Commuting difference operators, spinor bundles and the asymptotics of orthogonal polynomials with respect to varying complex weights

M. Bertola[†‡12] M.Y. Mo

[†] *Centre de recherches mathématiques, Université de Montréal*
C. P. 6128, succ. centre ville, Montréal, Québec, Canada H3C 3J7

[‡] *Department of Mathematics and Statistics, Concordia University*
7141 Sherbrooke W., Montréal, Québec, Canada H4B 1R6

Abstract

The paper has three parts. In the first part we apply the theory of commuting pairs of (pseudo) difference operators to the (formal) asymptotics of orthogonal polynomials: using purely geometrical arguments we show heuristically that the asymptotics, for large degrees, of orthogonal polynomial with respect to varying weights is intimately related to certain spinor bundles on a hyperelliptic algebraic curve reproducing formulæ appearing in the works of Deift et al. on the subject.

In the second part we show that given an arbitrary nodal hyperelliptic curve satisfying certain conditions of admissibility we can reconstruct a sequence of polynomials orthogonal with respect to semiclassical complex varying weights supported on several curves in the complex plane. The strong asymptotics of these polynomials will be shown to be given by the spinors introduced in the first part using a Riemann–Hilbert analysis.

In the third part we use Strebel theory of quadratic differentials and the procedure of welding to reconstruct arbitrary admissible hyperelliptic curves. As a result we can obtain orthogonal polynomials whose zeroes may become dense on a collection of Jordan arcs forming an arbitrary forest of trivalent loop-free trees.

Contents

1 Introduction and summary

The present paper deals with the asymptotics of certain (pseudo–)orthogonal polynomials, its formal properties and connections with algebraic geometry. In order to explain the framework, let us recall the main results for ordinary orthogonal polynomials [14] in a simple exemplifying case. Let $V(x)$ be an even-degree real polynomial bounded from below and consider the Hilbert space $L^2(\mathbb{R}, \mathrm{e}^{-NV(x)}\mathrm{d}x)$. Let $p_n(x)$ be the (real) orthogonal polynomials (OP) for this measure.

One of the main goals of modern asymptotic analysis is to describe their strong asymptotic as we let $n \to \infty$ while $N \to \infty$ at the same rate.

This problem has been brilliantly solved in [6, 7, 14, 15] using the associated Riemann–Hilbert problem (RHP); indeed the matrix

$$Y(x) := \left[\begin{array}{cc} p_n(x) & \phi_n(x) \\ p_{n-1}(x) & \phi_{n-1}(x) \end{array} \right] , \qquad \phi_n(x) = \frac{1}{2i\pi} \int_{\mathbb{R}} \frac{\mathrm{e}^{-NV(s)}p_n(s)}{s-x}\mathrm{d}s \ , z \in \mathbb{C} \setminus \mathbb{R} \qquad (1\text{-}1)$$

solves a RHP with jumps on the real axis and rather simple asymptotics at $z = \infty$ [18]. More importantly the solution of this RHP **characterizes** the orthogonal polynomials; therefore if one could solve the RHP, then he/she would immediately have access to all information on the corresponding OP.

Without entering now into the details it suffices to recall that the large parameter N and the degree n of the polynomials enter explicitly and in a simple way the Riemann–Hilbert data and the asymptotics at $x = \infty$. Therefore all the "complication" of the asymptotics is controlled by the (highly transcendental) solution of the RHP.

The technique developed in the last decade of the last millennium, called "nonlinear steepest descent" method, consists in transforming this RHP into a simpler, "asymptotic" one which differs from the exact one by small controllable errors, with suitable uniformly small bounds as $N \to \infty$.

The main character of the method is the so–called G-function (which we will discuss at length in the paper). Here we only indicate that it should satisfy some general properties which guarantee the amenability of the Deift–Zhou steepest descent method.

The main logic of approach in most literature is then:

- fix the potential V;

- try to find an appropriate G-function for the given potential;

- implement the steepest descent method

In a certain sense this point of view mixes a "forward problem" (finding the asymptotics of orthogonal polynomials for a given potential) with an "inverse problem" (reconstructing the matrix solution of a monodromy/jump problem).

Our approach is "purely inverse"; namely we assign certain asymptotic data in the large N-limit (which –of course– satisfy some consistency conditions) and remount the (class of) orthogonal polynomials and potentials whose asymptotics matches our given data. In a subsequent paper [1] it is shown that for any (polynomial) potential V (possibly complex) and any "Stokes' data" for the orthogonal polynomials, it is possible to construct an appropriate G-function. Together, the present paper and loc. cit. completely solve the problem of the large–N asymptotics for these *semiclassical* orthogonal polynomials.

1.1 Asymptotics of generalized pseudo-orthogonal polynomials

In order to explain the main theorem of the asymptotic analysis we need to introduce some notations. The setting in which we move is rather algebro–geometric; the main piece of data is a hyperelliptic (nodal) curve $\mathcal{L}$ of genus g

$$y^2 = M^2(x) \prod_{j=1}^{2g+2} (x - \alpha_j) \tag{1-2}$$

satisfying the conditions of **Boutroux** (we adopt the terminology of [24]) and **admissibility** (to be explained presently). Such curve is the **asymptotic spectral curve**; the reader acquainted with the asymptotic analysis of ordinary OPs may think of this curve as the hyperelliptic curve associated with the equilibrium measure. The Boutroux condition is a transcendental reality condition

$$\oint_\gamma y \mathrm{d}x \in i\mathbb{R} \tag{1-3}$$

where γ is any closed loop on the spectral curve $\mathcal{L}$. The Boutroux condition implies that (see Section 4) the set

$$\mathfrak{H}_0 := \left\{ \Re \int_{\alpha_1}^x y \mathrm{d}x = 0 \right\}$$

is well-defined independently of the branch-point α_j used in the integration; it also follows that it consists of a **forest of (open) trivalent trees** whose branches are Jordan arcs (possibly of infinite length).

It is then proved in Sec. 4 that this set uniquely defines a collection $\mathcal{B}$ of **finite** Jordan arcs joining the branchpoints α_j which can be used as branch-cuts for the algebraic function $y(x)$; one can then define uniquely (up to overall sign) the function

$$h(x) := \Re \int_{\alpha_1}^x y\mathrm{\underline{x}}$$

in such a way that it is **continuous** on $\mathbb{C}$ and **harmonic** on $\mathbb{C} \setminus \mathcal{B}$ (see figure 1).

The **admissibility** condition requires that the domains of negativity of (one of the two branches of) h be such that $h < 0$ on both sides of each branch-cut (and also that the zeroes of $M(x)$ do not belong to $\mathfrak{H}_0$).

An admissible Boutroux curve **defines** a **potential** via

$$V'(x) = 2(y(x))_{pol}$$

where the subscript means the polynomial part and the branch is such that $h(x) \sim \frac{1}{2}\Re V(x)$ at $x = \infty$.

1.1.1 Connection to the asymptotics of OPs

Such an admissible Boutroux curve is connected to the asymptotics of generalized OPs for the potential $V(x)$ above, in the following way (we omit some details for the sake of clarity; the full discussion is in Sec. 4).

Given an admissible Boutroux curve, we can define some contours Γ_j satisfying the properties

- They extend from ∞ to ∞ and approach it along directions for which $\Re V(x) \to +\infty$;

- The (admissible branch of the) function $h(x)$ is non-negative on these contours.

In a figurative way each contour must extend from ∞ to ∞ along different asymptotic directions and in order to do so is obliged to pass through the branchcuts $\mathcal{B}$ in order not to violate the second condition above.

To each contour we associate an arbitrary[3] complex number $\varkappa_j$ and then we can consider the generalized OPs as in [2]

Figure 1: An example of admissible Boutroux curve, with the set $\mathfrak{H}_0$ in evidence (solid-black contours). The branch-cuts defined by $\mathfrak{H}_0$ are the thick arcs. The domains of negativity of h are the shaded regions. In this case the potential has degree 5 (see Sec. 5 for an explanation of why). The contours in texture represent choices of the contours Γ_j. In this case there are three inequivalent choices. For example the dashed contour comes from the left, follows two arms of the cross of thick arcs and goes to infinity downwards. The dotted lines are critical levelsets of h passing through a saddle–point.

$$\sum \varkappa_j \int_{\Gamma_j} p_n(x) p_m(x) e^{-NV(x)} dx = \delta_{nm}. \tag{1-4}$$

It should be clear that if we fix N, the contours Γ_j can actually be deformed quite arbitrarily in the finite region of the complex plane; it is only in the large N-limit that we should require $h \le 0$ on these contours (this is the *steepest descent condition*).

The main result of this part of the paper is Thm. 4.1 which provides strong asymptotic results for p_{N+r} (where r is an arbitrary fixed integer) in terms of the solution of a *model Riemann Hilbert problem* (eqs. 4-21 in Lemma 4.6); This is the same logical approach used in [14, 15].

Such a RH problem is solved explicitly in terms of Theta functions and spinors on the Boutroux curve: this solution is essentially contained in [15] and is a special case of the more general approach in
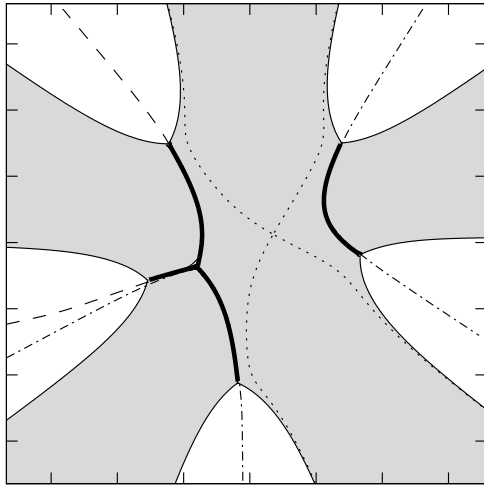
---

[3]These numbers must be sufficiently generic.

[25]. However our approach is yet different and contained in Sec. 3 where we solve rather an **inverse spectral problem** (with a condition of Serre duality) and we find that the solution solves the desired RHP (see the next section in this introduction for the conceptual approach). Note that the formulæ that we obtain are significantly more streamlined compared to those appearing in the seminal work [15] and also "explain" the natural appearance of spinor bundles in terms of Serre-duality.

The main point of this inverse asymptotic study is that Boutroux and admissibility conditions are sufficient (and also necessary as follows from the forward problem) for the amenability of the Deift–Zhou steepest descent method[4]

A consequence and feature of this inverse asymptotic study is that we can construct very easily very complicated and somewhat "surprising" examples: we mention that a consequence of the uniform asymptotics given by the steepest-descent method is that the zeroes of the OPs condense on the Jordan arcs of the branch-cuts $\mathcal{B}$ in the complex plane; these arcs may form an arbitrarily complicated **forest of loop-free trivalent trees**[5].

The implementation of the steepest descent method requires a modification of the usual Airy parametrix near the turning points where three edges of a tree are incident (see Sec. 4.5).

1.2 Boutroux and admissibility

The Boutroux condition is a transcendental constraint on the coefficients of the polynomial defining the (nodal) hyperelliptic curve. At first sight it may not be clear that such Boutroux curves do exist and even less clear how constrained is the topology of $\mathfrak{H}_0$.

This problem is addressed in Section 5; once more the approach is completely "backwards". We start with an relatively brief summary of Strebel theory of quadratic differentials [32] adapted to our simple situation. This allows us to introduce Strebel's coordinates in which the Boutroux condition is just a linear constraint, hence trivially satisfiable. The importance of quadratic differentials for the study of 2×2 Riemann–Hilbert problems has been pointed out in other papers (e.g. [27, 28]) but we are unaware of a systematic application to the steepest descent method as extensive as in the present paper.

These Strebel coordinates parametrize different "cells" of different topology of $\mathfrak{H}_0$ but of the same dimension; in particular we can –given a topological forest graph satisfying some simple requirements– **reconstruct** by cutting-and-pasting a Riemann surface with the desired properties. This process is known in the literature of quadratic differentials as **welding** [32]: in our simple application the welding merges some half-planes and infinite strips into a Riemann–sphere.

The condition of admissibility is also easily imposed in this approach since it simply singles out cells of specific topology. In a way in that section we are solving a more general problem of characterizing

---

[4] The nonlinear steepest descent method was suggested by Deift and Zhou in 1993 for oscillatory Riemann–Hilbert problems appearing in soliton theory. Some elements of the method appeared in earlier works of A. Its (1980-1982) [20] and also in [18]. An essential extension of the method, namely the introduction of the concept of G–function, allowed to include into the scheme in particular the RHP for orthogonal polynomials and it is due to Deift, Venakides and Zhou [14, 15].

[5] Here "forest" is in the sense of graph-theory.

arbitrary polynomials by the topology of the graph of their critical horizontal trajectories (see Sec. 5 for the details).

1.3   Spinor bundles and commuting difference operators

In this first part we deal with formal and heuristic aspect of the asymptotic analysis. The formulæ that we obtain will be shown *a posteriori* to represent the strong asymptotics of the orthogonal polynomials using the nonlinear steepest descent method. Of course the arguments used should be regarded as "temporary", pending a rigorous proof (that is contained in the second part of the paper).

The main idea is the following: it is well known that any sequence of OPs satisfies a three-term recurrence relation which can be written in semi–infinite matrix form with a tridiagonal symmetric matrix which we denote by $\mathbf{X}$. For generalized OPs w.r.t. complex moment functionals [2] the same is true but the matrix $\mathbf{X}$ has complex coefficients.

If the potential is "semiclassical" (i.e. if $V'(x)$ is a rational function and the integrations supported on arbitrary arcs, see [2] for details) the quasi-polynomials $\psi_n := p_n e^{-N\frac{V}{2}}$ satisfy a differential recurrence relation as well, with coefficients that form another matrix $\mathbf{Y}$ which has some closer relationship with $\mathbf{X}$. In the simplest case of polynomial potentials $V(x)$ then

$$-\frac{1}{N}\partial_x[\psi_0, \psi_1, \ldots]^t = \mathbf{Y}[\psi_0, \psi_1, \ldots]^t \tag{1-5}$$

with $\mathbf{Y}$ antisymmetric and finite–band

$$\mathbf{Y} = \frac{1}{2}\left(V'(\mathbf{X})_u - V'(\mathbf{X})_\ell\right) \tag{1-6}$$

the subscripts denoting the "upper" and "lower" part of the matrices. The commutation relations $[\frac{1}{N}\partial_x, x] = \frac{1}{N}$ imply the so–called **string equation**

$$[\mathbf{X}, \mathbf{Y}] = \frac{1}{N}\mathbf{1} . \tag{1-7}$$

An heuristic approach to the asymptotics is that of sending $N \to \infty$ and also "look" at these matrices down the diagonal; if we shift the in dices of the matrices by N and send N to infinity then these matrices become doubly infinite. Moreover the string equation turns into a commutativity equation $[\mathbf{Y}, \mathbf{X}] = 0$, while retaining their band-structure and symmetry.

In this regime these matrices can be regarded as **commuting finite difference operators**, a class of objects studied in several places in the literature. Indeed there exists a fairly general classification of commuting difference operators [26] which is the starting point of this part of the paper.

In the case at hand the main data are:

- a smooth hyperelliptic curve $\mathcal{L}$ of genus g with X being an invariant meromorphic function of degree 2 with (symmetric) poles at two points;

- an arbitrary meromorphic function Y with prescribed pole structure[6]

- a suitably generic effective divisor Γ of degree g.

These data allow us to reconstruct the matrices $\mathbf{X}$ and $\mathbf{Y}$ as the matrices expressing the multiplication operators by X and Y on a suitable basis of meromorphic functions.

We extend this approach by introducing the Serre dual basis (of meromorphic forms) and a pairing (implementing Serra's duality) which takes the simple form of a residue.

At this level the connection to OPs is still rather distant: in order to close the gap we need to consider a "self-dual" picture in which the meromorphic functions of [26] and our Serre–dual counterparts are replaced by sequences of **spinors** which are **self-dual** with respect to Serre–duality. This forces conditions on the divisor Γ and determines it up to the choice of a half–integer characteristic.

The actual application to the asymptotics of OP comes after tensoring this spinor bundle by a suitable **flat line bundle**; such bundle is provided automatically by the Riemann–Hilbert analysis and by the Boutroux condition.

Indeed the Boutroux condition in this setting is $\oint Y \mathrm{d}X \in i\mathbb{R}$ for all closed loops; this means that exponentiation of $\int Y \mathrm{d}X$ provides naturally transition functions for a **unitary** line bundle with characters given by the periods of the Boutroux differential $Y \mathrm{d}X$. This last step provides the backbone of the asymptotic analysis of Sec. 4

It should be added that this algebro–geometric part of the paper is more general and should yield the main ingredients for the asymptotic analysis of the most general semiclassical OPs, i.e. with potentials V whose derivative is rational and with integrals restricted to arbitrary arcs (**hard–edges**). However the technical details of the implementation of Deift–Zhou steepest descent method (Sec. 4) become rather more involved and we prefer to stay in the more standard setting of polynomial potentials for that part of the paper.

2   Notation and main tools

For a given smooth genus-g curve $\mathcal{L}$ with a fixed choice of symplectic homology basis of a and b-cycles, we denote by ω_j the normalized basis of holomorphic differentials

$$\oint_{a_j} \omega_\ell = \delta_{j\ell} , \qquad \oint_{b_j} \omega_\ell = \tau_{j\ell} = \tau_{\ell j} . \tag{2-1}$$

---

[6]This is actually a generalization of the setting of [26].

We will denote by Θ the theta function

$$\Theta(\mathbf{z}) := \sum_{\vec{n} \in \mathbb{Z}^g} e^{i\pi \vec{n} \cdot \tau \vec{n} - 2i\pi \mathbf{z} \cdot \vec{n}} \tag{2-2}$$

The Abel map (with base-point p_0) is defined by

$$\mathfrak{u} : \mathcal{L} \to \mathbb{C}^g \tag{2-3}$$

$$\mathfrak{u}(p) = \left[\int_{p_0}^p \omega_1, \ldots, \int_{p_0}^p \omega_g \right]^t \tag{2-4}$$

and is defined up to the *period lattice* $\mathbb{Z} + \tau \cdot \mathbb{Z}$. For brevity we will omit any symbolic reference to the Abel map when it appears as argument of a Theta function: namely if $p \in \mathcal{L}$ is a point and it appears as argument of a Theta-function, the Abel map (with a certain basepoint) will be understood.

We denote by $\mathcal{K}$ the vector of Riemann constants (also depending on the choice of the basepoint)

$$\mathcal{K}_j = -\sum_{\ell=1}^g \left[\oint_{a_\ell} \omega_\ell(p) \int_{p_0}^p \omega_j(q) - \delta_{j\ell} \frac{\tau_{jj}}{2} \right] \tag{2-5}$$

where in this expression the cycles a_j are realized as loops with basepoint p_0 and the inner integration is done along a path lying in the canonical dissection of the surface along the chosen representatives of the basis in the homology of the curve.

The Riemann constants have the crucial property that for a nonspecial divisor Γ of degree g, $\Gamma = \sum_{j=1}^g \gamma_j$, then the "function"

$$f(p) = \Theta(p - \Gamma - \mathcal{K}) \tag{2-6}$$

has zeroes precisely and only at $p = \gamma_j$, $j = 1 \ldots g$.

We will also have to use Theta functions with (complex) characteristics; for any two complex vectors $\vec{\epsilon}, \vec{\delta}$ the theta function with these (half) characteristics is defined via

$$\Theta \begin{bmatrix} \vec{\epsilon} \\ \vec{\delta} \end{bmatrix} (z) := \exp \left(2i\pi \left(\frac{\epsilon \cdot \tau \cdot \epsilon}{8} + \frac{1}{2}\epsilon \cdot z + \frac{1}{4}\epsilon \cdot \delta \right) \right) \Theta \left(z + \frac{\vec{\delta}}{2} + \tau \frac{\vec{\epsilon}}{2} \right) \tag{2-7}$$

Here the (half) characteristics of a point are defined by

$$2\mathbf{z} = \vec{\delta} + \tau \vec{\epsilon} \tag{2-8}$$

where the factor of 2 is purely conventional so that half integer characteristics have integer (half)-characteristics. In the sequel we will always use these half-characteristics. This modified Theta function has the following periodicity properties, for $\lambda, \mu \in \mathbb{Z}^g$

$$\Theta \begin{bmatrix} \vec{\epsilon} \\ \vec{\delta} \end{bmatrix} (z + \lambda + \tau\mu) = \exp \left[i\pi(\vec{\epsilon} \cdot \lambda - \vec{\delta} \cdot \mu) - i\pi\mu \cdot \tau \cdot \mu - 2i\pi z \cdot \mu \right] \Theta \begin{bmatrix} \vec{\epsilon} \\ \vec{\delta} \end{bmatrix} (z) \tag{2-9}$$

Definition 2.1 *The prime form $E(\zeta, \zeta')$ is the $(-1/2, -1/2)$ bi–differential on $\mathcal{L} \times \mathcal{L}$*

$$E(\zeta, \zeta') = \frac{\Theta_\Delta(\zeta - \zeta')}{h_\Delta(\zeta) h_\Delta(\zeta')} \tag{2-10}$$

$$h_\Delta(\zeta)^2 := \sum_{k=1}^{g} \partial_{\mathfrak{u}_k} \ln \Theta_\Delta \Big|_{\mathfrak{u}=0} \omega_k(\zeta) \,, \tag{2-11}$$

where ω_k are the normalized Abelian holomorphic differentials, $\mathfrak{u}$ is the corresponding Abel map and $\Delta = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$ is a half–integer odd characteristic (the prime form does not depend on which one).

The prime form $E(\zeta, \zeta')$ is antisymmetric in the argument and it is a section of an appropriate line bundle, i.e. it is multiplicatively multivalued on $\mathcal{L} \times \mathcal{L}$; indeed we have the multiplicative multivaluedness

$$E(\zeta + a_j, \zeta') = E(\zeta, \zeta') \tag{2-12}$$

$$E(\zeta + b_j, \zeta') = E(\zeta, \zeta') \exp\left(-\frac{\tau_{jj}}{2} - \int_\zeta^{\zeta'} \omega_j \right) \tag{2-13}$$

In our notation for the (half)-characteristics, the vectors α, β appearing in the definition of the prime form are actually integer valued. We also note for future reference that the half order differential h_Δ is in fact also multivalued according to

$$h_\Delta(p + a_j) = e^{i\pi\alpha_j} h_\Delta(p) \tag{2-14}$$

$$h_\Delta(p + b_j) = e^{-i\pi\beta_j} h_\Delta(p). \,. \tag{2-15}$$

Following the common usage in the literature, for a meromorphic function F (or section of some line–bundle) we will use the notation (F) for its divisor of zeroes/poles [16]. The writing

$$(F) \geq -kp + mq \tag{2-16}$$

means that F has at most a pole of order k at p and a zero of multiplicity at least m at q.

For $\mathcal{L}$ a hyperelliptic algebraic curve of genus g realized as a double cover of the plane, we will denote with $p^\star$ the image of the point $p \in \mathcal{L}$ under the hyperelliptic involution that interchanges the two sheets.

To conclude this section we recall some expressions which are well known [17] and will be used later. Suppose that the hyperelliptic surface is given by the equation $w^2 = \prod_{j=1}^{2g+2}(x - \alpha_j)$ and let us denote with $\infty_\pm$ the two points above $x = \infty$ where $w \sim \pm x^{g+1}$. If we write the normalized first–kind differentials ω_j as

$$\omega_j = \sum_{k=1}^{g} \sigma_{jk} \frac{x^{k-1} dx}{w} \tag{2-17}$$

then it is a direct check to verify that

$$\Theta_\Delta(p - \infty_+) = \frac{C_\Delta}{x}(1 + \mathcal{O}(x^{-1})) \,, \qquad C_\Delta := -\sum_{j=1}^{g} \sigma_{j,g} \partial_j \Theta_\Delta(0) \tag{2-18}$$

9

3   Part I: Jacobi matrices and difference operators

We follow [26] and consider the following setting:

1. An hyperelliptic Riemann surface $\mathcal{L}$ of genus g, with X invariant under the holomorphic involution and with divisor $(X) \geq -\infty_+ - \infty_-$.

2. A meromorphic function $\widehat{Y}$ with divisor

$$(\widehat{Y}) \geq -\sum k_\mu \infty_\mu - \sum \xi_j - k_- \infty_- + \infty_+ \tag{3-1}$$

   where the ξ_j's are chosen in some subset of the Weierstrass points $\mathrm{d}X(\xi_j) = 0$[7]. For later use we also point out that we could equivalently consider a meromorphic function such that $Y(p^\star) = -Y(p)$; indeed we may "antisymmetrize" the above function by adding a rational function of X. This symmetry will be of use later on so we will introduce the special notation

$$Y(p) := \frac{1}{2} \left(\widehat{Y}(p^\star) - \widehat{Y}(p) \right) \tag{3-2}$$

3. A nonspecial divisor Γ of degree g, $\Gamma = \sum_{j=1}^{g} \gamma_j$.

   The genericity assumption on Γ is that all the divisors $\Gamma + r(\infty_+ - \infty_-)$ are also non-special

$$\mathbf{i}(\Gamma + r(\infty_+ - \infty_-)) = 0 \tag{3-3}$$

where $\mathbf{i}(\mathfrak{D})$ denotes the dimension of the space of differentials with divisor exceeding $\mathfrak{D}$ [16].

   The genericity condition (3-3) implies also

$$\mathbf{i}(\Gamma + (r-1)\infty_+ - (r+1)\infty_-) = 1 \ . \tag{3-4}$$

Indeed, if we had $\mathbf{i}(\Gamma + (r-1)\infty_+ - (r+1)\infty_-) \geq 2$ then, by Riemann–Roch's theorem, there would exist a meromorphic function with $(f) \geq -\Gamma - r(\infty_+ - \infty_-) + \infty_+ + \infty_-$. But then we would have that both f and Xf (clearly linearly independent) would have divisor $\geq -\Gamma - r(\infty_+ - \infty_-)$ so that $\mathbf{r}(-\Gamma - r(\infty_+ - \infty_-)) \geq 2$ and –again by Riemann–Roch's thm.– we would have $\mathbf{i}(\Gamma + r(\infty_+ - \infty_-)) = \mathbf{r}(-\Gamma - r(\infty_+ - \infty_-)) - 1 \geq 1$, a contradiction with the assumption (3-3).

   By Riemann–Roch's theorem and condition (3-3) it follows that for each $r \in \mathbb{Z}$ there is a unique (up to multiplicative constant) meromorphic function with divisor

$$(P_r) \geq -\Gamma - r\infty_+ + r\infty_- \tag{3-5}$$

---

[7] Note that there is no loss of generality in assuming that $X(\infty_\mu) \neq X(\infty_\nu)$ (i.e. the poles of $\widehat{Y}$ have distinct X projection) because we may add to $\widehat{Y}$ a suitable rational function of X to reduce ourselves to this situation. This is why we did not put any pole at ∞_+; moreover we can always add a constant to Y so that it has a zero at ∞_+.

