



Cauchy biorthogonal polynomials

M. Bertola^{a,b,*}, M. Gekhtman^c, J. Szmiagielski^d

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

^b *Department of Mathematics and Statistics, Concordia University, 1455 de Maisonneuve W., Montréal, Québec, Canada H3G 1M8*

^c *Department of Mathematics 255 Hurley Hall, Notre Dame, IN 46556-4618, USA*

^d *Department of Mathematics and Statistics, University of Saskatchewan, 106 Wiggins Road, Saskatoon, Saskatchewan, S7N 5E6, Canada*

Received 17 April 2009; received in revised form 21 September 2009; accepted 29 September 2009
Available online 18 October 2009

Communicated by Guillermo Lopez Lagomasino

Abstract

The paper investigates the properties of certain biorthogonal polynomials appearing in a specific simultaneous Hermite–Padé approximation scheme. Associated with any totally positive kernel and a pair of positive measures on the positive axis we define biorthogonal polynomials and prove that their zeros are simple and positive. We then specialize the kernel to the Cauchy kernel $\frac{1}{x+y}$ and show that the ensuing biorthogonal polynomials solve a four-term recurrence relation, have relevant Christoffel–Darboux generalized formulas, and their zeros are interlaced. In addition, these polynomials solve a combination of Hermite–Padé approximation problems to a Nikishin system of order 2. The motivation arises from two distant areas; on the one hand, in the study of the inverse spectral problem for the peakon solution of the Degasperis–Procesi equation; on the other hand, from a random matrix model involving two positive definite random Hermitian matrices. Finally, we show how to characterize these polynomials in terms of a Riemann–Hilbert problem.

© 2009 Elsevier Inc. All rights reserved.

Keywords: Biorthogonal polynomials; Total positivity; Christoffel–Darboux identities; Riemann–Hilbert problems; Multiple Padé approximation

* Corresponding author at: Centre de recherches mathématiques, Université de Montréal, C. P. 6128, succ. centre ville, Montréal, Québec, Canada H3C 3J7.

E-mail addresses: bertola@crm.umontreal.ca (M. Bertola), Michael.Gekhtman.1@nd.edu (M. Gekhtman), szmiagiel@math.usask.ca (J. Szmiagielski).

1. Introduction and motivations

This paper mainly deals with a class of biorthogonal polynomials $\{p_n(x)\}_{\mathbb{N}}, \{q_n(y)\}_{\mathbb{N}}$ of degree n satisfying the biorthogonality relations

$$\int_{\mathbb{R}_+} \int_{\mathbb{R}_+} p_n(x)q_m(y) \frac{d\alpha(x)d\beta(y)}{x+y} = \delta_{mn}, \tag{1.1}$$

where $d\alpha, d\beta$ are positive measures supported on \mathbb{R}_+ with finite bimoments. These polynomials will be introduced in Section 2 in a more general context of polynomials associated with general *totally positive kernels* (Definition 2.1) with which they share some general properties in regard to their zeros.

While these properties are interesting in their own right, we wish to put the work in a more general context and explain the two main motivations behind it. They fall within two different and rather distant areas of mathematics : peakon solutions to nonlinear PDEs and Random Matrix theory.

Peakons for the Degasperis–Procesi equation. In the early 1990’s, Camassa and Holm [13] introduced the (CH) equation to model (weakly) dispersive shallow wave propagation. More generally, the CH equation belongs to the so-called b -family of PDEs

$$u_t - u_{xxt} + (b + 1)uu_x = bu_xu_{xx} + uu_{xxx}, \quad (x, t) \in \mathbb{R}^2, b \in \mathbb{R}, \tag{1.2}$$

Two cases, $b = 2$ and $b = 3$ within this family are now known to be integrable: the case $b = 2$ is the original CH equation whereas the case $b = 3$ is the Degasperis–Procesi [16] (DP) equation, which is more directly related to the present paper.

In all cases the b -family admits weak (distributional) solutions of the form:

$$u(x, t) = \sum_{i=1}^n m_i(t) e^{-|x-x_i(t)|}, \tag{1.3}$$

if and only if the positions $x_i(t)$ and the heights $m_i(t)$ satisfy the system of nonlinear ODEs:

$$\dot{x}_k = \sum_{i=1}^n m_i e^{-|x_k-x_i|}, \quad \dot{m}_k = (b - 1) \sum_{i=1}^n m_k m_i \operatorname{sgn}(x_k - x_i) e^{-|x_k-x_i|}, \tag{1.4}$$

for $k = 1, \dots, n$. The non-smooth character of the solution manifests itself by the presence of sharp peaks at $\{x_k\}$, hence the name *peakons*. For the CH equation the peakons solution were studied in [4,3], while for the DP equation in [25,26]; in both cases the solution is related to the **isospectral evolution** of an associated linear boundary value problem

$b = 2$ (CH)	$b = 3$ (DP)
$-\phi''(\xi, z) = zg(\xi)\phi(\xi, z)$	$-\phi'''(\xi, z) = zg(\xi)\phi(\xi, z)$
$\phi(-1) = \phi(1) = 0$	$\phi(-1) = \phi'(-1) = \phi(1) = 0$

(1.5)

The variables x, ξ and the quantities m, g, u are related by

$$\begin{aligned} \xi &= \tanh\left(\frac{x}{b-1}\right), & g(\xi) &= \left(\frac{1-\xi^2}{2}\right)^{-b} m(x), \\ m(x, t) &= u(x, t) - u_{xx}(x, t). \end{aligned} \tag{1.6}$$

Because of the similarity to the equation of an inhomogeneous classical string (after a separation of variables) we refer to the two linear ODEs as the *quadratic* and *cubic* string, respectively. The case of peakons corresponds to the choice

$$m(x, t) = 2 \sum_{j=1}^n \delta(x - x_j(t)) m_j(t). \tag{1.7}$$

The remarkable fact is that in both cases the associated spectral problems have a finite **positive** spectrum; this is not so surprising in the case of the quadratic string which is a self-adjoint problem, but it is quite unexpected for the cubic string, since the problem is not self-adjoint and there is no *a priori* reason for the spectrum to even be real [26].

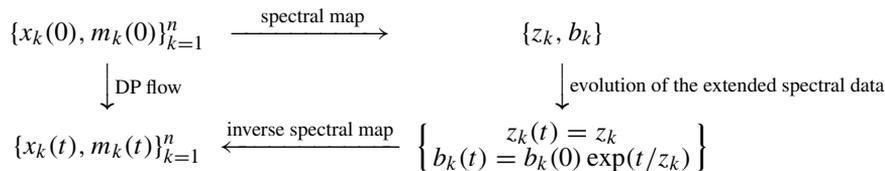
As it is natural within the Lax approach to integrable PDEs, the *spectral map* linearizes the evolution of the isospectral evolution: if $\{z_j\}$ are the eigenvalues of the respective boundary value problems and one introduces the appropriate spectral residues

$$b_j := \operatorname{res}_{z=z_j} \frac{W(z)}{z} dz, \quad W(z) := \frac{\phi_\xi(1, z)}{\phi(1, z)} \tag{1.8}$$

then one can show [25] that the evolution linearizes as follows (with the dot representing the time evolution)

$$\dot{z}_k = 0, \quad \frac{\dot{b}_k}{b_k} = \frac{1}{z_k}. \tag{1.9}$$

Since this is not the main focus of the paper, we are deliberately glossing over several interesting points; the interested reader is referred to [26] and our recent work [10] for further details. In short, the solution method for the DP equation can be illustrated by the diagram



In the inverse spectral map resides the role of the biorthogonal polynomials to be studied here, as we briefly sketch below. The inverse problem for the ordinary string with finitely many point masses is solved by the method of continued fractions of Stieltjes’ type as was pointed out by Krein [21]. The inverse problem for the cubic string with finitely many masses is solved with the help of the following simultaneous Hermite–Padé type approximation [26].

Definition 1.1 (*Padé-like Approximation Problem*). Let $d\mu(x)$ denote the spectral measure associated with the cubic string boundary value problem and $\frac{W(z)}{z} = \int \frac{1}{z-x} d\mu(x)$, $\frac{Z}{z} = \int \int \frac{x}{z-x} \frac{1}{x+y} d\mu(x) d\mu(y)$ denote the Weyl functions introduced in [26]. Then, given an integer

$1 \leq k \leq n$, we seek three polynomials (Q, P, \widehat{P}) of degree $k - 1$ satisfying the following conditions:

[Approximation]: $W = \frac{P}{Q} + O\left(\frac{1}{z^{k-1}}\right), Z = \frac{\widehat{P}}{Q} + O\left(\frac{1}{z^{k-1}}\right) (z \rightarrow \infty).$

[Symmetry]: $Z^* Q + W^* P + \widehat{P} = O\left(\frac{1}{z^k}\right) (z \rightarrow \infty)$ with $W^*(z) = -W(-z), Z^*(z) = Z(-z).$

[Normalization]: $P(0) = 1, \widehat{P}(0) = 0.$

This approximation problem has a unique solution [26] which, in turn, is used to solve the inverse problem for the cubic string. We point out that it is here in this approximation problem that the Cauchy kernel $\frac{1}{x+y}$ makes its, somewhat unexpected, appearance through the spectral representation of the second Weyl function.

Random matrix theory. The other source of our interest in biorthogonal polynomials comes from random matrix theory. It is well known [27] that the Hermitian matrix model is intimately related to (in fact, solved by) orthogonal polynomials (OPs). Not so much is known about the role of biorthogonal polynomials (BOPs). However, certain biorthogonal polynomials somewhat similar to the ones in the present paper appear prominently in the analysis of “the” two-matrix model after reduction to the spectrum of eigenvalues [7,9,8,18]; in that case the pairing is of the form

$$\iint p_n(x)q_m(y)e^{-xy}d\alpha(x)d\beta(y) = \delta_{mn}, \tag{1.10}$$

and the associated biorthogonal polynomials are sometimes called the Itzykson–Zuber BOPs, in short, the IZBOPs.

Several algebraic structural properties of these polynomials and their recurrence relation (both multiplicative and differential) have been thoroughly analyzed in the previously cited papers for densities of the form $d\alpha(x) = e^{-V_1(x)}dx, d\beta(y) = e^{-V_2(y)}dy$ for *polynomials potentials* $V_1(x), V_2(y)$ and for potentials with rational derivative (and hard-edges) in [5].

We recall that while ordinary OPs satisfy a multiplicative three-term recurrence relation, the BOPs defined by (1.10) solve a longer recurrence relation of length related to the degree of the differential $dV_j(x)$ over the Riemann sphere [5]; a direct (although not immediate) consequence of the finiteness of the recurrence relation is the fact that these BOPs (and certain integral transforms of them) are characterized by a Riemann–Hilbert problem for a matrix of size equal to the length of the recurrence relation (minus one). The BOPs introduced in this paper share all these features, although in some respects they are closer to the ordinary orthogonal polynomials than to the IZBOPs.

The relevant two-matrix model our polynomials are related to was introduced in [12]. We now give a brief summary of that work. Consider the set of pairs $\mathcal{H}_+^{(2)} := \{(M_1, M_2)\}$ of Hermitian *positive definite* matrices endowed with the $(U(N)$ -invariant) Lebesgue measure denoted by dM_1dM_2 . Define then the probability measure on this space by the formula:

$$d\mu(M_1, M_2) = \frac{1}{\mathcal{Z}_N^{(2)}} \frac{\alpha'(M_1)\beta'(M_2)dM_1dM_2}{\det(M_1 + M_2)^N} \tag{1.11}$$

where $\mathcal{Z}_N^{(2)}$ (the *partition function*) is a normalization constant, while $\alpha'(M_1), \beta'(M_2)$ stand for the product of the densities α', β' (the Radon–Nikodym derivatives of the measures $d\alpha, d\beta$ with respect to the Lebesgue measure) over the (positive) eigenvalues of M_j .

This probability space is similar to the two-matrix model discussed briefly above for which the coupling between matrices is $e^{N\text{Tr}M_1M_2}$ [19] instead of $\det(M_1 + M_2)^{-N}$. The connection with our BOPs (1.1) is analogous to the connection between ordinary orthogonal polynomials and the Hermitian Random matrix model [27], whose probability space is the set of Hermitian matrices \mathcal{H}_N equipped with the measure $d\mu_1(M) := \frac{1}{Z_N^{(1)}}\alpha'(M)dM$. In particular, we show in [12] how the statistics of the eigenvalues of the two matrices M_j can be described in terms of the biorthogonal polynomials we are introducing in the present work. A prominent role in the description of that statistics is played by the generalized Christoffel–Darboux identities we develop in Section 4.

We now summarize the main results of the paper:

- for an arbitrary totally positive kernel $K(x, y)$ and arbitrary positive measures $d\alpha, d\beta$ on \mathbb{R}_+^2 we prove that the matrix of bimoments $I_{ab} := \iint_{\mathbb{R}_+^2} x^a y^b K(x, y) d\alpha(x) d\beta(y)$ is totally positive (Theorem 2.1);
- this implies that there exist, unique, sequences of monic polynomials of degree n , $\tilde{p}_n(x), \tilde{q}_n(y)$ biorthogonal to each other as in (2.1); we prove that they have **positive and simple zeros** (Theorem 2.5);
- we then specialize to the kernel $K(x, y) = \frac{1}{x+y}$; in this case the **zeros** of $\tilde{p}_n(x)$ ($\tilde{q}_n(y)$) **are interlaced** with the zeros of the neighboring polynomials (Theorem 3.2);
- they solve a **four-term** recurrence relation as specified after (1.1) (Corollary 4.2);
- they satisfy **Christoffel–Darboux identities** (Proposition 4.3, Corollary 4.3, Theorems 5.3 and 5.5)
- they solve a **Hermite–Padé** approximation problem to a novel type of Nikishin systems (Section 5, Theorems 5.1 and 5.2);
- they can be characterized by a 3×3 **Riemann–Hilbert problems**, (Propositions 6.1 and 6.2);

In the follow-up paper we will explain the relation of the asymptotics of the BOPs introduced in this paper with a rigorous asymptotic analysis for continuous (varying) measures $d\alpha, d\beta$ using the nonlinear steepest descent method [11].

We should point out that polynomials solutions to higher order recurrence relations in the context of the inverse spectral method for nonlinear dynamical systems were also studied in [1], with applications to Bogoyavlensky lattices. Moreover, other examples of biorthogonal (or multiply orthogonal) polynomials related to various matrix models have been investigated in [24,5,7,8]. The first asymptotic results in this connection started to appear very recently [17].

2. Biorthogonal polynomials associated with a totally positive kernel

As one can see from the last section the kernel $K(x, y) = \frac{1}{x+y}$, $x, y > 0$, which we will refer to as the Cauchy kernel, plays a significant, albeit mysterious, role. We now turn to explaining the role of this kernel. We recall, following [23], the definition of the totally positive kernel.

Definition 2.1. A real function $K(x, y)$ of two variables ranging over linearly ordered sets \mathcal{X} and \mathcal{Y} , respectively, is said to be totally positive (TP) if for all

$$x_1 < x_2 < \cdots < x_m, \quad y_1 < y_2 < \cdots < y_m \quad x_i \in \mathcal{X}, y_j \in \mathcal{Y}, m \in \mathbb{N} \quad (2.1)$$

we have

$$\det [K(x_i, y_j)]_{1 \leq i, j \leq m} > 0. \quad (2.2)$$

We will also use a discrete version of the same concept.

Definition 2.2. A matrix $A := [a_{ij}]$, $i, j = 0, 1, \dots, n$ is said to be totally positive (TP) if all its minors are strictly positive. A matrix $A := [a_{ij}]$, $i, j = 0, 1, \dots, n$ is said to be totally nonnegative (TN) if all its minors are nonnegative. A TN matrix A is said to be oscillatory if some positive integer power of A is TP.

Since we will be working with matrices of infinite size we introduce a concept of the principal truncation.

Definition 2.3. A finite $n + 1$ by $n + 1$ matrix $B := [b_{i,j}]$, $i, j = 0, 1, \dots, n$ is said to be the principal truncation of an infinite matrix $A := [a_{ij}]$, $i, j = 0, 1, \dots$ if $b_{i,j} = a_{i,j}$, $i, j = 0, 1, \dots, n$. In such a case B will be denoted $A[n]$.

