Orthogonal Polynomials

Vilmos Totik

11/11/05

Abstract

In this survey, different aspects of the theory of orthogonal polynomials of one (real or complex) variable are reviewed. Orthogonal polynomials on the unit circle are not discussed.

MSC: 42C05, 33C47

Contents

Surveys in Approximation Theory
Volume 1, 2005. pp. 70–125.
Copyright © 2005 Surveys in Approximation Theory.
ISSN 1555-578X
All rights of reproduction in any form reserved.

1 Introduction

The theory of orthogonal polynomials can be divided into two main but only loosely related parts. The two parts have many things in common, and the division line is quite blurred, it is more or less along algebra vs. analysis. One of the parts is the algebraic aspect of the theory, which has close connections with special functions, combinatorics and algebra, and it is mainly devoted to concrete orthogonal systems or hierarchies of systems such as the Jacobi, Hahn, Askey-Wilson, ... polynomials. All the discrete polynomials and the q-analogues of classical ones belong to this theory. We will not treat this part; the interested reader can consult the three recent excellent monographs [39] by M. E. H. Ismail, [28] by W. Gautschi and [6] by G. E. Andrews, R. Askey and R. Roy. Much of the present state of the theory of orthogonal polynomials of several variables lies also close to this algebraic part of the theory. To discuss them would take us too far from our main direction; rather we refer the reader to the recent book [24] by C. F. Dunkl and Y. Xu.

The other part is the analytical aspect of the theory. Its methods are analytical, and it deals with questions that are typical in analysis, or questions that have emerged in and related to other parts of mathematical analysis. General properties fill a smaller part of the analytic theory, and the greater part falls into two main and extremely rich branches: orthogonal polynomials on the real line and on the circle. The richness is due to some special features of the real line and the circle. Classical real orthogonal polynomials, sometimes in other forms like continued fractions, can be traced back to the 18th century, but their rapid development occurred in the 19th and early 20th century. Orthogonal polynomials on the unit circle are much younger, and their existence is largely due to Szegő and Geronimus in the first half of the 20th century. B. Simon's recent treatise [80, 81] summarizes and greatly extends what has happened since then.

The connection of orthogonal polynomials with other branches of mathematics is truly impressive. Without even trying to be complete, we mention continued fractions, operator theory (Jacobi operators), moment problems, analytic functions (Bieberbach's conjecture), interpolation, Padé approximation, quadrature, approximation theory, numerical analysis, electrostatics, statistical quantum mechanics, special functions, number theory (irrationality and transcendence), graph theory (matching numbers), combinatorics, random matrices, stochastic processes (birth and death processes; prediction theory), data sorting and compression, Radon transform and computer tomography.

This work is a survey on orthogonal polynomials that do not lie on the unit circle. Orthogonal polynomials on the unit circle—both the classical theory and recent contributions—will be hopefully dealt with in a companion article.

This work is meant for non-experts, and it therefore contains introductory materials. We have tried to list most of the actively researched fields not directly connected with orthogonal polynomials on the unit circle, but because of space limitation we have only one or two pages on areas where dozens of papers and several books had been published. As a result, our account is necessarily incomplete. Also, the author's personal taste and interest is reflected in the survey, and the omission of a particular direction or a set of results reflects in no way on the importance or quality of the omitted works.

For further backgound on orthogonal polynomials, the reader can consult the books Szegő [91], Simon [80]-[81], Freud [27], Geronimus [34], Gautschi [28], Chicara [18], Ismail [39].

This is a largely extended version of the first part of the article Golinskii–Totik [36].

Acknowledgement. Research was supported by NSF grant DMS-040650 and OTKA T049448, TS44782, and was carried out within the Analysis Research Group of the Hungarian Academy of Sciences. The author participates in the project INTAS 03-51-6637 that supported a visit by Leonid Golinskii to Szeged, during which the outline of this paper was laid out.

2 Orthogonal polynomials

Orthogonal polynomials with respect to measures

Let μ be a positive Borel measure on the complex plane, with an infinite number of points in its support, for which

$$\int |z|^m d\mu(z) < \infty$$

for all $m > 0$. There are unique polynomials

$$p_n(z) = p_n(\mu, z) = \kappa_n z^n + \cdots, \qquad \kappa_n > 0,\ n = 0, 1, \ldots$$

that form an orthonormal system in $L^2(\mu)$, i.e.

$$\int p_m \overline{p_n} d\mu = \begin{cases} 0 & \text{if } m \neq n \\ 1 & \text{if } m = n. \end{cases}$$

These p_n's are called the **orthonormal polynomials** corresponding to μ. κ_n is the **leading coefficient**, and $p_n(z)/\kappa_n = z^n + \cdots$ is called the **monic orthogonal polynomial**. The leading coefficients play a special and important role in the theory, many properties depend on their behavior. When $d\mu(x) = w(x)dx$ on some interval, say, then we talk about orthogonal polynomials with respect to the weight function w.

The p_n's can be easily generated: all we have to do is to make sure that

$$\int \frac{p_n(z)}{\kappa_n} \overline{z^k} d\mu(z) = 0, \qquad k = 0, 1, \ldots, n-1, \tag{2.1}$$

which is an $n \times n$ system of equations for the non-leading coefficients of $p_n(z)/\kappa_n$ with matrix $(\sigma_{i,j})_{i,j=0}^{n-1}$, where

$$\sigma_{i,j} = \int z^i \overline{z^j} d\mu(z)$$

are the **complex moments** of μ. This matrix is nonsingular: if some linear combination with coefficients $c_0, \ldots, c_{n-1}$ of the rows is zero, then the polynomial $P_{n-1}(z) = c_0 + \cdots + c_{n-1}z^{n-1}$ is orthogonal to every z^j, $j < n$, and hence it is orthogonal to itself, i.e.,

$$\int |P_{n-1}|^2 d\mu = \int P_{n-1}\overline{P_{n-1}} d\mu = 0,$$

which implies $P_{n-1}(z) \equiv 0$. Thus, $c_0 = \cdots = c_{n-1} = 0$, which shows the nonsingularity of $(\sigma_{i,j})$. Therefore, the system (2.1) has a unique solution for the non-leading coefficients of $p_n(z)/\kappa_n$ (note that the leading coefficient is 1), and finally κ_n comes from normalization.

In particular, the complex moments already determine the polynomials. In terms of them one can write up explicit determinant formulae:

$$p_n(z) = \frac{1}{\sqrt{D_{n-1}D_n}} \begin{vmatrix} \sigma_{0,0} & \sigma_{0,1} & \cdots & \sigma_{0,n-1} & 1 \\ \sigma_{1,0} & \sigma_{1,1} & \cdots & \sigma_{1,n-1} & z \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \sigma_{n-1,0} & \sigma_{n-1,1} & \cdots & \sigma_{n-1,n-1} & z^{n-1} \\ \sigma_{n,0} & \sigma_{n,1} & \cdots & \sigma_{n,n-1} & z^n \end{vmatrix} \tag{2.2}$$

where

$$D_n = |\sigma_{i,j}|_{i,j=0}^n \tag{2.3}$$

are the so called **Gram determinants**.

Note that if μ is supported on the real line then

$$\sigma_{i,j} = \int x^{i+j} d\mu(x) =: \alpha_{i+j},$$

so $D_n = |\alpha_{i+j}|_{i,j=0}^n$ is a Hankel determinant, while if μ is supported on the unit circle then

$$\sigma_{i,j} = \int z^{i-j} d\mu(z) =: \beta_{i-j},$$

so $D_n = |\beta_{i-j}|_{i,j=0}^n$ is a Toeplitz determinant. In these two important cases the orthogonal polynomials have many special properties that are missing in the general theory.

The Riemann–Hilbert approach

Let μ be supported on the real line, and suppose that it is of the form $d\mu(t) = w(t)dt$ with some smooth function w. A new approach to generating orthogonal polynomials that has turned out to be of great importance was given in the early 1990's by Fokas, Its and Kitaev [26]. Consider 2×2 matrices

$$Y(z) = \begin{pmatrix} Y_{11}(z) & Y_{12}(z) \\ Y_{21}(z) & Y_{22}(z) \end{pmatrix}$$

where the Y_{ij} are analytic functions on $\mathbf{C}\setminus\mathbf{R}$, and solve for such matrices the following matrix-valued **Riemann–Hilbert problem**:

1. for all $x \in \mathbf{R}$

$$Y^{+}(x) = Y^{-}(x)\begin{pmatrix} 1 & w(x) \\ 0 & 1 \end{pmatrix}$$

where Y^{+}, resp. Y^{-}, is the limit of $Y(z)$ as z tends to x from the upper, resp. lower half plane, and

2.

$$Y(z) = \left(I + O\left(\frac{1}{z}\right)\right)\begin{pmatrix} z^n & 0 \\ 0 & z^{-n} \end{pmatrix}$$

at infinity, where I denotes the identity matrix.

Thus, one is looking for 4 functions $Y_{11}, ..., Y_{22}$ analytic on $\mathbf{C}\setminus\mathbf{R}$, such that if $Y_{ij}^{\pm}(x)$ denote the boundary limits of these functions at $x \in \mathbf{R}$ from the upper, resp. lower half plane, then

$$Y_{11}^{-}(x) = Y_{11}^{+}(x), \quad Y_{21}^{-}(x) = Y_{21}^{+}(x) \tag{2.4}$$

and

$$Y_{12}^{-}(x) = Y_{11}^{+}(x)w(x) + Y_{12}^{+}(x), \quad Y_{22}^{-}(x) = Y_{21}^{+}(x)w(x) + Y_{22}^{+}(x). \tag{2.5}$$

These connect the functions on the upper and lower half planes only very mildly, and what puts the problem into a rigid form is the second condition, namely it is required that for large z uniformly on the plane we have

$$Y_{11}(z) = z^n + O(|z|^{n-1}), \quad Y_{21}(z) = O(|z|^{n-1}) \tag{2.6}$$

and

$$Y_{12}(z) = O(z^{-n-1}), \quad Y_{22}(z) = z^{-n} + O(|z|^{-n-1}). \tag{2.7}$$

It can be shown that there is a unique solution $Y(z)$. The relevance of this to orthogonal polynomials is that the entry $Y_{11}(z)$ is precisely the monic polynomial $p_n(\mu, z)/\kappa_n$. Indeed, (2.4) shows that Y_{11} and Y_{12} are analytic everywhere, and if an entire function is $O(|z|^m)$ as $z \to \infty$, then it is a polynomial of degree at most m. Thus, we get from (2.6) that $Y_{11}(z) = z^n + \cdots$ is a monic polynomial of degree n, and $Y_{21}(z)$ is a polynomial of degree at most $n-1$. The relation (2.7) gives that the integral of $z^k Y_{12}(z)$ over the circle $|z| = R$ is $O(R^{k-n})$ for all $k < n$ and hence it tends to 0 as $R \to \infty$. By analyticity, the integral over the upper part of the circle can be deformed into an integral from R to $-R$ on the upper part of $\mathbf{R}$, i.e., into

$$\int_{R}^{-R} x^k Y_{12}^{+}(x)dx,$$

and similarly the integral over the lower part of the circle can be deformed into an integral from $-R$ to R on the lower part of $\mathbf{R}$, i.e., into

$$\int_{-R}^{R} x^k Y_{12}^{-}(x)dx.$$

The first relation in (2.5) implies

$$x^k Y_{12}^-(x) - x^k Y_{12}^+(x) = x^k Y_{11}(x)w(x),$$

therefore for $k = 0, 1, \dots, n-1$ we have

$$\int_{-R}^{R} x^k Y_{11}(x)w(x)dx = O(R^{k-n}) = O(R^{-1}) \to 0,$$

which implies

$$\int_{-\infty}^{\infty} x^k Y_{11}(x)w(x)dx = 0.$$

Thus, Y_{11} is indeed the monic n-th orthogonal polynomial with respect to w.

The other entries can also be explicitly written in terms of the orthogonal polynomials p_n and p_{n-1}: Y_{21} is a constant multiple of p_{n-1},

$$Y_{12}(z) = \frac{1}{2i\pi\kappa_n} \int \frac{p_n(x)w(x)}{x-z} dx$$

is the Cauchy transform of $p_n(x)w(x)/\kappa_n$, and Y_{22} is the Cauchy transform of Y_{21} ($=\mathrm{const}\cdot p_{n-1}$). Furthermore, κ_n and the recurrence coefficients a_n, b_n (see Section 4) can be expressed in terms of the entries of Y_1, where Y_1 is the matrix defined by

$$Y(z)\begin{pmatrix} z^{-n} & 0 \\ 0 & z^n \end{pmatrix} =: I + z^{-1}Y_1 + O\left(\frac{1}{z^2}\right).$$

For details on this Riemann–Hilbert approach, see Deift [20].

Orthogonal polynomials with respect to inner products

Sometimes one talks about orthogonal polynomials with respect to an inner product $\langle \cdot, \cdot \rangle$ which is defined on some linear space containing all polynomials, and orthogonality means $\langle p_n, p_m \rangle = 0$ for $m \neq n$. In this case the aforementioned orthogonalization process can be used, and with $\sigma_{i,j} = \langle x^i, x^j \rangle$, the determinantal formula (2.2) is still valid.

Sometimes one has an $\langle \cdot, \cdot \rangle$ with the standard inner product properties, except that positive definiteness may not hold (as an example consider non-Hermitian orthogonality from Section 14). Then the orthogonalization process and the determinantal formulae can still be used provided the Gram determinants (2.3) are different from zero. If this is not so, then we write

$$p_n(z) = \gamma_n z^n + \gamma_{n-1} z^{n-1} + \cdots,$$

and make sure that p_n is orthogonal to all powers z^k, $0 \le k < n$, i.e., solve the homogeneous system of equations

$$\sum_{j=0}^{n} \gamma_j \sigma_{j,k} = 0, \qquad k = 0, \dots, n-1,$$

for $\gamma_0, \gamma_1, \ldots, \gamma_n$. Since the number of unknowns is bigger than the number of equations, there is always a non-trivial solution, which gives rise to non-trivial orthogonal polynomials. However, then we cannot assert any more $\gamma_n \neq 0$, so the degree of p_n may be smaller than n, and there may be several choices for p_n. Still, in applications where non-Hermitian orthogonality is used, these p_n play the role of orthogonal polynomials.

Varying weights

In the last 25 years orthogonal polynomials with respect to **varying measures** have played a significant role in several problems, see e.g. the sections on exponential and Freud weights or on random matrices in Section 4. In forming them one has a sequence of measures μ_n (generally with some particular behavior), and for each n one forms the orthogonal system $\{p_k(\mu_n, z)\}_{k=0}^{\infty}$. In most cases one needs the behavior of $p_n(\mu_n, z)$ or that of $p_{n\pm k}(\mu_n, z)$ with some fixed k. We mention three examples.

The first example is that of Freud weights: $W(x) = e^{-|x|^\lambda}$, $\lambda > 0$. If one substitutes $x = n^{1/\lambda} y$, then with $P_n(y) = p_n(W, x)$ orthogonality takes the form

$$\int P_n(y) P_m(y) e^{-n|y|^\lambda} dy = 0, \qquad n \neq m,$$

and it turns out that this is just the right scaling, e.g. the zeros of P_n have an asymptotic zero distribution (while those of $p_n(W, z)$ are spreading out to infinity). Thus, studying orthogonal polynomials with respect to Freud weights W is equivalent to studying orthogonal polynomials with respect to the varying weights $w_n(x) = W(x)^n$, and actually, working with w_n turns out to be very natural.

For the second and third examples see multipoint Padé approximation and random matrix theory in Section 6.

Matrix orthogonal polynomials

Orthogonality of matrix polynomials (i.e., when the entries of the fixed size matrix are polynomials of degree $n = 0, 1, \ldots$ and orthogonality is with respect to a matrix measure) is a very active area which shows extreme richness compared to the scalar case. See Section 16 for a short discussion.

3 The L^2 extremal problem

One of the most useful tools in the study of orthogonal polynomials is the fact that they solve the following extremal problem: minimize the $L^2(\mu)$ norm for all monic polynomials $P_n(z) = z^n + \cdots$ of degree n. The minimum turns out to be $1/\kappa_n^2$, i.e., the n-th monic orthogonal polynomial is the (unique) extremal polynomial in

$$\inf_{P_n(z)=z^n+\cdots} \int |P_n|^2 d\mu = \frac{1}{\kappa_n^2}. \tag{3.1}$$

Indeed, any P_n is a linear combination $\sum_{k=1}^n c_k p_k$ with the orthonormal polynomials p_k, and here $c_n = 1/\kappa_n$ because P_n is monic, i.e., it has leading coefficient 1. Orthogonality gives

$$\int |P_n|^2 d\mu = \sum_{k=0}^n |c_k|^2,$$

from which we can see that this is always $\geq |c_n|^2 = 1/\kappa_n^2$, and equality occurs if and only if all the other c_k's are 0.

A related extremum problem leads to the so called **Christoffel functions** associated with μ. They are defined as

$$\lambda_n(\mu, z) = \inf_{P_n(z)=1,\ \deg(P_n)\leq n} \int |P_n|^2 d\mu. \tag{3.2}$$

If we write again $P_n = \sum_{k=0}^n c_k p_k(\mu, \cdot)$, then $P_n(z) = 1$ means

$$\sum_{k=0}^n c_k p_k(\mu, z) = 1,$$

and hence by Cauchy's inequality

$$1 \leq (\sum_{k=0}^n |c_k|^2)(\sum_{k=0}^n |p_k(\mu, z)|^2).$$

Therefore,

$$\int |P_n|^2 d\mu = \sum_{k=0}^n |c_k|^2 \geq (\sum_{k=0}^n |p_k(\mu, z)|^2)^{-1}$$

with equality if and only if

$$c_k = \frac{\overline{p_k(\mu, z)}}{\sum_{k=0}^n |p_k(\mu, z)|^2}.$$

Thus, we have arrived at the formula

$$\lambda_n(\mu, z)^{-1} = \sum_{k=0}^n |p_k(\mu, z)|^2 \tag{3.3}$$

for all $z \in \mathbf{C}$ for the Christoffel function $\lambda_n(\mu, z)$.

