© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

Chapter One

Introduction

1.1 MOTIVATING APPLICATIONS

The main aim of this monograph is to deduce asymptotic properties of polynomials that are orthogonal with respect to pure point measures supported on finite sets and use them to establish various statistical properties of discrete orthogonal polynomial ensembles, a special case of which yields new results for a random rhombus tiling of a large hexagon. Throughout this monograph, the polynomials that are orthogonal with respect to pure point measures will be referred to simply as *discrete orthogonal polynomials*. We begin by introducing several applications in which asymptotics of discrete orthogonal polynomials play an important role.

1.1.1 Discrete orthogonal polynomial ensembles

In order to illustrate some concrete applications of discrete orthogonal polynomials and also to provide some motivation for the scalings we study in this book, we give here a brief introduction to discrete orthogonal polynomial ensembles. More details can be found in Chapter 3.

General theory

In the theory of random matrices [Meh91, Dei99], the main object of study is the joint probability distribution of the eigenvalues. In unitary-invariant matrix ensembles, the eigenvalues are distributed as a Coulomb gas in the plane confined on the real line at the inverse temperature $\beta = 2$ subject to an external field. In recent years, various problems in probability theory have turned out to be representable in terms of the same Coulomb gas system with the condition that the particles are further confined to a discrete set. Such a system is called a *discrete orthogonal polynomial ensemble*. More precisely, consider the joint probability distribution of finding k particles at positions $x_1, \dots, x_k$ in a discrete set X to be given by the following expression (we are using the symbol $\mathbb{P}(\text{event})$ to denote the probability of an event):

$$\begin{aligned} p^{(k)}(x_1, \dots, x_k) :=& \mathbb{P}(\text{there are particles at each of the nodes } x_1, \dots, x_k) \\ =& \frac{1}{Z_k} \prod_{1 \le i < j \le k} (x_i - x_j)^2 \cdot \prod_{j=1}^{k} w(x_j)\,, \end{aligned} \tag{1.1}$$

where Z_k is a normalization constant (or *partition function*) chosen so that

$$\sum_{\substack{x_1 < \dots < x_k \\ x_j \in X}} p^{(k)}(x_1, \dots, x_k) = 1\,.$$

Note that the particles are all indistinguishable from each other.

Discrete orthogonal polynomial ensembles arise in a number of specific contexts (see, for example, [BorO01, Joh00, Joh01, Joh02]), with particular choices of the weight function $w(\cdot)$ related (in cases we are aware of) to classical discrete orthogonal polynomials. For instance:

- The Meixner weight

$$w(x) = \binom{x + M - N}{x} q^x\,, \qquad \text{for } x = 0, 1, 2, \dots,$$

For general queries, contact webmaster@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

arises in the directed last-passage site percolation model in the two-dimensional finite lattice $\mathbb{Z}_M \times \mathbb{Z}_N$ with independent geometric random variables as passage times for each site [Joh00]. The rightmost node occupied by a particle in the ensemble, $x_{\max} := \max_j x_j$, is a random variable having the same distribution as the last passage time to travel from the site $(0,0)$ to the site $(M-1, N-1)$.

- The Charlier weight
$$w(x) = \frac{t^x}{x!}, \qquad \text{for } x = 0, 1, 2, \ldots,$$
arises in the longest random word problem [Joh01].

- The Krawtchouk weight
$$w(x) = \binom{N-1}{x} p^x q^{N-1-x}, \qquad \text{for } x = 0, 1, \ldots, N-1,$$
arises in the random domino tiling of the Aztec diamond [Joh01, Joh02].

- The Hahn weight
$$w(x) = \binom{x+\alpha}{x}\binom{N-1+\beta-x}{N-1-x}, \qquad \text{for } x = 0, 1, 2, \ldots, N-1,$$
arises in the random rhombus tiling of a hexagon [Joh01, Joh02]. See also §3.4 for more details.

The first two cases (Meixner and Charlier) are examples of the *Schur measure* [BorO01, Oko01] on the set of partitions. On the other hand, in special limiting cases the Meixner and Charlier ensembles both become the *Plancherel measure*, which describes the longest increasing subsequence of a random permutation[1] [BaiDJ99, BorOO00, Joh01]. Clearly it would be of some theoretical interest to determine properties of the ensembles that are more or less independent of the particular choice of weight function, at least within some class. Such properties are said to support the conjecture of *universality* within the class of weight functions under consideration.

As is well known in random matrix theory [Meh91, TraW98], many quantities such as correlation functions and gap probabilities are expressible in terms of the polynomials that are orthogonal with respect to the weight w. In a similar way, the asymptotic study of discrete orthogonal polynomial ensembles is translated to the asymptotic study of discrete orthogonal polynomials. In many applications, the parameters of the weight are varying, and at the same time the location of the particle of interest also varies. For example, in the Hahn case, the interesting case turns out to be when $N \to \infty$ and the location of a particle is scaled as $x = aN + \xi$ or $x = aN + \xi N^{1/3}$ for some constant $a > 0$. Equivalently, upon scaling by $1/N$, the discrete set X/N asymptotically fills out an interval, while x/N scales as $x/N = a + \xi/N$ or $x/N = a + \xi N^{-2/3}$. Under this scaling limit (Plancherel-Rotach asymptotics), the asymptotics of Meixner, Charlier, and Krawtchouk polynomials were obtained using integral representations [IsmS98, Joh00, Joh02]. However, even though the Hahn polynomials are also classical polynomials, their integral representation does not seem to be so straightforward to analyze asymptotically using the classical steepest-descent method, and the asymptotics have not yet been obtained. The subject of this book is a general class of weights that contains Krawtchouk and Hahn weights (and a lot more) as special cases. For these general weights, we will compute the asymptotics of the associated discrete orthogonal polynomials for all values of the variable z in the complex plane. By specializing to suitable scaling regimes, the desired asymptotics of the associated discrete orthogonal polynomial ensembles are obtained.

Random tilings

The discrete orthogonal polynomial ensemble with the Hahn weight represents a probability distribution for certain events in a random rhombus tiling of a hexagon, and new results on the asymptotics of random

For general queries, contact webmaster@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

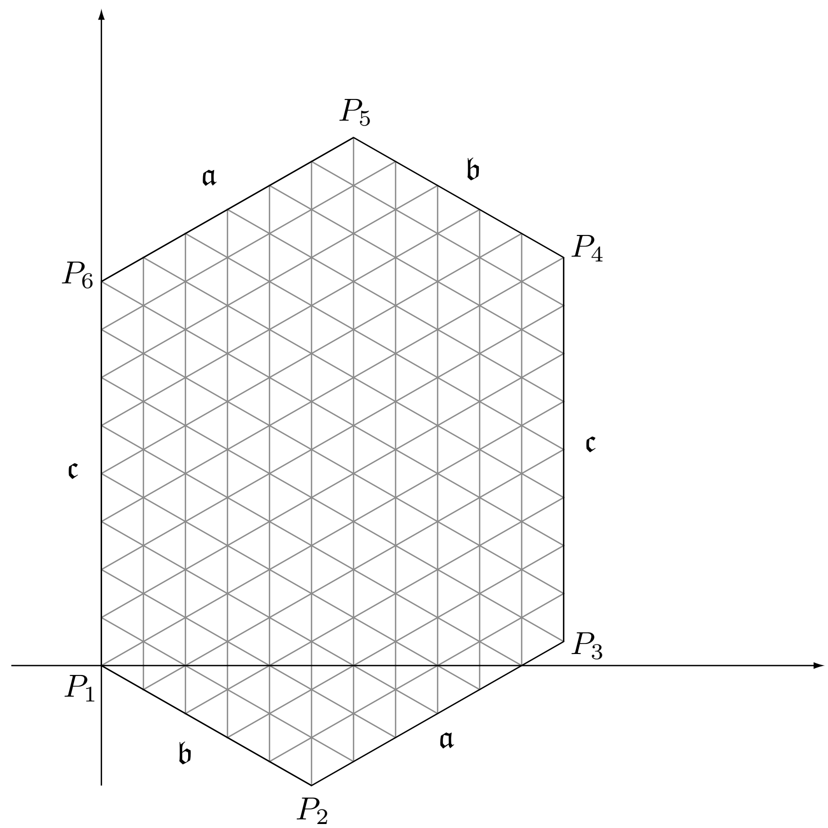

Figure 1.1 *The* $\mathfrak{abc}$*-hexagon with vertices* $P_1, \ldots, P_6$*, and the lattice* $\mathcal{L}$*.*

tilings will be presented in Chapter 3 using this connection. Here, we briefly introduce the subject of random rhombus tilings of a hexagon.

Let $\mathfrak{a}$, $\mathfrak{b}$, and $\mathfrak{c}$ be positive integers and consider the hexagon illustrated in Figure 1.1 having the following vertices (written as points in the complex plane):

$$\begin{aligned} P_1 &= 0\,, & P_2 &= \mathfrak{b}e^{-i\pi/6}\,, & P_3 &= P_2 + \mathfrak{a}e^{i\pi/6}\,, \\ P_4 &= P_3 + i\mathfrak{c}\,, & P_5 &= P_4 + \mathfrak{b}e^{5\pi i/6}\,, & P_6 &= i\mathfrak{c}\,. \end{aligned}$$

All interior angles of this hexagon are equal and measure $2\pi/3$ radian, and the lengths of the sides are, starting with the side (P_1, P_2) and proceeding in counterclockwise order, $\mathfrak{b}, \mathfrak{a}, \mathfrak{c}, \mathfrak{b}, \mathfrak{a}, \mathfrak{c}$. We call this the $\mathfrak{abc}$-*hexagon*. Denote by $\mathcal{L}$ the part of the set of lattice points

$$\left\{ke^{i\pi/6} + je^{-i\pi/6}\right\}_{k,j\in\mathbb{Z}} = \left\{\frac{\sqrt{3}}{2}n + \frac{i}{2}n'\right\}_{n,n'\in\mathbb{Z}}$$

that lies within the hexagon, including the sides (P_6, P_1), (P_1, P_2), (P_2, P_3), and (P_3, P_4) but excluding the sides (P_4, P_5) and (P_5, P_6). The lattice $\mathcal{L}$ can also be seen in Figure 1.1.

Consider covering the $\mathfrak{abc}$-hexagon with rhombus tiles having sides of unit length. The rhombus tiles come in three different types (orientations), which we refer to as type I, type II, and type III as shown in Figure 1.2. Tiles of types I and II are sometimes collectively called *horizontal rhombi*, while tiles of type III

[1] Strictly speaking, this is not a discrete orthogonal polynomial ensemble in the sense we have described because as a consequence of the limiting process involved in the definition, the number of particles k is not fixed in advance but is itself a random variable.

For general queries, contact webmaster@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

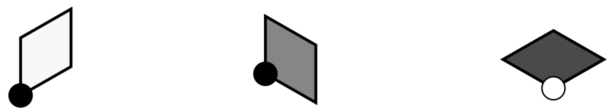

Figure 1.2 *The three types of rhombus tiles; the position of each tile is indicated with a dot.*

are sometimes called *vertical rhombi*. The position of each rhombus tile in the hexagon is a specific lattice point in $\mathcal{L}$ defined as indicated in Figure 1.2.

MacMahon's formula [Mac60] gives the total number of all possible rhombus tilings of the $\mathfrak{abc}$-hexagon as the expression

$$\prod_{i=1}^{\mathfrak{a}} \prod_{j=1}^{\mathfrak{b}} \prod_{k=1}^{\mathfrak{c}} \frac{i+j+k-1}{i+j+k-2} .$$

Consider the set of all rhombus tilings equipped with uniform probability. Hence we choose a tiling of the $\mathfrak{abc}$-hexagon at random. It is of some current interest to determine the behavior of various corresponding statistics of this ensemble in the limit as $\mathfrak{a}, \mathfrak{b}, \mathfrak{c} \to \infty$.