The second condition (3-4) states that for each $r \in \mathbb{Z}$ there is a unique (up to multiplicative constant) meromorphic differential F_n with divisor satisfying

$$(F_r) \geq \Gamma + (r-1)\infty_+ - (r+1)\infty_- \tag{3-6}$$

From these properties it follows that the product $F_r P_s$ is a differential with at most two simple poles when $r = s$ and otherwise it has only one pole. Therefore

$$\operatorname*{res}_{\infty_-} P_r F_s \propto \delta_{rs} \tag{3-7}$$

This condition is just a manifestation of Serre duality.

Since –for $r = s$– the product $P_r F_s$ is a third kind differential with simple poles, its residues cannot vanish and hence we can normalize the two dual sequence so that their their Serre pairing (3-7) is actually δ_{rs}.

Expressions for the functions P_r and differentials F_s in terms of Theta functions can be obtained following standard references, for example [26]. Since P_r, F_s play only a temporary role in our discussion, we delay explicit formulæ until we arrive at the final objects of interest.

The two sequences can be normalized in such a way that

$$\operatorname*{res}_{\infty_-} F_s P_r = \delta_{rs} \ . \tag{3-8}$$

Given the pole structure of $X(p)$ we see that $X P_r$ is a linear combination of P_{r+1}, P_r, P_{r-1}: indeed

$$(X P_r) \geq -\Gamma - (r+1)\infty_+ + (r-1)\infty_- \tag{3-9}$$

and the dimension (again generically) of the space of meromorphic functions with divisor exceeding the above one is $\mathbf{r} = 3$ and it is spanned by the above three meromorphic functions [26].

We can express the coefficients of this three-term recurrence relation

$$X P_r = \gamma_{r+1} P_{r+1} + \beta_r P_r + \widetilde{\gamma}_r P_{r-1} \tag{3-10}$$

in two distinct ways: first and foremost, using the duality (3-7)

$$\gamma_{r+1} = \operatorname*{res}_{\infty_-} F_{r+1} X P_r \ , \quad \beta_r = \operatorname*{res}_{\infty_-} F_r X P_r \ , \qquad \widetilde{\gamma}_r = \operatorname*{res}_{\infty_-} F_{r-1} X P_r \ . \tag{3-11}$$

This provides an explicit expression in terms of Theta functions if we express X as well in terms of them

$$X(p) = X_0 \frac{\Theta_\Delta(p - z_0) \Theta_\Delta(p - z_1)}{\Theta_\Delta(p - \infty_+) \Theta_\Delta(p - \infty_-)} \tag{3-12}$$

where z_0, z_1 are the two zeroes of X and X_0 is a constant. A second independent way is obtained as follows

$$X P_r(p) = \frac{c_r}{\det(P_{r+j}(z_k))_{j,k=1,2}} \det \begin{bmatrix} P_{r+1}(p) & P_r(p) & P_{r-1}(p) \\ P_{r+1}(z_0) & P_r(z_0) & P_{r-1}(z_0) \\ P_{r+1}(z_1) & P_r(z_1) & P_{r-1}(z_1) \end{bmatrix} \tag{3-13}$$

11

where z_0, z_1 are the two zeroes of X (interchanged by the hyperelliptic involution). The constant c_n is expressed in terms of Theta functions by matching the behaviors of both sides at one of the two infinities $\infty_\pm$ and using (3-12). The dual sequence F_n satisfies –by duality (3-7)– the transposed recurrence relation

$$X F_r = \widetilde{\gamma}_{r+1} F_{r+1} + \beta_r F_r + \gamma_r F_{r-1} \ . \tag{3-14}$$

3.1 Kernel and "Christoffel–Darboux" pairing

We want to define the sequence of "projectors" K_r $r \in \mathbb{Z}$ formally as the expressions

$$K_r(p, \xi) = \sum_{j=-\infty}^{r-1} P_j(p) F_j(\xi) \tag{3-15}$$

so that $\operatorname*{res}_{\infty_-} K_r(p, \xi) P_s(\xi) = P_s(p)$ if $s \leq r - 1$ and zero otherwise. Clearly these expressions make little sense as they stand since they involve infinite series whose convergence should then be proved: what we want to have is a kernel $K_r(p, \xi)$ which is a differential in ξ and a function in p that satisfies the following properties:

1. As a differential in ξ it has

    (a) a zero of order $r - 1$ at ∞_+ and a pole of order r at ∞_-

    (b) zeroes at Γ

    (c) simple pole at $\xi = p$ (the diagonal) with residue $+1$.

2. As a function of p it has

    (a) A pole of order $r - 1$ at ∞_+ and a zero of order r at ∞_-

    (b) Poles at Γ

    (c) A simple pole at $p = \xi$.

These properties define it uniquely as (see [17] pag 27. for similar kernels) and an explicit formula can also be written in terms of Theta–functions. Once more we will only write formulas for the final objects.

The kernel $K_n(p, \xi)$ enjoys the **Christoffel–Darboux** property (an exercise using the recursion relations or the divisor properties)

$$(X(p) - X(\xi)) K_r(p, \xi) = [F_{r-1}(\xi), F_r(\xi)] \begin{bmatrix} 0 & \gamma_r \\ -\widetilde{\gamma}_r & 0 \end{bmatrix} \begin{bmatrix} P_{r-1}(p) \\ P_r(p) \end{bmatrix} = \tag{3-16}$$

$$= \gamma_r P_r(p) F_{r-1}(\xi) - \widetilde{\gamma}_r P_{r-1}(p) F_r(\xi) \tag{3-17}$$

Therefore we have the

12

Theorem 3.1 ("Christoffel–Darboux" theorem)

$$K_r(p,\xi) = \frac{\gamma_r P_r(p) F_{r-1}(\xi) - \widetilde{\gamma}_r P_{r-1}(p) F_r(\xi)}{X(p) - X(\xi)} \qquad (3\text{-}18)$$

Taking the residue

$$1 = \operatorname*{res}_{\xi=p} K_r(p,\xi) = \frac{\gamma_r P_r(p) F_{r-1}(p) - \widetilde{\gamma}_r P_{r-1}(p) F_r(p)}{dX(p)} \qquad (3\text{-}19)$$

we have a representation of the differential dX as

$$dX = \gamma_r P_r F_{r-1} - \widetilde{\gamma}_r P_{r-1} F_r \qquad (3\text{-}20)$$

3.2 Flat line bundles

In the asymptotic analysis of orthogonal polynomials using the Riemann–Hilbert method that follows, we will need to tensor the line bundle whose sections correspond to the meromorphic functions P_n and the Serre-dual line bundle (the F_r's) by a suitable flat line bundle.

This line bundle can be described in much more general terms as associated to an arbitrary second–kind differential[8].

More precisely let η be a meromorphic differential such that all residues are zero (second kind differential) or integers. Let $\int \eta$ be its Abelian integral; it is defined on the universal covering of the curve $\mathcal{L}$ less the poles of η.

Definition 3.1 *Near a pole c of η we define the exponential singular part $E_{\eta,c}$ of $\int \eta$ as*

$$E_{\eta,c}(p) = \exp\left(\oint \eta(\xi)\ln(z(\xi) - z(p))\right) \qquad (3\text{-}21)$$

where $z(\xi)$ is a local coordinate $z(c) = 0$ and the integral is along loop surrounding $\xi = c$ such that p is outside the loop. The function is independent of the choice of local parameter up to multiplication of a holomorphic function with nonzero value at c.

Note that, since the residue of η at c is at most an integer, this "quantization" makes irrelevant which branch of the logarithm is used.

The twisted line bundle $\mathfrak{L}_\eta$. Associated to η there is a line–bundle $\mathfrak{L}_\eta$ with transition functions $E_{\eta,c}$ at the poles of η

Tensoring by $\mathfrak{L}_\eta$ the line bundle described by Γ means that the sections of the line bundle will be "functions" φ with the following properties

1. Poles at Γ

2. Near a pole c of η: $\varphi_n(p) E_{\eta,c}(p) = \mathcal{O}(1)$.

The formulæ for the dual sequences of wave functions/forms require minimal modifications and the expressions in terms of Theta function is an exercise that we delay to the next section.

---

[8]We could in fact use arbitrary meromorphic differentials but this would introduce some slight additional complication in the formulæ, and we leave this to another publication.

3.3 Spinors and the symmetric picture

In the applications stemming from the (formal) asymptotics of orthogonal polynomials the sequence of OP should be orthogonal to itself: i.e. we should put in some way the two dual sequences on a symmetric footing; moreover the hyperelliptic involution should yield the dual sequence directly. This "symmetry" requirement fixes the divisor Γ. A similar construction in the theory of the algebro-geometric solutions of soliton equations goes back to the works [9, 10]. Indeed, let us look for a spinor $\mathfrak{s}$ with the properties

1. it has a simple pole at ∞_+

2. it has simple zeroes at Γ

3. no other poles or zeroes.

Such a spinor exists *provided* that Γ satisfies some condition to be specified below (note that the degree of the above divisor is the correct one, $g-1$): the square of a spinor must be a differential (with divisor of degree $2g-2$) that has divisor

$$(\mathfrak{s}^2) = 2\Gamma - 2\infty_+ \tag{3-22}$$

Since the image in the Jacobian of the canonical class is $-2\mathcal{K}$ we must have

$$2\Gamma - 2\infty_+ = -2\mathcal{K} , \qquad \Rightarrow \Gamma + \mathcal{K} = \infty_+ + \nu \tag{3-23}$$

$$2\nu = 0 \tag{3-24}$$

[these equations are written understanding the Abel map and modulo the lattice of periods]. This determines the divisor Γ –up to the choice of a half-period ν– from Jacobi's inversion theorem. We then define

$$\pi_r :\propto P_r \mathfrak{s} , \qquad \pi_r^\star :\propto \frac{F_r}{\mathfrak{s}} \tag{3-25}$$

$$\mathfrak{K}_r(p,\xi) = \mathfrak{s}(p) K_r(p,\xi) \frac{1}{\mathfrak{s}(\xi)} \tag{3-26}$$

Exploiting the holomorphic equivalence between flat bundles and unitary flat bundles we can express $\mathfrak{s}$ as section of the tensor product with the unitary line bundle characterized by

$$\chi(a_j) = \mathrm{e}^{i\pi\mathcal{A}_j} , \qquad \chi(b_j) = \mathrm{e}^{i\pi\mathcal{B}_j} , \tag{3-27}$$

where $\mathcal{A}, \mathcal{B}$ are the half-characteristics of ν

$$2\nu = \mathcal{A} + \tau\mathcal{B}. \tag{3-28}$$

Since ν is a half period, then $\mathcal{A}, \mathcal{B} \in \mathbb{Z}^g$.

3.3.1 Θ–functional expressions

In this paragraph we provide the explicit expressions of the spinors π_n, since they will be the final object of interest in our application.

Both $\pi_r, \pi_r^\star$ are spinors (half integer differentials) belonging to a square-root of the canonical bundle of the curve tensored by the line–bundle $\mathfrak{L}_\eta$ with transition functions $E_{\eta,c.}$ and with divisors

$$(\pi_r) \geq -(r+1)\infty_+ + r\infty_- \ , \qquad (\pi_r^\star) \geq r\infty_+ - (r+1)\infty_- \tag{3-29}$$

$$\pi_r(p+\gamma) = \chi(\gamma)\pi_r(p) \ , \quad \pi_r^\star(p+\gamma) = \chi^{-1}(\gamma)\pi_r^\star(p) \ , \ \forall \gamma \in \pi_1(\mathcal{L}). \tag{3-30}$$

with $\chi(\gamma)$ defined (in a basis) by eq. 3-27.

Their explicit expressions are obtained using standard arguments and are given below

$$\pi_r := \frac{1}{\sqrt{h_r}} \frac{\Theta_\Delta^r(p-\infty_-)}{\Theta_\Delta^{r+1}(p-\infty_+)} \Theta \begin{bmatrix} \mathcal{A}+\vec{\epsilon} \\ \mathcal{B}+\vec{\delta} \end{bmatrix} (p + r\infty_- - (r+1)\infty_+) h_\Delta(p) e^{-\int_{\alpha_1}^p \eta}$$

$$\pi_r^\star := \frac{1}{\sqrt{h_r}} \frac{\Theta_\Delta^r(p-\infty_+)}{\Theta_\Delta^{r+1}(p-\infty_-)} \Theta \begin{bmatrix} -\mathcal{A}-\vec{\epsilon} \\ -\mathcal{B}-\vec{\delta} \end{bmatrix} (p - (r+1)\infty_- + r\infty_+) h_\Delta(p) e^{\int_{\alpha_1}^p \eta} \tag{3-31}$$

$$\epsilon_i := \frac{1}{2} \oint_{a_j} \eta \ , \qquad \delta_i := -\frac{1}{2} \oint_{b_j} \eta \ .$$

The constants h_n are just the suitable normalizations so that

$$\operatorname*{res}_{\infty_+} \pi_r \pi_s^\star = \delta_{rs} \tag{3-32}$$

They are

$$h_r = \frac{\Theta_\Delta(\infty_+ - \infty_-)}{\Theta \begin{bmatrix} \mathcal{A}+\vec{\epsilon} \\ \mathcal{B}+\vec{\delta} \end{bmatrix} (r(\infty_- - \infty_+)) \Theta \begin{bmatrix} -\mathcal{A}-\vec{\epsilon} \\ -\mathcal{B}-\vec{\delta} \end{bmatrix} ((r+1)(\infty_+ - \infty_-))} \tag{3-33}$$

and the spinorial kernel of the Christoffel–Darboux projector now reads

$$\mathfrak{K}_r(p,\xi) := \left[\frac{\Theta_\Delta(\xi - \infty_+)\Theta_\Delta(p - \infty_-)}{\Theta_\Delta(\xi - \infty_-)\Theta_\Delta(p - \infty_+)} \right]^r \frac{\Theta \begin{bmatrix} \mathcal{A}+\vec{\epsilon} \\ \mathcal{B}+\vec{\delta} \end{bmatrix} (r(\infty_- - \infty_+) + p - \xi)}{E(\xi,p)\Theta \begin{bmatrix} \mathcal{A}+\vec{\epsilon} \\ \mathcal{B}+\vec{\delta} \end{bmatrix} (r(\infty_- - \infty_+))} e^{-\int_\xi^p \eta} \tag{3-34}$$

In the following applications η will be an antisymmetric differential (under the hyperelliptic involution); in this case we have the symmetry

$$\pi_r(p^\star) = -\pi_r^\star(p) \tag{3-35}$$

which explains the notation; indeed this follows from the fact that the Abel map (based at a Weierstrass point) of $p^\star$ is the opposite of that of p and from the symmetries of the Theta functions with characteristics

$$\Theta \begin{bmatrix} A \\ B \end{bmatrix} (z) = \Theta \begin{bmatrix} -A \\ -B \end{bmatrix} (-z) \tag{3-36}$$

15

together with the antisymmetry of $\eta(p) = -\eta(p^\star)$. With these normalizations the recurrence relations become automatically symmetric because

$$\gamma_r = \operatorname*{res}_{\infty_+} \pi_r X \pi_{r+1}^\star = -\operatorname*{res}_{\infty_+ \star} \pi_r^\star X \pi_{r+1} = -\operatorname*{res}_{\infty_-} \pi_r^\star X \pi_{r+1} = \operatorname*{res}_{\infty_+} \pi_r^\star X \pi_{r+1} = \widetilde{\gamma}_r \tag{3-37}$$

and hence

$$X\pi_r = \gamma_r \pi_{r+1} + \beta_r \pi_r + \gamma_{r-1}\pi_{r-1} \tag{3-38}$$

Moreover it follows from the previous Christoffel–Darboux identities (Thm. 3.1) that

$$\mathfrak{K}_r(p,\xi) = \gamma_r \frac{\pi_{r-1}(p)\pi_r^\star(\xi) - \pi_r(p)\pi_{r-1}^\star(\xi)}{X(p) - X(\xi)} \tag{3-39}$$

$$dX(p) = \gamma_r \left(\pi_{r-1}(p)\pi_r^\star(p) - \pi_r(p)\pi_{r-1}^\star(p) \right) \tag{3-40}$$

3.4 Ladder matrices and Lax matrix

We preliminary point out that any spinorial Baker–Akhiezer function π_k can be written as a linear combination in terms of any other two consecutive wave-functions (forms) with polynomial coefficients in X: indeed the recurrence relations (3-38) can be rewritten in matrix form as

$$\mathbf{\Pi}_{r+1} := \begin{bmatrix} \pi_{r+1} \\ \pi_r \end{bmatrix} = \begin{bmatrix} \frac{X - \beta_r}{\gamma_r} & -\frac{\gamma_{r-1}}{\gamma_r} \\ 1 & 0 \end{bmatrix} \mathbf{\Pi}_r =: \underset{r}{\mathbf{a}}(X)\mathbf{\Pi}_r \tag{3-41}$$

The ladder matrix $\mathbf{a}_r(X)$ is invertible and the inverse is linear as well in X. Therefore

$$\mathbf{\Pi}_{r+k+1} = \underset{r+k}{\mathbf{a}} \cdots \underset{r}{\mathbf{a}} \mathbf{\Pi}_r =: \overset{r+k}{\underset{r}{\mathbf{a}}} \mathbf{\Pi}_r , \qquad k \geq 0 \tag{3-42}$$

$$\mathbf{\Pi}_{r-k-1} = \underset{r-k-1}{\mathbf{a}}{}^{-1} \cdots \underset{r-1}{\mathbf{a}}{}^{-1} \mathbf{\Pi}_r , \qquad k \geq 0 . \tag{3-43}$$

Even more directly, denoting by $\mathbb{A}_r = \begin{pmatrix} 0 & \gamma_r \\ -\gamma_r & 0 \end{pmatrix}$, the kernels K_r are written as

$$K_r(p,\xi) = \frac{\mathbf{\Pi}_r^\star(\xi)^t \mathbb{A}_r \mathbf{\Pi}_r(p)}{X(p) - X(\xi)} \tag{3-44}$$

Then we have the identity

$$\mathbf{\Pi}_j(p) = \operatorname*{res}_{\xi=p} \mathbf{\Pi}_j(\xi)\mathfrak{K}_r(p,\xi) = -\sum_{a=\pm} \operatorname*{res}_{\infty_a} \mathbf{\Pi}_j(\xi)\mathfrak{K}_r(p,\xi) = \sum_{a=\pm} \operatorname*{res}_{\infty_a} \frac{\mathbf{\Pi}_j(\xi)\mathbf{\Pi}_r^{\star,t}(\xi)\mathbb{A}_r}{X(p) - X(\xi)}\mathbf{\Pi}_r(p) . \tag{3-45}$$

which follows from the fact that –in any local parameter– $\mathfrak{K}_r(p,\xi) = \frac{\sqrt{dz}\sqrt{dz'}}{z-z'}(1 + \mathcal{O}(z - z')$, with $z = z(p), z' = z(\xi)$. It is easily seen that the matrix

$$\underset{r}{\overset{j}{\mathbf{a}}}(x) := \sum_{a=\pm} \operatorname*{res}_{\infty_a} \frac{\mathbf{\Pi}_j(\xi)\mathbf{\Pi}_r^{\star,t}(\xi)\mathbb{A}_r}{x - X(\xi)} \tag{3-46}$$

is a **polynomial** in x of degree $|r - j|$ and computes directly the product of the ladder matrices above.

16

3.4.1 Lax matrix

Although $Y\mathbf{\Pi}_r$ in general does not have the same pole structure as the sequence of the π_r's, nevertheless we can express it in terms of the same vector $\mathbf{\Pi}_r$.

Let us define the rational function V'_α of x, $\alpha = \mu, -$ such that

$$\widehat{Y}(p) - V'_\mu(X(p)) = \mathcal{O}(1) \quad \text{near } \infty_\mu \tag{3-47}$$

$$\widehat{Y}(p) - V'_-(X(p)) = \mathcal{O}(X^{-1}) \text{ near } \infty_- . \tag{3-48}$$

The function V'_- is a polynomial of degree k_- whereas V'_μ are polynomials in $\frac{1}{x - X(\infty_\mu)}$ of degree k_μ and without constant coefficient. Define then the total potential

$$V'(x) := \sum_{\alpha = \mu, -} V'_\alpha(x) . \tag{3-49}$$

This rational function of x has the property that $\widehat{Y}(p) - V'(X(p))$ is analytic near the points ∞_μ and ∞_-. Then we find (using that $\widehat{Y}$ has a zero at ∞_+)[9]

$$\widehat{Y}\mathbf{\Pi}_r(p) = \operatorname*{res}_{\xi=p} \widehat{Y}(\xi)\mathbf{\Pi}_r(\xi)\mathfrak{K}_r(p,\xi) = -\operatorname*{res}_{\mathfrak{U},\infty_-} \widehat{Y}\mathbf{\Pi}_r(\xi)\mathfrak{K}_r(p,\xi) = \tag{3-50}$$

$$= -\sum_{\alpha=\mu,-} \operatorname*{res}_{\xi=\infty_\alpha} \widehat{Y}(\xi)\mathbf{\Pi}_r(\xi)\mathfrak{K}_r(p,\xi) - \sum_j \operatorname*{res}_{\xi=\xi_j} \widehat{Y}(\xi)\mathbf{\Pi}_r(\xi)\mathfrak{K}_r(p,\xi) =$$

$$= -\sum_{\alpha=\mu,-} \operatorname*{res}_{\xi=\infty_\alpha} V'(X(\xi))\mathbf{\Pi}_r(\xi)\mathfrak{K}_r(p,\xi) - \sum_j \operatorname*{res}_{\xi=\xi_j} \widehat{Y}(\xi)\mathbf{\Pi}_r(\xi)\mathfrak{K}_r(p,\xi) =$$

$$= -\sum_{\alpha=\mu,-} \operatorname*{res}_{\xi=\infty_\alpha} V'(X(\xi))\mathbf{\Pi}_r(\xi)\frac{\mathbf{\Pi}_r^{\star,t}(\xi)\mathbb{A}\mathbf{\Pi}_r(p)}{X(p)-X(\xi)} - \sum_j \operatorname*{res}_{\xi=\xi_j} \widehat{Y}(\xi)\mathbf{\Pi}_r(\xi)\mathfrak{K}_r(p,\xi) =$$

$$= \begin{pmatrix} 0 & 0 \\ 0 & V'(X(p)) \end{pmatrix} \mathbf{\Pi}_r(p) - \sum_{\alpha=\mu,-} \operatorname*{res}_{\xi=\infty_\alpha} \frac{V'(X(\xi)) - V'(X(p))}{X(p)-X(\xi)}\mathbf{\Pi}_r^{\star,t}(\xi)\mathbb{A}\mathbf{\Pi}_r(p)+$$

$$- \sum_j \operatorname*{res}_{\xi=\xi_j} \widehat{Y}(\xi)\mathbf{\Pi}_r(\xi)\mathfrak{K}_r(p,\xi) = \tag{3-51}$$

$$\begin{pmatrix} 0 & 0 \\ 0 & V'(X(p)) \end{pmatrix} \mathbf{\Pi}_r(p) - \sum_{\alpha=\mu,-} \operatorname*{res}_{\xi=\infty_\alpha} \frac{V'(X(\xi)) - V'(X(p))}{X(p)-X(\xi)}\mathbf{\Pi}_r(\xi)\mathbf{\Pi}_r^{\star,t}(\xi)\mathbb{A}\mathbf{\Pi}_r(p)+$$

$$- \sum_j \operatorname*{res}_{\xi=\xi_j} \widehat{Y}(\xi)\mathbf{\Pi}_r(\xi)\frac{\mathbf{\Pi}_r^{\star,t}(\xi)\mathbb{A}\mathbf{\Pi}_r(p)}{X(p)-X(\xi)} \tag{3-52}$$

Summarizing we have –in matrix form–

$$\widehat{Y}\mathbf{\Pi}_r = \widehat{D}_r^{(\eta)}(X)\mathbf{\Pi}_r$$
$$\widehat{D}_r^{(\eta)}(x) = \begin{bmatrix} 0 & 0 \\ 0 & V'(x) \end{bmatrix} - \sum_{\alpha=\mu,+} \operatorname*{res}_{\infty_\alpha} \frac{V'(X) - V'(x)}{X - x}\mathbf{\Pi}_r(\xi)\mathbf{\Pi}_r^{\star,t}(\xi)\mathbb{A}_r+$$
$$- \sum_j \operatorname*{res}_{\xi=\xi_j} \frac{\widehat{Y}(\xi)\mathbf{\Pi}_r(\xi)\mathbf{\Pi}_r^{\star,t}(\xi)}{X - x}\mathbb{A}_r \tag{3-53}$$

---

[9]Recall that ξ_j represent those Weierstrass points (branchpoints for X) which coincide with a simple pole of Y (or $\widehat{Y}$).

Quite clearly, the matrix $\widehat{D}_r^{(\eta)}(x)$ is a rational matrix with the same pole structure as the derivative of the "potential" $V'(x)$: at the other poles (ξ_r) of $\widehat{Y}$, it has simple poles with a nilpotent residue.

Similarly we could repeat the computation for the antisymmetric function Y; however it is immediate to see that[10]

$$Y = \frac{1}{2}V'(X) - \widehat{Y} \tag{3-54}$$

so that the matrix $\widehat{D}_r^{(\eta)}(X)$ representing the multiplication by $\widehat{Y}$ differs from the matrix $D_r^{(\eta)}(X)$ only by the multiple of the identity $\frac{1}{2}V'(X)\mathbf{1}$

$$D_r^{(\eta)}(x) = \frac{V'(x)}{2}\mathbf{1} - \widehat{D}_r^{(\eta)}(x) \ , \qquad [\mathbf{\Pi}_r, \mathbf{\Pi}_r^\star]Y\sigma_3 = D_r^{(\eta)}(X)[\mathbf{\Pi}_r, \mathbf{\Pi}_r^\star] \ . \tag{3-55}$$

where we have used that $\mathbf{\Pi}_r^\star$ solves the eigenvector equation with eigenvalue $-Y$ since it corresponds to evaluation on the other sheet of the X–projection and Y is antisymmetric.

