Lectures on Random Matrix Theory for Course at SAMSI

Peter D. Miller

November 16 and 30, 2006*

1 Unitary Ensembles

1.1 Law on matrix entries.

Consider the set of $N \times N$ Hermitian matrices $\mathbf{M}$ ($\mathbf{M} = \mathbf{M}^\dagger =$ conjugate transpose of $\mathbf{M}$) equipped with a probability measure

$$dP(\mathbf{M}) := \frac{1}{Z_N} e^{-N\mathrm{Tr}(V(\mathbf{M}))} d\mathbf{M}\,, \tag{1}$$

where Z_N is a normalization constant (partition function) and $d\mathbf{M}$ denotes Lebesgue measure on the algebraically independent real components of $\mathbf{M}$: $\Re\{M_{ij}\}$ and $\Im\{M_{ij}\}$ for $i > j$ and $M_{ii} \in \mathbb{R}$. Some notes:

1. $V(x)$ is a potential increasing sufficiently rapidly for large $|x|$ to make the measure normalizable. $V(\mathbf{M})$ is defined through the spectral theorem: for each Hermitian matrix $\mathbf{M}$ there exists a unitary matrix $\mathbf{U}$ such that $\mathbf{M} = \mathbf{U}\mathrm{diag}(x_1, \dots, x_N)\mathbf{U}^\dagger$, where $x_1 \le x_2 \le \cdots \le x_N$ are the (real) eigenvalues of $\mathbf{M}$. Then $V(\mathbf{M})$ is the matrix
$$V(\mathbf{M}) := \mathbf{U}\mathrm{diag}(V(x_1), \dots, V(x_N))\mathbf{U}^\dagger\,. \tag{2}$$
As the trace is invariant under conjugation, note that
$$\mathrm{Tr}(V(\mathbf{M})) = \mathrm{Tr}(\mathrm{diag}(V(x_1), \dots, V(x_N))) = \sum_{k=1}^{N} V(x_k)\,. \tag{3}$$

2. Measures of this form are unitarily invariant, in the sense that if $\mathbf{U}$ is a fixed $N \times N$ unitary matrix then the map $\mathbf{M} \mapsto \mathbf{U}\mathbf{M}'\mathbf{U}^\dagger$ preserves the form:
$$dP(\mathbf{M}') = \frac{1}{Z_N} e^{-N\mathrm{Tr}(V(\mathbf{M}'))} d\mathbf{M}'\,. \tag{4}$$
This is important in quantum physics, where $\mathbf{M}$ represents an observable (a self-adjoint operator) written in terms of some basis. The unitary invariance means that the basis chosen to write $\mathbf{M}$ should not be important in any physically meaningful statistical theory of observables.

3. An important special case corresponds to the choice $V(x) = x^2$. This is the Gaussian Unitary Ensemble (GUE). Only in this case are the entries of $\mathbf{M}$ statistically independent random variables.

*Notes corrected May 15, 2008

1.2 Joint law for eigenvalues. Determinantal form.

Recall the idea of writing the measure $dP(\mathbf{M})$ not in terms of the matrix entries but rather in terms of the eigenvalues and eigenvectors of $\mathbf{M}$. The Jacobian of this transformation is proportional to the square of the determinant of the Vandermonde matrix of eigenvalues. Thus, if $\mathbf{M} = \mathbf{U}\mathrm{diag}(x_1, \dots, x_N)\mathbf{U}^\dagger$, then

$$dP(\mathbf{M}) = \frac{1}{Z_N} df(\mathbf{U}) \cdot \prod_{k=1}^{N} e^{-NV(x_k)} dx_k \cdot \prod_{j<k} (x_k - x_j)^2 \,, \tag{5}$$

for some measure df on the unitary group of eigenvector matrices. Thus, the eigenvalues and eigenvectors are statistically independent. The marginal distribution of eigenvalues comes from integrating out the eigenvectors:

$$dP(x_1, \dots, x_N) = \frac{1}{Z'_N} \prod_{k=1}^{N} e^{-NV(x_k)} dx_k \cdot \prod_{j<k} (x_k - x_j)^2 = P(x_1, \dots, x_N) \prod_{k=1}^{N} dx_k \,. \tag{6}$$

This law has determinantal form. Indeed, since

$$\prod_{j<k} (x_k - x_j) = \det \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ x_1 & x_2 & x_3 & \cdots & x_N \\ x_1^2 & x_2^2 & x_3^2 & \cdots & x_N^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_1^{N-1} & x_2^{N-1} & x_3^{N-1} & \cdots & x_N^{N-1} \end{bmatrix} = \det \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{N-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{N-1} \\ 1 & x_3 & x_3^2 & \cdots & x_3^{N-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_N & x_N^2 & \cdots & x_N^{N-1} \end{bmatrix} , \tag{7}$$

it is clear from simple properties of determinants that we can express P in the form

$$P(x_1, \dots, x_N) = \frac{1}{Z'_N} \det \left(\begin{bmatrix} e^{-NV(x_1)/2} & x_1 e^{-NV(x_1)/2} & \cdots & x_1^{N-1} e^{-NV(x_1)/2} \\ e^{-NV(x_2)/2} & x_2 e^{-NV(x_2)/2} & \cdots & x_2^{N-1} e^{-NV(x_2)/2} \\ \vdots & \vdots & \vdots & \vdots \\ e^{-NV(x_N)/2} & x_N e^{-NV(x_N)/2} & \cdots & x_N^{N-1} e^{-NV(x_N)/2} \end{bmatrix} \cdot \begin{bmatrix} e^{-NV(x_1)/2} & e^{-NV(x_2)/2} & \cdots & e^{-NV(x_N)/2} \\ x_1 e^{-NV(x_1)/2} & x_2 e^{-NV(x_2)/2} & \cdots & x_N e^{-NV(x_N)/2} \\ \vdots & \vdots & \vdots & \vdots \\ x_1^{N-1} e^{-NV(x_1)/2} & x_2^{N-1} e^{-NV(x_2)/2} & \cdots & x_N^{N-1} e^{-NV(x_N)/2} \end{bmatrix} \right) . \tag{8}$$

For that matter, we could also note that the determinant is unchanged by the operation of adding to any one row (column) multiples of the other rows (columns), so this formula can also be written as

$$P(x_1, \dots, x_N) = \frac{1}{Z''_N} \det \left(\begin{bmatrix} q_0(x_1) e^{-NV(x_1)/2} & q_1(x_1) e^{-NV(x_1)/2} & \cdots & q_{N-1}(x_1) e^{-NV(x_1)/2} \\ q_0(x_2) e^{-NV(x_2)/2} & q_1(x_2) e^{-NV(x_2)/2} & \cdots & q_{N-1}(x_2) e^{-NV(x_2)/2} \\ \vdots & \vdots & \vdots & \vdots \\ q_0(x_N) e^{-NV(x_N)/2} & q_1(x_N) e^{-NV(x_N)/2} & \cdots & q_{N-1}(x_N) e^{-NV(x_N)/2} \end{bmatrix} \cdot \begin{bmatrix} q_0(x_1) e^{-NV(x_1)/2} & q_0(x_2) e^{-NV(x_2)/2} & \cdots & q_0(x_N) e^{-NV(x_N)/2} \\ q_1(x_1) e^{-NV(x_1)/2} & q_1(x_2) e^{-NV(x_2)/2} & \cdots & q_1(x_N) e^{-NV(x_N)/2} \\ \vdots & \vdots & \vdots & \vdots \\ q_{N-1}(x_1) e^{-NV(x_1)/2} & q_{N-1}(x_2) e^{-NV(x_2)/2} & \cdots & q_{N-1}(x_N) e^{-NV(x_N)/2} \end{bmatrix} \right) . \tag{9}$$

where $q_k(x) = q_{k,k} x^k + \cdots$ is an arbitrary polynomial of exact degree k ($q_{k,k} \neq 0$ for all $k = 0, 1, 2, \dots$), and

$$Z''_N = Z'_N \prod_{k=0}^{N-1} q_{k,k}^2 \,. \tag{10}$$

Carrying out the matrix multiplication we get

$$P(x_1,\ldots,x_N)=\frac{1}{Z_N''}\det\begin{bmatrix} K_N(x_1,x_1) & K_N(x_1,x_2) & K_N(x_1,x_3) & \cdots & K_N(x_1,x_N)\\ K_N(x_2,x_1) & K_N(x_2,x_2) & K_N(x_2,x_3) & \cdots & K_N(x_2,x_N)\\ K_N(x_3,x_1) & K_N(x_3,x_2) & K_N(x_3,x_3) & \cdots & K_N(x_3,x_N)\\ \vdots & \vdots & \vdots & \vdots & \vdots\\ K_N(x_N,x_1) & K_N(x_N,x_2) & K_N(x_N,x_3) & \cdots & K_N(x_N,x_N)\end{bmatrix} \tag{11}$$

where

$$K_N(x,y):=e^{-NV(x)/2}e^{-NV(y)/2}\sum_{k=0}^{N-1}q_k(x)q_k(y)\,. \tag{12}$$

So $P(x_1,\ldots,x_N)$ is just a determinant involving polynomials $q_k(x)$ and a *weight* $w(x):=e^{-NV(x)}$. But the arbitrariness of the polynomials $q_k(x)$ should be an advantage, and a special choice may simplify calculations. Indeed, if we pick for $q_k(x)$ the orthonormal polynomials with respect to the weight $w(x)=e^{-NV(x)}$, *i.e.* the polynomials $q_k(x)=p_k(x)=p_{k,k}x^k+\cdots$ with $p_{k,k}>0$ such that

$$\int_{-\infty}^{+\infty}p_j(x)p_k(x)e^{-NV(x)}\,dx=\delta_{jk}\,,\qquad j,k=0,1,\ldots\,, \tag{13}$$

then the function $K_N(x,y)$ is a *reproducing kernel*:

$$p(x)e^{-NV(x)/2}=\int_{-\infty}^{+\infty}K_N(x,y)\left[p(y)e^{-NV(y)/2}\right]\,dy \tag{14}$$

holds for all polynomials $p(x)$ of degree $N-1$ (or less). Since $K_N(x,y)$ has the form of $p(x)e^{-NV(x)/2}$ for each y, we can put $p(y)e^{-NV(y)/2}=K_N(y,z)$ to get

$$K_N(x,z)=\int_{-\infty}^{+\infty}K_N(x,y)K_N(y,z)\,dy\,. \tag{15}$$

Some notes:

1. The orthogonality conditions uniquely determine the polynomials $p_k(x)$ as long as $V(x)$ grows fast enough as $|x|\to\infty$. A procedure for determining them is Gram-Schmidt orthogonalization.

2. Other notation: the monic orthogonal polynomials are denoted $\pi_k(x)$ and are simply given by

$$\pi_k(x):=\frac{1}{p_{k,k}}p_k(x)=x^k+\cdots\,. \tag{16}$$

3. The orthonormality conditions are equivalent to the conditions that

$$\int_{-\infty}^{+\infty}\pi_k(x)x^je^{-NV(x)}\,dx=0\,,\qquad 0\le j\le k-1 \tag{17}$$

and

$$\int_{-\infty}^{+\infty}\pi_k(x)x^ke^{-NV(x)}\,dx=\frac{1}{p_{k,k}^2}\,. \tag{18}$$

From now on we assume that the polynomials used to present the joint law $P(x_1,\ldots,x_N)$ in determinantal form are not arbitrary but are rather the orthogonal polynomials $q_k(x)=p_k(x)$ associated with the weight $w(x)=e^{-NV(x)}$. The identity (15) is then at our disposal and will have implications in simplifying formulae for correlation functions as we will now see.

1.3 Correlation functions. Definition and determinantal point processes.

The n-point correlation function is a kind of marginal distribution on a group of $n \leq N$ eigenvalues, and is defined by the formula

$$R_N^{(n)}(x_1, \ldots, x_n) := \frac{N!}{(N-n)!} \int_{\mathbb{R}(x_{n+1})} \cdots \int_{\mathbb{R}(x_N)} P(x_1, \ldots, x_N)\, dx_{n+1} \cdots dx_N \,. \tag{19}$$

The one-point function has the interpretation of N times the density of eigenvalues, in the sense that for any measurable set $I \subset \mathbb{R}$,

$$\int_I R_N^{(1)}(x)\, dx = \mathbb{E}(\text{number of eigenvalues in } I)\,. \tag{20}$$

More generally, the n-point function $R_N^{(n)}(x_1, \ldots, x_n)$ is the probability density for finding an eigenvalue near each of the points $x_1, \ldots, x_n$.

By definition, the correlation functions are multiple integrals, and for n fixed the number of dimensions over which to integrate is $N - n$ which grows with N. But, it is a consequence of the identity (15) that the correlation functions can be written alternately in a form that avoids all integration. For this we need the following "integrating out" lemma:

Proposition 1. *Let $\mathbf{J}_n$ be an $n \times n$ matrix with entries of the form $J_{ij} = f(x_i, x_j)$ for some measureable function $f : \mathbb{R}^2 \to \mathbb{C}$ with the property that*

$$\int f(x,y) f(y,z)\, d\mu(y) = f(x,z) \tag{21}$$

holds for some measure $d\mu$ on $\mathbb{R}$. Then,

$$\int \det(\mathbf{J}_n)\, d\mu(x_n) = \left[\int f(x,x)\, d\mu(x) - n + 1 \right] \det(\mathbf{J}_{n-1})\,. \tag{22}$$

Note that the matrix $\mathbf{J}_{n-1}$ has the same functional form as $\mathbf{J}_n$ but with one fewer row and column (omitting the variable x_n, which has been integrated out).

For a proof, see Deift [1] or Mehta [2]. To apply this result with $f(x,y) = K_N(x,y)$, which satisfies the hypotheses of the proposition according to (15), first we calculate

$$\int_{-\infty}^{+\infty} K_N(x,x)\, dx = \sum_{k=0}^{N-1} \int_{-\infty}^{+\infty} p_k(x)^2 e^{-NV(x)}\, dx = N\,, \tag{23}$$

and then we proceed by repeated integration, since

$$R_N^{(n-1)}(x_1, \ldots, x_{n-1}) = \frac{1}{N-n+1} \int_{-\infty}^{+\infty} R_N^{(n)}(x_1, \ldots, x_n)\, dx_n\,, \tag{24}$$

and

$$R_N^{(N)}(x_1, \ldots, x_N) = N! P(x_1, \ldots, x_N) = \frac{N!}{Z_N''} \det(\mathbf{J}_N) \tag{25}$$

where $\mathbf{J}_N$ is the $N \times N$ matrix with entries $J_{ij} = K_N(x_i, x_j)$. Since the factor $(N-n+1)$ that the proposition says to include with each integration is exactly the factor in the denominator of (24), by the proposition we have

$$R_N^{(n)}(x_1, \ldots, x_n) = \frac{N!}{Z_N''} \det(\mathbf{J}_n) = \frac{N!}{Z_N''} \det \begin{bmatrix} K_N(x_1,x_1) & K_N(x_1,x_2) & \cdots & K_N(x_1,x_n) \\ K_N(x_2,x_1) & K_N(x_2,x_2) & \cdots & K_N(x_2,x_n) \\ \vdots & \vdots & \vdots & \vdots \\ K_N(x_n,x_1) & K_N(x_n,x_2) & \cdots & K_N(x_n,x_n) \end{bmatrix}. \tag{26}$$

In particular, this shows that the scaled density of eigenvalues (the one-point function) is given by

$$R_N^{(1)}(x) = \frac{N!}{Z_N''} K_N(x,x)\,, \tag{27}$$

and since

$$N = \mathbb{E}(\text{number of eigenvalues in } \mathbb{R}) = \int_{-\infty}^{+\infty} R_N^{(1)}(x)\,dx = \frac{N!}{Z_N''}\int_{-\infty}^{+\infty} K_N(x,x)\,dx = \frac{N!}{Z_N''} N\,, \tag{28}$$

we have shown that the partition function $Z_n'' = N!$. Therefore, our formula for the correlation functions simplifies yet again:

$$R_N^{(n)}(x_1,\dots,x_n) = \det\begin{bmatrix} K_N(x_1,x_1) & K_N(x_1,x_2) & \cdots & K_N(x_1,x_n) \\ K_N(x_2,x_1) & K_N(x_2,x_2) & \cdots & K_N(x_2,x_n) \\ \vdots & \vdots & \vdots & \vdots \\ K_N(x_n,x_1) & K_N(x_n,x_2) & \cdots & K_N(x_n,x_n) \end{bmatrix}\,. \tag{29}$$

Therefore all multipoint correlation functions for the random variables (eigenvalues) $x_1,\dots,x_N$ have the form of simple determinants, and no integration is required to evaluate them once $K_N(x,y)$ is known. Correlation functions of this form are said to correspond to a *determinantal point process.*

Some notes:

1. Since $Z_N'' = N!$, with the use of (10), the "original" partition function of the joint law for eigenvalues is
$$Z_N' = N! \prod_{k=0}^{N-1} p_{k,k}^{-2}\,, \tag{30}$$
where $p_{k,k} > 0$ is the leading coefficient of the orthonormal polynomial $p_k(x)$.

2. More precise statistics can also be represented in terms of $K_N(x,y)$. For example, the probability that given measurable set I contains exactly k eigenvalues can be computed from the correlation functions by an inclusion/exclusion principle. The result is that such probabilities are expressed in terms of *Fredholm determinants* of operators on $L^2(I)$ of the form $I - t\mathcal{K}_N$ where $\mathcal{K}_N$ is the integral operator on I with kernel $K_N(x,y)$, and t is a "generating function" parameter. See Deift [1] or Mehta [2].

2 Asymptotic Behavior of Correlation Functions

We are interested in the way the statistics (correlation functions) behave as the matrices become larger and larger. Clearly this boils down to the study of the kernel $K_N(x,y)$ in the limit $N \to \infty$.

2.1 Christoffel-Darboux formula.

By its definition, the reproducing kernel $K_N(x,y)$ contains "more and more" information the larger N is, since it has the form of a partial sum of an infinite series. A fantastically useful result of the general theory of orthogonal polynomials (see Szegő [3]), is that, in a certain sense, the partial sum telescopes.

Proposition 2 (Christoffel-Darboux formula)**.** *Let $p_k(x)$ be the orthonormal polynomials with respect to any weight $w(x)\,dx$ on $\mathbb{R}$. Then, for any $n \geq 0$, and real x and y with $x \neq y$,*

$$\sum_{k=0}^{n-1} p_k(x)p_k(y) = \frac{p_{n-1,n-1}}{p_{n,n}} \cdot \frac{p_n(x)p_{n-1}(y) - p_{n-1}(x)p_n(y)}{x-y}\,. \tag{31}$$

Moreover, the same formula holds for $x = y$ by interpreting the right-hand side with l'Hôpital's rule:

$$\sum_{k=0}^{n-1} p_k(x)^2 = \frac{p_{n-1,n-1}}{p_{n,n}} \left[p_n'(x)p_{n-1}(x) - p_{n-1}'(x)p_n(x) \right] . \tag{32}$$

Using this result, we can write the reproducing kernel $K_N(x,y)$ as

$$K_N(x,y) = e^{-NV(x)/2} e^{-NV(y)/2} \cdot \frac{p_{N-1,N-1}}{p_{N,N}} \cdot \frac{p_N(x)p_{N-1}(y) - p_{N-1}(x)p_N(y)}{x-y} , \qquad x \neq y , \tag{33}$$

or

$$K_N(x,x) = e^{-NV(x)} \frac{p_{N-1,N-1}}{p_{N,N}} \left[p_N'(x)p_{N-1}(x) - p_{N-1}'(x)p_N(x) \right] . \tag{34}$$

Therefore, to study the large-N behavior of $K_N(x,y)$ it is enough to study the behavior of two polynomials.