In the scaling limit of $n \to \infty$, where

$$\mathfrak{a} = \mathfrak{A}n\,, \qquad \mathfrak{b} = \mathfrak{B}n\,, \qquad \mathfrak{c} = \mathfrak{C}n\,, \tag{1.2}$$

with fixed $\mathfrak{A}, \mathfrak{B}, \mathfrak{C} > 0$, the regions near the six corners are *frozen* or *polar zones* (*i.e.*, regions in which only one type of tile is present), while toward the center of the hexagon is a *temperate zone* (*i.e.*, a region containing all three types of tiles). The random tiling shown in Figure 1.3 dramatically illustrates the two types of regions, and the asymptotically sharp nature of the boundary between them, the *Arctic circle*. Indeed, it was shown by Cohn, Larsen, and Propp [CohLP98] that in such a limit, upon scaling by $1/n$, the expected shape of the boundary separating the polar zones from the temperate zone is given by the inscribed ellipse. The next interesting problem would be to compute the limiting fluctuation of the boundary. While this problem had remained unsolved until our work, an analogous problem had been solved in the context of rectangle tilings of Aztec diamonds; it is proved in [Joh01] that the fluctuation of the boundary between the polar zones and the temperate zone in the Aztec diamond tiling model is governed (in a proper scaling limit) by the Tracy-Widom law in random matrix theory [TraW94]. Indeed, Johansson [Joh01, Joh02] expressed the induced probability for certain configurations of rectangles in an Aztec diamond and of rhombi in the $\mathfrak{abc}$-hexagon in terms of particular discrete orthogonal polynomial ensembles. The weights corresponding to the Aztec diamond are Krawtchouk weights, and those corresponding to the $\mathfrak{abc}$-hexagon are Hahn or associated Hahn weights. By applying the classical steepest-descent method to the integral representation of the Krawtchouk polynomials, Johansson obtained the Tracy-Widom distribution for the Aztec diamond. One of the results implied by our analysis of general discrete orthogonal polynomial ensembles (see Theorem 3.14) is that the same Tracy-Widom law holds for rhombus tilings of the $\mathfrak{abc}$-hexagon. More details of the connection between the statistics of rhombus tilings and the Hahn discrete orthogonal polynomial ensembles are discussed along with our asymptotic results in §3.4 and, specifically, §3.4.2.

1.1.2 The continuum limit of the Toda lattice

In Flaschka's variables, the Toda lattice equations are the following coupled nonlinear ordinary differential equations:

$$\frac{da_k}{dt} = 2b_k^2 - 2b_{k-1}^2 \tag{1.3}$$

and

$$\frac{db_k}{dt} = (a_{k+1} - a_k)b_k\,. \tag{1.4}$$

For general queries, contact webmaster@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

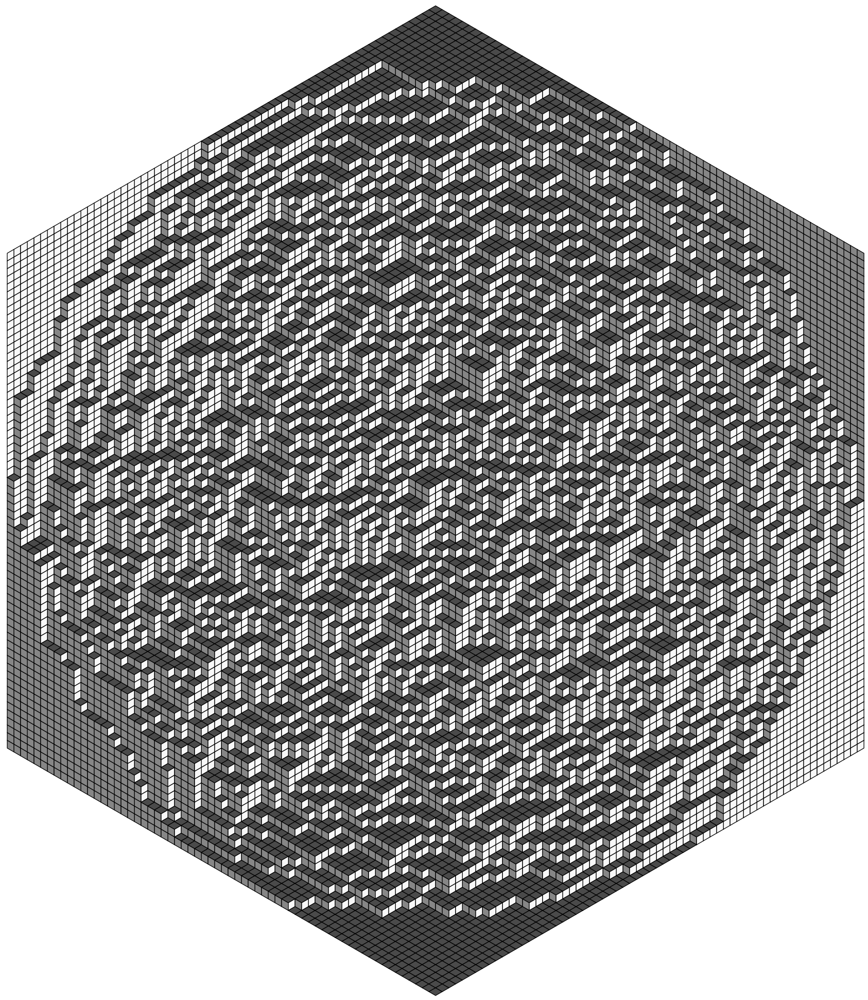

Figure 1.3 *A rhombus tiling of a large* $\mathfrak{abc}$*-hexagon with* $\mathfrak{a} = \mathfrak{b} = \mathfrak{c} = 64$. *The tiles are shaded as in Figure 1.2. Image provided by J. Propp.*

Here k is an integer index. The finite Toda lattices arise by "cutting the chain", with the imposition of boundary conditions at, say, $k = 0$ and $k = N - 1$. For one type of boundary condition, we may assume that with $a_k = a_{N,k}$ and $b_k = b_{N,k}$, (1.4) holds for $k = 0, \ldots, N - 2$ and (1.3) holds for $k = 1, \ldots, N - 2$, while

$$\frac{da_{N,0}}{dt} = 2b_{N,0}^2 \qquad \text{and} \qquad \frac{da_{N,N-1}}{dt} = -2b_{N,N-2}^2 \,. \tag{1.5}$$

Thus we obtain a first-order system of nonlinear differential equations for unknowns $a_{N,0}(t), \ldots, a_{N,N-1}(t)$ and $b_{N,0}(t), \ldots, b_{N,N-2}(t)$.

One way to view the Toda lattice equations (1.3) and (1.4) is as a numerical scheme for integrating the hyperbolic system

$$\frac{\partial A}{\partial T} = 4B\frac{\partial B}{\partial c} \qquad \text{and} \qquad \frac{\partial B}{\partial T} = B\frac{\partial A}{\partial c} \,. \tag{1.6}$$

Here $T = t/N$ is a rescaled time, $1/N$ is a small lattice spacing, and $c = k/N$. Of course, as (1.6) is a nonlinear hyperbolic system of partial differential equations, initially smooth solutions of (1.6) can develop

For general queries, contact webmaster@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

shocks (derivative singularities) in finite time. When this occurs, the Toda lattice equations should no longer be viewed as a viable numerical method for studying weak solutions of (1.6), as rapid oscillations develop in the numerical solution that are (it turns out) inconsistent even in an averaged sense with viscosity solutions of the hyperbolic system. See Figures 1.4 and 1.5.

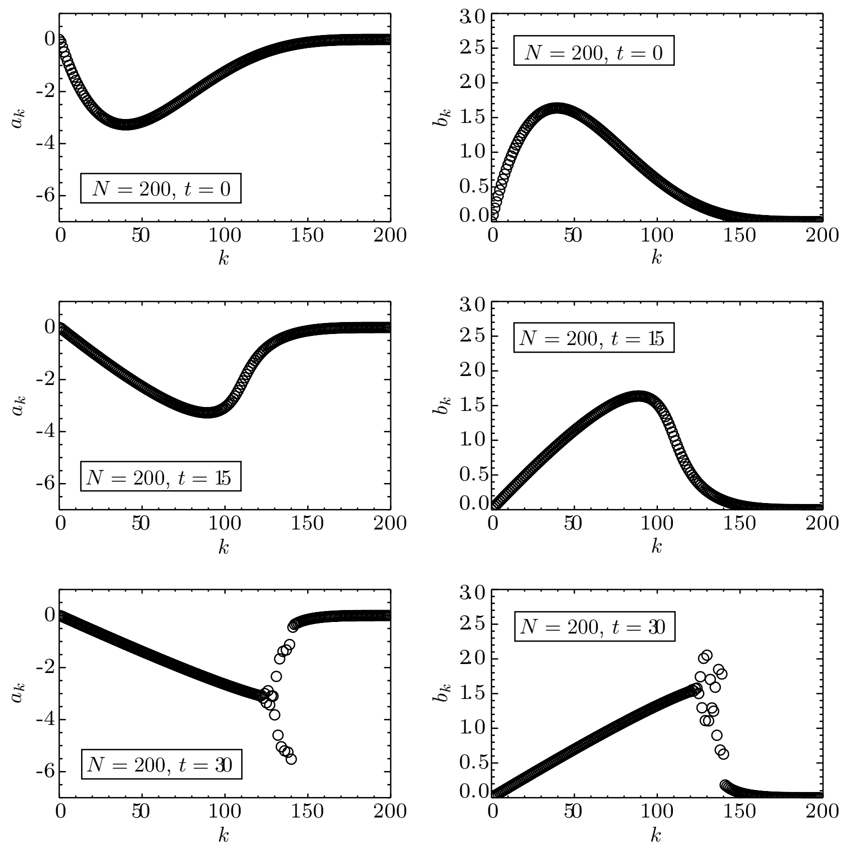

Figure 1.4 *A solution of the Toda lattice equations with smooth initial data on the interval* $c = k/N \in [0, 1]$ *sampled over* $N = 200$ *points.*

The proof of convergence of the Toda numerical scheme for (1.6), for times T less than the shock time, is one of the results of the analysis carried out by Deift and McLaughlin [DeiM98]. Their analysis of the Toda lattice equations with smooth initial data goes beyond the shock time and characterizes precisely the average properties of the rapid oscillations that subsequently occur. Indeed, it turns out that, while rapidly varying (on the fast time scale t and on the grid scale k), these oscillations can nonetheless be characterized by slowly varying quantities that satisfy an enlarged set of hyperbolic nonlinear partial differential equations (the Whitham equations) generalizing the hyperbolic system (1.6). The continuum limit of the Toda lattice has also been considered from a geometric point of view [BloGPU03] and from the point of view of orthogonal polynomials [AptV01].

The method used in [DeiM98] exploits the fact that the Toda lattice equations (1.3) and (1.4) comprise a completely integrable system. Specifically, this fact implies closed-form formulae for $a_{N,k}(t)$ and $b_{N,k}(t)$ in terms of initial data via ratios of Hankel-type determinants. Deift and McLaughlin analyzed these determinantal formulae in the continuum limit $N \to \infty$ for initial data sampling fixed smooth functions $A(c)$ and $B(c) > 0$ given on the interval $c \in [0, 1]$. Using the Lax-Levermore method, they showed that the large-N asymptotics are characterized by the solution of a constrained variational problem for a certain extremal measure on an interval, which they solved by converting it into a scalar Riemann-Hilbert problem.

