

# SPECTRAL CURVES, VARIATIONAL PROBLEMS AND THE HERMITIAN MATRIX MODEL WITH EXTERNAL SOURCE

ANDREI MARTÍNEZ-FINKELSHTEIN AND GUILHERME L. F. SILVA

ABSTRACT. We consider the hermitian random matrix model with external source and general polynomial potential, when the source has two distinct eigenvalues but is otherwise arbitrary.

All such models studied so far have a common feature: an associated cubic equation (“spectral curve”), one of whose solutions can be expressed in terms of the Cauchy (a.k.a. Stieltjes) transform of the limiting eigenvalue distribution  $\lambda$ .

This is our starting point: we show that to any such a spectral curve (not necessarily given by a random matrix ensemble) it corresponds a unique vector-valued measure with three components on the complex plane, characterized as a solution of a variational problem stated in terms of their logarithmic energy. We describe all possible geometries of the supports of these measures: the third component, if non-trivial, lives on a contour on the plane and separates the supports of the other two measures, both on the real line.

This general result is applied to the random matrix model with external source, under an additional assumption of uniform boundedness of the zeros of a sequence of average characteristic polynomials, when the size of the matrices goes to infinity (equivalently, uniform boundedness of certain recurrence coefficients). It is shown that any limiting zero distribution for such a sequence can be written in terms of a solution of a spectral curve, and thus admits the variational description obtained in the first part of the paper. As a consequence of our analysis, we obtain that the density of this limiting measure can have only a handful of local behaviors: sine, Airy and their higher order type behavior, Pearcey or yet the fifth power of the cubic (but no higher order cubics can appear).

We also compare our findings with the most general results available in the literature, showing that once an additional symmetry is imposed, our vector critical measure contains enough information to recover the solutions to the constrained equilibrium problem that was known to describe the limiting eigenvalue distribution in this symmetric situation.

## CONTENTS

1.	Introduction	2
2.	Statement of main results	6
3.	Construction of the vector critical measures	16
4.	From critical measures to constrained equilibrium	56
5.	Multiple orthogonal polynomials	65

---

2010 *Mathematics Subject Classification*. Primary: 15B52; Secondary: 30C15, 30F30, 31A15, 42C05.

*Key words and phrases*. Random matrix theory, logarithmic potential theory, vector energy, variational problems, critical measures, trajectories of quadratic differentials, recurrence coefficients.

6. Characterization of the eigenvalue distribution	75
Acknowledgments	83
References	83

## 1. INTRODUCTION

The main motivation of this work is the random matrix ensemble defined on the space  $\mathcal{M}_N$  of  $N \times N$  hermitian matrices equipped with the probability distribution

$$\frac{1}{Z_N} e^{-N \operatorname{Tr}(V(\mathbf{M}) - \mathbf{A}\mathbf{M})} d\mathbf{M}, \quad (1.1)$$

where  $V$  is a polynomial,  $\mathbf{A}$  is a fixed  $N \times N$  hermitian matrix (the *external source*),  $d\mathbf{M}$  is the Lebesgue measure on  $\mathcal{M}_N \simeq \mathbb{R}^{N^2}$ , and  $Z_N$  is the normalization constant, also called partition function. Due to the external source, this ensemble is not unitary invariant, but because of the invariance of  $d\mathbf{M}$  and the trace under unitary conjugation, without loss of generality we can assume that  $\mathbf{A}$  is diagonal and that  $V(0) = 0$ .

The most studied case is when  $\mathbf{A}$  has exactly two distinct eigenvalues,  $a_1$  and  $a_2$  with multiplicities  $n_1$  and  $n_2$ . Since for any  $c \in \mathbb{R}$ ,

$$V(\mathbf{M}) - \mathbf{A}\mathbf{M} = (V(\mathbf{M}) - c\mathbf{M}) + (c\mathbf{I} - \mathbf{A})\mathbf{M},$$

we can symmetrize the problem at the cost of changing the linear term of  $V$  by choosing

$$c = \frac{a_1 + a_2}{2}, \quad a = \frac{a_1 - a_2}{2},$$

so that the eigenvalues of the new external source  $\mathbf{A} - c\mathbf{I}$  are  $-a$  and  $a$ . In other words, without loss of generality we consider the situation when

$$\mathbf{A} = \operatorname{diag}(\underbrace{a_1, a_1, \dots, a_1}_{n_1 \text{ times}}, \underbrace{a_2, a_2, \dots, a_2}_{n_2 \text{ times}}), \quad a_1 = -a_2 = a \geq 0. \quad (1.2)$$

The asymptotic analysis of statistical properties (as  $N \rightarrow \infty$ ) of the spectra of random matrices perturbed by an additive external source can be traced back to Pastur [80]. In the late 1990's Zinn-Justin [95] showed that the eigenvalues of the model (1.1) are determinantal, and later Bleher and Kuijlaars [34] proved that the correlation kernel that appears in this determinantal expression can be written in terms of multiple orthogonal polynomials. The now well-known matrix-valued Riemann-Hilbert characterization for these polynomials (involving matrices whose size depends on the number of distinct eigenvalues of  $\mathbf{A}$ ) turned out to be crucial for the asymptotic analysis of external source models. Aptekarev, Bleher, and Kuijlaars were the first to develop this approach in a series of papers [33, 5, 35], where they analyzed the symmetric Gaussian case, when  $V(\mathbf{M}) = \mathbf{M}^2/2$  and  $\mathbf{A}$  is given by (1.2), with  $n_1 = n_2 \rightarrow \infty$ , recovering and extending results obtained a little earlier by Brézin and Hikami [43, 42] and Tracy and Widom [90]. It is worth mentioning that the Gaussian random matrix model with an external source is equivalent to the so-called deformed

GUE ensemble and with the Dyson Brownian motion (with parameter  $\beta = 2$ ), studied by Duits, Erdős, Johansson, Landon, Yau and others (see e.g. [46, 70, 52, 55, 56, 71]).

Riemann-Hilbert characterization of multiple orthogonal polynomials as a tool for the asymptotic analysis of the random matrix ensemble (1.1) has been successfully exploited in [31, 33, 34, 5, 10, 20, 19]. Further results have also been obtained by other authors, see for instance [14, 16, 15], and especially the recent monograph [44].

Let us summarize the key ingredients in some of the above mentioned asymptotic results.

Motivated by [33, 5, 35], McLaughlin [78] made an important observation that in an “ideal situation”, when the original Riemann-Hilbert problem can be appropriately normalized with the help of certain “ $g$ -functions”, there exists a function  $\xi(z)$  that encodes the Cauchy (or Stieltjes) transform of the limiting density of eigenvalues of (1.1) for the ray sequences of  $(n_1, n_2)$  with the slope  $\alpha$ , i.e. in the asymptotic regime

$$N := n_1 + n_2 \rightarrow \infty, \quad \frac{n_1}{N} \rightarrow \alpha \in (0, 1). \quad (1.3)$$

This function satisfies a cubic equation, also known as *spectral curve* or *master loop equation* (to be more precise, the spectral curve is the locus of solutions of equation (1.3), but we allow ourselves to use this term in a loose sense) of the form

$$\xi^3 - V'(z)\xi^2 + p_1(z)\xi + p_0(z) = 0, \quad (1.4)$$

where  $p_1$  and  $p_0$  are analytic functions of  $z$ . If  $V$  is a polynomial with real coefficients, then the  $p_j$ 's are also polynomials, and  $\xi = \xi(z)$  is an algebraic function. Furthermore, for a monic quartic potential  $V(z) = z^4/4$ , McLaughlin obtained an ansatz for (1.4) and used this to show that in this case his ideal situation indeed takes place. However, he also observed that for general  $V$  this is not always the case.

It is worth pointing out that spectral curves for matrix models other than (1.1) have also been obtained before. We refer the reader to the works [68, 50, 21, 17, 93, 64] for an account on the existence and relevance of spectral curves in random matrix theory.

With an extension of the Riemann-Hilbert approach of [33, 35, 5], Bleher, Delvaux and Kuijlaars [31] considered the asymptotic density of the eigenvalues of (1.1) with  $\mathbf{A}$  as in (1.2), imposing the additional symmetry that the potential  $V$  is an *even* polynomial and  $n_1 = n_2 \rightarrow \infty$  (which yields  $\alpha = 1/2$  in (1.3)). They showed that in this case the limiting density of eigenvalues can be described in terms of the solution of a variational problem involving two measures, one of them subject to a given upper bound (the so-called constrained vector equilibrium). They proved the existence and uniqueness of the asymptotic density, showing also the existence of the spectral curve (1.4). In addition, in the case of even quartic potential  $V$  they obtained the coefficients of (1.4) explicitly.

Independently, but almost simultaneously, Aptekarev, Lysov and Tulyakov [9, 10] tackled the same problem for the even quartic potential when  $n_1 = n_2 \rightarrow \infty$ . Their starting point was an ansatz for the spectral curve of the model, which was all they needed for their asymptotic analysis. Such ansatz is possible in virtue of the symmetry and low complexity of their potential  $V$ . In contrast with [31], they found a different variational description for the asymptotic density of the eigenvalues

of (1.1), now in terms of a vector equilibrium (no constraints) involving three measures, one of them living on a contour in the complex plane. Similar complex vector equilibrium with three components, as discussed in [10], has appeared before in the asymptotic analysis of a certain class of multiple orthogonal polynomials, see e.g. [7, 82, 8]. They also observed in [10] that the constrained equilibrium problem had no evident extension to the non-symmetric case of non-even potential  $V$ , and expressed their belief that their equilibrium with the complex contour should be more suitable for a general situation.

In short, in the limited number of cases where a characterization of the limiting density of eigenvalues of (1.1) with  $\mathbf{A}$  as in (1.2) has been established, it was described either in terms of a spectral curve of the form (1.4), or in terms of solutions of variational problems for vector-valued measures on the plane. Both characterizations, technically equivalent, can be used to build certain functions ( $g$ -functions) that are crucial for the asymptotic Riemann-Hilbert analysis. An advantage of the variational approach is that the components of the vector-valued measure can be used to describe the asymptotic eigenvalue distribution of the random matrix ensemble explicitly, but so far the most general rigorous analysis available is still restricted to a symmetric situation.

In our previous work [76] we showed that if the variational problem of the type we consider here has a solution (in an appropriate sense), then there exists an associated spectral curve. The first part of the present work (Sections 2.1 and 3) provides a converse statement in the setup motivated by the matrix model (1.1) but not necessarily restricted to it. Namely, assuming that  $V$  is an *arbitrary* real polynomial and that the cubic equation (1.4) satisfies certain natural conditions, we prove the *existence* of the solution of an associated variational problem involving three measures on the complex plane (vector-valued *critical measure*). It turns out that the components of this measure exhibit a symmetry property, known as the *S-property*, which should ultimately characterize the support of the critical measure.

When compared to the just mentioned works related to (1.1), the main novelty of our approach is the introduction of a quadratic differential on the Riemann surface associated to (1.4) and the study of its critical graph. This idea has been developed in our previous works [76, 77], and also in [36], and allows us to, starting solely from (1.4), extract the support for the measures that constitute the solution to our variational problem.

It turns out that in the situation presented here, the critical graph is sufficiently rigid so that

- there are only two qualitatively different configurations that allow us to classify two asymptotic regimes, described as *saturated* and *unsaturated* regimes, see Section 3.5.
- there are only two possible singular behaviors of the critical measures, corresponding to either square or cube roots in the density of its components, as explained in Theorem B in Section 2.1. The latter happens only in the unsaturated regime, and only with low powers of the cube root.

In the second part of our work (Sections 2.2.1 and 4), we apply our results to the analysis of the asymptotic density of the eigenvalues of the ensemble (1.1) with  $\mathbf{A}$  as in (1.2), comparing it with the totally symmetric situation considered by Bleher, Delvaux and Kuijlaars [31], when the potential

$V$  is an even polynomial and  $\alpha = 1/2$ . We show the equivalence of the variational problem for vector-valued measures with three components and the constrained vector equilibrium from [31]. This yields an alternative description of the asymptotic distribution found in [31], that extends the analogous result of [10] from the quartic to any even potential  $V$ . An advantage of our electrostatic model is that it is valid for arbitrary polynomial  $V$  and  $\alpha \in (0, 1)$ , making it natural to expect that vector-valued critical measures describe the eigenvalue distribution even in the most general case, in accordance with the conjecture in [10].

One of the main powers of the above-mentioned Riemann-Hilbert approach is that it provides a comprehensive asymptotic description in different scaling regimes, including the limiting eigenvalue distribution, the local scaling limits such as the classical Airy and Sine kernel universality-type results, etcetera. But this power can be seen as a drawback as well, as if one is interested in only one of such quantities, one still needs to complete all technical steps that are firmly tied to the other quantities as well.

As an alternative to the Riemann-Hilbert methods, several recent works have been developed to analyze the limiting eigenvalue distribution of random matrix models (in fact, these works deal with (subclasses of) determinantal point processes that encompass the eigenvalues of the model (1.1)) and its fluctuations [41, 69, 59, 40, 60]. In common, their analysis is based on the recurrence relation coefficients for the functions used to construct the correlation kernel of the particle system.

In the third and last part of this work (Sections 2.2.2 and 6), we use a combination of the just-mentioned approach and of an analogue of a WKB analysis to study the limiting eigenvalue distribution from the perspective of the recurrence coefficients. We start by introducing a differential equation for the wave functions associated to the model, and for which its characteristic equation becomes a “finite  $N$ ” version of the spectral curve. Assuming that the zeros of the average characteristic polynomial remain uniformly bounded (fact that we believe is always true, and needs not to be imposed as an additional hypothesis, see Remark 2.7), we show that the associated recurrence coefficients remain bounded as well. This, in turn, yields uniform boundedness of the coefficients in the finite  $N$  spectral curve. As a consequence, when  $N \rightarrow \infty$ , we establish existence along any convergent subsequence of the spectral curve studied in the first part, describing the limiting eigenvalue distribution of the matrix model.

In particular, this shows that in this situation the asymptotic eigenvalue distribution has either the same local behavior as in the one-matrix model (power of square-root behavior), corresponding to  $\mathbf{A} = \mathbf{0}$  and giving rise to Painlevé-type hierarchy limiting kernels, or vanishes as a cubic root or the fifth power of the cubic root in the interior of its support. The cubic root behavior is associated to the known Pearcey kernel (this situation appears, for instance, when  $n_1 = n_2$ ,  $a = 1$  and  $V(z) = z^2/2$ ), and although an example of a potential with limiting density of eigenvalues vanishing with power  $5/3$  is implicit in the literature (see the comments after Theorem B below), to our knowledge the corresponding limiting process has not yet been described in the literature. In short, our results mean that there is no higher-order Pearcey kernel at the bulk of the spectrum, other than the ones associated with  $1/3$  and  $5/3$  powers, nor powers of cubic root like behavior at the soft edges for the model (1.1).

Without further ado, let us discuss our main findings.

## 2. STATEMENT OF MAIN RESULTS

Throughout this paper, we use the notation  $\mathbb{R}_+ := (0, +\infty)$ ,  $i\mathbb{R} := \{ix : x \in \mathbb{R}\}$ ,

$$\mathbb{C}_\pm := \{z \in \mathbb{C} \mid \pm \operatorname{Re} z > 0\}, \quad \mathbb{H}_\pm := \{z \in \mathbb{C} \mid \pm \operatorname{Im} z > 0\}. \quad (2.1)$$

We define the zero counting measure of any polynomial  $q$  by

$$\mu(q) := \frac{1}{\deg q} \sum_{q(w)=0} \delta_w, \quad (2.2)$$

where each zero on the right-hand side is counted according to its multiplicity, so that  $|\mu(q)| := \mu(\mathbb{C}) = 1$ . We will understand by asymptotic zero distribution any limit of  $\mu(q)$  in the weak-\* sense (that we denote by  $\xrightarrow{*}$ ) as  $\deg q \rightarrow \infty$ .

The Cauchy transform of any finite Borel measure  $\nu$  on  $\mathbb{C}$  is given by

$$C^\nu(z) := \lim_{\varepsilon \rightarrow 0^+} \int_{|z-s|>\varepsilon} \frac{d\nu(s)}{s-z}, \quad z \in \mathbb{C}. \quad (2.3)$$

### 2.1. Vector critical measures from the spectral curve.

#### 2.1.1. The spectral curve.

As we pointed out in the Introduction, our primary motivation is the asymptotic behavior of the eigenvalues of the random matrices from the ensemble (1.1)–(1.2) in the asymptotic regime (1.3). This model is a particular case of a multiple orthogonal polynomial ensemble, and as discussed before, under mild conditions studying the global asymptotic regime of eigenvalues is equivalent to studying the weak-\* convergence, as  $N = n_1 + n_2 \rightarrow \infty$ , of the zero-counting measure (see the definition (2.2))  $\mu(\pi_{n_1, n_2})$  for the average characteristic polynomial [60]

$$\pi_{n_1, n_2}(z) := \mathbb{E} \det(z\mathbf{I} - \mathbf{M}). \quad (2.4)$$

For even polynomial potential  $V$  and for  $n_1 = n_2 = N/2$ , the authors of [31] found the spectral curve of the matrix model in the form of the algebraic equation (1.4), where  $p_0$  and  $p_1$  are polynomials that, in general, cannot be fully determined only from  $V$  and  $a$ . Moreover, they established that  $\mu(\pi_{n_1, n_2}) \xrightarrow{*} \nu$  as  $N \rightarrow \infty$ , and that  $\xi = \xi(z) = C^\nu(z) + V'(z)$  is one of the solutions of (1.4), where  $C^\nu$  is the Cauchy transform of  $\nu$ , defined in (2.3).

Such a spectral curve is a central object and starting point of our paper, so we turn it into a formal definition that synthesizes some general properties of such algebraic equations described in the existing literature.

**Definition 2.1.** Given a polynomial  $V$  of arbitrary degree  $m \geq 2$ ,

$$V(z) = \sum_{k=1}^m \frac{v_k}{k} z^k, \quad v_1, \dots, v_m \in \mathbb{R}, \quad v_m \neq 0, \quad (2.5)$$

and parameters  $(a, \alpha) \in \mathbb{R}_+ \times (0, 1)$ , an associated *admissible spectral curve* is an irreducible algebraic equation of the form

$$F(\xi, z) := \xi^3 + p_2(z)\xi^2 + p_1(z)\xi + p_0(z) = 0, \quad (2.6)$$

where  $p_0, p_1$  and  $p_2$  are real polynomials, with

$$\begin{aligned} p_2(z) &= -V'(z), \\ \deg p_1 &\leq m-2, \\ p_0(z) &= v_m a^2 z^{m-1} + (v_{m-1} a^2 + v_m a (2\alpha - 1)) z^{m-2} + \text{lower degree terms}, \end{aligned} \quad (2.7)$$

such that there exists a probability measure  $\lambda$  on  $\mathbb{R}$  for which the function

$$\xi(z) = C^\lambda(z) + V'(z) \quad (2.8)$$

is a solution to (2.6) for  $z \in \mathbb{C} \setminus \mathbb{R}$ .

*Remark 2.2.* Let us insist that Definition 2.1 focuses on the algebraic equation (2.6) only and does not assume a priori the existence of an underlying random matrix model. However, as the results from Sections 2.2.2 and 6 show, admissibility is a natural assumption in the context of Random Matrix Theory (see also the models studied in [33, 35, 31, 34, 5, 10, 20, 19, 9]).

Although playing a role in our approach, the conditions (2.7) should not be seen as a restriction but rather as a normalization, fixing the potential  $V$  and the parameters  $\alpha$  and  $a$ . The crucial condition is that one of the solutions of (2.6) can be written in terms of the Cauchy transform of a finite positive measure on  $\mathbb{R}$ . Algebraic equations admitting solutions in terms of Cauchy transforms of positive measures are rather special, appearing frequently in random matrix theory but also in other contexts, see e.g. [38, 39, 75, 87, 86, 85, 37, 83, 3].

Notice that the model (1.1) is only well-defined if  $m$  is even and  $v_m > 0$ , but some of our main findings do not need such a requirement. Obviously, when we apply our results to this random matrix model, we will then impose these conditions *en route*.

### 2.1.2. Critical measures.

Now we introduce the second ingredient relevant to the formulation of the main results of this section: the variational problem for vector measures. For that, we describe both the family of the measures and the energy functional acting on them.

The mutual (logarithmic) energy of two finite Borel measures  $\mu$  and  $\nu$  on  $\mathbb{C}$  is

$$I(\mu, \nu) := \iint \log \frac{1}{|s - z|} d\mu(s) d\nu(z);$$

we refer to  $I(\mu, \mu)$  simply as the energy of  $\mu$ .

For the real polynomial  $V$  as in (2.5) define

$$V_j(z) := V(z) - a_j z, \quad j = 1, 2, \quad V_3(z) := V_2(z) - V_1(z) = 2az, \quad (2.9)$$

where we use the convention that  $a_1 = -a_2 = a$ , and set

$$\phi_j(z) := \operatorname{Re} V_j(z), \quad z \in \mathbb{C}, \quad j = 1, 2, 3. \quad (2.10)$$

We define by  $\mathcal{T}$  the class of rectifiable contours  $\Gamma$  that are symmetric under complex conjugation, intersect  $\mathbb{R}$  at a single point and extend to  $\infty$  along their two ends in such a way that

$$\lim_{\substack{z \rightarrow \infty \\ z \in \Gamma}} \operatorname{Re} z = +\infty. \quad (2.11)$$

Associated to a contour  $\Gamma \in \mathcal{T}$  and a parameter  $\alpha \in (0, 1)$  we introduce  $\mathcal{M}_\alpha(\Gamma)$ , the set of vectors of non-negative measures  $\vec{\mu} = (\mu_1, \mu_2, \mu_3)$  satisfying the following conditions:

- Each component  $\mu_j$  is a Borel measure with finite energy and for which  $\int \phi_j d\mu_j$  is finite;
- $\operatorname{supp} \mu_1 \subset \mathbb{R}$ ,  $\operatorname{supp} \mu_2 \subset \mathbb{R}$  and  $\operatorname{supp} \mu_3 \subset \Gamma$ ;
- $|\mu_1| + |\mu_2| = 1$ ,  $|\mu_1| - |\mu_3| = \alpha \in (0, 1)$ .

Notice that any one of the two identities relating the masses of  $\mu_1, \mu_2$  and  $\mu_3$  can be equivalently replaced by

$$|\mu_2| + |\mu_3| = 1 - \alpha.$$

We introduce the *interaction matrix*

$$\mathbf{B} = (\mathbf{b}_{j,k})_{1 \leq j,k \leq 3} = \begin{pmatrix} 1 & 1/2 & -1/2 \\ 1/2 & 1 & 1/2 \\ -1/2 & 1/2 & 1 \end{pmatrix}. \quad (2.12)$$

Its structure shows that the interaction between  $\mu_1$  and  $\mu_2$  or between  $\mu_2$  and  $\mu_3$  is of repulsive or Angelesco character, while between  $\mu_1$  and  $\mu_3$  the interaction is of attractive or Nikishin type, see e.g. [65].

For  $\alpha \in (0, 1)$ ,  $\Gamma \in \mathcal{T}$  and for  $\vec{\mu} \in \mathcal{M}_\alpha(\Gamma)$ , the *vector energy*

$$E_S(\vec{\mu}) := \sum_{j,k=1}^3 \mathbf{b}_{j,k} I(\mu_j, \mu_k) + \sum_{j=1}^3 \int \phi_j d\mu_j, \quad (2.13)$$

is well defined and finite. The *vector equilibrium measure* of  $\Gamma$  is the vector  $\vec{\mu}_\Gamma \in \mathcal{M}_\alpha(\Gamma)$  for which

$$E_S(\vec{\mu}_\Gamma) = \inf_{\vec{\mu} \in \mathcal{M}_\alpha(\Gamma)} E_S(\vec{\mu});$$

if  $m$  is even and  $v_m > 0$  then our conditions on  $\mathcal{M}_\alpha(\Gamma)$  guarantee both existence and uniqueness of  $\vec{\mu}_\Gamma$ . Although the first two components of  $\vec{\mu}_\Gamma$  live on  $\mathbb{R}$ , we refer to  $\vec{\mu}_\Gamma$  as the vector equilibrium measure of  $\Gamma$  instead of  $(\mathbb{R}, \Gamma)$ . We stress also that  $\vec{\mu}_\Gamma$  does depend on  $V$ ,  $\alpha$  and  $a$ , but we do not make this dependence explicit in our notation.

Closely related to equilibrium measures is the notion of *vector critical measures*, i.e. saddle points of the energy functional  $E_S(\vec{\mu})$ . To be more precise, given a complex-valued function  $h \in C_c^2(\mathbb{C})$  (i.e., twice real differentiable with continuous second derivatives and compactly supported on  $\mathbb{C}$ ), we denote by  $\mu^t$  the pushforward of  $\mu$  by the map  $z \mapsto z + th(z)$ ,  $t \in \mathbb{R}$ , and set  $\vec{\mu}^t = (\mu_1^t, \mu_2^t, \mu_3^t)$ .



**Definition 2.3.** A vector of measures  $\vec{\mu} \in \mathcal{M}_\alpha(\Gamma)$  is called *critical* for the energy functional (2.13) if

$$\lim_{t \rightarrow 0} \frac{E_S(\vec{\mu}^t) - E_S(\vec{\mu})}{t} = 0, \quad (2.14)$$

for every  $h \in C_c^2(\mathbb{C})$ .

See [75] for the notion of scalar critical measure and its applications, and its extension in [76] to the vector setting.

Our first major result shows that each admissible spectral curve has a unique associated critical vector measure for the energy functional (2.13), and that its components satisfy certain symmetry conditions. In its formulation we use the notion of the *logarithmic potential* of a Borel measure  $\mu$  on  $\mathbb{C}$ ,

$$U^\mu(z) := \int \log \frac{1}{|s - z|} d\mu(s), \quad z \in \mathbb{C}.$$

**Theorem A.** Suppose that given  $V$ ,  $\alpha$  and  $a$ , (2.6) is an admissible spectral curve as described in Definition 2.1. Then there exists a unique vector critical measure  $\vec{\mu}_* = (\mu_1^*, \mu_2^*, \mu_3^*) \in \mathcal{M}_\alpha(\Gamma_*)$  for some  $\Gamma_* \in \mathcal{T}$ , such that the admissibility condition in Definition 2.1 is satisfied for

$$\lambda = \mu_1^* + \mu_2^*. \quad (2.15)$$

Furthermore, there exists  $x_* \in \mathbb{R}$  for which  $\text{supp } \mu_1 \subset [x_*, +\infty)$  and  $\text{supp } \mu_2 \subset (-\infty, x_*]$ . In particular, these supports intersect in at most one point.

If  $\mu_3^* \neq 0$ , then  $\text{supp } \mu_3^*$  is a bounded and connected arc with connected complement in  $\mathbb{C}$ , symmetric with respect to  $\mathbb{R}$ , and such that  $\text{supp } \mu_3^* \cap \mathbb{H}_+$  is an analytic arc. Still when  $\mu_3^* \neq 0$ , either  $\text{supp } \mu_1^* \cap \text{supp } \mu_2^* \cap \text{supp } \mu_3^*$  consists of a single point (necessarily  $x_*$  as above) or the supports of the measures are pairwise disjoint.

The components of the vector critical measure  $\vec{\mu}_*$  satisfy the following variational identities: for  $i = 1, 2, 3$ ,

$$\sum_{k=1}^3 \mathbf{b}_{i,k} U^{\mu_k^*}(z) + \frac{1}{2} \phi_i(z) \equiv \ell \quad (2.16)$$

along each connected component of  $\text{supp } \mu_i^*$ , for some constant  $\ell$  that might depend on the connected component.

Finally, on  $\text{supp } \mu_3^*$  the measure  $\vec{\mu}_*$  exhibits the S-property

$$\left( \frac{\partial}{\partial n_+} - \frac{\partial}{\partial n_-} \right) \left( \sum_{k=1}^3 \mathbf{b}_{3,k} U^{\mu_k^*}(z) + \frac{1}{2} \phi_3(z) \right) = 0, \quad z \in \text{supp } \mu_3^*, \quad (2.17)$$

where  $n_\pm$  denote the two unit normal vectors to  $\text{supp } \mu_3^*$  at  $z$  in the opposite directions.

Obviously, in the result just stated if  $\text{supp } \mu_3^* = \emptyset$ , the corresponding conditions in (2.16) and (2.17) are void.

The contour  $\Gamma_*$  will be more precisely described in Propositions 3.15 and 3.18 below, in terms of trajectories of an associated quadratic differential.

*Remark 2.4.* The fact that expressions in (2.16) are constant along each connected component of the support of the corresponding measure is a characterizing property of critical measures (see [76, Theorem 1.8]). In the case of the equilibrium measure on  $\Gamma_*$  we would have the equality of these constants on each connected component of  $\text{supp } \mu_i^*$ . The latter is equivalent to imposing an additional, the so-called *Boutroux condition* on the spectral curve [17], see the detailed discussion in Section 3.8, especially Remark 3.34.

On the other hand, not every critical measure is an equilibrium measure: there is an additional property that can be written as a set of inequalities satisfied by the left hand side of (2.16). For instance, it is possible to construct spectral curves, in the sense of Definition 2.1 but with  $a = 0$ , for which the critical measure in Theorem A does not coincide with the corresponding equilibrium measure on  $\mathbb{R}$ . It is natural to expect from our techniques that this situation can happen even when  $a > 0$ . In other words, the statement of Theorem A is sharp, and without a stronger condition on the spectral curve we cannot assure that the critical measure is in fact also an equilibrium measure.

When the matrix  $\mathbf{A} = \mathbf{0}$ , the model (1.1)–(1.2) reduces to the classical Hermitian random matrix model. For such models, it is known that the spectral curve exists as an algebraic equation of degree 2 (so that, actually, the expression (2.6) becomes reducible, see Proposition 3.29 below). In this situation, it follows from our arguments that  $\mu_2^* = \mu_3^* = 0$  and  $\mu_1^*$  is the standard critical or equilibrium measure in the external field  $V$ , see also Remark 3.30 below.

It turns out that the admissibility of a spectral curve, described in Definition 2.1, limits the variety of possible local behaviors of the measure  $\lambda$ .

**Theorem B.** *Suppose that given  $V$ ,  $\alpha$  and  $a$ , (2.6) is an admissible spectral curve as described in Definition 2.1. Then the measure  $\lambda = \mu_1^* + \mu_2^*$  is absolutely continuous with respect to the Lebesgue measure on  $\mathbb{R}$ , its support is a finite union of disjoint intervals, and its density vanishes on  $\text{supp } \lambda$  only at a finite number of points where it enjoys the following properties:*

(i) *If  $x_0$  is a zero of  $\frac{d\lambda}{dx}$  in the interior of  $\text{supp } \lambda$ , then for some  $c > 0$  either*

$$\frac{d\lambda}{dx}(x) = c|x - x_0|^k(1 + \mathcal{O}(x - x_0)), \quad x \rightarrow x_0, \quad (2.18)$$

*for a positive integer  $k$ , or for  $\nu \in \{1, 5\}$ ,*

$$\frac{d\lambda}{dx}(x) = c|x - x_0|^{\frac{\nu}{3}}(1 + \mathcal{O}(x - x_0)), \quad x \rightarrow x_0. \quad (2.19)$$

*Furthermore, the latter case happens only when  $\mu_3^* = 0$  and  $\text{supp } \mu_1^* \cap \text{supp } \mu_2^* = \{x_0\} = \{x_*\}$  with  $x_*$  as in Theorem A, which determines this point uniquely.*

(ii) If  $x_0$  is an endpoint of  $\text{supp } \lambda$  then for some  $c > 0$  and  $k \in \mathbb{N} \cup \{0\}$ ,

$$\frac{d\lambda}{dx}(x) = c|x - x_0|^{\frac{2k+1}{2}}(1 + \mathcal{O}(x - x_0)), \quad x \rightarrow x_0 \text{ along } \text{supp } \lambda. \quad (2.20)$$

In a slight abuse of terminology, in (ii) above and in the following, we will usually speak about “endpoints of  $\text{supp } \lambda$ ” meaning endpoints of a connected component of  $\text{supp } \lambda$ .

The local behaviors (2.18)–(2.20) are believed to determine the local statistics of eigenvalues near  $x_0$ . The behaviors (2.18) and (2.20) are known to arise in the one matrix model (when  $\mathbf{A} = \mathbf{0}$ ) and their “complexified” versions, and have been vastly studied in the literature [45, 30, 31, 35, 32, 27, 50, 94, 47, 12, 62, 53, 29].

The behavior (2.19) with  $\nu = 1$  is known to give rise to the Pearcey point process under suitable scaling of the eigenvalues, and as mentioned before is known to occur for  $V(z) = z^2/2$  [35, 90]. More recently, this type of behavior appeared in Wigner-type random matrix models as well [1, 2, 57]. The case  $\nu = 5$  is implicitly present but not discussed in the literature: for the quartic symmetric potential  $V$  discussed in [31, Figure 7.1] and the unique bifurcation point in the parameter space (that is, the unique choice of parameters that belongs to the boundary of cases I, II and II), the discriminant of the spectral curve (see equation (7.1) therein) has only 2 simple zeros and a zero at the origin of multiplicity 10, which shows that the solution has the desired local asymptotics.

The appearance of such behaviors in the theorem above is not surprising, but the fact that these are the only possible behaviors, in particular the non-existence of a criticality associated to the vanishing of the density with order  $\nu/3$  and  $\nu \neq 1, 5$ , is the new contribution in Theorem B.

## 2.2. Asymptotic eigenvalue distribution for the Hermitian matrix model with external source.

We apply the results above to the Hermitian matrix model with external source when  $\mathbf{A}$  is of the form (1.2). So for Sections 2.2.1–2.2.2, to ensure the model (1.1) is well defined, we also tacitly assume that  $m = \deg V$  is even and  $v_m > 0$ .

### 2.2.1. The totally symmetric case and the constrained equilibrium.

As it was mentioned in the Introduction, when  $V$  is an even polynomial, Bleher, Delvaux and Kuijlaars [31] showed that the limiting eigenvalue distribution of the random matrix ensemble (1.1)–(1.2) in the asymptotic regime (1.3) with  $\alpha = 1/2$  is given by a certain probability measure  $\nu_1$  on  $\mathbb{R}$ , characterizing this measure in terms of a constrained vector equilibrium that we describe next.

We denote by  $\sigma$  the measure supported on  $i\mathbb{R}$ , absolutely continuous with respect to the arc-length and with constant density

$$\frac{d\sigma}{|dz|} = \frac{a}{\pi}, \quad z \in i\mathbb{R}. \quad (2.21)$$

We then define  $\mathcal{M}^c$  as the set of pairs of measures  $\vec{\nu} = (\nu_1, \nu_2)$  satisfying additionally

- Each component  $\nu_j$  is a Borel measure with finite logarithmic energy;

- $\text{supp } \nu_1 \subset \mathbb{R}$ ,  $\text{supp } \nu_2 = i\mathbb{R}$ ;
- $|\nu_1| = 1$ ,  $|\nu_2| = 1/2$ , and  $\nu_2 \leq \sigma$  (i.e.,  $\sigma - \nu_2$  is a non-negative Boreal measure on  $i\mathbb{R}$ ).

For any  $\vec{\nu} = (\nu_1, \nu_2) \in \mathcal{M}^c$ , we consider the energy

$$E^c(\vec{\nu}) = I(\nu_1, \nu_1) + I(\nu_2, \nu_2) - I(\nu_1, \nu_2) + \int (V(x) - a|x|)d\nu_1(x).$$

Standard arguments from logarithmic potential theory show that there is a unique minimizer  $\vec{\nu}^* = (\nu_1^*, \nu_2^*)$  of  $E^c$ , called the constrained *vector equilibrium measure* in the class  $\mathcal{M}^c$ . One of the results of [31] for the case when  $V$  is an even polynomial and  $\alpha = 1/2$  is that the limiting zero distribution for the average characteristic polynomial (2.4) is precisely the component  $\nu_1^*$  of  $\vec{\nu}^*$ . They also found the spectral curve, with  $\nu_1^* = \lambda$ , writing expressions for the coefficients  $p_0$  and  $p_1$  in (2.6) in terms of  $\vec{\nu}^*$ .

As for  $\nu_2^*$ , the inequality constraint  $\nu_2^* \leq \sigma$  can become an equality on a subinterval  $[-y_*, y_*]$  of  $i\mathbb{R}$ , which is known as the *saturated* case or regime for this random matrix model.

In the following theorem we give an alternative description of  $\vec{\nu}^*$ , now in terms of the vector critical measure  $\vec{\mu}^*$ , see Definition 2.3.

**Theorem C.** *For an even polynomial  $V$ ,  $\alpha = 1/2$  and  $a > 0$ , let (2.6) be the admissible spectral curve corresponding to the totally symmetric random matrix model described above. If  $\vec{\mu}_* = (\mu_1^*, \mu_2^*, \mu_3^*) \in \mathcal{M}_\alpha(\Gamma_*)$ ,  $\Gamma_* \in \mathcal{T}$ , is the corresponding vector critical measure as in Theorem A, then  $\text{supp}(\mu_2 + \mu_3)$  is contained in the closure of the left half plane, and for the constrained equilibrium measure  $\vec{\nu}_* = (\nu_1^*, \nu_2^*)$  it holds that*

$$\nu_1^* = \mu_1^* + \mu_2^*,$$

*while  $\nu_2^*$  is the balayage of  $\mu_2^* + \mu_3^*$  onto the right half plane.*

For the special case when  $V$  is quartic, Theorem C was proven (for the vector equilibrium measures) by Aptekarev, Lysov and Tulyakov in [10, Proposition 2.2].

*Remark 2.5.* The proof of this theorem is carried out in Section 4 by constructing a pair of measures  $\vec{\nu}_* = (\nu_1^*, \nu_2^*)$  from the critical measure  $\vec{\mu}_* = (\mu_1^*, \mu_2^*, \mu_3^*)$ . This construction is valid for *any* symmetric spectral curve (see Definition 4.1), and once this symmetric spectral curve is assumed to be the one in [31] this constructed pair  $\vec{\nu}_* = (\nu_1^*, \nu_2^*)$  immediately reduces to the constrained equilibrium pair.

Curiously, this construction could also be extended even for general  $V$  and  $\alpha \in (0, 1)$ , with the modification that  $\nu_2^*$  would now be supported on a different, but asymptotically vertical contour, and the upper constraint is no longer a positive (not even real-valued) measure. This will be detailed in a future work.

For the notion of balayage in the logarithmic potential theory, see e.g. [72] or [88].

### 2.2.2. The general case.

The previous results show that our vector critical measures give an alternative description of the asymptotic eigenvalue distribution of the random matrix ensemble (1.1)–(1.2) in the limit (1.3) with  $V$  even and  $\alpha = 1/2$ . In this section we claim that under certain additional assumptions this description in fact still holds for general  $V$  and  $\alpha$ .

It is convenient to introduce a multi-index notation. We denote  $\vec{n} = (n_1, n_2) \in \mathbb{N}^2$ , where for us  $0 \in \mathbb{N}$ , and make  $|\vec{n}| = n_1 + n_2$ . Also, we set  $\vec{e}_1 = (1, 0)$  and  $\vec{e}_2 = (0, 1)$ . We consider sequences of multi-indices  $(\vec{n}_k)$  with an up-right path structure, that is,

- $\vec{n}_k = (n_{k,1}, n_{k,2})$  with  $|\vec{n}_k| = k$ ,
  - for every  $k$ ,  $\vec{n}_{k+1} = \vec{n}_k + \vec{e}_j$  for some  $j \in \{1, 2\}$ ,
  - $\lim_{k \rightarrow \infty} \frac{n_{k,1}}{k} = \alpha \in (0, 1)$ , and
  - there exists  $d \geq 2$  for which  $n_{k+d,j} \geq n_{k,j} + 1$ , for  $j = 1, 2$  and every  $k$ .
- (2.22)

The first three conditions above are the usual ones for an up-right path. The last condition says that the path stays within a fixed distance from the ray with slope  $\alpha$ ; it will ensure that the sequence of multiple orthogonal polynomials indexed by  $(\vec{n}_k)$  satisfies a recurrence relation with finitely many terms. The same condition also appeared in [59], see Equation (3.2) therein and also our comments after Theorem E below.

A remarkable fact established in [34] is that for the random matrix ensemble (1.1)–(1.2), the average characteristic polynomial  $\pi_{n_1, n_2} = \pi_{\vec{n}}$ , defined in (2.4), satisfies multiple orthogonality conditions. This means that  $\pi_{\vec{n}} = P_{\vec{n}}^{(N)}$ ,  $N = |\vec{n}|$ , where  $(P_{\vec{k}}^{(N)})$  is the sequence of multiple orthogonal polynomials (MOP) with respect to the exponential weights  $e^{-NV_1}$  and  $e^{-NV_2}$ :

$$\begin{aligned} \int_{\mathbb{R}} P_{\vec{k}}^{(N)}(x) x^l e^{-NV_j(x)} dx &= 0, \quad l = 0, \dots, k_j - 1, \quad j = 1, 2, \\ P_{\vec{k}}^{(N)}(x) &= x^{|\vec{k}|} + \text{lower order terms;} \end{aligned} \quad (2.23)$$

we remind the reader that  $V_1$  and  $V_2$  were defined in (2.9).