The WKB Ansatz. We now spend some words on the heuristics of the whole construction and on its logical and historical underpinnings. As pointed out in the introduction (at least for polynomial potentials) the OPs' satisfy a three term-recurrence relation of the same form as (3-38) and also a *differential* recurrence relation (1-5). The compatibility relation between the two is contained in eq. (1-7), often referred to as the "string equation" for the recurrence coefficients [31]. It can be rewritten in matrix form as the compatibility relation between the following differential/difference equations

$$-\frac{1}{N}\partial_x\Psi_n = D_n(x)\Psi_n \ , \Psi_{n+1} = \begin{bmatrix} \frac{x-\beta_n}{\gamma_n} & -\frac{\gamma_n}{\gamma_{n-1}} \\ 1 & 0 \end{bmatrix} \Psi_n \ , \quad \Psi_n(x) := \begin{bmatrix} p_n(x) & \phi_n \\ p_{n-1}(x) & \phi_{n-1} \end{bmatrix} e^{-\frac{N}{2}V(x)\sigma_3} \tag{3-56}$$

where $n = N + r$ (with $r \in \mathbb{Z}$ bounded as $N \to \infty$), ϕ_n are the Cauchy transforms of $e^{-NV}p_n$'s and γ_n, β_n denote the recurrence coefficients for the orthogonal polynomials (see for example [3] for explicit formulæ for the matrix D_n in a very general setting of semiclassical orthogonal polynomials). The heuristic approach is then to *assume* the validity of Wentzel-Kramers-Brillouin (WKB) ansatz for $\Psi_n = \Psi_{N+r}$ in the form

$$\Psi_{N+r} \sim \Xi_r e^{-N\sigma_3 \int^x Y(\xi)d\xi} \tag{3-57}$$

where Ξ_r (see eq. (3-68)) should be expanded (as well as Y) in formal series of N^{-1}. Plugging the Ansatz (3-57) into the differential equation (3-56) yields, to leading order

$$\Xi_r Y\sigma_3 = D_r^{(\eta)}(x)\Xi_r \tag{3-58}$$

which is nothing more than an *eigenvalue/eigenvector* equation precisely of the type appearing in (3-55) (with $\eta = NYdX$). In short the WKB approach amounts to substituting the ODE in (3-56) by its *symbol*, thus turning it into a spectral problem.

---

[10]Indeed $\widehat{Y}(p^\star) + \widehat{Y}(p)$ is invariant under the hyperelliptic involution and hence it is a rational function of X, namely $V'(X)$ in our notation.

This line of approach was developed in the works [30, 31], based in turn on the isomonodromic techniques in the theory of Painlevé equations [21]. In the context of OPs it was applied to the "one–cut" case in [6] and to multi–cut settings in [8].

The self–consistency of the WKB method calls for the periods of the exponent $\int^x \widehat{Y}(x)\mathrm{d}x$ to have no real part, for otherwise the asymptotic expansion in inverse powers of N for Ξ_N would not be valid under analytical continuation; such condition (**Boutroux condition**) will be the main requirement in the second part of the paper.

In particular, as anticipated in the introduction, the *asymptotic spectral curve* (1-2) with the Boutroux condition (1-3) should be thought of as the "limiting spectral curve" of the matrices $D_n(x)$ appearing in the differential equation of eqs. (3-56).

The matrix Ξ_r consists of genuine Baker–Akhiezer functions. The relation with spinors and the Serre duality (3-32) should be understood as a large–N relic of the orthogonality relations: indeed, at a very formal/suggestive level, the orthogonality reads $\int p_n p_m \mathrm{e}^{-NV(x)}\mathrm{d}x = \delta_{nm}$ which suggests that a symmetric way of considering the large-N limit should be by formally splitting $\mathrm{d}x = \sqrt{\mathrm{d}x}\sqrt{\mathrm{d}x}$ so that $p_{N+r}\mathrm{e}^{-\frac{N}{2}V(x)}\sqrt{\mathrm{d}x} \sim \pi_r$. The N–dependence will be in the asymptotic line bundle determined by the differential $\eta = NY\mathrm{d}x$ on the hyperelliptic spectral curve, while the finite perturbation r will give the lattice behavior (of Toda type).

Of course, the reader should regard these considerations as a mere back-of-the-mind motivation towards an educated guess, the full justification of which comes only *a posteriori* via a Deift–Zhou nonlinear steepest-descent analysis (contained in the second part of the paper) adapted to particular hyperelliptic curves satisfying the Boutroux condition. The existence/construction of these is the content of the third part of the paper.

Once the strict connection between the π_r's and the uniform asymptotics over compact sets of the orthogonal polynomials are made rigorous, we could conclude that indeed $\det(y\mathbf{1} - D_n(x))$ admits a *bona-fide* limit and this limit is a Boutroux curve.

Indeed it will be shown in Section 4 that the solution Ψ_{N+r} of the Riemann–Hilbert problem for the orthogonal polynomials (eq. (3-56) is indeed approximated by the matrix $\Xi_r(x)\mathrm{e}^{N\int Y\mathrm{d}x\sigma_3}$ (with Ξ_r given in (3-68)) *uniformly over compact sets not intersecting the branchcuts*.

Thus we have that the matrix of the differential equation $D_{N+r}(x) = \frac{1}{N}\Psi'_{N+r}(x)\Psi^{-1}_{N+r}(x)$ (which is a polynomial in x of constant degree equal to $\deg V'(x)$ for polynomial potentials) *also tends uniformly* to $D^\eta_r(x) = \Xi(x)^{-1}Y(x)\sigma_3\Xi(x)$, which is our eigenvalue-eigenvector equation (3-55) for the line-bundle $\eta = NY\mathrm{d}X$. Therefore *a fortiori* also the spectral curve must tend to the asymptotic Boutroux hyperelliptic curve. Since the convergence is uniform over compact sets and the objects D_{N+r} and its characteristic polynomials are all polynomials in x of bounded degree, the convergence is established in any bounded set of the plane.

3.5 The spinor h_Δ

Following ([17], page. 13) we can explicitly write the half order differential h_Δ (up to a multiplicative constant which is irrelevent in our application since it would then be reabsorbed in the normalizations). Let us denote by $\alpha_j = X(\xi_j)$, $j = 1, \ldots 2g+2$ the critical values of X (at the Weierstrass points) and

$$W^2 := \prod_{j=1}^{2g+2} (X - \alpha_j) \tag{3-59}$$

There are 4^g half periods in the Jacobian: they are all in one-to-one correspondence with the 4^g partitions $\{i_1, \ldots, i_{g+1-2m}\} \cup \{j_1, \ldots, j_{g+1+2m}\}$ of the set $\{1, 2, \ldots, 2g+2\}$ for $m \geq 0$ as follows:

1. All even *non-singular* half-periods are given by the choice $m = 0$ and are the image in the Jacobian of

$$\sum_{k=1}^{g+1} \alpha_{i_k} - \infty_+ - \infty_- - \mathcal{K} \tag{3-60}$$

2. all odd *non-singular* half-periods are given by the choice $m = 1$ and are the image in the Jacobian of

$$\sum_{j=1}^{g-1} \alpha_{i_j} - \mathcal{K} \tag{3-61}$$

3. All other half-periods are singular and they are even or odd according to the parity of m.

Any choice of odd, nonsingular half-integer characteristics corresponds to a choice of $g-1$ points amongst the Weierstrass points, $\xi_{i_1}, \ldots, \xi_{i_{g-1}}$, therefore

$$h_\Delta^2 \propto \prod_{k=1}^{g-1} (X - \alpha_{i_k}) \frac{dX}{W} \tag{3-62}$$

In other words, with proper understanding of the analytic continuation on the (double cover of the) hyperelliptic curve, the spinor h_Δ can be expressed as (up to overall constant)

$$h_\Delta = \frac{\prod_{k=1}^{g-1} (X - \alpha_{i_k})^{\frac{1}{4}}}{\prod_{k=1}^{g+3} (X - \alpha_{j_k})^{\frac{1}{4}}} \sqrt{dX} \tag{3-63}$$

3.6 Riemann Hilbert problem

The spinors (3-31) exhibit a multiplicative behavior which depends on the half period $\nu = \frac{1}{2}\vec{\mathcal{A}} + \frac{1}{2}\tau \cdot \vec{\mathcal{B}}$ ($\vec{\mathcal{A}}, \vec{\mathcal{B}} \in \mathbb{Z}^g$)

$$\pi_r(p + a_j) = e^{i\pi \mathcal{A}_j} \pi_r(p) , \qquad \pi_r(p + b_j) = e^{i\pi \mathcal{B}_j} \pi_r(p) \tag{3-64}$$

$$\pi_r^\star(p + a_j) = e^{-i\pi \mathcal{A}_j} \pi_r^\star(p) , \qquad \pi_r^\star(p + b_j) = e^{-i\pi \mathcal{B}_j} \pi_r^\star(p) \tag{3-65}$$

$$\tag{3-66}$$

These phases define a **character**

$$\chi_\nu : \pi_1(\mathcal{L}) \to \mathbb{Z}_2 \ , \tag{3-67}$$

and in general so does any half–period (or half-integer characteristics).

The matrix

$$\Xi_r := \frac{1}{\sqrt{\mathrm{d}X}} \begin{bmatrix} \pi_r & i\pi_r^\star \\ \pi_{r-1} & i\pi_{r-1}^\star \end{bmatrix} \tag{3-68}$$

satisfies

$$Y\Xi_r\sigma_3 = D_r^{(\eta)}(X)\Xi_r \tag{3-69}$$

for $D_r^{(\eta)}(X)$ given by the formula (3-55). Moreover, from (3-40) it follows

$$\det \Xi_r = \frac{i}{\gamma_r} \tag{3-70}$$

The matrix Ξ_r is a *bona-fide* multivalued matrix-valued function on the (desingularization of the) spectral curve

$$0 = H(X,Y) = \det(Y\mathbf{1} - D_r^{(\eta)}(X)) \ . \tag{3-71}$$

It can be thought of as a multivalued function of $x = X(p)$ with branchpoint singularities at the branch-points $\alpha_j = X(\xi_j)$.

On the sheet of the map $x = X$ containing the point ∞_+ it solves a certain Riemann–Hilbert problem which we describe below; the main complication arises from the proper understanding of the square–root of the differential $\mathrm{d}X$. We note that the entries of Ξ_n are proportional to (on each sheet)

$$Q_\Delta(x) := \frac{\prod_{k=1}^{g-1}(x - \alpha_{i_k})^{\frac{1}{4}}}{\prod_{k=1}^{g+3}(x - \alpha_{j_k})^{\frac{1}{4}}} = \frac{h_\Delta}{\sqrt{\mathrm{d}X}} \ . \tag{3-72}$$

Clearly this "function" makes sense only after a suitable surgery on the plane, which is what we describe in the paragraph below; moreover it depends on the choice of the characteristic Δ under the identification between half–periods and partitions of the Weierstrass point described earlier.

Each entry of Ξ is of the form (refer to eq. 3-31)

$$\begin{aligned}
\frac{\pi_r}{\sqrt{\mathrm{d}X}} &= \frac{1}{\sqrt{h_r}} \frac{\Theta_\Delta^r(p - \infty_-)}{\Theta_\Delta^{r+1}(p - \infty_+)} \Theta \begin{bmatrix} \mathcal{A} + \vec{\epsilon} \\ \mathcal{B} + \vec{\delta} \end{bmatrix} (p + r\infty_- - (r+1)\infty_+) e^{-\int_{\alpha_1}^p \eta} Q_\Delta(X(p)) \\
&=: F_r(p)Q_\Delta(X(p)) \\
\frac{\pi_r^\star}{\sqrt{\mathrm{d}X}} &:= \frac{1}{\sqrt{h_r}} \frac{\Theta_\Delta^r(p - \infty_+)}{\Theta_\Delta^{r+1}(p - \infty_-)} \Theta \begin{bmatrix} -\mathcal{A} - \vec{\epsilon} \\ -\mathcal{B} - \vec{\delta} \end{bmatrix} (p - (r+1)\infty_- + r\infty_+) e^{\int_{\alpha_1}^p \eta} Q_\Delta(X(p)) = \\
&=: F_r^\star(p)Q_\Delta(X(p))
\end{aligned} \tag{3-73}$$

where the functions F_r and $F_r^\star$ have the monodromy

$$\begin{aligned}
F_r(p + \gamma) &= \chi_{\nu+\Delta}(\gamma)F_r(p) \\
F_r^\star(p + \gamma) &= \chi_{\nu+\Delta}(\gamma)F_r^\star(p) \qquad \forall\gamma \in \pi_1(\mathcal{L}) \ ,
\end{aligned} \tag{3-74}$$

21

where we have used $\chi_\nu \chi_\Delta = \chi_{\nu+\Delta}$ and $\chi_{-\Delta} = \chi_\Delta$ (this last valid for half–periods only). It follows also from the above formulæ that each entry of Ξ_r has a singularity at the branchpoints α_j of type $(x - \alpha_j)^{-1/4}$. In fact $\Theta_\Delta(p - \infty_-)$ (and $\Theta_\Delta(p - \infty_+)$) has simple zeroes at the Weierstrass points $\alpha_{i_k} = X(\xi_{i_k})$, $k = 1, \ldots g - 1$ appearing in eq. (3-72); since the local coordinate is $\sqrt{x - \alpha_{i_k}}$ we see in eqs. 3-73 that at **all** branchpoints we have the advocated behavior.

When thinking of $F_r, F_r^\star$ as functions on a simply connected domain of $\mathbb{C}$ (one sheet of the X–projection), they define functions with singularities only at the branch-points and essential singularities at the X–projections of the poles of the twisting differential η.

If η is an antisymmetric differential (w.r.t. the hyperelliptic involution)[11] then there is the further symmetry (which follows directly from the explicit formula defining them)

$$F_r(x) = -F_r^\star(x^\star) . \tag{3-75}$$

(Nonstandard) surgery.   Let $\mathcal{B}$ be a set of branchcuts for the projection X: specifically, if α_i are the critical values of X then the smooth hyperelliptic curve is written as

$$W^2 = \prod_{i=1}^{2g+2} (X - \alpha_i) . \tag{3-76}$$

Then $\mathcal{B}$ is a collection of mutually non-intersecting oriented arcs Σ_j, joining two points $\alpha_{j_1}, \alpha_{j_2}$ in such a way that one branch of W can be defined as a single valued function on $\mathbb{C} \setminus \mathcal{B}$. The "standard" way of performing these cuts is to join α_{2i} to α_{2i+1} (whatever numbering has been chosen). We point out that there are "nonstandard" ways of performing an equally satisfying surgery; the only condition is that at each point α_i originates an **odd number of cuts** (we will need this generality in the following). We also require (which will be enough for our later application) that $\mathbb{C} \setminus \mathcal{B}$ is **connected** and that each connected component of $\mathcal{B}$ has an **even** number of vertices. We add some oriented arcs (called **gaps**) joining each connected component of $\mathcal{B}$ to the next, and the last one to ∞. We denote by Σ the collection of all oriented cuts $\mathcal{B}$ and oriented gaps (see for example Fig. 4). On the resulting simply connected domain $\mathbb{C} \setminus \Sigma$ we have the following Riemann–Hilbert problem.

Near the poles c we have

$$\Xi_r(p) \sim E_{\eta,c}(p)^{\sigma_3} \quad \text{near } c \tag{3-77}$$
$$\sigma_3 := \text{diag}(-1, 1), \quad X(p) = x.$$

The jumps discontinuities of this Riemann-Hilbert problem are given in the following paragraph.

---

[11] There is not much loss in generality in assuming that it is antisymmetric, because it can be always antisymmetrized by an exact differential which does not change its characteristics.

Jumps on Σ. For $x \in \Sigma \setminus \{\alpha_i\}_{i=1\ldots 2g}$ let $\widetilde{\gamma}(x)$ be a closed positively oriented loop in $\mathbb{C} \setminus \Sigma$ intersecting sigma only at x (the simple connectivity of $\mathbb{C} \setminus \Sigma$ implies that the homotopy class is unique).

Define $\sharp_\Delta(x)$ as the difference between the number of Weierstrass points (in the interior region cut by $\widetilde{\gamma}(x)$) entering in the numerator of Q_Δ (which *define* Δ) and the number of the ones in the denominator. It has the following **properties** which are easily proved:

- it is odd on the cuts;

- it is even on the gaps;

- if x is on the last gap extending to infinity, then $\sharp_\Delta(x) = -4 = (g-1) - (g+3)$.

The jumps of $Q_\Delta(x)$ on the cuts and the gaps are given by $e^{2\pi i(1 \pm \frac{1}{4}\sharp_\Delta(x))}$ depending on the orientations of the corresponding cut/gap. In particular the function $Q_\Delta(x)$ is continuous across the last gap.

Any $\mathbb{Z}_2$ (which we think of as the multiplicative group consists of 1 and -1) character on $\pi_1(\mathcal{L})$ induces an assignment of signs on the cuts and gaps of Σ as follows. Define a loop $\gamma(x)$ such that:

- $\gamma(x)$ is a closed loop on $\mathcal{L}$ based at ξ_1 $(X(\xi_1) = \alpha_1)$ the Weierstrass point chosen as base–point for the Abel map;

- the X–projection of $\gamma(x)$ is a **positively oriented loop** intersecting Σ **only** at $\alpha_1 = X(\xi_1)$ and at x.

The above recipe defines $\gamma(x)$ on $\mathcal{L}$ **up to orientation**; however this ambiguity is inessential for us because we will be evaluating $\mathbb{Z}_2$ characters only (and $(-1)^{-1} = -1$).

Given an arbitrary $\aleph : \pi_1(\mathcal{L}) \to \mathbb{Z}_2$ we assign to each *oriented* cut Σ_j and each *oriented* gap $\widetilde{\Sigma}_\ell$ the sign

$$\begin{aligned}
(-)^{\Sigma_j} &:= \aleph(\gamma(x))\langle X(\gamma(x)) \cdot \Sigma_j \rangle , \qquad \text{for } x \in \Sigma_j \\
(-)^{\widetilde{\Sigma}_j} &:= \aleph(\gamma(x)) \qquad \text{for } x \in \widetilde{\Sigma}_\ell
\end{aligned} \tag{3-78}$$

where $\langle X(\gamma(x)) \cdot \Sigma_j \rangle$ denotes the intersection number for oriented curves in $\mathbb{C}$. Some simple properties are worth pointing out:

- If we denote by Σ_1 the cut attached to α_1, then for $x \in \Sigma_1$ the loops $\gamma(x)$ are homotopically trivial (in $\mathcal{L}$), hence the character $\aleph(\gamma(x)) = 1$;

- there are $2g + 1$ amongst cuts and finite gaps (i.e. excluding the gap that extends to infinity);

- as x moves along the links of Σ then $\gamma(x)$ spans $2g$ homologically independent loops (defined up to orientation) in $\mathcal{L}$;

- any $\mathbb{Z}_2$ character on $\pi_1(\mathcal{L})$ corresponds (in one-to-one fashion) to a half-integer characteristic on the Jacobian of $\mathcal{L}$.

These facts imply that we can arbitrarily assign signs to all finite links of Σ (except Σ_1) and then **define** a $\mathbb{Z}_2$ character on $\pi_1(\mathcal{L})$ by using eqs. (3-78). This is consistent since the rank of $\pi_1(\mathcal{L})$ is $2g$ and there are (excluding Σ_1) the same number of finite links in Σ.

With this preparatory material we can formulate the jump relations of Ξ on Σ

Lemma 3.1 *The matrix $\Xi_r(x)$ on $\mathbb{C} \setminus \Sigma$ satisfies the following jump relations: on the* **cuts**

$$\Xi_r(x)_+ = (-1)^{\frac{\sharp_\Delta(x)-1}{2}} \chi_{\Delta+\nu} \langle X(\gamma(x)) \cdot \Sigma_j \rangle \Xi_r(x)_- \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \tag{3-79}$$

while on the **gaps**

$$\Xi_r(x)_+ = (-1)^{\frac{\sharp_\Delta(x)}{2}} \chi_{\nu+\Delta}(\gamma(x)) \Xi_r(x)_- \tag{3-80}$$

The proof is a simple inspection of the properties of $F_r(x)$, $F_r^\star(x)$ and Q_Δ (using the argument principle). The main point that we make here is that on the gaps the jump is at most a sign and on the cuts it is given by $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ up to a sign (which is computed in detail by the formula).

We have remarked above that we can assign arbitrary signs to all finite links of Σ except Σ_1 and lift this assignment to a $\mathbb{Z}_2$ character of $\pi_1(\mathcal{L})$. Therefore we have proved

Proposition 3.1 *If we choose the orientation of Σ_1 such that $(-1)^{\frac{\sharp_\Delta(x)-1}{2}} \langle X(\gamma(x)) \cdot \Sigma_1 \rangle = 1$ then, for any choice of Δ and of orientations of the other cuts/gaps, there is a corresponding (unique) choice of the half-period ν such that the matrix Ξ_n solves the* **sign–normalized** *Riemann–Hilbert problem*

$$\Xi_r(x)_+ = \Xi_r(x)_- \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \qquad on\ the\ \textbf{cuts} \tag{3-81}$$

$$\Xi_r(x)_+ = \Xi_r(x)_- \qquad on\ the\ \textbf{gaps} \tag{3-82}$$

$$\Xi_r(x) = \mathcal{O}\left(\frac{1}{(x-\alpha_j)^{\frac{1}{4}}}\right), \quad x \to \alpha_j\ , \tag{3-83}$$

namely it has no jumps on the gaps.

4 Part II: asymptotics of orthogonal polynomials

The construction in section 3.3 suggests that the spinors could be interpreted as the large N asymptotics of some orthogonal polynomials. In this section we will show that this is indeed the case. We will make use of the steepest decent method [11], [14], [15], [33], [19] to approximate the Riemann-Hilbert problem satisfied by orthogonal polynomials with semi-classical potentials [2] for large N. The solutions of these Riemann-Hilbert problems then represent the strong asymptotics of the orthogonal polynomials in the

large N limit. The main result of this section is that, when the meromorphic functions Y and C and the Riemann surface satisfy a certain condition (Definition 4.2), the spinorial matrix constructed in Sec. 3.2 provides the solution to this deformed Riemann-Hilbert problem and hence gives the strong asymptotics of the corresponding orthogonal polynomials away from the branch points of Y where Airy asymptotics must be used instead.

The setting is not as general as in the first part of the paper: the potential will be just polynomial, namely the divisor of poles of $y = Y(p)$ will coincide with that of $x = X(p)$. Additional complications would arise in presence of other singularities and hard-edges but very interesting and somewhat surprising features are already present for this simplest class of potentials: the general situation is addressed in a future publication.

Let $(\mathcal{L}, x, y)$ be a triple such that $\mathcal{L}$ is a smooth hyperelliptic surface of genus g, $x = X(p)$ is an involution-invariant meromorphic function with two simple poles and y is a meromorphic functions on the Riemann surface $\mathcal{L}$ represented as a (nodal) hyperelliptic relation between x and y

$$y^2 = \Pi_{i=1}^{2g+2}(x - \alpha_i)M^2(x)$$

where the α_i are the (distinct) x-values of the Weierstrass points and $M(x)$ is a polynomial with roots not coinciding with any of the α_j's.

Let $V(x) = \sum_1^{d+1} \frac{u_i}{i} x^i$ be a polynomial of order $d+1$ (without constant term) such that

$$Y(p) \sim \frac{1}{2} V'(X(p)) + O(1), \quad p \to \infty_+ \tag{4-1}$$

on $\mathcal{L}$. The degrees of V, M and the genus are related by

$$g + 1 + \deg(M) = \deg(V') . \tag{4-2}$$

We define the Stokes' ray for this triple $(\mathcal{L}, x, y)$ as follows

$$\mathcal{R}_k = \left\{ x \in \mathbb{C}, \arg(x) \in \vartheta + \frac{(2k-1)\pi}{2(d+1)} \right\}$$

$$\vartheta : = \frac{\arg(u_{d+1})}{d+1}, \quad k = 0, \ldots, 2d+1 \tag{4-3}$$

The situation which will be relevant for our discussion is the following, which we formalize in a definition

Definition 4.1 *The triple $(\mathcal{L}, x, y)$ is said to satisfy the **Boutroux condition** if all contour integrals $\oint_\gamma y\mathrm{d}x \in i\mathbb{R}$ are purely imaginary.*

Since y is anti-symmetric under the hyperelliptic involution and has only poles above $x = \infty$, there are $2g + 1$ homologically independent classes to consider. It will be proved later in Section 5 that such condition can be fulfilled. In fact these "Boutroux" triples will be constructed.

Let α_1 be one of the branchpoints: we can define the following function on the curve $\mathcal{L}$

$$h(p) = \Re \left(\int_{\alpha_1}^p y(s) \mathrm{d}s \right) . \tag{4-4}$$

The choice of α_1 or any other branchpoint does not affect the definition of h (see second bulleted item after Def. 4.2). Here we think of $h(p)$ as a function on the hyperelliptic curve itself: the contour of integration is immaterial because of the Boutroux condition (the only additive monodromy of the integral is purely imaginary). Therefore $h(p)$ is a well–defined harmonic function on $\mathcal{L} \setminus \{\infty_\pm\}$. If we consider it as a function on the x-plane, then it has two branches which differ only by a sign: in particular the harmonic continuation of one branch on the punctured x–plane $\mathbb{C} \setminus \{\alpha_j\}_{j=1,\ldots,2g+2}$ has the exact same multiplicative monodromy of the analytic continuation of y (again because of the Boutroux condition).

Lemma 4.1 *Under the Boutroux–condition, the function $h(x) = \Re \left(\int_{\alpha_1}^x y(s) \mathrm{d}s \right)$ has only multiplicative monodromy with values ± 1 and is otherwise independent of the choice of contour of integration where $y_1(x)$ is the branch of y that behaves like $\frac{1}{2} V'(x)$ near $x = \infty$. Its zero level set is well defined.*

The multivaluedness of $h(x)$ is the same as the multivaluedness of $y(x)$: therefore we can make appropriate cuts on the x-plane for which $y(x)$ becomes single valued on the resulting domain. On the same domain then h will be harmonic, possibly with jump-discontinuities across those cuts. If x belongs to one of these cuts and $h(x)_\pm$ denote the boundary values on the two sides, we have

$$h(x)_+ = -h(x)_- . \tag{4-5}$$

If we can choose the cuts within the zero-level set of h, then h will be **continuous** on the whole plane and **harmonic** away from the cuts. For the time being we formalize this into the following definition.

Definition 4.2 *Let $(\mathcal{L}, x, y)$ be a triple such that $\mathcal{L}$ is represented as a nodal hyperelliptic curve by*

$$y^2 = M^2(x) \Pi_{i=1}^{2g+2}(x - \alpha_i) \tag{4-6}$$

*where α_i are distinct and that $M(x)$ is a polynomial with roots not coinciding with any of the α_j's. This triple is called **admissible** (and noncritical) if*

1. *$\oint_\gamma y \mathrm{d}x$ is imaginary for any closed curve $\gamma \in \mathcal{L}$ and $\underset{p=\infty_+}{\mathrm{res}} \, y \mathrm{d}x = 1$ (**normalized Boutroux condition**[12]).*

---

[12] The adjective "normalized" refers to the normalization of the residue at ∞_+; clearly, given a non-normalized Boutroux triple one can get a normalized one by rescaling y by a real constant.

2. *It is possible to define branch cuts of y in such a way that*

    (a) all the cuts are **finite** *Jordan arcs denoted by Σ_i joining two branchpoints;*

    (b) $h(x)$ is continuous on the whole plane and harmonic away from the cuts;

    (c) for all cuts $h(x)$ (or $-h$) is negative on both sides of each cut.

