GLOBAL ASYMPTOTICS FOR THE CHRISTOFFEL-DARBOUX KERNEL OF RANDOM MATRIX THEORY

THOMAS KRIECHERBAUER, KRISTINA SCHUBERT, KATHARINA SCHÜLER, AND MARTIN VENKER

Dedicated to Leonid Pastur with admiration on the occasion of his 75th birthday

Abstract. The investigation of universality questions for local eigenvalue statistics continues to be a driving force in the theory of Random Matrices. For Matrix Models [51] the method of orthogonal polynomials can be used and the asymptotics of the Christoffel-Darboux kernel [57] become the key for studying universality. In this paper the existing results on the CD-kernel will be extended in two directions. Firstly, in order to analyze the transition from the universal to the non-universal regime, we provide leading order asymptotics that are global rather than local. This allows e.g. to describe the moderate deviations for the largest eigenvalues of unitary ensembles ($\beta = 2$), where such a transition occurs. Secondly, our asymptotics will be uniform under perturbations of the probability measure that defines the matrix ensemble. Such information is useful for the analysis of a different type of ensembles [23], which is not known to be determinantal and for which the method of orthogonal polynomials cannot be used directly. The just described applications of our results are formulated in this paper but will be proved elsewhere. As a byproduct of our analysis we derive first order corrections for the 1-point correlation functions of unitary ensembles in the bulk. Our proofs are based on the nonlinear steepest descent method [20]. They follow closely [17] and incorporate improvements introduced in [34, 62]. The presentation is self-contained except for a number of general facts from Random Matrix theory and from the theory of singular integral operators.

1. Introduction

It is a somewhat surprising but well established fact that eigenvalues of randomly chosen matrices provide useful statistical models in various areas of mathematics and physics ranging from number theory to quantum chaos. A common theme in many of these applications of Random Matrix Theory is that local eigenvalue statistics yield good asymptotic descriptions of point processes on the line where the random points are not distributed independently but display local repulsion of some form instead. Examples for local statistics are distributions of extremal values and of nearest neighbor spacings. See [2, Part III] and references therein for a recent collection of such applications and [27] for an elementary exposition of a specific example.

In general, the analysis of local statistics poses major technical difficulties for point processes with stochastic dependencies. However, in the special case of eigenvalue distributions of Gaussian ensembles Gaudin and Mehta [47] discovered that certain statistical quantities could be expressed in terms of Hermite polynomials (cf. [4, Section 3.10]). The then already known large degree asymptotics for these

2010 *Mathematics Subject Classification.* Primary 60B20; Secondary 42C05, 35Q15, 60F05.

Key words and phrases. Random Matrices, Orthogonal polynomials, Riemann-Hilbert problems.

The first, second and third author were partially supported by the SFB/TR 12. The fourth author was partially supported by the SFB 701. We gratefully acknowledge the support of the Deutsche Forschungsgemeinschaft provided through these grants.

polynomials eventually led to an efficient description of local eigenvalue statistics for matrices of large size. Starting with the work of Pastur and Shcherbina [49] this path of analysis was later extended to a larger class of random matrices that are now known as matrix models or as invariant matrix ensembles. We do not present a broad view on these ensembles in this paper and we refer the reader to the recent books [2, 4, 15, 22, 51] as well as to the classics [46, 13] instead.

In this paper we restrict our attention to unitarily invariant ensembles, which are commonly called unitary ensembles. Despite being slightly misleading we also use this abbreviation. More precisely, we consider probability measures $\hat{\mathbb{P}}_{N,V}$ on the set of $N \times N$ Hermitian matrices where $V : J \to \mathbb{R}$ denotes a function on $J = \{x \in \mathbb{R} \,|\, L_- \leq x \leq L_+\}$, which is a finite or infinite interval ($-\infty \leq L_- < L_+ \leq \infty$). Precise assumptions on V will be given below (see assumption $\mathbf{(GA)}_1$). The measure $\hat{\mathbb{P}}_{N,V}$ is of the form

$$\mathrm{d}\hat{\mathbb{P}}_{N,V}(M) = C_{N,V} e^{-N\mathrm{tr}(V(M))} \mathbb{1}_J(M)\,\mathrm{d}M, \tag{1.1}$$

where $C_{N,V} > 0$ is a normalizing constant and

$$\mathrm{d}M = \prod_{j=1}^{N} \mathrm{d}M_{jj} \prod_{1\leq j<k\leq N} \mathrm{d}M_{jk}^R \mathrm{d}M_{jk}^I$$

is called the Lebesgue measure on Hermitian matrices. It is defined as the product of Lebesgue measures on the real and imaginary parts of the entries $M_{jk} = M_{jk}^R + iM_{jk}^I$ of the upper triangular block. The matrix $V(M)$ is defined by functional calculus and tr denotes the trace. By a slight abuse of notation, the symbol $\mathbb{1}_J(M)$ takes on the value 1 if all eigenvalues of M lie in J and equals 0 otherwise.

$\hat{\mathbb{P}}_{N,V}$ induces a probability measure $\mathbb{P}_{N,V}$ on the vector $\lambda = (\lambda_1, \ldots, \lambda_N) \in \mathbb{R}^N$ of eigenvalues. Assuming that all orderings of the eigenvalues are equally likely, one obtains $\mathrm{d}\mathbb{P}_{N,V}(\lambda) = P_{N,V}(\lambda)\,\mathrm{d}\lambda$ with density (see e.g. [13, Sec. 5.3])

$$P_{N,V}(\lambda) = \frac{1}{Z_{N,V}} \prod_{1\leq j<k\leq N} (\lambda_k - \lambda_j)^2 \prod_{m=1}^{N} e^{-NV(\lambda_m)} \mathbb{1}_J(\lambda_m). \tag{1.2}$$

Of course, the constant $Z_{N,V}$ has to be chosen such that $\int_{\mathbb{R}^N} P_{N,V}(\lambda)\,\mathrm{d}\lambda = 1$. In order to analyze the distribution of the eigenvalues, one needs to understand the behavior of the k-point correlations

$$R_{N,V}^{(k)}(\lambda_1, \ldots, \lambda_k) := \frac{N!}{(N-k)!} \int_{\mathbb{R}^{N-k}} P_{N,V}(\lambda)\,\mathrm{d}\lambda_{k+1} \ldots \mathrm{d}\lambda_N \tag{1.3}$$

for all $1 \leq k \leq N$. Observe that the above mentioned examples for local statistics, i.e. the distribution of spacings as well as the distribution of the largest value $\lambda_{\max} := \max\{\lambda_1, \ldots, \lambda_N\}$, can be expressed in terms of k-point correlations (cf. [3]). For densities $P_{N,V}$ of the form (1.2) there are two key facts that allow to study k-point correlations in great detail.

Firstly, $P_{N,V}$ defines a determinantal point process, i.e.

$$R_{N,V}^{(k)}(\lambda_1, \ldots, \lambda_k) = \det\left[(K_{N,V}(\lambda_i, \lambda_j))_{1\leq i,j\leq k}\right] \tag{1.4}$$

for some kernel $K_{N,V}$ that is independent of k (see e.g. [13, (5.40)]). Moreover, the kernel is of such type that gap probabilities, i.e. probabilities p_I that no eigenvalue lies in I, are given by Fredholm determinants

$$p_I = \det\left(\mathbb{1} - \mathcal{K}_{N,V}|_{L^2(I)}\right),$$

where $\mathcal{K}_{N,V}$ denotes the integral operator associated with the kernel $K_{N,V}$ and $\mathbb{1}$ denotes the identity on $L^2(I)$ (see [60] for a nice derivation). This implies, for example,

$$\text{Probability}(\lambda_{\max} \geq t) = \det\left(\mathbb{1} - \mathcal{K}_{N,V}|_{L^2(t,\infty)}\right).$$

The second key fact is that the kernel $K_{N,V}$ can be expressed in terms of orthogonal polynomials. More precisely, denote by $p_j^{(N,V)}(x) = \gamma_j^{(N,V)} x^j + \ldots$ the unique polynomial of degree j with leading coefficient $\gamma_j^{(N,V)} > 0$ satisfying

$$\int_J p_i^{(N,V)}(x)\, p_j^{(N,V)}(x)\, e^{-NV(x)}\, \mathrm{d}x = \delta_{ij}.$$

Then

$$K_{N,V}(x,y) = \sum_{j=0}^{N-1} p_j^{(N,V)}(x) p_j^{(N,V)}(y) e^{-\frac{N}{2}(V(x)+V(y))}$$

and by the Christoffel-Darboux formula [57]

$$K_{N,V}(x,y) = \frac{\gamma_{N-1}^{(N,V)}}{\gamma_N^{(N,V)}} \frac{p_N^{(N,V)}(x) p_{N-1}^{(N,V)}(y) - p_{N-1}^{(N,V)}(x) p_N^{(N,V)}(y)}{x-y} e^{-\frac{N}{2}(V(x)+V(y))}. \tag{1.5}$$

This formula also gives rise to the naming of $K_{N,V}$ as the Christoffel-Darboux kernel or CD-kernel in short.

A driving force in the development of Random Matrix Theory has been the *universality conjecture*. Recent accounts of the state of the art can be found in the monographs [15, 51] and in the survey [31]. For unitary ensembles of type (1.1) it says, roughly speaking, that local eigenvalue statistics do not depend on V as the matrix size N tends to infinity. Of course, one needs to distinguish between regions where eigenvalues accumulate (bulk), regions where the appearance of eigenvalues is highly unlikely (void) and transitional regions (edge). One way to prove universality is to use relation (1.4) between k-point correlations and the CD-kernel. The task then becomes to show that the appropriately rescaled kernel $K_{N,V}$ does not depend on V in the limit $N \to \infty$. For example, if $x \in J$ lies in the bulk of the spectrum, then $K_{N,V}$ is well described by the sine kernel. More precisely,

$$\frac{1}{N\beta} K_{N,V}\left(x + \frac{s}{N\beta}, x + \frac{t}{N\beta}\right) = \frac{\sin(\pi(s-t))}{\pi(s-t)} + \mathcal{O}\left(N^{-1}\right), \tag{1.6}$$

where β denotes some positive constant, depending on x and V, such that $\frac{1}{N\beta}$ is the average spacing of eigenvalues that lie in a small vicinity of x. In other words: The universal aspect of (1.6) is that the V-dependency of the leading order ($N \to \infty$) of $K_{N,V}$ is all captured in a single rescaling factor β, which in addition defines a natural local length scale. One may not expect that the error bound $\mathcal{O}(N^{-1})$ in (1.6) is uniform for all $s, t \in \mathbb{R}$ since $x + \frac{s}{N\beta}$, $x + \frac{t}{N\beta}$ are required to lie in a vicinity of the point x. In fact, uniformity of the error bound in (1.6) has so far been formulated for s, t in arbitrary but fixed bounded sets (see [18] and [34, 62] for related unitary ensembles where orthogonal polynomials of Jacobi-type resp. of Laguerre-type are used). This means that (1.6) provides leading order asymptotics for $K_{N,V}(u,v)$ only if both u and v lie in some set U_N with shrinking length of order $\mathcal{O}(N^{-1})$. For some applications such information is not sufficient (see e.g. Application 3, [29, Theorem 1] and references therein) and better asymptotics, as stated e.g. in Theorem 1.7 below, are essential. Bulk universality with weaker bounds on the rate of convergence than in (1.6) but obtained under much weaker regularity assumptions can be found in

[49, 50, 45, 36, 37, 39, 40, 41, 42, 43, 58, 56]. See also [31] and references therein for a recent overview on universality results.

It is well known that the validity of the sine kernel description ends at spectral edges. There, the leading order description of $K_{N,V}$ becomes more complicated. In particular, one needs to distinguish between hard edges and soft edges. We do not give a precise definition of this terminology, but the reader may consult Remark 2.2 (b) below or e.g. [22, Subsections 7.1.3 and 7.2.1] for more information. In this paper we are only concerned with soft edges where the Airy kernel (see e.g. [51, Theorem 5.3.3] and [1, Section 10.4] for a definition and properties of the Airy function Ai)

$$\mathbb{A}\mathrm{i}(s,t) := \frac{\mathrm{Ai}(s)\mathrm{Ai}'(t) - \mathrm{Ai}'(s)\mathrm{Ai}(t)}{s-t} = \int_0^\infty \mathrm{Ai}(s+r)\mathrm{Ai}(t+r)\,\mathrm{d}r \tag{1.7}$$

replaces the sine kernel (see e.g. [14, 62, 45, 38] and [51, Section 13.1]). Let us have a closer look at the result [14, Theorem 1.1] which provides the strongest bounds on the rate of convergence. In [14] ensembles are considered that are slightly different from (1.1). Nevertheless, their result applies to ensembles of type (1.1) with $J = \mathbb{R}$ and $V(x) = x^{2m}$ for some positive integer m. It reads

$$\frac{1}{N^{\frac{2}{3}}\gamma} K_{N,V}\left(b + \frac{s}{N^{\frac{2}{3}}\gamma}, b + \frac{t}{N^{\frac{2}{3}}\gamma}\right) = \mathbb{A}\mathrm{i}(s,t) + \mathcal{O}\left(N^{-\frac{2}{3}} e^{-C(s+t)}\right), \tag{1.8}$$

where both b and γ depend on V. Again, the universal aspect of (1.8) is that the V-dependency of the leading asymptotics of $K_{N,V}$ only enters through these two numbers. While γ has no natural interpretation, b describes the almost sure limit of the largest eigenvalue as $N \to \infty$. The bound $\mathcal{O}\left(N^{-2/3} e^{-C(s+t)}\right)$ as well as the positive constant C appearing therein are both uniform for $s, t \in [L_0, \infty)$ with L_0 being arbitrary but fixed. This result appears to be of a more general type than the one stated above for the bulk, because s and t are not required to lie in a bounded set. However, due to the rapid decay of the Airy kernel for s, $t \to \infty$ (see e.g. (4.23)), the error term is dominant for $s, t \geq (\log N)^\alpha$ if $\alpha > \frac{2}{3}$. Although (1.8) is sufficient to prove the Tracy-Widom law for the rescaled largest eigenvalue $(\lambda_{\max} - b)N^{2/3}\gamma$ in the limit $N \to \infty$ (cf. [59], [14, Corollary 1.4]), relation (1.8) only provides estimates in the region $\left(b + N^{-2/3}(\log N)^\alpha, \infty\right)$, $\alpha > \frac{2}{3}$, that are far too weak to obtain moderate or large deviation principles for $\lambda_{\max}$. We will be able to derive such principles from Theorem 1.5 (see Application 1).

Another source of motivation to write this paper comes from a rather novel class of probability measures on $\mathbb{R}^N$, for which local bulk universality was proven recently in [23] (see also [63] for related results). In a more general setup they appeared first in [6], where macroscopic correlations have been studied. In our case, these ensembles have densities proportional to

$$\prod_{1 \leq j < k \leq N} \varphi(x_j - x_k) \prod_{m=1}^{N} e^{-NQ(x_m)}. \tag{1.9}$$

Apart from technical conditions it is assumed that φ satisfies

$$\varphi(0) = 0, \quad \varphi(t) > 0 \quad \text{for } t \neq 0 \quad \text{and} \quad \lim_{t \to 0} \frac{\varphi(t)}{|t|^2} = C > 0.$$

We prefer to think that (1.9) describes the distribution of the positions x_j of N particles on the real line rather than eigenvalues of some matrix ensemble. However, as the last author learned from G. Borot, the ensemble (1.9) in the form (1.29) can also be realized as eigenvalue distribution of a random Hermitian matrix ([5], see also [44]).

The motivation to study such ensembles comes from the fact that (1.9) constitutes a repulsive particle system which shares with (1.2) the repulsion strength between close particles. Proving local universality for these ensembles, i.e. showing that it has local limiting distributions not depending on Q and φ apart from the repulsion condition, might be a further step to an understanding of origin and range of the ubiquity of local random matrix distributions.

Generically, ensembles (1.9) do not carry a determinantal structure and the method of orthogonal polynomials, which led to the introduction of the Christoffel-Darboux kernel, cannot be applied directly. Nevertheless, by a stochastic linearization procedure any density of the form (1.9) with even φ can be related to a family of ensembles of type (1.2) with $V = V_N = Q + \frac{1}{N} f$, indexed by f. Indeed, the functions f are given as sample paths of a centered stationary Gaussian process on $\mathbb{R}$, whose covariance function is determined by φ. Using universality results for $K_{N,Q+f/N}$, pointwise for every f, Götze and Venker were able to prove universality for all k-point correlations in the bulk. Note that the rates of convergence for $K_{N,Q+f/N}$ will provide corresponding rates for the k-point correlations only if they are uniform in f. In view of Applications 2 and 3, we will therefore derive bounds on the rates of convergence for the Christoffel-Darboux kernel $K_{N,V}$ that are independent of V in some neighborhood of Q. At this point we can also explain why we do not only consider functions V that are defined on all of $\mathbb{R}$ but on some interval J instead. The main reason is that in the analysis of Götze and Venker a truncation is used so that $V = Q + \frac{1}{N} f$ is only defined on a suitable compact interval that is chosen independently of N. Note that the truncation is always performed in such a way that only soft edges (see paragraph below (1.10)) occur.

Before stating our results we summarize the main objectives of the present paper:

• With Theorem 1.3 we present a uniform leading order description of the Christoffel-Darboux kernel $K_{N,V}(x, y)$. Compared to the results in the literature the novel aspects are that the description is global in (x, y) rather than local and that it applies not only to a fixed function V but simultaneously to an open neighborhood of such functions.
• Having Applications 1, 2, and 3 in mind, we further evaluate the results of Theorem 1.3 in special cases. While Theorems 1.7 and 1.8 extend the universality results (1.6) and (1.8) for unitary ensembles to their maximal domains of validity, Theorem 1.5 contains first order corrections in the bulk as well as matching formulae at the edges for the 1-point correlation function that appear to be new, except for the Gaussian case [25].
• The proof of Theorem 1.3 uses the Riemann-Hilbert formulation for orthogonal polynomials [21]. The asymptotic analysis of the Riemann-Hilbert problem is essentially identical with [17] and is based on the nonlinear steepest descent method introduced by Deift-Zhou [20] and further developed in [19]. We have streamlined the exposition in view of our prime goal to analyze the global asymptotics of the Christoffel-Darboux kernel. In particular, we have incorporated the improvements that were introduced in [34] in the form presented in [62] (see (3.4) and (4.6)).
• We restrict our attention to unitary Hermite-type ensembles, where only soft edges occur, with varying weights e^{-NV} and convex V. Nevertheless, we hope that this paper may serve as a blueprint to extract global asymptotics of the Christoffel-Darboux kernel for other types of unitary ensembles (e.g. Laguerre-type, Jacobi-type) or in degenerate cases by adapting the analysis of this paper with the help of [12, 28, 33, 34, 35, 61, 7, 8, 9, 10, 11, 24, 52].

In order to specify our assumptions on V, we first need to introduce the Mhaskar-Rakhmanov-Saff numbers a, $b \in \mathbb{R}$ from the theory of orthogonal polynomials. In

the case of convex V and varying weights e^{-NV} they are defined implicitly by the conditions

$$\int_a^b \frac{V'(t)}{\sqrt{(b-t)(t-a)}}\,\mathrm{d}t = 0, \quad \int_a^b \frac{tV'(t)}{\sqrt{(b-t)(t-a)}}\,\mathrm{d}t = 2\pi. \tag{1.10}$$

The significance of these relations will become clear in Section 2. For twice differentiable functions $V : \mathbb{R} \to \mathbb{R}$ with V' strictly increasing and $\lim_{|x|\to\infty} V(x) = \infty$ one may show by elementary but not entirely trivial arguments that there exist unique real numbers $a < b$ satisfying (1.10). This also implies that for restrictions $\widetilde{V} = V\big|_J$ to any interval J, relations (1.10) can only hold if $[a, b] \subset J$. In our context such restrictions come into play because of the truncation procedure that is used in the analysis of ensembles of type (1.9), see [30], [54]. There the truncation is always performed in such a way that $[a, b]$ is contained in the interior of J. As explained in Remark 2.2 (b) below, this is the reason why we only need to consider soft edges.

We are now ready to state our general assumptions on V that are of two types:

(GA)$_1$ A function V is said to satisfy **(GA)**$_1$ if (1)–(4) hold

(1) $V : J \to \mathbb{R}$ is real analytic, $J = [L_-, L_+] \cap \mathbb{R}$ with $-\infty \le L_- < L_+ \le \infty$.

(2) V' is strictly increasing (convexity assumption).

(3) $\lim_{|x|\to\infty} V(x) = \infty$.

(4) There exist $L_- < a_V < b_V < L_+$ such that (1.10) holds with $a = a_V$, $b = b_V$.

Condition (1) allows to perform the nonlinear steepest descent method of Deift-Zhou [20] in its simplest form (see (3.8) and below). For an extension to the case of finite regularity we refer the reader to the work of McLaughlin and Miller [45]. Condition (2) ensures that the equilibrium measure (see Section 2) is of the most simple type as well, i.e. its support consists of only one interval with square-root vanishing of the density at the endpoints (see e.g. [18, 16, 32] and references therein for a more general picture). It is clear that some condition like (3) is needed to ensure integrability of $P_{N,V}$ given by (1.2) in the case of unbounded intervals J. On first sight, (3) appears to be too weak for that. However, using condition (2), it is not hard to see that V grows at least linearly for $x \to \pm\infty$.

The second type of assumptions is adapted to the analysis of ensembles of the form (1.9).

(GA)$_2$ Q is said to satisfy **(GA)**$_2$ if **(GA)**$_1$ holds for Q with (2), (3) replaced by

(2′) $\inf_{x\in J} Q''(x) > 0$.

(3′) J is compact, i.e. L_-, L_+ are both finite.

Remark 1.1. Given a function Q that satisfies **(GA)**$_2$, there exists a complex open neighborhood D of J such that Q can be extended analytically to a bounded function on D that we still denote by Q. In this sense, Q belongs to the real Banach space $X_D := \{f : D \to \mathbb{C} \,|\, f \text{ analytic and bounded, } f(D \cap \mathbb{R}) \subset \mathbb{R}\}$, equipped with the sup-norm $\|f\|_\infty := \sup_{z\in D}|f(z)|$.

Given a function V that satisfies **(GA)**$_1$, we now define a number of objects that are needed to state our results.