2.2 Example: the Gaussian unitary ensemble.

2.2.1 Hermite polynomials.

Here we do all of this with $V(x) = x^2$. The orthogonal polynomials in this case are the classical Hermite polynomials. One can lift all the results needed directly from Szegő's book [3], but to do so one needs the translation of the orthonormal polynomials with respect to e^{-Nx^2} which we call $p_k(x)$ into Szegő's notation. Szegő considers polynomials $H_k(x)$ with positive leading coefficient that satisfy

$$\int_{-\infty}^{\infty} H_j(x)H_k(x)e^{-x^2}\, dx = \sqrt{\pi}2^k k! \delta_{jk} , \tag{35}$$

and all of his results are written for $H_k(x)$. By a simple scaling argument,

$$p_k(x) = \left(\frac{N}{\pi}\right)^{1/4} \frac{1}{\sqrt{2^k k!}} H_k(\sqrt{N}x) . \tag{36}$$

The leading coefficient of $H_k(x)$ is 2^k, from which it follows that

$$p_{k,k} = \left(\frac{N}{\pi}\right)^{1/4} \sqrt{\frac{(2N)^k}{k!}} . \tag{37}$$

The Hermite polynomials have many, many special properties that make them easy to analyze by classical techniques. For example, (see §5.5 of [3]) $p_k(x)$ satisfies a second-order differential equation:

$$\frac{1}{N}\frac{d^2 p_k}{dx^2} - 2x\frac{dp_k}{dx} + 2kp_k = 0 . \tag{38}$$

The WKB method can be applied to study this differential equation when N and k are large. There is also a simple and explicit generating function for $p_k(x)$:

$$p_0(x) + p_1(x)w + \frac{1}{\sqrt{2!}}p_2(x)w^2 + \frac{1}{\sqrt{3!}}p_3(x)w^3 + \cdots = \left(\frac{N}{\pi}\right)^{1/4} e^{(2N)^{1/2}xw - w^2/2} . \tag{39}$$

More than being a formal power series relation, the left-hand side of this expression is a uniformly convergent series in any compact set of $\mathbb{C}^2$. Multiply through by $w^{-k-1}/(2\pi i)$ and integrate on a circle in the complex w-plane surrounding $w = 0$. By the Residue Theorem, only one term on the left-hand side survives:

$$\frac{1}{\sqrt{k!}}p_k(x) = \left(\frac{N}{\pi}\right)^{1/4} \frac{1}{2\pi i} \oint w^{-k-1} e^{(2N)^{1/2}xw - w^2/2}\, dw . \tag{40}$$

A complete analysis of $K_N(x,y)$ can be based upon this formula for $k=N$ and $k=N-1$. Rescaling the integration variable by setting $w=(2N)^{1/2}z$ and defining the analytic exponent function

$$h(z;x) := -\log(z) + 2xz - z^2\,, \tag{41}$$

we thus find

$$p_{N-1}(x) = \left(\frac{N}{\pi}\right)^{1/4}\sqrt{\frac{(N-1)!}{(2N)^{N-1}}}\frac{1}{2\pi i}\oint e^{Nh(z;x)}\,dz\,, \tag{42}$$

and

$$p_N(x) = \left(\frac{N}{\pi}\right)^{1/4}\sqrt{\frac{N!}{(2N)^{N}}}\frac{1}{2\pi i}\oint \frac{1}{z}e^{Nh(z;x)}\,dz\,. \tag{43}$$

These formulae may be differentiated with respect to x under the integral sign, resulting in integral representations for the derivatives as well:

$$p'_{N-1}(x) = \left(\frac{N}{\pi}\right)^{1/4}\sqrt{\frac{(N-1)!}{(2N)^{N-1}}}\frac{N}{\pi i}\oint ze^{Nh(z;x)}\,dz\,, \tag{44}$$

and

$$p'_N(x) = \left(\frac{N}{\pi}\right)^{1/4}\sqrt{\frac{N!}{(2N)^{N}}}\frac{N}{\pi i}\oint e^{Nh(z;x)}\,dz\,. \tag{45}$$

Therefore,

$$K_N(x,x) = \frac{e^{-Nx^2}(N-1)!}{2^{N+1}\pi^{5/2}N^{N-5/2}}\left[\oint ze^{Nh(z;x)}\,dz\oint\frac{1}{z'}e^{Nh(z';x)}\,dz' - \left(\oint e^{Nh(z;x)}\,dz\right)^2\right]\,. \tag{46}$$

and

$$K_N(x,y) = \frac{e^{-Nx^2/2}e^{-Ny^2/2}(N-1)!}{2^{N+2}\pi^{5/2}N^{N-3/2}(x-y)}\left[\oint\frac{1}{z}e^{Nh(z;y)}\,dz\oint e^{Nh(z';x)}\,dz' - \oint\frac{1}{z}e^{Nh(z;x)}\,dz\oint e^{Nh(z';y)}\,dz'\right]\,. \tag{47}$$

The way to analyze these formulae in the limit $N\to\infty$ is to use the method of steepest descent. This kind of calculation was first carried out for the Hermite polynomials by Plancherel and Rotach. More generally, one refers to the asymptotics of $p_k(x)$ orthogonal with respect to a weight $e^{-NV(x)}$ in the limit $k,N\to\infty$ with $k/N\to c$ and x fixed as *Plancherel-Rotach asymptotics.*

2.2.2 Stirling's formula.

Stirling's formula tells us how $(N-1)!$ behaves as $N\to+\infty$:

$$(N-1)! = \sqrt{2\pi}e^{-N}N^{N-1/2}(1+O(N^{-1}))\,. \tag{48}$$

Using this formula in our expressions for $K_N(x,x)$ and $K_N(x,y)$ gives

$$K_N(x,x) = e^{-N(x^2+1+\log(2))}\frac{N^2}{\pi^2\sqrt{2}}\left[\oint ze^{Nh(z;x)}\,dz\oint\frac{1}{z'}e^{Nh(z';x)}\,dz' - \left(\oint e^{Nh(z;x)}\,dz\right)^2\right](1+O(N^{-1}))\,, \tag{49}$$

and

$$K_N(x,y) = e^{-N(x^2/2+y^2/2+1+\log(2))}\frac{N}{2\pi^2\sqrt{2}(x-y)} \cdot\left[\oint\frac{1}{z}e^{Nh(z;y)}\,dz\oint e^{Nh(z';x)}\,dz' - \oint\frac{1}{z}e^{Nh(z;x)}\,dz\oint e^{Nh(z';y)}\,dz'\right](1+O(N^{-1}))\,. \tag{50}$$

2.2.3 Steepest descent analysis.

For an introduction in general context, see, *e.g.* [4]. The key result we need is the following.

Proposition 3. *Consider a contour integral of the form*

$$I := \int_C g(z) e^{Nh(z)}\, dz \tag{51}$$

where $g(z)$ and $h(z)$ are analytic functions and C is a smooth contour in the complex plane along which $\Im\{h(z)\} = \Im\{h(z_0)\}$ for some point $z_0 \in C$ which also corresponds to a local maximum for $\Re\{h(z)\}$ along C. (This implies that z_0 is a critical point of $h(z)$, that is, $h'(z_0) = 0$.) Then as long as $h''(z_0) \neq 0$,

$$I = e^{i\theta} \sqrt{\frac{2\pi}{N|h''(z_0)|}}\, g(z_0) e^{Nh(z_0)} (1 + O(N^{-1}))\,, \tag{52}$$

as $N \to +\infty$, where $\theta = -\arg(h''(z_0))/2 \pm \pi/2$ is the angle with which C traverses the point z_0.

The exponent function $h(z;x)$ has two critical points in the complex plane:

$$z = \frac{x}{2} \pm \frac{1}{2}\sqrt{x^2 - 2}\,, \tag{53}$$

and therefore we have two real critical points if $|x| > \sqrt{2}$ and a complex-conjugate pair of critical points if $|x| < \sqrt{2}$. We consider $x > 0$ ($x < 0$ is similar).

Outside the bulk: $x > \sqrt{2}$. Both critical points are real and positive. We have

$$h''((x + \sqrt{x^2 - 2})/2; x) < 0\,, \qquad h''((x - \sqrt{x^2 - 2})/2; x) > 0\,, \tag{54}$$

Figure 1 shows pictures of the complex z-plane illustrating the curves where $\Im\{h(z;x)\}$ is constant passing through the two critical points. Based on these figures, we see that the right thing to do is to deform the

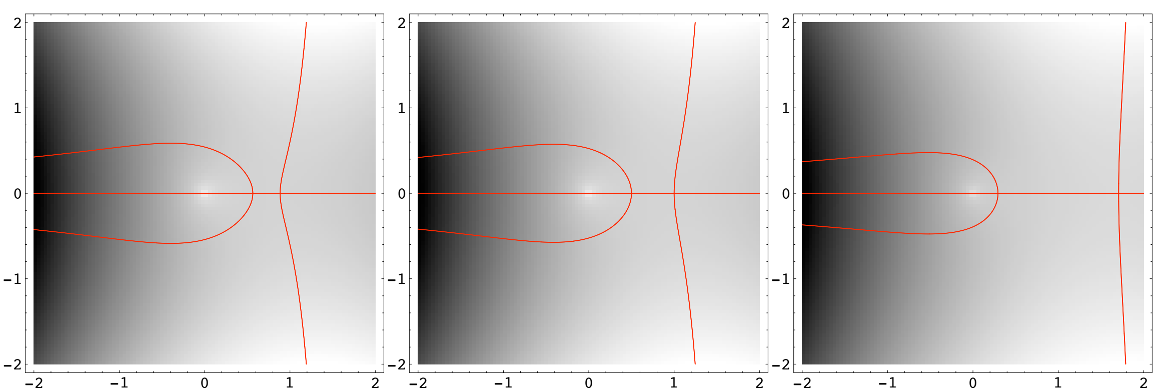

Figure 1: *The complex z-plane for $x = 1.45$, $x = 1.5$, and $x = 2$ (left to right). The lighter parts of the plot correspond to larger $\Re\{h(z;x)\}$, and the red curves are the levels of $\Im\{h(z;x)\}$ passing through the two real critical points.*

path of integration to a constant $\Im\{h(z;x)\}$ curve passing vertically over the leftmost critical point. It is an exercise to show that the one-point function $R_N^{(1)}(x) = K_N(x,x)$ is exponentially small as $N \to +\infty$ whenever x is fixed outside the bulk.

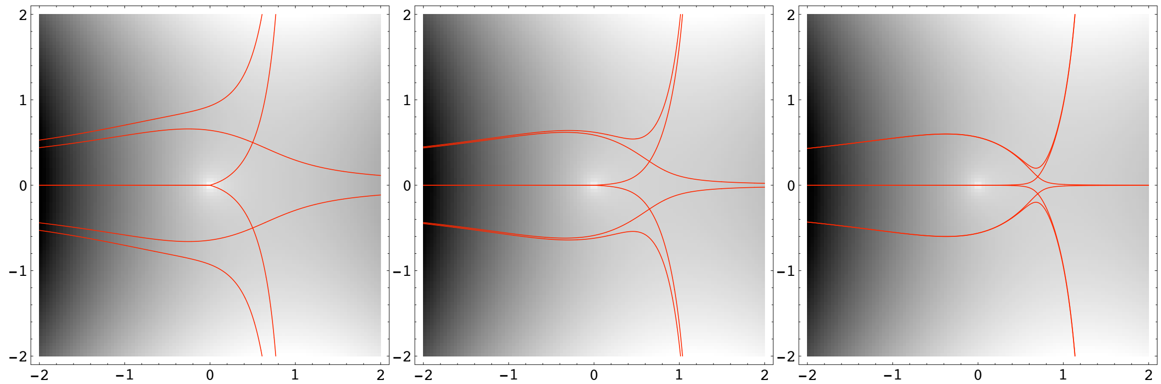

Figure 2: *The complex z-plane for $x = 1$, $x = 1.3$, and $x = 1.4$ (left to right). The lighter parts of the plot correspond to larger $\Re\{h(z;x)\}$, and the red curves are the levels of $\Im\{h(z;x)\}$ passing through the two real critical points.*

Inside the bulk: $0 < x < \sqrt{2}$**.** In this case the critical points form a complex conjugate pair, and figures showing the contours of $\Im\{h(z;x)\}$ passing through these two points are presented in Figure 2. These figures show that we should deform the circular path into two disjoint components, one in the upper and one in the lower half-plane, each passing from valley to valley over a saddle point. Thus the integral will asymptotically feel two contributions (of equal magnitude as $\Re\{h(z;x)\}$ is the same at the two critical points). That is, we write

$$\oint z^p e^{Nh(z;x)}\,dz = \int_C z^p e^{Nh(z;x)}\,dz - \int_{C^*} z^p e^{Nh(z;x)} = 2i\Im\left\{\int_C z^p e^{Nh(z;x)}\,dz\right\} \tag{55}$$

where C is the contour of constant $\Im\{h(z;x)\}$ passing from $-\infty$ to $+\infty$ over the critical point

$$z_-(x) := \frac{1}{2}\left(x - i\sqrt{2-x^2}\right) \tag{56}$$

in the lower half-plane as shown in Figure 2. Note that

$$h''(z_-(x);x) = -2(2-x^2) + 2ix\sqrt{2-x^2} = \sqrt{8(2-x^2)}e^{i\alpha(x)} \tag{57}$$

for some angle $\alpha(x)$ in the second quadrant (for $x > 0$). The angle $\theta(x)$ with which the contour C passes over the critical point $z_-(x)$ is then

$$\theta(x) := -\frac{1}{2}\alpha(x) + \frac{\pi}{2}\,, \tag{58}$$

which lies in the first quadrant. Also, define a phase angle $\phi(x)$ by writing

$$\begin{aligned} h(z_-(x);x) &= \Re\{h(z_-(x);x)\} + i\phi(x) \\ &= -\log|z_-(x)| + 2x\Re\{z_-(x)\} - \Re\{z_-(x)^2\} + i\phi(x) \\ &= \frac{1}{2}(x^2 + 1 + \log(2)) + i\phi(x)\,, \end{aligned} \tag{59}$$

and finally, note that

$$\frac{1}{z_-(x)} = 2z_-(x)^* = x + i\sqrt{2-x^2}\,. \tag{60}$$

Applying Proposition 3, we then find that

$$\begin{aligned}\oint \frac{1}{z}e^{Nh(z;x)}\,dz &= 2i\Im\left\{e^{i\theta(x)}\sqrt{\frac{2\pi}{N\sqrt{8(2-x^2)}}}\cdot(x+i\sqrt{2-x^2})\cdot e^{N(x^2+1+\log(2))/2}e^{iN\phi(x)}(1+O(N^{-1}))\right\}\\ &= 2i\sqrt{\frac{2\pi}{N\sqrt{8(2-x^2)}}}e^{N(x^2+1+\log(2))/2}\\ &\quad\cdot\left[x\sin(N\phi(x)+\theta(x))+\sqrt{2-x^2}\cos(N\phi(x)+\theta(x))+O(N^{-1})\right]\,,\end{aligned}\tag{61}$$

$$\begin{aligned}\oint ze^{Nh(z;x)}\,dz &= 2i\Im\left\{e^{i\theta(x)}\sqrt{\frac{2\pi}{N\sqrt{8(2-x^2)}}}\cdot\left(\frac{x}{2}-\frac{i}{2}\sqrt{2-x^2}\right)\cdot e^{N(x^2+1+\log(2))/2}e^{iN\phi(x)}(1+O(N^{-1}))\right\}\\ &= i\sqrt{\frac{2\pi}{N\sqrt{8(2-x^2)}}}e^{N(x^2+1+\log(2))/2}\\ &\quad\cdot\left[x\sin(N\phi(x)+\theta(x))-\sqrt{2-x^2}\cos(N\phi(x)+\theta(x))+O(N^{-1})\right]\,,\end{aligned}\tag{62}$$

and

$$\begin{aligned}\oint e^{Nh(z;x)}\,dz &= 2i\Im\left\{e^{i\theta(x)}\sqrt{\frac{2\pi}{N\sqrt{8(2-x^2)}}}\cdot e^{N(x^2+1+\log(2))/2}e^{iN\phi(x)}(1+O(N^{-1}))\right\}\\ &= 2i\sqrt{\frac{2\pi}{N\sqrt{8(2-x^2)}}}e^{N(x^2+1+\log(2))/2}\left[\sin(N\phi(x)+\theta(x))+O(N^{-1})\right]\,.\end{aligned}\tag{63}$$

Therefore,

$$\oint\frac{1}{z}e^{Nh(z;x)}\,dz\oint z'e^{Nh(z';x)}\,dz'-\left(\oint e^{Nh(z;x)}\,dz\right)^2=\frac{\pi}{N}\sqrt{\frac{2}{2-x^2}}e^{N(x^2+1+\log(2))}\left[2-x^2+O(N^{-1})\right]\,,\tag{64}$$

and so we see that if x is a fixed point in the bulk $|x|<\sqrt{2}$, then

$$R_N^{(1)}(x)=K_N(x,x)=\frac{N}{\pi}\sqrt{2-x^2}+O(1)\,,\tag{65}$$

as $N\to+\infty$. This completes (yet another, for this course) proof of the Wigner semicircle law.

What about multipoint correlations in the bulk? The asymptotic mean spacing between eigenvalues near a point x in the bulk is (from $R_N^{(1)}(x)$)

$$\frac{1}{N}\Delta(x):=\frac{1}{N}\lim_{N\to\infty}\frac{N}{R_N^{(1)}(x)}=\frac{\pi}{N\sqrt{2-x^2}}\,,\tag{66}$$

so to zoom in on a neighborhood of x, consider the expression

$$K_N^{\mathrm{loc}}(\xi,\eta):=\frac{K_N(x+N^{-1}\Delta(x)\xi,x+N^{-1}\Delta(x)\eta)}{K_N(x,x)}=\frac{K_N(x+N^{-1}\Delta(x)\xi,x+N^{-1}\Delta(x)\eta)}{N\Delta(x)^{-1}+O(1)}\,,\tag{67}$$

a "localized" and "renormalized" version of the kernel $K_N(x,y)$. Supposing that $x\in(-\sqrt{2},+\sqrt{2})$ is fixed and ν is confined to an arbitrary bounded subset of $\mathbb{R}$, we may replace x by $x+N^{-1}\Delta(x)\nu$ in (61)–(63), which remain valid under these conditions. Simplifying (61)–(63) under this substitution with the help of the fact that (direct calculation)

$$\phi'(x) = \frac{d}{dx}\Im\{h(z_-(x);x)\} = -\sqrt{2-x^2} \tag{68}$$

so that $\phi'(x)\Delta(x) \equiv -\pi$, we obtain

$$\begin{aligned}\oint \frac{1}{z}e^{Nh(z;x+N^{-1}\Delta(x)\nu)}\,dz &= 2i\sqrt{\frac{2\pi}{N\sqrt{8(2-x^2)}}}e^{N(x^2+1+\log(2))/2}e^{x\Delta(x)\nu}\\ &\cdot\left[x\sin(N\phi(x)+\theta(x)-\pi\nu)+\sqrt{2-x^2}\cos(N\phi(x)+\theta(x)-\pi\nu)+O(N^{-1})\right]\,,\end{aligned} \tag{69}$$

and

$$\begin{aligned}\oint e^{Nh(z;x+N^{-1}\Delta(x)\nu)}\,dz &= 2i\sqrt{\frac{2\pi}{N\sqrt{8(2-x^2)}}}e^{N(x^2+1+\log(2))/2}e^{x\Delta(x)\nu}\\ &\cdot\left[\sin(N\phi(x)+\theta(x)-\pi\nu)+O(N^{-1})\right]\,.\end{aligned} \tag{70}$$

Therefore, if x is replaced by $x+N^{-1}\Delta(x)\xi$ and y is replaced by $x+N^{-1}\Delta(x)\eta$, with ξ and η both in a bounded subset of $\mathbb{R}$, then

$$\begin{aligned}\oint \frac{1}{z}e^{Nh(z;x+N^{-1}\Delta(x)\eta)}\,dz\oint e^{Nh(z';x+N^{-1}\Delta(x)\xi)}\,dz' - \oint \frac{1}{z}e^{Nh(z;x+N^{-1}\Delta(x)\xi)}\,dz\oint e^{Nh(z';x+N^{-1}\Delta(x)\eta)}\,dz'\\ = \frac{2\sqrt{2}\pi}{N}e^{N(x^2+1+\log(2))}e^{x\Delta(x)(\xi+\eta)}\left[\sin(\pi(\xi-\eta))+O\left(\frac{\xi-\eta}{N}\right)\right]\,.\end{aligned} \tag{71}$$

It follows that

$$K_N^{\mathrm{loc}}(\xi,\eta) = S(\xi,\eta)+O(N^{-1})\,, \tag{72}$$

where the *sine kernel* is defined by

$$S(\xi,\eta) := \frac{\sin(\pi(\xi-\eta))}{\pi(\xi-\eta)}\,. \tag{73}$$

From the determinantal formula for $R_N^{(n)}(x_1,\ldots,x_n)$ it then follows easily that

$$\begin{aligned}&R_N^{(n)}(x+N^{-1}\Delta(x)\xi_1,\ldots,x+N^{-1}\Delta(x)\xi_n)\\ &\quad = K_N(x,x)^n\det\begin{bmatrix}K_N^{\mathrm{loc}}(\xi_1,\xi_1) & K_N^{\mathrm{loc}}(\xi_1,\xi_2) & \cdots & K_N^{\mathrm{loc}}(\xi_1,\xi_n)\\ K_N^{\mathrm{loc}}(\xi_2,\xi_1) & K_N^{\mathrm{loc}}(\xi_2,\xi_2) & \cdots & K_N^{\mathrm{loc}}(\xi_2,\xi_n)\\ \vdots & \vdots & \vdots & \vdots\\ K_N^{\mathrm{loc}}(\xi_n,\xi_1) & K_N^{\mathrm{loc}}(\xi_n,\xi_2) & \cdots & K_N^{\mathrm{loc}}(\xi_n,\xi_n)\end{bmatrix}\\ &\quad = \left(\frac{N}{\pi}\sqrt{2-x^2}\right)^n\det\begin{bmatrix}S(\xi_1,\xi_1) & S(\xi_1,\xi_2) & \cdots & S(\xi_1,\xi_n)\\ S(\xi_2,\xi_1) & S(\xi_2,\xi_2) & \cdots & S(\xi_2,\xi_n)\\ \vdots & \vdots & \vdots & \vdots\\ S(\xi_n,\xi_1) & S(\xi_n,\xi_2) & \cdots & S(\xi_n,\xi_n)\end{bmatrix}+O(N^{n-1})\,.\end{aligned} \tag{74}$$

In this way, all multipoint correlation functions for n eigenvalues near a point x in the bulk are expressed in terms of the sine kernel as $N\to+\infty$ in the Gaussian unitary ensemble.

At the edge: $x \approx \pm\sqrt{2}$**.** In this situation there are two critical points coalescing. Near the coalescing critical points, $h(z;x)$ may be approximated by a cubic polynomial in z, and integrals with exponents of the form $zx - z^3/3$ are Airy functions of x. These arguments can be made rigorous, showing that $K_N(x,y)$ can be represented in terms of the Airy function $Ai(x)$ when x and y are near the edge of the spectrum. An appropriate Fredholm determinant involving the (appropriately rescaled and renormalized) limiting kernel, called the *Airy kernel*

$$A(\xi,\eta) := \frac{Ai(\xi)Ai'(\eta) - Ai(\eta)Ai'(\xi)}{\xi - \eta}\,, \tag{75}$$

gives the *Tracy-Widom law* for the largest eigenvalue of a random matrix from the Gaussian unitary ensemble.