3. *For β_i any root of $M(x)$ then $h(\beta_i) \neq 0$ (noncriticality).*

Since the zero levelset

$$\mathfrak{H}_0 := \{x \mid h(x) = 0\}$$

is independent of the harmonic continuation it makes sense to study its topological properties.

Before proceeding, let us make a few observations of the set $\mathfrak{H}_0$ in the presence of the Boutroux condition.

- The set $\mathfrak{H}_0$ consists of a finite union of Jordan arcs.

- All branch-points α_j belong to $\mathfrak{H}_0$: indeed $\int_{\alpha_i}^{\alpha_j} y dx$ is half of a closed loop on the Riemann–surface $\mathcal{L}$ and –by the Boutroux condition– it is thus purely imaginary. Therefore $h(\alpha_j) = \Re \int_{\alpha_1}^{\alpha_j} y \mathrm{d}x = 0$.

- Since we assume all the branch points to be simple, there are 3 arcs originating from each branch point (as a simple computation in a local coordinate shows). They can either lead to another branch point or towards ∞. We call a branch that ends at another branch point a **closed arc** (or branch) and one that leads to ∞ an **open arc**. Moreover, the only self-intersection points of the set $\mathfrak{H}_0$ are the branch points.

- The set $\mathfrak{H}_0$ cannot contain a closed finite loop γ: if this happened, then there would necessarily be another closed loop within the region bounded by γ and without any branch point or singularity inside it. Since then (one branch of) h would be harmonic on this simply connected domain, continuous on its closure and with zero boundary value, $h(x)$ would vanish identically by the maximum modulus theorem, a contradiction.

- The function $h(x)$ is continuous across a cut Σ_i if and only if $\Sigma_i \subset \mathfrak{H}_0$.

- There can only be an odd number of branch cuts coming out of a branch point for otherwise the continuation of $y(x)$ in a neighborhood of the branch point would have no multivaluedness and the point could not be a branch–point. In our situation we can run either one or three cuts at each branch–points while remaining in $\mathfrak{H}_0$. If there is only one branch-cut then we cannot choose the branch of h so as to have a definite sign in a neighborhood of the branch-point. If there are 3 branch-cuts, then we can choose a branch of h such that h is **continuous** (not harmonic) with a semi-definite sign in a neighborhood of the branch-point.

These simple observations imply that $\mathfrak{H}_0$ is a **trivalent** graph with no closed loop with trivalent vertices at the branchpoints. It may contain possibly some open Jordan arcs not containing any branch-point (i.e. extending from ∞ to ∞). We will dwell at length on the topology of such graphs in Sec. 5.

The second condition in the definition of admissibility (Def 4.2) implies that all branch cuts belong to $\mathfrak{H}_0$ (because otherwise h would not be continuous across the cut): therefore none of the branch points between different connected components of $\mathfrak{H}_0$ can be connected by a branch cut. We then have

Lemma 4.2 *If $(\mathcal{L}, x, y)$ is admissible, then each connected component of $\mathfrak{H}_0$ contains an even number of branch points.*

Proof. We can form another graph $\mathcal{B}$ using the branch points as vertices and the branch cuts as edges. Let the connected components of this graph be $\mathcal{B}_i$. We will show that each $\mathcal{B}_i$ contains an even number of vertices.

First note that by the last bulleted item in the above list of facts, each branch point $\alpha_i \in \mathcal{B}_i$ can only be connected to an odd number (specifically 1 or 3) of branch points through branch cuts coming out of it. Let β be a branch point connected to α_i through a branch cut: then β can only be connected to an odd number of branch points. Apart from α_i, all of these points can only be connected to α_i through this branch point β, or a closed loop would form in the graph. Therefore the branch points that are connected to β will add an even number of new branch points to the connected component of $\mathcal{B}_i$ of the graph.

By repeating this argument, we see that the total number of branch points in this component must be even. **Q.E.D.**

Definition 4.3 *We will call a branch point with $0, 1, 2$ open arcs incident to it, a branch point (vertex) of* **type I, II or III** *respectively.*

In the set $\mathfrak{H}_0$ there are open arcs which do not contain any branch-points: removing those arcs yields a trivalent graph $\mathfrak{X}_0$ (the **critical graph**) with some **open edges** (i.e. edges attached to only one vertex). We can now define the branch cut structure of such a trivalent tree-like (or forest-like) graph.

Definition 4.4 *Let $\mathfrak{X}_0$ be a connected trivalent graph with no closed loop and open end-edges. Then the* **branch cut structure** *of the graph $\mathfrak{X}_0$ is a subgraph $\mathcal{B}$ of $\mathfrak{X}_0$ containing all its vertices and such that*

1. *Each vertex has either 1 or 3 edges incident to it,*

2. *$\mathcal{B}$ has no open-edges.*

3. *Each connected component of $\mathcal{B}$ has an even number of vertices.*

The edges of $\mathcal{B}$ will be called **branch cuts**. *The definition applies also to graphs $\mathfrak{X}_0$ with several connected components provided each component has the aforementioned properties.*
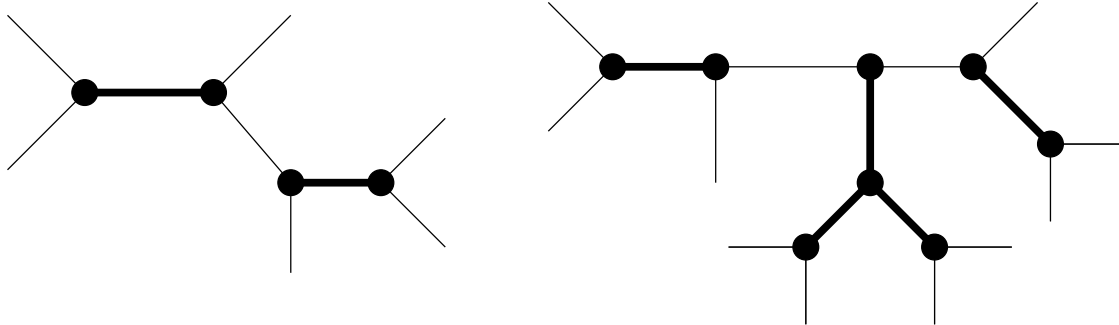
28

Figure 2: A trivalent open graph with the properties of Def. 4.4 and its branchcut structure (thick lines).

The following lemma shows that a branch cut structure for a graph with an even number of vertices exists and is unique.

Lemma 4.3 *Let $\mathfrak{X}_0$ be a graph with an even number of vertices $\alpha_1, \ldots, \alpha_{2k}$ as in definition 4.4. Then the branch cut structure $\mathcal{B}$ of $\mathfrak{X}_0$ exists and is unique.*

Proof. We can reason independently on each connected component of $\mathfrak{X}_0$, so there is no loss of generality in assuming $\mathfrak{X}_0$ to be connected.

First remove all open edges and call $\dot{\mathfrak{X}}_0$ the result of this first pruning: it is still connected and with only trivalent, univalent and bivalent vertices.

If the resulting graph $\dot{\mathfrak{X}}_0$ does not contain any bivalent vertices then $\mathcal{B} = \dot{\mathfrak{X}}_0$.

Viceversa, if $\dot{\mathfrak{X}}_0$ contains some bivalent vertices, then each of these belongs to a maximal chain of bivalent vertices; the chain connects two subgraphs of $\dot{\mathfrak{X}}_0$ which we denote by L and R (also with only uni/bi/trivalent vertices). They connect to the chain by an oddvalent vertex (if the vertex of –say– L is univalent, this means that L itself is a single vertex).

Now, if the chain has an even number of vertices, then either both L and R have an even number of vertices or they both have an odd number of vertices; we need to remove every second link in the chain so that each bivalent vertex is turned into an univalent one, and there is only one way of doing so such that all the resulting graphs contain an even number of vertices.

If the chain has an odd number of vertices, then either L or (exclusively) R have also an odd number of vertices; once more there is only one way of removing every second link in the chain so as to leave each subgraph with an even number of vertices.

The result of this second pruning is a collection of elementary segments (two univalent vertices joined by a link, which is what remains of the bivalent chain) and the two subgraphs L and R (possibly with one extra vertex on one of the two if the chain was odd). By repeating this procedure we could reduce the graph into a collection of subgraphs $\{\dot{\mathfrak{X}}_i\}$ which do not contain any bivalent vertex. The branch cut structure $\mathcal{B}$ of $\mathfrak{X}_0$ is then given by $\mathcal{B} = \cup_i \dot{\mathfrak{X}}_i$. **Q.E.D.**

In order to apply this construction to an admissible triple $(\mathcal{L}, x, y)$ we adopt the following strategy:

- Define $h(x) = \Re \int^x y_1 \mathrm{d}s$, where y_1 is the branch of y that behaves like $\frac{1}{2} V'(x)$ near $x = \infty$ with some arbitrarily chosen branch cuts. We can then define $\mathfrak{H}_0 = \{h(x) = 0\}$;

- remove the arcs of $\mathfrak{H}_0$ that contain no branch-point of y and call $\mathfrak{X}_0$ the resulting trivalent tree-like graph.

- The branch-cut structure $\mathcal{B}$ of $\mathfrak{X}_0$ can be used to *redefine* the branch-cuts of y_1 (whence the name of "branch-cut structure"). These branch-cuts may be "nonstandard". Indeed it may happen that a branchpoint has **three** branch-cuts connecting it to as many other branch-points. Although unusual this is consistent with the multi-valuedness of y.

The admissibility requirements then is the condition that the locally harmonic function h could be chosen such that

1. h is continuous on the x plane and harmonic away from the branch-cuts;

2. the sign on **both sides** of **each** branch-cut is negative.

At this point the reader may wonder how strong a condition is this and if there are any examples; in Sec. 5 we will show how to reconstruct an arbitrary admissible triple from the topological structure of its (admissible) graph $\mathfrak{X}_0$. For the time being we present only two figures exemplifying two curves of the same genus satisfying the Boutroux condition but one being admissible and the other non-admissible.

Example 4.1 *The first example is an admissible triple (left in Fig. 3, with*

$$y^2 = (x + 1 + r)(x - 1 - r)(x - e^{i\pi/3})(x - e^{2i\pi/3}) \ , \qquad , r \simeq 0.4144... \tag{4-7}$$

The second curve is not admissible, although it satisfies the Boutroux condition (rather approximately, on the left of Fig. 3)

$$y^2 \simeq (x - 2 - 0.15i)(x + 2 - 0.6i)(x - 1)(x - 0.88i) \tag{4-8}$$

For an admissible triple, we will choose branch cuts of y that satisfy condition 2 in Def. 4.2, which are the ones defined by the branch cut structure $\mathcal{B}$ of the graph $\mathfrak{X}_0$.

4.1 The G-function

The main character of the construction is (as in [19, 14]) the so-called G-function. In this context it is simply the Abelian integral $\int y \mathrm{d}x$; of course attention must be paid to its multivaluedness (which is only additive on the Riemann surface $\mathcal{L}$).

The construction of the previous section has yielded a forest graph $\mathcal{B} = \sqcup_i \mathcal{B}_i$ made of several connected components each of which is a loop-free treelike closed graph with only odd-valent (1 or 3) vertices.
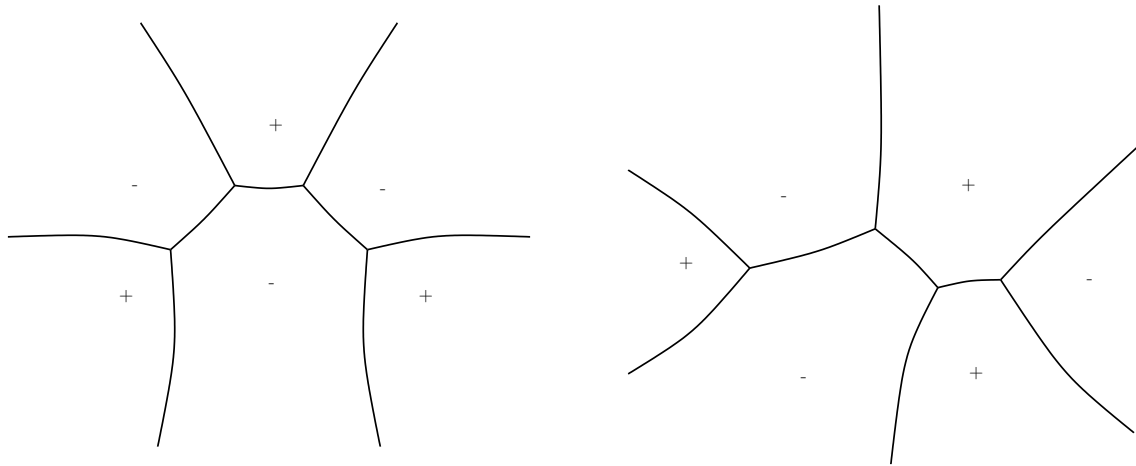
Figure 3: The critical graph $\mathfrak{H}_0 = \mathfrak{X}_0$ of two Boutroux curves: the one on the left is admissible (marked are the signs of h), the right is not admissible.
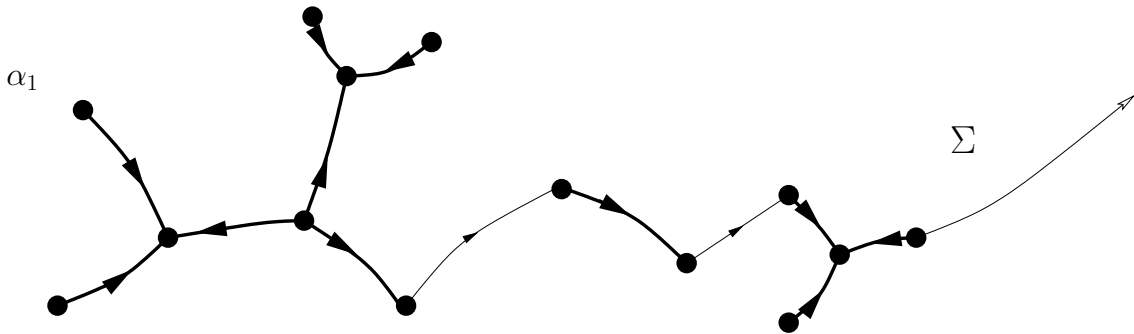

Figure 4: The set Σ connecting the different components of the branchcuts $\mathcal{B}$.

Let us order these components in some arbitrary fixed way; using some arcs not intersecting $\mathcal{B}$ we then connect a vertex of $\mathcal{B}_1$ with a vertex of $\mathcal{B}_2$, then one of $\mathcal{B}_2$ to one of $\mathcal{B}_3$ and so on and so forth, adding an open arc that extends to ∞ from the last connected component of $\mathcal{B}$: we call Σ the resulting set (see Fig. 4). We will call these arcs joining the consecutive $\mathcal{B}_i$'s the **complementary arcs** or **gaps** and denote them by $\widetilde{\Sigma}_j$, whereas we denote by Σ_i the edges of $\mathcal{B}$.

The set $\dot{\mathbb{C}} := \mathbb{C} \setminus \Sigma$ is simply connected and hence we can define unambiguously

$$G^{(i)}(x) = \int_{\alpha_1}^{x} y_i(s)\mathrm{d}s , \quad i = 1, 2, \tag{4-9}$$

where $y_1 = -y_2$ are the two branches of y and α_1 is a univalent vertex of $\mathcal{B}_1$ and $y_1 \sim \frac{1}{2}V'(x)$ at $x = \infty$. The contour of integration is taken to lie within the simply connected domain $\dot{\mathbb{C}}$.

We also assume that the edges of $\mathcal{B}$ have been oriented, so as to be able to distinguish the left from the right side. We then formulate

Lemma 4.4 *The function $G(x) := G^1(x, \Sigma)$ has jump discontinuities on Σ as follows.*

$$
\begin{aligned}
G(x)_+ &= G(x)_- + i\widetilde{\sigma}_j, & x \in \widetilde{\Sigma}_j \\
G(x)_+ &= -G(x)_- + i\sigma_j, & x \in \Sigma_j
\end{aligned}
\tag{4-10}
$$

where σ_i and $\widetilde{\sigma}_i$ are real.

Proof. Let us first consider a gap $\widetilde{\Sigma}_j$ and $x \in \widetilde{\Sigma}_j$: the two paths from α_1 to x on the left/right of the gap lift to a closed loop γ on the curve $\mathcal{L}$ because it encircles some number of connected components $\mathcal{B}_1, \ldots, \mathcal{B}_j$, each of which contains an even number of branchpoints. Therefore

$$G(x)_+ = G(x)_- + \int_{\gamma} y\mathrm{d}x , \tag{4-11}$$

which proves the first identity, with $\widetilde{\sigma}_j = -i \int_{\gamma} y\mathrm{d}x$.

In order to prove the second identity we first note that

$$G^1(x)_\pm = -G^2(x)_\pm \quad \text{on a branchcut.} \tag{4-12}$$

Let $x \in \Sigma_j$ (an edge of $\mathcal{B}$); a closed loop on the Riemann surface of y consists of a contour joining α_1 to x on the left of the cut in the first copy of $\dot{C}$ and a contour joining α_1 to x on the right of the cut on the other copy of $\dot{C}$. Therefore

$$G^1(x)_+ = G^2(x)_- + \int_{\gamma} y\mathrm{d}x . \tag{4-13}$$

From this the first assertion follows immediately with $\sigma = -i \int_{\gamma} y\mathrm{d}x$. **Q.E.D.**

Note that $h(x) = \Re G^1(x)$ and has no discontinuities (as we know already) on the gaps $\widetilde{\Sigma}_j$, where it is actually harmonic, and is continuous on the cuts Σ_j's where –however– is not differentiable (but it admits normal derivative on both sides).

From the definition of Stokes' ray (4-3), we can define sectors $\mathcal{S}_k$ in $U = \mathbb{C}/(\cup_1^{g+1}\Sigma_i)$ such that

1. As $x \to \infty$, the sector $\mathcal{S}_k$ is bounded by $\mathcal{R}_{2k}$ and $\mathcal{R}_{2k+1}$ (4-3)

2. Let the $\mathcal{Y}_+$ be the set

$$\mathcal{Y}_+ = \{x \in U, h(x) > 0\}$$

then each sector is asymptotically contained in $\mathcal{Y}_+$.

We will need later the following counting of edges and vertices

Lemma 4.5 *Let $\mathcal{B} = \sqcup_{i=1}^K \mathcal{B}_i$ be the decomposition of the branchcut structure into connected components. Let b_i be the number (which is even) of vertices and e_i the number of edges in $\mathcal{B}_i$. Then*

$$e_i = b_i - 1 \tag{4-14}$$

Therefore the total number of edges E is

$$E = 2g + 2 - K \ . \tag{4-15}$$

Proof. The first formula follows immediately by induction on the number of vertices; the second follows by considering that the sum $\sum_{i=1}^K b_i = 2g + 2$. **Q.E.D.**

4.2   The Stokes–Kirchoff normalized differential of second kind

A crucial rôle will be played by a suitable normalized differential of the second kind which we baptize "Stokes" because of its relation (to be shown) with the Stokes matrices.

Definition 4.5 *A **Stokes differential** is a second-kind differential dw with the following properties*

1. *it is antisymmetric w.r.t. the hyperelliptic involution;*

2. *it has poles (without residues) of degree at most $g + 1$ at $\infty_\pm$;*

3. *all periods around the connected components of the branchcut structure $\mathcal{B}$ are zero (or integer multiples of $2i\pi$).*

Some **properties** follow immediately:

- The Stokes differential belongs to a $\mathbb{Z}^{K-1}$ lattice of affine spaces of dimension $2g + 1 - K$. Indeed there are a span of $2g$ antisymmetric second-kind differentials with poles of that order at $\infty_\pm$ and without residue; imposing that the contour-integrals around each of the connected components is a multiple of $2i\pi$ (the residue at $\infty_\pm$ is zero so there are only as many loops as the finite gaps) imposes $K - 1$ affine constraints.

- Since the periods are zero (mod $2i\pi\mathbb{Z}$) around $\mathcal{B}_i$'s, its Abelian integral $w(x) = \int_{\alpha_1}^x \mathrm{d}w$ defines a **single valued** (mod $2i\pi\mathbb{Z}$) function on $\mathbb{C} \setminus \mathcal{B}$ with a pole of order at most $g + 1$ at infinity.

- By using similar arguments used for the G-function (together with the antisymmetry), we can show that it solves a certain scalar Riemann–Hilbert problem

$$w(x)_+ = -w(x)_- + i\mu_j \ , \qquad x \in \Sigma_j \tag{4-16}$$
$$\mu_1 = 0$$

on the edges of $\mathcal{B}$. Here Σ_1 is the edge attached to α_1 (the Weierstrass basepoint of the integration defining w); μ_1 is zero because α_1 and $x \in \Sigma_1$ can be joined by a *contractible* loop whose projection on the x-plane intersects Σ only at x and α_1.

The other important condition that we need to impose is specified in the next definition; note first that since there are no closed loops in the graph $\mathcal{B}$ we can always choose orientations of the edges of $\mathcal{B}$ in such a way that all the orientations of edges incident to any vertex are all incoming or outgoing. We assume one such orientation in the following definition.

Definition 4.6 *The Stokes differential* $\mathrm{d}w$ *is said to satisfy the* **(complex) Kirchoff's law** *if for each trivalent edge of* $\mathcal{B}$, *denoted* $\Sigma_{i_1}, \Sigma_{i_2}, \Sigma_{i_3}$ *the three incident edges (with their chosen orientation), then*

$$\mathrm{e}^{i\mu_{i_1}} + \mathrm{e}^{i\mu_{i_2}} + \mathrm{e}^{i\mu_{i_3}} = 0 \tag{4-17}$$

where all edges are oriented so that at each trivalent vertex they are all incoming or outgoing.

It should be clear that such Kirchoff-Stokes differentials do exist, since the Kirchoff's constraint poses only some number of nonlinear constraints. We need the exact count of these constraints.

If b_i is the number of vertices in $\mathcal{B}_i$ then there are $\frac{b_i}{2} + 1$ univalent vertices (of type I) and hence $\frac{b_i}{2} - 1$ trivalent vertices (recall that b_i is even). Summing up over all connected components of $\mathcal{B}$ we find the total number T of trivalent vertices

$$T = \sum_{i=1}^{K} \left(\frac{b_i}{2} - 1 \right) = g + 1 - K. \tag{4-18}$$

Therefore

Proposition 4.1 *The* **Kirchoff–Stokes** *differentials on a Boutroux curve form a* $\mathbb{Z}^{2g+1-K}$ *lattice of manifolds of dimension* $g = genus(\mathcal{L})$.

The lattice aspect is due to the obvious fact that we can arbitrarily add integer multiples of 2π to each μ_j.

4.3 Asymptotics

The following lemma shows that the spinors constructed in section 3.2 satisfy a Riemann-Hilbert problem closely related to the one satisfied by G. This will allow us to express the large N asymptotics of certain orthogonal polynomials in terms of these spinors.

The relation requires that we specify the choice of the flat line bundle $\mathfrak{L}_\eta$ of Sec. 3.2 associated to a Stokes–Kirchoff differential: we will set

$$\eta = \mathrm{d}w + Ny\mathrm{d}x . \tag{4-19}$$

Note that this η has residue $\pm N \in \mathbb{Z}$ at $\infty_\pm$ (because of the normalized Boutroux condition). We will call the line–bundle associated with this choice of Stokes–Kirchoff differential the **asymptotic line bundle** and denote it with $\mathfrak{L}_\infty$.

We note that it is the tensor product of a line bundle $\mathfrak{L}_{\mathrm{d}w}$ with the line–bundle $\mathfrak{L}_{Ny\mathrm{d}x}$; this last one, because of the Boutroux condition and since $N \in \mathbb{N}$, is a **unitary** line-bundle.

Lemma 4.6 *Let Σ be defined as in Sec. 4.1 and the half integer characteristics chosen as specified in Prop. 3.1 for the given Δ and choice of orientations of edges of $\mathcal{B}$[13]. Then the matrix*

$$\Psi := \Psi_{N,r} = \begin{bmatrix} \frac{1}{c_r} & 0 \\ 0 & -ic_{r-1} \end{bmatrix} \Xi_r(x)\mathrm{e}^{(NG+w(x))\sigma_3} \tag{4-20}$$

(where the constants c_r are specified in the proof: see eq. (3-68) for the definition of $\Xi_r(x)$) satisfies the following Riemann-Hilbert problem

$$\begin{aligned}
\Psi_+(x) &= \Psi_-(x) \begin{pmatrix} 0 & e^{-iN\sigma_j - i\mu_j} \\ -e^{iN\sigma_j + i\mu_j} & 0 \end{pmatrix}, & x \in \Sigma_j \\
\Psi_+(x) &= \Psi_-(x) \begin{pmatrix} e^{-iN\widetilde{\sigma}_j} & 0 \\ 0 & e^{iN\widetilde{\sigma}_j} \end{pmatrix}, & x \in \widetilde{\Sigma}_j \\
\Psi(x) &= \mathcal{O}\left(\frac{1}{(x-\alpha_j)^{\frac{1}{4}}} \right), & x \to \alpha_j \\
\Psi(x) &= (\mathbf{1} + \mathcal{O}(x^{-1}))x^{r\sigma_3}, & x \to \infty
\end{aligned} \tag{4-21}$$

and $\Psi(x)$ is holomorphic on $\dot{\mathbb{C}} = \mathbb{C} \setminus \Sigma$.

Remark 4.1 *An essentially identical RH problem appears in the work [12], where it was similarly reduced to the problem of Prop. 3.1. The idea perhaps was first used in [13].*

Proof of Lemma 4.6. The growth condition near the branchpoints follows from (3-83).

From the general definition in eq. (3-31) specialized to the differential (4-19) we see that the matrix Ψ is analytic in a punctured neighborhood of ∞. Indeed the essential singularity of the spinors $\pi_r, \pi_r^\star$ at

---

[13]The orientation of the edge Σ_1 attached to the basepoint of integration forces the orientation of the edges of the connected component of $\mathcal{B}$ to which it belongs (all edges should be either incoming or outgoing from the trivalent vertices).

