Chapter 8

Orthogonal Polynomials

> Although our intellect always longs for clarity and certainty, our nature often finds uncertainty fascinating.
>
> *On War*
> KARL VON CLAUSEWITZ

Orthogonal polynomials are an important example of orthogonal decompositions of Hilbert spaces. They are also of great practical importance: they play a central role in numerical integration using quadrature rules (Chapter 9) and approximation theory; in the context of UQ, they are also a foundational tool in polynomial chaos expansions (Chapter 11). There are multiple equivalent characterizations of orthogonal polynomials via their three-term recurrence relations, via differential operators, and other properties; however, since the primary use of orthogonal polynomials in UQ applications is to provide an orthogonal basis of a probability space, these notes take L^2-orthogonality as the primary definition, with the spectral properties as consequent theorems.

For the rest of this chapter, $\mathcal{N} = \mathbb{N}_0$ or $\{0, 1, \dots, N\}$ for some $N \in \mathbb{N}_0$. For simplicity, we work over $\mathbb{R}$ instead of $\mathbb{C}$, and so the L^2 inner product is a symmetric bilinear form rather than a conjugate-symmetric sesquilinear form.

8.1 Basic Definitions and Properties

Recall that a real *polynomial* in a single indeterminate x is an expression of the form

$$p(x) = c_n x^n + c_{n-1} x^{n-1} + \cdots + c_1 x + c_0,$$

where the coefficients $c_i \in \mathbb{R}$ are scalars. The greatest $n \in \mathbb{N}_0$ for which $c_n \neq 0$ is called the *degree* of p, $\deg(p)$; sometimes it is convenient to regard the zero polynomial as having degree -1. If p is of degree n and $c_n = 1$, then p is said to be *monic*. The space of all (real) polynomials in x is denoted $\mathbb{R}[x]$, and the space of polynomials of degree at most n is denoted $\mathbb{R}[x]_{\leq n}$.

Definition 8.1. Let μ be a non-negative measure on $\mathbb{R}$. A family of polynomials $\mathcal{Q} = \{q_n \mid n \in \mathcal{N}\} \subseteq \mathbb{R}[x]$ is called an *orthogonal system of polynomials* if $\deg(q_n) = n$, $q_n \in L^2(\mathbb{R}, \mu)$, and

$$\langle q_m, q_n \rangle_{L^2(\mu)} := \int_{\mathbb{R}} q_m(x) q_n(x) \, \mathrm{d}\mu(x) = 0 \iff m, n \in \mathcal{N} \text{ are distinct.}$$

That is, $\langle q_m, q_n \rangle_{L^2(\mu)} = \gamma_n \delta_{mn}$ for some some strictly positive constants

$$\gamma_n := \|q_n\|^2_{L^2(\mu)} = \int_{\mathbb{R}} q_n^2 \, \mathrm{d}\mu,$$

called the *normalization constants* of the system $\mathcal{Q}$. If $\gamma_n = 1$ for all $n \in \mathcal{N}$, then $\mathcal{Q}$ is an *orthonormal system*.

In other words, a system of orthogonal polynomials is nothing but a collection of non-trivial orthogonal elements of the Hilbert space $L^2(\mathbb{R}, \mu)$ that happen to be polynomials, with some natural conditions on the degrees of the polynomials. Note that, given μ, orthogonal (resp. orthonormal) polynomials for μ can be found inductively by using the Gram–Schmidt orthogonalization (resp. orthonormalization) procedure on the monomials $1, x, x^2, \dots$. In practice, however, the Gram–Schmidt procedure is numerically unstable, so it is more common to generate orthogonal polynomials by other means, e.g. the three-term recurrence relation (Theorem 8.10).

Example 8.2. 1. The *Legendre polynomials* Le_n (also commonly denoted by P_n in the literature), indexed by $n \in \mathbb{N}_0$, are orthogonal polynomials for uniform measure on $[-1, 1]$:

$$\int_{-1}^{1} \mathrm{Le}_m(x)\mathrm{Le}_n(x)\,\mathrm{d}x = \frac{2}{2n+1}\delta_{mn}.$$

2. The Legendre polynomials arise as the special case $\alpha = \beta = 0$ of the *Jacobi polynomials* $P_n^{(\alpha,\beta)}$, defined for $\alpha, \beta > -1$ and indexed by $n \in \mathbb{N}_0$, which are orthogonal polynomials for the beta distribution $(1-x)^\alpha(1-x)^\beta\,\mathrm{d}x$ on $[-1, 1]$:

$$\int_{-1}^{1} P_m^{(\alpha,\beta)}(x)P_n^{(\alpha,\beta)}(x)\,\mathrm{d}x = \frac{2^{\alpha+\beta+1}\Gamma(n+\alpha+1)\Gamma(n+\beta+1)}{n!(2n+\alpha+\beta+1)\Gamma(n+\alpha+\beta+1)}\delta_{mn}.$$

3. Other notable special cases of the Jacobi polynomials include the *Chebyshev polynomials of the first kind* T_n, which are the special case $\alpha = \beta = -\frac{1}{2}$, and the *Chebyshev polynomials of the second kind* U_n, which are the special case $\alpha = \beta = \frac{1}{2}$. The Chebyshev polynomials are intimately connected with trigonometric functions: for example,

$$T_n(x) = \cos(n \arccos(x)) \quad \text{for } |x| \leq 1,$$

and the n roots of T_n are $z_j := \cos\left(\frac{\pi}{2}\frac{2j-1}{n}\right)$ for $j = 1, \dots, n$.

4. The (*associated*) *Laguerre polynomials* $\mathrm{La}_n^{(\alpha)}$, defined for $\alpha > -1$ and indexed by $n \in \mathbb{N}_0$, are orthogonal polynomials for the gamma distribution $x^\alpha e^{-x}\,\mathrm{d}x$ on the positive real half-line:

$$\int_0^\infty \mathrm{La}_m^{(\alpha)}(x)\mathrm{La}_n^{(\alpha)}(x)x^\alpha e^{-x}\,\mathrm{d}x = \frac{\Gamma(1+\alpha+n)}{n!}\delta_{mn}.$$

The polynomials $\mathrm{La}_n := \mathrm{La}_n^{(0)}$ are known simply as the *Laguerre polynomials*.

5. The *Hermite polynomials* He_n, indexed by $n \in \mathbb{N}_0$, are orthogonal polynomials for standard Gaussian measure $\gamma := (2\pi)^{-1/2}e^{-x^2/2}\,\mathrm{d}x$ on $\mathbb{R}$:

$$\int_{-\infty}^{\infty} \mathrm{He}_m(x)\mathrm{He}_n(x)\frac{\exp(-x^2/2)}{\sqrt{2\pi}}\,\mathrm{d}x = n!\delta_{mn}.$$

Together, the Jacobi, Laguerre and Hermite polynomials are known as the *classical orthogonal polynomials*. They encompass the essential features of orthogonal polynomials on the real line, according to whether the (absolutely continuous) measure μ that generates them is supported on a bounded interval, a semi-infinite interval, or the whole real line. (The theory of orthogonal polynomials generated by discrete measures is similar, but has some additional complications.) The first few Legendre, Hermite and Chebyshev polynomials are given in Table 8.1 and illustrated in Figure 8.1. See Tables 8.2 and 8.3 at the end of the chapter for a summary of some other classical systems of orthogonal polynomials corresponding to various probability measures on subsets of the real line. See also Figure 8.4 for an illustration of the Askey scheme, which classifies the various limit relations among families of orthogonal polynomials.

n	$\mathrm{Le}_n(x)$	$\mathrm{He}_n(x)$	$T_n(x)$
0	1	1	1
1	x	x	x
2	$\frac{1}{2}(3x^2-1)$	x^2-1	$2x^2-1$
3	$\frac{1}{2}(5x^3-3x)$	x^3-3x	$4x^3-3x$
4	$\frac{1}{8}(35x^4-30x^2+3)$	x^4-6x^2+3	$8x^4-8x^2+1$
5	$\frac{1}{8}(63x^5-70x^3+15x)$	x^5-10x^3+15x	$16x^5-20x^3+5x$

Table 8.1: The first few Legendre polynomials Le_n, which are orthogonal polynomials for uniform measure $\mathrm{d}x$ on $[-1, 1]$; Hermite polynomials He_n, which are orthogonal polynomials for standard Gaussian measure $(2\pi)^{-1/2}e^{-x^2/2}\,\mathrm{d}x$ on $\mathbb{R}$; and Chebyshev polynomials of the first kind T_n, which are orthogonal polynomials for the measure $(1-x^2)^{-1/2}\,\mathrm{d}x$ on $[-1, 1]$.

Remark 8.3. Many sources, typically physicists' texts, use the weight function $e^{-x^2}\,\mathrm{d}x$ instead of probabilists' preferred $(2\pi)^{-1/2}e^{-x^2/2}\,\mathrm{d}x$ or $e^{-x^2/2}\,\mathrm{d}x$ for the Hermite polynomials. Changing from one normalization to the other is not difficult, but special care must be exercised in practice to see which normalization a source is using, especially when relying on third-party software packages.⟨8.1⟩ To convert integrals with respect to one Gaussian measure to integrals with respect to another (and hence get the right answers for Gauss–Hermite quadrature), use the following change-of-variables formula:

$$\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} f(x)e^{-x^2/2}\,\mathrm{d}x = \frac{1}{\pi}\int_{\mathbb{R}} f(\sqrt{2}x)e^{-x^2}\,\mathrm{d}x.$$

It follows from this that conversion between the physicists' and probabilists' Gauss–Hermite quadrature formulæ (see Chapter 9) is achieved by

$$w_i^{\mathrm{prob}} = \frac{w_i^{\mathrm{phys}}}{\sqrt{\pi}}, \quad x_i^{\mathrm{prob}} = \sqrt{2}x_i^{\mathrm{phys}}.$$

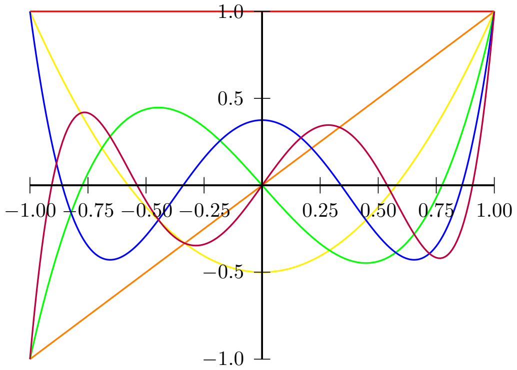
Existence of Orthogonal Polynomials. One thing that should be immediately obvious is that if the measure μ is supported on only $N \in \mathbb{N}$ points, then $\dim L^2(\mathbb{R}, \mu) = N$, and so μ admits only N orthogonal polynomials. This observation begs the question: what conditions on μ are necessary in order to ensure the existence of a desired number of orthogonal polynomials for μ? Recall that a matrix A is called a *Hankel matrix* if it has constant anti-diagonals, i.e. if a_{ij} depends only upon $i+j$. The definiteness of $L^2(\mu)$ inner products, and hence the existence of orthogonal polynomials, is intimately connected to determinants of Hankel matrices of moments of the measure μ:

Lemma 8.4. *The $L^2(\mu)$ inner product is positive definite on $\mathbb{R}[x]_{\leq d}$ if and only if the Hankel determinant $\det(H_n)$ is strictly positive for $n = 1, \dots, d+1$, where*

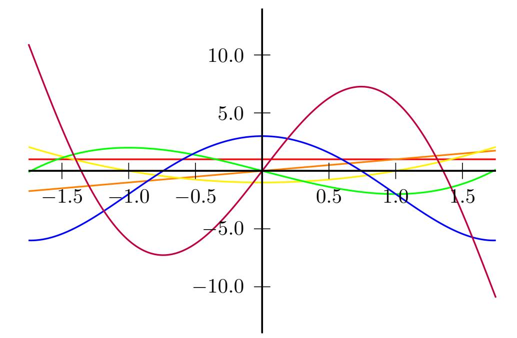
$$H_n := \begin{bmatrix} m_0 & m_1 & \cdots & m_{n-1} \\ m_1 & m_2 & \cdots & m_n \\ \vdots & \vdots & \ddots & \vdots \\ m_{n-1} & m_n & \cdots & m_{2n-2} \end{bmatrix}, \quad m_n := \int_{\mathbb{R}} x^n\,\mathrm{d}\mu(x). \tag{8.1}$$

Hence, the $L^2(\mu)$ inner product is positive definite on $\mathbb{R}[x]$ if and only if, for all $n \in \mathbb{N}$, $0 < \det(H_n) < \infty$.

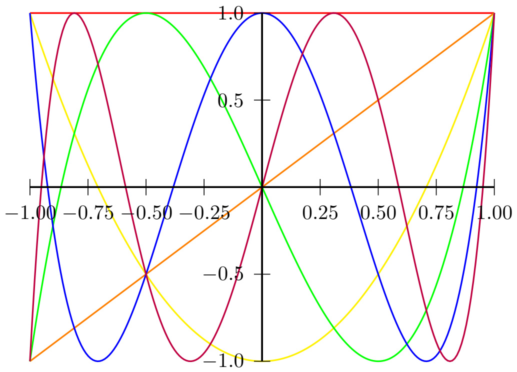
⟨8.1⟩ For example, the `GAUSSQ` Gaussian quadrature package from http://netlib.org/ uses the physicists' $e^{-x^2}\,\mathrm{d}x$ normalization. The `numpy.polynomial` package for Python provides separate interfaces to the physicists' and probabilists' Hermite polynomials, quadrature rules, &c. as `numpy.polynomial.hermite` and `numpy.polynomial.hermite_e` respectively.



(a) Legendre polynomials, Le_n



(b) Hermite polynomials, He_n



(c) Chebyshev polynomials of the first kind, T_n

Figure 8.1: The Legendre, Hermite and Chebyshev polynomials of degrees 0 (red), 1 (orange), 2 (yellow), 3 (green), 4 (blue) and 5 (purple) on $[-1, 1]$, $\mathbb{R}$ and $[-1, 1]$ respectively.

Proof. Let $p(x) := c_d x^d + \dots + c_1 x + c_0 \in \mathbb{R}[x]_{\leq d}$ be arbitrary. Note that

$$\|p\|_{L^2(\mu)}^2 = \int_{\mathbb{R}} \sum_{k,\ell=0}^{d} c_k c_\ell x^{k+\ell} \, \mathrm{d}\mu(x) = \sum_{k,\ell=0}^{d} c_k c_\ell m_{k+\ell},$$

and so $\|p\|_{L^2(\mu)} \in (0, \infty)$ if and only if H_{d+1} is a positive-definite matrix. This, in turn, is equivalent to having $\det(H_n) \in (0, \infty)$ for $n = 1, 2, \dots, d+1$. □

Theorem 8.5. *If the $L^2(\mu)$ inner product is positive definite on $\mathbb{R}[x]$, then there exists an infinite sequence of orthogonal polynomials for μ.*

Proof. Apply the Gram–Schmidt procedure to the monomials x^n, $n \in \mathbb{N}_0$. That is, take $q_0(x) = 1$, and for $n \in \mathbb{N}$ recursively define

$$q_n(x) := x^n - \sum_{k=0}^{n-1} \frac{\langle x^k, q_k \rangle}{\langle q_k, q_k \rangle} q_k(x).$$

Since the inner product is positive definite, $\langle q_k, q_k \rangle > 0$, and so each q_n is uniquely defined. By construction, each q_n is orthogonal to q_k for $k < n$. □

By Exercise 8.1, the hypothesis of Theorem 8.5 is satisfied if the measure μ has infinite support and all polynomials are μ-integrable. For example, there are infinitely many Legendre polynomials because polynomials are bounded on $[-1, 1]$, and hence integrable with respect to uniform (Lebesgue) measure; polynomials are unbounded on $\mathbb{R}$, but are integrable with respect to Gaussian measure by Fernique's theorem (Theorem 2.43), so there are infinitely many Hermite polynomials. In the other direction, there is the following converse result:

Theorem 8.6. *If the $L^2(\mu)$ inner product is positive definite on $\mathbb{R}[x]_{\leq d}$, but not on $\mathbb{R}[x]_{\leq n}$ for any $n > d$, then μ admits only $d + 1$ orthogonal polynomials.*

Proof. The Gram–Schmidt procedure can be applied so long as the denominators $\langle q_k, q_k \rangle$ are strictly positive and finite, i.e. for $k \leq d + 1$. The polynomial q_{d+1} is orthogonal to q_n for $n \leq d$; we now show that $q_{d+1} = 0$. By assumption, there exists a polynomial p of degree $d + 1$, having the same leading coefficient as q_{d+1}, such that $\|p\|_{L^2(\mu)}$ is 0, ∞, or even undefined; for simplicity, consider the case $\|p\|_{L^2(\mu)} = 0$, as the other cases are similar. Hence, $p - q_{d+1}$ has degree d, so it can be written in the orthogonal basis $\{q_0, \dots, q_d\}$ as

$$p - q_{d+1} = \sum_{k=0}^{d} c_k q_k$$

for some coefficients $c_0, \dots, c_d$. Hence,

$$0 = \|p\|_{L^2(\mu)}^2 = \|q_{d+1}\|_{L^2(\mu)}^2 + \sum_{k=0}^{d} c_k^2 \|q_k\|_{L^2(\mu)}^2,$$

which implies, in particular, that $\|q_{d+1}\|_{L^2(\mu)} = 0$. Hence, the normalization constant $\gamma_{d+1} = 0$, which is not permitted, and so q_{d+1} is not a member of a sequence of orthogonal polynomials for μ. □

Theorem 8.7. *If μ has finite moments only of degrees $0, 1, \dots, r$, then μ admits only a finite system of orthogonal polynomials $q_0, \dots, q_d$, where d is the minimum of $\lfloor r/2 \rfloor$ and $\#\operatorname{supp}(\mu) - 1$.*

Proof. Exercise 8.2. □

Theorem 8.8. *The coefficients of any system of orthogonal polynomials are determined, up to multiplication by an arbitrary constant for each degree, by the Hankel determinants of the polynomial moments. That is, if m_n and H_n are as in (8.1), then the n^{th} degree orthogonal polynomial q_n for μ is, for some $c_n \neq 0$,*

$$q_n = c_n \det \left[\begin{array}{ccc|c} & & & m_n \\ & H_n & & \vdots \\ & & & m_{2n-1} \\ \hline 1 & \cdots & x^{n-1} & x^n \end{array} \right]$$
$$= c_n \det \begin{bmatrix} m_0 & m_1 & m_2 & \cdots & m_n \\ m_1 & m_2 & m_3 & \cdots & m_{n+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ m_{n-1} & m_n & m_{n+2} & \cdots & m_{2n-1} \\ 1 & x & x^2 & \cdots & x^n \end{bmatrix}.$$

Proof. Exercise 8.6. □

Completeness of Orthogonal Polynomial Bases. A subtle point in the theory of orthogonal polynomials is that although an infinite family $\mathcal{Q}$ of orthogonal polynomials for μ forms an orthogonal set in $L^2(\mathbb{R}, \mu)$, it is *not* always true that $\mathcal{Q}$ forms a complete orthogonal basis for $L^2(\mathbb{R}, \mu)$, i.e. it is possible that

$$\overline{\operatorname{span} \mathcal{Q}} \subsetneq L^2(\mathbb{R}, \mu).$$

Examples of sufficient conditions for $\mathcal{Q}$ to form a complete orthogonal basis for $L^2(\mathbb{R}, \mu)$ include finite exponential moments (i.e. $\mathbb{E}_{X \sim \mu}[\exp(a|X|)]$ is finite for some $a > 0$), or the even stronger condition that the support of μ is a bounded set. A more detailed discussion can be found in the paper of Ernst & al. [50]. See Exercise 8.7 for the construction of an explicit example of an incomplete but infinite set of orthogonal polynomials, namely those corresponding to the probability distribution of a log-normal random variable.

