_2} \dots m_{i_n,j_n} m_{j_n,k_n} m_{k_n,\ell_n} m_{\ell_n,i_n} \tag{2.1}$$

where the sum is taken over all configurations, where by a *configuration* one means an assignment to each of the $4n$ labels appearing in the list $\{i_\nu, j_\nu, k_\nu, \ell_\nu\}_{\nu=1}^n$ a value from $\{1,\dots,N\}$. There are thus N^{4n} configurations. Each configuration corresponds to a selection of $4n$ matrix entries.

We compute the expectation of this quantity with respect to the original probability measure (1.13), interchanging integration with summation:

$$\begin{aligned} &\left\langle \left(\mathrm{Tr}\ M^4\right)^n \right\rangle \\ &= \sum_{\text{configs}} \langle m_{i_1,j_1} m_{j_1,k_1} m_{k_1,\ell_1} m_{\ell_1,i_1} m_{i_2,j_2} m_{j_2,k_2} \dots m_{i_n,j_n} m_{j_n,k_n} m_{k_n,\ell_n} m_{\ell_n,i_n} \rangle \end{aligned} \tag{2.2}$$

By the Wick calculus, the expectation of each term in the above sum can itself be evaluated as a sum over Wick couplings as they were described in Section 1.3.1. To appreciate this one should first observe that each of the matrix entries appearing in the expectation on the right hand side of (2.2) is a linear function. Next, we introduce the following notation:

$$\begin{aligned} &\langle m_{i_1,j_1} m_{j_1,k_1} m_{k_1,\ell_1} m_{\ell_1,i_1} m_{i_2,j_2} m_{j_2,k_2} \dots m_{i_n,j_n} m_{j_n,k_n} m_{k_n,\ell_n} m_{\ell_n,i_1} \rangle \\ &= \langle f_1 f_2 \cdots f_{4n} \rangle , \end{aligned} \tag{2.3}$$

where

$$f_1 = m_{i_1,j_1}, \quad f_2 = m_{j_1,k_1}, \quad f_3 = m_{k_1,\ell_1}, \quad f_4 = m_{\ell_1,i_1}, \tag{2.4}$$

and for $\nu = 2,\dots,n$,

$$f_{4(\nu-1)+1} = m_{i_\nu,j_\nu}, \tag{2.5}$$

$$f_{4(\nu-1)+2} = m_{j_\nu,k_\nu}, \tag{2.6}$$

$$f_{4(\nu-1)+3} = m_{k_\nu,\ell_\nu}, \tag{2.7}$$

$$f_{4(\nu-1)+4} = m_{\ell_\nu,i_\nu}. \tag{2.8}$$

The definitions (2.4)-(2.8) implicitly define an invertible mapping $(r,s)(\cdot) = (r(\cdot), s(\cdot))$ between the integers $\{1,\dots 4n\}$ and the matrix indices $\{(i_\nu,j_\nu),\ (j_\nu,k_\nu),\ (k_\nu,\ell_\nu),\ (\ell_\nu,i_\nu)\}$ which we will describe as follows:

$$\begin{aligned} &(r,s)(\cdot) = (r(\cdot), s(\cdot)) \ : \ \{1,\dots,4n\} \to \{(i_\nu,j_\nu),(j_\nu,k_\nu),(k_\nu,\ell_\nu),(\ell_\nu,i_\nu)\}_{\nu=1}^n, \\ &k \mapsto (r(k),s(k)),\ \text{so that } f_k = m_{r(k),s(k)} \ \text{(using (2.4)-(2.8))}. \end{aligned} \tag{2.9}$$

We now apply the Wick calculus:

$$\begin{aligned} &\sum_{\text{configs}} \langle f_1 \cdots f_{4n} \rangle \\ &= \sum_{\text{configs}} \sum_{\omega \in W_{4n}} \left\langle f_{\omega_{1,1}} f_{\omega_{1,2}} \right\rangle \left\langle f_{\omega_{2,1}} f_{\omega_{2,2}} \right\rangle \cdots \left\langle f_{\omega_{2n,1}} f_{\omega_{2n,2}} \right\rangle \end{aligned} \tag{2.10}$$

In this setting a Wick coupling, $\omega = \omega_{i,j}$, $i = 1,\dots,2n$, $j = 1,2$, is a partition of $1,2,\dots,4n$ into $2n$ couples $(\omega_{i,1},\omega_{i,2})$ so that $\omega_{1,1} < \omega_{2,1} < \cdots < \omega_{2n,1}$, and $\omega_{i,2} > \omega_{i,1}$ for $i = 1,\dots,2n$. We denote the set of all such Wick couplings of the integers $1,\dots,4n$ by W_{4n}.

Now the outer summation over configurations appearing on the right hand side of (2.10) is over the different realizations of the variables $f_1,\dots,f_{4n}$ determined by the configuration. We may interchange orders of summation:

$$\left\langle \left(\text{Tr } M^4\right)^n \right\rangle = \sum_{\text{configs}} \langle f_1 \cdots f_{4n} \rangle = \sum_{\omega \in W_{4n}} \sum_{\text{configs}} \left\langle f_{\omega_{1,1}} f_{\omega_{1,2}} \right\rangle \left\langle f_{\omega_{2,1}} f_{\omega_{2,2}} \right\rangle \cdots \left\langle f_{\omega_{2n,1}} f_{\omega_{2n,2}} \right\rangle . \tag{2.11}$$

The individual terms

$$\left\langle f_{\omega_{1,1}} f_{\omega_{1,2}} \right\rangle \left\langle f_{\omega_{2,1}} f_{\omega_{2,2}} \right\rangle \cdots \left\langle f_{\omega_{2n,1}} f_{\omega_{2n,2}} \right\rangle \tag{2.12}$$

are either 0 or 1. For each Wick coupling ω, certain configurations contribute a 1 to the inner summation, and other configurations do not contribute. The question of evaluating the original integral $\left\langle \left(\text{Tr } M^4\right)^n \right\rangle$ has been reduced to computing, for each Wick ordering $\omega \in W_{4n}$, the number of configurations providing unit contribution to the inner sum appearing in (2.11). For each different Wick coupling ω, it turns out that this number is of the form $N^{F(\omega)}$, and accepting this for the moment, we have the formula

$$\sum_{\text{configs}} \left\langle f_{\omega_{1,1}} f_{\omega_{1,2}} \right\rangle \left\langle f_{\omega_{2,1}} f_{\omega_{2,2}} \right\rangle \cdots \left\langle f_{\omega_{2n,1}} f_{\omega_{2n,2}} \right\rangle = N^{F(\omega)}. \tag{2.13}$$

To determine the contribution $N^{F(\omega)}$ for each Wick coupling, and to explain the quantity $F(\omega)$, let us first recall the direct relationship (2.4)-(2.8) between $\{f_j\}$ and the original variables of integration $\{m_{i_\nu,j_\nu}, m_{j_\nu,k_\nu}, m_{k_\nu,\ell_\nu}, m_{\ell_\nu,i_\nu}\}_{\nu=1}^n$, and the identities (1.14). Using the mapping $(r,s)(\cdot)$ defined in (2.9) , the inner sum in (2.11), over all configurations, is of the form

$$\sum_{\text{configs}} \left\langle m_{r(\omega_{1,1}),s(\omega_{1,1})} m_{r(\omega_{1,2}),s(\omega_{1,2})} \right\rangle \left\langle m_{r(\omega_{2,1}),s(\omega_{2,1})} m_{r(\omega_{2,2}),s(\omega_{2,2})} \right\rangle \cdots \cdots \left\langle m_{r(\omega_{2n,1}),s(\omega_{2n,1})} m_{r(\omega_{2n,2}),s(\omega_{2n,2})} \right\rangle . \tag{2.14}$$

Each different configuration appearing in (2.14) is a different choice of the matrix indices $\{(i_\nu, j_\nu)$, (j_ν, k_ν), (k_ν, ℓ_ν), $(\ell_\nu, i_\nu)\}$, and for each one we then ask if the contribution is 0 or 1. Using (1.14), the contribution will be 1 if and only if for each $\mu = 1, \ldots, 2n$,

$$\left\langle m_{r(\omega_{\mu,1}),s(\omega_{\mu,1})} m_{r(\omega_{\mu,2}),s(\omega_{\mu,2})} \right\rangle = 1, \tag{2.15}$$

and this is true if and only if

$$r(\omega_{\mu,1}) = s(\omega_{\mu,2}) \text{ and } s(\omega_{\mu,1}) = r(\omega_{\mu,2}) \text{ for each } \mu = 1, \ldots, 2n. \tag{2.16}$$

The sum (2.14) is precisely the number of configurations for which the $4n$ equalities (2.16) hold true, and since $r(\omega_{\mu,1})$ is, for each μ, one of $\{i_\nu, j_\nu, k_\nu, \ell_\nu\}_{\nu=1}^n$ (and similarly for $s(\omega_{\mu,2})$, $s(\omega_{\mu,1})$, and $r(\omega_{\mu,2})$), it follows that the sum over configurations, is now over

$$i_\nu, j_\nu, k_\nu, \ell_\nu = 1, \ldots N,$$
so that (2.16) holds.

The calculation of (2.14) now proceeds as follows. For each Wick coupling ω we write out equalities amongst the $4n$ indices of summation $\{i_\nu, j_\nu, k_\nu, \ell_\nu\}_{\nu=1}^n$ enforced by (2.16) and group them in "chains" (for example, one might have $i_3 = j_5 = \ell_7 \cdots$). Each index of summation appears in two equalities (because we are computing $\left\langle \left(\text{Tr } M^4\right)^n \right\rangle$, see (2.2)) and so each distinct chain is, fact, a closed cycle. For each closed cycle, the original summation is reduced to a summation over a single free parameter, ranging from 1 to N. Finally then, the summation (2.14) is $N^{F(\omega)}$ where $F(\omega)$ is precisely the number of closed cycles determined by the Wick coupling ω, and we have explained (2.13).

Summarizing, we have shown that

$$\left\langle \left(\text{Tr } M^4\right)^n \right\rangle = \sum_{\omega \in W_{4n}} N^{F(\omega)}, \tag{2.17}$$

where $F(\omega)$ is the number of closed cycles determined by ω via the equalities (2.16). Note that the right hand side of (2.17) is a finite sum, as W_{4n} is a finite set.

A key insight of the classic paper of [3] was to realize that counting the Wick couplings could be reduced to listing all possible diagrams (as defined in between (1.17) and (1.18) above). To see this one begins with a collection of n 4-valent vertex configurations labelled as in Figure 2.18. A given coupling determines a unique way of connecting the outgoing and incoming rays of each of the n 4-valent vertices, to form "roads". The result of such a glueing will be a diagram. Indeed, if we interpret each road in Figure 2.18 as an edge, and each edge is labelled by the "outgoing" side of the road, we clearly have a diagram. (The "faces" of the diagram correspond 1-1 to the closed index cycles of that coupling.) Conversely, given a diagram as defined, which includes the labellings, we can read off directly, from the way the incident edges are paired, what the couplings must be. This gives a bijection between diagrams and Wick couplings.

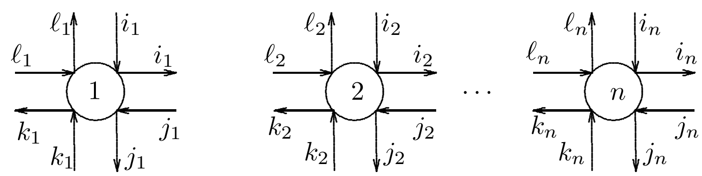

Figure 2.18.

The full count of non-vanishing contributions is then given by summing over diagrams with a weighting, $N^{F(\omega)}$, that counts the number of configurations that give rise to a nonvanishing contribution in the coupling, ω, associated to this diagram. Also the exponent of the weighting, $F(\omega)$, which was defined to be the number of closed index cycles in the coupling is now also seen to be the number of faces, F, in the associated diagram. The weighting factor is N^F. If the diagram is connected (i.e., if it is a map), this can be expressed in terms of the genus of the Riemann surface that the diagram maps to by using Euler's formula. Since the number of vertices of a diagram is n and the number of edges is $2n$ (= the number of couples in a Wick coupling), we have $2-2g = n-2n+F$ from which we deduce that the weighting factor for a 4-valent, n-vertex, g-map is N^{2-2g+n}. We can directly extend this to the case of *disconnected* diagrams since the Euler characteristic is additive with respect to disjoint unions. This generalization results in the possibility that g can become negative. Thus, to each diagram we can associate a unique integer g and then we will refer to that diagram as a g-diagram.

The beautiful connection between the quantity $\left\langle (\text{Tr } M^4)^n \right\rangle$ and the combinatorics of g-diagrams described in [3] (and discussed above) is summarized with the following formula:

$$\left\langle \left(\text{Tr } M^4\right)^n \right\rangle = \sum_g \#\{\text{4-valent, n-vertex, g-diagrams}\} N^{2-2g+n}. \tag{2.19}$$

We remark that this is a finite sum, and the reader may easily verify that the nonzero contributions to this sum come from $1-n \leq g \leq [(n+1)/2]$, where $[\ell]$ is the closest integer to ℓ less than or equal to ℓ. A straightforward extension of the above analysis to the case of mixed moments shows that

$$\left\langle \prod_{j=1}^{\nu} \left(\text{Tr } M^j\right)^{n_j} \right\rangle = \sum_g \#\{\text{g-diagrams with } n_j \text{ } j\text{-valent vertices; } j = 1,\dots,\nu\} N^{2-2g+\sum_{j=1}^{\nu}(\frac{j}{2}-1)n_j}. \tag{2.20}$$

(Again we observe that this is a finite sum.)

2.2. **Asymptotics of the partition function and enumeration of g-maps.** We turn now to the proof of Theorem 1.2. The starting point is the asymptotic expansion in Theorem 1.1, rewritten in the form

$$\log\left(\hat{Z}_N\right) - N^2 e_0(\mathbf{t}) - e_1(\mathbf{t}) = \sum_{g=2}^{\infty} e_g(\mathbf{t}) \left(\frac{1}{N}\right)^{2g-2}.$$

This is, as usual, shorthand: the right hand side is an asymptotic expansion in *inverse* powers of N^2. Exponentiating both sides of this equation we deduce that $Z_N e^{-N^2 e_0 - e_1}$ has an asymptotic expansion in inverse powers of N^2. Some algebraic manipulation shows that

$$\hat{Z}_N = \exp\left(N^2 e_0 + e_1\right)\left\{1 + \sum_{h=1}^{\infty} \rho_h \left(\frac{1}{N^2}\right)^h\right\} \tag{2.21}$$

where

$$\rho_h = \sum_{k=1}^{h} \frac{1}{k!} \sum_{\ell_1+\ldots+\ell_k=h, \ell_j>0} \prod_{j=1}^{k} e_{\ell_j+1}.$$

Formula (2.21) may appear to be nothing more than a simple calculational re-representation of $\hat{Z}_N$. However (2.21) explains concretely in what sense $\hat{Z}_N$ itself possesses an asymptotic representation.

It follows from Theorem 1.1 that the asymptotic series on the right hand side of (2.21) can be differentiated term-by-term for $\mathbf{t} \in \mathbb{T}(T,\gamma)$, and the ensuing derivatives may be evaluated at $\mathbf{t} = \mathbf{0}$. Having established this we now rewrite (2.21) as

$$\hat{Z}_N(\mathbf{t}) = \exp\left(N^2 e_0(\mathbf{t}) + e_1(\mathbf{t}) + \frac{1}{N^2} e_2(\mathbf{t}) + \cdots + \frac{1}{N^{2h-2}} e_h(\mathbf{t}) + \ldots\right). \tag{2.22}$$

In light of the above discussion, it is clear what is meant by this asymptotic expansion, which is valid for $\mathbf{t}$ satisfying the regularity conditions of Theorem 1.1.

Differentiating (2.22) n_j times in each t_j and taking the limit $\mathbf{t} \to \mathbf{0}$, we may now conclude, by comparison with (2.20), that

$$\begin{aligned} \frac{\partial^n}{\partial t_1^{n_1} \ldots \partial t_\nu^{n_\nu}} Z_N(\mathbf{t})\Big|_{\mathbf{t}=\mathbf{0}} &= \left\langle \prod_{j=1}^{\nu} \left(\mathrm{Tr}\ M^j\right)^{n_j} \right\rangle N^{-\sum_{j=1}^{\nu}(\frac{j}{2}-1)n_j} \\ &= \sum_g \#\{\text{g-diagrams with } n_j\ j\text{-valent vertices; } j = 1,\ldots,\nu\} N^{2-2g} \\ &= (-1)^n \frac{\partial^n}{\partial t_1^{n_1} \ldots \partial t_\nu^{n_\nu}} \left[\exp\left(\sum_{g=0}^{\infty} N^{2-2g} e_g(\mathbf{t})\right)\right]_{\mathbf{t}=\mathbf{0}}, \end{aligned} \tag{2.23}$$

where $n = n_1 + \cdots + n_\nu$.

The string of equalities in (2.23) states that the finite sum over g in the second line possesses an asymptotic expansion in inverse powers of N^2. We learn that all of the coefficients in this final asymptotic expansion must vanish from some point on. Much more information can be obtained in this manner from (2.23). For example, by examining (2.23) for the case of a single derivative, one may verify that for ℓ even,

$$\frac{\partial}{\partial t_\ell} e_0(\mathbf{0}) = -\#\{\text{ genus 0 diagrams with 1 } \ell\text{-valent vertex}\} \tag{2.24}$$

$$\frac{\partial}{\partial t_\ell} e_1(\mathbf{0}) = -\#\{\text{ genus 1 diagrams with 1 } \ell\text{-valent vertex}\} \tag{2.25}$$

$$\frac{\partial}{\partial t_\ell} e_g(\mathbf{0}) = 0 \quad \text{for } g > \frac{1}{2}\begin{cases} \ell/2 & \text{if } \ell/2 \text{ is even} \\ \ell/2 - 1 & \text{if } \ell/2 \text{ is odd} \end{cases}. \tag{2.26}$$

We observe that in (2.24) and (2.25), the diagrams have a single vertex, and so we may say that (2.24) and (2.25) enumerate 1-vertex g-maps.

Proceeding further, one may examine (2.23) for the case of higher order derivatives. Tedious calculations may be carried out which show that the e_g's are generating functions for g-maps (they enumerate those diagrams that are connected). However, this approach is quite complicated, and we adopt a different approach, which is to use (2.23) to discuss the relation between the logarithm of the partition function and connected diagrams, or g-maps, that was alluded to in (1.19). For the sake of clarity we will present this just in the case of $\mathbf{t} = (0, 0, 0, t_4 = t, 0, \ldots, 0)$; it will be evident from the discussion that the extension to more general valences is straightforward.