A closely related quantity is the vector of wave functions

$$\Psi_{\vec{n}}(z) = \begin{pmatrix} \pi_{\vec{n}}(z) \\ P_{\vec{n}-\vec{e}_1}^{(N)}(z) \\ P_{\vec{n}-\vec{e}_2}^{(N)}(z) \end{pmatrix} e^{-NV(z)} = \begin{pmatrix} P_{\vec{n}}^{(N)}(z) \\ P_{\vec{n}-\vec{e}_1}^{(N)}(z) \\ P_{\vec{n}-\vec{e}_2}^{(N)}(z) \end{pmatrix} e^{-NV(z)}, \quad |\vec{n}| = N = n_1 + n_2. \quad (2.24)$$

**Theorem D.** *The vector of wave functions  $\Psi_{\vec{n}}$  satisfies the first-order differential equation*

$$\frac{d}{dz} \Psi_{\vec{n}}(z) = N \mathbf{R}_{\vec{n}}(z) \Psi_{\vec{n}}(z), \quad z \in \mathbb{C}, \quad (2.25)$$

where  $\mathbf{R}_{\vec{n}}$  is a matrix-valued polynomial (depending on the multi-index  $\vec{n}$ ) of the form

$$\mathbf{R}_{\vec{n}}(z) = \begin{pmatrix} -V'(z) + \mathcal{O}(z^{m-2}) & \mathcal{O}(z^{m-2}) & \mathcal{O}(z^{m-2}) \\ \mathcal{O}(z^{m-2}) & \mathcal{O}(z^{m-3}) & \mathcal{O}(z^{m-3}) \\ \mathcal{O}(z^{m-2}) & \mathcal{O}(z^{m-3}) & \mathcal{O}(z^{m-3}) \end{pmatrix}, \quad z \rightarrow \infty, \quad (2.26)$$

and  $V$  was defined in (2.5). Moreover, the characteristic polynomial of  $\mathbf{R}_{\vec{n}}$  takes the form

$$\det(\xi \mathbf{I} + \mathbf{R}_{\vec{n}}) = \xi^3 + q_{\vec{n}}^{(2)}(z)\xi^2 + q_{\vec{n}}^{(1)}(z)\xi + q_{\vec{n}}^{(0)}(z), \quad (2.27)$$

where the coefficients  $q_{\vec{n}}^{(j)} = p_j$ ,  $j = 0, 1, 2$ , satisfy the conditions (2.7) with  $\alpha = n_1/N$ .

*Remark 2.6.* In the expressions (2.26)–(2.27) all terms of the form  $\mathcal{O}(z^k)$  may depend on  $\vec{n}$ .

For the quadratic potential (shifted Gaussian ensemble) obtained when  $m = 2$ , most terms in the expansion (2.26) can actually be computed exactly (see (6.15) below) and give

$$\mathbf{R}_{\vec{n}}(z) = \begin{pmatrix} -V'(z) & 0 & 0 \\ \mathcal{O}(1) & -a_1 & 0 \\ \mathcal{O}(1) & 0 & -a_2 \end{pmatrix}$$

which is enough to determine the coefficients of  $\det(\xi \mathbf{I} + \mathbf{R}_{\vec{n}}(z))$  uniquely.

The main consequence of our findings for the model (1.1) is our next result. To state it, recall that for the model (1.1) we assume  $V$  to be a polynomial of even degree  $m \geq 2$ ,  $\mathbf{A}$  to be of the form (1.2) and by  $N \rightarrow \infty$  we mean under the regime (1.3).

**Theorem E.** *Suppose that  $(\vec{n}_N)$  is a sequence of multi-indices satisfying (2.22) with corresponding value  $d$ , and denote by  $\tilde{\pi}_{\vec{n}_N}$  the average characteristic polynomial (2.4) for the modified random matrix ensemble (1.1)–(1.2), where  $N$  in (1.1) is replaced by  $N - \max\{m - 2, d\}$ . If all the zeros of the sequence  $(\tilde{\pi}_{\vec{n}_N})$  are contained in a fixed compact set of  $\mathbb{C}$ , then all the zeros of  $(\pi_{\vec{n}_N})$  remain uniformly bounded as  $N \rightarrow \infty$ , and for any limiting zero distribution  $\lambda$  of a subsequence  $(\pi_{\vec{n}_k})$  of  $(\pi_{\vec{n}_N})$ ,*

$$\lambda = \lim_{k \rightarrow \infty} \mu(\pi_{\vec{n}_k}),$$

*the function*

$$\xi(z) = C^\lambda(z) + V'(z)$$

*is a solution of a cubic equation (2.6) that constitutes an admissible spectral curve in the sense of Definition 2.1. Furthermore, this cubic equation is obtained as a limit of the finite  $N$  spectral curve (2.6) along a subsequence of  $(\vec{n}_k)$ .*

With (2.23) in mind, the average characteristic polynomial  $\tilde{\pi}_N$  is actually the MOP  $\tilde{\pi}_N = P_{\vec{n}_N}^{(N-\kappa)}$ , with  $\kappa = \max\{m - 2, d\}$ . So the assumption of Theorem E is that the zeros of the sequence of MOP's  $(P_{\vec{n}_N+\kappa}^{(N)})$  slightly off the given up-right path  $(\vec{n}_N)$  remain bounded. Nevertheless, we opted for the formulation above as it is written in terms of the average characteristic polynomial, which has a more concrete physical meaning.

As shown in Proposition 5.2 below, the zeros of  $\pi_{n_N}$  are all real, so the measure  $\lambda$  in the Theorem above is always real as well. Furthermore, the boundedness of zeros of  $P_{\vec{n}_{N+d}}^{(N)}$  implies that the coefficients for the recurrence relation satisfied by the sequence  $(P_{\vec{n}_k}^{(N)})$  remain appropriately bounded as  $N \rightarrow \infty$ . As a consequence of Hardy's result [59, Corollary 1.5] along the same subsequence  $k$  of  $N$  any sequence of spectral measures

$$\nu_k = \frac{1}{k} \sum_{j=1}^k \delta_{\lambda_k}$$

for the (random) eigenvalues  $\lambda_1, \dots, \lambda_k$  of the model (1.1) converges almost surely to  $\lambda$  as well. So the spectral curve is, in fact, describing not only the limiting measure for the average characteristic polynomial, but also the density of states for the random eigenvalues, again along subsequences at least.

As it follows from Theorem E above, the characteristic equation for  $\mathbf{R}_{\vec{n}}$  in (2.26), obtained by equating the expression in (2.27) to 0, can be seen as a finite  $N$  version of the spectral curve. Similar versions of it (and also of the ODE (2.25)) for the case  $\mathbf{A} = \mathbf{0}$  and also for two-matrix models have been obtained in the past by Bertola, Eynard and Harnad [21, 22, 24, 25, 23].

*Remark 2.7.* As it was mentioned above, the uniform boundedness of the zeros of  $(\pi_{\vec{N}})$  is equivalent to the boundedness of the recurrence coefficients of the associated sequence of multiple orthogonal polynomials  $(P_{\vec{n}_N})$ , and thus, we work in a framework and under hypotheses analogous to those in recent developments on fluctuations of eigenvalues of large random matrices [41, 69, 59, 40, 60].

We strongly believe that under our assumptions on  $V$  and  $\mathbf{A}$  and the explicit scaling with  $N$  in (1.1) the zeros of  $(\pi_{\vec{N}})$  remain bounded, so the conclusions of Theorem E should always take place. In fact, boundedness was established already in the Gaussian,  $V(x) = x^2/2$ , and in the symmetric cases,  $V(-x) = V(x)$  and  $\alpha = 1/2$ , at least under regularity assumptions, see [31].

### 2.3. Overview of the paper.

The core of the paper is Section 3, culminating in the proof of Theorems A and B. This is achieved in several steps, as follows:

- Section 3.1 is a preliminary discussion on the construction of the Riemann surface  $\mathcal{R}$  associated to the spectral curve (2.6).
- Sections 3.2–3.5 contain our most novel contributions, namely the construction of a quadratic differential  $\varpi$  on  $\mathcal{R}$  and a detailed study of its trajectories, which will ultimately describe the supports of the measures in Theorem A.
- In Section 3.6 we explore the consequences of the structure of trajectories of  $\varpi$  for the local behavior of the density of  $\lambda$  for the spectral curve (2.6).
- In Section 3.7, we return to the Riemann surface  $\mathcal{R}$ , now constructing its sheet structure in a very explicit way, once again thanks to the study of the trajectories of  $\varpi$ .

- Section 3.8 combines the results of the previous sections, proving Theorems A and B.

In Section 4 we analyze the spectral curve under symmetry conditions of Section 2.2.1; at the end of this section Theorem C is proved.

In Section 5 we prove some more or less known facts about zeros and recurrence relations satisfied by the polynomials  $P_k^{(N)}$  defined by (2.23).

Section 6 is devoted to the proof of our last two main theorems. In Section 6.1 we use the Riemann-Hilbert problem characterizing the orthogonality (2.23) as an algebraic tool to prove Theorem D. The calculations to get (2.25) are somewhat standard, but we push them a little further to obtain also the properties of the matrix of coefficients  $\mathbf{R}_{\vec{n}}$ , as claimed in Theorem D.

Last but not least, Section 6.2 contains the proof of Theorem E. It involves an asymptotic analysis of the coefficients in the finite  $N$  spectral curve (2.27), based on the relation between  $\mathbf{R}_{\vec{n}}$  and the recurrence coefficients, relation which is implicitly given through the Riemann-Hilbert problem for both type I and II multiple orthogonal polynomials.

### 3. CONSTRUCTION OF THE VECTOR CRITICAL MEASURES

In this section, our starting point is an admissible spectral curve (2.6), see Definition 2.1. Our main goal is to prove Theorems A and B, which will require several steps.

#### 3.1. The Riemann Surface associated to the Spectral Curve.

For each  $z \in \mathbb{C}$  the equation (2.6) has three solutions (not necessarily distinct) that we denote by  $\xi_1(z)$ ,  $\xi_2(z)$  and  $\xi_3(z)$ , recalling that  $\xi_1(z)$  is the fixed solution (2.8).

**Proposition 3.1.** *The functions  $\xi_1, \xi_2$  and  $\xi_3$  admit an expansion of the form*

$$\begin{aligned}\xi_1(z) &= V'(z) - \frac{1}{z} + \mathcal{O}(z^{-2}), \\ \xi_2(z) &= a + \frac{\alpha}{z} + \mathcal{O}(z^{-2}), \\ \xi_3(z) &= -a + \frac{1-\alpha}{z} + \mathcal{O}(z^{-2}),\end{aligned}\quad \text{as } z \rightarrow \infty. \tag{3.1}$$

*In particular, these solutions are meromorphic at  $z = \infty$ , and the measure  $\lambda$  from (2.8) has compact support.*

*Proof.* The discriminant of  $F$  with respect to  $\xi$  is

$$\text{Discr}_\xi(F) = 4(V')^3 p_0 - 27p_0^2 - 18V'p_0p_1 + (V')^2 p_1^2 - 4p_1^3 = 4v_m^4 a^2 z^{4m-4} + \mathcal{O}(z^{4m-3}),$$

so in particular  $\text{Discr}_\xi(F) > 0$  for  $z \in \mathbb{R}$  with sufficiently large absolute value, implying that  $\xi_1, \xi_2$  and  $\xi_3$  are not branched near  $z = \infty$ . This means that  $\xi_1$  is meromorphic at  $z = \infty$ , so we also



get that  $\lambda$  has to be compactly supported. Furthermore, these solutions then admit a Laurent expansion around  $z = \infty$ , which in virtue of (2.8) take the form

$$\begin{aligned}\xi_1(z) &= V'(z) - \frac{1}{z} + \mathcal{O}(z^{-2}), \\ \xi_2(z) &= \kappa_2 z^{m_2} + \eta_2 z^{m_2-1} + \mathcal{O}(z^{m_2-2}), \\ \xi_3(z) &= \kappa_3 z^{m_3} + \eta_3 z^{m_3-1} + \mathcal{O}(z^{m_3-2}),\end{aligned}\quad \text{as } z \rightarrow \infty. \quad (3.2)$$

for some  $m_2, m_3 \in \mathbb{Z}$  and real numbers  $\kappa_2, \kappa_3, \eta_2, \eta_3$ , with  $\kappa_2, \kappa_3 \neq 0$  (so, at this point, the second and the third equation in (3.2) are just to set up the notation).

Being the solutions to the algebraic equation (2.6), these functions  $\xi_1, \xi_2$  and  $\xi_3$  satisfy the relations

$$\begin{cases} \xi_1(z) + \xi_2(z) + \xi_3(z) = V'(z), \\ \xi_1(z)\xi_2(z) + \xi_1(z)\xi_3(z) + \xi_2(z)\xi_3(z) = p_1(z), \\ \xi_1(z)\xi_2(z)\xi_3(z) = -p_0(z). \end{cases} \quad (3.3)$$

Using (3.2) to expand the LHS of the last equation in (3.3) and then comparing with the leading coefficient of the RHS, we arrive at the identities

$$m_2 = -m_3, \quad \kappa_2 \kappa_3 = -a^2.$$

Now, from the first equation in (3.3) we get that the polynomial part of  $\xi_2$  near  $z = \infty$  equals minus the polynomial part of  $\xi_3$ , and consequently we must have  $m_2 = m_3 = 0$  and also  $\{\kappa_2, \kappa_3\} = \{a, -a\}$ . Thus, imposing w.l.g. that  $\kappa_2 = a$ , we get  $\kappa_3 = -a$ , and a further use of the first equation in (3.3) also provides the relation

$$\eta_2 + \eta_3 = 1. \quad (3.4)$$

Using again the last equation in (3.3) and comparing with (3.2), we see that we should have

$$p_0(z) = a^2 v_m z^{m-1} + (a^2 v_{m-1} + a v_m (\eta_2 - \eta_3)) z^{m-2} + \mathcal{O}(z^{m-3}),$$

so comparing with (2.7) this means that

$$\eta_2 - \eta_3 = 2\alpha - 1.$$

Combining it with (3.4) we finally obtain  $\eta_2 = \alpha$ ,  $\eta_3 = 1 - \alpha$ , as claimed.  $\square$

Because  $C^\lambda$  satisfies an algebraic equation and  $\text{supp } \lambda \subset \mathbb{R}$ , we get (see for instance [28]) that  $\text{supp } \lambda$  is the union of a finite number of bounded intervals, namely

$$\text{supp } \lambda = \bigcup_{k=1}^l [a_k, b_k], \quad a_1 < b_1 < \dots < a_l < b_l. \quad (3.5)$$

Since  $F(\xi, z)$  in (2.6) is monic in  $\xi$ , we conclude that  $\xi_1, \xi_2$  and  $\xi_3$  can have poles only at  $\infty$ , and actually by (3.1), only  $\xi_1$  has a pole at  $\infty$ . Hence, the measure  $\lambda$  is absolutely continuous with respect to the Lebesgue measure on  $\mathbb{R}$  with a bounded density (the latter follows from Plemelj's formula applied to (2.8), see also (3.12) below).

We construct the associated Riemann surface  $\mathcal{R}$  to (2.6) as a three-sheeted branched cover of  $\overline{\mathbb{C}}$

$$\mathcal{R} = \mathcal{R}_1 \cup \mathcal{R}_2 \cup \mathcal{R}_3,$$

where  $\mathcal{R}_k$ 's are copies of  $\overline{\mathbb{C}}$  cut along certain arcs (the branch cuts), chosen in such a way that  $\xi_k$  is meromorphic on  $\mathcal{R}_k$ . The branch cut structure for  $\mathcal{R}_1$  is completely determined by (2.8), so that

$$\mathcal{R}_1 = \overline{\mathbb{C}} \setminus \text{supp } \lambda = \overline{\mathbb{C}} \setminus \bigcup_{k=1}^l [a_k, b_k]. \quad (3.6)$$

The Equation (2.6) may have nonreal branch points, and in virtue of (3.8) these are necessarily branch points of  $\xi_2$  and  $\xi_3$  but not of  $\xi_1$ . Because (2.6) is an equation with real coefficients for  $z$  real, these nonreal branch points come in complex conjugate pairs. We connect each such pair by a simple and bounded real-symmetric contour; these contours will play the role of nonreal branch cuts. There are finitely many such pairs, so we can make sure that these branch cuts are chosen in such a way that the upper and lower half planes minus these branch cuts is simply connected. This way, the branches of solutions  $\xi_2$  and  $\xi_3$  labeled as the series (3.1) admit analytic continuations to the upper and lower half planes, and we claim that these satisfy

$$\xi_j(\bar{z}) = \overline{\xi_j(z)}, \quad j = 1, 2, 3, \quad (3.7)$$

whenever  $z$  is nonreal and does not belong to one of the branch cuts. Indeed, the Equation (2.6) is cubic in  $\xi$ , has real coefficients and its discriminant is

$$\begin{aligned} & -27p_0(z)^2 - 4p_1(z)^3 + 18p_0(z)p_1(z)p_2(z) + p_1(z)^2 + p_2(z)^2 - 4p_0(z)p_2(z)^3 \\ & = 4v_m^4 a^2 z^{4m-4} + \mathcal{O}(z^{4m-5}), \quad z \rightarrow \infty, \end{aligned}$$

where for the identity we used (2.7). This shows that the discriminant is positive for  $z$  real and sufficiently large, thus proving (3.7) for  $z$  large, and by analytic continuation (3.7) follows.

We construct the sheets  $\mathcal{R}_2$  and  $\mathcal{R}_3$  using the analytic continuation just explained. Hence, these two sheets are connected through nonreal branch cuts (if any), and also along intervals of the real line. At this stage, the branch cuts on the real line either connect two of the sheets  $\mathcal{R}_1, \mathcal{R}_2$  and  $\mathcal{R}_3$ , or possibly the three of them.

By (2.8), we have

$$\xi_1(\bar{z}) = \overline{\xi_1(z)}, \quad z \in \mathbb{C} \setminus \text{supp } \lambda, \quad (3.8)$$

and in combination with (3.7) this yields

$$\xi_{j,+}(x) = \overline{\xi_{j,-}(x)}, \quad j = 1, 2, 3, \quad (3.9)$$

for any  $x \in \mathbb{R}$  and not on the nonreal branch cuts; for the same values of  $x$  these boundary values  $\xi_{j,\pm}$  are continuous as function on  $\mathbb{R}$ . Finally, at points of intersection of the nonreal branch cuts with the real line, Equation (3.9) is understood in the sense of lateral limits.

But we can say more:

**Lemma 3.2.** *Let  $p \in \mathbb{R}$  be a regular point.*

*If  $p \in \mathbb{R} \setminus \text{supp } \lambda$  then*

$$\xi_1(p) \in \mathbb{R} \quad \text{and} \quad \text{Im } \xi_2(p) = -\text{Im } \xi_3(p). \quad (3.10)$$

*If  $p \in \text{supp } \lambda$  then there exists a permutation  $j, k$  of  $\{2, 3\}$  (depending on  $p$ ) such that*

$$\xi_{1\pm}(p) = \overline{\xi_{1\mp}(p)} = \xi_{j\mp}(p) = \overline{\xi_{j\pm}(p)}, \quad \xi_{k\pm}(p) \in \mathbb{R}, \quad (3.11)$$

*and*

$$\lambda'(p) := \frac{d\lambda}{dx}(p) = \frac{1}{2\pi i}(\xi_{1+}(p) - \xi_{1-}(p)) = \frac{1}{\pi} \text{Im } \xi_{1+}(p). \quad (3.12)$$

*Moreover, at such  $p \in \text{supp } \lambda$ ,*

$$\frac{1}{2\pi i} [(\xi_2 - \xi_3)_+(p) - (\xi_2 - \xi_3)_-(p)] = \frac{1}{\pi} \text{Im } (\xi_2 - \xi_3)_+(p) = \begin{cases} -\lambda'(p), & \text{if in (3.11), } j = 2, \\ \lambda'(p), & \text{if in (3.11), } j = 3. \end{cases} \quad (3.13)$$

*Proof.* (3.10) as well as the identity  $\xi_{1\pm}(p) = \overline{\xi_{1\mp}(p)}$  for  $p \in \text{supp } \lambda$  are direct consequences of (3.8) and that (2.6) has real coefficients. We have observed already that  $\lambda$  is absolutely continuous with respect to the Lebesgue measure on  $\mathbb{R}$  with a bounded density  $d\lambda/dx$  that can be recovered via the Sokhotsky-Plemelj's formula, which yields (3.12).

Next we verify (3.11). Because in this case we are assuming  $p \in \text{supp } \lambda$  is a regular point (that is, not a branch point of (2.6)), we learn from the already proven Formula (3.12) that the boundary values  $\xi_{1+}(p)$  and  $\xi_{1-}(p)$  do not coincide, so  $\xi_{1+}(p)$  coincides with at least one among  $\xi_{2-}$  and  $\xi_{3-}$ , say  $\xi_{1+}(p) = \xi_{2-}(p)$  (the case when  $\xi_{1+}(p) = \xi_{3-}(p)$  being analogous). Combined with the symmetry (3.9) we obtain that  $\xi_{1-}(p) = \xi_{2+}(p)$ . Using again Formula (3.12) we see that  $\xi_{1+}$  and  $\xi_{2+}$  have nonzero imaginary parts, and because the Equation (2.6) is real and cubic, we have to have the other solution real, that is,  $\xi_{3+} \in \mathbb{R}$ . This completes the proof of (3.11).

To conclude, using once more the symmetry (3.9) we see that if in (3.11)  $j = 2$  then  $\xi_{3+}(p) = \xi_{3-}(p)$ , and  $\xi_{2+}(p) = \xi_{2-}(p)$  otherwise. Using this in (3.13) we see that the left hand side expresses  $-\lambda'(p)$  in the former case, and  $\lambda'(p)$  otherwise.  $\square$

The global solution  $\xi$  to (2.6) is constructed by taking

$$\xi|_{\mathcal{R}_j} = \xi_j, \quad j = 1, 2, 3. \quad (3.14)$$

We denote by  $\pi : \mathcal{R} \rightarrow \overline{\mathbb{C}}$  the canonical projection from  $\mathcal{R}$  to  $\overline{\mathbb{C}}$ . For a point  $q \in \overline{\mathbb{C}}$ , we denote its preimage by  $\pi$  on  $\mathcal{R}_j$  by  $q^{(j)}$ . That is,  $q^{(j)}$  is uniquely determined by

$$\mathcal{R}_j \cap \pi^{-1}(q) = \{q^{(j)}\}.$$

Obviously, if  $q$  is not a branch point, then the three points  $q^{(1)}$ ,  $q^{(2)}$  and  $q^{(3)}$  are distinct, but if  $q$  is a branch point connecting the sheets  $\mathcal{R}_j$  and  $\mathcal{R}_k$  then  $q^{(j)} = q^{(k)}$ . Moreover, if  $U \subset \overline{\mathbb{C}}$ , we denote

$$U^{(k)} = \pi^{-1}(U) \cap \mathcal{R}_k, \quad k = 1, 2, 3,$$

and if  $a, b$  are any points on  $\overline{\mathbb{C}}$ , we denote by  $[a, b]$  the non-oriented straight line segment with endpoints  $a$  and  $b$ , so that  $[a, b]$  also makes sense for  $a, b \in \mathbb{R}$ ,  $b < a$ , and in our notation the sets  $[a, b]$  and  $[b, a]$  are the same. We also set

$$[a^{(k)}, b^{(k)}] := [a, b]^{(k)}.$$

Similar notation is used also for open intervals. Finally, if  $V \subset \mathcal{R}_k$  for some  $k$ , then we denote by  $V^*$  its complex conjugate on the same sheet  $\mathcal{R}_k$ .

### 3.2. The canonical quadratic differential $\varpi$ .

Associated to the Riemann surface  $\mathcal{R}$  we define the canonical quadratic differential  $\varpi$  through

$$\varpi = \begin{cases} -(\xi_2(z) - \xi_3(z))^2 dz^2, & \text{on } \mathcal{R}_1, \\ -(\xi_3(z) - \xi_1(z))^2 dz^2, & \text{on } \mathcal{R}_2, \\ -(\xi_1(z) - \xi_2(z))^2 dz^2, & \text{on } \mathcal{R}_3. \end{cases} \quad (3.15)$$

In the earlier 2000's, some works in the literature explored quadratic differentials in the context of asymptotic analysis, but in all of them they only defined quadratic differentials on (subdomains of) the complex plane [13, 67, 74, 73, 17, 26]. To our knowledge, this explicit construction of  $\varpi$  on the associated Riemann surface  $\mathcal{R}$  as above in terms of solutions to an algebraic equation first appeared, independently and almost simultaneously, in our previous work [76] and in the paper [11] by Aptekarev and Tulyakov. Since then, it has found applications in different contexts, see also [36, 77, 4]. This quadratic differential will play a substantial role in what comes next. A general account on the theory of quadratic differentials can be found in the books by Jenkins [63], Pommerenke [81, Chapter 8] and Strebel [89]. The general theory and background reviewed in [76, Appendix B] are sufficient for our purposes; in fact, we closely follow the notation and notions introduced therein.

Recall that a zero/pole of a quadratic differential defined in local coordinates as  $f(z)dz^2$  is simply a zero/pole of the function  $f$ , its multiplicity coincides with that as a zero/pole of  $f(z)$ , and this definition is independent of the choice of the local parameter  $z$ . Zeros and poles of a quadratic differential are its *critical* points, all the rest are its *regular* points. By the definition of  $\varpi$ , the projection by  $\pi$  of the zeros of  $\varpi$  onto  $\mathbb{C}$  coincides with the zeros of the discriminant of (2.6). As a consequence, the poles and some of the zeros of  $\varpi$ , together with their multiplicities, can actually be easily described.

- $\varpi$  has a pole of order 4 at  $\infty^{(1)}$  and poles of order  $2m + 2$  at  $\infty^{(2)}$  and  $\infty^{(3)}$ , and no other poles on  $\mathcal{R}$ ;
- It has zeros at each of the points  $a_j^{(k)}, b_j^{(k)}$ ,  $k = 1, 2, 3$ .

The multiplicities of the poles  $\infty^{(1)}, \infty^{(2)}$  and  $\infty^{(3)}$  are obtained from the asymptotic expansion (3.1). The multiplicity of each of the zeros at the endpoints of  $\text{supp } \lambda$  is determined by the local behavior of the density of  $\lambda$  (see Lemma 3.2). For instance, if

$$\frac{d\lambda}{dz}(z) = \text{const } |z - q|^{1/2}(1 + \mathcal{O}(z)), \quad z \rightarrow q \in \{a_1, b_1, \dots, a_l, b_l\},$$

then  $\xi_1(q) = \xi_j(q)$  for some  $j \in \{2, 3\}$ . If in addition  $\xi_1(q) \neq \xi_k(q)$  for  $k \in \{2, 3\} \setminus \{j\}$ , then in such a case  $q^{(k)}$  is a zero of multiplicity 1 of  $\varpi$  and the branch point  $q^{(1)} = q^{(j)}$  is a zero of multiplicity 2. Points where the density of  $\lambda$  vanish with a higher order, or also higher order branch points, can be analyzed in a similar manner. We will perform such calculations when the need arises, and refer the reader to [76, Section 4.4] for an explicit calculation as well.

### 3.3. Trajectories and orthogonal trajectories of $\varpi$ .

Locally, any quadratic differential is the square of a meromorphic differential, so we can talk about integration of  $\sqrt{-\varpi}$  along curves on  $\mathcal{R}$ . A curve  $\gamma \subset \mathcal{R}$  that does not contain critical points (poles or zeros) of  $\varpi$  is called an *arc of vertical trajectory* (or simply arc of trajectory) of  $\varpi$  if

$$\operatorname{Re} \int_{\tilde{\gamma}} \sqrt{-\varpi} \equiv \text{const} \quad (3.16)$$

along any subarc  $\tilde{\gamma} \subset \gamma$ . Similarly,  $\gamma \subset \mathcal{R}$  is an *arc of orthogonal (horizontal) trajectory* of  $\varpi$  if

$$\operatorname{Im} \int_{\tilde{\gamma}} \sqrt{-\varpi} \equiv \text{const} \quad (3.17)$$

along any subarc  $\tilde{\gamma} \subset \gamma$ . Maximal arcs of trajectories are simply called trajectories. A trajectory is called *critical* if it has well-defined endpoints on its two directions, and at least one of these endpoints is a zero or a simple pole of  $\varpi$ . To stress when a trajectory is not critical, we call it *regular*. *Ditto* for orthogonal trajectories and critical orthogonal trajectories. The orthogonal trajectories of a quadratic differential  $\varpi$  are the same as the trajectories of  $-\varpi$ , so the qualitative behavior of trajectories and orthogonal trajectories is essentially the same.

The union of critical trajectories of  $\varpi$  is called its *critical graph*, and the union of critical orthogonal trajectories goes by the name of *critical orthogonal graph*. The structure of both graphs of a quadratic differential can be quite rich, both at local and global level, and has found important applications in several areas of mathematics; we refer the reader to the book by Jenkins [63, Theorem 3.5], or also to our previous work [76, Theorem B1] which rephrases the former in terms closer to the present work. It should be noted though that describing the structure of a critical (or critical orthogonal) graph in each concrete case can be quite challenging.

A straightforward consequence of the assumption (2.8) involving the real measure  $\lambda$  is the following proposition:

**Lemma 3.3.** *Any interval in  $\mathbb{R}^{(1)} \setminus (\operatorname{supp} \lambda)^{(1)}$  is a finite union of arcs of critical trajectories and orthogonal trajectories of  $\varpi$ .*

*Proof.* Fix a small interval  $[a^{(1)}, b^{(1)}] \subset \mathbb{R}^{(1)} \setminus (\operatorname{supp} \lambda)^{(1)}$  that does not contain zeros of  $\varpi$ . By (3.10),  $\xi_1$  is real-valued on  $[a^{(1)}, b^{(1)}]$  and  $\xi_2, \xi_3$  are complex conjugate of each other. Thus, we have two possibilities:

- a)  $\xi_2$  and  $\xi_3$  are real on  $[a, b]$ . In this case  $(\xi_2(x) - \xi_3(x))dx$  is real-valued along  $[a, b]$ , hence  $[a^{(1)}, b^{(1)}]$  is an arc of orthogonal trajectory.

- b)  $\xi_2$  and  $\xi_3$  are non-real, and by (3.10), they are complex conjugate of each other. In this situation,  $(\xi_2(x) - \xi_3(x))dx$  is purely imaginary along  $[a, b]$ , so  $[a^{(1)}, b^{(1)}]$  is an arc of trajectory.

This concludes the proof of the Lemma.  $\square$

*Remark 3.4.* Later on (see Corollary 3.11) we will see that in fact  $\mathbb{R}^{(1)} \setminus (\text{supp } \lambda)^{(1)}$  is made only of orthogonal trajectories of  $\varpi$ , i.e. it is a subset of the critical orthogonal graph of  $\varpi$ ; thus, only the possibility a) takes place.

In our work we are interested in the detailed structure of the critical and critical orthogonal graphs of  $\varpi$  for two reasons. First, as we will see from Section 3.8, the support of the third component  $\mu_3^*$  of the vector critical measure  $\vec{\mu}_*$  (see Theorem A) is subset of the critical graph of  $\varpi$ . Second, the construction of the sets  $\Gamma_*$  and  $\sigma_*$  underlying Theorems A and C uses both the critical and the critical orthogonal graphs of  $\varpi$ . Thus, our goal for the rest of this section is to describe the relevant critical trajectories. In order to do so, we will invoke some basic facts or “principles” that we list next. These principles are very similar in spirit as the ones used in [76, Section 4.5.1], and as we will briefly point out in a moment, they follow easily from basic properties of trajectories.

- P1.** If  $D \subset \mathcal{R}$  is a simply connected domain without poles of  $\varpi$ , then any trajectory of  $\varpi$  in  $D$  either ends at a zero of  $\varpi$  in  $D$  or intersects the boundary  $\partial D$ .
- P2.** A regular trajectory and a regular orthogonal trajectory intersect at most once. If they are critical, the same holds true in any simply connected domain without poles of  $\varpi$ , and this intersection has to happen at one of their endpoints.
- P3.** If  $D \subset \mathcal{R}$  is a simply connected domain, not necessarily bounded, not containing poles of  $\varpi$ , whose boundary  $\partial D$  is a finite union of (arcs of) trajectories, then  $\partial D$  contains at least one pole of  $\varpi$ , and at this point arcs of  $\partial D$  meet at a non-zero inner angle. Same conclusion is true if  $\partial D$  is a union of orthogonal trajectories.
- P4.** Orthogonal trajectories of  $\varpi$  are asymptotically horizontal at  $\infty^{(1)}$ . That is, if  $\gamma$  is an orthogonal trajectory that extends to  $\infty^{(1)}$ , then

$$\lim_{\substack{z \rightarrow \infty \\ z \in \gamma \cap \mathbb{C}_+^{(1)}}} \arg z = 0 \pmod{2\pi}, \quad \lim_{\substack{z \rightarrow \infty \\ z \in \gamma \cap \mathbb{C}_-^{(1)}}} \arg z = \pi \pmod{2\pi}.$$

- P5.** Trajectories of  $\varpi$  are asymptotically vertical at  $\infty^{(1)}$ . That is, if  $\tau$  is a trajectory that extends to  $\infty^{(1)}$ , then

$$\lim_{\substack{z \rightarrow \infty \\ z \in \tau \cap \mathbb{H}_\pm^{(1)}}} \arg z = \pm \frac{\pi}{2} \pmod{2\pi}.$$

Principle **P1** above is the same as [81, Lemma 8.4]. Principle **P2** follows from [89, Theorem 5.5] and [81, Theorem 8.3]. Principle **P3** is a consequence of Teichmüller’s Lemma [89, Theorem 14.1], and Principles **P4** and **P5** follow from the local structure of trajectories near poles of order 4 [89, Theorem 7.4].

### 3.4. Critical Orthogonal Graph of $\varpi$ .

A priori it looks like for a general polynomial potential  $V$  and for a parameter  $\alpha \in (0, 1)$  the structure of the critical orthogonal graph of  $\varpi$  can be arbitrarily complicated. Our goal in this section is to show that the natural assumptions we imposed on the spectral curve in Definition 2.1 have a crucial consequence: the quadratic differential  $\varpi$  has at most one zero, and of order at most 1, in  $\mathbb{H}_+^{(1)}$  (see Proposition 3.9 below). The existence or not of such a zero characterizes the two possible scenarios or regimes of the asymptotic eigenvalue distribution, determined by the existence or not of the non-trivial third component  $\mu_3^*$  of the critical measure, and consequently also determines the occurrence of the Pearcey type behavior (2.19).

The main goal of this section is to describe the structure of orthogonal trajectories of  $\varpi$  on  $\mathcal{R}_1$ . Due to the real symmetry, it is enough to focus our attention on the upper half-plane in  $\mathcal{R}_1$ .

Thus, consider the collection  $\mathcal{Y}$  of zeros of  $\varpi$  in  $\mathbb{H}_+^{(1)}$  and also the collection  $\mathcal{X}$  of zeros of  $\varpi$  on  $\mathbb{R}^{(1)}$ . These are disjoint and finite collections of points. Furthermore, by construction, branch points of  $\mathcal{R}_1$  necessarily belong to  $\mathcal{X}$ , see the discussion after (3.15).

Given a set  $\gamma$  on  $\mathcal{R}_1$ , we denote by  $\bar{\gamma}$  its closure; in the case of a smooth arc with finite length,  $\bar{\gamma}$  is this arc together with its endpoints. Let  $\Xi$  be the union of maximal arcs of orthogonal trajectories  $\gamma$  of  $\varpi$  on  $\mathbb{H}_+^{(1)}$  such that  $\bar{\gamma} \cap \mathcal{Y} \neq \emptyset$ . In other words, any critical orthogonal trajectory  $\gamma \subset \Xi$  emanates from a certain zero  $y^{(1)} \in \mathcal{Y}$  along  $\mathbb{H}_+^{(1)}$ , and (applying Principle **P1** to  $\mathbb{H}_+^{(1)}$ ) exactly one of the following possibilities holds:

- (1)  $\gamma$  extends to  $\infty^{(1)}$ .
- (2)  $\gamma$  ends up at another zero  $\tilde{y}^{(1)} \in \mathcal{Y}$ .
- (3)  $\gamma$  ends up at a zero  $x^{(1)} \in \mathcal{X}$ .
- (4)  $\gamma$  connects  $y^{(1)}$  to a point  $d^{(1)} \in \mathbb{R}^{(1)}$ , which is not a zero of  $\varpi$ .

Consider also the set  $\tilde{\mathcal{X}} = \mathcal{X} \setminus \bar{\Xi} \subset \mathcal{X}$  of points in  $\mathcal{X}$  that do not belong to the closure of any contour in  $\Xi$ , and let  $\tilde{\Xi}$  be the union of maximal arcs of orthogonal trajectories  $\gamma$  of  $\varpi$  on  $\mathbb{H}_+^{(1)}$  such that  $\bar{\gamma} \cap \tilde{\mathcal{X}} \neq \emptyset$ .

From any  $x^{(1)} \in \tilde{\mathcal{X}}$  emanate at least three arcs of critical orthogonal trajectories, so at least one of them, say  $\tau$ , enters  $\mathbb{H}_+^{(1)}$ . By the definition of  $\tilde{\mathcal{X}}$ ,  $\tau$  does not intersect zeros in  $\mathbb{H}_+^{(1)}$ , so applying the Principle **P1** again we know that exactly one of the following is true:

- (5)  $\tau$  connects  $x^{(1)}$  to another point  $d^{(1)} \in \mathbb{R}^{(1)}$ , possibly with  $d^{(1)} \in \mathcal{X}$  (we will see that this possibility contradicts Lemma 3.6, and thus ultimately does not hold), or
- (6)  $\tau$  extends to  $\infty^{(1)}$ .

Finally, define

$$\mathcal{J}^{(1)} := \Xi \cup \tilde{\Xi}.$$

**Lemma 3.5.** *The set  $\mathcal{J}^{(1)}$  is a graph with finitely many vertices (that contain  $\mathcal{X} \cup \mathcal{Y}$ ) and edges and with the following properties.*

- (i) *All its edges are (arcs of) critical orthogonal trajectories in  $\mathbb{H}_+^{(1)}$ .*
- (ii) *Every point in  $\mathcal{Y}$  is a vertex of  $\mathcal{J}^{(1)}$ , and its degree is equal to its multiplicity as a zero of  $\varpi$  plus two.*
- (iii) *There are edges emanating from every vertex of  $\mathcal{J}^{(1)}$ .*
- (iv) *Every vertex  $y^{(1)} \in \mathcal{Y}$  of  $\mathcal{J}^{(1)}$  is connected to  $\mathbb{R}^{(1)} \cup \{\infty^{(1)}\}$ .*
- (v) *The connected components of  $\mathbb{H}_+^{(1)} \setminus \mathcal{J}^{(1)}$  are simply connected. If, in addition, the boundary of a connected component does not intersect  $\mathbb{R}^{(1)}$ , then it is also unbounded.*
- (vi) *Two different sides of an edge belong to the boundaries of two distinct connected components of  $\mathbb{H}_+^{(1)} \setminus \mathcal{J}^{(1)}$ .*

*Proof.* The properties (i)–(iii) follow immediately from the construction of  $\mathcal{J}^{(1)}$ .

If (iv) were not true, then  $\mathcal{J}^{(1)}$  would contain a bounded closed cycle  $\gamma \subset \mathbb{H}_+^{(1)}$ , which is a finite union of critical orthogonal trajectories and with simply connected interior  $D$ . Because  $\varpi$  has no poles in  $\mathbb{H}_+^{(1)}$ , the simply connected domain  $D$  contains no poles on its closure, thus contradicting the Principle **P3**.

Let us prove (vi) before (v). We first look at simple paths (that is, without self-intersections) contained in  $\mathcal{J}^{(1)}$ . There is a partial order on simple paths in  $\mathcal{J}^{(1)}$ , induced by inclusion. Therefore we can talk about maximal simple paths, that is, simple paths that are not contained in any other simple path with more edges. The endpoints of maximal simple paths are degree one vertices of  $\mathcal{J}^{(1)}$ , and therefore have to belong to  $\mathbb{R}^{(1)} \cup \{\infty^{(1)}\}$ . This is so because any vertex in  $\mathbb{H}_+^{(1)}$  has to belong to  $\mathcal{Y}$ , and as such, is a zero of  $\varpi$ . The claim then follows once we recall that from any zero emanate at least three orthogonal trajectories, i.e. edges of  $\mathcal{J}^{(1)}$ .