2.3 Asymptotics of correlation functions for more general $V(x)$.

In the more general case, all of the specialized methods we used to analyze $p_N(x)$ and $p_{N-1}(x)$ in the Hermite case are unavailable. It is not known whether orthogonal polynomials for a general weight of the form $e^{-NV(x)}$ posess any simple contour integral representations, or whether they satisfy any simple differential equations. These must be regarded as very special features of the Hermite weight e^{-Nx^2} (and a few other so-called "classical" weights, see Szegő). On the other hand, it is now known that many features of the Gaussian case are universal. The asymptotics of the one-point function are *not* universal (that is, for general $V(x)$ one does not have a semicircle law anymore) but nonetheless the correlations of eigenvalues in the bulk and at the edge follow, respectively, the sine kernel and the Airy kernel. We now move on to study the more general and quite recent techniques that allow one to prove universality results for eigenvalues in non-Gaussian but unitarily invariant Hermitian random matrix ensembles.

3 Riemann-Hilbert Problems for Orthogonal Polynomials

The breakthrough came from a paper of Fokas, Its, and Kitaev [5], which contained a way of characterizing orthogonal polynomials by means of a matrix-valued Riemann-Hilbert problem of analytic function theory. To prepare for understanding the Riemann-Hilbert problem for orthogonal polynomials (and Riemann-Hilbert problems more generally) we need to review some basics of analytic functions.

3.1 Analytic functions and Cauchy integrals.

3.1.1 Cauchy-Riemann equations and the Theorems of Liouville and Cauchy.

An analytic function $f(z) = u(x,y) + iv(x,y)$, $z = x + iy$, satisfies the Cauchy-Riemann equations:

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \qquad \text{and} \qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}\,. \tag{76}$$

We will use this fact later in the following context. Suppose $f(z)$ is known to be an analytic function that is real-valued on the real axis of the complex z-plane ($v(x,0) \equiv 0$). If $f(z)$ is increasing along the real axis ($u_x(x,0) > 0$) then by the Cauchy-Riemann equations we have $v_y(x,0) > 0$ also, which says that $v(x,y) = \Im\{f(z)\}$ is positive just above the real axis and negative just below the real axis. The reverse is true if $f(z)$ is decreasing.

A function $f(z)$ that is analytic for z in the whole complex plane is called *entire*.

Proposition 4 (Liouville's Theorem)**.** *An entire function $f(z)$ that is bounded in the limit $z \to \infty$ (from all directions) is a constant function $f(z) \equiv f_0$.*

It follows from this theorem that if $f(z)$ is an entire function that decays to zero as $z \to \infty$ then $f(z) \equiv 0$, and that if $f(z)$ is an entire function satisfying $f(z) = O(z^k)$ as $z \to \infty$ for some $k = 0, 1, 2, \ldots$, then $f(z)$ is a polynomial of degree (at most) k.

Finally, recall the following fundamental result, which we already used implicitly in applying the steepest descent method of contour integration.

Proposition 5 (Cauchy's Theorem). *Suppose C is a simple closed curve, and that $f(z)$ is a function analytic within C and continuous up to C. Then*

$$\int_C f(z)\,dz = 0\,. \tag{77}$$

3.1.2 Cauchy integrals.

The reference for this material is Muskhelishvili [6].

Cauchy transform. Given a contour Σ in the complex plane, the contour integral

$$(\mathcal{C}^\Sigma f)(z) := \frac{1}{2\pi i}\int_\Sigma \frac{f(w)\,dw}{w-z}\,, \qquad z \in \mathbb{C}\setminus\Sigma \tag{78}$$

is called the *Cauchy transform* of f relative to Σ, or in more general context, a *Cauchy integral.* The function $f(w)$, $w \in \Sigma$ is the *density* of the Cauchy integral. $\mathcal{C}^\Sigma f$ is an analytic function of z in its domain of definition. If Σ is bounded, or if Σ is unbounded but $f(w)$ decays to zero at a sufficiently rapid rate as $w \in \Sigma$ tends to infinity, then $(\mathcal{C}^\Sigma f)(z) = O(z^{-1})$ as $z \to \infty$.

Boundary values. Plemelj formula. While the Cauchy transform of f is only defined for values of z not on the contour Σ, it frequently happens that the integral has well-defined limiting values as z approaches a point $z_0 \in \Sigma$ from either side. As Σ is oriented, it has a left and right side at each point $z_0 \in \Sigma$, and we define

$$(\mathcal{C}_+^\Sigma f)(z_0) := \lim_{\substack{z\to z_0 \\ z \text{ on left side of } \Sigma}} (\mathcal{C}^\Sigma f)(z) \qquad \text{and} \qquad (\mathcal{C}_-^\Sigma f)(z_0) := \lim_{\substack{z\to z_0 \\ z \text{ on right side of } \Sigma}} (\mathcal{C}^\Sigma f)(z) \tag{79}$$

wherever these (nontangential to Σ) limits make sense. The left and right boundary values are not usually equal even if they both exist. They are, however, related.

Proposition 6 (Plemelj formula). *Suppose that $f(w)$ is a Hölder continuous function on Σ in a neighborhood U of a point $z_0 \in \Sigma$ at which Σ is smooth and orientable. Then the boundary values $(\mathcal{C}_\pm^\Sigma f)(z)$ both exist and are also Hölder continuous (with the same exponent) in $\Sigma \cap U$ and satisfy*

$$(\mathcal{C}_+^\Sigma f)(z_0) - (\mathcal{C}_-^\Sigma f)(z_0) = f(z_0)\,. \tag{80}$$

If $f(w)$ is not necessarily Hölder continuous, but is in $L^2(\Sigma)$, then it turns out that the boundary values $(\mathcal{C}_\pm^\Sigma f)(z)$ exist pointwise for almost all $z \in \Sigma$ and may be identified with functions in $L^2(\Sigma)$.

The operator viewpoint. As starting with a function on Σ, calculating the Cauchy transform $\mathcal{C}^\Sigma f$ and then taking limits from the complex plane to Σ again results in functions on Σ, one should view $\mathcal{C}_\pm^\Sigma$ as (obviously linear) operators acting on some space of functions on the contour Σ. A crucial fact is that on the Hölder spaces with exponent strictly less than one the *Cauchy operators* $\mathcal{C}_\pm^\Sigma$ are bounded operators: there is a constant $K_{H^\nu(\Sigma)} > 0$, $0 < \nu < 1$, such that

$$\|\mathcal{C}_\pm^\Sigma f\|_{H^\nu(\Sigma)} \leq K_{H^\nu(\Sigma)}\|f\|_{H^\nu(\Sigma)} \tag{81}$$

holds for all f defined on Σ for which the Hölder norm

$$\|f\|_{H^\nu(\Sigma)} := \sup_{w\in\Sigma}|f(w)| + \sup_{z,w\in\Sigma}\frac{|f(z)-f(w)|}{|z-w|^\nu} \tag{82}$$

is finite (such f make up the space $H^\nu(\Sigma)$). This result is attributed to Plemelj and Privalov. There is a corresponding result for the space $L^2(\Sigma)$ attributed to many people, some quite recent, including Coifman, McIntosh, and Meyer [7] in the case of Lipschitz curves Σ. There is a constant $K_{L^2(\Sigma)} > 0$ such that

$$\|\mathcal{C}_\pm^\Sigma f\|_{L^2(\Sigma)} \leq K_{L^2(\Sigma)}\|f\|_{L^2(\Sigma)} \tag{83}$$

holds for all $f \in L^2(\Sigma)$, that is, functions $\Sigma \to \mathbb{C}$ for which

$$\|f\|_{L^2(\Sigma)} := \left(\int_\Sigma |f(z)|^2 \, |dz| \right)^{1/2} \tag{84}$$

is finite.

3.2 The Riemann-Hilbert problem and its "solution".

Now we are ready to formulate the Riemann-Hilbert problem related to orthogonal polynomials.

Riemann-Hilbert Problem 1. *Let $n \geq 0$ be an integer. Find a 2×2 matrix-valued function $\mathbf{A}^n(z)$, $z \in \mathbb{C}$, with the following properties:*

> **Analyticity.** *$\mathbf{A}^n(z)$ is analytic for $z \in \mathbb{C} \setminus \mathbb{R}$, and takes continuous boundary values $\mathbf{A}^n_\pm(x)$ as z tends to $x \in \mathbb{R}$ from $z \in \mathbb{C}_\pm$.*
>
> **Jump Condition.** *The boundary values are connected by the relation*
>
> $$\mathbf{A}^n_+(x) = \mathbf{A}^n_-(x) \begin{bmatrix} 1 & e^{-NV(x)} \\ 0 & 1 \end{bmatrix}, \qquad x \in \mathbb{R}\,. \tag{85}$$
>
> **Normalization.** *The matrix $\mathbf{A}^n(z)$ is normalized at $z = \infty$ as follows:*
>
> $$\lim_{z \to \infty} \mathbf{A}^n(z) z^{-n\sigma_3} = \mathbb{I}\,, \tag{86}$$
>
> *where the limit may be taken in any direction.*

(Some notation: σ_3 is a *Pauli matrix* given by

$$\sigma_3 := \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \qquad \text{and} \qquad z^{-n\sigma_3} := \begin{bmatrix} z^{-n} & 0 \\ 0 & z^n \end{bmatrix} \tag{87}$$

and $\mathbb{I}$ denotes the 2×2 identity matrix.)

Proposition 7. *Suppose that $V(x)$ is a continuous function of $x \in \mathbb{R}$ that grows sufficiently rapidly as $|x| \to \infty$ (polynomial is enough), and that $n \geq 0$. Then Riemann-Hilbert Problem 1 has a unique solution, namely*

$$\mathbf{A}^n(z) = \begin{bmatrix} \pi_n(z) & (\mathcal{C}^{\mathbb{R}}\pi_n(\cdot)e^{-NV(\cdot)})(z) \\ -2\pi i p_{n-1,n-1}^2 \pi_{n-1}(z) & -2\pi i p_{n-1,n-1}^2 (\mathcal{C}^{\mathbb{R}}\pi_{n-1}(\cdot)e^{-NV(\cdot)})(z) \end{bmatrix}, \qquad z \in \mathbb{C} \setminus \mathbb{R} \tag{88}$$

for $n > 0$, and

$$\mathbf{A}^0(z) = \begin{bmatrix} 1 & (\mathcal{C}^{\mathbb{R}} e^{-NV(\cdot)})(z) \\ 0 & 1 \end{bmatrix}, \qquad z \in \mathbb{C} \setminus \mathbb{R}\,. \tag{89}$$

(Recall that $\pi_n(z) = p_n(z)/p_{n,n} = z^n + \cdots$ is the nth monic orthogonal polynomial with respect to the weight $e^{-NV(x)}$.)

Proof. Assume $n > 0$ (the proof for $n = 0$ is analogous and simpler but not a special case of general $n \geq 0$). The first column of the jump relation (85) reads

$$A^n_{11+}(x) = A^n_{11-}(x) \qquad \text{and} \qquad A^n_{21+}(x) = A^n_{21-}(x)\,, \qquad x \in \mathbb{R}\,. \tag{90}$$

So, $A^n_{11}(z)$ and $A^n_{21}(z)$ are analytic functions in $\mathbb{C} \setminus \mathbb{R}$ taking continuous boundary values on the real axis from the upper and lower half planes, and these boundary values agree. Thus $A^n_{11}(z)$ and $A^n_{21}(z)$ extend continuously to $\mathbb{R}$ and (these extensions) constitute continuous functions in the whole complex plane. It follows that analyticity also extends to $\mathbb{R}$, making the components of the first column of $\mathbf{A}^n(z)$ entire functions of z.

Now we consider what the normalization condition (86) says about the entire functions $A_{11}^n(z)$ and $A_{21}^n(z)$. The first column of (86) implies that

$$A_{11}^n(z) = z^n(1+o(1)) \qquad \text{and} \qquad A_{21}^n(z) = o(z^n) \tag{91}$$

as $z \to \infty$. By Liouville's Theorem, $A_{11}^n(z)$ has to be a monic polynomial of degree n, while $A_{21}^n(z)$ has to be a polynomial of degree (at most) $n-1$.

Now we look at the second column of the jump condition (85), which implies

$$A_{12+}^n(x) - A_{12-}^n(x) = A_{11}^n(x)e^{-NV(x)} \qquad \text{and} \qquad A_{22+}^n(x) - A_{22-}^n(x) = A_{21}^n(x)e^{-NV(x)}\,, \qquad x \in \mathbb{R}\,. \tag{92}$$

Although the polynomials $A_{11}^n(z)$ and $A_{21}^n(z)$ are not yet known, let's view these as equations to be solved for $A_{12}^n(z)$ and $A_{22}^n(z)$ in terms of $A_{11}^n(z)$ and $A_{21}^n(z)$. The Plemelj formula tells us how to solve these equations. Indeed, the general solution is given by

$$A_{j2}^n(z) = (\mathcal{C}^{\mathbb{R}} A_{j1}^n(\cdot)e^{-NV(\cdot)})(z) + e_j(z) \tag{93}$$

for $j = 1, 2$, where $e_j(z)$ are entire functions of z. (Check the jump condition!)

Finally, we consider the second column of the normalization condition (86), which reads

$$A_{12}^n(z) = o(z^{-n}) \qquad \text{and} \qquad A_{22}^n(z) = z^{-n} + o(z^{-n}) \tag{94}$$

as $z \to \infty$. In particular, these conditions require that $A_{j2}^n(z)$ tends to zero as $z \to \infty$ (because $n > 0$). Since the Cauchy integral component of $A_{j2}^n(z)$ already decays to zero for large z we therefore require that $e_j(z) \to 0$ as $z \to \infty$. By Liouville's Theorem it then follows that the entire functions $e_j(z)$ are identically zero. It remains to enforce the precise rate of decay $o(z^{-n})$ on the Cauchy integrals. To do this, we obtain asymptotic expansions of $A_{j2}^n(z)$ by expanding the Cauchy kernel under the integral sign:

$$\frac{1}{w-z} = -z^{-1} - wz^{-2} - w^2z^{-3} - \cdots\,, \tag{95}$$

from which it follows that $A_{j2}^n(z)$ has the asymptotic expansion

$$A_{j2}^n(z) \sim -\frac{1}{2\pi iz}\int_{\mathbb{R}} A_{j1}^n(x)e^{-NV(x)}\,dx - \frac{1}{2\pi iz^2}\int_{\mathbb{R}} A_{j1}^n(x)xe^{-NV(x)}\,dx - \frac{1}{2\pi iz^3}\int_{\mathbb{R}} A_{j1}^n(x)x^2e^{-NV(x)}\,dx - \cdots \tag{96}$$

as $z \to \infty$. To obtain the required decay it is therefore necessary that

$$\int_{\mathbb{R}} A_{11}^n(x)x^k e^{-NV(x)}\,dx = 0\,, \qquad k = 0, 1, 2, \ldots, n-1 \tag{97}$$

which identifies the monic polynomial $A_{11}^n(z)$ with $\pi_n(z)$, and that

$$\int_{\mathbb{R}} A_{21}^n(x)x^k e^{-NV(x)}\,dx = 0\,, \qquad k = 0, 1, 2, \ldots, n-2 \tag{98}$$

and

$$\int_{\mathbb{R}} A_{21}^n(x)x^{n-1}e^{-NV(x)}\,dx = \int_{\mathbb{R}} A_{21}^n(x)\pi_{n-1}(x)e^{-NV(x)}\,dx = -2\pi i\,, \tag{99}$$

which identifies the polynomial $A_{21}^n(z)$ with $-2\pi i p_{n-1,n-1}^2 \pi_{n-1}(z)$. $\square$

In the section title we called this the "solution" of the Riemann-Hilbert problem using quotation marks because while Proposition 7 provides a characterization of the components of the matrix $\mathbf{A}^n(z)$, the use of this "solution" formula in practice requires the construction of the orthogonal polynomials for $e^{-NV(x)}$ by some other means. The main idea here is to turn Proposition 7 on its head, and to use the conditions of Riemann-Hilbert Problem 1 as an indirect method to construct the orthogonal polynomials. This approach will be universally applicable in the limit when n and N are proportionately large, which is the case of Plancherel-Rotach asymptotics and the case relevant to random matrix theory.

Also note: the ingredients for the reproducing kernel $K_N(x,y)$ are explicitly contained in the first column of $\mathbf{A}^n(z)$ for $n = N$. Indeed,

$$K_N(x,y) = -\frac{e^{-N(V(x)+V(y))/2}}{2\pi i} \cdot \frac{A_{11}^N(x)A_{21}^N(y) - A_{21}^N(x)A_{11}^N(y)}{x-y}\,. \tag{100}$$

Finally, we note the following important fact:

Proposition 8. *The solution* $\mathbf{A}^n(z)$ *of Riemann-Hilbert Problem 1 satisfies* $\det(\mathbf{A}^n(z)) \equiv 1$.

Proof. We can prove this directly from the conditions of the Riemann-Hilbert problem itself. Indeed, $d(z) := \det(\mathbf{A}^n(z))$ is, by the analyticity condition, an analytic function for $z \in \mathbb{C}\backslash\mathbb{R}$ that takes continuous boundary values $d_\pm(x)$ for $x \in \mathbb{R}$. Taking determinants in (85) gives $d_+(x) = d_-(x)$, so $d(z)$ is an entire function. Finally, taking determinants in (86) shows that $d(z) \to 1$ as $z \to \infty$, so by Liouville's Theorem $d(z) \equiv 1$. □

4 Riemann-Hilbert Problems in General

4.1 General Riemann-Hilbert problems.

A contour Σ, possibly with self-intersection points, is called *complete* if it divides the complex plane into two complementary regions, which we may call $\Omega_\pm$. See Figure 3. A complete contour has a natural orientation

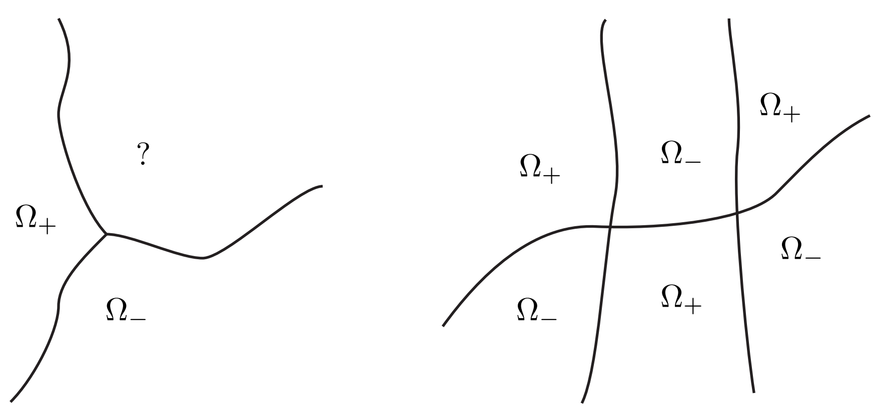

Figure 3: *Left: an incomplete contour. Right: a complete contour.*

on each arc, so that Ω_+ lies on the left. If Σ is a complete contour, then the following operator identities hold:

$$\mathcal{C}_+^\Sigma \circ \mathcal{C}_-^\Sigma = \mathcal{C}_-^\Sigma \circ \mathcal{C}_+^\Sigma = 0\,. \tag{101}$$

The proof is by Cauchy's Theorem, since for a complete contour Σ, $(\mathcal{C}_\pm^\Sigma f)(z)$ is a function on Σ that is the boundary value of a function analytic in $\Omega_\pm$ (and decaying at $z = \infty$ if $\Omega_\pm$ is unbounded). Therefore, to calculate $\mathcal{C}_-^\Sigma \circ \mathcal{C}_+^\Sigma f$, deform the integration path for $\mathcal{C}_-^\Sigma$ into Ω_+ and apply Cauchy's Theorem.

Applying (101) in the Plemelj formula shows that

$$\mathcal{C}_+^\Sigma - \mathcal{C}_-^\Sigma = 1 \qquad \text{implies} \qquad \mathcal{C}_+^\Sigma \circ \mathcal{C}_+^\Sigma - \mathcal{C}_+^\Sigma \circ \mathcal{C}_-^\Sigma = \mathcal{C}_+^\Sigma \qquad \text{implies} \qquad \mathcal{C}_+^\Sigma \circ \mathcal{C}_+^\Sigma = \mathcal{C}_+^\Sigma\,, \tag{102}$$

and likewise,

$$(-\mathcal{C}_-^\Sigma) \circ (-\mathcal{C}_-^\Sigma) = -\mathcal{C}_-^\Sigma\,. \tag{103}$$

This shows that for complete contours Σ, $P_\pm := \pm\mathcal{C}_\pm^\Sigma$ are *complementary projection operators.*

Let Σ be complete and let $\mathbf{V}(z)$, $z \in \Sigma$ be a 2×2 invertible (for each $z \in \Sigma$) matrix-valued function defined on Σ that decays sufficiently rapidly to $\mathbb{I}$ as $z \to \infty$ along any unbounded parts of Σ. Consider the following Riemann-Hilbert Problem.

Riemann-Hilbert Problem 2. *Find a 2×2 matrix-valued function $\mathbf{M}(z)$ with the following properties.*

Analyticity. *$\mathbf{M}(z)$ is analytic for $z \in \mathbb{C} \setminus \Sigma$ and takes boundary values $\mathbf{M}_\pm(w)$ in a certain function space as $z \to w \in \Sigma$ with $z \in \Omega_\pm$.*

Jump Condition. *The boundary values are connected by the relation*

$$\mathbf{M}_+(z) = \mathbf{M}_-(z)\mathbf{V}(z)\,, \qquad z \in \Sigma\,. \tag{104}$$

Normalization. *The matrix $\mathbf{M}(z)$ is normalized at $z = \infty$ as follows:*

$$\lim_{z\to\infty} \mathbf{M}(z) = \mathbb{I}\,, \tag{105}$$

where the limit may be taken in any direction.