We observe that, from our earlier discussion of the relation between diagrams and Wick orderings, that:

$$\sum_g \# \quad \{\text{4-valent, n-vertex, g-diagrams}\} N^{2-2g} = \sum_{\omega \in W_{4n}} N^{\chi(\omega)} := Q_n(N). \tag{2.27}$$

$Q_n(N)$ is a Laurent polynomial which we will refer to as the *diagram polynomial* for 4-valent, n-vertex diagrams. We introduce an *exponential generating function* [16] for diagram polynomials:

$$G(t) := \sum_{k=1}^{\infty} \frac{t^k}{k!} Q_k(N). \tag{2.28}$$

This is a formal generating function in the same sense as was discussed for (1.18). It is by no means a convergent series, but rather is interpreted in the sense that one is allowed to evaluate finitely many derivatives of both sides at $t = 0$. (Below we will indicate how to remove the formality of these arguments by truncating the series.) We also introduce an exponential generating function for *connected* 4-valent, diagrams (i.e. maps):

$$C(t) := \sum_{k=1}^{\infty} \frac{t^k}{k!} P_k(N),$$

where

$$P_n(N) = \sum_{\omega \in W_{4n}^{conn}} N^{\chi(\omega)} \tag{2.29}$$

and W_{4n}^{conn} is defined to be the subset of W_{4n} consisting of Wick couplings that correspond to connected diagrams. We observe that

$$\begin{aligned} C(t)C(t) &= \sum_{n=2}^{\infty} \sum_{k+\ell=n} \frac{t^n}{k!\ell!} P_k(N) P_\ell(N) \\ &= \sum_{n=2}^{\infty} \frac{t^n}{n!} \sum_{k=1}^{n-1} \begin{pmatrix} n \\ k \end{pmatrix} P_k(N) P_{n-k}(N). \end{aligned}$$

This series is a generating function for diagram polynomials of ordered pairs (ω_1, ω_2) of connected Wick couplings. There is a geometrical explanation for the appearance of the combinatorial coefficient $\begin{pmatrix} n \\ k \end{pmatrix}$ in this expression. It is a consequence of the fact that the labels $1, \ldots n$ get distributed over $\omega_1 \cup \omega_2$ and if $|\omega_1| = k$ there are $\begin{pmatrix} n \\ k \end{pmatrix}$ different ways of doing this.

Next, one may prove that $C(t)C(t)/2$ is the exponential generating function for 4-valent diagrams which consist of exactly two connected components. In other words, one may define an exponential generating function for 4-valent diagrams consisting of exactly two connected components, and prove that this generating function coincides, term-by-term, with $C(t)C(t)/2$. Subsequently, and in the same way, one may prove that $C^d(t)/d!$ is the exponential generating function for 4-valent, n-vertex diagrams with exactly d connected components. One may finally conclude that, in the sense of formal generating functions,

$$G(t) = \sum_{d=1}^{\infty} \frac{1}{d!} C^d(t) = \exp(C(t)) - 1. \tag{2.30}$$

To get around the formality of these generating functions, we repeat most of the argument from (2.28) to (2.30), truncating the infinite series involved. So we define

$$G_m(t) = \sum_{k=1}^{m} \frac{t^k}{k!} Q_k(N), \qquad C_m(t) = \sum_{k=1}^{m} \frac{t^k}{k!} P_k(N), \tag{2.31}$$

with Q_k and P_k defined in (2.27) and (2.29), respectively. Next we have that

$$\frac{1}{2}C_m(t)^2 = \sum_{k=2}^{m} \frac{t^k}{k!} P_k^{(2)}(N) + \mathcal{O}\left(t^{m+1}\right), \tag{2.32}$$

where the sum on the right hand side is the truncated exponential generating function for 4-valent diagrams consisting of exactly two connected components (i.e. $P_k^{(2)}(N)$ is the diagram polynomial for 4-valent, k-vertex diagrams with two connected components). Similarly, the reader may verify that

$$\frac{1}{d!}C_m(t)^d = \sum_{k=d}^{m} \frac{t^k}{k!} P_k^{(d)}(N) + \mathcal{O}\left(t^{m+1}\right), \tag{2.33}$$

where on the right hand side, $P_k^{(d)}(N)$ is the diagram polynomial for 4-valent, k-vertex diagrams with d connected components. Putting this all together, we have shown that

$$G_m(t) = e^{C_m(t)} + \mathcal{O}\left(t^{m+1}\right). \tag{2.34}$$

And since m was arbitrary, this is the proper sense in which one may interpret (2.30).

We have established the following set of relations:

$$1 + G_m(-t) = Z_N(t) + \mathcal{O}\left(t^{m+1}\right), \tag{2.35}$$

$$\exp C_m(-t) = 1 + G_m(-t) + \mathcal{O}\left(t^{m+1}\right). \tag{2.36}$$

We therefore have

$$C_m(-t) = \log\left(Z_N + \mathcal{O}\left(t^{m+1}\right)\right). \tag{2.37}$$

We may now differentiate this n times (for $n \leq m$) to obtain

$$\left.\frac{\partial^n}{\partial t^n} C_m(-t)\right|_{t=0} = P_n(N) = \left.\frac{\partial^n}{\partial t^n}\log Z_N(t)\right|_{t=0}. \tag{2.38}$$

So we have proven that $\log Z_N$ is the generating function for *connected* diagrams. Now using our asymptotic expansion for $\log Z_N$, we have

$$\begin{aligned} (-1)^n \frac{\partial^n}{\partial t^n} \sum_{g=0}^{\infty} N^{2-2g} e_g(t)\Big|_{t=0} &= P_n(N) \\ &= \sum_{\omega \in W_{4n}^{\text{conn}}} N^{\chi(\omega)} = \sum_g N^{2-2g} \#\{\text{4-valent, n-vertex, g-maps}\}. \end{aligned} \tag{2.39}$$

We have now shown that $e_g(t) = E_g(t)$ in the sense that

$$(-1)^n \frac{\partial^n}{\partial t^n} e_g(0) = \#\{\text{4-valent, n-vertex, g-maps}\}.$$

By similar reasoning one can fully establish Theorem 1.2, and we will not present these details here.

We refer the reader to [16] for a fuller discussion of the general cumulant relations as they pertain to graphical enumeration.

3. Asymptotics for orthogonal polynomials via Riemann–Hilbert methods

The amazing fact, due to Gaudin and Mehta [14], is that the mean density $\rho_N^{(1)}$ can be expressed in closed form in terms of the orthogonal polynomials $\{p_j(\cdot; N, \mathbf{t}\ )\}$.

$$\rho_N^{(1)}(\lambda) = \frac{1}{N} e^{-NV(\lambda)} \sum_{k=0}^{N-1} p_k(\lambda)^2 \tag{3.1}$$

$$= \frac{1}{N} e^{-NV}\left[p_N'(\lambda)\, p_{N-1}(\lambda) - p_N(\lambda)\, p_{N-1}'(\lambda)\right] \frac{\gamma_{N-1}^{(N)}}{\gamma_N^{(N)}}. \tag{3.2}$$

The second equality above follows from the first because of the Christoffel-Darboux formula. (See, for example, formula (3.2.3) in [24].)

We obtain global asymptotics for $\rho_N^{(1)}$ by carrying out to higher order the results of [6]-[8]. Deift, Kriecherbauer, McLaughlin, Venakides, and Zhou established very detailed Plancherel-Rotach type asymptotics for polynomials orthogonal with respect to $e^{-NV(x)}dx$, under some very general assumptions on the function V, including the functions V considered here. In [7], the authors demonstrated that their method can be used to obtain a complete asymptotic expansion for the polynomials. We have carried this procedure out to higher order.

Using the global asymptotic expansion of $\rho_N^{(1)}$, together with (1.34), we deduce an asymptotic expansion for $\log Z_N(\mathbf{t}\,)$.

In this section we review the Riemann–Hilbert approach for computing and establishing a global, uniform asymptotic expansion for the orthogonal polynomials. In [6]-[8], the authors began with a matrix valued function $Y(z)$, defined by

$$Y = \begin{pmatrix} \frac{1}{\gamma_N^{(N)}}\, p_N & \frac{1}{2\pi i \gamma_N^{(N)}} \int \frac{p_N(s) e^{-NV(s)}}{s-z} ds \\ -2\pi i \gamma_{N-1}^{(N)} p_{N-1} & -\gamma_{N-1}^{(N)} \int \frac{p_{N-1}(s)\, e^{-NV(s)}}{s-z} ds \end{pmatrix}, z \in \mathbb{C}\backslash\mathbb{R}.$$

The matrix Y solves the following Riemann–Hilbert problem:

- Y analytic in $\mathbb{C}\backslash\mathbb{R}$,
- $Y = \left(I + \mathcal{O}(\frac{1}{z})\right) z^{N\sigma_3}$,
- Y has Hölder continuous boundary values $Y_\pm$ for $z \in \mathbb{R}$,
- $Y_+ = Y_- \begin{pmatrix} 1 & e^{-NV} \\ 0 & 1 \end{pmatrix}$, for $z \in \mathbb{R}$.

The authors then make two explicit transformations: $Y \to M \to M_1$. Here we will explain these two transformations.

3.1. **Equilibrium measure and first transformation Y→M.** . A fundamental object in the theory of random matrices, as well as approximation theory, is the equilibrium measure, defined as follows:

$$\sup_{\text{Borel measures } \mu, \mu \geq 0, \int d\mu = 1} \left[-\int V(\lambda) d\mu(\lambda) + \int \int \log|\lambda - \mu|\, d\mu(\lambda)\, d\mu(\eta) \right].$$

There is a vast literature on this variational problem, originating in the work of Gauss (see [22]). Under general assumptions on V, the supremum is achieved at a unique measure μ_V, called the equilibrium measure. For the external field V, considered here, it is a well-known fact (see [5], or [13]) that the equilibrium measure is supported on finitely many intervals, with density that is analytic on the interior of each interval, behaving at worst like a square root at each endpoint.

We remind the reader of the definition of the set $\mathbb{T}(T, \gamma)$:

$$\mathbb{T}(T, \gamma) = \{\mathbf{t} \in \mathbb{R}^\nu : |\mathbf{t}| \leq T,\ t_\nu > \gamma \sum_{j=1}^{\nu-1} |t_j|\} \tag{3.3}$$

The following theorem is perhaps well understood. A proof can be constructed by combining the results of [5] and [19].

Theorem 3.1. *There is $T_0 > 0$ and $\gamma_0 > 0$ such that for all $0 < T < T_0$ and $\gamma > \gamma_0$, the following holds true: If $\mathbf{t} \in \mathbb{T}(T, \gamma)$, then*

$$d\mu_V = \psi\, d\lambda,$$
$$\psi(\lambda) = \frac{1}{2\pi} \chi_{(\alpha,\beta)}(\lambda) \sqrt{(\lambda - \alpha)(\beta - \lambda)}\, h(\lambda),$$

where $h(\lambda)$ is a polynomial of degree $\nu - 2$, which is strictly positive on the interval $[\alpha, \beta]$ (recall that the external field V is a polynomials of degree ν). The polynomial h is defined by

$$h(z) = \frac{1}{2\pi i} \oint \frac{V'(s)}{\sqrt{(s-\alpha)}\sqrt{(s-\beta)}} \frac{ds}{s-z}$$

where the integral is taken on a circle containing (α, β) and z in the interior, oriented counter-clockwise.

The endpoints α and β are determined by the equations

$$\int_\alpha^\beta \frac{V'(s)}{\sqrt{(s-\alpha)(\beta-s)}}\,ds = 0$$
$$\int_\alpha^\beta \frac{sV'(s)}{\sqrt{(s-\alpha)(\beta-s)}}\,ds = 2\pi.$$

The endpoints $\alpha(\mathbf{t}\,)$ and $\beta(\mathbf{t}\,)$ are analytic functions of $\mathbf{t}$ which possess smooth extensions to the closure of $\mathbb{T}(T,\gamma)$, and satisfy $-\alpha(\mathbf{0}) = \beta(\mathbf{0}) = 2$. In addition, the coefficients of the polynomial $h(\lambda)$ are also analytic functions of $\mathbf{t}$ with smooth extensions to $\overline{\mathbb{T}(T,\gamma)}$, with

$$h(\lambda, \mathbf{t} = \mathbf{0}) = 1.$$

In this paper we will assume that $\mathbf{t} \in \mathbb{T}(T,\gamma)$, for T small enough, and γ large enough, so that Theorem 3.1 holds true. In Section 4 we define T_β and γ_β (as well as T_α and γ_α) so that if $T < \min\{T_\beta, T_\alpha\}$ and $\gamma > \max\{\gamma_\alpha, \gamma_\beta\}$ then $\mathbf{t} \in \mathbb{T}(T,\gamma)$ implies more: the neighborhoods of $z = \beta$ and $z = \alpha$ defined in this section can be taken so large that their intersection contains a subinterval of the real axis. We will not present a proof of Theorem 3.1 here, but the basic idea is as follows: for $\mathbf{t} = \mathbf{0}$, the equilibrium measure is explicitly known, $\psi(\lambda) = \chi_{(-2,2)}(\lambda)\sqrt{4-\lambda^2}/(2\pi)$. Taking T small implies that the endpoints of the support cannot stray too far from ± 2; taking t_ν larger than $\gamma \sum_{j=1}^{\nu-1} |t_j|$ implies that h should be strictly positive on $\mathbb{R}$.

It will prove useful to adapt the following alternative representation for the function ψ:

$$\psi(\lambda) = \frac{1}{2\pi i} R_+(\lambda) h(\lambda), \ \ \lambda \in (\alpha,\beta), \tag{3.4}$$

where the function $R(\lambda)$ is defined via $R(\lambda)^2 = (\lambda-\alpha)(\lambda-\beta)$, with $R(\lambda)$ analytic in $\mathbb{C}\backslash[\alpha,\beta]$, and normalized so that $R(\lambda) \sim \lambda$ as $\lambda \to \infty$. The subscript $\pm$ in $R_\pm(\lambda)$ denotes the boundary value obtained from the upper (lower) half plane.

Following [6], we define a function $g(z)$ as follows,

$$g(z) = \int_\alpha^\beta \log(z-s)\,\psi(s)ds, \quad z \in \mathbb{C}\backslash(-\infty,\beta], \tag{3.5}$$

where for each s, $\log(z-s)$ is chosen to be branched along $(-\infty, s)$, $\log(z-s) > 0$ if $z > s$.

We now list several important properties of the function $g(z)$:

- 1 g is analytic for $z \in \mathbb{C}\backslash(-\infty,\beta]$, with continuous boundary values $g_\pm(z)$ on $(-\infty,\beta]$.
- 2 There is a constant ℓ such that for $z \in [\alpha,\beta]$,

$$g_+ + g_- - V(z) = \ell,$$

and for $z \in \mathbb{R}\backslash[\alpha,\beta]$,

$$g_+ + g_- - V < \ell.$$

Indeed, for $z > \beta$,

$$g_+ + g_- - V - \ell = 2g - V - \ell = -\int_\beta^z Rhds, \tag{3.6}$$

and for $z < \alpha$,

$$g_+ + g_- - V - \ell = -\int_\alpha^z Rhds.$$

- 3 For $z \in [\alpha,\beta]$,

$$g_+ - g_- = 2\pi i \int_z^\beta \psi(s)\,ds,$$

and this function possesses an analytic continuation to a neighborhood of (α,β).

- 4 For $z > \beta$,

$$g_+ - g_- = 0,$$

and for $z < \alpha$,

$$g_+ - g_- = 2\pi i.$$

Using the function $g(z)$, and the constant ℓ, we define (cf. Theorem 5.57 in [7]) $M(z)$ via

$$Y(z) = e^{N\frac{\ell}{2}\sigma_3}\, M(z)\, e^{N(g-\frac{\ell}{2})\sigma_3}, \quad z \in \mathbb{C}\backslash\mathbb{R}.$$

3.2. **Second Transformation $\mathbf{M \to M_1}$.** Next, we define (cf. Section 6 of [7]) a lens-shaped region around (α, β) (see Figure 3.7)

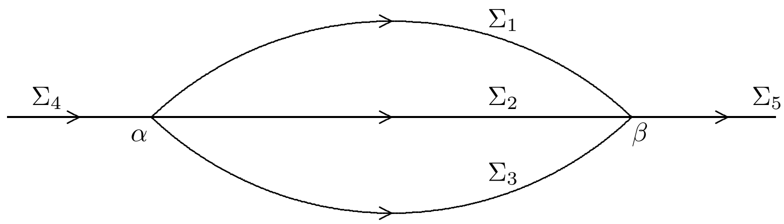

Figure 3.7. The contour Σ_S.

and define M_1 as follows:

- For z outside the upper and lower lenses,

$$M_1 = M. \tag{3.8}$$

- For z within the upper lens,

$$M_1 = M \begin{pmatrix} 1 & 0 \\ -e^{-N(g_+-g_-)} & 1 \end{pmatrix}. \tag{3.9}$$

- For z in the lower lens,

$$M_1 = M \begin{pmatrix} 1 & 0 \\ e^{N(g_+-g_-)} & 1 \end{pmatrix}. \tag{3.10}$$

(Recall that $g_+ - g_-$ possesses an analytic continuation to such a lens-shaped region around (α, β).)

With these two transformations, we arrive at a matrix-valued function M_1, satisfying the following Riemann–Hilbert problem.

The problem is to determine a 2×2 matrix valued function M_1 satisfying

(1.) M_1 analytic in $\mathbb{C}\backslash\Sigma_{M_1}$.
(2.) $M_1 = I + \mathcal{O}(\frac{1}{z})$, $z \to \infty$.
(3.) M_1 possesses continuous boundary values $(M_1)_\pm$ for $z \in \Sigma_{M_1}$.
(4.) (Jump relation) $(M_1)_+ = (M_1)_- V_M$,

where

$$V_M = \begin{pmatrix} 1 & e^{N(g_++g_--V-\ell)} \\ 0 & 1 \end{pmatrix}, \quad \text{for } z \in \Sigma_4 \cup \Sigma_5, \tag{3.11}$$

$$V_M = \begin{pmatrix} 1 & 0 \\ e^{-N(g_+-g_-)} & 1 \end{pmatrix}, \quad \text{for } z \in \Sigma_1, \tag{3.12}$$

$$V_M = \begin{pmatrix} 1 & 0 \\ e^{N(g_+-g_-)} & 1 \end{pmatrix}, \quad \text{for } z \in \Sigma_3, \tag{3.13}$$

$$V_M = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \text{for } z \in \Sigma_2, \tag{3.14}$$

3.3. **The construction of a global approximation to M_1.** The next step in the Riemann Hilbert approach is to construct an explicit approximation to M_1, which we will denote $M_1^{(A)}$. The definition of $M_1^{(A)}$ also is adapted directly from [6]-[8]. Let B_δ^α denote a disc of radius δ centered at α, and let B_δ^β denote a disc of radius δ centered at β. We will define $M_1^{(A)}$ outside $B_\delta^\alpha \cup B_\delta^\beta$ as follows:

$$M_1^{(A)}(z) = \begin{pmatrix} \frac{\gamma+\gamma^{-1}}{2} & \frac{\gamma-\gamma^{-1}}{2i} \\ \frac{\gamma-\gamma^{-1}}{-2i} & \frac{\gamma+\gamma^{-1}}{2} \end{pmatrix}, \tag{3.15}$$

$$\gamma(z) = \frac{(z-\beta)^{\frac{1}{4}}}{(z-\alpha)^{\frac{1}{4}}}. \tag{3.16}$$

To define $M_1^{(A)}(z)$ for z within B_δ^β, we require the following auxiliary function

$$\Phi_\beta(z) \equiv \left(\frac{3N}{4}\right)^{2/3} (-2g(z) + V(z) + \ell)^{2/3}, \quad z \in B_\delta^\beta. \tag{3.17}$$

As $h(\beta) \neq 0$, $-(2g(z) - V(z) - \ell)$ behaves like $(z-\beta)^{3/2}$ for z near β (cf. relations (3.6)), and we may choose the branch so that $\Phi_\beta(z)$ is analytic in a neighborhood of β, and

$$\Phi_\beta(z) = c(z-\beta) + O\left((z-\beta)^2\right), \ c > 0, \ z \to \beta. \tag{3.18}$$

Let the radius δ of the disc B_δ^β be chosen sufficiently small so that Φ_β maps B_δ^β injectively onto a neighborhood D' of 0. The decomposition of $D_{\epsilon,\beta}$ into four regions is as in [8]: $D_{\epsilon,\beta} = \mathrm{I} \cup \mathrm{II} \cup \mathrm{III} \cup \mathrm{IV}$, as shown in Figure 3.19 below.