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

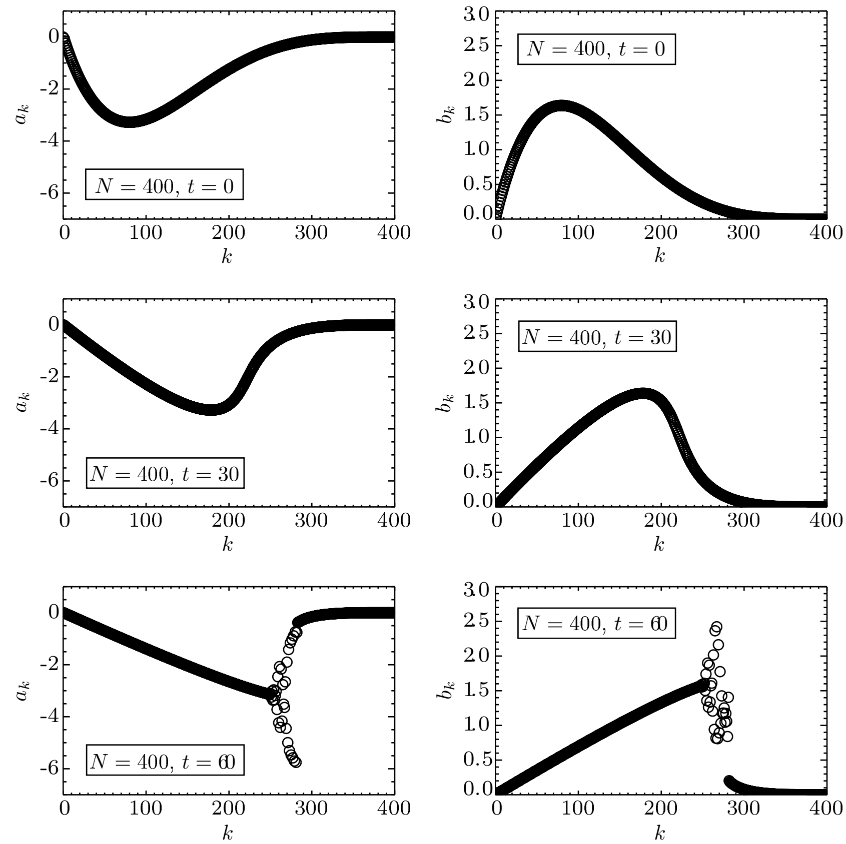

Figure 1.5 *A solution of the Toda lattice equations with smooth initial data on the interval* $c = k/N \in [0, 1]$ *sampled over* $N = 400$ *points.*

The number of subintervals where the extremal measure is unconstrained (this number depends generally on c and T) turns out to be related to the size of the system of Whitham equations needed to describe the slowly varying features of the microscopic oscillations. In particular, when there is only one such subinterval, the limit $N \to \infty$ is strong, and the hyperbolic system (1.6) governs the limit. Shock formation corresponds to the splitting of one unconstrained subinterval into two. Once this occurs, the analysis in [DeiM98] establishes the continuum limit only in a weak (averaged) sense. One of the applications that we will describe in this monograph is the strengthening of the asymptotics obtained in [DeiM98] after the shock time; we will obtain strong (locally uniform) asymptotics for the $a_{N,k}(t)$ and $b_{N,k}(t)$ even in the oscillatory region appearing after a shock has occurred, for example, in the irregular regions of the final plots in Figures 1.4 and 1.5.

Dynamical stability of solutions of the Toda lattice equations may be studied by means of the linearized Toda equations. Substituting $a_{N,k}(t) \to a_{N,k}(t) + \hat{a}_{N,k}(t)$ and $b_{N,k}(t) \to b_{N,k}(t)(1 + \hat{b}_{N,k}(t))$ into (1.3) and (1.4) and keeping only the linear terms gives

$$\frac{d\hat{a}_{N,k}}{dt} = 4b_{N,k}^2\hat{b}_{N,k} - 4b_{N,k-1}^2\hat{b}_{N,k-1} \tag{1.7}$$

and

$$\frac{d\hat{b}_{N,k}}{dt} = \hat{a}_{N,k+1} - \hat{a}_{N,k}\,. \tag{1.8}$$

Linear stability analysis is interesting in the large-N limit both before and after shock formation. Unfortunately, the detailed analysis in [DeiM98] does not directly provide information about a complete basis

For general queries, contact webmaster@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

for solutions of the linearized equations (1.7) and (1.8) corresponding to given smooth functions $A(c)$ and $B(c)$. One of the consequences of the analysis described in this book is an asymptotic description of the solutions of the linearized Toda equations in the continuum limit both before and after shock time. In some applications the linearized problem can have meaning as a linear system in its own right with prescribed time-dependent potentials $a_{N,k}(t)$ and $b_{N,k}(t)$; it is mathematically convenient then to imagine that the potentials originate from a Toda solution, in which case there is a complete basis for time-dependent modes (*i.e.*, a time-dependent spectral transform) for the linear problem. This approach is described for other completely integrable systems in [MilA98] and [MilC01], with corresponding physical applications described in references therein. See §3.5 for details of the results implied by application of the analysis in this monograph to the continuum limit of the Toda lattice.

1.2 DISCRETE ORTHOGONAL POLYNOMIALS

The common thread in all of these applications is the role played by systems of discrete orthogonal polynomials. Let $N \in \mathbb{N}$ and consider N distinct real *nodes* $x_{N,0} < x_{N,1} < \cdots < x_{N,N-1}$ to be given; together the nodes make up the support of the pure point measures we consider. We use the notation

$$X_N := \{x_{N,n}\}_{n=0}^{N-1}, \qquad \text{where } x_{N,j} < x_{N,k} \text{ whenever } j < k,$$

for the support set. We focus in this book on finite discrete sets of nodes; the analysis for infinite discrete node sets (as in the case of Meixner and Charlier weights) is not totally different, but details will be described elsewhere. Along with nodes we are given positive *weights* $w_{N,0}, w_{N,1}, \ldots, w_{N,N-1}$, which are the magnitudes of the point masses located at the corresponding nodes. We will occasionally use the alternate notation $w(x)$, $x \in X_N$, for a weight on the set of nodes X_N; thus

$$w(x_{N,n}) = w_{N,n}, \qquad n = 0, 1, 2, \ldots, N-1. \tag{1.9}$$

One should not infer from this notation that $w(x)$ has any meaning for any x, complex or real, other than for $x \in X_N$; even if w has a convenient functional form, we will evaluate $w(x)$ only when $x \in X_N$. The discrete orthogonal polynomials associated with this data are polynomials $\{p_{N,k}(z)\}_{k=0}^{N-1}$, where $p_{N,k}(z)$ is of degree exactly k with a positive leading coefficient and where

$$\sum_{n=0}^{N-1} p_{N,k}(x_{N,n}) p_{N,l}(x_{N,n}) w_{N,n} = \delta_{kl}. \tag{1.10}$$

Writing $p_{N,k}(z) = c_{N,k}^{(k)} z^k + \ldots + c_{N,k}^{(0)}$, we introduce distinguishing notation for the positive leading coefficient,

$$\gamma_{N,k} := c_{N,k}^{(k)},$$

and we denote by $\pi_{N,k}(z)$ the associated monic polynomial,

$$\pi_{N,k}(z) := \frac{1}{\gamma_{N,k}} p_{N,k}(z).$$

The discrete orthogonal polynomials exist and are uniquely determined by the orthogonality conditions because the inner product associated with (1.10) is positive-definite on $\operatorname{span}(1, z, z^2, \ldots, z^{N-1})$ but is degenerate on larger spaces of polynomials. The polynomials $p_{N,k}(z)$ may be built from the monomials by a Gram-Schmidt process. A general reference for properties of orthogonal polynomials specific to the discrete case is the book by Nikiforov, Suslov, and Uvarov [NikSU91].

One well-known elementary property of the discrete orthogonal polynomials is an exclusion principle for the zeros that forbids more than one zero from lying between adjacent nodes.

Proposition 1.1. *Each discrete orthogonal polynomial $p_{N,k}(z)$ has k simple real zeros. All these zeros lie in the range $x_{N,0} < z < x_{N,N-1}$, and no more than one zero lies in the closed interval $[x_{N,n}, x_{N,n+1}]$ between any two consecutive nodes.*

For general queries, contact webmaster@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

Proof. From the Gram-Schmidt process it follows that the coefficients of $p_{N,k}(z)$ are all real. Suppose that $p_{N,k}(z)$ vanishes to the nth order for some nonreal z_0. Then it follows that $p_{N,k}(z)$ also vanishes to the same order at z_0^* and thus that $p_{N,k}(z)/[(z-z_0)^n(z-z_0^*)^n]$ is a polynomial of lower degree, $k-2n \geq 0$. By orthogonality, we must have on the one hand,

$$\sum_{n=0}^{N-1} p_{N,k}(x_{N,n}) \cdot \frac{p_{N,k}(x_{N,n})}{|x_{N,n}-z_0|^{2n}} \cdot w_{N,n} = 0\,.$$

On the other hand, the left-hand side is strictly positive because $k < N$, so $p_{N,k}(z)$ cannot vanish at all of the nodes. So we have a contradiction and the roots must be real.

The necessarily real roots are simple for a similar reason. If z_0 is a real root of $p_{N,k}(z)$ of order greater than 1, the quotient $p_{N,k}(z)/(z-z_0)^2$ is a polynomial of degree $k-2 \geq 0$, which must be orthogonal to $p_{N,k}(z)$ itself,

$$\sum_{n=0}^{N-1} p_{N,k}(x_{N,n}) \cdot \frac{p_{N,k}(x_{N,n})}{(x_{N,n}-z_0)^2} \cdot w_{N,n} = 0\,,$$

but the left-hand side is manifestly positive, which gives the desired contradiction.

If a simple real zero z_0 of $p_{N,k}(z)$ satisfies either $z_0 \leq x_{N,0}$ or $z_0 \geq x_{N,N-1}$, then we repeat the above argument considering the polynomial $p_{N,k}(z)/(z-z_0)$ of degree $k-1 \geq 0$, to which $p_{N,k}(z)$ must be orthogonal but for which the inner product is strictly of one sign.

Finally, if more than one zero of $p_{N,k}(z)$ lies between the consecutive nodes $x_{N,n}$ and $x_{N,n+1}$, then we can certainly select two of them, say z_0 and z_1, and construct the polynomial $p_{N,k}(z)/[(z-z_0)(z-z_1)]$ of degree $k-2 \geq 0$. Again, this polynomial must be orthogonal to $p_{N,k}(z)$, but the corresponding inner product is of one definite sign, leading to a contradiction. □

Another well-known and important feature of all systems of orthogonal polynomials, which is present whether the weights are discrete or continuous, is the existence of a three-term recurrence relation. See [Sze91] for details. There are constants $a_{N,0}, a_{N,1}, \ldots, a_{N,N-2}$ and positive constants $b_{N,0}, b_{N,1}, \ldots, b_{N,N-2}$ such that

$$zp_{N,k}(z) = b_{N,k}p_{N,k+1}(z) + a_{N,k}p_{N,k}(z) + b_{N,k-1}p_{N,k-1}(z) \tag{1.11}$$

holds for $k = 1, \ldots, N-2$, while for $k = 0$ one has

$$zp_{N,0}(z) = b_{N,0}p_{N,1}(z) + a_{N,0}p_{N,0}(z)\,, \tag{1.12}$$

and for $k = N-1$,

$$zp_{N,N-1}(z) = \gamma_{N,N-1}\prod_{n=0}^{N-1}(z-x_{N,n}) + a_{N,N-1}p_{N,N-1}(z) + b_{N,N-2}p_{N,N-2}(z)\,. \tag{1.13}$$

Necessarily, one has $b_{N,k} = \gamma_{N,k}/\gamma_{N,k+1}$. Therefore the vectors $\mathbf{v}_j := (p_{N,0}(x_{N,j}), \ldots, p_{N,N-1}(x_{N,j}))^T$ form an orthonormal basis of eigenvectors (with corresponding eigenvalues $z = x_{N,j}$) of the symmetric tridiagonal $N \times N$ Jacobi matrix constructed from the sequences $\{a_{N,0}, \ldots, a_{N,N-1}\}$ and $\{b_{N,0}, \ldots, b_{N,N-2}\}$.

Our goal is to establish the asymptotic behavior of the polynomials $p_{N,k}(z)$ or their monic counterparts $\pi_{N,k}(z)$ in the limit of large degree, assuming certain asymptotic properties of the nodes and the weights. In particular, the number of nodes must necessarily increase to admit polynomials of arbitrarily large degree, and the weights we consider involve an exponential factor with the exponent proportional to the number of nodes (such weights are sometimes called *varying weights*). We will obtain pointwise asymptotics with a precise error bound uniformly valid in the whole complex plane. Our assumptions about the nodes and weights include as special cases all relevant classical discrete orthogonal polynomials but are significantly more general; in particular, we will consider nodes that are not necessarily equally spaced.