Note: the condition that Σ be complete can be dispensed with, since any contour can be completed by adding one or more arcs, and the jump matrix $\mathbf{V}(z)$ may be taken to be the identity matrix $\mathbb{I}$ on the added arcs. See Figure 4.

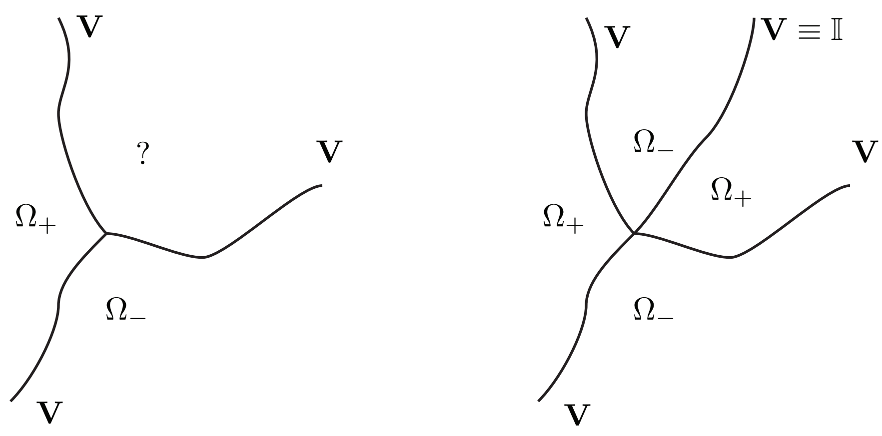

Figure 4: *Left: a Riemann-Hilbert problem on an incomplete contour. Right: a completion thereof.*

4.2 Singular integral equations.

Problems of the form of Riemann-Hilbert Problem 2 may be solved by rephrasing the problem in terms of the solution of a system of integral equations. Suppose that $\mathbf{V}(z)$ is can be factored for each $z \in \Sigma$ as

$$\mathbf{V}(z) = \mathbf{B}_-(z)^{-1}\mathbf{B}_+(z) \tag{106}$$

for some other invertible matrix functions $\mathbf{B}_\pm(z)$ defined on Σ. Note: one choice that is admissable for many applications is just $\mathbf{B}_+(z) := \mathbf{V}(z)$ and $\mathbf{B}_-(z) \equiv \mathbb{I}$. Then set

$$\mathbf{W}_+(z) := \mathbf{B}_+(z) - \mathbb{I} \qquad \text{and} \qquad \mathbf{W}_-(z) := \mathbb{I} - \mathbf{B}_-(z)\,, \tag{107}$$

and define a singular integral operator $\mathcal{C}^\Sigma_\mathbf{W}$ on Σ by the formula

$$(\mathcal{C}^\Sigma_\mathbf{W}\mathbf{F})(z) = (\mathcal{C}^\Sigma_+(\mathbf{F}\mathbf{W}_-))(z) + (\mathcal{C}^\Sigma_-(\mathbf{F}\mathbf{W}_+))(z)\,. \tag{108}$$

Proposition 9. *Suppose that $\mathbf{X}(z)$, $z \in \Sigma$, is the unique solution (in a well-chosen function space) of the singular integral equation*

$$\mathbf{X}(w) - (\mathcal{C}^\Sigma_\mathbf{W}\mathbf{X})(w) = \mathbb{I}\,, \qquad w \in \Sigma\,. \tag{109}$$

Then, the matrix

$$\mathbf{M}(z) := \mathbb{I} + (\mathcal{C}^\Sigma(\mathbf{X}\mathbf{W}_+))(z) + (\mathcal{C}^\Sigma(\mathbf{X}\mathbf{W}_-))(z)\,, \qquad z \in \mathbb{C} \setminus \Sigma \tag{110}$$

is the unique solution (with boundary values in the same function space) of Riemann-Hilbert Problem 2.

Proof. We just give the basic idea of the proof. Clearly the solution formula (110) represents an analytic function for $z \in \mathbb{C} \setminus \Sigma$. The Cauchy transforms in the solution formula (110) decay as $z \to \infty$, so the normalization condition (105) is satisfied.

Next, we verify the jump condition (104), which in view of the factorization $\mathbf{V}(z) = \mathbf{B}_-(z)^{-1}\mathbf{B}_+(z)$ can be written in the form $\mathbf{M}_+(z)\mathbf{B}_+(z)^{-1} = \mathbf{M}_-(z)\mathbf{B}_-(z)^{-1}$. It is this that we must show holds when $\mathbf{M}(z)$ is given by (110) where $\mathbf{X}(z)$ solves (109). Note that for $z \in \Sigma$,

$$\begin{aligned}\mathbf{M}_+(z) &= \mathbb{I} + (\mathcal{C}^\Sigma_+(\mathbf{X}\mathbf{W}_+))(z) + (\mathcal{C}^\Sigma_+(\mathbf{X}\mathbf{W}_-))(z)\\ &= \mathbb{I} + (\mathcal{C}^\Sigma_+(\mathbf{X}\mathbf{W}_+))(z) + \left[\mathbf{X}(z) - (\mathcal{C}^\Sigma_-(\mathbf{X}\mathbf{W}_+))(z) - \mathbb{I}\right]\\ &= \mathbf{X}(z) + (\mathcal{C}^\Sigma_+(\mathbf{X}\mathbf{W}_+))(z) - (\mathcal{C}^\Sigma_-(\mathbf{X}\mathbf{W}_+))(z)\end{aligned} \tag{111}$$

by using the integral equation (109), and similarly,

$$\mathbf{M}_-(z) = \mathbf{X}(z) + (\mathcal{C}^\Sigma_-(\mathbf{X}\mathbf{W}_-))(z) - (\mathcal{C}^\Sigma_+(\mathbf{X}\mathbf{W}_-))(z)\,. \tag{112}$$

Now we apply the Plemelj formula to these expressions:

$$\begin{aligned}\mathbf{M}_+(z) &= \mathbf{X}(z) + \left[\mathbf{X}(z)\mathbf{W}_+(z) + (\mathcal{C}^\Sigma_-(\mathbf{X}\mathbf{W}_+))(z)\right] - (\mathcal{C}^\Sigma_-(\mathbf{X}\mathbf{W}_+))(z)\\ &= \mathbf{X}(z)(\mathbb{I} + \mathbf{W}_+(z))\\ &= \mathbf{X}(z)\mathbf{B}_+(z)\,,\end{aligned} \tag{113}$$

and similarly,

$$\mathbf{M}_-(z) = \mathbf{X}(z)\mathbf{B}_-(z)\,. \tag{114}$$

So, it is indeed true that $\mathbf{M}_+(z)\mathbf{B}_+(z)^{-1} = \mathbf{M}_-(z)\mathbf{B}_-(z)^{-1}$ with the common value being $\mathbf{X}(z)$. $\square$

So, to solve Riemann-Hilbert Problem 2 it is enough to solve the integral equation (109). Let's consider this equation in the space $L^2(\Sigma)$. Suppose that $\mathbf{W}_\pm(z)$ are uniformly bounded matrix functions on Σ. Then $\mathcal{C}^\Sigma_\mathbf{W}$ is a bounded operator with norm

$$\|\mathcal{C}^\Sigma_\mathbf{W}\|_{L^2(\Sigma)} \le K_{L^2(\Sigma)} \cdot \left(\sup_{z\in\Sigma}|\mathbf{W}_+(z)| + \sup_{z\in\Sigma}|\mathbf{W}_-(z)|\right) \tag{115}$$

where $|\cdot|$ denotes any norm on 2×2 matrices. If $|\mathbf{W}_\pm(z)| \le \epsilon$ holds for some number $\epsilon > 0$ that can be (somehow) made small, then for sufficiently small ϵ we will have

$$\|\mathcal{C}^\Sigma_\mathbf{W}\|_{L^2(\Sigma)} \le 2\epsilon K_{L^2(\Sigma)} < 1\,, \qquad \text{for } \epsilon \text{ small enough.} \tag{116}$$

In this situation, the inverse operator $(1-\mathcal{C}^{\Sigma}_{\mathbf{W}})^{-1}$ exists as a bounded operator on $L^2(\Sigma)$, and is represented by the convergent Neumann series

$$(1-\mathcal{C}^{\Sigma}_{\mathbf{W}})^{-1} = 1 + \mathcal{C}^{\Sigma}_{\mathbf{W}} + \mathcal{C}^{\Sigma}_{\mathbf{W}}\circ\mathcal{C}^{\Sigma}_{\mathbf{W}} + \mathcal{C}^{\Sigma}_{\mathbf{W}}\circ\mathcal{C}^{\Sigma}_{\mathbf{W}}\circ\mathcal{C}^{\Sigma}_{\mathbf{W}} + \cdots\,. \tag{117}$$

This convergent series is also an asymptotic series in the limit $\epsilon\to 0$, and partial sums applied to $\mathbb{I}$ give accurate approximations to the unique solution $\mathbf{X}(z) = (1-\mathcal{C}^{\Sigma}_{\mathbf{W}})^{-1}\mathbb{I}$ of the integral equation (109) in this limit. Note that in this asymptotic situation we conclude that the solution $\mathbf{M}(z)$ of Riemann-Hilbert Problem 2 satisfies $\mathbf{M}(z)\to\mathbb{I}$ as $\epsilon\to 0$ for each fixed $z\in\mathbb{C}\setminus\Sigma$. We say that such a Riemann-Hilbert problem is a *near-identity problem.*

4.3 Equivalence of Riemann-Hilbert problems. Deformation.

Suppose that $\mathbf{G}(z)$ is a given 2×2 invertible matrix-valued function of $z\in\mathbb{C}$ that, moreover, is analytic for $z\in\mathbb{C}\setminus\Sigma_{\mathbf{G}}$ where $\Sigma_{\mathbf{G}}$ is some oriented contour. Then, if $\mathbf{M}(z)$ is an unknown satisfying Riemann-Hilbert Problem 2 we may form the product

$$\tilde{\mathbf{M}}(z) := \mathbf{M}(z)\mathbf{G}(z)\,,\qquad z\in\mathbb{C}\setminus\tilde{\Sigma}\,, \tag{118}$$

where $\tilde{\Sigma} := \Sigma\cup\Sigma_{\mathbf{G}}$ is an arcwise-oriented contour. It follows by direct calculation that $\mathbf{M}(z)$ satisfies Riemann-Hilbert Problem 2 if and only if $\tilde{\mathbf{M}}(z)$ satisfies another Riemann-Hilbert Problem:

Riemann-Hilbert Problem 3. *Find a 2×2 matrix-valued function $\tilde{\mathbf{M}}(z)$ with the following properties:*

> **Analyticity.** $\tilde{\mathbf{M}}(z)$ *is analytic for $z\in\mathbb{C}\setminus\tilde{\Sigma}$ and takes boundary values $\tilde{\mathbf{M}}_{\pm}(z)$ on $\tilde{\Sigma}$ from the left and right sides.*
>
> **Jump Condition.** *The boundary values of $\tilde{\mathbf{M}}(z)$ are related by*
>
> $$\tilde{\mathbf{M}}_{+}(z) = \tilde{\mathbf{M}}_{-}(z)\tilde{\mathbf{V}}(z)\,,\qquad z\in\tilde{\Sigma} \tag{119}$$
>
> *where $\tilde{\mathbf{V}}(z)$ is a given matrix-valued function on $\tilde{\Sigma}$ (see below for a definition).*
>
> **Normalization.** *The matrix $\tilde{\mathbf{M}}(z)$ is normalized at $z=\infty$ as follows:*
>
> $$\lim_{z\to\infty}\tilde{\mathbf{M}}(z)\mathbf{G}(z)^{-1} = \mathbb{I}\,. \tag{120}$$

What is $\tilde{\mathbf{V}}(z)$? There are three different formulae, all explicit. First suppose that $z\in\tilde{\Sigma}\setminus\Sigma_{\mathbf{G}}$. Then $\mathbf{G}(z)$ is analytic and hence continuous at z. Therefore, from (104),

$$\tilde{\mathbf{M}}_{+}(z) = \mathbf{M}_{+}(z)\mathbf{G}(z) = \mathbf{M}_{-}(z)\mathbf{V}(z)\mathbf{G}(z) = \tilde{\mathbf{M}}_{-}(z)\mathbf{G}(z)^{-1}\mathbf{V}(z)\mathbf{G}(z)\,,\qquad z\in\tilde{\Sigma}\setminus\Sigma_{\mathbf{G}}\,, \tag{121}$$

so for such z we have $\tilde{\mathbf{V}}(z) = \mathbf{G}(z)^{-1}\mathbf{V}(z)\mathbf{G}(z)$. Next, suppose that $z\in\tilde{\Sigma}\setminus\Sigma$. Then $\mathbf{M}(z)$ is continuous at z so

$$\tilde{\mathbf{M}}_{+}(z) = \mathbf{M}(z)\mathbf{G}_{+}(z) = \tilde{\mathbf{M}}_{-}(z)\mathbf{G}_{-}(z)^{-1}\mathbf{G}_{+}(z)\,,\qquad z\in\tilde{\Sigma}\setminus\Sigma\,, \tag{122}$$

so for such z we have $\tilde{\mathbf{V}}(z) = \mathbf{G}_{-}(z)^{-1}\mathbf{G}_{+}(z)$. Finally, suppose that $z\in\Sigma\cap\Sigma_{\mathbf{G}}$. Now both $\mathbf{M}(z)$ and $\mathbf{G}(z)$ are discontinuous at z so

$$\tilde{\mathbf{M}}_{+}(z) = \mathbf{M}_{+}(z)\mathbf{G}_{+}(z) = \mathbf{M}_{-}(z)\mathbf{V}(z)\mathbf{G}_{+}(z) = \tilde{\mathbf{M}}_{-}(z)\mathbf{G}_{-}(z)^{-1}\mathbf{V}(z)\mathbf{G}_{+}(z)\,,\qquad z\in\Sigma\cap\Sigma_{\mathbf{G}}\,, \tag{123}$$

so for such z we have $\tilde{\mathbf{V}}(z) = \mathbf{G}_{-}(z)^{-1}\mathbf{V}(z)\mathbf{G}_{+}(z)$.

The simplest example of this construction is in the case when the jump matrix $\mathbf{V}(z)$ and its inverse $\mathbf{V}(z)^{-1}$ are both analytic functions of $z\in\mathbb{C}$ themselves in some open set $U\subset\mathbb{C}$ with $U\cap\Sigma\neq\emptyset$. Then by defining the matrix $\mathbf{G}(z)$ as in Figure 5, we see that in going from $\mathbf{M}(z)$ to $\tilde{\mathbf{M}}(z)$, the jump on part of Σ has disappeared, and a new jump on $\Sigma_{\mathbf{G}}$ has appeared, but the new jump is given by the same formula:

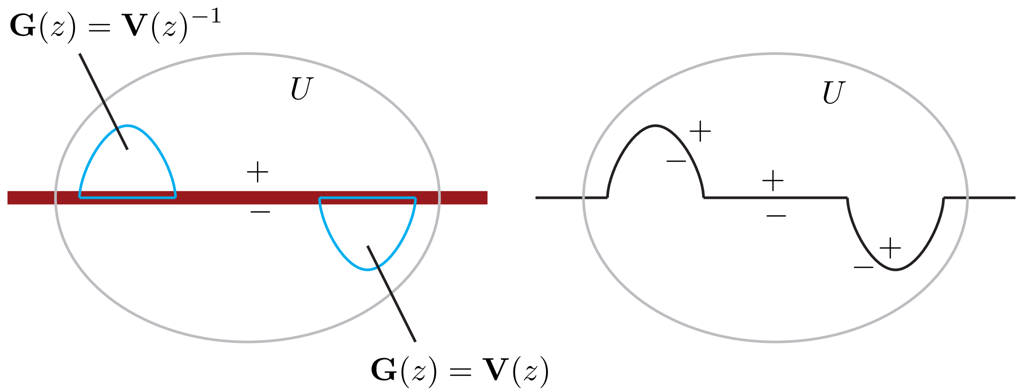

Figure 5: *Left: the matrix* $\mathbf{G}(z)$ *is set equal to* $\mathbb{I}$ *except where indicated.* Σ *is shown in red and* $\Sigma_{\mathbf{G}}$ *is shown in blue. Right: the resulting contour* $\tilde{\Sigma}$ *on which the jump matrix is* $\tilde{\mathbf{V}}(z) = \mathbf{V}(z)$ *(trivial parts of* $\tilde{\Sigma}$ *on which* $\tilde{\mathbf{V}}(z) \equiv \mathbb{I}$ *are not shown).*

$\tilde{\mathbf{V}}(z) = \mathbf{V}(z)$. This construction is the analogue for Riemann-Hilbert problems of Cauchy's Theorem for contour integration.

This *equivalence principle* for Riemann-Hilbert problems provides a way of solving a given Riemann-Hilbert problem of the form of Riemann-Hilbert Problem 3 where the jump matrices and/or the normalization condition involve a small parameter ϵ (for our application, we will have $\epsilon = 1/N$). You can try to find a known matrix $\mathbf{G}(z)$ that converts this Riemann-Hilbert problem into one of the form of Riemann-Hilbert Problem 2 that you can show is a near-identity problem in the limit $\epsilon \to 0$. Then you can rely on integral equation theory to provide an asymptotic expansion for $\mathbf{M}(z)$, and then asymptotics for the unknown of interest, $\tilde{\mathbf{M}}(z)$, are obtained through the relation $\tilde{\mathbf{M}}(z) = \mathbf{M}(z)\mathbf{G}(z)$. In the literature an explicit matrix $\mathbf{G}(z)$ that transforms your problem into a near-identity problem is called a *parametrix*. A parametrix may also be thought of as an asymptotically valid model for $\tilde{\mathbf{M}}(z)$.

5 Asymptotic Analysis of Riemann-Hilbert Problems for Orthogonal Polynomials

We will now consider in detail the asymptotic solution of Riemann-Hilbert Problem 1 characterizing the orthogonal polynomials with weight $e^{-NV(x)}$. According to (100), the situation interesting for random matrix theory is to obtain asymptotics for $\mathbf{A}^n(z)$ in the special case of $n = N$, and in the limit of $N \to +\infty$. What now follows is the modern version of the amazing calculation of Plancherel and Rotach based on contour integrals in the special case of $V(x) = x^2$. The modern version works for completely general $V(x)$.

5.1 Step 1: Repairing the normalization condition.

The most glaring difference between the Riemann-Hilbert problem for $\mathbf{A}^N(z)$ and a near-identity problem lies in the form of the normalization condition (86), which explicitly prevents $\mathbf{A}^N(z)$ from resembling the identity matrix at least near $z = \infty$. We can imagine trying to fix this problem by cooking up a scalar function $g(z)$ that looks like $\log(z)$ near $z = \infty$, because setting

$$\mathbf{B}^N(z) := \mathbf{A}^N(z)e^{-Ng(z)\sigma_3} = \mathbf{A}^N(z)\begin{bmatrix} e^{-Ng(z)} & 0 \\ 0 & e^{Ng(z)} \end{bmatrix}, \tag{124}$$

we see that the normalization condition (86) implies that $\mathbf{B}^N(z) \to \mathbb{I}$ as $z \to \infty$.

At this point it is important that in picking $g(z)$ we do not mess up (qualitatively speaking) the other properties that $\mathbf{A}^N(z)$ is supposed to have: analyticity in upper and lower half-planes and continuity of boundary values. For example, were we to choose $g(z) \equiv \log(z)$ then $e^{g(z)} \equiv z$ and while analyticity in half-planes is preserved upon going over to $\mathbf{B}^N(z)$ by (124), continuity of the boundary values is not (a pole of order N appears at $z = 0$).

One idea is that we could "smear out" the pole as follows: let $\psi(x)\,dx$ be a probability measure on $\mathbb{R}$ with a nice (say, compactly supported and Hölder continuous) density ψ. Now set

$$g(z) := \int_{-\infty}^{+\infty} \log(z - x)\,\psi(x)\,dx\,. \tag{125}$$

Then $g(z)$ is an analytic function for $z \in \mathbb{C} \setminus \mathbb{R}$ that takes continuous boundary values on $\mathbb{R}$ from $\mathbb{C}_\pm$, and moreover by expanding the integrand for large z,

$$g(z) = \log(z) \int_{-\infty}^{+\infty} \psi(x)\,dx + O(z^{-1}) = \log(z) + O(z^{-1}) \tag{126}$$

as $z \to \infty$, because $\psi(x)\,dx$ is a probability measure. Thus we have "smeared out" the pole by replacing a point mass by an absolutely continuous probability measure. (To be honest, this calculation works for any nice function $\psi(x)$ with integral equal to 1; the assumption that $\psi(x) \geq 0$ will, however, be useful later for a different reason.) Consequently, for such $g(z)$,

$$e^{-Ng(z)\sigma_3} = z^{-N\sigma_3}(\mathbb{I} + O(z^{-1}))\,, \qquad \text{as } z \to \infty \text{ for fixed } N \tag{127}$$

which is enough to conclude from the normalization condition (86) for $\mathbf{A}^N(z)$ that $\mathbf{B}^N(z)$ tends to the identity matrix as $z \to \infty$. At the end of step 1, we have obtained an equivalent Riemann-Hilbert problem for $\mathbf{B}^N(z)$ according to the principles outlined earlier:

Riemann-Hilbert Problem 4. *Find a 2×2 matrix-valued function $\mathbf{B}^N(z)$ with the following properties:*

> **Analyticity.** *$\mathbf{B}^N(z)$ is an analytic function of z for $z \in \mathbb{C} \setminus \mathbb{R}$ that takes continuous boundary values $\mathbf{B}^N_\pm(x)$ on $\mathbb{R}$ from the upper and lower half-planes $\mathbb{C}_\pm$.*
>
> **Jump Condition.** *The boundary values are related by*
>
> $$\mathbf{B}^N_+(x) = \mathbf{B}^N_-(x) \begin{bmatrix} e^{-N(g_+(x)-g_-(x))} & e^{-N(V(x)-g_+(x)-g_-(x))} \\ 0 & e^{N(g_+(x)-g_-(x))} \end{bmatrix}\,, \qquad x \in \mathbb{R}\,. \tag{128}$$
>
> *Here $g_\pm(x)$ denote the boundary values taken by $g(z)$ from $\mathbb{C}_\pm$.*
>
> **Normalization.** *The matrix $\mathbf{B}^N(z)$ is normalized at $z = \infty$ as follows:*
>
> $$\lim_{z\to\infty} \mathbf{B}^N(z) = \mathbb{I}\,. \tag{129}$$

Note that by the relation (124) of $\mathbf{B}^N(z)$ to $\mathbf{A}^N(z)$, the use of Proposition 8 gives $\det(\mathbf{B}^N(z)) \equiv 1$.