∞ is removed by the multiplication by the exponential $\exp \int \eta \sigma_3$. This leaves us with a power growth at ∞; specifically

$$\frac{\pi_r(p(x))\mathrm{e}^{\int_{\alpha_1}^{p(x)} Ny\mathrm{d}x+\mathrm{d}w}}{\sqrt{\mathrm{d}x}} \sim c_r x^r , \qquad x \to \infty \tag{4-22}$$

has a pole of degree r at ∞_+ (which is on the sheet chosen for $p = X^{-1}(x)$) and a zero of order r at ∞_-. This is so because (by their definition) the spinors $\pi_r \mathrm{e}^{\int \eta}$ have a pole of order $r+1$ at ∞_+ and a zero of order $r-1$ at ∞_-; since $\mathrm{d}X$ has a double pole at both $\infty_\pm$ and appears in the square–root, (4-22) follows. Similar considerations show that

$$\frac{\pi_r^\star(p(x))\mathrm{e}^{-\int_{\alpha_1}^{p(x)} Ny\mathrm{d}x+\mathrm{d}w}}{\sqrt{\mathrm{d}x}} \sim \frac{1}{c_r}x^{-r-1} , \qquad x \to \infty . \tag{4-23}$$

Formulas (4-22, 4-23) *define* the constants c_r; the reason why c_r appears in the denominator in (4-23) is simply that –by construction of Serre duality–

$$1 = \operatorname*{res}_{\infty_+} \pi_r \pi_r^\star = \operatorname*{res}_{\infty} \left(c_r x^r + \mathcal{O}(x^{r-1}) \right) \left(\frac{1}{c_r}x^{-r-1} + \mathcal{O}(x^{-r-2}) \right) \mathrm{d}x \tag{4-24}$$

If we remove the exponential part from Ψ then it satisfies the Riemann–Hilbert problem (3-81, 3-82); with the exponential part the character specified by

$$\begin{array}{ccc} \exp \oint_\bullet \eta : \pi_1(\mathcal{L}) & \longrightarrow & \mathbb{C}^\times \\ \gamma & \mapsto & \mathrm{e}^{\oint_\gamma Ny\mathrm{d}x+\mathrm{d}w} \end{array} . \tag{4-25}$$

appears as the jump relations of the Abelian integral $\int Ny\mathrm{d}x + \mathrm{d}w$, thus yielding the proof. **Q.E.D.**

For completeness we report the expressions for the entries of the matrix $\Psi_{N,r}$. The two vector complex characteristics ϵ_N and $(-\delta_N)$ are half of the a and b-periods (on the chosen dissection of the curve in terms of a homology basis) of the differential $Ny\mathrm{d}x + \mathrm{d}w$ following the notation of eq. (3-31)[14]:

$$[\epsilon_N]_j = \frac{1}{2} \oint_{a_j} \left(Ny\mathrm{d}x + \mathrm{d}w \right)$$
$$[\delta_N]_j = -\frac{1}{2} \oint_{b_j} \left(Ny\mathrm{d}x + \mathrm{d}w \right) \tag{4-26}$$

After straightforward simplifications, the entries of $\Psi_{N,r}$ read (using (3-73) specified to the case of line bundle associated to $\eta = Ny\mathrm{d}x + \mathrm{d}w$)

$$(\Psi_{N,r})_{11} = \widetilde{c}_r \frac{\Theta_\Delta^r(p - \infty_-)}{\Theta_\Delta^{r+1}(p - \infty_+)} \frac{\Theta \begin{bmatrix} \mathcal{A}+\epsilon_N \\ \mathcal{B}+\delta_N \end{bmatrix}(p + r\infty_- - (r+1)\infty_+)}{\Theta \begin{bmatrix} \mathcal{A}+\epsilon_N \\ \mathcal{B}+\delta_N \end{bmatrix}(r(\infty_- - \infty_+))} Q_\Delta(X(p)) ,$$

$$(\Psi_{N,r})_{12} = i\widetilde{c}_r \frac{\Theta_\Delta^r(p - \infty_+)}{\Theta_\Delta^{r+1}(p - \infty_-)} \frac{\Theta \begin{bmatrix} -\mathcal{A}-\epsilon_N \\ -\mathcal{B}-\delta_N \end{bmatrix}(p + r\infty_+ - (r+1)\infty_-)}{\Theta \begin{bmatrix} -\mathcal{A}-\epsilon_N \\ -\mathcal{B}-\delta_N \end{bmatrix}(r(\infty_+ - \infty_-))} Q_\Delta(X(p))$$

---

[14]We only add a subscript N to emphasize the dependence on the large parameter N.

$$(\Psi_{N,r})_{21} = \frac{i}{\widetilde{c}_{r-1}} \frac{\Theta_\Delta^{r-1}(p - \infty_-)}{\Theta_\Delta^r(p - \infty_+)} \frac{\Theta\begin{bmatrix} \mathcal{A}+\epsilon_N \\ \mathcal{B}+\delta_N \end{bmatrix}(p + (r-1)\infty_- - r\infty_+)}{\Theta\begin{bmatrix} \mathcal{A}+\epsilon_N \\ \mathcal{B}+\delta_N \end{bmatrix}(r(\infty_- - \infty_+))} Q_\Delta(X(p)) \ ,$$

$$(\Psi_{N,r})_{22} = \frac{1}{\widetilde{c}_{r-1}} \frac{\Theta_\Delta^{r-1}(p - \infty_+)}{\Theta_\Delta^r(p - \infty_-)} \frac{\Theta\begin{bmatrix} -\mathcal{A}-\epsilon_N \\ -\mathcal{B}-\delta_N \end{bmatrix}(p + (r-1)\infty_+ - r\infty_-)}{\Theta\begin{bmatrix} -\mathcal{A}-\epsilon_N \\ -\mathcal{B}-\delta_N \end{bmatrix}(r(\infty_+ - \infty_-))} Q_\Delta(X(p)) \ , \tag{4-27}$$

$$\widetilde{c}_r := \frac{C_\Delta^{\,r+1}}{\Theta_\Delta^r(\infty_+ - \infty_-)} \ , \tag{4-28}$$

where C_Δ was introduced in (2-18). The constants $\{\widetilde{c}_r\}_{r\in\mathbb{Z}}$ that appear in the formulæ above have been determined by the normalization that $\Psi_{1,1} = x^r(1 + \mathcal{O}(1/x))$ and $\Psi_{2,2} = x^{-r}(1 + \mathcal{O}(1/x))$.

4.4 Semiclassical generalized orthogonal polynomials

The main theorem of this section is that –broadly speaking– when $(\mathcal{L}, x, y)$ is an admissible triple the spinors π_n associated with it give the large N asymptotics of some orthogonal polynomials of the type considered in [2].

These polynomials are defined and related to our Boutroux-admissible curve as follows: let $(\mathcal{L}, x, y)$ be an admissible triple as in Def. 4.2 and let us choose a Kirchoff-Stokes differential dw (and corresponding Stokes function w) as in Def. 4.5 and with the same notations as in the remarks that follow that definition.

Each component $\mathcal{B}_i$ with b_i vertices has $b_i/2 + 1$ type-III (univalent) vertices on the boundary of as many distinct components $\mathcal{Y}_j$ of $\mathcal{Y}_+$, each of which contains (at least) one Stokes sector; we can choose $b_i/2 + 1$ distinct oriented contours $\Gamma_{i,\ell}, \ell = 1, \ldots, \frac{b_i}{2}$ connecting one Stokes sector to the remaining ones and passing through the edges of $\mathcal{B}_i$. It is easily seen that each edge belongs to at least one such contour.

We associate a complex weight $\varkappa_{i,\ell}$ to each oriented contour $\Gamma_{i,\ell}$ as follows; the periods of dw which enter its RHP (4-16) define "complex currents"

$$\rho_j := e^{-i\mu_j} \ , \tag{4-29}$$

on each edge Σ_j of $\mathcal{B}$. Because of the Kirchoff's condition in Def. 4.6 we can always find other complex currents $\varkappa_{i,\ell}$ to associate to the contours $\Gamma_{i,\ell}$ in such a way that the "net current" through each link is precisely ρ_j. Note that $\rho_1 = 1$ always[15] (see for example Fig. 5): note that the weights $\varkappa_{i,\ell}$ do not depend on the $2\pi\mathbb{Z}$ arbitrariness entering the definition of the μ_j's.

We repeat this procedure for all connected components $\mathcal{B}_1, \ldots \mathcal{B}_K$ of the branchcut structure $\mathcal{B}$. A visual example is contained in Fig. 6 and its detailed caption.

Define now the integral operator

$$\int_\varkappa := \sum_{i=1}^K \sum_{\ell=1}^{b_i/2} \varkappa_{i,\ell} \int_{\Gamma_{i,\ell}} \ . \tag{4-30}$$

---

[15]This will mean that one of the Stokes matrices is normalized, which we can always accomplish by a conjugation.
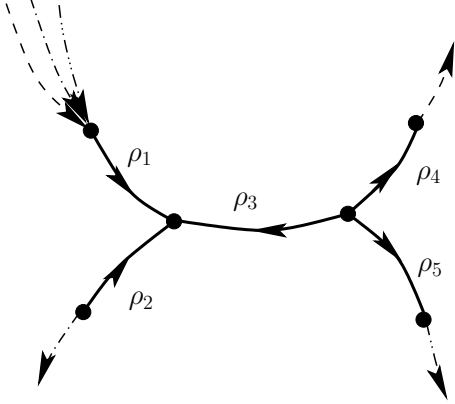
Figure 5: A connected component $\mathcal{B}_i$ with $b_i = 6$ vertices, $b_i - 1 = 5$ edges and $b_i/2 - 1 = 2$ trivalent vertices. Shown are the net currents $\rho_1, \ldots, \rho_5$ through the edges, satisfying the 2 Kirchoff constraints; we can find always appropriate complex currents to the contours (in textured linestyle) so as to give the desired net currents through the edges.

Recall that the Boutroux curve defines a "potential" $V(x)$ via (4-1). Let $p_n(x)$ be the **monic** orthogonal polynomials such that

$$\frac{1}{2\pi i} \int_{\varkappa} \mathrm{d}x\, p_m(x) p_n(x) e^{-NV(x)} = h_n\, \delta_{nm} \ . \tag{4-31}$$

Denote by $\phi_n(x)$ and $\tilde{\phi}_n(x)$ the following functions

$$
\begin{aligned}
\phi_n(x) &= p_n(x) e^{-\frac{1}{2}NV(x)} \\
\tilde{\phi}_n(x) &= \frac{e^{\frac{1}{2}NV(x)}}{2i\pi} \int_{\varkappa} \frac{e^{-\frac{1}{2}NV(z)} \phi_n(z)}{z - x} dz
\end{aligned}
$$

With these preparatory notations and remarks we are ready to formulate the main theorem of the section.

Theorem 4.1 (Strong asymptotics) *We have the following asymptotic estimates*

- *For any compact set not contained in $\mathbb{C} \setminus \mathcal{B}$, the following matrix-valued function*

$$\Phi_{N,r}(x) := \begin{pmatrix} \phi_{N+r}(x) & \tilde{\phi}_{N+r}(x) \\ \frac{2i\pi}{h_{N+r-1}} \phi_{N+r-1}(x) & \frac{2i\pi}{h_{N+r-1}} \tilde{\phi}_{N+r-1}(x) \end{pmatrix}$$

has the asymptotic behavior as $N \to \infty$ and $r \in \mathbb{Z}$ fixed

$$\Phi_{N,r} \sim \Psi_{N,r}(x) e^{-NG(x)\sigma_3} \ , \quad x \in \mathbb{C}/(\cup_{i=1}^{g+1} \Sigma_i) \tag{4-32}$$

away from the branch cuts. The matrix $\Psi_{N,r}$ was introduced in (4-20)) and the entries given by (4-27).
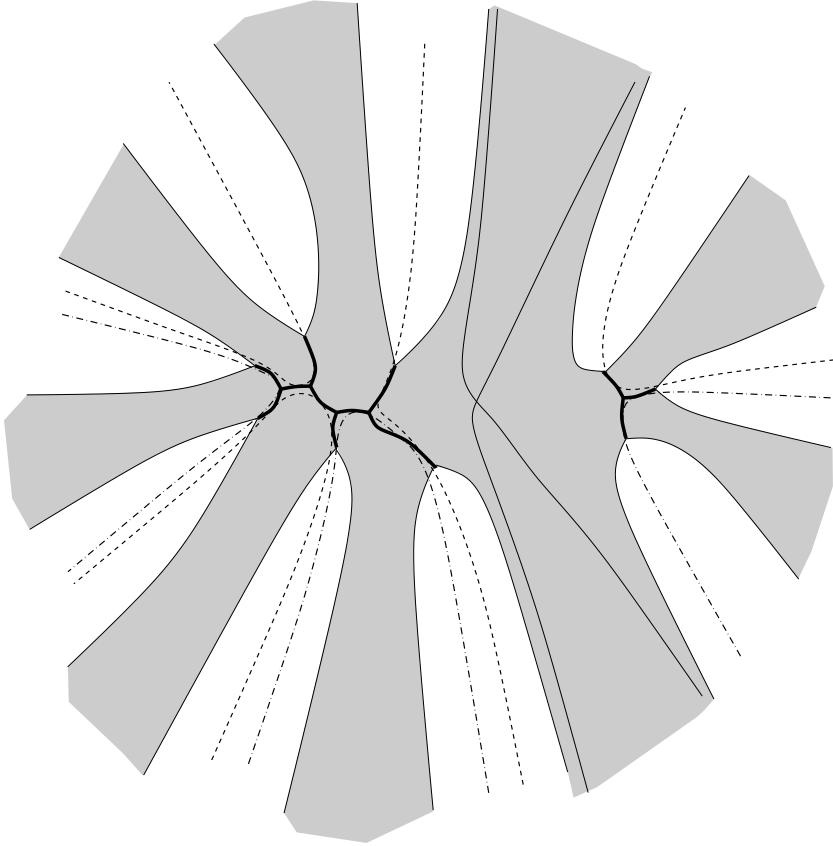
Figure 6: In this figure (which is only suggestive and not a numerical output) the grey area is where $h < 0$ and the black thick lines are where $h = 0$ and constitute the branchcut $\mathcal{B}$; here the function h is continuous but not harmonic. The thin line in the grey area denotes the levelset of $h(x) = c < 0$ passing through a saddle-point. Since there are 9 sectors within as many white areas the picture corresponds to some potential $V(x)$ of degree 9. One should conveniently think of the grey area as the "sea" and the white areas as the emerged "continents". The paths $\Gamma_{k,j}$ should connect the infinity within a continent to the infinity within another continent never going on the sea; for this purpose they can pass through the branchcuts as "causeways". In each (oriented) path runs a complex current $\varkappa_{k,j}$ and the net currents within the edges of $\mathcal{B}$ automatically satisfy Kirchoff's law. The only genericity requirement is that none of the net currents ought to vanish. In this picture there are only two components $\mathcal{B}_1$ and $\mathcal{B}_2$ (left and right respectively) with 10 and 4 branchpoints; $\mathcal{B}_1$ connects 6 continents, each of which contains a sector where $\Re(V) \to +\infty$, whereas $\mathcal{B}_2$ connects 3 such continents. The paths $\Gamma_{k,j}$ (dashed and dot-dashed) once defined can be deformed arbitrarily respecting the connectivity they provide between Stokes' sectors without affecting the values of (4-31) due to Cauchy's theorem (they are already depicted after the deformation but the reader should think of them at first as going exactly through the thick links). The existence of a situation with all the topological features of this example (and many others) will be proved in the last part of the paper.

- *On each edge Σ_i of $\mathcal{B}$ the asymptotic behavior as $N \to \infty$ is given by*

$$\Phi_{N,r} \sim \left(\Psi_{N,r}(x) \begin{pmatrix} e^{-NG(x)} & 0 \\ \pm \rho_i^{-1} e^{NG(x)} & e^{NG(x)} \end{pmatrix} \right)_\pm, \quad x \in \Sigma_i \tag{4-33}$$

Proof of Thm. 4.1. From [2], the matrix $\Phi_{N,r}(x)$ satisfies the following Riemann-Hilbert problem

$$(\Phi_{N,r})_+(x) = (\Phi_{N,r})_-(x) \begin{pmatrix} 1 & \kappa(x) \\ 0 & 1 \end{pmatrix}, \quad x \in \Gamma := \bigcup_{i,\ell} \Gamma_{i,\ell}$$

$$\Phi_{N,r}(x) = \exp \begin{pmatrix} -\frac{1}{2}NV(x) + (N+r)\log x & 0 \\ 0 & \frac{1}{2}NV(x) - (N+r)\log x \end{pmatrix}, \quad x \to \infty$$

where the $+$ and $-$ indices denote the values of the function on the left and right hand sides of the contour respectively. The piecewise constant function $\kappa(x)$ is given by

$$\kappa(x) = \begin{cases} \varkappa_{i,\ell} & x \in \Gamma_{i,\ell} \setminus \mathcal{B} \\ \rho_j & x \in \Sigma_j \subset \mathcal{B} \end{cases} \tag{4-34}$$

We can now transform the above Riemann-Hilbert problem by multiplying $e^{NG(x)\sigma_3}$ on the right of $\Phi_N(x)$, where σ_3 is the Pauli matrix $\sigma_3 = \operatorname{diag}(1, -1)$. Let

$$\Phi^1(x) := \Phi_N(x) e^{NG(x)\sigma_3} . \tag{4-35}$$

This new matrix has jump discontinuities on Σ and Γ (see Figure 4 for Σ). The Riemann–Hilbert problem is transformed into the following for $\Phi^1(x)$:

$$\Phi^1_+(x) = \Phi^1_-(x) \begin{pmatrix} e^{N(G_+ - G_-)} & \widetilde{\kappa}(x) e^{-N(G_- + G_+)} \\ 0 & e^{N(G_- - G_+)} \end{pmatrix}, \quad x \in \Gamma \supset \mathcal{B}$$

$$\Phi^1_+(x) = \Phi^1_-(x) \begin{pmatrix} e^{N(G_+ - G_-)} & 0 \\ 0 & e^{N(G_- - G_+)} \end{pmatrix}, \quad x \in \Sigma/(\Sigma \cap \Gamma)$$

$$\Phi^1(x) = \left(I + O\left(\frac{1}{x}\right) \right) x^{r\sigma_3}, \quad x \to \infty$$

Note that $\Sigma \setminus (\Sigma \cap \Gamma)$ is the union of all gaps and there is actually no jump on the last gap that extends to ∞ (because the residue of η at infinities is an integer).

Single cuts. Suppose $\mathcal{B}_i$ is a connected component of $\mathcal{B}$ with only one branch cut Σ_i (and hence $\mathcal{B}_i = \Sigma_i$)

The jump matrix of $\Phi^1(x)$ on the branch cut Σ_i is then

$$\Phi^1_+(x) = \Phi^1_-(x) \begin{pmatrix} e^{N(G_+ - G_-)} & \rho_i e^{-N(G_- + G_+)} \\ 0 & e^{N(G_- - G_+)} \end{pmatrix}, \quad x \in \Sigma_i$$

As in [14] and [33], we will make use of the factorization

$$\begin{pmatrix} a & -b \\ 0 & a^{-1} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -a^{-1}b^{-1} & 1 \end{pmatrix} \begin{pmatrix} 0 & -b \\ b^{-1} & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -ab^{-1} & 1 \end{pmatrix}$$

and write the jump matrix as

$$\begin{pmatrix} e^{N(G_+ - G_-)} & \rho_i e^{-N(G_- + G_+)} \\ 0 & e^{N(G_- - G_+)} \end{pmatrix} \tag{4-36}$$

$$= \begin{pmatrix} 1 & 0 \\ \rho_i^{-1} e^{2NG_-} & 1 \end{pmatrix} \begin{pmatrix} 0 & \rho_i e^{-N(G_- + G_+)} \\ -\rho_i^{-1} e^{N(G_- + G_+)} & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \rho_i^{-1} e^{2NG_+} & 1 \end{pmatrix} \tag{4-37}$$

We can now follow the same technique in [14], [33] and deform the contour into a 'lens' as follows. Let Γ_i^L and Γ_i^R be contours joining α_{2i-1} and α_{2i} on the left and right hand sides of Σ_i such that $h(x)$ is negative on both Γ_i^L and Γ_i^R (See figure 7). This is possible because of admissibility.

We can then deform the Riemann Hilbert problem near these branch cuts as follows. Let

$$\begin{aligned} \Phi^2(x) &= \Phi^1(x), &&\text{for } x \text{ outside the lens-shaped regions,} \\ \Phi^2(x) &= \Phi^1(x) \begin{pmatrix} 1 & 0 \\ -\rho_i^{-1} e^{2NG(x)} & 1 \end{pmatrix}, &&\text{for } x \text{ in the left hand side lens regions} \\ \Phi^2(x) &= \Phi^1(x) \begin{pmatrix} 1 & 0 \\ \rho_i^{-1} e^{2NG(x)} & 1 \end{pmatrix}, &&\text{for } x \text{ in the right hand side lens regions} \end{aligned} \tag{4-38}$$
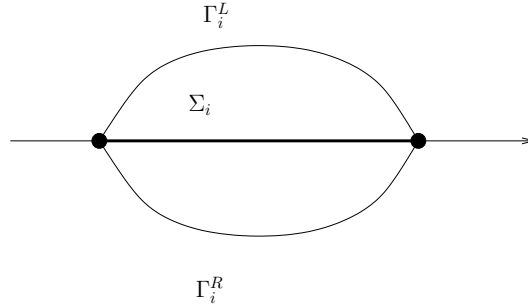

Figure 7: Lens–opening near an isolated branchcut.

Multiple cuts. Now suppose that $\mathcal{B}_j$ is a component of $\mathcal{B}$ that contains more than one branch cut. Then apart from the boundary points of $\mathcal{B}_j$, all the branch points are of type I. Let the boundary points be $\alpha_{\ell_1}, \ldots, \alpha_{\ell_n}$. Since this connected component of $\mathcal{B}$ is a loop–free trivalent tree (in a finite region of the plane) and a neighborhood of it (less $\mathcal{B}_j$) lies in the set $h(x) < 0$ by definition of admissibility, we can then join these points by curves that lies within $\{h(x) < 0\}$ as in figure 8.

Let the curve between α_{ℓ_i} and $\alpha_{\ell_{i+1}}$ be r_i. From each trivalent vertex we run three contours joining the vertex to the closest amongst the arcs r_i's without intersecting the cuts already made (see Fig. 9 for a self-explanatory exemplification). On each side of each edge Σ_m of $\mathcal{B}_j$ there are precisely two regions bounded by Σ_m a contour amongst the r_j's and the added contours from the trivalent vertices; we denote

these two regions $D_{m,L}$ and $D_{m,R}$; we then deform the Riemann-Hilbert problem near $\mathcal{B}_j$ to an equivalent one for a 2×2 matrix-valued function $\Phi^2(x)$ defined as follows in the various regions

$$
\begin{aligned}
\Phi^2(x) &= \Phi^1(x), \quad \text{for } x \text{ outside the lens-shaped regions,} \\
\Phi^2(x) &= \Phi^1(x) \begin{pmatrix} 1 & 0 \\ -\rho_m^{-1}e^{2NG(x)} & 1 \end{pmatrix}, \quad x \in D_{m,L} \\
\Phi^2(x) &= \Phi^1(x) \begin{pmatrix} 1 & 0 \\ \rho_m^{-1}e^{2NG(x)} & 1 \end{pmatrix}, \quad x \in D_{m,R}
\end{aligned}
$$

for each branch cut $\Sigma_m \in \mathcal{B}_j$.

The matrix $\Phi^2(x)$ then has jumps on $\Gamma \cup \Sigma$ and the contours shown in Fig. 10. On $(\Gamma \cup \Sigma) \setminus \mathcal{B}$ it has the same jumps as $\Phi^1(x)$ (eq. 4-37). On the contours in Fig. 10 its jumps are the following:

$$
\begin{aligned}
\nu^2(x) &= \begin{pmatrix} 1 & 0 \\ \rho_m^{-1}e^{2NG(x)} & 1 \end{pmatrix}, \quad x \in \Gamma_m^L \\
\nu^2(x) &= \begin{pmatrix} 1 & 0 \\ \rho_m^{-1}e^{2NG(x)} & 1 \end{pmatrix}, \quad x \in \Gamma_m^R \quad (4\text{-}39) \\
\nu^2(x) &= \begin{pmatrix} 0 & \rho_m e^{-N(G_-+G_+)} \\ -\rho_m^{-1}e^{N(G_-+G_+)} & 0 \end{pmatrix}, \quad x \in \Sigma_m \\
\nu^2(x) &= \begin{pmatrix} 1 & 0 \\ f_{ml}(x)e^{2NG(x)} & 1 \end{pmatrix}, \quad x \in r_{ml} \quad (4\text{-}40)
\end{aligned}
$$

where $f_{ml}(x)$ is either of the following, depending on the orientation of the contours

$$
f_{ml}(x) = \pm\rho_m^{-1} \pm \rho_l^{-1}
$$

in either of the above case, $f_{jk}(x)e^{2NG(x)} \to 0$ as $N \to \infty$ uniformly away from the branch points.

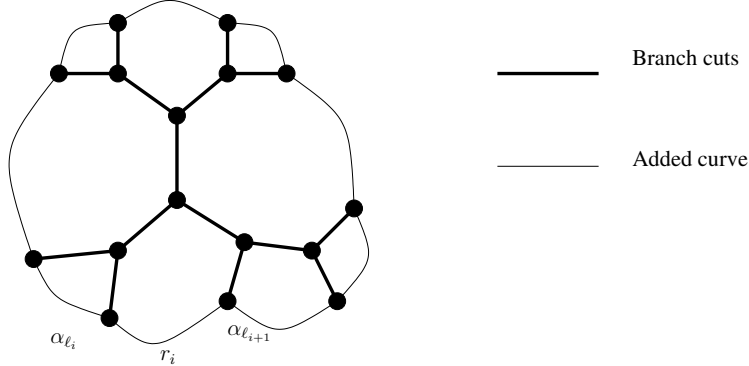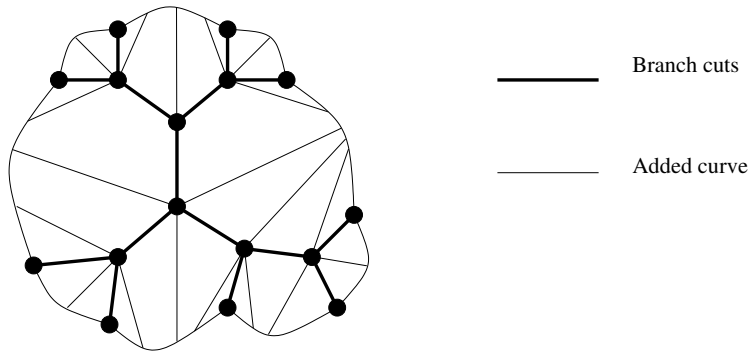Figure 8: Lens–opening near a tree of $\mathcal{B}$.

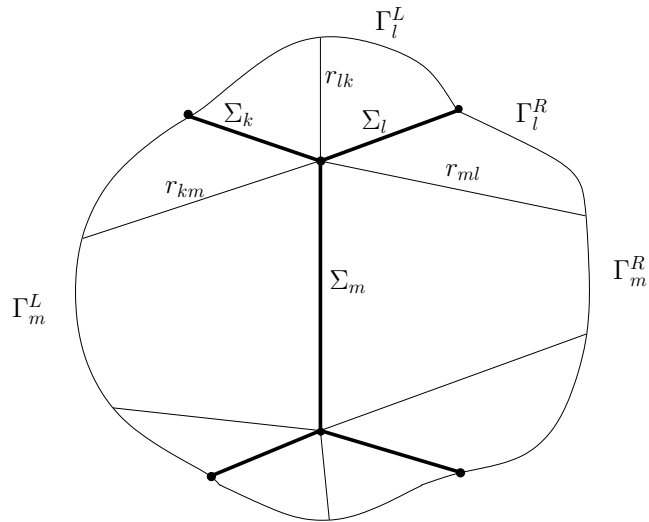Figure 9: Lens–opening near a tree of $\mathcal{B}$ and added jumps.