Now, suppose that  $D$  is a connected component of  $\mathbb{H}_+^{(1)} \setminus \mathcal{J}^{(1)}$  whose boundary contains an edge  $\tau$  whose both sides are part of  $\partial D$ , and consider a maximal simple path  $\gamma$  containing  $\tau$ . By the previous observation, the endpoints of  $\gamma$  cannot belong to  $\mathbb{H}_+^{(1)}$ . This means that both endpoints of the simple path  $\gamma$  belong to  $\mathbb{R}^{(1)} \cup \{\infty^{(1)}\}$ . Thus,  $\gamma$  has to split  $\mathbb{H}_+^{(1)}$  into two disjoint components, say  $D_1$  and  $D_2$ , and any edge in  $\gamma$ , in particular  $\tau$ , separates  $D_1$  and  $D_2$ . Because  $D$  is connected we must have either  $D \cap D_1 = \emptyset$  or  $D \cap D_2 = \emptyset$ , and as such only one side of  $\tau$  is in the interior of  $D$ , a contradiction.

Finally, we now prove (v). If a connected component  $D$  of  $\mathbb{H}_+^{(1)} \setminus \mathcal{J}^{(1)}$  were not simply connected, then its boundary  $\partial D$  would have a component  $\gamma$  other than the outer boundary of  $D$ . There are two possibilities for  $\gamma$ : either it is a tree, or it contains a closed path  $\tilde{\gamma}$ . The first possibility cannot hold, as in this case it would have an edge that contradicts (vi). As for the second possibility, the interior of  $\tilde{\gamma}$  would then be a simply connected domain  $\tilde{D}$  without poles on its closure. As before,



the domain  $\tilde{D}$  contradicts the Principle **P3**. This last reasoning also applies to show that  $D$  has to be unbounded when its boundary does not intersect  $\mathbb{R}^{(1)}$ .

□

The proof of Lemma 3.5 only made use of general principles **P1–P5** stated above. The next properties do rely on the crucial assumption (2.8) from the Definition 2.1.

**Lemma 3.6.** *If  $\gamma \subset \mathbb{H}_+^{(1)}$  is a union of arcs of critical orthogonal trajectories then at most one endpoint of  $\gamma$  belongs to  $\mathbb{R}^{(1)}$ .*

*Proof.* We start observing that the claim of the lemma is equivalent to assuring that there are no paths  $\tau$  on the graph  $\mathcal{J}^{(1)}$  that intersect  $\mathbb{R}^{(1)}$  more than once. Assuming the contrary, we must conclude that there exists a connected component  $D$  of  $\mathbb{H}_+^{(1)} \setminus \mathcal{J}^{(1)}$  with boundary of the form  $\partial D = \gamma \cup [c^{(1)}, d^{(1)}]$ ,  $c < d$ , where

- (i)  $\gamma$  is a path in  $\mathcal{J}^{(1)}$  without self-intersections connecting  $c^{(1)}$  and  $d^{(1)}$ , obtained as the union of finitely many orthogonal trajectories,
- (ii)  $(c^{(1)}, d^{(1)}) \cap \mathcal{X} = \emptyset$ , i.e. there are no zeros of  $\varpi$  on  $(c^{(1)}, d^{(1)})$ .

First property is self-evident, while (ii) follows from Lemma 3.5. An immediate consequence of this is that no endpoint  $q$  of  $\text{supp } \lambda$  can lie on  $(c, d)$ : by construction,  $q^{(1)}$  is a branch point of  $\mathcal{R}_1$  and consequently a zero of  $\varpi$ . In other words, either  $(c, d) \cap \text{supp } \lambda = \emptyset$  or  $[c, d] \subset \text{supp } \lambda$ .

We will now show that such a domain  $D$  cannot exist.

If  $(c, d) \cap \text{supp } \lambda = \emptyset$ , then the interval  $[c, d]$  is not a (part of a) branch cut for  $\xi_1$ . Consider the closed path  $\gamma \cup \gamma^*$  on  $\mathcal{R}$  (where  $\gamma^*$  is the complex conjugate of  $\gamma$ ): it is the boundary of a bounded and simply connected domain in  $\mathcal{R}_1$ , and this contradicts the Principle **P3**.

Assume that

$$[c, d] \subset \text{supp } \lambda. \quad (3.18)$$

The open set  $D$  is simply connected and does not contain vertices of  $\mathcal{J}^{(1)}$ . consequently it does not contain zeros of  $\varpi$  (see Lemma 3.5) and as such also no branch points of  $\xi_2$  and  $\xi_3$ . Hence,

$$\oint_{\partial D} \sqrt{-\varpi} = \oint_{\partial D} (\xi_2(s) - \xi_3(s)) = 0.$$

We can then use (i) above to get

$$\int_{\gamma} \sqrt{-\varpi} \in \mathbb{R}.$$

Thus,

$$\int_c^d (\xi_{2+}(x) - \xi_{3+}(x)) dx \in \mathbb{R}. \quad (3.19)$$

However, by (3.13) and by our assumption about  $[c, d]$ ,  $\text{Im}(\xi_{2+}(x) - \xi_{3+}(x))$  preserves sign on  $[c, d]$ , in contradiction with (3.19). □

*Remark 3.7.* Lemma 3.6 deals only with arcs obtained as union of critical orthogonal trajectories. However, more is true: there cannot be an arc  $\gamma$  of orthogonal trajectory on  $\mathbb{H}_+^{(1)}$  intersecting  $\mathbb{R}^{(1)}$  more than once, regardless whether  $\gamma$  is critical or regular. The proof of this fact follows along exactly the same lines as in the proof above: one considers  $D = [c^{(1)}, d^{(1)}] \cup \tau$  a domain contained in the region bounded by  $\gamma \cap \mathbb{R}^{(1)}$ , constructed as in (i)–(ii) above, and the rest of the argument remains the same.

**Lemma 3.8.** *Given  $y^{(1)} \in \mathcal{Y}$ , there is exactly one path  $\tau \subset \mathcal{J}^{(1)}$  that connects  $y^{(1)}$  with  $\mathbb{R}^{(1)}$ .*

*Proof.* The uniqueness follows from Lemma 3.6.

Let us prove existence by contradiction, assuming that no such path exists. There are at least three distinct arcs of orthogonal trajectories emanating from  $y^{(1)}$ , that form part of three different paths, which by our assumption and by Lemma 3.5 (iv), connect  $y^{(1)}$  to  $\infty^{(1)}$ . From Principle **P5** we know that there are only two allowed directions for these paths to go to  $\infty^{(1)}$ , and consequently two of them, say  $\gamma_1$  and  $\gamma_2$ , have to extend to  $\infty^{(1)}$  along the same angle. Hence,  $\gamma_1 \cup \gamma_2$  is the boundary of a domain with inner angle 0 at  $\infty^{(1)}$  and no poles on its interior, which contradicts Principle **P3**, concluding the proof.  $\square$

One of the crucial conclusions of our analysis is the following proposition.

**Proposition 3.9.** *The quadratic differential  $\varpi$  has at most one zero in  $\mathbb{H}_+^{(1)}$ , and if it exists, it is simple.*

In our previous notation, this means that either  $\mathcal{Y} = \emptyset$ , or it consists of a single point, which is necessarily a simple zero of  $\varpi$ .

*Proof.* Let  $y^{(1)} \in \mathcal{Y}$ . We first show by contradiction that zeros with higher multiplicity are not allowed. Suppose that  $y^{(1)}$  is a zero of multiplicity at least 2. Consequently there exist four arcs of orthogonal trajectories emanating from  $y^{(1)}$ , and from Lemma 3.5 (iv) there correspond four paths of orthogonal trajectories  $\gamma_1, \gamma_2, \gamma_3$  and  $\gamma_4$  in  $\mathcal{J}^{(1)}$ , which are all pairwise distinct.

Any two distinct paths  $\gamma_j$  and  $\gamma_k$  intersect at  $y^{(1)}$  and at no other point on  $\mathbb{H}_+^{(1)}$ , otherwise  $\mathbb{H}_+^{(1)} \setminus \mathcal{J}^{(1)}$  would have bounded connected components, contradicting Lemma 3.5 (v). Also, at most one of them intersects  $\mathbb{R}^{(1)}$  (see Lemma 3.6). Recalling again Lemma 3.5 (iv), we conclude that there are at least three among these paths that go to  $\infty^{(1)}$ . Because of Principle **P4**, two of such arcs, say  $\gamma_j$  and  $\gamma_k$ , have to extend to  $\infty^{(1)}$  in the same asymptotic direction. But then  $\gamma_k \cup \gamma_j$  is the boundary of a simply connected domain with zero interior angle at  $\infty^{(1)}$  and no poles in its interior, contradicting Principle **P3**.

The analysis above, combined with Lemma 3.8, also gives the behavior of orthogonal trajectories emanating from a simple zero: there are three arcs of orthogonal trajectories, two of them extend to  $\infty^{(1)}$  horizontally (and with opposite angles) and the remaining one necessarily connects the zero to the real axis.

The case of two simple distinct zeros  $y^{(1)}, \tilde{y}^{(1)} \in \mathbb{H}_+^{(1)}$ , connected by a path  $\gamma$ , is topologically equivalent to a double zero, already discarded.

To conclude the proof, it remains to analyze the case of two simple distinct zeros  $y^{(1)}, \tilde{y}^{(1)} \in \mathbb{H}_+^{(1)}$ ,  $y^{(1)} \neq \tilde{y}^{(1)}$ , not connected by a path in  $\mathcal{J}^{(1)}$ .

The behavior of paths of orthogonal trajectories we just observed tells us that one of the three arcs of orthogonal trajectories emanating from  $y^{(1)}$  is part of a path connecting  $y^{(1)}$  to  $\mathbb{R}^{(1)}$ , and the other two arcs extend to  $\infty^{(1)}$  in the opposite asymptotic directions. Hence, the complement of these three paths in  $\mathbb{H}_+^{(1)}$  is the union of three disjoint domains. By assumption,  $\tilde{y}^{(1)}$  cannot belong to the boundary of these domains, so it belongs to one of these domains. But since the topology of the three arcs of orthogonal trajectories emanating from  $y^{(1)}$  and  $\tilde{y}^{(1)}$  is the same, a simple analysis shows that these two sets of paths of orthogonal trajectories necessarily intersect, which contradicts our assumption that  $y^{(1)}$  and  $\tilde{y}^{(1)}$  are not connected by a path in  $\mathcal{J}^{(1)}$ .

The proof is finally complete.  $\square$

**Proposition 3.10.** *If  $q^{(1)}$  is a regular point of  $\varpi$  on  $\mathbb{R}^{(1)} \setminus (\text{supp } \lambda)^{(1)}$ , then the arc of orthogonal trajectory that passes through  $q^{(1)}$  is a real interval. In consequence, each interval in  $(\mathbb{R} \setminus \text{supp } \lambda)^{(1)}$  that does not contain singular points of  $\varpi$ , is an arc of orthogonal trajectory of  $\varpi$ .*

*Proof.* Suppose that the orthogonal arc through  $q^{(1)}$ , say  $\sigma$ , is not an interval. Symmetry tells us that the orthogonal trajectory  $\sigma$  is symmetric with respect to complex conjugation. Furthermore, as a consequence of Remark 3.7, it stays within  $\mathcal{R}_1$  and either extends to  $\infty^{(1)}$  in its two ends, and in such a case necessarily horizontally and with the same angle in both ends, or it ends at the critical point  $y_*^{(1)}$ . Without loss of generality, we can assume the latter does not happen. Indeed, in a neighborhood of  $q^{(1)}$  on the real line, every point is regular. If  $\sigma$  happens to contain  $y_*^{(1)}$  then the same would not happen for the orthogonal arc emanating from any other point in this neighborhood, so we could take an arbitrarily small perturbation of  $q^{(1)}$  and proceed forward with the orthogonal trajectory emanating from the perturbed point.

Moving forward with the proof, pick a point  $p$ ,  $\text{Re } p \geq 0$ , such that  $p^{(1)} \in \sigma$ . Then also  $\bar{p}^{(1)} \in \sigma$ , and we can define

$$p \mapsto \int_{p^{(1)}}^{\bar{p}^{(1)}} (\xi_2(s) - \xi_3(s)) ds$$

where we integrate along  $\sigma$ . By the definition of orthogonal trajectory,  $(\xi_2(s) - \xi_3(s))ds$  is real along  $\sigma$ . Furthermore, by our previous discussion,  $\sigma$  does not contain critical points of  $\varpi$ , so  $(\xi_2(s) - \xi_3(s))ds$  also has constant sign along  $\sigma$ . Thus, the just defined function is real-valued, vanishing at  $p = q \in \mathbb{R}$ , and strictly monotone along  $\sigma \cap \mathbb{H}_+^{(1)}$ . In particular, there exists  $\delta > 0$  and  $M > 0$  such that

$$\left| \text{Re} \int_{p^{(1)}}^{\bar{p}^{(1)}} \sqrt{-\varpi} \right| = \left| \text{Re} \int_p^{\bar{p}} (\xi_2(s) - \xi_3(s)) ds \right| \geq \delta > 0, \quad \text{whenever } |p| \geq M. \quad (3.20)$$

On the other hand, now that we know that  $\varpi$  has exactly one zero on  $\mathbb{H}_+^{(1)}$  (and its conjugate on the lower half plane, these two particular points are the only possible branch points of  $\xi_2$  and  $\xi_3$  outside the real axis, so we can deform integrals and write

$$\int_p^{\bar{p}} (\xi_2(s) - \xi_3(s)) ds = \sum_{\gamma} \oint_{\tilde{\gamma}} (\xi_2(s) - \xi_3(s)) ds + \int_{\gamma} (\xi_2(s) - \xi_3(s)) ds$$

where  $\gamma$  is any other real-symmetric contour intersecting  $\mathbb{R}$  only once, at a point which is not a branch point of  $\xi_2$  nor  $\xi_3$ , and the  $\tilde{\gamma}$ 's are closed loops encircling pairs of branch points of  $\xi_2$  and  $\xi_3$  on  $\mathbb{R}$ , and possibly also  $\infty$ . We stress that this way we can make sure none of the  $\tilde{\gamma}$ 's encircle the non-real branch points of  $\xi_2$  and  $\xi_3$ .

Now, by (3.13) (and the asymptotics (3.1) in case we have to encircle  $\infty$ ) all integrals over the  $\tilde{\gamma}$ 's are purely imaginary, so that for sufficiently large  $M$  (such that  $\text{supp } \lambda \subset \{|z| < M\}$ ), the contour  $\gamma$  above can be replaced by the arc of a circle  $|s| = |p|$  in the right half plane that joins  $p$  and  $\bar{p}$ . Since the asymptotic expansion of  $\xi_2$  and  $\xi_3$  in (3.1) is uniform in a neighborhood of  $\infty$ , we can integrate it term-by-term and conclude that as  $|p| \rightarrow +\infty$ ,

$$\text{Re} \int_{p^{(1)}}^{\bar{p}^{(1)}} (\xi_2(s) - \xi_3(s)) ds = \text{Re} (2a(\bar{p} - p) + (2\alpha - 1) \log(p/\bar{p})) + \mathcal{O}(|p|^{-1}) = \mathcal{O}(|p|^{-1})$$

which contradicts (3.20) for  $|p|$  sufficiently large.

Finally, the last statement follows from the fact that every regular point  $q^{(1)} \in \mathbb{R}^{(1)} \setminus (\text{supp } \lambda)^{(1)}$  of  $\varpi$  belongs to an arc of an orthogonal trajectory.  $\square$

This result shows that in fact only one of the two possibilities mentioned in Lemma 3.3 holds.

**Corollary 3.11.** *Any gap  $(b_k^{(1)}, a_{k+1}^{(1)}) \subset \mathbb{R}^{(1)} \setminus (\text{supp } \lambda)^{(1)}$  is a union of orthogonal trajectories and only contains zeros of  $\varpi$  of even multiplicity.*

*Proof.* The fact that the whole interval  $(b_k^{(1)}, a_{k+1}^{(1)})$  is a union of orthogonal trajectories is contained in Proposition 3.10. Any zero  $p^{(1)}$  of  $\varpi$  in  $(b_k^{(1)}, a_{k+1}^{(1)})$  has to have even multiplicity, because the angle between the two orthogonal trajectories  $(p^{(1)}, (p + \varepsilon)^{(1)})$  and  $((p - \varepsilon)^{(1)}, p^{(1)})$  is  $\pi$ , which has to be an integer multiple of  $2\pi/(\text{multiplicity of } p^{(1)} + 2)$ .  $\square$

### 3.5. Saturated and unsaturated regimes.

As will be concluded later, the existence or not of the simple zero claimed by Proposition 3.9 determines whether the measure  $\mu_3^*$  in Theorem A is nontrivial or vanishes identically. In terms of the constrained equilibrium problem in Theorem C, it will determine whether the upper constraint for  $\nu_2^*$  is active or not. These facts and the traditional terminology explained in Section 2.2.1 motivate the next definition.

**Definition 3.12.** If  $\varpi$  does not vanish in  $\mathbb{H}_+^{(1)}$ , we say that we are in the *unsaturated regime*. Otherwise, when  $\varpi$  has a simple zero in  $\mathbb{H}_+^{(1)}$  that we denote by  $y_*^{(1)}$ , we say that we are in the *saturated regime*.

In this section, we collect some extra information on the zeros and trajectories of  $\varpi$  in  $\mathcal{R}_1$ . Later on, these results will help in the construction of a more appropriate and explicit branch cut structure for the remaining sheets  $\mathcal{R}_2$  and  $\mathcal{R}_3$  of  $\mathcal{R}$ , also providing a solid ground to the proof of our main results.

The general theory of quadratic differentials (see e.g. [63], [81, Chapter 8] or [89]) tells us that the complement of the critical (vertical or orthogonal) graph of a quadratic differential  $\varpi$  on a compact Riemann surface  $\mathcal{R}$  is a finite union of canonical domains. These are the so-called half-plane, strip, ring and circle domains, as well as density domains, filled with recurrent trajectories. We refrain from giving a systematic explanation of this theory and instead refer to the literature above or to our previous work [76, Appendix B]. Here we only describe the half-plane domains that we will need next, and whose existence is assured by this general theory.

Associated to the pole  $\infty^{(1)}$  of order 4, there are two half-plane domains bounded by orthogonal trajectories. In virtue of the symmetry of the problem, the projection of these half-plane domains to  $\mathbb{C}$  are related by complex conjugation, so we denote them by  $\mathcal{H}$  and  $\mathcal{H}^*$ . They enjoy the following properties:

$\mathcal{H}1)$  The boundary  $\partial\mathcal{H}$  is a finite union of critical orthogonal trajectories, contains the pole  $\infty^{(1)}$  and at least one zero of  $\varpi$ , but no other poles. Furthermore,  $\partial\mathcal{H}$  is a loop on  $\mathcal{R}$  that starts and ends at  $\infty^{(1)}$ , creating an inner angle  $\pi$  there, and it is asymptotically horizontal in its both ends, see Principle **P.4** in Section 3.3.

$\mathcal{H}2)$  For any fixed point  $p \in \partial\mathcal{H}$  and a choice of sign  $\epsilon \in \{1, -1\}$ , the canonical map

$$z \mapsto \epsilon \int_p^z \sqrt{-\varpi}$$

is a conformal map from  $\mathcal{H}$  to  $\mathbb{C}_+$ .

$\mathcal{H}3)$  The domain  $\mathcal{H}$  does not contain critical orthogonal trajectories of  $\varpi$ .

$\mathcal{H}4)$  Given any point  $p \in \mathcal{H}$  (necessarily regular), the orthogonal trajectory that passes through  $p$  is an analytic arc that extends to  $\infty^{(1)}$  along its two ends, with angles 0 and  $\pi$ , and it is entirely contained in  $\mathcal{H}$ .

By symmetry, only one among  $\mathcal{H}$  and  $\mathcal{H}^*$  contains points  $p \in \mathbb{H}_+^{(1)}$  with  $\pi(p)$  having positive and large imaginary part; we fix this one to be  $\mathcal{H}$ .

For each regime (saturated or not) our goal is:

- a) to describe the domain  $\mathcal{H}$ ;
- b) to construct an appropriate contour  $\Gamma_* \in \mathcal{T}$ , see Theorem A.

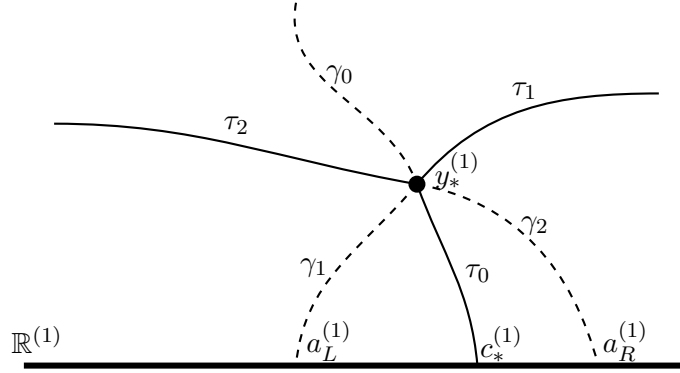


FIGURE 1. Saturated regime: the structure of trajectories (dashed lines, labeled as  $\gamma_0, \gamma_1, \gamma_2$ ) and orthogonal trajectories (solid lines, labeled as  $\tau_0, \tau_1, \tau_2$ ) emanating from  $y_*^{(1)}$  in  $\mathbb{H}_+^{(1)}$ . The trajectories on the lower half plane can be obtained by complex conjugation.

### 3.5.1. Saturated regime.

We start by analyzing the case of the saturated regime, i.e. of the existence of the simple zero  $y_*^{(1)}$  of  $\varpi$  in  $\mathbb{H}_+^{(1)}$ , see Definition 3.12.

**Proposition 3.13.** *In the saturated regime, the trajectories and orthogonal trajectories emanating from  $y_*^{(1)}$  are as follows:*

- (a) *Exactly one of the orthogonal trajectories from  $y_*^{(1)}$  intersects the real axis; we call the intersection point  $c_*$ . The remaining two trajectories extend to  $\infty^{(1)}$  in opposite asymptotically horizontal directions.*

*If  $c_* \in \mathbb{R} \setminus \text{supp } \lambda$  then  $c_*$  is a zero of  $\varpi$  of even multiplicity.*

- (b) *Exactly one of the trajectories from  $y_*^{(1)}$  extends to  $\infty^{(1)}$  with the asymptotic angle  $\pi/2$ . The remaining two trajectories intersect the real axis; we call the leftmost and the rightmost intersection points  $a_L$  and  $a_R$ , respectively.*

We have

$$a_L < c_* < a_R \tag{3.21}$$

and

$$(a_L, c_*) \cap \text{supp } \lambda \neq \emptyset, \quad (c_*, a_R) \cap \text{supp } \lambda \neq \emptyset. \tag{3.22}$$

The structure of trajectories emanating from  $y_*^{(1)}$  in  $\mathbb{H}_+^{(1)}$  are depicted on Figure 1.

*Proof.* The description of the orthogonal trajectories emanating from  $y_*^{(1)}$  follows exactly as in the proof of Proposition 3.9, we label these orthogonal trajectories by  $\tau_0, \tau_1$  and  $\tau_2$  as displayed in Figure 1.

Between any two consecutive orthogonal trajectories  $\tau_j$  and  $\tau_{j+1}$  emanates exactly one trajectory from  $y_*^{(1)}$ , say  $\gamma_{j+2}$  (with the convention that  $\gamma_{j+3} = \gamma_j$ ), see Figure 1. Observe also that  $\gamma_j$  is the analytic continuation of  $\tau_j$  through  $y_*^{(1)}$ . An application of Principle **P2** formulated in Section 3.3 tells us that  $\gamma_j$  has to remain in the domain bounded by  $\tau_j$  and  $\tau_{j+1}$ . Having in mind that trajectories cannot extend to  $\infty^{(1)}$  horizontally, we then apply the Principle **P1** to conclude that  $\gamma_1$  and  $\gamma_2$  connect  $y_*^{(1)}$  to the real line, whereas  $\gamma_0$  extends to  $\infty^{(1)}$  with angle  $\pi/2$ .

If  $c_* \in \mathbb{R} \setminus \text{supp } \lambda$  then from  $c_*$  emanate at least 4 orthogonal trajectories ( $\tau_0, \tau_0^*$ , as well as the real line, see Corollary 3.11), and hence  $c_*$  must be a zero of  $\varpi$  of even multiplicity.

We have (3.21) by construction. Furthermore,  $\gamma_1 \cup [a_L, c_*] \cup \tau_0$  encloses a bounded domain without branch points of  $\xi_2$  and  $\xi_3$  in its interior; if  $(a_L, c_*) \cap \text{supp } \lambda = \emptyset$  then  $[a_L, c_*] \cup \tau_0$  are orthogonal trajectories, while  $\gamma_1$  is a trajectory of  $\varpi$ , so that

$$\text{Im} \int_{\gamma_1} \sqrt{-\varpi} = \frac{1}{i} \int_{\gamma_1} \sqrt{-\varpi} \neq 0,$$

which leads us into a contradiction. The other assertion in (3.22) is proved in the same manner.  $\square$

**Corollary 3.14.** *Suppose we are in the saturated regime. From any zero  $p^{(1)} \in \mathbb{R}^{(1)}$  of  $\varpi$  emanates at most one orthogonal trajectory on  $\mathbb{H}_+^{(1)}$ . In particular, any such a zero  $p^{(1)}$  satisfies:*

- $p$  lies in the convex hull of the support of  $\lambda$ , and
- if  $p^{(1)} \in \mathbb{R}^{(1)} \setminus (\text{supp } \lambda)^{(1)}$ , then it is a zero of  $\varpi$  of order exactly 2.

*In particular, if  $c_* \notin \text{supp } \lambda$  then  $c_*^{(1)}$  is a zero of  $\varpi$  of order exactly 2.*

*Proof.* If  $p = c_*$ , defined in Proposition 3.13, then such an orthogonal trajectory is exactly one, that we denoted  $\tau_0$ . Indeed, any other orthogonal trajectory from  $c_*^{(1)}$  into  $\mathbb{H}_+^{(1)}$  must diverge to  $\infty$ , either in the direction of  $\tau_1$  or  $\tau_2$ , see Figure 2, which contradicts Principle **P3**.

Suppose that  $p^{(1)} \neq c_*^{(1)}$  is a zero of  $\varpi$  on  $\mathbb{R}^{(1)}$  from which emanate at least two orthogonal trajectories on  $\mathbb{H}_+^{(1)}$ , say  $\gamma_1$  and  $\gamma_2$ . Then  $\gamma_1$  and  $\gamma_2$  both have to stay within the closure of the same connected component of  $\mathbb{H}_+^{(1)} \setminus (\tau_0 \cup \overline{\mathcal{H}})$ . However, we know also that these two trajectories have to extend to  $\infty^{(1)}$  along different angles (see Lemma 3.6), which leads us into a contradiction (see Figure 2).

If, in addition,  $p^{(1)} \in \mathbb{R}^{(1)} \setminus (\text{supp } \lambda)^{(1)}$ , then the angle between each orthogonal trajectory emanating from  $p^{(1)}$  has to be  $2\pi/(\text{order of } p^{(1)})$ , and the order must be even as assured in Corollary 3.11. If this order is at least four then at least two such orthogonal trajectories have to emerge on  $\mathbb{H}_+^{(1)}$ , contradicting the first part. This shows that zeros on  $\mathbb{R}^{(1)} \setminus \text{supp } \lambda^{(1)}$  have to have order 2.

If  $p$  is not in the convex hull of  $\text{supp } \lambda$ , say  $p \in (-\infty, a_1)$ , then the orthogonal trajectory emerging from  $p^{(1)}$  vertically has to extend to  $\infty^{(1)}$  along the angle  $\pi$ , and the domain on  $\mathbb{H}^{(1)}$  bounded by

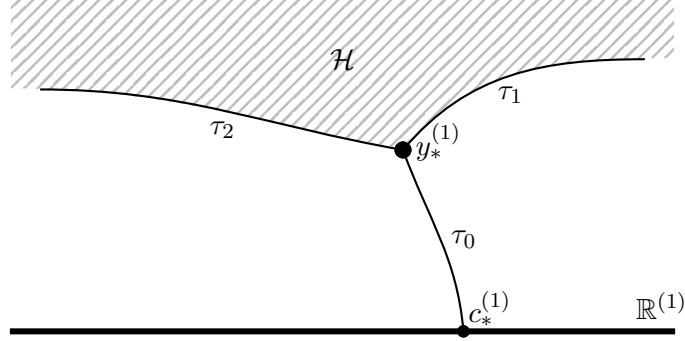


FIGURE 2. Saturated regime: the orthogonal trajectories emanating from  $y_*$  in  $\mathbb{H}_+^{(1)}$  and the associated orthogonal half-plane domain  $\mathcal{H}$  (shaded area). The point  $c_*$  is the intersection of  $\tau_0$  with  $\mathbb{R}^{(1)}$ . We dropped the superscript  $^{(1)}$  for ease of notation.

this trajectory and the orthogonal  $(-\infty^{(1)}, a_1^{(1)})$  violates Principle **P3**. This shows the first bullet of the Corollary, concluding the proof.  $\square$

In the saturated regime, the description of  $\mathcal{H}$  is an immediate consequence of Proposition 3.13 and it is displayed in Figure 2.

Now we describe the contour  $\Gamma_*$  used in the definition of the vector critical measures, see Section 2.1:

**Proposition 3.15.** *Suppose we are in the saturated regime. Define the contour  $\Gamma_*$  as*

$$\Gamma_* := \Delta_3 \cup \tau_*, \quad \tau_* := \pi(\tau_1 \cup \tau_1^*), \quad \Delta_3 := \pi((\gamma_1 \cup \gamma_1^*) \cap \mathcal{R}_1).$$

*Then  $\Gamma_*$  is piece-wise analytic, belongs to the class  $\mathcal{T}$  (see (2.11)) and satisfies*

$$\begin{aligned} (\xi_2(z) - \xi_3(z))dz &\in \mathbb{R} \quad \text{along } \tau_*, \\ (\xi_2(z) - \xi_3(z))dz &\in i\mathbb{R} \quad \text{along } \Delta_3. \end{aligned}$$

*Furthermore,  $\Gamma_*$  intersects  $\mathbb{R}$  at exactly one point that we denote by  $x_*$ .*

*Finally, on  $\Gamma_*$  the difference  $\xi_2 - \xi_3$  vanishes only at the points  $y_*$  and  $\overline{y_*}$ , where  $(\xi_2 - \xi_3)^2$  has a simple zero, as well as possibly at  $x_*$ .*

The contour  $\Gamma_*$  is displayed in Figure 3, and the proof of Proposition 3.15 is an immediate consequence of Proposition 3.13 and Corollary 3.14. Notice that  $x_*$  is the point that was denoted initially by  $a_L$  in Proposition 3.13, see also Figure 1.

### 3.5.2. Unsaturated regime.

The construction of  $\Gamma_*$  in Proposition 3.15 is only meaningful in the saturated regime. Now we address the unsaturated regime, for which we again start by describing the structure of the half-plane domain  $\mathcal{H}$ . Recall that now all the zeros of  $\varpi$  on the first sheet  $\mathcal{R}_1$  are actually on the real line  $\mathbb{R}^{(1)}$ .



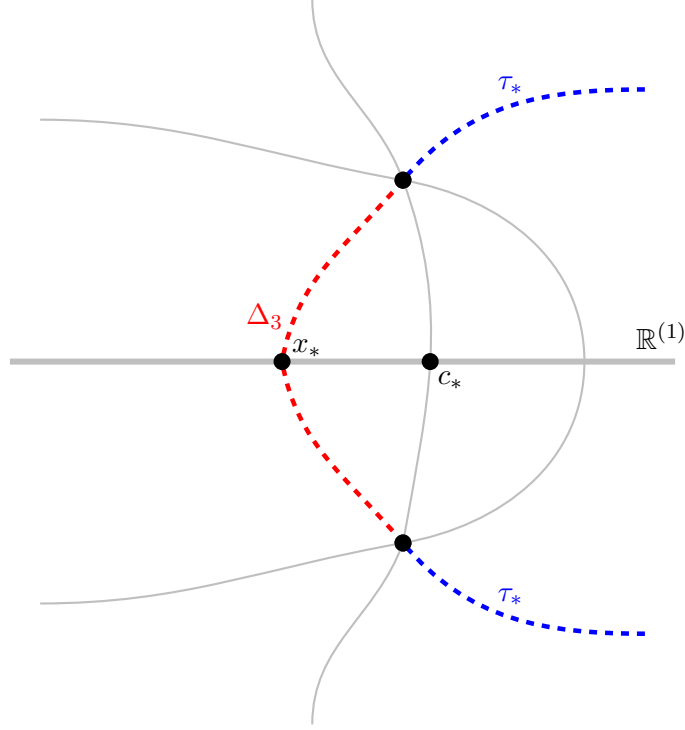


FIGURE 3. Saturated regime: contour  $\Gamma_* = \tau_* \cup \Delta_3$  (dashed), as described in Proposition 3.15. For convenience, we kept the remaining trajectories from  $y_*^{(1)}$  and  $\bar{y}_*^{(1)}$  in light gray, compare with Figure 1.

**Proposition 3.16.** *Suppose we are in the unsaturated regime. Then exactly one of the following has to occur.*

- (I)  $\partial\mathcal{H} \cap \mathbb{R}^{(1)}$  consists of exactly one point  $c_*^{(1)}$ , which is not an endpoint of  $\text{supp } \lambda^{(1)}$  and it is necessarily a zero of  $\varpi$ . In this case  $\partial\mathcal{H}$  is a union of two critical orthogonal trajectories  $\tau_1$  and  $\tau_2$  intersecting at  $c_*^{(1)}$  and extending to  $\infty^{(1)}$  with angles 0 and  $\pi$ , respectively. And finally,  $c_* \in (a_1, b_l) = \text{conv}(\text{supp } \lambda)$  and both  $\tau_1$  and  $\tau_2$  are analytic arcs contained in  $\mathbb{H}_+^{(1)}$  (except, obviously, their endpoint  $c_*^{(1)}$ ), see Figure 4. In particular, in this case  $\mathcal{H} \subset \mathbb{H}_+^{(1)}$ .
- (II)  $\partial\mathcal{H}$  contains exactly two distinct zeros of  $\varpi$ , say  $c_1^{(1)}, c_2^{(1)} \in \mathbb{R}^{(1)}$ ,  $c_2 < c_1$ , see Figure 5. Then  $\partial\mathcal{H}$  is a union of two critical orthogonal trajectories  $\tau_1$  and  $\tau_2$  on  $\mathbb{H}_+^{(1)}$ , with behavior as in (I), and the real interval  $(c_2^{(1)}, c_1^{(1)})$ . The points  $c_1^{(1)}$  and  $c_2^{(1)}$  are the unique zeros of  $\varpi$  on  $\partial\mathcal{H}$ , and

$$(c_2, c_1) \subset \text{conv}(\text{supp } \lambda) \setminus \text{supp } \lambda.$$

In particular, in this case  $\mathcal{H} \subset \mathbb{H}_+^{(1)}$  as well.

- (III)  $\partial\mathcal{H}$  contains an unbounded interval. In this case, there are two further possibilities:

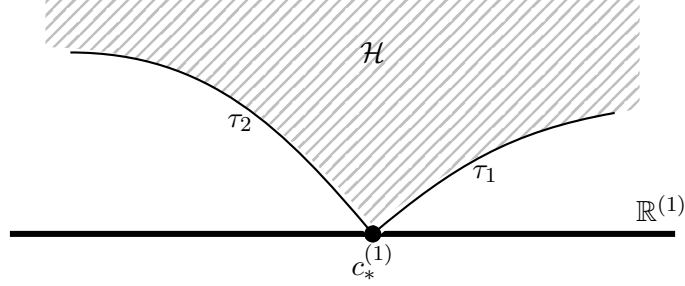


FIGURE 4. Unsaturated regime: the orthogonal trajectories emanating from  $x_*^{(1)}$  (solid lines) associated orthogonal half-plane domain  $\mathcal{H}$  (shaded area) in the situation (I) of Proposition 3.16.

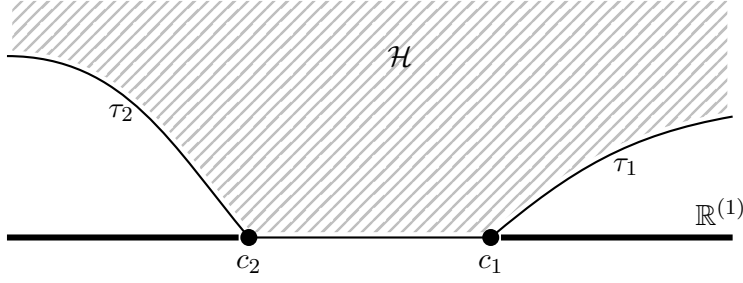


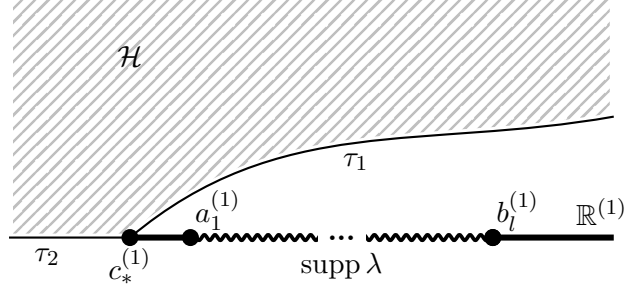
FIGURE 5. Unsaturated regime: the orthogonal half-plane domain  $\mathcal{H}$ , which is the shaded region with boundary  $\tau_1 \cup [c_1, c_2] \cup \tau_2$  as described by the situation (II) of Proposition 3.16. We emphasize that in this case  $[c_2, c_1]$  does not intersect  $\text{supp } \lambda$ .

- (a) This interval is of the form  $\tau_2 = (-\infty^{(1)}, c_*^{(1)})$  with  $c_* \leq a_1$ , and  $\partial\mathcal{H} = \tau_1 \cup \tau_2$ , with  $\tau_1$  is as in the case (I), see Figure 6 .
- (b) This interval is of the form  $\tau_1 = (c_*^{(1)}, \infty^{(1)})$  with  $c_* \geq b_l$ ,  $\partial\mathcal{H} = \tau_1 \cup \tau_2$ , with  $\tau_2$  is as in the case (I), see Figure 6 .

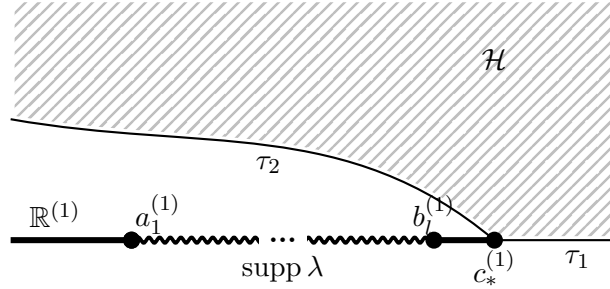
We will see below in Proposition 3.29 that, in fact, the situation (III) would correspond to a reducible spectral curve, happening only when  $a = 0$ . So under our conditions, the situation (III) is irrelevant.

*Proof.* Assume first that  $\partial\mathcal{H}$  does not contain unbounded intervals of  $\mathbb{R}^{(1)}$ . We will show that under this situation either (I) or (II) occurs.

Our initial assumption means that the boundary of  $\mathcal{H}$  near infinity consists of critical orthogonal trajectories  $\tau_1$  and  $\tau_2$  on  $\mathbb{H}_+^{(1)}$  emerging from  $\infty^{(1)}$  at angles 0 and  $\pi$  respectively, and those trajectories are not real intervals. Because  $\partial\mathcal{H}$  has to contain at least one zero (see property  $\mathcal{H}1$ ) above), and in the unsaturated regime there are no zeros on  $\mathbb{H}_+^{(1)}$ , the orthogonal trajectories  $\tau_1$  and  $\tau_2$



Situation (III)-(a)



Situation (III)-(b)

FIGURE 6. Unsaturated regime: the orthogonal half-plane domain  $\mathcal{H}$  and the orthogonal critical trajectories  $\tau_1$  and  $\tau_2$ , in the situations (III)-(a) and (III)-(b) described in Proposition 3.16. The wiggled segments represent the bounded interval of the real axis containing  $\text{supp } \lambda$ .

have to meet the real axis. Denote by  $c_1^{(1)}$  and  $c_2^{(1)}$  the first points of intersection of  $\tau_1$  and  $\tau_2$  with  $\mathbb{R}^{(1)}$ , respectively, and by  $\tilde{\tau}_1$  and  $\tilde{\tau}_2$  the subarcs of  $\tau_1$  and  $\tau_2$  that go from  $\infty^{(1)}$  to  $c_1^{(1)}$  and  $c_2^{(1)}$ .

If  $c_1 = c_2 =: c_*$ , then we are immediately in case (I), as long as we also verify that  $c_* \in (a_1, b_l)$ . To see that the former is indeed true, suppose to the contrary that  $c_*$  belongs, say, to  $(-\infty, a_1]$ , the case when  $c_* \in [b_l, \infty)$  is analogous. Because the interval  $(-\infty^{(1)}, a_1^{(1)})$  is also a union of arcs of orthogonal trajectories (see Proposition 3.10), then the domain on  $\mathcal{H}^{(1)}$  bounded by  $(-\infty^{(1)}, a_1^{(1)})$  and  $\tau_2$  violates the Principle **P3**.