Finally,

Definition 2.4. An infinite matrix $A := [a_{ij}]$, $i, j = 0, 1, \dots$ is said to be TP (TN) if $A[n]$ is TP (TN) for every $n = 0, 1, \dots$

Definition 2.5 (Basic Setup). Let $K(x, y)$ be a **totally positive kernel** on $\mathbb{R}_+ \times \mathbb{R}_+$ and let $d\alpha, d\beta$ be two Stieltjes measures on \mathbb{R}_+ . We make two simplifying assumptions to avoid degenerate cases:

1. 0 is not an atom of either of the measures (i.e. $\{0\}$ has zero measure).
2. α and β have infinitely many points of increase.

We furthermore assume:

3. the polynomials are dense in the corresponding Hilbert spaces $H_\alpha := L^2(\mathbb{R}_+, d\alpha)$, $H_\beta := L^2(\mathbb{R}_+, d\beta)$,
4. the map $K : H_\beta \rightarrow H_\alpha$, $Kq(x) := \int K(x, y)q(y)d\beta(y)$ is bounded, injective and has a dense range in H_α .

Under these assumptions K provides a non-degenerate pairing between H_β and H_α :

$$\langle a|b \rangle = \iint a(x)b(y)K(x, y)d\alpha d\beta, \quad a \in H_\alpha, b \in H_\beta. \tag{2.3}$$

Remark 2.1. The points 3 and 4 above could be weakened, especially the density assumption, but we believe the last two assumptions are the most natural to work with in the Hilbert space setup of the theory.

Now, let us consider the matrix \mathcal{I} of generalized bimoments

$$[\mathcal{I}]_{ij} = I_{ij} := \iint x^i y^j K(x, y)d\alpha(x)d\beta(y). \tag{2.4}$$

Theorem 2.1. *The semi-infinite matrix \mathcal{I} is TP.*

Proof. According to a theorem of Fekete, (see Chapter 2, Theorem 3.3 3.3 in [23]), we only need to consider minors of consecutive rows/columns. Writing out the determinant,

$$\Delta_n^{ab} := \det[I_{a+i,b+j}]_{0 \leq i,j \leq n-1}$$

we find

$$\begin{aligned} \Delta_n^{ab} &= \sum_{\sigma \in S_n} \epsilon(\sigma) \iint \prod_{j=1}^n x_j^a y_j^b \prod_{j=1}^n x_j^{\sigma_j-1} y_j^{j-1} K(x_j, y_j) d^n \alpha(X) d^n \beta(Y) \\ &= \iint C(X)^a C(Y)^b \Delta(X) \prod_{j=1}^n y_j^{j-1} \prod_{j=1}^n K(x_j, y_j) d^n \alpha d^n \beta. \\ C(X) &:= \prod_{j=1}^N x_j, \quad \Delta(X) := \prod_{i < j} (x_i - x_j) = \det[x_i^{j-1}]_{1 \leq i, j \leq n} \end{aligned}$$

where S_n stands for the symmetric group on n elements, and the integrals are understood as n -fold integrals over \mathbb{R}_+ . Since our intervals are subsets of \mathbb{R}_+ we can absorb the powers of $C(X)$, $C(Y)$ into the measures to simplify the notation. Moreover, the function $S(X, Y) := \prod_{j=1}^n K(x_j, y_j)$ enjoys the following simple property

$$S(X, Y_\sigma) = S(X_{\sigma^{-1}}, Y)$$

for any $\sigma \in S_n$. Finally, the product measures $d^n \alpha = d^n \alpha(X)$, $d^n \beta = d^n \beta(Y)$ are clearly permutation invariant.

Thus, without any loss of generality, we only need to show that

$$D_n := \iint \Delta(X) \prod_{j=1}^n y_j^{j-1} S(X, Y) d^n \alpha d^n \beta > 0,$$

which is tantamount to showing positivity for $a = b = 0$. First, we symmetrize D_n with respect to the variables X ; this produces

$$\begin{aligned} D_n &= \frac{1}{n!} \sum_{\sigma \in S_n} \iint \Delta(X_\sigma) \prod_{j=1}^n y_j^{j-1} S(X_\sigma, Y) d^n \alpha d^n \beta \\ &= \frac{1}{n!} \sum_{\sigma \in S_n} \iint \Delta(X) \epsilon(\sigma) \prod_{j=1}^n y_j^{j-1} S(X, Y_{\sigma^{-1}}) d^n \alpha d^n \beta \\ &= \frac{1}{n!} \sum_{\sigma \in S_n} \iint \Delta(X) \epsilon(\sigma) \prod_{j=1}^n y_{\sigma_j}^{j-1} S(X, Y) d^n \alpha d^n \beta \\ &= \frac{1}{n!} \iint \Delta(X) \Delta(Y) S(X, Y) d^n \alpha d^n \beta. \end{aligned}$$

Subsequent symmetrization over the Y variables does not change the value of the integral and we obtain (after restoring the definition of $S(X, Y)$)

$$\begin{aligned} D_n &= \frac{1}{(n!)^2} \sum_{\sigma \in S_n} \epsilon(\sigma) \iint \Delta(X) \Delta(Y) \prod_{j=1}^n K(x_j, y_{\sigma_j}) d^n \alpha d^n \beta \\ &= \frac{1}{(n!)^2} \iint \Delta(X) \Delta(Y) \det[K(x_i, y_j)]_{i, j \leq n} d^n \alpha d^n \beta. \end{aligned}$$

Finally, since $\Delta(X) \Delta(Y) \det[K(x_i, y_j)]_{i, j \leq n} d^n \alpha d^n \beta$ is permutation invariant, it suffices to integrate over the region $0 < x_1 < x_2 < \dots < x_n \times 0 < y_1 < y_2 < \dots < y_n$, and, as a result

$$D_n = \iint_{\substack{0 < x_1 < x_2 < \dots < x_n \\ 0 < y_1 < y_2 < \dots < y_n}} \Delta(X) \Delta(Y) \det[K(x_i, y_j)]_{i, j \leq n} d^n \alpha d^n \beta. \tag{2.5}$$

Due to the total positivity of the kernel $K(x, y)$ the integrand is a positive function of all variables and so the integral must be strictly positive. \square

To simplify future computations we define $[x] := (1, x, x^2, \dots)^T$ so that the matrix of generalized bimoments (2.4) is simply given by:

$$\mathcal{I} = ([x][y]^T). \tag{2.6}$$

Now, let Λ denote the semi-infinite upper shift matrix. Then we observe that multiplying the measure $d\alpha(x)$ by x^i or, multiplying $d\beta(y)$ by y^j , is tantamount to multiplying \mathcal{I} on the left by Λ^i , or on the right by $(\Lambda^T)^j$ respectively, which gives us a whole family of bimoment matrices associated with the same $K(x, y)$ but different measures. Thus we have

Corollary 2.1. *For any nonnegative integers i, j the matrix of generalized bimoments $\Lambda^i \mathcal{I} (\Lambda^T)^j$ is TP.*

We conclude this section with a few comments about the scope of [Theorem 2.1](#).

Remark 2.2. Provided that the negative moments are well defined, the theorem then applies to the doubly infinite matrix $I_{i,j}, i, j \in \mathbb{Z}$.

Remark 2.3. If the intervals are \mathbb{R} and $K(x, y) = e^{xy}$ then the proof above fails because we cannot re-define the measures by multiplying by powers of the variables, since they become then signed measures, so in general the matrix of bimoments is **not** totally positive. Nevertheless the proof above shows (with $a = b = 0$ or $a, b \in 2\mathbb{N}$) that the matrix of bimoments is positive definite and – in particular – the biorthogonal polynomials always exist, which is known and proved in [18].

2.1. Biorthogonal polynomials

Due to the total positivity of the matrix of bimoments in our setting, there exist uniquely defined two sequences of monic polynomials

$$\tilde{p}_n(x) = x^n + \dots, \quad \tilde{q}_n(y) = y^n + \dots$$

such that

$$\iint \tilde{p}_n(x) \tilde{q}_m(y) K(x, y) d\alpha(x) d\beta(y) = h_n \delta_{mn}.$$

Standard considerations (Cramer’s Rule) show that they are provided by the following formulas

$$\tilde{p}_n(x) = \frac{1}{D_n} \det \left[\begin{array}{ccc|c} I_{00} & \dots & I_{0n-1} & 1 \\ \vdots & & \vdots & \vdots \\ I_{n0} & \dots & I_{nn-1} & x^n \end{array} \right]$$

$$\tilde{q}_n(y) = \frac{1}{D_n} \det \left[\begin{array}{ccc} I_{00} & \dots & I_{0n} \\ \vdots & & \vdots \\ I_{n-10} & \dots & I_{n-1n} \\ \hline 1 & \dots & y^n \end{array} \right] \tag{2.7}$$

$$h_n = \frac{D_{n+1}}{D_n} > 0, \quad D_n := \Delta_n^{00} = \det[I_{ij}]_{0 \leq i, j \leq n-1} \tag{2.8}$$

where $D_j > 0$ by Eq. (2.5). For convenience we re-define the sequence in such a way that they are also **normalized** (instead of monic), by dividing them by the square root of h_n ;

$$p_n(x) = \frac{1}{\sqrt{D_n D_{n+1}}} \det \left[\begin{array}{ccc|c} I_{00} & \dots & I_{0n-1} & 1 \\ \vdots & & \vdots & \vdots \\ I_{n0} & \dots & I_{nn-1} & x^n \end{array} \right], \tag{2.9}$$

$$q_n(y) = \frac{1}{\sqrt{D_n D_{n+1}}} \det \left[\begin{array}{ccc} I_{00} & \dots & I_{0n} \\ \vdots & & \vdots \\ I_{n-10} & \dots & I_{n-1n} \\ \hline 1 & \dots & y^n \end{array} \right]. \tag{2.10}$$

Thus $\langle p_n | q_m \rangle = \delta_{nm}$.

We note also that the BOPs can be obtained by triangular transformations of $[x], [y]$

$$\mathbf{p}(x) = S_p[x], \quad \mathbf{q}(y) = S_q[y], \quad [x] = [1, x, x^2, \dots]^t \tag{2.11}$$

where $S_{p,q}$ are (formally) invertible lower triangular matrices such that $S_p^{-1}(S_q^{-1})^T = \mathcal{I}$, where, we recall, \mathcal{I} is the generalized bimoment matrix. Moreover, our BOPs satisfy, by construction, the recursion relations:

$$x p_i(x) = X_{i,i+1} p_{i+1}(x) + X_{i,i} p_i(x) + \dots + X_{i,0} p_0(x),$$

$$y q_i(y) = Y_{i,i+1} q_{i+1}(y) + Y_{i,i} q_i(y) + \dots + Y_{i,0} q_0(y),$$

which will be abbreviated as

$$x \mathbf{p}(x) = \mathbf{X} \mathbf{p}(x), \quad y \mathbf{q}(y)^T = \mathbf{q}(y)^T \mathbf{Y}^T, \tag{2.12}$$

where \mathbf{X} and \mathbf{Y} are Hessenberg matrices with positive entries on the supradiagonal, and $\mathbf{p}(x) \mathbf{q}(y)$ are infinite column vectors $\mathbf{p}(x)^T := (p_0(x), p_1(x), p_2(x), \dots)^t$, $\mathbf{q}(y)^T := (q_0(y), q_1(y), q_2(y), \dots)^t$ respectively.

The biorthogonality can now be written as $\langle \mathbf{p} | \mathbf{q}^T \rangle = Id$ where Id denotes the semi-infinite identity matrix. Moreover

$$\langle x \mathbf{p} | \mathbf{q}^T \rangle = \mathbf{X}, \quad \langle \mathbf{p} | y \mathbf{q}^T \rangle = \mathbf{Y}^T. \tag{2.13}$$

Remark 2.4. The significance of the last two formulas lies in the fact that the operator of multiplication is no longer symmetric with respect to the pairing $\langle \bullet | \bullet \rangle$ and as a result the matrices \mathbf{X} and \mathbf{Y}^T are distinct.

2.2. Simplicity of the zeros

In this section we will use the concept of a Chebyshev system of order n and a closely related concept of a Markov sequence. We refer to [28,29,21] for more information. The following theorem is a convenient restatement of Lemma 2 in [21], p.137. For easy display we replace determinants with wedge products.

Theorem 2.2. Given a system of continuous functions $\{u_i(x)|i = 0, \dots, n\}$ let us define the vector field

$$\mathbf{u}(x) = [u_0(x), u_1(x), \dots, u_n(x)]^T, \quad x \in U \subset \mathbb{R}. \tag{2.14}$$

Then $\{u_i(x)|i = 0, \dots, n\}$ is a Chebyshev system of order n on U iff the top exterior power

$$\mathbf{u}(x_0) \wedge \mathbf{u}(x_1) \wedge \dots \wedge \mathbf{u}(x_n) \neq 0 \tag{2.15}$$

for all $x_0 < x_1 < \dots < x_n$ in U . More generally for a sequence of functions $\{u_i(x)|i = 0, \dots, n\}$, if we denote the truncation of $\mathbf{u}(x) := [u_0(x), u_1(x), \dots]$ to the first $n + 1$ components by $\mathbf{u}_n(x)$, then $\{u_i(x)|i = 0 \dots n\}$ is a Markov system iff the top exterior power

$$\mathbf{u}_n(x_0) \wedge \mathbf{u}_n(x_1) \wedge \dots \wedge \mathbf{u}_n(x_n) \neq 0 \tag{2.16}$$

for all $x_0 < x_1 < \dots < x_n$ in U and all $n \in \mathbb{N}$.

The following well known theorem is now immediate.

Theorem 2.3. Suppose $\{u_i(x)|i = 0, \dots, n\}$ is a Chebyshev system of order n on U , and suppose we are given n distinct points x_1, \dots, x_n in U . Then, up to a multiplicative factor, the only generalized polynomial $P(x) = \sum_{i=0}^n a_i u_i(x)$, which vanishes precisely at x_1, \dots, x_n in U is given by

$$P(x) = \mathbf{u}(x) \wedge \mathbf{u}(x_1) \wedge \dots \wedge \mathbf{u}(x_n). \tag{2.17}$$

Theorem 2.4. Denote by $u_i(x) = \int K(x, y)y^i d\beta(y)$, $i = 0, \dots, n$. Then $\{u_i(x)|i = 0, \dots, n\}$ is a Chebyshev system of order n on \mathbb{R}_+ . Moreover, $P(x)$ as defined in Theorem 2.3 changes sign each time x passes through any of the zeros x_j .

Proof. It is instructive to look at the computation. Let $x_0 < x_1 < \dots < x_n$, then using multilinearity of the exterior product,

$$\begin{aligned} P(x_0) &= \mathbf{u}(x_0) \wedge \mathbf{u}(x_1) \wedge \dots \wedge \mathbf{u}(x_n) \\ &= \int K(x_0, y_0)K(x_1, y_1) \dots K(x_n, y_n)[y_0]_n \wedge [y_1]_n \wedge \dots \wedge [y_n]_n d\beta(y_0) \dots d\beta(y_n) \\ &= \frac{1}{n!} \int \det[K(x_i, y_j)]_{i,j=0}^n \Delta(Y) d\beta(y_0) \dots d\beta(y_n) \\ &= \int_{y_0 < y_1 < \dots < y_n} \det[K(x_i, y_j)]_{i,j=0}^n \Delta(Y) d\beta(y_0) \dots d\beta(y_n), \end{aligned}$$

where $[y]_n = [y^0, y^1, \dots, y^n]^T$. Thus $P(x_0) > 0$. The rest of the proof is the argument about the sign of the integrand. To see how the sign changes we observe that the sign of P depends only on the ordering of x, x_1, x_2, \dots, x_n , in view of the total positivity of the kernel. In other words, the sign of P is $\text{sgn}(\pi)$ where π is the permutation rearranging x, x_1, x_2, \dots, x_n in an increasing sequence. \square

Corollary 2.2. Let $\{f_i(x) := \int K(x, y)q_i(y)d\beta(y), |i = 0, \dots, n\}$. Then $\{f_i(x)|i = 0, \dots, n\}$ is a Markov sequence on \mathbb{R}_+ .

Proof. Indeed, [Theorem 2.2](#) implies that the group $GL(n + 1)$ acts on the set of Chebyshev systems of order n . It suffices now to observe that q_j are obtained from $\{1, y, \dots, y^n\}$ by an invertible transformation. \square

Remark 2.5. Observe that $\{f_i(x)|i = 0, \dots, n\}$ is a Markov sequence regardless of biorthogonality.

Biorthogonality enters however in the main theorem.