For general queries, contact webmaster@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

1.3 ASSUMPTIONS

1.3.1 Basic assumptions

We will establish rigorous asymptotics for the discrete orthogonal polynomials subject to the following fundamental assumptions.

The nodes

We suppose the existence of a *node density function* $\rho^0(x)$ that is real-analytic in a complex neighborhood of a closed interval $[a,b]$ and satisfies

$$\int_a^b \rho^0(x)\,dx = 1$$

and

$$\rho^0(x) > 0 \text{ strictly,} \quad \text{for all } x \in [a,b]\,. \tag{1.14}$$

The nodes are then defined precisely in terms of the density function $\rho^0(x)$ by the quantization rule

$$\int_a^{x_{N,n}} \rho^0(x)\,dx = \frac{2n+1}{2N}\,, \tag{1.15}$$

for $N \in \mathbb{N}$ and $n = 0,1,2,\ldots,N-1$. Thus the nodes lie in a bounded open interval (a,b) and are distributed with density $\rho^0(x)$.

The weights

Without loss of generality, we write the weights in the form

$$w_{N,n} = (-1)^{N-1-n} e^{-NV_N(x_{N,n})} \prod_{\substack{m=0\\m\neq n}}^{N-1} (x_{N,n} - x_{N,m})^{-1} = e^{-NV_N(x_{N,n})} \prod_{\substack{m=0\\m\neq n}}^{N-1} |x_{N,n} - x_{N,m}|^{-1}\,. \tag{1.16}$$

No generality has been sacrificed with this representation because the family of functions $\{V_N(x)\}$ is *a priori* specified only at the nodes; in other words, given positive weights $\{w_{N,n}\}$, one may solve (1.16) uniquely for the N quantities $\{V_N(x_{N,n})\}$. However, we now assume that for each sufficiently large N, $V_N(x)$ may be taken to be a real-analytic function defined in a complex neighborhood G of the closed interval $[a,b]$ and that

$$V_N(x) = V(x) + \frac{\eta(x)}{N}\,, \tag{1.17}$$

where $V(x)$ is a fixed real-analytic *potential function* defined in G and

$$\limsup_{N\to\infty} \sup_{z\in G} |\eta(z)| < \infty\,. \tag{1.18}$$

Note that in general the correction $\eta(z)$ may depend on N, although $V(x)$ may not. In some cases (*e.g.*, Krawtchouk polynomials; see §2.4.1) it is possible to take $V_N(x) \equiv V(x)$ for all N, in which case $\eta(x) \equiv 0$. However, the freedom of assuming $\eta(x) \not\equiv 0$ is useful in handling other cases (*e.g.*, the Hahn and associated Hahn polynomials; see §2.4.2). While (1.16) may be written for any system of positive weights, the condition that (1.17) should hold restricts attention to systems of weights that have analytic continuum limits in a certain precise sense.

◁ **Remark:** The familiar examples of classical discrete orthogonal polynomials correspond to nodes that are equally spaced, say, on $(a,b) = (0,1)$ (in which case we have $\rho^0(x) \equiv 1$). In this special case, the product factor on the right-hand side of (1.16) becomes simply

$$\prod_{\substack{m=0\\m\neq n}}^{N-1} |x_{N,n} - x_{N,m}|^{-1} = \frac{N^{N-1}}{n!(N-n-1)!}\,.$$

For general queries, contact webmaster@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

Using Stirling's formula to take the continuum limit of this factor (*i.e.*, considering $N \to \infty$ with $n/N \to x$) shows that in these cases the leading term in formula (1.16) is a continuous weight on $(0, 1)$:

$$w_{N,n} \sim w(x) := C \left(\frac{e^{-V(x)}}{x^x(1-x)^{1-x}} \right)^N \tag{1.19}$$

as $N \to \infty$ and $n/N \to x \in (0, 1)$, where C is independent of x. However, the process of taking the continuum limit of the weight first to arrive at a formula like (1.19) and then obtaining asymptotics of the polynomials of degree proportional to N as $N \to \infty$ is not equivalent to the scaling limit process we will consider here. Our results will display new phenomena because we simultaneously take the continuum limit as the degree of the polynomial grows. ▷

Our choice of the form (1.16) for the weights is motivated by several specific examples of classical discrete orthogonal polynomials. The form (1.16) is sufficiently general for us to carry out useful calculations related to proofs of universality conjectures arising in certain types of random tiling problems, random growth models, and last-passage percolation problems. Also, the form (1.16) is appropriate for study of the Toda lattice in the continuum limit by inverse spectral theory.

The degree

We assume that the degree k of the polynomial of interest is tied to the number N of nodes by a relation of the form

$$k = cN + \kappa \,,$$

where $c \in (0, 1)$ is a fixed parameter and κ remains bounded as $N \to \infty$.

1.3.2 Simplifying assumptions of genericity

In order to keep our exposition as simple as possible, we make further assumptions that exclude certain nongeneric triples $(\rho^0(x), V(x), c)$. These assumptions depend on the functions $\rho^0(x)$ and $V(x)$, and on the parameter c, in an implicit manner that is easier to describe once some auxiliary quantities have been introduced. They will be given in §2.1.2.

In regard to these particular assumptions, we want to stress two points. First, the excluded triples are nongeneric in the sense that any perturbation of, say, the parameter c will immediately return us to the class of triples for which all of our results are valid. The discussion at the beginning of §5.1.2 provides some insight into the generic nature of our assumptions. Second, the discrete orthogonal polynomials corresponding to nongeneric triples can be analyzed by the same basic method that we use here, with many of the same results. To do this, the proofs we present will require modifications to include additional local analysis near certain isolated points in the complex z-plane. Some such modifications have already been described in detail in the context of asymptotics for polynomials orthogonal with respect to continuous weights in §5 of [DeiKMVZ99b]. The remaining modifications have to do with nongeneric behavior near the endpoints of the interval $[a, b]$, and while the corresponding local analysis has not been done before, it can be expected to be of a similar character.

1.4 GOALS AND METHODOLOGY

Given an interval $[a, b]$, appropriate fixed functions $\rho^0(x)$ and $V(x)$, appropriate sequences $\eta(x) = \eta_N(x)$ and $\kappa = \kappa_N$, and a constant $c \in (0, 1)$, we wish to find accurate asymptotic formulae, valid in the limit $N \to \infty$ with rigorous error bounds, for the polynomial $\pi_{N,k}(z)$. These formulae should be uniformly valid in overlapping regions of the complex z-plane. We will also require asymptotic formulae for related quantities, like the zeros of $\pi_{N,k}(z)$, the three-term recurrence coefficients, and the reproducing kernels $K_{N,k}(x, y)$.

For general queries, contact webmaster@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

1.4.1 The basic interpolation problem

Given a natural number N, a set X_N of nodes, and a set of corresponding weights $\{w_{N,n}\}$, consider the possibility of finding the matrix $\mathbf{P}(z;N,k)$ solving the following problem, where k is an integer.

Interpolation Problem 1.2. *Find a 2×2 matrix $\mathbf{P}(z;N,k)$ with the following properties:*

1. **Analyticity***: $\mathbf{P}(z;N,k)$ is an analytic function of z for $z \in \mathbb{C} \setminus X_N$.*
2. **Normalization***: As $z \to \infty$,*

$$\mathbf{P}(z;N,k)\begin{pmatrix} z^{-k} & 0 \\ 0 & z^k \end{pmatrix} = \mathbb{I} + O\left(\frac{1}{z}\right). \tag{1.20}$$

3. **Singularities***: At each node $x_{N,n} \in X_N$, the first column of $\mathbf{P}(z;N,k)$ is analytic and the second column of $\mathbf{P}(z;N,k)$ has a simple pole, where the residue satisfies the condition*

$$\operatorname*{Res}_{z=x_{N,n}} \mathbf{P}(z;N,k) = \lim_{z\to x_{N,n}} \mathbf{P}(z;N,k)\begin{pmatrix} 0 & w_{N,n} \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & w_{N,n}P_{11}(x_{N,n};N,k) \\ 0 & w_{N,n}P_{21}(x_{N,n},N,k) \end{pmatrix}, \tag{1.21}$$

for $n = 0, \ldots, N-1$.

This problem is a discrete version of the Riemann-Hilbert problem appropriate for orthogonal polynomials with continuous weights that was first used by Fokas, Its, and Kitaev in [FokIK91] (see also [DeiKMVZ99a, DeiKMVZ99b]). The discrete version was first studied by Borodin [Bor00] (see also [Bor03] and [BorB03]). The solution of this problem encodes all quantities of relevance to a study of the discrete orthogonal polynomials, as we will now see.

Proposition 1.3. *Interpolation Problem 1.2 has a unique solution when $0 \le k \le N-1$. The solution is*

$$\mathbf{P}(z;N,k) = \begin{pmatrix} \pi_{N,k}(z) & \displaystyle\sum_{n=0}^{N-1} \frac{w_{N,n}\pi_{N,k}(x_{N,n})}{z - x_{N,n}} \\ \gamma_{N,k-1}p_{N,k-1}(z) & \displaystyle\sum_{n=0}^{N-1} \frac{w_{N,n}\gamma_{N,k-1}p_{N,k-1}(x_{N,n})}{z - x_{N,n}} \end{pmatrix}, \tag{1.22}$$

for $0 < k \le N-1$, while

$$\mathbf{P}(z;N,0) = \begin{pmatrix} 1 & \displaystyle\sum_{n=0}^{N-1} \frac{w_{N,n}}{z - x_{N,n}} \\ 0 & 1 \end{pmatrix}. \tag{1.23}$$

Proof. Consider the first row of $\mathbf{P}(z;N,k)$. According to (1.21), the function $P_{11}(z;N,k)$ is an entire function of z. Because $k \ge 0$, it follows from the normalization condition (1.20) that in fact $P_{11}(z;N,k)$ is a monic polynomial of degree exactly k. Similarly, from the characterization (1.21) of the simple poles of $P_{12}(z;N,k)$, we see that $P_{12}(z;N,k)$ is necessarily of the form

$$P_{12}(z;N,k) = e_1(z) + \sum_{n=0}^{N-1} \frac{w_{N,n}P_{11}(x_{N,n};N,k)}{z - x_{N,n}},$$

where $e_1(z)$ is an entire function. The normalization condition (1.20) for $k \ge 0$ immediately requires, via Liouville's Theorem, that $e_1(z) \equiv 0$, and then when $|z| > \max_n |x_{N,n}|$, we have by geometric series expansion that

$$P_{12}(z;N,k) = \sum_{m=0}^{\infty} \left(\sum_{n=0}^{N-1} P_{11}(x_{N,n};N,k)x_{N,n}^m w_{N,n} \right) \frac{1}{z^{m+1}}.$$

For general queries, contact webmaster@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

According to the normalization condition (1.20), $P_{12}(z;N,k) = o(z^{-k})$ as $z \to \infty$; therefore it follows that the monic polynomial $P_{11}(z;N,k)$ of degree exactly k must satisfy

$$\sum_{n=0}^{N-1} P_{11}(x_{N,n};N,k)x_{N,n}^m w_{N,n} = 0\,, \qquad \text{for } m = 0,1,2,\ldots,k-1\,.$$

As long as $k \le N-1$, these conditions uniquely identify $P_{11}(z;N,k)$ with the monic discrete orthogonal polynomial $\pi_{N,k}(z)$. The existence and uniqueness of $\pi_{N,k}(z)$ for such k is guaranteed given distinct orthogonalization nodes and positive weights (which implies that the inner product is a positive-definite quadratic form).