Definition 1.2. Assume that **(GA)**$_1$ holds for V. The linear rescaling that maps $[-1, 1]$ onto $[a_V, b_V]$ (cf. **(GA)**$_1$(4)) is denoted by

$$\lambda_V : \mathbb{R} \to \mathbb{R},\ \lambda_V(s) := \frac{b_V - a_V}{2}s + \frac{b_V + a_V}{2}. \tag{1.11}$$

Hence, $[-1,1] \subset \hat{J} := \lambda_V^{-1}(J)$ (cf. $\mathbf{(GA)}_1(1)$). Moreover, set

$$h_V : \hat{J} \times \hat{J} \to \mathbb{R},\ h_V(t,x) := \int_0^1 (V \circ \lambda_V)''(x + u(t-x))\,\mathrm{d}u = \frac{(V \circ \lambda_V)'(t) - (V \circ \lambda_V)'(x)}{t-x}, \tag{1.12}$$

$$G_V : \hat{J} \to \mathbb{R},\ G_V(x) := \frac{1}{\pi} \int_{-1}^1 \frac{h_V(t,x)}{\sqrt{1-t^2}}\,\mathrm{d}t, \tag{1.13}$$

$$\rho_V : \mathbb{R} \to \mathbb{R},\ \rho_V(x) := \begin{cases} \frac{1}{2\pi}\sqrt{1-x^2}\,G_V(x) & , \text{ if } |x| \le 1, \\ 0 & , \text{ else}, \end{cases} \tag{1.14}$$

$$\xi_V : \mathbb{R} \to \mathbb{R},\ \xi_V(x) := 2\pi \int_x^1 \rho_V(t)\,\mathrm{d}t, \tag{1.15}$$

$$\eta_V : \hat{J} \to \mathbb{R},\ \eta_V(x) := \begin{cases} \int_1^x \sqrt{t^2-1}\,G_V(t)\,\mathrm{d}t & , \text{ if } x > 1, \\ 0 & , \text{ if } |x| \le 1, \\ \int_x^{-1} \sqrt{t^2-1}\,G_V(t)\,\mathrm{d}t & , \text{ if } x < -1, \end{cases} \tag{1.16}$$

$$a : \mathbb{R} \setminus [-1,1] \to \mathbb{R},\ a(x) := \left(\frac{x-1}{x+1}\right)^{1/4}, \tag{1.17}$$

$$\hat{a} : (-1,1) \to \mathbb{R},\ \hat{a}(x) := \left(\frac{1-x}{1+x}\right)^{1/4}. \tag{1.18}$$

As argued in Remark A.2 below, there exists a real analytic function $\hat{f}_V : \mathcal{E}_V \to \mathbb{R}$ with $\mathcal{E}_V := [-1-\delta_V, -1+\delta_V] \cup [1-\delta_V, 1+\delta_V]$ for some $0 < \delta_V < 1$ such that

$$(|x-1|\gamma_V^+ \hat{f}_V(x))^{3/2} = \begin{cases} \frac{3}{4}\xi_V(x) & , \text{ if } 1-\delta_V \le x \le 1, \\ \frac{3}{4}\eta_V(x) & , \text{ if } 1 \le x \le 1+\delta_V, \end{cases} \tag{1.19}$$

$$(|x+1|\gamma_V^- \hat{f}_V(x))^{3/2} = \begin{cases} \frac{3}{4}(2\pi - \xi_V(x)) & , \text{ if } -1 \le x \le -1+\delta_V, \\ \frac{3}{4}\eta_V(x) & , \text{ if } -1-\delta_V \le x \le -1. \end{cases} \tag{1.20}$$

Here

$$\gamma_V^{\pm} := 2^{-1/3} G_V(\pm 1)^{2/3} \tag{1.21}$$

is chosen in such a way that $\hat{f}_V(\pm 1) = 1$. Finally,

$$f_{N,V} : \mathcal{E}_V \to \mathbb{R},\ f_{N,V}(x) := \begin{cases} N^{2/3}\gamma_V^+(x-1)\hat{f}_V(x) & , \text{ if } |x-1| \le \delta_V, \\ -N^{2/3}\gamma_V^-(x+1)\hat{f}_V(x) & , \text{ if } |x+1| \le \delta_V, \end{cases} \tag{1.22}$$

$$d_V : \mathcal{E}_V \to \mathbb{R},\ d_V(x) := \begin{cases} ((x+1)\hat{f}_V(x))^{1/4} & , \text{ if } |x-1| \le \delta_V, \\ ((1-x)\hat{f}_V(x))^{1/4} & , \text{ if } |x+1| \le \delta_V. \end{cases} \tag{1.23}$$

Our theorem on the global asymptotics of the Christoffel-Darboux kernel reads

Theorem 1.3.

(a) Assume that the function V satisfies $\mathbf{(GA)}_1$. Let λ_V, δ_V, $\hat{J}$, η_V, a, $\hat{a}$, ξ_V, $\gamma_V^{\pm}$, $f_{N,V}$, d_V be given as in Definition 1.2. Then there exists a positive $\delta_0 < \delta_V$ such

that for all $\delta \in (0, \delta_0]$ the following holds: Define $k : \hat{J} \to \mathbb{R}^2$ by

$$k(x) := \begin{cases} \frac{1}{\sqrt{4\pi}} \operatorname{sgn}(x)^N e^{-\frac{N}{2}\eta_V(x)} \begin{pmatrix} a(x) \\ [a(x)]^{-1} \end{pmatrix} & , \text{ if } x \in \hat{J} \setminus [-1-\delta, 1+\delta], \\ \frac{1}{\sqrt{\pi}} \begin{pmatrix} \hat{a}(x) \cos\left(\frac{N}{2}\xi_V(x) + \frac{\pi}{4}\right) \\ [\hat{a}(x)]^{-1} \cos\left(\frac{N}{2}\xi_V(x) - \frac{\pi}{4}\right) \end{pmatrix} & , \text{ if } x \in (-1+\delta, 1-\delta), \\ \begin{pmatrix} -\operatorname{Ai}'(f_{N,V}(x))[N^{1/6}\gamma_V^+ d_V(x)]^{-1} \\ \operatorname{Ai}(f_{N,V}(x))[N^{1/6}\gamma_V^+ d_V(x)] \end{pmatrix} & , \text{ if } x \in [1-\delta, 1+\delta], \\ (-1)^N \begin{pmatrix} \operatorname{Ai}(f_{N,V}(x))[N^{1/6}\gamma_V^- d_V(x)] \\ -\operatorname{Ai}'(f_{N,V}(x))[N^{1/6}\gamma_V^- d_V(x)]^{-1} \end{pmatrix} & , \text{ if } x \in [-1-\delta, -1+\delta]. \end{cases}$$

Then,

(1.24)
$$\frac{b_V - a_V}{2} K_{N,V}(\lambda_V(x), \lambda_V(y)) = \frac{k_1(x)k_2(y) - k_2(x)k_1(y)}{x-y} + k(y)^T \mathcal{O}\left(N^{-1}\right) k(x).$$

Here $\mathcal{O}(N^{-1})$ denotes a real-valued 2×2 matrix whose entries are of order N^{-1}. The bound is uniform for δ in compact subsets of $(0, \delta_0]$ and for x, y in bounded subsets of $\hat{J}$.

(b) *Assume that **(GA)**$_2$ holds for some function Q. Let D, X_D be defined as in Remark 1.1. Then there exists an open neighborhood $\mathcal{U}$ of Q in $(X_D, \|\cdot\|_\infty)$ such that $V\big|_J$ satisfies **(GA)**$_1$ for all $V \in \mathcal{U}$. Hence claim (a) holds for $V\big|_J$. The error bound in (1.24) is also uniform in $V \in \mathcal{U}$ and δ_0 can be chosen to be independent of $V \in \mathcal{U}$.*

Remark 1.4.

(a) The definition of k in the bulk, i.e. for $|x| < 1 - \delta$ in Theorem 1.3, could also be given by

$$k(x) = \frac{1}{\sqrt{2\pi}(1-x^2)^{1/4}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \cos\left(\frac{N}{2}\xi_V(x) - \frac{1}{2}\arcsin(x)\right) \\ \sin\left(\frac{N}{2}\xi_V(x) + \frac{1}{2}\arcsin(x)\right) \end{pmatrix},$$

which is more common in the literature. Using $\sin\left(\frac{1}{2}\arccos(x)\right) = \sqrt{1-x}$, the identity of both formulae is easily verified. We find the formula stated in Theorem 1.3 more suitable for deriving sine-kernel asymptotics.

(b) Of course, for $x = y \in \hat{J}$, the leading term of $K_{N,V}$ in (1.24) reads $k_1'(x)k_2(x) - k_1(x)k_2'(x)$.

The evaluation of $K_{N,V}(x, y)$ on the diagonal $x = y$ is of particular interest because it equals the 1-point correlation function (see (1.4)) which agrees with the expected density of eigenvalues up to a factor N. The following theorem describes the large N behavior of $K_{N,V}(x, x)$. Note that particular care has been taken to extend bulk resp. void asymptotics to the edge region. The first order correction in statement (i) was already derived in [25] for Gaussian ensembles $V(x) = x^2$ using Hermite polynomials.

Theorem 1.5. *Let the assumptions of Theorem 1.3 (a) resp. (b) be satisfied. Then there exists a (small) positive constant c such that the following representations for $D(x) := \frac{b_V - a_V}{2} K_{N,V}(\lambda_V(x), \lambda_V(x))$ hold.*

(i) For $x \in [-1,1]$ with $|x| < 1 - c^{-1}N^{-2/3}$:

$$D(x) = \left(N\rho_V(x) - \frac{1}{2\pi(1-x^2)} \cos(N\xi_V(x)) \right) \left(1 + \mathcal{O}\left(\frac{1}{N^2(1-|x|)^3} \right) \right).$$

(ii) For $x \in \hat{J} \setminus [-1,1]$ with $|x| > 1 + c^{-1}N^{-2/3}$:

$$D(x) = \frac{1}{4\pi} e^{-N\eta_V(x)} \left[\frac{1}{x^2-1} + \mathcal{O}\left(\frac{1}{N(|x|-1)^{5/2}} \right) + \mathcal{O}\left(\frac{1}{N} \right) \right].$$

(iii) For $1 - c < x < 1 + cN^{-2/5}$:

$$D(x) = N^{2/3}\gamma_V^+ \, \mathbb{A}\mathrm{i}\left(\gamma_V^+ N^{2/3}(x-1), \gamma_V^+ N^{2/3}(x-1) \right) (1 + r(x)), \tag{1.25}$$

$$\text{where } r(x) = \begin{cases} \mathcal{O}(1-x) + \mathcal{O}\left(N^{-2/3}\right) & , \text{ if } x \le 1, \\ \mathcal{O}\left(N(x-1)^{5/2}\right) + \mathcal{O}\left(N^{-2/3}\right) & , \text{ if } x \ge 1. \end{cases}$$

(iv) In the case $-1 - cN^{-2/5} < x < -1 + c$, statement (iii) holds with x being replaced by $-x$ on the r.h.s. of equation (1.25) and with γ_V^+ replaced by γ_V^-.

In the situation of part (b) the error bounds as well as the constant c can be chosen independently of $V \in \mathcal{U}$ with $\mathcal{U}$ as given by Theorem 1.3.

It will be shown in [55] that the information provided in Theorem 1.5 (ii) suffices to derive moderate and large deviations principles for the distribution of the largest eigenvalue $\lambda_{\max}$:

Application 1. *Assume that V satisfies **(GA)**$_1$ and let $\hat{\mathbb{P}}_{N,V}$ be the associated unitarily invariant matrix ensemble (see (1.1)). The random variable $\alpha := \lambda_V^{-1}(\lambda_{\max})$ denotes the linearly rescaled maximal eigenvalue (cf. Definition 1.2). Moreover, we introduce the distribution function of α on a local scale,*

$$O_N(s) := \hat{\mathbb{P}}_{N,V}\left(\alpha > 1 + \frac{s}{\gamma_V^+ N^{2/3}} \right) \quad (\textit{cf. } (1.21)).$$

Then we have in the regime of moderate deviations, i.e. $q_N \le s \le p_N$ with $q_N \to \infty$ and $N^{-2/3}p_N \to 0$ for $N \to \infty$, that

$$\frac{\log O_N(s)}{s^{3/2}} = -\frac{4}{3} - \frac{\log\left(16\pi s^{3/2}\right)}{s^{3/2}} + \mathcal{O}\left(\frac{s}{N^{2/3}} \right). \tag{1.26}$$

Under the more restrictive assumption that $N^{-4/15}p_N$ remains bounded, this implies

$$O_N(s) = \frac{1}{16\pi} s^{-\frac{3}{2}} e^{-\frac{4}{3}s^{3/2}} \left(1 + \mathcal{O}\left(\frac{s^{5/2}}{N^{2/3}} \right) \right). \tag{1.27}$$

Remark 1.6 (see [55]). As mentioned above, it follows from the already known asymptotics of the Christoffel-Darboux kernel that $O_N(s) \to F_{\mathrm{TW},2}(s)$ for s fixed as $N \to \infty$, where $F_{\mathrm{TW},2}$ denotes the Tracy-Widom distribution with parameter $\beta = 2$. The asymptotics of the Tracy-Widom distribution are given by

$$F_{\mathrm{TW},2}(s) = \frac{1}{16\pi} s^{-\frac{3}{2}} e^{-\frac{4}{3}s^{3/2}} \left(1 + \mathcal{O}\left(s^{-\frac{3}{2}} \right) \right) \quad \text{as } s \to \infty.$$

Thus, $F_{\mathrm{TW},2}(s)$ continues to provide the correct leading order description of $O_N(s)$ for all values of s up to order $o(N^{4/15})$.

In case one is interested in the asymptotics of $O_N(s)$ for even larger values of s, one may use the formula that is valid in the large deviations regime: Choose $(q_N)_N$

with $q_N \to \infty$ for $N \to \infty$ and $L > 1$ such that $[1 + \frac{q_N}{N^{2/3}}, L]$ is contained in the interior of $\hat{J}$. Then for all $x \in [1 + \frac{q_N}{N^{2/3}}, L]$ (see [55])

$$\hat{\mathbb{P}}_{N,V}(\alpha > x) = \frac{e^{-N\eta_V(x)}}{4\pi N(x^2-1)\eta_V'(x)}\left(1 + \mathcal{O}\left(\frac{1}{N(x-1)^{3/2}}\right)\right). \tag{1.28}$$

In fact, one may derive the statements contained in Application 1 from (1.28).

There is a fundamental difference between the results in Theorems 1.3, 1.5, and (1.28) as compared to those stated in Application 1. In the latter case the dependency of the leading term on V is all included in the three numbers a_V, b_V and γ_V^+. Therefore Application 1 is called a universality result, whereas all the other leading order formulae stated above depend on V through all of the information that is encoded in the function G_V (and derived quantities such as ρ_V, η_V, ξ_V, $\hat{f}_V$, $f_{N,V}$, d_V).

As advertised earlier, we will now present universality results that are consequences of Theorems 1.3 and 1.5 and that extend the previously known range of applicability. We start with the bulk:

Theorem 1.7. *Let the assumptions of Theorem 1.3 (a) resp. (b) be satisfied and let $0 < \delta_0 < \delta_V$ be given as in Theorem 1.3. For all $\delta \in (0, \delta_0]$ there exists a positive constant c_δ (independent of $V \in \mathcal{U}$ in the situation of (b) with $\mathcal{U}$ as given by Theorem 1.3) such that for all $x \in [-1+\delta, 1-\delta]$, and $|s|$, $|t| < c_\delta N$:*

$$\begin{aligned}&\frac{b_V - a_V}{2N\rho_V(x)} K_{N,V}\left(\lambda_V\left(x + \frac{s}{N\rho_V(x)}\right), \lambda_V\left(x + \frac{t}{N\rho_V(x)}\right)\right)\\ &= \frac{\sin(\pi(s-t))}{\pi(s-t)} + \mathcal{O}\left(\frac{1+|s|+|t|}{N}\right).\end{aligned}$$

The error bound is uniform in x (and in $V \in \mathcal{U}$ in part (b)) for δ chosen from an arbitrary but fixed compact subset of $(0, \delta_0]$.

The formulation of universality at the edge is more subtle. The reason is the rapid decay of the Airy-kernel as the arguments tend to $+\infty$, whereas oscillations occur for negative arguments. In view of applications in Random Matrix Theory, we present our results on $K_{N,V}(x,y)$ for $b_V - \mathcal{O}(N^{-2/3}) \le x, y \le b_V + \mathcal{O}(N^{-2/5})$, i.e. in a region which is not symmetric around b_V.

Theorem 1.8. *Let the assumptions of Theorem 1.3 (a) resp. (b) be satisfied and assume that $q < 0 < p$ are given. Then*

$$\begin{aligned}&\frac{b_V - a_V}{2N^{2/3}\gamma_V^+} K_{N,V}\left(\lambda_V\left(1 + \frac{s}{N^{2/3}\gamma_V^+}\right), \lambda_V\left(1 + \frac{t}{N^{2/3}\gamma_V^+}\right)\right)\\ &= \begin{cases} \mathbb{A}\mathrm{i}(s,t) + \mathcal{O}\left(N^{-2/3}\right) & , \text{ if } q \le s, t \le 2,\\ \mathbb{A}\mathrm{i}(s,t)\left(1 + \mathcal{O}\left(\frac{s^{5/2}+t^{5/2}}{N^{2/3}}\right)\right) & , \text{ if } 1 \le s, t \le pN^{4/15},\\ \mathbb{A}\mathrm{i}(s,t) + \mathrm{Ai}'(t)\mathcal{O}\left(\frac{t^{3/2}}{N^{2/3}}\right) & , \text{ if } q \le s \le 1,\ 2 \le t \le pN^{4/15}.\end{cases}\end{aligned}$$

The error bounds may depend on the choice of q and p, but they are uniform in V close to Q in part (b).

Observe that in the case $s \le 1$ and $t \ge 2$ the asymptotics of the Airy function [1, 10.4.59, 10.4.61] imply that the Airy-kernel $\mathbb{A}\mathrm{i}(s,t)$ is of order $\mathrm{Ai}'(t)t^{-1}$, except near zeros of $\mathbb{A}\mathrm{i}$, and it is therefore of leading order for $t = o(N^{4/15})$.

The second application is concerned with edge universality for models of the form (1.9). The following results will appear in [30].

Application 2. *Let us define the density*

$$P^h_{N,Q}(x) := \frac{1}{Z^h_{N,Q}} \prod_{i<j} |x_i - x_j|^2 e^{-\sum_{i<j} h(x_i - x_j) - N \sum_{j=1}^N Q(x_j)}. \tag{1.29}$$

Note that $P^h_{N,Q}$ is of the form (1.9) *with $\varphi(t) = |t|^2 e^{-h(t)}$. We assume that Q is even and satisfies* ***(GA)***$_1$ *and that $\alpha_Q := \inf_{t\in\mathbb{R}} Q''(t) > 0$, i.e. that Q is strictly convex. Furthermore, let $R^{h,k}_{N,Q}$ denote the k-th correlation function of $P^h_{N,Q}$ which is defined analogously to* (1.3)*. Then, given any even, real-analytic Schwartz function h, there is an $0 < \alpha^h < \infty$ such that for all Q as above with $\alpha_Q \geq \alpha^h$, we have for some b^h_Q and γ^h_Q*

(1.30)

$$\left| \left(\frac{1}{N^{2/3}\gamma^h_Q} \right)^k R^{h,k}_{N,Q} \left(b^h_Q + \frac{t_1}{N^{2/3}\gamma^h_Q}, \dots, b^h_Q + \frac{t_k}{N^{2/3}\gamma^h_Q} \right) - \det[\mathbb{A}\mathrm{i}(t_i, t_j)]_{1\leq i,j\leq k} \right| \to 0,$$

as $N \to \infty$, uniformly for $t_1, \dots, t_k$ from intervals of the form $[s, \mathcal{O}(N^{4/15})]$.

Moreover, the rescaled largest particle converges in law towards the Tracy-Widom distribution with $\beta = 2$, as $N \to \infty$:

$$\lim_{N\to\infty} P^h_{N,Q} \left(\gamma^h_Q N^{2/3} \left(x_{\max} - b^h_Q \right) \leq s \right) = F_{\mathrm{TW},2}(s) \tag{1.31}$$

for any $s \in \mathbb{R}$. Here $F_{\mathrm{TW},2}$ is the distribution function of the Tracy-Widom distribution with $\beta = 2$ already mentioned in Remark 1.6.

To explain the origin of b^h_Q and γ^h_Q, we recall from [23] the result that under the conditions above there is a probability measure μ^h_Q on $\mathbb{R}$ which describes the global behavior of particles distributed according to (1.29). This measure is the equilibrium measure in the sense of Section 2, minimizing the energy functional $I_{\widetilde{Q}}$ with the implicitly defined external field $\widetilde{Q}(t) := Q(t) + \int h(t-s)\,\mathrm{d}\mu^h_Q(s)$. It is given by $\mu^h_Q = \mu_{\widetilde{Q}}$ in the notation of (2.1) (see also (1.14)) and we may define $b^h_Q := b_{\widetilde{Q}}$ resp. $\gamma^h_Q := \gamma^+_{\widetilde{Q}}$ in the sense of $\mathbf{(GA)}_1(4)$ resp. (1.21).

Remark 1.9. Recall that ensemble (1.29) is not known to be determinantal, therefore convergence of the correlation functions does not follow directly from the convergence of a kernel as provided by Theorem 1.8. The limiting correlations, however, are of determinantal form and coincide in particular with the ones known from the more classical random matrix ensembles.

The analysis of local universality for ensembles of this type has been started in [23], where bulk universality was proved for $\beta = 2$ and the macroscopic correlations have been identified. Bulk universality for general $\beta > 0$ has been addressed in [63].

In the proofs of (1.30) and (1.31), truncations of $\widetilde{Q}$ to compact sets J are performed. In this truncated setting universality results and rates of convergence for ensembles of type (1.2) with external fields of the form $V \equiv \widetilde{Q} + \frac{1}{N} f$ are essential. It is here that general assumptions $\mathbf{(GA)}_2$ are needed, eventually yielding uniformity in the functions f.

The third application deals with universality of the nearest-neighbor spacings for the ensembles (1.29).

Usually, universality is understood in the sense of convergence of the correlation functions to some universal limit. In general, many local statistics of interest can be expressed in terms of correlation functions, but to deduce universality of these statistics from universality of correlation functions, some care is needed. This is

no mere mathematical problem, because in a statistical experiment, one does not observe correlations but e.g. spacings instead. Then the relevant quantity is the counting measure of the nearest-neighbor spacings and it is this object for which limit theorems are needed.