Theorem 2.5. *The zeros of p_n, q_n are all simple and positive. They fall within the convex hull of the support of the measure $d\alpha$ (for p_n 's) and $d\beta$ (for the q_n 's).*

Proof. We give first a proof for p_n . The theorem is trivial for $n = 0$. For $1 \leq n$, let us suppose p_n has $r < n$ zeros of odd order in the convex full of $\text{supp}(d\alpha)$. In full analogy with the classical case, $1 \leq r$, since

$$\int p_n(x)f_0(x)d\alpha(x) = \iint p_n(x)K(x, y)d\alpha(x)d\beta(y) = 0$$

by biorthogonality, forcing, in view of positivity of $K(x, y), p_n(x)$ to change sign in the convex hull of $\text{supp}(d\alpha)$. In the general case, denote the zeros by $x_1 < x_2 < \dots < x_r$. Using a Chebyshev system $f_i(x), i = 0, \dots, r$ on \mathbb{R}_+ we can construct a unique, up to a multiplicative constant, generalized polynomial which vanishes exactly at those points, namely

$$R(x) = F(x) \wedge F(x_1) \wedge F(x_2) \wedge \dots \wedge F(x_r) \tag{2.18}$$

where

$$F(x) = [f_0(x) \quad f_1(x) \quad \dots \quad f_r(x)]^t, \quad x \in \mathbb{R}.$$

It follows then directly from biorthogonality that

$$\int p_n(x)F(x) \wedge F(x_1) \wedge F(x_2) \wedge \dots \wedge F(x_r)d\alpha(x) = 0. \tag{2.19}$$

On the other hand, $R(x)$ is proportional to $P(x)$ in [Theorem 2.3](#) which, by [Theorem 2.4](#), changes sign at each of its zeros, so the product $p_n(x)R(x)$ is nonzero and of fixed sign over $\mathbb{R}_+ \setminus \{x_1, x_2, \dots, x_r\}$. Consequently, the integral is nonzero, since α is assumed to have infinitely many points of increase. Thus, in view of the contradiction, $r \geq n$, hence $r = n$, for p_n is a polynomial of degree n . The case of q_n follows by observing that the adjoint K^* is also a TP kernel and hence it suffices to switch α with β throughout the argument given above. \square

Lemma 2.1. *In the notation of [Corollary 2.2](#) $f_n(x)$ has n zeros and n sign changes in the convex hull of $\text{supp}(d\alpha)$.*

Proof. Clearly, since $\{u_i(x)|i = 0, \dots, n\}$ is a Chebyshev system of order n on \mathbb{R}_+ , the number of zeros of f_n cannot be greater than n . Again, from

$$\int f_n(x)p_0(x)d\alpha(x) = 0,$$

we conclude that f_n changes sign at least once within the convex hull of $\text{supp}(d\alpha)$. Let then $x_1 < x_2 < \dots < x_r, 1 \leq r \leq n$ be all zeros of f_n within the convex hull of $\text{supp}(d\alpha)$ at which

f_n changes its sign. Thus, on one hand,

$$\int \epsilon \prod_{i=1}^r (x - x_i) f_n(x) d\alpha(x) > 0, \quad \epsilon = \pm,$$

while, on the other hand, using biorthogonality we get

$$\int \epsilon \prod_{i=1}^r (x - x_i) f_n(x) d\alpha(x) = 0, \quad \epsilon = \pm,$$

which shows that $r = n$. \square

In view of [Theorem 2.3](#) the statement about the zeros of f_n has the following corollary.

Corollary 2.3 (*Heine-like Representation for f_n*).

$$f_n(x) = C u(x) \wedge u(x_1) \wedge u(x_2) \cdots \wedge u(x_n) \tag{2.20}$$

where x_j are the zeros of f_n and C is a constant.

3. Cauchy BOPs

From now on we restrict our attention to the particular case of the totally positive kernel, namely, the Cauchy kernel

$$K(x, y) = \frac{1}{x + y} \tag{3.1}$$

whose associated biorthogonal polynomials will be called Cauchy BOPs. Thus, from this point onward, we will be studying the general properties of BOPs for the pairing

$$\iint p_n(x) q_m(y) \frac{d\alpha(x) d\beta(y)}{x + y} = \langle p_n | q_m \rangle. \tag{3.2}$$

Until further notice, we do not assume anything about the relationship between the two measures $d\alpha, d\beta$, other than what is in the basic setup of [Definition 2.5](#).

3.1. Rank one shift condition

It follows immediately from [Eq. \(3.1\)](#) that

$$I_{i+1,j} + I_{i,j+1} = \langle x^{i+1} | y^j \rangle + \langle x^i | y^{j+1} \rangle = \int x^i d\alpha \int y^j d\beta, \tag{3.3}$$

which, with the help of the shift matrix A and the matrix of bimoments \mathcal{I} , can be written as:

$$\begin{aligned} A\mathcal{I} + \mathcal{I}A^T &= \alpha\beta^T, \\ \alpha &= (\alpha_0, \alpha_1, \dots)^T, & \alpha_j &= \int x^j d\alpha(x) > 0, \\ \beta &= (\beta_0, \beta_1, \dots)^T, & \beta_j &= \int y^j d\beta(y) > 0. \end{aligned}$$

Moreover, by linearity and Eq. (2.13), we have

$$\mathbf{X} + \mathbf{Y}^T = \boldsymbol{\pi} \boldsymbol{\eta}^T, \quad \boldsymbol{\pi} := \int \mathbf{p} d\alpha, \quad \boldsymbol{\eta} := \int \mathbf{q} d\beta,$$

$$\mathbf{p}(x) := (p_0(x), p_1(x), \dots)^t, \quad \mathbf{q}(y) := (q_0(y), q_1(y), \dots)^t \tag{3.4}$$

which connects the multiplication operators in H_α and H_β . Before we elaborate on the nature of this connection we need to clarify one aspect of Eq. (3.4).

Remark 3.1. One needs to exercise a great deal of caution using the matrix relation given by Eq. (3.4). Its only rigorous meaning is in action on vectors with finitely many nonzero entries or, equivalently, this equation holds for all principal truncations.

Proposition 3.1. *The vectors $\boldsymbol{\pi}, \boldsymbol{\eta}$ are strictly positive (have nonvanishing positive coefficients).*

Proof. We prove the assertion only for $\boldsymbol{\pi}$, the one for $\boldsymbol{\eta}$ being obtained by interchanging the roles of $d\alpha$ and $d\beta$.

From the expressions (2.10) for $p_n(x)$ we immediately have

$$\pi_n = \sqrt{\frac{1}{D_n D_{n+1}}} \det \left[\begin{array}{ccc|c} I_{00} & \dots & I_{0n-1} & \alpha_0 \\ \vdots & & \vdots & \vdots \\ I_{n0} & \dots & I_{nn-1} & \alpha_n \end{array} \right]. \tag{3.5}$$

Since we know that $D_n > 0$ for any $n \geq 0$ we need to prove the positivity of the other determinant. Determinants of this type were studied in Lemma 4.10 in [26].

We nevertheless give a complete proof of positivity. First, we observe that

$$\begin{aligned} \pi_n \sqrt{D_{n+1} D_n} &= \sum_{\sigma \in S_{n+1}} \epsilon(\sigma) \int \prod_{j=1}^{n+1} x_j^{\sigma_j-1} \prod_{j=1}^n y_j^{j-1} \frac{d^{n+1} \alpha d^n \beta}{\prod_{j=1}^n (x_j + y_j)} \\ &= \int \Delta(X_1^{n+1}) \prod_{j=1}^n y_j^{j-1} \frac{d^{n+1} \alpha d^n \beta}{\prod_{j=1}^n (x_j + y_j)}. \end{aligned} \tag{3.6}$$

Here the symbol X_1^{n+1} is to remind that the vector consists of $n + 1$ entries (whereas Y consists of n entries) and that the Vandermonde determinant is taken accordingly. Note also that the variable x_{n+1} never appears in the product in the denominator. Symmetrizing the integral in the x_j 's with respect to labels $j = 1, \dots, n$, but leaving x_{n+1} fixed, gives

$$\pi_n \sqrt{D_{n+1} D_n} = \frac{1}{n!} \int \Delta(X_1^{n+1}) \Delta(Y) \frac{d^{n+1} \alpha d^n \beta}{\prod_{j=1}^n (x_j + y_j)}. \tag{3.7}$$

Symmetrizing now with respect to the whole set x_1, \dots, x_{n+1} we obtain

$$\pi_n \sqrt{D_{n+1} D_n} = \frac{1}{n!(n+1)!} \int \Delta(X_1^{n+1}) \Delta(Y) \times \det \begin{bmatrix} K(x_1, y_1) & \dots & K(x_{n+1}, y_1) \\ \vdots & & \vdots \\ K(x_1, y_n) & \dots & K(x_{n+1}, y_n) \\ 1 & \dots & 1 \end{bmatrix} d^{n+1} \alpha d^n \beta. \tag{3.8}$$

Moreover, since the integrand is permutation invariant, it suffices to integrate over the region $0 < x_1 < x_2 < \dots < x_n < x_{n+1} \times 0 < y_1 < y_2 < \dots < y_n$, and, as a result

$$\pi_n \sqrt{D_{n+1} D_n} = \iint_{0 < x_1 < x_2 < \dots < x_{n+1} \times 0 < y_1 < y_2 < \dots < y_n} \Delta(X_1^{n+1}) \Delta(Y) \times \det \begin{bmatrix} K(x_1, y_1) & \dots & K(x_{n+1}, y_1) \\ \vdots & & \vdots \\ K(x_1, y_n) & \dots & K(x_{n+1}, y_n) \\ 1 & \dots & 1 \end{bmatrix} d^{n+1} \alpha d^n \beta. \tag{3.9}$$

We thus need to prove that the determinant containing the Cauchy kernel $\frac{1}{x+y}$ is positive for $0 < x_1 < x_2 < \dots < x_{n+1}$ and $0 < y_1 < y_2 < \dots < y_n$. It is not difficult to prove that

$$\det \begin{bmatrix} \frac{1}{x_1 + y_1} & \dots & \frac{1}{x_{n+1} + y_1} \\ \vdots & & \vdots \\ \frac{1}{x_1 + y_n} & \dots & \frac{1}{x_{n+1} + y_n} \\ 1 & \dots & 1 \end{bmatrix} = \frac{\Delta(X_1^{n+1}) \Delta(Y)}{\prod_{j=1}^{n+1} \prod_{k=1}^n (x_j + y_k)} \tag{3.10}$$

and this function is clearly positive in the above range. \square

3.2. Interlacing properties of the zeros

From (2.6) and (2.11) and (2.12) the following factorizations are valid for all principal truncations:

$$\mathcal{I} = S_p^{-1} (S_q^{-1})^T, \quad \mathbf{X} = S_p \Lambda (S_p)^{-1}, \quad \mathbf{Y} = S_q \Lambda S_q^{-1}.$$

Moreover, since \mathcal{I} is TP, the triangular matrices S_p^{-1} and S_q^{-1} are totally nonnegative (TN) [15] and have the same diagonal entries: the n th diagonal entry being $\sqrt{D_n/D_{n-1}}$. Furthermore, one can amplify the statement about S_p^{-1} and S_q^{-1} using another result of Cryer [14] which implies that both triangular matrices are in fact triangular TP matrices (all nontrivial in the sense defined in [14] minors are strictly positive). This has the immediate consequence.

Lemma 3.1. *All principal truncations $\mathbf{X}[n], \mathbf{Y}[n]$ are invertible.*

Proof. From the factorization $\mathbf{X} = S_p \Lambda (S_p)^{-1}$ we conclude that it suffices to prove the claim for $\Lambda S_p^{-1}[n]$ which in matrix form reads:

$$\begin{bmatrix} (S_p^{-1})_{10} & (S_p^{-1})_{11} & & & \\ (S_p^{-1})_{20} & (S_p^{-1})_{21} & (S_p^{-1})_{22} & & \\ & & & \ddots & \\ \vdots & \vdots & & & (S_p^{-1})_{n+1,n+1} \\ (S_p^{-1})_{n+1,0} & (S_p^{-1})_{n+1,1} & \dots & & (S_p^{-1})_{n+1,n} \end{bmatrix} \mathbf{0}.$$

However, the determinant of this matrix is strictly positive, because S_p^{-1} is a triangular TP. \square

Remark 3.2. This lemma is not automatic, since $\Lambda[n]$ is not invertible.

We now state the main theorem of this section.

Theorem 3.1. \mathbf{X} and \mathbf{Y} are TN.

Proof. We need to prove the theorem for every principal truncation. Let $n \geq 0$ be fixed. We will suppress the dependence on n , for example \mathbf{X} in the body of the proof means $\mathbf{X}[n]$ etc. First, we claim that \mathbf{X} and \mathbf{Y} admit the L-U factorization: $\mathbf{X} = \mathbf{X}_- \mathbf{X}_+$, $\mathbf{Y} = \mathbf{Y}_- \mathbf{Y}_+$, where A_+ denotes the upper triangular factor and A_- is the unipotent lower triangular factor in the Gauss factorization of a matrix A . Indeed, $\mathbf{X}_+ = (\Lambda S_p^{-1})_+$, $\mathbf{Y}_+ = (\Lambda S_q^{-1})_+$ are upper triangular components of TN matrices ΛS_p^{-1} and ΛS_q^{-1} and thus are totally nonnegative invertible bi-diagonal matrices by Lemma 3.1.

From $\mathbf{X} + \mathbf{Y}^T = \boldsymbol{\pi} \boldsymbol{\eta}^T$ we then obtain

$$(\mathbf{Y}_+^T)^{-1} \mathbf{X}_- + \mathbf{Y}_- \mathbf{X}_+^{-1} = \left((\mathbf{Y}_+^T)^{-1} \boldsymbol{\pi} \right) \left(\boldsymbol{\eta}^T \mathbf{X}_+^{-1} \right) := \boldsymbol{\rho} \boldsymbol{\mu}^T.$$

We need to show that vectors $\boldsymbol{\rho}$, $\boldsymbol{\mu}$ have positive entries. For this, notice that

$$\begin{aligned} \boldsymbol{\rho} &= ((\mathbf{Y}_+)^T)^{-1} S_p \boldsymbol{\alpha} = (((\Lambda S_q^{-1})_+)^T)^{-1} S_p \boldsymbol{\alpha}, \\ \boldsymbol{\mu} &= ((\mathbf{X}_+)^T)^{-1} S_q \boldsymbol{\beta} = (((\Lambda S_p^{-1})_+)^T)^{-1} S_q \boldsymbol{\beta}. \end{aligned}$$

Now, it is easy to check that if the matrix of generalized bimoments \mathcal{I} is replaced by $\mathcal{I} \Lambda^T$ (see Corollary 2.1) then $S_p \rightarrow (((\Lambda S_q^{-1})_+)^T)^{-1} S_p$, while $\boldsymbol{\alpha}$ is unchanged, which implies that $\boldsymbol{\rho}$ is a new $\boldsymbol{\pi}$ in the notation of Proposition 3.1 and hence positive by the same Proposition. Likewise, considering the matrix of generalized bimoments $\Lambda \mathcal{I}$, for which $\boldsymbol{\beta}$ is unchanged, $S_q \rightarrow (((\Lambda S_p^{-1})_+)^T)^{-1} S_q$ and $\boldsymbol{\mu}$ is a new $\boldsymbol{\eta}$ in the notation of Proposition 3.1 implying the claim.

Thus

$$\boldsymbol{\rho} = D_\rho \mathbf{1}, \quad \boldsymbol{\mu} = D_\mu \mathbf{1},$$

where D_ρ, D_μ are diagonal matrices with positive entries and $\mathbf{1}$ is a vector of 1s.

We have

$$D_\rho^{-1} (\mathbf{Y}_+^T)^{-1} \mathbf{X}_- D_\mu^{-1} + D_\rho^{-1} \mathbf{Y}_- \mathbf{X}_+^{-1} D_\mu^{-1} = \mathbf{1} \mathbf{1}^T.$$

The first (resp. second) term on the left that we can call $\tilde{\mathbf{X}}$ (resp. $\tilde{\mathbf{Y}}^T$) is a lower (resp. upper) triangular matrix with positive diagonal entries. The equality above then implies that

(i) $\tilde{X}_{ij} = \tilde{Y}_{ij} = 1$ for all $i > j$ and (ii) $\tilde{X}_{ii} + \tilde{Y}_{ii} = 1$ for all i . In particular, both \tilde{X}_{ii} and \tilde{Y}_{ii} are positive numbers strictly less than 1.