The second row of $\mathbf{P}(z;N,k)$ is studied similarly. The matrix element $P_{21}(z;N,k)$ is seen from (1.21) to be an entire function of z, which according to the normalization condition (1.20) must be a polynomial of degree at most $k-1$ (for the special case of $k=0$ these conditions immediately imply that $P_{21}(z;N,0) \equiv 0$). The characterization (1.21) implies that $P_{22}(z;N,k)$ can be expressed in the form

$$P_{22}(z;N,k) = e_2(z) + \sum_{n=0}^{N-1} \frac{w_{N,n}P_{21}(x_{N,n};N,k)}{z - x_{N,n}}\,,$$

where $e_2(z)$ is an entire function. If $k=0$, then $P_{22}(z;N,0) = e_2(z)$, and then according to the normalization condition (1.20) we must take $e_2(z) \equiv 1$. On the other hand, if $k>0$, then (1.20) implies that $P_{22}(z;N,k)$ decays for large z, and therefore we must take $e_2(z) \equiv 0$ in this case. Expanding the denominator in a geometric series for $|z| > \max_n |x_{N,n}|$, we find

$$P_{22}(z;N,k) = \sum_{m=0}^{\infty}\left(\sum_{n=0}^{N-1} P_{21}(x_{N,n};N,k)x_{N,n}^m w_{N,n}\right)\frac{1}{z^{m+1}}\,.$$

Imposing the normalization condition (1.20) amounts to insisting that $P_{22}(z;N,k) = z^{-k} + O(z^{-k-1})$ as $z \to \infty$; therefore

$$\sum_{n=0}^{N-1} P_{21}(x_{N,n};N,k)x_{N,n}^m w_{N,n} = 0\,, \qquad \text{for } m = 0,1,2,\ldots,k-2\,, \tag{1.24}$$

and

$$\sum_{n=0}^{N-1} P_{21}(x_{N,n};N,k)x_{N,n}^{k-1} w_{N,n} = 1\,. \tag{1.25}$$

With the use of (1.24), the condition (1.25) can be replaced by

$$\sum_{n=0}^{N-1} P_{21}(x_{N,n};N,k)\pi_{N,k-1}(x_{N,n})w_{N,n} = 1$$

or, equivalently,

$$\sum_{n=0}^{N-1}\left[\frac{1}{\gamma_{N,k-1}}P_{21}(x_{N,n};N,k)\right]p_{N,k-1}(x_{N,n})w_{N,n} = 1\,. \tag{1.26}$$

The conditions (1.24) and (1.26) therefore uniquely identify the quotient $P_{21}(z;N,k)/\gamma_{N,k-1}$ with the orthogonal polynomial $p_{N,k-1}(z)$.

The interpolation problem is thus solved uniquely by the matrix explicitly given by (1.22) for $k>0$ and by (1.23) for $k=0$. □

◁ **Remark:** In fact, Interpolation Problem 1.2 can also be solved for $k=N$, with a unique solution of the form (1.22), if we define

$$\pi_{N,N}(z) := \prod_{n=0}^{N-1}(z - x_{N,n})\,,$$

which of course is not in the finite family of orthogonal polynomials as it is not normalizable. ▷

The constants in the three-term recurrence relations (see equations (1.11)–(1.13)) are also encoded in the matrix $\mathbf{P}(z;N,k)$ solving Interpolation Problem 1.2.

For general queries, contact webmaster@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

Corollary 1.4. *Let k be fixed with $1 \le k \le N-2$ and let $s_{N,k}$, $y_{N,k}$, $r_{N,k}$, and $u_{N,k}$ denote certain terms in the large-z expansion of the matrix elements of $\mathbf{P}(z;N,k)$,*

$$z^k P_{12}(z;N,k) = \frac{s_{N,k}}{z} + \frac{y_{N,k}}{z^2} + O\left(\frac{1}{z^3}\right),$$
$$\frac{1}{z^k} P_{11}(z;N,k) = 1 + \frac{r_{N,k}}{z} + O\left(\frac{1}{z^2}\right),$$
$$\frac{1}{z^k} P_{21}(z;N,k) = \frac{u_{N,k}}{z} + O\left(\frac{1}{z^2}\right),$$

as $z \to \infty$. Then

$$\begin{aligned} \gamma_{N,k} &= \frac{1}{\sqrt{s_{N,k}}}, & \gamma_{N,k-1} &= \sqrt{u_{N,k}}, \\ a_{N,k} &= r_{N,k} + \frac{y_{N,k}}{s_{N,k}}, & b_{N,k} &= \sqrt{s_{N,k+1}u_{N,k+1}}. \end{aligned} \tag{1.27}$$

Also, $a_{N,k} = r_{N,k} - r_{N,k+1}$.

Proof. By expansion of the explicit solution given by (1.22) in Proposition 1.3 in the limit of large z, we deduce the identity

$$s_{N,k} = \sum_{n=0}^{N-1} \pi_{N,k}(x_{N,n}) x_{N,n}^k w_{N,n} = \frac{1}{\gamma_{N,k}^2} \tag{1.28}$$

(because $z^k = \pi_{N,k}(z) + O(z^{k-1})$, and using the definition of the normalization constants $\gamma_{N,k}$), the identity

$$y_{N,k} = \sum_{n=0}^{N-1} \pi_{N,k}(x_{N,n}) x_{N,n}^{k+1} w_{N,n} = -\frac{c_{N,k+1}^{(k)}}{\gamma_{N,k}^2 \gamma_{N,k+1}} \tag{1.29}$$

(because $z^{k+1} = \pi_{N,k+1}(z) - c_{N,k+1}^{(k)} \gamma_{N,k+1}^{-1} \pi_{N,k}(z) + O(z^{k-1})$), the identity

$$r_{N,k} = \frac{c_{N,k}^{(k-1)}}{\gamma_{N,k}}, \tag{1.30}$$

and the identity

$$u_{N,k} = \gamma_{N,k-1}^2. \tag{1.31}$$

Similarly, by expansion of the three-term recurrence relation in the limit of large z, we deduce the identity

$$\gamma_{N,k} z^{k+1} + c_{N,k}^{(k-1)} z^k = b_{N,k}\gamma_{N,k+1} z^{k+1} + (b_{N,k} c_{N,k+1}^{(k)} + a_{N,k}\gamma_{N,k}) z^k + O(z^{k-1})$$

as $z \to \infty$. Therefore

$$b_{N,k} = \frac{\gamma_{N,k}}{\gamma_{N,k+1}} \quad \text{and} \quad a_{N,k} = \frac{c_{N,k}^{(k-1)}}{\gamma_{N,k}} - \frac{c_{N,k+1}^{(k)}}{\gamma_{N,k+1}}.$$

Comparing with (1.28)–(1.31) completes the proof. □

1.4.2 Exponentially deformed weights and the Toda lattice

Let the nodes X_N be fixed. Suppose now that the weights $\{w_{N,n}\}$ are allowed to depend smoothly on a parameter t so as to remain positive for all t. By the solution of Interpolation Problem 1.2, this deformation of the weights induces a corresponding deformation in the orthogonal polynomials and all related quantities (*e.g.*, three-term recurrence coefficients $a_{N,k}$, $b_{N,k}$ and norming constants $\gamma_{N,k}$). In particular, the dynamics induced on the recurrence coefficients can be written as a system of nonlinear differential equations for the $a_{N,k}$ and $b_{N,k}$. The collection of all possible differential equations obtainable in this way is the *Toda lattice hierarchy.*

For general queries, contact webmaster@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

The Toda lattice hierarchy is spanned by the flows

$$w_{N,n}(t) := w_{N,n}\exp(2x_{N,n}^p t)\,, \qquad n = 0,\ldots,N-1\,, \tag{1.32}$$

for $p = 1, 2, 3, \ldots$. Here we show how to derive the differential equations satisfied by the recurrence coefficients in the simplest case of $p = 1$. From the solution $\mathbf{P}(z; N, k, t)$ of Interpolation Problem 1.2 with weights of the form (1.32), define the *Jost matrix*

$$\mathbf{M}(z; N, k, t) := \mathbf{P}(z; N, k, t)\begin{pmatrix} e^{zt} & 0 \\ 0 & e^{-zt} \end{pmatrix}.$$

A simple calculation then shows that

$$\operatorname*{Res}_{z=x_{N,n}} \mathbf{M}(z; N, k, t) = \lim_{z\to x_{N,n}} \mathbf{M}(z; N, k, t)\begin{pmatrix} 0 & w_{N,n} \\ 0 & 0 \end{pmatrix}, \tag{1.33}$$

for $n = 0, \ldots, N-1$. In other words, the relations that constrain the residues at the simple poles of $\mathbf{M}(z; N, k, t)$ are independent of t as well as k. Note that this does not imply that $\mathbf{M}(z; N, k, t)$ is independent of t and k; indeed, as $z \to \infty$, $\mathbf{M}(z; N, k, t)$ exhibits exponential behavior in both t and k according to the normalization condition satisfied by $\mathbf{P}(z; N, k, t)$. On the other hand, (1.33) does imply that the matrices

$$\mathbf{L}(z; N, k, t) := \mathbf{M}(z; N, k+1, t)\mathbf{M}(z; N, k, t)^{-1}$$

and

$$\mathbf{B}(z; N, k, t) := \frac{d\mathbf{M}}{dt}(z; N, k, t)\mathbf{M}(z; N, k, t)^{-1}$$

are entire analytic functions of z for each k and t. Moreover, both of these matrix-valued functions are, by Liouville's Theorem, polynomials in z because they have polynomial asymptotics as $z \to \infty$. Writing

$$\begin{pmatrix} r_{N,k}(t) & s_{N,k}(t) \\ u_{N,k}(t) & v_{N,k}(t) \end{pmatrix} := \lim_{z\to\infty} z\left[\mathbf{P}(z; N, k, t)\begin{pmatrix} z^{-k} & 0 \\ 0 & z^k \end{pmatrix} - \mathbb{I}\right],$$

a direct calculation using the normalization condition satisfied by the matrix $\mathbf{P}(z; N, k, t)$ shows that

$$\mathbf{L}(z; N, k, t) = \begin{pmatrix} z + r_{N,k+1}(t) - r_{N,k}(t) & -s_{N,k}(t) \\ u_{N,k+1}(t) & 0 \end{pmatrix} + O(z^{-1})$$

and

$$\mathbf{B}(z; N, k, t) = \begin{pmatrix} z & -2s_{N,k}(t) \\ 2u_{N,k}(t) & -z \end{pmatrix} + O(z^{-1})$$

as $z \to \infty$. By Liouville's Theorem, we therefore have, exactly,

$$\mathbf{L}(z; N, k, t) = \begin{pmatrix} z + r_{N,k+1}(t) - r_{N,k}(t) & -s_{N,k}(t) \\ u_{N,k+1}(t) & 0 \end{pmatrix}$$

and

$$\mathbf{B}(z; N, k, t) = \begin{pmatrix} z & -2s_{N,k}(t) \\ 2u_{N,k}(t) & -z \end{pmatrix}.$$

The simultaneous linear equations

$$\begin{aligned} \mathbf{M}(z; N, k+1, t) &= \mathbf{L}(z; N, k, t)\mathbf{M}(z; N, k, t)\,, \\ \frac{d\mathbf{M}}{dt}(z; N, k, t) &= \mathbf{B}(z; N, k, t)\mathbf{M}(z; N, k, t)\,, \end{aligned} \tag{1.34}$$

satisfied by $\mathbf{M}(z; N, k, t)$, are said to make up a *Lax pair* for the Toda lattice. By computing the "shifted derivative" $d\mathbf{M}(z; N, k+1, t)/dt$ two different ways using the Lax pair and equating the results, one finds that

$$\left[\frac{d\mathbf{L}}{dt}(z; N, k, t) + \mathbf{L}(z; N, k, t)\mathbf{B}(z; N, k, t) - \mathbf{B}(z; N, k+1, t)\mathbf{L}(z; N, k, t)\right]\mathbf{M}(z; N, k, t) = \mathbf{0}\,. \tag{1.35}$$

For general queries, contact webmaster@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