Application 3. *For the ensemble $P^h_{N,Q}$ (see (1.29)), let $a \in \operatorname{supp}(\mu^h_Q)^\circ$, where μ^h_Q is the limiting measure for $P^h_{N,Q}$ as described in Application 2. Let $I_N(a)$ denote an interval with center a such that $|I_N(a)| \to 0$ and $N|I_N(a)| \to \infty$ for $N \to \infty$ hold. Furthermore, we consider the measure $P^h_{N,Q}$ on the Weyl chamber, i.e. we have $x_1 \leq \cdots \leq x_N$. Let $\psi := \mathrm{d}\mu^h_Q/\mathrm{d}x$ denote the density of μ^h_Q and introduce the rescaled position $\tilde{x}_j := N\psi(a)x_j$ of the j-th particle. Then we define the counting measure of spacings around a as*

$$S(I_N(a), x) := \sum_{x_j, x_{j+1} \in I_N(a)} \delta_{\tilde{x}_{j+1} - \tilde{x}_j}.$$

We will prove in [54] *that*

$$\lim_{N\to\infty} \mathbb{E}^h_{N,Q} \left(\sup_{s\in\mathbb{R}} \left| \int_0^s \frac{1}{N|I_N(a)|\psi(a)} \, \mathrm{d}S(I_N(a), x) - F_2(s) \right| \right) = 0, \tag{1.32}$$

where F_2 is the distribution function of the limiting spacing distribution of the Gaussian Unitary Ensemble and $\mathbb{E}^h_{N,Q}$ denotes expectation w.r.t. $P^h_{N,Q}$.

Assertion (1.32) expresses the uniform convergence of the distribution function of the empirical spacings to its universal limit. Here $N|I_N(a)|\psi(a)$ asymptotically normalizes $S(I_N(a), x)$.

The convergence analog to (1.32), initially shown in [26] for circular unitary ensembles, will be proved in [53] for invariant ensembles ($\beta = 1, 2, 4$) (see also [29] for a survey on spacings in invariant ensembles). Prior to [53], for unitarily invariant ensembles only convergence pointwise in s has been shown (see [18], [13]), reading

$$\lim_{N\to\infty} \mathbb{E}_{N,V} \int_0^s \frac{1}{N|I_N(a)|\psi(a)} \, \mathrm{d}S(I_N(a), \lambda) = F_2(s) \tag{1.33}$$

for each $s \in \mathbb{R}$ and with corresponding definitions of $\mathbb{E}_{N,V}$ and ψ. The advantage of uniform convergence as in (1.32) over pointwise convergence as in (1.33), is that it allows for quantile comparison of some empirical spacing distribution to the spacing distribution F_2.

The paper is organized as follows. In Section 2 we collect all information about the equilibrium measure that is needed in our analysis of the Riemann-Hilbert problem which is then performed in Section 3. A somewhat involved construction, which allows to control the singularities that arise at the edges, is shifted to the Appendix. Section 4 is devoted to the proofs of our main results, Theorems 1.3, 1.5, 1.7, and 1.8.

2. The equilibrium measure and its log-transform

The function ρ_V, that we have defined in (1.14), can be viewed as a rescaled version of the density of the equilibrium measure of logarithmic potential theory with respect to the external field V. We do not use this fact directly in our arguments. Nevertheless, let us briefly recall what it means. Assume that V satisfies $\mathbf{(GA)}_1$. It is well known, also under far less restrictive assumptions on V (see e.g. [6]), that the energy functional

$$I_V(\mu) := \int_J \int_J \log|t-s|^{-1} \, \mathrm{d}\mu(s)\mathrm{d}\mu(t) + \int_J V(t) \, \mathrm{d}\mu(t),$$

defined on the Borel measures μ on J with $\int_J \mathrm{d}\mu = 1$, has a unique minimizer μ_V. In the situation of $\mathbf{(GA)}_1$ it can be written in the form

$$\mathrm{d}\mu_V(x) = \hat{\rho}_V(x)\,\mathrm{d}x, \text{ with } \hat{\rho}_V(x) := \frac{2}{b_V - a_V}\rho_V(\lambda_V^{-1}(x)) \tag{2.1}$$

and with λ_V, a_V, b_V, ρ_V as introduced in Definition 1.2. Uniqueness of the minimizer follows from the positive definiteness of the quadratic part of the functional I_V (see e.g. [13, Lemma 6.41]), so that μ_V is characterized as the solution of the Euler-Lagrange equation:

$$\exists l \in \mathbb{R}: \qquad 2\int_J \log|x-t|^{-1}\,\mathrm{d}\mu(t) + V(x) \begin{cases} = l & , \text{ if } x \in \operatorname{supp}(\mu), \\ \geq l & , \text{ if } x \in J \setminus \operatorname{supp}(\mu) \end{cases} \tag{2.2}$$

For the convenience of the reader we sketch the proof that μ_V as defined in (2.1) satisfies (2.2). We reemphasize that it is only (2.2), in the form of Corollary 2.3 below, which is used in the analysis of the subsequent sections. The fact that μ_V minimizes I_V is not relevant in the context of this paper and no proof will be given.

Observe that $\hat{\rho}_V$ as given by (2.1), see also Definition 1.2, is Hölder continuous and therefore

$$\frac{\mathrm{d}}{\mathrm{d}x}\int_J \log|x-t|^{-1}\,\mathrm{d}\mu_V(t) = \pi(H\hat{\rho}_V)(x), \; x \in \mathbb{R}, \tag{2.3}$$

with the Hilbert transform H being defined by

$$Hf(x) := \frac{1}{\pi}\mathrm{PV}\int_J \frac{f(t)}{t-x}\,\mathrm{d}t, \; x \in \mathbb{R}.$$

In the following Lemma 2.1, together with Remark 2.2, we derive an explicit formula for $2\pi(H\hat{\rho}_V) + V'$ on J, from which (2.2) is immediate. We also prove $\int_J \mathrm{d}\mu_V = 1$. The method of proof is taken from [16]; see also [13, Section 6.7] for some motivation of the auxiliary functions that appear in the arguments below. A different method to obtain explicit representations for the equilibrium measure was introduced in [48], see also [51, Section 11.2].

Lemma 2.1. *Let V satisfy assumption $\mathbf{(GA)}_1$ and let ρ_V, λ_V, G_V be given as in Definition 1.2. Then*

(a)

$$\int_{-1}^{1} \rho_V(x)\,\mathrm{d}x = 1 \qquad \textit{and} \qquad \rho_V(x) > 0 \quad \textit{for all } x \in (-1,1).$$

(b)

$$2\pi(H\rho_V)(x) + (V \circ \lambda_V)'(x) = \begin{cases} 0 & , \textit{ if } x \in [-1,1], \\ \operatorname{sgn}(x) G_V(x)\sqrt{x^2-1} & , \textit{ if } x \in \hat{J} \setminus [-1,1]. \end{cases}$$

Remark 2.2. (a) It follows from Lemma 2.1 (a) and from the definition of ρ_V that μ_V as given by (2.1) is a Borel measure on J with support $[a_V, b_V]$ and $\int_J \mathrm{d}\mu_V = 1$. Moreover,

$$2\pi(H\hat{\rho}_V) + V' = \frac{2}{b_V - a_V}\left[2\pi(H\rho_V) + (V \circ \lambda_V)'\right] \circ \lambda_V^{-1}$$

vanishes on $[a_V, b_V]$ and is strictly positive resp. negative to the right resp. left of $[a_V, b_V]$ by statement (b) of the above Lemma. Together with (2.3) this implies that μ_V satisfies the Euler-Lagrange equation (2.2).

(b) Observe that the functional I_V has a unique minimizer also for those V that do not satisfy condition (4) of $\mathbf{(GA)}_1$. In this case the minimizer μ_V still has a density but it is unbounded in the vicinity of at least one endpoint of J. Such an endpoint is called a hard edge. In all cases considered in the present paper, however, the density of μ_V vanishes at the endpoints which is the defining property of a soft edge.

Proof of Lemma 2.1. In the proof we simplify the notation by using $\rho \equiv \rho_V$, $h \equiv h_V$, $G \equiv G_V$ (see (1.14), (1.12), (1.13)). In addition, we introduce the auxiliary function

$$W := V \circ \lambda_V \tag{2.4}$$

that is defined on $\hat{J}$ (see Definition 1.2). For example, we have (1.12)

$$h(t,x) = \frac{W'(t) - W'(x)}{t - x} = \int_0^1 W''(x + u(t-x))\,\mathrm{d}u.$$

It follows from the strict monotonicity of V', which is inherited by W', that h is positive except possibly on the diagonal. Hence $G > 0$ on $\hat{J}$ and $\rho > 0$ on $(-1,1)$. In order to prove the remaining claims, we introduce the unique function q that is analytic on $\mathbb{C} \setminus [-1,1]$ satisfying in addition

$$q(z)^2 = (z^2 - 1) \quad \text{and} \quad q(x) = \sqrt{x^2 - 1} \quad \text{for } x > 1. \tag{2.5}$$

Moreover, restricting q to the upper resp. lower half-plane, we denote by $q_\pm$ the extension to the real line, i.e.

$$q_\pm(x) := \lim_{\varepsilon \searrow 0} q(x \pm i\varepsilon) = \begin{cases} \sqrt{x^2-1} & , \text{ if } x > 1\,, \\ \pm i\sqrt{1-x^2} & , \text{ if } x \in [-1,1]\,, \\ -\sqrt{x^2-1} & , \text{ if } x < -1\,. \end{cases}$$

Furthermore, recall the definition of the Cauchy-transform with respect to $\mathbb{R}$:

$$(Cf)(z) = \frac{1}{2\pi i} \int_{\mathbb{R}} \frac{f(t)}{t - z}\,\mathrm{d}t\,, \quad z \in \mathbb{C} \setminus \mathbb{R}\,. \tag{2.6}$$

The key step in the proof is to relate ρ to the auxiliary function

$$F := q \cdot C\left(\frac{iW'}{\pi q_+} \mathbb{1}_{[-1,1]}\right),$$

extended analytically to $\mathbb{C} \setminus [-1,1]$, where $\mathbb{1}_A$ denotes the characteristic function corresponding to A. Using

$$W'(t) = W'(x) + h(t,x)(t-x),$$

we obtain for $z \in \mathbb{C} \setminus [-1,1]$ that

$$F(z) = \frac{q(z)}{2\pi i}\left[W'(x) \int_{-1}^1 \frac{i}{\pi q_+(t)} \frac{1}{t-z}\,\mathrm{d}t + \int_{-1}^1 \frac{ih(t,x)}{\pi q_+(t)} \frac{t-x}{t-z}\,\mathrm{d}t\right].$$

A straightforward residue calculation (with a vanishing residue at ∞) yields

$$\int_{-1}^1 \frac{i}{\pi q_+(t)} \frac{1}{t-z}\,\mathrm{d}t = -\frac{1}{q(z)}.$$

Choosing $x = \mathrm{Re}(z)$, we arrive by dominated convergence at

$$F_\pm(x) := \lim_{\varepsilon \searrow 0} F(x \pm i\varepsilon) = -\frac{W'(x)}{2\pi i} + \frac{q_\pm(x)}{2\pi i} G(x) \quad \text{for } x \in \hat{J}. \tag{2.7}$$

Recalling the definition of ρ (see (1.14)), we conclude

$$\rho = \mathrm{Re}(F_+) \quad \text{and} \quad 2\rho = F_+ - F_- \quad \text{on } \mathbb{R}, \tag{2.8}$$

where we have used in addition that ρ, $\mathrm{Re}(F_+)$, and $F_+ - F_-$ vanish identically on $\mathbb{R}\setminus[-1,1] \supset \mathbb{R}\setminus\hat{J}$. Now we have reached the only point in the proof where condition (4) of $\mathbf{(GA)}_1$ comes into play. Translated to W it reads

$$\int_{-1}^{1} \frac{W'(s)}{\sqrt{1-s^2}}\,\mathrm{d}s = 0 \quad \text{and} \quad \int_{-1}^{1} \frac{s\,W'(s)}{\sqrt{1-s^2}}\,\mathrm{d}s = 2\pi.$$

As a consequence we can determine the leading order behavior of $F(z)$ for $|z|\to\infty$. It follows from

$$\frac{1}{t-z} = -\frac{1}{z} - \frac{t}{z^2} - \frac{t^2}{z^3} + \frac{t^3}{z^3(t-z)} \tag{2.9}$$

that

$$C\left(\frac{iW'}{\pi q_+}\mathbb{1}_{[-1,1]}\right)(z) = -\frac{1}{2\pi^2 i}\left(\frac{0}{z} + \frac{2\pi}{z^2} + \mathcal{O}\left(|z|^{-3}\right)\right) \quad \text{as } |z|\to\infty.$$

Since $q(z) = z(1+\mathcal{O}(|z|^{-2}))$, for $|z|\to\infty$, we have

$$i\pi F(z) = -\frac{1}{z} + \mathcal{O}\left(|z|^{-2}\right) \quad \text{as } |z|\to\infty. \tag{2.10}$$

A first conclusion of (2.10) is

$$\frac{1}{2\pi i}\int_{\mathbb{R}} \frac{F_+(x)-F_-(x)}{x-z}\,\mathrm{d}x = F(z),\ z\in\mathbb{C}\setminus[-1,1],$$

by a residue calculation (again with a vanishing residue at ∞). Recalling (2.8), we have proved

$$F = 2C\rho. \tag{2.11}$$

Since

$$2\pi i(C\rho)(z) = -\frac{1}{z}\left(\int_{-1}^{1}\rho(x)\,\mathrm{d}x\right) + \mathcal{O}\left(|z|^{-2}\right), \quad |z|\to\infty$$

by (2.9), claim (a) follows from (2.10).

Recall that the Hölder continuity of ρ implies the existence of the pointwise limits

$$(C_\pm\rho)(x) := \lim_{\varepsilon\searrow 0}(C\rho)(x\pm i\varepsilon) = \frac{1}{2}\left[\pm\rho(x) - i(H\rho)(x)\right]. \tag{2.12}$$

Thus, by (2.11) and (2.7),

$$H\rho = -2\,\mathrm{Im}(C_+\rho) = -\,\mathrm{Im}(F_+) = -\frac{W'}{2\pi} + \frac{\mathrm{Re}(q_+)}{2\pi}G$$

on $\hat{J}$. As $\mathrm{Re}(q_+)$ vanishes on $[-1,1]$ and is equal to $\mathrm{sgn}(x)\sqrt{x^2-1}$ on $\mathbb{R}\setminus[-1,1]$, we have proved statement (b).

□

The results of Lemma 2.1 provide information on the log-transform

$$g:\ \mathbb{C}\setminus(-\infty,1] \to \mathbb{C},\ g(z) := \int_{-1}^{1}\log(z-t)\rho_V(t)\,\mathrm{d}t \tag{2.13}$$

of ρ_V that will be essential for the analysis of the Riemann-Hilbert problem in the subsequent sections. This information is summarized in the following corollary.

Corollary 2.3. *Let V satisfy $\mathbf{(GA)}_1$ and let ξ_V, η_V, ρ_V, W, g be given as in Definition 1.2,* (2.4)*, and* (2.13)*. Then g is analytic and there exists $l\in\mathbb{R}$ such that*

$$g_+(x) - g_-(x) = i\xi_V(x) \quad \textit{for } x\in\mathbb{R}, \tag{2.14}$$

$$g_+(x) + g_-(x) = W(x) + l - \eta_V(x) \quad \textit{for } x\in\hat{J}, \tag{2.15}$$

$$\xi_V(x) = 2\pi \quad \textit{for } x \le -1 \qquad \textit{and} \qquad \xi_V(x) = 0 \quad \textit{for } x \ge 1, \tag{2.16}$$

$$g(z) = \log z + \mathcal{O}\left(|z|^{-1}\right) \quad \textit{as } |z| \to \infty. \tag{2.17}$$

Moreover, the function e^g has an analytic extension onto $\mathbb{C}\setminus[-1,1]$.

Proof. The limits $g_\pm$ of g on $\mathbb{R}$ are given by

$$g_\pm(x) = \int_{-1}^{1} \log|x-t|\,\rho_V(t)\,\mathrm{d}t \pm i\pi \int_x^1 \rho_V(t)\,\mathrm{d}t, \ x \in \mathbb{R}.$$

Relations (2.14) and (2.15) now follow from statement (b) of Lemma 2.1 and from

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\int_{-1}^{1} \log|x-t|\,\rho_V(t)\,\mathrm{d}t\right) = -\pi(H\rho_V)(x) \ \text{ for } x \in \mathbb{R} \quad \text{(cf. (2.3))}.$$

The relations in (2.16) follow from Lemma 2.1 (a) and (1.15). The first relation of (2.16) explains the analytic extendability of e^g across $(-\infty,-1)$. The asymptotic formula (2.17) is again a consequence of Lemma 2.1 (a).

□

We conclude this section by proving the first claim of statement (b) in our Theorem 1.3.

Lemma 2.4. *Assume that **(GA)**$_2$ holds for some function Q. Let D, X_D be defined as in Remark 1.1. Then there exists an $\epsilon > 0$ such that any $V \in B_\epsilon(Q) \subset X_D$, restricted to J, satisfies **(GA)**$_1$. Moreover, the maps $V \mapsto a_V$, $V \mapsto b_V$, defined on $B_\epsilon(Q)$ by **(GA)**$_1$ (4) are C^1 with bounded derivatives.*

Proof. Observe first that condition (1) of **(GA)**$_1$ holds because of the definition of X_D and that **(GA)**$_1$ (3) does not pose any condition due to the boundedness of J (see **(GA)**$_2$ (3′)). Since $\|V''\big|_J\|_\infty \le C\|V\|_\infty$ for some $C > 0$ that only depends on the distance between J and $\mathbb{C}\setminus D$, we can ensure **(GA)**$_1$ (2) on $B_\epsilon(Q)$ for ϵ sufficiently small by using **(GA)**$_2$ (2′). In order to investigate **(GA)**$_1$ (4) we introduce

$$\begin{aligned}
&U := \{(a,b) \in (L_-, L_+)^2 \mid a < b\},\\
&\lambda_{a,b}(s) := \frac{b-a}{2}s + \frac{b+a}{2}, \ s \in \mathbb{R}, \ \text{ for } (a,b) \in U \quad \text{(cf. (1.11))},\\
&A: X_D \times U \to \mathbb{R}^2, \ A(V,a,b) := \begin{pmatrix} \int_{-1}^{1} \frac{V'(\lambda_{a,b}(s))}{\sqrt{1-s^2}}\,\mathrm{d}s \\ (b-a)\int_{-1}^{1} \frac{sV'(\lambda_{a,b}(s))}{\sqrt{1-s^2}}\,\mathrm{d}s - 4\pi \end{pmatrix}.
\end{aligned}$$

A short calculation shows that (1.10) holds iff $A(V,a,b) = 0$. Hence, by **(GA)**$_2$, we have $A(Q,a_Q,b_Q) = 0$. Using again that derivatives of $V\big|_J$ are uniformly bounded by $\|V\|_\infty$, it is straightforward to verify that A is a continuously differentiable map with bounded derivatives on bounded domains. By the implicit function theorem we only need to show that $[\partial_{(a,b)}A](Q,a_Q,b_Q)$ has a non-vanishing determinant. This can be seen as follows: Introducing the measure α on $[-1,1]$ with density $\mathrm{d}\alpha/\mathrm{d}s := Q''\left(\lambda_{a_Q,b_Q}(s)\right)/\sqrt{1-s^2}$, denoting its moments by $m_k := \int_{-1}^{1} s^k\,\mathrm{d}\alpha(s)$, and keeping in mind that $A_2(Q,a_Q,b_Q) = 0$ one obtains

$$\left(\partial_{(a,b)}A\right)(Q,a_Q,b_Q) = \begin{pmatrix} \frac{1}{2}(m_0 - m_1) & \frac{1}{2}(m_0 + m_1) \\ -\frac{4\pi}{b_Q - a_Q} + \frac{b_Q - a_Q}{2}(m_1 - m_2) & \frac{4\pi}{b_Q - a_Q} + \frac{b_Q - a_Q}{2}(m_1 + m_2) \end{pmatrix}$$

with determinant $\frac{4\pi}{b_Q-a_Q}m_0+\frac{b_Q-a_Q}{2}\left(m_0m_2-m_1^2\right)$. Since $m_1^2\leq m_0m_2$ by the Cauchy-Schwarz inequality we have derived

$$\det\left[\left(\partial_{(a,b)}A\right)(Q,a_Q,b_Q)\right]\geq\frac{4\pi}{b_Q-a_Q}m_0>0.$$

□

Finally, we note that the proof of Lemma 2.4 requires less restrictive assumptions than $\mathbf{(GA)}_2$. On the one hand, analyticity could be replaced by a finite regularity assumption (e.g. C^3 would easily suffice). Analyticity will only be used in the next section in a crucial way. On the other hand, the strict positivity of Q'' as formulated in $\mathbf{(GA)}_2$ $(2')$ could be replaced by the weaker $\mathbf{(GA)}_1$ (2) and $m_0>0$ used in the proof would still hold. However, in this situation not all $V\in B_\epsilon(Q)$ necessarily satisfy $\mathbf{(GA)}_1$ (2) no matter how small one chooses ϵ. Then it is not a priori clear whether the corresponding functions ρ_V are non-negative. In case there exists a point x_0 with $\rho_V(x_0)<0$, a gap in the support of the equilibrium measure μ_V opens and entirely different formulae for ρ_V, G_V, etc. are needed (see e.g. [16]).

3. The Riemann-Hilbert problem for the Christoffel-Darboux kernel

This section combined with the Appendix contains everything about Riemann-Hilbert problems that is used in this paper. We follow closely the path laid out in [17] (see also [13, Section 7]) and start with the Fokas-Its-Kitaev characterization [21] of orthogonal polynomials by Riemann-Hilbert problems (RHP). The asymptotic analysis is then performed using the Deift-Zhou nonlinear steepest descent method. In all of this section we assume that V satisfies $\mathbf{(GA)}_1$. As in the previous section we find it convenient to begin our analysis with the linear rescaling λ_V (see (1.11)). As in (2.4), let

$$W:=V\circ\lambda_V,\ W:\hat{J}\to\mathbb{R},\ \hat{J}=\lambda_V^{-1}(J)=\{x\in\mathbb{R}\,|\,\hat{L}_-\leq x\leq\hat{L}_+\}. \tag{3.1}$$

From the defining orthogonality relations for $p_j^{(N,V)}$ resp. $p_j^{(N,W)}$ and from the definition of the corresponding Christoffel-Darboux kernels (cf. formulae above (1.5)), one obtains

$$p_j^{(N,W)}=\sqrt{\frac{b_V-a_V}{2}}p_j^{(N,V)}\circ\lambda_V,\ \text{and}$$

$$K_{N,W}(x,y)=\frac{b_V-a_V}{2}K_{N,V}\left(\lambda_V(x),\lambda_V(y)\right),\quad\text{for } x,y\in\hat{J}. \tag{3.2}$$

The following theorem provides the Riemann-Hilbert formulation for the Christoffel-Darboux kernel $K_{N,W}$ (see e.g. [13, 17, 34, 62]).