For example, for measures μ lying on the real line it is easy to see from this formula that μ has a point mass at x_0, i.e., $\mu(\{x_0\}) > 0$ if and only if $\sum_k p_k(\mu, x_0)^2 < \infty$, and then

$$\mu(\{x_0\}) = (\sum_{k=0}^\infty p_k(\mu, x_0)^2)^{-1}.$$

4 Orthogonal polynomials on the real line

Let μ be supported on the real line. In this case orthogonalization leads to real polynomials (i.e., all the coefficients are real). The most remarkable property of this real case is that the p_n's obey a **three-term recurrence formula**

$$xp_n(x) = a_n p_{n+1}(x) + b_n p_n(x) + a_{n-1} p_{n-1}(x), \tag{4.1}$$

where

$$a_n = \frac{\kappa_n}{\kappa_{n+1}} > 0, \qquad b_n = \int x p_n^2(x) d\mu(x) \tag{4.2}$$

are the so called **recurrence coefficients**. Indeed, if we write $xp_n(x)$ as a linear combination $\sum_{k=0}^{n+1} c_k p_k(z)$ with

$$c_k := \int x p_n(x) p_k(x) d\mu(x),$$

then all the c_k's for $k < n-1$ vanish by orthogonality:

$$c_k = \int x p_n(x) p_k(x) d\mu(x) = \int (x p_k(x)) p_n(x) d\mu(x) = 0$$

because $xp_k(x)$ is a polynomial of degree smaller than n. Comparison of the leading coefficients on both sides gives that $c_{n+1} = \kappa_n/\kappa_{n+1}$, but since c_{n+1} is also the integral of $xp_n(x)p_{n+1}(x)$ against μ, we get that

$$c_{n-1} = \int x p_n(x) p_{n-1}(x) d\mu(x) = \frac{\kappa_{n-1}}{\kappa_n}.$$

Finally, c_n is the integral given in (4.2).

We emphasize that the three-term recurrence is a very special property of real orthogonal polynomials, and it is due to the fact that in this case the polynomials are real, hence

$$\int x p_n(x) \overline{p_k(x)} d\mu(x) = \int p_n(x) \overline{(x p_k(x))} d\mu(x) = 0$$

for $k < n-1$. In the non-real case the two sides here are totally different. The three-term recurrence is missing in the general case, and it is replaced by a different recurrence for polynomials on the circle.

Conversely, any system of polynomials satisfying (4.1) with real $a_n > 0$, b_n is an orthonormal system with respect to a (not necessarily unique) measure on the real line (Favard's theorem). The unicity of the measure in question is the same as the determinacy of the moment problem, which in turn is again closely related to the behavior of orthogonal polynomials; see Section 6.

In the real case the zeros of p_n are real and simple and the zeros of p_n and p_{n+1} interlace, i.e., in between any two zeros of p_{n+1} there is a zero of

p_n. In fact, p_n must have n sign changes, for if it had only $m < n$, say at the points $y_1, \ldots, y_n \in \mathbf{R}$, then it could not be orthogonal to the polynomial $q(x) = \prod_{j=1}^{m}(x - y_j)$ of degree $m < n$, for then $q(x)p_n(x)$ would be of constant sign. Let now $x_n < x_{n-1} < \ldots < x_1$ be the zeros of p_n, and suppose that we already know that the zeros of p_n and p_{n-1} interlace, which implies $\mathrm{sign}(p_{n-1}(x_k)) = (-1)^{k-1}$. If we substitute x_k into the recurrence formula (4.1) then $a_n > 0$ gives that $p_{n+1}(x_k)$ and $p_{n-1}(x_k)$ are of opposite signs at x_k, i.e., $\mathrm{sign}(p_{n+1}(x_k)) = (-1)^k$, and this gives that the zeros of p_n and p_{n+1} also interlace. Thus, the interlacing property follows by induction.

The three-term recurrence implies for the so called **reproducing kernel** the **Christoffel-Darboux formula**

$$\sum_{k=0}^{n} p_k(x)p_k(t) = \frac{\kappa_n}{\kappa_{n+1}} \frac{p_{n+1}(x)p_n(t) - p_n(x)p_{n+1}(t)}{x - t}. \tag{4.3}$$

Indeed, use the recurrence formula for p_{n+1} on the right and $a_n = \kappa_n/\kappa_{n+1}$; then induction gives (4.3). The special case

$$\lambda_n(\mu, x)^{-1} = \sum_{k=0}^{n} p_k(x)^2 = \frac{\kappa_n}{\kappa_{n+1}} \left(p'_{n+1}(x)p_n(x) - p'_n(x)p_{n+1}(x)\right) \tag{4.4}$$

is worth mentioning.

The starting values of the recurrence (4.1) are $p_{-1} \equiv 0$, $p_0 = (\mu(\mathbf{C})))^{-1/2}$. If one starts from $q_{-1} = -1$, $q_0 \equiv 0$ and uses the same recurrence (with $a_{-1} = 1$)

$$xq_n(x) = a_n q_{n+1}(x) + b_n q_n(x) + a_{n-1}q_{n-1}(x), \tag{4.5}$$

then q_n is of degree $n - 1$, and by Favard's theorem the different q_n's are orthogonal with respect to some measure. The q_n's are called **orthogonal polynomials of the second kind** (sometimes for p_n we say that they are **of the first kind**). They can also be written in the form

$$q_n(z) = (\mu(\mathbf{C}))^{-1/2} \int \frac{p_n(z) - p_n(x)}{z - x} d\mu(x).$$

5 Classical orthogonal polynomials

These are

- **Jacobi polynomials** $P_n^{(\alpha,\beta)}$, $\alpha, \beta > -1$, orthogonal with respect to the weight $(1 - x)^\alpha (1 + x)^\beta$ on $[-1, 1]$,

- **Laguerre polynomials** $L_n^{(\alpha)}$, $\alpha > -1$, with orthogonality weight $x^\alpha e^{-x}$ on $[0, \infty)$,

- **Hermite polynomials** H_n orthogonal with respect to e^{-x^2} on the real line $(-\infty, \infty)$.

In the literature various normalizations are used for them.

They are very special, for they possess many properties that no other orthogonal polynomial system does. In particular,

- they have derivatives which form again an orthogonal polynomial system, e.g. the derivative of $P_n^{(\alpha,\beta)}$ is a constant multiple of $P_{n-1}^{(\alpha+1,\beta+1)}$:

$$(P_n^{(\alpha,\beta)})'(x) = \frac{1}{2}(n+\alpha+\beta+1)P_{n-1}^{(\alpha+1,\beta+1)}(x),$$

- they all possess a Rodrigues type formula

$$P_n(x) = \frac{1}{d_n w(x)} \frac{d^n}{dx^n}\{w(x)\sigma(x)^n\},$$

where w is the weight function and σ is a polynomial that is independent of n, for example,

$$L_n^{(\alpha)}(x) = e^x x^{-\alpha} \frac{1}{n!} \frac{d^n}{dx^n}\left(e^{-x}x^{n+\alpha}\right),$$

- they satisfy a differential-difference relation of the form

$$\pi(x)P_n'(x) = (\alpha_n x + \beta_n)P_n(x) + \gamma_n P_{n-1}(x),$$

e.g.

$$x(L_n^{(\alpha)})'(x) = nL_n^{(\alpha)}(x) - (n+\alpha)L_{n-1}^{(\alpha)}(x),$$

- they satisfy a non-linear equation of the form

$$\begin{aligned} \sigma(x)\left(P_n(x)P_{n-1}(x)\right)' &= (\alpha_n x + \beta_n)P_n(x)P_{n-1}(x) \\ &\quad + \gamma_n P_n^2(x) + \delta_n P_{n-1}^2(x), \end{aligned}$$

with some constants $\alpha_n, \beta_n, \gamma_n, \delta_n$, and σ a polynomial of degree at most 2, e.g.

$$(H_n(x)H_{n-1}(x))' = 2xH_n(x)H_{n-1}(x) - H_n^2(x) + 2nH_{n-1}^2(x).$$

Every one of these properties has a converse, namely if a system of orthogonal polynomials possesses any of these properties, then it is (up to a change of variables) one of the classical systems, see Al-Salam [3]. See also Bochner's result in the next section claiming that the classical orthogonal polynomials are essentially the only polynomial (not just orthogonal polynomial) systems that satisfy a certain second order differential equation.

Classical orthogonal polynomials are also special in the sense that they possess a relatively simple

- second order differential equation, e.g.

$$xy'' + (\alpha + 1 - x)y' + ny = 0$$

  for $L_n^{(\alpha)}$,

- generating function, e.g.

$$\sum_n \frac{H_n(x)}{n!} w^n = \exp(2xw - w^2),$$

- integral representation, e.g.

$$(1-x)^\alpha (1+x)^\beta P_n^{(\alpha,\beta)}(x) = \frac{(-1)^n}{2^{n+1}\pi i} \int (1-t)^{n+\alpha}(1+t)^{n+\beta}(t-x)^{-n-1}dt$$

  over an appropriate contour,

and these are powerful tools to study their behavior.

For all these results see Szegő [91].

6 Where do orthogonal polynomials come from?

In this section we mention a few selected areas where orthogonal polynomials naturally arise.

Continued fractions

Continued fractions played an extremely important role in the development of several branches of mathematics, but their significance has unjustly diminished in modern mathematics. A **continued fraction** is of the form

$$\frac{B_1}{A_1 + \frac{B_2}{A_2 + \cdots}},$$

and its **n-th convergent** is

$$\frac{S_n}{R_n} = \frac{B_1}{A_1 + \frac{B_2}{A_2 + \cdots \frac{B_n}{A_n}}}, \quad n = 1, 2, \ldots.$$

The value of the continued fraction is the limit of its convergents. The denominators and numerators of the convergents satisfy the three-term recurrence relations

$$\begin{aligned} R_n &= A_n R_{n-1} + B_n R_{n-2}, \qquad R_0 \equiv 1,\ R_{-1} \equiv 0 \\ S_n &= A_n S_{n-1} + B_n S_{n-2}, \qquad S_0 \equiv 0,\ S_{-1} \equiv 1, \end{aligned}$$

which immediately connects continued fractions with three-term recurrences and hence with orthogonal polynomials.

But the connection is deeper than just this formal observation. Many elementary functions (like $z - \sqrt{z^2 - 1}$) have a continued fraction development where the B_n's are constants while the A_n's are linear functions, in which case the convergents are ratios of some orthogonal polynomials of the second and first kind. An important example is that of Cauchy transforms of measures μ with compact support on the real line (so called **Markov functions**):

$$f(z) = \int \frac{d\mu(x)}{x - z} = -\frac{\alpha_0}{z} - \frac{\alpha_1}{z^2} - \dots . \tag{6.1}$$

The coefficients α_j in the development of (6.1) are the moments

$$\alpha_j = \int x^j \, d\mu(x), \quad j = 0, 1, \dots$$

of the measure μ. The continued fraction development

$$f(z) \sim \frac{B_1}{z - A_1 + \frac{B_2}{z - A_2 + \cdots}}$$

of f at infinity converges locally uniformly outside the smallest interval that contains the support of μ (A. Markov's theorem).

As has been mentioned, the numerators $S_n(z)$ and the denominators $R_n(z)$ of the n-th convergents

$$\frac{S_n(z)}{R_n(z)} = \frac{B_1}{z - A_1 + \frac{B_2}{z - A_2 + \cdots \frac{B_n}{z - A_n}}}, \quad n = 1, 2, \dots$$

satisfy the recurrence relations

$$\begin{aligned} R_n(z) &= (z - A_n)R_{n-1}(z) + B_n R_{n-2}(z), \quad R_0 \equiv 1,\ R_{-1} \equiv 0 \qquad (6.2) \\ S_n(z) &= (z - A_n)S_{n-1}(z) + B_n S_{n-2}(z), \quad S_0 \equiv 0,\ S_{-1} \equiv 1. \end{aligned}$$

These are precisely the recurrence formulae for the monic orthogonal polynomials of the first and second kind with respect to μ, hence the n-th convergent is $cq_n(z)/p_n(z)$ with $c = \mu(\mathbf{C})^{1/2}$.

See Szegő [91, pp. 54–57] as well as Kruschev [42] and the numerous references there.

Padé approximation and rational interpolation

One easily gets from the recurrence relations (6.2) that

$$\frac{S_m(z)}{R_m(z)} - \frac{S_{m+1}(z)}{R_{m+1}(z)} = (-1)^n \frac{B_1 B_2 \cdots B_{n+1}}{R_n(z) R_{n+1}(z)},$$

and summation of these for $m = n, n+1, \dots$ yields that

$$\frac{S_n(z)}{R_n(z)} = \sum_{k=0}^{2n} \frac{-\alpha_k}{z^{k+1}} + O(z^{-2n-1}),$$

i.e., with the preceding notation the rational function

$$S_n(z)/R_n(z) = cq_n(z)/p_n(z) \quad \text{with } c = \mu(\mathbf{C})^{1/2}$$

of numerator degree $n-1$ and of denominator degree n interpolates $f(z)$ at infinity to order $2n$. This is the analogue (called $[n-1/n]$ **Padé approximant**) of the n-th Taylor polynomial (which interpolates the function to order n) for rational functions. The advantage of Padé approximation over Taylor polynomials lies in the fact that the poles of Padé approximants may imitate the singularities of the function in question, while Taylor polynomials are good only up to the first singularity. The error in $[n-1/n]$ Padé approximation has the form

$$f(z) - c\frac{q_n(z)}{p_n(z)} = \frac{1}{p_n^2(z)} \int \frac{p_n^2(x)}{x-z} d\mu(x).$$

Orthogonal polynomials appear in more general rational interpolation (called **multipoint Padé approximation**) to Markov functions, see e.g. Stahl–Totik [87, Sec. 6.1]. For every n select a set $A_n = \{x_{0,n}, \dots, x_{2n,n}\}$ of $2n+1$ interpolation points from $\overline{\mathbf{C}} \setminus I$ where I is the smallest interval that contains the support of μ. The points need not be distinct, but we assume that A_n is symmetric with respect to the real line. Put

$$\omega_n(z) := \prod_{\substack{j=0 \\ x_{jn} \neq \infty}}^{2n} (z - x_{jn}).$$

The degree of ω_n is equal the number of finite points in A_n. By $r_n(z) = u_n(z)/Q_n(z)$ we denote a rational function of numerator and denominator degree at most n that interpolates the function f at the $2n+1$ points of the set $A_n = \{x_{0,n}, \dots, x_{2n,n}\}$ in the sense that

$$\frac{f(z) - r_n(f, A_n; z)}{\omega_n(z)} = O(z^{-(2n+1)}) \quad \text{as} \quad |z| \to \infty;$$

the expression on the left is bounded at every finite point of A_n, and at infinity it has the indicated behavior. Now for Markov functions this rational interpolant uniquely exists, Q_n is the n-th orthogonal polynomial with respect to the varying weight $d\mu(x)/\omega_n(x)$, and the remainder term of the interpolation has the representation

$$(f - r_n(f, A_n; \cdot))(z) = \frac{\omega_n(z)}{Q_n^2(z)} \int \frac{Q_n^2(x)}{\omega_n(x)(x-z)} d\mu(x)$$

for all z outside the support of μ. Thus, the rate of convergence of the rational interpolants is intimately connected with the behavior of the orthogonal polynomials with respect to the varying weight $d\mu(x)/\omega_n(x)$.

Moment problem

The **moments** of a measure μ, $\mu(\mathbf{C}) = 1$, supported on the real line, are

$$\alpha_n = \int x^n d\mu(x), \qquad n = 0, 1, \ldots.$$

The Hamburger moment problem is to determine if a sequence $\{\alpha_n\}$ (with normalization $\alpha_0 = 1$) of real numbers is the moment sequence of a measure with infinite support, and if this measure is unique (the Stieltjes moment problem asks the same, but for measures on $[0, \infty)$). The existence is easy: $\{\alpha_n\}$ are the moments of some measure supported on $\mathbf{R}$ if and only if all the Hankel determinants $|\alpha_{i+j}|_{i,j=0}^m$, $m = 0, 1, \ldots$, are positive. The unicity (usually called determinacy) depends on the behavior of the orthogonal polynomials (2.2) defined from the moments $\sigma_{i,j} = \alpha_{i+j}$. In fact, there are different measures with the same moments α_j if and only if there is a non-real z_0 with $\sum_n |p_n(z_0)|^2 < \infty$, which in turn is equivalent to $\sum_n |p_n(z)|^2 < \infty$ for all $z \in \mathbf{C}$. Furthermore, the Cauchy transforms of all solutions ν of the moment problem have the parametric form

$$\int \frac{d\nu(x)}{z - x} = \frac{C(z)F(z) + A(z)}{D(z)F(z) + B(z)},$$

where F is an arbitrary analytic function (the parameter) mapping the upper half plane $\mathbf{C}_+$ into $\overline{\mathbf{C}}_+ \cup \{\infty\}$, and A, B, C and D have explicit representations in terms of the first and second kind orthogonal polynomials p_n and q_n:

$$A(z) = z \textstyle\sum_n q_n(0) q_n(z); \qquad B(z) = -1 + z \textstyle\sum_n q_n(0) p_n(z);$$

$$C(z) = 1 + z \textstyle\sum_n p_n(0) q_n(z); \quad D(z) = z \textstyle\sum_n p_n(0) p_n(z).$$

For all these results see Akhiezer [2], and for an operator theoretic approach to the moment problem see Simon [78] (in particular, Theorems 3 and 4.14).