8.2 Recurrence Relations

An æsthetically pleasing fact about orthogonal polynomials, and one that is of vital importance in numerical methods, is that every system of orthogonal polynomials satisfies a *three-term recurrence relation* of the form

$$q_{n+1}(x) = (A_n x + B_n) q_n(x) - C_n q_{n-1}(x) \tag{8.2}$$

for some sequences (A_n), (B_n), (C_n), with the initial terms $q_0(x) = 1$ and $q_{-1}(x) = 0$. There are many variations in the way that this three-term recurrence is presented: another one, which is particularly commonly used for orthogonal polynomials arising from discrete measures, is

$$-xq_n(x) = A_n q_{n+1}(x) - (A_n + C_n) q_n(x) + C_n q_{n-1}(x) \tag{8.3}$$

Example 8.9. The Legendre, Hermite and Chebyshev polynomials satisfy the recurrence relations

$$\begin{aligned} \mathrm{Le}_{n+1}(x) &= \frac{2n+1}{n+1} x \mathrm{Le}_n(x) - \frac{n}{n+1} \mathrm{Le}_{n-1}(x), \\ \mathrm{He}_{n+1}(x) &= x\mathrm{He}_n(x) - n\mathrm{He}_{n-1}(x), \\ T_{n+1}(x) &= 2xT_n(x) - T_{n-1}(x). \end{aligned}$$

These relations can all be verified by direct substitution and an integration by parts with respect to the appropriate generating measure μ on $\mathbb{R}$. The Jacobi polynomials also satisfy the three-term recurrence (8.2) with

$$\begin{aligned} A_n &= \frac{(2n+1+\alpha+\beta)(2n+2+\alpha+\beta)}{2(n+1)(n+1+\alpha+\beta)} \\ B_n &= \frac{(\alpha^2-\beta^2)(2n+1+\alpha+\beta)}{2(n+1)(2n+\alpha+\beta)(n+1+\alpha+\beta)} \\ C_n &= \frac{(n+\alpha)(n+\beta)(2n+2+\alpha+\beta)}{(n+1)(n+1+\alpha+\beta)(2n+\alpha+\beta)}. \end{aligned} \tag{8.4}$$

The coefficients for the three-term recurrence relation are determined (up to multiplication by a constant for each degree) by the following theorem, which gives the coefficients for the *monic* orthogonal polynomials associated to a measure μ:

Theorem 8.10. *Let $\mathcal{Q} = \{q_n \mid n \in \mathcal{N}\}$ be the monic orthogonal polynomials for a measure μ. Then*

$$\begin{aligned} q_{n+1}(x) &= (x-\alpha_n)q_n(x) - \beta_n q_{n-1}(x), \\ q_0(x) &= 1, \\ q_{-1}(x) &= 0, \end{aligned} \tag{8.5}$$

where

$$\begin{aligned} \alpha_n &:= \frac{\langle xq_n, q_n\rangle_{L^2(\mu)}}{\langle q_n, q_n\rangle_{L^2(\mu)}}, \quad \text{for } n \geq 0, \\ \beta_n &:= \frac{\langle q_n, q_n\rangle_{L^2(\mu)}}{\langle q_{n-1}, q_{n-1}\rangle_{L^2(\mu)}}, \quad \text{for } n \geq 1 \\ \beta_0 &:= \langle q_0, q_0\rangle_{L^2(\mu)} \equiv \int_{\mathbb{R}} \mathrm{d}\mu. \end{aligned}$$

Hence, the orthonormal polynomials $\{p_n \mid n \in \mathcal{N}\}$ for μ satisfy

$$\begin{aligned} \sqrt{\beta_{n+1}} p_{n+1}(x) &= (x-\alpha_n)p_n(x) - \sqrt{\beta_n} p_{n-1}(x), \\ p_0(x) &= \beta_0^{-1/2}, \\ p_{-1}(x) &= 0, \end{aligned} \tag{8.6}$$

Proof. First, note that the L^2 inner product[8.2] satisfies the *shift property*

$$\langle xf, g\rangle_{L^2(\mu)} = \langle f, xg\rangle_{L^2(\mu)} \tag{8.7}$$

for all $f, g \colon \mathbb{R} \to \mathbb{R}$ for which either side exists.

Since $\deg(q_{n+1} - xq_n) \leq n$, it follows that

$$q_{n+1}(x) - xq_n(x) = -\alpha_n q_n(x) - \beta_n q_{n-1}(x) + \sum_{j=0}^{n-2} c_{nj} q_j(x) \tag{8.8}$$

for suitable scalars α_n, β_n and c_{nj}. Taking the inner product of both sides of (8.8) with q_n yields, by orthogonality,

$$-\langle xq_n, q_n\rangle_{L^2(\mu)} = -\alpha_n \langle q_n, q_n\rangle_{L^2(\mu)},$$

so that $\alpha_n = \langle xq_n, q_n\rangle_{L^2(\mu)} / \langle q_n, q_n\rangle_{L^2(\mu)}$ as claimed. The expression for β_n is obtained similarly, by taking the inner product of (8.8) with q_{n-1} instead of with q_n. Finally, taking the inner product of (8.8) with q_j for $j < n-1$ yields

$$-\langle xq_n, q_j\rangle_{L^2(\mu)} = c_{nj} \langle q_j, q_j\rangle_{L^2(\mu)}. \tag{8.9}$$

[8.2] The Sobolev inner product, for example, does not satisfy the shift property (8.7). Hence, the recurrence theory for Sobolev orthogonal polynomials is more complicated than the L^2 case considered in these notes.

It follows from the shift property (8.7) that $\langle xq_n, q_j\rangle_{L^2(\mu)} = \langle q_n, xq_j\rangle_{L^2(\mu)}$. Since $\deg(xq_j) \leq n-1$, it follows that the left-hand side of (8.9) vanishes, so $c_{nj} \equiv 0$, and the recurrence relation (8.5) is proved. □

Furthermore, there is a converse theorem, which provides a characterization of precisely which sequences (A_n), (B_n), (C_n) arise from systems of orthogonal polynomials:

Theorem 8.11 (Favard). *Let (A_n), (B_n), (C_n) be real sequences and let $\mathcal{Q} = \{q_n \mid n \in \mathcal{N}\}$ be defined by*

$$\begin{aligned} q_{n+1}(x) &= (A_n x + B_n)q_n(x) - C_n q_{n-1}(x), \\ q_0(x) &= 1, \\ q_{-1}(x) &= 0. \end{aligned}$$

Then $\mathcal{Q}$ is a system of orthogonal polynomials for some measure μ if and only if, for all $n \in \mathcal{N}$,

$$A_n \neq 0, \quad C_n \neq 0, \quad C_n A_n A_{n-1} > 0.$$

The proof of Favard's theorem is beyond the scope of these notes.

A useful consequence of the three-term recurrence is the following *Christoffel–Darboux formula* for sums of products of orthogonal polynomial values at two points:

Theorem 8.12 (Christoffel–Darboux formula). *The orthonormal polynomials $\{p_n \mid n \in \mathcal{N}\}$ for a measure μ satisfy*

$$\sum_{k=0}^{n} p_k(y)p_k(x) = \sqrt{\beta_{n+1}}\frac{p_{n+1}(y)p_n(x) - p_n(y)p_{n+1}(x)}{y-x}, \tag{8.10}$$

and

$$\sum_{k=0}^{n} |p_k(x)|^2 = \sqrt{\beta_{n+1}}\big(p'_{n+1}(x)p_n(x) - p'_n(x)p_{n+1}(x)\big). \tag{8.11}$$

Proof. Multiply the recurrence relation (8.6), i.e.

$$\sqrt{\beta_{k+1}}p_{k+1}(x) = (x-\alpha_k)p_k(x) - \sqrt{\beta_k}p_{k-1}(x),$$

by $p_k(y)$ on both sides and subtract the corresponding expression with x and y interchanged to obtain

$$\begin{aligned} (y-x)p_k(y)p_k(x) = {} & \sqrt{\beta_{k+1}}\big(p_{k+1}(y)p_k(x) - p_k(y)p_{k+1}(x)\big) \\ & - \sqrt{\beta_k}\big(p_k(y)p_{k-1}(x) - p_{k-1}(y)p_k(x)\big). \end{aligned}$$

Sum both sides from $k=0$ to $k=n$ and use the telescoping nature of the sum on the right to obtain (8.10). Take the limit as $y \to x$ to obtain (8.11). □

Corollary 8.13. *The orthonormal polynomials $\{p_n \mid n \in \mathcal{N}\}$ for a measure μ satisfy*

$$p'_{n+1}(x)p_n(x) - p'_n(x)p_{n+1}(x) > 0.$$

Proof. Since $\beta_n > 0$ for all n, (8.11) implies that $p'_{n+1}(x)p_n(x) - p'_n(x)p_{n+1}(x) \geq 0$, and it vanishes if and only if the sum on the left-hand side of (8.11) vanishes. However, since

$$\sum_{k=0}^{n} |p_k(x)|^2 \geq |p_0(x)|^2 = \beta_0^{-1} > 0,$$

the claim follows. □

8.3 Differential Equations

In addition to their orthogonality and recurrence properties, the classical orthogonal polynomials are eigenfunctions for second-order differential operators. In particular, these operators take the form

$$\mathcal{L} = Q(x)\frac{\mathrm{d}^2}{\mathrm{d}x^2} + L(x)\frac{\mathrm{d}}{\mathrm{d}x},$$

where $Q(x) \in \mathbb{R}[x]_{\leq 2}$ and $L(x) \in \mathbb{R}[x]_{\leq 1}$, and the degree-$n$ orthogonal polynomial q_n satisfies

$$(\mathcal{L}q_n)(x) \equiv Q(x)q_n''(x) + L(x)q_n'(x) = \lambda_n q_n(x), \tag{8.12}$$

where the eigenvalue λ_n is $n\big(\frac{n-1}{2}Q'' + L'\big)$. Note that it makes sense for Q to be quadratic and L to be linear, since then (8.12) is an equality of two degree-n polynomials.

Example 8.14. 1. The Jacobi polynomials satisfy $\mathcal{L}P_n^{(\alpha,\beta)} = \lambda_n P_n^{(\alpha,\beta)}$, where

$$\begin{aligned}\mathcal{L} &:= (1-x^2)\frac{\mathrm{d}^2}{\mathrm{d}x^2} + (\beta - \alpha - (\alpha+\beta+2)x)\frac{\mathrm{d}}{\mathrm{d}x},\\ \lambda_n &:= -n(n+\alpha+\beta+1).\end{aligned}$$

2. The Hermite polynomials satisfy $\mathcal{L}\mathrm{He}_n = \lambda_n \mathrm{He}_n$, where

$$\begin{aligned}\mathcal{L} &:= \frac{\mathrm{d}^2}{\mathrm{d}x^2} - x\frac{\mathrm{d}}{\mathrm{d}x},\\ \lambda_n &:= -n.\end{aligned}$$

3. The Laguerre polynomials satisfy $\mathcal{L}\mathrm{La}_n^{(\alpha)} = \lambda_n \mathrm{La}_n^{(\alpha)}$, where

$$\begin{aligned}\mathcal{L} &:= x\frac{\mathrm{d}^2}{\mathrm{d}x^2} - (1+\alpha-x)\frac{\mathrm{d}}{\mathrm{d}x},\\ \lambda_n &:= -n.\end{aligned}$$

It is not difficult to verify that if $\mathcal{Q} = \{q_n \mid n \in \mathcal{N}\}$ is a system of monic orthogonal polynomials, which therefore satisfy the three-term recurrence

$$q_{n+1}(x) = (x - \alpha_n)q_n(x) - \beta_n q_{n-1}(x),$$

then q_n is an eigenfunction for $\mathcal{L}$ with eigenvalue $n\big(\frac{n-1}{2}Q'' + L'\big)$ — simply apply the three-term recurrence to the claimed equation $\mathcal{L}q_{n+1} = \lambda_{n+1}q_{n+1}$ and examine the highest-degree terms. What is more difficult to show is the converse result (which uses results from Sturm–Liouville theory and is beyond the scope of these notes) that, subject to suitable conditions on Q and L, the *only* eigenfunctions of $\mathcal{L}$ are polynomials of the correct degrees, with the claimed eigenvalues, orthogonal under the measure $\mathrm{d}\mu = w(x)\,\mathrm{d}x$, where

$$w(x) \propto \frac{1}{Q(x)}\exp\left(\int \frac{L(x)}{Q(x)}\,\mathrm{d}x\right).$$

Furthermore, the degree-n orthogonal polynomial q_n is given by *Rodrigues' formula*

$$q_n(x) \propto \frac{1}{w(x)}\frac{\mathrm{d}^n}{\mathrm{d}x^n}\big(w(x)Q(x)^n\big).$$

(Naturally, w and the resulting polynomials are only unique up to choices of normalization.) For our purposes, the main importance of the differential properties of orthogonal polynomials is that, as a consequence, the convergence rate of orthogonal polynomial approximations to a given function f is improved when f has a high degree of differentiability; see Theorem 8.24 later in this chapter.