Theorem 3.1. *Assume that V satisfies $\mathbf{(GA)}_1$ and define W as in (3.1). There exists a unique analytic map*

$$Y:\mathbb{C}\backslash\hat{J}\to\mathbb{C}^{2\times2}$$

satisfying (i)-(iii), where

(i) The restrictions $Y\big|_{\mathbb{C}_\pm}$ of Y to the upper resp. lower half-plane can be extended continuously to maps $Y_\pm$ defined on $\overline{\mathbb{C}}_\pm\backslash\{\hat{L}_\pm\}$ satisfying

$$Y_+(x)=Y_-(x)\begin{pmatrix}1&e^{-NW(x)}\\0&1\end{pmatrix},\quad\text{for } x\in\left(\hat{L}_-,\hat{L}_+\right).$$

(ii) $\lim\limits_{|z|\to\infty}Y(z)\begin{pmatrix}z^{-N}&0\\0&z^N\end{pmatrix}=\begin{pmatrix}1&0\\0&1\end{pmatrix}=:I$

(iii) $Y(z) = \begin{pmatrix} \mathcal{O}(1) & \mathcal{O}\left(|\log|z-\hat{L}_\pm||\right) \\ \mathcal{O}(1) & \mathcal{O}\left(|\log|z-\hat{L}_\pm||\right) \end{pmatrix}$, *for* $z \to \hat{L}_\pm$, *in case* $\hat{L}_\pm$ *is finite.*

The unique solution is given by

$$Y = \begin{pmatrix} \left[\gamma_N^{(N,W)}\right]^{-1} p_N^{(N,W)} & \left[\gamma_N^{(N,W)}\right]^{-1} C\left(p_N^{(N,W)} e^{-NW}\right) \\ -2\pi i \gamma_{N-1}^{(N,W)} p_{N-1}^{(N,W)} & -2\pi i \gamma_{N-1}^{(N,W)} C\left(p_{N-1}^{(N,W)} e^{-NW}\right) \end{pmatrix}, \tag{3.3}$$

where C *denotes the Cauchy transform (see Remark 3.2 below).*

Moreover, $\det Y(z) = 1$ *for all* $z \in \mathbb{C}\backslash\hat{J}$ *and*

$$K_{N,W}(x,y) = e^{-\frac{N}{2}(W(x)+W(y))} \frac{1}{2\pi i (x-y)} \begin{pmatrix} 0 & 1 \end{pmatrix} Y_+^{-1}(y) Y_+(x) \begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{3.4}$$

for $x, y \in \hat{J}$.

Remark 3.2. In order to use the definition of the Cauchy transform as given in (2.6), we extend e^{-NW} by 0 on $\mathbb{R}\backslash\hat{J}$.

A detailed proof of the above theorem in the case $\hat{J} = \mathbb{R}$ can be found in [13]. For the convenience of the reader we sketch the main ideas, emphasizing those aspects that are different in the case of finite $\hat{L}_\pm$.

Sketch of proof of Theorem 3.1. Let us begin with the first column $\begin{pmatrix} f \\ g \end{pmatrix} := \begin{pmatrix} Y_{11} \\ Y_{21} \end{pmatrix}$ of Y. From (i) and (iii) it follows that f and g can be extended to entire functions. By (ii) we learn from a version of Liouville's theorem that f is a polynomial of degree N with leading coefficient 1 and that g is a polynomial of degree $< N$. Applying the basic relation $h = C_+h - C_-h$ (cf. (2.12)) to $h = fe^{-NW}\mathbb{1}_{\hat{J}}$ and $h = ge^{-NW}\mathbb{1}_{\hat{J}}$, it is straightforward to verify that

$$Y = \begin{pmatrix} f & C\left(fe^{-NW}\right) \\ g & C\left(ge^{-NW}\right) \end{pmatrix}$$

has all desired properties, except possibly relation (ii) for the second column, i.e.

$$\lim_{|z|\to\infty} Y(z) \begin{pmatrix} 0 \\ z^N \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{3.5}$$

Using the summation formula for geometric series, one derives

$$C\left(fe^{-NW}\right) = \sum_{j=0}^{N} \frac{i}{2\pi z^{j+1}} \int_{\hat{J}} f(x) x^j e^{-NW(x)} \, \mathrm{d}x + \frac{1}{z^{N+1}} C\left(x^{N+1} fe^{-NW}\right).$$

By the assumptions on V, transferred to W, $C\left(x^{N+1}fe^{-NW}\right)$ is bounded except for neighborhoods of $\hat{L}_\pm$, in case they are finite. Thus, the first component of (3.5) implies $\int_{\hat{J}} f(x) x^j e^{-NW(x)} \, \mathrm{d}x = 0$ for each $0 \le j \le N-1$. As already noted, f is a polynomial of degree N with leading coefficient 1 and therefore $f = \left[\gamma_N^{(N,W)}\right]^{-1} p_N^{(N,W)}$. A similar reasoning leads to $g = \alpha p_{N-1}^{(N)}$, where $\alpha = -2\pi i \gamma_{N-1}^{(N,W)}$ follows from

$$\begin{aligned} 1 = \lim_{|z|\to\infty} Y_{22}(z) z^N &= \frac{i}{2\pi} \int_{\hat{J}} \alpha p_{N-1}^{(N,W)}(x) x^{N-1} e^{-NW(x)} \, \mathrm{d}x \\ &= \frac{i\alpha}{2\pi\gamma_{N-1}^{(N,W)}} \int_{\hat{J}} \left(p_{N-1}^{(N,W)}(x)\right)^2 e^{-NW(x)} \, \mathrm{d}x = \frac{i\alpha}{2\pi\gamma_{N-1}^{(N,W)}}. \end{aligned}$$

So far we have argued that Y as defined by (3.3) indeed solves the RHP (i)-(iii).

To prove uniqueness one may appeal to Liouville's theorem. Let Z denote any solution of RHP (i)-(iii) and set $d(z) := \det(Z(z))$ for $z \in \mathbb{C}\backslash\hat{J}$. Since $d_+(x) = d_-(x)$ on $(\hat{L}_-, \hat{L}_+)$ by (i), and $d(z) = \mathcal{O}(|\log|z - \hat{L}_\pm||)$ for z near $\hat{L}_\pm$ by (iii), d must have an entire extension. Condition (ii) yields $\lim_{|z|\to\infty} d(z) = 1$ from which $d(z) = 1$ follows for all $z \in \mathbb{C}\backslash\hat{J}$. Therefore $M := YZ^{-1}$ defines an analytic function on $\mathbb{C}\backslash\hat{J}$ (Y as given in (3.3)) with $M_+ = M_-$ on $(\hat{L}_-, \hat{L}_+)$, $M(z) = \mathcal{O}(|\log|z - \hat{L}_\pm||)$ near $\hat{L}_\pm$ and $\lim_{|z|\to\infty} M(z) = I$. Using Liouville's theorem again, we see that $M(z) = I$ and $Y(z) = Z(z)$ for all $z \in \mathbb{C}\backslash\hat{J}$.

Finally, we need to verify the formula for the Christoffel-Darboux kernel $K_{N,W}$. From (3.3) and (1.5) we learn

$$K_{N,W}(x,y) = \frac{(Y_{11})_+(x)(Y_{21})_+(y) - (Y_{21})_+(x)(Y_{11})_+(y)}{-2\pi i(x-y)} e^{-\frac{N}{2}(W(x)+W(y))}$$

for $x, y \in \hat{J}$ (recall, that Y_{11} and Y_{21} are polynomials and therefore $Y_{j1} = (Y_{j1})_+$). As observed in the proof of uniqueness above, $\det Y_+(y) = 1$ holds for all $y \in \hat{J}$ and therefore $\begin{pmatrix} 0 & 1 \end{pmatrix} Y_+^{-1}(y) = \begin{pmatrix} (-Y_{21})_+(y) & (Y_{11})_+(y) \end{pmatrix}$, completing the proof. □

Let us now turn to the asymptotic analysis of the Riemann-Hilbert problem formulated in Theorem 3.1. Again, we follow closely [13, 17]. The basic strategy is to perform explicit transformations that lead to an equivalent RHP with jump relations $R_+ = R_-\left(I + \mathcal{O}\left(\frac{1}{N}\right)\right)$ and asymptotic condition $R(z) \to I$ for $|z| \to \infty$. By basic functional analytic considerations one obtains $R(z) = I + \mathcal{O}\left(\frac{1}{N}\right)$, which is sufficient for proving our results.

In order to simplify notation, we suppress the dependency on V when we discuss the above mentioned transformations, i.e. $G \equiv G_V$, $\rho \equiv \rho_V$, $\xi \equiv \xi_V$, $\eta \equiv \eta_V$ (see (1.13) to (1.16)) as well as dependencies on N. Denoting the unique solution of the RHP in Theorem 3.1 by Y, we define

$$T(z) := e^{-N\frac{l}{2}\sigma_3} Y(z) e^{-N\left(g(z)-\frac{l}{2}\right)\sigma_3}, \quad z \in \mathbb{C} \setminus \hat{J}. \tag{3.6}$$

Here l is the Lagrange multiplier (see (2.2), (2.15)), g the log-transform of ρ (2.13), and $\sigma_3 := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ denotes the third Pauli matrix, i.e. $e^{a\sigma_3} = \begin{pmatrix} e^a & 0 \\ 0 & e^{-a} \end{pmatrix}$ for $a \in \mathbb{C}$. Recall also from Corollary 2.3 that e^{Ng} can be viewed as an analytic function on $\mathbb{C} \setminus [-1,1] \supset \mathbb{C} \setminus \hat{J}$. A straightforward calculation yields that T also satisfies a RHP, similar to the one in Theorem 3.1, but with (i) and (ii) replaced by

$$\text{(i)}_T \quad T_+ = T_- v_T \quad \text{on} \quad (\hat{L}_-, \hat{L}_+), \quad \text{where } v_T = \begin{pmatrix} e^{-N(g_+-g_-)} & e^{N(g_++g_--W-l)} \\ 0 & e^{N(g_+-g_-)} \end{pmatrix}$$

$$\text{(ii)}_T \quad \lim_{|z|\to\infty} T(z) = I \quad \text{(use (2.17))}.$$

Observe that T already has the desired asymptotic behavior for $|z| \to \infty$. Note in addition that the results of Section 2, summarized in Corollary 2.3, provide the following representation for the jump matrix v_T,

$$v_T = \begin{cases} \begin{pmatrix} 1 & e^{-N\eta} \\ 0 & 1 \end{pmatrix} & \text{on } \hat{J} \setminus [-1,1], \\ \\ \begin{pmatrix} e^{-iN\xi} & 1 \\ 0 & e^{iN\xi} \end{pmatrix} & \text{on } [-1,1]. \end{cases} \tag{3.7}$$

Since $\eta > 0$ by definition, v_T is of the form $I + \textit{small}$ away from $[-1,1]$. On the critical interval $[-1,1]$ the jump matrix has rapidly oscillating terms on the diagonal.

The nonlinear steepest descent method was designed to handle such jump conditions. In our situation this method is based on the factorization

$$v_T = \begin{pmatrix} e^{-iN\xi} & 1 \\ 0 & e^{iN\xi} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ e^{iN\xi} & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ e^{-iN\xi} & 1 \end{pmatrix} =: v_l v_0 v_u. \tag{3.8}$$

As shown in Corollary 3.6 below (see also Definition 3.5 and Lemma 3.3), ξ has an analytic continuation to a neighborhood of $(-1,1)$ with $(\mathrm{Im}(\xi(z)))\cdot(\mathrm{Im}(z)) < 0$, i.e.

$$v_l = I + \textit{small} \quad \text{below } (-1,1), \qquad v_u = I + \textit{small} \quad \text{above } (-1,1).$$

The basic idea is as follows: Define S as displayed in Figure 1. Observe that S then

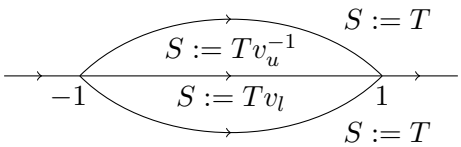

FIGURE 1. Definition of the matrix-valued function S

has jumps across all of the oriented contour displayed in Figure 1. The orientation is indicated by the arrows and it defines the + side of the contour to lie on its left-hand side by standard convention. The jump relations read

$$\begin{aligned} &S_+ = T = (Tv_u^{-1})v_u = S_- v_u \quad \text{on the upper lip} \\ &S_+ = Tv_l = S_- v_l \quad \text{on the lower lip} \\ &S_+ = T_+ v_u^{-1} = T_- v_T v_u^{-1} = T_- v_l v_0 = S_- v_0 \quad \text{on } [-1,1], \end{aligned}$$

where we used (3.8) to derive the last relation. Ignoring the jumps on the two lips and on $\hat{J} \setminus [-1,1]$, because there the corresponding jump matrices are close to I, we may hope that S is close to the matrix M that satisfies

$$M_+ = M_- v_0 \quad \text{on } [-1,1] \quad \text{and} \quad M(z) \to I \quad \text{as } |z| \to \infty. \tag{3.9}$$

This matrix is given by [13, (7.66)-(7.72)]

$$M(z) := \frac{1}{2} \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix} \begin{pmatrix} a(z) & 0 \\ 0 & a^{-1}(z) \end{pmatrix} \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix}, \quad z \in \mathbb{C} \setminus [-1,1], \tag{3.10}$$

where a now denotes the analytic continuation $a(z) = (z-1)^{1/4}(z+1)^{-1/4}$ of the real function defined in (1.17) to $\mathbb{C} \setminus [-1,1]$. As a consequence, T should be close to $\mathbb{T}$ with $\mathbb{T} = Mv_u$ on the upper lens, $\mathbb{T} = Mv_l^{-1}$ on the lower lens, and $\mathbb{T} = M$ outside the lens shaped region of Figure 1. In order to prove that $\mathbb{T}$ is indeed a parametrix for T, i.e. $\mathbb{T}$ is close to T for large values of N, one introduces $R = T\,\mathbb{T}^{-1}$ and shows that the jump matrices corresponding to R are sufficiently close to I so that one can deduce $R = I + \textit{small}$ everywhere by functional analytic means. Then $T = (I + \textit{small})\,\mathbb{T}$ which turns out to be a useful formula for T and via (3.6) also for Y. As we show now, the above described procedure works in principle, but requires a more detailed analysis near ± 1 and near $\hat{L}_\pm$.

We begin by discussing the analytic extensions of some of the quantities of Definition 1.2. One novel aspect of our paper is that we provide formulae and estimates that are uniform for functions V from some open set. This is the reason, why we state the following auxiliary results in the setting of $\mathbf{(GA)}_2$. The simpler case of a single function V, for which $\mathbf{(GA)}_1$ is assumed, is left for remarks thereafter.

Lemma 3.3. *Let Q satisfy* **(GA)**$_2$ *and denote $\hat{J} := \lambda_Q^{-1}(J)$ (see(1.11)). Then there exist $\sigma, d > 0$ and an open neighborhood $\mathcal{U}$ of Q in $(X_D, \|\cdot\|_\infty)$ (see Remark 1.1) such that the analytic continuation of G_V (1.13) satisfies*

$$G_V(z) \in \{w \in \mathbb{C} \,|\, |w| \geq d, \arg(w) \leq \tfrac{\pi}{8}\} \tag{3.11}$$

for all $V \in \mathcal{U}$ and all

$$z \in \hat{J}_\sigma := \{z \in \mathbb{C} \,|\, \mathrm{dist}(z, \hat{J}) \leq \sigma\}.$$

Proof. Since $\hat{J}$ is compact and Q' strictly increasing, it follows that G_Q attains a positive minimum on $\hat{J}$, which we denote by d_0. By Lemma 2.4 we may choose some open neighborhood $\mathcal{U}_0$ of Q and $\tau_0 > 0$ such that

$$\hat{J}_{2\tau_0} \subset \lambda_V^{-1}(D) \quad \text{for all } V \in \mathcal{U}_0.$$

Thus definitions (1.12), (1.13) of h_V and G_V can be extended to $\hat{J}_{2\tau_0}$ simultaneously for all $V \in \mathcal{U}_0$. Denoting by D' the convex hull of $\lambda_V(\hat{J}_{2\tau_0}) \cup \lambda_Q(\hat{J}_{2\tau_0})$, we have

$$\|(V \circ \lambda_V) - (Q \circ \lambda_Q)\|_{\hat{J}_{2\tau_0},\infty} \leq \|V - Q\|_{D,\infty} + \|Q'\|_{D',\infty} \cdot \|\lambda_V - \lambda_Q\|_{\hat{J}_{2\tau_0},\infty}.$$

Using Lemma 2.4 again, one may choose $\mathcal{U} \subset \mathcal{U}_0$ with

$$\|(V \circ \lambda_V) - (Q \circ \lambda_Q)\|_{\hat{J}_{2\tau_0},\infty} < \frac{\tau_0^2 d_0}{8} \quad \text{for all } V \in \mathcal{U}.$$

From the definition of h_V and from the analyticity of $V \circ \lambda_V$ it follows that

$$\|G_Q - G_V\|_{\hat{J}_{\tau_0},\infty} < \frac{d_0}{8}. \tag{3.12}$$

For $z \in \hat{J}_{\tau_0}$ there exists $x \in \hat{J}$ with $|x - z| \leq \tau_0$. Thus,

$$|G_V(z) - G_Q(x)| \leq \frac{d_0}{8} + \|G_Q'\|_{\hat{J}_{\tau_0},\infty} \cdot \tau_0.$$

As $G_Q(x) \in [d_0, \infty)$ by the definition of d_0, the claim follows with

$$\sigma := \min\left(\tau_0, \frac{d_0}{8}\|G_Q'\|^{-1}_{\hat{J}_{\tau_0},\infty}\right) \quad \text{and} \quad d := \frac{3}{4}d_0.$$

□

Remark 3.4. In the case that some function V satisfies **(GA)**$_1$ with a bounded domain of definition J, the proof of Lemma 3.3 implies that the claim of the lemma holds for this particular function V, i.e. with $\mathcal{U}$ being replaced by $\{V\}$. If J is unbounded, the situation is different. Statement (3.11) then holds for all $z \in \tilde{J}_\sigma = \{z \in \mathbb{C} \,|\, \mathrm{dist}(z, \tilde{J}) \leq \sigma\}$, where $\tilde{J}$ is an arbitrary but fixed bounded subset of $\hat{J}$. Observe that $d = d(\tilde{J})$ and $\sigma = \sigma(\tilde{J})$ may depend on the choice of $\tilde{J}$ if no additional assumptions on V are being made. In order to be able to treat the cases of bounded and unbounded sets J simultaneously, we define a standard compact subinterval $\tilde{J}_0$ of $\hat{J}$ by replacing any infinite endpoint $\pm\infty$ of $\hat{J}$ by ± 2. Furthermore, set $\sigma := \sigma(\tilde{J}_0)$ and $d := d(\tilde{J}_0)$.

We are now ready to define analytic extensions of ξ_V and η_V.

Definition 3.5. Denote by

$$\tilde{q}(z) := (1 - z)^{\frac{1}{2}}(1 + z)^{\frac{1}{2}} \tag{3.13}$$

the analytic extension of $t \mapsto \sqrt{1-t^2}$, $t \in (-1,1)$, to $\mathbb{C} \setminus ((-\infty,-1] \cup [1,\infty))$ and recall (2.5) for the definition of the related function q. Set

$$\xi_V(z) := \int_z^1 \tilde{q}(t) G_V(t)\,\mathrm{d}t \quad \text{for } |\mathrm{Re}(z)| < 1,$$

$$\eta_V(z) := \begin{cases} \int_1^z q(t) G_V(t)\,\mathrm{d}t & , \text{ if } \mathrm{Re}(z) > 1, \\ \int_{-1}^z q(t) G_V(t)\,\mathrm{d}t & , \text{ if } \mathrm{Re}(z) < -1. \end{cases}$$

Of course, the domains of ξ_V and η_V are further restricted by the domain of definition of G_V as described in Lemma 3.3 resp. Remark 3.4 above. The reader should verify that ξ_V and η_V agree with the functions defined in (1.15) resp. (1.16) on their common domains, justifying that the same symbols are used.

Corollary 3.6. *Let Q satisfy **(GA)**$_2$ and let $\mathcal{U}$, σ, d, $\hat{J} = [\hat{L}_-, \hat{L}_+]$ and $\hat{J}_\sigma$ be given as in Lemma 3.3. Then for all $V \in \mathcal{U}$ the following holds:*

(a) We have

$$\mathrm{Re}(\eta_V(z)) \geq \frac{2}{3} d |z-1|^{\frac{3}{2}} \quad \textit{for } z \in \hat{J}_\sigma \textit{ with } |\arg(z-1)| \leq \frac{\pi}{16},$$

$$\mathrm{Re}(\eta_V(z)) \geq \frac{2}{3} d |z+1|^{\frac{3}{2}} \quad \textit{for } z \in \hat{J}_\sigma \textit{ with } |\arg(-z-1)| \leq \frac{\pi}{16}.$$

(b) For any compact $K \subset (0,\sigma]$ there exists $c(K) > 0$ such that for all $\delta \in K$:

$$\mathrm{Im}(\xi_V(z)) \begin{Bmatrix} \leq -c(K)|\mathrm{Im}(z)| \\ \geq \quad c(K)|\mathrm{Im}(z)| \end{Bmatrix} \quad \textit{for } |\mathrm{Re}(z)| \leq 1-\delta \textit{ and } \begin{Bmatrix} 0 \leq \mathrm{Im}(z) \leq \delta \\ -\delta \leq \mathrm{Im}(z) \leq 0 \end{Bmatrix}.$$

Proof. (a) We consider the case $\mathrm{Re}(z) > 1$. Using the straight line parametrization $\gamma(t) = 1 + te^{i\arg(z-1)}$, Lemma 3.3 yields the following estimates on the integrand $f(t) := q(\gamma(t))\, G_V(\gamma(t))\, \gamma'(t)$ in the definition of η_V

$$|f(t)| \geq \sqrt{2t}\, d, \qquad |\arg(f(t))| \leq \frac{2\pi}{16} + \frac{\pi}{8} = \frac{\pi}{4}$$

for all $0 \leq t \leq |z-1|$. From this the claim can be derived without effort.