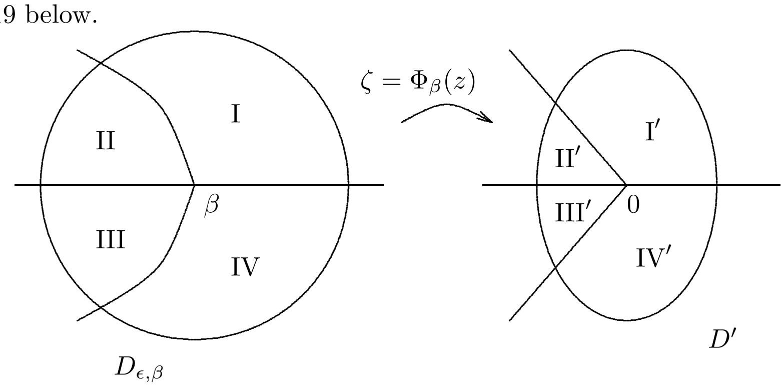

Figure 3.19. Decompositions of $D_{\epsilon,\beta}$ and D'.

The four rays on the right of Figure 3.19 divide the ζ–plane into four sectors, $0 < \arg(\zeta) < \frac{2\pi}{3}$, $\frac{2\pi}{3} < \arg(\zeta) < \pi$, $-\pi < \arg(\zeta) < \frac{-2\pi}{3}$ and $\frac{-2\pi}{3} < \arg(\zeta) < 0$. The intersection of these sectors with D' defines the regions I′, II′, III′ and IV′. (The precise angles separating the sectors are not important: we could, for example, replace $\frac{2\pi}{3}$ by any angle strictly between 0 and π, etc.) Finally, the regions I, II, III and IV are defined to be the pre–images under $\Phi(z)$ of I′, II′, III′ and IV′. We define the contour γ_σ in the ζ-plane to be the four rays shown in Figure 3.19, with orientation inherited from the original contour Σ_S and the mapping Φ_β.

We will need the following auxiliary functions in order to define the global approximation.

Denote

$$\omega := e^{\frac{2\pi i}{3}}, \tag{3.20}$$

and define

$$\Psi^\sigma : \mathbb{C} \setminus \gamma_\sigma \to \mathbb{C}^{2\times 2}, \tag{3.21}$$

$$\Psi^\sigma(\zeta) = \begin{cases} \begin{pmatrix} Ai(\zeta) & Ai(\omega^2\zeta) \\ Ai'(\zeta) & \omega^2 Ai'(\omega^2\zeta) \end{pmatrix} e^{-\frac{\pi i}{6}\sigma_3} & , \text{ for } \zeta \in I, \\ \begin{pmatrix} Ai(\zeta) & Ai(\omega^2\zeta) \\ Ai'(\zeta) & \omega^2 Ai'(\omega^2\zeta) \end{pmatrix} e^{-\frac{\pi i}{6}\sigma_3} \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} & , \text{ for } \zeta \in II, \\ \begin{pmatrix} Ai(\zeta) & -\omega^2 Ai(\omega\zeta) \\ Ai'(\zeta) & -Ai'(\omega\zeta) \end{pmatrix} e^{-\frac{\pi i}{6}\sigma_3} \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} & , \text{ for } \zeta \in III, \\ \begin{pmatrix} Ai(\zeta) & -\omega^2 Ai(\omega\zeta) \\ Ai'(\zeta) & -Ai'(\omega\zeta) \end{pmatrix} e^{-\frac{\pi i}{6}\sigma_3} & , \text{ for } \zeta \in IV. \end{cases} \tag{3.22}$$

Here σ_3 is the Pauli matrix $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$.

The next auxiliary quantity which we define is the analytic matrix valued function $E_\beta(z)$, defined as follows.

$$E_\beta(z) = \sqrt{\pi} e^{i\pi/6} \begin{pmatrix} \gamma^{-1} & -\gamma \\ -i\gamma^{-1} & -i\gamma \end{pmatrix} \Phi_\beta^{\sigma_3/4}. \tag{3.23}$$

(At first blush, this matrix valued function is analytic on a domain which consists of a disc centered at $z = \beta$, with the real axis deleted. However, a straightforward argument (as in [7, Section 7]) shows that E extends to an analytic function on the disc B_δ^β.)

We are now ready to define $M_1^{(A)}$ within B_δ^β:

$$M_1^{(A)}(z) = E_\beta(z)\ \Psi^\sigma(\Phi_\beta(z))\ e^{2\sigma_3 \Phi_\beta^{3/2}/3}, \quad \text{for } z \in B_\delta^\beta. \tag{3.24}$$

We now proceed to the definition of the approximation $M_1^{(A)}(z)$ within B_δ^α. As the construction is entirely analogous to the construction of $M_1^{(A)}$ in B_δ^β, we will only present the formulae. More details on this construction can be found in [7, Section 7] , or [8, Subsection 4.4]. We define the transformation $\Phi_\alpha(z)$, analogous to Φ_β defined in (3.17) above,

$$\Phi_\alpha(z) \equiv \left(\frac{3N}{4}\right)^{2/3} \left(2g(\alpha) - V(\alpha) - \ell - (2g(z) - V(z) - \ell)\right)^{2/3}, \quad z \in B_\delta^\alpha. \tag{3.25}$$

As $h(\alpha) \neq 0$, $2g(\alpha) - V(\alpha) - \ell - (2g(z) - V(z) - \ell)$ behaves like $(z-\alpha)^{3/2}$ for z near α (cf. relations (3.6)), and we may choose the branch so that $\Phi_\alpha(z)$ is analytic in a neighborhood of α, and

$$\Phi_\alpha(z) = -c(z-\alpha) + O\left((z-\alpha)^2\right),\ c > 0,\ z \to \alpha. \tag{3.26}$$

Next we define $E_\alpha(z)$ as follows:

$$E_\alpha(z) = \sqrt{\pi} e^{i\pi/6} \begin{pmatrix} \gamma & -\gamma^{-1} \\ i\gamma & i\gamma^{-1} \end{pmatrix} \Phi_\alpha^{\sigma_3/4}. \tag{3.27}$$

We now define the approximation $M_1^{(A)}$ within B_δ^α:

$$M_1^{(A)}(z) = E_\alpha(z)\ \Psi^\sigma\ (\Phi_\alpha(z))\ e^{2\sigma_3 \Phi_\alpha^{3/2}/3} \sigma_3, \quad \text{for } z \in B_\delta^\alpha. \tag{3.28}$$

3.4. **The Riemann–Hilbert problem for the error.** The next step in the Riemann–Hilbert approach is to define the ratio of M_1 to $M_1^{(A)}$:

$$S = M_1(z) \left(M_1^{(A)}(z)\right)^{-1}, \quad z \in \mathbb{C} \backslash \Sigma_{M_1}. \tag{3.29}$$

It then follows that $R(z)$ satisfies the following Riemann–Hilbert problem:

Problem 3.30. The problem is to determine a 2×2 matrix $S(z)$ satisfying

(a) $S(z)$ analytic in $\mathbb{C} \backslash \Sigma_S$,
(b) $S(z)$ possesses continuous boundary values $S_\pm(z)$ for $z \in \Sigma_S$,

(c) $S = I + \mathcal{O}(1/z),\ z \to \infty,$
(d) $S_+ = S_- V_S,\ z \in \Sigma_S,$

where Σ_S is defined in Figure 3.31.

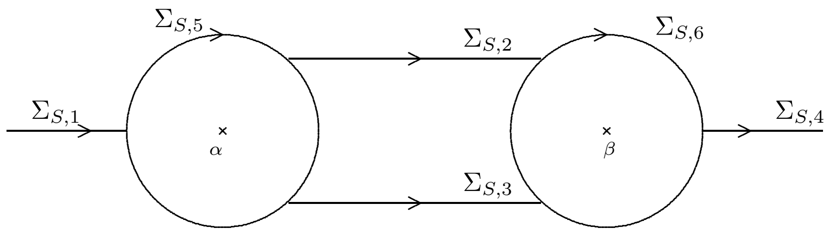

Figure 3.31. The contour $\hat{\Sigma}_S$.

The jump matrix V_S is defined as follows:

$$V_{S,1} = M_1^{(A)} \begin{pmatrix} 1 & e^{N(g_+ + g_- - V - \ell)} \\ 0 & 1 \end{pmatrix} \left(M_1^{(A)}\right)^{-1}, \quad z \in \Sigma_{S,1}, \tag{3.32}$$

$$V_{S,2} = M_1^{(A)} \begin{pmatrix} 1 & 0 \\ e^{-N(g_+ - g_-)} & 1 \end{pmatrix} \left(M_1^{(A)}\right)^{-1}, \quad z \in \Sigma_{S,2}, \tag{3.33}$$

$$V_{S,3} = M_1^{(A)} \begin{pmatrix} 1 & 0 \\ e^{N(g_+ - g_-)} & 1 \end{pmatrix} \left(M_1^{(A)}\right)^{-1}, \quad z \in \Sigma_{S,3}, \tag{3.34}$$

$$V_{S,4} = M_1^{(A)} \begin{pmatrix} 1 & e^{N(g_+ + g_- - V - \ell)} \\ 0 & 1 \end{pmatrix} \left(M_1^{(A)}\right)^{-1}, \quad z \in \Sigma_{S,4}, \tag{3.35}$$

$$V_{S,5} = \left(M_1^{(A)}\right)_- \left(M_1^{(A)}\right)_+^{-1}, \quad z \in \Sigma_{S,5}, \tag{3.36}$$

$$V_{S,6} = \left(M_1^{(A)}\right)_- \left(M_1^{(A)}\right)_+^{-1}, \quad z \in \Sigma_{S,6}. \tag{3.37}$$

The fact of the matter is that

1. On the upper lens, lower lens, and the real axis component of Σ_S

$$\|V_S - I\|_{L^\infty},\ \|V_S - I\|_{L^2} \le Ce^{-dN},\ C, d > 0. \tag{3.38}$$

This can be deduced from (3.32)-(3.36), because (1) the global approximation $M_1^{(A)}$, as well as its inverse, is uniformly bounded on these portions of the contour, (2) the off-diagonal entries of the middle matrices on the right hand side of (3.32)-(3.36) are uniformly exponentially decaying in N on these portions of the contour, and (3) for $z \to \infty$, $z \in \Sigma_{S,1} \cup \Sigma_{S,4}$, one has $e^{N(g_+ + g_- - V - \ell)} \le Ce^{-dN(|z|+1)}$ (for the complete argument, see the proof of [7, Proposition 7.64]).

2. On the circle $\Sigma_{S,6}$, we have the following complete asymptotic expansion:

$$V_{S,6} = I + \frac{1}{N} V_1^\beta + \frac{1}{N^2} V_2^\beta + \cdots \tag{3.39}$$

$$V_1^\beta = \frac{5(z-\alpha)^{1/2}}{72(z-\beta)^{1/2} \int_\beta^z R(s)h(s)ds} \begin{pmatrix} -1 & i \\ i & 1 \end{pmatrix} + \frac{7(z-\beta)^{1/2}}{72(z-\alpha)^{1/2} \int_\beta^z R(s)h(s)ds} \begin{pmatrix} 1 & i \\ i & -1 \end{pmatrix}, \tag{3.40}$$

$$V_2^\beta = \frac{35}{2592} \left(\int_\beta^z R(s)h(s)ds \right)^{-2} \begin{pmatrix} -1 & 12i \\ -12i & -1 \end{pmatrix}. \tag{3.41}$$

More generally, we have for k even, $k \ge 2$:

$$V_k^\beta = \left(\frac{1}{2} \int_\beta^z R(s)h(s)ds \right)^{-k} \begin{pmatrix} \frac{t_k + s_k}{2} & \frac{i(s_k - t_k)}{2} \\ \frac{-i(s_k - t_k)}{2} & \frac{t_k + s_k}{2} \end{pmatrix}, \tag{3.42}$$

and for k odd, $k \geq 1$:

$$V_k^{\beta} = \frac{2^{k-1} t_k (z-\alpha)^{1/2}}{(z-\beta)^{1/2}\left(\int_\beta^z R(s)h(s)ds\right)^k} \begin{pmatrix} -1 & i \\ i & 1 \end{pmatrix} + \frac{2^{k-1} s_k (z-\beta)^{1/2}}{(z-\alpha)^{1/2}\left(\int_\beta^z R(s)h(s)ds\right)^k} \begin{pmatrix} 1 & i \\ i & -1 \end{pmatrix}, \tag{3.43}$$

where

$$s_0 = t_0 = 1, \tag{3.44}$$

$$s_k = \frac{\Gamma(3k+\frac{1}{2})}{54^k k! \Gamma(k+\frac{1}{2})}, \; t_k = -\frac{6k+1}{6k-1} s_k \;\text{ for } k \geq 1, \tag{3.45}$$

3. On the circle ∂B_δ^α, we have the following complete asymptotic expansion:

$$V_{S,5} = I + \frac{1}{N} V_1^\alpha + \frac{1}{N^2} V_2^\alpha + \cdots \tag{3.46}$$

$$V_1^\alpha = \frac{5(z-\beta)^{1/2}}{72(z-\alpha)^{1/2}\int_\alpha^z R(s)h(s)ds} \begin{pmatrix} -1 & -i \\ -i & 1 \end{pmatrix} + + \frac{7(z-\alpha)^{1/2}}{72(z-\beta)^{1/2}\int_\alpha^z R(s)h(s)ds} \begin{pmatrix} 1 & -i \\ -i & -1 \end{pmatrix}, \tag{3.47}$$

$$V_2^\alpha = \frac{35}{2592}\left(\int_\alpha^z R(s)h(s)ds\right)^{-2} \begin{pmatrix} -1 & -12i \\ 12i & -1 \end{pmatrix}. \tag{3.48}$$

More generally, we have for k even, $k \geq 2$,

$$V_k^\alpha = \frac{2^{k-1}}{\left(\int_\alpha^z R(s)h(s)ds\right)^k} \begin{pmatrix} s_k + t_k & i(t_k - s_k) \\ i(s_k - t_k) & s_k + t_k \end{pmatrix}, \tag{3.49}$$

and for k odd, $k \geq 1$, we have

$$V_k^\alpha = \frac{2^{k-1} s_k (z-\beta)^{1/2}}{(z-\alpha)^{1/2}\left(\int_\alpha^z R(s)h(s)ds\right)^k} \begin{pmatrix} -1 & -i \\ -i & 1 \end{pmatrix} + \frac{2^{k-1} t_k (z-\alpha)^{1/2}}{(z-\beta)^{1/2}\left(\int_\alpha^z R(s)h(s)ds\right)^k} \begin{pmatrix} -1 & i \\ i & 1 \end{pmatrix}, \tag{3.50}$$

The reader may verify these asymptotic expansions using the definitions (3.36)-(3.37) of V_S, the definitions (3.24) and (3.28) of $M_1^{(A)}$, together with the well-known asymptotic expansions for the Airy function and its derivative. In [7], the authors present a complete proof of this fact (see, Lemma 7.34 and its proof), and we leave the reader to carry out the analogous arguments in the present case.

3.5. **The solution S of the Riemann–Hilbert problem 3.30 and its asymptotic expansion.** In this subsection we will present a formula for the solution S of the Riemann–Hilbert problem 3.30, as well as an asymptotic expansion for S in powers of N^{-1}. To do so, we will require the integral operators $C_\pm$, defined as follows:

$$C_\pm : L^2(\Sigma_S) \to L^2(\Sigma_S), \tag{3.51}$$

$$C_\pm(f)(z) = \lim_{z' \to z,\; z \in \pm \text{ side of } \Sigma_S} \frac{1}{2\pi i} \int_{\Sigma_S} \frac{f(s)}{s - z'} ds, \quad z \in \Sigma_S. \tag{3.52}$$

The relevant integral operator $C_{V_S} : L^2(\Sigma_S) \to L^2(\Sigma_S)$ is defined via

$$C_{V_S}(f) = C_-\left[f\left(V_S - I\right)\right]. \tag{3.53}$$

The asymptotic expansions described in (3.39)-(3.48), together with the estimates (3.38), and the well known boundedness of the integral operators $C_\pm$ imply that the integral operator C_{V_S} is bounded, with operator norm satisfying

$$\|C_{V_S}\|_{L^2(\Sigma_S)\to L^2(\Sigma_S)} \le \frac{C}{N}. \tag{3.54}$$

This implies that $\mathbb{I} - C_{V_S}$ can be inverted by Neumann series for N sufficiently large. Therefore, we may define

$$\mu_S = \left(\mathbb{I} - C_{V_S}\right)^{-1} C_{V_S}(I). \tag{3.55}$$

Then, as in [7, Theorem 7.70], the explicit formula for the solution S to the Riemann–Hilbert problem (3.30) is

$$S(z) = I + \frac{1}{2\pi i}\int_{\Sigma_S} \frac{(I+\mu_S)(V_S - I)}{s-z} ds, \quad z \in \mathbb{C}\setminus\Sigma_S. \tag{3.56}$$

Furthermore, as shown in Corollary 7.77 and Theorem 7.81 in [7], S is uniformly bounded in $\mathbb{C}\setminus\Sigma_S$, and possesses a complete asymptotic expansion in powers of N^{-1} (in Section 7 of [7], the error matrix is called $R(z)$, whereas here the error matrix is called S).

The terms in the asymptotic expansion may be obtained in a number of ways. For example, one may just compute the Neumann expansion of μ_S, and keep only those terms which arise due to integrations around the circles ∂B_δ^α and ∂B_δ^β. Alternatively, one may posit an asymptotic expansion for S in powers of N^{-1}, plug this expansion into the Riemann–Hilbert problem 3.30, and obtain a sequence of Riemann–Hilbert problems for the coefficients in the expansion, which may be solved iteratively. The proof of Theorem 7.81 in [7] involves (1) defining an integral operator $C_{V_S}^{(l)}$ associated only to the circles ∂B_δ^α and ∂B_δ^β, using only the first l terms in the expansion of V_S on these circles, (2) building an explicit approximation $S_l(z)$ by using the first l terms in a Neumann series for $\left(\mathbb{I} - C_{V_S}^{(l)}\right)^{-1}$, and (3) comparing this approximation to the true solution $S(z)$. The ensuing Riemann–Hilbert problem for the error is shown to have a solution which is uniformly bounded by the first neglected term in the asymptotic expansion, etc.

The result of these calculations is the following asymptotic expansion for the solution S:

$$S(z) \sim I + \sum_{j=1}^{\infty} N^{-j} S_j(z), \tag{3.57}$$

where $S_j(z)$ is piecewise analytic, and can be defined iteratively via the following sequence of additive Riemann–Hilbert problems.

Problem 3.58. The first problem is to determine a 2×2 matrix $S_1(z)$ satisfying

(a) $S_1(z)$ analytic in $\mathbb{C}\backslash(\partial B_\delta^\alpha \cup \partial B_\delta^\beta)$,
(b) $S_1(z)$ possesses continuous boundary values $(S_1)_\pm(z)$ for $z \in \partial B_\delta^\alpha \cup \partial B_\delta^\beta$,
(c) $S_1 = \mathcal{O}(1/z),\ z\to\infty$,
(d) $(S_1)_+ - (S_1)_- = V_1,\ z \in \partial B_\delta^\alpha \cup \partial B_\delta^\beta$,

Problem 3.59. Having determined S_1, we now determine S_k, $k \ge 2$ satisfying

(a) $S_k(z)$ analytic in $\mathbb{C}\backslash(\partial B_\delta^\alpha \cup \partial B_\delta^\beta)$,
(b) $S_k(z)$ possesses continuous boundary values $(S_k)_\pm(z)$ for $z \in \partial B_\delta^\alpha \cup \partial B_\delta^\beta$,
(c) $S_k = \mathcal{O}(1/z),\ z\to\infty$,
(d) $(S_k)_+ - (S_k)_- = V_k + \sum_{j=1}^{k-1}(S_j)_-\ V_{k-j}\quad,\ z \in \partial B_\delta^\alpha \cup \partial B_\delta^\beta$.