8.4 Roots of Orthogonal Polynomials

The points x at which an orthogonal polynomial $q_n(x) = 0$ are its *roots*, or *zeros*, and enjoy a number of useful properties. To begin with, they can be found as the eigenvalues of a suitable matrix:

Definition 8.15. The *Jacobi matrix* of a measure μ is the infinite, symmetric, tridiagonal matrix

$$J_\infty(\mu) := \begin{bmatrix} \alpha_0 & \sqrt{\beta_1} & 0 & \cdots \\ \sqrt{\beta_1} & \alpha_1 & \sqrt{\beta_2} & \ddots \\ 0 & \sqrt{\beta_2} & \alpha_2 & \ddots \\ \vdots & \ddots & \ddots & \ddots \end{bmatrix}$$

where α_k and β_k are as in Theorem 8.10. The upper-left $n \times n$ minor of $J_\infty(\mu)$ is denoted $J_n(\mu)$.

Theorem 8.16. *Let $p_0, p_1, \dots$ be the orthonormal polynomials for μ. The zeros of p_n are all real, are the eigenvalues of $J_n(\mu)$, and the eigenvector of $J_n(\mu)$ corresponding to the zero at z is*

$$\boldsymbol{p}(z) := \begin{bmatrix} p_0(z) \\ \vdots \\ p_{n-1}(z) \end{bmatrix}.$$

Proof. Let $\boldsymbol{p}(x) := [p_0(x), \dots, p_{n-1}(x)]^\top$ as above. Then the first n recurrence relations for the orthonormal polynomials, as given in Theorem 8.10, can be summarized as

$$x\boldsymbol{p}(x) = J_n(\mu)\boldsymbol{p}(x) + \sqrt{\beta_n}p_n(x)[0, \dots, 0, 1]^\top. \tag{8.13}$$

Now let $x = z$ be any zero of p_n. Note that $\boldsymbol{p}(z) \neq [0, \dots, 0]^\top$, since $\boldsymbol{p}(z)$ has $1/\sqrt{\beta_0}$ as its first component $p_0(z)$. Hence, (8.13) immediately implies that $\boldsymbol{p}(z)$ is an eigenvector of $J_n(\mu)$ with eigenvalue z. Finally, since $J_n(\mu)$ is a symmetric matrix, its eigenvalues (the zeros of p_n) are all real. □

All that can be said about the roots of an arbitrarily polynomial p of degree n is that, by the Fundamental Theorem of Algebra, p has n roots in $\mathbb{C}$ when counted with multiplicity. Since the zeros of orthogonal polynomials are eigenvalues of a symmetric matrix (the Jacobi matrix), these zeros must be real. In fact, though, orthogonal polynomials are guaranteed to have *simple* real roots, and the roots of successive orthogonal polynomials alternate with one another:

Theorem 8.17 (Zeros of orthogonal polynomials). *Let μ be supported in a non-degenerate interval $I \subseteq \mathbb{R}$, and let $\mathcal{Q} = \{q_n \mid n \in \mathcal{N}\}$ be a system of orthogonal polynomials for μ*

1. *For each $n \in \mathcal{N}$, q_n has exactly n distinct real roots $z_1^{(n)}, \dots, z_n^{(n)} \in I$.*
2. *If (a, b) is an open interval of μ-measure zero, then (a, b) contains at most one root of any orthogonal polynomial q_n for μ.*
3. *The zeros $z_i^{(n)}$ of q_n and $z_i^{(n+1)}$ of q_{n+1} alternate:*

$$z_1^{(n+1)} < z_1^{(n)} < z_2^{(n+1)} < \cdots < z_n^{(n+1)} < z_n^{(n)} < z_{n+1}^{(n+1)};$$

hence, whenever $m > n$, between any two zeros of q_n there is a zero of q_m.

Proof. 1. First observe that $\langle q_n, 1 \rangle_{L^2(\mu)} = 0$, and so q_n changes sign in I. Since q_n is continuous, the Intermediate Value Theorem implies that q_n has at least one real root $z_1^{(n)} \in I$. For $n > 1$, there must be another root $z_2^{(n)} \in I$ of q_n distinct from $z_1^{(n)}$, since if q_n were to vanish only at $z_1^{(n)}$, then $\big(x - z_1^{(n)}\big)q_n$ would not change sign in I, which would contradict the orthogonality relation $\langle x - z_1^{(n)}, q_n \rangle_{L^2(\mu)} = 0$. Similarly, if $n > 2$, consider $\big(x - z_1^{(n)}\big)\big(x - z_2^{(n)}\big)q_n$ to deduce the existence of yet a third distinct root $z_3^{(n)} \in I$. This procedure terminates when all the n complex roots of q_n are shown to lie in I.

2. Suppose that (a, b) contains two distinct zeros $z_i^{(n)}$ and $z_j^{(n)}$ of q_n. Let $a_n \neq 0$ denote the coefficient of x^n in $q_n(x)$. Then

$$\begin{aligned} \left\langle q_n, \prod_{k\neq i,j} \big(x - z_k^{(n)}\big) \right\rangle_{L^2(\mu)} &= \int_{\mathbb{R}} q_n(x) \prod_{k\neq i,j} \big(x - z_k^{(n)}\big) \, \mathrm{d}\mu(x) \\ &= a_n \int_{\mathbb{R}} \big(x - z_i^{(n)}\big)\big(x - z_j^{(n)}\big) \prod_{k\neq i,j} \big(x - z_k^{(n)}\big)^2 \, \mathrm{d}\mu(x) \\ &> 0, \end{aligned}$$

since the integrand is positive outside of (a, b). However, this contradicts the orthogonality of q_n to all polynomials of degree less than n.

3. As usual, let p_n be the normalized version of q_n. Let σ, τ be consecutive zeros of p_n, so that $p_n'(\sigma)p_n'(\tau) < 0$. Then Corollary 8.13 implies that p_{n+1} has opposite signs at σ and τ, and so the Intermediate Value Theorem implies that p_{n+1} has at least one zero between σ and τ. This observation accounts for $n - 1$ of the $n + 1$ zeros of p_{n+1}, namely $z_2^{(n+1)} < \cdots < z_n^{(n+1)}$. There are two further zeros of p_{n+1}, one to the left of $z_1^{(n)}$ and one to the right of $z_n^{(n)}$. This follows because $p_n'\big(z_n^{(n)}\big) > 0$, so Corollary 8.13 implies that $p_{n+1}\big(z_n^{(n)}\big) < 0$. Since $p_{n+1}(x) \to +\infty$ as $x \to +\infty$, the Intermediate Value Theorem implies the existence of $z_{n+1}^{(n+1)} > z_n^{(n)}$. A similar argument establishes the existence of $z_1^{(n+1)} < z_1^{(n)}$. □

8.5 Polynomial Interpolation

The existence of a unique polynomial $p(x) = \sum_{i=0}^{n} c_i x^i$ of degree at most n that interpolates the values $y_0, \dots, y_n \in \mathbb{R}$ at $n + 1$ distinct points $x_0, \dots, x_n \in \mathbb{R}$ follows from the invertibility of the *Vandermonde matrix*

$$V_n(x_0, \dots, x_n) := \begin{bmatrix} 1 & x_0 & x_0^2 & \cdots & x_0^n \\ 1 & x_1 & x_1^2 & \cdots & x_1^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^n \end{bmatrix} \in \mathbb{R}^{(n+1)\times(n+1)} \tag{8.14}$$

and hence the unique solvability of the system of simultaneous linear equations

$$V_n(x_0, \dots, x_n) \begin{bmatrix} c_0 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} y_0 \\ \vdots \\ y_n \end{bmatrix}. \tag{8.15}$$

In practice, a polynomial interpolant would never be constructed in this way since, for nearly-coincident nodes, the Vandermonde matrix is notoriously ill-conditioned: the determinant is given by

$$\det(V_n) = \prod_{0\leq i<j\leq n} (x_i - x_j)$$

and, while the condition number of the Vandermonde matrix is hard to calculate exactly, there are dishearteningly large lower bounds such as

$$\kappa_{n,\infty} := \|V_n\|_\infty \|V_n^{-1}\|_\infty \gtrsim 2^{n/2} \tag{8.16}$$

for sets of nodes that are symmetric about the origin, with respect to the matrix (operator) ∞-norm.

However, there is another — and better-conditioned — way to express the polynomial interpolation problem, the so-called *Lagrange form*, which amounts to a clever choice of basis for $\mathbb{R}[x]_{\leq n}$ (instead of the usual monomial basis $\{1, x, x^2, \dots, x^n\}$) so that the matrix in (8.15) in the new basis is the identity matrix.