5.2 Logarithmic potential theory and equilibrium measures.

Now we are faced with a question: how should we choose the measure $\psi(x)\,dx$ to our advantage in considering the limit $N \to +\infty$? To study this problem, let's write out what $g_+(x) - g_-(x)$ and $V(x) - g_+(x) - g_-(x)$ are in terms of $\psi(x)$. To do this, we note that for fixed $y \in \mathbb{R}$,

$$[\log(x-y)]_\pm := \lim_{\epsilon\downarrow 0} \log(x + i\epsilon - y) = \begin{cases} \log|x-y|\,, & \text{for } x > y \\ \log|x-y| \pm i\pi & \text{for } x < y\,, \end{cases} \tag{130}$$

so that

$$g_{\pm}(x) = \int_{-\infty}^{+\infty} \log|x-y|\,\psi(y)\,dy \pm i\pi \int_{x}^{+\infty} \psi(y)\,dy\,, \tag{131}$$

and therefore

$$g_+(x) - g_-(x) = i\theta(x)\,, \qquad \theta(x) := 2\pi \int_x^{+\infty} \psi(y)\,dy \tag{132}$$

and

$$V(x) - g_+(x) - g_-(x) = -2\int_{-\infty}^{+\infty} \log|x-y|\,\psi(x)\,dx + V(x)\,. \tag{133}$$

5.2.1 The energy functional and its minimizer.

The right-hand side of (133) is the *variational derivative* or *Frechét derivative* of a functional of $\psi(x)\,dx$:

$$V(x) - g_+(x) - g_-(x) = \frac{\delta E}{\delta \psi}(x)\,, \tag{134}$$

where

$$E[\psi] := \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \log\left(\frac{1}{|x-y|}\right)\psi(x)\,dx\,\psi(y)\,dy + \int_{-\infty}^{+\infty} V(x)\,\psi(x)\,dx\,. \tag{135}$$

Physically, $E[\psi]$ is the Coulomb energy of a distribution $\psi(x)\,dx$ of positive electric charge confined to the real line in a two-dimensional universe (the Green's function for Laplace's equation in two dimensions is the kernel in the double integral). The charges are influenced both by mutual repulsion (the first term) and by being trapped in an externally applied electrostatic potential $V(x)$ (the second term). The *equilibrium measure* is the probability measure $\psi(x)\,dx$ that minimizes $E[\psi]$. It exists and is unique when $V(x)$ grows fast enough as $|x| \to \infty$ by virtue of general convexity arguments (the definitive statement is the Gauss-Frostman Theorem; see Saff and Totik [8]).

The equilibrium measure is characterized by its *variational conditions.* There is a constant $\ell \in \mathbb{R}$, the Lagrange multiplier used to enforce the constraint

$$\int_{-\infty}^{+\infty} \psi(x)\,dx = 1\,, \tag{136}$$

such that when $\psi(x)$ is the density of the equilibrium measure,

$$\frac{\delta E}{\delta \psi}(x) \equiv \ell\,, \qquad x \in \operatorname{supp}(\psi) \tag{137}$$

while

$$\frac{\delta E}{\delta \psi}(x) > \ell\,, \qquad x \notin \operatorname{supp}(\psi)\,. \tag{138}$$

Physically speaking once again, it seems reasonable that if $V(x)$ is a strictly convex function (that is, $V''(x) > 0$ everywhere) then we have a "single-well" rather than a "multiple-well" potential, and we expect the charges to all be lying in a single lump. This guess holds water mathematically as this result shows:

Proposition 10. *Suppose that $V : \mathbb{R} \to \mathbb{R}$ is a strictly convex function that grows sufficiently rapidly as $|x| \to \infty$. Then the equilibrium measure is supported on a single interval $[\alpha, \beta] \subset \mathbb{R}$.*

Proof. Suppose there were a gap in the support. Then the function $f(x)$ defined by

$$f(x) := \frac{\delta E}{\delta \psi}(x) - \ell \tag{139}$$

has to have a graph like that shown in Figure 6. For x in the gap we may differentiate under the integral

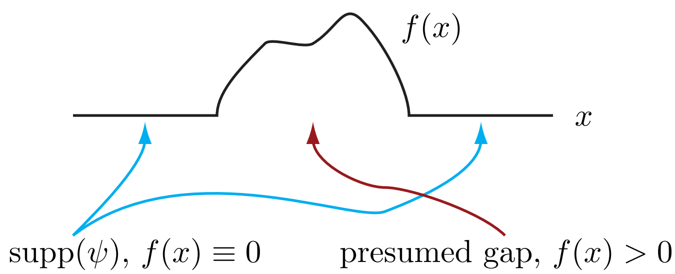

Figure 6: *The graph of $f(x)$ near a gap in supp(ψ).*

sign to find:

$$f'(x) = \frac{d}{dx}\frac{\delta E}{\delta \psi}(x) = -2\int_{\text{supp}(\psi)} \frac{\psi(y)\,dy}{x-y} + V'(x)\,, \qquad \text{for } x \text{ in the presumed gap} \tag{140}$$

and again:

$$f''(x) = 2\int_{\text{supp}(\psi)} \frac{\psi(y)\,dy}{(x-y)^2} + V''(x)\,, \qquad \text{for } x \text{ in the presumed gap.} \tag{141}$$

This is strictly positive for convex V. But to have $f(x) > 0$ in the gap and $f(x) = 0$ on either side requires $f''(x_0) < 0$ for some x_0 in the gap, which gives a contradiction. □

From now on we assume that $V(x)$ is convex for simplicity, although almost all of the theory goes through in the nonconvex case too (one has to deal with the possibility of multiple components of supp(ψ)).

5.2.2 How to find the equilibrium measure.

Equations governing $g'(z)$. We proceed somewhat indirectly, by looking for $g'(z)$ instead of $\psi(x)$. For $z \in \mathbb{C} \setminus \mathbb{R}$ differentiation of the "logarithmic transform" formula for $g(z)$ in terms of $\psi(x)$ gives

$$g'(z) = \int_{-\infty}^{+\infty} \frac{\psi(x)\,dx}{z-x} = -2\pi i (\mathcal{C}^{\mathbb{R}}\psi)(z)\,, \tag{142}$$

so $g'(z)$ is a factor of $-2\pi i$ away from the Cauchy transform of ψ. If we can find $g'(z)$, then we will know $\psi(x)$, since from the Plemelj formula,

$$\psi(x) = -\frac{1}{2\pi i}(g'_+(x) - g'_-(x))\,, \qquad x \in \mathbb{R}\,. \tag{143}$$

Differentiation of (133) and using the variational condition (137) gives

$$g'_+(x) + g'_-(x) = V'(x)\,, \qquad x \in \text{supp}(\psi)\,, \tag{144}$$

while on the other hand, from (143),

$$g'_+(x) - g'_-(x) = 0\,, \qquad x \notin \text{supp}(\psi)\,. \tag{145}$$

Moreover, from (142) we see that

$$g'(z) = z^{-1} + O(z^{-2})\,, \tag{146}$$

as $z \to \infty$. We view the system (144)–(146) as equations to be solved for a function $g'(z)$ analytic in $\mathbb{C} \setminus \mathbb{R}$. Once $g'(z)$ is known from these conditions, the equilibrium measure will be given by (143).

Solving for $g'(z)$. The "square-root trick". In the convex V case the support is an unknown interval $[\alpha, \beta]$. We can easily make the two equations (144)–(145) look more similar by the following "square-root trick". Consider the function $R(z)$ defined by the following properties:

1. $R(z)^2 = (z-\alpha)(z-\beta)$ for all $z \in \mathbb{C}$.
2. $R(z)$ is analytic for $z \in \mathbb{C} \setminus [\alpha, \beta]$.
3. $R(z) = z + O(1)$ as $z \to \infty$.

In terms of the principal branch of the function $w^{1/2}$, we can even write a formula: $R(z) = (z-\alpha)^{1/2}(z-\beta)^{1/2}$. (Complex variables exercise: why is this not the same as $((z-\alpha)(z-\beta))^{1/2}$?) It is a consequence of this definition that

$$R_+(x) = R_-(x)\,, \qquad x \in \mathbb{R} \setminus [\alpha, \beta]\,, \qquad \text{while} \qquad R_+(x) = -R_-(x)\,, \qquad x \in [\alpha, \beta]\,. \tag{147}$$

Now make a change of variables by looking for $h(z) = g'(z)/R(z)$ instead of $g'(z)$ itself. Then $h(z)$ is analytic for $z \in \mathbb{C} \setminus \mathbb{R}$, and using (147) together with (144)–(145) we see that its boundary values satisfy

$$h_+(x) - h_-(x) = \begin{cases} \dfrac{V'(x)}{R_+(x)}\,, & x \in (\alpha, \beta) \\ \\ 0\,, & x \in \mathbb{R} \setminus [\alpha, \beta]\,. \end{cases} \tag{148}$$

Next, from the asymptotic condition (146) and the fact that $R(z) = z + O(1)$, we see that we must require that $h(z) = z^{-2} + O(z^{-3})$. In particular, $h(z)$ decays to zero as $z \to \infty$, so we may solve explicitly for $h(z)$ by the Plemelj formula:

$$h(z) = \frac{1}{2\pi i} \int_\alpha^\beta \frac{V'(x)\,dx}{R_+(x)(x-z)} = (\mathcal{C}^{[\alpha,\beta]}(V'/R_+))(z)\,. \tag{149}$$

Expanding this formula for large z gives

$$h(z) = -\frac{1}{2\pi i z} \int_\alpha^\beta \frac{V'(x)\,dx}{R_+(x)} - \frac{1}{2\pi i z^2} \int_\alpha^\beta \frac{xV'(x)\,dx}{R_+(x)} + O(z^{-3}) \tag{150}$$

as $z \to \infty$. Therefore the required decay at infinity requires

$$\int_\alpha^\beta \frac{V'(x)\,dx}{R_+(x)} = 0 \qquad \text{and} \qquad \int_\alpha^\beta \frac{xV'(x)\,dx}{R_+(x)} = -2\pi i \tag{151}$$

which are equations that determine the endpoints α and β of the support interval!

Further simplification for entire $V(x)$. This whole procedure simplifies if the convex function V is also an entire function. Indeed, we may first rewrite (149) as

$$h(z) = \frac{1}{4\pi i} \int_\alpha^\beta \frac{V'(x)\,dx}{R_+(x)(x-z)} + \frac{1}{4\pi i} \int_\alpha^\beta \frac{V'(x)\,dx}{R_+(x)(x-z)} = \frac{1}{4\pi i} \int_\alpha^\beta \frac{V'(x)\,dx}{R_+(x)(x-z)} - \frac{1}{4\pi i} \int_\alpha^\beta \frac{V'(x)\,dx}{R_-(x)(x-z)}\,, \tag{152}$$

and then through the sequence of contour deformations shown in Figure 7, we get

$$h(z) = \frac{V'(z)}{2R(z)} + \frac{1}{4\pi i} \oint_L \frac{V'(w)\,dw}{R(w)(w-z)} \tag{153}$$

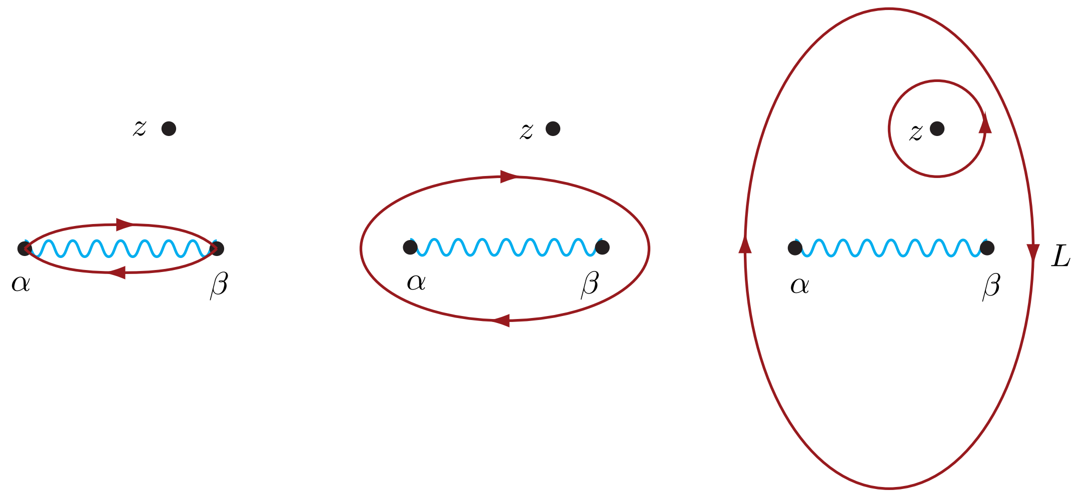

Figure 7: *Left to right: the steps of contour deformation for $h(z)$ when $V(x)$ is entire.*

where L is the large contour in the right-hand diagram of Figure 7. There are no singularities of the integrand outside of L, so the final integral may be calculated by residues at $w = \infty$. Indeed,

$$h(z) = \frac{V'(z)}{2R(z)} - \frac{1}{2} \operatorname*{Res}_{w=\infty} \frac{V'(w)}{R(w)(w-z)} \tag{154}$$

where the indicated residue is just the coefficient of w^{-1} in the Taylor/Laurent series expansion about $w = \infty$. Here the following expansions are useful:

$$\frac{1}{w-z} = w^{-1} + zw^{-2} + z^2w^{-3} + \cdots \tag{155}$$

and

$$\frac{1}{R(w)} = w^{-1} + \frac{1}{2}(\alpha+\beta)w^{-2} + \frac{1}{8}(3\alpha^2 + 2\alpha\beta + 3\beta^2)w^{-3} + \cdots . \tag{156}$$

Example: $V(x) = x^2$. This is convex and entire. We have the expansion

$$\begin{aligned} \frac{V'(w)}{R(w)(w-z)} &= 2w \cdot \left[\frac{1}{w} + \cdots\right] \cdot \left[\frac{1}{w} + \cdots\right] \\ &= \frac{2}{w} + \cdots , \end{aligned} \tag{157}$$

so the required residue is just 2, and therefore

$$h(z) = \frac{z}{R(z)} - 1 . \tag{158}$$

Expanding this formula at $z = \infty$ gives

$$h(z) = \frac{1}{2z}(\alpha+\beta) + \frac{1}{8z^2}(3\alpha^2 + 2\alpha\beta + 3\beta^2) + O(z^{-3}) , \tag{159}$$

so since we require that $h(z) = z^{-2} + O(z^{-3})$, the endpoints α and β are determined from the equations

$$\alpha + \beta = 0 \qquad \text{and} \qquad 3\alpha^2 + 2\alpha\beta + 3\beta^2 = 8 . \tag{160}$$

The unique solution with $\alpha \leq \beta$ is $\beta = -\alpha = \sqrt{2}$. With the endpoints known, we may find the equilibrium measure $\psi(x)$ by using (143) and

$$g'(z) = R(z)h(z) = z - R(z)\,. \tag{161}$$

since z has no jump on the real line and for $x \in [\alpha, \beta]$ we have $R_\pm(x) = \pm i\sqrt{(x-\alpha)(\beta-x)} = \pm i\sqrt{2-x^2}$, we get

$$\psi(x) = \frac{1}{\pi}\sqrt{2-x^2}\,, \qquad x \in [-\sqrt{2}, \sqrt{2}] \tag{162}$$

which in view of what is to come may be viewed as yet another (for this course) proof of the Wigner semicircle law.

Example: $V(x) = x^4$. This is convex and entire, and it is the first concrete example we have considered of a $V(x)$ that is outside the class of what can be handled by classical techniques. In this case, we have the expansion

$$\begin{aligned}\frac{V'(w)}{R(w)(w-z)} &= 4w^3 \cdot \left[\frac{1}{w} + \frac{z}{w^2} + \frac{z^2}{w^3} + \cdots\right] \cdot \left[\frac{1}{w} + \frac{1}{2w^2}(\alpha+\beta) + \frac{1}{8w^3}(3\alpha^2 + 2\alpha\beta + 3\beta^2) + \cdots\right] \\ &= 4w + [4z + 2\alpha + 2\beta] + \left[4z^2 + 2(\alpha+\beta)z + \frac{1}{2}(3\alpha^2 + 2\alpha\beta + 3\beta^2)\right]\frac{1}{w} + \cdots\,,\end{aligned} \tag{163}$$

from which we easily peel off the required residue. Therefore,

$$h(z) = \frac{2z^3}{R(z)} - 2z^2 - (\alpha+\beta)z - \frac{1}{4}(3\alpha^2 + 2\alpha\beta + 3\beta^2)\,. \tag{164}$$

Expanding this formula for large z gives

$$h(z) = \frac{1}{8z}(5\alpha^3 + 3\alpha^2\beta + 3\alpha\beta^2 + 5\beta^3) + \frac{1}{64z^2}(35\alpha^4 + 20\alpha^3\beta + 18\alpha^2\beta^2 + 20\alpha\beta^3 + 35\beta^4) + O(z^{-3}) \tag{165}$$

so since $5\alpha^3 + 3\alpha^2\beta + 3\alpha\beta^2 + 5\beta^3 = (5\alpha^2 - 2\alpha\beta + 5\beta^2)(\alpha+\beta)$, the equations that determine the endpoints in this case are

$$(5\alpha^2 - 2\alpha\beta + 5\beta^2)(\alpha+\beta) = 0 \qquad \text{and} \qquad 35\alpha^4 + 20\alpha^3\beta + 18\alpha^2\beta^2 + 20\alpha\beta^3 + 35\beta^4 = 64\,. \tag{166}$$

The only real solutions of the first equation correspond to $\alpha = -\beta$, and with this choice the second equation becomes simply $\beta^4 = 4/3$, so the support interval in this case is

$$[\alpha, \beta] = \left[-\left(\frac{4}{3}\right)^{1/4}, \left(\frac{4}{3}\right)^{1/4}\right]. \tag{167}$$

Finally, since

$$g'(z) = R(z)h(z) = 2z^3 - \left[2z^2 + \left(\frac{4}{3}\right)^{1/2}\right]R(z)\,, \tag{168}$$

we find the equilibrium measure for the case $V(x) = x^4$ to be:

$$\psi(x) = \frac{1}{\pi}\left[2x^2 + \left(\frac{4}{3}\right)^{1/2}\right]\sqrt{\left(\frac{4}{3}\right)^{1/2} - x^2}\,, \qquad x \in \left[-\left(\frac{4}{3}\right)^{1/4}, \left(\frac{4}{3}\right)^{1/4}\right]. \tag{169}$$

The density $\psi(x)$ is plotted along with the semicircle law for comparison in Figure 8. Note that unlike the semicircle law, the equilibrium measure for $V(x) = x^4$ is a bimodal distribution.

Final note: a similar procedure can be employed in the case of analytic but nonconvex V, but one needs to guess at the number of intervals of support and add "consistency tests" to the problem to confirm that the

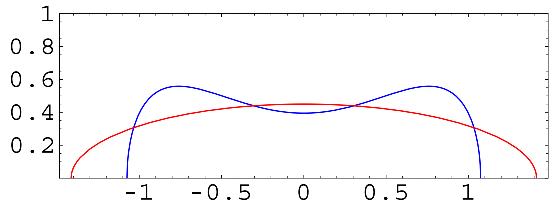

Figure 8: *Blue curve: the density $\psi(x)$ of the equilibrium measure for $V(x) = x^4$. Red curve: the Wigner semicircle law (equilibrium measure for $V(x) = x^2$).*

guess was correct (the number of support intervals is known to be finite in the analytic case). For nonanalytic but convex V the residue method does not apply but the integral formula (149) still holds "in place" on the real axis, leading to a formula for $\psi(x)$ as a principal-value Cauchy integral. One cannot carry out the deformation illustrated in Figure 7 without analyticity of $V(x)$. For nonconvex and nonanalytic V it is an open problem to formulate conditions under which the support consists of a finite number of intervals. There are concrete examples for which the number of support intervals is infinite (they shrink and accumulate at one or more points).