Suppose now that  $c_1 \neq c_2$ , so that  $\mathcal{H} \cap \mathbb{H}_+^{(1)}$  is as displayed in Figure 7. The argument just presented above excludes  $c_j \in (-\infty, a_1]$  or  $c_j \in [b_l, \infty)$ . We will now show that  $(c_2, c_1) \subset \mathbb{R} \setminus \text{supp } \lambda$ , so that (II) is taking place.

As a consequence of Lemma 3.6, we obtain that in this situation  $\partial\mathcal{H} \cap \mathbb{H}_+^{(1)} = \tilde{\tau}_1 \cup \tilde{\tau}_2$ . Indeed, if there were a point in the intersection  $\partial\mathcal{H} \cap \mathbb{H}_+^{(1)} \notin \tilde{\tau}_1 \cup \tilde{\tau}_2$ , then there would be an arc  $\tilde{\tau}$  in  $\partial\mathcal{H}$  that does not intersect  $\tilde{\tau}_1$  and  $\tilde{\tau}_2$ . Having in mind that the whole boundary  $\partial\mathcal{H}$  is a connected and

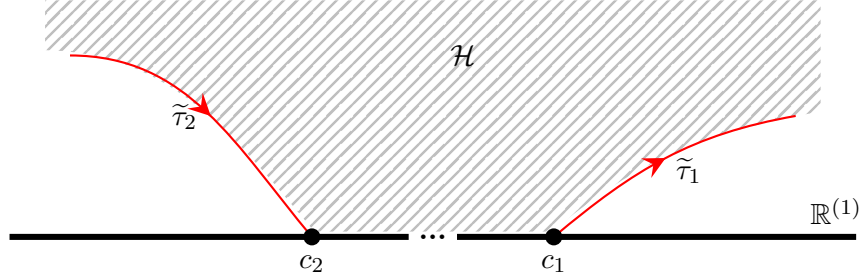


FIGURE 7. The structure of  $\mathcal{H}$  on  $\mathbb{H}_+^{(1)}$  used in the proof of Proposition 3.16, in the situation when  $c_1 \neq c_2$ .

piecewise analytic curve, this would mean that  $\tilde{\tau}$  would have to extend to the real line on its two ends, contradicting Lemma 3.6.

The orthogonal trajectories on a half-plane domain are analytic loops starting and ending at  $\infty^{(1)}$  that provide a foliation of this domain. If we assume that

$$(c_2, c_1) \cap \text{supp } \lambda \neq \emptyset,$$

then there must exist orthogonal trajectories inside  $\mathcal{H}$  that move to another sheet across one of the cuts  $[a_j, b_j]$ , as well as trajectories on the upper half plane that stay completely on  $\mathbb{H}_+^{(1)}$ , check the dashed lines in Figure 8.

Let us orient the orthogonal trajectories on  $\mathcal{H}$  from  $-\infty$  to  $+\infty$  and travel along such curves with this orientation. The sketched behavior of trajectories implies in particular that we can choose some orthogonal trajectory  $\gamma$  that leaves  $\mathbb{H}_+^{(1)}$  only for an arbitrarily small and short period, in the sense that the length of its subarcs that are not in  $\mathbb{H}_+^{(1)}$  can be assumed to be arbitrarily small. This means that this  $\gamma$  can be written in the form  $\gamma = \gamma_2 \cup \tilde{\gamma} \cup \gamma_1$ , where  $\tilde{\gamma}$  is the piece of  $\gamma$  not in  $\mathbb{H}_+^{(1)}$ . As it was mentioned, it can be assumed that  $\tilde{\gamma}$  has arbitrarily small length, and when we walk along  $\gamma$  we observe the following properties:

- Starting from  $\infty^{(1)}$  with an angle  $\pi$ , we meet the real axis at a first point  $p^{(1)}$ . This determines  $\gamma_2$ .
- We then move to another sheet, so  $p^{(1)} \in [a_j^{(1)}, b_j^{(1)}]$  for some  $j$ . Because  $\gamma$  can be chosen to make the length of  $\tilde{\gamma}$  as small as needed, we return to  $\mathbb{H}_+^{(1)}$  at a point  $q^{(1)} \in [a_j^{(1)}, b_j^{(1)}]$  for the same  $j$ . The walk from  $p^{(1)}$  to  $q^{(1)}$  determines  $\tilde{\gamma}$ .
- From  $q^{(1)}$ , we move to  $\infty^{(1)}$  with angle 0, determining then  $\gamma_1$  and finishing our walk along  $\gamma$ .

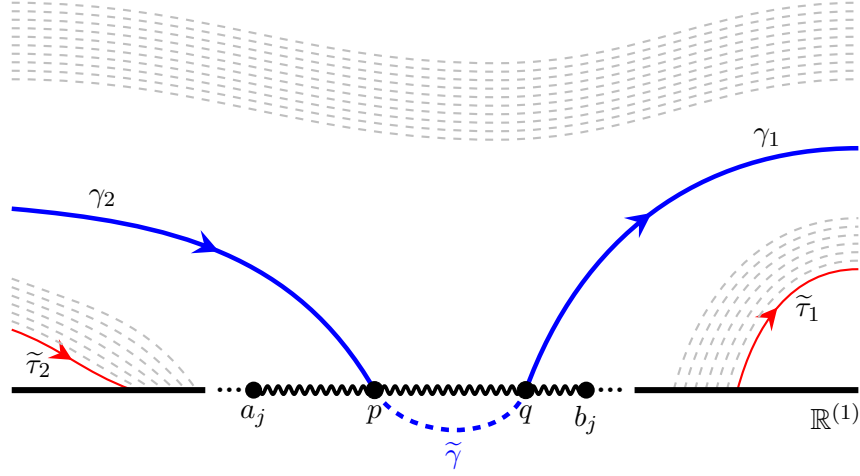


FIGURE 8. Part of the orthogonal half plane  $\mathcal{H}$  on  $\mathbb{H}_+^{(1)}$ , along with its boundary curves  $\tilde{\gamma}_1$  and  $\tilde{\gamma}_2$  (in solid gray), and part of the foliation by its orthogonal trajectories (dashed gray). On  $\mathbb{R}$ , the wiggled line represents the interval  $[a_j^{(1)}, b_j^{(1)}]$  on which the orthogonal trajectory  $\gamma = \gamma_1 \cup \tilde{\gamma} \cup \gamma_3$  intersects  $\mathbb{R}^{(1)}$ . The subarc  $\tilde{\gamma}$  (thick dashed) is the part of  $\gamma$  which is on another sheet (when labeling the points, we have dropped the upper index  $^{(1)}$  for convenience).

Such a trajectory  $\gamma$  is displayed in Figure 8. From the definition of an orthogonal trajectory, we know that

$$\operatorname{Im} \int_{\tilde{\gamma}} \sqrt{-\varpi} = 0.$$

Because  $\tilde{\gamma}$  can be chosen arbitrarily small, we can actually deform the path  $\tilde{\gamma}$  in the integration above without crossing any critical points of  $\varpi$  nor branch cuts of  $\mathcal{R}$  other than  $[a_j^{(1)}, b_j^{(1)}]$ , and then conclude that

$$\int_p^q (\xi_{2+}(x) - \xi_{3+}(x)) dx \in \mathbb{R},$$

which is in contradiction with (3.13). Hence,  $(c_2, c_1) \cap \operatorname{supp} \lambda = \emptyset$ . Since  $\partial\mathcal{H}$  has to be connected, there is a bounded arc  $\tilde{\gamma} \subset \partial\mathcal{H}$  connecting  $c_1^{(1)}$  and  $c_2^{(1)}$  and, as a consequence of Lemma 3.6,  $\tilde{\gamma} = [c_2^{(1)}, c_1^{(1)}]$ , thus concluding that (II) takes place, as explained above. The fact that the interval  $[c_2^{(1)}, c_1^{(1)}]$  does not contain other zeros is true because, if there were any zero there, then there would necessarily be an orthogonal trajectory emerging from it on the upper half place, contradicting the property  $\mathcal{H}3$ ) above.

In summary, assuming that  $\partial\mathcal{H}$  does not have unbounded real intervals we concluded that either  $c_1 = c_2 (= x_*)$ , corresponding to the case (I), or  $c_2 < c_1$ , corresponding to case (II)

Finally, suppose now that  $\partial\mathcal{H}$  does contain an unbounded real interval. Arguments similar to the ones carried out above in the case  $c_1 \neq c_2$  also exclude the following situations:

- (1) Both intervals  $(-\infty^{(1)}, a_1^{(1)})$  and  $(b_l^{(1)}, \infty^{(1)})$  belong to  $\partial\mathcal{H}$ .
- (2) The interval  $(-\infty^{(1)}, a_1^{(1)})$  is contained in  $\partial\mathcal{H}$ , and the remaining arc from  $\partial\mathcal{H}$  emerges from  $a_1^{(1)}$  in another sheet. Same goes for  $(b_l^{(1)}, \infty^{(1)})$ .

To conclude the proof, let us assume that  $\partial\mathcal{H}$  contains a maximal subinterval  $(-\infty^{(1)}, c_*^{(1)})$  of  $(-\infty^{(1)}, a_1^{(1)})$ . As we are assuming this interval is maximal, if  $c_* < a_1$  there should be an arc of orthogonal trajectory  $\tau_1 \subset \partial\mathcal{H}$  emerging from  $c_*$  on  $\mathbb{H}_+^{(1)}$ , whereas if  $c_* = a_1$  then such  $\tau_1$  also has to exist but now by (2) above. Either way, by Lemma 3.6 this arc  $\tau_1$  cannot return to  $\mathbb{R}^{(1)}$  anymore, so it has to extend to  $\infty^{(1)}$  (recall that we are under the unsaturated regime, so no zeros on  $\mathbb{H}_+^{(1)}$ ), and thus  $\partial\mathcal{H} = (-\infty^{(1)}, c_*^{(1)}] \cup \tau_1 = \tau_2 \cup \tau_1$  as claimed by situation (III)–(a).

The case that  $\partial\mathcal{H}$  contains a subinterval of  $(b_l^{(1)}, \infty^{(1)})$  implies (III)–(b) follows along exactly the same lines, we skip the details.  $\square$

The following is an analogue of Corollary 3.14, now in the unsaturated case:

**Corollary 3.17.** *Suppose we are in the unsaturated regime. If  $p^{(1)} \in \mathbb{R}^{(1)}$  is any zero of  $\varpi$ , different from the point  $c_*^{(1)}$  of the case (I) of Proposition 3.16, then from  $p^{(1)}$  emanates at most one orthogonal trajectory on  $\mathbb{H}_+^{(1)}$ . In particular, any such zero on  $\mathbb{R}^{(1)} \setminus (\text{supp } \lambda)^{(1)}$  must be of order exactly 2.*

*Proof.* A simple inspection in the description of  $\partial\mathcal{H}$  shows that other than  $c_*$ , from no other zero on the boundary emanates two trajectories. Thus, suppose that  $p^{(1)}$  is a zero on  $\mathbb{R}^{(1)} \setminus \partial\mathcal{H}$  from which emanates at least two orthogonal trajectories, say  $\gamma_1$  and  $\gamma_2$ .

Without loss of generality, assume  $p > c_1$  (in the case (II)) or  $p > c_*$  (in the case (I)), the other cases follow analogously. This means that we are in the situations covered in one of Figures 4 or 5, and as a consequence of Lemma 3.6 we see that  $\gamma_1$  and  $\gamma_2$  have to extend to  $\infty^{(1)}$  along the same connected component of  $\mathbb{H}_+^{(1)} \setminus \overline{\mathcal{H}}$  (this is the white region between  $\partial\mathcal{H}$  and  $\mathbb{R}^{(1)}$  in Figures 4, 5 or 6). In such a way, the domain delimited by  $\gamma_1 \cup \gamma_2$  then violates Principle **P3**.

In the case (II), if  $p \in [c_2, c_1]$  we get the same contradiction with Principle **P3** by considering now the domain delimited by, for instance,  $\gamma_1 \cup [p, c_1] \cup \tau_1$ .

The assertion about the order of  $p^{(1)}$  as a zero of  $\varpi$  is proved exactly as in Corollary 3.14.  $\square$

We are finally ready to construct the contour  $\Gamma_* \in \mathcal{T}$  for the unsaturated regime.

**Proposition 3.18.** *Suppose we are in the unsaturated regime, in one of the cases (I) or (II) from Proposition 3.16. Then there exists a contour  $\Gamma_* \in \mathcal{T}$  that intersects the real axis at a single point, that we denote by  $c_*$ , and for which*

$$(\xi_2(z) - \xi_3(z))dz \in \mathbb{R} \quad \text{along } \Gamma_* \setminus \{c_*\}. \quad (3.23)$$

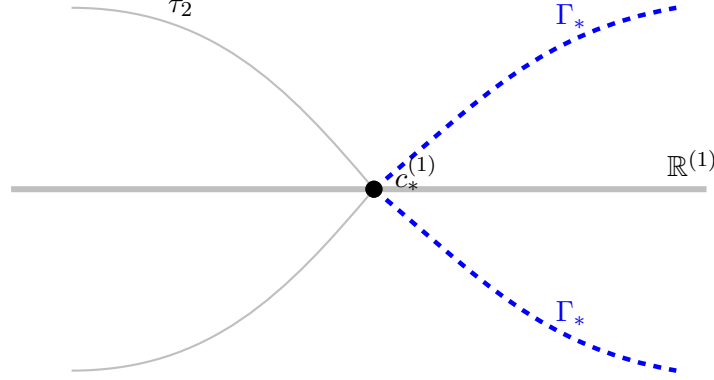


FIGURE 9. Unsaturated regime: contour  $\Gamma_*$  (dashed) as described in Proposition 3.18 and corresponding to the situation (I) of Proposition 3.16. Arcs of  $\Gamma_*$  correspond to the orthogonal trajectory  $\tau_1$  and its reflection with respect to  $\mathbb{R}$ , see Figure 4. Orthogonal trajectory  $\tau_2$  is indicated in gray.

Moreover, the difference  $\xi_2 - \xi_3$  does not vanish along  $\Gamma_*$  away from the real line.

*Remark 3.19.* As we mentioned before, in Proposition 3.29 we exclude the occurrence of the case (III) from Proposition 3.16, so the contour  $\Gamma_*$  in fact always exists.

The contour  $\Gamma_*$  in the unsaturated regime is depicted in Figure 9, corresponding to configuration (I) of Proposition 3.16, in which case  $c_*$  is the point for which we used the same notation there. In the situation (II) its geometry is analogous, and in such a case  $c_* = c_1$ . At least at the formal level, it is natural to expect that  $\Gamma_*$  in the unsaturated regime can be obtained from the corresponding contour in the saturated regime (see Figure 3) by continuous deformation of the parameter data  $(V, a, \alpha)$ .

*Proof.* For the structure of  $\partial\mathcal{H}$  covered by either one of the Figures 4 or 5, we take

$$\Gamma_* = \pi(\tau_1 \cup \tau_1^*),$$

and from the definition of the orthogonal trajectory  $\tau_1$  it follows that (3.23) is satisfied.

The fact that  $\Gamma_* \in \mathcal{T}$  follows from the geometry of the orthogonal trajectories involved in the construction, whereas properties of the zeros of  $\xi_2 - \xi_3$  are consequences of the definition of an orthogonal trajectory, see (3.17), and the properties of the zeros of  $\varpi$  on  $\partial\mathcal{H}$ , see for instance Proposition 3.16.  $\square$

### 3.6. Local behavior of the density of $\lambda$ .

The results in the previous section describe the behavior of certain trajectories of  $\varpi$ . Although some of the main findings therein, namely Propositions 3.15 and 3.18, contain, in principle, everything needed for the proof of Theorem A, there are several almost immediate consequences that not only provide more accurate information about the density  $\lambda$  but also give a more concrete geometric

picture of the sheets of  $\mathcal{R}$ , and as such, facilitate the proof of the main results. So without further ado, let us analyze the key consequences.

**Proposition 3.20.** *If  $\frac{d\lambda}{dx}(p) > 0$  for some  $p \in \mathbb{R}$ , then none of the functions  $\xi_1$ ,  $\xi_2$  and  $\xi_3$  has a branch point at  $p$ .*

*Proof.* If  $\lambda'(p) > 0$  then by (3.12),  $\xi_{1\pm}(p) \notin \mathbb{R}$  and  $\xi_{1+}(p) \neq \xi_{1-}(p)$ , so that  $p$  cannot be a branch point of  $\xi_1$ . But then by (3.11), one of the values  $\xi_{2\pm}(p)$ ,  $\xi_{3\pm}(p)$  is real and the other not, so that again  $\xi_2$  and  $\xi_3$  cannot be simultaneously branched at  $p$ .  $\square$

**Proposition 3.21.** *If  $p$  is an endpoint of one of the intervals comprising  $\text{supp } \lambda$ , then  $p$  cannot be a branch point of both  $\xi_2$  and  $\xi_3$ .*

*Proof.* We carry out the proof assuming that  $p \in \{b_1, \dots, b_l\}$ . The case when  $p \in \{a_1, \dots, a_l\}$  is similar.

To get to a contradiction, assume that  $\xi_2$  and  $\xi_3$  are branched at  $p$ . Because  $p$  is an endpoint of the support of  $\lambda$ , we get from (2.8) that  $\xi_1$  is branched at  $p$  as well, so  $p$  is a common branch point to the three functions  $\xi_1$ ,  $\xi_2$  and  $\xi_3$ . Hence, for some  $\varepsilon > 0$  the functions  $\xi_2$  and  $\xi_3$  admit an analytic continuation to  $U := D_\varepsilon(p) \setminus (p - \varepsilon, p]$ , and in this neighborhood we can expand

$$\xi_j(z) = \widehat{\xi}(z) + \beta\omega_j(z-p)^{\nu/3} + \mathcal{O}((z-p)^{(\nu+1)/3}), \quad z \in U, \quad j = 1, 2, 3,$$

for some function  $\widehat{\xi}$  analytic near  $p$  and some integer  $\nu > 0$  not a multiple of 3,  $\beta > 0$ , and  $|w_j| = 1$ , with the  $\omega_j$ 's pairwise distinct and satisfying  $\omega_1^3 = \omega_2^3 = \omega_3^3$ . The branch of cubic root in the expansion above is the principal one, that is, it is real-valued on  $(p, +\infty)$  and with a branch cut on  $(-\infty, p)$ .

The function  $\xi_1$  is real-valued on  $(p, p + \varepsilon)$  (see (2.8)), and thus we must have  $\omega_1 = 1$ , and consequently  $\xi_{2+}(x), \xi_{3+}(x)$  must be nonreal on  $(p, p + \varepsilon)$ . Because the equation (2.6) has real coefficients, we must then have  $\xi_{2+}(x) = \overline{\xi_{3+}(x)}$ , meaning that  $\xi_{2+}(x) - \xi_{3+}(x) \in i\mathbb{R} \setminus \{0\}$  for  $x \in (p, p + \varepsilon)$ . But this is in contradiction to Proposition 3.10.  $\square$

**Proposition 3.22.** *If  $p$  is an interior point of  $\text{supp } \lambda$  with  $\frac{d\lambda}{dx}(p) = 0$ , then  $p$  is either*

- a) *a regular point of all three branches  $\xi_1, \xi_2$  and  $\xi_3$ , or*
- b) *a branch point of all three branches  $\xi_1, \xi_2$  and  $\xi_3$  simultaneously. Then, for  $j = 1, 2, 3$ ,*

$$\xi_j(z) = \frac{V'(p)}{3} + \beta\omega_j(z-p)^{\nu/3} + \mathcal{O}(z-p)^{(\nu+1)/3}, \quad z \rightarrow p, \quad \text{Im } z > 0, \quad (3.24)$$

*where the cubic root has a branch cut along  $(-\infty, p)$ . Here  $\beta > 0$ ,  $\omega_1^3 = \omega_2^3 = \omega_3^3$  are three distinct values, and either one of the following possibilities takes place:*

- $\nu = 1$  and  $\omega_1 = e^{\pi i/3}$ , or
- $\nu = 5$  and  $\omega_1 = e^{2\pi i/3}$ .



Moreover, this situation can occur only in the unsaturated regime, and only at  $p_*^{(1)} = c_*^{(1)}$  described in Proposition 3.16, case (I).

*Proof.* By Lemma 3.2, because  $\frac{d\lambda}{dx}(p) = 0$  we obtain  $\xi_{1+}(p) = \xi_{1-}(p)$ . Also by the same Lemma, for any  $x \neq p$  in a small interval  $(p - \varepsilon, p + \varepsilon)$  there exists a permutation  $j, k$  of  $\{2, 3\}$  (depending on  $x$ ) such that

$$\xi_{1\pm}(x) = \overline{\xi_{1\mp}(x)} = \xi_{j\mp}(x) = \overline{\xi_{j\pm}(x)}, \quad \xi_{k\pm}(x) \in \mathbb{R}.$$

If  $\xi_1$  is not branched at  $p$  then without loss of generality we can assume that it is not branched on  $(p - \varepsilon, p + \varepsilon)$  either. Using continuity this means that in the identities above the indices  $j$  and  $k$  are independent of  $x$ . For instance, let  $j = 2$ , so that

$$\xi_{1\pm}(x) = \overline{\xi_{1\mp}(x)} = \xi_{2\mp}(x) = \overline{\xi_{2\pm}(x)}, \quad \xi_{3\pm}(x) \in \mathbb{R},$$

which yields that all three functions have a trivial monodromy at  $p$ . In other words, functions

$$\begin{cases} \xi_1(z), & \text{if } \operatorname{Im} z > 0, \\ \xi_2(z), & \text{if } \operatorname{Im} z < 0, \end{cases} \quad \text{and} \quad \begin{cases} \xi_3(z), & \text{if } \operatorname{Im} z > 0, \\ \xi_3(z), & \text{if } \operatorname{Im} z < 0, \end{cases}$$

have an analytic and single-valued continuation to a small neighborhood of  $p$ , so none of them could be branched at  $p$ . The other case, when  $j = 3$ , is analyzed identically.

This shows that we are in the option a).

Assume now that all three functions  $\xi_1, \xi_2$  and  $\xi_3$  are branched at a point  $p$  in the interior of  $\operatorname{supp} \lambda$ . From the general theory, the Puiseux expansion of the solutions to (2.6) at a branch point  $p$  of multiplicity three is of the form

$$\xi_j(z) = \frac{V'(p)}{3} + \beta \omega_j (z - p)^{\nu/3} + \mathcal{O}(z - p)^{(\nu+1)/3}, \quad z \rightarrow p, \quad \operatorname{Im} z > 0, \quad (3.25)$$

where the constant term comes from (2.8),  $\beta > 0$  and  $\omega_j$ 's satisfy  $\omega_1^3 = \omega_2^3 = \omega_3^3$  and  $|\omega_1| = 1$ . Let us analyze the possibilities for  $\nu \in \mathbb{N}$ . For the initial screening, we will use the consequence of Corollaries 3.14 and 3.17, according to which from any zero  $p^{(1)} \in \mathbb{R}^{(1)}$  of  $\varpi$  emanates at most one orthogonal trajectory on the upper half plane  $\mathbb{H}_+^{(1)}$ , unless we are in the unsaturated regime and  $p_*^{(1)} = c_*^{(1)}$  (see Proposition 3.16, case (I)), when there are exactly two.

With our assumptions, when  $p$  is a triple branch point, the local parameter at  $p$  is

$$z = p + u^3, \quad dz^2 = u^4 du^2, \quad (3.26)$$

and

$$\varpi = -(\xi_j - \xi_k)^2 dz^2 = \operatorname{const} \times u^{2\nu+4} du^2. \quad (3.27)$$

This means that  $\varpi$  has a zero of order  $2\nu + 4$  at this point, and there are  $2\nu + 6$  orthogonal trajectories emanating from  $p$  on  $\mathcal{R}$ , with the identical angle (in the local parameter  $u$ )

$$\frac{2\pi}{2\nu + 6} = \frac{\pi}{\nu + 3}$$

between two consecutive trajectories. By (3.26), this angle projects on the  $z$  plane as  $3\pi/(\nu+3)$ . Hence, if

$$\frac{3\pi}{\nu+3} < \frac{\pi}{3}, \quad \text{which is equivalent to} \quad \nu > 6,$$

then a winding counting implies that we must have at least three critical orthogonal trajectories emanating from  $p^{(1)}$  on the upper half plane  $\mathbb{H}_+^{(1)}$ , which is impossible.

Since by assumption  $\nu \not\equiv 0 \pmod{3}$ , we are only left with the possibilities  $\nu \in \{1, 2, 4, 5\}$ .

We now can discriminate further looking at the sign of  $\text{Im } \xi_{1+}(x)$  in a neighborhood of  $p$ . Indeed, by (3.12),

$$\text{Im } \xi_{1+}(x) = \pi \frac{d\lambda}{dx}(x) > 0 \quad \text{on } (p - \varepsilon, p + \varepsilon) \setminus \{p\},$$

so that  $\xi_{1+}(x) \notin \mathbb{R}$  there, and by (3.11), at least one of the solutions has to be real in each of the intervals  $(p - \varepsilon, p)$  and  $(p, p + \varepsilon)$ . Thus, necessarily either  $\omega_2$  or  $\omega_3$  are real (or equivalently, either  $\omega_2 \in \{-1, 1\}$  or  $\omega_3 \in \{-1, 1\}$ ). Furthermore, by (3.25),

$$\text{Im } \xi_{1+}(x) = \begin{cases} \beta \text{Im}(\omega_1)|x - p|^{\nu/3} + \mathcal{O}(|x - p|^{\frac{\nu+1}{3}}) & \text{on } (p, p + \varepsilon), \\ \beta \text{Im}(\omega_1 e^{\pi i \nu/3})|x - p|^{\nu/3} + \mathcal{O}(|x - p|^{\frac{\nu+1}{3}}) & \text{on } (p - \varepsilon, p). \end{cases}$$

Hence, by assumption, both  $\text{Im}(\omega_1) \geq 0$  and  $\text{Im}(\omega_1 e^{\pi i \nu/3}) \geq 0$ . Moreover, since either  $\xi_{2+}(x) \in \mathbb{R}$  or  $\xi_{3+}(x) \in \mathbb{R}$  in  $(p - \varepsilon, p)$ , we also need that either  $\omega_2 e^{\pi i \nu/3} \in \mathbb{R}$  or  $\omega_3 e^{\pi i \nu/3} \in \mathbb{R}$ .

Taking into account that  $\omega_1^3 = \omega_2^3 = \omega_3^3$  and they are pairwise distinct, this leaves us only two possibilities:

- a) if, without loss of generality,  $\omega_3 = 1$  then  $\omega_1 = e^{2\pi i/3} = \overline{\omega_2}$ . A priori,  $\text{Im}(\omega_1 e^{\pi i \nu/3}) \geq 0$  implies that  $\nu = 1, 4$  or  $5$ . But recalling that either  $\omega_2 e^{\pi i \nu/3} \in \mathbb{R}$  or  $\omega_3 e^{\pi i \nu/3} \in \mathbb{R}$ , we conclude that the only possibility is  $\nu = 5$ .
- b) if, without loss of generality,  $\omega_3 = -1$  then  $\omega_1 = e^{\pi i/3} = \overline{\omega_2}$ . A priori,  $\text{Im}(\omega_1 e^{\pi i \nu/3}) \geq 0$  implies that  $\nu = 1, 2$  or  $5$ . But recalling that either  $\omega_2 e^{\pi i \nu/3} \in \mathbb{R}$  or  $\omega_3 e^{\pi i \nu/3} \in \mathbb{R}$ , we conclude that the only possibility is  $\nu = 1$ .

This leaves us only with the two possibilities indicated in the case b) of the Proposition. In order to establish the last statement, we return to the local parameter (3.26) and count the actual number of orthogonal trajectory emanating from  $p^{(1)}$  on the upper half plane  $\mathbb{H}_+^{(1)}$ .

If, again without loss of generality,  $\omega_2 = \pm 1$ , then  $(p, p + \varepsilon)^{(2)}$  is an arc of trajectory of  $\varpi$ . According to the local structure of trajectories and orthogonal trajectories, there is an orthogonal trajectory emanating from  $p^{(2)}$  on  $\mathcal{R}^{(2)}$  in the direction  $e^{i\theta_0}$ ,  $\theta_0 = 3\pi/(2(\nu+3))$ , and the rest are equally spaced, emanating from  $p$  (on the corresponding sheet) with the directions  $e^{i\theta_k}$ , where

$$\theta_k = \frac{3\pi}{2(\nu+3)} + k \frac{3\pi}{\nu+3}, \quad k = 0, 1, \dots, 2\nu+5.$$

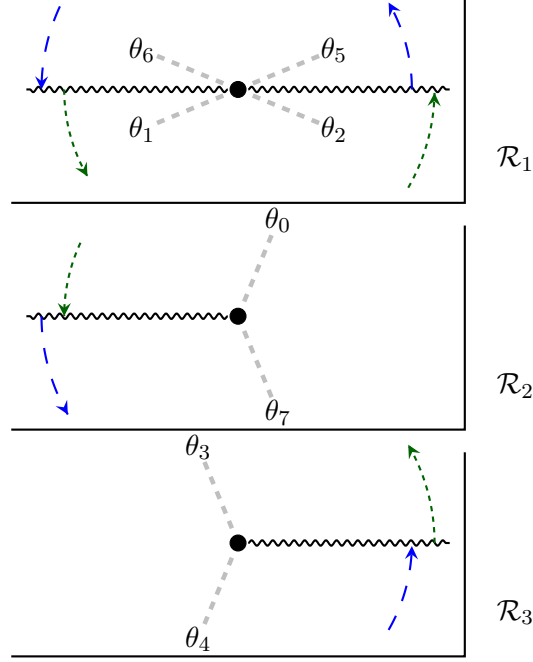


FIGURE 10. The sheet structure of  $\mathcal{R}$  and behavior of local trajectories near  $p = p^{(1)}$  in the case when the expansion (3.24) is taking place with  $\nu = 1$ . The wiggled lines correspond to the branch cuts connecting the three different sheets at  $p$ , with the arrow lines (long dashed and short dashed) indicating the orientation of the branch cut connections. The dashed gray lines correspond to the angles  $\theta_0, \dots, \theta_7$  indicating the directions in which the orthogonal trajectories emanate from  $p$ .

A simple analysis shows that in both cases there are exactly two orthogonal trajectories in  $\mathbb{H}_+^{(1)}$ : for  $\nu = 1$  these correspond to the directions  $\theta_5, \theta_6$  (see Figure 10), and for  $\nu = 5$ , to the directions  $\theta_{11}, \theta_{12}$ .

It remains to apply Corollaries 3.14 and 3.17 and Proposition 3.16 to conclude the proof. □

*Remark 3.23.* Looking at the possible deformations of our problem, we can describe, rather informally, the dynamical origin of both types of triple branch points stated in Proposition 3.22. The transition from the saturated to unsaturated regime takes place when both non-real zeros of  $\varpi$ ,  $y_*^{(1)}$  and  $\overline{y_*^{(1)}}$  (square root type branch points), merge on the real line. If this happens in the bulk of  $\text{supp } \lambda$ , the triple branching with  $\nu = 1$  occurs. Otherwise, the zero formed from the coalescence of  $y_*^{(1)}$  and  $\overline{y_*^{(1)}}$  on  $\mathbb{R}^{(1)}$  is “trapped” between the colliding endpoints of  $\text{supp } \mu_1$  and  $\text{supp } \mu_2$  (also

square root type branch points), giving origin to  $\nu = 5$  in (3.24). In random matrix theory language, the latter would correspond to the closure of a gap happening at the place where the support of the third measure  $\text{supp } \mu_3^*$  (which will be constructed later) degenerates to a single point.

We finish this section with a definition that is motivated by the conclusions of Proposition 3.15:

**Definition 3.24.** We say that we are in the *regular unsaturated regime* if we are in the saturated regime but there is no point for which the expansion (3.24) is valid. Otherwise, we say that we are in the *singular unsaturated regime*.

The words “regular” and “singular” in the definition above come from their consequence on the local behavior of the density of  $\lambda$  claimed by Theorem A. As we will see in a moment, the Pearcey-type singular behavior exists precisely when we are in the singular unsaturated regime.

### 3.7. Riemann surface for the spectral curve: take II.

At the beginning of Section 3 we mentioned the three-sheeted Riemann surface  $\mathcal{R} = \mathcal{R}_1 \cup \mathcal{R}_2 \cup \mathcal{R}_3$  associated to the spectral curve (2.6). In that moment, we chose  $\mathcal{R}_1 = \mathbb{C} \setminus \text{supp } \lambda$ , but the remaining sheets  $\mathcal{R}_2$  and  $\mathcal{R}_3$ , apart from minor considerations, were taken rather arbitrarily.

Now we can use the information from the previous sections to construct  $\mathcal{R}_2$  and  $\mathcal{R}_3$  in a more precise manner.

Recall that the saturated and unsaturated regimes are described in Definition 3.12, and the unsaturated regime was divided further into two subregimes in Definition 3.24.

**Theorem 3.25.** *The branch points of  $\xi_1, \xi_2$  and  $\xi_3$  are as follows.*

- (a) *For any  $p \in \{a_1, b_1, \dots, a_l, b_l\}$ , the function  $\xi_1$  and exactly one of the other two solutions  $\xi_2$  or  $\xi_3$  are branched (“square root” branching).*
- (b) *In the singular unsaturated regime, the three functions  $\xi_1, \xi_2$  and  $\xi_3$  are branched at  $p = c_*$ , which is a point in the interior of  $\text{supp } \lambda$ . Furthermore, this is the unique point where the expansion (3.24) is valid.*
- (c) *In the saturated regime, the functions  $\xi_2$  and  $\xi_3$  are branched also at  $y_* \in \mathbb{H}_+$  and  $\bar{y}_*$  and  $(\xi_2 - \xi_3)^2$  has a simple zero at these points.*

*The points listed above are the only branch points of these functions.*

*Proof.* From the definition of  $\xi_1$  in (2.8), it follows that it has to be branched at every point  $p \in \{a_1, b_1, \dots, a_l, b_l\}$ . By Proposition 3.21 exactly one among  $\xi_2$  and  $\xi_3$  is also branched at this point.

In the singular unsaturated regime, the point  $p = c_*$  is a branch point of the three functions by Definition 3.24 and Proposition 3.22. Uniqueness of such  $c_*$  follows also from Proposition 3.22.

In the saturated regime, the quadratic differential  $\varpi$  has a zero at  $y_*^{(1)}$ . By the construction of the sheet  $\mathcal{R}_1$  in (3.6) we know that the difference  $(\xi_2 - \xi_3)^2$  has to have a simple zero at this point, and consequently  $\xi_2$  and  $\xi_3$  have to be branched with square root behavior for their difference. Symmetry under conjugation gives us the same conclusion for  $\bar{y}_*$ .

To see that these are the only possible branch points, fix  $p \in \mathbb{C}$ . We proceed case by case

- If  $p \in \mathbb{C} \setminus \mathbb{R}$ , then from (2.8) we know that  $\xi_1$  cannot be branched at this point. As for the other two solutions  $\xi_2$  and  $\xi_3$ , proceeding as above and using Proposition 3.9 we see that they can only vanish at this  $p$  if we are in the saturated regime and  $p = y_*$  or  $p = \bar{y}_*$ , that is, if we are in case (c).
- If  $p \in \mathbb{R} \setminus \text{supp } \lambda$ , then by the representation (2.8) the function  $\xi_1$  cannot be branched at  $p$ . Also, from Corollary 3.11 we learn that if  $\xi_2 - \xi_3$  vanishes at this point, it has to do so with integer order, so these functions are not branched at  $p$  either.
- The case when  $p$  is an endpoint of  $\text{supp } \lambda$  falls into (a), so it was already dealt with.
- In the case when  $p$  is in the interior of  $\text{supp } \lambda$  with  $\frac{d\lambda}{dx}(p) > 0$ , then none of the functions  $\xi_1, \xi_2$  and  $\xi_3$  is branched, as proved in Proposition 3.20.
- Finally, as a last case suppose that  $p$  is in the interior of  $\text{supp } \lambda$  but  $\frac{d\lambda}{dx}(p) = 0$  and  $p$  is a branch point for one of the solutions. Then by Proposition 3.22 it has to be a branch point of all three solutions. By Proposition 3.22 and Definition 3.24, we are in the case (b).

□

The previous proposition takes care of the book-keeping of branch points that will be necessary to construct the sheet structure of  $\mathcal{R}$  in a very explicit manner. Our next step is to construct the sets that will support the measures  $\mu_1^*$  and  $\mu_2^*$  in Theorem A, which will also provide a natural branch cut structure for  $\mathcal{R}$ .

In the unsaturated regime, an inspection of the possible geometries of  $\partial\mathcal{H}$  in Figures 4, 5 and 6 shows that  $\mathbb{H}_+^{(1)} \setminus \bar{\mathcal{H}}$  has at most two connected components, which are necessarily unbounded. Likewise, in the saturated regime an inspection of Figure 2 shows that  $\mathbb{H}_+^{(1)} \setminus (\bar{\mathcal{H}} \cup \tau_0)$  also has two unbounded components.

In either case, we denote these components by  $\mathcal{G}_1$  and  $\mathcal{G}_2$ , with the convention that they are possibly empty if the situation (III) of Proposition 3.16 takes place, and labeled in such a way that  $\mathcal{G}_1$  is unbounded near  $\infty$  for large and positive values of  $\text{Re } z$ , whereas  $\mathcal{G}_2$  remains unbounded for large negative values of  $\text{Re } z$ . These sets are displayed in Figures 11 and 12, in the saturated and unsaturated regimes, respectively.

The solution  $\xi_1$  of the spectral curve is, from the very beginning, fixed on the whole plane by (2.8), while the other two solutions are well defined analytic functions in any simply connected domain  $D$  of  $\mathbb{C}$  not containing branch points. However, their labeling is fixed only by their asymptotic behavior

(3.1), and is an issue to solve in the case of a bounded  $D$ . Up to this point, this distinction of labels was not relevant simply because we only cared about the *unsigned* difference of solutions  $\xi_2$  and  $\xi_3$ .

The sets  $\mathcal{G}_1$  and  $\mathcal{G}_2$  are simply connected and do not contain branch points of the spectral curve. Thus, we can *uniquely* define  $\xi_2$  and  $\xi_3$  as analytic functions on  $\mathcal{G}_j$ ,  $j = 1, 2$ , by performing analytic continuation from  $\infty$  throughout the whole set  $\mathcal{G}_j$  with the help of the asymptotics (3.1). With this convention in mind, for the next result, for points  $z, w \in \mathcal{G}_k$  set

$$\Upsilon_k(z, w) := \int_z^w (\xi_2(s) - \xi_3(s)) ds, \quad z, w \in \mathcal{G}_k,$$

where the path of integration is contained in  $\mathcal{G}_k$ ; we extend  $\Upsilon_k$  for finite points on the boundary of  $\mathcal{G}_k$  by continuity.

**Lemma 3.26.** *For  $k = 1, 2$  and any points  $x, y \in \mathbb{R} \cap \partial\mathcal{G}_k$ ,  $x > y$ ,*

$$(-1)^k \operatorname{Im} \Upsilon_k(x, y) \geq 0. \quad (3.28)$$

*Proof.* Fix  $x > y$  as in the proposition. Clearly the path of integration defining  $\Upsilon_k$  is irrelevant, as long as it stays within  $\overline{\mathcal{G}_k}$ . Also, denoting  $x_\varepsilon = x + i\varepsilon$ , continuity tells us that it is enough to verify that  $(-1)^k \operatorname{Im} \Upsilon_k(x_\varepsilon, y_\varepsilon) \geq 0$  for any  $\varepsilon > 0$  sufficiently small.

Since  $y_\varepsilon^{(1)}$  is a regular point of  $\varpi$ , there is a unique orthogonal trajectory passing through  $y_\varepsilon^{(1)}$ ; in other words, there are two rays of orthogonal trajectories emanating from  $y_\varepsilon^{(1)}$  in opposite directions. From Remark 3.7 we see that one of these trajectories, say  $\tau$ , has to extend to  $\infty^{(1)}$  without intersecting  $\mathbb{R}^{(1)}$ . We orient  $\tau$  from  $y_\varepsilon^{(1)}$  to  $\infty$ . In a similar manner, there is an orthogonal trajectory  $\gamma$  from  $x_\varepsilon^{(1)}$  to  $\infty$ .

Either  $\gamma$  or  $\tau$  contain both  $x_\varepsilon^{(1)}$  and  $y_\varepsilon^{(1)}$  (and then  $\operatorname{Im} \Upsilon_k(x_\varepsilon, y_\varepsilon) = 0$ , which is consistent with (3.28)), or  $\gamma$  and  $\tau$  are disjoint.

In the latter case, both contours  $\tau$  and  $\gamma$  diverge to  $\infty$  within  $\mathcal{G}_k$  in the same asymptotic direction (with  $\arg(\cdot) = 0$  for  $k = 1$  and  $\arg(\cdot) = \pi$  for  $k = 2$ ).