Definition 8.18. Let $x_0, \dots, x_n \in \mathbb{R}$ be distinct. The associated *nodal polynomial* is defined to be

$$\prod_{j=0}^{n} (x - x_j) \in \mathbb{R}[x]_{\leq n+1}.$$

For $0 \leq j \leq n$, the associated *Lagrange basis polynomial* $\ell_j \in \mathbb{R}[x]_{\leq n}$ is defined by

$$\ell_j(x) := \prod_{\substack{0 \leq k \leq n \\ k \neq j}} \frac{x - x_k}{x_j - x_k}.$$

Given also arbitrary values $y_0, \dots, y_n \in \mathbb{R}$, the associated *Lagrange interpolation polynomial* is

$$L(x) := \sum_{j=0}^{n} y_j \ell_j(x).$$

Theorem 8.19. *Given distinct $x_0, \dots, x_n \in \mathbb{R}$ and any $y_0, \dots, y_n \in \mathbb{R}$, the associated Lagrange interpolation polynomial L is the unique polynomial of degree at most n such that $L(x_k) = y_k$ for $k = 0, \dots, n$.*

Proof. Observe that each Lagrange basis polynomial $\ell_j(x) \in \mathbb{R}[x]_{\leq n}$, and so $L(x) \in \mathbb{R}[x]_{\leq n}$. Observe also that $\ell_j(x_k) = \delta_{jk}$. Hence,

$$L(x_k) = \sum_{j=0}^{n} y_j \ell_j(x_k) = \sum_{j=0}^{n} y_j \delta_{jk} = y_k.$$

For uniqueness, consider the basis $\{\ell_0, \dots, \ell_n\}$ of $\mathbb{R}[x]_{\leq n}$ and suppose that $p = \sum_{j=0}^{n} c_j \ell_j$ is any polynomial that interpolates the values $\{y_k\}_{k=0}^{n}$ at the points $\{x_k\}_{k=0}^{n}$. But then, for each $k = 0, \dots, n$,

$$y_k = \sum_{j=0}^{n} c_j \ell_j(x_k) = \sum_{j=0}^{n} c_j \delta_{jk} = c_k,$$

and so $p = L$, as claimed. □

Runge's Phenomenon. Given the task of choosing nodes $x_k \in [a, b]$ between which to interpolate functions $f : [a, b] \to \mathbb{R}$, it might seem natural to choose the nodes x_k to be equally spaced. Runge's phenomenon [143] shows that this is not always a good choice of interpolation scheme. Consider the function $f : [-1, 1] \to \mathbb{R}$ defined by

$$f(x) := \frac{1}{1 + 25x^2}, \tag{8.17}$$

and let L_n be the degree-n (Lagrange) interpolation polynomial for f on the equally-spaced nodes $x_k := \frac{2k}{n} - 1$. As illustrated in Figure 8.2, L_n oscillates wildly near the endpoints of the interval $[-1, 1]$. Even worse, as n increases, these oscillations do not die down but increase without bound: it can be shown that

$$\lim_{n \to \infty} \sup_{x \in [-1,1]} |f(x) - L_n(x)| = \infty.$$

As a consequence, polynomial interpolation and numerical integration using uniformly spaced nodes — as in the Newton–Cotes formula (Definition 9.5) — can in general be very inaccurate. The oscillations near ± 1 can be controlled by using a non-uniform set of nodes, in particular one that is denser near ± 1 than near 0; the standard example is the set of *Chebyshev nodes* defined by

$$x_k := \cos\left(\frac{2k - 1}{2n}\pi\right),$$

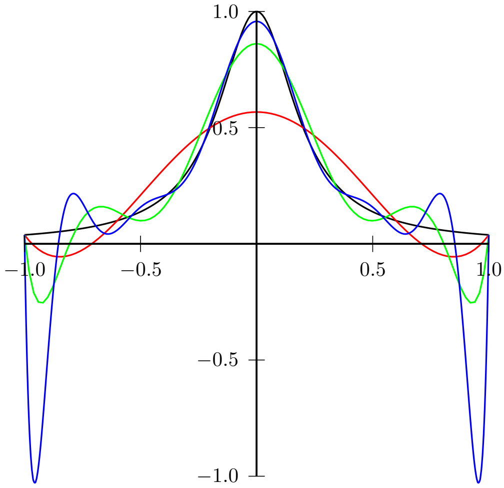


Figure 8.2: Runge's phenomenon. The function $f(x) := (1 + 25x^2)^{-1}$ in black, and polynomial interpolations of degrees 5 (red), 9 (green), and 13 (blue) on evenly-spaced nodes.

i.e. the roots of the Chebyshev polynomials of the first kind T_n, which are orthogonal polynomials for the measure $(1 - x^2)^{-1/2}\,\mathrm{d}x$ on $[-1, 1]$. Indeed, for every *absolutely* continuous function $f\colon [1, 1] \to \mathbb{R}$, the sequence of interpolating polynomials through the Chebyshev nodes converges uniformly to f.

However, Chebyshev nodes are not a panacea. Indeed, for every predefined set of interpolation nodes there is a continuous function for which the interpolation process on those nodes diverges. For every continuous function there is a set of nodes on which the interpolation process converges. In practice, in the absence of guarantees of convergence, one should always perform 'sanity checks' to see if an interpolation scheme has given rise to potentially spurious Runge-type oscillations. One should also check whether or not the interpolant depends sensitively upon the nodal set and data.

Norms of Interpolation Operators. The convergence and optimality of interpolation schemes can be quantified using the norm of the corresponding interpolation operator. From an abstract functional-analytic point of view, interpolation is the result of applying a suitable projection operator Π to a function f in some space $\mathcal{V}$ to yield an interpolating function Πf in some prescribed subspace $\mathcal{U}$ of $\mathcal{V}$. For example, in the above discussion, given $n + 1$ distinct nodes $x_0, \dots, x_n$, the interpolation subspace $\mathcal{U}$ is $\mathbb{R}[x]_{\leq n}$ and the operator Π is

$$\Pi\colon f \mapsto \sum_{i=0}^{n} f(x_i)\ell_i,$$

or, in terms of pointwise evaluation functionals (Dirac measures) δ_a, $\Pi = \sum_{i=1}^{n} \delta_{x_i}\ell_i$. Note, in particular, that Π is a projection operator that acts as the identity function on the interpolation subspace, i.e. the degree-n polynomial interpolation of a polynomial $p(x) \in \mathbb{R}[x]_{\leq n}$ is just p itself. The following general lemma gives an upper bound on the error incurred by any interpolation scheme that can be written as a projection operator:

Lemma 8.20 (Lebesgue's approximation lemma)**.** *Let $(\mathcal{V}, \|\cdot\|)$ be a normed space, $\mathcal{U} \subseteq \mathcal{V}$, and $\Pi\colon \mathcal{V} \to \mathcal{U}$ a linear projection onto $\mathcal{U}$ (i.e. for all $u \in \mathcal{U}$, $\Pi u = u$) with finite operator norm $\|\Pi\|_{\mathrm{op}}$. Then, for all $v \in \mathcal{V}$,*

$$\|v - \Pi v\| \leq (1 + \|\Pi\|_{\mathrm{op}}) \inf_{u \in \mathcal{U}} \|v - u\|. \tag{8.18}$$

Proof. Let $\varepsilon > 0$ be arbitrary, and let $u^* \in \mathcal{U}$ be ε-suboptimal for the infimum on the right-hand side of (8.18), i.e.

$$\|v - u^*\| \leq \varepsilon + \inf_{u \in \mathcal{U}} \|v - u\|. \tag{8.19}$$

Now

$$\begin{aligned}
\|v - \Pi v\| &\leq \|v - u^*\| + \|u^* - \Pi v\| \\
&= \|v - u^*\| + \|\Pi u^* - \Pi v\| && \text{since } \Pi|_{\mathcal{U}} = \mathrm{id}_{\mathcal{U}} \\
&\leq \|v - u^*\| + \|\Pi\|_{\mathrm{op}} \|u^* - v\| && \text{by the definition of } \|\Pi\|_{\mathrm{op}} \\
&= (1 + \|\Pi\|_{\mathrm{op}}) \|v - u^*\| \\
&= (1 + \|\Pi\|_{\mathrm{op}}) \inf_{u \in \mathcal{U}} \|v - u\| + \varepsilon (1 + \|\Pi\|_{\mathrm{op}}) && \text{by (8.19).}
\end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, (8.18) follows. □

Thus, with respect to a given norm $\|\cdot\|$, polynomial interpolation is quasi-optimal up to a constant factor given by the operator norm of the interpolation operator in that norm; in this context, $\|\Pi\|_{\mathrm{op}}$ is often called the *Lebesgue constant* of the interpolation scheme. For the maximum norm, the Lebesgue constant has a convenient expression in terms of the Lagrange basis polynomials; see Exercise 8.10. The next section considers optimal approximation with respect to L^2 norms, which amounts to orthogonal projection.

8.6 Polynomial Approximation

The following theorem on the uniform approximation (on compact sets) of continuous functions by polynomials should hopefully be familiar:

Theorem 8.21 (Weierstrass). *Let $[a, b] \subset \mathbb{R}$ be a bounded interval, let $f : [a, b] \to \mathbb{R}$ be continuous, and let $\varepsilon > 0$. Then there exists a polynomial p such that*

$$\sup_{a \leq x \leq b} |f(x) - p(x)| < \varepsilon.$$

Remark 8.22. Note well that Weierstrass' theorem only ensures uniform approximation of continuous functions on *compact* sets. The reason is simple: since any polynomial of finite degree tends to $\pm\infty$ at the extremes of the real line $\mathbb{R}$, no polynomial can be uniformly close, over all of $\mathbb{R}$, to any non-constant bounded function.

As a consequence of standard results on orthogonal projection in Hilbert spaces, we have the following:

Theorem 8.23. *Let $\mathcal{Q} = \{q_n \mid n \in \mathcal{N}\}$ be a system of orthogonal polynomials for a measure μ on a subinterval $I \subseteq \mathbb{R}$. For any $f \in L^2(I, \mu)$ and any $d \in \mathbb{N}_0$, the orthogonal projection $\Pi_d f$ of f onto $\mathbb{R}[x]_{\leq d}$ is the best degree-d polynomial approximation of f in the $L^2(I, \mu)$ norm, i.e.*

$$\Pi_d f = \operatorname*{arg\,min}_{p(x) \in \mathbb{R}[x]_{\leq d}} \|p - f\|_{L^2(\mu)},$$

where, denoting the orthogonal polynomials for μ by $\{q_k \mid k \geq 0\}$,

$$\Pi_d f := \sum_{k=0}^{d} \frac{\langle f, q_k \rangle_{L^2(\mu)}}{\|q_k\|^2_{L^2(\mu)}} q_k,$$

and the residual is orthogonal to the projection subspace:

$$\langle f - \Pi_d f, p \rangle_{L^2(\mu)} = 0 \quad \text{for all } p(x) \in \mathbb{R}[x]_{\leq d}.$$

An important property of polynomial expansions of functions is that the quality of the approximation (i.e. the rate of convergence) improves as the regularity of the function to be approximated increases. This property is referred to as *spectral convergence* and is easily quantified by using the machinery of Sobolev spaces. Recall that, given $k \in \mathbb{N}_0$ and a measure μ on a subinterval $I \subseteq \mathbb{R}$, the Sobolev inner product and norm are defined by

$$\begin{aligned} \langle u, v \rangle_{H^k(\mu)} &:= \sum_{m=0}^{k} \left\langle \frac{\mathrm{d}^m u}{\mathrm{d}x^m}, \frac{\mathrm{d}^m v}{\mathrm{d}x^m} \right\rangle_{L^2(\mu)} = \sum_{m=0}^{k} \int_I \frac{\mathrm{d}^m u}{\mathrm{d}x^m} \frac{\mathrm{d}^m v}{\mathrm{d}x^m} \, \mathrm{d}\mu \\ \|u\|_{H^k(\mu)} &:= \langle u, v \rangle_{H^k(\mu)}^{1/2}. \end{aligned}$$

The Sobolev space $H^k(\mu)$ consists of all $L^2(\mu)$ functions that have weak derivatives of all orders up to k in $L^2(\mu)$, and is equipped with the above inner product and norm. (As usual, we abuse terminology and confuse functions with their equivalence classes modulo equality μ-almost everywhere.)