Jacobi matrices and spectral theory of self-adjoint operators

Tridiagonal, so called **Jacobi matrices**

$$J = \begin{pmatrix} b_0 & a_0 & 0 & 0 & \cdots \\ a_0 & b_1 & a_1 & 0 & \cdots \\ 0 & a_1 & b_2 & a_2 & \cdots \\ 0 & 0 & a_2 & b_2 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

with bounded $a_n > 0$ and bounded real b_n define a bounded self-adjoint operator J in l_2, a so called **Jacobi operator**. These are the discrete analogues of second order linear differential operators of Schrödinger type on the half line. Every bounded self adjoint operator with a cyclic vector is a Jacobi operator in an appropriate base.

The formal eigen-equation $J\pi = \lambda\pi$ is equivalent to the three-term recurrence

$$a_{n-1}\pi_{n-1} + b_n\pi_n + a_n\pi_{n+1} = \lambda\pi_n, \qquad n = 1, 2, \ldots,$$

$$b_0\pi_0 + a_0\pi_1 = \lambda\pi_0, \qquad \pi_0 = 1.$$

Thus, $\pi_n(\lambda)$ is of degree n in λ.

By the spectral theorem, J, as a self-adjoint operator having a cyclic vector $((1, 0, 0, \ldots))$, is unitarily equivalent to multiplication by x in some $L^2(\mu)$ with some probability measure μ having compact support on the real line. This μ is called the spectral measure associated with J (and with its spectrum). More precisely, if $p_n(x) = p_n(\mu, x)$ are the orthonormal polynomials with respect to μ, and U maps the unit vector $e_n = (0, \ldots, 0, 1, 0, \ldots)$ into p_n, then U can be extended to a unitary operator from l_2 onto $L^2(\mu)$, and if $Sf(x) = xf(x)$ is the multiplication operator by x in $L^2(\mu)$, then $J = U^{-1}SU$. The recurrence coefficients for $p_n(\mu, x)$ are precisely the a_n's and b_n's from the Jacobi matrix, i.e., $p_n(x) = c\pi_n(x)$ with some fixed constant c. These show that Jacobi operators are equivalent to multiplication by x in $L^2(\mu)$ spaces if the particular basis $\{p_n(\mu, \cdot)\}$ is used (see e.g. Deift [20, Ch. 2]).

The truncated $n \times n$ matrix

$$J_n = \begin{pmatrix} b_0 & a_0 & 0 & 0 & \cdots \\ a_0 & b_1 & a_1 & 0 & \cdots \\ 0 & a_1 & b_2 & a_2 & \cdots \\ \vdots & \vdots & \vdots & \ddots & \ldots \\ 0 & 0 & 0 & a_{n-2} & b_{n-1} \end{pmatrix}$$

has n real and distinct eigenvalues, which turn out to be the zeros of p_n, i.e., the monic polynomial $p_n(z)/\kappa_n$ is the characteristic polynomial of J_n.

Quadrature

For a measure μ, an **n-point quadrature** (rule) is a sequence of n points $x_{n,1}, \ldots, x_{n,n}$ and a sequence of associated numbers $\lambda_{n,1}, \ldots, \lambda_{n,n}$. It is expected that

$$\int f d\mu \sim \sum_{k=1}^{n} \lambda_{n,k} f(x_{n,k})$$

in some sense for as large a class of functions as possible. Often the accuracy of the quadrature is measured by its exactness, which is defined as the largest m

such that the quadrature is exact for all polynomials of degree at most m, i.e., m is the largest number with the property that

$$\int x^j \, d\mu(x) = \sum_{k=1}^{n} \lambda_{n,k} x_{n,k}^j \qquad \text{for all } 0 \le j \le m.$$

For μ with support on the real line and for quadrature based on n points this exactness m cannot be larger than $2n-1$, and this optimal value $2n-1$ is attained if and only if $x_{n,1}, \ldots, x_{n,n}$ are the zeros of the orthonormal polynomial $p_n(\mu, x)$ corresponding to μ, and the so called Cotes numbers $\lambda_{n,k}$ are chosen to be

$$\lambda_{n,k} = \lambda_n(\mu, x_{n,k}) = \left(\sum_{j=0}^{n} p_j(\mu, x_{n,k})^2 \right)^{-1},$$

where $\lambda_n(\mu, z)$ is the Christoffel function (3.2) associated with μ.

See Szegő [91, Ch. XV].

Random matrices

Some statistical-mechanical models in quantum systems use random matrices. Let $\mathcal{H}_n$ be the set of all $n \times n$ Hermitian matrices $M = (m_{i,j})_{i,j=1}^n$, and let there be given a probability distribution on $\mathcal{H}_n$ of the form

$$P_n(M) dM = D_n^{-1} \exp(-n \mathrm{Tr}\{V(M)\}) dM,$$

where $V(\lambda)$, $\lambda \in \mathbf{R}$, is a real-valued function that increases sufficiently fast at infinity (typically an even polynomial in quantum field theory applications), $\mathrm{Tr}\{H\}$ denotes the trace of the matrix H,

$$dM = \prod_{k=1}^{n} dm_{k,k} \prod_{k<j} d\,\Re m_{k,j} \; d\,\Im m_{k,j}$$

is the "Lebesgue" measure for Hermitian matrices, and D_n is a normalizing constant so that the total integral of $P_n(M)dM$ is one.

Every matrix $M \in \mathcal{H}_n$ has n real eigenvalues which carry physical information on the system when it is in the state described by M. The quantity

$$N_n(\mathbf{D}) = \frac{\#\{\text{eigenvalues in } \mathbf{D}\}}{n}$$

is the random variable that equals the normalized number of eigenvalues in the interval $\mathbf{D}$. This model is known as the unitary ensemble associated with V.

Let $p_j(w^n, x)$ be the orthonormal polynomials with respect to the varying weight $w^n(x)$, $w(x) = \exp(-V(x))$. Then the joint probability density of the eigenvalues can be written in the form

$$d_n \left| p_{i-1}(w^n, \lambda_j) w^{n/2}(\lambda_j) \right|^2_{1 \le i,j \le n},$$

where d_n is a normalizing constant built up from the leading coefficients of the $p_j(w^n, \cdot)$. With the so called weighted reproducing kernel

$$K_n(t,s) = \sum_{j=0}^{n-1} p_j(w^n, t) w^{n/2}(t)\, p_j(w^n, s) w^{n/2}(s),$$

it can also be written in the form

$$\frac{1}{n!} \left|K_n(\lambda_i, \lambda_j)\right|_{1\le i,j\le n}.$$

In particular, for the expected number of eigenvalues in an interval $\mathbf{D}$ we have

$$EN_n(\mathbf{D}) = \int_{\mathbf{D}} \frac{K_n(\lambda,\lambda)}{n}\, d\lambda,$$

where $1/K_n(\lambda,\lambda)$ is known in the theory of orthogonal polynomials as the n-th (weighted) Christoffel function associated with the weight w^n, while the limit of the left hand side (as $n \to \infty$) is known as the density of states.

See, e.g., Mehta [60] and Pastur–Figotin [68].

7 Some questions leading to classical orthogonal polynomials

There are almost an infinite number of problems where classical orthogonal polynomials emerge. Let us just mention a few.

Electrostatics

Put at 1 and -1 two positive charges p and q, and with these fixed charges put n positive unit charges on $[-1, 1]$ at the points $x_1, \ldots, x_n$. On the plane the Coulomb force is proportional with the reciprocal of the distance, and so a charge generates a logarithmic potential field. Therefore, the mutual energy of all these charges is

$$I(x_1,\ldots,x_n) = p\sum_{j=1}^{n} \log\frac{1}{|1-x_j|} + q\sum_{j=1}^{n}\log\frac{1}{|1+x_j|} + \sum_{i<j}\log\frac{1}{|x_i-x_j|},$$

and the equilibrium problem asks for finding $x_1, \ldots, x_n$ for which this energy is minimal. The unique minimum occurs (see Szegő [91, Section 6.7]) for the zeros of the Jacobi polynomial $P_n^{(2p-1,2q-1)}$ orthogonal with respect to the weight $(1-x)^{2p-1}(1+x)^{2q-1}$.

There is a similar characterization of the zeros of Laguerre and Hermite polynomials, and even of more general orthogonal polynomials (for the latter see Ismail [39, Section 3.5]).

Polynomial solutions of eigenvalue problems

Consider the eigenvalue problem

$$f(x)\frac{d^2}{dx^2}y(x) + g(x)\frac{d}{dx}y(x) + h(x)y(x) = \lambda y(x),$$

where f, g, h are fixed polynomials and λ is a free constant, and it is required that this have a polynomial solution of exact degree n for all $n = 0, 1, \ldots$, for which the corresponding λ and $y(x)$ will be denoted by λ_n and $y_n(x)$, respectively. Bochner's theorem from [16] states that, except for some trivial solutions of the form $y(x) = ax^n + bx^m$ and for some polynomials related to Bessel functions, the only solutions are (in all of them we can take $h(x) = 0$)

- Jacobi polynomials $P_n^{(\alpha,\beta)}$ ($f(x) = 1 - x^2$, $g(x) = \beta - \alpha - x(\alpha + \beta + 2)$, $\lambda_n = -n(n + \alpha + \beta + 1)$)
- Laguerre polynomials $L_n^{(\alpha)}$ ($f(x) = x$, $g(x) = 1 + \alpha - x$, $\lambda_n = -n$) and
- Hermite polynomials $H_n(x)$ ($f(x) = 1$, $g(x) = -2x$, $\lambda_n = -2n$).

Harmonic analysis on spheres and balls

Harmonic analysis on spheres and balls in $\mathbf{R}^d$ is based on spherical harmonics, i.e., harmonic homogeneous polynomials. In this theory, special Jacobi polynomials, the so called **ultraspherical** or **Gegenbauer polynomials** $P_n^{(\alpha)}$, play a fundamental role – they are orthogonal with respect to the weight $(1-x^2)^{\alpha-1/2}$.

Let S^{d-1} be the unit sphere in $\mathbf{R}^d$ and let $\mathcal{H}_n^d$ be the restriction to S^{d-1} of all harmonic polynomials $Q(x_1, \ldots, x_n)$ of d variables that are homogeneous of degree n, i.e.,

$$\sum_{k=1}^{n} \frac{\partial^2}{\partial x_k^2} Q = 0, \qquad Q(\lambda x_1, \ldots, \lambda x_n) = \lambda^n Q(x_1, \ldots, x_n), \quad \lambda > 0.$$

The dimension of $\mathcal{H}_n^d$ is

$$\binom{n+d-1}{d-1} - \binom{n+d-3}{d-1},$$

and an orthogonal basis in it can be produced as follows. With $\rho = x_{d-1}^2 + x_d^2$ let $g_{s,0} = \rho^s P_s^{(0)}(x_{d-1}/\rho)$ and $g_{s,1} = x_d \rho^s P_s^{(1)}(x_{d-1}/\rho)$. With $n_d = 0$ or $n_d = 1$ consider all multiindices $\mathbf{n} = (n_1, n_2, \ldots, n_d)$ such that $n_1 + \cdots + n_d = n$, and if for such a multiindex we define the function $Y_{\mathbf{n}}(x_1, \ldots, x_d)$ as

$$g_{n_{d-1},n_d} \prod_{j=1}^{d-2} \left((x_j^2 + \cdots + x_d^2)^{n_j} P_{n_j}^{(\lambda_j)}(x_j(x_j^2 + \cdots + x_d^2)^{-1/2}) \right),$$

then these $Y_{\mathbf{n}}$ constitute an orthogonal basis in $\mathcal{H}_n^d$ (see e.g. Dunkl–Xu [24, p. 35]).

If $\underline{x} = (x_1, \ldots, x_n)$ and $\langle \underline{x}, \underline{y} \rangle = \sum_k x_k y_k$ is the inner product in $\mathbf{R}^d$, then the reproducing kernel for these spherical polynomials is $P_n^{(d-2)/2}(\langle \underline{x}, \underline{y} \rangle)$ in the sense that for all $Q \in \mathcal{H}_n^d$ and for all $x \in S^{d-1}$ we have

$$c_{n,d} \int P_n^{(d-2)/2}(\langle \underline{x}, \underline{y} \rangle) Q(\underline{y}) d\sigma(\underline{y}) = Q(\underline{x}),$$

where integration is with respect to surface area, and $c_{n,d}$ is an explicit normalizing constant (see e.g. Dunkl–Xu [24, p. 37]).

As a result, Gegenbauer polynomials are all over the theory of spherical harmonics, as well as in the corresponding theory for the unit ball.

Approximation theory

In the literature, expansions of functions into classical orthogonal polynomial series are second only to trigonometric expansions, and numerous works have been devoted to their convergence and approximation properties, see e.g. Szegő [91, Ch. XIII].

The **Chebyshev polynomials** $\cos(n \arccos x)$ are orthogonal on $[-1, 1]$ with respect to the weight $w(x) = (1 - x^2)^{-1/2}$. These directly correspond to trigonometric functions, and expansions into them have virtually the same properties as trigonometric Fourier expansions. But there are many other aspects of approximation where Chebyshev polynomials appear. If one considers, for example, the best approximation on $[-1, 1]$ of x^n in the uniform norm by polynomials $P_{n-1}(x)$ of smaller degree then the smallest error appears when $x^n - P_{n-1}(x) = 2^{1-n} \cos(n \arccos x)$ is the monic n-th Chebyshev polynomial. Actually, monic Chebyshev polynomials minimize all $L^p(w)$, $p > 0$, norms among monic polynomials of a given degree.

As we have seen in (3.1), the monic orthogonal polynomials $p_n(\mu)/\kappa_n$ are the solutions to the extremal problem

$$\int |P_n|^2 d\mu \to \min, \tag{7.3}$$

where the minimum is taken for all monic polynomials of degree n. This extremal property makes orthogonal polynomials, in particular Chebyshev polynomials, indispensable tools in approximation theory.

Lagrange interpolation and its various generalizations like Hermite-Fejér or Hermite interpolation etc. is mostly done on the zeros of some orthogonal polynomials. In fact, these nodes are often close to optimal in the sense that the Lebesgue constant increases at the optimal rate. In many cases interpolation on zeros of orthogonal polynomials has special properties due to explicitly calculable expressions. Recall e.g. Fejér's result that if P_{2n-1} is the unique polynomial of degree at most $2n-1$ that interpolates a continuous function f at the nodes of the n-th Chebyshev polynomial and that has zero derivative at each of these nodes, then P_{2n-1} uniformly converges to f on $[-1,1]$ as $n \to \infty$. For the role of orthogonal polynomials in interpolation see the books Szabados–Vértesi [90] and Mastroianni–Milovanovic [55].

8 Heuristics

In this section we do not state precise results. We just want to indicate some heuristics on the behavior of orthogonal polynomials. For the concepts below, as well as for a more precise form of some of the heuristics see the following sections, in particular Section 9.

As we have seen, the monic orthogonal polynomials $p_n(\mu)/\kappa_n$ minimize the $L^2(\mu)$ norm in (7.3). Therefore, these polynomials try to be small where the measure is large, e.g. one expects the zeros to cluster at the support $S(\mu)$ of μ. The example of arc measure on the unit circle, for which the orthogonal polynomials are z^n, shows however, that this is not true (due to the fact that the complement of the support is not connected). The statement is true when the support lies on $\mathbf{R}$ or on some systems of arcs, and also in the general case when instead of the support one considers the polynomial convex hull of the support of μ (for the definition, see the next section): on any compact set outside the polynomial convex hull there can only be a fixed number of zeros of $p_n(\mu)$ for every n. When the complement of $S(\mu)$ is connected and $S(\mu)$ has no interior, then the distribution of the zeros shows a remarkable universality and indifference to the size of μ. In many situations the distribution of the zeros is the equilibrium distribution of the support $S(\mu)$. When $S(\mu) = [-1,1]$, this means that under very weak assumptions the zero distribution is always the arcsine distribution $dx/\pi\sqrt{1-x^2}$.

The $L^2(\mu)$ minimality of $p_n(\mu)/\kappa_n$ in the sense of (7.3) is something like minimality in the L^∞ norm on $S(\mu)$. Therefore, $p_n(\mu)/\kappa_n$ should behave like the monic polynomial T_n minimizing the L^∞ norm on $S(\mu)$ (so called Chebyshev polynomials for $S(\mu)$). Since

$$\frac{1}{n}\log|T_n(z)| = \int \log|z-t|d\nu_n(t)$$

where ν_n has mass $1/n$ at each zero of T_n, in the limit the behavior should be like

$$U^\nu(z) = \int \log|z-t|d\nu(t), \tag{8.1}$$

where ν is the probability measure on $S(\mu)$ for which the maximum of U^ν on $S(\mu)$ is as small as possible (this is the so called equilibrium measure of $S(\mu)$). More generally, if $d\nu = d\nu_n = w^n(x)dx$ is a varying weight in the specified way, then the same reasoning leads to a behavior like (8.1), but now ν is a measure for which the supremum of $U^\nu(z) + \log w(z)$ is as small as possible (weighted equilibrium measure).

Universal behavior can also be seen for the polynomials themselves. Usually they obey

$$\frac{1}{n}\log|p_n(\mu,z)| \to g_{\mathbf{C}\setminus S(\mu)}(z,\infty), \qquad z \notin S(\mu), \tag{8.2}$$

where $g_{\mathbf{C}\setminus S(\mu)}(z,\infty)$ is the Green function with pole at infinity associated with the complement of the support. When the unbounded component of the complement of $S(\mu)$ is simply connected, then in that component often there is a finer asymptotic behavior of $p_n(\mu)$ of the form

$$p_n(z) \sim d_n g_\mu(z)\Phi(z)^n, \qquad z \notin S(\mu), \tag{8.3}$$

where Φ is the mapping function that maps $\mathbf{C}\setminus S(\mu)$ conformally onto the outside of the unit disk, and g_μ is a function (might be called generalized Szegő function) that depends on μ. Such a fine asymptotic is restricted to the simply connected case, see e.g. Section 10.