This means that $\tilde{\mathbf{X}}, \tilde{\mathbf{Y}}$ admit factorizations

$$\tilde{\mathbf{X}} = (Id - \Lambda^T)^{-1}L_X, \quad \tilde{\mathbf{Y}} = (Id - \Lambda^T)^{-1}L_Y,$$

where

$$L_X = \sum_{i=0}^{\infty} \tilde{X}_{ii}E_{ii} + (1 - \tilde{X}_{ii})E_{i+1 i}, \quad L_Y = \sum_{i=0}^{\infty} \tilde{Y}_{ii}E_{ii} + (1 - \tilde{Y}_{ii})E_{i+1 i}.$$

Since all entries of bi-diagonal matrices L_X, L_Y are positive, these matrices are totally nonnegative and so are

$$\mathbf{X} = \mathbf{Y}_+^T (Id - \Lambda^T)^{-1} L_X \mathbf{X}_+, \quad \mathbf{Y} = \mathbf{X}_+^T (Id - \Lambda^T)^{-1} L_Y \mathbf{Y}_+. \quad \square \tag{3.11}$$

Corollary 3.1. *\mathbf{X} and \mathbf{Y} are oscillatory matrices.*

Proof. We give a proof for \mathbf{X} . The factorization (3.11) we have just obtained shows that \mathbf{X} is the product of an invertible lower triangular TN matrix $\mathbf{Y}_+^T (Id - \Lambda^T)^{-1}$ and a tri-diagonal matrix $J = L_X \mathbf{X}_+$. Note that L_X has all positive values on the main diagonal and the first sub-diagonal. Entries on the first super-diagonal of \mathbf{X}_+ coincide with corresponding entries of \mathbf{X} and thus are strictly positive by construction. Moreover, leading principal minors of \mathbf{X} are strictly positive (see the proof of Lemma 3.1), which implies that all diagonal entries of \mathbf{X}_+ are strictly positive too. Thus J is a tri-diagonal matrix with all nontrivial entries strictly positive.

Since diagonal entries of $\mathbf{Y}_+^T (Id - \Lambda^T)^{-1}$ are strictly positive and all other entries are nonnegative, every zero entry of \mathbf{X} implies that the corresponding entry of J is zero. In view of that all entries on the first super- and sub-diagonals of \mathbf{X} must be strictly positive, which, by a fundamental criterion of Gantmacher and Krein (Theorem 10, II, [21]), ensures that \mathbf{X} is oscillatory. \square

The interlacing properties for the zeros of polynomials p_n, q_n , as well as other properties of Sturm sequences, follow then from Gantmacher–Krein theorems on spectral properties of oscillatory matrices (see II, Theorem 13, in [21]). We summarize the most important properties implied by Gantmacher–Krein theory.

Theorem 3.2. *The sequences of BOPs $\{q_n\}$ and $\{p_n\}$ are Sturm sequences. Moreover,*

1. *their respective zeros are positive and simple,*
2. *the roots of adjacent polynomials in the sequences are interlaced,*
3. *the following alternative representations of the biorthogonal polynomials hold*

$$p_n(x) = \sqrt{\frac{D_n}{D_{n+1}}} \det(x - X[n - 1]), \quad 1 \leq n,$$

$$q_n(y) = \sqrt{\frac{D_n}{D_{n+1}}} \det(y - Y[n - 1]), \quad 1 \leq n.$$

Remark 3.3. The fact that the roots are positive and simple follows indeed from the fact that X and Y are oscillatory. Theorem 2.5, however, indicates that this property is true even for a more general case when the totally positive kernel $K(x, y)$ is not necessarily the Cauchy kernel.

It should be mentioned that analogous results about the interlacing properties of roots of special classes of polynomials associated with Hermite–Padé approximations of Nikishin systems have appeared recently in [20,2].

4. Four-term recurrence relations and Christoffel–Darboux identities

We establish in this section a basic form of recurrence relations and an analog of classical Christoffel–Darboux identities satisfied by $\{q_n\}$ and $\{p_n\}$. First, we introduce the following notation for semi-infinite, finite-band matrices.

Definition 4.1. Given two integers $a \leq b$, a semi-infinite matrix A is said to have the support in $[a, b]$ if

$$j - i < a \quad \text{or} \quad j - i > b \quad \text{imply} \quad A_{ij} = 0 \tag{4.1}$$

The set of all matrices with supports in $[a, b]$ is denoted $M_{[a,b]}$.

The content of this section relies heavily on the relation (3.4) which we recall for convenience:

$$\mathbf{X} + \mathbf{Y}^T = \pi \boldsymbol{\eta}^T = D_\pi \mathbf{1}^T D_\eta$$

where D_π, D_η respectively, are diagonal matrices of averages of \mathbf{p} and \mathbf{q} . Since the vector $\mathbf{1}$ is a null vector of $\Lambda - Id$ we obtain.

Proposition 4.1. \mathbf{X} and \mathbf{Y} satisfy:

1. $(\Lambda - Id)D_\pi^{-1}\mathbf{X} + (\Lambda - Id)D_\pi^{-1}\mathbf{Y}^T = 0.$
2. $A := (\Lambda - Id)D_\pi^{-1}\mathbf{X} \in M_{[-1,2]}.$
3. $\mathbf{X}D_\eta^{-1}(\Lambda^T - Id) + \mathbf{Y}^T D_\eta^{-1}(\Lambda^T - Id) = 0.$
4. $\widehat{A} := \mathbf{X}D_\eta^{-1}(\Lambda^T - Id) \in M_{[-2,1]}.$

As an immediate corollary we obtain the factorization property for X and Y .

Corollary 4.1. Let A, \widehat{A} and

$$L := (\Lambda - Id)D_\pi^{-1}, \quad \widehat{L} := D_\eta^{-1}(\Lambda^T - Id),$$

respectively, denote matrices occurring in Proposition 4.1. Then

$$L\mathbf{X} = A, \quad \mathbf{X}\widehat{L} = \widehat{A}, \quad A \in M_{[-1,2]}, \quad \widehat{A} \in M_{[-2,1]}.$$

Likewise, \mathbf{Y} admits a similar factorization:

$$\mathbf{Y}L^T = B, \quad (\widehat{L}^T)\mathbf{Y} = \widehat{B},$$

where $B = -A^T, \widehat{B} = -\widehat{A}^T.$

Hence,

Corollary 4.2. \mathbf{p} and \mathbf{q} satisfy four-term recurrence relations of the form

$$x \left(\frac{p_n(x)}{\pi_n} - \frac{p_{n-1}(x)}{\pi_{n-1}} \right) = A_{n-1,n+1}p_{n+1}(x) + A_{n-1,n}p_n(x) \\ + A_{n-1,n-1}p_{n-1}(x) + A_{n-1,n-2}p_{n-2}(x),$$

$$y \left(\frac{q_n(y)}{\eta_n} - \frac{q_{n-1}(y)}{\eta_{n-1}} \right) = \widehat{B}_{n-1,n+1}q_{n+1}(y) + \widehat{B}_{n-1,n}q_n(y) + \widehat{B}_{n-1,n-1}q_{n-1}(y) + \widehat{B}_{n-1,n-2}q_{n-2}(y),$$

for $1 \leq n$ with the proviso that $p_{-1} = q_{-1} = 0$.

Proof. We give the proof for $\mathbf{p}(x)$ in matrix form. Indeed, from

$$x\mathbf{p}(x) = \mathbf{X}\mathbf{p}(x),$$

it follows that

$$xL\mathbf{p}(x) = L\mathbf{X}\mathbf{p}(x),$$

hence the claim, since $L \in M_{[0,1]}$ and $L\mathbf{X} = A \in M_{[-1,2]}$. \square

Let us observe that \widehat{L} has a unique formal inverse, represented by a lower triangular matrix. Let us then define

$$\widehat{\mathbf{p}}(x) = \widehat{L}^{-1}\mathbf{p}(x).$$

Theorem 4.1 (Christoffel–Darboux Identities for \mathbf{q} and \mathbf{p}).

$$(x + y) \sum_{j=0}^{n-1} q_j(y)p_j(x) = \mathbf{q}^T(y)[\Pi, (y - \mathbf{Y}^T)\widehat{L}]\widehat{\mathbf{p}}(x) \tag{4.2}$$

where $\Pi := \Pi_n$ is the diagonal matrix $\text{diag}(1, 1, \dots, 1, 0, \dots)$ with n ones (the entries are labeled from 0 to $n - 1$). The explicit form of the commutators is:

$$[\Pi, (y - \mathbf{Y}^T)\widehat{L}] = \widehat{A}_{n-1,n}E_{n-1,n} - \left(\frac{y}{\eta_n} + \widehat{A}_{n,n-1} \right) E_{n,n-1} - \widehat{A}_{n,n-2}E_{n,n-2} - \widehat{A}_{n+1,n-1}E_{n+1,n-1}, \tag{4.3}$$

where $A_{i,j}, \widehat{A}_{i,j}$ respectively, denote the (i, j) th entries of A, \widehat{A} , occurring in Proposition 4.1.

Proof. We give the proof of Eq. (4.2). Since $(y - \mathbf{Y})\mathbf{q} = 0$ it suffices to prove that the left hand side equals $\mathbf{q}^T \Pi (y - \mathbf{Y}^T)\widehat{L}\widehat{\mathbf{p}}(x)$. From the definition of $\widehat{\mathbf{p}}$ and Eq. (2.12) we obtain

$$(x + y)\mathbf{q}^T(y)\Pi\mathbf{p}(x) = \mathbf{q}^T(y)\Pi y\widehat{L}\widehat{\mathbf{p}}(x) + \mathbf{q}^T(y)\Pi\mathbf{X}\mathbf{p}(x) = \mathbf{q}^T(y)\Pi y\widehat{L}\widehat{\mathbf{p}}(x) + \mathbf{q}^T(y)\Pi\mathbf{X}\widehat{L}\widehat{\mathbf{p}}(x),$$

which, after switching $\mathbf{X}\widehat{L}$ with $-\mathbf{Y}^T\widehat{L}$ in view of Proposition 4.1, gives Eq. (4.2). To get the commutator equation (4.3) one needs to perform an elementary computation using the definition of \widehat{A} . \square

We establish now basic properties of $\widehat{\mathbf{p}}$ and its biorthogonal partner $\widehat{\mathbf{q}}$ defined below.

Proposition 4.2. The sequences of polynomials

$$\widehat{\mathbf{p}} = \widehat{L}^{-1}\mathbf{p}, \quad \widehat{\mathbf{q}}^T = \mathbf{q}^T\widehat{L} \tag{4.4}$$

are characterized by the following properties

1. $\deg \widehat{q}_n = n + 1, \deg \widehat{p}_n = n;$
2. $\int \widehat{q}_n d\beta = 0;$

- 3. $\iint \widehat{p}_n(x)\widehat{q}_m(y) \frac{d\alpha d\beta}{x+y} = \delta_{mn}$;
- 4. $\widehat{q}_n(y) = \frac{1}{\eta_{n+1}} \sqrt{\frac{D_{n+1}}{D_{n+2}}} y^{n+1} + \mathcal{O}(y^n)$.

In addition

- a. $\widehat{\mathbf{q}}$ and $\widehat{\mathbf{p}}$ satisfy the intertwining relations with \mathbf{q} and \mathbf{p}

$$\begin{aligned} y\widehat{\mathbf{q}}^T &= -\mathbf{q}^T \widehat{A}, \\ x\mathbf{p} &= \widehat{A}\widehat{\mathbf{p}}; \end{aligned} \tag{4.5}$$

- b. $\widehat{\mathbf{q}}$ and $\widehat{\mathbf{p}}$ admit the determinantal representations:

$$\widehat{q}_n(y) = \frac{1}{\eta_n \eta_{n+1} \sqrt{D_n D_{n+2}}} \det \begin{bmatrix} I_{00} & \dots & I_{0n+1} \\ \vdots & & \vdots \\ I_{n-10} & \dots & I_{n-1n+1} \\ \beta_0 & \dots & \beta_{n+1} \\ 1 & \dots & y^{n+1} \end{bmatrix} \tag{4.6}$$

$$\widehat{p}_n(x) = \frac{1}{D_{n+1}} \det \begin{bmatrix} I_{00} & \dots & I_{0n} & 1 \\ \vdots & & \vdots & \vdots \\ I_{n-10} & \dots & I_{n-1n} & x^{n-1} \\ I_{n0} & \dots & I_{nn} & x^n \\ \beta_0 & \dots & \beta_n & 0 \end{bmatrix} \tag{4.7}$$

- c. $\beta_0 \iint \widehat{p}_n(x)y^j \frac{d\alpha d\beta}{x+y} = \beta_j \iint \widehat{p}_n(x) \frac{d\alpha d\beta}{x+y}$, $j \leq n$.

Proof. Assertions (1), (2) and (4) follow directly from the shape of the matrix \widehat{L} . Assertion (3) follows from $(\mathbf{p}, \mathbf{q}^t) = \mathbf{1}$ by multiplying it by \widehat{L} on the right and by \widehat{L}^{-1} on the left. Assertion (c) follows from assertions (1), (2) and (3); indeed from (2) and (3), it follows that the polynomial \widehat{p}_n is biorthogonal to all polynomials of degree $\leq n$ with zero $d\beta$ -average and $\{\beta_0 y^j - \beta_j : 0 \leq j \leq n\}$ is a basis for such polynomials.

The intertwining relations follow from the definitions of the matrices \widehat{L} , \widehat{A} and of the polynomials $\widehat{\mathbf{p}}$, $\widehat{\mathbf{q}}$.

The determinantal expression for \widehat{q}_n follows by inspection since the proposed expression has the defining properties (1) and (2) and is biorthogonal to all powers $1, x, \dots, x^{n-1}$. The normalization is found by comparing the leading coefficients of $\widehat{q}_n = \frac{1}{\eta_{n+1}} q_{n+1} + \mathcal{O}(y^n)$. The determinantal expression for $\widehat{p}_n(x)$ follows again by inspection; indeed if $F(x)$ is the determinant in (4.7) then

$$\langle F(x) | y^j \rangle = \det \begin{bmatrix} I_{00} & \dots & I_{0n} & I_{0j} \\ \vdots & & \vdots & \vdots \\ I_{n-10} & \dots & I_{n-1n} & I_{n-1j} \\ I_{n0} & \dots & I_{nn} & I_{nj} \\ \beta_0 & \dots & \beta_n & 0 \end{bmatrix} = -\beta_j D_{n+1} = \frac{\beta_j}{\beta_0} \langle F(x) | 1 \rangle \tag{4.8}$$

where the determinants are computed by expansion along the last row. The proportionality constant is again found by comparison. \square

One easily establishes a counterpart to Theorem 4.1 valid for $\widehat{\mathbf{q}}$ and $\widehat{\mathbf{p}}$.

Proposition 4.3 (Christoffel–Darboux Identities for $\widehat{\mathbf{q}}$ and $\widehat{\mathbf{p}}$). We have

$$(x + y) \sum_{j=0}^{n-1} \widehat{q}_j(y) \widehat{p}_j(x) = \mathbf{q}^T(y) [(x - X)\widehat{L}, \Pi] \widehat{\mathbf{p}}(x) = \mathbf{q}^T(y) [\Pi, (-x - Y^T)\widehat{L}] \widehat{\mathbf{p}}(x). \tag{4.9}$$

Remark 4.1. Observe that the commutators occurring in both theorems have identical structure; they only differ in the variable y in **Theorem 4.1** being now replaced by $-x$. We will denote by $\mathbb{A}(x)$ the commutator $[\Pi, (-x - \mathbf{Y}^T)\widehat{L}]$ and by $\mathbb{A}_n(x)$ its nontrivial 3×3 block. Thus the nontrivial block in **Proposition 4.3** reads:

$$\mathbb{A}_n(x) = \left[\begin{array}{cc|c} 0 & 0 & \widehat{A}_{n-1,n} \\ \hline -\widehat{A}_{n,n-2} & \frac{x}{\eta_n} - \widehat{A}_{n,n-1} & 0 \\ 0 & -\widehat{A}_{n+1,n-1} & 0 \end{array} \right] \tag{4.10}$$

while the block appearing in **Theorem 4.1** is simply $\mathbb{A}_n(-y)$.

With this notation in place we can present the Christoffel–Darboux identities in a unified way.