Each of these Riemann–Hilbert problems possesses a unique solution, explicitly representable by contour integration. It is interesting and useful to note that in fact each of these contour integral representations can be evaluated explicitly. Thus, for example, we include an explicit representation for $S_1(z)$ in Appendix A.

Some very important properties of the sequence $\{S_k\}_{k\geq 0}$ are established in the following Lemma. We will need the Pauli matrices

$$\sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{3.60}$$

Lemma 3.2. *For each $k \in \mathbb{Z}_{\geq 0}$, there are complex valued functions $s_k(z)$ so that the following holds.*

1. *For k even, the function $S_k(z)$, defined to be the unique solution to the Riemann-Hilbert problem 3.59 satisfies the symmetry conditions*

$$S_k(z) = s_k^{(1)}(z) I + s_k^{(2)}(z)\sigma_2. \tag{3.61}$$

2. *For k odd, the function S_k satisfies the symmetry conditions*

$$S_k(z) = s_k^{(1)}(z)\sigma_3 + s_k^{(2)}(z)\sigma_1. \tag{3.62}$$

The functions $s_k^{(j)}(z)$, $j = 1, 2$, are piecewise analytic functions of z, explicitly computable in terms of the parameters $\mathbf{t}$.

In words: for k odd, $S_k(z)$, is a symmetric, trace 0 matrix. For k even, S_k is the sum of a multiple of the identity matrix, and a skew-symmetric matrix.

Proof. For $k = 1$, the Lemma is seen to be true by studying the solution to the Riemann-Hilbert problem 3.58:

$$S_1(z) = \frac{1}{2\pi i} \int \frac{V_1(s)}{s - z} ds, \tag{3.63}$$

where the integral is taken along the boundaries of the discs B_δ^α and B_δ^β, oriented clockwise. Since V_1, as defined in (3.40) and (3.47) is clearly symmetric trace zero, S_1 inherits the same property.

We now argue inductively. Suppose that k is even, and the Lemma holds true up to $k-1$. The reader may verify that the right hand side of the jump condition (d) in the Riemann–Hilbert problem 3.59 is of the form

$$\nu_k^{(1)}(z) I + \nu_k^{(2)}(z)\sigma_2, \tag{3.64}$$

where $\nu_k^{(j)}(z)$, $j = 1, 2$, are scalar complex valued functions of z. Indeed, for k even, $V_k(z)$, as defined in (3.42) and (3.49) are of this form. Moreover, each term appearing in the sum on the right hand side of the jump condition (d) in the Riemann–Hilbert problem 3.59 is also of the same form (this can be deduced from the form of S_j and V_{k-j} for each value of $j \in \{1, \ldots, k-1\}$, given that k is even). There is an explicit formula for the solution to this additive Riemann–Hilbert problem, which is given by the Cauchy transform of the right hand side of jump condition (d). Since every term is of the form (3.64), the Cauchy transform is also of this form, and so (3.61) holds true for k even.

The argument for k odd follows the same line of reasoning, and we leave the details to the reader. □

At this point we have established the following formula for the solution Y to the original Riemann–Hilbert problem stated just before Subsection 2.1:

$$Y(z) = e^{N\ell\sigma_3/2} S(z) M_1^{(A)}(z) T(z) e^{N(g(z)-\ell/2)\sigma_3}, \tag{3.65}$$

where ℓ is the Lagrange multiplier associated to the variational problem, $S(z)$ is the solution to the Riemann–Hilbert problem 3.30 for the error, $M_1^{(A)}(z)$ is the explicit approximation defined in (3.15), (3.24), and (3.28), $T(z)$ is either the identity matrix I for z outside the lens shaped regions, or one of two explicit triangular factors appearing in (3.9) and (3.10), and $g(z)$ is the function defined in (3.5). It is straightforward to verify that with the exception of $S(z)$, all terms appearing in (3.65) depend analytically on the times $\mathbf{t}$ for $\mathbf{t}$ in any open subset of $\mathbb{T}(T, \gamma)$, and in fact extend as infinitely differentiable functions to all of $\mathbb{T}(T, \gamma)$. It is also true that the function $S(z)$ enjoys the same analyticity properties. This is made precise in the following Theorem.

Theorem 3.3. *The matrix valued function S is infinitely differentiable in the times $\mathbf{t}$ for $\mathbf{t} \in \mathbb{T}(T,\gamma)$, and derivatives of S again possess an asymptotic expansion in N, obtained by differentiating the original asymptotic expansion for S. The coefficients in this asymptotic expansion depend analytically on $\mathbf{t}$ for $\mathbf{t} \in \mathbb{T}(T,\gamma)$ (with smooth extension to the closure), for T sufficiently small.*

Proof. The proof hinges on the following fundamental facts: the jump matrix V_S is analytic in a vicinity of every point z on the contour Σ_S, it possesses an asymptotic expansion in powers of N^{-1} in which each term is a piecewise analytic function of z and depends analytically on $\mathbf{t}$ for $\mathbf{t} \in \mathbb{T}(T,\gamma)$; this asymptotic expansion may be differentiated term by term with respect to the times $\mathbf{t} \in \mathbb{T}(T,\gamma)$, yielding asymptotic expansions for arbitrary derivatives of V_S. From this information, it is straightforward to prove that μ_S defined in (3.55) is an analytic function of $\mathbf{t}$ for $\mathbf{t} \in \mathbb{T}(T,\gamma)$, has an analytic continuation to a neighborhood of the contour Σ_S in the z-variable, and possesses an asymptotic expansion (which can be differentiated term by term) obtained directly from a Neumann series representation. An outline of the argument is as follows.

First, observe that the terms in the Neumann series representation of μ_S are individually analytic functions of z in a vicinity of every point of the contour Σ_S, and they are analytic function of $\mathbf{t}$ for $\mathbf{t} \in \mathbb{T}(T,\gamma)$. Second, the series is uniformly convergent. This follows because (1) each Neumann iterate can be expressed as an integral on a *deformed* contour which is bounded away from the contour Σ_S and (2) the j-th iterate may then be bounded by $C(D/N)^j$, where the constants C and D are independent of j. For derivatives of μ_S, similar calculations, using the fact that derivatives of V_S possess asymptotic expansions, show that an Lth order derivative of the jth iterate may be bounded by $\tilde{C} j^L (\tilde{D}/N)^j$, where $\tilde{C}$ and $\tilde{D}$ are independent of j, but may depend on L. This shows that the Neumann series for μ_S may be differentiated term by term.

The proof that derivatives of S possess expansions which may be obtained by differentiating the original expansion for S term by term now follows from the uniform convergence of the Neumann series for μ_S and its derivatives. □

4. The mean density of eigenvalues: exact formula and asymptotics

The previous section explains (and extends to higher orders) the work in [6]-[8], in which techniques for the asymptotic analysis of Riemann-Hilbert problems, originally developed for the analysis of singular limits of integrable partial differential equations by Deift and Zhou (and further developed by many researchers), are used to establish an asymptotic description for the polynomials $p_N(z)$ and $p_{N-1}(z)$ orthogonal with respect to the N-dependent measure (1.23). There are many consequences of the complete asymptotic description for the orthogonal polynomials. For a number of such applications, we refer the reader to [6]-[8], where detailed asymptotic expansions for (1) the leading coefficients of of the polynomials, (2) the recurrence coefficients, (3) the zeros of the polynomials were obtained, along with a proof of universality of spacings of eigenvalues of random matrices for a wide class of unitarily invariant probability measures on Hermitian matrices. In this section our goal is to establish formula (4.4), and prove that this formula for the mean density of eigenvalues holds true for z in a fixed sized neighborhood of $[z^*,\beta]$ for N sufficiently large. The parameter z^* will be chosen below. The formula (4.4) holds true in a large neighborhood containing the point β, and there is an analogous formula, (4.15) which holds true in a fixed size neighborhood of the interval $[\alpha, z^*]$, and the two formulae hold simultaneously on an overlap region. In order to achieve this overlap region, we will need to introduce T_β, γ_β, T_α, and γ_α, which are values of T and γ from Theorem 3.1 so that if $T < \min\{T_\beta, T_\alpha\}$ and $\gamma > \max\{\gamma_\beta,\gamma_\alpha\}$, then the neighborhoods $D_{\epsilon,\beta}$ and $D_{\epsilon,\alpha}$ may be taken large enough to have nontrivial overlap.

Using formula (4.4) we will obtain a complete asymptotic expansion for the mean density of eigenvalues, and then in the following section we will evaluate an asymptotic expansion of integrals of the form

$$\int_{\mathbb{R}} f(\lambda)\rho_N^{(1)}(\lambda)d\lambda, \tag{4.1}$$

for general C^∞ functions that grow no worse than polynomial at ∞.

We begin with two important observations.

Observation 1: It is possible to use the results of the previous section to compute an asymptotic expansion for the mean density $\rho_N^{(1)}$ which is valid for all $\lambda \in (\alpha, \beta)$. That expansion is as follows,

$$\rho_N^{(1)}(\lambda) = \psi(\lambda) + \frac{1}{4\pi N}\left(\frac{1}{\lambda-\beta} - \frac{1}{\lambda-\alpha}\right)\cos\left\{N\int_\lambda^\beta \psi(s)ds\right\} + \frac{1}{N^2}\left[H(\lambda) + G(\lambda)\sin\left\{N\int_\lambda^\beta \psi(s)ds\right\}\right] + \cdots \tag{4.2}$$

in which $H(\lambda)$ and $G(\lambda)$ are locally analytic functions which are explicitly computable in terms of the original external field $V(\lambda)$. However, for the computation of integrals of the form (4.1), this expansion provides a piece of the puzzle, but clearly cannot be used in a vicinity of the endpoints α and β. Indeed, the asymptotic expansion fails in a vicinity of these endpoints, because the second term in (4.2) possesses poles, whereas the mean density $\rho_N^{(1)}$ does not. From this observation we see that to compute asymptotic expansions of integrals of the form (4.1), we will need to to have an asymptotic expansion for $\rho_N^{(1)}$ which is valid for λ near these endpoints. It is a piece of good fortune that the representations which we obtain are actually valid on regions which cover the entire interval $[\alpha, \beta]$, and we will not need to evaluate integrals "in the bulk".

Observation 2: It is straightforward to use the results of the preceding section to prove that $\rho_N^{(1)}(\lambda)$ is exponentially small (in N), and exponentially decaying (in λ) for $\lambda \in \mathbb{R} \setminus (\alpha - \delta, \beta + \delta)$, for any fixed $\delta > 0$. We will not present the details of those calculations, but leave them to the avid reader. The implication of this is that the contribution to an integral of the form (4.1) from the set $\mathbb{R}\setminus(\alpha-\delta, \beta+\delta)$ is exponentially small in N, and hence negligible to all orders in N. Therefore we must only compute an asymptotic expansion for integrals of the form

$$\int_{\alpha-\delta}^{\beta+\delta} f(\lambda)\rho_N^{(1)}(\lambda)d\lambda, \tag{4.3}$$

in which the allowable functions f are C^∞ smooth, and compactly supported within $(\alpha - \delta, \beta + \delta)$.

We now turn to the formula (3.1) for $\rho_N^{(1)}$, and re-express this fundamental quantity using the Riemann–Hilbert analysis of the previous section. We present two functions, $\rho_N^{(1,\beta)}$ and $\rho_N^{(1,\alpha)}$ representing $\rho_N^{(1)}$; the first one is a valid representation for $\rho_N^{(1)}$ in an interval containing $z = \beta$, and the other, in an interval containing $z = \alpha$. The first such formula is the following explicit, exact expression:

$$\begin{aligned} N\rho_N^{(1,\beta)}(z) = &\left(\frac{\Phi_\beta'(z)}{4\Phi_\beta(z)} - \frac{\gamma'(z)}{\gamma(z)}\right)\left[2\ Ai\left(\Phi_\beta(z)\right)\ Ai'\left(\Phi_\beta(z)\right)\right] \\ &+\Phi_\beta'(z)\left[\Big(Ai'\left(\Phi_\beta(z)\right)\Big)^2 - \Phi_\beta(z)\ \Big(Ai\left(\Phi_\beta(z)\right)\Big)^2\right] \\ &+\frac{i}{2}\Big[\Big(S'\ B\ \Psi\left(\Phi_\beta(z)\right)\Big)_{11}\ \Big(S\ B\ \Psi\left(\Phi_\beta(z)\right)\Big)_{21} \\ &\qquad - \Big(S\ B\ \Psi\left(\Phi_\beta(z)\right)\Big)_{11}\ \Big(S'\ B\ \Psi\left(\Phi_\beta(z)\right)\Big)_{21}\Big], \end{aligned} \tag{4.4}$$

where

$$\Psi(\zeta) = \begin{cases} \begin{pmatrix} Ai(\zeta) & Ai(\omega^2\zeta) \\ Ai'(\zeta) & \omega^2 Ai'(\omega^2\zeta) \end{pmatrix}, & \text{for } \zeta \in \mathbb{C}_+, \\ \begin{pmatrix} Ai(\zeta) & -\omega^2 Ai(\omega\zeta) \\ Ai'(\zeta) & -Ai'(\omega\zeta) \end{pmatrix}, & \text{for } \zeta \in \mathbb{C}_-. \end{cases} \tag{4.5}$$

The transformation $\Phi_\beta(z)$ was defined in (3.17), $\gamma(z)$ was defined in (3.16), and the matrix B is defined by

$$B(z) = \begin{pmatrix} \gamma^{-1}(z) & -\gamma(z) \\ -i\gamma^{-1}(z) & -i\gamma(z) \end{pmatrix}\Phi_\beta(z)^{\sigma_3/4}, \tag{4.6}$$

and $S(z)$, defined in (3.29), is the solution to the Riemann–Hilbert problem 3.30 for the error, which possesses a complete asymptotic expansion (3.57), with coefficients S_k solving the Riemann–Hilbert problems 3.58 and 3.59, and satisfying the symmetry conditions described in Lemma 3.2.

Shortly we will present an analogous representation of $\rho_N^{(1)}$ near the left endpoint α. The following lemma will be instrumental in describing the relation between these two representations as well as their respective domains of validity.

Lemma 4.1. *There is a unique point z^* in the interval (α, β) at which $(\Phi_\beta(z^*)) = (\Phi_\alpha(z^*))$*

Proof. It follows from formulas (3.17), (3.25) and (3.6) that for $z < \beta$,

$$(-\Phi_\beta(z))^{3/2} = \left(\frac{3N}{4}\right)\int_z^\beta \frac{R_+}{i} h ds; \tag{4.7}$$

and for $z > \alpha$,

$$(-\Phi_\alpha(z))^{3/2} = \left(\frac{3N}{4}\right)\int_\alpha^z \frac{R_+}{i} h ds. \tag{4.8}$$

It then follows from comparison to (3.4) that the integrals in (4.8) and (4.7) are both positive.

By construction (see (3.18) and (3.26)) we know that $\Phi_\beta(z)$ and $\Phi_\alpha(z)$ are negative real for $\alpha < z < \beta$. The z^* we seek must satisfy

$$\int_\alpha^{z^*} Rhds \pm \int_{z^*}^\beta Rhds = 0.$$

We already know that

$$\int_\alpha^z Rhds + \int_z^\beta Rhds = \int_\alpha^\beta Rhds = 2\pi i.$$

It follows that z^* is uniquely determined, equivalently, by either of the two equations:

$$\int_\alpha^{z*} Rhds = \pi i = \int_{z*}^\beta Rhds$$

□

Lemma 4.2. *There is a $T_\beta > 0$ sufficiently small and γ_β sufficiently large (both independent of N) so that for $\mathbf{t} \in \mathbb{T}(T_\beta, \gamma_\beta)$, there exists a complex neighborhood $\mathbf{B}_\beta$ of the interval $[z^*, 2]$, independent of N, so that formula (4.4) holds true for all $z \in \mathbf{B}_\beta$.*

Proof. The proof is a straightforward, though lengthy algebraic manipulation of the explicit relationships between the original quantity $Y(z)$, and the error matrix $S(z)$. The only analytical issue is this: is the transformation $\Phi_\beta(z)$ an invertible analytic transformation from some fixed size neighborhood of the interval $[z^*, 2]$ into its range, for $\mathbf{t}$ sufficiently small? But this is clear, because for $\mathbf{t} = 0$, the function Φ_β is quite simple,

$$\Phi_\beta(z)|_{\mathbf{t=0}} = \left(\frac{3N}{4}\int_2^z \sqrt{s-2}\sqrt{s+2}\, ds\right)^{2/3}.$$

The reader may easily verify that there exists a branch of this function which is analytic on a fixed open neighborhood of the interval $(-2, 2]$. Now since $\Phi_\beta(z)$ depends analytically on $\mathbf{t}$, it is clear that $\Phi_\beta(z)$ is certainly an invertible analytic transformation from some fixed size neighborhood of $[0, 2]$ into its range.

The reader may verify that formula (4.4) is true, by starting with the formula (3.1) for $\rho_1^{(N)}$, and re-expressing this formula in terms of the solution Y of the original Riemann–Hilbert problem

$$\rho_1^{(N)}(z) = \frac{e^{-NV(z)}}{-2\pi i N}\left[Y_{11}'(z)Y_{21}(z) - Y_{11}(z)Y_{21}'(z)\right]. \tag{4.9}$$

Next, observe that for $z \in \mathbb{C}_\pm \cap \{\text{lens shaped region in Figure 3.7}\}$, we have the following representation for the solution Y of the original Riemann–Hilbert problem:

$$\begin{aligned} Y(z) \;&= \\ & e^{\frac{N\ell}{2}\sigma_3} S(z)\ E_\beta(z)\ \Psi^\sigma(\Phi_\beta(z)) e^{\frac{2}{3}\sigma_3 \Phi_\beta(z)^{3/2}} \begin{pmatrix} 1 & 0 \\ \pm e^{\mp N(g_+ - g_-)} & 1 \end{pmatrix} e^{N(g - \frac{\ell}{2})\sigma_3} \\ &= e^{\frac{n\ell}{2}\sigma_3} S(z)\ E_\beta(z)\ \Psi(\Phi_\beta(z))\ e^{\left(\frac{N}{2}V(z) - \pi i/6\right)\sigma_3}. \end{aligned} \tag{4.10}$$

Here $E_\beta(z)$ was defined in (3.23), and is related to the matrix valued function $B(z)$ defined in (4.6) via

$$E_\beta(z) = \sqrt{\pi} e^{i\pi/6} B(z). \tag{4.11}$$

The simplification in (4.10) follows from the identity

$$\begin{pmatrix} 1 & 0 \\ \mp 1 & 1 \end{pmatrix} e^{\frac{2}{3}\sigma_3 \Phi_\beta(z)^{3/2}} \begin{pmatrix} 1 & 0 \\ \pm e^{\mp N(g_+ - g_-)} & 1 \end{pmatrix} e^{N(g - \frac{\ell}{2})\sigma_3} = e^{\frac{N}{2} V(z)\sigma_3}, \tag{4.12}$$

which can be deduced from the properties of $g(z)$ and $\Phi_\beta(z)$. Next, from the definition (4.6), one easily obtains the identity

$$B'(z) = \left(\frac{\Phi_\beta'(z)}{4\Phi_\beta(z)} - \frac{\gamma'(z)}{\gamma(z)} \right) B(z)\sigma_3. \tag{4.13}$$

The following identity for Ψ is also equally easy to prove, using the definition (4.5) and the differential equation satisfied by the Airy function:

$$\Psi'(\zeta) = \begin{pmatrix} 0 & 1 \\ \zeta & 0 \end{pmatrix} \Psi(\zeta). \tag{4.14}$$

Formula (4.4) may now be verified by differentiating (4.10), substituting into (4.9) and using the differential relations (4.13) and (4.14).