(b) The proof is similar to (a). However, one should split the path of integration into two line segments, from z to $x := \mathrm{Re}(z)$ and from x to 1. Since $\mathrm{Im}(\xi_V(x)) = 0$, one has

$$\mathrm{Im}(\xi_V(z)) = \mathrm{Re}\left(\int_{\mathrm{Im}(z)}^0 f(t)\,\mathrm{d}t \right), \quad \text{with } f(t) = \tilde{q}(x+it) G_V(x+it).$$

The claim now follows from Lemma 3.3 and from the estimates $|\arg \tilde{q}(x+it)| \leq \frac{\pi}{8}$ and $|\tilde{q}(x+it)| \geq \sqrt{1-(1-\delta_K)^2}$ with $\delta_K := \min(K) > 0$. □

Remark 3.7. If V satisfies **(GA)**$_1$ the following version of Corollary 3.6 holds true and is used in the subsequent analysis. For statement (b) σ is chosen as explained in Remark 3.4. For statement (a) and unbounded sets J we replace $\hat{J}$ by arbitrary bounded subsets $\tilde{J}$ of $\hat{J}$. In this situation σ and d depend on $\tilde{J}$ (cf. Remark 3.4).

Our next step is to define the parametrix $\mathbb{T}$ explicitly. Recall from the discussion below (3.10) that different formulae are needed for different parts of $\mathbb{C}$. The relevant subdivision of the complex plane into regions I, II_u, II_l, $\mathrm{III}^\pm$, $\mathrm{IV}^\pm$ is displayed in Figure 2. Note that regions $\mathrm{IV}^\pm$ only come into play if $\hat{L}_\pm$ is finite.

The construction of $\mathbb{T}$ depends on two parameters δ and ε which denote the radii of the circles around ± 1 resp. around $\hat{L}_\pm$. The line segments Σ_2^u, Σ_2^l are chosen to be parallel to the real axis and to begin at $-1 + \delta e^{\pi i/4}$ resp. $-1 + \delta e^{-\pi i/4}$. Dotted lines are used for those parts of the boundaries where T and $\mathbb{T}$ satisfy exactly the

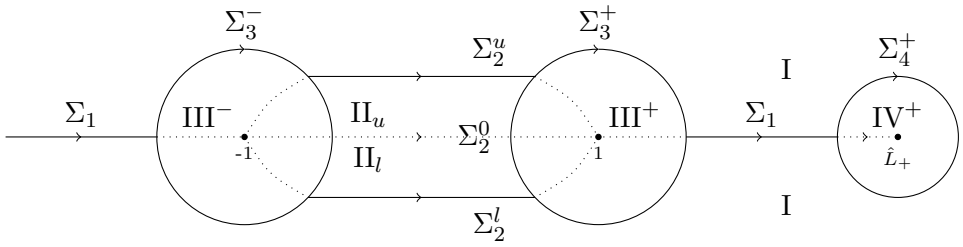

FIGURE 2. Subdivision of $\mathbb{C}$ for the definition of the parametrix $\mathbb{T}$ in the case that J is bounded above and unbounded below.

same jump relation implying that $R = T\,\mathbb{T}^{-1}$ as defined in (3.21) has an analytic continuation across those lines. The definitions of the locations of the curved dotted lines within regions $\mathrm{III}^{\pm}$ and the definition of the parametrix $\mathbb{T}$ in these regions are somewhat involved. They are transferred to the Appendix (see (A.16), (A.20), and Figure 3).

We begin the definition of $\mathbb{T}$ by fixing a common upper bound σ_0 for the radii δ and ε of the circles that appear in the construction. Let σ be given by Lemma 3.3 resp. as in Remark 3.4 and let $0 < \hat{\sigma} < \sigma$ be defined through Lemma A.1. Set

$$\sigma_0 := \min\{\hat{\sigma}/10,\ (\hat{L}_+ - 1)/6,\ (-1 - \hat{L}_-)/6\}. \tag{3.14}$$

From now on we always assume $\delta,\ \varepsilon \in (0, \sigma_0]$. Moreover, we suppress the V-dependency in the notation, i.e. $\xi \equiv \xi_V$, $\eta \equiv \eta_V$. The definition of the parametrix $\mathbb{T}$ in the regions I, II_u, II_l has already been motivated in the paragraph below (3.10). Accordingly, we set

$$\mathbb{T}(z) := M(z) \quad \text{for } z \in \mathrm{I}\,, \tag{3.15}$$

$$\mathbb{T}(z) := M(z) v_u(z) = M(z) \begin{pmatrix} 1 & 0 \\ e^{-iN\xi(z)} & 1 \end{pmatrix} \quad \text{for } z \in \mathrm{II}_u\,, \tag{3.16}$$

$$\mathbb{T}(z) := M(z) v_l^{-1}(z) = M(z) \begin{pmatrix} 1 & 0 \\ e^{iN\xi(z)} & 1 \end{pmatrix}^{-1} \quad \text{for } z \in \mathrm{II}_l\,. \tag{3.17}$$

As mentioned above, the choice for the parametrix in $\mathrm{III}^{\pm}$ is presented in the Appendix, see (A.16), (A.20). For regions $\mathrm{IV}^{\pm}$ we first define the function

$$b(z) := \begin{cases} \dfrac{1}{2\pi i} \displaystyle\int_{\hat{L}_-}^{\hat{L}_- + 2\sigma_0} \frac{e^{-N\eta(t)}}{t - z}\,\mathrm{d}t & , \text{ if } z \in \mathrm{IV}^- := B_\varepsilon(\hat{L}_-) \setminus [\hat{L}_-, \hat{L}_- + \varepsilon)\,, \\ \dfrac{1}{2\pi i} \displaystyle\int_{\hat{L}_+ - 2\sigma_0}^{\hat{L}_+} \frac{e^{-N\eta(t)}}{t - z}\,\mathrm{d}t & , \text{ if } z \in \mathrm{IV}^+ := B_\varepsilon(\hat{L}_+) \setminus (\hat{L}_+ - \varepsilon, \hat{L}_+]\,. \end{cases} \tag{3.18}$$

It is a basic property of the Cauchy transform (cf. (2.12)) that

$$b_+ - b_- = e^{-N\eta} \tag{3.19}$$

on $(\hat{L}_-, \hat{L}_- + \varepsilon) \cup (\hat{L}_+ - \varepsilon, \hat{L}_+)$. Set

$$\mathbb{T}(z) := M(z) \begin{pmatrix} 1 & b(z) \\ 0 & 1 \end{pmatrix} \quad \text{for } z \in \mathrm{IV}^+ \cup \mathrm{IV}^-. \tag{3.20}$$

As argued above, T and $\mathbb{T}$ are compared by deriving a RHP for

$$R := T\,\mathbb{T}^{-1}. \tag{3.21}$$

The essential information is provided by the following

Lemma 3.8. *(a) Let V satisfy* **(GA)**$_1$ *and let R be defined as in* (3.21)*, see also* (3.14) *to* (3.20), (A.16), *and* (A.20)*. Then R has an analytic extension to $\mathbb{C}\setminus\Sigma_R$, where*

$$\Sigma_R := \Sigma_1 \cup \Sigma_2^u \cup \Sigma_2^l \cup \Sigma_3^+ \cup \Sigma_3^- \cup \Sigma_4^+ \cup \Sigma_4^- \,, \quad \textit{see Figure 2.}$$

Moreover,

$$\text{(i)}_R \qquad R_+ = R_- v_R, \quad \textit{with} \quad v_R = I + \Delta_R \textit{ and}$$

$$\|\Delta_R\|_{L^1(\Sigma_R)} + \|\Delta_R\|_{L^\infty(\Sigma_R)} = \mathcal{O}\left(N^{-1}\right), \tag{3.22}$$

$$\text{(ii)}_R \qquad \lim_{|z|\to\infty} R(z) = I.$$

The error bound $\mathcal{O}(N^{-1})$ in (3.22) *is uniform for (δ,ε) from an arbitrary but fixed compact subset of $(0,\sigma_0]^2$.*

(b) Assume that Q satisfies **(GA)**$_2$*. Then there exists an open neighborhood $\mathcal{U}$ of Q in $(X_D, \|\cdot\|_\infty)$ (see Remark 1.1) such that statement (a) holds for all $V\in\mathcal{U}$. The error bound in* (3.22) *is in addition uniform in $V\in\mathcal{U}$.*

Proof. The claim that R has an analytic continuation within $\mathrm{III}^+\cup\mathrm{III}^-$, as well as the estimate

$$\|\Delta_R\|_{L^\infty(\Sigma_3^+\cup\Sigma_3^-)} = \mathcal{O}\left(N^{-1}\right),$$

together with the stated uniformity of the error bound are proved in Lemma A.4. The following representations for

$$v_R = R_-^{-1}R_+ = \mathbb{T}_- T_-^{-1} T_+ \mathbb{T}_+^{-1} = \mathbb{T}_- v_T \mathbb{T}_+^{-1} \tag{3.23}$$

are straightforward from (3.7) and from the explicit definition of $\mathbb{T}$ above. Observe that $v_T = I$ on those parts of Σ_R, where T has no jumps. We obtain

$$\text{on } \Sigma_1: \quad v_R = M v_T M^{-1} = I + M \begin{pmatrix} 0 & e^{-N\eta} \\ 0 & 0 \end{pmatrix} M^{-1}, \tag{3.24}$$

$$\text{on } \Sigma_2^u: \quad v_R = M v_u M^{-1} = I + M \begin{pmatrix} 0 & 0 \\ e^{-iN\xi} & 0 \end{pmatrix} M^{-1}, \tag{3.25}$$

$$\text{on } \Sigma_2^0: \quad v_R = M_- v_l^{-1} v_T v_u^{-1} M_+^{-1} = M_- v_0 M_+^{-1} = I \text{ by } (3.8), (3.9), \tag{3.26}$$

$$\text{on } \Sigma_2^l: \quad v_R = M v_l M^{-1} = I + M \begin{pmatrix} 0 & 0 \\ e^{iN\xi} & 0 \end{pmatrix} M^{-1}, \tag{3.27}$$

$$\text{on } \Sigma_4^+\cup\Sigma_4^-: \quad v_R = M \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} M^{-1} = I + M \begin{pmatrix} 0 & b \\ 0 & 0 \end{pmatrix} M^{-1}, \tag{3.28}$$

on $(\hat{L}_-, \hat{L}_- + \varepsilon) \cup (\hat{L}_+ - \varepsilon, \hat{L}_+)$:

$$\begin{aligned} v_R &= M \begin{pmatrix} 1 & b_- \\ 0 & 1 \end{pmatrix} v_T \begin{pmatrix} 1 & b_+ \\ 0 & 1 \end{pmatrix}^{-1} M^{-1} \\ &= M \begin{pmatrix} 1 & e^{-N\eta} - (b_+ - b_-) \\ 0 & 1 \end{pmatrix} M^{-1} = I \quad \text{by (3.19).} \end{aligned} \tag{3.29}$$

The analytic extendibility of R across Σ_2^0 follows from (3.26). Relation (3.29) implies that R has an analytic continuation on $B_\varepsilon(\hat{L}_\pm)\setminus\{\hat{L}_\pm\}$. Since T and $\mathbb{T}$ are both bounded in the first column and are bounded by $\mathcal{O}(|\log|z-\hat{L}_\pm||)$ in the second column, we have $R(z) = \mathcal{O}(|\log|z-\hat{L}_\pm||)$ for $z\to\hat{L}_\pm$. The singularities at $\hat{L}_\pm$ must therefore be removable by the Riemann Continuation Theorem.

The L^∞-part of estimate (3.22) is a consequence of the boundedness of M on $\mathbb{C}\setminus(B_\delta(1)\cup B_\delta(-1))$ (with a bound that depends on δ), Corollary 3.6, and Remark 3.7. For $s\in\Sigma_4^\pm$ close to $\hat{L}_\pm \mp \varepsilon$ one may resolve the difficulty due to the singularity of the Cauchy kernel by deforming the path of integration (in the definition of the

function b) away from the real axis so that its distance to s is at least, say $\frac{\varepsilon}{2}$. Here we use that the estimates of Corollary 3.6 (a) also apply away from the real axis. Observe that it follows from the definition of σ_0 that the deformed contour lies in the range of applicability of Corollary 3.6 (a).

The L^1-part of estimate (3.22) is implied by the L^∞-estimate, except in the case that Σ_1 contains an unbounded component. This only occurs in the situation of $\mathbf{(GA)}_1$ with unbounded J. Suppose e.g. that $L_+ = \infty$. It follows from the strict monotonicity of W' that for $x \geq 2$ and $t \in [-1, 1]$ one has (see (1.12), (1.13), and (1.16))

$$h(t,x) \geq \frac{2}{3x}(W'(2) - W'(1)) \quad \text{and} \quad \eta(x) \geq \eta(2) + \frac{W'(2) - W'(1)}{2}(x-2)\,,$$

yielding the desired L^1-bound.

Claim $(\mathrm{ii})_R$ follows from $(\mathrm{ii})_T$ below (3.6), from $\lim_{|z|\to\infty} M(z) = I$, and from the definition of R in (3.21). □

We are now ready to formulate the main result of this section. As demonstrated in the subsequent section, it provides estimates on $R - I$ that suffice to prove all the theorems stated in the Introduction. The theorem follows from Lemma 3.8 and from some functional analytic argeuments that we quote from [13, Section 7.5].

Theorem 3.9. *(a) Let V satisfy $\mathbf{(GA)}_1$ and let R be defined as in* (3.21)*, see also* (3.14)–(3.20)*,* (A.16)*, and* (A.20)*. Then for all x, $y \in \hat{J}$,*

$$R_+(x) - I = \mathcal{O}\left(N^{-1}\right), \quad R_+'(x) = \mathcal{O}\left(N^{-1}\right), \text{ and} \tag{3.30}$$

$$R_+(y)^{-1}R_+(x) = I + \mathcal{O}\left(\frac{|x-y|}{N}\right). \tag{3.31}$$

The error bounds are uniform for (δ, ε) in compact subsets of $(0, \sigma_0/2]^2$ and for x, y in bounded subsets of $\hat{J}$.

(b) Assume that Q satisfies $\mathbf{(GA)}_2$. Then there exists an open neighborhood $\mathcal{U}$ of Q in $(X_D, \|\cdot\|_\infty)$ (see Remark 1.1) such that statement (a) holds for all $V \in \mathcal{U}$. The error bounds in (3.30)*,* (3.31) *are in addition uniform in $V \in \mathcal{U}$.*

Proof. The RHP for R that is stated in Lemma 3.8 is equivalent to a singular integral equation (see Section 7.5 in [13] for a detailed description). We apply [13, Theorem 7.103] with $m \equiv R$, $b_- \equiv I$, $b_+ \equiv v_R$, $w_- \equiv 0$, $w = w_+ \equiv \Delta_R$ (see Lemma 3.8). The Cauchy operator C has to be defined on Σ_R, i.e.

$$(Cf)(z) = \frac{1}{2\pi i}\int_{\Sigma_R} \frac{f(t)}{t-z}\,\mathrm{d}t.$$

Theorem 7.103 in [13] relates the RHP for m to the integral operator C_w that reads

$$C_w : L^2(\Sigma_R; \mathbb{C}^{2\times 2}) \to L^2(\Sigma_R; \mathbb{C}^{2\times 2}), \quad f \mapsto C_-(f\Delta_R)$$

in our situation. Since C_- is a bounded operator on $L^2(\Sigma_R) \equiv L^2(\Sigma_R; \mathbb{C}^{2\times 2})$ and as $\|\Delta_R\|_{L^\infty(\Sigma_R)} = \mathcal{O}(N^{-1})$ by Lemma 3.8, we have

$$\|C_w\|_{\mathrm{Op}} = \mathcal{O}\left(N^{-1}\right), \tag{3.32}$$

where $\|\cdot\|_{\mathrm{Op}}$ denotes the operator norm. Albeit the constant function $z \mapsto I$ might not lie in $L^2(\Sigma_R)$, $C_w(I) = C_-(\Delta_R)$ exists, because $\Delta_R \in L^2(\Sigma_R)$. Moreover,

$$\|\Delta_R\|_{L^2} \leq \left(\|\Delta_R\|_{L^1}\|\Delta_R\|_{L^\infty}\right)^{1/2} \tag{3.33}$$

together with (3.22) implies

$$\|C_-(\Delta_R)\|_{L^2(\Sigma_R)} = \mathcal{O}\left(N^{-1}\right). \tag{3.34}$$

The relevant singular integral equation [13, (7.104)] then reads $(\mathbb{1}-C_w)\mu = I$, where $\mathbb{1}$ denotes the identity on $L^2(\Sigma_R)$. In fact, μ is of the form $\mu = I + \tilde{\mu}$, $\tilde{\mu} \in L^2(\Sigma_R)$, and the proper equation for $\tilde{\mu}$ is given by

$$(\mathbb{1} - C_w)\tilde{\mu} = C_w(I) = C_-(\Delta_R), \qquad \text{see [13, (7.105)].} \tag{3.35}$$

From (3.32) it follows that $\|C_w\|_{\mathrm{Op}} \leq \frac{1}{2}$ for N sufficiently large which implies the invertibility of $\mathbb{1} - C_w$. Using also (3.34), we conclude that (3.35) has a unique solution $\tilde{\mu} \in L^2(\Sigma_R)$ with

$$\|\tilde{\mu}\|_{L^2(\Sigma_R)} = \mathcal{O}\left(N^{-1}\right). \tag{3.36}$$

Theorem 7.103 in [13] provides a formula for R in terms of the solution $\tilde{\mu}$ of integral equation (3.35):

$$R = I + C(\Delta_R + \tilde{\mu}\Delta_R).$$

Observe that (3.22), (3.36), and (3.33) yield

$$\|\Delta_R + \tilde{\mu}\Delta_R\|_{L^1(\Sigma_R)} = \mathcal{O}\left(N^{-1}\right).$$

By the definition of the Cauchy operator C we have

$$|R(z) - I| \leq \frac{1}{2\pi}\frac{\|\Delta_R + \tilde{\mu}\Delta_R\|_{L^1(\Sigma_R)}}{\operatorname{dist}(z, \Sigma_R)},$$
$$|R'(z)| \leq \frac{1}{2\pi}\frac{\|\Delta_R + \tilde{\mu}\Delta_R\|_{L^1(\Sigma_R)}}{\operatorname{dist}(z, \Sigma_R)^2}$$

and claim (3.30) follows for $x \in \hat{J}$ that have at least distance, say $\frac{\delta}{10}$, from Σ_R.

We still need to consider those $x \in \hat{J}$ with $\operatorname{dist}(x, \Sigma_R) < \frac{\delta}{10}$. For example, let $x = 1 - \frac{101}{100}\delta$ with distance $\frac{\delta}{100}$ from Σ_3^+. In this case we change the parameter δ in the construction, i.e. the radius of the circles around ± 1, to $\tilde{\delta} := \frac{9}{10}\delta$. Note that by definition the corresponding parametrices $\mathbb{T}$ and $\tilde{\mathbb{T}}$ agree at the given point x, thus $\tilde{R}(x) := T(x)\tilde{\mathbb{T}}^{-1}(x) = T(x)\mathbb{T}^{-1}(x) = R(x)$. We may apply Lemma 3.8 to $\tilde{R}$ to derive

$$\|\Delta_{\tilde{R}} + \tilde{\mu}_{\tilde{R}}\Delta_{\tilde{R}}\|_{L^1(\Sigma_{\tilde{R}})} = \mathcal{O}\left(N^{-1}\right)$$

with the desired uniformity properties of the error bound. As $\operatorname{dist}(x, \Sigma_{\tilde{R}}) \geq \frac{\delta}{10}$, we obtain the desired estimates on $|\tilde{R} - I|$ and on $|\tilde{R}'(x)|$. Since R agrees with $\tilde{R}$ on a small neighborhood of x, we have $\tilde{R}'(x) = R'(x)$ and (3.30) is established for $x = 1 - \frac{101}{100}\delta$. More generally, such a procedure of shrinking or enlarging the disks around ± 1, $\hat{L}_\pm$ allows to prove (3.30) for all $x \in \hat{J}\backslash\Sigma_1$ (see Figure 2) by ensuring that $\operatorname{dist}(x, \Sigma_{\tilde{R}}) \geq \min(\varepsilon, \delta)/10$. Observe that the restriction of δ, ε to the size $\sigma_0/2$ in the statement of the theorem allows to enlarge circles without leaving the domain that is covered by Lemma 3.8.

In the case $x \in \Sigma_1$ we may assume that again the distance between x and any of the circles is at least $\min(\epsilon, \delta)/10 =: \kappa_0$ by shrinking the circles if necessary. Recall that v_R is of the form (3.24) on $[x - \kappa_0, x + \kappa_0]$. It is claimed in the theorem that the error bounds are uniform for x in bounded sets $B \subset J$. This is relevant only in the case of unbounded sets J in the situation of $\mathbf{(GA)}_1$, otherwise we choose $B := J$. Denote further $\tilde{J} := \lambda_V^{-1}(B)$, i.e. $\tilde{J} = \hat{J}$ for bounded sets J. According to Remark 3.7 the estimates of Corollary 3.6 (a) will hold in all cases considered, if $z \in \hat{J}_\sigma$ is replaced by $z \in \tilde{J}_{\tilde{\sigma}}$ with $\tilde{\sigma} := \sigma(\tilde{J})$ (cf. Remark 3.4). Set $\kappa := \min(\kappa_0, \tilde{\sigma})/2$. Thus, v_R possesses an analytic continuation on a neighborhood of $\overline{B_\kappa(x)}$. We may deform

the contour of jumps away from x by defining

$$\tilde{R}(z) := \begin{cases} R(z)\, v_R(z) & \text{, if } |z-x| < \kappa \text{ and } \operatorname{Im}(z) < 0, \\ R(z) & \text{, else.} \end{cases}$$

Then $\tilde{R}$ has an analytic continuation across $(x-\kappa, x+\kappa)$ and the jump relation $\tilde{R}_+ = \tilde{R}_- v_R$ is moved to the lower half-circle $\partial B_\kappa(x) \cap \{z \in \mathbb{C} | \operatorname{Im}(z) < 0\}$ (cf. [17, Figure 7.8]). Observe further that any $z \in \overline{B_\kappa(x)}$ satisfies $|\arg(z-1)| < \frac{\pi}{16}$ if $x > 0$, resp. $|\arg(-1-z)| < \frac{\pi}{16}$ if $x < 0$ by construction. Using the lower bounds for $\operatorname{Re}(\eta_V(z))$ provided by statement (a) of Corollary 3.6 in combination with Remark 3.7 concludes the proof of (3.30). Relation (3.31) follows from (3.30) and

$$R_+(y)^{-1} R_+(x) = I + R_+(y)^{-1}(R_+(x) - R_+(y)).$$

□

4. Proofs of Main Results

We begin with the proof of our basic Theorem 1.3 about the Christoffel-Darboux kernel from which the other main results of this paper, Theorems 1.5, 1.7, and 1.8, will be derived thereafter.