Corollary 4.3 (Christoffel–Darboux Identities for \mathbf{q} , \mathbf{p} , and $\widehat{\mathbf{q}}$, $\widehat{\mathbf{p}}$). The biorthogonal polynomials \mathbf{q} , \mathbf{p} , and $\widehat{\mathbf{q}}$, $\widehat{\mathbf{p}}$ satisfy

$$(x + y) \sum_{j=0}^{n-1} q_j(y) p_j(x) = \mathbf{q}^T(y) \mathbb{A}(-y) \widehat{\mathbf{p}}(x), \tag{4.11}$$

$$(x + y) \sum_{j=0}^{n-1} \widehat{q}_j(y) \widehat{p}_j(x) = \mathbf{q}^T(y) \mathbb{A}(x) \widehat{\mathbf{p}}(x). \tag{4.12}$$

5. Approximation problems and perfect duality

We will associate a chain of Markov functions associated with measures $d\alpha$ and $d\beta$ by taking the Stieltjes’ transforms of the corresponding measures as well as their reflected, with respect to the origin, images.

Definition 5.1. Define

$$\begin{aligned} W_\beta(z) &= \int \frac{1}{z - y} d\beta(y), & W_{\alpha^*}(z) &= \int \frac{1}{z + x} d\alpha(x), \\ W_{\alpha^*\beta}(z) &= - \iint \frac{1}{(z + x)(x + y)} d\alpha(x) d\beta(y), \\ W_{\beta\alpha^*}(z) &= \iint \frac{1}{(z - y)(y + x)} d\alpha(x) d\beta(y). \end{aligned} \tag{5.1}$$

We recall now an important notion of a Nikishin system associated with two measures (see [28], p. 142, called there a MT system of order 2).

Definition 5.2. Given two measures $d\mu_1$ and $d\mu_2$ with disjoint supports Δ_1, Δ_2 respectively, a Nikishin system of order 2 is a pair of functions

$$f_1(z) = \int_{\Delta_1} \frac{d\mu_1(x_1)}{z - x_1}, \quad f_2(z) = \int_{\Delta_1} \frac{d\mu_1(x_1)}{z - x_1} \int_{\Delta_2} \frac{d\mu_2(x_2)}{x_1 - x_2}.$$

Remark 5.1. The definition of a Nikishin system depends on the order in which one “folds” measures. If one starts from $d\mu_2$, rather than $d\mu_1$ one obtains a priori a different system. As we show below the relation between these two Nikishin systems is in fact of central importance to the theory we are developing.

The following elementary observation provides the proper framework for our discussion.

Lemma 5.1. Let $d\alpha^*$ denote the measure obtained from $d\alpha$ by reflecting the support of $d\alpha$ with respect to the origin. Then $W_\beta, W_{\beta\alpha^*}$ and $W_{\alpha^*}, W_{\alpha^*\beta}$ are Nikishin systems associated with measures $d\beta$ and $d\alpha^*$ with no predetermined ordering of measures.

The relation between these two Nikishin systems can now be readily obtained.

Lemma 5.2.

$$W_\beta(z)W_{\alpha^*}(z) = W_{\beta\alpha^*}(z) + W_{\alpha^*\beta}(z). \tag{5.2}$$

Proof. Elementary computation gives:

$$\begin{aligned} W_\beta(z)W_{\alpha^*}(z) &= \iint \frac{1}{(z - y)(z + x)} d\alpha(x)d\beta(y) \\ &= \iint \frac{1}{(x + y)} \left[\frac{1}{z - y} - \frac{1}{z + x} \right] d\alpha(x)d\beta(y), \end{aligned}$$

which implies the claim. \square

Remark 5.2. Eq. (5.2) was introduced in [26] for the DP peakons (see Lemma 4.7 there). Observe that this formula is valid for any Nikishin system of order 2.

We formulate now the main approximation problem, modeled after that of [26].

Definition 5.3. Let $n \geq 1$. Given two Nikishin systems $W_\beta, W_{\beta\alpha^*}$ and $W_{\alpha^*}, W_{\alpha^*\beta}$ we seek polynomials $Q(z)$, $\deg Q = n$, $P_\beta(z)$, $\deg P_\beta = n - 1$ and $P_{\beta\alpha^*}(z)$, $\deg P_{\beta\alpha^*} = n - 1$, which satisfy Padé-like approximation conditions as $z \rightarrow \infty, z \in \mathbb{C}_\pm$:

$$Q(z)W_\beta(z) - P_\beta(z) = \mathcal{O}\left(\frac{1}{z}\right), \tag{5.3a}$$

$$Q(z)W_{\beta\alpha^*}(z) - P_{\beta\alpha^*}(z) = \mathcal{O}\left(\frac{1}{z}\right), \tag{5.3b}$$

$$Q(z)W_{\alpha^*\beta}(z) - P_\beta(z)W_{\alpha^*}(z) + P_{\beta\alpha^*}(z) = \mathcal{O}\left(\frac{1}{z^{n+1}}\right). \tag{5.3c}$$

Remark 5.3. In the case that both measures have compact support we can remove the condition that $z \in \mathbb{C}_\pm$ since all the functions involved are then holomorphic around $z = \infty$.

Remark 5.4. In the terminology used for example in [33] the triplets of polynomials $Q, P_\beta, P_{\beta\alpha^*}$ provide a Hermite–Padé approximation of type *I* to the Nikishin system $W_\beta, W_{\beta\alpha^*}$ and, simultaneously, a Hermite–Padé approximation of type *II* to the Nikishin system $W_{\alpha^*}, W_{\alpha^*\beta}$.

Simultaneous approximations were first studied in [30]. Their algebraic properties were investigated in [32,31] via an approach similar to the one we employed above.

Definition 5.4. We call the left hand sides of approximation problems (5.3) $R_\beta, R_{\beta\alpha^*}$ and $R_{\alpha^*\beta}$ respectively, referring to them as remainders.

The relation of the approximation problem (5.3) to the theory of biorthogonal polynomials \mathbf{q} and \mathbf{p} is the subject of the next theorem.

Theorem 5.1. Let $q_n(y)$ be defined as in (2.10), and let us set $Q(z) = q_n(z)$. Then $Q(z)$ is the unique, up to a multiplicative constant, solution of the approximation problem (5.3). Moreover, $P_\beta, P_{\beta\alpha^*}$ and all the remainders $R_\beta, R_{\beta\alpha^*}$ and $R_{\alpha^*\beta}$ are uniquely determined from Q with the help of the formulas:

$$P_\beta(z) = \int \frac{Q(z) - Q(y)}{z - y} d\beta(y), \quad P_{\beta\alpha^*}(z) = \iint \frac{Q(z) - Q(y)}{(z - y)(x + y)} d\alpha(x) d\beta(y), \tag{5.4a}$$

$$R_\beta(z) = \int \frac{Q(y)}{z - y} d\beta(y), \quad R_{\beta\alpha^*}(z) = \iint \frac{Q(y)}{(z - y)(x + y)} d\alpha(x) d\beta(y), \tag{5.4b}$$

$$R_{\alpha^*\beta}(z) = - \iint \frac{Q(y)}{(z + x)(x + y)} d\alpha(x) d\beta(y) = \int \frac{R_\beta(x)}{z - x} d\alpha^*(x). \tag{5.4c}$$

Proof. We start with the first approximation problem involving $Q(z)W_\beta(z)$. Writing explicitly its first term we get:

$$\int \frac{Q(z)}{z - y} d\beta(y) = \int \frac{Q(z) - Q(y)}{z - y} d\beta(y) + \int \frac{Q(y)}{z - y} d\beta(y).$$

Since $\int \frac{Q(z) - Q(y)}{z - y} d\beta(y)$ is a polynomial in z of degree $n - 1$, while $\int \frac{Q(y)}{z - y} d\beta(y) = \mathcal{O}(\frac{1}{z})$, we get the first and the third formulas. The second and fourth formulas are obtained in an analogous way from the second approximation problem. Furthermore, to get the last formula we compute P_β and $P_{\beta\alpha^*}$ from the first two approximation problems and substitute into the third approximation problem, using on the way Lemma 5.2, to obtain:

$$R_\beta W_{\alpha^*} - R_{\beta\alpha^*} = R_{\alpha^*\beta}.$$

Substituting explicit formulas for R_β and $R_{\beta\alpha^*}$ gives the final formula. To see that $Q(z)$ is proportional to $q_n(z)$ we rewrite $-R_{\alpha^*\beta}$ as:

$$\begin{aligned} \iint \frac{Q(y)}{(z + x)(x + y)} d\alpha(x) d\beta(y) &= \iint \frac{Q(y)}{(x + y)} \left[\frac{1}{z + x} - \frac{1 - (-\frac{x}{z})^n}{z + x} \right] d\alpha(x) d\beta(y) \\ &+ \iint \sum_{j=0}^{n-1} \frac{(-x)^j}{z^{j+1}} \frac{Q(y)}{(x + y)(z + x)} d\alpha d\beta \\ &= \iint \frac{Q(y)}{(x + y)} \left[\frac{(-x/z)^n}{z + x} \right] d\alpha(x) d\beta(y) + \iint \sum_{j=0}^{n-1} \frac{(-x)^j}{z^{j+1}} \frac{Q(y)}{(x + y)(z + x)} d\alpha d\beta \end{aligned}$$

To finish the argument we observe that the first term is already $\mathcal{O}(\frac{1}{z^{n+1}})$, hence the second term must vanish. This gives:

$$\iint \frac{x^j Q(y)}{x + y} d\alpha(x) d\beta(y) = 0, \quad 0 \leq j \leq n - 1,$$

which characterizes uniquely (up to a multiplicative constant) the polynomial q_n . □

Remark 5.5. In the body of the proof we used an equivalent form of the third approximation condition, namely

$$R_\beta W_{\alpha^*}(z) - R_{\beta\alpha^*}(z) = R_{\alpha^*\beta}(z) = \mathcal{O}\left(\frac{1}{z^{n+1}}\right). \tag{5.5}$$

By symmetry, we can consider the Nikishin systems associated with measures α and β^* with the corresponding Markov functions $W_\alpha, W_{\alpha\beta^*}$ and $W_{\beta^*}, W_{\beta^*\alpha}$. We then have an obvious interpretation of the polynomials p_n .

Theorem 5.2. *Let $p_n(x)$ be defined as in (2.10), and let us set $Q(z) = p_n(z)$. Then $Q(z)$ is the unique, up to a multiplicative constant, solution of the approximation problem for $z \rightarrow \infty$, $z \in \mathbb{C}_\pm$:*

$$Q(z)W_\alpha(z) - P_\alpha(z) = \mathcal{O}\left(\frac{1}{z}\right), \tag{5.6a}$$

$$Q(z)W_{\alpha\beta^*}(z) - P_{\alpha\beta^*}(z) = \mathcal{O}\left(\frac{1}{z}\right), \tag{5.6b}$$

$$Q(z)W_{\beta^*\alpha}(z) - P_\alpha(z)W_{\beta^*}(z) + P_{\alpha\beta^*}(z) = \mathcal{O}\left(\frac{1}{z^{n+1}}\right), \tag{5.6c}$$

where $P_\alpha, P_{\alpha\beta^*}$ are given by formulas of Theorem 5.1 after switching α with β .

Clearly, one does not need to go to four different types of Nikishin systems in order to characterize q_n and p_n . The following corollary is an alternative characterization of biorthogonal polynomials which uses only the first pair of Nikishin systems.

Corollary 5.1. *Consider the Nikishin systems $W_\beta, W_{\beta\alpha^*}$ and $W_{\alpha^*}, W_{\alpha^*\beta}$. Then the pair of biorthogonal polynomials $\{q_n, p_n\}$ solves:*

1. $Q(z) = q_n(z)$ solves Hermite–Padé approximations given by Eq. (5.3),

$$Q(z)W_\beta(z) - P_\beta(z) = \mathcal{O}\left(\frac{1}{z}\right),$$

$$Q(z)W_{\beta\alpha^*}(z) - P_{\beta\alpha^*}(z) = \mathcal{O}\left(\frac{1}{z}\right),$$

$$Q(z)W_{\alpha^*\beta}(z) - P_\beta(z)W_{\alpha^*}(z) + P_{\beta\alpha^*}(z) = \mathcal{O}\left(\frac{1}{z^{n+1}}\right).$$

2. $Q(z) = p_n(-z)$ solves switched (Type I with Type II) Hermite–Padé approximations

$$Q(z)W_{\alpha^*}(z) - P_{\alpha^*}(z) = \mathcal{O}\left(\frac{1}{z}\right), \tag{5.7a}$$

$$Q(z)W_{\alpha^*\beta}(z) - P_{\alpha^*\beta}(z) = \mathcal{O}\left(\frac{1}{z}\right), \tag{5.7b}$$

$$Q(z)W_{\beta\alpha^*}(z) - P_{\alpha^*}(z)W_{\beta}(z) + P_{\alpha^*\beta}(z) = \mathcal{O}\left(\frac{1}{z^{n+1}}\right). \tag{5.7c}$$

We finish this section with a few results needed for the Riemann–Hilbert problem approach to biorthogonal polynomials $\{q_n, p_n\}$ which will be presented in the next section.

Definition 5.5. We define the auxiliary vectors in addition to the main polynomial vectors $\mathbf{q}_0(w) := \mathbf{q}(w)$ and $\mathbf{p}_0(z) := \mathbf{p}(z)$, as

$$\mathbf{q}_1(w) := \int \mathbf{q}(y) \frac{d\beta(y)}{w - y}, \quad \mathbf{q}_2(w) := \int \frac{\mathbf{q}_1(x)}{w - x} d\alpha^*(x), \tag{5.8}$$

$$\mathbf{p}_1(z) := \int \frac{\mathbf{p}(x)d\alpha(x)}{z - x}, \quad \mathbf{p}_2(z) := \int \frac{\mathbf{p}_1(y)}{z - y} d\beta^*(y). \tag{5.9}$$

Moreover,

$$\widehat{\mathbf{p}}_1(z) := \widehat{L}^{-1} \left(\mathbf{p}_1(z) + \frac{1}{\beta_0} (\mathbf{p}|1) \right) = \widehat{L}^{-1} \mathbf{p}_1(z) - \mathbf{1}, \tag{5.10}$$

$$\widehat{\mathbf{p}}_2(z) := \int \frac{\widehat{\mathbf{p}}_1(y)}{z - y} d\beta^*(y). \tag{5.11}$$

Here $\mathbf{1}$ is the vector of ones.¹

Remark 5.6. Note that the definition above unifies the approximants and their respective remainders (see Theorem 5.1), thus, for example, $\mathbf{q}_1(w) = \mathbf{R}_{\beta}(w)$, $\mathbf{q}_2(w) = \mathbf{R}_{\alpha^*\beta}(w)$ etc. The definition of “hatted” quantities is justified below.

Theorem 5.3 (Extended Christoffel–Darboux Identities). Let $a, b = 0, \dots, 2$. Then

$$(w + z)\mathbf{q}_a^T(w) \Pi \mathbf{p}_b(z) = \mathbf{q}_a^T(w) \mathbb{A}(-w) \widehat{\mathbf{p}}_b(z) - \mathbb{F}(w, z)_{ab} \tag{5.12}$$

where

$$\mathbb{F}(w, z) = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & W_{\beta^*}(z) + W_{\beta}(w) \\ 1 & W_{\alpha}(z) + W_{\alpha^*}(w) & W_{\alpha^*}(w)W_{\beta^*}(z) + W_{\alpha^*\beta}(w) + W_{\beta^*\alpha}(z) \end{bmatrix}. \tag{5.13}$$

¹ The formula $\beta_0^{-1}(\widehat{\mathbf{p}}_n, \mathbf{1}) = -1$ follows directly from the determinantal expression in Proposition 4.2.

Proof. The proof goes by repeated applications of the Christoffel–Darboux Identities given by Theorem 4.1 and Padé approximation conditions (5.3). The details have been relegated to Appendix A. □

We point out that if we set $w = -z$ in the CDI’s contained in Theorem 5.3, the left hand side vanishes identically and the right hand side contains terms of the form $\mathbf{q}_a(-z)\mathbb{A}(z)\widehat{\mathbf{p}}_b(z)$ minus $\mathbf{F}_{ab}(-z, z)$. The main observation is that the second term is **constant**, independent of both z and n , and hence one ends up with the **perfect pairing** (see [5]) between the auxiliary vectors. For the reader’s convenience we recall the definition of $\mathbb{A}(z)$ to emphasize the implicit dependence on the index n hidden in the projection Π .

Theorem 5.4. (Perfect Duality)

Let

$$\mathbb{J} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

Then

$$\mathbf{q}_a^T(-z)\mathbb{A}(z)\widehat{\mathbf{p}}_b(z) = \mathbb{J}_{ab}, \quad \text{where } \mathbb{A}(z) = [(z - \mathbf{X})\widehat{L}, \Pi].$$

Proof. The only nontrivial entry to check is (2, 2). In this case, after one substitutes $w = -z$ into $W_{\alpha^*}(w)W_{\beta^*}(z) + W_{\alpha^*\beta}(w) + W_{\beta^*\alpha}(z)$, one obtains the identity of Lemma 5.2. □

There also exists an analog of the extended Christoffel–Darboux identities of Theorem 5.3 for the “hatted” quantities.