Legendre expansions of Sobolev functions on $[-1, 1]$ satisfy the following spectral convergence theorem; the analogous result for Hermite expansions of Sobolev functions on $\mathbb{R}$ is Exercise 8.13, and the general result is Exercise 8.14.

Theorem 8.24 (Spectral convergence of Legendre expansions)**.** *There is a constant $C_k \geq 0$ that may depend upon k but is independent of d and f such that, for all $f \in H^k([-1, 1], \mathrm{d}x)$,*

$$\|f - \Pi_d f\|_{L^2(\mathrm{d}x)} \leq C_k d^{-k} \|f\|_{H^k(\mathrm{d}x)}. \tag{8.20}$$

Proof. As a special case of the Jacobi polynomials (or by Exercise 8.11), the Legendre polynomials satisfy $\mathcal{L}\mathrm{Le}_n = \lambda_n \mathrm{Le}_n$, where the differential operator $\mathcal{L}$ and eigenvalues λ_n are

$$\mathcal{L} = \frac{\mathrm{d}}{\mathrm{d}x}\left((1 - x^2) \frac{\mathrm{d}}{\mathrm{d}x} \right) = (1 - x^2) \frac{\mathrm{d}^2}{\mathrm{d}x^2} - 2x \frac{\mathrm{d}}{\mathrm{d}x}, \quad \lambda_n = -n(n+1).$$

If $f \in H^k([-1, 1], \mathrm{d}x)$, then, by the definition of the Sobolev norm and the operator $\mathcal{L}$, $\|\mathcal{L}f\|_{L^2} \leq C\|f\|_{H^2}$ and, indeed, for any $m \in \mathbb{N}$ such that $2m \leq k$,

$$\|\mathcal{L}^m f\|_{L^2} \leq C\|f\|_{H^{2m}}. \tag{8.21}$$

The key ingredient of the proof is integration by parts:

$$\begin{aligned} \langle f, \mathrm{Le}_n \rangle_{L^2} &= \lambda_n^{-1} \int_{-1}^{1} (\mathcal{L}\mathrm{Le}_n)(x) f(x) \, \mathrm{d}x \\ &= \lambda_n^{-1} \int_{-1}^{1} \big((1 - x^2) \mathrm{Le}_n''(x) f(x) - 2x \mathrm{Le}_n'(x) f(x) \big) \, \mathrm{d}x \\ &= -\lambda_n^{-1} \int_{-1}^{1} \big(((1 - x^2) f)'(x) \mathrm{Le}_n'(x) + 2x \mathrm{Le}_n'(x) f(x) \big) \, \mathrm{d}x && \text{by IBP} \\ &= -\lambda_n^{-1} \int_{-1}^{1} (1 - x^2) f'(x) \mathrm{Le}_n'(x) \, \mathrm{d}x \\ &= \lambda_n^{-1} \int_{-1}^{1} \big((1 - x^2) f' \big)'(x) \mathrm{Le}_n(x) \, \mathrm{d}x && \text{by IBP} \\ &= \lambda_n^{-1} \langle \mathcal{L}f, \mathrm{Le}_n \rangle_{L^2}. \end{aligned}$$

Hence, for all $m \in \mathbb{N}_0$ for which f has $2m$ weak derivatives,

$$\langle f, \mathrm{Le}_n \rangle_{L^2} = \frac{\langle \mathcal{L}^m f, \mathrm{Le}_n \rangle_{L^2}}{\lambda_n^m}. \tag{8.22}$$

Hence,

$$
\begin{aligned}
\|f - \Pi_d f\|_{L^2}^2 &= \sum_{n=d+1}^{\infty} \frac{|\langle f, \mathrm{Le}_n \rangle_{L^2}|^2}{\|\mathrm{Le}_n\|_{L^2}^2} \\
&= \sum_{n=d+1}^{\infty} \frac{|\langle \mathcal{L}^m f, \mathrm{Le}_n \rangle_{L^2}|^2}{\lambda_n^{2m} \|\mathrm{Le}_n\|_{L^2}^2} && \text{by (8.22)} \\
&\leq \frac{1}{\lambda_d^{2m}} \sum_{n=d+1}^{\infty} \frac{|\langle \mathcal{L}^m f, \mathrm{Le}_n \rangle_{L^2}|^2}{\|\mathrm{Le}_n\|_{L^2}^2} \\
&\leq \frac{1}{\lambda_d^{2m}} \sum_{n=0}^{\infty} \frac{|\langle \mathcal{L}^m f, \mathrm{Le}_n \rangle_{L^2}|^2}{\|\mathrm{Le}_n\|_{L^2}^2} \\
&= \frac{1}{\lambda_d^{2m}} \|\mathcal{L}^m f\|_{L^2}^2 && \text{by Parseval (Theorem 3.19)} \\
&\leq C^2 d^{-4m} \|f\|_{H^{2m}}^2 && \text{by (8.21)}
\end{aligned}
$$

since $|\lambda_d| \geq d^2$. Setting $k = 2m$ and taking square roots yields (8.20). □

Gibbs' Phenomenon. However, in the other direction, poor regularity can completely ruin the nice convergence of spectral expansions. The classic example of this is *Gibbs' phenomenon*, in which one tries to approximate the sign function

$$
\operatorname{sgn}(x) := \begin{cases} -1, & \text{if } x < 0, \\ 0, & \text{if } x = 0, \\ 1, & \text{if } x > 0, \end{cases}
$$

on $[-1, 1]$ by its expansion with respect to a system of orthogonal polynomials such as the Legendre polynomials $\mathrm{Le}_n(x)$ or the Fourier polynomials e^{inx}. The degree-$(2N + 1)$ Legendre expansion of the sign function is

$$
(\Pi_{2N+1} \operatorname{sgn})(x) = \sum_{n=0}^{N} \frac{(-1)^n (4n + 3)(2n)!}{2^{2n+1}(n + 1)! n!} \mathrm{Le}_{2n+1}(x). \tag{8.23}
$$

See Figure 8.3 for an illustration. Although $\Pi_{2N+1} \operatorname{sgn} \to \operatorname{sgn}$ as $N \to \infty$ in the L^2 sense, there is no hope of uniform convergence: the oscillations at the discontinuity at 0, and indeed at the endpoints ± 1, do not decay to zero as $N \to \infty$. The inability of globally smooth basis functions such as Legendre polynomials to accurately resolve discontinuities naturally leads to the consideration of non-smooth basis functions such as wavelets.

Remark 8.22 Revisited. To repeat, even though smoothness of f improves the rate of convergence of the orthogonal polynomial expansion $\Pi_d f \to f$ as $d \to \infty$ in the L^2 sense, the uniform convergence and pointwise predictive value of an orthogonal polynomial expansion $\Pi_d f$ are almost certain to be poor on unbounded (non-compact) domains, and no amount of smoothness of f can rectify this problem.

8.7 Multivariate Orthogonal Polynomials

For working with polynomials in d variables, we will use standard multi-index notation. Multi-indices will be denoted by Greek letters $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}_0^d$. For $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ and $\alpha \in \mathbb{N}_0^d$, the monomial x^α is defined by

$$
x^\alpha := x_1^{\alpha_1} x_2^{\alpha_2} \ldots x_d^{\alpha_d},
$$

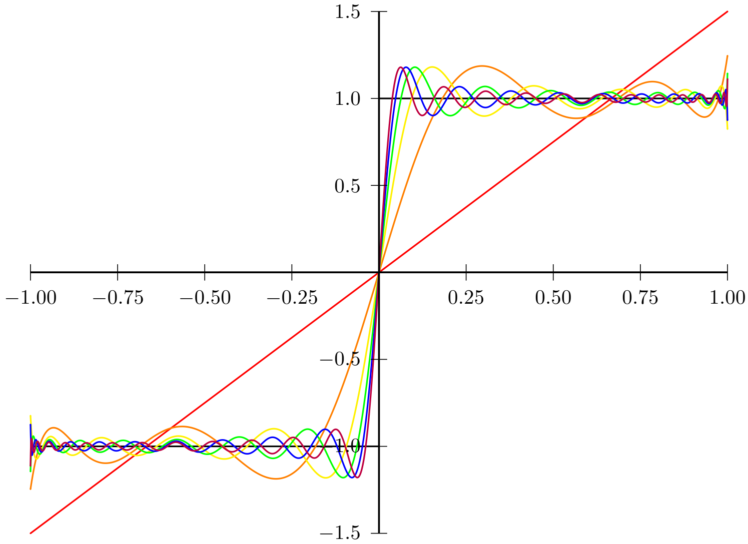


Figure 8.3: Legendre expansions of the sign function on $[-1, 1]$ exhibit Gibbsian oscillations at 0 and at ± 1. The sign function is shown in black, as are Legendre expansions (8.23) to degree $2N - 1$ for $N = 1$ (red), 5 (orange), 10 (yellow), 15 (green), 20 (blue), and 25 (purple).

and $|\alpha| := \alpha_1 + \cdots + \alpha_d$ is called the *total degree* of x^α. The total degree of a polynomial p (i.e. a finite linear combination of such monomials) is denoted $\deg(p)$ and is the maximum of the total degrees of the summands. The space of all polynomials in $x_1, \dots, x_d$ is denoted $\mathbb{R}[x_1, \dots, x_d]$, while the subset consisting of those multivariate polynomials of total degree at most k is denoted $\mathbb{R}[x_1, \dots, x_d]_{\leq k}$.