Asymptotics of orthogonal polynomials have a hierarchy, and the different types of asymptotics usually require the measure to be sufficiently strong with different degree on its support. Consider first the case of compact support $S(\mu)$. The weakest is **n-th root asymptotics** stating the behavior (8.2) for $|p_n(\mu,z)|^{1/n}$ outside the support of the measure. It is mostly equivalent to a corresponding distribution of the zeros, as well as asymptotical minimal behavior of $\kappa_n^{1/n}$. It holds under very weak assumptions on the measure, roughly stating that the logarithmic capacity of the points where $\mu' > 0$ (derivative with respect to equilibrium measure), be the same as the capacity of $S(\mu)$. **Ratio asymptotics**, i.e., asymptotic behavior of $p_{n+1}(\mu,z)/p_n(\mu,z)$, is stronger, and is equivalent with asymptotics for the ratio κ_{n+1}/κ_n of consecutive leading coefficients. It can only hold when $\mathbf{C}\setminus S(\mu)$ (more precisely its unbounded component) is simply connected, and in this case it is enough that $\mu' > 0$ almost everywhere with respect to the equilibrium measure of the support of μ (see Section 10). Finally, **strong asymptotics** of the form (8.3) needs roughly that $\log\mu'$ be integrable (Szegő condition, see Section 10).

All these are outside the support. On the support the orthogonal polynomials are of oscillatory behavior, and in the real case under smoothness assumptions on the measure often a so called Plancherel-Rotach type asymptotic formula

$$p_n(\mu,x) \sim d_n g(x)\sin(nh(x) + H(x))$$

holds, where g, h, H are fixed functions. Here $h(x)$ is directly linked with the zeros, h'/π is precisely the distribution of the zeros. When $S(\mu) = [-1,1]$ and the measure is smooth, then $h(x) = \arccos x$.

When $S(\mu)$ is not of compact support (like Laguerre, Hermite or Freud weights), then usually the zeros are spreading out, and one has to rescale them to $[-1,1]$ (or to $[0,1]$) to get a distribution, which is mostly NOT the arcsine distribution. In a similar fashion, various asymptotics hold for the polynomials only after a corresponding rescaling.

The Christoffel function (3.2) or, what is the same, the square sum

$$\frac{1}{\lambda_n(\mu, z)} = \sum_{k=0}^{n} |p_k(\mu, z)|^2,$$

behaves much more regularly than the orthogonal polynomials. Outside the support the behavior of λ_n is what one gets from the heuristics above on the polynomials (after square summation). On $S(\mu)$ the typical behavior of $\lambda_n(\mu, z)$ is like $\mu(\mathbf{D}_n(z))$, where $\mathbf{D}_n(z)$ is the disk about z with equilibrium measure $1/n$ (equilibrium measure of the support $S(\mu)$). In particular, $n\lambda_n(\mu, z)$ tends pointwise to the Radon-Nikodym derivative of μ with respect to the equilibrium measure. As a rule of thumb the estimate $|p_n(\mu, x)|^2 \leq C/n\lambda_n(\mu, x)$ holds in many cases.

If $f \in L^2(\mu)$, its Fourier expansion into $\{p_n(\mu, \cdot)\}$ is

$$f(x) \sim \sum_{k=0}^{\infty} c_k p_k(x), \qquad c_k = \int f\overline{p_k} d\mu.$$

The n-th partial sum has the closed form

$$\int f(t)K_n(x, t)d\mu(t), \qquad K_n(x, t) = \sum_{k=0}^{n} p_k(x)p_k(t).$$

In the real case for the reproducing kernel $K_n(x, t)$ we have the Christoffel-Darboux formula

$$K_n(x, t) = \frac{\kappa_n}{\kappa_{n+1}} \frac{p_{n+1}(x)p_n(t) - p_n(x)p_{n+1}(t)}{x - t},$$

which suggests a singular integral-type behavior for the partial sums. In general, Fourier expansions into orthogonal polynomials are sensitive to the weight (recall e.g. Pollard's theorem that Legendre expansions are bounded in $L^p[-1, 1]$ only for $4/3 < p < 4$), but sometimes convergence properties are equivalent to those of a related trigonometric Fourier series (so called transplantation theorems, see e.g. Askey [11]).

9 General orthogonal polynomials

In this section μ always has compact support $S(\mu)$. For all the results below see Stahl–Totik [87] and the references therein.

Lower and upper bounds

The **energy** $V(K)$ of a compact set K is defined as the infimum of

$$I(\nu) = \int\int \log \frac{1}{|x-t|} d\nu(x) d\nu(t) \tag{9.1}$$

where the infimum is taken for all positive Borel measures on K with total mass 1. The **logarithmic capacity** is then $\mathrm{cap}(K) = e^{-V(K)}$. For the leading coefficients κ_n of the orthonormal polynomials $p_n(\mu)$ we have

$$\frac{1}{\mathrm{cap}(S(\mu))} \le \liminf_{n\to\infty} \kappa_n^{1/n}. \tag{9.2}$$

To get an upper bound we need the concept of carrier: a Borel set E is a **carrier** for μ if $\mu(\mathbf{C} \setminus E) = 0$. The capacity of a Borel set is the supremum of the capacities of its compact subsets, and the **minimal carrier capacity** c_μ associated with μ is the infimum of the capacities of all carriers. With this

$$\limsup_{n\to\infty} \kappa_n^{1/n} \le \frac{1}{c_\mu}. \tag{9.3}$$

When $\mathrm{cap}(K)$ is positive, then there is a unique measure $\nu = \omega_K$ minimizing the energy in (9.1), and this measure is called the **equilibrium measure** of K. **Green's function** $g_{\mathbf{C}\setminus K}(z, \infty)$ with pole at infinity of $\mathbf{C} \setminus K$ can then be defined as

$$g_{\mathbf{C}\setminus K}(z,\infty) = \log \frac{1}{\mathrm{cap}(K)} - \int \log \frac{1}{|z-t|} d\omega_K(t). \tag{9.4}$$

We have for all μ (with $\mathrm{cap}(S(\mu)) > 0$) the estimate

$$\liminf_{n\to\infty} \frac{1}{n} \log |p_n(\mu,z)|^{1/n} \ge g_{\mathbf{C}\setminus S(\mu)}(z,\infty) \tag{9.5}$$

locally uniformly outside the convex hull of $S(\mu)$, while in the convex hull but outside the so called polynomial convex hull $\mathrm{Pc}(S(\mu))$ (for the definition see below) (9.5) is true **quasi-everywhere** (i.e., with the exception of a set of zero capacity). The same is true on the **outer boundary** of $S(\mu)$, which is defined as the boundary $\partial\Omega$ of the unbounded component Ω of the complement $\mathbf{C}\setminus S(\mu)$, namely for quasi-every $z \in \partial\Omega$

$$\liminf_{n\to\infty} |p_n(\mu,z)|^{1/n} \ge 1.$$

The **minimal carrier Green function** $g_\mu(z,\infty)$ is the supremum for all carriers E of the Green function of $\mathbf{C} \setminus E$, where the latter is defined as the infimum of $g_{\mathbf{C}\setminus K}$ for all compact subsets K of E. With this,

$$\limsup_{n\to\infty} \frac{1}{n} \log |p_n(\mu,z)|^{1/n} \le g_\mu(z,\infty) \tag{9.6}$$

locally uniformly on the whole plane.

All these estimates are sharp.

When the bounds in (9.2) and (9.3) coincide we have convergence for $\kappa_n^{1/n}$, and these bounds coincide precisely when the bounds in (9.5) and (9.6) do so.

Zeros

The zeros of $p_n(\mu)$ always lie in the convex hull of the support $S(\mu)$ of the measure μ. This is a consequence of the L^2 extremal property (3.1) of orthogonal polynomials. In fact, if there was a zero z_0 of $p_n(\mu, z)$ outside the convex hull of the support, then we could move that zero towards the convex hull (along a line segment that is perpendicular to a line separating z_0 and $S(\mu)$). During this move the absolute value of the polynomial decreases at all points of $S(\mu)$ and hence so does its $L^2(\mu)$ norm, but that is impossible by (3.1).

To say somewhat more on the location of zeros we need the concept of the polynomial convex hull. When Ω is the unbounded component of the complement $\mathbf{C} \setminus S(\mu)$, then $\mathrm{Pc}(S(\mu)) = \mathbf{C} \setminus \Omega$ is called the **polynomial convex hull** of $S(\mu)$ (it is the union of $S(\mu)$ with all the "holes" in it, i.e., with the bounded components of $\mathbf{C} \setminus S(\mu)$). Now the zeros cluster on $\mathrm{Pc}(S(\mu))$ in the sense that for any compact subset K of Ω there is a number N_K independent of n, such that $p_n(\mu)$ can have at most N_K zeros in K. The proof of this is based on the following lemma: *Let $V, S \subseteq \mathbf{C}$ be two compact sets. If V and $\mathrm{Pc}(S)$ are disjoint, then there exist $a < 1$ and $m \in \mathbf{N}$ such that for arbitrary m points $x_1, \ldots, x_m \in V$ there exist m points $y_1, \ldots, y_m \in \mathbf{C}$ for which the rational function*

$$r_m(z) := \prod_{j=1}^{m} \frac{z - y_j}{z - x_j} \tag{9.7}$$

has on S a sup-norm satisfying

$$\|r_m\|_S \le a. \tag{9.8}$$

Taking this for granted, assume that V is a compact set contained in Ω. We apply the lemma with $S = S(\mu)$, and let $a < 1$ and $m \in \mathbf{N}$ be the numbers in the lemma. Let us assume that $p_n(\mu; z)$ has at least m zeros $x_1, \ldots, x_m$ on V. By the lemma there exist m points $y_1, \ldots, y_m \in \mathbf{C}$ such that the rational function r_m defined as in (9.7) by the points $x_1, \ldots, x_m$ and $y_1, \ldots, y_m$ satisfies the inequality (9.8). With r_m we define the modified monic polynomial

$$q_n(z) := r_m(z) p_n(\mu; z) = z^n + \ldots,$$

For the $L^2(\mu)$ norm of this polynomial we have the estimate

$$\|q_n\|_{L^2(\mu)} \le \|r_m\|_{S(\mu)} \|p_n(\mu; \cdot)\|_{L^2(\mu)} < \|p_n(\mu; \cdot)\|_{L^2(\mu)},$$

which contradicts the minimality (3.1) of the monic orthogonal polynomial $p_n(\mu; z)$. Hence, we have proved that $p_n(\mu; z)$ has at most $m-1$ zeros on V, as was stated.

What we have said about the zeros can be sharpened for measures on the real line. For example, if μ is supported on the real line, then $\mathrm{Pc}(S(\mu)) = S(\mu)$, and if K is a closed interval disjoint from the support, then there is at most one zero in K. It was shown in Denison–Simon [23] that if $x_0 \in \mathbf{R}$ is not in the support, then for some $\delta > 0$ and all n either p_n or p_{n+1} has no zero in $(x_0 - \delta, x_0 + \delta)$. Note that if μ is a symmetric measure on $[-1, -1/2] \cup [1/2, 1]$, then $p_{2n+1}(0) = 0$ for all n, so the result is sharp.

Any isolated point in the support that lies on the outer boundary attracts precisely one zero. Let z_0 be an isolated point of $S(\mu)$, such that its distance from the convex hull of $S(\mu) \setminus \{z_0\}$ is $\delta > 0$. Then p_n has at most one zero in the disk $\{|z - z_0| < \delta/3\}$ (Simon [80, Section 8.1]). It is also clear that for any symmetric measure μ with $S(\mu) = [-1, -1/2] \cup \{0\} \cup [1/2, 1]$ the polynomials $p_{2n}(\mu)$ have 2 zeros near 0, so the result is sharp (in this case $\delta = 0$). Moreover, if μ lies on the unit circle, then there exist two positive constants C and a and a zero z_n of p_n such that $|z_n - z_0| \le Ce^{-an}$.

In general, each component of the polynomial convex hull consisting of more than one point attracts infinitely many zeros: if γ is a Jordan curve in Ω such that $S(\mu) \cap \gamma$ is infinite, then the number of zeros of p_n that lie inside γ tends to infinity (Saff–Totik [76]). Mass points of μ do not necessarily attract zeros (above we have mentioned that they do if they lie on the outer boundary). In fact, it was shown in Saff–Totik [76] that if ρ is the measure on the unit circle given by the density function $\sin^2(\theta/2)$, then for any measure σ that is supported in the open unit disk there is a $\lambda > 0$ such that all zeros of the n-th orthogonal polynomials with respect to $\mu = \rho + \lambda\sigma$ tend to the unit circle as $n \to \infty$.

Next put a unit mass at every zero of $p_n(\mu)$ (counting multiplicity). This gives the so called **counting measure** $\nu_{p_n(\mu)}$ on the zero set. **Zero distribution** amounts to finding the limit behavior of $\frac{1}{n}\nu_{p_n(\mu)}$. The normalized arc measure on the unit circle (for which $p_n(\mu, z) = z^n$) shows that if the interior of the polynomial convex hull $\mathrm{Pc}(S(\mu))$ is not empty, then the zeros may be far away from the outer boundary $\partial\Omega$, where the equilibrium measure $\omega_{S(\mu)}$ is supported. Thus, assume that $\mathrm{Pc}(S(\mu))$ has empty interior and also that there is no Borel set of capacity zero and full μ-measure, i.e., the minimal carrier capacity c_μ is positive (the $c_\mu = 0$ case is rather pathological, almost anything can happen with the zeros then). In this case

$$\lim_{n\to\infty} \kappa_n^{1/n} = \log \frac{1}{\mathrm{cap}(S(\mu))} \tag{9.9}$$

if and only if

$$\lim \frac{1}{n}\nu_{p_n(\mu)} = \omega_{S(\mu)}$$

in weak* sense, i.e., asymptotically minimal behavior of $\kappa_n^{1/n}$ (see (9.2)) is equivalent to the fact that the zero distribution is the equilibrium distribution. In a similar way, asymptotic maximal behavior (see (9.3)), i.e.,

$$\lim_{n\to\infty} \kappa_n^{1/n} = \frac{1}{c_\mu} \tag{9.10}$$

holds precisely when

$$\lim_{n\to\infty} \frac{1}{n}\nu_{p_n(\mu)} = \omega_\mu,$$

where ω_μ is the so called minimal carrier equilibrium measure, for which a representation like (9.4) is true, but for the minimal carrier Green function g_μ.

Regularity

(9.9) is called **regular limit behavior**, and in this case we write $\mu \in \mathbf{Reg}$. Thus, the important class **Reg** is defined by the property (9.9). $\mu \in \mathbf{Reg}$ is equivalent to either of

- $\lim_{n\to\infty} |p_n(\mu, z)|^{1/n} = \exp(g_{\mathbf{C}\setminus S(\mu)}(z, \infty)), \quad z \notin \mathrm{Con}(S(\mu))$
- $\limsup_{n\to\infty} |p_n(\mu, z)|^{1/n} = 1$ for quasi-every $z \in \partial\Omega$.

If Ω is a regular set with respect to the Dirichlet problem, then $\mu \in \mathbf{Reg}$ is equivalent to either of

- $\lim_{n\to\infty} \|p_n(\mu)\|_{\sup,S(\mu)}^{1/n} = 1$
- For any sequence $\{P_n\}$ of polynomials of degree $n = 1, 2, \ldots$

$$\lim_{n\to\infty} \left(\frac{\|P_n\|_{\sup,S(\mu)}}{\|P_n\|_{L^2(\mu)}} \right)^{1/n} = 1.$$

The last statement expresses the fact that in the n-th root sense the $L^2(\mu)$ and L^∞ norms (on $S(\mu)$) are asymptotically the same.

All equivalent formulations of $\mu \in \mathbf{Reg}$ point to a certain “thickness” of μ on its support. Regularity is an important property, and it is desirable to know “thickness” conditions under which it holds. Several regularity criteria are known, e.g. either of the conditions

- all Borel sets $B \subseteq S(\mu)$ with full measure (i.e with $\mu(B) = \mu(S(\mu))$) have capacity $\mathrm{cap}(B) = \mathrm{cap}(S(\mu))$, i.e., $c_\mu = \mathrm{cap}(S(\mu))$ or
- $d\mu/d\omega_{S(\mu)} > 0$ (Radon-Nikodym derivative) $\omega_{S(\mu)}$-almost everywhere

is sufficient for $\mu \in \mathbf{Reg}$. Regularity holds under fairly weak assumptions on the measure, e.g. if $S(\mu) = [0,1]$, and

$$\liminf_{r\to 0} r \log \mu([x-r, x+r]) \geq 0$$

for almost every $x \in [0,1]$ (i.e., if μ is not exponentially small around almost every point), then $\mu \in \mathbf{Reg}$.

No necessary and sufficient condition for regularity in terms of the size of the measure μ is known. The only existing necessary condition is for the case $S(\mu) = [0,1]$, and it reads that for every $\eta > 0$

$$\lim_{n\to\infty} \operatorname{cap}\left(\left\{x \,\middle|\, \mu([x-1/n, x+1/n]) \geq e^{-\eta n}\right\}\right) = \frac{1}{4}$$

(here $1/4$ is the capacity of $[0,1]$).