Now it also follows from the conditions of Interpolation Problem 1.2 that $\det(\mathbf{P}(z;N,k,t)) \equiv 1$, so the matrix $\mathbf{M}(z;N,k,t)$ is a fundamental solution matrix for the Lax pair. Therefore we deduce from (1.35) the *compatibility condition*

$$\frac{d\mathbf{L}}{dt}(z;N,k,t) + \mathbf{L}(z;N,k,t)\mathbf{B}(z;N,k,t) - \mathbf{B}(z;N,k+1,t)\mathbf{L}(z;N,k,t) = \mathbf{0}\,. \tag{1.36}$$

This condition is the *zero-curvature representation* of the Toda lattice equations. Although the matrix elements for $\mathbf{L}(z;N,k,t)$ and $\mathbf{B}(z;N,k,t)$ depend on z, it is easy to check that the combination on the left-hand side of (1.36) is independent of z and is a matrix with three nonzero elements. The result of these observations is that (1.36) is equivalent to the following three differential equations:

$$\begin{aligned} \frac{d}{dt}(r_{N,k} - r_{N,k+1}) &= 2s_{N,k+1}u_{N,k+1} - 2s_{N,k}u_{N,k}\,, \\ \frac{ds_{N,k}}{dt} &= 2(r_{N,k} - r_{N,k+1})s_{N,k}\,, \\ \frac{du_{N,k+1}}{dt} &= -2(r_{N,k} - r_{N,k+1})u_{N,k+1}\,. \end{aligned} \tag{1.37}$$

According to Corollary 1.4, the quantities appearing in these equations are related to the three-term recurrence coefficients

$$a_{N,k} = r_{N,k} - r_{N,k+1} \qquad \text{and} \qquad b_{N,k}^2 = s_{N,k+1}u_{N,k+1}\,,$$

so from (1.37) we obtain a closed set of equations governing the dynamics of the recurrence coefficients:

$$\frac{da_{N,k}}{dt} = 2b_{N,k}^2 - 2b_{N,k-1}^2 \qquad \text{and} \qquad \frac{db_{N,k}}{dt} = (a_{N,k+1} - a_{N,k})b_{N,k}\,. \tag{1.38}$$

These are the *Toda lattice equations.* We have derived these equations assuming that $k = 1, \dots, N-2$, but a similar analysis for $k = 0$ yields

$$\frac{da_{N,0}}{dt} = 2b_{N,0}^2 \qquad \text{and} \qquad \frac{db_{N,0}}{dt} = (a_{N,1} - a_{N,0})b_{N,0}\,,$$

and for $k = N-1$ yields

$$\frac{da_{N,N-1}}{dt} = -2b_{N,N-2}^2\,.$$

In other words, (1.38) holds for $k = 0, \dots, N-1$ if we define $b_{N,-1} = b_{N,N-1} = 0$.

Now we consider the *squared eigenfunctions* associated with the Lax pair (1.34). Let $\mathbf{C}$ be a constant matrix with trace zero and define the matrix of squared eigenfunctions as

$$\mathbf{W}(z;N,k,t) := \mathbf{M}(z;N,k,t)\mathbf{C}\mathbf{M}(z;N,k,t)^{-1}\,.$$

It follows that $\mathbf{W}(z;N,k,t)$ has trace zero as well and therefore may be written as

$$\mathbf{W}(z;N,k,t) = \begin{pmatrix} f_{N,k}(z;t) & g_{N,k}(z;t) \\ h_{N,k}(z;t) & -f_{N,k}(z;t) \end{pmatrix}. \tag{1.39}$$

From the Lax pair, it follows that $\mathbf{W}(z;N,k,t)$ also obeys a system of linear simultaneous equations even though its elements consist of quadratic forms in the elements of $\mathbf{M}(z;N,k,t)$:

$$\begin{aligned} \mathbf{W}(z;N,k+1,t) &= \mathbf{L}(z;N,k,t)\mathbf{W}(z;N,k,t)\mathbf{L}(z;N,k,t)^{-1}\,, \\ \frac{d\mathbf{W}}{dt}(z;N,k,t) &= [\mathbf{B}(z;N,k,t), \mathbf{W}(z;N,k,t)]\,, \end{aligned} \tag{1.40}$$

where $[\mathbf{A},\mathbf{B}] := \mathbf{AB} - \mathbf{BA}$ denotes the matrix commutator. Substituting the form (1.39) into the difference equation in (1.40) (first multiplying on the right by $\mathbf{L}(z;N,k,t)$) yields four equations:

$$\begin{aligned} (z - a_{N,k})f_{N,k+1} + u_{N,k+1}g_{N,k+1} &= (z - a_{N,k})f_{N,k} - s_{N,k}h_{N,k}\,, \\ -s_{N,k}f_{N,k+1} &= (z - a_{N,k})g_{N,k} + s_{N,k}f_{N,k}\,, \\ (z - a_{N,k})h_{N,k+1} - u_{N,k+1}f_{N,k+1} &= u_{N,k+1}f_{N,k}\,, \\ -s_{N,k}h_{N,k+1} &= u_{N,k+1}g_{N,k}\,. \end{aligned}$$

For general queries, contact webmaster@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

Here we have used $a_{N,k} = r_{N,k} - r_{N,k+1}$. The last of these equations may be solved by introducing a sequence $\phi_{N,k}$ and writing

$$g_{N,k} := s_{N,k}\phi_{N,k} \qquad \text{and} \qquad h_{N,k} := -u_{N,k}\phi_{N,k-1}\,. \tag{1.41}$$

This choice also makes two of the remaining three equations identical, so only two equations remain:

$$\begin{aligned} (z - a_{N,k})(f_{N,k+1} - f_{N,k}) + b_{N,k}^2\phi_{N,k+1} - b_{N,k-1}^2\phi_{N,k-1} = 0\,,\\ f_{N,k+1} + f_{N,k} + (z - a_{N,k})\phi_{N,k} = 0\,. \end{aligned} \tag{1.42}$$

Here we have used $b_{N,k}^2 = s_{N,k+1}u_{N,k+1}$. Similarly, using (1.39) and (1.41) in the differential equation in (1.40) yields only two distinct scalar equations (up to shifts in k), namely,

$$\frac{df_{N,k}}{dt} = -2b_{N,k-1}^2(\phi_{N,k} - \phi_{N,k-1}) \qquad \text{and} \qquad \frac{d\phi_{N,k}}{dt} = 2(z - a_{N,k})\phi_{N,k} + 4f_{N,k}\,.$$

Using the second equation in (1.42) to eliminate $z - a_{N,k}$ yields the linear system

$$\frac{df_{N,k}}{dt} = -2b_{N,k-1}^2(\phi_{N,k} - \phi_{N,k-1}) \qquad \text{and} \qquad \frac{d\phi_{N,k}}{dt} = 2f_{N,k} - 2f_{N,k+1}\,.$$

By taking finite differences of these equations with respect to k and introducing new variables

$$\hat{a}_{N,k} := f_{N,k+1} - f_{N,k} \qquad \text{and} \qquad \hat{b}_{N,k} := -\frac{1}{2}(\phi_{N,k+1} - \phi_{N,k})\,,$$

the equations become

$$\frac{d\hat{a}_{N,k}}{dt} = 4b_{N,k}^2\hat{b}_{N,k} - 4b_{N,k-1}^2\hat{b}_{N,k-1} \qquad \text{and} \qquad \frac{d\hat{b}_{N,k}}{dt} = \hat{a}_{N,k+1} - \hat{a}_{N,k}\,.$$

We recognize these as the linearized Toda lattice equations. Although the linearized Toda lattice equations are themselves independent of z, we note that the functions $\hat{a}_{N,k}(z;t)$ and $\hat{b}_{N,k}(z;t)$ that satisfy them do indeed depend nontrivially on z through the solution of Interpolation Problem 1.2, and so by variation of z we obtain an infinite family of solutions of the linearized problem. Of course, only $2N - 1$ of them can be linearly independent, but there is sufficient freedom in the choice of the parameter z to construct a complete basis of solutions of the linearized problem for each N.

1.4.3 Triangularity of residue matrices and dual polynomials

The matrices that encode the residues in Interpolation Problem 1.2 are upper-triangular. An essential aspect of our methodology will be to modify the matrix $\mathbf{P}(z;N,k)$ in order to selectively reverse the triangularity of the residue matrices near certain individual nodes $x_{N,n}$. Let $\Delta \subset \mathbb{Z}_N$, where

$$\mathbb{Z}_N := \{0, 1, 2, \ldots, N-1\}\,,$$

and denote the number of elements in Δ by $\#\Delta$. We will reverse the triangularity for nodes $x_{N,n}$ for which $n \in \Delta$. Consider the matrix $\mathbf{Q}(z;N,k)$ related to the solution $\mathbf{P}(z;N,k)$ of Interpolation Problem 1.2 as follows:

$$\mathbf{Q}(z;N,k) := \mathbf{P}(z;N,k)\left[\prod_{n\in\Delta}(z - x_{N,n})\right]^{-\sigma_3} = \mathbf{P}(z;N,k)\begin{pmatrix} \displaystyle\prod_{n\in\Delta}(z - x_{N,n})^{-1} & 0 \\ 0 & \displaystyle\prod_{n\in\Delta}(z - x_{N,n}) \end{pmatrix}. \tag{1.43}$$

Here σ_3 is a Pauli matrix:

$$\sigma_3 := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

It is easy to check that the matrix $\mathbf{Q}(z;N,k)$ so defined is an analytic function of z for $z \in \mathbb{C} \setminus X_N$ that satisfies the normalization condition

$$\mathbf{Q}(z;N,k)\begin{pmatrix} z^{\#\Delta-k} & 0 \\ 0 & z^{k-\#\Delta} \end{pmatrix} = \mathbb{I} + O\left(\frac{1}{z}\right) \qquad \text{as } z \to \infty.$$

For general queries, contact webmaster@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

Furthermore, at each node $x_{N,n}$, the matrix $\mathbf{Q}(z;N,k)$ has a simple pole. If n belongs to the complementary set

$$\nabla := \mathbb{Z}_N \setminus \Delta\,,$$

then the first column is analytic at $x_{N,n}$ and the pole is in the second column such that the residue satisfies the condition

$$\operatorname*{Res}_{z=x_{N,n}} \mathbf{Q}(z;N,k) = \lim_{z\to x_{N,n}} \mathbf{Q}(z;N,k) \begin{pmatrix} 0 & w_{N,n} \displaystyle\prod_{m\in\Delta}(x_{N,n}-x_{N,m})^2 \\ 0 & 0 \end{pmatrix}, \tag{1.44}$$

for $n \in \nabla$. If $n \in \Delta$, then the second column is analytic at $x_{N,n}$ and the pole is in the first column such that the residue satisfies the condition

$$\operatorname*{Res}_{z=x_{N,n}} \mathbf{Q}(z;N,k) = \lim_{z\to x_{N,n}} \mathbf{Q}(z;N,k) \begin{pmatrix} 0 & 0 \\ \dfrac{1}{w_{N,n}} \displaystyle\prod_{\substack{m\in\Delta \\ m\neq n}}(x_{N,n}-x_{N,m})^{-2} & 0 \end{pmatrix}, \tag{1.45}$$

for $n \in \Delta$. Thus the triangularity of the residue matrices has been reversed for nodes in $\Delta \subset X_N$.