Figure 10: Another tree and lens-opening with added jumps.

From the jump conditions (4-39) and (4-10), we see that, away from the branch points, the Riemann-Hilbert problem for $\Phi^2(x)$ approaches the following uniformly as $N \to \infty$.

$$\Phi^2(x) \to \Phi^\infty(x)$$

where $\Phi^\infty(x)$ satisfies the following jump conditions

$$\begin{aligned}
\nu^2(x) &= \begin{pmatrix} 0 & \rho_j e^{-iN\sigma_j} \\ -\rho_j^{-1} e^{iN\sigma_j} & 0 \end{pmatrix}, & x \in \Sigma_j \\
\nu^2(x) &= \begin{pmatrix} e^{iN\tilde{\sigma}_j} & 0 \\ 0 & e^{-iN\tilde{\sigma}_j} \end{pmatrix}, & x \in \tilde{\Sigma}_j
\end{aligned} \tag{4-41}$$

and $\Phi^\infty(x) \to I$ as $x \to \infty$.

By lemma 4.6, the matrix $\Psi_{N,r}$ satisfies such the Riemann-Hilbert problem (4-21) and has the same behavior as $x \to \infty$. Hence $\Phi^\infty(x) = \Psi_{N,r}(x)$. The asymptotic formula (4-32) and (4-33) can now be seen by reversing the sequence of transformations. To complete the proof, we need to show that there exists parametrices near the branch points. We will discuss this in the next section.

4.5 Parametrix near branch points

Since the Riemann-Hilbert problem for $\Phi^2(x)$ does not tend to the one satisfied by $\Phi^\infty(x)$ near the branch points, we still need to solve the Riemann-Hilbert problem exactly near the branch points. The solution of the Riemann-Hilbert problem (4-39) for $\Phi^2(x)$ near the branch points are the local parametrices and they can be constructed from the Airy function. This problem has been discussed in many places in the literature [11], [19], [14] and we shall not repeat the details of the analysis here. Instead, we would illustrate the main idea and provide the solution to such a problem. The only relatively new feature –in this respect– is the presence of turning point (the branch-points) of type I, namely where three branchcuts are connected.

In order to complete the asymptotic analysis and obtain a uniform estimate, we need to find a solution $\Phi^p(x)$ of the Riemann-Hilbert problem (4-39) in a neighborhood O_α of a branch point α such that as $N \to \infty$, we have $\Phi^p(x) \to \Phi^2(x)$ near the boundary of such neighborhood. We transform the RH problem a first time by multiplying the solution $\Phi^2(x)$ by $\exp(-NG(x)\sigma_3)$ on the right

$$Z^p(x) = \Phi^2(x) \exp(-NG(x)\sigma_3), \quad x \in O_\alpha.$$

As a consequence, the Riemann–Hilbert problem satisfied by the matrix $Z^p(x)$ is transformed into a problem with constant jumps.

Univalent turning point (type III). If α is a turning point with only one incident branch-cut, say Σ_m, then the Riemann–Hilbert problem for Z^p in O_α consists of the following jumps on the contours γ_i

$$
\begin{aligned}
\nu^p(x) &= \begin{pmatrix} 1 & 0 \\ \rho_m^{-1} & 1 \end{pmatrix}, & x \in \gamma_1 \\
\nu^p(x) &= \begin{pmatrix} 0 & \rho_m \\ -\rho_m^{-1} & 0 \end{pmatrix}, & x \in \gamma_2 \\
\nu^p(x) &= \begin{pmatrix} 1 & 0 \\ \rho_m^{-1} & 1 \end{pmatrix}, & x \in \gamma_3 \\
\nu^p(x) &= \begin{pmatrix} 1 & \rho_m \\ 0 & 1 \end{pmatrix}, & x \in \gamma_4
\end{aligned}
\tag{4-42}
$$

where ρ_m is a constant depending on which branch cut is incident to the turning point under scrutiny. (See figure 11)

Trivalent turning point (type I). If α has 3 branch cuts attached, the function $Z^p(x)$ then satisfies the following Riemann-Hilbert problem

$$
\begin{aligned}
\nu^p(x) &= \begin{pmatrix} 1 & 0 \\ \rho_{m_1}^{-1} + \rho_{m_2}^{-1} & 1 \end{pmatrix}, & x \in \gamma_1 \\
\nu^p(x) &= \begin{pmatrix} 0 & \rho_{m_2} \\ -\rho_{m_2}^{-1} & 0 \end{pmatrix}, & x \in \gamma_2 \\
\nu^p(x) &= \begin{pmatrix} 1 & 0 \\ \rho_{m_2}^{-1} + \rho_{m_3}^{-1} & 1 \end{pmatrix}, & x \in \gamma_3 \\
\nu^p(x) &= \begin{pmatrix} 0 & \rho_{m_3} \\ -\rho_{m_3}^{-1} & 0 \end{pmatrix}, & x \in \gamma_4 \\
\nu^p(x) &= \begin{pmatrix} 1 & 0 \\ \rho_{m_3}^{-1} + \rho_{m_1}^{-1} & 1 \end{pmatrix}, & x \in \gamma_5 \\
\nu^p(x) &= \begin{pmatrix} 0 & \rho_{m_1} \\ -\rho_{m_1}^{-1} & 0 \end{pmatrix}, & x \in \gamma_6
\end{aligned}
\tag{4-43}
$$

where the orientations of the rays γ_i are chosen such that all of them are pointing away from the branch point.

A different choice of orientations will only result in a change of sign of some of these constants. The crucial observation is that with this choice of orientations we have $\rho_{m_1} + \rho_{m_2} + \rho_{m_3} = 0$, which implies that the product of all the above jump-matrices is the identity and is precisely the Kirchoff's law for our "currents" ρ_j's.

We would like to construct solutions $Z^p(x)$ to the Riemann-Hilbert problems (4-42) and (4-43) near the branch point α such that

$$
Z^p(x)e^{NG(x)\sigma_3} = \Phi^\infty(x)(\mathbf{1} + \mathcal{O}(N^{-1})), \quad x \in \partial O_\alpha, \quad N \to \infty
\tag{4-44}
$$

The solutions near these branch points can be constructed by using the Airy function. The Airy function $Ai(x)$ is the unique solution of the ODE

$$Ai''(\xi) = \xi Ai(\xi)$$

with the following asymptotics

$$Ai(\xi) = \frac{1}{2\sqrt{\pi}}\xi^{-\frac{1}{4}}e^{-\frac{2}{3}\xi^{\frac{3}{2}}}\left(1+O\left(\frac{1}{\xi^{\frac{3}{2}}}\right)\right)$$

$$Ai'(\xi) = -\frac{1}{2\sqrt{\pi}}\xi^{\frac{1}{4}}e^{-\frac{2}{3}\xi^{\frac{3}{2}}}\left(1+O\left(\frac{1}{\xi^{\frac{3}{2}}}\right)\right)$$

as $\xi \to \infty$ and $|\arg \xi| < \pi$. The branches of $\xi^{\frac{1}{4}}$ and $\xi^{\frac{3}{2}}$ in the above are principal branches with branch cut at the negative real axis.

Since $G(x) - G(\alpha)$ behaves like $cx^{\frac{3}{2}}$ near a branch point for some constant c, we can choose a function $\xi = (\frac{3}{2}N(G(x) - G(\alpha)))^{\frac{2}{3}}$ that maps the branch cut onto the negative real axis.[16] This is possible because the function $\frac{3}{2}N(G(x) - G(\alpha))$ maps the branch cut onto the imaginary axis. The function ξ then defines a one-to-one mapping between the neighborhood O_α of the branch point α and a neighborhood of the origin in the complex ξ plane. (See Figures 11, and 12.)
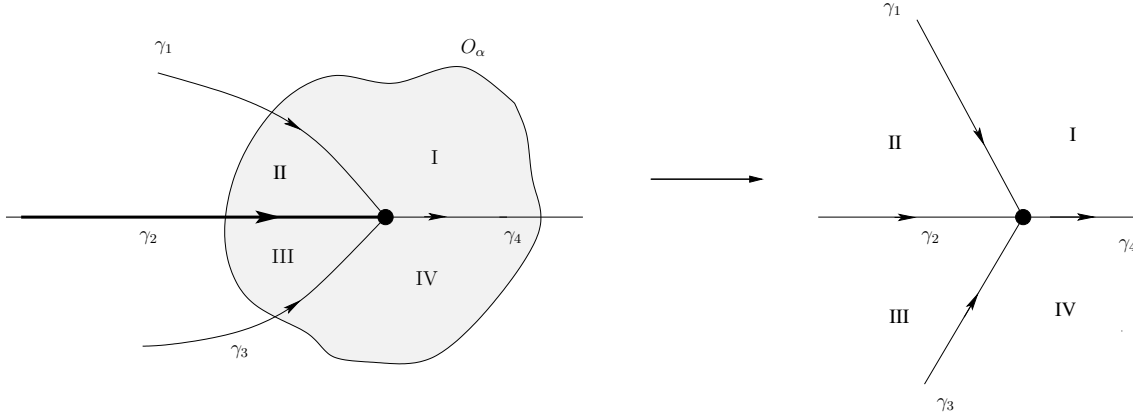

Figure 11: The function $(\frac{3}{2}N(G(x) - G(\alpha)))^{\frac{3}{2}}$ maps the neighborhood O_α of α to a neighborhood of the origin.

As in [11], a solution of the Riemann-Hilbert problem (4-42) is given by the following. Let $\Psi^p(x)$ be the matrix given by

$$Z^p(\xi) = \begin{pmatrix} Ai(\xi) & Ai(\omega^2\xi) \\ Ai'(\xi) & \omega^2 Ai'(\omega^2\xi) \end{pmatrix} e^{-\frac{i\pi}{6}\sigma_3}\rho_m^{-\frac{\sigma_3}{2}}, x \in I$$

---

[16]This requirement is due to the fact that in defining the Airy function $Ai(\xi)$, we assume $|\arg \xi| < \pi$. The requirement can be dropped if we use different branches of the Airy function. See [19]
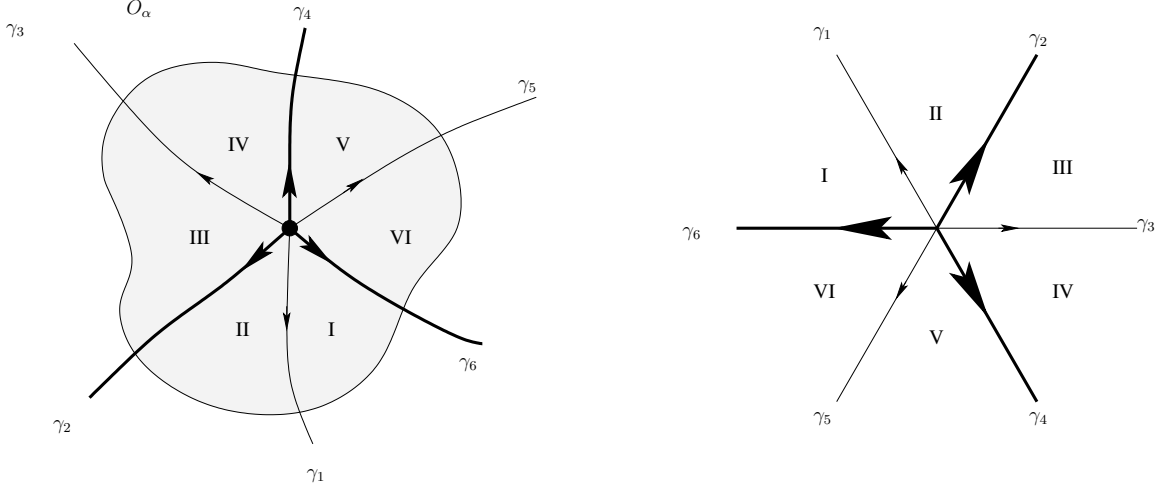
Figure 12: The function $(\frac{3}{2}N(G(x) - G(\alpha)))^{\frac{3}{2}}$ maps the neighborhood O_α of α to a neighborhood of the origin.

$$
\begin{aligned}
Z^p(\xi) &= \begin{pmatrix} Ai(\xi) & Ai(\omega^2\xi) \\ Ai'(\xi) & \omega^2 Ai'(\omega^2\xi) \end{pmatrix} e^{-\frac{i\pi}{6}\sigma_3} \rho_m^{-\frac{\sigma_3}{2}} \begin{pmatrix} 1 & 0 \\ -\rho_m^{-1} & 1 \end{pmatrix}, x \in II \\
Z^p(\xi) &= \begin{pmatrix} Ai(\xi) & -\omega^2 Ai(\omega\xi) \\ Ai'(\xi) & -Ai'(\omega\xi) \end{pmatrix} e^{-\frac{i\pi}{6}\sigma_3} \rho_m^{-\frac{\sigma_3}{2}} \begin{pmatrix} 1 & 0 \\ \rho_m^{-1} & 1 \end{pmatrix}, x \in III \\
Z^p(\xi) &= \begin{pmatrix} Ai(\xi) & -\omega^2 Ai(\omega\xi) \\ Ai'(\xi) & -Ai'(\omega\xi) \end{pmatrix} e^{-\frac{i\pi}{6}\sigma_3} \rho_m^{-\frac{\sigma_3}{2}}, x \in IV
\end{aligned}
\tag{4-45}
$$

where $\omega = e^{\frac{2\pi i}{3}}$. Then by using the relations

$$
\begin{aligned}
Ai(\xi) + \omega Ai(\omega\xi) + \omega^2 Ai(\xi\omega^2) &= 0 \\
Ai'(\xi) + \omega^2 Ai'(\xi\omega) + \omega Ai'(\xi\omega^2) &= 0
\end{aligned}
$$

one can show that $Z^p(\xi)$ defined by (4-45) does satisfy the jump conditions (4-42).

Since the only jump discontinuity of $\xi^{\frac{3}{2}}$ is at the branch cut where it changes sign, while $N(G(x) - G(\alpha))$ has jumps at the gap $\tilde{\Sigma}_m$ and the branch cut Σ_m, these two functions are not identical. However, a branch of $\xi^{\frac{3}{2}}$ can be chosen such that the difference between $\frac{2}{3}\xi^{\frac{3}{2}}$ and $N(G(x) - G(\alpha))$ remains bounded in O_α.

Let ζ be the function $NG(x) - \frac{2}{3}\xi^{\frac{3}{2}}$, where the branch of $\xi^{\frac{3}{2}}$ is chosen such that ζ remains bounded in O_α: in fact this difference is a locally constant function on $\mathcal{O}_\alpha \setminus \Sigma$.

By considering the jump discontinuities of $NG(x)$ and $\xi^{\frac{3}{2}}$, we see that $\zeta(x)$ is bounded throughout O_α and has the following jump discontinuities

$$
\begin{aligned}
\zeta(x)_+ &= \zeta(x)_- + iN\tilde{\sigma}_m, & x \in \tilde{\Sigma}_m \cap O_\alpha \\
\zeta(x)_+ &= -\zeta(x)_- + iN\sigma_m, & x \in \Sigma_m \cap O_\alpha
\end{aligned}
\tag{4-46}
$$

To fix the boundary conditions such that $Z^p(\xi)e^{NG(x)\sigma_3} \to \Phi^\infty(x)$ as $N \to \infty$ on the boundary of O_α, first note that $Z^p(\xi)e^{NG(x)\sigma_3}$ has the following asymptotic form

$$Z^p(\xi)e^{NG(x)\sigma_3} = \frac{e^{\frac{i\pi}{12}}}{2\sqrt{\pi}} \begin{pmatrix} \xi^{-\frac{1}{4}} & 0 \\ 0 & \xi^{\frac{1}{4}} \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \left(I + O\left(\frac{1}{N}\right)\right) e^{-\frac{\pi i}{4}\sigma_3} e^{\zeta(x)\sigma_3} \rho_m^{\frac{\sigma_3}{2}}, \quad \xi \to \infty$$

Let $Z_{as}(\xi)$ be the leading term of the above asymptotic expansion. Then from (4-46) and the fact that $\xi^{\frac{1}{4}} \to i\xi^{\frac{1}{4}}$ across the branch cut of $G(x)$, we see that $Z_{as}(\xi)$ has the following jump discontinuities

$$Z_{as}(\xi)_+ = Z_{as}(\xi)_- \begin{pmatrix} 0 & \rho_m e^{-iN\sigma_m} \\ -\rho_m^{-1}e^{iN\sigma_m} & 0 \end{pmatrix}, \quad x \in \Sigma_m \cap O_\alpha$$

$$Z_{as}(\xi)_+ = Z_{as}(\xi)_- \begin{pmatrix} e^{iN\tilde{\sigma}_m} & 0 \\ 0 & e^{-iN\tilde{\sigma}_m} \end{pmatrix}, \quad x \in \Sigma_m \cap O_\alpha$$

Therefore the matrix

$$E(x) = \Phi^\infty(x)Z_{as}^{-1}(x)$$

is single-valued in O_α. It is also holomorphic in O_α because it can at worst have a square-root singularity at α, but since it is single-valued, this cannot happen and hence $E(x)$ is holomorphic in O_α.

Hence the function $E(x)Z^p(x)$ is bounded in O_α, tends uniformly to $\Phi^\infty(x)$ near the boundary of O_α and satisfies the jump condition (4-42).

For a branch point with 3 branch cuts attaching to it we choose $\xi = (\frac{3}{2}N(G(x) - G(\alpha)))^{\frac{2}{3}}$ that maps the branch cut Σ_{m_1} onto the negative real axis. The following $Z^p(\xi)$ would satisfy the jump conditions (4-43)

$$Z^p(\xi) = \begin{pmatrix} Ai(\xi) & Ai(\omega^2\xi) \\ Ai'(\xi) & \omega^2 Ai'(\omega^2\xi) \end{pmatrix} e^{-\frac{i\pi}{6}\sigma_3} \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \left(\frac{\rho_{m_3}}{\rho_{m_1}\rho_{m_2}}\right)^{\frac{\sigma_3}{2}}, \quad \xi \in I$$

$$Z^p(\xi) = \begin{pmatrix} Ai(\xi) & Ai(\omega^2\xi) \\ Ai'(\xi) & \omega^2 Ai'(\omega^2\xi) \end{pmatrix} e^{-\frac{i\pi}{6}\sigma_3} \left(\frac{\rho_{m_3}}{\rho_{m_1}\rho_{m_2}}\right)^{\frac{\sigma_3}{2}}, \quad \xi \in II$$

$$Z^p(\xi) = -\begin{pmatrix} Ai(\xi) & Ai(\omega^2\xi) \\ Ai'(\xi) & \omega^2 Ai'(\omega^2\xi) \end{pmatrix} e^{-\frac{i\pi}{6}\sigma_3} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \left(\frac{\rho_{m_1}}{\rho_{m_2}\rho_{m_3}}\right)^{\frac{\sigma_3}{2}}, \quad \xi \in III$$

$$Z^p(x) = -\begin{pmatrix} Ai(\xi) & -\omega^2 Ai(\omega\xi) \\ Ai'(\xi) & -Ai'(\omega\xi) \end{pmatrix} e^{-\frac{i\pi}{6}\sigma_3} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \left(\frac{\rho_{m_1}}{\rho_{m_2}\rho_{m_3}}\right)^{\frac{\sigma_3}{2}}, \quad \xi \in IV \qquad (4\text{-}47)$$

$$Z^p(\xi) = -\begin{pmatrix} Ai(\xi) & -\omega^2 Ai(\omega\xi) \\ Ai'(\xi) & -Ai'(\omega\xi) \end{pmatrix} e^{-\frac{i\pi}{6}\sigma_3} \left(\frac{\rho_{m_2}}{\rho_{m_1}\rho_{m_3}}\right)^{\frac{\sigma_3}{2}}, \quad \xi \in V$$

$$Z^p(\xi) = -\begin{pmatrix} Ai(\xi) & -\omega^2 Ai(\omega\xi) \\ Ai'(\xi) & -Ai'(\omega\xi) \end{pmatrix} e^{-\frac{i\pi}{6}\sigma_3} \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \left(\frac{\rho_{m_2}}{\rho_{m_1}\rho_{m_3}}\right)^{\frac{\sigma_3}{2}}, \quad \xi \in VI$$

By using $\rho_{m_1} + \rho_{m_2} + \rho_{m_3} = 0$, one can check that $Z^p(\xi)$ does indeed satisfy the jump conditions (4-43).

Let $\zeta(x)$ be the following function

$$\zeta(x) = NG(x) - \xi^{\frac{3}{2}}, \quad x \in I \cup II \cup V \cup VI$$

$$\zeta(x) = NG(x) + \xi^{\frac{3}{2}}, \quad x \in III \cup IV$$

48

where the branch of $\xi^{\frac{3}{2}}$ is chosen such that $\zeta(x)$ is finite in O_α. Since $\xi^{\frac{3}{2}}$ has only one branch cut at γ_6 while $NG(x)$ has 3 branch cuts at γ_2, γ_4 and γ_6 respectively, $\zeta(x)$ has the following jump discontinuities

$$
\begin{aligned}
\zeta(x)_+ &= -\zeta(x)_- + iN\sigma_{m_1}, \quad x \in \gamma_6 \\
\zeta(x)_+ &= -\zeta(x)_- + iN\sigma_{m_2}, \quad x \in \gamma_2 \\
\zeta(x)_+ &= -\zeta(x)_- + iN\sigma_{m_3}, \quad x \in \gamma_4
\end{aligned}
\tag{4-48}
$$

To fix the boundary condition in this case, first observe that the function $Z^p(x)$ has the following asymptotic expansions

$$
Z^p(\xi) \sim \frac{e^{\frac{i\pi}{12}}}{2\sqrt{\pi}}\xi^{-\frac{1}{4}\sigma_3}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}\left(I + O\left(\frac{1}{N}\right)\right)e^{-\frac{\pi i}{4}\sigma_3}e^{-\frac{2}{3}\xi^{\frac{3}{2}}\sigma_3}\left(\frac{\rho_{m_3}}{\rho_{m_1}\rho_{m_2}}\right)^{\frac{\sigma_3}{2}}, \quad \xi \in I \cup II
$$

$$
Z^p(\xi) \sim -\frac{e^{\frac{i\pi}{12}}}{2\sqrt{\pi}}\xi^{-\frac{1}{4}\sigma_3}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}\left(I + O\left(\frac{1}{N}\right)\right)\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}e^{(\frac{2}{3}\xi^{\frac{3}{2}}+\frac{\pi i}{4})\sigma_3}\left(\frac{\rho_{m_1}}{\rho_{m_2}\rho_{m_3}}\right)^{\frac{\sigma_3}{2}},
$$
$$
\xi \in III \cup IV
$$

$$
Z^p(\xi) \sim -\frac{e^{\frac{i\pi}{12}}}{2\sqrt{\pi}}\xi^{-\frac{1}{4}\sigma_3}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}\left(I + O\left(\frac{1}{N}\right)\right)e^{-\frac{\pi i}{4}\sigma_3}e^{-\frac{2}{3}\xi^{\frac{3}{2}}\sigma_3}\left(\frac{\rho_{m_2}}{\rho_{m_1}\rho_{m_3}}\right)^{\frac{\sigma_3}{2}}, \quad \xi \in V \cup VI
$$

as $\xi \to \infty$. If we now take take $Z_{as}(\xi)$ to be

$$
\begin{aligned}
Z_{as}(x) &= \frac{e^{\frac{i\pi}{12}}}{2\sqrt{\pi}}\xi^{-\frac{1}{4}\sigma_3}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}e^{-\frac{\pi i}{4}\sigma_3}e^{\zeta(x)\sigma_3}\left(\frac{\rho_{m_3}}{\rho_{m_1}\rho_{m_2}}\right)^{\frac{\sigma_3}{2}}, \quad \xi \in I \cup II \\
Z^p(x) &= -\frac{e^{\frac{i\pi}{12}}}{2\sqrt{\pi}}\xi^{-\frac{1}{4}\sigma_3}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}e^{-\frac{\pi i}{4}\sigma_3}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}e^{\zeta(x)\sigma_3}\left(\frac{\rho_{m_1}}{\rho_{m_2}\rho_{m_3}}\right)^{\frac{\sigma_3}{2}}, \quad \xi \in III \cup IV \\
Z_{as}(x) &= -\frac{e^{\frac{i\pi}{12}}}{2\sqrt{\pi}}\xi^{-\frac{1}{4}\sigma_3}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}e^{-\frac{\pi i}{4}\sigma_3}e^{\zeta(x)\sigma_3}\left(\frac{\rho_{m_2}}{\rho_{m_1}\rho_{m_3}}\right)^{\frac{\sigma_3}{2}}, \quad \xi \in V \cup VI
\end{aligned}
$$

then from (4-48), we see that $Z_{as}(x)$ has the same jump discontinuities as $\Phi^2(x)$.

Therefore the matrix

$$
E(x) = \Phi^\infty(x)Z_{as}^{-1}(x)
$$

is holomorphic in O_α and $E(x)Z^p(x)$ will be bounded in O_α and will tend uniformly (to within $\mathcal{O}(1/N)$) to $\Phi^\infty(x)$ near the boundary of O_α. Also, it satisfies the jump condition (4-43).

Finally, if we let $\Phi^3(x)$ be the following matrix-valued function

$$
\begin{aligned}
\Phi^3(x) &= \Phi^\infty(x), \quad x \in \mathbb{C}/\cup_{i=1}^{2g+2} O_{\alpha_i} \\
\Phi^3(x) &= \Phi^p(x), \quad x \in \cup_{i=1}^{2g+2} O_{\alpha_i}
\end{aligned}
$$

where $\Phi^p(x) = E(x)Z^p(x)e^{NG(x)\sigma_3}$. Then as in [11], [14] and [19], one can show that the Riemann-Hilbert problem satisfied by $\Phi^2(x)$ tends to the one satisfied by $\Phi^3(x)$ uniformly at all points as $N \to \infty$. Therefore the function $\Phi^3(x)$ would give the asymptotics of the orthogonal polynomials as $N \to \infty$.