We first define:

Definition 5.6. For $a = 0, 1, 2$,

$$\widehat{\mathbf{q}}_a^T := \mathbf{q}_a^T \widehat{L}. \tag{5.14}$$

The following identities follow directly from the respective definitions.

Lemma 5.3.

$$w\widehat{\mathbf{q}}_a^T(w) = \begin{cases} \mathbf{q}_a^T(w)\mathbf{Y}^T\widehat{L}, & a = 0, 1 \\ \mathbf{q}_2^T(w)\mathbf{Y}^T\widehat{L} - \langle 1|\widehat{\mathbf{q}}_0^T \rangle, & a = 2. \end{cases}$$

$$(z - \mathbf{X})\widehat{L}\widehat{\mathbf{p}}_b(z) = \begin{cases} 0, & b = 0, \\ \frac{\langle \mathbf{p}_0|z + y \rangle}{\beta_0}, & b = 1, \\ -\langle \mathbf{p}_0|1 \rangle + \frac{\langle \mathbf{p}_0|z + y \rangle W_{\beta^*}(z)}{\beta_0}, & b = 2. \end{cases}$$

Theorem 5.5 (Extended Christoffel–Darboux Identities for $\widehat{\mathbf{q}}_a, \widehat{\mathbf{p}}_b$). Let $a, b = 0, \dots, 2$. Then

$$(w + z)\widehat{\mathbf{q}}_a^T(w)\Pi\widehat{\mathbf{p}}_b(z) = \mathbf{q}_a^T(w)\mathbb{A}(z)\widehat{\mathbf{p}}_b(z) - \widehat{\mathbb{F}}(w, z)_{ab} \tag{5.15}$$

where

$$\widehat{\mathbb{F}}(w, z) = \mathbb{F}(w, z) - \frac{w + z}{\beta_0} \begin{bmatrix} 0 & 1 & W_{\beta^*}(z) \\ 0 & W_{\beta}(z) & W_{\beta}(w)W_{\beta^*}(z) \\ 1 & W_{\alpha^*\beta^*}(w) & W_{\alpha^*\beta^*}(w)W_{\beta^*}(z) \end{bmatrix}. \tag{5.16}$$

Proof. We give an outline of the proof. For $a = 0, 1$, in view of Lemma 5.3

$$(w + z)\widehat{\mathbf{q}}_a^T(w) \Pi \widehat{\mathbf{p}}_b(z) = \mathbf{q}_a^T(w) \mathbb{A}(z) \widehat{\mathbf{p}}_b(z) + \mathbf{q}_a^T(w) \Pi(z - \mathbf{X}) \widehat{L} \widehat{\mathbf{p}}_b(z).$$

The second term equals, again by Lemma 5.3,

$$\mathbf{q}_a^T(w) \Pi \begin{cases} 0, & b = 0, \\ \frac{\langle \mathbf{p}_0 | z + y \rangle}{\beta_0}, & b = 1, \\ -\langle \mathbf{p}_0 | 1 \rangle + \frac{\langle \mathbf{p}_0 | z + y \rangle W_{\beta^*}(z)}{\beta_0}, & b = 2. \end{cases}$$

Now, one goes case by case, using biorthogonality of \mathbf{q}_0^T and \mathbf{p}_0 , and the definition of $\mathbf{q}_1^T(w)$. After a few elementary steps one arrives at the claimed result. The computation for $a = 2$ is only slightly more involved. From Lemma 5.3 we obtain:

$$(w + z)\widehat{\mathbf{q}}_2^T(w) \Pi \widehat{\mathbf{p}}_b(z) = \mathbf{q}_2^T(w) \mathbb{A}(z) \widehat{\mathbf{p}}_b(z) - \langle 1 | \widehat{\mathbf{q}}_0 \rangle \Pi \widehat{\mathbf{p}}_b(z) + \mathbf{q}_2^T(w) \Pi(z - \mathbf{X}) \widehat{L} \widehat{\mathbf{p}}_b(z).$$

In view of biorthogonality of $\widehat{\mathbf{q}}_0^T$ and $\widehat{\mathbf{p}}$, after some intermediate computations, one obtains:

$$\langle 1 | \widehat{\mathbf{q}}_0 \rangle \Pi \widehat{\mathbf{p}}_b(z) = \begin{cases} 1, & b = 0 \\ W_\alpha(z) + \frac{\langle 1 | 1 \rangle}{\beta_0}, & b = 1, \\ W_{\beta^* \alpha}(z) + \frac{\langle 1 | 1 \rangle}{\beta_0} W_{\beta^*}(z), & b = 2. \end{cases}$$

Likewise,

$$\mathbf{q}_2^T(w) \Pi(z - \mathbf{X}) \widehat{L} \widehat{\mathbf{p}}_b(z) = \begin{cases} 0, & b = 0 \\ \frac{w + z}{\beta_0} W_{\alpha^* \beta}(w) - W_{\alpha^*}(w) + \frac{\langle 1 | 1 \rangle}{\beta_0}, & b = 1, \\ \frac{w + z}{\beta_0} W_{\beta^*}(z) W_{\alpha^*}(w) - W_{\alpha^* \beta}(w) \\ - W_{\beta^*}(z) W_{\alpha^*}(w) + \frac{\langle 1 | 1 \rangle}{\beta_0} W_{\beta^*}(z), & b = 2, \end{cases}$$

and the claim follows. \square

6. Riemann–Hilbert problems

In this section we set up two Riemann–Hilbert problems characterizing the Cauchy BOPs that enter the Christoffel–Darboux identities of the previous section. This is done in anticipation of possible applications to the study of universality for the corresponding two-matrix model. Moreover, since the Christoffel–Darboux kernels contain also the hatted polynomials, it is useful to formulate the Riemann–Hilbert problems for those polynomials as well. Similar Riemann–Hilbert problems were posed in [1,33].

We will also make the **assumption** (confined to this section) that the measures $d\alpha, d\beta$ are *absolutely continuous with respect to Lebesgue’s measure* on the respective axes. Thus one can write $\frac{d\alpha}{dx} = e^{-\frac{U(x)}{\hbar}}$, $\frac{d\beta}{dy} = e^{-\frac{V(y)}{\hbar}}$, for the respective (positive!) densities on the respective supports: the signs in the exponents are conventional so as to have (in the case of an unbounded support) the *potentials* U, V bounded from below. The constant \hbar is only for convenience when studying the asymptotics of biorthogonal polynomials for large degrees (small \hbar).

Since the Christoffel–Darboux identities involve the expressions $\mathbf{q}_a \mathbb{A} \widehat{\mathbf{p}}_b$, we are naturally led to characterize the sequences \mathbf{q} and $\widehat{\mathbf{p}}$. However, the other sequences can be characterized in a similar manner by swapping the roles of the relevant measures and symbols.

6.1. Riemann–Hilbert problem for the \mathbf{q} -BOPs

We will be describing here only the RHP characterizing the polynomials $q_n(y)$, where the characterization of the polynomials $p_n(x)$ is obtained by simply interchanging α with β (see for example Theorem 5.2).

We consider the real axis \mathbb{R} oriented as usual and define

$$\begin{aligned} \vec{\mathbf{q}}_0^{(n)}(w) &:= [q_{n-2}(w) \ q_{n-1}(w) \ q_n(w)]^t, & \vec{\mathbf{q}}_1^{(n)}(w) &:= \int \vec{\mathbf{q}}^{(n)}(y) \frac{d\beta(y)}{w-y}, \\ \vec{\mathbf{q}}_2^{(n)}(w) &:= \int \vec{\mathbf{q}}_1^{(n)}(x) \frac{d\alpha^*(x)}{w-x}. \end{aligned} \tag{6.1}$$

For simplicity of notation we will suppress the superscript (n) in most of the following discussions, only to restore it when necessary for clarity; the main point is that an arrow on top of the corresponding vector will denote a “window” of three consecutive entries of either the ordinary vector \mathbf{q} (index $a = 0$), or the auxiliary vectors \mathbf{q}_a (index $a = 1, 2$, see Definition 5.5) which, as we might recall at this point, combine the polynomials and the corresponding remainders in the Hermite–Padé approximation problem given by Theorem 5.1. Some simple observations are in order. The vector $\vec{\mathbf{q}}_1(w)$ is an analytic vector which has a jump discontinuity on the support of $d\beta$ contained in the positive real axis. As $w \rightarrow \infty$ (away from the support of $d\beta$) it decays as $\frac{1}{w}$. Its jump discontinuity is (using Plemelj formula)

$$\vec{\mathbf{q}}_1(w)_+ = \vec{\mathbf{q}}_1(w)_- - 2\pi i \frac{d\beta}{dw} \vec{\mathbf{q}}_0(w), \quad w \in \text{supp}(d\beta). \tag{6.2}$$

Looking at the leading term at $w = \infty$ we see that

$$\vec{\mathbf{q}}_1(w) = \frac{1}{w} [\eta_{n-2} \ \eta_{n-1} \ \eta_n]^t + \mathcal{O}(1/w^2). \tag{6.3}$$

The vector $\vec{\mathbf{q}}_2(w)$ is also analytic with a jump discontinuity on the **reflected support** of $d\alpha$ (i.e. on $\text{supp}(d\alpha^*)$). In view of Theorem 5.1, recalling that \mathbf{q}_2 are remainders of the Hermite–Padè approximation problem of type II, we easily see that

$$\begin{aligned} \vec{\mathbf{q}}_2(w) &= \left[\frac{c_{n-2}}{(-w)^{n-1}} \quad \frac{c_{n-1}}{(-w)^n} \quad \frac{c_n}{(-w)^{n+1}} \right]^t (1 + \mathcal{O}(1/w)), \\ c_n &:= \langle x^n | q_n \rangle = \sqrt{\frac{D_{n+1}}{D_n}} > 0. \end{aligned} \tag{6.4}$$

The jump discontinuity of $\vec{\mathbf{q}}_2$ is

$$\vec{\mathbf{q}}_2(w)_+ = \vec{\mathbf{q}}_2(w)_- - 2\pi i \frac{d\alpha^*}{dw} \vec{\mathbf{q}}_1(w) \quad w \in \text{supp}(d\alpha^*). \tag{6.5}$$

The behavior of $\vec{\mathbf{q}}_0(w)$ at infinity is

$$\vec{\mathbf{q}}_0(w) = \begin{bmatrix} \frac{w^{n-2}}{c_{n-2}} & \frac{w^{n-1}}{c_{n-1}} & \frac{w^n}{c_n} \end{bmatrix}^t (1 + \mathcal{O}(1/w)), \tag{6.6}$$

with the same c_n 's as in (6.4).

Define the matrix

$$\Gamma(w) := \overbrace{\begin{bmatrix} 1 & -c_n \eta_n & 0 \\ 0 & 1 & 0 \\ 0 & \frac{(-1)^{n-1} \eta_{n-2}}{c_{n-2}} & 1 \end{bmatrix}}{=: \mathcal{N}_q} \begin{bmatrix} 0 & 0 & c_n \\ 0 & \frac{1}{\eta_{n-1}} & 0 \\ \frac{(-1)^n}{c_{n-2}} & 0 & 0 \end{bmatrix} \times [\vec{\mathbf{q}}_0^{(n)}(w), \vec{\mathbf{q}}_1^{(n)}(w), \vec{\mathbf{q}}_2^{(n)}(w)] \tag{6.7}$$

Proposition 6.1. *The matrix $\Gamma(w)$ is analytic on $\mathbb{C} \setminus \text{supp}(d\beta) \cup \text{supp}(d\alpha^*)$. Moreover, it satisfies the jump conditions*

$$\Gamma(w)_+ = \Gamma(w)_- \begin{bmatrix} 1 & -2\pi i \frac{d\beta}{dw} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad w \in \text{supp}(d\beta) \subset \mathbb{R}_+ \tag{6.8}$$

$$\Gamma(w)_+ = \Gamma(w)_- \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -2\pi i \frac{d\alpha^*}{dw} \\ 0 & 0 & 1 \end{bmatrix}, \quad w \in \text{supp}(d\alpha^*) \subset \mathbb{R}_-$$

and its asymptotic behavior at $w = \infty$ is

$$\Gamma(w) = (\mathbf{1} + \mathcal{O}(w^{-1})) \begin{bmatrix} w^n & 0 & 0 \\ 0 & w^{-1} & 0 \\ 0 & 0 & w^{-n+1} \end{bmatrix}. \tag{6.9}$$

Moreover, $\Gamma(w)$ can be written as:

$$\Gamma(w) = \begin{bmatrix} c_n \eta_n & 0 & 0 \\ 0 & \frac{1}{\eta_{n-1}} & 0 \\ 0 & 0 & \frac{(-1)^{n-1} \eta_{n-2}}{c_{n-2}} \end{bmatrix} \begin{bmatrix} \widehat{q}_{n-1} & \widehat{q}_{1,n-1} & \widehat{q}_{2,n-1} \\ q_{n-1} & q_{1,n-1} & q_{2,n-1} \\ \widehat{q}_{n-2} & \widehat{q}_{1,n-2} & \widehat{q}_{2,n-2} \end{bmatrix}. \tag{6.10}$$

Proof. All the properties listed are obtained from elementary matrix computations. □

Remark 6.1. An analogous problem with the roles of α, β , etc., interchanged, characterizes the monic orthogonal polynomials $p_{n-1}(x)$ of degree $n - 1$ in x .

Corollary 6.1. *Given $n \in \mathbb{N}$, the absolutely continuous measures $d\beta \subset \mathbb{R}_+$ and $d\alpha^* \subset \mathbb{R}_-$, and assuming the existence of all the bimoments I_{ij} there exists a unique matrix $\Gamma(w)$ solving the*

RHP specified by Eqs. (6.8), (6.9). The solution characterizes uniquely the polynomials q_{n-1} as well as \widehat{q}_{n-1} . In particular, the normalization constants c_{n-1}, η_{n-1} (i.e. the “norm” of the monic orthogonal polynomials and the β average of the q_{n-1}) are read off the following expansions

$$\begin{aligned} \Gamma_{2,1}(w) &= \frac{1}{c_{n-1}\eta_{n-1}} w^{n-1} + \mathcal{O}(w^{n-2}), \\ \Gamma_{2,3}(w) &= (-1)^n \frac{c_{n-1}}{\eta_{n-1}w^n} + \mathcal{O}(w^{-n-1}) \end{aligned} \tag{6.11}$$

or, equivalently,

$$\frac{1}{\eta_{n-1}^2} = (-1)^n \lim_{w \rightarrow \infty} w \Gamma_{2,1}(w) \Gamma_{2,3}(w), \quad c_{n-1}^2 = (-1)^n \lim_{w \rightarrow \infty} w^{2n-1} \frac{\Gamma_{2,3}(w)}{\Gamma_{2,1}(w)}. \tag{6.12}$$

Proof. Given $d\beta$ and $d\alpha^*$ it suffices to construct the Nikishin systems $W_\beta, W_{\beta\alpha^*}$ and $W_{\alpha^*}, W_{\alpha^*\beta}$ followed by solving the Hermite–Padé approximation problems given by Eqs. (5.3). The existence of the solution is ensured by the existence of all bimoments I_{ij} (see Eq. (2.4) for the definition). Then one constructs the polynomials \widehat{q}_j , finally the matrix $\Gamma(w)$ using Eq. (6.10). By construction $\Gamma(w)$ satisfies the Riemann–Hilbert factorization problem specified by Eqs. (6.8) and (6.9). Since the determinant of $\Gamma(w)$ is constant in w (and equal to one), the solution of the Riemann–Hilbert problem is unique. The formulas for η_{n-1} and c_{n-1} follow by elementary matrix computations. \square

Remark 6.2. By multiplication on the right with a diagonal matrix $\mathbb{Y}(w) := \Gamma(w)\text{diag}(\exp(-\frac{2V+U^*}{3\hbar}), \exp(\frac{V-U^*}{3\hbar}), \exp(\frac{2U^*+V}{3\hbar}))$ one can reduce the RHP to an equivalent one with constant jumps. It then follows that $\mathbb{Y}(w)$ solves a linear ODE with the same singularities as $V', U^{*\prime}$; for example if U', V' are rational functions then so is the coefficient matrix of the ODE and the orders of poles do not exceed those of V', U' . In this case it can be shown [6] that the principal minors of the matrix of bimoments are isomonodromic tau-functions in the sense of Jimbo–Miwa–Ueno [22].