The relation between the solution $\mathbf{P}(z;N,k)$ of Interpolation Problem 1.2 and the matrix $\mathbf{Q}(z;N,k)$ obtained therefrom by selective reversal of residue triangularity gives rise in a special case to a remarkable duality between pairs of weights $\{w_{N,n}\}$ defined on the same set of nodes and their corresponding families of discrete orthogonal polynomials that comes up in applications. Given nodes X_N and weights $\{w_{N,n}\}$, the dual polynomials arise by taking $\Delta = \mathbb{Z}_N$ in the change of variables (1.43) and then defining

$$\overline{\mathbf{P}}(z;N,\overline{k}) := \sigma_1 \mathbf{Q}(z;N,k)\sigma_1\,, \qquad \text{where } \overline{k} := N-k\,.$$

Here σ_1 is another Pauli matrix:

$$\sigma_1 := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Thus we are reversing the triangularity at all of the nodes and swapping rows and columns of the resulting matrix. It is easy to check that $\overline{\mathbf{P}}(z;N,\overline{k})$ satisfies

$$\overline{\mathbf{P}}(z;N,\overline{k}) \begin{pmatrix} z^{-\overline{k}} & 0 \\ 0 & z^{\overline{k}} \end{pmatrix} = \mathbb{I} + O\left(\frac{1}{z}\right) \qquad \text{as } z\to\infty$$

and is a matrix with simple poles in the second column at all nodes such that

$$\operatorname*{Res}_{z=x_{N,n}} \overline{\mathbf{P}}(z;N,\overline{k}) = \lim_{z\to x_{N,n}} \overline{\mathbf{P}}(z;N,\overline{k}) \begin{pmatrix} 0 & \overline{w}_{N,n} \\ 0 & 0 \end{pmatrix}$$

holds for $n \in \mathbb{Z}_N$, where the *dual weights* $\{\overline{w}_{N,n}\}$ are defined by the identity

$$w_{N,n}\overline{w}_{N,n} \prod_{\substack{m=0 \\ m\neq n}}^{N-1} (x_{N,n}-x_{N,m})^2 = 1\,. \tag{1.46}$$

Comparing with Interpolation Problem 1.2, we see that $\overline{P}_{11}(z;N,\overline{k})$ is the monic orthogonal polynomial $\overline{\pi}_{N,\overline{k}}(z)$ of degree $\overline{k}$ associated with the dual weights $\{\overline{w}_{N,j}\}$ (and the same set of nodes X_N). In this sense, families of discrete orthogonal polynomials always come in dual pairs. An explicit relation between the dual polynomials comes from the representation of $\mathbf{P}(z;N,k)$ given by Proposition 1.3:

$$\begin{aligned} \overline{\pi}_{N,\overline{k}}(z) &= \overline{P}_{11}(z;N,\overline{k}) \\ &= P_{22}(z;N,k) \prod_{n=0}^{N-1}(z-x_{N,n}) \\ &= \sum_{m=0}^{N-1} w_{N,m}\gamma_{N,k-1}^2 \pi_{N,k-1}(x_{N,m}) \prod_{\substack{n=0 \\ n\neq m}}^{N-1}(z-x_{N,n})\,. \end{aligned} \tag{1.47}$$

For general queries, contact webmaster@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

Since the left-hand side is a monic polynomial of degree $\overline{k} = N - k$ and the right-hand side is apparently a polynomial of degree $N-1$, equation (1.47) furnishes k relations involving the weights and the normalization constants $\gamma_{N,k}$.

In particular, if we evaluate (1.47) for $z = x_{N,l}$ for some $l \in \mathbb{Z}_N$, then only one term from the sum on the right-hand side survives, and we find

$$\overline{\pi}_{N,\overline{k}}(x_{N,l}) = \gamma^2_{N,k-1} w_{N,l} \prod_{\substack{n=0 \\ n \neq l}}^{N-1} (x_{N,l} - x_{N,n}) \cdot \pi_{N,k-1}(x_{N,l}) \,, \tag{1.48}$$

an identity relating values of each discrete orthogonal polynomial and a corresponding dual polynomial at any given node. The identity (1.48) has also been derived by Borodin [Bor02].

Furthermore, by using (1.47) twice, along with the fact that $\overline{\overline{\pi}}_k(z) \equiv \pi_k(z)$ (*i.e.*, duality is an involution), we can obtain some additional identities involving the discrete orthogonal polynomials and their duals. By involution, (1.47) implies that

$$\pi_{N,k}(z) = \overline{\gamma}^2_{N,\overline{k}-1} \gamma^2_{N,k} \sum_{m=0}^{N-1} \pi_{N,k}(x_{N,m}) \prod_{\substack{n=0 \\ n \neq m}}^{N-1} \frac{z - x_{N,n}}{x_{N,m} - x_{N,n}} \,.$$

The sum on the right-hand side is the Lagrange interpolating polynomial of degree $N-1$ (at most) that agrees with $\pi_{N,k}(z)$ at all N nodes. Of course this identifies the sum with $\pi_{N,k}(z)$ itself, and we therefore deduce the relation

$$\overline{\gamma}_{N,\overline{k}-1} = \frac{1}{\gamma_{N,k}}$$

between the leading coefficients of the discrete orthogonal polynomials and their duals.

◁ **Remark:** We want to point out that the notion of duality described here is different from that explained in [NikSU91]. The latter generally involves relationships between families of discrete orthogonal polynomials with two different sets of nodes of orthogonalization. For example, the Hahn polynomials are orthogonal on a lattice of equally spaced points, and the polynomials dual to the Hahn polynomials by the scheme of [NikSU91] are orthogonal on a quadratic lattice for which $x_{N,n} - x_{N,n-1}$ is proportional to n. However, the polynomials dual to the Hahn polynomials under the scheme described above are the associated Hahn polynomials, which are orthogonal on the same equally spaced nodes as the Hahn polynomials themselves. The notion of duality described above coincides with that described in [Bor02] and is also equivalent to the "hole-particle transformation" considered by Johansson [Joh02]. ▷

1.4.4 Overview of the key steps

The characterization of the discrete orthogonal polynomials in terms of Interpolation Problem 1.2 is the starting point for our asymptotic analysis. Our rigorous analysis of $\mathbf{P}(z; N, k)$ consists of three steps:

1. We introduce a change of variables, transforming $\mathbf{P}(z; N, k)$ into $\mathbf{X}(z)$, another matrix function of z. The transformation mediating between $\mathbf{P}(z; N, k)$ and $\mathbf{X}(z)$ is explicit and exact. The matrix $\mathbf{X}(z)$ is shown to satisfy a matrix Riemann-Hilbert problem that is equivalent to Interpolation Problem 1.2.

2. We construct an explicit model $\hat{\mathbf{X}}(z)$ for $\mathbf{X}(z)$ on the basis of formal asymptotics. We call $\hat{\mathbf{X}}(z)$ a *global parametrix* for $\mathbf{X}(z)$.

3. We compare $\mathbf{X}(z)$ to the global parametrix $\hat{\mathbf{X}}(z)$ by considering the error $\mathbf{E}(z) := \mathbf{X}(z)\hat{\mathbf{X}}(z)^{-1}$, which should be close to the identity matrix if the formally obtained global parametrix $\hat{\mathbf{X}}(z)$ is indeed a good approximation of $\mathbf{X}(z)$. We rigorously analyze $\mathbf{E}(z)$ by viewing its definition in terms of $\mathbf{X}(z)$ as another change of variables since $\hat{\mathbf{X}}(z)$ is known explicitly from step 2. This means that we may pose an equivalent Riemann-Hilbert problem for $\mathbf{E}(z)$. We prove that this Riemann-Hilbert problem

For general queries, contact webmaster@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

may be solved by a convergent Neumann series if N is sufficiently large. The series for $\mathbf{E}(z)$ is also an asymptotic series whose first term is the identity matrix, such that $\mathbf{E}(z) - \mathbb{I}$ is of order $1/N$ in a suitable precise sense. This gives an asymptotic formula for the unknown matrix $\mathbf{X}(z) = \mathbf{E}(z)\hat{\mathbf{X}}(z)$. Inverting the explicit change of variables from step 1 linking $\mathbf{X}(z)$ with $\mathbf{P}(z; N, k)$, we finally arrive at an asymptotic formula for $\mathbf{P}(z; N, k)$.

The first step in this process is the most crucial since the explicit transformation from $\mathbf{P}(z; N, k)$ to $\mathbf{X}(z)$ has to result in a problem that has been properly prepared for asymptotic analysis. The transformation is best presented as a composition of several subsequent transformations:

1(a). A transformation (1.43) is introduced from $\mathbf{P}(z; N, k)$ to a new unknown matrix $\mathbf{Q}(z; N, k)$, having the effect of moving poles at some of the nodes in X_N from the second column of $\mathbf{P}(z; N, k)$ to the first column of $\mathbf{Q}(z; N, k)$. This transformation turns out to be necessary in our approach to take into account subintervals of $[a, b]$ that are saturated with zeros of $\pi_{N,k}(z)$ in the sense that there is a zero between each pair of neighboring nodes (recall Proposition 1.1). The saturated regions are not known in advance but are detected by the equilibrium measure (see step 1(c) below).

1(b). The matrix $\mathbf{Q}(z; N, k)$ is transformed into $\mathbf{R}(z)$, a matrix that has, instead of polar singularities, a jump discontinuity across a contour in the complex z-plane along which $\mathbf{R}(z)$ takes continuous boundary values. To see how a pole may be removed at the cost of a jump across a contour, consider a point x_0 at which a matrix function $\mathbf{M}(z)$ is meromorphic, having a simple pole in the second column such that, for some given constant w_0,

$$\operatorname*{Res}_{z=x_0} \mathbf{M}(z) = \lim_{z \to x_0} \mathbf{M}(z) \begin{pmatrix} 0 & w_0 \\ 0 & 0 \end{pmatrix} . \tag{1.49}$$

If $f(z)$ is a scalar function analytic in the region $0 < |z - x_0| < \epsilon$ for some $\epsilon > 0$ having a simple pole at x_0 with residue w_0 (obviously there are many such functions and consequently significant freedom in making a choice), then we may try to define a new matrix function $\mathbf{N}(z)$ by choosing some positive $\delta < \epsilon$ sufficiently small and setting

$$\mathbf{N}(z) = \begin{cases} \mathbf{M}(z) , & \text{for } |z - x_0| > \delta , \\ \mathbf{M}(z) \begin{pmatrix} 1 & -f(z) \\ 0 & 1 \end{pmatrix} , & \text{for } |z - x_0| < \delta . \end{cases} \tag{1.50}$$

It follows that the singularity of $\mathbf{N}(z)$ at $z = x_0$ is removable. Therefore $\mathbf{N}(z)$ may be considered to be analytic in the region $|z - x_0| < \delta$ and also at each point of the region $|z - x_0| > \delta$ where additionally $\mathbf{M}(z)$ is known to be analytic. In place of the residue condition (1.49), we now have a known jump discontinuity across the circle $|z - x_0| = \delta$ along which $\mathbf{N}(z)$ takes continuous boundary values from the inside (denoted $\mathbf{N}_+(z)$) and the outside (denoted $\mathbf{N}_-(z)$):

$$\mathbf{N}_+(z) = \mathbf{N}_-(z) \begin{pmatrix} 1 & -f(z) \\ 0 & 1 \end{pmatrix} , \qquad \text{for } |z - x_0| = \delta . \tag{1.51}$$

Obviously, the disc $|z - x_0| < \delta$ can be replaced by another domain D containing x_0. This technique of removing poles was first introduced in [DeiKKZ96].

The problem at hand is more complicated because the number of poles grows in the limit of interest; in this limit the poles accumulate on a fixed set, and thus it is not feasible to surround each with its own circle of fixed size. In [KamMM03] a generalization of the technique described above was developed precisely to allow for the simultaneous removal of a large number of poles in a way that is asymptotically advantageous as the number of poles increases. This generalization employs a single function $f(z)$ with simple poles at $x_{N,n}$, for $n = 0, \dots, N-1$, having corresponding residues $w_{N,n}$, and makes the change of variables (1.50) in a common domain D containing all of the points $x_{N,0}, \dots, x_{N,N-1}$. The essential asymptotic analysis is then related to the nature of the jump condition that generalizes (1.51) for z on the boundary of D. This jump condition can have different asymptotic properties in the limit $N \to \infty$ according to the placement of the boundary of D in the complex plane.

For general queries, contact webmaster@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

Further difficulties arise here because it turns out that the correct location for the boundary of D needed to facilitate the asymptotic analysis in the limit $N \to \infty$ coincides in part with the interval $[a, b]$ that contains the poles, and in the context of the method in [KamMM03] this leads to singularities both in the boundary values of the matrix unknown and also in the jump matrix relating the boundary values. These singularites are an obstruction to further analysis. Therefore the transformation we will introduce from $\mathbf{Q}(z; N, k)$ to $\mathbf{R}(z)$ uses a further variation of the pole removal technique developed in [Mil02] in which two different residue-interpolating functions $f_1(z)$ and $f_2(z)$ are used in respective disjoint domains D_1 and D_2 such that all of the poles $x_{N,n}$ are common boundary points of both domains. This version of the pole removal technique ultimately enables subsequent detailed analysis in the neighborhood of the interval $[a, b]$ in which $\mathbf{Q}(z; N, k)$ has poles.