□

Formula (4.4) may seem unwieldy due to its length. However, this is not quite the case. Indeed, the first two lines of this formula do not contain $S(z)$ at all, and are explicit combinations of Airy functions (composed with $\Phi_\beta(z)$) together with other functions defined explicitly in Section 3. Moreover, we will show that the third and fourth line of (4.4) possess asymptotic expansions in powers of N^{-1} in which each term is of the same explicit form as the first two lines. This is the content of Lemma 4.4 below.

There is, of course, an analogous formula which is valid on a fixed size complex neighborhood of the interval $[-2, z^*]$. We summarize the analogous result in the following Lemma.

Lemma 4.3. *There is $T_\alpha > 0$ sufficiently small and γ_α sufficiently large (both independent of N) so that for $\mathbf{t} \in \mathbb{T}(T_\alpha, \gamma_\alpha)$, there exists a complex neighborhood of $\mathbf{B}_\alpha$ of the interval $[-2, z^*]$, independent of N, so that the formula (4.15) holds true for all $z \in \mathbf{B}_\alpha$.*

$$\begin{aligned} N\rho_N^{(1,\alpha)}(z) = &-\left(\frac{\Phi_\alpha'(z)}{4\Phi_\alpha(z)} + \frac{\gamma'(z)}{\gamma(z)} \right) \left[2\ Ai\left(\Phi_\alpha(z)\right)\ Ai'\left(\Phi_\alpha(z)\right) \right] \\ &- \Phi_\alpha'(z) \left[\left(Ai'\left(\Phi_\alpha(z)\right)\right)^2 - \Phi_\alpha(z) \left(Ai\left(\Phi_\alpha(z)\right)\right)^2 \right] \\ &+ \frac{i}{2} \left[\left(S'\ B\ \Psi\left(\Phi_\alpha(z)\right)\right)_{11} \left(S\ B\ \Psi\left(\Phi_\alpha(z)\right)\right)_{21} \right. \\ &\qquad \left. - \left(S\ B\ \Psi\left(\Phi_\alpha(z)\right)\right)_{11} \left(S'\ B\ \Psi\left(\Phi_\alpha(z)\right)\right)_{21} \right], \end{aligned} \tag{4.15}$$

where in this case

$$B(z) = \begin{pmatrix} \gamma(z) & -\gamma^{-1}(z) \\ i\gamma(z) & i\gamma^{-1}(z) \end{pmatrix} \Phi_\alpha(z)^{\sigma_3/4}. \tag{4.16}$$

Next, we will compute an asymptotic expansion for $\rho_N^{(1)}$, by using formulae (4.4), (4.15), and the asymptotic expansion for $S(z)$ described in Section 3.

Lemma 4.4. *If* $\mathbf{t} \in \mathbb{T}(T_\beta, \gamma_\beta)$*, then there is a fixed size neighborhood of the interval* $[z^*, \beta]$*, the following asymptotic expansion holds true:*

$$\begin{aligned}
(4.17) \qquad & \frac{i}{2N}\Big[\Big(S'\ B\ \Psi\left(\Phi_\beta(z)\right)\Big)_{11}\Big(S\ B\ \Psi\left(\Phi_\beta(z)\right)\Big)_{21} \\
&\qquad - \Big(S\ B\ \Psi\left(\Phi_\beta(z)\right)\Big)_{11}\Big(S'\ B\ \Psi\left(\Phi_\beta(z)\right)\Big)_{21}\Big] \\
&= \sum_{j \text{ even},\ j\geq 2} N^{-j}\tilde{a}_j(z)\Psi_{11}^2(\Phi_\beta)\frac{\sqrt{\Phi_\beta}}{\gamma(z)^2} \\
&\quad + \sum_{j \text{ even},\ j\geq 2} N^{-j}\tilde{b}_j(z)\frac{\Psi_{21}^2(\Phi_\beta(z))\gamma(z)^2}{\sqrt{\Phi_\beta(z)}} \\
&\quad + \sum_{j \text{ odd},\ j\geq 3}^{\infty} N^{-j}\tilde{c}_j(z)\Psi_{11}(\Phi_\beta(z))\Psi_{21}(\Phi_\beta(z)),
\end{aligned}$$

where for each j*, the coefficient functions* $\tilde{a}_j(z)$*,* $\tilde{b}_j(z)$*, and* $\tilde{c}_j(z)$ *are jointly analytic in* z *and* $\mathbf{t}$*, for* z *in a fixed sized complex neighborhood of the interval* $[z^*, \beta]$*, and for* $\mathbf{t} \in \mathbb{T}(T_\beta, \gamma_\beta)$*.*

Proof. From Lemma 3.2, we have an asymptotic expansion for the function $S(z)$:

$$\begin{aligned}
(4.18) \qquad S(z) = I &+ \sum_{k \text{ odd},\ k\geq 1}\left(s_k^{(1)}(z)\sigma_3 + s_k^{(2)}(z)\sigma_1\right)N^{-k} \\
&+ \sum_{k \text{ even},\ k\geq 2}\left(s_k^{(1)}(z)I + s_k^{(2)}(z)\sigma_2\right)N^{-k},
\end{aligned}$$

in which $s_k^{(j)}(z)$ are functions of z that are piecewise analytic functions of z, which are, in particular, analytic on a fixed size neighborhood of the interval $[0, \beta]$, and depend analytically on $\mathbf{t}$ for $\mathbf{t} \in \mathbb{T}(T_\beta, \gamma_\beta)$. An asymptotic expansion for $S'(z)$ is obtained by differentiating (4.18) term by term. This, together with the explicit formula (4.6) for $B(z)$, may be substituted into the left hand side of (4.17). Now straightforward algebraic manipulation yields an asymptotic expansion of the form displayed in the right hand side of (4.17). The quantities $\tilde{a}_j$ and $\tilde{b}_j$ may be computed via these calculations, but all that we will require is that these functions are analytic in a neighborhood of $[0, \beta]$, and depend analytically on $\mathbf{t}$ for $\mathbf{t} \in \mathbb{T}(T_\beta, \gamma_\beta)$, which is manifestly true. ☐

There is, of course, an analogous result which holds for the interval $[\alpha - \delta, z^*]$, which we will only state, as the proof follows by the same arguments as the proof of Lemma 4.4

Lemma 4.5. *If* $\mathbf{t} \in \mathbb{T}(T_\alpha, \gamma_\alpha)$*, then there is a fixed size neighborhood of the interval* $[\alpha, z^*]$*, on which the following asymptotic expansion holds true:*

$$\begin{aligned}
(4.19) \qquad & \frac{i}{2N}\Big[\Big(S'\ B\ \Psi\left(\Phi_\alpha(z)\right)\Big)_{11}\Big(S\ B\ \Psi\left(\Phi_\alpha(z)\right)\Big)_{21} \\
&\qquad - \Big(S\ B\ \Psi\left(\Phi_\alpha(z)\right)\Big)_{11}\Big(S'\ B\ \Psi\left(\Phi_\alpha(z)\right)\Big)_{21}\Big] \\
&= \sum_{j \text{ even},\ j\geq 2} N^{-j}\tilde{a}_j(z)\Psi_{11}^2(\Phi_\alpha)\gamma(z)^2\sqrt{\Phi_\alpha} \\
&\quad + \sum_{j \text{ even},\ j\geq 2} N^{-j}\tilde{b}_j(z)\frac{\Psi_{21}^2(\Phi_\alpha(z))}{\gamma(z)^2\sqrt{\Phi_\alpha(z)}} \\
&\quad + \sum_{j \text{ odd},\ j\geq 3}^{\infty} N^{-j}\tilde{c}_j(z)\Psi_{11}(\Phi_\alpha(z))\Psi_{21}(\Phi_\alpha(z)),
\end{aligned}$$

where for each j*, the coefficient functions* $\tilde{a}_j(z)$*,* $\tilde{b}_j(z)$*, and* $\tilde{c}_j(z)$ *are jointly analytic in* z *and* $\mathbf{t}$*, for* z *in a fixed size complex neighborhood of the interval* $[\alpha, z^*]$*, and for* $\mathbf{t} \in \mathbb{T}(T_\alpha, \gamma_\alpha)$*.*

Remark: Observe that if we take $T < \min\{T_\beta, T_\alpha\}$ and $\gamma > \max\{\gamma_\beta, \gamma_\alpha\}$ then the conclusions of Lemmas 4.2, 4.3, 4.4, and 4.5 hold true simultaneously. From now on we will take $\mathbf{t} \in \mathbb{T}(T,\gamma)$ for such values of T and γ.

5. Proof of the Main Theorem

We are now in a position to analyze the structure of the fundamental moments (1.33) which will enable us to establish the $1/N^2$ expansion of $\log Z_N$. To do this we will be using the representations of the one-point function developed in the previous section. Precisely, we use

$$\rho_N^{(1)}(z) = \chi_\alpha(z)\rho_N^{(1,\alpha)} + \chi_\beta(z)\rho_N^{(1,\beta)} \tag{5.1}$$

where $\{\chi_\alpha, \chi_\beta\}$ is a partition of unity for $\mathbb{R}$ defined by

$$\begin{aligned} \chi_\beta(z) \quad & \text{is} \quad C^\infty \text{ with} \\ 0 \leq \chi_\beta(z) \quad & \leq \quad 1 \\ \overline{supp \chi_\beta} \quad & \subset \quad (z^* - \epsilon, \infty) \\ \chi_\beta(z) \quad & \equiv 1 \qquad \text{for } z \in (z^* + \epsilon, \infty). \end{aligned} \tag{5.2}$$

We then define χ_α by

$$\chi_\alpha(z) = 1 - \chi_\beta(z)$$

which has similar properties to χ_β but with support in $(-\infty, z^* + \epsilon)$

The expression (5.1) is merely a different representation of the one-point function $\rho_N^{(1)}$, and the partition of unity allows us to carry out our asymptotic analysis separately on $\chi_\alpha(z)\rho_N^{(1,\alpha)}(z)$ in $\mathbf{B}_\alpha$ and on $\chi_\beta(z)\rho_N^{(1,\beta)}(z)$ in $\mathbf{B}_\beta$, and then, at the end, to add the two contributions together. This partition of unity will prove especially useful in avoiding the evaluation of boundary terms during repeated integrations by parts. Because of the complete similarity of the analysis for each of these separate pieces it will suffice to demonstrate our claims for $\chi_\beta(z)\rho_N^{(1,\beta)}(z)$.

We now begin the analysis of integrals of the form

$$\int_{\alpha-\delta}^{\beta+\delta} f(\lambda)\rho_N^{(1)}(\lambda)d\lambda = \int_{\alpha-\delta}^{z^*+\epsilon} \chi_\alpha(\lambda)\rho_N^{(1)}(\lambda)d\lambda + \int_{z^*-\epsilon}^{\beta+\delta} \chi_\beta(\lambda)\rho_N^{(1)}(\lambda)d\lambda. \tag{5.3}$$

(Recall from Observation 2 and formula (4.3) that we need only consider integrals over the bounded set $(\alpha - \delta, \beta + \delta)$ for some sufficiently small (but independent of N) $\delta > 0$.) We will prove that if the function f is C^∞ smooth and compactly supported within $(\alpha - \delta, \beta + \delta)$, the second integral in (5.3) above possesses an asymptotic expansion in even powers of N:

$$\int_{z^*-\epsilon}^{\beta+\delta} \chi_\beta(\lambda)\rho_N^{(1)}(\lambda)d\lambda = f_0 + N^{-2}f_1 + N^{-4}f_2 + \cdots. \tag{5.4}$$

As we have explained, this will complete the proof that

$$\frac{\partial}{\partial t_\ell}\frac{1}{N^2}\log Z_N(\mathbf{t}) = \tilde{e}_0(\mathbf{t}) + \frac{1}{N^2}\tilde{e}_1(\mathbf{t}) + \cdots \tag{5.5}$$

i.e. that $\partial_{t_\ell} \log Z_N$ possesses a complete asymptotic expansion in even powers of N. Since this asymptotic expansion is uniformly valid for all $\mathbf{t} \in \mathbb{T}(T,\gamma)$, we may integrate (5.5) from $\mathbf{t} = \mathbf{0}$, and thus we will complete the proof of Theorem 1.1.

From the analysis of the previous section, in particular lemma 4.4, one observes that the terms appearing in an expression of the form $\int_{z^*-\epsilon}^{\beta+\delta} \lambda^\ell \chi_\beta(\lambda)\rho_N^{(1,\beta)}(\lambda)$ are of one of the following four types of integrals:

$$(0') \qquad \frac{1}{N}\int_{z^*-\epsilon}^{\beta+\delta} g(z)F_0(\Phi_\beta)\Phi'_\beta dz,$$

$$(1') \qquad N^{-j}\int_{z^*-\epsilon}^{\beta+\delta} g(z)Ai(\Phi_\beta)Ai'(\Phi_\beta)dz \quad (j \text{ odd}),$$

$$(2') \qquad N^{-j}\int_{z^*-\epsilon}^{\beta+\delta} g(z)\frac{\sqrt{\Phi_\beta}}{\gamma(z)^2}Ai^2(\Phi_\beta)dz = N^{1/3-j}\int_{z^*-\epsilon}^{\beta+\delta}\tilde{g}(z)Ai^2(\Phi_\beta)dz \ (\ j \text{ even}),$$

$$(3') \ N^{-j}\int_{z^*-\epsilon}^{\beta+\delta} g(z)\frac{\gamma(z)^2}{\sqrt{\Phi_\beta}}[Ai']^2(\Phi_\beta)dz = N^{-1/3-j}\int_{z^*-\epsilon}^{\beta+\delta}\tilde{g}(z)[Ai']^2(\Phi_\beta)dz, \quad (j \text{ even}),$$

where $g(z)$ and $\tilde{g}(z)$ are general infinitely differentiable functions of z, and are compactly supported within $(z^*-\epsilon, \beta+\delta)$. The function F_0 appearing in (0') above is defined by

$$F_0(\zeta) := [Ai']^2(\zeta) - \zeta Ai^2(\zeta). \tag{5.6}$$

The main result would be established if we could show that all of these quantities possess asymptotic expansions in even powers of N. We therefore must analyze the following four quantities:

$$(0) \quad \frac{1}{N}\int_{z^*-\epsilon}^{\beta+\delta} g(z)F_0(\Phi_\beta)\Phi'_\beta dz, \qquad (1) \quad \int_{z^*-\epsilon}^{\beta+\delta} g(z)Ai(\Phi_\beta)Ai'(\Phi_\beta)dz,$$

$$(2) \quad N^{1/3}\int_{z^*-\epsilon}^{\beta+\delta}\tilde{g}(z)Ai^2(\Phi_\beta)dz, \qquad (3) \quad N^{-1/3}\int_{z^*-\epsilon}^{\beta+\delta}\tilde{g}(z)[Ai']^2(\Phi_\beta)dz,$$

and prove the following statements true:

(0), (2), and (3) have asymptotic expansions in even powers of N.
(1) has an asymptotic expansion in odd powers of N.

In fact, once we show that integrals of the form (0) have asymptotic expansions in even powers of N, the following three basic observations take care of the remaining types of integrals.

Observation 1:

(1) has an asymptotic expansion in odd powers of N if and only if (2) has an asymptotic expansion in even powers of N:

$$\begin{aligned} 2\int_{z^*-\epsilon}^{\beta+\delta} gAi(\Phi_\beta)Ai'(\Phi_\beta)dz &= -\int_{z^*-\epsilon}^{\beta+\delta}\left[\frac{g}{\Phi'_\beta}\right]' Ai^2(\Phi_\beta)dz \\ &= -\frac{1}{N}\int_{z^*-\epsilon}^{\beta+\delta} N^{1/3}\left[\frac{g}{\phi'_\beta}\right]' Ai^2(\Phi_\beta)dz \end{aligned}$$

where in the first equality we've integrated by parts. (No boundary terms are present as the integrand is compactly supported within $(z^*-\epsilon, \beta+\delta)$.) We now observe that $\Phi'_\beta(z)$ is analytic and nonvanishing throughout the region of integration so that the integral is well-defined. Lastly, the notation $\phi'_\beta(z) := N^{-2/3}\Phi'_\beta(z)$ defines an analytic, nonvanishing function which is independent of N. The quantity $\left[g/\phi'_\beta\right]'$ is a new infinitely differentiable function of z, compactly supported within $(z^*-\epsilon, \beta+\delta)$, and we have shown that integrals of the form (2) possess asymptotic expansions in even powers of N if and only if integrals of the form (1) possess expansions in odd powers of N.

Observation 2:

Integrals of the form (2) have asymptotic expansions in even powers of N if and only if integrals of the form (0) have asymptotic expansions in even powers of N:

$$\begin{aligned} N^{1/3}\int_{z^*-\epsilon}^{\beta+\delta} gAi^2(\Phi_\beta)dz &= N^{1/3}\int_{z^*-\epsilon}^{\beta+\delta}\frac{g}{\Phi_\beta'}Ai^2(\Phi_\beta)d\Phi_\beta \\ &= \frac{1}{N}\int_{z^*-\epsilon}^{\beta+\delta}\left[\frac{g}{\phi_\beta'}\right]'\frac{1}{\phi_\beta'}F_0(\Phi_\beta)\Phi_\beta' dz \end{aligned}$$

where once again we have integrated by parts in the second equality. Now we have arrived at an integral of the form (0), since the quantity $\left(1/\phi_\beta'\right)\left[g/\phi_\beta'\right]'$ is infinitely differentiable on $[z^*-\epsilon, \beta+\delta]$.

Observation 3:

Finally, if integrals of the forms (0) and (1) respectively have even and odd asymptotic expansions, then integrals of the form (3) have even asymptotic expansions:

$$\begin{aligned} N^{-1/3}\int_{z^*-\epsilon}^{\beta+\delta} g[Ai']^2(\Phi_\beta)dz &= \\ & -N^{-1}\int_{z^*-\epsilon}^{\beta+\delta}\left[\frac{g}{\phi_\beta'}\right]'\{\frac{2}{3}Ai(\Phi_\beta)Ai'(\Phi_\beta)+\frac{1}{3}\Phi_\beta F_0(\Phi_\beta)\}dz \\ &= -\frac{2}{3}N^{-1}\int_{z^*-\epsilon}^{\beta+\delta}\left[\frac{g}{\phi_\beta'}\right]' Ai(\Phi_\beta)Ai'(\Phi_\beta)dz \\ & \quad -\frac{1}{3}N^{-1}\int_{z^*-\epsilon}^{\beta+\delta}\left[\frac{g}{\phi_\beta'}\right]'\frac{\phi_\beta}{\phi_\beta'}F_0(\Phi_\beta)\Phi_\beta' dz. \end{aligned}$$

There is an integration by parts in the first equation. The two terms in the last equation are integrals of the form (1) and (0), respectively, and this establishes our third observation.