Thus, for any  $R$  sufficiently large, both trajectories  $\tau$  and  $\gamma$  intersect  $\overline{\mathcal{G}_k} \cap \{|z| = R\}$ , say at points  $u$  and  $v$  respectively. We can write

$$\Upsilon_k(x_\varepsilon, y_\varepsilon) = \int_{y_\varepsilon}^u (\xi_2(s) - \xi_3(s)) ds + \int_u^v (\xi_2(s) - \xi_3(s)) ds + \int_v^{x_\varepsilon} (\xi_2(s) - \xi_3(s)) ds.$$

The first and the last integrals above can be computed along  $\tau$  and  $\gamma$ , which are orthogonal trajectories, so

$$\operatorname{Im} \Upsilon_k(x_\varepsilon, y_\varepsilon) = \operatorname{Im} \int_u^v (\xi_2(s) - \xi_3(s)) ds.$$

From the geometry of  $\mathcal{G}_k$  and the fact that  $\gamma$  and  $\tau$  cannot intersect, we can take  $R$  sufficiently large and make sure that  $\operatorname{Im} u > \operatorname{Im} v$  when  $k = 1$  and  $\operatorname{Im} u < \operatorname{Im} v$  when  $k = 2$ . We then use the

expansion (3.1) to get that

$$\int_u^v (\xi_2(s) - \xi_3(s))ds = 2a(v - u)(1 + \mathcal{O}(1)),$$

and this yields the result.  $\square$

*Remark 3.27.* The claim in Lemma 3.26 has a natural interpretation in terms of the quadratic differential  $\varpi$ . Each set  $\mathcal{G}_k$  can be decomposed as  $\mathcal{G}_k = \mathbb{H}_+^{(1)} \cap (\cup \mathcal{S})$ , where the union is over a finite number of sets  $\mathcal{S}$ , each one being a strip domain of the critical graph of  $\varpi$ . That is, the boundary of  $\partial \mathcal{S}$  contains exactly two poles of  $\varpi$  and finitely many zeros, and fixing a finite point  $q \in \mathcal{S}$  the function  $z \mapsto \Upsilon_k(q, z)$  is a conformal map from  $\mathcal{S}$  to a domain bounded by two vertical lines on  $\mathbb{C}$  - hence the name *strip domain*.

Lemma 3.26 is then saying that the (signed) height (or moduli parameter) associated to each  $\mathcal{S}$  is the same, and as such one can extend  $\Upsilon_k(q, \cdot)$  conformally to the whole union  $\cup \mathcal{S}$  by gluing together some subarcs of the boundaries of these domains.

In the unsaturated regime, Theorem 3.25 assures us that  $\xi_2$  and  $\xi_3$  can be analytically continued from  $\infty$ , using the expansion (3.1), to the upper half plane  $\mathbb{H}_+$ .

In the saturated regime, however, the upper half plane contains the branch point  $y_*$  of  $\xi_2$  and  $\xi_3$ . Thus, in this situation from now on we speak about the analytic continuation of these functions  $\xi_2$  and  $\xi_3$  from  $\infty$  using (3.1) to the simply connected domain  $\mathbb{H}_+ \setminus \Delta_3$ , where  $\Delta_3$  is the contour defined in Proposition 3.15. The only branch point of these functions on  $\mathbb{H}_+$  is an endpoint of  $\Delta_3$ , so this analytic continuation is well defined.

Observe also that this definition does not agree with the convention for  $\xi_2$  and  $\xi_3$  we used for Lemma 3.26. To help in the discussion regarding this distinction, for simplicity let us denote by  $\widehat{\xi}_2$  and  $\widehat{\xi}_3$  the branches of these analytic functions which are obtained with analytic continuation from  $\infty$  with (3.1) to the domain  $\mathbb{H}_+ \setminus \tau_0$ , where  $\tau_0 = \overline{\mathcal{G}_1} \cap \overline{\mathcal{G}_2}$ , as displayed in Figure 11, and we continue denoting by  $\xi_2, \xi_3$  the analytic continuations to  $\mathbb{H}_+ \setminus \Delta_3$ . Also, denote by  $\mathcal{D}$  the domain on  $\mathcal{G}_2$  bounded by  $\tau_0$  and  $\Delta_3$ .

The branches  $\widehat{\xi}_2$  and  $\widehat{\xi}_3$  are the ones for which (3.26) is applicable, so for them this Lemma implies that

$$\operatorname{Im}(\widehat{\xi}_{2+}(x) - \widehat{\xi}_{3+}(x))dx < 0, \quad x < c_*, \quad \text{and} \quad \operatorname{Im}(\widehat{\xi}_{2+}(x) - \widehat{\xi}_{3+}(x))dx > 0, \quad x > c_*.$$

On the sets  $\mathcal{G}_1$  and  $\mathcal{G}_2 \setminus \overline{\mathcal{D}}$ , the branches  $\xi_2$  and  $\widehat{\xi}_2$  coincide, as well as the branches  $\xi_3$  and  $\widehat{\xi}_3$ . This means that the inequalities above immediately transfer to

$$\operatorname{Im}(\xi_{2+}(x) - \xi_{3+}(x))dx < 0, \quad x < x_*, \quad \text{and} \quad \operatorname{Im}(\xi_{2+}(x) - \xi_{3+}(x))dx > 0, \quad x > x_*.$$

On the interval  $(x_*, c_*)$ , however, there is a change. In the domain  $\mathcal{D}$  we actually have  $\widehat{\xi}_3 = \xi_2$  and  $\widehat{\xi}_2 = \xi_3$ , and the previous inequality updates to

$$\operatorname{Im}(\xi_{2+}(x) - \xi_{3+}(x))dx < 0, \quad x < x_*, \quad \text{and} \quad \operatorname{Im}(\xi_{2+}(x) - \xi_{3+}(x))dx > 0, \quad x > x_*. \quad (3.29)$$

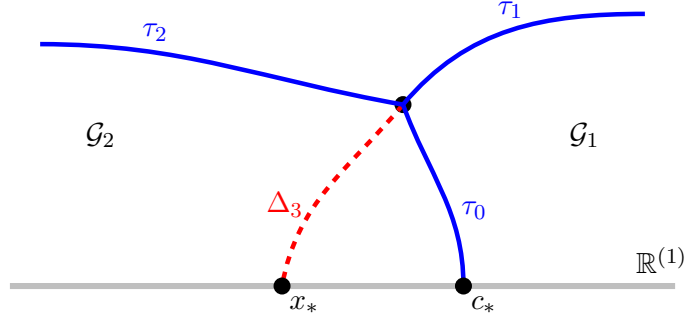


FIGURE 11. Saturated regime: the partition of the half plane  $\mathbb{H}_+^{(1)}$  induced by the orthogonal trajectories  $\tau_0, \tau_1$  and  $\tau_2$ , the corresponding sets  $\mathcal{G}_1$  and  $\mathcal{G}_2$  and also the set  $\Delta_3$ , which is a vertical trajectory. This is an updated version of previous figures, compare for instance with Figures 2 and 3.

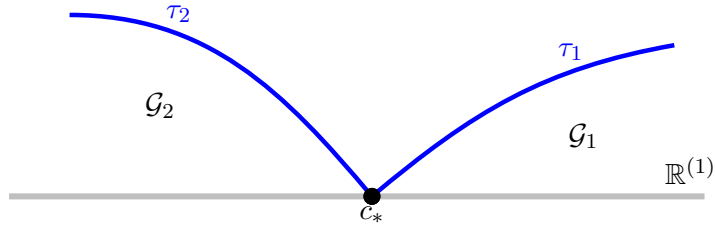


FIGURE 12. Unsaturated regime: the partition of the half plane  $\mathbb{H}_+^{(1)}$  induced by the orthogonal trajectories  $\tau_1$  and  $\tau_2$  and the corresponding sets  $\mathcal{G}_1$  and  $\mathcal{G}_2$ . This is an updated version of previous figures, compare for instance with Figures 4 and 9, and corresponds only to the situation in Proposition 3.16–(I), but the geometry of  $\mathcal{G}_1$  and  $\mathcal{G}_2$  in the situations (II)–(III) is analogous.

These calculations were valid in the saturated regime, but in the unsaturated regime one can simply take  $\mathcal{D} = \emptyset$  and  $x_* = c_*$  in the calculations above, and the inequalities (3.29) are still valid.

Inspired by the boundary value behavior (3.13), let us define the closed subsets  $\Delta_1$  and  $\Delta_2$  of  $\mathbb{R}$  as follows:

$$\Delta_j := \overline{\{x \in \mathbb{R} : \text{sign}(\text{Im}(\xi_2 - \xi_3)_+(x)) = (-1)^j\}}, \quad j = 1, 2. \quad (3.30)$$

Each  $\Delta_j$  is a finite union of real intervals; by (3.11) from Lemma 3.2,  $\xi_{1\pm}(x) = \xi_{(j+1)\mp}(x)$  on  $\Delta_j$ ,

$$\text{supp } \lambda = \Delta_1 \cup \Delta_2, \quad (3.31)$$

and  $\Delta_1 \cap \Delta_2$  a priori can have at most a finite number of points. Recall also that in the saturated regime we defined the real-symmetric and piece-wise analytic curve  $\Delta_3$ , see Proposition 3.15. Our next step is to understand the relative position between  $\Delta_1, \Delta_2$  and  $\Delta_3$ .

The inequalities (3.29) immediately imply the next result.



**Proposition 3.28.** *The sets  $\Delta_1$  and  $\Delta_2$  satisfy*

$$\Delta_1 \subset [x_*, +\infty) \quad \text{and} \quad \Delta_2 \subset (-\infty, x_*].$$

*Thus, either  $\Delta_1 \cap \Delta_2 = \emptyset$  or  $\Delta_1 \cap \Delta_2 = \{x_*\}$ , and these situations can be further classified as follows.*

- (a) *In the regular unsaturated regime, we always have  $\Delta_1 \cap \Delta_2 = \emptyset$ , and we can take  $x_* = c_*$  to be the unique critical point on the boundary of  $\mathcal{H}$ .*
- (b) *In the singular unsaturated regime, we always have  $\Delta_1 \cap \Delta_2 = \{x_*\}$  and  $x_* = c_*$  is the unique point where the expansion (3.24) holds (see Proposition 3.25). In particular,  $x_*$  is in the interior of  $\text{supp } \lambda$ .*
- (c) *In the saturated regime,  $\Delta_1 \cap \Delta_2 = \emptyset$  or  $\Delta_1 \cap \Delta_2 = \{x_*\}$ ; in case the latter takes place,  $x_*$  is the point of intersection of  $\text{supp } \lambda$  with  $\Delta_3$  (see Proposition 3.15).*

After all this work, we can finally eliminate the case (III) that appeared in Proposition 3.16.

**Proposition 3.29.** *Suppose that we are in the unsaturated regime and case (III) of Proposition 3.16 takes place. Then the spectral curve (2.6) is reducible and  $\alpha \in \{0, 1\}$ .*

*Proof.* Suppose, for instance, that situation (III)–(a) is taking place. In this case  $c_* = x_* \leq a_1$  and the construction just carried out assures that  $\Delta_2 \subset (-\infty, a_1]$ . However, in this interval the spectral curve has no branch points (see Theorem 3.25), and consequently  $\Delta_2 = \emptyset$ . That is, we have

$$\text{Im}(\xi_{2+}(x) - \xi_{3+}(x)) \geq 0, \quad x \in \mathbb{R}.$$

But a combination of (3.11) and (3.13) then tells us that  $\xi_1$  and  $\xi_2$  never share a branch point. Once again recalling Theorem 3.25, we also know that in the unsaturated regime  $\xi_2$  and  $\xi_3$  do not share branch points. This means that  $\xi_2$  does not have branch points at all, and thus it is entire. So the algebraic equation is reducible and we can compute from (3.1) that  $\alpha = 0$  and  $\xi_2 \equiv a$ .

The case (III)–(b) leads to reducibility and  $\alpha = 1$  in a similar way. □

*Remark 3.30.* Actually, the argument outlined above says that if we were to allow  $\alpha = [0, 1]$  in Definition 2.1, then the spectral curve (2.6) would be reducible if, and only if,  $\alpha \in \{0, 1\}$ , and in such a situation the unsaturated regime, case (III), would be taking place.

With the help of Theorem 3.25 and Proposition 3.28 we are now ready to explicitly construct the Riemann surface  $\mathcal{R} = \mathcal{R}_1 \cup \mathcal{R}_2 \cup \mathcal{R}_3$ . The construction will be based on whether we are in the regular unsaturated, singular unsaturated or saturated regimes. In either case, as a final result we will find that each  $\xi_j$  is analytic in  $\mathcal{R}_j$ , and the global meromorphic solution to (2.6) is defined on the whole surface  $\mathcal{R}$  as in (3.14).

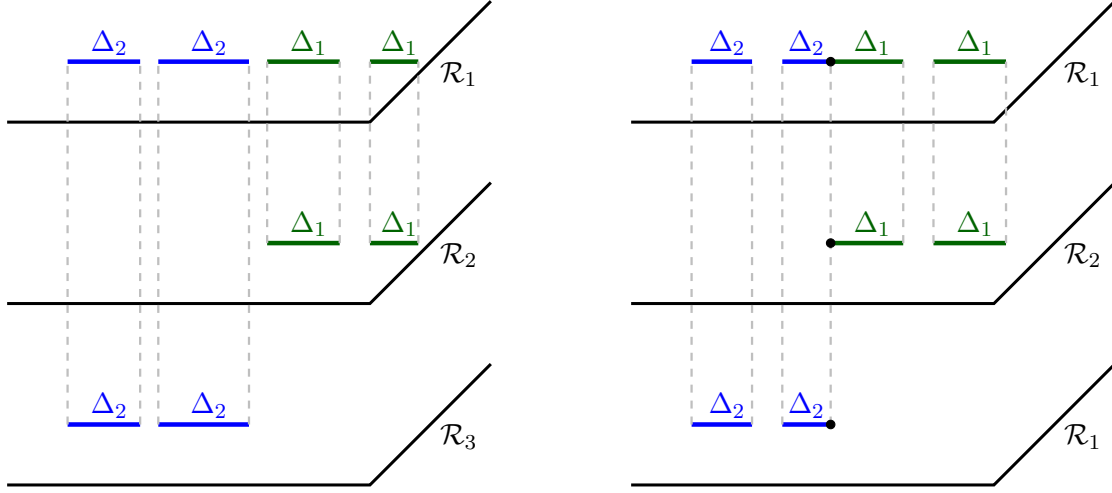


FIGURE 13. The sheet configurations in the regular unsaturated regime (left) and singular unsaturated regime (right). In both figures, the solid lines correspond to the cuts and the dashed gray lines indicate the common branch points to the sheets. In the singular unsaturated regime there is a branch point common to the three sheets, this point is indicated by the black dot on the three sheets.

**Riemann surface in the regular unsaturated regime.** By Theorem 3.25, in this case the collection of branch points of  $\xi_1, \xi_2$  and  $\xi_3$  is  $\{a_1, b_1, \dots, a_l, b_l\}$ , and by Proposition 3.28,

$$\Delta_1 \cap \Delta_2 = \emptyset.$$

We take  $\mathcal{R}_1$  as in (3.6),

$$\mathcal{R}_2 := \overline{\mathbb{C}} \setminus \Delta_1, \quad \mathcal{R}_3 := \overline{\mathbb{C}} \setminus \Delta_2,$$

and  $\mathcal{R}$  is realized as the three-sheeted branched cover

$$\mathcal{R} = \mathcal{R}_1 \cup \mathcal{R}_2 \cup \mathcal{R}_3$$

of  $\overline{\mathbb{C}}$ , with  $\mathcal{R}_1$  glued with  $\mathcal{R}_j$  along  $\Delta_{j-1}$  in the usual crosswise manner,  $j = 2, 3$ . In particular, in this case there is no direct connection between the sheets  $\mathcal{R}_2$  and  $\mathcal{R}_3$ . This construction is illustrated in Figure 13, left.

By the Riemann-Hurwitz formula, the genus of  $\mathcal{R}$  in this case is

$$g = l - 2,$$

where  $l$  is the number of disjoint components of the support  $\text{supp } \lambda$ .

**Riemann surface in the singular unsaturated regime.** By Theorem 3.25, the branch points are  $a_1, b_1, \dots, a_l, b_l$  and  $x_*$ , and  $\Delta_1 \cap \Delta_2 = \{x_*\}$ , where  $x_*$  is described by the option b) of Proposition 3.28.

As in the previous case,  $\mathcal{R} = \mathcal{R}_1 \cup \mathcal{R}_2 \cup \mathcal{R}_3$ , observing that  $x_*$  is common to the three sheets, and it is also the only common point to  $\mathcal{R}_2$  and  $\mathcal{R}_3$ . This construction is shown in Figure 13, right.

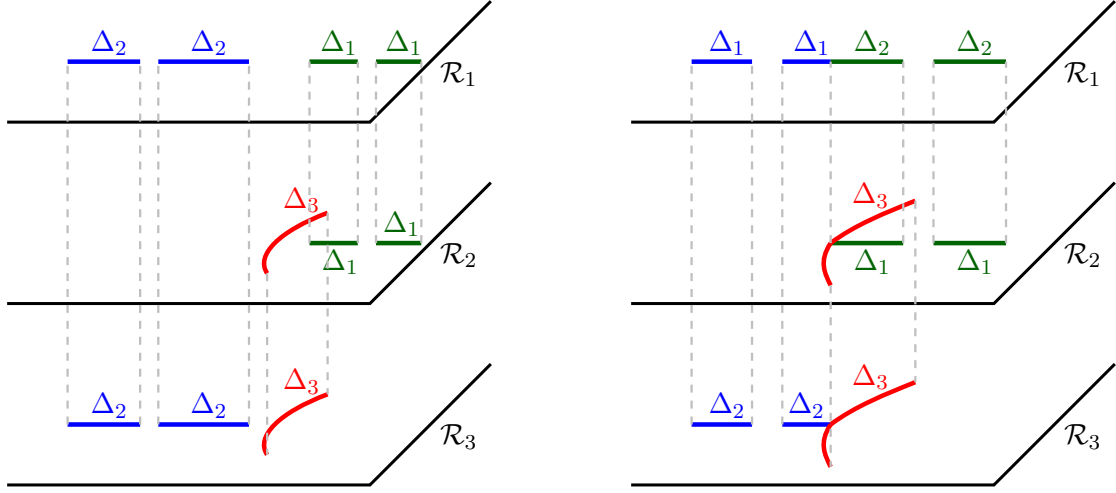


FIGURE 14. The sheet configurations in the saturated regime. On the left, the case when  $\Delta_3 \cap \text{supp } \lambda = \emptyset$  and on the right when  $\Delta_3 \cap \text{supp } \lambda = \{x_*\}$ . In both figures, the solid lines correspond to the cuts and the dashed gray lines indicate the common branch points to the sheets.

The genus of  $\mathcal{R}$  in this case is still

$$g = l - 2,$$

where  $l$  is the number of disjoint components of the support  $\text{supp } \lambda$ . Notice that in the transition from the regular to the singular unsaturated regime the number of components decreases in 1, but the genus of  $\mathcal{R}$  remains unchanged.

**Riemann surface in the saturated regime.** Once again by Theorem 3.25, the branch points are  $a_1, b_1, \dots, a_l, b_l$ , and also  $y_*$  and  $\bar{y}_*$ , the latter two being common branch points to  $\xi_2$  and  $\xi_3$ .

In any case, by Proposition 3.28 either  $\Delta_1 \cap \Delta_2 = \{x_*\}$  or  $\Delta_1 \cap \Delta_2 = \emptyset$  (when  $\Delta_3 \cap \mathbb{R}$  is not in  $\text{supp } \lambda$ ).

We set  $\mathcal{R}_1$  as in (3.6),

$$\mathcal{R}_2 := \overline{\mathbb{C}} \setminus (\Delta_1 \cup \Delta_3), \quad \mathcal{R}_3 := \overline{\mathbb{C}} \setminus (\Delta_2 \cup \Delta_3),$$

and glue these sheets together in

$$\mathcal{R} = \mathcal{R}_1 \cup \mathcal{R}_2 \cup \mathcal{R}_3,$$

with  $\mathcal{R}_1$  connected to  $\mathcal{R}_j$  along  $\Delta_{j-1}$ ,  $j = 2, 3$ , and  $\mathcal{R}_2$  and  $\mathcal{R}_3$  connected along  $\Delta_3$ , always in the usual crosswise manner. This sheet configuration is illustrated in Figure 14.

Again the Riemann-Hurwitz formula gives us that now the genus of  $\mathcal{R}$  is

$$g = l - 1,$$

where  $l$  is the number of disjoint components of the support  $\text{supp } \lambda$ .

### 3.8. From the Riemann surface to critical measures.

We are finally in the position to construct the measures  $\mu_1^*$ ,  $\mu_2^*$  and  $\mu_3^*$  for Theorem A. To do so, we will make use of the sets  $\Delta_j$  that were already introduced in the previous section. Recall also the set  $\Gamma_*$  formally defined in Proposition 3.15 (in the saturated regime) and Proposition 3.18 (in the unsaturated regime). We orient  $\Gamma_*$  from the lower to the upper half plane, and in the saturated regime the set  $\Delta_3 \subset \Gamma_*$  inherits the orientation from  $\Gamma_*$ . In the unsaturated regime we set  $\Delta_3 = \emptyset$ .

Define measures  $\mu_1^*$ ,  $\mu_2^*$  and  $\mu_3^*$  through their densities as

$$\begin{aligned} d\mu_1^*(s) &= d\lambda(s)|_{\Delta_1} = \frac{1}{2\pi i}(\xi_{1+}(s) - \xi_{1-}(s))ds = \frac{1}{2\pi i}(\xi_1 - \xi_2)_+(s)ds, \quad s \in \Delta_1, \\ d\mu_2^*(s) &= d\lambda(s)|_{\Delta_2} = \frac{1}{2\pi i}(\xi_{1+}(s) - \xi_{1-}(s))ds = \frac{1}{2\pi i}(\xi_1 - \xi_3)_+(s)ds, \quad s \in \Delta_2, \\ d\mu_3^*(s) &= \frac{1}{2\pi i}(\xi_{2+}(s) - \xi_{3+}(s))ds, \quad s \in \Delta_3, \end{aligned} \quad (3.32)$$

with the convention that  $\mu_3^* = 0$  if  $\Delta_3 = \emptyset$ , that is, if we are in the unsaturated regime. Notice that using (3.13) we can alternatively express all three  $\mu_j^*$ 's in terms of the boundary values of  $\xi_{2+} - \xi_{3+}$  on  $\Delta_j$ 's.

**Proposition 3.31.** *The measures  $\mu_1^*$ ,  $\mu_2^*$  and  $\mu_3^*$  defined in (3.32) are positive and their Cauchy transforms satisfy the identities*

$$\begin{aligned} \xi_1(z) &= C^{\mu_1^*}(z) + C^{\mu_2^*}(z) + V'(z), \\ \xi_2(z) &= C^{\mu_3^*}(z) - C^{\mu_1^*}(z) + a, \\ \xi_3(z) &= -C^{\mu_2^*}(z) - C^{\mu_3^*}(z) - a, \end{aligned} \quad z \in \mathbb{C} \setminus (\Delta_1 \cup \Delta_2 \cup \Delta_3). \quad (3.33)$$

In particular, the following relations between their total masses holds:

$$|\mu_1^*| + |\mu_2^*| = 1, \quad |\mu_1^*| - |\mu_3^*| = \alpha, \quad |\mu_2^*| + |\mu_3^*| = 1 - \alpha.$$

Moreover, if  $\Delta$  is a connected component of  $\Delta_1 \cup \Delta_2 \cup \Delta_3$  then

$$(-\mu_1^* + \mu_2^* + 2\mu_3^*)(\Delta) = \frac{1}{2\pi i} \oint_{\Delta(1)} \sqrt{-\varpi} = \frac{1}{2\pi i} \oint_{\Delta} (\xi_2(s) - \xi_3(s))ds, \quad (3.34)$$

where the contour integral is understood as a small loop encircling  $\Delta$  in the clockwise direction, and separating it from the other components.

*Proof.* Measures  $\mu_1^*$  and  $\mu_2^*$  are positive because they are restrictions of a positive measure  $\lambda$  (see Equation (3.32)). Also, from Proposition 3.15 we learn that the density of  $\mu_3^*$ , whenever  $\mu_3^* \neq 0$ , is real and does not change sign on  $\Delta_3 \cap \mathbb{H}_+$ , so in order to prove that  $\mu_3^*$  is positive on  $\Delta_3 \cap \mathbb{H}_+$  it is enough to show that this density is positive near  $y_*$ . To see the latter, denote

$$\Phi(z) = \int_{y_*}^z (\xi_2(s) - \xi_3(s))ds, \quad z \in \mathbb{H}_+ \setminus \Delta_3,$$

This function  $\Phi$  is holomorphic in its domain of definition, and according to Proposition 3.15 it satisfies

$$\Phi(z) \in \mathbb{R} \text{ when } z \in \tau_* \cap \mathbb{H}_+, \quad \text{and} \quad \Phi(z) \in i\mathbb{R} \text{ when } z \in \Delta_3 \cap \mathbb{H}_+. \quad (3.35)$$

Now, the contour  $\tau_*$ , by its definition, extends to  $\infty$  with angle 0 (see Figure 3). From the asymptotics (3.1) we also get that

$$\Phi(z) = 2az(1 + o(1)), \quad z \rightarrow \infty,$$

which shows that actually  $\Phi(z) > 0$  for  $z \in \tau_* \cap \mathbb{H}_+$  in a neighborhood of  $\infty$ . Because  $\xi_2 - \xi_3$  has no zeros along  $\tau_* \cap \mathbb{H}_+$  (check again Proposition 3.15), we actually get that

$$\Phi(\tau_* \cap \mathbb{H}_*) = (0, +\infty).$$

At  $z = y_*$ , the difference  $\xi_2 - \xi_3$  vanishes as square root, so  $\Phi$  vanishes with power  $3/2$  at  $z = y_*$ . To rotate from  $\tau_*$  to  $\Delta_3$  at  $y_*$  in the counter-clockwise direction, we change the angle by  $\pi$ , so in the image of  $\Phi$  this becomes a rotation by  $3\pi/2$  in the counter-clockwise direction. Combining with (3.35), this means that for some  $\delta > 0$ ,

$$\Phi_+(z) \in i\mathbb{R}_-, \quad z \in D_\delta(y_*) \cap \Delta_3.$$

This is equivalent to saying that  $d\mu_3^*$  is positive along  $D_\delta(y_*) \cap \mathbb{H}_+$ , and by the previous observation we consequently conclude that  $\mu_3^*$  is positive everywhere on  $\Delta_3 \cap \mathbb{H}_+$ .

The proof that  $\mu_3^*$  is positive on  $\Delta_3 \cap \mathbb{H}_-$  follows along the same lines, or alternatively by symmetry under complex conjugation. We omit the details.

The first equation in (3.33) is easily derived once we observe that  $\mu_1^* + \mu_2^* = \lambda$  and recall (2.8). To get the second one, fix  $x \in \mathbb{C} \setminus (\Delta_1 \cup \Delta_2 \cup \Delta_3)$ . For a contour  $\gamma$  enclosing  $\Delta_1 \cup \Delta_2 \cup \Delta_3$  in the counter-clockwise orientation and with  $z$  in its exterior, deformation of contours shows that

$$\begin{aligned} C^{\mu_1^*}(z) - C^{\mu_3^*}(z) &= \frac{1}{2\pi i} \int_{\Delta_1} \frac{\xi_{1+}(s) - \xi_{1-}(s)}{s - z} ds - \frac{1}{2\pi i} \int_{\Delta_3} \frac{\xi_{2+}(s) - \xi_{3+}(s)}{s - z} ds \\ &= \frac{1}{2\pi i} \int_{\Delta_1} \frac{\xi_{2-}(s) - \xi_{2+}(s)}{s - z} ds + \frac{1}{2\pi i} \int_{\Delta_3} \frac{\xi_{2-}(s) - \xi_{2+}(s)}{s - z} ds \\ &= \frac{1}{2\pi i} \oint_{\gamma} \frac{\xi_2(s)}{s - z} ds, \end{aligned}$$

and by the residues theorem and the expansion (3.1),

$$C^{\mu_1^*}(z) - C^{\mu_3^*}(z) = -\text{Res} \left( \frac{\xi_2(s)}{s - z}, s = z \right) - \text{Res} \left( \frac{\xi_2(s)}{s - z}, s = \infty \right) = -\xi_2(z) + a,$$

which is equivalent to the second equation in (3.33). For the third equation, simply add the first two and recall that  $\xi_1 + \xi_2 = V' - \xi_3$ , see (3.3).

The relations between the masses then follow from (3.33) and the expansion (3.1), having in mind that for any finite complex measure  $\sigma$  with compact support, the asymptotics

$$C^\sigma(z) = -\frac{\sigma(\mathbb{C})}{z} + \mathcal{O}(z^{-2}), \quad z \rightarrow \infty,$$

holds.

Finally, (3.34) is obtained using the explicit expressions (3.33) and interchanging the order of integrals in the right hand side.  $\square$

Set

$$\vec{\mu}_* := (\mu_1^*, \mu_2^*, \mu_3^*)$$

with components being the measures in (3.32), and let  $\Gamma_*$  be as in Propositions 3.15 and 3.18. The fact that  $\vec{\mu} \in \mathcal{M}(\Gamma_*)$  is an immediate consequence of Proposition 3.31.

We now show that  $\vec{\mu}_*$  is critical. The transformation  $\xi \mapsto \zeta = \xi - \frac{V'}{3}$  transforms (2.6) in an equation of the form

$$\zeta^3 - R(z)\zeta + D(z) = 0,$$

for some polynomials  $R$  and  $D$ . The solutions to this equation are

$$\begin{aligned} \zeta_1(z) &= \xi_1(z) - \frac{V'(z)}{3} = C^{\mu_1}(z) + C^{\mu_2}(z) + Q'_1(z), & Q_1(z) &:= \frac{2V(z)}{3}, \\ \zeta_2(z) &= \xi_3(z) - \frac{V'(z)}{3} = -C^{\mu_2}(z) - C^{\mu_3}(z) + Q'_2(z), & Q_2(z) &:= -az - \frac{V(z)}{3}, \\ \zeta_3(z) &= \xi_2(z) - \frac{V'(z)}{3} = C^{\mu_3}(z) - C^{\mu_1}(z) + Q'_3(z), & Q_3(z) &:= az - \frac{V(z)}{3}. \end{aligned}$$

By [76, Theorem 1.3] the vector of measures  $(\mu_2^*, \mu_1^*, \mu_3^*)$  is critical for the energy with interaction matrix

$$\begin{pmatrix} 1 & 1/2 & 1/2 \\ 1/2 & 1 & -1/2 \\ 1/2 & -1/2 & 1 \end{pmatrix}$$

and potentials

$$\begin{aligned} \Phi_1(z) &= Q_1(z) - Q_2(z) = V(z) + az = V_2(z), \\ \Phi_2(z) &= Q_1(z) - Q_3(z) = V(z) - az = V_1(z), \\ \Phi_3(z) &= Q_3(z) - Q_2(z) = 2az = V_3(z). \end{aligned}$$

Interchanging the roles of  $\mu_1^*$  and  $\mu_2^*$ , we thus get that  $\vec{\mu}_* = (\mu_1^*, \mu_2^*, \mu_3^*)$  is critical for our energy  $E_S(\cdot)$ .

The  $S$ -property (2.17) is an immediate consequence of the fact that  $\vec{\mu}_*$  is critical, see [76, Theorem 1.8].

Finally, the classification of possible singular behaviors of  $\lambda$  follows from the sheet structure for  $\xi_1$  and (3.33).

For a Borel measure  $\mu$  on  $\mathbb{C}$  its logarithmic potential  $U^\mu$  and its Cauchy transform  $C^\mu$  are related through

$$\frac{\partial U^\mu}{\partial z}(z) = \frac{1}{2}C^\mu(z), \quad z \in \mathbb{C} \setminus \text{supp } \mu; \quad (3.36)$$

under suitable assumptions on  $\mu$  and its support, this identity can be further extended to boundary points of  $\text{supp } \mu$  through appropriate limits, as well as to  $\text{supp } \mu$  in the distributional sense. Using this relation, an immediate consequence of (3.33) are the following expressions:

**Lemma 3.32.** *Let  $y \in \Delta_1 \cup \Delta_2 \cup \Delta_3$ . There are real constants  $l_1, l_2$  and  $l_3$ , possibly depending on  $y$ , such that for  $z \in \mathbb{C} \setminus (\mathbb{R} \cup \Delta_3)$ , the logarithmic potentials of the measures  $\mu_1^*$ ,  $\mu_2^*$  and  $\mu_3^*$  satisfy*

$$\begin{aligned} U^{\mu_1^*}(z) + U^{\mu_2^*}(z) + \operatorname{Re} V(z) &= \operatorname{Re} \int_y^z \xi_1(s) ds + l_1, \\ U^{\mu_3^*}(z) - U^{\mu_1^*}(z) + a \operatorname{Re} z &= \operatorname{Re} \int_y^z \xi_2(s) ds + l_2, \\ U^{\mu_2^*}(z) + U^{\mu_3^*}(z) + a \operatorname{Re} z &= -\operatorname{Re} \int_y^z \xi_3(s) ds + l_3. \end{aligned}$$

We can finally establish the following result, which completes the proof of Theorem A.

**Proposition 3.33.** *The components of the vector critical measure  $\vec{\mu}_*$  satisfy the variational identities (2.16) from Theorem A. In particular, there exists a real constant  $\ell_3$  such that in the saturated case,*

$$2U^{\mu_3^*}(z) - U^{\mu_1^*}(z) + U^{\mu_2^*}(z) + \phi_3(z) = \ell_3, \quad z \in \Delta_3. \quad (3.37)$$

Additionally,

i) *if for any closed loop  $\gamma$  on the Riemann surface  $\mathcal{R}$ , visiting the sheets  $\mathcal{R}_1$  and  $\mathcal{R}_2$ ,*

$$\operatorname{Re} \oint_{\gamma} \sqrt{-\varpi} = 0, \quad (3.38)$$

*then there exists a real constant  $\ell_1$  such that*

$$2U^{\mu_1^*}(z) + U^{\mu_2^*}(z) - U^{\mu_3^*}(z) + \phi_1(z) = \ell_1, \quad z \in \Delta_1. \quad (3.39)$$

ii) *if for any closed loop  $\gamma$  on the Riemann surface  $\mathcal{R}$ , visiting the sheets  $\mathcal{R}_1$  and  $\mathcal{R}_3$ ,*

$$\operatorname{Re} \oint_{\gamma} \sqrt{-\varpi} = 0, \quad (3.40)$$

*then there exists a real constant  $\ell_2$  such that*

$$2U^{\mu_2^*}(z) + U^{\mu_1^*}(z) + U^{\mu_3^*}(z) + \phi_2(z) = \ell_2, \quad z \in \Delta_2. \quad (3.41)$$

Finally, if both conditions i) and ii) are satisfied, then  $\ell_1 = \ell_2$ .

Recall that

$$\phi_1(z) = \operatorname{Re}(V(z) - az), \quad \phi_2(z) = \operatorname{Re}(V(z) + az), \quad \phi_3(z) = 2a \operatorname{Re}(z).$$

*Proof.* Let  $p \in \Delta_1$  be any point on  $\Delta_1$ . Then Lemma 3.32, there exist real constants  $l_1, l_2$  such that for  $z \in \mathbb{H}_+ \setminus \Delta_3$ ,

$$2U^{\mu_1^*}(z) + U^{\mu_2^*} - U^{\mu_3^*}(z) + \phi_1(z) = \operatorname{Re} \int_p^z (\xi_1(s) - \xi_2(s)) ds + l_1 - l_2.$$

The assertion that the right hand side remains constant along the connected component of  $\Delta_1$  that contains  $p$  is a direct consequence of (3.32). Furthermore, if  $q \in \Delta_1$  is a point on a different connected component of  $\Delta_1$ , then

$$\left(2U^{\mu_1^*} + U^{\mu_2^*} - U^{\mu_3^*} + \phi_1\right)(q) = \left(2U^{\mu_1^*} + U^{\mu_2^*} - U^{\mu_3^*} + \phi_1\right)(p) + \operatorname{Re} \int_p^q (\xi_1(s) - \xi_2(s))ds,$$

where we integrate along any path in  $\mathbb{H}_+ \setminus \Delta_3$ . But the integral in the right hand side above can be written as in (3.38), so that condition (3.38) implies equality of the left hand side in (3.39) along the whole set  $\Delta_1$  (with the same constant). The remaining variational equalities follow in a similar manner. We omit the details.  $\square$

*Remark 3.34.* The condition that all the periods of  $\xi dz$  are purely imaginary (compare with conditions (3.38), (3.40)) is commonly called as the *Boutroux condition* for the spectral curve. It has appeared several times in the past as transcendental conditions to determine the associated spectral curve of a matrix model, see for instance [17, 68, 26, 27].

*Proof of Theorem B.* If  $x_0$  is an endpoint of  $\operatorname{supp} \lambda$ , then by Theorem 3.25 (a) only  $\xi_1$  and exactly one of  $\xi_2$  or  $\xi_3$  can be branched. Having in mind that this density is a multiple of  $\xi_{1+} - \xi_{j+}$ , the expansion (2.20) has to be valid.

If  $x_0$  is a point in the interior of  $\operatorname{supp} \lambda$ , again by Theorem 3.25 either none of the functions  $\xi_j$  is branched, leading to (2.18) or the three are branched at the unique point  $x_*$ , giving (2.19).  $\square$

#### 4. FROM CRITICAL MEASURES TO CONSTRAINED EQUILIBRIUM

Our goal in this section is to compare the constrained equilibrium problem of Bleher, Delvaux and Kuijlaars [31] with our vector critical measure  $\vec{\mu}_* = (\mu_1^*, \mu_2^*, \mu_3^*)$ .

In [31], the authors work under the symmetry assumptions  $\alpha = 1/2$  and  $V(z) = V(-z)$ . These in turn imply additional symmetries on the spectral curve they found. These symmetries are summarized in the next definition.

**Definition 4.1.** We say a spectral curve as in Definition 2.1 is *symmetric* if

$$F(-\xi, -z) = -F(\xi, z), \quad \xi, z \in \mathbb{C}, \quad (4.1)$$

or in other words, if the coefficients  $p_0, p_1$  and  $p_2$  satisfy  $p_k(-z) = (-1)^{k+1}p_k(z)$ ,  $k = 0, 1, 2$ .

The spectral curve found in [31] is admissible (in the sense of Definition 2.1) and symmetric; although the coefficients  $p_0$  and  $p_1$  are not explicit, their expressions in terms of the constrained equilibrium measure are enough to establish this fact. Furthermore, the corresponding Riemann surface  $\mathcal{R}$  for  $F(\xi, z) = 0$  was also described in [31], and the Boutroux condition, although not explicitly mentioned there, follows easily from their construction of the spectral curve from the constrained equilibrium problem.

To be a little more precise, in the aforementioned work the authors actually describe  $\mathcal{R}$  under the additional assumption that their constrained vector equilibrium problem is regular (not to be



confused with our notion of regular saturated regime!). Their notion of regularity only affects their asymptotic analysis, and consequently their proof of convergence for the limiting eigenvalue distribution is only valid in that case. But their result on the *existence* of the spectral curve for a symmetric potential, and the geometry of  $\mathcal{R}$ , is valid even for their singular cases.

As an outline of this section, we will study in depth the symmetric spectral curves (4.1), extracting from the corresponding critical measures  $(\mu_1^*, \mu_2^*, \mu_3^*)$ , whose existence is now established, a new pair of measures  $(\nu_1, \nu_2)$ . This construction will be based in the geometry of the Riemann surface  $\mathcal{R}$  already established in this paper, combined with techniques from potential theory. At the end of the section, we then verify that if the symmetric spectral curve under consideration is the one in [31], then this new pair  $(\nu_1, \nu_2)$  reduces to the solution of the constrained equilibrium problem they considered.

#### 4.1. Geometry of trajectories.

For a symmetric spectral curve,  $V(z) = -V(-z)$ , and from the expansion of  $\xi_2$  and  $\xi_3$  given in (3.1) we thus get that  $\alpha = 1/2$  in this case. Another immediate consequence of this definition is that the critical graph of the quadratic differential  $\varpi$  is symmetric with respect to the imaginary axis, and Propositions 3.13, 3.15 and 3.18 are supplemented by the following one (see also the notation in Definition 3.12):

**Proposition 4.2.** *Suppose the spectral curve (2.6) is symmetric. Then the trajectories of  $\varpi$  are symmetric with respect to the imaginary axis. In particular,  $\operatorname{Re} y_* = 0$ . In the saturated regime,  $y_* \in i\mathbb{R}_+$ ,  $x_* < 0$ ,  $(\Delta_3 \setminus \{-y_*, y_*\}) \subset \mathbb{C}_-$ , and*

$$(\xi_2(z) - \xi_3(z))dz \begin{cases} \in \mathbb{R} & \text{along the vertical segment } (-y_*, y_*), \\ \in i\mathbb{R} & \text{along } i\mathbb{R} \setminus [-y_*, y_*]. \end{cases} \quad (4.2)$$

Moreover, in the variational identity (3.37) of Proposition 3.33,

$$\ell_3 = 0, \quad (4.3)$$

whenever (3.38) and (3.40) are met.