10 Strong, ratio and weak asymptotics

Strong asymptotics

Let μ be supported on $[-1,1]$ and suppose that the so called **Szegő condition**

$$\int_{-1}^{1} \frac{\log \mu'(t)}{\sqrt{1-t^2}} dt > -\infty \tag{10.1}$$

holds, where μ' is the Radon-Nikodym derivative of μ with respect to linear Lebesgue measure. Note that this condition means that the integral is finite, for it cannot be ∞. It expresses a certain denseness of μ, and under this condition G. Szegő proved several asymptotics for the corresponding orthonormal polynomials $p_n(\mu)$. This theory was developed on the unit circle and then was translated into the real line. The **Szegő function** associated with μ is

$$D_\mu(z) := \exp\left(\sqrt{z^2-1}\frac{1}{2\pi}\int_{-1}^{1} \frac{\log \mu'(t)}{z-t}\frac{dt}{\sqrt{1-t^2}}\right) \tag{10.2}$$

and it is the outer function in the Hardy space on $\mathbf{C} \setminus [-1,1]$ with boundary values $|D_\mu(x)|^2 = \mu'(x)$. Outside $[-1,1]$ the asymptotic formula

$$p_n(\mu, z) = (1+o(1))\frac{1}{\sqrt{2\pi}}(z+\sqrt{z^2-1})^n D_\mu(z)^{-1} \tag{10.3}$$

holds locally uniformly. In particular, the leading coefficient κ_n of $p_n(\mu)$ is of the form

$$\kappa_n = (1+o(1))\frac{2^n}{\sqrt{2\pi}} \exp\left(\frac{-1}{2\pi}\int_{-1}^{1} \frac{\log \mu'(t)}{\sqrt{1-t^2}} dt\right). \tag{10.4}$$

If $d\mu(x) = w(x)dx$ and $h(t) = w(\cos t)\sin t$ satisfies a Dini-Lipshitz condition

$$|h(t+\delta) - h(t)| \le \frac{C}{|\log\delta|^{1+\varepsilon}}, \qquad \varepsilon > 0,$$

then with

$$\Gamma_w(x) := \frac{1}{2\pi}\int_{-1}^{1} \frac{\log w(\xi) - \log w(x)}{\xi - x}\left(\frac{1-x^2}{1-\xi^2}\right)^{1/2} d\xi,$$

we have uniformly on $[-1,1]$

$$\begin{aligned}(1-x^2)^{1/4}w(x)^{1/2}p_n(x) &= \left(\frac{2}{\pi}\right)^{1/2}\cos\left((n+\frac{1}{2})\arccos x + \Gamma_w(x) - \frac{\pi}{4}\right)\\ &+ O((\log n)^{-\varepsilon}).\end{aligned}$$

For all these results see Szegő [91], Chapter 6. The Szegő condition is also necessary for these results, e.g. an asymptotic formula like (10.3) and (10.4) is equivalent to (10.1).

Ratio asymptotics

If one assumes weaker conditions then necessarily weaker results will follow. A large and important class of measures is the **Nevai class** $M(b,a)$ (see Nevai [62]), for which the coefficients in the three-term recurrence

$$xp_n(x) = a_n p_{n+1}(x) + b_n p_n(x) + a_{n-1}p_{n-1}(x)$$

satisfy $a_n \to a$, $b_n \to b$. This is equivalent to ratio asymptotics

$$\lim_{n\to\infty} \frac{p_{n+1}(z)}{p_n(z)} = \frac{z - b + \sqrt{(z-b)^2 - 4a^2}}{2}$$

for large z (actually, away from the support of μ), and the monograph Nevai [62] contains a very detailed treatment of orthogonal polynomials in this class. It is also true that if the limit of $p_{n+1}(z)/p_n(z)$ exists at a single non-real z, then $\mu \in M(b,a)$ for some a, b (Simon [79]).

The classes $M(b,a)$ are scaled versions of each other, and the most important condition ensuring $M(0,1/2)$ is given in Rakhmanov's theorem from [75]: if μ is supported in $[-1,1]$ and $\mu' > 0$ almost everywhere on $[-1,1]$, then $\mu \in M(0,1/2)$. Conversely, Blumenthal's theorem from [15] states that $\mu \in M(0,1/2)$ implies that the support of μ is $[-1,1]$ plus at most countably many points that converge to ± 1. Thus, in this respect the extension of Rakhmanov's theorem given in [22] by Denisov is of importance: if $\mu' > 0$ almost everywhere on $[-1,1]$ and outside $[-1,1]$ the measure μ has at most countably many mass points converging to ± 1, then $\mu \in M(0,1/2)$. However, $M(0,1/2)$

contains many other measures not just those that are in these theorems, e.g. in Delyon–Simon–Souillard [21] a continuous singular measure in the Nevai class was exhibited, and the result in Totik [92] shows that the Nevai class contains practically all types of measures allowed by Blumenthal's theorem.

Weak and relative asymptotics

Under Rahmanov's condition $\text{supp}(\mu) = [-1,1]$, $\mu' > 0$ a.e., some parts of Szegő's theory can be proven in a weaker form (see e.g. Máté–Nevai–Totik [57, 58]). In these the **Turán determinants**

$$T_n(x) := p_n^2(x) - p_{n-1}(x)p_{n+1}(x)$$

play a significant role. In fact, then given any interval $\mathbf{D} \subset (-1,1)$ the Turán determinant T_n is positive on $\mathbf{D}$ for all large n, and $T_n(x)^{-1}dx$ converges in the weak* sense to $d\mu$ on $\mathbf{D}$. Furthermore, the absolutely continuous part μ' can be also separately recovered from T_n:

$$\lim_{n\to\infty} \int \left| T_n(x)\mu'(x) - \frac{2}{\pi}(1-x^2)^{1/2} \right| dx = 0.$$

Under Rahmanov's condition we also have weak convergence, for example,

$$\lim_{n\to\infty} \int f(x)p_n^2(x)\mu'(x)dx = \frac{1}{\pi}\int_{-1}^{1} \frac{f(x)}{\sqrt{1-x^2}}dx \tag{10.5}$$

for any continuous function f. Pointwise we only know a highly oscillatory behavior: for almost all $x \in [-1,1]$

$$\limsup_{n\to\infty} p_n(x) \geq \frac{2}{\pi}(\mu'(x))^{-1/2}(1-x^2)^{-1/4},$$

$$\liminf_{n\to\infty} p_n(x) \leq -\frac{2}{\pi}(\mu'(x))^{-1/2}(1-x^2)^{-1/4},$$

and if $E_n(\varepsilon)$ is the set of points $x \in [-1,1]$ where

$$|p_n(x)| \geq (1+\varepsilon)\frac{2}{\pi}(\mu'(x))^{-1/2}(1-x^2)^{-1/4},$$

then $|E_n(\varepsilon)| \to 0$ for all $\varepsilon > 0$. However, it is not true that the sequence $\{p_n(\mu,x)\}$ is pointwise bounded, since for every $\varepsilon > 0$ there is a weight function $w > 1$ on $[-1,1]$ such that $p_n(0)/n^{1/2-\varepsilon}$ is unbounded (see Rakhmanov [74]).

Simon [79] extended (10.5) by showing that if the recurrence coefficients satisfy $b_n \to b$, $a_{2n+1} \to a'$ and $a_{2n} \to a''$, then there is an explicitly calculated measure ρ depending only on b, a', a'' such that

$$\lim_{n\to\infty} \int f(x)p_n^2(x)\mu'(x)dx = \int_{-1}^{1} f(x)d\rho(x) \tag{10.6}$$

for any continuous function f, and conversely, if (10.6) exists for $f(x) = x, x^2, x^4$, then $b_n \to b$, $a_{2n+1} \to a'$ and $a_{2n} \to a''$ with some b, a', a''.

For measures in Nevai's class, part of Szegő's theory can be extended to **relative asymptotics**, i.e., when sequences of orthogonal polynomials corresponding to two measures are compared. Here is a sample theorem: let α be supported in $[-1, 1]$ and in Nevai's class $M(0, 1/2)$, and let $d\beta = gd\alpha$, where g is a function such that for some polynomial R both Rg and R/g are Riemann integrable. Then

$$\lim_{n\to\infty} \frac{p_n(\beta, z)}{p_n(\alpha, z)} = D_g(z)^{-1}$$

uniformly on $\mathbf{C}$ away from $[-1, 1]$, where D_g is Szegő's function with respect to the measure $g(x)dx$.

Widom's theory

Szegő's theory can be extended to measures lying on a single Jordan curve or arc J (see Kaliaguine [40] where also additional outside lying mass points are allowed), in which case the role of $z + \sqrt{z^2 - 1}$ in (10.3) is played by the conformal map Φ of $\mathbf{C} \setminus J$ onto the exterior of the unit disk, and the role of 2^n in (10.4) is played by the reciprocal of the logarithmic capacity of J (see Section 9). Things change considerably if the measure is supported on a set J consisting of two or more smooth curve or arc components $J_1, \ldots, J_m$. A general feature of this case is that $\kappa_n \text{cap}(J)^n$ does not have a limit, its limit points fill a whole interval (though if some associated harmonic measures are all rational then the limit points may form a finite set). The polynomials themselves have asymptotic form

$$\frac{p_n(z)}{\kappa_n} = \text{cap}(J)^n \Phi(z)^n (F_n(z) + o(1))$$

uniformly away from J, where Φ is the (multi-valued) complex Green function of the complement $\mathbf{C} \setminus J$, and where F_n is the solution of an L^2-extremal problem involving analytic functions belonging to some class Γ_n. The functions F in Γ_n are determined by an H^2 condition plus an argument condition, namely if the change of the argument of Φ as we go around J_k is $\gamma_k 2\pi$ modulo 2π, then in Γ_n we consider functions whose change of the argument around J_k is $-n\gamma_k 2\pi$ modulo 2π. Now the point is that these function classes Γ_n change with n, and hence so does F_n, and this is the reason that a single asymptotic formula like (10.4) or (10.3) does not hold. The fundamentals of the theory were laid out in H. Widom's paper [97]; and since then many results have been obtained by F. Peherstorfer and his collaborators, as well as A. I. Aptekarev, J. Geronimo, S. P. Suetin and W. Van Assche. The theory has deep connections with function theory, the theory of Abelian integrals and the theory of elliptic functions. We refer the reader to the papers Aptekarev [7], Geronimo–Van Assche [31], Peherstorfer [69]–[72] and Suetin [88]–[89].

Asymptotics for Christoffel functions

The Christoffel functions

$$\lambda_n(\mu,x)^{-1}=\sum_{k=0}^{n}p_k(\mu,x)^2$$

behave somewhat more regularly than the orthogonal polynomials. In Máté–Nevai–Totik [59] it was shown that if μ is supported on $[-1,1]$, it belongs to the **Reg** class there (see Section 9) and $\log\mu'$ is integrable over an interval $I\subset[-1,1]$, then for almost all $x\in I$

$$\lim_{n\to\infty}n\lambda_n(\mu,x)=\pi\sqrt{1-x^2}\mu'(x).$$

This result is true (see Totik [93]) in the form

$$\lim_{n\to\infty}n\lambda_n(\mu,x)=\frac{d\mu(x)}{d\omega_{\mathrm{supp}(\mu)}(x)},\qquad \text{a.e. } x\in I$$

when the support is a general compact subset of $\mathbf{R}$, $\mu\in\mathbf{Reg}$ and $\log\mu'\in L^1(I)$.

Often only a rough estimate is needed for Christoffel functions, and such a one is provided in Mastroianni–Totik [56]: if μ is supported on $[-1,1]$ and it is a doubling measure, i.e.,

$$\mu(2I)\le L\mu(I)$$

for all $I\subset[-1,1]$, where $2I$ is the twice enlarged I, then uniformly on $[-1,1]$

$$\lambda_n(\mu,x)\sim\mu\left(\Delta_n(x)\right);\qquad \Delta_n(x)=\frac{\sqrt{1-x^2}}{n}+\frac{1}{n^2}.$$

11 Recurrence coefficients and spectral measures

Let μ be a unit measure of compact support on the real line, and

$$xp_n(x)=a_np_{n+1}(x)+b_np_n(x)+a_{n-1}p_{n-1}(x),$$

the recurrence relation for the corresponding orthogonal polynomials. We have already mentioned in Section 6 that μ is the spectral measure for the Jacobi matrix

$$J=\begin{pmatrix} b_0 & a_0 & 0 & 0 & \cdots\\ a_0 & b_1 & a_1 & 0 & \cdots\\ 0 & a_1 & b_2 & a_2 & \cdots\\ 0 & 0 & a_2 & b_2 & \cdots\\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

and this gives a one-to-one correspondence between unit measures with compact (and infinite) support on the real line and Jacobi operators with bounded entries. Every such Jacobi operator is a bounded self-adjoint operator on l_2, hence operator theory and orthogonal polynomials meet at this point, and techniques and questions from both areas are relevant. Information between the measure μ and the recurrence coefficients might be called the spectral analysis of orthogonal polynomials. In this, a special role is played by the Chebyshev case $a_n = 1/2$ and $b_n = 0$, where the Jacobi matrix is denoted by J_0. If the sequences $\{a_n\}$ and $\{b_n\}$ have limits then we may assume $a_n \to 1/2$, $b_n \to 0$ (this is just rescaling, and this was the Nevai class $M(0, 1/2)$), hence in this case the Jacobi operator is a compact perturbation of J_0, and one of the main questions of the theory is how properties of $J - J_0$ are reflected in the spectral measure μ.

We have already mentioned in Section 10 that $\mu \in M(0, 1/2)$ implies that the support of μ is $[-1, 1]$ plus some additional mass points converging to ± 1 (called Blumenthal's theorem in orthogonal polynomials; it is a special case of Weyl's theorem in operator theory on the invariance of the essential spectrum under compact perturbation). Conversely, if the support of μ is $[-1, 1]$ plus some additional mass points converging to ± 1 and $\mu'(x) > 0$ for almost all $x \in [-1, 1]$, then $\mu \in M(a, b)$ (Denisov [22], Rakhmanov [75], Nevai–Totik [63]). No spectral characterization of $\mu \in M(0, 1/2)$ is known; this important class seems to contain all sorts of measures. For example, if ν is any measure with support $[-1, 1]$ then there is a $\mu \in M(0, 1/2)$ which is absolutely continuous with respect to ν. In particular, $M(0, 1/2)$ contains discrete measures, continuously singular measures or measures that are given by a continuous density which is positive on a set of measure $< \varepsilon$.

Strengthening the condition $\mu \in M(0, 1/2)$ can be done in several ways. After numerous works in the subject by Szegő, Shohat, Geronimus, Krein, Kolmogorov and others, a complete characterization for $J - J_0$ being a Hilbert-Schmidt operator was given in [43] by R. Killip and B. Simon (note that μ is assumed to have total mass 1):

$$\sum_n (a_n - 1/2)^2 + \sum_n b_n^2 < \infty \tag{11.1}$$

if and only if the following conditions hold:

(i) the support of μ is $[-1, 1]$ plus some additional mass points $E_j^{\pm}$ converging to ± 1,

(ii) if μ' is the absolutely continuous part of μ on $[-1, 1]$, then

$$\int_{-1}^{1} (\log \mu'(t)) \sqrt{1 - t^2}\, dt > -\infty,$$

(iii) for the mass points $E_j^{\pm}$ lying outside $[-1, 1]$ we have

$$\sum_j |E_j^+ - 1|^{3/2} + \sum_j |E_j^- + 1|^{3/2} < \infty.$$

It was also shown in Killip–Simon [43] that if $J - J_0$ is trace class, i.e.,

$$\sum_n |a_n - 1/2| + \sum_n b_n < \infty,$$

then Szegő's condition

$$\int_{-1}^{1} \frac{\log \mu'(t)}{\sqrt{1-t^2}} dt > -\infty \tag{11.2}$$

holds. The conclusion is also true if $\mu \in M(0, 1/2)$, for the mass points $E_j^{\pm}$ lying outside $[-1, 1]$ we have

$$\sum_j |E_j^+ - 1|^{1/2} + \sum_j |E_j^- + 1|^{1/2} < \infty,$$

and

$$\limsup_n (2^n a_1 \cdots a_n) > 0. \tag{11.3}$$

If the support of μ is contained in $[-1, 1]$, then Szegő's condition automatically holds if (11.3) is true. Actually, when $\mathrm{supp}(\mu) = [-1, 1]$ then Szegő's condition (11.2) is equivalent to (11.1) and to the (conditional) convergence of the series $\sum_n (a_n - 1/2)$ and $\sum_n b_n$.

There is also an extended theory of orthogonal polynomials with several different applications when the recurrence coefficients do not converge, but they are asymptotically periodic in the sense that for some k all the sequences $(a_{kn+j})_{n=1}^{\infty}$ and $(b_{kn+j})_{n=1}^{\infty}$, $j = 1, \ldots, k$ converge. These are related to so called sieved orthogonal polynomials and to orthogonal polynomials generated by polynomial mappings. In this case the essential support of the spectral measure lies on several intervals. There are numerous papers on this subject by M. E. H. Ismail, N. A. Al-Salam, J. A. Charris, J. Wimp, J. Bustoz, J. Geronimo, W. Van Assche, F. Peherstorfer, R. Steinbauer, N. I. Akhiezer, B. P. Osilenker and others; see e.g. Charris–Ismail [17], Geronimo–Van Assche [31], Peherstorfer [70], Peherstorfer–Steinbauer [73], Akhiezer [1] for details and for further references.

12 Exponential and Freud weights

These are weight functions of the form $e^{-2Q(x)}$, where x is on the real line or on some subinterval thereof. For simplicity we shall first assume that Q is even. We get **Freud weights** when $Q(x) = |x|^{\alpha}$, $\alpha > 0$, $x \in \mathbf{R}$, and **Erdős weights** if Q tends to infinity faster than any polynomial as $|x| \to \infty$. G. Freud started to investigate these weights in the sixties and seventies, but they independently appeared also in the Russian literature and in statistical physics. One can safely say that some of Freud's problems and the work of P. Nevai and E. A. Rahmanov were the primary cause of the sudden revitalization of the theory of orthogonal polynomials since the early 1980's. In the last 20 years D. Lubinsky with coauthors have conducted systematic studies on exponential weights, see e.g. Levin–Lubinsky [46, 45], Lubinsky [49], Lubinsky–Saff [50], Van Assche [94]; we should mention the names E. Levin, E. B. Saff, W. Van Assche, E. A. Rahmanov and H. N. Mhaskar. In the mid 1990's a new stimulus came from the Riemann–Hilbert approach that was used together with the steepest descent method by P. Deift, T. Kriecherbauer, K. T.-R. McLaughlin, S. Venakides and X. Zhou ([20]) to give complete asymptotics when Q is analytic.