We turn now to the heart of the matter. We will give an inductive proof that

$$\frac{1}{N}\int_{z^*-\epsilon}^{\beta+\delta} g(z)F_0(\Phi_\beta)\Phi_\beta' dz \tag{5.7}$$

has an asymptotic expansion in even powers of N. We will integrate by parts repeatedly, peeling off contributing factors as we go along. The proof proceeds in four steps. The first three steps involve setting up an integration by parts inductive argument.

Step 1: Show that the function F_0, which is analytic in ζ, possesses the following asymptotic expansion for $\zeta \to -\infty$:

$$\begin{aligned} F_0(\zeta) &= \sqrt{-\zeta}\left(c_0 + c_1(-\zeta)^{-3} + c_2(-\zeta)^{-6}\cdots\right) \\ &+\sqrt{-\zeta}\left(d_0(-\zeta)^{-3} + d_1(-\zeta)^{-6} + d_2(-\zeta)^{-9} + \cdots\right)\sin\left(\frac{4}{3}(-\zeta)^{3/2}\right) \\ &+\left(f_0(-\zeta)^{-1} + f_1(-\zeta)^{-4} + f_2(-\zeta)^{-7} + \cdots\right)\cos\left(\frac{4}{3}(-\zeta)^{3/2}\right). \end{aligned} \tag{5.8}$$

Step 2: Show that Step 1 implies that there are functions $G_1(\zeta)$, $G_2(\zeta)$, and $G_3(\zeta)$ which are analytic in ζ, satisfy

$$G_1'(\zeta) = \left(F_0(\zeta) - c(-\zeta)^{1/2}\right), \tag{5.9}$$

$$G_2'(\zeta) = G_1(\zeta), \tag{5.10}$$

$$G_3'(\zeta) = G_2(\zeta), \tag{5.11}$$

and possesses the following asymptotic expansion for $\zeta \to -\infty$:

$$\begin{aligned} G_1(\zeta) &= \left(c_0^{(1)}(-\zeta)^{-3/2} + c_1^{(1)}(-\zeta)^{-9/2} + \cdots\right) \\ &\quad + \left(d_0^{(1)}(-\zeta)^{-3/2} + d_1^{(1)}(-\zeta)^{-9/2} + \cdots\right)\sin\left(\frac{4}{3}(-\zeta)^{3/2}\right) \\ &\quad + \left(f_0^{(1)}(-\zeta)^{-3} + f_1^{(1)}(-\zeta)^{-6} + \cdots\right)\cos\left(\frac{4}{3}(-\zeta)^{3/2}\right) \end{aligned} \tag{5.12}$$

$$\begin{aligned} G_2(\zeta) &= \left(c_0^{(2)}(-\zeta)^{-1/2} + c_1^{(2)}(-\zeta)^{-7/2} + \cdots\right) \\ &\quad + \left(d_0^{(2)}(-\zeta)^{-7/2} + d_1^{(2)}(-\zeta)^{-13/2} + \cdots\right)\sin\left(\frac{4}{3}(-\zeta)^{3/2}\right) \\ &\quad + \left(f_0^{(2)}(-\zeta)^{-2} + f_1^{(2)}(-\zeta)^{-5} + \cdots\right)\cos\left(\frac{4}{3}(-\zeta)^{3/2}\right) \end{aligned} \tag{5.13}$$

$$\begin{aligned} G_3(\zeta) &= c_0^{(3)}(-\zeta)^{1/2} + c_1^{(3)}(-\zeta)^{-5/2} + \cdots \\ &\quad + \left(d_0^{(3)}(-\zeta)^{-5/2} + d_1^{(3)}(-\zeta)^{-11/2} + \cdots\right)\sin\left(\frac{4}{3}(-\zeta)^{3/2}\right) \\ &\quad + \left(f_0^{(3)}(-\zeta)^{-4} + f_1^{(3)}(-\zeta)^{-7} + \cdots\right)\cos\left(\frac{4}{3}(-\zeta)^{3/2}\right). \end{aligned} \tag{5.14}$$

This defines antiderivatives of F_0 which we will also need in what follows:

$$F_1 = G_1 - \frac{2c_0}{3}(-\zeta)^{3/2}, \ (\ F_1' = F_0\ )\ , \tag{5.15}$$

$$F_2 = G_2 + \frac{4c_0}{15}(-\zeta)^{5/2}, \ (\ F_2' = F_1\ )\ , \tag{5.16}$$

$$F_3 = G_3 - \frac{8c_0}{105}(-\zeta)^{7/2}, \ (\ F_3' = F_2\ )\ , \tag{5.17}$$

and we will show that these three anti-derivatives are all exponentially decaying for $\zeta \to +\infty$:

$$|F_\ell| \leq C_\ell e^{-\frac{4}{3}\zeta^{3/2}} \tag{5.18}$$

Step 3: We express the integral (5.7) as follows

$$\frac{1}{N}\int_{z^*-\epsilon}^{\beta+\delta} g(z)F_0(\Phi_\beta)\Phi_\beta' dz = \hat{e}_0 + A, \tag{5.19}$$

$$\hat{e}_0 = \frac{1}{N}\int_{z^*-\epsilon}^{\beta} g(z)c_0\left(-\Phi_\beta(z)\right)^{1/2}\Phi_\beta' dz \tag{5.20}$$

$$A = \frac{1}{N}\int_{z^*-\epsilon}^{\beta} g(z)\left(F_0(\Phi_\beta) - c_0\left(-\Phi_\beta(z)\right)^{1/2}\right)\Phi_\beta' dz + \frac{1}{N}\int_{\beta}^{\beta+\delta} g(z)F_0(\Phi_\beta)\Phi_\beta' dz \tag{5.21}$$

Now $\hat{e}_0$ is in fact independent of N:

$$\hat{e}_0 = \frac{-c_0}{N}\int_{z^*-\epsilon}^{\beta} g(z)\frac{d}{dz}\left(\frac{2}{3}\left(-\Phi_\beta(z)\right)^{3/2}\right)dz = \frac{-ic_0}{2}\int_{z^*-\epsilon}^{\beta} g(z)R_+(z)h(z)dz \tag{5.22}$$

and so we must show that A possesses an asymptotic expansion in even powers of N. Integrating both terms in (5.21) by parts, we obtain

$$A = -\frac{1}{N}\int_{z^*-\epsilon}^{\beta} g'(z)G_1\left(\Phi_\beta(z)\right)dz - \frac{1}{N}\int_{\beta}^{\beta+\delta} g'(z)F_1\left(\Phi_\beta(z)\right)dz. \tag{5.23}$$

Observe that in (5.23), the boundary terms from $z^*-\epsilon$ and $\beta+\delta$ do not contribute because g is compactly supported within $(z^*-\epsilon, \beta+\delta)$. Furthermore, the boundary terms from the integrals at $z=\beta$ cancel, because $F_1(\Phi_\beta(z))$ and $G_1(\Phi_\beta(z))$ coincide at $z=\beta$. Since G_1 and F_1 are both uniformly bounded, it follows immediately that $|A| \leq CN^{-1}$. However, it is not yet straightforward to deduce the asymptotic behavior for A as $N \to \infty$, and that is our next task.

Two more integrations by parts yield

$$A = -\frac{1}{N}\int_{z^*-\epsilon}^{\beta}\left[\frac{1}{\Phi_\beta'}\left(\frac{1}{\Phi_\beta'}g'(z)\right)'\right]' G_3\left(\Phi_\beta(z)\right)dz - \frac{1}{N}\int_{\beta}^{\beta+\delta}\left[\frac{1}{\Phi_\beta'}\left(\frac{1}{\Phi_\beta'}g'(z)\right)'\right]' F_3\left(\Phi_\beta(z)\right)dz. \tag{5.24}$$

Recalling the asymptotic expansion (5.14) for G_3 valid for $\zeta \to -\infty$, we may rewrite (5.24) as follows:

$$A = \hat{e}_1\, N^{-2} + A_2, \tag{5.25}$$

$$\hat{e}_1 = -N\int_{z^*-\epsilon}^{\beta}\left[\frac{1}{\Phi_\beta'}\left(\frac{1}{\Phi_\beta'}g'(z)\right)'\right]' c_0^{(3)}\left(-\Phi_\beta(z)\right)^{1/2}dz, \tag{5.26}$$

$$A_2 = -\frac{1}{N}\int_{z^*-\epsilon}^{\beta}\left[\frac{1}{\Phi_\beta'}\left(\frac{1}{\Phi_\beta'}g'(z)\right)'\right]'\left(G_3\left(\Phi_\beta(z)\right) - c_0^{(3)}\left(-\Phi_\beta(z)\right)^{1/2}\right)dz - \frac{1}{N}\int_{\beta}^{\beta+\delta}\left[\frac{1}{\Phi_\beta'}\left(\frac{1}{\Phi_\beta'}g'(z)\right)'\right]' F_3\left(\Phi_\beta(z)\right)dz. \tag{5.27}$$

As with $\hat{e}_0$, $\hat{e}_1$ is independent of N. Indeed, using (4.7), we have

$$\hat{e}_1 = -\int_{z^*-\epsilon}^{\beta}\left[\frac{1}{\phi_\beta'}\left(\frac{1}{\phi_\beta'}g'(z)\right)'\right]'\tilde{c}_0\left(-\phi_\beta(z)\right)^{1/2}dz, \tag{5.28}$$

$$\text{where for } z \in (z^*-\epsilon, \beta),\ \phi_\beta(z) = -\left(\frac{-3i}{4}\int_z^{\beta} R_+(s)h(s)ds\right)^{2/3}. \tag{5.29}$$

Observe that once we have established Steps 1 and 2, then Step 3, carried out above, shows that

$$\frac{1}{N}\int_{z^*-\epsilon}^{\beta+\delta} g(z)F_0(\Phi_\beta)\Phi_\beta' dz = \hat{e}_0 + \frac{1}{N^2}\hat{e}_1 + A_2, \tag{5.30}$$

with $\hat{e}_0$ and $\hat{e}_1$ independent of N (see (5.22) and (5.28), respectively), and with A_2 defined in (5.27). Furthermore, it is straightforward to prove that $|A_2| \le CN^{-7/3}$, and hence (5.30) shows that the first two terms in the asymptotic expansion of (5.7) are in even powers of N, as desired. The above calculations must now be automated to establish the existence of a complete asymptotic expansion in even powers of N.

Step 4: The procedure should now be clear to establish the induction. We must prove that we can integrate by parts thrice, peel off the next term in the asymptotic expansion, and repeat.

We must prove that there exists G_m, F_m, $m = 4, 5, \ldots$, satisfying a number of relations, all appearing in triples. First the differential relations:

$$G_{3j+1}' = G_{3j} - c_0^{(3j)}(-\zeta)^{1/2},\ G_{3j+2}' = G_{3j+1},\ G_{3j+3}' = G_{3j+2}, \tag{5.31}$$

$$F_{3j+1}' = F_{3j},\ F_{3j+2}' = F_{3j+1},\ F_{3j+3}' = F_{3j+2}, \tag{5.32}$$

then the asymptotic prescription

$$G_{3j+3} = (-\zeta)^{1/2}\left(c_0^{(3j+3)} + c_1^{(3j+3)}(-\zeta)^{-3} + \cdots\right) + G_S^{(3j+3)}\sin\left(\frac{4}{3}(-\zeta)^{3/2}\right) + G_C^{(3j+3)}\cos\left(\frac{4}{3}(-\zeta)^{3/2}\right), \tag{5.33}$$

where $G_S^{(3j+3)}$ and $G_C^{(3j+3)}$ are asymptotic series which behave as follows:

$$\text{For } j \text{ even:} \begin{cases} G_S^{(3j+3)} = (-\zeta)^{-(3j+5)/2}\left(d_0^{(3j+3)} + d_1^{(3j+3)}(-\zeta)^{-3} + \cdots\right) \\ G_C^{(3j+3)} = (-\zeta)^{-(3j+8)/2}\left(f_0^{(3j+3)} + f_1^{(3j+3)}(-\zeta)^{-3} + \cdots\right) \end{cases} \tag{5.34}$$

$$\text{For } j \text{ odd:} \begin{cases} G_S^{(3j+3)} = (-\zeta)^{-(3j+8)/2}\left(d_0^{(3j+3)} + d_1^{(3j+3)}(-\zeta)^{-3} + \cdots\right) \\ G_C^{(3j+3)} = (-\zeta)^{-(3j+5)/2}\left(f_0^{(3j+3)} + f_1^{(3j+3)}(-\zeta)^{-3} \cdots\right) \end{cases} \tag{5.35}$$

(note that this implies a similar asymptotic behavior for G_{3j+1} and G_{3j+2}). The quantities F_ℓ must satisfy the following estimates for $\zeta \to +\infty$:

$$|F_{3j+1}|, |F_{3j+2}|, |F_{3j+3}| \leq C_{3j} e^{-\frac{4}{3}\zeta^{3/2}}, \tag{5.36}$$

and finally we require the following relation between F_ℓ and G_ℓ:

$$G_m(0) = F_m(0), \quad m = 3j+1,\ 3j+2,\ 3j+3. \tag{5.37}$$

Once we have the existence of the functions F_ℓ and G_ℓ out of the way, then we may proceed with the inductive integration by parts. Suppose we have arrived at the situation

$$\frac{1}{N}\int_{z^*-\epsilon}^{\beta+\delta} g(z) F_0(\Phi_\beta)\Phi_\beta' dz = \hat{e}_0 + N^{-2}\hat{e}_1 + \cdots + N^{-2\ell}\hat{e}_\ell + A_{\ell+1}, \tag{5.38}$$

where $\hat{e}_j$ are all independent of N, and

$$\begin{aligned} A_{\ell+1} = \frac{(-1)^\ell}{N}\int_{z^*-\epsilon}^{\beta+\delta} \frac{d}{dz}\left((\Phi_\beta')^{-1}\frac{d}{dz}\left((\Phi_\beta')^{-1} \cdots \frac{d}{dz}\left((\Phi_\beta')^{-1} g'(z)\right)\cdots\right)\right) \times \\ \times \left(G_{3\ell}(\Phi_\beta(z)) - c_0^{(3\ell)}(-\Phi_\beta)^{1/2}\right) dz \\ + \frac{(-1)^\ell}{N}\int_{z^*-\epsilon}^{\beta+\delta} \frac{d}{dz}\left((\Phi_\beta')^{-1}\frac{d}{dz}\left((\Phi_\beta')^{-1} \cdots \frac{d}{dz}\left((\Phi_\beta')^{-1} g'(z)\right)\cdots\right)\right) \times \\ \times F_{3\ell}(\Phi_\beta) dz, \end{aligned} \tag{5.39}$$

and in each nested set of derivatives appearing in the integrals in (5.39), the differential operator

$$\frac{d}{dz}\left((\Phi_\beta')^{-1}\ \cdot\right) \tag{5.40}$$

appears $3\ell - 1$ times. We further assume that in (5.39), $G_{3\ell}$ satisfies (5.33), $F_{3\ell}$ satisfies (5.36), (both with $j = \ell - 1$) and (5.37) holds with $m = 3\ell$. In order to integrate by parts, we note that from (5.31) and (5.32),

$$\left(G_{3\ell}(\Phi_\beta(z)) - c_0^{(3\ell)}(-\Phi_\beta)^{1/2}\right) = \frac{1}{\Phi_\beta'}\frac{d}{dz} G_{3\ell+1}(\Phi_\beta(z)), \tag{5.41}$$

$$F_{3\ell}(\Phi_\beta(z)) = \frac{1}{\Phi_\beta'(z)}\frac{d}{dz} F_{3\ell+1}(\Phi_\beta(z)). \tag{5.42}$$

Using (5.37) and integrating by parts, we find

$$\begin{aligned} A_{\ell+1} = \\ -\frac{(-1)^\ell}{N}\int_{z^*-\epsilon}^{\beta+\delta} \frac{d}{dz}\left((\Phi_\beta')^{-1}\frac{d}{dz}\left((\Phi_\beta')^{-1} \cdots \frac{d}{dz}\left((\Phi_\beta')^{-1} g'(z)\right)\cdots\right)\right) \times \\ \times (G_{3\ell+1})\ dz \\ -\frac{(-1)^\ell}{N}\int_{z^*-\epsilon}^{\beta+\delta} \frac{d}{dz}\left((\Phi_\beta')^{-1}\frac{d}{dz}\left((\Phi_\beta')^{-1} \cdots \frac{d}{dz}\left((\Phi_\beta')^{-1} g'(z)\right)\cdots\right)\right) \times \\ \times F_{3\ell+1}(\Phi_\beta) dz, \end{aligned} \tag{5.43}$$

where now the differential operator $(d/dz(\Phi_\beta')^{-1}\ \cdot)$ appears 3ℓ times. At this point, using (4.7), it is straightforward to prove that $|A_{\ell+1}| \leq CN^{-2\ell-2/3}$. However, we may integrate by parts two more times, and we then arrive at

$$
\begin{aligned}
A_{\ell+1} = &\\
&-\frac{(-1)^\ell}{N}\int_{z^*-\epsilon}^{\beta+\delta}\frac{d}{dz}\left((\Phi_\beta')^{-1}\ \frac{d}{dz}\left((\Phi_\beta')^{-1}\ \cdots\frac{d}{dz}\left((\Phi_\beta')^{-1}\ g'(z)\right)\cdots\right)\right)\times\\
&\qquad\times\left(G_{3\ell+3}\right)\ dz\\
&-\frac{(-1)^\ell}{N}\int_{z^*-\epsilon}^{\beta+\delta}\frac{d}{dz}\left((\Phi_\beta')^{-1}\ \frac{d}{dz}\left((\Phi_\beta')^{-1}\ \cdots\frac{d}{dz}\left((\Phi_\beta')^{-1}\ g'(z)\right)\cdots\right)\right)\times\\
&\qquad\times F_{3\ell+3}(\Phi_\beta)dz,
\end{aligned}
\tag{5.44}
$$

where now the differential operator $(d/dz(\Phi_\beta')^{-1}\ \cdot)$ appears $3\ell+2$ times in each integral. Expression (5.44) may now be rewritten as follows:

$$
A_{\ell+1} = N^{-2\ell-2}\hat{e}_{\ell+1} + A_{\ell+2},
\tag{5.45}
$$

$$
\begin{aligned}
\hat{e}_{\ell+1} = &\\
&(-1)^{\ell+1}N^{2\ell+1}\int_{z^*-\epsilon}^{\beta+\delta}\frac{d}{dz}\left((\Phi_\beta')^{-1}\ \frac{d}{dz}\left((\Phi_\beta')^{-1}\ \cdots\frac{d}{dz}\left((\Phi_\beta')^{-1}\ g'(z)\right)\cdots\right)\right)\times\\
&\qquad\times\left(c_0^{(3\ell+3)}\left(-\Phi_\beta\right)^{1/2}\right)\ dz,
\end{aligned}
\tag{5.46}
$$

$$
\begin{aligned}
A_{\ell+2} = &\\
&\frac{(-1)^{\ell+1}}{N}\int_{z^*-\epsilon}^{\beta+\delta}\frac{d}{dz}\left((\Phi_\beta')^{-1}\ \frac{d}{dz}\left((\Phi_\beta')^{-1}\ \cdots\frac{d}{dz}\left((\Phi_\beta')^{-1}\ g'(z)\right)\cdots\right)\right)\times\\
&\qquad\times\left(G_{3\ell+3} - c_0^{(3\ell+3)}\left(-\Phi_\beta\right)^{1/2}\right)\ dz\\
&\frac{(-1)^{\ell+1}}{N}\int_{z^*-\epsilon}^{\beta+\delta}\frac{d}{dz}\left((\Phi_\beta')^{-1}\ \frac{d}{dz}\left((\Phi_\beta')^{-1}\ \cdots\frac{d}{dz}\left((\Phi_\beta')^{-1}\ g'(z)\right)\cdots\right)\right)\times\\
&\qquad\times F_{3\ell+3}(\Phi_\beta)dz,
\end{aligned}
\tag{5.47}
$$

where the differential operator $(d/dz(\Phi_\beta')^{-1}\ \cdot)$ appears $3\ell+2$ times in each integral. The quantity $\hat{e}_{\ell+1}$ is easily seen to be independent of N using the representation (4.7) of Φ_β:

$$
\begin{aligned}
\hat{e}_{\ell+1} = &\\
&(-1)^{\ell+1}\int_{z^*-\epsilon}^{\beta+\delta}\frac{d}{dz}\left((\phi_\beta')^{-1}\ \frac{d}{dz}\left((\phi_\beta')^{-1}\ \cdots\frac{d}{dz}\left((\phi_\beta')^{-1}\ g'(z)\right)\cdots\right)\right)\times\\
&\qquad\times\left(c_0^{(3\ell+3)}\left(-\phi_\beta\right)^{1/2}\right)\ dz.
\end{aligned}
\tag{5.48}
$$

Therefore we have deduced that

$$
\frac{1}{N}\int_{z^*-\epsilon}^{\beta+\delta} g(z)F_0(\Phi_\beta)\Phi_\beta' dz = \hat{e}_0 + N^{-2}\hat{e}_1 + \cdots + N^{-2\ell-2}\hat{e}_{\ell+1} + A_{\ell+2},
\tag{5.49}
$$

with $\hat{e}_{\ell+1}$ defined in (5.48), and $A_{\ell+2}$ defined in (5.47). Summarizing, once we establish the existence of the functions F_ℓ and G_ℓ, we have established by induction that (5.7) has an asymptotic expansion in even powers of N.