We follow and further streamline the formalism introduced in [62]. In most of the section we again suppress the V-dependency of various quantities for the sake of readability.

Proof of Theorem 1.3: Recall relations (3.2), (3.4) from the previous section, which imply

$$\frac{b_V - a_V}{2} K_{N,V}(\lambda_V(x), \lambda_V(y)) = \frac{e^{-\frac{N}{2}(W(x)+W(y))}}{2\pi i (x-y)} \begin{pmatrix} 0 & 1 \end{pmatrix} Y_+^{-1}(y)\, Y_+(x) \begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{4.1}$$

for $x, y \in \hat{J}$ with $x \neq y$. Moreover, (3.6) and (3.21) yield a formula for the solution Y of the corresponding Riemann-Hilbert problem, which reads

$$Y_+ = e^{N\frac{l}{2}\sigma_3} R_+ \mathbb{T}_+ e^{N\left(g_+ - \frac{l}{2}\right)\sigma_3} \quad \text{on } \hat{J},$$

where the parameters in the construction of $\mathbb{T}$ are chosen to be δ as given by the statement of Theorem 1.3 with $\delta_0 := \sigma_0/2$ and $\varepsilon := \sigma_0/2$ (see (3.14)). We find it convenient to rewrite

$$Y_+ = e^{\frac{N}{2}W} e^{N\frac{l}{2}\sigma_3} R_+ A F_0 \quad \text{with} \tag{4.2}$$

$$A := \frac{1}{\sqrt{2}} e^{\frac{\pi i}{4}} \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix} \quad \text{and} \quad F_0 := e^{-\frac{N}{2}W} A^{-1} \mathbb{T}_+ e^{N\left(g_+ - \frac{l}{2}\right)\sigma_3}. \tag{4.3}$$

Recall further that $\det Y = 1$ everywhere (see Theorem 3.1) and that $\begin{pmatrix} 0 & 1 \end{pmatrix} Y_+^{-1}$ and formula (4.1) for the Christoffel-Darboux kernel only depend on the entries of the first column of Y_+. This observation allows us to replace F_0 in (4.2) by any F that has the same first column as F_0. Defining

$$F(x) := \begin{cases} A^{-1}\mathbb{T}_+(x)\, e^{\frac{N}{2}(2g_+(x) - l - W(x))\sigma_3} & \text{, if } x \in \hat{J} \setminus \left(\mathrm{IV}^+ \cup \mathrm{IV}^-\right), \\ A^{-1}\mathbb{T}_+(x) \begin{pmatrix} 1 & -b(x) \\ 0 & 1 \end{pmatrix} e^{\frac{N}{2}(2g_+(x) - l - W(x))\sigma_3} & \text{, if } x \in \hat{J} \cap \left(\mathrm{IV}^+ \cup \mathrm{IV}^-\right), \end{cases} \tag{4.4}$$

(see Figure 2 and (3.18)) one obtains from (4.2)

$$Y_+ \begin{pmatrix} 1 \\ 0 \end{pmatrix} = e^{\frac{N}{2}W} e^{N\frac{l}{2}\sigma_3} R_+ A F \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \tag{4.5}$$

For the next computation it is essential that the determinants of A, R_+, Y_+, and F are all equal to 1 on $\hat{J}$. For A and Y_+ this follows directly from (4.3) resp. from Theorem 3.1. In the case of F and R_+ one concludes from (4.4), (3.21), (3.6), and Theorem 3.1 that it suffices to show that $\det \mathbb{T} = 1$. Except for the regions $\mathrm{III}^{\pm}$ this is immediate from the definitions (3.10), (3.15) to (3.17), and (3.20). In the disks $\mathrm{III}^{\pm}$ relations (A.9), (A.16), (A.18), (A.20), and (A.11) imply that one needs to verify that $\det \psi_0 = (2\pi)^{-1} e^{\pi i/6}$. To this end observe first that by definition (A.5) ψ_0 is given as Wronskians for the linear second order differential equation $w'' = zw$ (see [1, 10.4.1]) and its determinant is therefore constant on the upper and lower half-plane. The desired result then follows from the asymptotic formula (A.7).

Using the relation

$$X^{-1} = \frac{1}{\det X} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} X^T \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

for invertible 2×2 matrices X twice, one obtains from (4.5)

$$\begin{aligned} \begin{pmatrix} 0 & 1 \end{pmatrix} Y_+^{-1} &= \left(Y_+ \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right)^T \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \\ &= e^{\frac{N}{2} W} \left(F \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right)^T \left(e^{N \frac{l}{2} \sigma_3} R_+ A \right)^T \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \\ &= e^{\frac{N}{2} W} \left(F \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right)^T \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} A^{-1} R_+^{-1} e^{-N \frac{l}{2} \sigma_3}. \end{aligned} \tag{4.6}$$

Finally, we set

$$k := \frac{1}{\sqrt{2\pi}} e^{\frac{\pi i}{4}} F \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{on } \hat{J}. \tag{4.7}$$

Then (4.1), (4.6), (4.5), and Theorem 3.9 imply

$$\begin{aligned} &\frac{b_V - a_V}{2} K_{N,V} \left(\lambda_V(x), \lambda_V(y) \right) \\ &= \frac{1}{x-y} k(y)^T \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} A^{-1} R_+(y)^{-1} R_+(x) A k(x) \\ &= \frac{k_1(x) k_2(y) - k_2(x) k_1(y)}{x-y} + k(y)^T \mathcal{O}\left(N^{-1}\right) k(x). \end{aligned}$$

Moreover, the bounds on the 2×2 matrix that are denoted by $\mathcal{O}\left(N^{-1}\right)$ in the above formula have all the uniformity properties stated in Theorem 1.3 since they agree with those formulated in Theorem 3.9. Observe further, that R_+ is differentiable by Theorem 3.9. This allows to extend the last relation to the diagonal $x = y$.

In order to complete the proof of Theorem 1.3 one needs to verify that k as defined by (4.7) and (4.4) is the same as the definition of k in the statement of Theorem 1.3. This can be done by straightforward computations, using definitions (4.3), (3.10), (3.15) to (3.18), (3.20), (A.1), (A.5), (A.11), (A.9), (A.14), (A.16), (A.17), (A.18), (A.20) together with (A.2) that relates φ to ξ and η. Note also, that Corollary 2.3 implies

$$e^{\frac{N}{2}(2g_+(x) - l - W(x))\sigma_3} = \begin{cases} e^{i \frac{N}{2} \xi(x) \sigma_3} & \text{for } x \in [-1, 1], \\ \operatorname{sgn}(x)^N e^{-\frac{N}{2} \eta(x) \sigma_3} & \text{for } x \in \hat{J} \setminus [-1, 1]. \end{cases}$$

□

The following proposition provides useful formulae for the leading order term of the Christoffel-Darboux kernel. Its proof is entirely elementary using relations such as $\frac{a}{b} - \frac{b}{a} = (\frac{1}{a} + \frac{1}{b})(a - b)$ and $2\cos(\alpha + \frac{\pi}{4})\cos(\beta - \frac{\pi}{4}) = \cos(\alpha + \beta) + \sin(\beta - \alpha)$.

Proposition 4.1. *Let k be defined as in Theorem 1.3. Then*

$$\frac{k_1(x)k_2(y) - k_2(x)k_1(y)}{x-y}$$

equals

(a) $$\frac{\operatorname{sgn}(xy)^N}{4\pi} e^{-\frac{N}{2}(\eta(x)+\eta(y))} \left(\frac{1}{a(x)} + \frac{1}{a(y)}\right) \frac{a(x)-a(y)}{x-y}$$
for $x, y \in \hat{J} \setminus [-1-\delta, 1+\delta]$,

(b) $$\frac{1}{2\pi}\left(\frac{\hat{a}(x)}{\hat{a}(y)} + \frac{\hat{a}(y)}{\hat{a}(x)}\right) \frac{\sin\left(N\pi \int_y^x \rho(s)\,\mathrm{d}s\right)}{x-y}$$
$$+ \frac{1}{2\pi}\cos\left(\frac{N}{2}\left(\xi(x)+\xi(y)\right)\right)\left(\frac{1}{\hat{a}(x)} + \frac{1}{\hat{a}(y)}\right)\frac{\hat{a}(x)-\hat{a}(y)}{x-y}$$
for $x, y \in (-1+\delta, 1-\delta)$,

(c) $$\mathbb{A}\mathrm{i}\left(f_N(x), f_N(y)\right)\frac{f_N(x)-f_N(y)}{x-y}$$
$$+ \left(\frac{\mathrm{Ai}(f_N(x))\,\mathrm{Ai}'(f_N(y))}{d(y)} + \frac{\mathrm{Ai}(f_N(y))\,\mathrm{Ai}'(f_N(x))}{d(x)}\right)\frac{d(x)-d(y)}{x-y}$$
for $x, y \in [1-\delta, 1+\delta]$,

(d) (-1) *times the formula in (c)*
for $x, y \in [-1-\delta, -1+\delta]$.

In the remaining part of the present section we evaluate the formulae for the Christoffel-Darboux kernel given by Theorem 1.3 and by Propostion 4.1 in special cases.

Proof of Theorem 1.7: For $x \in [-1+\delta, 1-\delta]$ and real s, t denote

$$u := x + \frac{s}{N\rho(x)}, \qquad v := x + \frac{t}{N\rho(x)}.$$

Choosing d as defined in Lemma 3.3 resp. as at the end of Remark 3.4 one learns from (1.14) that $\rho(x) \geq \frac{d}{2\pi}\sqrt{1-(1-\delta)^2}$. Thus u, $v \in (-1+\frac{\delta}{2}, 1-\frac{\delta}{2})$ for $|s|$, $|t| < c_\delta N$ with $c_\delta := \frac{\delta d}{4\pi}\sqrt{1-(1-\delta)^2}$. We apply Theorem 1.3 with δ replaced by $\frac{\delta}{2}$ and Proposition 4.1 with (x, y) replaced by (u, v). Since the second summand in Proposition 4.1 (b) is uniformly bounded for u, $v \in (-1+\frac{\delta}{2}, 1-\frac{\delta}{2})$, one may conclude from

$$\frac{\hat{a}(u)}{\hat{a}(v)} + \frac{\hat{a}(v)}{\hat{a}(u)} = 2 + \mathcal{O}\left(|u-v|^2\right)$$

that

$$\frac{b_V - a_V}{2N\rho_V(x)} K_{N,V}\left(\lambda_V(u), \lambda_V(v)\right) = \frac{\sin\left(N\pi\int_v^u \rho(r)\,\mathrm{d}r\right)}{\pi(s-t)} + \mathcal{O}\left(N^{-1}\right).$$

The claim now follows from

$$N\pi\int_v^u \rho(r)\,\mathrm{d}r = \pi(s-t) + \mathcal{O}\left(|s-t|\cdot\frac{|s|+|t|}{N}\right)$$

and the crude estimate $|\sin(\alpha+\beta) - \sin(\alpha)| \leq |\beta|$. □

Proof of Theorem 1.5: We apply Theorem 1.3 and Proposition 4.1 with $\delta = \delta_0$. Statements (i) and (ii) for values of $x \in \hat{J}$ that have distance greater than δ_0 from

± 1 follow from formulae (a) and (b) of Proposition 4.1, from $|k(x)| = \mathcal{O}(e^{-N\eta(x)/2})$ for $|x| \geq 1+\delta_0$, from $|k(x)| = \mathcal{O}(1)$ for $|x| \leq 1-\delta_0$, and from

$$2\frac{a'(x)}{a(x)} = \frac{1}{x^2-1} \quad \text{for } |x|>1, \qquad 2\frac{\hat{a}'(x)}{\hat{a}(x)} = \frac{1}{x^2-1} \quad \text{for } |x|<1.$$

Note that statement (iii) is not relevant for such values of x, i.e. for $x \in \hat{J}$ satisfying $||x|-1| > \delta_0$, as we choose the constant $c \leq \delta_0$. In fact, starting with $c = \delta_0$ the value of c may be decreased a number of times as we proceed in the proof so that the same constant c can be used simultaneously in all statements of Theorem 1.5.

We turn to the case $x \in [1-\delta_0, 1+\delta_0]$. From Theorem 1.3 and Proposition 4.1 (c) it is clear that

$$\begin{aligned}
&\frac{b_V - a_V}{2} K_{N,V}\left(\lambda_V(x), \lambda_V(x)\right) \\
&= \mathbb{A}\mathrm{i}\left(f_N(x), f_N(x)\right) f_N'(x) + 2\,\mathrm{Ai}(f_N(x))\,\mathrm{Ai}'(f_N(x))\frac{d'(x)}{d(x)} + k(x)^T \mathcal{O}\left(N^{-1}\right) k(x) \\
&=: A(x) + B(x) + C(x).
\end{aligned} \tag{4.8}$$

For $x \in (1, 1+\delta_0]$ the following asymptotic results on the Airy function and the Airy kernel are relevant. They are immediate from [1, 10.4.59, 10.4.61] and from $\mathbb{A}\mathrm{i}(\zeta,\zeta) = \mathrm{Ai}'(\zeta)^2 - \zeta\,\mathrm{Ai}(\zeta)^2$. For $\zeta \geq 1$ and $u := \frac{2}{3}\zeta^{3/2}$ one has

$$\mathrm{Ai}(\zeta)^2 = \frac{1}{4\pi\sqrt{\zeta}} e^{-2u}\left(1 + \mathcal{O}\left(u^{-1}\right)\right), \quad \mathrm{Ai}'(\zeta)^2 = \frac{\sqrt{\zeta}}{4\pi} e^{-2u}\left(1 + \mathcal{O}\left(u^{-1}\right)\right), \tag{4.9}$$

$$\mathrm{Ai}(\zeta)\,\mathrm{Ai}'(\zeta) = -\frac{1}{4\pi} e^{-2u}\left(1 + \mathcal{O}\left(u^{-1}\right)\right), \tag{4.10}$$

$$\mathbb{A}\mathrm{i}(\zeta,\zeta) = \frac{1}{8\pi\zeta} e^{-2u}\left(1 + \mathcal{O}\left(u^{-1}\right)\right). \tag{4.11}$$

Using in addition $f_N(x) = \mathcal{O}(N^{2/3}(x-1))$ and $f_N^{-1}(x) = \mathcal{O}(N^{-2/3}(x-1)^{-1})$ (see Lemmas A.1 (iii) and 3.3), we obtain for $N^{2/3}(x-1)$ sufficiently large that the left hand side of (4.8) is of the form

$$f_N\,\mathbb{A}\mathrm{i}(f_N, f_N)\left[\frac{f_N'}{f_N} - 4\frac{d'}{d}\left(1 + \mathcal{O}\left(\frac{1}{N(x-1)^{3/2}}\right)\right) + \mathcal{O}\left(\frac{1}{N(x-1)^{1/2}}\right)\right], \tag{4.12}$$

where the three summands correspond to A, B and C in (4.8). Observe further that $\frac{2}{3}(f_N(x))^{3/2} = \frac{1}{2}N\eta(x)$ and $f_N = N^{2/3}\gamma^+(da)^4$ (see (1.17), (1.19), (1.22), (1.23)), implying

$$\frac{f_N'(x)}{f_N(x)} = 4\frac{d'(x)}{d(x)} + 4\frac{a'(x)}{a(x)} = 4\frac{d'(x)}{d(x)} + \frac{2}{x^2-1}. \tag{4.13}$$

Now, statement (ii) of Theorem 1.5 follows for $x \in (1+c^{-1}N^{-2/3}, 1+\delta_0]$ from (4.12), from $d'/d = \mathcal{O}(1)$, and from (4.11). The just presented arguments also work in the case $x \in [-1-\delta_0, -1-c^{-1}N^{-2/3})$. Except for replacing $x-1$ by $|x|-1$ in the $\mathcal{O}$-terms, only (4.13) needs to be modified by

$$\frac{f_N'}{f_N} = 4\frac{d'}{d} - 4\frac{a'}{a} \quad \text{(see (1.22) and (1.23))}. \tag{4.14}$$

This neutralizes the change of sign stated in part (d) of Proposition 4.1.

Next, we consider $x \in [1-\delta_0, 1-c^{-1}N^{-2/3})$. Setting $u := \frac{2}{3}(-\zeta)^{3/2}$, the following asymptotics for $\zeta \le -1$, derived from [1, 10.4.60, 10.4.62], are useful.

$$\mathrm{Ai}(\zeta)^2 = \mathcal{O}((-\zeta)^{-1/2}), \quad \mathrm{Ai}'(\zeta)^2 = \mathcal{O}((-\zeta)^{1/2}), \tag{4.15}$$

$$\mathrm{Ai}(\zeta)\,\mathrm{Ai}'(\zeta) = -\frac{1}{2\pi}\left(\cos(2u) + \mathcal{O}\left(u^{-1}\right)\right), \tag{4.16}$$

$$\mathbb{A}\mathrm{i}(\zeta,\zeta) = \frac{1}{\pi}(-\zeta)^{1/2}\left(1 - \frac{1}{6u}\cos(2u) + \mathcal{O}\left(u^{-2}\right)\right). \tag{4.17}$$

Together with Corollary A.3, one obtains in the notation of (4.8) and with $u = \frac{2}{3}(-f_N)^{3/2}$:

$$\begin{aligned} A(x) + C(x) &= \mathbb{A}\mathrm{i}(f_N(x), f_N(x)) f_N'(x)\left(1 + \mathcal{O}\left(\frac{1}{N^2(1-x)}\right)\right) \\ &= \frac{f_N'(x)}{f_N(x)}\left(-\frac{3}{2\pi}u + \frac{1}{4\pi}\cos(2u) + \mathcal{O}\left(u^{-1}\right)\right)\left(1 + \mathcal{O}\left(\frac{1}{N^2(1-x)}\right)\right), \\ B(x) &= -\frac{1}{\pi}\cos(2u)\frac{d'(x)}{d(x)} + \mathcal{O}\left(u^{-1}\right). \end{aligned}$$

In addition to (4.13), we observe from (1.19) and (1.15) that

$$u = \frac{2}{3}(-f_N)^{3/2} = \frac{1}{2}N\xi, \quad \text{thus } \frac{f_N'}{f_N} = \frac{2\xi'}{3\xi} \text{ with } \xi' = -2\pi\rho. \tag{4.18}$$

In summary, $A + B + C$ is of the form

$$N\rho(x)\left(1 + \mathcal{O}\left(\frac{1}{N^2(1-x)}\right)\right) + \frac{1}{2\pi(x^2-1)}\cos(N\xi(x)) + \mathcal{O}\left(\frac{1}{N(1-x)^{5/2}}\right).$$

Since $\rho^{-1}(x) = \mathcal{O}((1-x)^{-1/2})$ by Lemma 3.3, statement (i) of Theorem 1.5 follows near 1 by choosing the value of c so small such that, say, $(2\pi(1-x^2))^{-1} < \frac{1}{2}N\rho(x)$ on $[1-\delta_0, 1-c^{-1}N^{-2/3})$.

For $x \in (-1+c^{-1}N^{-2/3}, -1+\delta_0]$, one proceeds in exactly the same way, replacing in Proposition 4.1 statement (c) by (d), (4.13) by (4.14), and (4.18) by

$$u = \frac{2}{3}(-f_N)^{3/2} = \frac{1}{2}N(2\pi - \xi), \quad \text{thus } \frac{f_N'}{f_N} = -\frac{2\xi'}{3(2\pi - \xi)}.$$

In order to prove statement (iii) of Theorem 1.5, let us denote

$$s := N^{2/3}\gamma^+(x-1), \qquad \zeta := f_N(x) = s\hat{f}(x)$$

for $x \in [1-\delta_0, 1+\delta_0]$. Since $\mathbb{A}\mathrm{i}(t,t) = \int_t^\infty \mathrm{Ai}(r)^2\,\mathrm{d}r$, we have

$$\frac{\mathbb{A}\mathrm{i}(\zeta,\zeta)}{\mathbb{A}\mathrm{i}(s,s)} = 1 - \left(\int_0^1 \frac{\mathrm{Ai}(s + r(\zeta - s))^2}{\mathbb{A}\mathrm{i}(s,s)}\,\mathrm{d}r\right)(\zeta - s). \tag{4.19}$$

It follows from Lemma A.1(iii) that

$$\zeta - s = s(\hat{f}(x) - 1) = s\mathcal{O}\left(|x-1|\right) = \mathcal{O}\left(N^{2/3}(x-1)^2\right). \tag{4.20}$$

Using (4.9), (4.11), (4.15), (4.17), and $\mathbb{A}\mathrm{i}(t,t)^{-1} = \mathcal{O}(1)$ for $-1 \le t \le 1$, one obtains

$$\frac{\mathrm{Ai}(s + r(\zeta - s))^2}{\mathbb{A}\mathrm{i}(s,s)} = \begin{cases} \mathcal{O}\left(s^{1/2}\right) & \text{, if } 1 \le s \le N^{4/15} \\ \mathcal{O}\left(1\right) & \text{, if } -1 \le s \le 1 \\ \mathcal{O}\left(|s|^{-1}\right) & \text{, if } s \le -1 \end{cases} \tag{4.21}$$

uniformly for $r \in [0,1]$ and $x \in [1-\delta_0, 1+\delta_0]$. Applying in addition (4.10) and (4.16), we find $B(x) + C(x) = N^{2/3}\gamma^+ \,\mathbb{A}\mathrm{i}(\zeta,\zeta)\left(r_B(x) + \mathcal{O}\left(N^{-1}\right)\right)$ with

$$r_B(x) = \begin{cases} \mathcal{O}\left(N^{-2/3}\right) + \mathcal{O}(|x-1|) & \text{for } x \in [1, 1+\delta_0], \\ \mathcal{O}\left(N^{-2/3}\right) & \text{for } x \in [1-\delta_0, 1]. \end{cases} \tag{4.22}$$

Finally, claim (iii) of Theorem 1.5 follows from (4.19), (4.20), (4.21), (4.22) and

$$A + B + C = N^{2/3}\gamma^+ \,\mathbb{A}\mathrm{i}(s,s) \cdot \frac{\mathbb{A}\mathrm{i}(\zeta,\zeta)}{\mathbb{A}\mathrm{i}(s,s)} \left(\hat{f} + (x-1)\hat{f}' + r_B + \mathcal{O}\left(N^{-1}\right)\right).$$

□

Before turning to the proof of Theorem 1.8 we state an asymptotic result for a special type of integrals, which can be verified using integration by parts.