One can roughly say that because of the fast vanishing of the weight around infinity, things happen on a finite subinterval $[-a_n, a_n]$ (depending on the degree of the polynomials), and on $[-a_n, a_n]$ techniques developed for $[-1, 1]$ are applied. For Freud weights one can also make the substitution $x \to n^{1/\lambda} x$ and go to orthogonality with respect to the varying weight $e^{-n|x|^\lambda}$, in which case things are automatically reduced to a finite interval which is the support of a weighted energy problem.

The a_n are the so called **Mhaskar-Rahmanov-Saff numbers** defined by

$$n = \frac{2}{\pi} \int_0^1 \frac{a_n t Q'(a_n t)}{\sqrt{1-t^2}}\, dt. \tag{12.1}$$

The zeros of $p_n(w^2)$, $w(x) = \exp(-Q(x))$ are spreading out and the largest zero is very close to a_n, which tends to ∞.

To describe the distribution of the zeros and the behavior of the polynomials one has to make appropriate contractions. Let us consider first the case of Freud weight $w(x) = \exp(-|x|^\alpha)$, and let p_n be the n-th orthogonal polynomial with respect to w^2 (on $(-\infty, \infty)$). In this case

$$a_n = n^{1/\alpha}\gamma_\alpha, \qquad \gamma_\alpha^\alpha := \Gamma\left(\frac{\alpha}{2}\right)\Gamma\left(\frac{1}{2}\right) \Big/ 2\Gamma\left(\frac{\alpha}{2} + \frac{1}{2}\right).$$

Thus, for the largest zero $x_{n,n}$ we have $x_{n,n}/n^{1/\alpha} \to \gamma_\alpha$ as $n \to \infty$, and to describe zero distribution we divide (contract) all zeros $x_{n,i}$ by $n^{1/\alpha}\gamma_\alpha$. These contracted zeros asymptotically have the **Ullman distribution**

$$\frac{d\mu_w(t)}{dt} := \frac{\alpha}{\pi} \int_{|t|}^1 \frac{u^{\alpha-1}}{\sqrt{u^2 - t^2}}\, du, \quad t \in [-1, 1]. \tag{12.2}$$

This measure μ_w minimizes the weighted energy

$$\int\int \log \frac{1}{|x-t|} d\mu(x) d\mu(t) + 2 \int Q d\mu \tag{12.3}$$

among all probability measures compactly supported on $\mathbf{R}$. It is a general feature of exponential weights that the behavior of zeros of the polynomials is governed by the solution of a weighted energy problem (weighted equilibrium measures, see Saff–Totik [77]). If κ_n is the leading coefficient of p_n, i.e., $p_n(z) = \kappa_n z^n + \cdots$, then (Lubinsky–Saff [50])

$$\lim_{n\to\infty} \kappa_n \pi^{1/2} 2^{-n} e^{-n/\alpha} n^{(n+1/2)/\alpha} = 1,$$

and we have

$$\lim_{n\to\infty} |p_n(n^{1/\alpha}\gamma_\alpha z)|^{1/n}$$
$$= \exp\left(\log|z+\sqrt{z^2-1}| + \mathrm{Re}\int_0^1 \frac{zu^{\alpha-1}}{\sqrt{z^2-u^2}}\,du\right)$$

locally uniformly outside $[-1,1]$. This latter is so called n-th root asymptotics, while the former is strong asymptotics. Strong asymptotics for $p_n(z)$ on different parts of the complex plane was given using the Riemann–Hilbert approach, see Deift [20] and Kriecherbauer–McLaughlin [44] and the references there. On the real line we have a Plancherel–Rotach type formula

$$n^{1/2\alpha}p_n(w_\alpha; n^{1/\alpha}\gamma_\alpha x)\exp(-n\gamma_\alpha^\alpha|x|^\alpha)-$$
$$-\sqrt{\frac{2}{\pi}}\frac{1}{\sqrt[4]{1-x^2}}\cos\left(\frac{1}{2}\arccos x + n\pi\mu_w([x,1]) - \frac{\pi}{4}\right) \to 0$$

uniformly on any subinterval of $(-1,1)$.

Things become more complicated for non-Freud weights, but the corresponding results are of the same flavor. In this case the weight is not necessarily symmetric, but under some conditions (like Q being convex or $xQ'(x)$ being increasing for $x>0$ and an analogous condition for $x<0$) the relevant weighted equilibrium measure's support is an interval, and the definition of the Mhaskar–Rahmanov–Saff numbers $a_{\pm n}$ is

$$\begin{aligned} n &= \frac{1}{\pi}\int_{a_{-n}}^{a_n} \frac{xQ'(x)}{\sqrt{(x-a_{-n})(a_n-x)}}dx, \\ 0 &= \frac{1}{\pi}\int_{a_{-n}}^{a_n} \frac{Q'(x)}{\sqrt{(x-a_{-n})(a_n-x)}}dx. \end{aligned}$$

Now one solves the weighted equilibrium problem (12.3) for all measures μ *with total mass* n, and if μ_n is the solution then $[a_{-n}, a_n]$ is the support of μ_n and μ_n/n will play the role of the measure μ_w from (12.2) above.

The weight does not even have to be defined on all $\mathbf{R}$, e.g. in [46] a theory was developed by Levin and Lubinsky that simultaneously includes far reaching generalizations of non-symmetric Freud, Erdős and Pollaczek weights such as

(a) nonsymmetric Freud-type weights

$$Q(x) = \begin{cases} |x|^\alpha, & x\in[0,\infty) \\ |x|^\beta, & x\in(-\infty,0), \end{cases}$$

(b) nonsymmetric Erdős weights such as

$$Q(x) = \begin{cases} \exp_l(|x|^\alpha) - \exp_l(0), & x\in[0,\infty) \\ \exp_k(|x|^\beta) - \exp_k(0), & x\in(-\infty,0) \end{cases}$$

with $\exp_l$ is the l-times iterated exponential function, or

(c) nonsymmetric Pollaczek type weights that vanish fast at ± 1 such as

$$Q(x) = \begin{cases} \exp_l((1-x)^{-\alpha}) - \exp_l(1), & x \in [0,1) \\ \exp_k((1-x)^{-\beta}) - \exp_k(0), & x \in (-1,0). \end{cases}$$

In all cases the interval $[a_{-n}, a_n]$ is where things happen, e.g. this is the shortest interval on which the supremum norm of a weighted polynomial is attained:

$$\|wP_n\|_{\sup} = \|wP_n\|_{\sup,[-a_{-n},a_n]}$$

for all polynomials of degree at most n. These numbers $a_{\pm n}$ are everywhere in the theory, e.g.

$$\sup_x |p_n(x)|w(x)|x-a_n|^{1/4}|x-a_{-n}|^{1/4} \sim 1.$$

13 Sobolev orthogonality

In **Sobolev orthogonality** we consider orthogonality with respect to an inner product

$$(f,g) = \sum_{k=0}^{r} \int f^{(k)} \overline{g^{(k)}} d\mu_k, \tag{13.1}$$

where μ_k are given positive measures. There are several motivations for this kind of orthogonality. Perhaps the most natural one is smooth data fitting. The Spanish school around F. Marcellán, G. Lopez and A. Martinez-Finkelshtein has been particularly active in developing this area (see the surveys Marcellán–Alfaro–Rezola [51] and Martinez-Finkelshtein [53, 52] and the references therein).

In this section let $Q_n(z) = z^n + \cdots$ denote the *monic* orthogonal polynomial with respect to the Sobolev inner product (13.1), and $q_n(\mu_k)$ the monic orthogonal polynomials with respect to the measure μ_k.

Most arguments for the standard theory fail in this case, e.g. it is no longer true that the zeros lie in the convex hull of the support of the measures μ_k, $k = 0, 1, \ldots, r$. It is not even known if the zeros are bounded if all the measures μ_k have compact support. Nonetheless, for the case $r = 1$, and $\mu_0, \mu_1 \in \mathbf{Reg}$ (see Section 9) it was shown in Gautschi–Kuijlaars [29] that the asymptotic distribution of the zeros of the *derivative* Q_n' is the equilibrium measure $\omega_{E_0 \cup E_1}$, where E_i is the support of μ_i, $i = 0, 1$ (which also have to be assumed to be regular). Furthermore, if, in addition, $E_0 \subseteq E_1$, then the asymptotic zero distribution of Q_n is ω_{E_0}.

In general, both the algebraic and the asymptotic/analytic situation is quite complicated, and there are essentially two important cases which have been understood to a satisfactory degree.

Case I: The discrete case. In this case μ_0 is some “strong” measure, e.g. from the Nevai class $M(b, a)$ (see Section 10), and $\mu_1, \ldots, \mu_k$ are finite discrete measures.

It turns out that then the situation is similar to adding these discrete measures to μ_0 (the new measure will also be in the same Nevai class), and considering standard orthogonality with respect to this new measure. For ecample, if $r = 1$, then

$$\lim_{n\to\infty} \frac{Q_n(z)}{q_n(\mu_0 + \mu_1, z)} = 1$$

holds uniformly on compact subsets of $\mathbf{C} \setminus \operatorname{supp}(\mu_0 + \mu_1)$. Thus, the Sobolev orthogonal polynomials differ from those of the measure μ_0, but not more than what happens when adding mass points to μ_0.

In this discrete case the Q_n's satisfy a higher order recurrence relation, hence this case is also related to matrix orthogonality (see the end of the Section 16).

Case II: The Szegő case. Suppose now that $\mu_0, \dots, \mu_k$ are all supported on the same smooth curve or arc J, and they satisfy Szegő's condition there (see Section 10). In this case the k-th derivative of Q_n satisfies, locally uniformly in the complement of J, the asymptotic formula

$$\lim_{n\to\infty} \frac{Q_n^{(k)}(z)}{n^k q_{n-k}(\mu_r, z)} = \frac{1}{[\Phi'(z)]^{m-k}},$$

where Φ is the conformal map that maps $\mathbf{C} \setminus J$ onto the complement of the unit disk. That is, in this case the measures $\mu_0, \dots, \mu_{r-1}$ do not appear in the asymptotic formula, only μ_r matters. The reason for this is the following: $Q = Q_n$ minimizes

$$(Q, Q) = \sum_{k=0}^{r} \int |Q^{(k)}|^2 d\mu_k \tag{13.2}$$

among all monic polynomials of degree n, while $q = q_{n-k}(\mu_k)$ minimizes

$$\int |q|^2 d\mu_k$$

among all monic polynomials of degree $n - k$. But the polynomial $Q_n^{(k)}(t) = n(n-1)\cdots(n-k+1)t^{n-k} + \cdots$ is a monic polynomial times the factor $n(n-1)\cdots(n-k+1) \sim n^k$, and this factor is dominant for $k = r$, so everything else will be negligible. There are results for compensation of this n^k factor which lead to Sobolev orthogonality with respect to varying measures.

Under the much less restrictive assumption that $\mu_0 \in \mathbf{Reg}$ (see Section 9) and the other measures μ_k are supported in the support E of μ_0 it is true (López–Pijeira-Cabrera–Izquierdo [47]) that the asymptotic zero distribution of $Q_n^{(k)}$ is the equilibrium measure ω_E for all k,

$$\lim_{n\to\infty} \|Q_n^{(k)}\|_{\sup,E}^{1/n} = \operatorname{cap}(E),$$

and hence, away from the zeros in the unbounded component of the complement of E, we have

$$\lim_{n\to\infty} |Q_n^k(z)|^{1/n} = e^{g_{\mathbf{C}\setminus E}(z)}$$

where $g_{\mathbf{C}\setminus E}$ is the Green function for this unbounded component.

The techniques developed for exponential weights and for Sobolev orthogonality were combined in Geronimo–Lubinsky–Marcellan [30] to prove strong asymptotics for Sobolev orthogonal polynomials when $r = 1$ and $\mu_0 = \mu_1$ are exponential weights.

14 Non-Hermitian orthogonality

We refer to **non-Hermitian orthogonality** in either of these cases:

- the measure μ is non-positive or even complex-valued and we consider p_n with
$$\int p_n(z)\overline{z^k}d\mu = 0, \qquad k = 0, 1, \ldots, n-1, \tag{14.1}$$
- μ is again non-positive or complex-valued, or positive but lies on a complex curve or arc and orthogonality is considered without complex conjugation, i.e.,
$$\int p_n(z)z^k d\mu = 0, \qquad k = 0, 1, \ldots, n-1. \tag{14.2}$$

More generally, one could consider non-positive inner products, but we shall restrict our attention to complex measures and orthogonality (14.2).

As an example, consider the diagonal Padé approximant to the Cauchy transform

$$f(z) = \int \frac{d\mu(t)}{z-t}$$

of a signed or complex-valued measure, i.e., consider polynomials p_n and q_n of degree at most n such that

$$f(z)p_n(z) - q_n(z) = O(z^{-n-1})$$

at infinity. Then p_n satisfies the non-Hermitian orthogonality relation

$$\int p_n(x)x^j d\mu(x) = 0, \qquad j = 0, 1, \ldots, n-1. \tag{14.3}$$

In this non-Hermitian case even the Gram-Schmidt orthogonalization process may fail, and then p_n is defined as the solution of the orthogonality condition (14.1), resp. (14.2), which give a system of homogeneous equations for the coefficients of p_n. Thus, p_n may have smaller degree than n, and things can get pretty wild with this kind of orthogonality. For example, in the simple case

$$d\mu(x) = (x - \cos\pi\alpha_1)(x - \cos\pi\alpha_2)(1-x^2)^{-1/2}dx, \quad x \in [-1, 1],$$

with $0 < \alpha_1 < \alpha_2 < 1$ rationally independent algebraic numbers, the zeros of p_n from (14.3) are dense on the whole complex plane (compare this with the fact that for positive μ all zeros lie in $[-1,1]$). In Stahl [86] it was shown that it is possible to construct a complex measure μ on $[-1,1]$, such that for an arbitrary prescribed asymptotic behavior some subsequence $\{p_{n_k}\}$ will have this zero behavior. Nonetheless, the asymptotic distribution of the zeros is again the equilibrium distribution of the support of μ under regularity conditions on μ. For example, this is the case if

- $|\mu|$ belongs to the **Reg** class (see Section 9), and the argument of μ, i.e., $d\mu(t)/d|\mu|(t)$, is of bounded variation (Baratchart–Küstner–Totik [12]), or
- $d\mu(x) = g(\arccos x)(1-x^2)^{-1/2}dx$, $x \in [-1,1]$, g is bounded away from zero and infinity, and satisfies $|g(\theta+\delta) - g(\theta)| \le K|\log\delta|^{-1-\delta}$, or
- μ is supported on finitely many intervals, the argument of μ is uniformly continuous and for $a(\delta) = \inf_{x\in\mathrm{supp}(\mu)} |\mu|([x-\delta, x+\delta])$ the property $\lim_{\delta\to 0}\log a(\delta) = 0$ holds (Stahl [86]).

In [84]–[86] H. Stahl obtained asymptotics for non-Hermitian orthogonal polynomials even for varying measures and gave several applications of them to Padé approximation. When the measure μ is of the form $d\mu(x) = g(x)(1-x^2)^{-1/2}dx$, $x \in [-1,1]$, with an analytic g, for $z \in \mathbf{C}\setminus[-1,1]$, a strong asymptotic formula of the form

$$\frac{p_n(z)}{\kappa_n} = (1+o(1))\frac{(z+\sqrt{z^2-1})^n}{2^n} D_\mu(z)^{-1}\exp\left(\frac{1}{2\pi}\int_{-1}^{1}\frac{\log\mu'(t)}{\sqrt{1-t^2}}dt\right)$$

(with D_μ the Szegő function (10.2)) was proved by J. Nuttall [66], [67], A. A. Gonchar and S. P. Suetin [38]. For a recent Riemann–Hilbert approach see the paper [10] by A. I. Aptekarev and W. Van Assche. A similar result holds on the support of the measure, as well as for the case of varying weights, see Aptekarev–Van Assche [10].

15 Multiple orthogonality

Multiple orthogonality comes from simultaneous Padé approximation. It is a relatively new area where we have to mention the names of E. M. Nikishin, V. N. Sorokin, A. A. Gonchar and E. A. Rahmanov, A. I. Aptekarev, A. B. J. Kuijlaars, J. Geronimo and W. Van Assche (see the survey [95] by W. Van Assche and the references there and the paper Gonchar–Rakhmanov [37]). The analogues of many classical concepts and properties have been found, and also the analogues of the classical orthogonal polynomials are known, e.g. in the multiple Hermite case the measures are $d\mu(x) = e^{-x^2+c_jx}dx$.

Asymptotic behavior of multiple orthogonal polynomials is not fully understood yet due to the interaction of the different measures. For the existing results see Aptekarev [8], Van Assche [95], Van Assche's Chapter 23 in [39] and the references therein.

Types and normality

On $\mathbf{R}$ let there be given r measures $\mu_1, \ldots, \mu_r$ with finite moments and infinite support, and consider multiindices $\underline{n} = (n_1, \ldots, n_r)$ of nonnegative integers with norm $|\underline{n}| = n_1 + \cdots + n_r$. There are two types of multiple orthogonality corresponding to the appropriate Hermite-Padé approximation.

In **type I multiple orthogonality** we are looking for polynomials $Q_{\underline{n},j}$ of degree $n_j - 1$ for each $j = 1, \ldots, r$, such that

$$\sum_{j=1}^{r} \int x^k Q_{\underline{n},j}(x) d\mu_j(x) = 0, \qquad k = 0, 1, \ldots, |\underline{n}| - 2.$$

These orthogonality relations give $|\underline{n}| - 1$ homogeneous linear equations for the $|\underline{n}|$ coefficients of the r polynomials $Q_{\underline{n},j}$, so there is a non-trivial solution. If the rank of the system is $|\underline{n}| - 1$, then the solution is unique up to a multiplicative factor, in which case the index $\underline{n}$ is called **normal**. This happens precisely if each $Q_{\underline{n},j}$ is of exact degree $n_j - 1$.