Proof of Step 1: We establish Step 1 using well-known [1] asymptotic representations of the Airy functions for large negative values of the argument:

$$Ai(-z) = \frac{1}{\sqrt{\pi}z^{1/4}}\left(\sin(\zeta+\frac{\pi}{4})w_1(\zeta) - \cos(\zeta+\frac{\pi}{4})w_2(\zeta)\right), |\arg z| < \frac{2}{3}\pi \tag{5.50}$$

$$Ai'(-z) = \frac{z^{1/4}}{\sqrt{\pi}}\left(-\cos(\zeta+\frac{\pi}{4})w_3(\zeta) - \sin(\zeta+\frac{\pi}{4})w_4(\zeta)\right), |\arg z| < \frac{2}{3}\pi, \tag{5.51}$$

where

$$\zeta = \frac{2}{3}(z)^{3/2} \tag{5.52}$$

and the $w_i(\zeta)$ are asymptotic series in large ζ defined recursively in terms of the coefficients

$$\begin{aligned} u_s &= \frac{(2s+1)(2s+3)(2s+5)\dots(6s-1)}{(216)^s s!} \\ v_s &= -\frac{6s+1}{6s-1}u_s \\ \text{with} \quad & u_0 = 1, \; v_0 = 1. \end{aligned}$$

Explicitly the asymptotic series are given by

$$w_1 = \sum_{s=0}^{\infty}(-1)^s\frac{u_{2s}}{\zeta^{2s}}, \qquad w_2 = \sum_{s=0}^{\infty}(-1)^s\frac{u_{2s+1}}{\zeta^{2s+1}}, \tag{5.53}$$

$$w_3 = \sum_{s=0}^{\infty}(-1)^s\frac{v_{2s}}{\zeta^{2s}}, \qquad w_4 = \sum_{s=0}^{\infty}(-1)^s\frac{v_{2s+1}}{\zeta^{2s+1}}. \tag{5.54}$$

We caution the reader that in formulae (5.50)-(5.54) the variable $z > 0$ is a dummy variable, as is ζ appearing in (5.52) above. We will replace z appearing in (5.50)-(5.54) with $-\zeta$ (where $\zeta < 0$). Consequently, the parameter ζ appearing in (5.50)-(5.54) becomes $(2/3)(-\zeta)^{3/2}$.

One can now substitute the expansions of the Airy functions (5.50) into the expression (5.6) for $F_0(\zeta)$ and collect separately the coefficients of $\cos(4(-\zeta)^{3/2}/3)$, $\sin(4(-\zeta)^{3/2}/3)$ and the "non-oscillatory" terms. One finds that the expansion is of the form

$$\begin{aligned} F_0(\zeta) = (-\zeta)^{1/2}\; S_0^{(0)} &+ \\ &+(-\zeta)^{1/2}\; S_1^{(0)} \sin\left(\frac{4}{3}(-\zeta)^{3/2}\right) + \; S_2^{(\ell)} \cos\left(\frac{4}{3}(-\zeta)^{3/2}\right), \end{aligned} \tag{5.55}$$

where $S_0^{(0)}$, $S_1^{(0)}$ and $S_2^{(0)}$ are asymptotic series valid for $\zeta \to -\infty$:

$$S_0^{(0)} = \frac{1}{2\pi}(w_1^2 + w_2^2 + w_3^2 + w_4^2), \tag{5.56}$$

$$S_1^{(0)} = \frac{1}{2\pi}\left(w_1^2 - w_2^2 - w_3^2 + w_4^2\right), \tag{5.57}$$

$$S_2^{(0)} = \frac{(-\zeta)^{1/2}}{\pi}\left(w_3 w_4 - w_1 w_2\right). \tag{5.58}$$

Now using the definitions (5.53)-(5.54) of the asymptotic series $\{w_j\}_{j=1}^4$ (with ζ replaced by $(-\zeta)^{3/2}$), we see that $S_0^{(0)}$ is an asymptotic expansion in powers of $(-\zeta)^{-3}$, starting with the constant term. Similarly, $S_1^{(0)}$ is an asymptotic expansion in powers of $(-\zeta)^{-3}$ starting with $(-\zeta)^{-3}$, and $S_2^{(0)}$ is an asymptotic expansion in powers of the form $(-\zeta)^{-3p-1}$, $p = 0, 1, \dots$. This establishes the form of the expansion specified in Step 1, formula (5.8)

Proof of Steps 2 and 4: Now we must establish the existence of F_ℓ and G_ℓ, $\ell \geq 1$, satisfying (5.9)-(5.18) (for $\ell = 1, 2, 3$), and (5.31)-(5.37) (for $\ell \geq 4$).

We observe that if $\{F_\ell\}_{\ell\geq 1}$ satisfies (5.15)-(5.17) and (5.32), then clearly F_ℓ is the ℓ-th antiderivative of the function F_0. Now $F_0(\zeta) = \left(Ai'(\zeta)\right)^2 - \zeta\left(Ai(\zeta)\right)^2$, and it straightforward to prove that the ℓ-th antiderivative of F_0 is given by

$$F_\ell(\zeta) = q_0^{(\ell)}(\zeta)\left(Ai'(\zeta)\right)^2 + q_1^{(\ell)}(\zeta)\left(Ai(\zeta)\right)^2 + q_3^{(\ell)}(\zeta)Ai(\zeta)Ai'(\zeta) \tag{5.59}$$

where $\{q_\mu^{(\ell)}\}$ are polynomials in ζ. Now (5.59), together with the asymptotics for $\zeta \to \infty$ of the Airy function $Ai(\zeta)$ imply that for each ℓ, F_ℓ satisfies (5.18) and (5.36). Furthermore, using the asymptotics (5.50) and (5.51) (again replacing z with $-\zeta$, and ζ with $(2/3)(-\zeta)^{3/2}$), and letting $\zeta \to -\infty$, we find that for each ℓ, F_ℓ possesses an expansion of the following form

$$F_\ell(\zeta) = (-\zeta)^{1/2}\ S_0^{(\ell)} + \\ +(-\zeta)^{1/2}\ S_1^{(\ell)} \sin\left(\frac{4}{3}(-\zeta)^{3/2}\right) + S_2^{(\ell)} \cos\left(\frac{4}{3}(-\zeta)^{3/2}\right), \tag{5.60}$$

where S_0, S_1, and S_2 are asymptotic series valid for $\zeta \to -\infty$. We will refine our knowledge of these series in what follows, but for now we observe the following important basic property of these asymptotic series:

Property 5.61: The asymptotic series $S_0^{(\ell)}$, $S_1^{(\ell)}$ and $S_2^{(\ell)}$ are all in integer powers of $(-\zeta)$.

In order to obtain more detailed information about the asymptotic series $S_0^{(\ell)}$, $S_1^{(\ell)}$ and $S_2^{(\ell)}$, we will make use of the following Lemma.

Lemma 5.1. *If a function $G(\zeta)$ is C^∞ on $(-\infty, a)$ for some $a > 0$, and possesses the following asymptotic expansion (which may be differentiated repeatedly to obtain asymptotic expansions for derivatives),*

$$G(\zeta) = \left(a(-\zeta)^{\frac{-j}{2}} + b(-\zeta)^{\frac{-(j+3)}{2}} + \dots\right) trig\left(\frac{4}{3}(-\zeta)^{3/2}\right), \tag{5.62}$$

in which the asymptotic expansion in parentheses descends in powers of $(-\zeta)^{3/2}$, $j \geq 1$, and $trig(\cdot)$ denotes either $\sin(\cdot)$ or $\cos(\cdot)$, then the function $\int_{-\infty}^s G(\zeta)d\zeta$ possesses the following asymptotic expansion for $s \to -\infty$:

$$\int_{-\infty}^s G(\zeta)d\zeta = \left(\tilde{a}(-s)^{\frac{-(j+1)}{2}} + \tilde{b}(-s)^{\frac{-(j+4)}{2}} + \dots\right) trig\left(\frac{4}{3}(-s)^{3/2}\right)$$

where the expression in parentheses descends in powers of $(-s)^{3/2}$.

Proof.

$$\begin{aligned}
&\int_{-\infty}^s \left(a(-\zeta)^{\frac{-j}{2}} + b(-\zeta)^{\frac{-(j+3)}{2}} + \dots\right) trig\left(\frac{4}{3}(-\zeta)^{3/2}\right) d\zeta \\
= &\int_{-\infty}^s \left(a(-\zeta)^{\frac{-j}{2}} + b(-\zeta)^{\frac{-(j+3)}{2}} + \dots\right) \frac{-1}{2(-\zeta)^{1/2}} \frac{d}{d\zeta} trig\left(\frac{4}{3}(-\zeta)^{3/2}\right) d\zeta \\
= &-\left(a(-s)^{\frac{-j}{2}} + b(-s)^{\frac{-(j+3)}{2}} + \dots\right) \frac{1}{2(-s)^{1/2}} trig\left(\frac{4}{3}(-s)^{3/2}\right) \\
&+\frac{1}{2}\int_{-\infty}^s \left(a(-\zeta)^{\frac{-(j+1)}{2}} + b(-\zeta)^{\frac{-(j+4)}{2}} + \dots\right)' trig\left(\frac{4}{3}(-\zeta)^{3/2}\right) d\zeta \\
= &-\frac{1}{2}\left(a(-s)^{\frac{-(j+1)}{2}} + b(-s)^{\frac{-(j+4)}{2}} + \dots\right) trig\left(\frac{4}{3}(-s)^{3/2}\right) \\
&+\frac{1}{2}\int_{-\infty}^s \left(a\frac{j+1}{2}(-\zeta)^{\frac{-(j+3)}{2}} + b\frac{j+4}{2}(-\zeta)^{\frac{-(j+6)}{2}} + \dots\right) \times \\
&\qquad \times \frac{-1}{2(-\zeta)^{1/2}} \frac{d}{d\zeta} trig\left(\frac{4}{3}(-\zeta)^{3/2}\right) d\zeta,
\end{aligned}$$

and continuing in this way establishes the lemma. □

Let us write F_1 in the following convoluted way:

$$F_1(\zeta) = \int_{-\infty}^\zeta \left(F_0(s) - c_0(-s)^{1/2}\right) ds - \frac{2}{3}c_0(-\zeta)^{3/2}. \tag{5.63}$$

One sees that F_1 so defined is an antiderivative of F_0 immediately. To verify that F_1 so defined is exponentially decaying for $\zeta \to +\infty$, one uses the asymptotic expansion of F_0 obtained in Step 1 together with Lemma 5.1 to compute the asymptotic expansion of the integral in (5.63), and compares to the asymptotic expansion

(5.60). Since there is no constant term in (5.60), neither in the asymptotic expansion of (5.63), F_1 defined by (5.63) must be expressible in the form (5.59), and hence it is exponentially decaying for $\zeta \to +\infty$.

So we have learned that (1) F_1 exists satisfying $F_1' = F_0$, (2) F_1 satisfies (5.18) for $\zeta \to +\infty$, and (3) F_1 possesses an asymptotic expansion of the form (5.60) for $\zeta \to -\infty$. Furthermore, we learn that the asymptotic series $(-\zeta)^{1/2}S_0^{(1)}$ appearing in (5.60) with $\ell = 1$ is obtained by computing the term-by-term anti-derivative of the asymptotic series $(-\zeta)^{1/2}S_0^{(0)}$ appearing in (5.55). This provides us with more detailed information about the asymptotic series $S_0^{(1)}$:

$$(-\zeta)^{1/2}S_0^{(1)} = -\frac{2}{3}c_0(-\zeta)^{3/2} + \frac{2}{3}c_1(-\zeta)^{-3/2} + \frac{2}{9}c_2(-\zeta)^{-9/2} + \cdots . \tag{5.64}$$

If, rather than using Lemma 5.1, one is more careful integrating by parts repeatedly, it is straightforward to prove the following two properties:

Property 5.65: The expansion $S_1^{(1)}$ starts at $(-\zeta)^{-2}$ and descends in powers of $(-\zeta)^3$.

Property 5.66: The expansion $S_2^{(1)}$ starts at $(-\zeta)^{-3}$ and descends in powers of $(-\zeta)^3$.

Setting $G_1 = F_1 + (2c_0/3)(-\zeta)^{3/2}$, we have proven the existence of G_1 satisfying (5.9) and (5.12), with $c_0^{(1)} = -2c_1/3$ (the constant c_1 appears in (5.8)). We may now define

$$G_2 = \int_{-\infty}^{\zeta} G_1(s)ds, \tag{5.67}$$

The function G_2 clearly satisfies (5.10), and using the asymptotic expansion (5.12) satisfied by G_1 together with careful integration by parts shows that G_2 satisfies the asymptotic expansion (5.13) for $\zeta \to -\infty$. We next define F_2 via

$$F_2 = G_2 + \frac{4c_0}{15}(-\zeta)^{5/2} \tag{5.68}$$

which, it turns out, is equivalent to the definition $F_2 = \int_\infty^\zeta F_1(s)ds$. The function F_2 then clearly satisfies (5.16) and (5.18).

The functions G_3 and F_3 are defined in similar fashion. First we define

$$G_3 = \int_{-\infty}^{\zeta}\left(G_2(s) - c_0^{(2)}(-s)^{-1/2}\right)ds - 2c_0^{(2)}(-\zeta)^{1/2}. \tag{5.69}$$

Second, using the asymptotic expansion (5.13) satisfied by G_2 and careful integration by parts in the manner used to prove Lemma 5.1, we deduce that G_3 possesses the asymptotic expansion (5.14). Third, we define F_3 in the obvious way:

$$F_3 = G_3 - \frac{8c_0}{105}(-\zeta)^{7/2}. \tag{5.70}$$

Uniqueness of asymptotic expansions then implies that $F_3 = \int_\infty^\zeta F_2(s)ds$, and hence F_3 satisfies (5.17) and (5.18).

The inductive argument goes as follows: suppose we have the existence of F_ℓ and G_ℓ for $\ell = 1, \ldots, 3k$, satisfying (5.9)-(5.18), and possibly (5.31) - (5.37) with $j = k-1$ (if $k \geq 2$). We define G_{3k+1} and G_{3k+2} as follows:

$$G_{3k+1}(\zeta) = \int_{-\infty}^{\zeta}\left(G_{3k}(s) - c_0^{(3k)}(-s)^{1/2}\right)ds \tag{5.71}$$

$$G_{3j+2} = \int_{-\infty}^{\zeta} G_{3k+1}(s)ds. \tag{5.72}$$

The asymptotic expansion satisfied by G_{3k} (either (5.14) if $k = 1$, or (5.33) with $j = k-1$ if $k \geq 2$), together with a careful integration by parts argument as outlined in the proof of Lemma 5.1 shows that G_{3k+1} satisfies the following asymptotic expansion:

$$G_{3k+1} = (-\zeta)^{1/2}\left(c_0^{(3k+1)}(-\zeta)^{-2} + c_1^{(3k+1)}(-\zeta)^{-5} + \cdots\right) + G_S^{(3k+1)} \sin\left(\frac{4}{3}(-\zeta)^{3/2}\right) + G_C^{(3k+1)} \cos\left(\frac{4}{3}(-\zeta)^{3/2}\right), \tag{5.73}$$

where $G_S^{(3k+1)}$ and $G_C^{(3k+1)}$ are asymptotic expansions:

(5.74) For $3k+1$ even:

$$\begin{cases} G_S^{(3k+1)} = (-\zeta)^{-(3k+6)/2}\left(d_0^{(3k+1)} + d_1^{(3k+1)}(-\zeta)^{-3} + \cdots\right) \\ G_C^{(3k+1)} = (-\zeta)^{-(3k+3)/2}\left(f_0^{(3k+1)} + f_1^{(3k+1)}(-\zeta)^{-3} + \cdots\right) \end{cases}$$

(5.75) For $3k+1$ odd:

$$\begin{cases} G_S^{(3k+1)} = (-\zeta)^{-(3k+3)/2}\left(d_0^{(3k+1)} + d_1^{(3k+1)}(-\zeta)^{-3} + \cdots\right) \\ G_C^{(3k+1)} = (-\zeta)^{-(3k+6)/2}\left(f_0^{(3k+1)} + f_1^{(3k+1)}(-\zeta)^{-3} \cdots\right) \end{cases}.$$

Similarly, G_{3k+2} satisfies the following asymptotic description for $\zeta \to -\infty$:

$$G_{3k+2} = (-\zeta)^{1/2}\left(c_0^{(3k+2)}(-\zeta)^{-1} + c_1^{(3k+2)}(-\zeta)^{-4} + \cdots\right) + G_S^{(3k+2)} \sin\left(\frac{4}{3}(-\zeta)^{3/2}\right) + G_C^{(3k+2)} \cos\left(\frac{4}{3}(-\zeta)^{3/2}\right), \tag{5.76}$$

where $G_S^{(3k+2)}$ and $G_C^{(3k+2)}$ are asymptotic expansions:

(5.77) For $3k+2$ even:

$$\begin{cases} G_S^{(3k+2)} = (-\zeta)^{-(3k+7)/2}\left(d_0^{(3k+2)} + d_1^{(3k+2)}(-\zeta)^{-3} + \cdots\right) \\ G_C^{(3k+2)} = (-\zeta)^{-(3k+4)/2}\left(f_0^{(3k+4)} + f_1^{(3k+4)}(-\zeta)^{-3} + \cdots\right) \end{cases}$$

(5.78) For $3k+2$ odd:

$$\begin{cases} G_S^{(3k+2)} = (-\zeta)^{-(3k+4)/2}\left(d_0^{(3k+2)} + d_1^{(3k+2)}(-\zeta)^{-3} + \cdots\right) \\ G_C^{(3k+2)} = (-\zeta)^{-(3k+7)/2}\left(f_0^{(3k+2)} + f_1^{(3k+2)}(-\zeta)^{-3} \cdots\right) \end{cases}.$$

We may now define G_{3k+3}:

$$G_{3k+3}(\zeta) = \int_{-\infty}^{\zeta} \left(G_{3k+2}(s) - c_0^{(3k+2)}(-s)^{-1/2}\right) ds - 2c_0^{(3k+2)}(-\zeta)^{1/2}. \tag{5.79}$$

It is by now a straightforward exercise to establish that G_{3k+3} so defined satisfies (5.31) as well as the asymptotic description (5.33) for $\zeta \to -\infty$.