5.2.3 Effect of the equilibrium measure on the Riemann-Hilbert problem.

From now on we assume that the probability measure used to build the function $g(z)$ is not arbitrary, but rather is chosen to be the equilibrium measure associated with the function V. Then, in terms of quantities defined earlier, the jump condition for $\mathbf{B}^N(z)$ can be written in the form

$$\mathbf{B}^N_+(x) = \mathbf{B}^N_-(x) \begin{bmatrix} e^{-iN\theta(x)} & e^{-N\delta E/\delta\psi(x)} \\ 0 & e^{iN\theta(x)} \end{bmatrix}, \qquad x \in \mathbb{R}\,. \tag{170}$$

To see the relevance of the variational conditions (137)–(138), we need to bring in the Lagrange multiplier ℓ. Define a new unknown matrix $\mathbf{C}^N(z)$ in terms of $\mathbf{B}^n(z)$ by the formula

$$\mathbf{C}^N(z) := e^{N\ell\sigma_3/2}\mathbf{B}^N(z)e^{-N\ell\sigma_3/2}\,. \tag{171}$$

Then $\mathbf{C}^N(z)$ is easily seen (on the basis of the conditions on $\mathbf{B}^N(z)$ stemming from Riemann-Hilbert Problem 4) to satisfy the following slightly-modified Riemann-Hilbert problem:

Riemann-Hilbert Problem 5. *Find a 2×2 matrix-valued function $\mathbf{C}^N(z)$ with the following properties:*

> **Analyticity.** *$\mathbf{C}^N(z)$ is an analytic function of z for $z \in \mathbb{C} \setminus \mathbb{R}$ taking continuous boundary values $\mathbf{C}^N_\pm(x)$ as $z \to x \in \mathbb{R}$ from $\mathbb{C}_\pm$.*
>
> **Jump Condition.** *The boundary values are related by*
>
> $$\mathbf{C}^N_+(x) = \mathbf{C}^N_-(x) \begin{bmatrix} e^{-iN\theta(x)} & e^{N(\ell-\delta E/\delta\psi(x))} \\ 0 & e^{iN\theta(x)} \end{bmatrix}, \qquad x \in \mathbb{R}\,. \tag{172}$$
>
> **Normalization.** *The matrix $\mathbf{C}^N(z)$ is normalized at $z = \infty$ as follows:*
>
> $$\lim_{z\to\infty} \mathbf{C}^N(z) = \mathbb{I}\,. \tag{173}$$

Note that according to the definition (171) of $\mathbf{C}^N(z)$ in terms of $\mathbf{B}^N(z)$ which satisfies $\det(\mathbf{B}^N(z)) \equiv 1$, we have $\det(\mathbf{C}^N(z)) \equiv 1$.

Recalling the variational conditions (137)–(138) and the fact that by definition $\theta(x)$ is real-valued, we see that the main point of introducing the equilibrium measure is that all entries of the jump matrix for $\mathbf{C}^N(z)$ are bounded in modulus by 1. In other words, the exponential growth as $N \to \infty$ has been completely eliminated by the proper choice of $g(z)$. Bringing in the equilibrium measure has "stabilized" this Riemann-Hilbert problem as we will now see in detail.

5.3 Step 2: Steepest descent for Riemann-Hilbert problems.

To get even more explicit, we stick with our working assumption that V is a convex analytic function, which implies that the support of $\psi(x)$ is an interval $[\alpha, \beta]$ and that $\psi(x)$ is analytic in (α, β). Then, since

$$\theta(x) \equiv 2\pi \text{ for } x < \alpha \qquad \text{and} \qquad \theta(x) \equiv 0 \text{ for } x > \beta, \tag{174}$$

the jump condition (172) for $\mathbf{C}^N(z)$ takes the form

$$\mathbf{C}_+^N(x) = \mathbf{C}_-^N(x) \begin{bmatrix} 1 & e^{N(\ell - \delta E/\delta\psi(x))} \\ 0 & 1 \end{bmatrix}, \qquad x \in \mathbb{R} \setminus [\alpha, \beta] \tag{175}$$

which in view of the variational condition (138) involves an exponentially small perturbation of the identity matrix in the limit $N \to \infty$ as long as x is fixed. The variational condition (137) then says that

$$\mathbf{C}_+^N(x) = \mathbf{C}_-^N(x) \begin{bmatrix} e^{-iN\theta(x)} & 1 \\ 0 & e^{iN\theta(x)} \end{bmatrix}, \qquad x \in [\alpha, \beta], \tag{176}$$

and because $\psi(x) > 0$ in the support, $\theta(x)$ is a real analytic function that is strictly decreasing.

According to the Cauchy-Riemann equations, $e^{-iN\theta(z)}$ is exponentially small as $N \to \infty$ when z is fixed in the upper half-plane just above (α, β). Similarly $e^{iN\theta(z)}$ is exponentially small as $N \to \infty$ when z is fixed in the lower half-plane just below (α, β). Unfortunately, the jump matrix contains both exponentials, and if we deform the contour either up or down from the real axis we will introduce exponential growth in one jump matrix entry even though another will be exponentially small. And perhaps equally annoying is the fact that the 1 never goes away at all.

We need a way to separate all of these effects. The key is to notice the following factorization:

$$\begin{bmatrix} e^{-iN\theta(x)} & 1 \\ 0 & e^{iN\theta(x)} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ e^{iN\theta(x)} & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ e^{-iN\theta(x)} & 1 \end{bmatrix}. \tag{177}$$

This separates the various effects at least algebraically. But now we may make a change of variables that will allow us to convert oscillations on the real interval (α, β) into exponential decay on some nonreal contours. Consider the definition of a new unknown matrix $\mathbf{D}^N(z)$ in terms of $\mathbf{C}^N(z)$ as shown in Figure 9. The matrix $\mathbf{D}^N(z)$ defined in this way satisfies its own Riemann-Hilbert problem (direct calculation) relative to the oriented contour $\Sigma_{\mathbf{D}} = \mathbb{R} \cup L_+ \cup L_-$:

Riemann-Hilbert Problem 6. *Find a 2×2 matrix-valued function $\mathbf{D}^N(z)$ with the following properties:*

> **Analyticity.** $\mathbf{D}^N(z)$ *is an analytic function of z for $z \in \mathbb{C} \setminus \Sigma_{\mathbf{D}}$ taking continuous boundary values $\mathbf{D}_\pm^N(z)$ for $z \in \Sigma_{\mathbf{D}}$ from each side.*
>
> **Jump Condition.** *The boundary values are related by the following equations:*

$$\mathbf{D}_+^N(x) = \mathbf{D}_-^N(x) \begin{bmatrix} 1 & e^{N(\ell - \delta E/\delta\psi(x))} \\ 0 & 1 \end{bmatrix}, \qquad x \in \mathbb{R} \setminus [\alpha, \beta], \tag{178}$$

$$\mathbf{D}_+^N(x) = \mathbf{D}_-^N(x) \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \qquad x \in (\alpha, \beta), \tag{179}$$

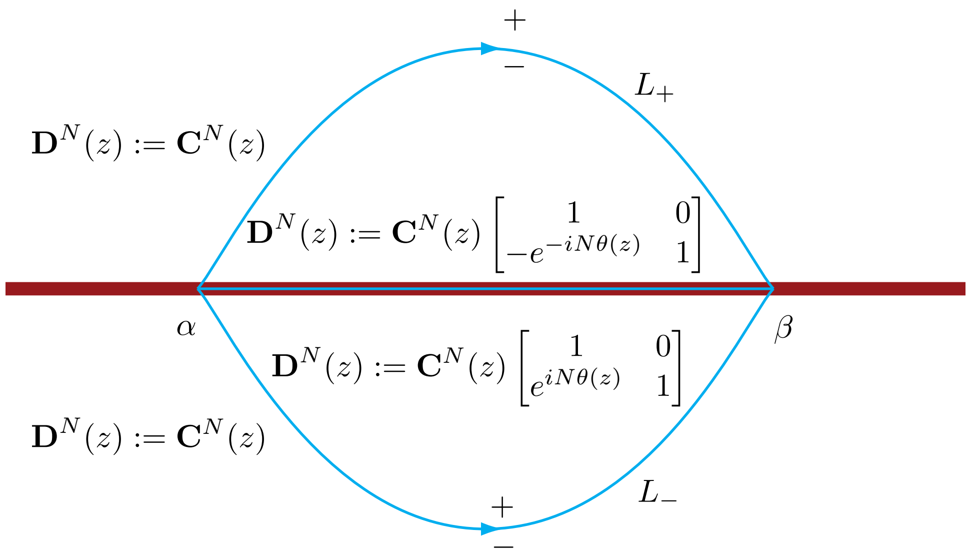

Figure 9: *The definition of* $\mathbf{D}^N(z)$ *in terms of* $\mathbf{C}^N(z)$ *and two "lens-shaped" regions bounded by contours* $L_\pm$*.*

$$\mathbf{D}_+^N(z) = \mathbf{D}_-^N(z) \begin{bmatrix} 1 & 0 \\ e^{\pm iN\theta(z)} & 1 \end{bmatrix}, \qquad z \in L_\mp\,. \tag{180}$$

Normalization. *The matrix* $\mathbf{D}^N(z)$ *is normalized at* $z = \infty$ *as follows:*

$$\lim_{z\to\infty} \mathbf{D}^N(z) = \mathbb{I}\,. \tag{181}$$

Note that by the explicit relation between $\mathbf{D}^N(z)$ and $\mathbf{C}^N(z)$ which satisfies $\det(\mathbf{C}^N(z)) \equiv 1$, we have $\det(\mathbf{D}^N(z)) \equiv 1$.

5.4 Step 3: Construction of a parametrix for $\mathbf{D}^N(z)$.

Here we will make a model for $\mathbf{D}^N(z)$, called a *parametrix*.

5.4.1 Outer parametrix.

It is a fact that as a consequence of our preparation, the jump matrix relating the boundary values of $\mathbf{D}^N(z)$ on the contour $\Sigma_{\mathbf{D}}$ has a pointwise limit as $N \to +\infty$. The only part of $\Sigma_{\mathbf{D}}$ on which the limit is not the identity matrix (no discontinuity of $\mathbf{D}^N(z)$ in the limit) is the support interval (α, β). This looks simple enough to solve explicitly.

Indeed, let us look for a 2×2 matrix-valued function $\hat{\mathbf{D}}_{\rm out}(z)$ that is analytic for $z \in \mathbb{C} \setminus [\alpha, \beta]$, tends to $\mathbb{I}$ as $z \to \infty$, and satisfies

$$\hat{\mathbf{D}}_{\rm out+}(x) = \hat{\mathbf{D}}_{\rm out-}(x) \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \qquad x \in (\alpha, \beta)\,. \tag{182}$$

The eigenvalues of the jump matrix are $\pm i$:

$$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} = \mathbf{U} \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix} \mathbf{U}^\dagger, \qquad \mathbf{U} := \begin{bmatrix} e^{-i\pi/4}/\sqrt{2} & e^{i\pi/4}/\sqrt{2} \\ e^{i\pi/4}/\sqrt{2} & e^{-i\pi/4}/\sqrt{2} \end{bmatrix}. \tag{183}$$

(Note that $\mathbf{U}$ is a unitary eigenvector matrix.) Therefore, the matrix $\mathbf{F}(z)$ related explicitly to $\hat{\mathbf{D}}_{\text{out}}(z)$ by the invertible similarity transformation

$$\mathbf{F}(z) := \mathbf{U}^\dagger \hat{\mathbf{D}}_{\text{out}}(z)\mathbf{U} \tag{184}$$

also should be analytic for $z \in \mathbb{C} \setminus [\alpha, \beta]$ and tend to $\mathbb{I}$ as $z \to \infty$, but the jump condition on (α, β) is now diagonal:

$$\mathbf{F}_+(x) = \mathbf{F}_-(x) \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}, \qquad x \in (\alpha, \beta)\,. \tag{185}$$

Let's seek $\mathbf{F}(z)$ as a diagonal matrix: $\mathbf{F}(z) = \text{diag}(F_1(z), F_2(z))$, where $F_j(z)$ is a scalar complex function analytic for $z \in \mathbb{C} \setminus [\alpha, \beta]$ with a limiting value of 1 at $z = \infty$ and satisfying

$$F_{1+}(x) = F_{1-}(x)e^{i\pi/2} \qquad \text{and} \qquad F_{2+}(x) = F_{2-}(x)e^{-i\pi/2}\,, \qquad x \in (\alpha, \beta)\,. \tag{186}$$

These conditions constitute a pair of *scalar* Riemann-Hilbert problems for $F_j(z)$.

Scalar Riemann-Hilbert problems can basically be handled by taking a logarithm (there are extra steps required if the "index" is nonzero, but that is not the case here). Indeed, if we write $F_j(z)$ in the form

$$F_j(z) = e^{L_j(z)}\,, \tag{187}$$

where we suppose that the $L_j(z)$ are analytic functions for $z \in \mathbb{C} \setminus [\alpha, \beta]$ that tend to *zero* as $z \to \infty$, then the jump conditions for $F_j(z)$ are converted from multiplicative form to additive form:

$$L_{1+}(x) - L_{1-}(x) = \frac{i\pi}{2} \qquad \text{and} \qquad L_{2+}(x) - L_{2-}(x) = -\frac{i\pi}{2}\,, \qquad x \in (\alpha, \beta)\,. \tag{188}$$

Suitable functions $L_j(z)$ can then be found with the help of the Plemelj formula:

$$L_1(z) = -L_2(z) = (\mathcal{C}^{[\alpha,\beta]}(i\pi/2))(z) = \frac{1}{4}\int_\alpha^\beta \frac{dx}{x - z} = \frac{1}{4}\log(z - \beta) - \frac{1}{4}\log(z - \alpha)\,. \tag{189}$$

These are principal branches of the logarithm. Therefore

$$F_1(z) = F_2(z)^{-1} = \gamma(z) := \left(\frac{z - \beta}{z - \alpha}\right)^{1/4}\,, \tag{190}$$

and consequently the matrix

$$\hat{\mathbf{D}}_{\text{out}}(z) = \mathbf{U}\gamma(z)^{\sigma_3}\mathbf{U}^\dagger \tag{191}$$

is a solution to the "pointwise asymptotic" of Riemann-Hilbert Problem 6. We call this function $\hat{\mathbf{D}}_{\text{out}}(z)$ the *outer parametrix* for $\mathbf{D}^N(z)$. Note that it does not depend on N at all, and it satisfies $\det(\hat{\mathbf{D}}_{\text{out}}(z)) \equiv 1$.

Note: it is this step of the analysis that becomes substantially more complicated if the equilibrium measure $\psi(x)\,dx$ associated with V turns out to have more than one support interval. In the general multi-interval case the outer parametrix must be constructed from the Riemann theta functions of a hyperelliptic Riemann surface of genus $G > 0$. The genus G is one less than the number of intervals of support. When $G > 0$, the full force of (19th century) algebraic geometry is required to solve for the outer parametrix $\hat{\mathbf{D}}_{\text{out}}(z)$.

5.4.2 Inner parametrices.

We might hope that $\hat{\mathbf{D}}_{\text{out}}(z)$ is a good approximation of $\mathbf{D}^N(z)$, but unfortunately this cannot possibly be true near the endpoints of $[\alpha, \beta]$ because the outer parametrix $\hat{\mathbf{D}}_{\text{out}}(z)$ blows up like a $-1/4$ power near these two points, while the analyticity condition of Riemann-Hilbert Problem 6 insists that $\mathbf{D}^N(z)$ should take continuous (and hence bounded for each N) boundary values.

So, we have to make a better "local" model for $\mathbf{D}^N(z)$ near $z = \alpha$ and $z = \beta$. For simplicity, we make the further assumption (generically true for analytic convex V) that the density $\psi(x)$ of the equilibrium measure vanishes exactly like a square root and no faster at $z = \alpha$ and $z = \beta$ (our explicit formulae show that this holds for $V(x) = x^2$ and $V(x) = x^4$, for example). We will give details of the construction near $z = \beta$; that near $z = \alpha$ is similar.

Writing the jump conditions for $\mathbf{D}^N(z)$ in a good form. The idea is to look carefully at the jump matrix for $\mathbf{D}^N(z)$ near $z = \beta$. The function $\theta(z)$ is positive and decreasing to zero for real $z < \beta$, and it vanishes like a multiple of $(\beta - z)^{3/2}$ according to our working assumption about $\psi(x)$. In fact, it can be shown that the function

$$w(z) := -\theta(z)^{2/3} \tag{192}$$

is an analytic function of z in a full complex neighborhood of $z = \beta$. But more is true: $w(\beta) = 0$ and $w'(\beta) > 0$, which means that the relation

$$w = w(z) \tag{193}$$

defines a *conformal mapping* between a neighborhood of $z = \beta$ and a neighborhood of $w = 0$. Next, consider the analytic function

$$F(z) := \ell - V(z) + 2g(z)\,, \qquad \Im\{z\} > 0\,. \tag{194}$$

Taking a boundary value on the real line for $x > \beta$ gives

$$\begin{aligned} F_+(x) &= \ell - V(x) + 2g_+(x) \\ &= \ell - V(x) + g_+(x) + g_-(x) \qquad (\text{because } g_+(x) = g_-(x) \text{ for } x > \beta) \\ &= \ell - \frac{\delta E}{\delta \psi}(x) \qquad x > \beta\,, \end{aligned} \tag{195}$$

while taking a boundary value on the support interval gives

$$\begin{aligned} F_+(x) &= \ell - V(x) + 2g_+(x) \\ &= g_+(x) - g_-(x) \qquad (\text{because } \ell - V(x) + g_+(x) + g_-(x) = 0 \text{ for } x \in (\alpha, \beta)) \\ &= i\theta(x)\,, \qquad x \in (\alpha, \beta) \\ &= i(-w)^{3/2}\,, \qquad w < 0\,. \end{aligned} \tag{196}$$

Therefore, by analytic continuation through the upper half-plane for w we may also write

$$\ell - \frac{\delta E}{\delta \psi}(x) = -w^{3/2}\,, \qquad w > 0\,. \tag{197}$$

In this way, *all jump matrices for $\mathbf{D}^N(z)$ in a neighborhood of $z = \beta$ can be written in terms of a single variable w*. Moreover, we can even scale out the N by setting

$$\zeta := N^{2/3} w\,, \tag{198}$$

upon which we find that the jump conditions for $\mathbf{D}^N(z)$ near $z = \beta$ can be written in the form

$$\mathbf{D}^N_+ = \mathbf{D}^N_- \begin{bmatrix} 1 & e^{-\zeta^{3/2}} \\ 0 & 1 \end{bmatrix}, \qquad \zeta > 0\,, \tag{199}$$

$$\mathbf{D}^N_+ = \mathbf{D}^N_- \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \qquad \zeta < 0\,, \tag{200}$$

and

$$\mathbf{D}^N_+ = \mathbf{D}^N_- \begin{bmatrix} 1 & 0 \\ e^{\zeta^{3/2}} & 1 \end{bmatrix}, \qquad \zeta \in N^{2/3} w(L_\pm)\,. \tag{201}$$

As the contours $L_\pm$ are somewhat arbitrary, we choose them so that in a fixed neighborhood of $z = \beta$ the images of $L_\pm$ under the conformal map w are segments meeting $w = 0$ at the angles $\pm 2\pi/3$. See Figure 10.