*Proof.* We invite the reader to take a look at Figures 15 and 16, where the trajectories and orthogonal trajectories of relevance are updated for the symmetric situation considered here. The proof of almost all assertions is straightforward; we only establish (4.3). In what follows, it is convenient to define

$$y_* = 0$$

in the unsaturated regime, in order to handle both cases simultaneously.

By Lemma 3.32, for  $z \in \mathbb{H}_+ \setminus \Delta_3$ ,

$$2U^{\mu_3^*}(z) - U^{\mu_1^*}(z) + U^{\mu_2^*}(z) + 2a \operatorname{Re} z = \ell_3 + \operatorname{Re} \int_{y_*}^z (\xi_2(s) - \xi_3(s)) ds.$$

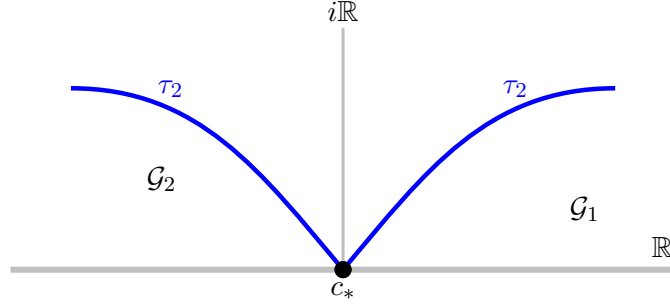


FIGURE 15. Symmetric unsaturated regime: The orthogonal trajectories that determine the half plane domain  $\mathcal{H}$ . This is an update version, taking into account symmetry, of previous figures, compare for instance with Figure 12.

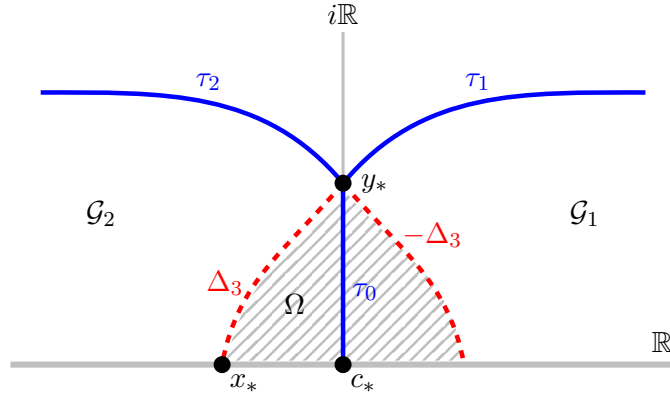


FIGURE 16. Symmetric saturated regime: The orthogonal trajectories on  $\mathbb{H}_+$  that emanate from  $y_*^{(1)}$  and the trajectory that we defined to be  $\Delta_3$ . For convenience, we also added the set  $-\Delta_3$ . The domain  $\Omega$  is the whole shaded region, so in particular  $\tau_0 \subset \Omega$ . This is an update version, taking into account symmetry, of previous figures, compare for instance with Figures 2, 3 and 11.

Taking into account (4.2),

$$2U^{\mu_3^*}(z) - U^{\mu_1^*}(z) + U^{\mu_2^*}(z) = \ell_3, \quad z \in i\mathbb{R}_+ \setminus [0, y_*]. \quad (4.4)$$

Recall that for a (signed) compactly supported Borel measure  $\mu$  on  $\mathbb{C}$ ,

$$U^\mu(z) = -|\mu| \log |z| + o(1), \quad z \rightarrow \infty. \quad (4.5)$$

Notice that for  $\alpha = 1/2$ ,  $(2\mu_3^* - \mu_1^* + \mu_2^*)$  is a neutral measure. Thus, making  $z \rightarrow +i\infty$  in (4.4) we get (4.3).  $\square$

For a symmetric spectral curve,  $\xi(z)$  is a solution if and only if  $-\xi(-z)$  is a solution also. Taking into account the branch cuts for  $\xi_1$  we have that

$$\xi_1(-z) = -\xi_1(z), \quad z \in \mathbb{C} \setminus \text{supp } \lambda. \quad (4.6)$$

In the unsaturated case, when  $[a_k, b_k]$ ,  $k = 1, \dots, l$ , are the only branch cuts for  $\xi_2$  and  $\xi_3$ , we also have

$$\xi_2(-z) = -\xi_3(z), \quad z \in \mathbb{C} \setminus \text{supp } \lambda. \quad (4.7)$$

This implies, in particular, that for the unsaturated regime the picture is totally symmetric:  $x \in \Delta_1$  if and only if  $-x \in \Delta_2$ , so that in particular,  $|\mu_1| = |\mu_2| = 1/2$ . In fact, as in this case  $\partial\mathcal{H}$  has to be symmetric w.r.t. the imaginary axis, its unique boundary critical point  $c_* = x_*$  has to be at the origin. This is the point of separating  $\Delta_1$  and  $\Delta_2$  in Proposition 3.28, so

$$\Delta_1 \subset \mathbb{C}_+, \quad \Delta_2 \subset \mathbb{C}_-.$$

In the saturated regime we have two additional branch points,  $\pm y_* \in i\mathbb{R}$ , and an additional branch cut  $\Delta_3$  for  $\xi_2 - \xi_3$ . This means that (4.7) is now valid only outside of the bounded domain, say  $\Omega$  (see Figure 16), delimited by  $\Delta_3$  and  $-\Delta_3 = \{-z : z \in \Delta_3\}$ . Inside such a domain, we have

$$\xi_2(-z) = -\xi_2(z), \quad \xi_3(-z) = -\xi_3(z), \quad z \in \mathbb{C} \setminus \text{supp } \lambda.$$

As a consequence, we do have the symmetry outside  $\Omega$ :  $x \in \Delta_1 \setminus \Omega$  if and only if  $-x \in \Delta_2 \setminus \Omega$ . But the asymmetry of  $\Delta_3$  implies that now

$$\Delta_2 \cap \mathbb{C}_+ = \emptyset \quad \text{and} \quad \Delta_1 \cap \mathbb{C}_- \subset \Omega.$$

In addition, certainly  $\Delta_1 \cap \mathbb{C}_- \neq \emptyset$ , otherwise the union of trajectories  $\Delta_3 \cup (-\Delta_3)$  would encircle a bounded domain, symmetric for  $\mathbb{R}$ , that violates the Principle **P3**.

On a connected component  $[a_k, b_k]$  of  $\text{supp } \lambda$  that intersects  $\Delta_3$  we have that it carries  $\mu_2$  on the left of the intersection point  $x_*$ , and  $\mu_1$  on the right.

#### 4.2. Construction of the constrained measure.

In the saturated regime,  $\Delta_3 \cup [-y_*, y_*]$  enclose a bounded domain that we call  $\mathcal{H}_3$ . In the unsaturated regime, we set  $\mathcal{H}_3 := \emptyset$ . Let also

$$\mathcal{H}_1 := \mathbb{C}_- \setminus \overline{\mathcal{H}_3}, \quad \mathcal{H}_2 := \mathbb{C}_+,$$

see the notation in (2.1). In the saturated regime, the sets  $\mathcal{H}_1, \mathcal{H}_2$  and  $\mathcal{H}_3$  are shown in Figure 17.

With

$$\begin{aligned} H_1(z) &:= U^{\mu_1^*}(z) - U^{\mu_3^*}(z), \\ H_2(z) &:= U^{\mu_2^*}(z) + U^{\mu_3^*}(z), \\ H_3(z) &:= U^{\mu_2^*}(z) + U^{\mu_3^*}(z) + 2a \operatorname{Re} z, \end{aligned} \quad (4.8)$$

(in the unsaturated regime, only  $H_1$  and  $H_2$  are relevant), let  $H : \mathcal{H}_1 \cup \mathcal{H}_2 \cup \mathcal{H}_3 \rightarrow \mathbb{R}$  be defined by

$$H|_{\mathcal{H}_j} := H_j, \quad j = 1, 2, 3. \quad (4.9)$$

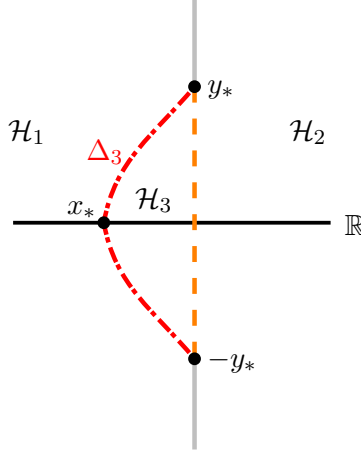


FIGURE 17. Saturated regime: the partition of the plane into the domains  $\mathcal{H}_1$ ,  $\mathcal{H}_2$  and  $\mathcal{H}_3$  used in the definition of the function  $H$  in (4.9).

By the discussion above,  $\Delta_1 \subset \overline{\mathcal{H}_2 \cup \mathcal{H}_3}$ ,  $\Delta_2 \subset \overline{\mathcal{H}_1}$ , so that  $H$  is continuous and harmonic in  $\mathcal{H}_1 \cup \mathcal{H}_2 \cup \mathcal{H}_3$ .

**Lemma 4.3.** *The function  $H$  extends continuously to  $\mathbb{C}$ .*

*Proof.* Observe that

$$H_3(z) - H_2(z) = 2a \operatorname{Re} z = \phi_3(z), \quad (4.10)$$

and  $\operatorname{Re} z = 0$  along  $[-y_*, y_*]$ , so that  $H$  is continuous across this segment. Furthermore,

$$H_3(z) - H_1(z) = 2U^{\mu_3^*}(z) + U^{\mu_2^*}(z) - U^{\mu_1^*}(z) + \phi_3(z),$$

and by (4.3) and Proposition 3.33 this is zero along  $\Delta_3$ :  $H$  is continuous across  $\Delta_3$  as well.

Finally, for  $z \in i\mathbb{R} \setminus [-y_*, y_*]$ ,

$$\begin{aligned} H_2(z) - H_1(z) &= 2U^{\mu_3^*}(z) + U^{\mu_2^*}(z) - U^{\mu_1^*}(z) \\ &= 2U^{\mu_3^*}(z) + U^{\mu_2^*}(z) - U^{\mu_1^*}(z) + \phi_3(z) = 0, \end{aligned} \quad (4.11)$$

where we have used again (4.3) and Proposition 3.33. Continuity along  $i\mathbb{R} \setminus [-y_*, y_*]$  is established also.  $\square$

**Lemma 4.4.** *Suppose that we are in the saturated regime. Then the function  $H$  admits a harmonic continuation through  $\Delta_3$ .*

*Proof.* According to Lemma 4.3,  $H$  is continuous across  $\Delta_3$ , so it is enough to verify that

$$\frac{\partial H}{\partial n_+}(z) = -\frac{\partial H}{\partial n_-}(z), \quad z \in \Delta_3 \setminus \{\pm y_*\},$$

for  $z$  along  $\Delta_3$ , where  $n_{\pm}$  are the normal vectors to  $\Delta_3$  at  $z$ . This identity follows immediately from the explicit expression for  $H$  in  $\mathcal{H}_1$  and  $\mathcal{H}_3$ , combined with the  $S$ -property (2.17).  $\square$

Our final ingredient is the following

**Lemma 4.5.** *There exists a neighborhood  $\Omega_2$  of  $i\mathbb{R} \setminus [-y_*, y_*]$  such that*

$$\min(H_1(z), H_2(z)) = H_j(z), \quad z \in \Omega_2 \cap \mathcal{H}_j, \quad j = 1, 2. \quad (4.12)$$

*Moreover, in the saturated regime there exists a neighborhood  $\Omega_1$  of  $(-y_*, y_*)$  such that*

$$\min(H_2(z), H_3(z)) = H_j(z), \quad z \in \Omega_1 \cap \mathcal{H}_j, \quad j = 2, 3. \quad (4.13)$$

*Proof.* Clearly, (4.13) follows immediately from (4.10). So, let us prove (4.12).

Recalling (4.8) and that in the present case  $\ell_3 = 0$ , we proceed as before (see Lemma 3.32) and get that for  $z$  in a neighborhood of  $i\mathbb{R} \setminus [-y_*, y_*]$ ,

$$\begin{aligned} H_2(z) - H_1(z) &= 2U^{\mu_3^*}(z) - U^{\mu_1^*}(z) + U^{\mu_2^*}(z) = \operatorname{Re} \int_{y_*}^z (\xi_2(s) - \xi_3(s)) ds - \phi_3(z) \\ &= \operatorname{Re} \int_{y_*}^z (\xi_2(s) - \xi_3(s) - 2a) ds, \end{aligned} \quad (4.14)$$

where in the last step we used the explicit form of  $\phi_3$ . This shows that in order to conclude the proof we need to analyze the structure of trajectories on  $\mathcal{R}_1$  of the following auxiliary quadratic differential, defined by

$$\tilde{\varpi} = \begin{cases} -(\xi_2(z) - \xi_3(z) - 2a)^2 dz^2, & \text{on } \mathcal{R}_1, \\ -(\xi_3(z) - \xi_1(z) - 2a)^2 dz^2, & \text{on } \mathcal{R}_2, \\ -(\xi_1(z) - \xi_2(z) - 2a)^2 dz^2, & \text{on } \mathcal{R}_3. \end{cases} \quad (4.15)$$

From the expansion (3.1) it follows that  $\tilde{\varpi}$  is either regular or has a zero at  $\infty^{(1)}$ , whose order depends on the first nonzero term in the error term in (3.1). Furthermore, since  $a \in \mathbb{R}$ , arcs of orthogonal trajectories of  $\varpi$  and  $\tilde{\varpi}$  on  $\mathbb{R}$  coincide, as well as arcs of trajectories on  $i\mathbb{R}$ .

We claim that  $\tilde{\varpi}$  does not vanish on  $\mathbb{H}_+^{(1)}$ . The proof follows the ideas already used to establish Proposition 3.9. Suppose that there is a zero, say  $p^{(1)} \in \mathbb{H}_+^{(1)}$ , of  $\tilde{\varpi}$ . Emanating from it, there are at least three distinct orthogonal trajectories, and because  $\infty^{(1)}$  is either a regular point or a zero of  $\varpi$ , at most one of these orthogonal trajectories diverges to  $\infty^{(1)}$ , in the asymptotically vertical direction (see Principle **P5** in Section 3.3).

This means that there is a path  $\gamma \subset \mathbb{H}_+^{(1)}$ , comprised of a finite union of orthogonal critical trajectories, with endpoints  $c$  and  $d$  on the real axis; without loss of generality, we assume that there is no other zero of  $\tilde{\varpi}$  in  $(c, d)^{(1)}$ . If we assume that  $(c, d) \cap \operatorname{supp} \lambda = \emptyset$ , we obtain that the union of  $\gamma$  and of its reflection with respect to  $\mathbb{R}$  enclose a bounded domain on  $\mathcal{R}_1$ , contradicting Principle **P3**. Thus, the other option to discard is when  $[c, d] \subset \operatorname{supp} \lambda$ . Since necessarily

$$\int_c^d (\xi_{1-}(x) - \xi_{1+}(x) - 2a) dx \in \mathbb{R},$$

and because  $a$  is real, we get that

$$\int_c^d (\xi_{1-}(x) - \xi_{1+}(x)) dx \in \mathbb{R},$$

which leads to a contradiction with (3.12).

The just established fact implies that

$$h(z) := \operatorname{Im} \int_{y_*}^z (\xi_2(s) - \xi_3(s) - 2a) ds < 0, \quad z \in i\mathbb{R}_+ \setminus [0, y_*]. \quad (4.16)$$

Indeed, notice that on  $i\mathbb{R}_+$  we could have defined equivalently

$$h(z) = \frac{1}{i} \int_{y_*}^z (\xi_2(s) - \xi_3(s) - 2a) ds.$$

By definition  $h(y_*) = 0$ , and for  $x > 0$

$$h'(ix) = -2ax + (\xi_2 - \xi_3)(ix).$$

Since  $(\xi_2 - \xi_3)(y_*) = 0$  (see Propositions 3.15 and 3.18), we have that  $h'(y_*) < 0$ , so that  $h$  is strictly decreasing on  $i\mathbb{R}_+ \setminus [0, y_*]$  in a neighborhood of  $y_*$ . Since  $\tilde{\omega}$  does not vanish on  $\mathbb{H}_+^{(1)}$ , this means that  $h'(y) \neq 0$  there, thus proving (4.16).

We can return now to (4.14). The real part of the integral in the right hand side is 0 along  $[y_*, i\infty)$ , while the imaginary part is  $h(z)$ , and we just proved that it is strictly negative there. Since conformal maps preserve orientation, we conclude that  $H_2(z) - H_1(z) > 0$  for  $z$  immediately to the left of  $[y_*, i\infty)$  and  $H_2(z) - H_1(z) < 0$  immediately to the right, concluding the proof of (4.12).  $\square$

Our main conclusion is that  $H$  is actually superharmonic on  $\mathbb{C}$ , so up to an additive harmonic term, it is a logarithmic potential of a positive measure. We can say more:

**Theorem 4.6.** *There exists a positive measure  $\nu_2$  with  $|\nu_2| = 1/2$ ,  $\operatorname{supp} \nu_2 \subset i\mathbb{R}$ , for which*

$$U^{\nu_2}(z) = H(z), \quad z \in \mathbb{C}.$$

*The measure  $\nu_2$  is absolutely continuous with respect to the arc length on  $i\mathbb{R}$ , has continuous density and*

$$\begin{aligned} \nu_2 &= \sigma, & \text{on } [-y_*, y_*], \\ \nu_2 &< \sigma, & \text{on } i\mathbb{R} \setminus [-y_*, y_*], \end{aligned} \quad (4.17)$$

*where  $\sigma$  is given in (2.21).*

*Proof.* A combination of Lemmas 4.3 and 4.4 tells us that  $H$  is harmonic in  $\mathbb{C} \setminus i\mathbb{R}$ , and by Lemma 4.5, it is superharmonic in a neighborhood of  $i\mathbb{R}$ .

From the Riesz Decomposition Theorem [84, Theorem II.3.1, page 100] there exist a measure  $\nu_2$  and a harmonic function  $u : \mathbb{C} \rightarrow \mathbb{R}$  such that

$$H(z) = u(z) + U^{\nu_2}(z), \quad z \in \mathbb{C}, \quad \operatorname{supp} \nu_2 \subset i\mathbb{R}.$$

From the definition of  $H$ , the relations between the masses of  $\mu_1^*$ ,  $\mu_2^*$  and  $\mu_3^*$  and from (4.5) it follows that

$$u(z) = \left( \frac{1}{2} - |\nu_2| \right) \log \frac{1}{|z|} + o(1), \quad z \rightarrow \infty.$$

Since  $u$  is harmonic on  $\mathbb{C}$ , from the maximum principle it follows that  $|\nu_2| = 1/2$  and  $u \equiv 0$ .

The partial derivatives of the potential  $H = U^{\nu_2}$  are continuous up to the boundary of each of the domains  $\mathcal{H}_1$ ,  $\mathcal{H}_2$  and  $\mathcal{H}_3$ , so from [84, Thm. 1.5, p. 92] we get that  $\nu_2$  is absolutely continuous with continuous density.

For (4.17) we use the definition of the functions  $H_j$ 's in (4.8) and also (4.9) to express

$$\begin{aligned} \frac{\partial U_+^{\nu_2}}{\partial z}(z) - \frac{\partial U_-^{\nu_2}}{\partial z}(z) &= \frac{\partial H_+}{\partial z}(z) - \frac{\partial H_-}{\partial z}(z) \\ &= \begin{cases} 2a \frac{\partial}{\partial z} \operatorname{Re} z, & z \in (-y_*, y_*), \\ -\frac{\partial}{\partial z} \left( U^{\mu_3^*}(z) + U^{\mu_2^*}(z) - U^{\mu_1^*}(z) \right), & z \in i\mathbb{R} \setminus [-y_*, y_*], \end{cases} \end{aligned}$$

where we oriented  $i\mathbb{R}$  from  $-i\infty$  to  $i\infty$ . In virtue of Sokhotsky-Plemelj's formula and (3.36),

$$\frac{d\nu_2}{ds}(z) = \frac{1}{2\pi i} (C_+^{\nu_2}(z) - C_-^{\nu_2}(z)) = \frac{1}{\pi i} \frac{\partial}{\partial z} (U_+^{\nu_2}(z) - U_-^{\nu_2}(z)) = \frac{a}{\pi}, \quad z \in (-y_*, y_*).$$

In particular, this leads to the first equation in (4.17).

Similarly,

$$\begin{aligned} \frac{d\nu_2}{ds}(z) &= \frac{1}{2\pi i} (-2C^{\mu_3^*}(z) - C^{\mu_2^*}(z) + C^{\mu_1^*}(z)) \\ &= -\frac{1}{2\pi i} (\xi_2(z) - \xi_3(z) - 2a), \quad z \in i\mathbb{R} \setminus [-y_*, y_*]. \end{aligned}$$

which can be rearranged to

$$\frac{d\sigma}{ds}(z) - \frac{d\nu_2}{ds}(z) = \frac{1}{2\pi i} (\xi_2(z) - \xi_3(z)), \quad z \in i\mathbb{R} \setminus [-y_*, y_*].$$

In virtue of Proposition 3.9, this identity means that the density of the difference  $\sigma - \nu_2$  does not vanish on  $i\mathbb{R} \setminus [-y_*, y_*]$ . Also, because  $\nu_2$  is finite, its density has to vanish as  $z \rightarrow \pm i\infty$ . This is enough to conclude that the density of  $\sigma - \nu_2$  is always positive on  $i\mathbb{R} \setminus [-y_*, y_*]$ , giving us the inequality in (4.17) and concluding the proof.  $\square$

As we mentioned, in [31] the authors proved the existence of a symmetric spectral curve for any even potential  $V$  and  $\alpha = 1/2$ . With  $\bar{\nu}^* = (\nu_1^*, \nu_2^*)$  being the constrained equilibrium measure discussed at the beginning of (2.2.1), the measure  $\lambda$  in (2.8) is

$$\lambda = \nu_1^*. \tag{4.18}$$

In addition, their spectral curve also satisfies the Boutroux conditions (3.38) and (3.40), and the variational conditions for the constrained equilibrium problem say that  $\nu = \nu_2^*$  is the unique measure supported on  $i\mathbb{R}$  with  $|\nu| = 1/2$ ,  $\nu \leq \sigma$ , and for which

$$2U^\nu(z) - U^{\nu_1^*}(z) = 0, \quad z \in \text{supp}(\sigma - \nu), \quad 2U^\nu(z) - U^{\nu_1^*}(z) \leq 0, \quad z \in i\mathbb{R}. \quad (4.19)$$

We now compare their constrained equilibrium measure  $\bar{\nu}^*$  with our construction in the symmetric case, that is, we now prove Theorem C.

*Proof of Theorem C.* So from now on, let us assume we are working with the symmetric spectral curve given in [31], denoting their constrained equilibrium measure by  $\bar{\nu}^* = (\nu_1^*, \nu_2^*)$ , and our pair of measures by  $\nu_1 = \mu_1^* + \mu_2^* = \lambda$  and  $\nu_2$ , the latter given in Theorem 4.6. From (4.18), we immediately get

$$\nu_1^* = \lambda = \mu_1^* + \mu_2^* = \nu_1,$$

and our goal is thus to show that  $\nu_2^* = \nu_2$ .

To do so, let us start by summarizing our findings in the symmetric situation so far: we have a measure  $\nu_2$  on  $i\mathbb{R}$ , satisfying the constraint (4.17), and such that by its construction and (4.8),

$$\begin{aligned} U^{\nu_2}(z) &= U^{\mu_1^*}(z) - U^{\mu_3^*}(z), \quad z \in \mathcal{H}_1, \\ U^{\nu_2}(z) &= U^{\mu_2^*}(z) + U^{\mu_3^*}(z), \quad z \in \mathcal{H}_2, \\ U^{\nu_2}(z) &= U^{\mu_2^*}(z) + U^{\mu_3^*}(z) + 2a \operatorname{Re} z, \quad z \in \mathcal{H}_3. \end{aligned} \quad (4.20)$$

In particular, recalling that  $\Delta_2 \subset \mathbb{C}_-$ , variational identities from Propositions 3.33 and 4.2 imply that

$$\begin{aligned} 2U^{\mu_1^* + \mu_2^*}(x) - U^{\nu_2}(x) + V(x) - a \operatorname{Re} x &= \ell_1, \quad x \in \Delta_1 \cap \mathbb{C}_+, \\ 2U^{\mu_1^* + \mu_2^*}(x) - U^{\nu_2}(x) + V(x) + a \operatorname{Re} x &= \ell_2, \quad x \in \Delta_2 \subset \mathbb{C}_-, \\ 2U^{\mu_3^*}(z) - U^{\mu_1^*}(z) + U^{\mu_2^*}(z) + \phi_3(z) &= 0, \quad z \in \Delta_3. \end{aligned}$$

In principle, the constants above may depend on the connected components of the sets  $\Delta_1, \Delta_2$  and  $\Delta_3$ . However, as we are now working under the assumption of the symmetric spectral curve in [31], (3.38) and (3.40) hold true and the constant  $\ell_j$  is the same for any connected component of  $\Delta_j$ .

As for  $x \in \Delta_1 \cap \mathbb{C}_-$ , we have that

$$2U^{\mu_1^* + \mu_2^*}(x) - U^{\mu_2^* + \mu_3^*}(x) + V(x) - a \operatorname{Re} x = \ell_1,$$

and by the third identity in (4.20), it can be written as

$$2U^{\mu_1^* + \mu_2^*}(x) - U^{\nu_2}(x) + V(x) + a \operatorname{Re} x = \ell_1,$$

which shows that  $\ell_1 = \ell_2 =: \ell$ . Recalling that  $\nu_1^* = \mu_1^* + \mu_2^*$ , we summarize these identities as

$$2U^{\nu_1^*}(x) - U^{\nu_2}(x) + V(x) - a|\operatorname{Re} x| = \ell, \quad x \in \text{supp } \lambda = \Delta_1 \cap \Delta_2.$$

On the other hand, adding the first two equalities in (4.20) we get that

$$2U^{\nu_2}(z) = U^{\nu_1^*}(z), \quad z \in i\mathbb{R} \setminus [-y_*, y_*], \quad (4.21)$$



and

$$2U^{\nu_2}(z) = 2U^{\nu_1^*}(z), \quad z \in [-y_*, y_*].$$

By symmetry,  $U^{\nu_1^*}(0) = 0$ . Also, having in mind that  $\text{supp } \nu_1^* \subset \mathbb{R}$ ,

$$\frac{d}{dy} \left( 2U^{\nu_1^*}(iy) \right) = \frac{d}{dy} \int \log \frac{1}{x^2 + y^2} d\nu_1^*(x) = -2y \int \frac{1}{x^2 + y^2} d\nu_1^*(x),$$

showing that  $U^{\nu_1^*}(z)$  is strictly decreasing as  $z$  traverses  $i\mathbb{R}_+$  upwards. Thus,  $U^{\nu_1^*}(z) < 0$  for  $z \in i\mathbb{R} \setminus \{0\}$ , and

$$2U^{\nu_2}(z) = 2U^{\nu_1^*}(z) \leq U^{\nu_1^*}(z), \quad z \in [-y_*, y_*]. \quad (4.22)$$

Equations (4.21)–(4.22) show that  $\nu_2$  satisfies (4.19), and consequently  $\nu_2 = \nu_2^*$ , finishing the proof of Theorem C.  $\square$

## 5. MULTIPLE ORTHOGONAL POLYNOMIALS

From this section on, we will also tacitly assume that  $V$  is a polynomial of even degree and positive leading coefficient, so that the measures  $e^{-N(V(x) \pm ax)} dx$  have all moments finite, and the matrix model (1.1) is well defined.

As it was mentioned in Section 2.2.2, we can embed the average characteristic polynomial  $\pi_{n_1, n_2} = \pi_{\vec{n}}$ , defined in (2.4), in a family of multiple orthogonal polynomials (MOP)  $P_{\vec{k}}^{(N)}$ , satisfying (2.23), in such a way that  $\pi_{\vec{n}} = P_{\vec{n}}^{(N)}$ ,  $N = n_1 + n_2$ . However, in order to prove Theorem E we will need to treat the indices  $N$  and  $\vec{k}$  as independent parameters.

### 5.1. Zeros of MOP.

It is well known that the zeros of standard orthogonal polynomials on  $\mathbb{R}$  are real and simple, and that the zeros of two consecutive polynomials interlace. Such a result for multiple orthogonal polynomials is not immediate; however, in our case it still holds, and we prove it taking advantage of a certain feature of our weights.

Namely, a system of functions  $\{w_1, \dots, w_s\}$  forms an **AT system** on a bounded or unbounded interval or the real axis if any linear combination  $a_1 w_1 + \dots + a_s w_s$  has at most  $s - 1$  zeros on this interval, counted with account of multiplicity (see e.g. [79, Chapter 4]).

**Proposition 5.1.** *For any non-negative integers  $k_1, k_2$  the system*

$$\left\{ e^{Nax}, \dots, x^{k_1-1} e^{Nax}, e^{-Nax}, \dots, x^{k_2-1} e^{-Nax} \right\}$$

*is an AT-system on  $\mathbb{R}$ .*

*Proof.* Suppose to the contrary that a function of the form

$$f(x) = p_1(x) e^{Nax} + p_2(x) e^{-Nax},$$

where  $p_1$  and  $p_2$  are polynomials of degree up to  $k_1 - 1$  and  $k_2 - 1$ , respectively, has at least  $k_1 + k_2$  zeros on  $\mathbb{R}$ . Then the same is true for the function

$$g(x) = p_1(x) + p_2(x)e^{-2Nax},$$

and thus, the  $k_1$ -th derivative  $g^{(k_1)}$  of  $g$  has to have at least  $k_2$  zeros. However,  $g^{(k_1)}(x) = e^{-2Nax} \times (\text{polynomial degree} \leq k_2 - 1)$ , which trivially can have at most  $k_2 - 1$  zeros.  $\square$

The previous proposition establishes that

$$\left\{ e^{-NV_1(x)}, \dots, x^{k_1-1} e^{-NV_1(x)}, e^{-NV_2(x)}, \dots, x^{k_2-1} e^{-NV_2(x)} \right\}$$

is an AT-system on  $\mathbb{R}$ . By [58, Theorem 2.1], the following is true:

**Proposition 5.2.** *For any index  $\vec{k}$ , the zeros of  $P_{\vec{k}}^{(N)}$  are real and simple; furthermore, for  $j = 1, 2$ , the zeros of  $P_{\vec{k}}^{(N)}$  and  $P_{\vec{k}-\vec{e}_j}^{(N)}$  interlace.*

Interlacing has straightforward consequences on the limits of the zero-counting measures  $\mu(P_{\vec{n}}^{(N)})$  (recall the definition in (2.2)):

**Proposition 5.3.** *Suppose that for some path  $(\vec{n}_j)$  and a sequence of integers  $(N_j)$  we have the convergence*

$$\mu(P_{\vec{n}_j}^{(N_j)}) \xrightarrow{*} \lambda \quad \text{as } j \rightarrow \infty.$$

*for some compactly supported measure  $\lambda$ . Then for any fixed vector  $\vec{v} \in \mathbb{Z}^2$  we have the convergence*

$$\mu(P_{\vec{n}_j+\vec{v}}^{(N_j)}) \xrightarrow{*} \lambda \quad \text{as } j \rightarrow \infty,$$

*for the same measure  $\lambda$ .*

*Proof.* Clearly, it is sufficient to prove the assertion for  $P_{\vec{n}_j \pm \vec{e}_k}^{(N_j)}$ ,  $k = 1, 2$ . By the interlacing property, for any interval  $[a, b] \subset \mathbb{R}$ ,

$$\left| \int_a^b d\mu(P_{\vec{n}_j}^{(N_j)}) - \frac{|\vec{n}_j \pm \vec{e}_k|}{|\vec{n}_j|} \int_a^b d\mu(P_{\vec{n}_j \pm \vec{e}_k}^{(N_j)}) \right| \leq \frac{2}{|\vec{n}_j|},$$

and the result follows.  $\square$

## 5.2. Recurrence coefficients of MOP.

Closely related to the multiple orthogonal polynomials  $(P_{\vec{k}}^{(N)})$  in (2.23) is the sequence of associated biorthogonal functions  $(Q_{\vec{k}}^{(N)})$ , also known as *type I multiple orthogonal polynomials*, of the form

$$Q_{\vec{k}}^{(N)}(x) = A_{\vec{k}}^{(N,1)}(x)e^{-NV_1(x)} + A_{\vec{k}}^{(N,2)}(x)e^{-NV_2(x)}, \quad (5.1)$$

where  $A_{\vec{k}}^{(N,1)}$  and  $A_{\vec{k}}^{(N,2)}$  are polynomials with

$$A_{\vec{k}}^{(N,j)}(x) = \gamma_{\vec{k}}^{(N,j)} x^{k_j-1} + \text{lower order terms}, \quad (5.2)$$

uniquely determined by the orthogonality conditions

$$\int x^k Q_{\vec{j}}^{(N)}(x) dx = \begin{cases} 0, & \text{if } k \leq |\vec{j}| - 2, \\ 1, & \text{if } k = |\vec{j}| - 1, \end{cases}$$

(see e.g. [61, Section 23.1.3] or [91]). Taking into account (2.23) we see that  $(P_{\vec{k}}^{(N)})$  and  $(Q_{\vec{k}}^{(N)})$  satisfy the biorthogonality relations

$$\int P_{\vec{k}}^{(N)}(x) Q_{\vec{j}}^{(N)}(x) dx = \begin{cases} 1, & \text{if } |\vec{k}| = |\vec{j}| - 1, \\ 0, & \text{if } j_1 \leq k_1 \text{ and } j_2 \leq k_2, \text{ or } |\vec{k}| \leq |\vec{j}| - 2, \end{cases} \quad (5.3)$$

and also that

$$\frac{1}{\gamma_{\vec{k}}^{(N,j)}} = \int x^{k_j-1} P_{\vec{k}-\vec{e}_j}^{(N)}(x) e^{-NV_j(x)} dx, \quad j = 1, 2. \quad (5.4)$$

It is well known [91, Equation (1.7)] that multiple orthogonal polynomials  $(P_{\vec{k}}^{(N)})$  satisfy a system of linear recurrence relations only involving nearest neighbor multi-indices  $\vec{k} \pm \vec{e}_j$ ,

$$x P_{\vec{k}}^{(N)}(x) = P_{\vec{k}+\vec{e}_j}^{(N)}(x) + b_{\vec{k}}^{(j)} P_{\vec{k}}^{(N)}(x) + a_{\vec{k}}^{(1)} P_{\vec{k}-\vec{e}_1}^{(N)}(x) + a_{\vec{k}}^{(2)} P_{\vec{k}-\vec{e}_2}^{(N)}(x), \quad j = 1, 2, \quad (5.5)$$

where  $a_{\vec{k}}^{(j)} = a_{\vec{k}}^{(N,j)}$  and  $b_{\vec{k}}^{(j)} = b_{\vec{k}}^{(N,j)}$  vary with  $N$ . Additionally, the coefficients  $a_{\vec{k}}^{(N,j)}$  have the following expression in terms of  $\gamma_{\vec{k}}^{(N,j)}$  in (5.2),

$$a_{\vec{k}}^{(N,j)} = \frac{\gamma_{\vec{k}}^{(N,j)}}{\gamma_{\vec{k}+\vec{e}_j}^{(N,j)}}, \quad j = 1, 2. \quad (5.6)$$

The sequence of type I MOP's also satisfy a similar recurrence, although it is not as widely explored. It takes the form [91, Equation (1.10)]

$$x Q_{\vec{k}}^{(N)}(x) = Q_{\vec{k}-\vec{e}_j}^{(N)}(x) + b_{\vec{k}-\vec{e}_j}^{(j)} Q_{\vec{k}}^{(N)}(x) + a_{\vec{k}}^{(1)} Q_{\vec{k}+\vec{e}_1}^{(N)}(x) + a_{\vec{k}}^{(2)} Q_{\vec{k}+\vec{e}_2}^{(N)}(x), \quad (5.7)$$

where the coefficients are the same as the ones that appear in (5.5), although here the  $b$ 's come with shifted index.

Theorem E will be proved under the assumption of boundedness of zeros of MOP's slightly off a given up-right path. This control of zeros of polynomials off the path, in essence, can be traced back to the different recurrence relations satisfied by the MOP's: even though they satisfy a four term (nearest neighbor) recurrence relation, under general conditions one cannot assure that the coefficients in this recurrence are bounded (see for instance [49, Theorem 2.3.1]). On the other hand, under the assumption of boundedness of zeros, the recurrence relation along a given path

has bounded coefficients, but has the drawback of having more terms, as shown in the next three results, moving us further away from the path.

**Proposition 5.4.** *Let  $(\vec{n}_k)$  be an up-right path as in (2.22), with  $d$  defined in (2.22) and  $N > 0$  given. There exist coefficients  $\theta_{\vec{n}_k}^{(N,0)}, \dots, \theta_{\vec{n}_k}^{(N,d)}$  such that the sequence  $(P_{\vec{n}_k}^{(N)})_{k \geq 0}$  satisfies a  $d$ -term recurrence relation of the form*

$$x P_{\vec{n}_k}^{(N)}(x) = P_{\vec{n}_{k+1}}^{(N)}(x) + \sum_{\ell=0}^d \theta_{\vec{n}_k}^{(N,\ell)} P_{\vec{n}_{k-\ell}}^{(N)}(x). \quad (5.8)$$

Furthermore, these recurrence coefficients  $\theta_{\vec{n}_k}^{(N,0)}, \dots, \theta_{\vec{n}_k}^{(N,d)}$  are given by polynomial expressions in the variables  $a_{\vec{n}_{k-j}}^{(N,1)}, a_{\vec{n}_{k-j}}^{(N,2)}, b_{\vec{n}_{k-j}}^{(N,1)}$  and  $b_{\vec{n}_{k-j}}^{(N,2)}$ ,  $j = 0, \dots, d$ , and with coefficients and degrees that remain uniformly bounded as  $N, k \rightarrow \infty$ .

*Proof.* We follow standard arguments, see e.g. [59, Lemma 3.9] and also [51, Proposition 2.1]. Let us write

$$m_k = j, \quad \text{with } j \text{ determined by } \vec{n}_k - \vec{n}_{k-1} = \vec{e}_j.$$

Evaluating (5.5) with  $j = 1$  and  $j = 2$  and taking the difference, we get the compatibility condition

$$P_{\vec{n}_k + \vec{e}_1}^{(N)}(x) = P_{\vec{n}_k + \vec{e}_2}^{(N)}(x) + (b_{\vec{n}_k}^{(N,2)} - b_{\vec{n}_k}^{(N,1)}) P_{\vec{n}_k}^{(N)}(x).$$

We apply this to (5.5) recursively: whenever  $m_\ell = 1$  we use the above to get rid of  $P_{\vec{n}_\ell - \vec{e}_2}^{(N)}$ , whereas when  $m_\ell = 2$  we eliminate  $P_{\vec{n}_\ell - \vec{e}_1}^{(N)}$ . As a result, (5.5) becomes

$$x P_{\vec{n}_k}^{(N)}(x) = P_{\vec{n}_{k+1}}^{(N)}(x) + \sum_{\ell=0}^k \theta_{\vec{n}_k}^{(N,\ell)} P_{\vec{n}_{k-\ell}}^{(N)}(x),$$

where

$$\theta_{\vec{n}_k}^{(N,0)} = b_{\vec{n}_k}^{(N,m_k)}, \quad \theta_{\vec{n}_k}^{(1)} = a_{\vec{n}_k}^{(N,1)} + a_{\vec{n}_k}^{(N,2)}$$

and

$$\theta_{\vec{n}_k}^{(N,\ell)} = a_{\vec{n}_k}^{(N,1)} \prod_{j=k-\ell+1}^k (b_{\vec{n}_j - \vec{e}_1}^{(N,1)} - b_{\vec{n}_j - \vec{e}_1}^{(N,m_j)}) + a_{\vec{n}_k}^{(N,2)} \prod_{j=k-\ell+1}^k (b_{\vec{n}_j - \vec{e}_2}^{(N,2)} - b_{\vec{n}_j - \vec{e}_2}^{(N,m_j)}), \quad 2 \leq \ell \leq k.$$

From the last condition in (2.22), we see that when  $\ell > d$ , both products above are zero, concluding the proof.  $\square$

**Corollary 5.5.** *Let  $(\vec{n}_k)$  be an up-right path as in (2.22), and  $N > 0$  fixed and set*

$$\Theta_{\vec{n}_k}^{(N)} := (\theta_{\vec{n}_k}^{(N,0)}, \dots, \theta_{\vec{n}_k}^{(N,d)})$$

to be the vector of coefficients of the recurrence relation (5.8). For any given  $\ell > 0$ , there exist polynomials  $F_{-d(\ell-1)}^{(\ell)}, F_{-d\ell+1}^{(\ell)}, \dots, F_{\ell-1}^{(\ell)}$  in  $(\ell+1)(d+1)$ -variables such that the sequence  $(P_{\vec{n}_k}^{(N)})_{k \geq 0}$  satisfies the recurrence relation

$$x^\ell P_{\vec{n}_k}^{(N)}(x) = P_{\vec{n}_{k+\ell}}^{(N)}(x) + \sum_{j=-d(\ell-1)}^{\ell-1} F_j^{(\ell)} \left( \Theta_{\vec{n}_k - (\ell-1)d}^{(N)}, \Theta_{\vec{n}_k - (\ell-1)d+1}^{(N)}, \dots, \Theta_{\vec{n}_{k+\ell-1}}^{(N)} \right) P_{\vec{n}_{k+j}}^{(N)}(x), \quad k \geq d\ell, \quad (5.9)$$

The polynomials  $F_j^{(\ell)}$ 's are universal, in the sense that they are independent of the measures of multiple orthogonality and independent of  $N$  and of the up-right path  $(\vec{n}_k)$ . In particular, their coefficients and degrees remain bounded as  $N, k \rightarrow \infty$ .