Proposition 4.2. *Let $\alpha, \beta \in \mathbb{R}$. Then for $s, t \geq 1$:*

$$\int_0^\infty (s+r)^\alpha (t+r)^\beta e^{-\frac{2}{3}[(s+r)^{3/2} + (t+r)^{3/2}]} \,\mathrm{d}r$$
$$= \frac{s^\alpha\, t^\beta}{s^{1/2} + t^{1/2}}\, e^{-\frac{2}{3}(s^{3/2} + t^{3/2})} \left(1 + \mathcal{O}\left(s^{-\frac{3}{2}}\right) + \mathcal{O}\left(t^{-\frac{3}{2}}\right)\right).$$

Proof of Theorem 1.8: For $q \leq s, t \leq pN^{4/15}$ denote:

$$x := 1 + \frac{s}{N^{2/3}\gamma^+}, \qquad \hat{s} := f_N(x) = s\hat{f}(x),$$
$$y := 1 + \frac{t}{N^{2/3}\gamma^+}, \qquad \hat{t} := f_N(y) = t\hat{f}(y).$$

It follows from Theorem 1.3 and Proposition 4.1 (c) that the Christoffel-Darboux kernel can be written in the form

$$\frac{b_V - a_V}{2N^{\frac{2}{3}}\gamma^+} K_{N,V}\left(\lambda_V(x), \lambda_V(y)\right) = \hat{A} + \hat{B} + \hat{C}$$

with

$$\hat{A} := \mathbb{A}\mathrm{i}(\hat{s},\hat{t}) \left(\hat{f}(x) + (y-1)\frac{\hat{f}(x) - \hat{f}(y)}{x-y}\right)$$
$$\hat{B} := \frac{1}{N^{2/3}\gamma^+} \left(\frac{\mathrm{Ai}(\hat{s})\,\mathrm{Ai}'(\hat{t})}{d(y)} + \frac{\mathrm{Ai}'(\hat{s})\,\mathrm{Ai}(\hat{t})}{d(x)}\right) \frac{d(x) - d(y)}{x-y}$$
$$\hat{C} := k(y)^T \mathcal{O}\left(N^{-5/3}\right) k(x).$$

In the case $q \leq s, t \leq 2$ one may use the boundedness of the derivatives of $\mathbb{A}\mathrm{i}$, $\hat{f}$, and d together with the bounds $d^{-1}(x)$, $d^{-1}(y) = \mathcal{O}(1)$, and $k(x)$, $k(y) = \mathcal{O}(N^{1/6})$ to derive

$$\hat{A} = \mathbb{A}\mathrm{i}(s,t)\left(1 + \mathcal{O}(|x-1|) + \mathcal{O}(|y-1|)\right) = \mathbb{A}\mathrm{i}(s,t) + \mathcal{O}\left(N^{-2/3}\right),$$
$$\hat{B} = \mathcal{O}\left(N^{-2/3}\right), \qquad \text{and} \qquad \hat{C} = \mathcal{O}\left(N^{-4/3}\right).$$

This yields the desired result.

Next, we assume $1 \leq s, t \leq pN^{4/15}$. Recall from (1.7) that

$$\mathbb{A}\mathrm{i}(u,v) = \int_0^\infty \mathrm{Ai}(u+r)\,\mathrm{Ai}(v+r)\,\mathrm{d}r.$$

The asymptotics [1, 10.4.59, 10.4.61] of Ai resp. Ai′ and Proposition 4.2 lead to

$$\mathbb{A}\mathrm{i}(\hat{s},\hat{t}) = (\hat{s}\hat{t})^{-\frac{1}{4}} (\hat{s}^{\frac{1}{2}} + \hat{t}^{\frac{1}{2}})^{-1} e^{-\frac{2}{3}(\hat{s}^{3/2} + \hat{t}^{3/2})} \left(1 + \mathcal{O}\left(\hat{s}^{-\frac{3}{2}}\right) + \mathcal{O}\left(\hat{t}^{-\frac{3}{2}}\right)\right). \tag{4.23}$$

Using the asymptotics of the Airy function again, we obtain

$$\hat{B} = \mathbb{A}\mathrm{i}(\hat{s},\hat{t}) \cdot \mathcal{O}\left(N^{-2/3}\left(\hat{s}^{\,1/2}+\hat{t}^{\,1/2}\right)^2\right) = \mathbb{A}\mathrm{i}(\hat{s},\hat{t}) \cdot \mathcal{O}\left(N^{-1}(s+t)\right)$$

$$\begin{aligned}\hat{C} &= \mathbb{A}\mathrm{i}(\hat{s},\hat{t}) \cdot \mathcal{O}\left(N^{-2}\left(s^{3/2}+t^{3/2}\right)+N^{-5/3}(s+t)+N^{-4/3}\left(s^{1/2}+t^{1/2}\right)\right)\\ &= \mathbb{A}\mathrm{i}(\hat{s},\hat{t}) \cdot \mathcal{O}\left(N^{-1}\right)\end{aligned}$$

and consequently

$$\hat{A}+\hat{B}+\hat{C} = \mathbb{A}\mathrm{i}(\hat{s},\hat{t})\left(1+\mathcal{O}\left(\frac{s+t}{N^{2/3}}\right)\right). \tag{4.24}$$

Differentiating the integral representation in (1.7) and applying Proposition 4.2 yields the bound on the derivative

$$D\,\mathbb{A}\mathrm{i}(\tilde{s},\tilde{t}) = \mathbb{A}\mathrm{i}(s,t) \cdot \mathcal{O}\left(s^{1/2}+t^{1/2}\right),$$

uniformly for $\tilde{s},\tilde{t}$ between s and $\hat{s}$ resp. between t and $\hat{t}$. Since $s-\hat{s}=\mathcal{O}(s^2N^{-2/3})$ and $t-\hat{t}=\mathcal{O}(t^2N^{-2/3})$, this implies

$$\mathbb{A}\mathrm{i}(\hat{s},\hat{t}) = \mathbb{A}\mathrm{i}(s,t)\left(1+\mathcal{O}\left(\frac{s^{5/2}+t^{5/2}}{N^{2/3}}\right)\right).$$

In view of (4.24) the claim follows.

Finally, we turn to the case $q \le s \le 1$ and $2 \le t \le pN^{4/15}$. Elementary calculations together with [1, 10.4.59, 10.4.61] give

$$\begin{aligned}&\mathrm{Ai}(\hat{t}) = \mathrm{Ai}(t)\left(1+\mathcal{O}\left(\frac{t^{5/2}}{N^{2/3}}\right)\right), && \mathrm{Ai}'(\hat{t}) = \mathrm{Ai}'(t)\left(1+\mathcal{O}\left(\frac{t^{5/2}}{N^{2/3}}\right)\right),\\ &\mathrm{Ai}(\hat{s}) = \mathrm{Ai}(s)+\mathcal{O}\left(N^{-2/3}\right), && \mathrm{Ai}'(\hat{s}) = \mathrm{Ai}'(s)+\mathcal{O}\left(N^{-2/3}\right),\\ &\hat{s}-\hat{t} = (s-t)\left(1+\mathcal{O}\left(\frac{t}{N^{2/3}}\right)\right), && \frac{1}{s-t} = \mathcal{O}\left(t^{-1}\right).\end{aligned}$$

These relations suffice to derive

$$\hat{A} = \left(\mathbb{A}\mathrm{i}(s,t)+\mathrm{Ai}'(t)\cdot\mathcal{O}\left(\frac{t^{3/2}}{N^{2/3}}\right)\right)\left(1+\mathcal{O}\left(\frac{t}{N^{2/3}}\right)\right) = \mathbb{A}\mathrm{i}(s,t)+\mathrm{Ai}'(t)\cdot\mathcal{O}\left(\frac{t^{3/2}}{N^{2/3}}\right),$$

$$\hat{B} = \mathrm{Ai}'(t)\mathcal{O}\left(N^{-2/3}\right), \qquad \text{and} \qquad \hat{C} = \mathrm{Ai}'(t)\mathcal{O}\left(N^{-4/3}\right),$$

which completes the proof. □

A. Appendix: Construction of the parametrix near ±1

Since the definition of the parametrix $\mathbb{T}$ near ±1 is somewhat involved, we first motivate the construction. The validation of the non-obvious claims made along the way begins with Lemma A.1. Our presentation is similar to [13, Section 7.6] and [31, Subsection 6.4.6].

Assume that V satisfies $\mathbf{(GA)}_1$ and let q, $\tilde{q}$, $G \equiv G_V$, $\xi \equiv \xi_V$, $\eta \equiv \eta_V$, $\rho \equiv \rho_V$ be given as in (2.5), (3.13), and (1.13) to (1.16). We begin the construction by introducing another auxiliary function. Let $\kappa > 0$ be chosen such that G has an analytic extension to $B_\kappa(-1) \cup B_\kappa(1)$. Set

$$\varphi(z) := \begin{cases} \frac{1}{2}\int_1^z q(\zeta)\,G(\zeta)\,\mathrm{d}\zeta & \text{, if } z \in B_\kappa(1)\setminus(1-\kappa,1],\\ \frac{1}{2}\int_{-1}^z q(\zeta)\,G(\zeta)\,\mathrm{d}\zeta & \text{, if } z \in B_\kappa(-1)\setminus[-1,-1+\kappa).\end{cases} \tag{A.1}$$

The map φ is analytic. According to Definition 3.5, we have

$$\varphi(z) = \frac{1}{2}\eta(z) \quad \text{and} \quad \varphi(z) = \begin{cases} \mp\frac{i}{2}\xi(z) & \text{near } 1,\ \mathrm{Im}(z) \gtrless 0, \\ \mp\frac{i}{2}(\xi(z) - 2\pi) & \text{near } -1,\ \mathrm{Im}(z) \gtrless 0, \end{cases} \tag{A.2}$$

on the corresponding common domains of definition. In order to see this use $q(z) = \pm i\tilde{q}(z)$ for $\mathrm{Im}(z) \gtrless 0$, $\int_{-1}^{z} \tilde{q}G = \int_{-1}^{1} 2\pi\rho + \int_{1}^{z} \tilde{q}G$, and Lemma 2.1 (a).

The reason for defining φ is the relation

$$e^{N\varphi_-\sigma_3} v_T e^{-N\varphi_+\sigma_3} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \text{on } (-1-\kappa, -1+\kappa) \cup (1-\kappa, 1+\kappa), \tag{A.3}$$

which is immediate from (A.2) and (3.7). Recall from Lemma 3.8 that it is one of the desired properties of the parametrix $\mathbb{T}$ that $R = T\,\mathbb{T}^{-1}$ has no jumps in regions $\mathrm{III}^{\pm}$. By (3.23) this implies $\mathbb{T}_+ = \mathbb{T}_- v_T$ on $\mathrm{III}^{\pm} \cap \mathbb{R}$.

Writing $\mathbb{T}$ in the form $\mathbb{T} = \hat{\mathbb{T}} e^{N\varphi\sigma_3}$, we obtain for $\hat{\mathbb{T}}$ the constant jump

$$\hat{\mathbb{T}}_+ = \hat{\mathbb{T}}_- \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \text{on } (-1-\kappa, -1+\kappa) \cup (1-\kappa, 1+\kappa). \tag{A.4}$$

The next ingredient in the construction is the observation that for $\omega := e^{2\pi i/3}$ the matrix-valued function

$$\psi_0(\zeta) := \begin{cases} \begin{pmatrix} \mathrm{Ai}(\zeta) & \mathrm{Ai}(\omega^2\zeta) \\ \mathrm{Ai}'(\zeta) & \omega^2\mathrm{Ai}'(\omega^2\zeta) \end{pmatrix} e^{-\frac{\pi i}{6}\sigma_3} & , \text{ if } \zeta \in \{z \in \mathbb{C} \mid \mathrm{Im}\, z > 0\}, \\ \begin{pmatrix} \mathrm{Ai}(\zeta) & -\omega^2\mathrm{Ai}(\omega\zeta) \\ \mathrm{Ai}'(\zeta) & -\mathrm{Ai}'(\omega\zeta) \end{pmatrix} e^{-\frac{\pi i}{6}\sigma_3} & , \text{ if } \zeta \in \{z \in \mathbb{C} \mid \mathrm{Im}\, z < 0\}, \end{cases} \tag{A.5}$$

satisfies the same jump condition as $\hat{\mathbb{T}}$,

$$(\psi_0)_+ = (\psi_0)_- \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \text{on } \mathbb{R}, \tag{A.6}$$

due to the relation $\mathrm{Ai}(\zeta) + \omega\mathrm{Ai}(\omega\zeta) + \omega^2\mathrm{Ai}(\omega^2\zeta) = 0$ for all $\zeta \in \mathbb{R}$ (see [1, 10.4.7]).

Note that there are plenty of maps $A : \mathbb{C}\backslash\mathbb{R} \to \mathbb{C}^{2\times 2}$ with jump relation (A.6). What leads to the choice of ψ_0 is the additional condition $v_R = I + \mathcal{O}(N^{-1})$ on $\Sigma_3^{\pm}$ from Lemma 3.8 that is essential for proving Theorem 3.9. In order to obtain the required jump relation for v_R across $\Sigma_3^{\pm}$ we need more freedom in the construction. Observe first that for any entire function f mapping the upper resp. lower half-plane into itself the composition $\psi_0 \circ f$ also satisfies (A.6), because the jump matrix is constant. The crucial property of ψ_0 is its asymptotic behavior (cf. (A.21) below)

$$\psi_0(\zeta)\, e^{\frac{2}{3}\zeta^{3/2}\sigma_3} \sim \frac{e^{\frac{\pi i}{12}}}{2\sqrt{\pi}} \begin{pmatrix} \zeta^{-\frac{1}{4}} & 0 \\ 0 & \zeta^{\frac{1}{4}} \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} e^{-\frac{\pi i}{4}\sigma_3} \tag{A.7}$$

as $\zeta \to \infty$. We may use this as follows to define the parametrix $\mathbb{T}$ on, say $B_\kappa(1)$. Suppose there exists an analytic function $f_N : B_\kappa(1) \to \mathbb{C}$ satisfying $\frac{2}{3} f_N(z)^{3/2} = N\varphi(z)$. Define $\mathbb{T}_0 := (\psi_0 \circ f_N)\, e^{N\varphi\sigma_3}$. By the above discussion we have

$$(\mathbb{T}_0)_+ = (\mathbb{T}_0)_-\, v_T \quad \text{on } \mathrm{III}^+ \cap \mathbb{R}. \tag{A.8}$$

By (3.23) the jump matrix v_R will be equal to $\mathbb{T}_0 M^{-1}$ (see (3.10)) on the arc $A := \Sigma_3^+ \cap \{z \in \mathbb{C} \mid |\arg(z-1)| < \frac{3\pi}{4}\}$ if we choose $\mathbb{T} = \mathbb{T}_0$ in III^+, because $\mathbb{T}_+ = M$ on A. As $N \to \infty$, $f_N(z)$ also tends to infinity and (A.7) implies

$$\mathbb{T}_0(s)\, M^{-1}(s) \sim \frac{1}{2\sqrt{\pi}} e^{-\frac{\pi i}{6}} \begin{pmatrix} 0 & a(s) f_N^{-1/4}(s) \\ -f_N^{1/4}(s) a^{-1}(s) & 0 \end{pmatrix} \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix}$$

for $s \in A$. These are not the desired asymptotics for $v_R \sim I$ on A. However, defining

$$E_N := \sqrt{\pi} e^{\frac{\pi i}{6}} \begin{pmatrix} 1 & -1 \\ -i & -i \end{pmatrix} \begin{pmatrix} f_N^{1/4} a^{-1} & 0 \\ 0 & f_N^{-1/4} a \end{pmatrix} \quad \text{and} \quad \mathbb{T} := E_N \mathbb{T}_0 \quad \text{in } \mathrm{III}_3^+, \tag{A.9}$$

one obtains $v_R = \mathbb{T}_- \mathbb{T}_+^{-1} \sim I$ on A. Moreover, it turns out that E_N has an analytic extension in III_3^+ so that jump relation (A.8) still holds for $\mathbb{T}$. Finally, we need to consider the jumps of v_R on $\Sigma_3^+ \backslash A$. Here, $\mathbb{T}_+ = M v_u$ resp. $\mathbb{T}_+ = M v_l^{-1}$. Recall that the definition of $\mathbb{T}$ in (3.16), (3.17) is based on the factorization of v_T as given in (3.8). Applying the same procedure inside the disk III_3^+ with the factorization

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \tag{A.10}$$

for the jump matrix of $\hat{\mathbb{T}}$ (see (A.4)), one is led to a modification in the definition of ψ_0. It consists of replacing ψ_0 by

$$\psi_\beta(z) = \begin{cases} \psi_0(z) & \text{for } z \in \Omega_1 \cup \Omega_4, \\ \psi_0(z) \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}^{-1} & \text{for } z \in \Omega_2, \\ \psi_0(z) \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} & \text{for } z \in \Omega_3, \end{cases} \tag{A.11}$$

where the unbounded sectors Ω_j depend on the angle β as displayed in Figure 3 below. The value for β will be specified in (A.15). The following Lemma allows to define the function f_N satisfying $\frac{2}{3} f_N^{3/2} = N\varphi$ and to verify all the properties that were used in the construction described above.

Lemma A.1. *(a) Let V satisfy **(GA)**$_1$ and let φ, $\gamma^\pm \equiv \gamma_V^\pm$ be defined as in (A.1), (1.21). Then there exist $\hat{\sigma} > 0$ and an analytic function $\hat{f} : B_{\hat{\sigma}}(-1) \cup B_{\hat{\sigma}}(1) \to \mathbb{C}$ such that (i) to (iii) hold.*

(i) $\left[\gamma^+ (z-1) \hat{f}(z) \right]^{3/2} = \frac{3}{2} \varphi(z)$ *for* $z \in B_{\hat{\sigma}}(1) \setminus (1-\hat{\sigma}, 1]$,

$\left[-\gamma^- (z+1) \hat{f}(z) \right]^{3/2} = \frac{3}{2} \varphi(z)$ *for* $z \in B_{\hat{\sigma}}(-1) \setminus [-1, -1+\hat{\sigma})$,

(ii) $\hat{f}(s) \in \mathbb{R}$ *for all* $s \in \mathbb{R} \cap (B_{\hat{\sigma}}(1) \cup B_{\hat{\sigma}}(-1))$,

(iii) $\hat{f}(\pm 1) = 1$ *and* $\hat{f}(B_{\hat{\sigma}}(\pm 1)) \subset B_{1/5}(1)$.

*(b) Assume that Q satisfies **(GA)**$_2$. Then there exists an open neighborhood $\mathcal{U}$ of Q in $(X_D, \|\cdot\|_\infty)$ (see Remark 1.1) such that statement (a) holds for all $V \in \mathcal{U}$ with $\hat{\sigma}$ independent of V.*

Proof. Let σ be defined as in (the end of) Remark 3.4 resp. as in Lemma 3.3 in the situation of (a) resp. (b). This implies in particular that $G \equiv G_V$ possesses an analytic extension on $B_\sigma(\pm 1)$. Note that in the case of (b) Lemma 3.3 also provides a neighborhood $\mathcal{U}$ of Q such that the same value of σ can be used for all $V \in \mathcal{U}$ simultaneously. We introduce the auxiliary function

$$h(z) := \frac{1}{2} G(z) (1 \pm z)^{1/2},$$

which is analytic on $B_\sigma(\pm 1)$. By definition (A.1) we have

$$\varphi(z) = \pm \int_{\pm 1}^{z} (\pm s - 1)^{1/2} h(s) \, \mathrm{d}s \quad \text{for } z \in B_\sigma(\pm 1) \setminus [-1, 1].$$

We write $h(z) = h(\pm 1) + \tilde{h}(z)(z \mp 1)$ with $\tilde{h}$ analytic on $B_\sigma(\pm 1)$. Reducing σ and $\mathcal{U}$, if necessary, we may ensure that G and G' are both bounded on $B_\sigma(\pm 1)$ (uniform in V in the situation of (b), cf. arguments leading to (3.12)). Thus, such a uniform bound also exists for $\tilde{h} \equiv \tilde{h}_V$, i.e.

$$\exists C > 0\, \forall z \in B_\sigma(\pm 1)\ (\forall V \in \mathcal{U}) : |\tilde{h}_V(z)| \le C. \tag{A.12}$$

A short calculation gives

$$\frac{3}{2}\varphi(z) = h(\pm 1)(\pm z - 1)^{3/2}(1 + r(z)) \qquad \text{with}$$
$$r(z) := \frac{3}{2h(\pm 1)(\pm z - 1)^{3/2}}\left(\int_{\pm 1}^{z} (\pm s - 1)^{3/2}\tilde{h}(s)\,\mathrm{d}s\right)$$

for $z \in B_\sigma(\pm 1) \setminus [-1, 1]$. Observe that r has an analytic extension to all of $B_\sigma(\pm 1)$. Moreover, the upper bound (A.12) on $\tilde{h}$ together with the lower bound on $h(\pm 1)$ provided by Lemma 3.3 resp. by Remark 3.4 imply

$$\exists C_0 > 0\, \forall z \in B_\sigma(\pm 1)\ (\forall V \in \mathcal{U}) : |r_V(z)| \le C_0|z \mp 1|. \tag{A.13}$$

Choosing $\hat{\sigma} := \min(\sigma, (5C_0)^{-1})$ it follows from (A.13) that

$$\hat{f} : B_{\hat{\sigma}}(1) \cup B_{\hat{\sigma}}(-1) \to \mathbb{C}, \quad \hat{f}(z) := (1 + r(z))^{2/3}$$

defines an analytic function that satisfies claim (iii). To see property (ii), one first verifies from the construction that $\hat{f}$ is real on $(-1 - \hat{\sigma}, -1] \cup [1, 1 + \hat{\sigma})$. The claim is then a consequence of the identity principle. The relations in (i) are immediate from the construction and from $\gamma^\pm = (h(\pm 1))^{2/3}$ (see (1.21)). □

Remark A.2. In view of Definition 3.5 and relation (A.2) it follows from statement (i) of Lemma A.1 that the restriction of $\hat{f}$ to the reals has all the properties that are claimed in Definition 1.2 with, say, $\delta_V = \hat{\sigma}/2$.