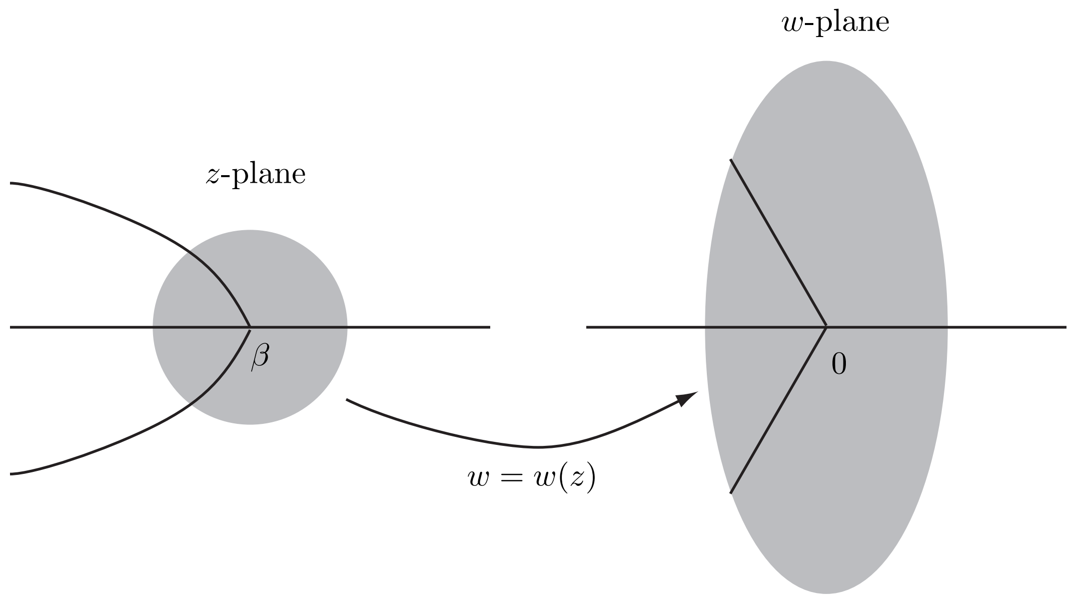

Figure 10: *The conformal map $w = w(z)$ maps a fixed disc centered at $z = \beta$ to a fixed neighborhood of the origin, straightening out the contours $L_\pm$.*

Matching condition. Local model Riemann-Hilbert problem. We will find an *exact* solution of these jump conditions to use as an inner parametrix for $\mathbf{D}^N(z)$ in a fixed z-neighborhood of $z = \beta$. We need to figure out how this mystery function of ζ should behave as $\zeta \to \infty$. For this, the condition we have in mind is that the inner parametrix we build should be a "match well" onto the outer parametrix. We can easily write the outer parametrix in terms of ζ too: since

$$\gamma(z) = \tilde{\gamma}(z) w^{1/4} = \tilde{\gamma}(z) N^{-1/6} \zeta^{1/4} , \tag{202}$$

where $\tilde{\gamma}(z)$ is a function analytic and nonvanishing in a fixed neighborhood of $z = \beta$, we may write $\hat{\mathbf{D}}_{\mathrm{out}}(z)$ in the form

$$\hat{\mathbf{D}}_{\mathrm{out}}(z) = \mathbf{H}^N(z) \zeta^{\sigma_3/4} \mathbf{U}^\dagger , \tag{203}$$

where

$$\mathbf{H}^N(z) := \mathbf{U} \tilde{\gamma}(z)^{\sigma_3} N^{-\sigma_3/6} \tag{204}$$

is a matrix factor that is analytic in a neighborhood of $z = \beta$. Temporarily ignoring the analytic prefactor (we'll put it back later) we seek a matrix-valued function of ζ that satisfies the exact jump conditions for $\mathbf{D}^N(z)$ and also matches well onto the remaining two factors of $\hat{\mathbf{D}}_{\mathrm{out}}(z)$ as $\zeta \to \infty$. Thus, we seek a matrix $\mathbf{Z}(\zeta)$ that satisfies a Riemann-Hilbert problem relative to a contour $\Sigma_{\mathbf{Z}}$ shown in Figure 11:

Riemann-Hilbert Problem 7. *Find a 2×2 matrix-valued function $\mathbf{Z}(\zeta)$ with the following properties:*

> **Analyticity.** $\mathbf{Z}(\zeta)$ *is analytic for $z \in \mathbb{C} \setminus \Sigma_{\mathbf{Z}}$ and takes continuous boundary values on $\Sigma_{\mathbf{Z}}$.*
>
> **Jump Condition.** *The boundary values are related as follows:*

$$\mathbf{Z}_+(\zeta) = \mathbf{Z}_-(\zeta) \begin{bmatrix} 1 & e^{-\zeta^{3/2}} \\ 0 & 1 \end{bmatrix} , \qquad \zeta > 0 , \tag{205}$$

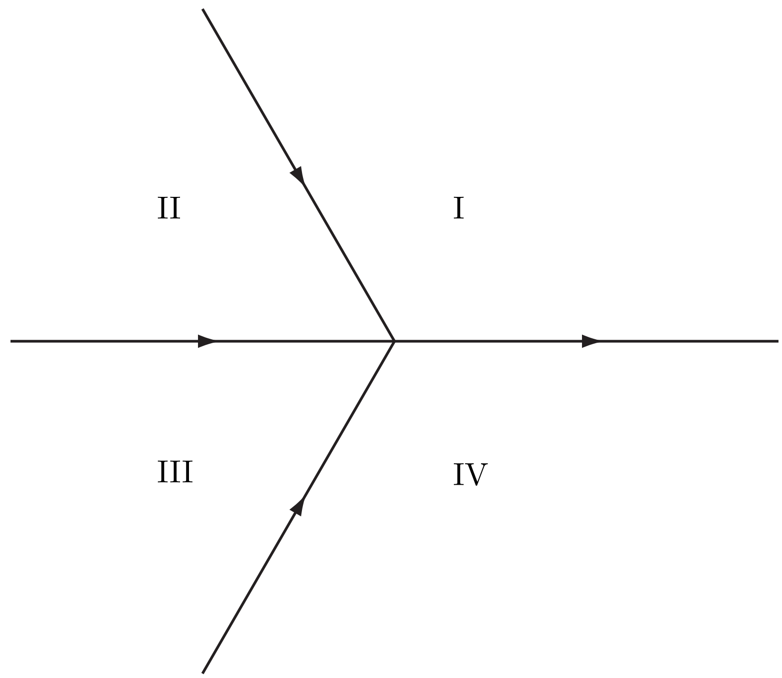

Figure 11: *The (infinite) contour $\Sigma_{\mathbf{Z}}$ in the ζ-plane.*

$$\mathbf{Z}_+(\zeta) = \mathbf{Z}_-(\zeta)\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \qquad \zeta < 0\,, \tag{206}$$

$$\mathbf{Z}_+(\zeta) = \mathbf{Z}_-(\zeta)\begin{bmatrix} 1 & 0 \\ e^{\zeta^{3/2}} & 1 \end{bmatrix}, \qquad \arg(\zeta) = \pm\frac{2\pi}{3}\,. \tag{207}$$

Normalization. *The matrix $\mathbf{Z}(\zeta)$ is normalized at $\zeta = \infty$ as follows:*

$$\lim_{\zeta\to\infty} \mathbf{Z}(\zeta)\mathbf{U}\zeta^{-\sigma_3/4} = \mathbb{I}\,. \tag{208}$$

It turns out that Riemann-Hilbert Problem 7 has a unique solution, and moreover, we may build it explicitly. This will require some use of special functions as we will now see. First, let's make the simple observation that if we make the substitution

$$\mathbf{W}(\zeta) = \mathbf{Z}(\zeta)e^{-\zeta^{3/2}\sigma_3/2} = \mathbf{Z}(\zeta)\begin{bmatrix} e^{-\zeta^{3/2}/2} & 0 \\ 0 & e^{\zeta^{3/2}/2} \end{bmatrix}, \tag{209}$$

then, miraculously, the jump conditions become piecewise constant:

$$\mathbf{W}_+(\zeta) = \mathbf{W}_-(\zeta)\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \qquad \zeta > 0\,, \tag{210}$$

$$\mathbf{W}_+(\zeta) = \mathbf{W}_-(\zeta)\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \qquad \zeta < 0\,, \tag{211}$$

and

$$\mathbf{W}_+(\zeta) = \mathbf{W}_-(\zeta)\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \qquad \arg(\zeta) = \pm\frac{2\pi}{3}\,. \tag{212}$$

The price we pay for this simplification is that the exponential factor appears in the normalization condition:

$$\lim_{\zeta\to\infty} \mathbf{W}(\zeta)e^{\zeta^{3/2}\sigma_3/2}\mathbf{U}\zeta^{-\sigma_3/4} = \mathbb{I}\,. \tag{213}$$

Finding $\mathbf{W}(\zeta)$ in terms of special functions. Now, we observe the following: since the jump matrices are all constant, the matrix $d\mathbf{W}/d\zeta$ also satisfies exactly the same jump conditions as does $\mathbf{W}$ itself. From this we can expect that the product

$$\mathbf{Y}(\zeta) := \frac{d\mathbf{W}}{d\zeta}(\zeta)\cdot\mathbf{W}(\zeta)^{-1} \tag{214}$$

is an entire function of ζ. If we suppose that the condition (213) holds in the sense that

$$\mathbf{W}(\zeta)e^{\zeta^{3/2}\sigma_3/2}\mathbf{U}\zeta^{-\sigma_3/4} = \mathbb{I} + \zeta^{-1}\mathbf{B} + o(\zeta^{-1}) \tag{215}$$

with some constant matrix $\mathbf{B}$ as $\zeta\to\infty$, with this relation being differentiable with respect to ζ, then we can easily extract the asymptotic behavior of $\mathbf{Y}(\zeta)$ in this limit. Indeed, we find

$$\mathbf{Y}(\zeta) = \frac{3i}{4}\begin{bmatrix} B_{21} & B_{22}-B_{11}-\zeta \\ 1 & -B_{21}\end{bmatrix} + o(1) \tag{216}$$

as $\zeta\to\infty$. But, since $\mathbf{Y}(\zeta)$ is entire, it follows from Liouville's Theorem that $\mathbf{Y}(\zeta)$ is exactly equal to its polynomial part at infinity:

$$\mathbf{Y}(\zeta) \equiv \frac{3i}{4}\begin{bmatrix} B_{21} & B_{22}-B_{11}-\zeta \\ 1 & -B_{21}\end{bmatrix} \tag{217}$$

Therefore, the matrix $\mathbf{W}(\zeta)$ satisfies the first-order differential equation

$$\frac{d\mathbf{W}}{d\zeta} = \frac{3i}{4}\begin{bmatrix} B_{21} & B_{22}-B_{11}-\zeta \\ 1 & -B_{21}\end{bmatrix}\mathbf{W}\,. \tag{218}$$

Setting

$$\xi = \left(\frac{3}{4}\right)^{2/3}(\zeta - c)\,, \qquad c := B_{21}^2 + B_{22} - B_{11}\,, \tag{219}$$

it follows by elimination that each element of the second row of $\mathbf{W}(\zeta)$ satisfies the second-order linear differential equation

$$\frac{d^2w_2}{d\xi^2} - \xi w_2 = 0\,. \tag{220}$$

The differential equation (220) is called the *Airy equation* and its solutions are special functions generally called *Airy functions*. One particular solution of the Airy equation is given by a contour integral (obtained by solving the Airy equation by Laplace transform):

$$w_2 = Ai(\xi) := \frac{1}{2\pi i}\int_C e^{\xi z - z^3/3}\,dz \tag{221}$$

where C is any contour in the complex z-plane that begins at infinity with $\arg(z) = -2\pi i/3$ and ends at infinity with $\arg(z) = 2\pi i/3$. This solution is called the *Airy function* (strangely, the other solutions are also Airy functions, but only $Ai(\xi)$ gets to be called *the* Airy function) and is real-valued for real ξ. A plot of $Ai(\xi)$ and its derivative $Ai'(\xi)$ for real ξ is shown in Figure 12. Given that $Ai(\xi)$ is a solution, it is easy to check that $Ai(\xi e^{\pm 2\pi i/3})$ are also solutions; however only two of these three are linearly independent because by Cauchy's Theorem we can easily deduce that

$$Ai(\xi) + e^{-2\pi i/3}Ai(\xi e^{-2\pi i/3}) + e^{2\pi i/3}Ai(\xi e^{2\pi i/3}) \equiv 0 \tag{222}$$

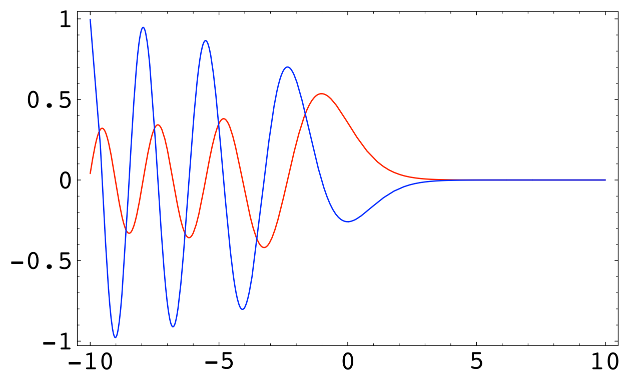

Figure 12: *The Airy function $Ai(\xi)$ (red) and its derivative $Ai'(\xi)$ (blue) for real ξ.*

(this is the only independent linear relation among the three solutions). By the (classical) method of steepest descent for integrals,

$$Ai(\xi) = \frac{1}{2\xi^{1/4}\sqrt{\pi}}e^{-2\xi^{3/2}/3}(1 + O(|\xi|^{-3/2})) \qquad Ai'(\xi) = -\frac{\xi^{1/4}}{2\sqrt{\pi}}e^{-2\xi^{3/2}/3}(1 + O(|\xi|^{-3/2})) \tag{223}$$

as $\xi \to \infty$ in any direction of the complex plane *except* in the negative real direction. (There are other formulae that hold for negative ξ going to infinity; the apparent discontinuity along the negative real ξ-axis is "fake" and introduced by the asymptotics because *all solutions of the Airy equation are entire functions of* ξ. This phenomenon is typical for a big class of differential equations and is called *Stokes' phenomenon.*)

Let's first construct $\mathbf{W}(\zeta)$ in the region $0 < \arg(\zeta) < 2\pi/3$ (labelled "I" in Figure 11) using Airy functions. As a basis of solutions of the Airy equation we choose $w_2 = Ai(\xi)$ and $w_2 = Ai(\xi e^{-2\pi i/3})$, since our asymptotic formulae (223) hold in each case as $\zeta \to \infty$ in this sector. The second row of the general solution of the system (218) may then be written in the form

$$\begin{bmatrix} W_{21}(\zeta) & W_{22}(\zeta) \end{bmatrix} = \begin{bmatrix} a_1 Ai(\xi) + b_1 Ai(\xi e^{-2\pi i/3}) & a_2 Ai(\xi) + b_2 Ai(\xi e^{-2\pi i/3}) \end{bmatrix} \tag{224}$$

for some constants a_1, a_2, b_1, and b_2 that are free to be chosen to satisfy any auxiliary conditions. Now, we use (223) to calculate the large-ζ behavior of the product

$$\begin{aligned} \mathbf{e}(\zeta)^T &:= \begin{bmatrix} W_{21}(\zeta) & W_{22}(\zeta) \end{bmatrix} e^{\zeta^{3/2}\sigma_3/2}\mathbf{U}\zeta^{-\sigma_3/4} \\ &= \frac{1}{2\sqrt{2\pi}}\left(\frac{4}{3}\right)^{1/6} \cdot \Big[a_1 e^{-i\pi/4} e^{3c\zeta^{1/2}/4}\zeta^{-1/2}(1+o(1)) + b_1 e^{-i\pi/12} e^{\zeta^{3/2}} e^{-3c\zeta^{1/2}/4}\zeta^{-1/2}(1+o(1)) \\ &\qquad + a_2 e^{i\pi/4} e^{-\zeta^{3/2}} e^{3c\zeta^{1/2}/4}\zeta^{-1/2}(1+o(1)) + b_2 e^{5\pi i/12} e^{-3c\zeta^{1/2}/4}\zeta^{-1/2}(1+o(1)) \,, \\ &\quad a_1 e^{i\pi/4} e^{3c\zeta^{1/2}/4}(1+o(1)) + b_1 e^{5\pi i/12} e^{\zeta^{3/2}} e^{-3c\zeta^{1/2}/4}(1+o(1)) \\ &\qquad + a_2 e^{-i\pi/4} e^{-\zeta^{3/2}} e^{3c\zeta^{1/2}/4}(1+o(1)) + b_2 e^{-i\pi/12} e^{-3c\zeta^{1/2}/4}(1+o(1)) \Big] \,, \end{aligned} \tag{225}$$

as $\zeta \to \infty$ with $0 < \arg(\zeta) < 2\pi/3$. According to the normalization condition imposed on $\mathbf{W}(\zeta)$ we need to arrive at a result of the form

$$\mathbf{e}(\zeta)^T = \begin{bmatrix} 0 & 1 \end{bmatrix} + \zeta^{-1}\begin{bmatrix} B_{21} & B_{22} \end{bmatrix} + o(\zeta)^{-1} \tag{226}$$

as $\zeta \to \infty$ with $0 < \arg(\zeta) < 2\pi/3$. Equating these two expressions has several implications:

1. The terms proportional to $e^{\pm\zeta^{3/2}}$ grow exponentially as $\zeta \to \infty$ in different parts of the sector $0 < \arg(\zeta) < 2\pi/3$, so these terms are inconsistent with the desired asymptotics. The only way to resolve this difficulty is to make the choice that

$$b_1 = 0 \qquad \text{and} \qquad a_2 = 0\,. \tag{227}$$

2. The next troublesome terms are those proportional to $e^{\pm 3c\zeta^{1/2}/4}$ as these terms all grow exponentially as $\zeta \to \infty$ somewhere in the sector $0 < \arg(\zeta) < 2\pi/3$ *unless* $c = 0$. We could remove these terms by choosing the remaining coefficients a_1 and b_2 to vanish, but this would leave us with $\mathbf{e}(\zeta)^T \equiv 0$, which also is inconsistent with the desired asymptotics. Therefore, we are forced to accept the condition that

$$c = B_{21}^2 + B_{22} - B_{11} = 0\,. \tag{228}$$

With $b_1 = a_2 = c = 0$, the asymptotic formula for $\mathbf{e}(\zeta)^T$ simplifies and the error estimates become more precise:

$$\mathbf{e}(\zeta)^T = \frac{1}{2\sqrt{2\pi}}\left(\frac{4}{3}\right)^{1/6}\left[(e^{-i\pi/4}a_1 + e^{5\pi i/12}b_2)\zeta^{-1/2} + O(|\zeta|^{-2})\,,\quad (e^{i\pi/4}a_1 + e^{-i\pi/12}b_2) + O(|\zeta|^{-3/2})\right]. \tag{229}$$

3. The final "offensive" term is the one proportional to $\zeta^{-1/2}$ as this is large compared with the desired rate of decay of ζ^{-1}. Therefore, we make the choice that

$$e^{-i\pi/4}a_1 + e^{5\pi i/12}b_2 = 0\,. \tag{230}$$

Then, to get the normalization constant correct, we make the choice that

$$e^{i\pi/4}a_1 + e^{-i\pi/12}b_2 = 2\sqrt{2\pi}\left(\frac{3}{4}\right)^{1/6}. \tag{231}$$

Taken together, equations (230) and (231) imply that

$$a_1 = e^{-i\pi/4}\sqrt{2\pi}\left(\frac{3}{4}\right)^{1/6} \qquad \text{and} \qquad b_2 = e^{i\pi/12}\sqrt{2\pi}\left(\frac{3}{4}\right)^{1/6}. \tag{232}$$

Therefore, we have determined the second row of $\mathbf{W}(\zeta)$ in the sector $0 < \arg(\zeta) < 2\pi/3$:

$$\begin{bmatrix} W_{21}(\zeta) & W_{22}(\zeta) \end{bmatrix} = \left[e^{-i\pi/4}\sqrt{2\pi}\left(\tfrac{3}{4}\right)^{1/6} Ai((\tfrac{3}{4})^{2/3}\zeta) \quad e^{i\pi/12}\sqrt{2\pi}\left(\tfrac{3}{4}\right)^{1/6} Ai((\tfrac{3}{4})^{2/3}\zeta e^{-2\pi i/3}) \right], \tag{233}$$

and we have also learned (because the error terms in (229) are $o(\zeta^{-1})$ in each case) that

$$B_{21} = B_{22} = 0\,. \tag{234}$$

Now, since $B_{21} = 0$, the second equation of the system (218) implies that

$$\begin{aligned} \begin{bmatrix} W_{11}(\zeta) & W_{12}(\zeta) \end{bmatrix} &= \begin{bmatrix} -\tfrac{4i}{3}W_{21}'(\zeta) & -\tfrac{4i}{3}W_{22}'(\zeta) \end{bmatrix} \\ &= \left[e^{-3\pi i/4}\sqrt{2\pi}\left(\tfrac{4}{3}\right)^{1/6} Ai'((\tfrac{3}{4})^{2/3}\zeta) \quad e^{11\pi i/12}\sqrt{2\pi}\left(\tfrac{4}{3}\right)^{1/6} Ai'((\tfrac{3}{4})^{2/3}\zeta e^{-2\pi i/3}) \right]. \end{aligned} \tag{235}$$

Applying the asymptotic formulae (223) shows that

$$\begin{bmatrix} W_{11}(\zeta) & W_{12}(\zeta) \end{bmatrix} e^{\zeta^{3/2}\sigma_3/2}\mathbf{U}\zeta^{-\sigma_3/4} = \begin{bmatrix} 1 + O(|\zeta|^{-3/2})\,, & O(|\zeta|^{-1}) \end{bmatrix} \tag{236}$$

as $\zeta \to \infty$ with $0 < \arg(\zeta) < 2\pi/3$. This confirms that $B_{11} = 0$ (which together with (234) is consistent with (228)) and shows that the only element of $\mathbf{B}$ that might not be zero is B_{12}.

The result of our work is that

$$\mathbf{W}(\zeta)=\mathbf{W}_{\mathrm{I}}(\zeta):=\begin{bmatrix} e^{-3\pi i/4}\sqrt{2\pi}\left(\frac{4}{3}\right)^{1/6}Ai'((\frac{3}{4})^{2/3}\zeta) & e^{11\pi i/12}\sqrt{2\pi}\left(\frac{4}{3}\right)^{1/6}Ai'((\frac{3}{4})^{2/3}\zeta e^{-2\pi i/3}) \\ e^{-i\pi/4}\sqrt{2\pi}\left(\frac{3}{4}\right)^{1/6}Ai((\frac{3}{4})^{2/3}\zeta) & e^{i\pi/12}\sqrt{2\pi}\left(\frac{3}{4}\right)^{1/6}Ai((\frac{3}{4})^{2/3}\zeta e^{-2\pi i/3}) \end{bmatrix} \tag{237}$$

for $0<\arg(\zeta)<2\pi/3$. This completes the construction of $\mathbf{W}(\zeta)$ in the sector labeled "I" in Figure 11.

To build $\mathbf{W}(\zeta)$ in the other three regions of the complex plane, it is easiest to take advantage of the constant jump conditions this matrix satisfies. For example, to get $\mathbf{W}(\zeta)$ in the sector $2\pi/3<\arg(\zeta)<\pi$ (labeled "II" in Figure 11), use the jump for $\mathbf{W}(\zeta)$ between sectors I and II to get

$$\mathbf{W}_{\mathrm{II}}(\zeta)=\mathbf{W}_{\mathrm{I}}(\zeta)\begin{bmatrix}1 & 0\\ 1 & 1\end{bmatrix}^{-1} \tag{238}$$

and then use the identity (222) to write the result in terms of $Ai((\frac{3}{4})^{2/3}\zeta e^{\pm 2\pi i/3})$ and their derivatives (this basis of solutions has no Stokes phenomenon in the sector of interest). The result is that

$$\mathbf{W}(\zeta)=\mathbf{W}_{\mathrm{II}}(\zeta):=\begin{bmatrix} e^{-5\pi i/12}\sqrt{2\pi}\left(\frac{4}{3}\right)^{1/6}Ai'(\left(\frac{3}{4}\right)^{2/3}\zeta e^{2\pi i/3}) & e^{11\pi i/12}\sqrt{2\pi}\left(\frac{4}{3}\right)^{1/6}Ai'((\frac{3}{4})^{2/3}\zeta e^{-2\pi i/3}) \\ e^{-7\pi i/12}\sqrt{2\pi}\left(\frac{3}{4}\right)^{1/6}Ai(\left(\frac{3}{4}\right)^{2/3}\zeta e^{2\pi i/3}) & e^{i\pi/12}\sqrt{2\pi}\left(\frac{3}{4}\right)^{1/6}Ai((\frac{3}{4})^{2/3}\zeta e^{-2\pi i/3}) \end{bmatrix} \tag{239}$$

holds for $2\pi/3<\arg(\zeta)<\pi$.