4.6 Density of zeroes of the orthogonal polynomials

The content of Thm. 4.1 is that the (monic) OPs p_{N+r} behave uniformly on compact sets not intersecting the branchcut structure $\mathcal{B}$ as the following expressions (see 4-27)

$$p_{N+r}(x)e^{-\frac{N}{2}V(x)} \sim (\Psi_{N,r})_{1,1}e^{-Ng(x)} = \widetilde{c}_r \frac{\Theta_\Delta^r(p-\infty_-)}{\Theta_\Delta^{r+1}(p-\infty_+)} \frac{\Theta\left[\substack{\mathcal{A}+\epsilon_N \\ \mathcal{B}+\delta_N}\right](p+r\infty_--(r+1)\infty_+)}{\Theta\left[\substack{\mathcal{A}+\epsilon_N \\ \mathcal{B}+\delta_N}\right](r(\infty_--\infty_+))} Q_\Delta(X(p))e^{-Ng(x)} \,,$$

where $\widetilde{c}_r$ is the constant (independent of N) computed after (4-27), p is point on the curve $\mathcal{L}$ with $X(p) = x$ and on the same sheet as ∞_+ and the vector of complex characteristics ϵ_N, δ_N are given in eq. (4-26).

In particular (using that the Abel map of ∞_+ is opposite of that of ∞_-) it follows that

$$
\begin{aligned}
\frac{1}{N}\ln|p_{N+r}(x)| \quad &\sim \frac{1}{2}\Re(V(x)) - h(x)+ \\
&+ \frac{1}{N}\ln\left| \frac{Q_\Delta(X(p))\Theta_\Delta^r(p-\infty_-)}{\Theta_\Delta^{r+1}(p-\infty_+)} \frac{\Theta\left[\substack{\epsilon_N \\ \delta_N}\right](p+r\infty_--(r+1)\infty_+)}{\Theta\left[\substack{\epsilon_N \\ \delta_N}\right](r(\infty_--\infty_+))} \right| \quad (4\text{-}49) \\
&+ \frac{\ln|\widetilde{c}_r|}{N}, \quad h(x) := \Re g(x) \quad\quad\quad\quad\quad\quad\quad\quad\quad\quad (4\text{-}50)
\end{aligned}
$$

uniformly over compact sets in $\mathbb{C} \setminus \mathcal{B}$.

In the RHS of (3-73) there are at most g zeroes (on the first sheet) which do not belong to $\mathcal{B}$; therefore we have immediately

Proposition 4.2 *All but at most g zeroes of the polynomials p_{N+r} are contained in an arbitrary neighborhood of $\mathcal{B}$ for N large enough.*

In different terms this means that the almost all zeroes accumulate on $\mathcal{B}$.

It should be evident that the zeroes do not actually lie on $\mathcal{B}$ but they get closer as N increases. From Thm. 4.1, formula (4-33) it follows that in a neighborhood of $\mathcal{B}$ we have

$$p_{N+r}(x)e^{-NV(x)} \sim Ae^{NG(x)} - Be^{-NG(x)} \,, \quad\quad\quad\quad\quad\quad (4\text{-}51)$$

where A, B are expressions depending in x but with at most g zeroes (whose count is asymptotically irrelevant).

Solutions are of the form

$$G(x) = \frac{1}{2N}\ln\left|\frac{B}{A}\right| + \frac{i}{2N}\arg(B/A) + 2i\pi\frac{k}{N} \quad\quad\quad\quad (4\text{-}52)$$

which shows (in a slightly heuristic way) that the zeroes have asymptotically vanishing $\Re G$ (which means that they are "close" to the branchcuts $\mathcal{B}$ where $h = \Re G$ vanishes) and $\Delta\Im G \sim \frac{1}{N}$.

Using Cauchy–Riemann's equations for G one finds that the tangential density along a cut is

$$\frac{1}{\pi}\frac{\partial \Im G}{\partial s} = -\frac{1}{\pi}\frac{\partial h}{\partial n_+} \tag{4-53}$$

where ∂s and ∂n denote the tangential and normal derivatives along the (smooth parts of the) branchcuts $\mathcal{B}$.

To put it differently the density of zeroes of p_{N+r} per unit length along an arc of $\mathcal{B}$ as $N \to \infty$ tends to

$$\nu_\infty(x) := -\frac{1}{2\pi}\left(\frac{\partial h}{\partial n_+} + \frac{\partial h}{\partial n_-}\right) = -\frac{1}{\pi}\frac{\partial h}{\partial n_\pm}, \quad x \in \mathcal{B}\setminus\bigcup\{\alpha_j\} \tag{4-54}$$

where $\partial n_\pm$ denote the normal derivatives on the two sides of the cut.

Note that this expression is positive because the function h is zero on the cut and negative (strictly) on its left/right neighborhoods (by definition of admissibility of our triple $(\mathcal{L}, x, y)$), hence the normal derivatives are strictly negative on the cuts (but away from the turning points). By elementary harmonic-function theory (or using electrostatic analogy) we then have

$$h(z) = \frac{1}{2}\Re V(z) + \int_{\mathcal{B}}\ln|z - \zeta|\nu_\infty(\zeta)\mathrm{d}s \tag{4-55}$$

(where $\mathrm{d}s$ denotes the ordinary arc–length) and hence the total mass of $\int_{\mathcal{B}}\nu_\infty(x)\mathrm{d}s$ is necessarily 1 since that corresponds to the residue of $y\mathrm{d}x$ (and to the constant in front of the logarithmic term of $G(x)$).

5 Part III: Reconstruction of (admissible) Boutroux curves

In this section we prove that any admissible graph is the graph of an admissible triple. In order to do this we take a detour in the theory of Strebel differentials, which we state here in a simple form suitable for our application; all the general statements can be found in [32] and [22].

Let $P(z) = \prod_{i=1}^{k}(z - a_i)^{\mu_i}$ be an arbitrary polynomial of degree $n = \sum_{i=1}^{k}\mu_i$. It defines a quadratic differential

$$\phi(z) := P(z)\mathrm{d}z^2 . \tag{5-1}$$

The **metric** associated to a quadratic differential ϕ, denoted by $|\phi|$, is (quite literally) $|P(z)||dz|^2$; in our case this reads

$$|\phi| = |P(z)|\left(dx^2 + dy^2\right) . \tag{5-2}$$

It is an easy exercise to see that this metric is flat, with conical singularities at the zeroes of $P(z)$ (and cuspidal singularities at the poles of order higher than 1). The flat coordinates are given by the real/imaginary parts of

$$w := \int^{z}\sqrt{P(s)}\mathrm{d}s \tag{5-3}$$

which is a locally defined parameter away from the zeroes (poles) of P.

Definition 5.1 *The horizontal lines are the lines defined by $\Im(w) = const$ (or equivalently $\arg(\sqrt{P(z)}dz) = 0, \pi$).*

Since $\sqrt{P}$ is defined up to a sign, so is w (and up to translations as well), but the notion of horizontality is well–defined.

Definition 5.2 *The* **critical horizontal lines** *are the horizontal lines which contain any of the critical points (zeroes) of P.*

The **critical horizontal graph** *(critical graph for short) is the union of all critical horizontal lines.*

It is a simple check in a local coordinate that from a zero $z = a$ of multiplicity μ there originate $\mu + 2$ critical horizontal lines, with relative angles of $2\pi/(\mu + 2)$ at a.

Clearly the critical graph of P consists of the union of a finite number of Jordan arcs.

In the general theory of trajectories of quadratic differentials one may encounter trajectories that fill domains with non-empty interior, called **recurrent** trajectories: by definition a trajectory $\gamma : \mathbb{R} \to \mathcal{L}$ ($\mathcal{L}$ an arbitrary Riemann surface) is recurrent if it belongs to its **limiting set**

$$L_\gamma := \bigcap_{t \in \mathbb{R}} \bigcup_{s > t} \{\gamma(s)\} \ . \tag{5-4}$$

Visually these are curves filling "ergodically" some region. This is not the case in the situation at hands because of the following

Proposition 5.1 (Thm 15.2 [32]) *No trajectory ray of a holomorphic quadratic differential on a domain of connectivity ≤ 3 is recurrent.*

In our case the quadratic differential is holomorphic in $\mathbb{C}P^1 \setminus \{\infty\} = \mathbb{C}$: therefore (as follows also from [22]), our critical trajectories can only connect two critical points or one critical point to ∞.

We first dwell a bit on the topology of the critical graph, $\mathfrak{X}$.

The statements summarized in the following lemmas can be found (with different notation) in [22] but are not difficult to show directly; here we restrict the attention to polynomial quadratic differentials, but more general statements can be found ibidem.

Lemma 5.1 *The graph $\mathfrak{X}$ contains no loop in the finite part of the plane.*

Let $\mathbb{C} \setminus \mathfrak{X} = \sqcup \Gamma_j$ be the decomposition in connected components.

Lemma 5.2 *Each connected component Γ_j is simply connected.*

Lemma 5.3 *Each (simply) connected component Γ_j has at most two boundaries, each consisting of an infinite piecewise Jordan curve. Each of the boundary components contain at least one critical point (since it is constituted of critical lines).*

In each Γ_j we can choose one boundary critical point z_0 as basepoint for integration and define w_j to have $\Im(w_j) \geq 0$, $w_j(z_0) = 0$.

Lemma 5.4 *The function $w_j = \int_{z_0}^z \sqrt{P(s)}\mathrm{d}s$ is a uniformization function from Γ_j to an infinite horizontal strip or to the upper half-plane.*

Lemma 5.4 shows that $\mathbb{C} \setminus \mathfrak{X}$ is the union of halfplanes and strips (topologically); the complex z-plane itself, $\mathbb{C}_z$ is then realized as a union of closed half-w-planes or horizontal w-strips with appropriate identification of the boundary points; each of these halfplanes/strips has at least one marked point on each boundary.

Definition 5.3 *A marked halfplane/strip is a copy of $\Im(w) \geq 0$ $(0 \leq \Im w \leq \ell$, respectively) together with a collection of marked points on boundary (at least one for each boundary) and up to horizontal translations.*

From this definition it is clear that a marked halfplane with $K + 1$ marked points $(K \geq 0)$ has K **real** moduli, whereas a marked strip with $K + 2$ points (at least one for each boundary) has $K + 2$ real moduli or, better, one complex modulus and K real moduli. These are simply the differences of two chosen critical points on different boundaries (which gives a complex parameter with nonzero imaginary part) and the K remaining relative positions of the other marked points on the two boundaries (which are real).

Near $z = \infty$ which is a pole of order $n + 4$ for the quadratic differential ϕ, the general theory [32] shows that

1. any horizontal line approaches ∞ asymptotic to $n + 2$ directions forming relative angles $\frac{2\pi}{n+2}$.

2. any noncritical horizontal line in a neighborhood of the pole is a topological circle, approaching the pole along two consecutive asymptotic rays.

This in particular forces the following

Lemma 5.5 *The left of the two rims of any strip approach ∞ asymptotically along the same direction. Ditto for the right.*

Proof. Suppose that –say– the right upper/lower rims approach infinity along different critical directions; then, by 2 above, there would be a noncritical horizontal trajectory within the strip that is topologically a circle with one point at ∞ and confined in $\Re(w) \ll 0$ or $\Re(w) \gg 0$. Such a trajectory would not span the whole strip from one side to
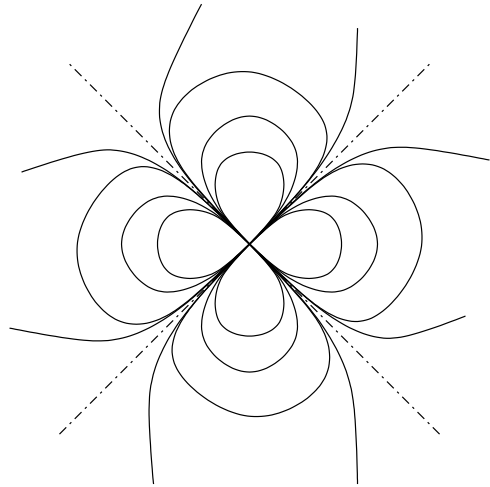
53

Figure 13: The structure of the horizontal lines near infinity (in the local parameter $\zeta = \frac{1}{z}$), here for a polynomial of degree 2.

the other. This is a contradiction, since horizontal noncritical trajectories within a strip span the whole strip from left to right. **Q.E.D.**

Lemma 5.6 *Two strips cannot have a complete boundary in common.*

Proof. If this were the case there would have to be at least one critical point on the separating boundary; such point would have only two critical lines, but then it would be necessarily regular (the critical points are all at least trivalent). **Q.E.D.**

We can depict the topological structure of the graph $\mathfrak{X}$ as follows: we draw a disk, whose boundary represents the asymptotic directions at infinity. On the boundary we mark $n+2$ points representing the asymptotic directions of approach. Inside the disk we mark the critical points and connect them according to the connectivity dictated by the graph $\mathfrak{X}$ (see the example in Fig. 14).

It follows from Lemma 5.5 that the two left (right) rims of a strip approach ∞ along the same direction, therefore they join at the same vertex on the clock diagram.

Decorations. Each strip has at least one critical point on each rim; we chose two such points α, β and associate to this pair the integral $\rho := \int_\alpha^\beta \sqrt{P(x)}\mathrm{d}x$, where the path of integration lies within said strip and the branch of the square-root is chosen so as to have a result with positive imaginary part. We call this number the **modulus** of the strip (associated to the given choice of pair of points).

If the boundary between regions has more than one critical point belonging to it then we assign to each arc between two adjacent critical points the Strebel length of that arc, namely $\int_\alpha^\beta \sqrt{P(x)}\mathrm{d}x$, where the integral is performed along said arc and the branch is chosen so as to have a **positive** result.

The clock-diagram, together with these parameters (the complex moduli of the strips and the positive Strebel lengths of consecutive critical points on the same boundary) will be called the **decorated clock diagram**.

The critical graph $\mathfrak{X}$ can be considered (from a topological point of view) as a loop–free forest made of vertices of different valencies and edges connecting vertices with either other vertices or infinity (along a given direction). In the following lemma we study some elementary enumerative properties of this graph.

Lemma 5.7 *Let $\mu_i + 2$ be the valencies of the vertices, V the number of vertices, P the number of halfplanes and S the number of strips. Let o_i be the number of **open** critical lines from the $i-$th vertex (i.e. lines that go to infinity), and c_i the number of critical lines that are closed (i.e. go to some other vertex). Then*

$$2S + P = \sum o_i \tag{5-5}$$

$$S = V - 1 - \frac{1}{2}\sum c_i \tag{5-6}$$

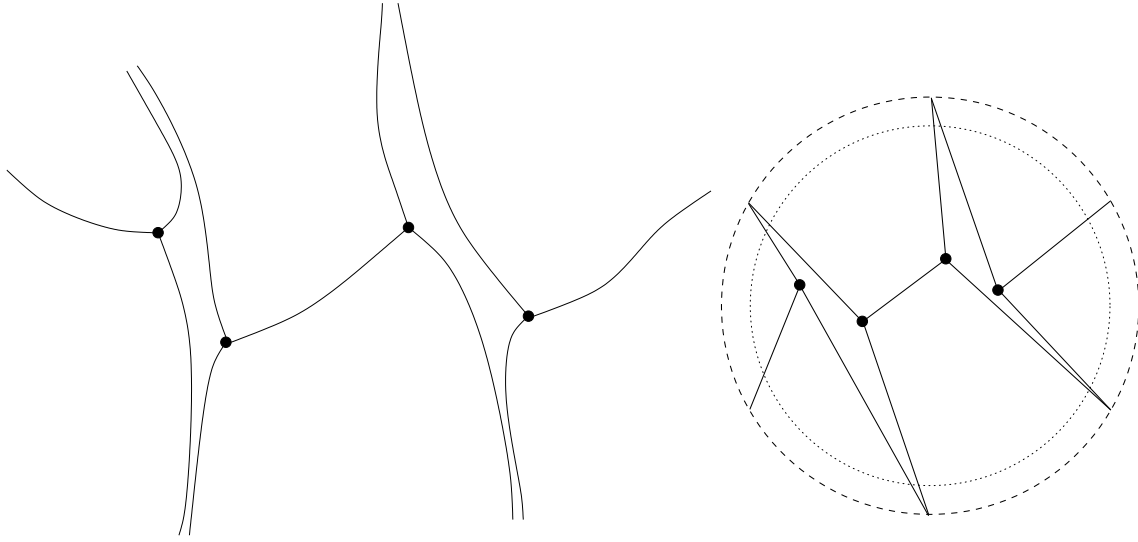*As a consequence, the number halfplanes is $n + 2 = 2 + \sum \mu_i$.*

Figure 14: An example of a critical graph for a polynomial of degree 4 and its "clock" diagram.

Proof. First of all $o_i + c_i = \mu_i + 2$ because this counts the valency of each vertex. The sum $\sum c_i$ is even because each closed line appears in exactly two vertices (i.e. is counted twice). Note that each halfplane has at least one critical point (vertex) on its boundary, and there must be two open critical half-lines on its boundary. On the other hand each strip has 4 open trajectories (two for each side). So $4S + 2P$ is the number of *sides* of open trajectories, i.e. twice the number of open trajectories $2\sum o_i$. From this we have the first equation (5-5).

The second formula is proved as follows. Assume first that there are no closed trajectories; in this case each vertex is on the boundary of some strip, unless there is only one vertex (in which case the formula holds trivially). This is so because the i-th vertex splits the plane into v_i sectors, one of which must contain another vertex. Then the two rays bounding this sector are one side of a strip. Each vertex then contributes one strip and hence the number of strips is $V - 1$.

Suppose now that the i-th vertex has c_i closed trajectories; this means that there are c_i other vertices connected to this one (necessarily distinct, since there are no closed loops): they do not contribute *per se* to the number of strips. The formula follows.

Finally, by substituting (5-6) into (5-5), one has $P = 2 + \sum \mu_i = n + 2$. **Q.E.D.**

Collecting Lemmas 5.2,5.3,5.4, 5.7,5.6 together with the idea of the clock diagram we see that any clock-diagram must follow the rules formalized in the following definition.

Definition 5.4 *A (topological) clock diagram is a regular $n + 2$-gon called the* **clock** *(whose vertices are*

"infinite") containing V *vertices (the "finite vertices") of valencies $\mu_i + 2$ such that*

$$n = \sum_{i=1}^{V} \mu_i$$

Moreover

1. *each edge connects two finite vertices or one finite vertex and an infinite one;*

2. *there are no loops in the interior of the clock;*

3. *each component Γ_j in which the interior of the clock is separated by the edges is topologically a marked halfplane or strip.*

4. *each half–plane borders exactly one of the sides of the clock.*

5.1   Inverse problem

We want to show that for any graph $\mathfrak{X}$ giving rise to a **decorated** clock-diagram with the topological properties of Def. 5.4, there is a polynomial whose critical graph corresponds to the given graph. This is essentially an (interesting) exercise in **welding** [32].

In fact the problem is essentially solved and in the more general setting of rational quadratic differentials in [22], so we basically rephrase the contents of [22] adding the details that are needed for our application.

Suppose we have a clock-diagram; by Def. 5.4 the interior of the clock is partitioned in regions Γ_j each of which is either a halfplane or a strip (topologically). This implies that the given diagram tells us how to glue (topologically for the time being) half-planes and strips so as to have a simply connected topological space.

We now choose an arbitrary decoration of the clock-diagram: for each component of a boundary with $K + 1$ marked points we assign (arbitrarily) K real positive numbers, representing the relative distances between consecutive marked points; for each strip, we choose two marked points on the two distinct boundaries and assign an arbitrary complex parameter ρ, $\Im(\rho) > 0$ to this pair.

By virtue of this construction we will have **decorated** marked halfplanes/strips in such a way that the decorations on the boundaries (i.e. the relative distances of the marked points) match between neighboring Γ_j's.

We think of these marked abstract halfplanes/strips Γ_j as realized in copies of the w-upper-half-plane and introduce a **flat** coordinate w_j for each halfplane/strip, normalized so that w_j vanishes at one of the marked points on one boundary.

We form the topological surface $X = \sqcup \overline{\Gamma_j}/\sim$, where the equivalence relation is the **metric** identification of the boundaries according to the coordinates w_j and the topological structure dictated by the clock diagram.

5.1.1 Conformal structure

The construction of the conformal structure follows [22], with some minor deviations.

Our topological surface X is connected and simply connected since it is a model of the interior of the clock (and hence of the plane): the conformal structure is defined in an interior point of Γ_j by the coordinate w_j itself. A neighborhood O of a critical vertex a of multiplicity μ intersects $\mu + 2$ halfplanes/strips which we simply denote by $\Gamma_1, \ldots, \Gamma_{\mu+2}$. The local coordinate z in this neighborhood O is then defined by

$$z(p) = \mathrm{e}^{\frac{2i\pi\ell}{\mu+2}} w_\ell^{\frac{2}{\mu+2}}, \quad p \in O \cap \Gamma_\ell, \qquad w_\ell = \int_a^z \sqrt{P(s)}\mathrm{d}s, \quad \ell = 1, \ldots, \mu+2. \tag{5-7}$$

Near a smooth common boundary points between –say– Γ_1, Γ_2 of coordinates $s, \widetilde{s}$, the coordinates are $(w_1 - s)$ and $\pm(w_2 - \widetilde{s})$ on the two sides (respectively), where the sign depends on the relative orientations of the strips/halfplanes at that point.

It follows from these definitions of the conformal structure that $(\mathrm{d}w_j)^2$ lifts to a well–defined **holomorphic** quadratic differential on the (now) Riemann surface X; this differential has zeroes precisely at the marked points and are of multiplicities μ.

In order to conclude that it defines a polynomial, we need to compactify the surface and show that the quadratic differential has a pole at the point of compactification.

Compactification We first topologically compactify X using one-point compactification (Alexandrov). The ensuing topological space is connected, compact and simply connected.

Let $n = \sum \mu_i$ be the sum of all multiplicities of zeroes of the quadratic differential $\mathrm{d}w^2$ which –as discussed above– is globally defined on $\overline{X} \setminus \{\infty\}$.

We need to define a local coordinate ζ at the compactification point ∞; since a neighborhood of ∞ intersects all regions Γ_j's, we need to express the to-be-defined coordinate ζ in terms of the coordinates w_j naturally defined in each region; moreover we must do so in such a way that boundaries of adjacent regions are mapped to the same line in the ζ-plane near $\zeta = \zeta(\infty) = 0$.

Let U_∞ be a neighborhood of ∞ and let V_j be the intersections of U_∞ with all the domains Γ_ℓ's. To have a pictorial idea we can turn the clock-diagram inside-out so that the inside of the $(n + 2)$-polygon represents the point ∞ and the lines from the vertices the critical trajectories (see figure 15 for an example).

Let us proceed clockwise on such an inverted-clock diagram; because of our convention $\Im(w_j)$'s the orientation of two consecutive half-planes (the regions that border one edge of the inverted-clock diagram) must be opposite, independently on the number and orientations of the subregions of strips that may be incident on the vertex between the two planes.
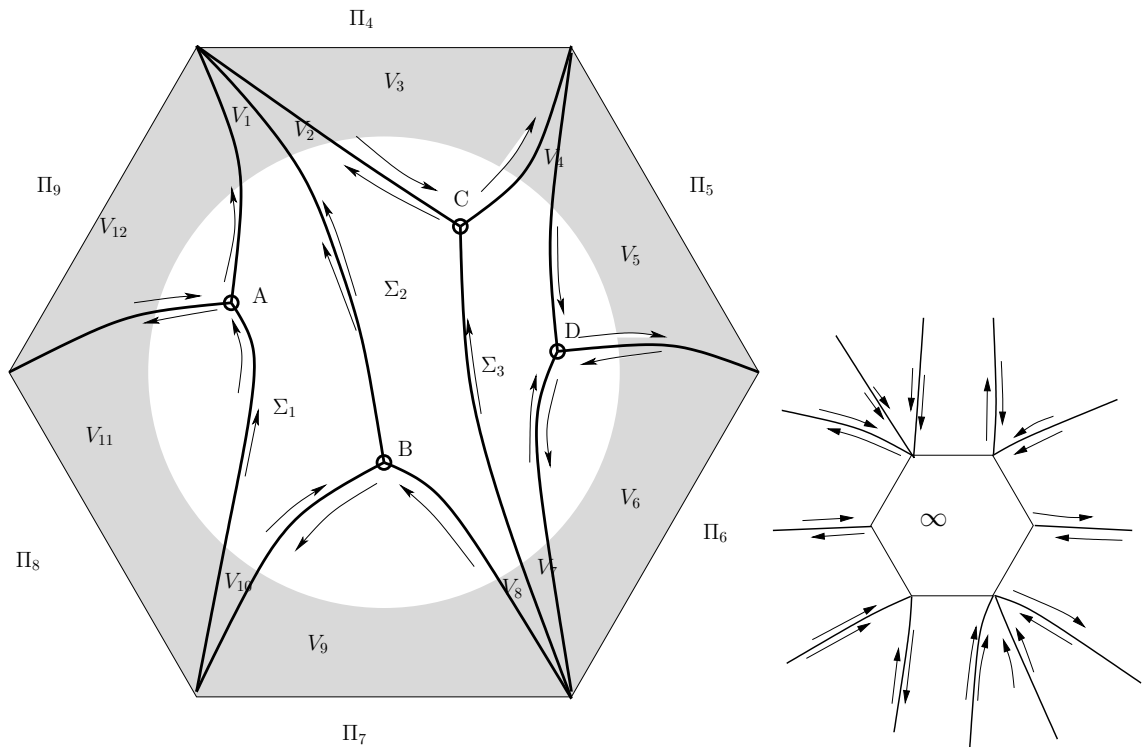
Figure 15: An example of the process of compactification and the corresponding inverted clock-diagram at the point of compactification.

Starting from region V_1 we can continue the differential $\mathrm{d}w = \mathrm{d}w_1$ to the region V_2, where $\mathrm{d}w = \pm \mathrm{d}w_2$. If V_2 is one end of a strip then we may have one or the other sign, depending on our choice of uniformizer; if V_2 is another half-plane then necessarily we have a minus sign according to our observation above.

Recall that there are $n + 2 = 2 + \sum \mu_j$ half-planes (edges); therefore after going around ∞ once we will have monodromy sign of $+1$ if n is even, -1 if n is odd.

In other words we have

Lemma 5.8 *If $n = 2k$ then the differential $\mathrm{d}w$ is well-defined (up to overall sign) in $U_\infty \setminus \{\infty\}$. If $n = 2k + 1$ then the differential $\mathrm{d}w$ is well-defined on the double-cover of $U_\infty \setminus \{\infty\}$.*

In the case of even n (the case of interest to us) we can define a function w by integrating (one branch of) $\mathrm{d}w$; this function is not single-valued in general but has additive monodromy

$$ w \mapsto w + \oint \mathrm{d}w = w + 2i\pi\beta . \tag{5-8} $$

In fact it is not difficult to see that β has a geometrical meaning of the alternating sum over all vertices of the clock of the total modular width of the strips incident to each vertex.

If n is odd it is also not difficult to see that $\oint \mathrm{d}w$ (where the integral winds twice around the compactification point) is always zero due to the monodromy $\mathrm{d}w \mapsto -\mathrm{d}w$ after one loop.

A uniformizer $\zeta = \zeta(w)$ must have the following properties

- $\zeta(w)$ is a locally analytic function for $|w| \gg 0$;

- $\zeta(w + \beta) = \zeta(w)$;

- $\zeta(w)$ has a singularity of type $w^{-\frac{n}{2} - 1}$ (so as to be able to accommodate $n + 2$ halfplanes in one).