Corollary A.3. *Let the assumptions of Lemma A.1 (a) resp. (b) be satisfied. Define $f : B_{\hat{\sigma}}(-1) \cup B_{\hat{\sigma}}(1) \to \mathbb{C}$ by $f(z) := (z \mp 1)\hat{f}(z)$ for $z \in B_{\hat{\sigma}}(\pm 1)$, with $\hat{\sigma}$, $\hat{f}$ as provided by Lemma A.1. Then $f'\left(B_{\hat{\sigma}/10}(\pm 1)\right) \subset B_{1/3}(1)$.*

Proof. From Lemma A.1 (iii) it follows for all $z \in B_{\hat{\sigma}/10}(\pm 1)$ that $|\hat{f}(z) - 1| < \frac{1}{5}$ and that $|\hat{f}'(z)| \le \frac{6}{5}(\frac{9}{10}\hat{\sigma})^{-1}$, thus $|f'(z) - 1| < \frac{1}{5} + \frac{6}{45} = \frac{1}{3}$. □

We are now ready to define the parametrix $\mathbb{T}$ on $\mathrm{III}^\pm$. We begin with the disk $\mathrm{III}^+ = B_\delta(1)$ for $0 < \delta \le \sigma_0 \le \frac{\hat{\sigma}}{10}$ (see (3.14) for a definition of σ_0 and Lemma A.1 for $\hat{\sigma}$). Let $\hat{f}$, f be given by Lemma A.1 resp. by Corollary A.3 and set in accordance with Definition 1.2

$$f_N(z) := N^{2/3}\gamma^+ f(z) = N^{2/3}\gamma^+(z - 1)\hat{f}(z), \quad z \in B_\delta(1). \tag{A.14}$$

Observe that E_N of (A.9) is now defined on the domain $B_\delta(1) \setminus (1 - \delta, 1]$. Moreover, E_N can be extended analytically to $B_\delta(1)$ since the factor $(z - 1)^{1/4}$ cancels and $\hat{f}^{\pm 1/4}$ is analytic by statement (iii) of Lemma A.1. Note in addition, that Corollary A.3 implies that $f : B_\delta(1) \to f(B_\delta(1))$ is biholomorphic and statement (iii) of Lemma A.1 yields

$$\beta_\delta^+ := \arg\left(f\left(1 + \delta e^{\frac{3\pi i}{4}}\right)\right) \in \left(\frac{5\pi}{8}, \frac{7\pi}{8}\right). \tag{A.15}$$

The significance of the point $1 + \delta e^{3\pi i/4}$ is that Σ_2^u and Σ_3^+ of Figure 2 intersect here. By Lemma A.1 (ii), we have $\arg(f(1 + \delta e^{-3\pi i/4})) = -\beta_\delta^+$. Splitting up $f(B_\delta(1))$ as displayed in Figure 3 with $\beta = \beta_\delta^+$, the regions III_j^+ are defined by the relations

$f(\mathrm{III}_j^+) = \Omega_j \cap f(B_\delta(1))$ for $j = 1,2,3,4$. Properties (ii) and (iii) of Lemma A.1 imply that $\mathrm{III}_1^+, \mathrm{III}_2^+$ are located in the upper half-plane, whereas $\mathrm{III}_3^+, \mathrm{III}_4^+$ lie below $\mathbb{R}$. Moreover, the joint boundary of III_1^+ and III_2^+ resp. III_3^+ and III_4^+ are curves that connect 1 with $1+\delta e^{3\pi i/4}$ (resp. $1+\delta e^{-3\pi i/4}$). They can also be found in Figure 2 as dotted lines.

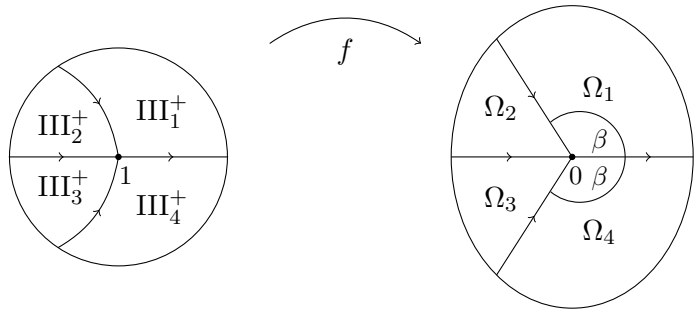

FIGURE 3. Definition of subregions III_j^+ via f as given in Corollary A.3

As motivated above (see in particular the discussion between (A.7) and (A.11)) the parametrix is given by

(A.16)
$$\mathbb{T}(z) := \begin{cases} E_N(z)\,\psi_{\beta_\delta^+}(f_N(z))\,e^{N\varphi(z)\sigma_3} & , \text{ if } z \in \mathrm{III}_1^+ \cup \mathrm{III}_4^+, \\ E_N(z)\,\psi_{\beta_\delta^+}(f_N(z))\,e^{N\varphi(z)\sigma_3}\begin{pmatrix} 1 & 0 \\ e^{-iN\xi(z)} & 1 \end{pmatrix} & , \text{ if } z \in \mathrm{III}_2^+, \\ E_N(z)\,\psi_{\beta_\delta^+}(f_N(z))\,e^{N\varphi(z)\sigma_3}\begin{pmatrix} 1 & 0 \\ e^{iN\xi(z)} & 1 \end{pmatrix}^{-1} & , \text{ if } z \in \mathrm{III}_3^+. \end{cases}$$

The definition of $\mathbb{T}$ on III_3^- can be summarized as follows (cf. Definition 1.2):

$$f_N(z) := -N^{2/3}\gamma^- f(z) = -N^{2/3}\gamma^-(z+1)\,\hat{f}(z)\,,\ z \in B_\delta(-1)\,, \tag{A.17}$$

$$E_N := \sqrt{\pi}e^{\frac{\pi i}{6}}\begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix}\begin{pmatrix} f_N^{1/4}a & 0 \\ 0 & f_N^{-1/4}a^{-1} \end{pmatrix}, \text{ extended to } B_\delta(-1)\,, \tag{A.18}$$

$$\beta_\delta^- := \arg\left(-f\left(-1+\delta e^{-\frac{\pi i}{4}}\right)\right) \in \left(\frac{5\pi}{8}, \frac{7\pi}{8}\right). \tag{A.19}$$

Splitting $-f(B_\delta(-1))$ into wedges Ω_j with $\beta = \beta_\delta^-$, we define III_j^- via $-f(\mathrm{III}_j^-) = \Omega_j \cap [-f(B_\delta(-1))]$. Observe that the regions above $\mathbb{R}$ are now III_3^- and III_4^-. Then

(A.20)
$$\mathbb{T}(z) := \begin{cases} E_N(z)\,\sigma_3\psi_{\beta_\delta^-}(f_N(z))\,\sigma_3 e^{N\varphi(z)\sigma_3} & , \text{ if } z \in \mathrm{III}_1^- \cup \mathrm{III}_4^-, \\ E_N(z)\,\sigma_3\psi_{\beta_\delta^-}(f_N(z))\,\sigma_3 e^{N\varphi(z)\sigma_3}\begin{pmatrix} 1 & 0 \\ e^{iN\xi(z)} & 1 \end{pmatrix}^{-1} & , \text{ if } z \in \mathrm{III}_2^-, \\ E_N(z)\,\sigma_3\psi_{\beta_\delta^-}(f_N(z))\,\sigma_3 e^{N\varphi(z)\sigma_3}\begin{pmatrix} 1 & 0 \\ e^{-iN\xi(z)} & 1 \end{pmatrix} & , \text{ if } z \in \mathrm{III}_3^-. \end{cases}$$

Finally, we state the main result of the Appendix. It is crucial for proving Lemma 3.8.

Lemma A.4. *Let the assumptions of Lemma 3.8 (a) resp. (b) be satisfied and let R be given by* (3.21)*, see also* (3.14) *to* (3.20)*,* (A.16)*, and* (A.20)*. Then R has an*

analytic continuation on $III_3^+ \cup III_3^-$. *For* $\Delta_R \equiv v_R - I$ *we have*

$$\|\Delta_R\|_{L^\infty\left(\Sigma_3^+\cup\Sigma_3^-\right)} = \mathcal{O}\left(N^{-1}\right).$$

The error bound is uniform for (δ, ε) *from an arbitrary but fixed compact subset of* $(0, \sigma_0]^2$. *In the situation of Lemma 3.8 (b) the error bound is also uniform in* $V \in \mathcal{U}$ *for some open neighborhood* $\mathcal{U}$ *of* Q.

Proof. Using (3.23), (A.1), (A.2), and (A.6) it follows from the definition of $\mathbb{T}$ that all jump matrices v_R equal I within $\mathrm{III}_3^\pm$. Hence, R has an analytic continuation in these discs. Moreover, one easily derives

$$v_R = E_N[\psi_{\beta_\delta^+} \circ f_N]e^{N\varphi\sigma_3}M^{-1} \quad \text{on } \Sigma_3^+.$$

Inserting formulae [1, 10.4.7, 10.4.59, 10.4.61] into definitions (A.5), (A.11) gives

(A.21)

$$\psi_\beta(\zeta)\, e^{\frac{2}{3}\zeta^{3/2}\sigma_3} = \frac{e^{\frac{\pi i}{12}}}{2\sqrt{\pi}} \begin{pmatrix} \zeta^{-\frac{1}{4}} & 0 \\ 0 & \zeta^{\frac{1}{4}} \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} e^{-\frac{\pi i}{4}\sigma_3} \left[I + \mathcal{O}\left(|\zeta|^{-1}\right)\right] \text{ for } |\zeta| \to \infty$$

with a uniform bound $\mathcal{O}(|\zeta|^{-1})$ for β in compact subsets of $(\frac{\pi}{3}, \pi)$. Setting $\zeta = f_N(z)$ and recalling the choice $\beta_\delta^+ \in [\frac{5\pi}{8}, \frac{7\pi}{8}]$ for β, we obtain from the boundedness of M, M^{-1} on Σ_3^+ (uniform in δ for δ in compact subsets of $(0, \sigma_0]$) that

$$\|\Delta_R\|_{L^\infty\left(\Sigma_3^+\right)} = \mathcal{O}\left(\frac{1}{N}\left\|\left(\gamma^+ f\right)^{-3/2}\right\|_{L^\infty\left(\Sigma_3^+\right)}\right) = \mathcal{O}\left(\frac{1}{N}\left(\frac{4}{5}\gamma^+\delta\right)^{-3/2}\right),$$

where we have also used Lemma A.1 (iii). In the situation of Lemma 3.8 (b) we appeal to Lemma 3.3 to justify that γ^+ has a lower bound uniform in $V \in \mathcal{U}$. In summary, the bound $\|\Delta_R\|_{L^\infty(\Sigma_3^+)} = \mathcal{O}(N^{-1})$ has been established with the desired uniformity properties. A similar argument proves $\|\Delta_R\|_{L^\infty(\Sigma_3^-)} = \mathcal{O}(N^{-1})$.

□

Acknowledgements

The first author is grateful to Leonid Pastur for good advice and for sharing his deep insights into the field of Mathematical Physics on a number of occasions. The fourth author would like to thank Leonid Pastur for several helpful suggestions while he was preparing his PhD thesis.

References

[1] M. Abramowitz and I. A. Stegun. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, volume 55 of *National Bureau of Standards Applied Mathematics Series*. For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C., 1964.

[2] G. Akemann, J. Baik, and P. Di Francesco, editors. *The Oxford handbook of random matrix theory*. Oxford University Press, Oxford, 2011.

[3] G. W. Anderson. Spectral statistics of unitary ensembles. In *The Oxford handbook of random matrix theory*, pages 66–85. Oxford Univ. Press, Oxford, 2011.

[4] G. W. Anderson, A. Guionnet, and O. Zeitouni. *An introduction to random matrices*, volume 118 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010.

[5] G. Borot, B. Eynard, and N. Orantin. Abstract loop equations, topological recursion, and applications. 2013. arXiv:1303.5808.

[6] A. Boutet de Monvel, L. Pastur, and M. Shcherbina. On the statistical mechanics approach in the random matrix theory: integrated density of states. *J. Statist. Phys.*, 79(3-4):585–611, 1995.

[7] T. Claeys. Birth of a cut in unitary random matrix ensembles. *Int. Math. Res. Not. IMRN*, (6):Art. ID rnm166, 40, 2008.

[8] T. Claeys and A. B. J. Kuijlaars. Universality of the double scaling limit in random matrix models. *Comm. Pure Appl. Math.*, 59(11):1573–1603, 2006.

[9] T. Claeys and A. B. J. Kuijlaars. Universality in unitary random matrix ensembles when the soft edge meets the hard edge. In *Integrable systems and random matrices*, volume 458 of *Contemp. Math.*, pages 265–279. Amer. Math. Soc., Providence, RI, 2008.

[10] T. Claeys, A. B. J. Kuijlaars, and M. Vanlessen. Multi-critical unitary random matrix ensembles and the general Painlevé II equation. *Ann. of Math. (2)*, 168(2):601–641, 2008.

[11] T. Claeys and M. Vanlessen. Universality of a double scaling limit near singular edge points in random matrix models. *Comm. Math. Phys.*, 273(2):499–532, 2007.

[12] D. Dai and A. B. J. Kuijlaars. Painlevé IV asymptotics for orthogonal polynomials with respect to a modified Laguerre weight. *Stud. Appl. Math.*, 122(1):29–83, 2009.

[13] P. Deift. *Orthogonal polynomials and random matrices: a Riemann-Hilbert approach*, volume 3 of *Courant Lecture Notes in Mathematics*. New York University Courant Institute of Mathematical Sciences, New York, 1999.

[14] P. Deift and D. Gioev. Universality at the edge of the spectrum for unitary, orthogonal, and symplectic ensembles of random matrices. *Comm. Pure Appl. Math.*, 60(6):867–910, 2007.

[15] P. Deift and D. Gioev. *Random matrix theory: invariant ensembles and universality*, volume 18 of *Courant Lecture Notes in Mathematics*. Courant Institute of Mathematical Sciences, New York, 2009.

[16] P. Deift, T. Kriecherbauer, and K. T.-R. McLaughlin. New results on the equilibrium measure for logarithmic potentials in the presence of an external field. *J. Approx. Theory*, 95(3):388–475, 1998.

[17] P. Deift, T. Kriecherbauer, K. T-R McLaughlin, S. Venakides, and X. Zhou. Strong asymptotics of orthogonal polynomials with respect to exponential weights. *Comm. Pure Appl. Math.*, 52(12):1491–1552, 1999.

[18] P. Deift, T. Kriecherbauer, K. T.-R. McLaughlin, S. Venakides, and X. Zhou. Uniform asymptotics for polynomials orthogonal with respect to varying exponential weights and applications to universality questions in random matrix theory. *Comm. Pure Appl. Math.*, 52(11):1335–1425, 1999.

[19] P. Deift, S. Venakides, and X. Zhou. New results in small dispersion KdV by an extension of the steepest descent method for Riemann-Hilbert problems. *Internat. Math. Res. Notices*, (6):286–299, 1997.

[20] P. Deift and X. Zhou. A steepest descent method for oscillatory Riemann-Hilbert problems. Asymptotics for the MKdV equation. *Ann. of Math. (2)*, 137(2):295–368, 1993.

[21] A. S. Fokas, A. R. Its, and A. V. Kitaev. The isomonodromy approach to matrix models in 2D quantum gravity. *Comm. Math. Phys.*, 147(2):395–430, 1992.

[22] P. J. Forrester. *Log-gases and random matrices*, volume 34 of *London Mathematical Society Monographs Series*. Princeton University Press, Princeton, NJ, 2010.

[23] F. Götze and M. Venker. Local universality of repulsive particle systems and random matrices. 2013. arXiv:1205.0671, to appear in *Ann. Probab.*

[24] A. R. Its, A. B. J. Kuijlaars, and J. Östensson. Critical edge behavior in unitary random matrix ensembles and the thirty-fourth Painlevé transcendent. *Int. Math. Res. Not. IMRN*, (9):Art. ID rnn017, 67, 2008.

[25] A. Kamenev and M. Mézard. Wigner - Dyson statistics from the replica method. *J. Phys. A: Math. Gen.*, 32:4373–4388, 1999.

[26] N. M. Katz and P. Sarnak. *Random matrices, Frobenius eigenvalues, and monodromy*, volume 45 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 1999.

[27] T. Kriecherbauer and J. Krug. A pedestrian's view on interacting particle systems, KPZ universality and random matrices. *J. Phys. A*, 43(40):403001, 41, 2010.

[28] T. Kriecherbauer and K. T.-R. McLaughlin. Strong asymptotics of polynomials orthogonal with respect to Freud weights. *Internat. Math. Res. Notices*, (6):299–333, 1999.

[29] T. Kriecherbauer and K. Schubert. Spacings: An example for universality in random matrix theory. In Gerold Alsmeyer and Matthias Löwe, editors, *Random Matrices and Iterated Random Functions*. Springer Berlin Heidelberg, 2013.

[30] T. Kriecherbauer and M. Venker. Edge universality for a class of repulsive particle systems. In preparation.

[31] A. B. J. Kuijlaars. Universality. In *The Oxford handbook of random matrix theory*, pages 103–134. Oxford Univ. Press, Oxford, 2011.

[32] A. B. J. Kuijlaars and K. T-R McLaughlin. Generic behavior of the density of states in random matrix theory and equilibrium problems in the presence of real analytic external fields. *Comm. Pure Appl. Math.*, 53(6):736–785, 2000.

[33] A. B. J. Kuijlaars, K. T.-R. McLaughlin, W. Van Assche, and M. Vanlessen. The Riemann-Hilbert approach to strong asymptotics for orthogonal polynomials on $[-1, 1]$. *Adv. Math.*, 188(2):337–398, 2004.

[34] A. B. J. Kuijlaars and M. Vanlessen. Universality for eigenvalue correlations from the modified Jacobi unitary ensemble. *Int. Math. Res. Not.*, (30):1575–1600, 2002.

[35] A. B. J. Kuijlaars and M. Vanlessen. Universality for eigenvalue correlations at the origin of the spectrum. *Comm. Math. Phys.*, 243(1):163–191, 2003.

[36] E. Levin and D. S. Lubinsky. Universality limits in the bulk for varying measures. *Adv. Math.*, 219(3):743–779, 2008.

[37] E. Levin and D. S. Lubinsky. Universality limits for exponential weights. *Constr. Approx.*, 29(2):247–275, 2009.

[38] E. Levin and D. S. Lubinsky. Universality limits at the soft edge of the spectrum via classical complex analysis. *Int. Math. Res. Not. IMRN*, (13):3006–3070, 2011.

[39] D. S. Lubinsky. Universality limits in the bulk for arbitrary measures on compact sets. *J. Anal. Math.*, 106:373–394, 2008.

[40] D. S. Lubinsky. A new approach to universality limits involving orthogonal polynomials. *Ann. of Math. (2)*, 170(2):915–939, 2009.

[41] D. S. Lubinsky. Universality in the bulk holds close to given points. *J. Approx. Theory*, 163(7):904–922, 2011.

[42] D. S. Lubinsky. Bulk universality holds in measure for compactly supported measures. *J. Anal. Math.*, 116:219–253, 2012.

[43] D. S. Lubinsky. Bulk universality holds pointwise in the mean for compactly supported measures. *Michigan Math. J.*, 61(3):631–649, 2012.

[44] A. Lytova and L. Pastur. On asymptotic behavior of multilinear eigenvalue statistics of random matrices. *J. Stat. Phys.*, 133(5):871–882, 2008.

[45] K. T.-R. McLaughlin and P. D. Miller. The $\overline{\partial}$ steepest descent method for orthogonal polynomials on the real line with varying weights. *Int. Math. Res. Not. IMRN*, pages Art. ID rnn 075, 66, 2008.

[46] M. L. Mehta. *Random matrices*, volume 142 of *Pure and Applied Mathematics (Amsterdam)*. Elsevier/Academic Press, Amsterdam, third edition, 2004.

[47] M. L. Mehta and M. Gaudin. On the density of eigenvalues of a random matrix. *Nuclear Phys.*, 18:420–427, 1960.

[48] L. Pastur. Spectral and probabilistic aspects of matrix models. In *Algebraic and geometric methods in mathematical physics (Kaciveli, 1993)*, volume 19 of *Math. Phys. Stud.*, pages 207–242. Kluwer Acad. Publ., Dordrecht, 1996.

[49] L. Pastur and M. Shcherbina. Universality of the local eigenvalue statistics for a class of unitary invariant random matrix ensembles. *J. Statist. Phys.*, 86(1-2):109–147, 1997.

[50] L. Pastur and M. Shcherbina. Bulk universality and related properties of Hermitian matrix models. *J. Stat. Phys.*, 130(2):205–250, 2008.

[51] L. Pastur and M. Shcherbina. *Eigenvalue distribution of large random matrices*, volume 171 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2011.

[52] B. Rider and X. Zhou. Janossy densities for unitary ensembles at the spectral edge. *Int. Math. Res. Not. IMRN*, (13):Art. ID rnn037, 51, 2008.

[53] K. Schubert. Spacings in orthogonal and symplectic random matrix ensembles. In preparation.

[54] K. Schubert and M. Venker. Spacings in repulsive particle systems - uniform convergence to the universal GUE spacing distribution. In preparation.

[55] K. Schüler. Moderate and large deviations for extremal eigenvalues of unitarily invariant ensembles. PhD thesis, in preparation.

[56] B. Simon. Two extensions of Lubinsky's universality theorem. *J. Anal. Math.*, 105:345–362, 2008.

[57] G. Szegő. *Orthogonal polynomials*. American Mathematical Society, Providence, R.I., fourth edition, 1975. American Mathematical Society, Colloquium Publications, Vol. XXIII.

[58] V. Totik. Universality and fine zero spacing on general sets. *Ark. Mat.*, 47(2):361–391, 2009.

[59] C. A. Tracy and H. Widom. Level-spacing distributions and the Airy kernel. *Comm. Math. Phys.*, 159(1):151–174, 1994.

[60] C. A. Tracy and H. Widom. Correlation functions, cluster functions, and spacing distributions for random matrices. *J. Statist. Phys.*, 92(5-6):809–835, 1998.

[61] M. Vanlessen. Strong asymptotics of the recurrence coefficients of orthogonal polynomials associated to the generalized Jacobi weight. *J. Approx. Theory*, 125(2):198–237, 2003.
[62] M. Vanlessen. Strong asymptotics of Laguerre-type orthogonal polynomials and applications in random matrix theory. *Constr. Approx.*, 25(2):125–175, 2007.
[63] M. Venker. Particle systems with repulsion exponent β and random matrices. *Electron. Commun. Probab.*, 18:no. 83, 1–12, 2013.

Inst. for Mathematics, Univ. Bayreuth, 95440 Bayreuth, Germany
E-mail address: thomas.kriecherbauer@uni-bayreuth.de

Inst. for Math. Stat., Univ. Münster, Einsteinstr. 62, 48149 Münster, Germany
E-mail address: kristina.schubert@uni-muenster.de

Inst. for Mathematics, Univ. Bayreuth, 95440 Bayreuth, Germany
E-mail address: katharina.schueler@uni-bayreuth.de

Fac. of Mathematics, Univ. Bielefeld, P.O.Box 100131, 33501 Bielefeld, Germany
E-mail address: mvenker@math.uni-bielefeld.de