Turning now to the existence of F_{3k+1}, F_{3k+2} and F_{3k+3}, we define them through

$$F_{3k+1} = \int_{\infty}^{\zeta} F_{3k}(s)ds, \quad F_{3k+2} = \int_{\infty}^{\zeta} F_{3k+1}(s)ds, \quad F_{3k+3} = \int_{\infty}^{\zeta} F_{3k+2}(s)ds. \tag{5.80}$$

It is immediately clear from these definitions that $\{F_{3k+\mu}\}_{\mu=1}^{3}$ satisfy (5.18) for $\zeta \to \infty$.

Because of the recursive definitions of F_ℓ and G_ℓ, it follows that

$$F_{3k+1} = G_{3k+1} - \frac{2c_0^{(3k)}}{3}(-\zeta)^{3/2} + \frac{2^4 c_0^{(3k-3)}}{945}(-\zeta)^{9/2} + \cdots + \frac{(-1)^{3k+1}2^{3k+1}c_0}{\prod_{j=1}^{3k+1}(2j+1)}(-\zeta)^{(6k+3)/2}, \tag{5.81}$$

$$F_{3k+2} = G_{3k+2} + \frac{4c_0^{(3k)}}{15}(-\zeta)^{5/2} + \frac{2^5 c_0^{(3k-3)}}{10395}(-\zeta)^{11/2} + \cdots + \frac{(-1)^{3k+2}2^{3k+2}c_0}{\prod_{j=1}^{3k+2}(2j+1)}(-\zeta)^{(6k+5)/2}, \tag{5.82}$$

$$F_{3k+3} = G_{3k+3} - \frac{8c_0^{(3k)}}{105}(-\zeta)^{7/2} + \frac{2^6 c_0^{(3k-3)}}{135135}(-\zeta)^{13/2} + \cdots + \frac{(-1)^{3k+3}2^{3k+3}c_0}{\prod_{j=1}^{3k+3}(2j+1)}(-\zeta)^{(6k+7)/2}. \tag{5.83}$$

we observe that there are no constant terms in the relations (5.81)-(5.83). Indeed, this is so because (1) for each ℓ, F_ℓ has the representation (5.59) for some polynomials $\{q_\mu^{(\ell)}\}_{\mu=0}^2$, and consequently the asymptotic expansion for F_ℓ valid for $\zeta \to -\infty$ is of the form (5.60) with no constant term and (2) G_ℓ as defined possesses an asymptotic expansion with no constant term for $\zeta \to -\infty$. Therefore the arbitrary constants of integration which may appear in (5.81)-(5.83) are all 0.

Finally, (5.81)-(5.83) imply that $\{F_{3k+\mu}\}_{\mu=1}^3$ and $\{G_{2k+\mu}\}_{\mu=1}^3$ satisfy (5.37) with $j = k$.

This completes the proof of Step 2 and Step 4, and we have established the first claim of Theorem 1.3, that asymptotic expansions of the form (1.35) hold true. The explanation for why the coefficients in such an asymptotic expansion are analytic is as follows.

We return to the integral (5.4), and show that each term in *its* asymptotic expansion depends analytically on $\mathbf{t}$. Since the same is true of the other integral appearing in (5.3), this will complete the proof of Theorem 1.3. We begin by observing that each term in the asymptotic expansion of (5.4) is of the form

$$\int_{z^*-\epsilon}^{\beta} \chi_\beta^{(j)}(\lambda) Q(\lambda,\mathbf{t})\,(-\phi_\beta)^{1/2}\,d\lambda,$$

where

- $\chi_\beta^{(j)}(\lambda)$ represents some finite number of derivatives of $\chi_\beta(\lambda)$;
- $Q(\lambda,\mathbf{t})$ represents terms obtained from the integration by parts procedure but do not depend on the partition of unity χ_β (these terms depend analytically on λ and on $\mathbf{t}$);
- $(-\phi_\beta(\lambda))^{1/2}$ may be taken to be analytic with branch cut emanating from $\lambda = \beta$, down the real axis, and passing in particular through $z^* + \epsilon$. For λ away from the cut, this function also depends analytically on $\mathbf{t}$.

We now decompose the integral into a sum:

$$\int_{z^*-\epsilon}^{z^*+\epsilon} \chi_\beta^{(j)}(\lambda) Q(\lambda,\mathbf{t})\,(-\phi_\beta)^{1/2}\,d\lambda = \int_{z^*-\epsilon}^{\beta} \chi_\beta^{(j)}(\lambda) Q(\lambda,\mathbf{t})\,(-\phi_\beta)^{1/2}\,d\lambda + \int_{z^*+\epsilon}^{\beta} Q(\lambda,\mathbf{t})\,(-\phi_\beta)^{1/2}\,d\lambda. \tag{5.84}$$

The first integral on the right hand side of (5.84) is clearly analytic in $\mathbf{t}$, because the integrand is analytic in $\mathbf{t}$. For the second integral, note that χ_β does not appear. This is because the partition of unity is such that $\chi_\beta \equiv 1$ on $(z^* + \epsilon, \beta)$. Now this second integral is seen to be analytic by expressing it as $1/2$ the value of a contour integral encircling the interval $(z^* + \epsilon, \beta)$, passing through $z^* + \epsilon$. This can be done because the integrand is analytic, with a square-root branch point at $\lambda = \beta$. This completes the proof of Theorem 1.3.

6. Appendix

Here we present a set of explicit formulae for $S_1(z)$, the solution of Riemann–Hilbert problem 3.58.

1. for $z \in \mathbb{C} \setminus (B_\delta^\alpha \cup B_\delta^\beta)$

$$
\begin{aligned}
S_1(z) = &\\
&\frac{5}{144h(\alpha)}\left\{3(z-\alpha)^{-2} + \frac{3}{5}\left[\frac{1}{\alpha-\beta} - 3\frac{h'(\alpha)}{h(\alpha)}\right](z-\alpha)^{-1}\right\}\begin{pmatrix} -1 & -i \\ -i & 1 \end{pmatrix}\\
&+\frac{7}{48(\alpha-\beta)h(\alpha)}(z-\alpha)^{-1}\begin{pmatrix} -1 & i \\ i & 1 \end{pmatrix}\\
&+\frac{5}{144h(\beta)}\left\{3(z-\beta)^{-2} + \frac{3}{5}\left[\frac{1}{\beta-\alpha} - 3\frac{h'(\beta)}{h(\beta)}\right](z-\beta)^{-1}\right\}\begin{pmatrix} -1 & i \\ i & 1 \end{pmatrix}\\
&+\frac{7}{48(\beta-\alpha)h(\beta)}(z-\beta)^{-1}\begin{pmatrix} -1 & i \\ i & 1 \end{pmatrix}.
\end{aligned} \tag{6.1}
$$

2. For $z \in B_\delta^\beta$,

$$
\begin{aligned}
S_1(z) = &\\
&\frac{5}{144h(\alpha)}\left\{3(z-\alpha)^{-2} + \frac{3}{5}\left[\frac{1}{\alpha-\beta} - 3\frac{h'(\alpha)}{h(\alpha)}\right](z-\alpha)^{-1}\right\}\begin{pmatrix} -1 & -i \\ -i & 1 \end{pmatrix}\\
&+\frac{7}{48(\alpha-\beta)h(\alpha)}(z-\alpha)^{-1}\begin{pmatrix} -1 & i \\ i & 1 \end{pmatrix}\\
&+\frac{5}{72}\left\{\frac{3}{2h(\beta)}(z-\beta)^{-2} + \frac{3}{10h(\beta)}\left[\frac{1}{\beta-\alpha} - \frac{3h'(\beta)}{h(\beta)}\right](z-\beta)^{-1}\right.\\
&\qquad\left. - \frac{(z-\alpha)^{1/2}}{(z-\beta)^{1/2}\int_\beta^z R(s)h(s)ds}\right\}\begin{pmatrix} -1 & i \\ i & 1 \end{pmatrix}\\
&+\frac{7}{72}\left\{\frac{3}{2(\beta-\alpha)h(\beta)}(z-\beta)^{-1} - \frac{(z-\beta)^{1/2}}{(z-\alpha)^{1/2}\int_\beta^z R(s)h(s)ds}\right\}\begin{pmatrix} 1 & i \\ i & -1 \end{pmatrix}.
\end{aligned} \tag{6.2}
$$

3. For $z \in B_\delta^\alpha$,

$$
\begin{aligned}
S_1(z) = &\\
&\frac{5}{144h(\beta)}\left\{3(z-\beta)^{-2} + \frac{3}{5}\left[\frac{1}{\beta-\alpha} - 3\frac{h'(\beta)}{h(\beta)}\right](z-\beta)^{-1}\right\}\begin{pmatrix} -1 & i \\ i & 1 \end{pmatrix}\\
&+\frac{7}{48(\beta-\alpha)h(\beta)}(z-\beta)^{-1}\begin{pmatrix} 1 & i \\ i & -1 \end{pmatrix}\\
&+\frac{5}{72}\left\{\frac{3}{2h(\alpha)}(z-\alpha)^{-2} + \frac{3}{10h(\alpha)}\left[\frac{1}{\alpha-\beta} - \frac{3h'(\alpha)}{h(\alpha)}\right](z-\alpha)^{-1}\right.\\
&\qquad\left. - \frac{(z-\beta)^{1/2}}{(z-\alpha)^{1/2}\int_\alpha^z R(s)h(s)ds}\right\}\begin{pmatrix} -1 & -i \\ -i & 1 \end{pmatrix}\\
&+\frac{7}{72}\left\{\frac{3}{2(\alpha-\beta)h(\alpha)}(z-\alpha)^{-1} - \frac{(z-\alpha)^{1/2}}{(z-\beta)^{1/2}\int_\alpha^z R(s)h(s)ds}\right\}\begin{pmatrix} 1 & -i \\ -i & -1 \end{pmatrix}.
\end{aligned} \tag{6.3}
$$

For comparative purposes, if $t_j = 0, j \neq 4$, and if we set $t_4 = t$, then we have

1. for $z \in \mathbb{C} \setminus (B_\delta^\alpha \cup B_\delta^\beta)$

$$
S_1(z) = \tag{6.4}
$$

$$
\left(\frac{-7\beta}{96(8-\beta^2)(\beta-z)} + \frac{\beta(-56z-136\beta+15\beta^2 z+25\beta^3)}{96(8-\beta^2)^2(z+\beta)^2}\right)\begin{pmatrix} 1 & i \\ i & -1 \end{pmatrix} \tag{6.5}
$$

$$
+\left(\frac{7\beta}{96(8-\beta^2)(z+\beta)} - \frac{\beta(-136\beta+56z+25\beta^3-15\beta^2 z)}{96(8-\beta^2)^2(\beta-z)^2}\right)\begin{pmatrix} -1 & i \\ i & 1 \end{pmatrix} \tag{6.6}
$$

2. For $z \in B_\delta^\beta$,

$$S_1(z) = \left(\frac{-7\beta}{96(8-\beta^2)(z-\beta)} + \right. \tag{6.7}$$

$$\left. \frac{\beta(-56z - 136\beta + 15\beta^2 z + 25\beta^3)}{96(8-\beta^2)^2(z+\beta)^2} - \frac{7\gamma^2}{72\int_\beta^z R(s)h(s)ds} \right) \begin{pmatrix} 1 & i \\ i & -1 \end{pmatrix} + \tag{6.8}$$

$$+ \left(\frac{7\beta}{96(8-\beta^2)(z+\beta)} + \right. \tag{6.9}$$

$$\left. - \frac{\beta(-136\beta + 56\beta z + 25\beta^3 - 15\beta^2 z)}{96(8-\beta^2)^2(\beta - z)^2} - \frac{5}{72\gamma^2 \int_\beta^z R(s)h(s)ds} \right) \begin{pmatrix} -1 & i \\ i & 1 \end{pmatrix} \tag{6.10}$$

3. For $z \in B_\delta^{(-\beta)}$, we have

$$S_1(z) = \left(\frac{-\beta(-136\beta + 56z + 25\beta^3 - 15\beta^2 z)}{96(8-\beta^2)^2(\beta - z)^2} \right. \tag{6.11}$$

$$\left. - \frac{7}{96} \left[\frac{4\gamma^{-2}}{3\int_{-\beta}^z R(s)h(s)ds} + \frac{\beta(z+\beta)^{-1}}{(8-\beta^2)} \right] \right) \begin{pmatrix} -1 & i \\ i & 1 \end{pmatrix}$$

$$+ \left(\frac{\beta \left(25\,\beta^3 - 136\,\beta + 15\,\beta^2 z - 56\,z\right)}{96\left(\beta^2 - 8\right)^2 \left(z+\beta\right)^2} + \frac{\gamma^2}{72\int_{-\beta}^z R(s)h(s)ds} \right.$$

$$\left. - \frac{7\beta}{96} \left[\frac{(z-\beta)^{-1}}{8-\beta^2} \right] \right) \begin{pmatrix} -1 & -i \\ -i & 1 \end{pmatrix}$$

7. Conclusions

This paper provides the first mathematical derivation of the form of the asymptotic expansion of the Hermitian random matrix partition function, conjectured in [3]. This uses the connection to non-classical orthogonal polynomials, together with the Deift-Zhou steepest descent and stationary phase method for the asymptotic analysis of Riemann-Hilbert problems. In particular, the asymptotic analysis of (1.1) presented here demonstrates that

- $\log \left(\hat{Z}_N \right)$ has an asymptotic expansion of the form of (1.5);
- For arbitrary integers $g \geq 0$, $e_g(t_1, \ldots, t_\nu)$ is an analytic function of the (complex) vector $\mathbf{t} := (t_1, \ldots, t_\nu)$, in a neighborhood of $(0, \ldots, 0)$;
- For times, $\mathbf{t}$, within the domains of holomorphy for the e_g, the limiting mean density $\rho_N^{(1)}$ of eigenvalues for the Hermitian random matrix ensemble is supported on a single interval (α, β). The computed asymptotic expansion of $\rho_N^{(1)}$ is valid for all $\lambda \in (\alpha, \beta)$ and in fact depends only on the equilibrium measure;
- the coefficients of the e_g are generating functions for graphical enumeration in the sense that

$$(-1)^n \frac{\partial^n}{\partial t_k^n} e_g(\mathbf{0}) = \#\{\text{k-valent, n-vertex, g-maps}\}.$$

As a consequence of these results one is now situated to be able to combine asymptotic analytical techniques with the results of many ingenious ideas, contributions, and calculations which have arisen over the past 20 years from the theory of 2D quantum gravity [13], to investigate the connection between the asymptotic expansion of the partition function and solutions of integrable systems such as the KdV hierarchy, the Toda lattice, and singular limits of these equations described by modulation equations [25, 15, 9].

References

1. M. Abramowitz and I. Stegun. Handbook of Mathematical Functions, U.S. Govt. Printing Office, Washington, D.C., 1972.
2. S. Albeverio, L. Pastur, and M. Shcherbina. On the $1/n$ expansion for some unitary invariant ensembles of random matrices. *Comm. Math. Phys.* **224** (2001), 271-305.
3. D. Bessis, X. Itzykson, and J.B. Zuber. Quantum Field Theory Techniques in Graphical Enumeration. Adv. Appl. Math. 1 (1980) 109-157.

4. P. Bleher and A. Its. Semiclassical asymptotics of orthogonal polynomials, Riemann–Hilbert problem, and the universality in the matrix model. *Ann. Math.* **50** (1) (1999) 185-266.
5. P. Deift, T. Kriecherbauer and K. T-R McLaughlin. New results on the equilibrium measure for logarithmic potentials in the presence of an external field. *J. Approx. Thry.*, —bf 95 (1998), 388-475.
6. P. Deift, T. Kriecherbauer, K.T.-R. McLaughlin, S. Venakides, and X. Zhou. Asymptotics for polynomials orthogonal with respect to varying exponential weights. *Int. Math. Res. Notices* **16** (1997) 759-782.
7. P. Deift, T. Kriecherbauer, K.T.-R. McLaughlin, S. Venakides, and X. Zhou. Strong asymptotics of orthogonal polynomials with respect to exponential weights via Riemann–Hilbert techniques. *Commun. Pure Appl. Math.* **52** (12) (1999) 1491–1552.
8. P. Deift, T. Kriecherbauer, K.T.-R. McLaughlin, S. Venakides, and X. Zhou. Uniform asymptotics for polynomials orthogonal with respect to varying exponential weights and applications to universality questions in random matrix theory. *Commun. Pure Appl. Math.* **52** (11) (1999) 1335–1425.
9. P. Deift and K. T-R McLaughlin. A Continuum Limit of the Toda Lattice. *Memoirs of the AMS*, **131**, no. 624, January 1998.
10. P. Deift, S. Venakides, and X. Zhou. New results in small dispersion KdV by an extension of the steepest descent method for Riemann–Hilbert problems. *Internat. Math. Research Notices* **6**, 286–299 (1997).
11. P. Deift and X. Zhou. A Steepest descent method for oscillatory Riemann-Hilbert problems, asymptotics for the mKDV equation, *Ann. Math.* **137** (1993), 295-368.
12. P. Deift and X. Zhou. Asymptotics for the Painleve II equation. *Comm. Pure Appl. Math.* **48** (1995), no. 3, 277–337.
13. P. Di Francesco, P. Ginsparg and J. Zinn-Justin. 2D gravity and random matrices. *Physics Reports* **254** (1995) 1-133.
14. Gaudin and M. L. Mehta. M. L. Mehta and M. Gaudin. On the density of eigenvalues of a random matrix. *Nuclear Phys.* **18** (1960) 420–427.
15. T. Grava and F. R. Tian. The generation, propagation, and extinction of multiphases in the KdV zero dispersion limit. Preprint (2002).
16. F. Harary and E.M. Palmer. Graphical Enumeration, Academic Press, New York, (1973).
17. J. Harris and I. Morrison. Moduli of Curves, Springer Graduate Texts in Mathematics, New York, vol. 187, (1998).
18. J. Harer and D. Zagier. The Euler characteristic of the moduli space of curves. *Invent. Math.* **85** (1986), no. 3, 457–485.
19. A. Kuijlaars and K. T-R McLaughlin. Generic behavior of the density of states in random matrix theory and equilibrium problems in the presence of real analytic external fields. *Comm. Pure and Appl. Math.* Vol. LIII (2000), 736-785.
20. A. Magnus. On Freud's equations for exponential weights. *J. Approx. Theory*, **46**, 65-99, (1986).
21. M.L. Mehta. Random Matrices, 2nd Edition, Academic Press, San Diego, CA, 1991.
22. E. B. Saff and V. Totik. Logarithmic Potentials with External Fields. New York: Springer-Verlag, 1997.
23. B. Simon, The $P(\phi)_2$ Euclidean (Quantum) Field Theory, Princeton Series in Physics, Princeton University Press (1974).
24. G. Szegő, Orthogonal Polynomials, AMS Colloquium Publications, Vol. 23, 1939.
25. E. Witten. Two-Dimensional Gravity and Intersection Theory on Moduli Space. *Surveys in Differential Geometry*, vol 1, (1991), 243-310.

Dept. of Math, Univ. of Arizona
E-mail address: ercolani@math.arizona.edu

Dept. of Math., Univ. of Arizona and Dept. of Math., Univ. of North Carolina
E-mail address: mcl@amath.unc.edu