6.2. Riemann–Hilbert problem for the $\widehat{\mathbf{p}}$ -BOPs

Referring to the defining properties of $\widehat{p}_n(x)$ as indicated in Proposition 4.2 we are going to define a second 3×3 local RHP that characterizes them.

Define

$$\widetilde{\mathbf{p}}_0(z) := [\widehat{p}_{n-2}(z) \quad \widehat{p}_{n-1}(z) \quad \widehat{p}_n(z)]^t \tag{6.13}$$

and $\widetilde{\mathbf{p}}_{1,2}(z)$ as the same **windows** of the auxiliary vectors $\widehat{\mathbf{p}}_{1,2}$ introduced in Definition 5.5. We first study the large z asymptotic behavior of $\widehat{p}_{0,n}(z), \widehat{p}_{1,n}(z), \widehat{p}_{2,n}(z)$.

Lemma 6.1. *The asymptotic behavior at $z \rightarrow \infty, z \in \mathbb{C}_\pm$ is given by:*

$$\widehat{p}_{0,n}(z) = -\frac{\eta_n}{c_n} z^n (1 + \mathcal{O}(1/z)), \tag{6.14}$$

$$\widehat{p}_{1,n}(z) = -1 + \mathcal{O}(1/z), \tag{6.15}$$

$$\widehat{p}_{2,n}(z) = (-1)^n \frac{c_{n+1}\eta_{n+1}}{z^{n+2}} (1 + \mathcal{O}(1/z)). \tag{6.16}$$

Proof. We give a proof for $\widehat{p}_{1,n}(z) = \int \frac{\widehat{p}_{0,n}(x)}{z-x} d\alpha(x) + \frac{1}{\beta_0} \langle \widehat{p}_{0,n} | 1 \rangle$. The first term is $\mathcal{O}(\frac{1}{z})$, while the second term can be computed using biorthogonality and the fact that $\widehat{p}_{0,n} = -(\eta_n p_{0,n} + \eta_{n-1} p_{0,n-1} + \dots + \eta_0 p_{0,0})$. Thus the second term equals $-\frac{\eta_0}{\beta_0} \langle p_{0,0} | 1 \rangle = -1$, since $\eta_0 = q_0 \beta_0$, hence the claim for $\widehat{p}_{1,n}(z)$ follows. The remaining statements are proved in a similar manner. \square

For reasons of normalization, and in full analogy with Eq. (6.7), we arrange the window of all \widehat{p}_s wave vectors into the matrix

$$\widehat{\Gamma}(z) = \overbrace{\begin{bmatrix} 0 & 0 & -\frac{c_n}{\eta_n} \\ 0 & -1 & 0 \\ \frac{(-1)^n}{c_{n-1}\eta_{n-1}} & 0 & 0 \end{bmatrix}}{=: \mathcal{N}_{\widehat{p}}} \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} \widehat{\mathbf{p}}(z), \widehat{\mathbf{p}}_1(z), \widehat{\mathbf{p}}_2(z) \end{bmatrix}. \tag{6.17}$$

Proposition 6.2. *The matrix $\widehat{\Gamma}(z)$ is analytic in $\mathbb{C} \setminus (\text{supp}(d\alpha) \cup \text{supp}(d\beta^*))$. Moreover, it satisfies the jump conditions*

$$\widehat{\Gamma}(z)_+ = \widehat{\Gamma}(z)_- \begin{bmatrix} 1 & -2\pi i \frac{d\alpha}{dz} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad z \in \text{supp}(d\alpha) \subseteq \mathbb{R}_+ \tag{6.18}$$

$$\widehat{\Gamma}(z)_+ = \widehat{\Gamma}(z)_- \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -2\pi i \frac{d\beta^*}{dz} \\ 0 & 0 & 1 \end{bmatrix}, \quad z \in \text{supp}(d\beta^*) \subseteq \mathbb{R}_-,$$

and its asymptotic behavior at $z = \infty$ is

$$\widehat{\Gamma}(z) = \left(\mathbf{1} + \mathcal{O}\left(\frac{1}{z}\right) \right) \begin{bmatrix} z^n & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{z^n} \end{bmatrix}. \tag{6.19}$$

$\widehat{\Gamma}(z)$ can be written as:

$$\widehat{\Gamma}(z) = \begin{bmatrix} c_n & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & \frac{(-1)^n}{c_{n-1}} \end{bmatrix} \begin{bmatrix} p_{0,n} & p_{1,n} & p_{2,n} \\ \widehat{p}_{0,n-1} & \widehat{p}_{1,n-1} & \widehat{p}_{2,n-1} \\ p_{0,n-1} & p_{1,n-1} & p_{2,n-1} \end{bmatrix}. \tag{6.20}$$

The existence and uniqueness of the solution of the Riemann–Hilbert problem (6.18), (6.19) is proved in a similar way to the proof of Corollary 6.1.

Corollary 6.2. *Given $n \in \mathbb{N}$, the absolutely continuous measures $d\alpha \subset \mathbb{R}_+$ and $d\beta^* \subset \mathbb{R}_-$, and assuming the existence of all the bimoments I_{ij} there exists a unique matrix $\Gamma(z)$ solving the RHP specified by Eqs. (6.18), (6.19). The solution characterizes uniquely the polynomials \widehat{p}_{n-1} and p_n .*

Acknowledgments

M.B. would like to thank the Department of Mathematics of the University of Notre Dame for their hospitality during which the project was initiated and J. Harnad for insight into the relationship of Cauchy biorthogonal polynomials with matrix models.

While working on this project, M.B. and M.G. enjoyed the hospitality of the Department of Mathematics, University of Saskatchewan and M.G. and J.S. enjoyed the hospitality of the Centre de recherches mathématiques, Université de Montréal.

J.S. would also like to thank H. Lundmark for an ongoing collaboration on the cubic string problem which motivated many of the questions addressed in this paper.

The first author’s work was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC), Grant. No. 261229-03 and by the Fonds FCAR du Québec No. 88353. The second author’s work was supported in part by NSF Grant DMD-0400484. The third author’s work was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC), Grant. No. 138591-04.

Appendix. Proof of extended Christoffel–Darboux identities

Theorem A.1 (Extended Christoffel–Darboux Identities Theorem 5.3). *Let $a, b = 0, \dots, 2$. Then*

$$(w + z)\mathbf{q}_a^T(w)\Pi\mathbf{p}_b(z) = \mathbf{q}_a^T(w)\mathbb{A}(-w)\widehat{\mathbf{p}}_b(z) - \mathbb{F}(w, z)_{ab} \tag{A.1}$$

where

$$\mathbb{F}(w, z) = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & W_{\beta^*}(z) + W_{\beta}(w) \\ 1 & W_{\alpha}(z) + W_{\alpha^*}(w) & W_{\alpha^*}(w)W_{\beta^*}(z) + W_{\alpha^*\beta}(w) + W_{\beta^*\alpha}(z) \end{bmatrix}. \tag{A.2}$$

Proof. The proof goes by repeated applications of the Christoffel–Darboux Identities given by Theorem 4.1 and Padé approximation conditions (5.3). We observe that all quantities with labels $a = 1, 2$ have asymptotic expansions around ∞ in the open half-planes \mathbb{C}_{\pm} (they are holomorphic expansions in the case of compactly supported measures $d\alpha, d\beta$). We will subsequently call the part of the expansion corresponding to negative powers of z or w , of a function $f(z, w)$ the *regular* part of f and denote it $(f(z, w))_{-,z}, (f(z, w))_{-,w}$ respectively. In all cases the *regular* parts are obtained by subtracting certain polynomial expressions from functions holomorphic in \mathbb{C}_{\pm} and as such the *regular* parts are holomorphic in these half-planes with vanishing limits at ∞ approach from within the respective half-planes.

We will only indicate the main steps in computations for each entry, denoted below by (a, b) . (0, 1): With the help of the first approximation condition, we have

$$\mathbf{q}_1^T(w)\Pi\mathbf{p}_0(z) = \left(\int \frac{\mathbf{q}_0^T(w)\Pi\mathbf{p}_0(z)}{w - y} d\beta(y) \right)_{-,w}.$$

Using the Christoffel–Darboux Identities and the notation of Corollary 4.3 we get

$$\begin{aligned} \mathbf{q}_1^T(w)\Pi\mathbf{p}_0(z) &= \left(\int \frac{\mathbf{q}_0^T(w)\mathbb{A}(-w)\widehat{\mathbf{p}}_0(z)}{(w + z)(w - y)} d\beta(y) \right)_{-,w} \\ &= \int \frac{\mathbf{q}_0^T(y)\mathbb{A}(-w)\widehat{\mathbf{p}}_0(z)}{(w + z)(w - y)} d\beta(y) + \left(\int \frac{(\mathbf{q}_0^T(w) - \mathbf{q}_0^T(y))\mathbb{A}(-w)\widehat{\mathbf{p}}_0(z)}{(w + z)(w - y)} d\beta(y) \right)_{-,w}, \end{aligned}$$

where we dropped the projection sign in the first term because $\mathbb{A}(-w)$ is a polynomial of degree one. Using now the partial fraction decomposition

$$\frac{1}{(w+z)(w-y)} = \frac{1}{z+y} \left(\frac{1}{w-y} - \frac{1}{w+z} \right),$$

we get that

$$\begin{aligned} & \left(\int \frac{(\mathbf{q}_0^T(w) - \mathbf{q}_0^T(y))\mathbb{A}(-w)\widehat{\mathbf{p}}_0(z)}{(w+z)(w-y)} d\beta(y) \right)_{-,w} \\ &= - \left(\int \frac{(\mathbf{q}_0^T(-z) - \mathbf{q}_0^T(y))[\Pi, (-z - \mathbf{Y}^T)\widehat{\mathbf{L}}]\widehat{\mathbf{p}}_0(z)}{(w+z)(z+y)} d\beta(y) \right)_{-,w}. \end{aligned}$$

Observe that $(-z - \mathbf{Y}^T)\widehat{\mathbf{L}}\widehat{\mathbf{p}}_0(z) = 0$, $\mathbf{q}_0^T(-z)(-z - \mathbf{Y}^T)\widehat{\mathbf{L}} = 0$ and $\mathbf{q}_0^T(y)(-z - \mathbf{Y}^T)\widehat{\mathbf{L}} = -(y+z)\mathbf{q}_0^T(y)\widehat{\mathbf{L}}$ so

$$\begin{aligned} & \left(\int \frac{(\mathbf{q}_0^T(w) - \mathbf{q}_0^T(y))\mathbb{A}(-w)\widehat{\mathbf{p}}_0(z)}{(w+z)(w-y)} d\beta(y) \right)_{-,w} \\ &= \left(\int \frac{(\mathbf{q}_0^T(y))(z + \mathbf{Y}^T)\widehat{\mathbf{L}}\Pi\widehat{\mathbf{p}}_0(z)}{(w+z)(z+y)} d\beta(y) \right)_{-,w} = \int \frac{\mathbf{q}_0^T(y)\widehat{\mathbf{L}}\Pi\widehat{\mathbf{p}}_0(z)}{w+z} d\beta(y) = 0, \end{aligned}$$

because the β averages of $\widehat{\mathbf{q}}$ are zero. Thus

$$(w+z)\mathbf{q}_1^T(w)\Pi\mathbf{p}_0(z) = \mathbf{q}_1^T(w)\mathbb{A}(-w)\widehat{\mathbf{p}}_0(z).$$

(2, 0): Using the second Padè approximation condition and biorthogonality we easily obtain

$$\mathbf{R}_{\beta\alpha^*}^T(w)\Pi\mathbf{p}_0(z) = \frac{\mathbf{R}_{\beta\alpha^*}^T(w)\mathbb{A}(-w)\widehat{\mathbf{p}}_0(z) + 1}{w+z}.$$

Now, substituting this formula into the formula for the third approximation condition, written as in Eq. (5.5), gives:

$$\mathbf{R}_{\alpha^*\beta}^T(w)\Pi\mathbf{p}_0(z) = \frac{\mathbf{R}_{\alpha^*\beta}^T(w)\mathbb{A}(-w)\widehat{\mathbf{p}}_0(z) - 1}{w+z}.$$

Restoring the collective notation of $\mathbf{q}_a, \mathbf{p}_a$ we obtain :

$$(w+z)\mathbf{q}_2^T(w)\Pi\mathbf{p}_0(z) = \mathbf{q}_2^T(w)\mathbb{A}(-w)\widehat{\mathbf{p}}_0(z) - 1.$$

(0, 1): To compute $\mathbf{q}_0^T(w)\Pi\mathbf{p}_1(z)$ we use the Padè approximation conditions (5.6), in particular the first condition gives us:

$$\mathbf{q}_0^T(w)\Pi\mathbf{p}_0(z)W_\alpha(z) - \mathbf{q}_0^T(w)\Pi\mathbf{P}_\alpha(z) = \mathbf{q}_0^T(w)\Pi\mathbf{R}_\alpha(z).$$

We observe that this time we have to project on the negative powers of z . Thus the goal is to compute $(\mathbf{q}_0^T(w) \Pi \mathbf{p}_0(z) W_\alpha(z))_{-,z}$. We have

$$\begin{aligned} \left(\int \frac{\mathbf{q}_0^T(w) \Pi \mathbf{p}_0(z) d\alpha(x)}{z-x} \right)_{-,z} &= \left(\int \frac{\mathbf{q}_0^T(w) \mathbb{A}(-w) \widehat{\mathbf{p}}_0(z) d\alpha(x)}{(z-x)(w+z)} \right)_{-,z} \\ &= \left(\int \frac{\mathbf{q}_0^T(w) \mathbb{A}(-w) \widehat{\mathbf{p}}_0(x) d\alpha(x)}{(z-x)(w+z)} \right)_{-,z} \\ &\quad + \left(\int \frac{\mathbf{q}_0^T(w) \mathbb{A}(-w) (\widehat{\mathbf{p}}_0(z) - \widehat{\mathbf{p}}_0(x)) d\alpha(x)}{(z-x)(w+z)} \right)_{-,z}. \end{aligned}$$

We see that the first term is already *regular* in z . To treat the second term we perform the partial fraction expansion $\frac{1}{(z-x)(w+z)} = \frac{1}{w+x} [\frac{1}{z-x} - \frac{1}{w+z}]$ and observe that the term with $\frac{1}{z-x}$ does not contribute, while the second term

$$\begin{aligned} & - \left(\int \frac{\mathbf{q}_0^T(w) \mathbb{A}(-w) (\widehat{\mathbf{p}}_0(z) - \widehat{\mathbf{p}}_0(x)) d\alpha(x)}{(w+x)(w+z)} \right)_{-,z} \\ &= - \left(\int \frac{\mathbf{q}_0^T(w) \mathbb{A}(-w) (\widehat{\mathbf{p}}_0(-w) - \widehat{\mathbf{p}}_0(x)) d\alpha(x)}{(w+x)(w+z)} \right)_{-,z} = \int \frac{\mathbf{q}_0^T(w) \mathbb{A}(-w) \widehat{\mathbf{p}}_0(x) d\alpha(x)}{(w+x)(w+z)}. \end{aligned}$$

Thus

$$\mathbf{q}_0^T(w) \Pi \mathbf{p}_1(z) = \frac{\mathbf{q}_0^T(w) \mathbb{A}(-w) \widehat{L}^{-1} \mathbf{p}_1(z)}{w+z} - \frac{\mathbf{q}_0^T(w) \mathbb{A}(-w) \widehat{L}^{-1} \mathbf{p}_1(-w)}{w+z}.$$

In other words,

$$(w+z) \mathbf{q}_0^T(w) \Pi \mathbf{p}_1(z) = \mathbf{q}_0^T(w) \mathbb{A}(-w) \widehat{L}^{-1} (\mathbf{p}_1(z) - \mathbf{p}_1(-w)).$$

More explicitly, the second term above can be rewritten as

$$-\mathbf{q}_0^T(w) \mathbb{A}(-w) \widehat{L}^{-1} \mathbf{p}_1(-w) = \mathbf{q}_0^T(w) \Pi \int \mathbf{p}(x) d\alpha(x).$$

On the other hand

$$\begin{aligned} \mathbf{q}_0^T(w) \mathbb{A}(-w) \iint \frac{\widehat{\mathbf{p}}(x) d\alpha(x) d\beta(y)}{\beta_0(x+y)} &= \mathbf{q}_0^T(w) \Pi \iint \frac{(w+x) \mathbf{p}(x) d\alpha(x) d\beta(y)}{\beta_0(x+y)} \\ &= \mathbf{q}_0^T(w) \Pi \int \mathbf{p}(x) d\alpha(x) + \mathbf{q}_0^T(w) \Pi \iint \frac{(w-y) \mathbf{p}(x) d\alpha(x) d\beta(y)}{\beta_0(x+y)}. \end{aligned}$$

Now the second term $\mathbf{q}_0^T(w) \Pi \iint \frac{(w-y) \mathbf{p}(x) d\alpha(x) d\beta(y)}{\beta_0(x+y)} = 0$ because $\mathbf{q}_0^T(w) \Pi \langle \mathbf{p}(x) | \bullet \rangle$ is a projector on polynomials of degree $\leq n-1$ and thus $w \mathbf{q}_0^T(w) \Pi \langle \mathbf{p}(x) | 1 \rangle - \mathbf{q}_0^T(w) \Pi \langle \mathbf{p}(x) | y \rangle = w - w = 0$, hence

$$(w+z) \mathbf{q}_0^T(w) \Pi \mathbf{p}_1(z) = \mathbf{q}_0^T(w) \mathbb{A}(-w) \widehat{\mathbf{p}}_1(z),$$

where $\widehat{\mathbf{p}}_1(z) = \widehat{L}^{-1} (\mathbf{p}_1(z) + \frac{1}{\beta_0} \langle \mathbf{p} | 1 \rangle)$ as advertised earlier.