In **type II multiple orthogonality** we are looking for a single polynomial $P_{\underline{n}}$ of degree $|\underline{n}|$ such that

$$\begin{aligned} \int x^k P_{\underline{n}}(x) d\mu_1(x) &= 0, \qquad k = 1, \ldots, n_1 - 1 \\ &\vdots \\ \int x^k P_{\underline{n}}(x) d\mu_r(x) &= 0, \qquad k = 1, \ldots, n_r - 1. \end{aligned}$$

These are $|\underline{n}|$ homogeneous linear equations for the $|\underline{n}| + 1$ coefficients of $P_{\underline{n}}$, and again if the solution is unique up to a multiplicative constant, then $\underline{n}$ is called normal. This is again equivalent to $P_{\underline{n}}$ being of exact degree $\underline{n}$.

$\underline{n}$ is normal for type I orthogonality precisely when it is normal for type II, so we just speak of normality. This is the case, for example, if the μ_j's are supported on intervals $[a_j, b_j]$ that are disjoint except perhaps for their endpoints; in fact, in this case $P_{\underline{n}}$ has n_j simple zeros on (a_j, b_j). Normality also holds if $d\mu_j = w_j d\mu$ with a common μ supported on some interval $[a, b]$, and for all $m_j \leq n_j$, $j = 1, \ldots, r$, every non-trivial linear combination of the functions

$$w_1(x), xw_1(x), \ldots, x^{m_1-1}w_1(x), w_2(x), xw_2(x), \ldots, x^{m_r-1}w_r(x)$$

has at most $m_1 + \cdots + m_r - 1$ zeros on $[a, b]$ (this means that these functions form a so called Chebyshev system there). In this case $P_{\underline{n}}$ has $|\underline{n}| - 1$ zeros on $[a, b]$.

Recurrence formulae

To describe recurrence formulae, let $\underline{e}_j = (0, \dots, 1, \dots, 0)$ where the single 1 entry is at position j. Under the normality assumption if $P_{\underline{n}}$ is the monic orthogonal polynomial, then for any k

$$xP_{\underline{n}}(x) = P_{\underline{n}+\underline{e}_k}(x) + a_{\underline{n},0}P_{\underline{n}}(x) + \sum_{j=1}^{r} a_{\underline{n},j}P_{\underline{n}-\underline{e}_j}(x).$$

Another recurrence formula is

$$xP_{\underline{n}}(x) = P_{\underline{n}+\underline{e}_k}(x) + b_{\underline{n},0}P_{\underline{n}}(x) + \sum_{j=1}^{r} b_{\underline{n},j}P_{\underline{n}-\underline{e}_{\pi(1)}-\cdots-\underline{e}_{\pi(j)}}(x),$$

where $\pi(1), \dots, \pi(r)$ is an arbitrary, but fixed, permutation of $1, 2, \dots, r$. The orthogonal polynomials with different indices are strongly related to one another, e.g. $P_{\underline{n}+\underline{e}_k}(x) - P_{\underline{n}+\underline{e}_l}(x)$ is a constant multiple of $P_{\underline{n}}(x)$.

If $d\mu_j = w_j d\mu$, then similar recurrence relations hold in case of type I orthogonality for

$$Q_{\underline{n}}(x) = \sum_{j=1}^{r} Q_{\underline{n},j}(x)w_j(x).$$

Also, type I and type II are related by a **biorthogonality** property:

$$\int P_{\underline{n}}Q_{\underline{m}}d\mu = 0$$

except for the case when $\underline{m} = \underline{n} + \underline{e}_k$ for some k, and then the previous integral is not zero (under the normality condition).

To describe an analogue of the Christoffel-Darboux formula let $\{\underline{m}_j\}$ be a sequence of multiindices such that $\underline{m}_0$ is the identically 0 multiindex, and $\underline{m}_{j+1}$ coincides with $\underline{m}_j$ except for one component which is 1 larger than the corresponding component of $\underline{m}_j$. Set $P_j = P_{\underline{m}_j}$, $Q_j = Q_{\underline{m}_{j+1}}$ and with $\underline{m} = \underline{m}_n$

$$h_{\underline{m}}^{(j)} := \int P_{\underline{m}}(x)x^{(\underline{m})_j}d\mu_j(x),$$

where $(\underline{m})_j$ denotes the j-th component of the multiindex $\underline{m}$. Then (see Daems–Kuijlaars [19]), again with $\underline{m} = \underline{m}_n$,

$$(x-y)\sum_{k=0}^{n-1} P_k(x)Q_k(y) = P_{\underline{m}}(x)Q_{\underline{m}}(y) - \sum_{j=1}^{r} \frac{h_{\underline{m}}^{(j)}}{h_{\underline{m}-\underline{e}_j}^{(j)}} P_{\underline{m}-\underline{e}_j}(x)Q_{\underline{m}+\underline{e}_j}(y).$$

Thus, the left hand side depends only on $\underline{m} = \underline{m}_n$ and not on the particular choice of the sequence $\underline{m}_j$ leading to it.

The Riemann–Hilbert problem

There is an approach (see Van Assche–Geronimo–Kuijlaars [96]) to both types of multiple orthogonality in terms of matrix-valued Riemann–Hilbert problem for $(r+1)\times(r+1)$ matrices $Y=(Y_{ij}(z))_{i,j=0}^r$.

If $d\mu_j(x)=w_j dx$, then one requires that

- Y is analytic on $\mathbf{C}\setminus\mathbf{R}$,
- if $Y^{\pm}(x)$ denote the limit of $Y(z)$ as $z\to x\in\mathbf{R}$ from the upper, respectively the lower, half plane, then we have $Y^+(x)=Y^-(x)S(x)$, where

$$S(x):=\begin{bmatrix} 1 & w_1(x) & w_2(x) & \cdots & w_r(x) \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix},$$

- as $z\to\infty$

$$Y(z)=\left(I+O\left(\frac{1}{z}\right)\right)\begin{bmatrix} z^{|\underline{n}|} & 0 & 0 & \cdots & 0 \\ 0 & z^{-n_1} & 0 & \cdots & 0 \\ 0 & 0 & z^{-n_2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & z^{-n_r} \end{bmatrix}.$$

The first entry $Y_{11}(z)$ is precisely the orthogonal polynomial $P_{\underline{n}}$ of type II, and the other entries are also explicit in terms of the $P_{\underline{n}}$'s and w_j's (all other entries are either a constant multiple of $P_{\underline{n}-\underline{e}_k}$ or a Cauchy transform of its multiple with w_j). For type I orthogonality the transfer matrix is

$$S(x):=\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -w_1(x) & 1 & 0 & \cdots & 0 \\ -w_2(x) & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -w_r(x) & 0 & 0 & \cdots & 1 \end{bmatrix},$$

the behavior at infinity is of the form

$$Y(z)=\left(I+O\left(\frac{1}{z}\right)\right)\begin{bmatrix} z^{-|\underline{n}|} & 0 & 0 & \cdots & 0 \\ 0 & z^{n_1} & 0 & \cdots & 0 \\ 0 & 0 & z^{n_2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & z^{n_r} \end{bmatrix},$$

and the multiple orthogonal polynomials $Q_{\underline{n},j}$ are $Y_{1,j+1}/2\pi i$.

16 Matrix orthogonal polynomials

In the last 20 years the fundamentals of matrix orthogonal polynomials have been developed mainly by A. Durán and his coauthors (see also the work [9] by A. I. Aptekarev and E. M. Nikishin). The theory shows many similarities with the scalar case, but there is an unexpected richness which is still to be explored.

For all the results in this section see López-Rodriguez–Durán [48] and Durán–Grünbaum [25] and the numerous references there.

Matrix orthogonal polynomials

An $N \times N$ matrix

$$P(t) = \begin{pmatrix} p_{11}(t) & \cdots & p_{1N}(t) \\ \vdots & \ddots & \vdots \\ p_{N1}(t) & \cdots & p_{NN}(t) \end{pmatrix}$$

with polynomial entries $p_{ij}(t)$ of degree at most n is called a **matrix polynomial** of degree at most n. Alternatively, one can write

$$P(t) = C_n t^n + \cdots + C_0$$

with numerical matrices $C_n, \dots, C_0$ of size $N \times N$.

The number $t = a$ is called a **zero** of P if $P(a)$ is singular, and the **multiplicity** of a is the multiplicity of a as a zero of $\det P(a)$. When the **leading coefficient matrix** C_n is non-singular, then P has nN zeros counting multiplicity.

From now on we fix the dimension to be N, but the degree n can be any natural number. I will denote the $N \times N$ unit matrix and 0 stands for all kinds of zeros (numerical or matrix).

A **matrix**

$$W(t) = \begin{pmatrix} \mu_{11}(t) & \cdots & \mu_{1N}(t) \\ \vdots & \ddots & \vdots \\ \mu_{N1}(t) & \cdots & \mu_{NN}(t) \end{pmatrix}$$

of complex measures defined on (or part of) the real line is positive definite if for any Borel set E the numerical matrix $W(E)$ is positive semidefinite. We assume that all moments of W are finite. With such a matrix we can define a matrix inner product on the space of $N \times N$ matrix polynomials via

$$(P, Q) = \int P(t) dW(t) Q^*(t),$$

and if (P, P) is nonsingular for any P with nonsingular leading coefficient, then just as in the scalar case one can generate a sequence $\{P_n\}_{n=0}^{\infty}$ of matrix polynomials of degree $n = 0, 1, \dots$ which are orthonormal with respect to W:

$$\int P_n(t) dW(t) P_m^*(t) = \begin{cases} 0 & \text{if } n \neq m \\ I & \text{if } n = m, \end{cases}$$

and here P_n has nonsingular leading coefficient matrix. The sequence $\{P_n\}$ is determined only up to left multiplication by unitary matrices, i.e., if U_n are unitary matrices, then the polynomials U_nP_n also form an orthonormal system with respect to W.

Three-term recurrence and quadrature

Just as in the scalar case, these orthogonal polynomials satisfy a three-term recurrence relation

$$tP_n(t) = A_{n+1}P_{n+1}(t) + B_nP_n(t) + A_n^*P_{n-1}(t), \qquad n \geq 0, \tag{16.1}$$

where A_n are nonsingular matrices, and B_n are Hermitian. Conversely, the analogue of Favard's theorem is also true: if a sequence of matrix polynomials $\{P_n\}$ of corresponding degree $n = 0, 1, 2, \ldots$, satisfy (16.1) with nonsingular A_n and Hermitian B_n, then there is a positive definite measure matrix W such that the P_n are orthonormal with respect to W.

The three-term recurrence formula easily yields the Christoffel-Darboux formula:

$$(w-z)\sum_{k=0}^{n-1} P_k^*(z)P_k(w) = P_{n-1}^*(z)A_nP_n(w) - P_n^*(z)A_n^*P_{n-1}(w),$$

from which for example it follows that

$$P_{n-1}^*(z)A_nP_n(z) - P_n^*(z)A_n^*P_{n-1}(z) = 0,$$

$$\sum_{k=0}^{n-1} P_k^*(z)P_k(z) = P_{n-1}^*(z)A_nP_n'(z) - P_n^*(z)A_n^*P_{n-1}'(z).$$

The orthogonal polynomials Q_n of the second kind

$$Q_n(t) = \int \frac{P_n(t) - P_n(x)}{t-x}\,dW(x), \qquad n = 1, 2, \ldots,$$

also satisfy the same recurrence and are orthogonal with respect to some other matrix measure. For them we have

$$P_{n-1}^*(t)A_nQ_n(t) - P_n^*(z)A_n^*Q_{n-1}(t) \equiv I,$$

and

$$Q_n(t)P_{n-1}^*(t) - P_n(t)Q_{n-1}^*(t) \equiv A_n^{-1}.$$

With the recurrence coefficient matrices A_n, B_n one can form the **block Jacobi matrix**

$$J = \begin{pmatrix} B_0 & A_0 & 0 & 0 & \cdots \\ A_0^* & B_1 & A_1 & 0 & \cdots \\ 0 & A_1^* & B_2 & A_2 & \cdots \\ 0 & 0 & A_2^* & B_2 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

The zeros of P_n are real and they are the eigenvalues (with the same multiplicity) of the N-truncated block Jacobi matrix (which is of size nN). If a is a zero then its multiplicity p is at most N, the rank of $P_n(a)$ is $N-p$, and the space of those vectors v for which $P_n(a)v = 0$ is of dimension p. If we write $x_{n,k}$, $1 \le k \le m$, for the different zeros of P_n, and l_k is the multiplicity of $x_{n,k}$, then the matrices

$$\Gamma_k = \frac{1}{(\det(P_n(t)))^{(l_k)}(x_{n,k})} \left(\mathrm{Adj}(P_n(t))\right)^{(l_k-1)}(x_{n,k})Q_n(x_{n,k}), \quad 1 \le k \le m$$

are positive semidefinite of rank l_k, and with them the **matrix quadrature formula**

$$\int P(t)dW(t) = \sum_{k=1}^{m} P(x_{n,k})\Gamma_{n,k}$$

holds for every matrix polynomial P of degree at most $2n-1$.

If we assume that $A_n \to A$, $B_n \to B$ where A is non-singular, then

$$P_n(z)P_{n-1}^{-1}(z)A_n^{-1} \to \int \frac{dW_{A,B}(t)}{z-t}$$

locally uniformly outside the cluster set of the zeros, where $W_{A,B}(t)$ is the measure matrix of orthogonality for the sequence of matrix orthogonal polynomials S_n with recurrence coefficients A, B for all n, i.e., which satisfy the three-term recurrence

$$tS_n(t) = A^*S_{n+1}(t) + BS_n(t) + AS_{n-1}(t).$$

The distribution of the zeros themselves will be $1/N$-times the trace of the matrix measure of orthogonality for another sequence of matrix orthogonal polynomials R_n satisfying

$$tR_n(t) = AR_{n+1}(t) + BR_n(t) + AR_{n-1}(t), \qquad n \ge 2,$$

with appropriate modifications for $n < 2$.

Families of orthogonal polynomials

If the matrix of orthogonality is diagonal (or similar to a diagonal matrix) with diagonal entries μ_i, then the orthogonal matrix polynomials are also diagonal with i-th entry equal to $p_n(\mu_i)$, the n-th orthogonal polynomial with respect to μ_i. Many matrix orthogonal polynomials in the literature can be reduced to this scalar case. Recently however, some remarkably rich non-reducible families have been obtained by A. Duran and F. Grünbaum (see [25] and the references therein), which may play the role of the classical orthogonal polynomials in higher dimension. They found families of matrix orthogonal polynomials that satisfy second order (matrix) differential equations just like the classical orthogonal polynomials. Their starting point was a symmetry property between the orthogonality measure matrix and a second order differential operator. They worked out several explicit examples. Here is one of them: $N = 2$, the measure matrix (more precisely its density) is

$$H(t) := e^{-t^2} \begin{pmatrix} 1 + |a|^2 t^4 & at^2 \\ \overline{a}t^2 & 1 \end{pmatrix}, \qquad t \in \mathbf{R},$$

where $a \in \mathbf{C} \setminus \{0\}$ is a free parameter. The corresponding $P_n(t)$ satisfies

$$P_N''(t) + P_n'(t) \begin{pmatrix} -2t & 4at \\ 0 & -2t \end{pmatrix} \quad + \quad P_n(t) \begin{pmatrix} -4 & 2a \\ 0 & 0 \end{pmatrix}$$

$$= \quad \begin{pmatrix} -2n-4 & 2a(2n+1) \\ 0 & -2n \end{pmatrix} P_n(t).$$

There is an explicit Rodrigues type representation for the polynomials themselves, and the three-term recurrence (16.1) holds with $B_n = 0$,

$$A_{n+1} := \sqrt{\frac{n+1}{2}} \begin{pmatrix} \gamma_{n+3}/\gamma_{n+2} & a\gamma_{n+2}\gamma_{n+1} \\ 0 & \gamma_n/\gamma_{n+1} \end{pmatrix},$$

where

$$\gamma_n^2 := 1 + \frac{|a|^2}{2} \binom{n}{2}.$$

Connection with higher order scalar recurrence

Matrix orthogonality is closely connected to $(2N+1)$**-term recurrence**s for scalar polynomials. To describe this we need the following operators on polynomials p: if $p(t) = \sum_k a_k t^k$, then

$$R_{N,m}(p) = \sum_s a_{sN+m} t^s,$$

i.e., from a polynomial the operator $R_{N,m}$ takes those powers where the exponent is congruent to m modulo N, removes the common factor t^m and changes t^N to t.

Now suppose that $\{p_n\}_{n=0}^{\infty}$ is a sequence of scalar polynomials of corresponding degree $n = 0, 1, \ldots$, and suppose that this sequence satisfies a $(2N+1)$-term recurrence relation

$$t^N p_n(t) = c_{n,0} p_n(t) + \sum_{k=1}^{N} (\overline{c_{n,k}} p_{n-k}(t) + c_{n+k,k} p_{n+k}(t)),$$

where $c_{n,0}$ is real, $c_{n,N} \neq 0$ (and $p_k(t) \equiv 0$ for $k < 0$). Then

$$P_n(t) = \begin{pmatrix} R_{N,0}(p_{nN}) & \cdots & R_{N,N-1}(p_{nN}) \\ R_{N,0}(p_{nN+1}) & \cdots & R_{N,N-1}(p_{nN}+1) \\ \vdots & \ddots & \vdots \\ R_{N,0}(p_{nN+N-1}) & \cdots & R_{N,N-1}(p_{nN}+N-1) \end{pmatrix}$$

is a sequence of matrix orthogonal polynomials with respect to a positive definite measure matrix. Conversely, if $P_n = (P_{n,m,j})_{m,j=0}^{N-1}$ is a sequence of orthonormal matrix polynomials, then the scalar polynomials

$$p_{nN+m}(t) = \sum_{j=0}^{N-1} t^j P_{n,m,j}(t^N), \qquad 0 \leq m < N, \; n = 0, 1, 2, \ldots,$$

satisfy a $(2N+1)$-recurrence relation of the above form.