We sketch the main steps of the construction following the ideas in [23, 32].

Consider the intersection of a halfplane, denoted by Π_1, with a neighborhood of infinity; Π_1 comes equipped by definition with a coordinate w_1. We now proceed clockwise around infinity and consider the next region of intersection; this might be a strip or another halfplane which borders Π_1 on the boundary $\Im(w_1) = 0$, $\Re(w_1) \gg 1$.

Using now the coordinate of this region (possibly up to translations) we can extend the coordinate w_1 to this strip; if the other edge of this strip borders another strip we repeat the procedure until we reach a strip which borders with the next clockwise halfplane Π_2.

The region we obtain looks like the one in Fig. 16

Let's denote H_{ρ_1} this region, where ρ_1 is the sum of all complex moduli of the parts of strips that we have attached to the right of the halfplane: we repeat this construction for all $n + 2$ half-planes, obtaining thus regions $H_{\rho_2}, \ldots, H_{\rho_{n+2}}$.
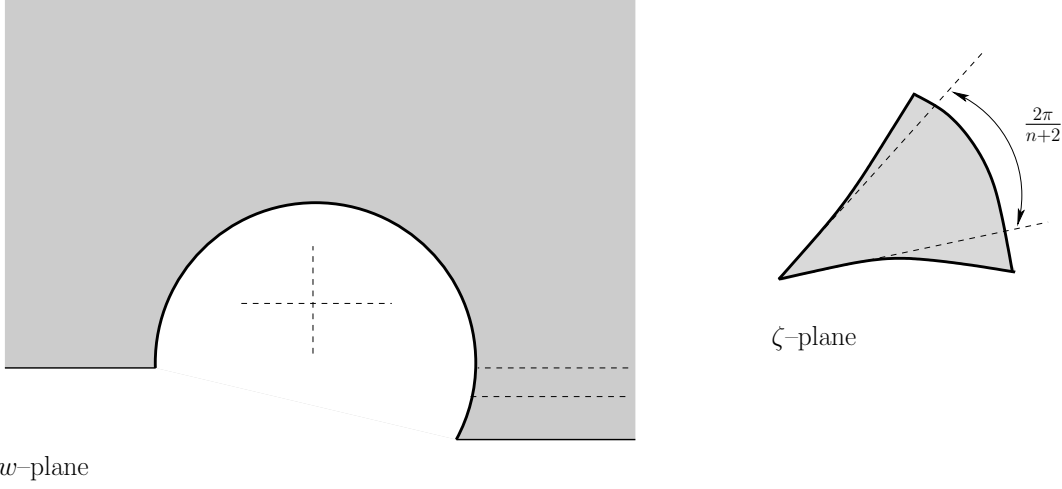
w–plane

ζ–plane

Figure 16: The regions used for compactification; on the left an example of a shifted region $H_\rho + \rho'$ and on the right the image through $\zeta(w)$ of that region; the asymptotic directions have an angular separation of $2\pi/(n+2)$. Subsequent regions are mapped to the ζ plane by choosing different branches so as to fill all sectors of a neighborhood of $\zeta = 0$. Also shown are the separations between the added half-strips on the bottom right of the w halfplane (in this case two strips).

In order to glue one such region to the next, keeping in mind that consecutive half-planes have upside-down matching, we have to glue H_{ρ_1} with $-H_{\rho_2} + \rho_1$ and so on and so forth. We obtain a chain of regions (we assume for definiteness n even)

$$H_{\rho_1} \to -H_{\rho_2} + \rho_1 \to H_{\rho_3} - \rho_1 + \rho_2 \to -H_{\rho_4} + \rho_1 - \rho_2 + \rho_3 \cdots H_{\rho_{n+2}} + \sum_{j=1}^{n+1} (-1)^j \rho_j \ . \qquad (5\text{-}9)$$

The resulting Riemann surface $\mathcal{V}$ has in particular two boundaries corresponding to a half-line in the negative real w axis of H_{ρ_1} and and a half-line in the line $\{w + \rho_{n+2}, \ \Im(w) = 0, \ \Re(w) \gg 0\}$ of the last region; we denote them by Σ_L, Σ_R respectively.

In addition on $\mathcal{V}$ the function w is now single valued and realizes a $n/2 + 1$ cover of a neighborhood of ∞; the two boundaries Σ_L and Σ_R are mapped to

$$w(\Sigma_L) \subset \{\Im(w) = 0 \ \ \Re(w) \ll 0\} \qquad (5\text{-}10)$$

$$w(\Sigma_R) \subset \{\Im(w) = \Im(\beta) \ \ \Re(w) \ll 0\} \qquad (5\text{-}11)$$

where

$$\beta := \frac{1}{2i\pi} \sum_j^{n+2} (-1)^j \rho_j \qquad (5\text{-}12)$$

and the points on these lines are identified by $w \sim w + \beta$. Now consider the map $\zeta(w)$ defined implicitly (up to overall additive constant) by

$$\zeta^{-\frac{n+2}{2}} + \beta \ln \zeta = w \qquad (5\text{-}13)$$

60

For $|w|$ large there is a univalent branch of $\zeta(w)$ which behaves as any given branch of $w^{-\frac{2}{n+2}}$.

Under this map, the line Σ_L is mapped by (one branch of) $\zeta(w)$ to a Jordan arc $\zeta(\Sigma_L)$ approaching $\zeta = 0$. Consider a neighborhood U of $\zeta = 0$ with a cut along this arc, so as to have a simply connected domain. The image $w(U \setminus \zeta(\Sigma_L))$ covers a punctured neighborhood of $w = \infty$ precisely $n/2 + 1$ times and realizes a biholomorphic equivalence with the Riemann surface $\mathcal{V}$; because of the periodicity of $\zeta(w)$, the two sides of the cut $\zeta(\Sigma_L)$ are mapped to Σ_L and Σ_R.

Thus the function ζ realizes the above identification of Σ_L and Σ_R and gives a conformal uniformization of $\mathcal{X}$; in addition ζ is continuous at the compactification point ∞ and therefore defines a conformal structure at this point[17].

This construction also shows that the differential $\mathrm{d}w^2$ has a pole of order $n+4$ at the compactification point ∞ in the local uniformizer ζ.

Since $\overline{X}$ is holomorphically equivalent to the Riemann-sphere the differential dw^2 is represented in a global uniformizing coordinate z as

$$\mathrm{d}w^2 = P(z)\mathrm{d}z^2 \ , \tag{5-14}$$

with $P(z)$ a polynomial of degree n: we have thus succeeded in proving the equivalence of polynomials and decorated admissible graphs.

Remark 5.1 *In a nutshell the above discussion boils down to the following statement: the space of polynomials of given multiplicities of zeroes*

$$P(z) = \prod_{j=1}^{K}(z - a_j)^{\mu_j} \ , \tag{5-15}$$

can be (locally) parametrized by decorated admissible Strebel graphs. The "coordinates" are (essentially)

$$E_{j-1} := \int_{a_1}^{a_j} \sqrt{P(z)}\mathrm{d}z \ , \qquad j = 2, \dots, \tag{5-16}$$

up to translations and dilations that leave these integrals invariant (and for some choice of the contour of integrations and branch of the square-root).

A naïve parameter counting confirms this fact: polynomials of given multiplicities of zeroes are parametrized by the K positions of the zeroes (distinct) and an overall multiplicative constant, in total $K + 1$ complex parameters. If we factor out the group of invariance of the above integrals, namely the action of translations $z \mapsto z + c$ and dilations

$$\widetilde{z} = \lambda z \ , \qquad \widetilde{P}(\widetilde{z}) = \lambda^2 P(\lambda z) \tag{5-17}$$

we have the same number of parameters, $K - 1$.

---

[17]If n is odd we have already shown that (on the double cover) w has no monodromy and the construction works identically with the caveat that we need to use the double cover of the neighborhood of the compactification point.

For our application, however, it is essential to control the global topological structure of the critical graph, something the above coordinates tell nothing about. One obvious reason is that the zeroes of the polynomials are on the same footing and hence –even for the same polynomial– we could assign different coordinates. More importantly, we could have two inequivalent clock diagrams with decorations given by the same numbers: the corresponding polynomials would then be different but with the same Strebel lengths between zeroes[18].

5.2 (Admissible) Boutroux curves

In the case of relevance to our paper we must restrict the parametrization given by the decorated clock–diagram of a polynomial to a suitable submanifold; this is the submanifold of polynomials of the form

$$P(z) = -M^2(z) \prod_{j=1}^{2g+2} (z - \alpha_j) = -y^2 \tag{5-18}$$

(where the sign is just conventional for our application so that the function h introduced in Def. 4.2 is the imaginary part of the Strebel flat coordinate w) with additional constraints on the clock–diagram and on its decoration described below.

The corresponding clock–diagram has only trivalent vertices (the α_j's) and even-valent ones (the zeroes of M). The Boutroux condition (Def. 4.1), as discussed implies that the **critical** 0-level set[19] $\mathfrak{X}_0$ of $\Im \int_{\alpha_1}^x \sqrt{P(z)} \mathrm{d}z$ is well defined independently of the choice of critical point α_j and independently of the choice of contour of integration.

In this case the **height** function $h = \Im w$ can be defined as a continuous global function with smooth saddle points at the zeroes of $M(z)$: indeed it is easily seen that there is a consistent choice of signs for the Strebel coordinate near each even critical point (the zeroes of M all appear with even multiplicity in the quadratic differential) in such a way that the height function $h = \Im(w)$ is harmonic there.

The critical horizontal graph of $P(z)$ thus naturally splits into $\mathfrak{X}_0 \sqcup \mathfrak{X}_1$, where all the α_j's belong to $\mathfrak{X}_0$. The distances between critical points α_j belonging to the same connected
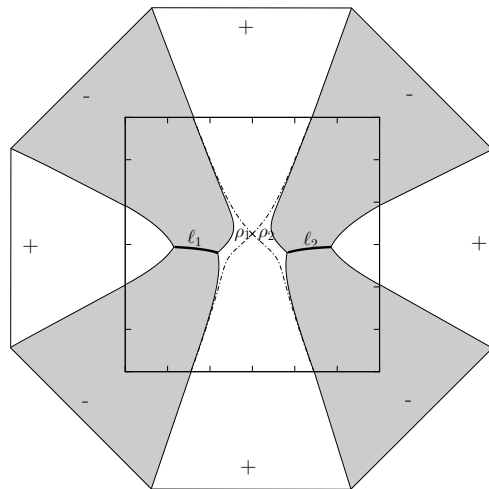

Figure 17: An example of reconstruction of admissible triple. The part within the graduated square is an actual numerical output.

---

[18]To put it differently, clock diagrams with given number of finite vertices represent **cells** in the space of polynomials; the decorations give coordinates to each cell, but it is meaningless on a global level to consider only the decoration.

[19]The **critical level set** of a function is the union of all connected components of the level-set that contain at least one critical point.

component of $\mathfrak{X}_0$ are arbitrary real numbers. The other mod-
uli are the complex moduli of the strips: they must satisfy the real constraint that the heights of the strips "between" two connected components of $\mathfrak{X}_0$ must add up to zero.

The reconstruction theorems proved in the previous sections could be rephrased by saying "if you can draw the graph you want, then it exists (with the same topological features)"; instead of giving an abstract account of the procedure we find it more instructive and transparent to analyze some examples.

Let us consider the examples in figures (17, 18, 20, 19). In Fig. 17 the black contours make up $\mathfrak{X}_0$, whereas the complete Strebel critical graph is the collection of solid and dashed curves. The numbers ℓ_1, ℓ_2 are arbitrary real (positive) numbers; they are $\int_\alpha^{\alpha'} \sqrt{P} \mathrm{d}x$ (where α, α' denote the two critical points. The two numbers ρ_i are instead complex (with nonzero imaginary part) and correspond to $\int_\alpha^c \sqrt{P} \mathrm{d}x$, where c is the critical point in the middle. They satisfy the only constraint that $\Im(\rho_1) = \Im(\rho_2)$. Of course there could be more strips between the two connected components of $\mathfrak{X}_0$. (The normalization $\operatorname*{res}_\infty y \mathrm{d}x = 1$ imposes the constraint $\ell_1 + \ell_2 = 1$).

The topology of this example is the "usual" one in the *two-cut* potentials: there are two arcs supporting the asymptotic distributions of zeroes of the orthogonal polynomials and the differential $y \mathrm{d}x$ has a critical point between them. The degree of the potential V is here quartic.

The second example (Fig. 18) is "unusual" but it can happen for this class of semiclassical orthogonal polynomials: it is a cubic potential without double zero for the differential $y \mathrm{d}x$ (similarly, one could have also a quartic potential without double zero). The support of the zeroes of the polynomials corresponds to the thick lines. There are three free real parameters ℓ_1, ℓ_2, ℓ_3 (subject to $\ell_1 + \ell_3 = 1$, which is the normalization condition).

The third example (Fig. 20) is the case of two double points and double support (something that cannot happen for ordinary orthogonal polynomials)

The fourth example (Fig. 19) shows that it can happen that not all Stokes sectors could be joined by curves Γ satisfying the requirements for the steepest–descent method: here $\ell_1 \in \mathbb{R}_+$ is arbitrary, and $\rho_1, \rho_2 \in \mathcal{H}_+$ are also arbitrary, but with the condition $\Im(\rho_2) > \Im(\rho_1)$. The Stokes sector on the right (represented by the edge of the hexagon bordering the white area) cannot be joined to the other sectors because there is a river of $h < 0$ "underwater" in between (the critical value at the point c of intersection of the dashed lines has $h(c) < 0$).

All these examples are admissible for our asymptotic study in the sense of Def. 4.2; even without a formalization of the rules that make an admissible triple, the reader
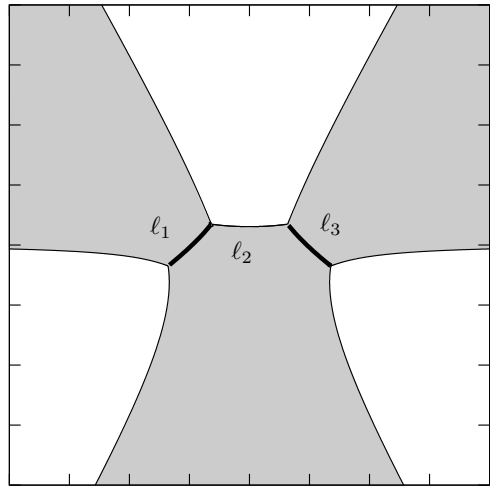

Figure 18: An example of reconstruction of admissible triple of unusual topology (numerical output).
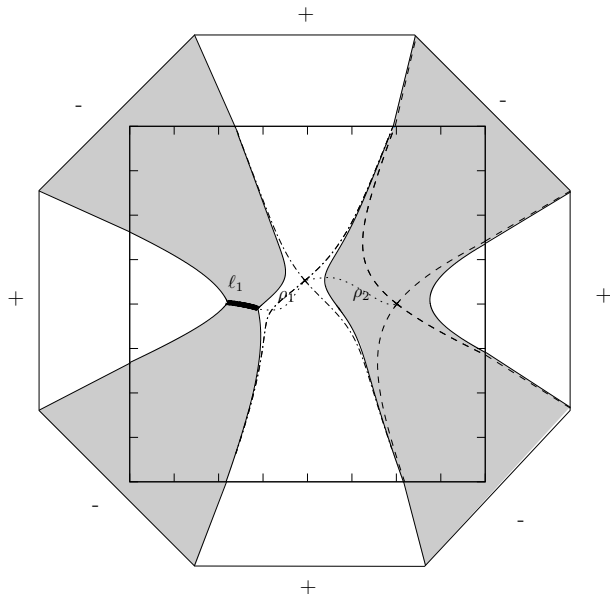
Figure 19: An example of admissible triple where the Stokes sectors are not all connectible by admissible paths (the part within the graduated square is a numerical output). Here $\Im(\rho_2) > \Im(\rho_1)$.

should have no difficulty in imagining and drawing on paper
even very complicated decorated clock-diagrams that correspond to a situation as in Def. 4.2.

6 Conclusion

The first two parts (Sec. 3 and 4) have shown that any admissible triple (Def. 4.2) is associated to the asymptotics with respect to varying weights of certain pseudo–orthogonal polynomials. Some concluding remarks that follow from the construction are in order;

- there might be different orthogonal polynomials that have the same asymptotic. This can happen if the potential is the same (necessarily) and the asymptotic Boutroux admissible curve has two Stokes sectors that belong to the same connected component of $\{h(x) > 0\}$. Indeed in this case the jump on the contour joining them and remaining in the positive-h region will be exponentially close (and uniformly) to the identity in the large N limit, thus becoming irrelevant.

- If two Stokes sectors cannot be joined by a curve in $\mathcal{Y}_+ \cup \mathcal{B}$ (which happens for example in Fig 19) then the corresponding finite-n RHP for the orthogonal polynomials cannot have a jump on a contour joining them.

- In the asymptotic regime the change of Stokes' parameters (the $\varkappa$'s) is **isospectral**; indeed it simply corresponds to choice of a different Stokes-Kirchoff's differential η, hence of a different twisting of

the line–bundle of the s. It is easy (but we don't do it here for brevity) to write explicitly the conjugating matrix for the spectral –problem associated to the spinors (eq. 3-55)

The last part (Sec. 5) has shown that there can be admissible triples whose branch-cut structure $\mathcal{B}$ can be basically as complicated as one may wish and it is (topologically) a forest of loop-free trivalent trees; the zeroes of the pseudo–orthogonal polynomials then accumulate on $\mathcal{B}$ and we have (not completely rigorously) specified the asymptotic density of zeroes along these arcs in Sec. 4.6.

6.1 Outlook

The definition of admissibility (Def. 4.2) is forced upon us because the jump matrices (Stokes' matrices) in the Riemann–Hilbert problem for orthogonal polynomials are all **upper triangular**. In a more general setting, e.g. as in the study of Painlevé II equations in the asymptotic regime [19], there can be also lower triangular Stokes matrices: in this case we could relax the notion of admissibility. The key fact that drives the steepest-descent method is that the upper-lower triangular form of the jump matrix should be related to the signs of h on the two sides of the branch-cut (i.e. if $h < 0$ on both sides then the jump should be upper triangular, if $h > 0$ then the jump should be lower triangular).

This means that one could use any Boutroux curve to model the asymptotics of *some* RHP by choosing the jump-matrices of the appropriate form according to the signs of h. This seems a promising avenue of research that we intend to pursue in a different publication.

Another extension of the present setting that we set out to explore is the inclusion of hard-edges and more general potentials with rational derivative; while the heuristic approach remains unchanged (and in fact the first part of this paper is already developed to the full extent of this generality) several changes need to be made in the steepest descent part and in the analysis of critical trajectories.
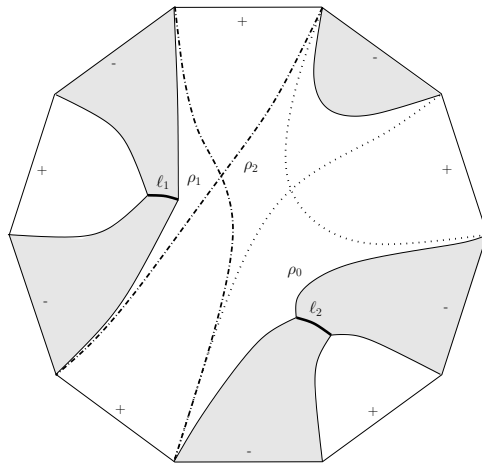

Figure 20: An example of reconstruction of admissible triple of unusual topology with two double points: the constraint is that $\Im(\rho_1) = \Im(\rho_2) + \Im(\rho_0)$. The shaded area is the where $h < 0$ (the "underwater"). The dashed curves are at positive equal height h (the dash-dot one is higher). Not a numerical output.

It seems that the introduction of hard-edges is the simplest generalization; this requires the use of a different local parametrix near each hard-edge (built out of Bessel functions) and the modification of the construction of admissible curves (some ingredients for this generalization can be found in [28]). All this will be dealt with in a forthcoming publication.

References

[1] M. Bertola, "Boutroux curves with external field: equilibrium measures without a minimization problem", arXiv:0705.3062, submitted.

[2] M. Bertola, B. Eynard, J. Harnad, "Partition functions for matrix models and isomonodromic tau functions", Random matrix theory. J. Phys. A **36** (2003), no. 12, 3067–3083.

[3] M. Bertola, B. Eynard, J. Harnad, "Semiclassical orthogonal polynomials, matrix models and isomonodromic tau functions", nlin.SI/0410043, Comm. Math. Phys. to appear.

[4] M. Bertola, M. Gekhtman, "Biorthogonal Laurent polynomials, Toeplitz determinants, minimal Toda orbits and isomonodromic tau functions", nlin.SI/0503050.

[5] M. Bertola, M. Y. Mo, "Isomonodromic deformation of resonant rational connections", IMRP, no.11, (2005), 565-635.

[6] P. Bleher, A. R. Its, "Semiclassical asymptotics of orthogonal polynomials, Riemann-Hilbert problem, and universality in the matrix model", Ann. of Math. (2) **150** (1999), no. 1, 185–266.

[7] P. Bleher, A. R. Its, "On asymptotic analysis of orthogonal polynomials via the Riemann-Hilbert method", Symmetries and integrability of difference equations (Canterbury, 1996), 165–177 London Math. Soc. Lecture Note Ser., **255**.

[8] V. Buslaev, L. Pastur, "A class of the multi–interval eigenvalue distributions of matrix models and related structures", in *Asymptotic combinatorics with application to mathematical physics* (St. Petersburg, 2001), NATO Sci. Ser. II Math. Phys. Chem., **77**, 51–70.

[9] I. V. Cherednik, "Differential equations for the Baker-Akhiezer functions of algebraic curves", Funct. Anal. Appl. **12** (1978), no. 3, 195–203 (1979).

[10] I. V. Cherednik, "On the reality conditions in the "finite gap integration" ", Dokl. Acad. Nauk. USSR, **252** (1980), no. 5, 1104–1108.

[11] P. A. Deift, "Orthogonal polynomials and random matrices: a Riemann-Hilbert approach", Courant lecture notes in mathematics, 3. New York University, Courant Institute of Mathematical Sciences, New York;American Mathematical Society, Providence, RI, 1999.

[12] P. A. Deift, A. Its, A. Kapaev, X. Zhou, "On the algebro–geometric integration of the Schlesinger equations", Commun. Math. Phys., **203** (1999), 613–633.

[13] P. A. Deift, X. Zhou, "A Riemann–Hilbert approach to asymptotic problems arising in the theory of random matrix models and also in the theory of integrable statistical mechanics", Ann. of Math. (2) **146** (1997), no.1, 149–235.

[14] P. A. Deift, T. Kriecherbauer, K. T. McLaughlin, S. Venakides, X. Zhou. "Strong asymptotics of orthogonal polynomials with respect to exponential weights", Comm. Pure Appl. Math. **52** (1999), no. 12, 1491–1552.

[15] P. A. Deift, T. Kriecherbauer, K. T. McLaughlin, S. Venakides, X. Zhou. "Uniform asymptotics for polynomials orthogonal with respect to varying exponential weights and applications to universality questions in random matrix theory", Comm. Pure Appl. Math. **52** (1999), no. 11, 1335–1425.

[16] H. M. Farkas, I. Kra, "Riemann surfaces", volume 71 of *Graduate Texts in Mathematics*. Springer–Verlag, New York, second edition, 1992.

[17] J. Fay, "Theta Functions on Riemann Surfaces", Lect. Notes in Math. **352**, Springer Verlag, (1973).

[18] A. Fokas, A. Its, A. Kitaev, "The isomonodromy approach to matrix models in 2D quantum gravity", *Commun. Math. Phys.* **147**, 395–430 (1992).

[19] A. R. Its, A. A. Kapaev. "The nonlinear steepest descent approach to the asymptotics of the second Painlevè transcendent in the complex domain", MathPhys odyssey, 2001, 273–311.

[20] A. R. Its, V. B. Matveev, "Algebro-geometric integration of the modified nonlinear Schrödinger equation, finite zone solutions and their degeneration" (Russian) Questions in quantum field theory and statistical physics, **2** Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI) 101 (1981), 64–76, 206.

[21] A. R. Its, V. Y. Novokshenov "The isomonodromic deformation method in the theory of Painlevé equations", Lecture Notes in Mathematics, 1191. Springer-Verlag, Berlin, 1986. iv+313 pp. ISBN: 3-540-16483-9

[22] J. A. Jenkins, D. C. Spencer, "Hyperelliptic Trajectories", Annals of Math., **53** Vol. 53, No. 1 (1951), 4–35.

[23] G. Jensen, "Quadratic differentials", Chapter 8 in C. Pommerenke, "Univalent functions", Göttingen : Vandenhoeck und Ruprecht, 1975

[24] A. A. Kapaev, "Monodromy approach to scaling limits in isomonodromic systems", 16th International Conference on Nonlinear Evolution Equations and Dynamical Systems (Russian), (also available in English at nlin.SI/0211022). Teoret. Mat. Fiz. **137** (2003), no. 3, 393–407.

[25] D. Korotkin, "Solution of matrix Riemann-Hilbert problems with quasi-permutation monodromy matrices", Math. Ann. **329** (2004), no. 2, 335–364.

[26] I. M. Krichever, S. P. Novikov, "A two-dimensionalized Toda chain, commuting difference operators, and holomorphic vector bundles", (Russian) Uspekhi Mat. Nauk **58** (2003), no. 3(351), 51–88; translation in Russian Math. Surveys **58** (2003), no. 3, 473–510

[27] A. B. J. Kuijlaars, A. Martinez-Finkelshtein, "Strong asymptotics for Jacobi polynomials with varying nonstandard parameters" J. Anal. Math. **94** (2004), 195–234.

[28] A. B. J. Kuijlaars, K. T-R McLaughlin, "Asymptotic Zero Behavior of Laguerre Polynomials with Negative Parameter", Constr. Approx. **20** (2004), 497–523.

[29] F. Marcellán, I. A. Rocha, "Complex Path Integral Representation for Semiclassical Linear Functionals", J. Appr. Theory **94**, 107–127, (1998).

[30] G. Moore "Matrix models of 2D gravity and isomonodromic deformation" Common trends in mathematics and quantum field theories (Kyoto, 1990). Progr. Theoret. Phys. Suppl. No. 102, (1990), 255–285 (1991).

[31] G. Moore "Geometry of the string equations" Comm. Math. Phys. **133** (1990), no. 2, 261–304.

[32] K. Strebel, "Quadratic Differentials", Modern Surveys in Mathematics, Springer–Verlag (1984).

[33] A. Tovbis, S. Venakides, X. Zhou. "On semiclassical (zero dispersion limit) solutions of the focusing nonlinear Schrödinger equation", Comm. Pure Appl. Math. **57** (2004), no. 7, 877–985.