Moving in this way from one sector to the next using the constant jump conditions for $\mathbf{W}(\zeta)$ gives

$$\mathbf{W}(\zeta)=\mathbf{W}_{\mathrm{IV}}(\zeta):=\begin{bmatrix} e^{-3\pi i/4}\sqrt{2\pi}\left(\frac{4}{3}\right)^{1/6}Ai'(\left(\frac{3}{4}\right)^{2/3}\zeta) & e^{7\pi i/12}\sqrt{2\pi}\left(\frac{4}{3}\right)^{1/6}Ai'(\left(\frac{3}{4}\right)^{2/3}\zeta e^{2\pi i/3}) \\ e^{-i\pi/4}\sqrt{2\pi}\left(\frac{3}{4}\right)^{1/6}Ai(\left(\frac{3}{4}\right)^{2/3}\zeta) & e^{5\pi i/12}\left(\frac{3}{4}\right)^{1/6}Ai(\left(\frac{3}{4}\right)^{2/3}\zeta e^{2\pi i/3}) \end{bmatrix} \tag{240}$$

holds for $-2\pi/3<\arg(\zeta)<0$ (in the sector labeled "IV" in Figure 11). Finally,

$$\mathbf{W}(\zeta)=\mathbf{W}_{\mathrm{III}}(\zeta):=\begin{bmatrix} e^{11\pi i/12}\sqrt{2\pi}\left(\frac{4}{3}\right)^{1/6}Ai'(\left(\frac{3}{4}\right)^{2/3}\zeta e^{-2\pi i/3}) & e^{7\pi i/12}\sqrt{2\pi}\left(\frac{4}{3}\right)^{1/6}Ai'(\left(\frac{3}{4}\right)^{2/3}\zeta e^{2\pi i/3}) \\ e^{i\pi/12}\sqrt{2\pi}\left(\frac{3}{4}\right)^{1/6}Ai(\left(\frac{3}{4}\right)^{2/3}\zeta e^{-2\pi i/3}) & e^{5\pi i/12}\left(\frac{3}{4}\right)^{1/6}Ai(\left(\frac{3}{4}\right)^{2/3}\zeta e^{2\pi i/3}) \end{bmatrix} \tag{241}$$

holds for $-\pi<\arg(\zeta)<-2\pi/3$ (in the sector labeled "III" in Figure 11.

It is easy to verify from these formulae that as $\zeta\to\infty$ in every direction of the complex plane,

$$\mathbf{W}(\zeta)e^{\zeta^{3/2}\sigma_3/2}\mathbf{U}\zeta^{-\sigma_3/4}=\mathbb{I}+\begin{bmatrix}O(|\zeta|^{-3/2}) & O(|\zeta|^{-1})\\ O(|\zeta|^{-2}) & O(|\zeta|^{-3/2})\end{bmatrix}. \tag{242}$$

also, it can be shown (this requires knowing the Wronskians of pairs of $Ai(\xi)$, $Ai(\xi e^{2\pi i/3})$, and $Ai(\xi e^{-2\pi i/3})$) that $\det(\mathbf{W}(\zeta))\equiv 1$.

The inner parametrix for $\mathbf{D}^N(z)$. As our model for $\mathbf{D}^N(z)$ near $z=\beta$, we set

$$\hat{\mathbf{D}}^N_\beta(z):=\mathbf{H}^N(z)\mathbf{Z}(N^{2/3}w(z))=\mathbf{H}^N(z)\mathbf{W}(N^{2/3}w(z))e^{Nw(z)^{3/2}\sigma_3/2}. \tag{243}$$

A similar procedure, carried out near $z=\alpha$, results in a second inner parametrix $\hat{\mathbf{D}}^N_\alpha(z)$. Both of these matrices have determinants equal to 1.

5.4.3 Global parametrix.

Let U_α and U_β be two small discs of radius independent of N that are centered at $z=\alpha$ and $z=\beta$ respectively. We are now ready to define our global parametrix, a matrix-valued function of z that we expect to be a good approximation to $\mathbf{D}^N(z)$ uniformly in the complex plane. We now set

$$\hat{\mathbf{D}}^N(z):=\begin{cases}\hat{\mathbf{D}}_{\mathrm{out}}(z), & z\in\mathbb{C}\setminus(U_\alpha\cup U_\beta)\\ \hat{\mathbf{D}}^N_\alpha(z), & z\in U_\alpha\\ \hat{\mathbf{D}}^N_\beta(z), & z\in U_\beta.\end{cases} \tag{244}$$

This matrix satisfies $\det(\hat{\mathbf{D}}^N(z)) \equiv 1$ at each point of its definition in the complex plane.

5.5 Step 4: Comparison of $\mathbf{D}^N(z)$ with its parametrix. Arrival at near-identity target Riemann-Hilbert problem.

The global parametrix $\hat{\mathbf{D}}^N(z)$ has now been explicitly constructed, and we may compare it with the (unknown) solution $\mathbf{D}^N(z)$ of Riemann-Hilbert Problem 6 by defining a new unknown $\mathbf{E}^N(z)$ in terms of $\mathbf{D}^N(z)$:

$$\mathbf{E}^N(z) := \mathbf{D}^N(z)\hat{\mathbf{D}}^N(z)^{-1} . \tag{245}$$

Note that this relation implies that $\det(\mathbf{E}^N(z)) \equiv 1$. Moreover, by the general equivalence principles described earlier, the matrix $\mathbf{E}^N(z)$ satisfies a Riemann-Hilbert problem because $\mathbf{D}^N(z)$ does and because the global parametrix is given. The oriented contour $\Sigma_{\mathbf{E}}$ on which $\mathbf{E}^N(z)$ has discontinuities is shown in Figure 13. The fact that all of the contours within the discs U_α and U_β as well as the rest of the support

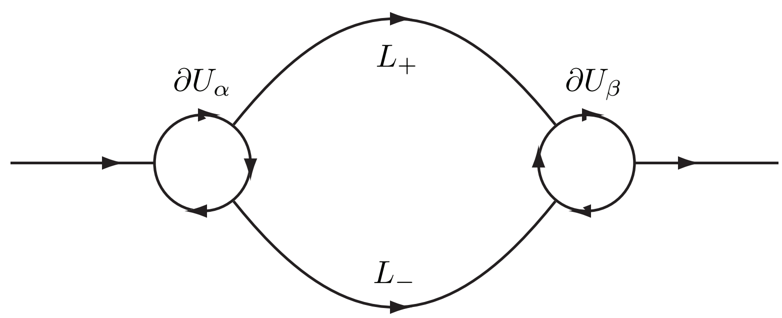

Figure 13: *The contour $\Sigma_{\mathbf{E}}$ of discontinuity for $\mathbf{E}^N(z)$. This contour is independent of N.*

interval (α, β) have "disappeared" follows from the facts that

1. The outer parametrix $\hat{\mathbf{D}}_{\text{out}}(z)$ satisfies *exactly* the same jump condition on the support interval as does $\mathbf{D}^N(z)$.
2. The inner parametrix $\hat{\mathbf{D}}_\beta^N(z)$ satisfies *exactly* the same jump conditions on all four contours within U_β as does $\mathbf{D}^N(z)$. A similar statement holds for $\hat{\mathbf{D}}_\alpha^N(z)$.

Furthermore, it is a consequence of our definition of the global parametrix that the Riemann-Hilbert problem satisfied by $\mathbf{E}^N(z)$ is a near-identity problem in the limit $N \to +\infty$. This Riemann-Hilbert problem has the following form:

Riemann-Hilbert Problem 8. *Find a 2×2 matrix-valued function $\mathbf{E}^N(z)$ with the following properties:*

> **Analyticity.** *$\mathbf{E}^N(z)$ is an analytic function for $z \in \mathbb{C} \setminus \Sigma_{\mathbf{E}}$, taking continuous boundary values $\mathbf{E}_\pm^N(z)$ from each side of $\Sigma_{\mathbf{E}}$.*
>
> **Jump Condition.** *The boundary values are related as follows. At each point z of $\Sigma \setminus (\partial U_\alpha \cup \partial U_\beta)$,*
>
> $$\mathbf{E}_+^N(z) = \mathbf{E}_-^N(z) \left(\mathbb{I} + \textit{uniformly exponentially small as } N \to +\infty\right) , \tag{246}$$
>
> *and at each point z of ∂U_α or ∂U_β,*
>
> $$\mathbf{E}_+^N(z) = \mathbf{E}_-^N(z) \left(\mathbb{I} + O(N^{-1})\right) . \tag{247}$$

Normalization. *The matrix $\mathbf{E}^N(z)$ is normalized at $z = \infty$ as follows:*

$$\lim_{z\to\infty} \mathbf{E}^N(z) = \mathbb{I}\,. \tag{248}$$

Let's see why the jump matrix is a small perturbation of the identity matrix, as claimed above. To prove (246), just note that on these parts of $\Sigma_{\mathbf{E}}$, the jump matrix for $\mathbf{D}^N(z)$ is already a small perturbation of $\mathbb{I}$ while the outer parametrix $\hat{\mathbf{D}}_{\rm out}(z)$ and its inverse are uniformly bounded (because z is bounded away from $z = \alpha$ and $z = \beta$ on these parts of $\Sigma_{\mathbf{E}}$) and independent of N. And to prove (247), note that there is *no discontinuity* of $\mathbf{D}^N(z)$ on the two circles ∂U_α and ∂U_β, so the mismatch just comes from the discontinuity of the parametrix. On ∂U_β with "+" indicating the outside and "−" the inside:

$$\mathbf{E}_+(z) = \mathbf{D}^N(z)\hat{\mathbf{D}}_{\rm out}(z)^{-1} = \mathbf{D}^N(z)\hat{\mathbf{D}}_\beta^N(z)^{-1}\cdot\hat{\mathbf{D}}_\beta^N(z)\hat{\mathbf{D}}_{\rm out}(z)^{-1} = \mathbf{E}_-(z)\hat{\mathbf{D}}_\beta^N(z)\hat{\mathbf{D}}_{\rm out}(z)^{-1}\,, \tag{249}$$

and the jump matrix is

$$\begin{aligned}\hat{\mathbf{D}}_\beta^N(z)\hat{\mathbf{D}}_{\rm out}(z)^{-1} &= \mathbf{H}^N(z)\mathbf{W}(N^{2/3}w(z))e^{-Nw(z)^{3/2}\sigma_3/2}\cdot\mathbf{U}\zeta^{-\sigma_3/4}\mathbf{H}^N(z)^{-1}\\ &= \mathbf{H}^N(z)\left(\mathbb{I} + \begin{bmatrix} O(N^{-1}) & O(N^{-2/3})\\ O(N^{-4/3}) & O(N^{-1})\end{bmatrix}\right)\mathbf{H}^N(z)^{-1}\\ &= \mathbf{U}\tilde{\gamma}(z)^{\sigma_3/4}(\mathbb{I}+O(N^{-1}))\tilde{\gamma}(z)^{-\sigma_3/4}\mathbf{U}^\dagger\\ &= \mathbb{I} + O(N^{-1})\,.\end{aligned} \tag{250}$$

Thus, $\mathbf{E}^N(z)$ satisfies a near-identity Riemann-Hilbert problem, so by series solution of the system of associated integral equations, we learn that for each fixed z not on the contour $\Sigma_{\mathbf{E}}$,

$$\mathbf{E}^N(z) = \mathbb{I} + O(N^{-1}) \tag{251}$$

as $N \to +\infty$. This implies that the unknown $\mathbf{D}^N(z)$ is close to the parametrix $\hat{\mathbf{D}}^N(z)$, because

$$\mathbf{D}^N(z) = \mathbf{E}^N(z)\hat{\mathbf{D}}^N(z) = (\mathbb{I}+O(N^{-1}))\hat{\mathbf{D}}^N(z)\,. \tag{252}$$

Furthermore, as $\mathbf{A}^N(z)$ is related to $\mathbf{D}^N(z)$ by explicit steps, we obtain at last asymptotically valid formulae for all of the matrix elements of $\mathbf{A}^N(z)$. In particular, we obtain the first column from which the kernel $K_N(x,y)$ can be extracted. For example, if we suppose z to lie in the upper half-plane just above a point in (α,β), then by unraveling the relation between $\mathbf{A}^N(z)$ and $\mathbf{D}^N(z)$ in this region,

$$\mathbf{A}^N(z) = e^{-N\ell\sigma_3/2}\mathbf{E}^N(z)\hat{\mathbf{D}}_{\rm out}(z)\begin{bmatrix}1 & 0\\ e^{-iN\theta(z)} & 1\end{bmatrix}e^{N(g(z)+\ell/2)\sigma_3}\,, \tag{253}$$

where we recall that $\hat{\mathbf{D}}_{\rm out}(z)$ denotes the outer parametrix obtained explicitly in §5.4.1.

Note: in fact we obtain asymptotic information about derivatives of $\mathbf{A}^N(z)$ because if C is a positively oriented circle of radius R centered at z and C does not intersect $\Sigma_{\mathbf{E}}$, then by Cauchy's formula

$$\frac{d^k\mathbf{E}^N}{dz^k}(z) = \frac{k!}{2\pi i}\oint_C (w-z)^{-(k+1)}\mathbf{E}^N(w)\,dw = \frac{k!}{2\pi i}\oint_C (w-z)^{-(k+1)}(\mathbf{E}^N(w)-\mathbb{I})\,dw = O(N^{-1})\,. \tag{254}$$

(The constant in the bound depends on k and R.)

6 Implications for Random Matrix Theory.

The asymptotic formulae for $K_N(x,y)$ obtained by "Riemann-Hilbert" analysis show that for completely arbitrary convex analytic $V(x)$ (and in fact an even broader class of potentials),

1. The one-point function satisfies
$$R_N^{(1)}(x) = N\psi(x) + O(1) \tag{255}$$
as $N \to +\infty$, where $\psi(x)$ is the density of the equilibrium measure associated with the potential $V(x)$. As this depends on V, the limiting distribution of eigenvalues is *not* universal. One only gets the Wigner semicircle law if $V(x)$ is proportional to x^2.

   The proof of (255) begins with the exact formula
$$R_N^{(1)}(x) = -\frac{e^{-NV(x)}}{2\pi i}\left[A_{11}^{N\prime}(x)A_{21}^N(x) - A_{11}^N(x)A_{21}^{N\prime}(x)\right] . \tag{256}$$
   Note that taking a limit from the upper half-plane in (253),
$$\begin{aligned} A_{11+}^N(x) &= e^{Ng_+(x)}(\mathbf{E}^N\hat{\mathbf{D}}_{\text{out}})_{11+}(x) + e^{Ng_+(x)}e^{-iN\theta(x)}(\mathbf{E}^N\hat{\mathbf{D}}_{\text{out}})_{12+}(x) , \\ A_{21+}^N(x) &= e^{N(V(x)-g_+(x))}e^{iN\theta(x)}(\mathbf{E}^N\hat{\mathbf{D}}_{\text{out}})_{21+}(x) + e^{N(V(x)-g_+(x))}(\mathbf{E}^N\hat{\mathbf{D}}_{\text{out}})_{22+}(x) . \end{aligned} \tag{257}$$
   Differentiation of these formulae, using only the fact that derivatives of $\mathbf{E}^N(z)\hat{\mathbf{D}}_{\text{out}}(z)$ remain uniformly bounded as $N \to +\infty$, gives
$$\begin{aligned} A_{11+}^{N\prime}(x) &= Ne^{Ng_+(x)}\left[g_+'(x)(\mathbf{E}^N\hat{\mathbf{D}}_{\text{out}})_{11+}(x) + (g_+'(x) - i\theta'(x))e^{-iN\theta(x)}(\mathbf{E}^N\hat{\mathbf{D}}_{\text{out}})_{12+}(x) + O(N^{-1})\right] , \\ A_{21+}^{N\prime}(x) &= Ne^{N(V(x)-Ng_+(x))}\Big[(V'(x) - g_+'(x) + i\theta'(x))e^{iN\theta(x)}(\mathbf{E}^N\hat{\mathbf{D}}_{\text{out}})_{21+}(x) \\ &\qquad + (V'(x) - g_+'(x))(\mathbf{E}^N\hat{\mathbf{D}}_{\text{out}})_{22+}(x) + O(N^{-1})\Big] . \end{aligned} \tag{258}$$
   Assembling the ingredients,
$$\begin{aligned} R_N^{(1)}(x) = -\frac{N}{2\pi i}\Big[&(2g_+'(x) - V'(x) - i\theta'(x))\Big[e^{iN\theta(x)}(\mathbf{E}^N\hat{\mathbf{D}}_{\text{out}})_{11+}(x)(\mathbf{E}^N\hat{\mathbf{D}}_{\text{out}})_{21+}(x) \\ &+ e^{-iN\theta(x)}(\mathbf{E}^N\hat{\mathbf{D}}_{\text{out}})_{12+}(x)(\mathbf{E}^N\hat{\mathbf{D}}_{\text{out}})_{22+}(x) + (\mathbf{E}^N\hat{\mathbf{D}}_{\text{out}})_{11+}(x)(\mathbf{E}^N\hat{\mathbf{D}}_{\text{out}})_{22+}(x) \\ &+ (\mathbf{E}^N\hat{\mathbf{D}}_{\text{out}})_{12+}(x)(\mathbf{E}^N\hat{\mathbf{D}}_{\text{out}})_{21+}(x)\Big] + i\theta'(x)\det(\mathbf{E}^N\hat{\mathbf{D}}_{\text{out}})_+(x) + O(N^{-1})\Big] . \end{aligned} \tag{259}$$
   But $\det(\mathbf{E}^N(z)) \equiv 1$ and $\det(\hat{\mathbf{D}}_{\text{out}}(z)) \equiv 1$, and in the support interval (α, β),
$$2g_+'(x) = (g_+'(x) + g_-'(x)) + (g_+'(x) - g_-'(x)) = V'(x) + i\theta'(x) \tag{260}$$
   so
$$R_N^{(1)}(x) = -\frac{N}{2\pi i}\left(i\theta'(x) + O(N^{-1})\right) , \tag{261}$$
   and differentiating the definition of $\theta(x)$ we obtain $\theta'(x) = -2\pi\psi(x)$ which gives (255).

2. If $x \in (\alpha, \beta)$ is fixed and $\xi_1, \ldots, \xi_n$ all lie in a bounded set, then
$$\begin{aligned} &R_N^{(n)}(x + N^{-1}\Delta(x)\xi_1, \ldots, x + N^{-1}\Delta(x)\xi_n) \\ &\qquad = N^n\psi(x)^n \det\begin{bmatrix} S(\xi_1,\xi_1) & S(\xi_1,\xi_2) & \cdots & S(\xi_1,\xi_n) \\ S(\xi_2,\xi_1) & S(\xi_2,\xi_2) & \cdots & S(\xi_2,\xi_n) \\ \vdots & \vdots & \vdots & \vdots \\ S(\xi_n,\xi_1) & S(\xi_n,\xi_2) & \cdots & S(\xi_n,\xi_n) \end{bmatrix} + O(N^{n-1}) \end{aligned} \tag{262}$$
   as $N \to +\infty$ (with n fixed), where $S(\xi, \eta)$ is the sine kernel and $\Delta(x) = \psi(x)^{-1}$ is proportional to the asymptotic mean spacing. This is again obtained by expanding the kernel $K_N(x, y)$ written in terms of the first column of $\mathbf{A}^N(z)$ as above; it is a "higher-order" calculation.

3. A similar result gives Airy kernel correlations near the edge, and universality of the Tracy-Widom law for extreme eigenvalues. Here one needs to use the inner parametrix, which explains the Airy functions.

References

[1] P. Deift, *Orthogonal Polynomials and Random Matrices: A Riemann-Hilbert Approach*, Courant Lecture Notes in Mathematics, Volume 3, 1999.

[2] M. Mehta, *Random Matrices*, Second Edition, Academic Press, 1991.

[3] G. Szegő, *Orthogonal Polynomials*, American Mathematical Society Colloquium Publications, Volume 23, 1991 (original edition from 1939).

[4] P. D. Miller, *Applied Asymptotic Analysis*, Graduate Studies in Mathematics Volume 75, American Mathematical Society, 2006.

[5] A. S. Fokas, A. R. Its, and A. V. Kitaev, "Discrete Painlevé equations and their appearance in quantum gravity," *Comm. Math. Phys.*, **142**, 313–344, 1991.

[6] N. I. Muskhelishvili, *Singular Integral Equations*, Second Edition, Dover, 1992 (original edition from 1953).

[7] R. R. Coifman, A. McIntosh, and Y. Meyer, "L'intégrale de Cauchy définit un opérateur borné sur L^2 pur les courbes lipschitziennes" [The Cauchy integral defines a bounded operator on L^2 for Lipschitz curves], *Ann. Math.* **116**, 361–387, 1982.

[8] E. B. Saff and V. Totik, *Logarithmic Potentials with External Fields*, Springer Verlag, 1997.