*Proof.* For  $\ell = 1$  the assertion is trivial. Let  $\ell \geq 2$ ; we can rewrite the recurrence relation (5.8) as

$$x^\ell P_{\vec{n}_k}^{(N)}(x) = x^{\ell-2} \left( x P_{\vec{n}_{k+1}}^{(N)}(x) + \sum_{j=0}^d \theta_{\vec{n}_k}^{(N,j)} x P_{\vec{n}_{k-j}}^{(N)}(x) \right)$$

and use (5.8) again with each polynomial between parentheses. Iterating this process we get (5.9).  $\square$

We also need to control the growth of the recurrence coefficients appearing in the previous results.

**Proposition 5.6.** *Let  $(\vec{n}_k)$  be an up-right path as in (2.22) and  $(N_k)$  be a sequence of integers with  $N_k \rightarrow \infty$  as  $k \rightarrow \infty$ . If the zeros of  $P_{\vec{n}_{k+1}}^{(N_k)}$  remain uniformly bounded as  $k \rightarrow \infty$ , then the corresponding recurrence coefficients  $\theta_{\vec{n}_k}^{(N_k,0)}, \dots, \theta_{\vec{n}_k}^{(N_k,d)}$  in (5.8) also remain bounded as  $k \rightarrow \infty$ .*

*Proof.* We follow the standard scheme of proof, see e.g. [6, Lemma 2.2]. For ease of notation, we omit the  $k$ -dependence on  $N_k$  and simply write  $N_k = N$ .

Recall that the zeros of  $P_{\vec{n}_j}^{(N)}$  are all real and simple, and the zeros of  $P_{\vec{n}_j}^{(N)}$  and  $P_{\vec{n}_{j+1}}^{(N)}$  interlace, for any  $j$ . If they are additionally contained in an interval  $\Delta \subset \mathbb{R}$ , then for any compact  $K \subset \mathbb{C} \setminus \Delta$  there exists a constant  $C(K) > 0$  such that

$$\left| \frac{P_{\vec{n}_{j-k}}^{(N)}(x)}{P_{\vec{n}_{j-i}}^{(N)}(x)} \right| \leq C, \quad x \in K, \quad j \in \mathbb{N}, \quad 0 \leq i \leq d, \quad -1 \leq k < i.$$

We rewrite (5.8) as

$$x = \frac{P_{\vec{n}_{k+1}}^{(N)}(x)}{P_{\vec{n}_k}^{(N)}(x)} + \theta_{\vec{n}_k}^{(N,0)} + \sum_{j=1}^d \theta_{\vec{n}_k}^{(N,j)} \frac{P_{\vec{n}_{k-j}}^{(N)}(x)}{P_{\vec{n}_k}^{(N)}(x)}.$$

With  $r > 0$  sufficiently large for which  $\Delta \subset \{|z| < r\}$ , Cauchy's integral formula yields

$$\theta_k^{(N,0)} = -\frac{1}{2\pi i} \oint_{|z|=r} \frac{P_{\vec{n}_{k+1}}^{(N)}(z) dz}{P_{\vec{n}_k}^{(N)}(z) z},$$

and this proves that  $|\theta_k^{(N,0)}| \leq C(\{|z| \leq r\})$ .

In the same vein, for any  $\ell$  with  $1 \leq \ell \leq d$ ,

$$x \frac{P_{\vec{n}_k}^{(N)}(x)}{P_{\vec{n}_{k-\ell}}^{(N)}(x)} = \frac{P_{\vec{n}_{k+1}}^{(N)}(x)}{P_{\vec{n}_{k-\ell}}^{(N)}(x)} + \theta_k^{(N,\ell)} + \sum_{\substack{j=1 \\ j \neq \ell}}^d \theta_k^{(N,j)} \frac{P_{\vec{n}_{k-j}}^{(N)}(x)}{P_{\vec{n}_{k-\ell}}^{(N)}(x)}$$

so that

$$\theta_k^{(N,\ell)} = \frac{1}{2\pi i} \oint_{|z|=r} \left( z \frac{P_{\vec{n}_k}^{(N)}(z)}{P_{\vec{n}_{k-\ell}}^{(N)}(z)} - \frac{P_{\vec{n}_{k+1}}^{(N)}(z)}{P_{\vec{n}_{k-\ell}}^{(N)}(z)} - \sum_{j=1}^{\ell-1} \theta_k^{(N,j)} \frac{P_{\vec{n}_{k-j}}^{(N)}(z)}{P_{\vec{n}_{k-\ell}}^{(N)}(z)} \right) \frac{dz}{z}.$$

Thus, assuming inductively that for some constant  $M > 0$

$$|\theta_k^{(N,j)}| \leq M, \quad j = 0, \dots, \ell-1,$$

we obtain

$$|\theta_k^{(N,\ell)}| \leq r C(\{|z| \leq r\}) + C(\{|z| \leq r\}) + (\ell-1)M C(\{|z| \leq r\}),$$

which is enough to conclude the proof.  $\square$

The last results established relations between type II MOP's along a given up-right path. But we also need to relate both type I and II MOP's slightly off the given path to the ones along the path.

For a given sequence of multi-indices  $(\vec{n}_k)$  with an up-right path structure satisfying (2.22) we define the symbols

$$\|_k, \perp_k \in \{1, 2\} \tag{5.10}$$

by the property

$$\vec{e}_{\|_k} := \vec{n}_{k+1} - \vec{n}_k, \quad \vec{e}_{\|_k} \perp \vec{e}_{\perp_k}. \tag{5.11}$$

In other words, the vector  $\vec{e}_{\|_k}$  gives us the direction from  $\vec{n}_k$  to  $\vec{n}_{k+1}$ , while  $\vec{e}_{\perp_k}$  is the only vector among  $\vec{e}_1$  and  $\vec{e}_2$  which is perpendicular to  $\vec{e}_{\|_k}$ .

**Proposition 5.7.** *Fix  $N$  and an up-right path  $(\vec{n}_k)$  with corresponding value  $d$  as in (2.22) and let  $(Q_{\vec{m}}) = (Q_{\vec{m}}^{(N)})$  and  $(P_{\vec{m}}) = (P_{\vec{m}}^{(N)})$  be type I and type II MOP's for the given  $N$ . Then there exist  $\ell_k \in \{1, \dots, d-1\}$  such that with the notation (5.10)–(5.11),*

$$Q_{\vec{n}_k + \vec{e}_{\perp_k}}(x) = \sum_{i=0}^{\ell_k} q_{\vec{n}_k}^{(i)} Q_{\vec{n}_{k+i+1}}(x), \tag{5.12}$$

where the coefficients  $q_{\vec{n}_k}^{(i)}$  are defined recursively by the formula

$$q_{\vec{n}_k}^{(0)} = 1, \quad q_{\vec{n}_k}^{(j+1)} = q_{\vec{n}_k}^{(j)} \left( b_{\vec{n}_k+j+\vec{e}_{\parallel_k}}^{(\perp_k)} - b_{\vec{n}_k+j+\vec{e}_{\perp_k}}^{(\parallel_k)} \right), \quad j = 0, 1, \dots, \ell_k - 1, \quad (5.13)$$

and  $b_k^{(j)}$  are the recurrence coefficients in (5.5).

Analogously, there exist  $\nu_k \in \{1, \dots, d-1\}$  such that

$$P_{\vec{n}_k-\vec{e}_{\perp_{k-1}}}(x) = \sum_{i=0}^{\nu_k} p_{\vec{n}_k}^{(i)} P_{\vec{n}_k-i-1}(x), \quad (5.14)$$

where the coefficients above are defined recursively by

$$p_{\vec{n}_k}^{(0)} = 1, \quad p_{\vec{n}_k}^{(j+1)} = p_{\vec{n}_k}^{(j)} \left( b_{\vec{n}_k-j-1-\vec{e}_{\perp_{k-1}}}^{(\perp_{k-1})} - b_{\vec{n}_k-j-1-\vec{e}_{\perp_{k-1}}}^{(\parallel_{k-1})} \right), \quad j = 0, 1, \dots, \nu_k - 1. \quad (5.15)$$

*Proof.* Taking the difference of the recurrence relations (5.7) with  $j = 1$  and  $j = 2$  gives the identity

$$Q_{\vec{n}-\vec{e}_2}(x) - Q_{\vec{n}-\vec{e}_1}(x) = \left( b_{\vec{n}-\vec{e}_1}^{(1)} - b_{\vec{n}-\vec{e}_2}^{(2)} \right) Q_{\vec{n}}(x),$$

which is valid for any  $\vec{n} \in \mathbb{N}^2$ . Applying it to  $\vec{n} = \vec{n}_k + \vec{e}_1 + \vec{e}_2 = \vec{n}_{k+1} + \vec{e}_{\perp_k}$  we obtain the relation

$$\begin{aligned} Q_{\vec{n}_k+\vec{e}_{\perp_k}}(x) &= Q_{\vec{n}_k+\vec{e}_{\parallel_k}}(x) + \left( b_{\vec{n}_k+\vec{e}_{\parallel_k}}^{(\perp_k)} - b_{\vec{n}_k+\vec{e}_{\perp_k}}^{(\parallel_k)} \right) Q_{\vec{n}_{k+1}+\vec{e}_{\perp_k}}(x) \\ &= Q_{\vec{n}_{k+1}}(x) + q_{\vec{n}_k}^{(1)} Q_{\vec{n}_{k+1}+\vec{e}_{\perp_k}}(x). \end{aligned}$$

If  $\vec{e}_{\perp_k} = \vec{e}_{\parallel_{k+1}}$  so that  $\vec{n}_{k+1} + \vec{e}_{\perp_k} = \vec{n}_{k+2}$ , then we get (5.12)–(5.13) with  $\ell_k = 1$ . Otherwise,  $\perp_k = \perp_{k+1}$ ,  $\parallel_k = \parallel_{k+1}$ , and we replace  $Q_{\vec{n}_{k+1}+\vec{e}_{\perp_k}}$  in the right hand side using this identity with  $k \mapsto k+1$ :

$$\begin{aligned} Q_{\vec{n}_k+\vec{e}_{\perp_k}}(x) &= Q_{\vec{n}_{k+1}}(x) + q_{\vec{n}_k}^{(1)} \left( Q_{\vec{n}_{k+2}}(x) + \left( b_{\vec{n}_{k+1}+\vec{e}_{\parallel_{k+1}}}^{(\perp_{k+1})} - b_{\vec{n}_{k+1}+\vec{e}_{\perp_{k+1}}}^{(\parallel_{k+1})} \right) Q_{\vec{n}_{k+2}+\vec{e}_{\perp_k}}(x) \right) \\ &= Q_{\vec{n}_{k+1}}(x) + q_{\vec{n}_k}^{(1)} Q_{\vec{n}_{k+2}}(x) + q_{\vec{n}_k}^{(1)} \left( b_{\vec{n}_{k+1}+\vec{e}_{\parallel_k}}^{(\perp_k)} - b_{\vec{n}_{k+1}+\vec{e}_{\perp_k}}^{(\parallel_k)} \right) Q_{\vec{n}_{k+2}+\vec{e}_{\perp_k}}(x) \\ &= Q_{\vec{n}_{k+1}}(x) + q_{\vec{n}_k}^{(1)} Q_{\vec{n}_{k+2}}(x) + q_{\vec{n}_k}^{(2)} Q_{\vec{n}_{k+2}+\vec{e}_{\perp_k}}(x). \end{aligned}$$

We can iterate this process until  $Q_{\vec{n}_{k+j-1}+\vec{e}_1+\vec{e}_2} = Q_{\vec{n}_{k+j}+\vec{e}_{\perp_k}} = Q_{\vec{n}_{k+j+1}}$ . By the last assumption in (2.22), this will happen in at most  $d-1$  steps.

The proof of (5.14) goes along the same line. From the nearest neighbor recurrence relation (5.5) we obtain

$$P_{\vec{n}+\vec{e}_2}(x) - P_{\vec{n}+\vec{e}_1}(x) = (b_{\vec{n}}^{(1)} - b_{\vec{n}}^{(2)}) P_{\vec{n}}(x),$$

valid for any  $\vec{n} \in \mathbb{N}^2$ , and apply it to  $\vec{n} = \vec{n}_k - \vec{e}_1 - \vec{e}_2 = \vec{n}_{k-1} - \vec{e}_{\perp_{k-1}}$ . This yields

$$P_{\vec{n}_{k-1}-\vec{e}_{\perp_{k-1}}}(x) = P_{\vec{n}_{k-1}}(x) + \left( b_{\vec{n}_{k-1}-\vec{e}_{\perp_{k-1}}}^{(\perp_{k-1})} - b_{\vec{n}_{k-1}-\vec{e}_{\perp_{k-1}}}^{(\parallel_{k-1})} \right) P_{\vec{n}_{k-1}-\vec{e}_{\perp_{k-1}}}(x).$$

If  $\vec{e}_{\perp k-1} = \vec{e}_{\parallel k-2}$  so that  $\vec{n}_{k-1} - \vec{e}_{\perp k-1} = \vec{n}_{k-2}$ , then we get (5.14)–(5.15) with  $\nu_k = 1$ . Otherwise,  $\vec{e}_{\perp k-1} = \vec{e}_{\perp k-2}$ ,  $\parallel_{k-1} = \parallel_{k-2}$ , and we replace  $P_{\vec{n}_{k-1} - \vec{e}_{\perp k-1}}$  in the right hand side using this identity with  $k \mapsto k-1$ :

$$\begin{aligned} P_{\vec{n}_{k-1} - \vec{e}_{\perp k-1}}(x) &= P_{\vec{n}_{k-1}}(x) + p_{\vec{n}_k}^{(1)} \left( P_{\vec{n}_{k-2}}(x) + \left( b_{\vec{n}_{k-2} - \vec{e}_{\perp k-2}}^{(\perp_{k-2})} - b_{\vec{n}_{k-2} - \vec{e}_{\perp k-2}}^{(\parallel_{k-2})} \right) P_{\vec{n}_{k-2} - \vec{e}_{\perp k-2}}(x) \right) \\ &= P_{\vec{n}_{k-1}}(x) + p_{\vec{n}_k}^{(1)} P_{\vec{n}_{k-2}}(x) + p_{\vec{n}_k}^{(1)} \left( b_{\vec{n}_{k-2} - \vec{e}_{\perp k-1}}^{(\perp_{k-1})} - b_{\vec{n}_{k-2} - \vec{e}_{\perp k-1}}^{(\parallel_{k-1})} \right) P_{\vec{n}_{k-2} - \vec{e}_{\perp k-2}}(x) \\ &= P_{\vec{n}_{k-1}}(x) + p_{\vec{n}_k}^{(1)} P_{\vec{n}_{k-2}}(x) + p_{\vec{n}_k}^{(2)} P_{\vec{n}_{k-2} - \vec{e}_{\perp k-2}}(x). \end{aligned}$$

We can iterate this process until  $P_{\vec{n}_{k-j} - \vec{e}_{\perp k-1}} = P_{\vec{n}_{k-j-1}}$ . By the last assumption in (2.22), this will happen in at most  $d-1$  steps.  $\square$

Proposition 5.6 allows us to control the recurrence coefficients for the type II MOP's along an up-right path. The next result is its analogue for the coefficients that appeared in Proposition 5.7.

**Proposition 5.8.** *Consider an up-right path  $(\vec{n}_N)$ , with corresponding sequence of type I and II MOP's with  $N$ -varying weights  $(Q_{\vec{n}_N}^{(N)})$  and  $(P_{\vec{n}_N}^{(N)})$  and coefficients  $(p_{\vec{n}_N}^{(j)}) = (p_{\vec{n}_N}^{(N,j)})$  and  $(q_{\vec{n}_N}^{(j)}) = (q_{\vec{n}_N}^{(N,j)})$  as in Proposition 5.7. If the zeros of  $P_{\vec{n}_{N+d}}^{(N)}$  remain bounded as  $N \rightarrow \infty$ , then the sequences  $(p_{\vec{n}_N}^{(N,j)})$  and  $(q_{\vec{n}_N}^{(N,j)})$  also remain bounded as  $N \rightarrow \infty$ .*

*Proof.* By standard arguments already used in the proof of Proposition 5.6 we obtain that the nearest neighbor recurrence coefficients (5.5) satisfy the identity

$$b_{\vec{n}}^{(N,j)} = -\frac{1}{2\pi i} \oint_{|z|=r} \frac{P_{\vec{n}+\vec{e}_j}^{(N)}(z)}{P_{\vec{n}}^{(N)}(z)} \frac{dz}{z}, \quad j = 1, 2, \quad (5.16)$$

where  $r$  is chosen so that the disk  $\{|z| = r\}$  encloses all the roots of  $P_{\vec{n}}^{(N)}$ . By Proposition 5.2, the zeros of  $P_{\vec{n}+\vec{e}_j}^{(N)}$  and  $P_{\vec{n}}^{(N)}$  interlace, and by the same arguments used to prove (5.5) we see that boundedness of the zeros of  $P_{\vec{n}+\vec{e}_j}^{(N)}$  implies boundedness of  $b_{\vec{n}}^{(N,j)}$ .

In particular, taking  $\vec{n} = \vec{n}_{N+k} + \vec{e}_i$ ,  $i \neq j$ ,  $i \in \{1, 2\}$ ,  $k \geq 0$ , we get

$$b_{\vec{n}_{N+k} + \vec{e}_i}^{(N,j)} = -\frac{1}{2\pi i} \oint_{|z|=r} \frac{P_{\vec{n}_{N+k} + \vec{e}_1 + \vec{e}_2}^{(N)}(z)}{P_{\vec{n}_{N+k} + \vec{e}_i}^{(N)}(z)} \frac{dz}{z}.$$

By Proposition 5.2, the zeros of  $P_{\vec{n}_{N+d}}^{(N)}$  bound the zeros of  $P_{\vec{n}_{N+k} + \vec{e}_1 + \vec{e}_2}^{(N)}$ ,  $0 \leq k \leq \ell_N - 2 \leq d-3$ . Moreover, by construction,  $P_{\vec{n}_{N+\ell_N-1} + \vec{e}_1 + \vec{e}_2}^{(N)} = P_{\vec{n}_{N+d}}^{(N)}$ . This shows that all  $b_{\vec{n}_{N+k} + \vec{e}_i}^{(N,j)}$  taking part in the formulas (5.13) are uniformly bounded.



We apply similar arguments to bound the  $p$ 's, except that we use (5.16) with  $\vec{n} = \vec{n}_{N-k} - \vec{e}_j$ ,  $j \in \{1, 2\}$  and  $k \geq 0$ , which gives

$$b_{\vec{n}_{N-k}-\vec{e}_j}^{(N,j)} = -\frac{1}{2\pi i} \oint_{|z|=r} \frac{P_{\vec{n}_{N-k}}^{(N)}(z)}{P_{\vec{n}_{N-k}-\vec{e}_j}^{(N)}(z)} \frac{dz}{z}.$$

The argument is even simpler now, since because of the up-right path structure and the interlacing, all zeros of  $P_{\vec{n}_{N-k}}^{(N)}$ ,  $k \geq 0$ , are bounded by the zeros of  $P_{\vec{n}_{N+d}}^{(N)}$ .  $\square$

*Remark 5.9.* Observe that the assumption of Proposition 5.8 that all the zeros of  $P_{\vec{n}_{N+d}}^{(N)}$  remain bounded as  $N \rightarrow \infty$  is equivalent to assume boundedness of zeros of the sequence  $P_{\vec{n}_N}^{(N-d)}$ . This will be used in the formulation of the statement of Theorem E.

We conclude this section with a result that will be needed to prove Theorem E.

**Proposition 5.10.** *Let  $P_{\vec{k}}^{(N)}$  and  $Q_{\vec{k}}^{(N)}$  be the biorthogonal functions in (5.3) constructed for the weights in (2.23), where we recall that  $V$  is a polynomial with  $\deg V = m$ . Let  $(\vec{n}_N)$  be a up-right path as in (2.22) with corresponding value  $d$ . Set  $\kappa = \max\{m-2, d\}$ . If the zeros of  $P_{\vec{n}_N+\kappa}^{(N)}$  remain bounded as  $N \rightarrow \infty$ , then the integrals*

$$\int x^s P_{\vec{n}_N}^{(N)}(x) Q_{\vec{n}_N}^{(N)}(x) dx, \quad s = 1, \dots, m-2$$

*all remain bounded as  $N \rightarrow \infty$ . The same conclusion holds true if in the integral above we replace  $P_{\vec{n}_N}^{(N)}(x) Q_{\vec{n}_N}^{(N)}(x)$  by either one of*

$$P_{\vec{n}_N-\vec{e}_j}^{(N)} Q_{\vec{n}_N}^{(N)}, \quad P_{\vec{n}_N}^{(N)} Q_{\vec{n}_N+\vec{e}_j}^{(N)}, \quad \text{or} \quad P_{\vec{n}_N-\vec{e}_j}^{(N)} Q_{\vec{n}_N+\vec{e}_i}^{(N)}, \quad i, j = 1, 2.$$

*Proof.* Equation (5.3) says that for an up-right path  $(\vec{n}_k)$ ,

$$\int P_{\vec{n}_k}^{(N)}(x) Q_{\vec{n}_j}^{(N)}(x) dx = \delta_{k,j-1}.$$

Using (5.9) we then get that if  $k \geq ds$ ,

$$\begin{aligned} & \int x^s P_{\vec{n}_N}^{(N)}(x) Q_{\vec{n}_N}^{(N)}(x) dx \\ &= \sum_{j=-d(s-1)}^s F_j^{(s)} \left( \Theta_{\vec{n}_{N-(s-1)d}}^{(N)}, \Theta_{\vec{n}_{N-(s-1)d+1}}^{(N)}, \dots, \Theta_{\vec{n}_{N+s-1}}^{(N)} \right) \int P_{\vec{n}_{N+j}}^{(N)}(x) Q_{\vec{n}_N}^{(N)}(x) dx \\ &= F_{-1}^{(s)} \left( \Theta_{\vec{n}_{N-(s-1)d}}^{(N)}, \Theta_{\vec{n}_{N-(s-1)d+1}}^{(N)}, \dots, \Theta_{\vec{n}_{N+s-1}}^{(N)} \right), \end{aligned}$$

where in the above and for the rest of the proof we have set

$$F_s^{(s)} = 1 \quad \text{and} \quad F_j^{(s)} = 0 \quad \text{for } j > s.$$

The conclusion is now a consequence of Corollary 5.5 and Proposition 5.6.

Consider the second integral of interest, namely

$$\int x^s P_{\vec{n}_N - \vec{e}_j}^{(N)}(x) Q_{\vec{n}_N}^{(N)}(x) dx, \quad s = 1, \dots, m-2.$$

If  $\vec{n}_N - \vec{e}_j = \vec{n}_{N-1}$ , the result follows from what we just proved. So, we only need to consider the case (see notation (5.10)–(5.11))

$$\int x^s P_{\vec{n}_N - \vec{e}_{\perp_{N-1}}}^{(N)}(x) Q_{\vec{n}_N}^{(N)}(x) dx,$$

which by Proposition 5.7 and the biorthogonality reduces to

$$\sum_{i=0}^{\nu_N} p_{\vec{n}_N}^{(i)} \int x^s P_{\vec{n}_N - i-1}^{(N)}(x) Q_{\vec{n}_N}^{(N)}(x) dx = \sum_{i=0}^{\nu_N} p_{\vec{n}_N}^{(i)} F_i^{(s)} \left( \Theta_{\vec{n}_N - i-1-(s-1)d}^{(N)}, \dots, \Theta_{\vec{n}_N - i+s-2}^{(N)} \right).$$

Proposition 5.8 and interlacing then tells us that the  $p$ 's in the sum above are bounded as long as the zeros of  $P_{\vec{n}_N+d}^{(N)}$  are bounded, whereas Proposition 5.6 assures the  $F$ 's above are bounded as long as the zeros of  $P_{\vec{n}_N+m-3}^{(N)}$  are bounded.

Analogously, for

$$\int x^s P_{\vec{n}_N}^{(N)}(x) Q_{\vec{n}_N + \vec{e}_j}^{(N)}(x) dx$$

we again only need to analyze

$$\begin{aligned} \int x^s P_{\vec{n}_N}^{(N)}(x) Q_{\vec{n}_N + \vec{e}_{\perp_N}}^{(N)}(x) dx &= \sum_{i=0}^{\ell_N} q_{\vec{n}_N}^{(i)} \int x^s P_{\vec{n}_N}^{(N)}(x) Q_{\vec{n}_N + i+1}^{(N)}(x) dx \\ &= \sum_{i=0}^{\ell_N} q_{\vec{n}_N}^{(i)} F_i^{(s)} \left( \Theta_{\vec{n}_N - (s-1)d}^{(N)}, \dots, \Theta_{\vec{n}_N + s-1}^{(N)} \right), \end{aligned}$$

which is bounded in the same manner as in the previous case.

Finally, for the last integral,

$$\int x^s P_{\vec{n}_N - \vec{e}_j}^{(N)}(x) Q_{\vec{n}_N + \vec{e}_i}^{(N)}(x) dx$$

we need only to consider the case

$$\int x^s P_{\vec{n}_N - \vec{e}_{\perp_{N-1}}}^{(N)}(x) Q_{\vec{n}_N + \vec{e}_{\perp_{N+1}}}^{(N)}(x) dx,$$

as all the other possible choices for  $\vec{e}_j$  and  $\vec{e}_i$  reduce to one of the previous cases. This last integral, in turn, equals

$$\begin{aligned} \sum_{j=0}^{\nu_N} \sum_{i=0}^{\ell_N} p_{\vec{n}_N}^{(j)} q_{\vec{n}_N}^{(i)} \sum_{k=-d(s-1)}^s F_k^{(s)} \left( \Theta_{\vec{n}_N - j-1-(s-1)d}^{(N)}, \dots, \Theta_{\vec{n}_N - j+s-2}^{(N)} \right) \int P_{\vec{n}_N - j-1+k}^{(N)}(x) Q_{\vec{n}_N + i+1}^{(N)}(x) dx \\ = \sum_{j=0}^{\nu_N} \sum_{i=0}^{\ell_N} p_{\vec{n}_N}^{(j)} q_{\vec{n}_N}^{(i)} F_{i+j+1}^{(s)} \left( \Theta_{\vec{n}_N - j-1-(s-1)d}^{(N)}, \dots, \Theta_{\vec{n}_N - j+s-2}^{(N)} \right). \end{aligned}$$

With the same argument as before, we are assured that the  $p$ 's and  $q$ 's are bounded because the zeros of  $P_{\vec{n}_{N+d}}^{(N)}$  remain bounded, whereas the  $F$ 's remain bounded because the zeros of  $P_{\vec{n}_{N+m-3}}^{(N)}$  remain bounded, completing the proof.  $\square$

## 6. CHARACTERIZATION OF THE EIGENVALUE DISTRIBUTION

Our main goal in this final section is to prove Theorems D and E from Section 2.2.2. We thus emphasize that now we will always assume that  $V$  is a polynomial of even degree  $m$  and positive leading coefficient.

### 6.1. The associated Riemann-Hilbert problems and the proof of Theorem D.

The starting point is the Riemann-Hilbert problem characterizing the MOP's in (2.23), obtained for the first time by Geronimo, Kuijlaars and Van Assche [92]: find a  $3 \times 3$  matrix-valued function  $\mathbf{Y} = \mathbf{Y}_{\vec{k}} = \mathbf{Y}_{\vec{k}}^{(N)}$  such that

- i)  $\mathbf{Y} : \mathbb{C} \setminus \mathbb{R} \rightarrow \mathbb{C}^{3 \times 3}$  is analytic;
- ii)  $\mathbf{Y}$  has continuous boundary values  $\mathbf{Y}_{\pm}$  on  $\mathbb{R}$  from  $\mathbb{H}_{\pm}$ , and they satisfy the jump condition

$$\mathbf{Y}_{+}(x) = \mathbf{Y}_{-}(x)(\mathbf{I} + e^{-NV_1(x)} \mathbf{E}_{12} + e^{-NV_2(x)} \mathbf{E}_{13}), \quad x \in \mathbb{R},$$

where  $\mathbf{E}_{ij} = \mathbf{e}_i \mathbf{e}_j^T$  is the matrix whose only non-zero entry is 1 in the position  $(i, j)$  and  $V_1$  and  $V_2$  were defined in (2.9);

- iii) the following asymptotic expansion is valid,

$$\mathbf{Y}(z) = (\mathbf{I} + \mathcal{O}(z^{-1})) \text{diag}(z^{k_1+k_2}, z^{-k_1}, z^{-k_2}),$$

as  $z \rightarrow \infty$  in any direction non-tangential to  $\mathbb{R}$ .

The unique solution  $\mathbf{Y}$  satisfying i)–iii) is given by

$$\mathbf{Y}(z) = \mathbf{\Lambda}_{\vec{k}}^{(N)} \begin{pmatrix} P_{\vec{k}}^{(N)}(z) & \mathcal{C}_1[P_{\vec{k}}^{(N)}](z) & \mathcal{C}_2[P_{\vec{k}}^{(N)}](z) \\ -2\pi i P_{\vec{k}-\vec{e}_1}^{(N)}(z) & \mathcal{C}_1[P_{\vec{k}-\vec{e}_1}^{(N)}](z) & \mathcal{C}_2[P_{\vec{k}-\vec{e}_1}^{(N)}](z) \\ -2\pi i P_{\vec{k}-\vec{e}_2}^{(N)}(z) & \mathcal{C}_1[P_{\vec{k}-\vec{e}_2}^{(N)}](z) & \mathcal{C}_2[P_{\vec{k}-\vec{e}_2}^{(N)}](z) \end{pmatrix} \text{diag}(1, 2\pi i, 2\pi i)^{-1} \quad (6.1)$$

with

$$\mathbf{\Lambda}_{\vec{k}}^{(N)} := \text{diag}\left(1, \gamma_{\vec{k}}^{(N,1)}, \gamma_{\vec{k}}^{(N,2)}\right), \quad (6.2)$$

where the coefficients above are as in (5.4).

In the formula above and in what follows,  $\mathcal{C}$  and  $\mathcal{C}_j$  denote the Cauchy and weighted Cauchy operators, defined for a suitable function  $f$  by

$$\mathcal{C}[f](z) = \int \frac{f(x)}{x-z} dx, \quad \mathcal{C}_j[f](z) := \mathcal{C}[f e^{-NV_j}](z), \quad z \in \mathbb{C} \setminus \mathbb{R}, \quad j = 1, 2.$$

The biorthogonal functions (5.1) can also be encoded in a Riemann-Hilbert problem, usually denoted by the letter  $\mathbf{X} = \mathbf{X}_{\vec{k}}$ . This Riemann-Hilbert problem formulation is not needed here, but

rather its explicit solution, which is known to satisfy  $\mathbf{X}^T = \mathbf{Y}^{-1}$ , and consequently  $\mathbf{Y}^{-1}$  can be explicitly computed in terms of the biorthogonal functions  $(Q_{\vec{k}}^{(N)})$ , taking the form [92]

$$\mathbf{Y}^{-1}(z) = \begin{pmatrix} -\mathcal{C}[Q_{\vec{k}}^{(N)}](z) & -\frac{1}{2\pi i}\mathcal{C}[Q_{\vec{k}+\vec{e}_1}^{(N)}](z) & -\frac{1}{2\pi i}\mathcal{C}[Q_{\vec{k}+\vec{e}_2}^{(N)}](z) \\ 2\pi i A_{\vec{k}}^{(N,1)}(z) & A_{\vec{k}+\vec{e}_1}^{(N,1)}(z) & A_{\vec{k}+\vec{e}_2}^{(N,1)}(z) \\ 2\pi i A_{\vec{k}}^{(N,2)}(z) & A_{\vec{k}+\vec{e}_1}^{(N,2)}(z) & A_{\vec{k}+\vec{e}_2}^{(N,2)}(z) \end{pmatrix} \text{diag} \left( 1, \gamma_{\vec{k}+\vec{e}_1}^{(N,1)}, \gamma_{\vec{k}+\vec{e}_2}^{(N,2)} \right)^{-1}. \quad (6.3)$$

*Remark 6.1.* We want to stress that we are dealing with *varying weights*, so all the quantities just introduced depend on  $N$ ,  $k_1$  and  $k_2$ , and these are *independent parameters*. Of particular interest for us is the choice  $k_1 = n_1, k_2 = n_2$  and  $N = k_1 + k_2 = n_1 + n_2$  simultaneously, as this leads to the average characteristic polynomial (2.4). In what follows, we always denote  $N = n_1 + n_2$ , but we need to treat  $k_1$  and  $k_2$  as independent of  $N$ . To avoid a more cumbersome notation, we will drop temporarily the  $N$  dependence and simply write  $P_{\vec{k}}^{(N)} = P_{\vec{k}}$ ,  $A_{\vec{k}}^{(N,1)} = A_{\vec{k}}^{(1)}$  etc, and reserve the indices  $n_1$  and  $n_2$  for when  $N = n_1 + n_2$ . Whenever the  $N$ -dependence becomes relevant, we write it explicitly. We will also deliberately drop the argument  $z$  from these functions when the variable is irrelevant.

The Riemann-Hilbert problem for  $\mathbf{Y}$  and the corresponding non-linear steepest descent method of Deift and Zhou are powerful tools in proving the macro and microscopic scale asymptotic results for (1.1) and several other random matrix models [9, 54, 66, 18, 31, 32, 48]. We will not explore this asymptotic side of Riemann-Hilbert problems, but rather use it as an algebraic tool that nicely encodes many quantities on the model.

Define

$$\mathbf{T}(z) = \mathbf{T}_{\vec{k}}(z) := \mathbf{Y}(z)\mathbf{F}(z)^{-1}, \quad z \in \mathbb{C} \setminus \mathbb{R}, \quad (6.4)$$

with

$$\mathbf{F}(z) := \text{diag}(e^{NV(z)}, e^{Na_1z}, e^{Na_2z}), \quad z \in \mathbb{C}, \quad \text{so} \quad \mathbf{F}(z)^{-1} = \text{diag}(e^{-NV(z)}, e^{-Na_1z}, e^{-Na_2z}).$$

Further denoting

$$\mathbf{D}(z) := \text{diag}(V'(z), a_1, a_2), \quad \mathbf{K} = \mathbf{K}_{\vec{k}}^{(N)} := \frac{1}{N} \text{diag}(k_1 + k_2, -k_1, -k_2),$$

the matrix  $\mathbf{F}$  satisfies

$$\mathbf{F}'(z) = N\mathbf{D}(z)\mathbf{F}(z), \quad (\mathbf{F}(z)^{-1})' = -N\mathbf{D}(z)\mathbf{F}(z)^{-1}. \quad (6.5)$$

**Proposition 6.2.** *The matrix-valued function  $\mathbf{T}$  satisfies the first-order differential equation*

$$\frac{d}{dz} \mathbf{T}(z) = N\mathbf{R}_{\vec{k}}(z)\mathbf{T}(z), \quad z \in \mathbb{C} \setminus \mathbb{R}, \quad (6.6)$$

where  $\mathbf{R}_{\vec{k}}$  is a matrix-valued polynomial of the form

$$\mathbf{R}_{\vec{k}}(z) = \begin{pmatrix} -V'(z) + \mathcal{O}(z^{m-2}) & \mathcal{O}(z^{m-2}) & \mathcal{O}(z^{m-2}) \\ \mathcal{O}(z^{m-2}) & \mathcal{O}(z^{m-3}) & \mathcal{O}(z^{m-3}) \\ \mathcal{O}(z^{m-2}) & \mathcal{O}(z^{m-3}) & \mathcal{O}(z^{m-3}) \end{pmatrix}, \quad (6.7)$$

and  $V$  was defined in (2.5). Moreover,

$$\mathrm{Tr} \mathbf{R}_{\vec{k}}(z) = -V'(z), \quad (6.8)$$

$$\mathrm{Tr} \mathbf{R}_{\vec{k}}^2(z) = (V'(z))^2 + \mathcal{O}(z^{m-2}), \quad (6.9)$$

$$\det \mathbf{R}_{\vec{k}}(z) = \det \left( \frac{1}{N} \mathbf{T}'(z) \mathbf{T}(z)^{-1} \right) = v_m a^2 z^{m-1} + \left( v_{m-1} a + v_m \frac{k_1 - k_2}{N} \right) a z^{m-2} + \mathcal{O}(z^{m-3}). \quad (6.10)$$

*Proof.* It is straightforward to check that  $\mathbf{T}$  is the only solution of the following Riemann-Hilbert problem:

- (i)  $\mathbf{T}$  is analytic on  $\mathbb{C} \setminus \mathbb{R}$ ;
- (ii)  $\mathbf{T}$  has continuous boundary values  $\mathbf{T}_{\pm}$  on  $\mathbb{R}$  from  $\mathbb{H}_{\pm}$ , and they satisfy the jump condition

$$\mathbf{T}_+(x) = \mathbf{T}_-(x)(\mathbf{I} + \mathbf{E}_{12} + \mathbf{E}_{13}), \quad x \in \mathbb{R};$$

- (iii) we have that

$$\mathbf{T}(z) = (\mathbf{I} + \mathbf{S}(z)) \mathrm{diag}(z^{k_1+k_2}, z^{-k_1}, z^{-k_2}) \mathbf{F}(z)^{-1} \quad (6.11)$$

where

$$\mathbf{S}(z) = \mathcal{O}(z^{-1})$$

as  $z \rightarrow \infty$  in any direction non-tangential to  $\mathbb{R}$ .

Since the jump matrix for  $\mathbf{T}$  across  $\mathbb{R}$  is constant, standard arguments yield (6.6), that is,

$$\frac{d}{dz} \mathbf{T}(z) = N \mathbf{R}_{\vec{k}}(z) \mathbf{T}(z), \quad z \in \mathbb{C} \setminus \mathbb{R}, \quad (6.12)$$

where  $\mathbf{R}_{\vec{k}}(z)$  is an entire matrix-valued function. We can find  $\mathbf{R}_{\vec{k}}$  by comparing the asymptotics at  $z \rightarrow \infty$  of both sides. Indeed, with the notation in (6.4)–(6.5) we have that

$$\begin{aligned} \frac{d}{dz} \mathbf{T}(z) &= N \left( \frac{1}{N} \mathbf{Y}'(z) - \mathbf{Y}(z) \mathbf{D}(z) \right) \mathbf{F}(z)^{-1} \\ &= N \left( \frac{1}{N} \mathbf{S}'(z) + (\mathbf{I} + \mathbf{S}(z)) \left( \frac{1}{z} \mathbf{K} - \mathbf{D}(z) \right) \right) \mathrm{diag} \left( z^{k_1+k_2}, z^{-k_1}, z^{-k_2} \right) \mathbf{F}(z)^{-1}. \end{aligned}$$

Using it in (6.12) together with (6.11), and recalling that  $\det \mathbf{T}(z) \equiv 1$  for all  $z \in \mathbb{C} \setminus \mathbb{R}$  so that  $\mathbf{T}(z)^{-1}$  always exists, we get that

$$\mathbf{R}_{\vec{k}}(z) = \frac{1}{N} \mathbf{T}'(z) \mathbf{T}^{-1}(z) = \left[ \frac{1}{N} \mathbf{S}'(z) + (\mathbf{I} + \mathbf{S}(z)) \left( \frac{1}{z} \mathbf{K} - \mathbf{D}(z) \right) \right] (\mathbf{I} + \mathbf{S}(z))^{-1}. \quad (6.13)$$

Liouville's theorem implies that  $\mathbf{R}_{\vec{k}}$  is a polynomial; since

$$(\mathbf{I} + \mathbf{S}(z))^{-1} = \mathbf{I} + \mathcal{O}(z^{-1}), \quad z \rightarrow \infty,$$

we have in fact that

$$\begin{aligned} \mathbf{R}_{\vec{k}}(z) &= \text{Polyn} \left[ (\mathbf{I} + \mathbf{S}(z)) \left( \frac{1}{z} \mathbf{K} - \mathbf{D}(z) \right) (\mathbf{I} + \mathbf{S}(z))^{-1} \right] \\ &= -(\mathbf{I} + \mathbf{S}(z)) \mathbf{D}(z) (\mathbf{I} + \mathbf{S}(z))^{-1} + \tilde{\mathbf{S}}(z), \end{aligned} \quad (6.14)$$

where

$$\tilde{\mathbf{S}}(z) = \mathcal{O}(z^{-1}), \quad z \rightarrow \infty,$$

and  $\text{Polyn}[\cdot]$  stands for the polynomial part of the expansion of the argument at infinity. Direct computations yield that

$$\mathbf{R}_{\vec{k}}(z) = -\mathbf{D}(z) + \begin{pmatrix} \mathcal{O}(z^{m-2}) & \mathcal{O}(z^{m-2}) & \mathcal{O}(z^{m-2}) \\ \mathcal{O}(z^{m-2}) & \mathcal{O}(z^{m-3}) & \mathcal{O}(z^{m-3}) \\ \mathcal{O}(z^{m-2}) & \mathcal{O}(z^{m-3}) & \mathcal{O}(z^{m-3}) \end{pmatrix}, \quad (6.15)$$

giving us (6.7). Using the cyclic property of the trace in (6.14) we get that

$$\text{Tr } \mathbf{R}_{\vec{k}}(z) = -\text{Tr } \mathbf{D}(z) = -V'(z),$$

and

$$\text{Tr } \mathbf{R}_{\vec{k}}^2 = \text{Tr } \mathbf{D}^2(z) - 2 \text{Tr}(\mathbf{D}(z) \tilde{\mathbf{S}}(z)) + \mathcal{O}(z^{m-2}).$$

This proves (6.8)–(6.9).