Given a measure μ on $\mathbb{R}^d$, it is tempting to apply the Gram–Schmidt process with respect to the inner product

$$\langle f, g \rangle_{L^2(\mu)} := \int_{\mathbb{R}^d} f(x) g(x) \, \mathrm{d}\mu(x)$$

to the monomials $\{x^\alpha \mid \alpha \in \mathbb{N}_0^d\}$ to obtain a system of orthogonal polynomials for the measure μ. However, there is an immediate problem, in that orthogonal polynomials of several variables are not unique. In order to apply the Gram–Schmidt process, we need to give a linear order to multi-indices $\alpha \in \mathbb{N}_0^d$. There are many choices of well-defined total order (for example, the lexicographic order or the graded lexicographic order); but there is no natural choice and different orders will give different sequences of orthogonal polynomials. Instead of fixing such a total order, we relax Definition 8.1 slightly:

Definition 8.25. Let μ be a non-negative measure on $\mathbb{R}^d$. A family of polynomials $\mathcal{Q} = \{q_\alpha \mid \alpha \in \mathbb{N}_0^d\}$ is called a *weakly orthogonal system of polynomials* if q_α is such that

$$\langle q_\alpha, p \rangle_{L^2(\mu)} = 0 \quad \text{for all } p(x) \in \mathbb{R}[x_1, \dots, x_d] \text{ with } \deg(p) < |\alpha|.$$

The system $\mathcal{Q}$ is called a *strongly orthogonal system of polynomials* if

$$\langle q_\alpha, q_\beta \rangle_{L^2(\mu)} = 0 \iff \alpha \neq \beta.$$

Hence, in the many-variables case, an orthogonal polynomial of total degree n, while it is required to be orthogonal to all polynomials of strictly lower total degree, may be non-orthogonal to other polynomials of the same total degree n. However, the meaning of orthonormality is unchanged: a system of polynomials $\{p_\alpha \mid \alpha \in \mathbb{N}_0^d\}$ is *orthonormal* if

$$\langle p_\alpha, p_\beta \rangle_{L^2(\mu)} = \delta_{\alpha\beta}.$$

While the computation of orthogonal polynomials of many variables is, in general, a difficult task, it is substantially simpler if the measure is a product measure: multivariate orthogonal polynomials can be obtained as products of univariate orthogonal polynomials.

Theorem 8.26. *Suppose that* $\mu = \bigotimes_{i=1}^{d} \mu_i$ *is a product measure on* $\mathbb{R}^d$ *and that, for each* $i = 1, \dots, d$, $\mathcal{Q}^{(i)} = \{q_{\alpha_i}^{(i)} \mid \alpha_i \in \mathcal{N}_i\}$ *is a system of orthogonal polynomials for the marginal measure* μ_i *on* $\mathbb{R}$*. Then*

$$\mathcal{Q} = \bigotimes_{i=1}^{d} \mathcal{Q}^{(i)} = \left\{ q_\alpha := \prod_{i=1}^{d} q_{\alpha_i}^{(i)} \,\middle|\, \alpha \in \mathcal{N}_1 \times \cdots \times \mathcal{N}_d \right\}$$

is a strongly orthogonal system of polynomials for μ *in which* $\deg(q_\alpha) = |\alpha|$.

Proof. It is clear that q_α, as defined above, has total degree $|\alpha|$. Let q_α and q_β be distinct polynomials in the proposed orthogonal system $\mathcal{Q}$. Since $\alpha \neq \beta$, it follows that α and β differ in at least one component, so suppose without loss of generality that $\alpha_1 \neq \beta_1$. By Fubini's theorem,

$$\langle q_\alpha, q_\beta \rangle_{L^2(\mu)} = \int_{\mathbb{R}^d} q_\alpha q_\beta \,\mathrm{d}\mu = \int_{\mathbb{R}^{d-1}} \prod_{j=2}^{d} q_{\alpha_j}^{(j)} q_{\beta_j}^{(j)} \left[\int_{\mathbb{R}} q_{\alpha_1}^{(1)} q_{\beta_1}^{(1)} \,\mathrm{d}\mu_1 \right] \mathrm{d}\mu_2 \otimes \cdots \otimes \mu_d.$$

But, since $\mathcal{Q}^{(1)}$ is a system of orthogonal univariate polynomials for μ_1, and since $\alpha_1 \neq \beta_1$,

$$\int_{\mathbb{R}} q_{\alpha_1}^{(1)}(x_1) q_{\beta_1}^{(1)}(x_1) \,\mathrm{d}\mu_1(x_1) = 0.$$

Hence, $\langle q_\alpha, q_\beta \rangle_{L^2(\mu)} = 0$.

On the other hand, for each polynomial $q_\alpha \in \mathcal{Q}$,

$$\|q_\alpha\|_{L^2(\mu)}^2 = \big\|q_{\alpha_1}^{(1)}\big\|_{L^2(\mu_1)}^2 \big\|q_{\alpha_2}^{(2)}\big\|_{L^2(\mu_2)}^2 \cdots \big\|q_{\alpha_d}^{(d)}\big\|_{L^2(\mu_d)}^2,$$

which is strictly positive by the assumption that each $\mathcal{Q}^{(i)}$ is a system of orthogonal univariate polynomials for μ_i.

Hence, $\langle q_\alpha, q_\beta \rangle_{L^2(\mu)} = 0$ if and only if α and β are distinct, so $\mathcal{Q}$ is a system of strongly orthogonal polynomials for μ. ☐

Bibliography

W At Warwick, MA228 Numerical Analysis covers some aspects of the theory of orthogonal polynomials, including polynomial approximation and interpolation.

Many important properties of orthogonal polynomials, and standard examples, are given in reference form in Chapter 22 of Abramowitz & Stegun [1] and in Chapter 18 of the NIST Handbook [122, 129]. The classic detailed reference on orthogonal polynomial theory is the 1939 monograph of Szegő [163]. The book of Chihara [33] is also well regarded. An excellent more modern reference is the book of Gautschi [61]; some topics covered in that book that are not treated here include orthogonal polynomials on the semicircle, Sobolev orthogonal polynomials, computational methods, and applications to spline approximation and slowly convergent series. Note that Gautschi uses the physicists' e^{-x^2} normalization for Hermite polynomials, not the probabilists' $e^{-x^2/2}$ normalization that is used in these notes. See Xu's lecture notes [187] for further treatment of multivariate orthogonal polynomial theory. Theorem 8.11 is ostensibly due to Favard [52], but the result was already known to authors such as Stieltjes working in the theory of continued fractions.

Interpolation and approximation theory are discussed thoroughly in Chapter 2 of Gautschi [62], and Chapter 6 of Kincaid & Cheney [91]. The lower bound (8.16) for the condition number of the Vandermonde matrix for symmetric nodes is due to Gautschi & Inglese [63]. The Gibbs phenomenon and methods for its resolution are discussed at length in the article of Gottlieb & Shu [70]. Wavelet bases are discussed in Chapter 8 of the book of Le Maître & Knio [102] and in articles of Le Maître & al. [103, 104, 105].

Exercises

Exercise 8.1. Prove that the $L^2(\mathbb{R}, \mu)$ inner product is positive definite on the space $\mathbb{R}[x]$ of all polynomials if all polynomials are μ-integrable and the measure μ has infinite support.

Exercise 8.2. Prove Theorem 8.7. That is, show that if μ has finite moments only of degrees $0, 1, \dots, r$, then μ admits only a finite system of orthogonal polynomials $q_0, \dots, q_d$, where $d = \min\{\lfloor r/2 \rfloor, \# \operatorname{supp}(\mu) - 1\}$.

Exercise 8.3. Define a Borel measure, the *Cauchy–Lorentz distribution*, μ on $\mathbb{R}$ by

$$\frac{\mathrm{d}\mu}{\mathrm{d}x}(x) = \frac{1}{\pi} \frac{1}{1 + x^2}.$$

Show that μ is a probability measure, that $\dim L^2(\mathbb{R}, \mu; \mathbb{R}) = \infty$, find all orthogonal polynomials for μ, and explain your results.

Exercise 8.4. Following the example of the Cauchy–Lorentz distribution, given $\ell \in [0, \infty)$, construct an explicit example of a probability measure $\mu \in \mathcal{M}_1(\mathbb{R})$ with moments of orders up to k but no higher, i.e. such that

$$\int_{\mathbb{R}} x^k \, \mathrm{d}\mu(x) \in \mathbb{R} \text{ for } 0 \leq k \leq \ell,$$

and

$$\int_{\mathbb{R}} x^k \, \mathrm{d}\mu(x) \text{ is infinite or does not exist for } k > \ell.$$

Exercise 8.5. Calculate orthogonal polynomials for the *generalized Maxwell distribution* $\mathrm{d}\mu(x) = x^\alpha e^{-x^2} \, \mathrm{d}x$ on the half-line $[0, \infty)$, where $\alpha > -1$ is a constant. The case $\alpha = 2$ is known as the *Maxwell distribution* and the case $\alpha = 0$ as the *half-range Hermite distribution*.

Exercise 8.6. Prove Theorem 8.8.

Exercise 8.7. Let μ be the probability distribution of $Y := e^X$, where $X \sim \mathcal{N}(0, 1)$ is a standard normal random variable, i.e. let μ be the standard *log-normal distribution*. The following exercise shows that the system $\mathcal{Q} = \{q_k \mid k \in \mathbb{N}_0\}$ of orthogonal polynomials for μ is not a complete orthogonal basis for $L^2((0, \infty), \mu; \mathbb{R})$.

1. Show that μ has the Lebesgue density function $\rho \colon \mathbb{R} \to \mathbb{R}$ given by

$$\rho(y) := \mathbb{1}[y > 0] \frac{1}{y\sqrt{2\pi}} \exp\left(-\frac{1}{2}(\log y)^2\right).$$

2. Let $f \in L^1(\mathbb{R}, \mu; \mathbb{R})$ be odd and 1-periodic, i.e. $f(x) = -f(-x) = f(x + 1)$ for all $x \in \mathbb{R}$. Show that, for all $k \in \mathbb{N}_0$,

$$\int_0^\infty y^k f(\log y) \, \mathrm{d}\mu(y) = 0.$$

3. Let $g := f \circ \log$ and suppose that $g \in L^2((0, \infty), \mu; \mathbb{R})$. Show that the expansion of g in the orthogonal polynomials $\{q_k \mid k \in \mathbb{N}_0\}$ has all coefficients equal to zero, and thus that this expansion does not converge to g when $g \neq 0$.

Exercise 8.8. Complete the proof of Theorem 8.10 by deriving the formula for β_n.