References

[1] N. I. Ahiezer, Orthogonal polynomials on several intervals, *Dokl. Akad. Nauk SSSR*, **134**(1960), 9–12. (Russian); translated as *Soviet Math. Dokl.*, **1**(1960), 989–992.

[2] N. I. Akhiezer, *The Classical Moment Problem and Some Related Questions in Analysis*, Oliver and Boyd, Edinburgh, 1965

[3] W. A. Al-Salam, Characterization theorems for orthogonal polynomials, in: Orthogonal Polynomials: Theory and Practice, M. Ismail and P. Nevai (eds), NATO ASI Series C, **294**, Kluwer, Dordrecht, 1990.

[4] M. P. Alfaro, M. Bello, J. M. Montaner and J. L. Varona, Some asymptotic properties for orthogonal polynomials with respect to varying measures, *J. Approx. Theory*, **135**(2005), 22–34.

[5] M. P. Alfaro and L. Vigil, Solution of a problem of P. Turán, *J. Approx. Theory*, **53**(1988), 195–197.

[6] G. E. Andrews, R. Askey and R. Roy, *Special Functions*, Encyclopedia of Mathematics and its Applications, Cambridge University Press, New York, 1999.

[7] A. I. Aptekarev, Asymptotic properties of polynomials orthogonal on a system of contours, and periodic motions of Toda chains. (Russian) *Mat. Sb. (N.S.)*, **125(167)**(1984), 231–258.

[8] A. I. Aptekarev, Multiple orthogonal polynomials, *J. Comput. Appl. Math.*, **99**(1988), 423–447.

[9] A. I. Aptekarev and E. M. Nikishin, The scattering problem for a discrete Sturm-Liouville operator, *Mat. Sb.* (N.S.), **121**(1983), 327–358. (Russian)

[10] A. I. Aptekarev and W. Van Assche, Scalar and matrix Riemann–Hilbert approach to the strong asymptotics of Padé approximants and complex orthogonal polynomials with varying weights, *J. Approx. Theory*, **129**(2004), 129–166.

[11] R. Askey, A transplantation theorem for Jacobi series. *Illinois J. Math.*, **13**(1969), 583–590.

[12] L. Baratchart, R. Küstner and V. Totik, Zero distribution via orthogonality, *Annales de l'Institut Fourier*, **55**(2005), 1455-1499.

[13] D. Barrios and G. López, Ratio asymptotics for polynomials orthogonal on arcs of the unit circle, *Constr. Approx.*, **15**(1999), 1–31.

[14] M. Bello and G. López, Ratio and relative asymptotics of polynomials orthogonal on an arc of the unit circle, *J. Approx. Theory*, **92**(1998), 216–244.

[15] O. Blumenthal, *Über die Entwicklung einer willkürlichen Funktion nach den Nennern des Kettenbruches für* $\int_{-\infty}^{0}[\phi(\xi)/(z-\xi)]\,d\xi$, Inaugural Dissertation, Göttingen, 1898.

[16] S. Bochner, Über Sturm-Liouvillesche Polynomsysteme, *Math. Zeitschrift*, **29**(1929), 730–736.

[17] J. A. Charris and M. E. H. Ismail, Sieved orthogonal polynomials. VII. Generalized polynomial mappings, *Trans. Amer. Math. Soc.*, **340**(1993), 71–93.

[18] T. S. Chihara, *An introduction to orthogonal polynomials.* Mathematics and its Applications, Vol. **13**, Gordon and Breach Science Publishers, New York-London-Paris, 1978.

[19] E. Daems and A. B. J. Kuijlaars, A Christoffel-Darboux formula for multiple orthogonal polynomials, *J. Approx. Theory*, **130**(2004), 190–202.

[20] P. Deift, *Orthogonal Polynomials and Random Matrices: A Riemann–Hilbert Approach*, Courant Lecture Notes, 1999.

[21] F. Delyon, B. Simon, and B. Souillard, From power pure point to continuous spectrum in disordered systems, *Annales Inst. Henri Poincaré*, **42**(1985), 283–309.

[22] S. Denisov, On Rakhmanov's theorem for Jacobi matrices, *Proc. Amer. Math. Soc.*, **132**(2004), 847–852.

[23] S. Denisov and B. Simon, Zeros of orthogonal polynomials on the real line, *J. Approx. Theory*, **121**(2003), 357–364.

[24] C. F. Dunkl and Y. Xu, *Orthogonal polynomials of several variables*, Encyclopedia of Mathematics and its Applciations, Cambridge University Press, New York, 2001.

[25] A. J. Durán and F. A. Grünbaum, A survey on orthogonal matrix polynomials satisfying second order differential equations, *J. Comput. Appl. Math.*, **178**(2005), 169-190.

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

[27] G. Freud, *Orthogonal Polynomials.* Akadémiai Kiadó/Pergamon Press, Budapest, 1971.

[28] W. Gautschi, *Orthogonal Polynomials: Computation and Approximation*, Numerical Mathematics and Scientific Computation, Oxford University Press, New York, 2004.

[29] W. Gautschi and A. B. J. Kuijlaars, Zeros and critical points of Sobolev orthogonal polynomials, *J. Approx. Theory*, **91**(1997), 117–137.

[30] J. Geronimo, D. S. Lubinsky and F. Marcellan, Asymptotics for Sobolev orthogonal polynomials for exponential weights, to appear in *Constructive Approx.*

[31] J. S. Geronimo and W. Van Assche, Orthogonal polynomials on several intervals via a polynomial mapping, *Trans. Amer. Math. Soc.*, **308**(1988), 559–581.

[32] Ya. L. Geronimus, On the character of the solution of the moment problem in the case of the periodic in the limit associated fraction, *Bull. Acad. Sci. USSR Math.*, **5**(1941), 203–210. [Russian]

[33] Ya. L. Geronimus, On polynomials orthogonal on the circle, on trigonometric moment problem, and on allied Carathéodory and Schur functions, *Mat. Sb.*, **15**(1944), 99–130. [Russian]

[34] Ya. L. Geronimus, On the trigonometric moment problem, *Ann. of Math.*, **47**(1946), 742–761.

[35] Ya. L. Geronimus, *Orthogonal polynomials on the unit circle and their applications*, Amer. Math. Soc. Translation, 1954, Vol. 104.

[36] L. Golinskii and V. Totik, Orthogonal polynomials from Stieltjes to Simon, Proc. Conf. honoring B. Simon, AMS Conference Series (to appear).

[37] A. A. Gonchar and E. A. Rakhmanov, On the convergence of simultaneous Padé approximants for systems of functions of Markov type, *Number theory, mathematical analysis and their applications. Trudy Mat. Inst. Steklov.*, **157**(1981), 31–48. (Russian)

[38] A. A. Gonchar and S. P. Suetin, On Padé approximation of Markov type meromorphic functions, *Current Problems in Mathematics, 5.*, Rossiiskaya Akademiya Nauk, Matematicheskii Institut im. V. A. Steklova, Moscow, 2004. 68 pp. (electronic). (Russian)

[39] M. E. H. Ismail, *Classical and quantum orthogonal polynomials in one variable*, Encyclopedia in Mathematics, Cambridge University Press, 2005.

[40] V. Kaliaguine, A note on the asymptotics of orthogonal polynomials on a complex arc: the case of a measure with a discrete part. , *J. Approx. Theory*, **80**(1995), 138–145.

[41] S. Khrushchev, A singular Riesz product in the Nevai class and inner functions with the Schur parameters in $\cap_{p>2}\ell^p$, *J. Approx. Theory*, **108**(2001), 249–255.

[42] S. Khrushchev, Continued fractions and orthogonal polynomials on the unit circle, Proc. Conf. OPSFA, Copenhagen, 2003.

[43] R. Killip and B. Simon, Sum rules for Jacobi matrices and their applications to spectral theory, *Annals of Math.*, **158**(2003), 253-321

[44] T. Kriecherbauer and K. T-R. McLaughlin, Strong asymptotics of polynomials orthogonal with respect to a Freud weight, *Internat. Math. Research Notices*, **6**(1999), 299–333.

[45] A. Levin and D. S. Lubinsky, *Christoffel Functions and Orthogonal Polynomials for Exponential Weights on* $[-1,1]$, Memoirs Amer. Math. Soc. , **535**, (1994)

[46] A. Levin and D. S. Lubinsky, *Orthogonal polynomials for exponential weights*, CMS Books in Mathematics/Ouvrages de Mathmatiques de la SMC, **4**, Springer-Verlag, New York, 2001.

[47] G. López, H. Pijeira-Cabrera and I. P. Izquierdo, Sobolev orthogonal polynomials in the complex plane, *J. Comput. Appl. Math.*, **127**(2001), 219-230.

[48] P. López-Rodriguez and A. J. Durán, Orthogonal matrix polynomials, Laredo Lecture Notes (Lecture notes of the 1st SIAG Summer School on Orthogonal Polynomials and Special Functions, Laredo, Spain, July 24-29 2000). Nova Science Publisher (2003). R. Álvarez-Nodarse, F. Marcellán and W. Van Assche (Eds.)

[49] D. S. Lubinsky, *Strong Asymptotics for Extremal Errors Associated with Erdös-type Weights*, Pitman Res. Notes, Vol. **202**, Longmans, Nurn Mill, Essex, 1989.

[50] D. S. Lubinsky and E. B. Saff, *Strong Asymptotics for Extremal Polynomials Associated with Weights on* $(-\infty, \infty)$, Springer Lecture Notes, Vol. **1305**, Springer Verlag, Berlin, 1988.

[51] F. Marcellán, M. Alfaro and M. L. Rezola, Orthogonal polynomials on Sobolev spaces: old and new directions, *J. Comput. Appl. Math.,* **48**(1993), 113–131.

[52] A. Martinez-Finkelshtein, Analytic aspects of Sobolev orthogonal polynomials, *J. Comput. Appl. Math.*, **99**(1998), 491–510.

[53] A. Martinez-Finkelshtein, Analytic properties of Sobolev orthogonal polynomials revisited, *J. Comput. Appl. Math.*, **127**(2001), 255-266.

[54] A. Martinez-Finkelshtein, Szegő polynomials: a view from the Riemann–Hilbert window, preprint arXiv:math. CA/0508117, 2005.

[55] G. Mastroianni and G. Milovanovic, *Approximation of Functions and Functionals with Application*, World Scientific, to appear.

[56] G. Mastroianni and V. Totik, Polynomial inequalities with doubling and A_∞ weights, *Constructive Approx.*, **16**(2000), 37–71.

[57] A. Máté, P. Nevai and V. Totik, Extensions of Szegő's theory of orthogonal polynomials, *Constructive Approx.*, **3**(1987), 51–72.

[58] A. Máté, P. Nevai and V. Totik, Strong and weak convergence of orthogonal polynomials, *Amer. J. Math.*, **109**(1987), 239–282.

[59] A. Máté, P. Nevai and V. Totik, Szegő's extremum problem on the unit circle, *Annals of Math.*, **134**(1991), 433–453.

[60] M. Mehta, *Random matrices*, second ed., Academic Press, Inc. , Boston, 1991.

[61] H. N. Mhaskar and E. B. Saff, On the distribution of zeros of polynomials orthogonal on the unit circle, *J. Approx. Theory*, **63**(1990), 30–38.

[62] P. Nevai, *Orthogonal Polynomials,* Memoirs Amer. Math. Soc., **213**(1979)

[63] P. Nevai and V. Totik, Orthogonal polynomials and their zeros, *Acta Sci. Math.*(Szeged), **53**(1989), 99–104.

[64] P. Nevai and V. Totik, Denisov's theorem on recurrent coefficients, *J. Approx. Theory*, **127**(2004), 240–245.

[65] E. M. Nikishin, An estimate for orthogonal polynomials, *Acta Sci. Math.*(Szeged), **48**(1985), 395–399. [Russian]

[66] J. Nuttall, Asymptotics for diagonal Hermite-Padé approximants, *J. Approx. Theory*, **42**(1984), 299–386.

[67] J. Nuttall, Padé polynomial asymptotics from a singular integral equation, *Constructive Approx.*, **6**(1990), 157–166.

[68] L. Pastur and A. Figotin, *Spectra of Random and Almost-Periodic Operators*, Grundlehren der mathematischen Wissenschaften, Vol. **297**, Springer Verlag, Berlin, 1992.

[69] F. Peherstorfer, On Bernstein-Szegő orthogonal polynomials on several intervals, *SIAM J. Math. Anal.*, **21**(1990), 461–482.

[70] F. Peherstorfer, On Bernstein-Szegő orthogonal polynomials on several intervals. II. Orthogonal polynomials with periodic recurrence coefficients, *J. Approx. Theory*, **64**(1991), 123–161.

[71] F. Peherstorfer, Orthogonal and extremal polynomials on several intervals, *J. Comput. Appl. Math.*, **48**(1993), 187–205.

[72] F. Peherstorfer, Zeros of polynomials orthogonal on several intervals. *Int. Math. Res. Notes*, **7**(2003), 361–385.

[73] F. Peherstorfer and R. Steinbauer, On polynomials orthogonal on several intervals, Special functions (Torino, 1993), *Ann. Numer. Math.*, **2**(1995), 353–370.

[74] E. A. Rakhmanov, On V. A. Steklov's problem in the theory of orthogonal polynomials, (Russian) *Dokl. Akad. Nauk SSSR*, **254**(1980), 802–806.

[75] E. A. Rakhmanov, On the asymptotics of the ratio of orthogonal polynomials, *Math. USSR Sbornik,* **46**(1983), 105–117; Russian original, *Mat. Sb.*, **118**(1982), 104–117.

[76] E. B. Saff and V. Totik, What parts of a measure's support attract zeros of the corresponding orthogonal polynomials? *Proc. Amer. Math. Soc.*, **114**(1992), 185–190.

[77] E. B. Saff and V. Totik, *Logarithmic Potentials with External Fields*, Grundlehren der mathematischen Wissenschaften, **316**, Springer-Verlag, Berlin, 1997.

[78] B. Simon, The classical moment problem as a self-adjoint finite difference operator, *Advances in Math.*, **137**(1998), 82–203.

[79] B. Simon, Ratio asymptotics and weak asymptotic measures for orthogonal polynomials on the real line, *J. Approx. Theory*, **126**(2004), 198–217.

[80] B. Simon, *Orthogonal Polynomials on the Unit Circle*, V. 1: Classical Theory, AMS Colloquium Series, American Mathematical Society, Providence, RI, 2005.

[81] B. Simon, *Orthogonal Polynomials on the Unit Circle*, V. 2: Spectral Theory, AMS Colloquium Series, American Mathematical Society, Providence, RI, 2005.

[82] B. Simon and A. Zlatoš, Sum rules and the Szegő condition for orthogonal polynomials on the real line, *Comm. Math. Phys.*, **242**(2003), 393–423.

[83] B. Simon and A. Zlatoš, Higher-order Szegő theorems with two singular points, *J. Approx. Theory*, **134**(2005), 114–129.

[84] H. Stahl, Divergence of diagonal Padé approximants and the asymptotic behavior of orthogonal polynomials associated with nonpositive measures, *Constr. Approx.*, **1**(1985), 249–270.

[85] H. Stahl, Orthogonal polynomials with complex-valued weight function. I, II. *Constr. Approx.*, **2**(1986), 225–240, 241–251.

[86] H. Stahl, Orthogonal polynomials with respect to complex-valued measures. Orthogonal polynomials and their applications (Erice, 1990), *IMACS Ann. Comput. Appl. Math.*, **9**(1991), 139–154.

[87] H. Stahl and V. Totik, *General Orthogonal Polynomials*, Encyclopedia of Mathematics and its Applications, **43**, Cambridge University Press, New York, 1992.

[88] S. P. Suetin, On the uniform convergence of diagonal Padé approximants for hyperelliptic functions, *Mat. Sb.*, **191**(2000), 81–114; (Russian), translation in *Sb. Math.*, **191**(2000), 1339–1373.

[89] S. P. Suetin, Padé approximants and the effective analytic continuation of a power series, *Uspekhi Mat. Nauk,* **57**(2002), 45–142; (Russian), translation in *Russian Math. Surveys*, **57**(2002), 43–141.

[90] J. Szabados and P. Vértesi, *Interpolation of Functions*, World Scientific, Singapore, 1990.

[91] G. Szegő, *Orthogonal Polynomials*, Coll. Publ. , **XXIII**, Amer. Math. Soc. , Providence, 1975.

[92] V. Totik, Orthogonal polynomials with ratio asymptotics, *Proc. Amer. Math. Soc.*, **114**(1992), 491–495.

[93] V. Totik, Asymptotics for Christoffel functions for general measures on the real line, *J. D'Analyse Math.*, **81**(2000), 283–303.

[94] W. Van Assche, *Asymptotics for Orthogonal Polynomials*, Spriger Lecture Notes, Vol. **1265**, Springer Verlag, Berlin, 1987.

[95] W. Van Assche, Hermite-Padé approximation and multiple orthogonal polynomials, preprint.

[96] W. Van Assche, J. S. Geronimo, A. B. J. Kuijlaars, Riemann–Hilbert problems for multiple orthogonal polynomials, in: J. Bustoz et al. (eds), Special Functions 2000: Current Perspectives and Future Directions, Kluwer, Dordrecht, 2001, 23–59.

[97] H. Widom, Extremal polynomials associated with a system of curves in the complex plane, *Adv. Math.*, **3**(1969), 127–232.

Bolyai Institute, University of Szeged
Szeged
Aradi v. tere 1, 6720
Hungary, and

Department of Mathematics
University of South Florida
Tampa, FL, 33620
USA