Finally, by (6.13),

$$\det \mathbf{R}_{\vec{k}}(z) = \det \left( \frac{1}{z} \mathbf{K} - \mathbf{D}(z) \right) \det \left( \mathbf{I} + \frac{1}{z} \mathbf{E}(z) \right),$$

where

$$\mathbf{E}(z) := \frac{z}{N} \left( \frac{1}{z} \mathbf{K} - \mathbf{D}(z) \right)^{-1} (\mathbf{I} + \mathbf{S}(z))^{-1} \mathbf{S}'(z).$$

Notice that

$$\left( \frac{1}{z} \mathbf{K} - \mathbf{D}(z) \right)^{-1} = \text{diag} \left( \frac{1}{-V'(z) + \frac{k_1+k_2}{Nz}}, \frac{-1}{a_1 + \frac{k_1}{Nz}}, \frac{-1}{a_2 + \frac{k_2}{Nz}} \right) = \mathcal{O}(1), \quad z \rightarrow \infty,$$

so that

$$\text{Tr } \mathbf{E}(z) = \mathcal{O}(z^{-1}), \quad z \rightarrow \infty.$$

Using that

$$\det \left( \mathbf{I} + \frac{1}{z} \mathbf{E}(z) \right) = 1 + \frac{1}{z} \text{Tr } \mathbf{E}(z) + \mathcal{O}(z^{-2}) = 1 + \mathcal{O}(z^{-2}),$$

and since

$$\det \left( \frac{1}{z} \mathbf{K} - \mathbf{D}(z) \right) = \left( -V'(z) + \frac{k_1+k_2}{Nz} \right) \left( a_1 + \frac{k_1}{Nz} \right) \left( a_2 + \frac{k_2}{Nz} \right),$$

we conclude that

$$\det \mathbf{R}_{\vec{k}}(z) = \text{Polyn} \left[ \left( -V'(z) + \frac{k_1+k_2}{Nz} \right) \left( a_1 + \frac{k_1}{Nz} \right) \left( a_2 + \frac{k_2}{Nz} \right) (1 + \mathcal{O}(z^{-2})) \right],$$

which gives (6.10).  $\square$

*Proof of Theorem D.* With  $k_1 = n_1$  and  $k_2 = n_2$ , the explicit expression for  $\mathbf{Y}$  in (6.1) and relation (6.4) shows that the first column of  $\mathbf{T}$  is the vector of wave functions  $\Psi_{\vec{n}}$ , and (2.25) is a consequence of (6.6).

Equation (2.26) is the same as (6.7).

To verify that the coefficients in (2.27) satisfy (2.7), recall that for any  $3 \times 3$  matrix  $\mathbf{M}$ ,

$$\det(\xi \mathbf{I} + \mathbf{M}) = \xi^3 + (\text{Tr } \mathbf{M}) \xi^2 + \frac{1}{2} ((\text{Tr } \mathbf{M})^2 - \text{Tr}(\mathbf{M}^2)) \xi + \det \mathbf{M} \quad (6.16)$$

and

$$\det \mathbf{M} = \frac{1}{6} ((\text{Tr } \mathbf{M})^3 - 3 \text{Tr } \mathbf{M} \text{Tr}(\mathbf{M}^2) + 2 \text{Tr}(\mathbf{M}^3)). \quad (6.17)$$

The conditions on the coefficients of (2.27) are then immediate from (6.8)–(6.10).  $\square$

## 6.2. Asymptotic distribution.

In order to prove Theorem E, we need to analyze the entries of the coefficient matrix  $\mathbf{R}_{\vec{k}}$  in (2.25) in more detail. This analysis is split into some lemmas.

**Lemma 6.3.** *The biorthogonal functions  $P_{\vec{k}}$  and  $Q_{\vec{j}}$  in (5.3) enjoy the following properties.*

(i) Whenever  $|\vec{k}| < |\vec{j}|$ ,

$$P_{\vec{k}} \mathcal{C}[Q_{\vec{j}}] = \mathcal{C}[P_{\vec{k}} Q_{\vec{j}}],$$

and

$$P_{\vec{k}} \mathcal{C}[Q_{\vec{k}}] = -1 + \mathcal{C}[P_{\vec{k}} Q_{\vec{k}}].$$

(ii) For any  $j_2, k_2$ ,

$$A_{\vec{j}}^{(1)} \mathcal{C}_1[P_{\vec{k}}] = \begin{cases} \mathcal{C}_1[A_{\vec{j}}^{(1)} P_{\vec{k}}], & \text{if } j_1 \leq k_1 + 1, \\ -1 + \mathcal{C}_1[A_{\vec{j}}^{(1)} P_{\vec{k}}], & \text{if } j_1 = k_1 + 2. \end{cases}$$

(iii) For any  $j_1, k_1$ ,

$$A_{\vec{j}}^{(2)} \mathcal{C}_2[P_{\vec{k}}] = \begin{cases} \mathcal{C}_2[A_{\vec{j}}^{(2)} P_{\vec{k}}], & \text{if } j_2 \leq k_2 + 1, \\ -1 + \mathcal{C}_2[A_{\vec{j}}^{(2)} P_{\vec{k}}], & \text{if } j_2 = k_2 + 2. \end{cases}$$

*Proof.* Write

$$P_{\vec{k}}(z) \mathcal{C}[Q_{\vec{j}}](z) = - \int Q_{\vec{j}}(x) \frac{P_{\vec{k}}(x) - P_{\vec{k}}(z)}{x - z} dx + \mathcal{C}[P_{\vec{k}} Q_{\vec{j}}](z).$$

The fraction on the integrand is a monic polynomial in  $x$  of degree  $k_1 + k_2 - 1$ , and part (i) then follows from (5.3).

To prove (ii), we write similarly

$$A_{\vec{j}}^{(1)}(z) \mathcal{C}_1[P_{\vec{k}}](z) = - \int P_{\vec{k}}(x) \frac{A_{\vec{j}}^{(1)}(x) - A_{\vec{j}}^{(1)}(z)}{x - z} e^{-NV_1(x)} dx + \mathcal{C}_1[A_{\vec{j}}^{(1)} P_{\vec{k}}](z)$$

From (5.2) the quotient in the integrand is a polynomial of degree  $j_1 - 2$  with leading coefficient  $\gamma_{j-\vec{e}_1}^{(1)}$ . The orthogonality relations (2.23) imply the first identity in (ii), and also that

$$A_{k_1+2,j_2}^{(1)}(z)\mathcal{C}_1[P_{\vec{k}}](z) = -\gamma_{\vec{k}+\vec{e}_1}^{(1)} \int x^{k_1} P_{\vec{k}}(x) e^{-NV_1(x)} dx + \mathcal{C}_1[A_{\vec{j}}^{(1)} P_{\vec{k}}](z),$$

and the second identity then follows from (5.4). Part (iii) follows analogously.  $\square$

**Lemma 6.4.** *If  $(\vec{n}_N)$  is a sequence of multi-indices that satisfies the assumption of Theorem E, then the coefficients of the polynomial (of degree  $\leq m - 2$ )*

$$\text{Polyn}(V'(z)\mathcal{C}[P_{\vec{n}_N} Q_{\vec{n}_N}](z))$$

*remain bounded as  $N \rightarrow \infty$ .*

*The same holds true if we replace  $P_{\vec{n}_N} Q_{\vec{n}_N}$  in the expression above either by*

$$P_{\vec{n}_N-\vec{e}_j} Q_{\vec{n}_N}, \quad P_{\vec{n}_N} Q_{\vec{n}_N+\vec{e}_j}, \quad \text{or} \quad P_{\vec{n}_N-\vec{e}_j} Q_{\vec{n}_N+\vec{e}_i}, \quad i, j = 1, 2. \quad (6.18)$$

*Proof.* For simplicity we omit the  $N$ -dependence on the multi-indices and write  $\vec{n}_N = \vec{n}$ . We start with  $P_{\vec{n}} Q_{\vec{n}}$ . As the coefficients of  $V$  are independent of  $N, n_1, n_2$ , it is enough to verify that the coefficients of the polynomial part of  $z^j \mathcal{C}[P_{\vec{n}} Q_{\vec{n}}]$  are bounded for  $j = 0, \dots, m - 1$ . The case  $j = 0$  is obvious because the Cauchy transform is  $\mathcal{O}(z^{-1})$  as  $z \rightarrow \infty$ .

For  $1 \leq j \leq m - 1$ , we proceed similarly as in the previous proof and write

$$z^j \mathcal{C}[P_{\vec{n}} Q_{\vec{n}}](z) = - \int P_{\vec{n}}(x) Q_{\vec{n}}(x) \frac{x^j - z^j}{x - z} dx + \mathcal{C}[x^j P_{\vec{n}} Q_{\vec{n}}](z), \quad (6.19)$$

so

$$\begin{aligned} \text{Polyn}(z^j \mathcal{C}[P_{\vec{n}} Q_{\vec{n}}](z)) &= - \sum_{s=0}^{j-1} z^{j-1-s} \int x^s P_{\vec{n}}(x) Q_{\vec{n}}(x) dx \\ &= - \sum_{s=1}^{j-1} z^{j-1-s} \int x^s P_{\vec{n}}(x) Q_{\vec{n}}(x) dx, \end{aligned} \quad (6.20)$$

where we have used also biorthogonality (5.3). By Proposition 5.10, integrals

$$\int x^s P_{\vec{n}_k}(x) Q_{\vec{n}_k}(x) dx, \quad s = 1, \dots, m - 2$$

are uniformly bounded (at least, for  $k$  large enough, e.g.  $k \geq 2m$ ).

Finally, for the expressions in (6.18) we apply the same procedure and reduce to integrals also of the forms considered in Proposition 5.10; we skip the details.  $\square$



The entries of the matrix  $\mathbf{R}_{\vec{k}}$  involve norm constants for  $P_{\vec{k}}$  and as such they may not be bounded. So instead of looking at  $\mathbf{R}_{\vec{k}}$ , we modify it into

$$\mathbf{W}_{\vec{k}}(z) = \mathbf{\Lambda}_{\vec{k}}^{-1} \mathbf{R}_{\vec{k}}(z) \mathbf{\Lambda}_{\vec{k}}, \quad (6.21)$$

where  $\mathbf{\Lambda}_{\vec{k}}$  is as in (6.2).

Consequently, all the equations (6.7)–(6.10) remain valid if we replace  $\mathbf{R}_{\vec{k}}$  by  $\mathbf{W}_{\vec{k}}$ . In particular, the entries of  $\mathbf{W}_{\vec{k}}$  are polynomials of degree at most  $m - 1$ .

**Proposition 6.5.** *Under the assumptions of Theorem E, the coefficients of the polynomial entries of the matrix  $\mathbf{W}_{\vec{n}_N}$  remain bounded as  $N \rightarrow \infty$ .*

*Proof.* For ease of notation, during this proof we again suppress the  $N$  dependence of the multi-indices and simply denote  $\vec{n}_N = \vec{n}$ .

We start rewriting the first identity in (6.13), which gives us

$$\mathbf{W}_{\vec{n}}(z) = \frac{1}{N} \mathbf{\Lambda}_{\vec{n}}^{-1} (\mathbf{Y}(z) \mathbf{F}(z))' \mathbf{F}(z)^{-1} \mathbf{Y}(z)^{-1} \mathbf{\Lambda}_{\vec{n}} = \mathbf{\Lambda}_{\vec{n}}^{-1} \text{Polyn}(\mathbf{Y}(z) \mathbf{D}(z) \mathbf{Y}^{-1}(z)) \mathbf{\Lambda}_{\vec{n}},$$

where we also used that the entries of  $\mathbf{W}_{\vec{n}}$  are polynomials in  $z$ . Using (6.1), (6.3) and (5.6), we compute

$$\begin{aligned} \mathbf{W}_{\vec{n}} = \text{Polyn} \left[ \begin{pmatrix} P_{\vec{n}} & \mathcal{C}_1[P_{\vec{n}}] & \mathcal{C}_2[P_{\vec{n}}] \\ -2\pi i P_{\vec{n}-\vec{e}_1} & \mathcal{C}_1[P_{\vec{n}-\vec{e}_1}] & \mathcal{C}_2[P_{\vec{n}-\vec{e}_1}] \\ -2\pi i P_{\vec{n}-\vec{e}_2} & \mathcal{C}_1[P_{\vec{n}-\vec{e}_2}] & \mathcal{C}_2[P_{\vec{n}-\vec{e}_2}] \end{pmatrix} \text{diag}(1, 2\pi i, 2\pi i)^{-1} \mathbf{D} \right. \\ \left. \times \begin{pmatrix} -\mathcal{C}[Q_{\vec{n}}] & -\frac{1}{2\pi i} \mathcal{C}[Q_{\vec{n}+\vec{e}_1}] & -\frac{1}{2\pi i} \mathcal{C}[Q_{\vec{n}+\vec{e}_2}] \\ 2\pi i A_{\vec{n}}^{(1)} & A_{\vec{n}+\vec{e}_1}^{(1)} & A_{\vec{n}+\vec{e}_2}^{(1)} \\ 2\pi i A_{\vec{n}}^{(2)} & A_{\vec{n}+\vec{e}_1}^{(2)} & A_{\vec{n}+\vec{e}_2}^{(2)} \end{pmatrix} \text{diag}(1, a_{\vec{n}}^{(1)}, a_{\vec{n}}^{(2)}) \right]. \end{aligned}$$

From this identity we can compute the entries of  $\mathbf{W}_{\vec{n}} = (\mathbf{W}(i, j))_{i,j=0}^2$ . The  $(1, 1)$  entry, denoted  $\mathbf{W}_{\vec{n}}(0, 0)$ , reads

$$\mathbf{W}(0, 0) = \text{Polyn} \left[ -V' P_{\vec{n}} \mathcal{C}[Q_{\vec{n}}] + a_1 A_{\vec{n}}^{(1)} \mathcal{C}_1[P_{\vec{n}}] + a_2 A_{\vec{n}}^{(2)} \mathcal{C}_2[P_{\vec{n}}] \right] = -V' - \text{Polyn} [V' P_{\vec{n}} Q_{\vec{n}}],$$

where in the last step we used Lemma 6.3. By Lemma 6.4 this last expression has bounded polynomial coefficients as  $N \rightarrow \infty$ . The other entries are dealt with similarly: with

$$(\mathbf{c}_0, \mathbf{c}_1, \mathbf{c}_2) = (1, -2\pi i, -2\pi i), \quad a_{\vec{n}}^{(0)} = 1,$$

the other entries are

$$\mathbf{W}(i, j) = \frac{a_{\vec{n}}^{(j)}}{\mathbf{c}_j} \text{Polyn} \left( -\mathbf{c}_i V' P_{\vec{n}-\vec{e}_i} \mathcal{C}[Q_{\vec{n}+\vec{e}_j}] + a_1 A_{\vec{n}+\vec{e}_j}^{(1)} \mathcal{C}_1[P_{\vec{n}-\vec{e}_i}] + a_2 A_{\vec{n}+\vec{e}_j}^{(2)} \mathcal{C}_2[P_{\vec{n}-\vec{e}_i}] \right)$$

which can be shown to be bounded using Lemmas 6.3 and 6.4 again.  $\square$

We are finally ready to prove Theorem E.

*Proof of Theorem E.* We start by rewriting (2.25) in terms of the matrix  $\mathbf{W}_{\vec{n}_N}$  in (6.21), namely

$$\frac{d}{dz} \Phi_{\vec{n}_N}(z) = N \mathbf{W}_{\vec{n}_N}(z) \Phi_{\vec{n}_N}(z), \quad \Phi_{\vec{k}}(z) := \Lambda_{\vec{k}}^{-1} \Psi_{\vec{k}}(z). \quad (6.22)$$

Using (2.24) we obtain that for any  $z$  for which the entries of  $\Phi_{\vec{n}_N}$  do not vanish

$$\frac{d}{dz} \Phi_{\vec{n}_N}(z) = N (\mathbf{Q}_{\vec{n}_N}(z) - V'(z) \mathbf{I}) \Phi_{\vec{n}_N}(z) \quad (6.23)$$

where

$$\mathbf{Q}_{\vec{k}}(z) := \text{diag} \left( \frac{P'_{\vec{k}}(z)}{NP_{\vec{k}}(z)}, \frac{P'_{\vec{k}-\vec{e}_1}(z)}{NP_{\vec{k}-\vec{e}_1}(z)}, \frac{P'_{\vec{k}-\vec{e}_2}(z)}{NP_{\vec{k}-\vec{e}_2}(z)} \right).$$

Comparing (6.22) and (6.23), we conclude that

$$(\mathbf{Q}_{\vec{n}_N}(z) - V'(z) \mathbf{I} - \mathbf{W}_{\vec{n}_N}(z)) \Phi_{\vec{n}_N}(z) = \mathbf{0}.$$

which is valid for  $z \in \mathbb{C}$  for which the entries of  $\Phi_{\vec{n}_N}$  are nonzero. Thus, for such  $z$  we conclude that

$$\det(V'(z) \mathbf{I} - \mathbf{Q}_{\vec{n}_N}(z) + \mathbf{W}_{\vec{n}_N}(z)) = 0. \quad (6.24)$$

The last step is to take the limit in this equation. First, Proposition 6.5 tells us that we can extract a subsequence  $(\vec{n}_k) \subset (\vec{n}_N)$  for which all the entries of  $\mathbf{W}_{\vec{n}_k}$  converge, say to  $\mathbf{W}_\infty$ . Because  $\mathbf{W}_{\vec{n}_k}$  also satisfies (6.8)–(6.10), for any  $\xi, z \in \mathbb{C}$  in compacts we get the uniform convergence

$$\det(\xi \mathbf{I} + \mathbf{W}_{\vec{n}_k}(z)) \xrightarrow{k \rightarrow \infty} \det(\xi \mathbf{I} + \mathbf{W}_\infty(z)) = \xi^3 + p_2(z) \xi^2 + p_1(z) \xi + p_0(z),$$

where the polynomial coefficients satisfy (2.7).

Next, under the assumptions of Theorem E,

$$\lim \mathbf{Q}_{\vec{n}_k}(z) = -C^\lambda(z) \mathbf{I}, \quad z \in \mathbb{C} \setminus \mathbb{R},$$

and (6.24) combined with continuity of the determinant gives

$$\det((C^\lambda(z) + V'(z)) \mathbf{I} + \mathbf{W}_\infty(z)) = 0, \quad z \in \mathbb{C} \setminus \mathbb{R},$$

which by analytic continuation extends to  $z \in \mathbb{C} \setminus \text{supp } \lambda$ . This proves the claim that  $\xi(z)$  is indeed a solution to a spectral curve in the sense of Definition 2.6.

Finally, as  $\mathbf{Q}_{\vec{n}_k}$  and  $\Lambda_{\vec{n}_k}$  are diagonal matrices, the relation (6.21) tells us that

$$\det(\xi \mathbf{I} + \mathbf{R}_{\vec{n}_k}(z)) = \det(\xi \mathbf{I} + \mathbf{W}_{\vec{n}_k}(z)) \rightarrow \det(\xi \mathbf{I} + \mathbf{W}_\infty(z)),$$

showing that the spectral curve for  $\xi(z) = C^\lambda(z) + V'(z)$  indeed is a limit of (2.27), concluding the proof.  $\square$

## ACKNOWLEDGMENTS

The first author was partially supported Simons Foundation Collaboration Grants for Mathematicians (grant 710499) and by the European Regional Development Fund along with Spanish Government (grant MTM2017-89941-P) and Junta de Andalucía (grant UAL18-FQM-B025-A, as well as research group FQM-229 and Instituto Interuniversitario Carlos I de Física Teórica y Computacional), and by the University of Almería (Campus de Excelencia Internacional del Mar CEIMAR).

Most of this work was carried out while the second author was a Postdoctoral Assistant Professor in the Department of Mathematics at the University of Michigan, and also while he was visiting Fudan University, China. He thanks the hospitality and excellent research atmosphere at these institutions. The second author also acknowledges his current support by São Paulo Research Foundation under grants # 2019/16062-1 and # 2020/02506-2.

Both authors have benefited from fruitful discussions with several colleagues, such as Andrew Celsus, Adrien Hardy, Arno Kuijlaars, Ken McLaughlin, Walter Van Assche, Maxim Yattselev and Lun Zhang, to mention a few. Last but not least, we are deeply indebted to the anonymous referee, who did an outstanding job.

## REFERENCES

- [1] O. Ajanki, L. Erdos, and T. Krüger. Singularities of solutions to quadratic vector equations on the complex upper half-plane. *Comm. Pure Appl. Math.*, 70(9):1672–1705, 2017.
- [2] O. H. Ajanki, L. Erdos, and T. Krüger. Universality for general Wigner-type matrices. *Probab. Theory Related Fields*, 169(3-4):667–727, 2017.
- [3] G. W. Anderson and O. Zeitouni. A law of large numbers for finite-range dependent random matrices. *Comm. Pure Appl. Math.*, 61(8):1118–1154, 2008.
- [4] A. I. Aptekarev, W. Van Assche, and M. L. Yattselev. Hermite-Padé approximants for a pair of Cauchy transforms with overlapping symmetric supports. *ArXiv:1505.03993*, page 52 pp.
- [5] A. I. Aptekarev, P. M. Bleher, and A. B. J. Kuijlaars. Large  $n$  limit of Gaussian random matrices with external source. II. *Comm. Math. Phys.*, 259(2):367–389, 2005.
- [6] A. I. Aptekarev, V. Kalyagin, G. López Lagomasino, and I. A. Rocha. On the limit behavior of recurrence coefficients for multiple orthogonal polynomials. *J. Approx. Theory*, 139(1-2):346–370, 2006.
- [7] A. I. Aptekarev, A. B. J. Kuijlaars, and W. Van Assche. Asymptotics of Hermite-Padé rational approximants for two analytic functions with separated pairs of branch points (case of genus 0). *Int. Math. Res. Pap. IMRP*, 2007(4):128 pp., 2008.
- [8] A. I. Aptekarev and V. G. Lysov. Systems of Markov functions generated by graphs and the asymptotics of their Hermite-Padé approximants. *Mat. Sb.*, 201(2):29–78, 2010.
- [9] A. I. Aptekarev, V. G. Lysov, and D. N. Tulyakov. The global eigenvalue distribution regime of random matrices with an anharmonic potential and an external source. *Teoret. Mat. Fiz.*, 159(1):34–57, 2009.
- [10] A. I. Aptekarev, V. G. Lysov, and D. N. Tulyakov. Random matrices with an external source and the asymptotics of multiple orthogonal polynomials. *Mat. Sb.*, 202(2):3–56, 2011.
- [11] A. I. Aptekarev and D. N. Tulyakov. Nuttall’s Abelian integral on the Riemann surface of the cube root of a polynomial of degree 3. *Izv. Ross. Akad. Nauk Ser. Mat.*, 80(6):5–42, 2016.
- [12] M. R. Atkin, T. Claeys, and F. Mezzadri. Random matrix ensembles with singularities and a hierarchy of Painlevé III equations. *Int. Math. Res. Not. IMRN*, (8):2320–2375, 2016.
- [13] J. Baik, P. Deift, K. T.-R. McLaughlin, P. Miller, and X. Zhou. Optimal tail estimates for directed last passage site percolation with geometric random variables. *Adv. Theor. Math. Phys.*, 5(6):1207–1250, 2001.
- [14] J. Baik and D. Wang. On the largest eigenvalue of a Hermitian random matrix model with spiked external source I. Rank 1 case. *Int. Math. Res. Not. IMRN*, (22):5164–5240, 2011.

- [15] J. Baik and D. Wang. On the largest eigenvalue of a Hermitian random matrix model with spiked external source I. Rank 1 case. *Int. Math. Res. Not. IMRN*, (22):5164–5240, 2011.
- [16] J. Baik and D. Wang. On the largest eigenvalue of a Hermitian random matrix model with spiked external source II: Higher rank cases. *Int. Math. Res. Not. IMRN*, (14):3304–3370, 2013.
- [17] M. Bertola. Boutroux curves with external field: equilibrium measures without a variational problem. *Anal. Math. Phys.*, 1(2-3):167–211, 2011.
- [18] M. Bertola and T. Bothner. Universality conjecture and results for a model of several coupled positive-definite matrices. *Comm. Math. Phys.*, 337(3):1077–1141, 2015.
- [19] M. Bertola, R. Buckingham, S. Y. Lee, and V. Pierce. Spectra of random Hermitian matrices with a small-rank external source: the critical and near-critical regimes. *J. Stat. Phys.*, 146(3):475–518, 2012.
- [20] M. Bertola, R. Buckingham, S. Y. Lee, and V. Pierce. Spectra of random Hermitian matrices with a small-rank external source: the supercritical and subcritical regimes. *J. Stat. Phys.*, 153(4):654–697, 2013.
- [21] M. Bertola, B. Eynard, and J. Harnad. Duality, biorthogonal polynomials and multi-matrix models. *Comm. Math. Phys.*, 229(1):73–120, 2002.
- [22] M. Bertola, B. Eynard, and J. Harnad. Differential systems for biorthogonal polynomials appearing in 2-matrix models and the associated Riemann-Hilbert problem. *Comm. Math. Phys.*, 243(2):193–240, 2003.
- [23] M. Bertola, B. Eynard, and J. Harnad. The duality of spectral curves that arises in two-matrix models. *Teoret. Mat. Fiz.*, 134(1):32–45, 2003.
- [24] M. Bertola, B. Eynard, and J. Harnad. Partition functions for matrix models and isomonodromic tau functions. *J. Phys. A*, 36(12):3067–3083, 2003. Random matrix theory.
- [25] M. Bertola, B. Eynard, and J. Harnad. Semiclassical orthogonal polynomials, matrix models and isomonodromic tau functions. *Comm. Math. Phys.*, 263(2):401–437, 2006.
- [26] M. Bertola and M. Y. Mo. Commuting difference operators, spinor bundles and the asymptotics of orthogonal polynomials with respect to varying complex weights. *Adv. Math.*, 220(1):154–218, 2009.
- [27] M. Bertola and A. Tovbis. Asymptotics of orthogonal polynomials with complex varying quartic weight: global structure, critical point behavior and the first Painlevé equation. *Constr. Approx.*, 41(3):529–587, 2015.
- [28] J.-E. Björk, J. Borcea, and R. Bøgvad. Subharmonic configurations and algebraic Cauchy transforms of probability measures. In *Notions of positivity and the geometry of polynomials*, Trends Math., pages 39–62. Birkhäuser/Springer Basel AG, Basel, 2011.
- [29] P. Bleher and A. Deaño. Painlevé I double scaling limit in the cubic random matrix model. *Random Matrices Theory Appl.*, 5(2):1650004, 58, 2016.
- [30] P. M. Bleher and A. Deaño. Topological expansion in the cubic random matrix model. *Int. Math. Res. Not. IMRN*, (12):2699–2755, 2013.
- [31] P. M. Bleher, S. Delvaux, and A. B. J. Kuijlaars. Random matrix model with external source and a constrained vector equilibrium problem. *Comm. Pure Appl. Math.*, 64(1):116–160, 2011.
- [32] P. M. Bleher and A. R. Its. Semiclassical asymptotics of orthogonal polynomials, Riemann-Hilbert problem, and universality in the matrix model. *Ann. of Math.*, 150(1):185–266, 1999.
- [33] P. M. Bleher and A. B. J. Kuijlaars. Large  $n$  limit of Gaussian random matrices with external source. I. *Comm. Math. Phys.*, 252(1-3):43–76, 2004.
- [34] P. M. Bleher and A. B. J. Kuijlaars. Random matrices with external source and multiple orthogonal polynomials. *Int. Math. Res. Not.*, (3):109–129, 2004.
- [35] P. M. Bleher and A. B. J. Kuijlaars. Large  $n$  limit of Gaussian random matrices with external source. III. Double scaling limit. *Comm. Math. Phys.*, 270(2):481–517, 2007.
- [36] P. M. Bleher and G. L. F. Silva. The mother body phase transition in the normal matrix model. *Mem. Amer. Math. Soc.*, 265(1289):v+144, 2020.
- [37] R. Bøgvad and B. Shapiro. On mother body measures with algebraic Cauchy transform. *Enseign. Math.*, 62(1-2):117–142, 2016.
- [38] J. Borcea, R. Bøgvad, and B. Shapiro. Homogenized spectral problems for exactly solvable operators: asymptotics of polynomial eigenfunctions. *Publ. Res. Inst. Math. Sci.*, 45(2):525–568, 2009.
- [39] J. Borcea, R. Bøgvad, and B. Shapiro. Corrigendum: “Homogenized spectral problems for exactly solvable operators: asymptotics of polynomial eigenfunctions”. *Publ. Res. Inst. Math. Sci.*, 48(1):229–233, 2012.

- [40] J. Breuer and M. Duits. Universality of mesoscopic fluctuations for orthogonal polynomial ensembles. *Comm. Math. Phys.*, 342(2):491–531, 2016.
- [41] J. Breuer and M. Duits. Central limit theorems for biorthogonal ensembles and asymptotics of recurrence coefficients. *J. Amer. Math. Soc.*, 30(1):27–66, 2017.
- [42] E. Brézin and S. Hikami. Level spacing of random matrices in an external source. *Phys. Rev. E*, 58:7176–7185, 1998.
- [43] E. Brézin and S. Hikami. Universal singularity at the closure of a gap in a random matrix theory. *Phys. Rev. E*, 57:4140–4149, 1998.
- [44] E. Brézin and S. Hikami. *Random matrix theory with an external source*, volume 19 of *Springer Briefs in mathematical physics*. Springer Singapore, 2016.
- [45] T. Claeys, A. B. J. Kuijlaars, and M. Vanlessen. Multi-critical unitary random matrix ensembles and the general Painlevé II equation. *Ann. of Math.*, 168:601–642, 2008.
- [46] T. Claeys, T. Neuschel, and M. Venker. Boundaries of sine kernel universality for Gaussian perturbations of Hermitian matrices. *Random Matrices Theory Appl.*, 8(3):1950011, 50, 2019.
- [47] D. Dai. Asymptotics of orthogonal polynomials and the Painlevé transcendents. In *Frontiers in orthogonal polynomials and q-series*, volume 1 of *Contemp. Math. Appl. Monogr. Expo. Lect. Notes*, pages 189–212. World Sci. Publ., Hackensack, NJ, 2018.
- [48] P. Deift. *Orthogonal Polynomials and Random Matrices: A Riemann-Hilbert Approach*. Number 3 in Courant Lecture Notes. American Mathematical Society, 2000.
- [49] S. A. Denisov and M. L. Yattselev. Spectral theory of jacobi matrices on trees whose coefficients are generated by multiple orthogonality. <https://arxiv.org/abs/2008.08210>, 2020.
- [50] M. Duits. Painlevé kernels in Hermitian matrix models. *Constr. Approx.*, 39(1):173–196, 2014.
- [51] M. Duits, B. Fahs, and R. Kozhan. Global fluctuations for multiple orthogonal polynomial ensembles. *ArXiv:1912.04599*, 2019.
- [52] M. Duits and K. Johansson. On mesoscopic equilibrium for linear statistics in Dyson’s Brownian motion. *Mem. Amer. Math. Soc.*, 255(1222):v+118, 2018.
- [53] M. Duits and A. B. J. Kuijlaars. Painlevé I asymptotics for orthogonal polynomials with respect to a varying quartic weight. *Nonlinearity*, 19(10):2211–2245, 2006.
- [54] M. Duits and A. B. J. Kuijlaars. Universality in the two-matrix model: a Riemann-Hilbert steepest-descent analysis. *Comm. Pure Appl. Math.*, 62(8):1076–1153, 2009.
- [55] L. Erdős and K. Schnelli. Universality for random matrix flows with time-dependent density. *Ann. Inst. Henri Poincaré Probab. Stat.*, 53(4):1606–1656, 2017.
- [56] L. Erdős and H.-T. Yau. *A dynamical approach to random matrix theory*, volume 28 of *Courant Lecture Notes in Mathematics*. Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, RI, 2017.
- [57] L. Erdos, T. Krüger, and D. Schröder. Cusp universality for random matrices I: local law and the complex hermitian case. *ArXiv:1809.03971*, page 50 pages, 2018.
- [58] M. Haneczok and W. Van Assche. Interlacing properties of zeros of multiple orthogonal polynomials. *J. Math. Anal. Appl.*, 389(1):429–438, 2012.
- [59] A. Hardy. Average characteristic polynomials of determinantal point processes. *Ann. Inst. Henri Poincaré Probab. Stat.*, 51(1):283–303, 2015.
- [60] A. Hardy. Polynomial ensembles and recurrence coefficients. *Constr. Approx.*, 48(1):137–162, 2018.
- [61] M. E. H. Ismail. *Classical and quantum orthogonal polynomials in one variable*, volume 98 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2005. With two chapters by Walter Van Assche, With a foreword by Richard A. Askey.
- [62] 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.
- [63] J. A. Jenkins. *Univalent functions and conformal mapping*. Ergebnisse der Mathematik und ihrer Grenzgebiete. Neue Folge, Heft 18. Reihe: Moderne Funktionentheorie. Springer-Verlag, Berlin-Göttingen-Heidelberg, 1958.
- [64] K. Johansson. On fluctuations of eigenvalues of random hermitian matrices. *Duke Math. J.*, 91(1):151–204, 1998.

- [65] A. B. J. Kuijlaars. Multiple orthogonal polynomial ensembles. In *Recent trends in orthogonal polynomials and approximation theory*, volume 507 of *Contemp. Math.*, pages 155–176. Amer. Math. Soc., Providence, RI, 2010.
- [66] A. B. J. Kuijlaars. Multiple orthogonal polynomials in random matrix theory. In *Proceedings of the International Congress of Mathematicians. Volume III*, pages 1417–1432, New Delhi, 2010. Hindustan Book Agency.
- [67] A. B. J. Kuijlaars, H. Stahl, W. Van Assche, and F. Wielonsky. Asymptotique des approximants de Hermite-Padé quadratiques de la fonction exponentielle et problèmes de Riemann-Hilbert. *C. R. Math. Acad. Sci. Paris*, 336(11):893–896, 2003.
- [68] A. B. J. Kuijlaars and A. Tovbis. The supercritical regime in the normal matrix model with cubic potential. *Adv. Math.*, 283:530–587, 2015.
- [69] G. Lambert. Limit theorems for biorthogonal ensembles and related combinatorial identities. *Adv. Math.*, 329:590–648, 2018.
- [70] B. Landon, P. Sosoe, and H.-T. Yau. Fixed energy universality of Dyson Brownian motion. *Adv. Math.*, 346:1137–1332, 2019.
- [71] B. Landon and H.-T. Yau. Convergence of local statistics of Dyson Brownian motion. *Comm. Math. Phys.*, 355(3):949–1000, 2017.
- [72] A. Martínez-Finkelshtein. Equilibrium problems of potential theory in the complex plane. In *Orthogonal polynomials and special functions*, volume 1883 of *Lecture Notes in Math.*, pages 79–117. Springer, Berlin, 2006.
- [73] A. Martínez-Finkelshtein, P. Martínez-González, and F. Thabet. Trajectories of quadratic differentials for Jacobi polynomials with complex parameters. *Comput. Methods and Function Theory*, 16(3):347–364, 2016.
- [74] A. Martínez-Finkelshtein and R. Orive. Riemann-Hilbert analysis of Jacobi polynomials orthogonal on a single contour. *J. Approx. Theory*, 134(2):137–170, 2005.
- [75] A. Martínez-Finkelshtein and E. A. Rakhmanov. Critical measures, quadratic differentials, and weak limits of zeros of Stieltjes polynomials. *Comm. Math. Phys.*, 302(1):53–111, 2011.
- [76] A. Martínez-Finkelshtein and G. L. F. Silva. Critical measures for vector energy: Global structure of trajectories of quadratic differentials. *Adv. Math.*, 302:1137–1232, 2016.
- [77] A. Martínez-Finkelshtein and G. L. F. Silva. Critical measures for vector energy: asymptotics of non-diagonal multiple orthogonal polynomials for a cubic weight. *Adv. Math.*, 349:246–315, 2019.
- [78] K. T.-R. McLaughlin. Asymptotic analysis of random matrices with external source and a family of algebraic curves. *Nonlinearity*, 20(7):1547–1571, 2007.
- [79] E. M. Nikishin and V. N. Sorokin. *Rational approximations and orthogonality*, volume 92 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI, 1991. Translated from the Russian by Ralph P. Boas.
- [80] L. A. Pastur. The spectrum of random matrices. *Teoret. Mat. Fiz.*, 10(1):102–112, 1972.
- [81] C. Pommerenke. *Univalent functions*. Vandenhoeck & Ruprecht, Göttingen, 1975. With a chapter on quadratic differentials by Gerd Jensen, *Studia Mathematica/Mathematische Lehrbücher*, Band XXV.
- [82] E. A. Rakhmanov. On the asymptotics of Hermite-Padé polynomials for two Markov functions. *Mat. Sb.*, 202(1):133–140, 2011.
- [83] N. R. Rao and A. Edelman. The polynomial method for random matrices. *Found. Comput. Math.*, 8(6):649–702, 2008.
- [84] E. B. Saff and V. Totik. *Logarithmic potentials with external fields*, volume 316 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1997. Appendix B by Thomas Bloom.
- [85] B. Shapiro and A. Solynin. Root-counting measures of Jacobi polynomials and topological types and critical geodesics of related quadratic differentials. In *Analysis meets geometry*, Trends Math., pages 369–438. Birkhäuser/ Springer, 2017.
- [86] B. Shapiro, K. Takemura, and M. Tater. On spectral polynomials of the Heun equation. II. *Comm. Math. Phys.*, 311(2):277–300, 2012.
- [87] B. Shapiro and M. Tater. On spectral polynomials of the Heun equation. I. *J. Approx. Theory*, 162(4):766–781, 2010.
- [88] H. Stahl and V. Totik. *General orthogonal polynomials*, volume 43 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1992.

- [89] K. Strebel. *Quadratic differentials*, volume 5 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer-Verlag, Berlin, 1984.
- [90] C. A. Tracy and H. Widom. The Pearcey process. *Comm. Math. Phys.*, 263(2):381–400, 2006.
- [91] W. Van Assche. Nearest neighbor recurrence relations for multiple orthogonal polynomials. *J. Approx. Theory*, 163(10):1427–1448, 2011.
- [92] W. Van Assche, J. S. Geronimo, and A. B. J. Kuijlaars. Riemann-Hilbert problems for multiple orthogonal polynomials. In *Special functions 2000: current perspective and future directions (Tempe, AZ)*, volume 30 of *NATO Sci. Ser. II Math. Phys. Chem.*, pages 23–59. Kluwer Acad. Publ., Dordrecht, 2001.
- [93] J. Xu, E. Fan, and Y. Chen. Large  $N$ -limit for random matrices with external source with three distinct eigenvalues. *Random Matrices Theory Appl.*, 5(2):1650005, 30, 2016.
- [94] S.-X. Xu and D. Dai. Tracy-Widom distributions in critical unitary random matrix ensembles and the coupled Painlevé II system. *Comm. Math. Phys.*, 365(2):515–567, 2019.
- [95] P. Zinn-Justin. Random Hermitian matrices in an external field. *Nuclear Phys. B*, 497(3):725–732, 1997.

(AMF) DEPARTMENT OF MATHEMATICS, BAYLOR UNIVERSITY, WACO TX, USA, AND DEPARTMENT OF MATHEMATICS, UNIVERSITY OF ALMERÍA, ALMERÍA, SPAIN

*Email address:* `A.Martinez-Finkelshtein@baylor.edu`

(GS) INSTITUTO DE CIÊNCIAS MATEMÁTICAS E DE COMPUTAÇÃO, UNIVERSIDADE DE SÃO PAULO (ICMC-USP), SÃO CARLOS, SP, BRAZIL

*Email address:* `silvag@usp.br`