1(c). $\mathbf{R}(z)$ is transformed into $\mathbf{S}(z)$ by a change of variables that is written explicitly in terms of the equilibrium measure. The equilibrium measure is the solution of a variational problem of logarithmic potential theory that is posed in terms of the functions $\rho^0(x)$ and $V(x)$ given on the interval $[a, b]$ and the constant $c \in (0, 1)$. The fundamental properties of the equilibrium measure are well known in general, and for particular cases of $\rho^0(x)$, $V(x)$, and c, it is not difficult to calculate the equilibrium measure explicitly. The purpose of introducing the equilibrium measure is that the variational problem it satisfies entails some constraints that impose strict inequalities on variational derivatives. These variational derivatives ultimately appear in the problem with a factor of N in certain exponents, and the inequalities lead to desirable exponential decay as $N \to \infty$.

The technique of preparing a matrix Riemann-Hilbert problem for subsequent asymptotic analysis with the introduction of an appropriate equilibrium measure first appeared in the paper [DeiVZ97] and was subsequently applied to the computation of asymptotics for orthogonal polynomials with continuous weights in [DeiKMVZ99a, DeiKMVZ99b]. The key quantity in all of these papers is the complex logarithmic potential of the equilibrium measure, the g-function. In order to apply these methods in the discrete weights context, we need to modify the relationship between the g-function and the equilibrium measure (see (4.5) and (4.7) below) to reflect the local reversal of triangularity described in 1(a) above. This amounts to a further generalization of the technique introduced in [DeiVZ97].

1(d). The final transformation explicitly relates $\mathbf{S}(z)$ to a matrix $\mathbf{X}(z)$. The matrix $\mathbf{S}(z)$ is apparently difficult to analyze in the neighborhood of subintervals of $[a, b]$ where constraints in the variational problem are not active and consequently exponential decay is not obvious. A model for this kind of situation is a matrix $\mathbf{M}(z)$ that takes continuous boundary values on an interval I of the real axis from above (denoted $\mathbf{M}_+(z)$) and below (denoted $\mathbf{M}_-(z)$) that satisfy a jump relation of the form

$$\mathbf{M}_+(z) = \mathbf{M}_-(z) \begin{pmatrix} e^{iN\theta(z)} & 1 \\ 0 & e^{-iN\theta(z)} \end{pmatrix},$$

where $\theta(z)$ is a real-analytic function that is strictly increasing for $z \in I$. This is therefore a rapidly oscillatory jump relation that has no obvious limit as $N \to \infty$. However, noting the algebraic factorization

$$\begin{pmatrix} e^{iN\theta(z)} & 1 \\ 0 & e^{-iN\theta(z)} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ e^{-iN\theta(z)} & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ e^{iN\theta(z)} & 1 \end{pmatrix}$$

and using the analyticity of $\theta(z)$, we may choose some sufficiently small $\epsilon > 0$ and define a new unknown by setting

$$\mathbf{N}(z) := \begin{cases} \mathbf{M}(z) \begin{pmatrix} 1 & 0 \\ -e^{iN\theta(z)} & 1 \end{pmatrix}, & \text{for } \Re(z) \in I \text{ and } 0 < \Im(z) < \epsilon\,, \\ \mathbf{M}(z) \begin{pmatrix} 1 & 0 \\ e^{-iN\theta(z)} & 1 \end{pmatrix}, & \text{for } \Re(z) \in I \text{ and } -\epsilon < \Im(z) < 0\,, \\ \mathbf{M}(z)\,, & \text{otherwise}\,. \end{cases}$$

For general queries, contact webmaster@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

The matrix $\mathbf{N}(z)$ has jump discontinuities along the three parallel contours I, $I + i\epsilon$, and $I - i\epsilon$. If on any of these we indicate the boundary value taken by $\mathbf{N}(z)$ from above as $\mathbf{N}_+(z)$ and from below as $\mathbf{N}_-(z)$, then the oscillatory jump condition for $\mathbf{M}(z)$ in I is replaced by the three different formulae:

$$\mathbf{N}_+(z) = \mathbf{N}_-(z) \begin{pmatrix} 1 & 0 \\ e^{iN\theta(z)} & 1 \end{pmatrix}, \qquad z \in I + i\epsilon\,, \tag{1.52}$$

$$\mathbf{N}_+(z) = \mathbf{N}_-(z) \begin{pmatrix} 1 & 0 \\ e^{-iN\theta(z)} & 1 \end{pmatrix}, \qquad z \in I - i\epsilon\,, \tag{1.53}$$

$$\mathbf{N}_+(z) = \mathbf{N}_-(z) \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \qquad z \in I\,. \tag{1.54}$$

The Cauchy-Riemann equations satisfied by the function $\theta(z)$ in I imply that $\Re(i\theta(z))$ is negative for $\Im(z) = \epsilon$ and positive for $\Im(z) = -\epsilon$. Thus the jump conditions (1.52)–(1.54) all have obvious asymptotics as $N \to \infty$.

The replacement of an oscillatory jump matrix by an exponentially decaying one on the basis of algebraic factorization is the essence of the steepest-descent method for Riemann-Hilbert problems first proposed in [DeiZ93]. Our transformation from $\mathbf{S}(z)$ to $\mathbf{X}(z)$ will be based on this key idea but will involve more complicated factorizations of both upper- and lower-triangular matrices.

These three steps of our analysis of the matrix $\mathbf{P}(z; N, k)$ solving Interpolation Problem 1.2 will be carried out in Chapters 4 and 5.

1.5 OUTLINE OF THE REST OF THE BOOK

Our main results are presented in Chapter 2 (for the discrete orthogonal polynomials themselves) and Chapter 3 (for corresponding applications). The subsequent chapters concern the proof of the main results: Chapters 4, 5, and 6 contain the proof of results stated in Chapter 2, and the results stated in Chapter 3 are proven in Chapter 7.

The detailed asymptotic behavior in the limit $N \to \infty$ of the discrete orthogonal polynomials in overlapping sets that cover the entire complex plane will be discussed in Chapter 2. After some important definitions and notation are established in §2.1 and §2.2, the results themselves will be given in in §2.3. In §2.4 we show how the general theory applies in some classical cases, specifically the Krawtchouk polynomials and two types of polynomials in the Hahn family. The equilibrium measures for the Hahn polynomials are also described in Theorem 2.17.

Further results of our analysis in the context of statistical ensembles associated with families of discrete orthogonal polynomials are discussed in Chapter 3. First, we introduce the notion of a discrete orthogonal polynomial ensemble in §3.1 and describe the ensembles associated with dual polynomials in §3.2. In §3.4, we discuss rhombus tilings of a hexagon as a specific application of discrete orthogonal polynomial ensembles and their duals. Our general results on the universality of various statistics in the limit $N \to \infty$ are explained in §3.3. The specific results implied by the general ones in the context of the hexagon tiling problem are described in §3.4.2. We also obtain new results on the continuum limit of the Toda lattice, which we explain in §3.5.

As mentioned above, Chapters 4 and 5 contain the complete asymptotic analysis of the matrix $\mathbf{P}(z; N, k)$ in the limit $N \to \infty$. This analysis is then used in Chapter 6 to establish the results presented in §2.3 and used again in Chapter 7 to establish the results presented in §3.3.

In Chapter 4 we describe all details of a sequence of algebraic transformations of the interpolation problem for the discrete orthogonal polynomials to arrive at a simpler Riemann-Hilbert problem to which a formal asymptotic analysis can be applied. For this purpose, we exploit a transformation from a Riemann-Hilbert problem with pole conditions (see Interpolation Problem 1.2) to a Riemann-Hilbert problem on a contour, a doubly constrained equilibrium measure, and hole-particle duality. This chapter is the technical core of the analysis of Riemann-Hilbert problems with pole conditions. Chapter 5 concerns the construction of a

For general queries, contact webmaster@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

global parametrix and rigorous error estimates. By combining the calculations in Chapters 4 and 5, we prove the theorems stated in §2.3 in Chapter 6. Using the asymptotic analysis of the Riemann-Hilbert problem discussed in Chapters 4 and 5, we prove in Chapter 7 the theorems stated in §3.3.

Appendix A summarizes construction of the solution of a limiting Riemann-Hilbert problem by means of hyperelliptic function theory. Appendix B gives a proof of the determination of the equilibrium measure of the Hahn weight presented in §2.4. Finally, Appendix C contains a list of some important symbols used frequently throughout the book.

For the asymptotic results given here that correspond to theorems already stated in our announcement [BaiKMM03], we generally obtain significantly sharper error estimates. Since we published that paper, we have learned how to circumvent certain technical difficulties related to the continuum limit of the discrete orthogonality measures and the possibility of transition points where triangularity of residue matrices changes abruptly. In our opinion, these technical innovations do more than make the error estimates sharper; they also make the proofs more elegant.

1.6 RESEARCH BACKGROUND

The work described in this mongraph is connected with three different themes of current research.

First, in the context of approximation theory, there has been recent activity [DeiKMVZ99a, DeiKMVZ99b] in the study of polynomials orthogonal on the real axis with respect to *general* continuous varying weights and the corresponding large-degree pointwise asymptotics. The significance of the work [DeiKMVZ99a, DeiKMVZ99b] is that the method is not at all particular to any special classical formulae for weights; they are completely general. Thus a natural question is whether it is possible to further generalize the method in [DeiKMVZ99a, DeiKMVZ99b] to handle the discrete weights. However, it has turned out that discrete weights are of such a fundamentally different character than their continuous counterparts that this would require the development of new tools for asymptotic analysis. The setting for the work [DeiKMVZ99a, DeiKMVZ99b] is the characterization of the orthogonal polynomials in terms of the solution of a certain matrix-valued Riemann-Hilbert problem [FokIK91] with jump conditions on contours. For discrete weights, the corresponding Riemann-Hilbert problem is defined by constraints on residues of poles. Under the conditions that we consider in this work, each point mass added to the weight amounts to a pole in the matrix solution of the Riemann-Hilbert problem, so analyzing the asymptotics of an accumulation of poles becomes the main difficulty.

Second, there has been some recent progress [KamMM03, Mil02] in the integrable systems literature concerning the problem of computing asymptotics for solutions of integrable nonlinear partial differential equations (*e.g.*, the nonlinear Schrödinger equation) in the limit where the spectral data associated with the solution via the inverse scattering transform is made up of a large number of discrete eigenvalues. Significantly, inverse scattering theory also exploits much of the theory of matrix Riemann-Hilbert problems, and it turns out that the discrete eigenvalues appear as poles in the corresponding matrix-valued unknown. So, the methods recently developed in the context of inverse scattering actually suggest a general scheme by means of which an accumulation of poles in the matrix unknown can be analyzed.

Finally, a number of problems in probability theory have recently been identified that are in some sense solved in terms of discrete orthogonal polynomials, and certain statistical questions can be translated into corresponding questions about the asymptotic behavior of the polynomials. The particular problems we have in mind are related to statistics of random tilings of various shapes, to last-passage percolation models, and also to certain natural measures on sets of partitions. The joint probability distributions in these problems are examples of discrete orthogonal polynomial ensembles [Joh00, Joh01]. Roughly speaking, the analogy is that the relationship between universal asymptotic properties of discrete orthogonal polynomials and universal statistics for discrete orthogonal polynomial ensembles is the same as the relationship between universal asymptotic properties of polynomials orthogonal with respect to continuous weights and universal eigenvalue statistics of certain random matrix ensembles. The techniques required for computing asymptotics of discrete orthogonal polynomials with general weights have become available at just the time when questions that can be answered with these tools are appearing in the applied literature.

For general queries, contact webmaster@press.princeton.edu