(1, 1): We use again the Padé approximation conditions (5.6), this time multiplying on the left by $\mathbf{q}_1^T(w) \Pi$ and projecting on the negative powers of z , to obtain:

$$\left(\mathbf{q}_1^T(w) \Pi \mathbf{p}_0(z) W_\alpha(z) \right)_{-,z} = \mathbf{q}_1^T(w) \Pi \mathbf{p}_1(z).$$

With the help of the result for the (0, 1) entry, after carrying out the projection, we obtain

$$(w + z)\mathbf{q}_1^T(w)\Pi\mathbf{p}_1(z) = \mathbf{q}_1^T(w)\mathbb{A}(-w)\widehat{\mathbf{p}}_1(z) + \mathbf{q}_1^T(w)\mathbb{A}(-w)\left(\int\frac{\widehat{\mathbf{p}}(x)d\alpha(x)}{w+x}-\frac{1}{\beta_0}\langle\widehat{\mathbf{p}}|1\rangle\right).$$

We claim that

$$\mathbf{q}_1^T(w)\mathbb{A}(-w)\left(\int\frac{\widehat{\mathbf{p}}(x)d\alpha(x)}{w+x}-\frac{1}{\beta_0}\langle\widehat{\mathbf{p}}|1\rangle\right) = -1.$$

Indeed, the left hand side of the equation equals:

$$\begin{aligned} \frac{1}{\beta_0}\mathbf{q}_1^T(w)\Pi\int\int\frac{(y-w)\mathbf{p}(x)d\alpha(x)d\beta(y)}{x+y} &= \frac{1}{\beta_0}\int\frac{\mathbf{q}_0^T(\xi)}{w-\xi}\Pi(\langle\mathbf{p}|y\rangle-w\langle\mathbf{p}|1\rangle)d\beta(\xi) \\ &= \frac{1}{\beta_0}\int\frac{\xi-w}{w-\xi}d\beta(\xi) = -1. \end{aligned}$$

Thus

$$(w + z)\mathbf{q}_1^T(w)\Pi\mathbf{p}_1(z) = \mathbf{q}_1^T(w)\mathbb{A}(-w)\widehat{\mathbf{p}}_1(z) - 1.$$

(2, 1): This time we use projections in both variables, one at a time, and compare the results. First, let us use the projections in z . Thus

$$\mathbf{q}_2^T(w)\Pi\mathbf{p}_1(z) = \left(\mathbf{q}_2^T(w)\Pi\mathbf{p}_0(z)W_\alpha(z)\right)_{-,z}.$$

Carrying out all the projections we obtain an expression of the form:

$$\mathbf{q}_2^T(w)\Pi\mathbf{p}_1(z) = \frac{\mathbf{q}_2^T(w)\mathbb{A}(-w)\widehat{\mathbf{p}}_1(z)}{w+z} - \frac{W_\alpha(z) + F(w)}{w+z}.$$

Observe that, since $\mathbf{q}_2^T(w)$ is $\mathcal{O}(1/w)$ and the first term on the right is much smaller, $F(w) = \mathcal{O}(1)$. More precisely, by comparing the terms at $1/w$ on both sides, we conclude that in fact, $F(w) = \mathcal{O}(1/w)$. Now, we turn to the projection in w , resulting in an expression of the form:

$$\mathbf{q}_2^T(w)\Pi\mathbf{p}_1(z) = \frac{\mathbf{q}_2^T(w)\mathbb{A}(-w)\widehat{\mathbf{p}}_1(z)}{w+z} - \frac{W_{\alpha^*}(w) + G(z)}{w+z}.$$

This, and the fact that $F(w) = \mathcal{O}(1/w)$, implies that $F(w) = W_{\alpha^*}(w)$, $G(z) = W_\alpha(z)$. Hence

$$(w + z)\mathbf{q}_2^T(w)\Pi\mathbf{p}_1(z) = \mathbf{q}_2^T(w)\mathbb{A}(-w)\widehat{\mathbf{p}}_1(z) - (W_\alpha(z) + W_{\alpha^*}(w)).$$

(0, 2): We use the projection in the z variable and the fact that by the Padè approximation condition (5.5), after exchanging α with β , $\mathbf{p}_2(z) = \mathbf{p}_1(z)W_{\beta^*}(z) - \mathbf{R}_{\alpha\beta^*}(z)$. Using the result for the (0, 1) entry we obtain:

$$\mathbf{q}_0^T(w)\Pi\mathbf{p}_2(z) = \frac{\mathbf{q}_0^T(w)\mathbb{A}(-w)\mathbf{p}_1(z)W_{\beta^*}(z)}{w+z} - \left(\frac{\mathbf{q}_0^T(w)\mathbb{A}(-w)\mathbf{p}_0(z)W_{\alpha\beta^*}(z)}{w+z}\right)_{-,z}.$$

Carrying out the projection and reassembling terms according to the definition of $\widehat{\mathbf{p}}_2(z)$ we obtain:

$$\mathbf{q}_0^T(w)\Pi\mathbf{p}_2(z) = \frac{\mathbf{q}_0^T(w)\mathbb{A}(-w)\widehat{\mathbf{p}}_2(z)}{w+z} - \frac{\mathbf{q}_0^T(w)\Pi\langle\mathbf{p}_0|1\rangle}{w+z} = \frac{\mathbf{q}_0^T(w)\mathbb{A}(-w)\widehat{\mathbf{p}}_2(z)}{w+z} - \frac{1}{w+z}.$$

(1, 2): We use the projection in the z variable and the Padè approximation condition $\mathbf{p}_2 = \mathbf{p}_1(z)W_{\beta^*}(z) - \mathbf{R}_{\alpha\beta^*}(z)$. Consequently,

$$\begin{aligned} \mathbf{q}_1^T(w) \Pi \mathbf{p}_2(z) &= \mathbf{q}_1^T(w) \Pi \mathbf{p}_1(z) W_{\beta^*}(z) - \mathbf{q}_1^T(w) \Pi \mathbf{R}_{\alpha\beta^*}(z) \\ &= \left(\frac{\mathbf{q}_1^T(w) \mathbb{A}(-w) \widehat{\mathbf{p}}_1(z) - 1}{w+z} \right) W_{\beta^*}(z) - \left(\mathbf{q}_1^T(w) \Pi \mathbf{p}_0(z) W_{\alpha\beta^*}(z) \right)_{-,z}. \end{aligned}$$

Using the existing identities and carrying out the projection in the second term we obtain:

$$(w+z)\mathbf{q}_1^T(w) \Pi \mathbf{p}_2(z) = \mathbf{q}_1^T(w) \mathbb{A}(-w) \widehat{\mathbf{p}}_2(z) - W_{\beta^*}(z) - W_{\beta}(w).$$

(2, 2): The computation is similar to the one for (1, 2) entry; we use both projections. The projection in the z variable gives:

$$\mathbf{q}_2^T(w) \Pi \mathbf{p}_2(z) = \frac{\mathbf{q}_2^T(w) \mathbb{A}(-w) \widehat{\mathbf{p}}_2(z)}{w+z} + \frac{F(w) - (W_{\alpha^*}(w) + W_{\alpha}(z))W_{\beta^*}(z) + W_{\alpha\beta^*}(z)}{w+z}.$$

On the other hand, carrying out the projection in the w variable we obtain:

$$\mathbf{q}_2^T(w) \Pi \mathbf{p}_2(z) = \frac{\mathbf{q}_2^T(w) \mathbb{A}(-w) \widehat{\mathbf{p}}_2(z)}{w+z} + \frac{G(z) - (W_{\beta}(w) + W_{\beta^*}(z))W_{\alpha^*}(w) + W_{\beta\alpha^*}(w)}{w+z}.$$

Upon comparing the two expressions and using Lemma 5.2 we obtain $F(w) = -W_{\alpha^*\beta}(w)$, hence

$$\begin{aligned} (w+z)\mathbf{q}_2^T(w) \Pi \mathbf{p}_2(z) &= \mathbf{q}_2^T(w) \mathbb{A}(-w) \widehat{\mathbf{p}}_2(z) - W_{\alpha^*\beta}(w) \\ &\quad - (W_{\alpha^*}(w) + W_{\alpha}(z))W_{\beta^*}(z) + W_{\alpha\beta^*}(z) \\ &= \mathbf{q}_2^T(w) [\Pi, \mathbb{A}(-w)] \widehat{\mathbf{p}}_2(z) - (W_{\alpha^*}(w)W_{\beta^*}(z) + W_{\alpha^*\beta}(w) + W_{\beta^*\alpha}(z)), \end{aligned}$$

where in the last step we used again Lemma 5.2. \square

References

[1] A. Aptekarev, V. Kaliaguine, J. Van Iseghem, The genetic sums’ representation for the moments of a system of Stieltjes functions and its application, *Constr. Approx.* 16 (4) (2000) 487–524.
 [2] A.I. Aptekarev, G. Lopes Lagomasino, I.A. Rocha, Asymptotic behavior of the ratio of Hermite-Padé polynomials for Nikishin systems, *Mat. Sb.* 196 (8) (2005) 3–20.
 [3] R. Beals, D. Sattinger, J. Szmigielski, Multipeakons and the classical moment problem, *Adv. Math.* 154 (2000) 229–257.
 [4] R. Beals, D.H. Sattinger, J. Szmigielski, Multi-peakons and a theorem of Stieltjes, *Inverse Problems* 15 (1) (1999) L1–L4.
 [5] M. Bertola, Biorthogonal polynomials for two-matrix models with semiclassical potentials, *J. Approx. Theory* 144 (2) (2007) 162–212. CRM preprint CRM-3205 (2005).
 [6] M. Bertola, Moment determinants as isomonodromic tau functions, *Nonlinearity* 22 (1) (2009) 29–50.
 [7] M. Bertola, B. Eynard, J. Harnad, Duality, biorthogonal polynomials and multi-matrix models, *Comm. Math. Phys.* 229 (1) (2002) 73–120.
 [8] M. Bertola, B. Eynard, J. Harnad, Differential systems for biorthogonal polynomials appearing in 2-matrix models and the associated Riemann-Hilbert problem, *Comm. Math. Phys.* 243 (2) (2003) 193–240.
 [9] M. Bertola, B. Eynard, J. Harnad, The duality of spectral curves that arises in two-matrix models, *Teoret. Mat. Fiz.* 134 (1) (2003) 32–45.
 [10] M. Bertola, M. Gekhtman, J. Szmigielski, Cubic String Boundary Value Problems and Cauchy Biorthogonal polynomials [arXiv:0904.2188](https://arxiv.org/abs/0904.2188).
 [11] M. Bertola, M. Gekhtman, J. Szmigielski, Strong asymptotics of Cauchy biorthogonal polynomials. 2008 (in preparation).

- [12] M. Bertola, M. Gekhtman, J. Szmigielski, The Cauchy two-matrix model, *Comm. Math. Phys.* 287 (3) (2009) 983–1014.
- [13] R. Camassa, D.D. Holm, An integrable shallow water equation with peaked solitons, *Phys. Rev. Lett.* 71 (11) (1993) 1661–1664.
- [14] C.W. Cryer, The LU -factorization of totally positive matrices, *Linear Algebra Appl.* 7 (1973) 83–92.
- [15] C.W. Cryer, Some properties of totally positive matrices, *Linear Algebra Appl.* 15 (1) (1976) 1–25.
- [16] A. Degasperis, M. Procesi, Asymptotic integrability, in: A. Degasperis, G. Gaeta (Eds.), *Symmetry and Perturbation Theory* (Rome, 1998), World Scientific Publishing, River Edge, NJ, 1999, pp. 23–37.
- [17] M. Duits, A. Kuijlaars, Universality in the two matrix model: A Riemann–Hilbert steepest descent analysis, *Comm. Pure Appl. Math.* 62 (2009) 1076–1153.
- [18] N.M. Ercolani, K.T.-R. McLaughlin, Asymptotics and integrable structures for biorthogonal polynomials associated to a random two-matrix model, *Physica D* 152/153 (2001) 232–268. *Advances in nonlinear mathematics and science.*
- [19] B. Eynard, M.L. Mehta, Matrices coupled in a chain. I. Eigenvalue correlations, *J. Phys. A* 31 (19) (1998) 4449–4456.
- [20] U. Fidalgo Prieto, J. Illán, G. López Lagomasino, Hermite-Padé approximation and simultaneous quadrature formulas, *J. Approx. Theory* 126 (2) (2004) 171–197.
- [21] F.P. Gantmacher, M.G. Krein, *Oscillation Matrices and Kernels and Small Vibrations of Mechanical Systems*, revised edition, AMS Chelsea Publishing, Providence, RI, 2002, Translation based on the 1941 Russian original, edited and with a preface by Alex Eremenko.
- [22] M. Jimbo, T. Miwa, K. Ueno, Monodromy preserving deformation of linear ordinary differential equations with rational coefficients. i. general theory, *Physica D* 2 (2) (1981) 306–352.
- [23] S. Karlin, *Total positivity*. Vol. I, Stanford University Press, Stanford, Calif, 1968.
- [24] A.B.J. Kuijlaars, K.T.-R. McLaughlin, A Riemann-Hilbert problem for biorthogonal polynomials, *J. Comput. Appl. Math.* 178 (1-2) (2005) 313–320.
- [25] H. Lundmark, J. Szmigielski, Multi-peakon solutions of the Degasperis–Procesi equation, *Inverse Problems* 19 (2003) 1241–1245.
- [26] H. Lundmark, J. Szmigielski, Degasperis-Procesi peakons and the discrete cubic string, *IMRP Int. Math. Res. Pap.* (2) (2005) 53–116.
- [27] M.L. Mehta, *Random Matrices*, third edition, in: *Pure and Applied Mathematics* (Amsterdam), vol. 142, Elsevier/Academic Press, Amsterdam, 2004.
- [28] E.M. Nikishin, V.N. Sorokin, *Rational Approximations and Orthogonality*, in: *Translations of Mathematical Monographs*, vol. 92, American Mathematical Society, Providence, RI, 1991, Translated from the Russian by Ralph P. Boas.
- [29] E.M. Nikišin, Simultaneous Padé approximants, *Mat. Sb. (N.S.)* 113(155) (4(12)) (1980) 499–519, 637.
- [30] V.N. Sorokin, Simultaneous approximation of several linear forms, *Vestnik Moskov. Univ. Ser. I Mat. Mekh.* (1) (1983) 44–47, 101.
- [31] V.N. Sorokin, Hermite-Padé approximants of polylogarithms, *Izv. Vyssh. Uchebn. Zaved. Mat.* (5) (1994) 49–59.
- [32] Vladimir N. Sorokin, Jeannette Van Iseghem, Algebraic aspects of matrix orthogonality for vector polynomials, *J. Approx. Theory* 90 (1) (1997) 97–116.
- [33] W. Van Assche, J.S. Geronimo, A.B.J. Kuijlaars, Riemann-Hilbert problems for multiple orthogonal polynomials, in: *Special functions 2000: current perspective and future directions* (Tempe, AZ), in: *NATO Sci. Ser. II Math. Phys. Chem.*, vol. 30, Kluwer Acad. Publ, Dordrecht, 2001, pp. 23–59.