Exercise 8.9. Calculate the orthogonal polynomials of Table 8.2 by hand for degree at most 5, and write a numerical program to compute them for higher degree.

Exercise 8.10. Let $x_0, \dots, x_n$ be distinct nodes in an interval I and let $\ell_i(x) \in \mathbb{R}[x]_{\leq n}$ be the associated Lagrange basis polynomials. Define $\lambda \colon I \to \mathbb{R}$ by

$$\lambda(x) := \sum_{i=0}^n |\ell_i(x)|.$$

Show that, with respect to the supremum norm $\|f\|_\infty := \sup_{x\in I} |f(x)|$, the polynomial interpolation operator Π from $\mathcal{C}^0(I;\mathbb{R})$ to $\mathbb{R}[x]_{\leq n}$ has operator norm given by

$$\|\Pi\|_{\mathrm{op}} = \|\lambda\|_\infty,$$

i.e. the Lebesgue constant of the interpolation scheme is the supremum norm of the sum of absolute values of the Lagrange basis polynomials.

Exercise 8.11. Using the three-term recurrence relation $(n+1)\mathrm{Le}_{n+1}(x) = (2n+1)x\mathrm{Le}_n(x) - n\mathrm{Le}_{n-1}(x)$, prove by induction that, for all $n \in \mathbb{N}_0$,

$$\frac{\mathrm{d}}{\mathrm{d}x}\mathrm{Le}_n(x) = \frac{n}{x^2-1}(x\mathrm{Le}_n(x) - \mathrm{Le}_{n-1}(x)),$$

and

$$\frac{\mathrm{d}}{\mathrm{d}x}\left((1-x^2)\frac{\mathrm{d}}{\mathrm{d}x}\right)\mathrm{Le}_n(x) = -n(n+1)\mathrm{Le}_n(x).$$

Exercise 8.12. Let $\gamma = \mathcal{N}(0,1)$ be standard Gaussian measure on $\mathbb{R}$. Establish the integration-by-parts formula

$$\int_{\mathbb{R}} f(x)g'(x)\,\mathrm{d}\gamma(x) = -\int_{\mathbb{R}} (f'(x) - xf(x))g(x)\,\mathrm{d}\gamma(x).$$

Using the three-term recurrence relation $\mathrm{He}_{n+1}(x) = x\mathrm{He}_n(x) - n\mathrm{He}_{n-1}(x)$, prove by induction that, for all $n \in \mathbb{N}_0$,

$$\frac{\mathrm{d}}{\mathrm{d}x}\mathrm{He}_n(x) = n\mathrm{He}_{n-1}(x),$$

and

$$\left(\frac{\mathrm{d}^2}{\mathrm{d}x^2} - x\frac{\mathrm{d}}{\mathrm{d}x}\right)\mathrm{He}_n(x) = -n\mathrm{He}_n(x).$$

Exercise 8.13 (Spectral convergence of Hermite expansions). Let $\gamma = \mathcal{N}(0,1)$ be standard Gaussian measure on $\mathbb{R}$. Use Exercise 8.12 to mimic the proof of Theorem 8.24 to show that there is a constant $C_k \geq 0$ that may depend upon k but is independent of d and f such that, for all $f \in H^k(\mathbb{R},\gamma)$, f and its degree d expansion in the Hermite orthogonal basis of $L^2(\mathbb{R},\gamma)$ satisfy

$$\|f - \Pi_d f\|_{L^2(\gamma)} \leq C_k d^{-k/2}\|f\|_{H^k(\gamma)}.$$

Exercise 8.14 (Spectral convergence for classical orthogonal polynomial expansions). Let $\mathcal{Q} = \{q_n \mid n \in \mathbb{N}_0\}$ be orthogonal polynomials for an absolutely continuous measure $\mathrm{d}\mu = w(x)\,\mathrm{d}x$ on $\mathbb{R}$, where the weight function w is proportional to $\frac{1}{Q(x)}\exp\bigl(\int \frac{L(x)}{Q(x)}\,\mathrm{d}x\bigr)$ with L linear and Q quadratic, which are eigenfunctions for the differential operator $\mathcal{L} = Q(x)\frac{\mathrm{d}^2}{\mathrm{d}x^2} + L(x)\frac{\mathrm{d}}{\mathrm{d}x}$ with eigenvalues $\lambda_n = n\bigl(\frac{n-1}{2}Q'' + L'\bigr)$.

1. Show that μ has an integration-by-parts formula of the following form: for all smooth functions f and g with compact support in the interior of $\mathrm{supp}(\mu)$,

   $$\int_{\mathbb{R}} f(x)g'(x)\,\mathrm{d}\mu(x) = -\int_{\mathbb{R}} (Tf)(x)g(x)\,\mathrm{d}\mu(x),$$

   where

   $$(Tf)(x) = f'(x) + f(x)\frac{L(x) - Q'(x)}{Q(x)}.$$

2. Hence show that, for smooth enough f, $\mathcal{L}f = T^2(Qf) - T(Lf)$.
3. Hence show that, whenever f has $2m$ derivatives,

   $$\langle f, q_n\rangle_{L^2(\mu)} = \frac{\langle \mathcal{L}^m f, q_n\rangle_{L^2(\mu)}}{\lambda_n^m}$$

   Show also that $\mathcal{L}$ is a symmetric and negative semi-definite operator (i.e. $\langle \mathcal{L}f, g\rangle_{L^2(\mu)} = \langle f, \mathcal{L}g\rangle_{L^2(\mu)}$ and $\langle \mathcal{L}f, f\rangle_{L^2(\mu)} \leq 0$), so that $(-\mathcal{L})$ has a square root $(-\mathcal{L})^{1/2}$, and $\mathcal{L}$ has a square root $\mathcal{L}^{1/2} = i(-\mathcal{L})^{1/2}$.

4. Conclude that there is a constant $C_k \geq 0$ that may depend upon k but is independent of d and f such that $f \colon \mathbb{R} \to \mathbb{R}$ and its degree d expansion $\Pi_d f$ in the basis $\mathcal{Q}$ of $L^2(\mathbb{R}, \mu)$ satisfy

$$\|f - \Pi_d f\|_{L^2(\mu)} \leq C_k |\lambda_d|^{-k/2} \big\|\mathcal{L}^{k/2} f\big\|_{L^2(\mu)}.$$

Polynomial	Parameters	Support	Distribution, μ: Name	Distribution, μ: Density, $w(x)$	Normalization i.e. $\|q_n\|^2_{L^2(\mu)}$
Continuous					
Chebyshev (first)		$[-1,1]$		$(1-x^2)^{-1/2}$	$\pi - \frac{\pi}{2}\delta_{0n}$
Chebyshev (second)		$[-1,1]$	Wigner semicircular	$(1-x^2)^{1/2}$	$\frac{\pi}{2}$
Hermite (phys.)		$\mathbb{R}$	Gaussian	e^{-x^2}	$\sqrt{\pi}2^n n!$
Hermite (prob.)		$\mathbb{R}$	Gaussian	$(2\pi)^{-1/2}e^{-x^2/2}$	$n!$
Jacobi	$\alpha, \beta > 0$	$[-1,1]$	beta	$(1-x)^\alpha(1+x)^\beta$	$\frac{2^{\alpha+\beta+1}\Gamma(n+\alpha+1)\Gamma(n+\beta+1)}{n!(2n+\alpha+\beta+1)\Gamma(n+\alpha+\beta+1)}$
Laguerre	$\alpha > -1$	$[0,\infty)$	gamma	$x^\alpha e^{-x}$	$\frac{\Gamma(n+\alpha+1)}{n!}$
Legendre		$[-1,1]$	uniform	1	$\frac{2}{2n+1}$
Discrete					
Charlier	$\alpha > 0$	$\mathbb{N}_0$	Poisson	$e^{-\alpha}\alpha^x/x!$	$\alpha^{-n}n!$
Hahn	$\alpha, \beta > -1$ or $< -N$	$\{0,\dots,N\}$	hypergeometric	$\frac{(\alpha+1)_x(\beta+1)_{N-x}}{x!(N-x)!}$	$\frac{(-1)^n(n+\alpha+\beta+1)_{N+1}(\beta+1)_n n!}{(2n+\alpha+\beta+1)(\alpha+1)_n(-N)_n N!}$
Kravchuk	$p \in (0,1)$	$\{0,\dots,N\}$	binomial	$\binom{n}{x}p^x(1-p)^x$	$\left(\frac{1-p}{p}\right)^n/\binom{N}{n}$
Meixner	$\beta > 0, c \in (0,1)$	$\mathbb{N}_0$	negative binomial	$(\beta)_x c^x/x!$	$\frac{c^{-n}n!}{(\beta)_n(1-c)^\beta}$

Table 8.2: Summary of some commonly-used orthogonal polynomials, their associated probability distributions, and key properties. Here, $(x)_n := x(x+1)\dots(x+n-1)$ denotes the *rising factorial* or *Pochhammer symbol*.

Polynomial	Differential Equation (8.12)			Recurrence Coefficients		
	$Q(x)$	$L(x)$	λ_n	A_n	B_n	C_n
Continuous				Recurrence (8.2)		
Chebyshev (first)	$1-x^2$	$-x$	$-n^2$	$2-\delta_{0n}$	0	1
Chebyshev (second)	$1-x^2$	$-3x$	$-n(n+2)$	2	0	1
Hermite (phys.)	1	$-2x$	$-2n$	2	0	$2n$
Hermite (prob.)	1	$-x$	$-n$	1	0	n
Jacobi	$1-x^2$	$\beta-\alpha-(\alpha+\beta+2)x$	$-n(n+\alpha+\beta+1)$	See (8.4)		
Laguerre	x	$1+\alpha-x$	$-n$	$-\frac{1}{n+1}$	$\frac{2n+\alpha+1}{n+1}$	$\frac{n+\alpha}{n+1}$
Legendre	$1-x^2$	$-2x$	$-n(n+1)$	$\frac{2n+1}{n+1}$	0	$\frac{n}{n+1}$
Discrete				Recurrence (8.3)		
Charlier				α		n
Hahn				$\frac{(n+\alpha+\beta+1)(n+\alpha+1)(N-n)}{(2n+\alpha+\beta+1)(2n+\alpha+\beta+2)}$		$\frac{n(n+\alpha+\beta+1)(n+\beta)}{(2n+\alpha+\beta)(2n+\alpha+\beta+1)}$
Kravchuk				$(N-n)p$		$(1-p)n$
Meixner				$\frac{(n+\beta)c}{1-c}$		$\frac{n}{1-c}$

Table 8.3: Continuation of Table 8.2.

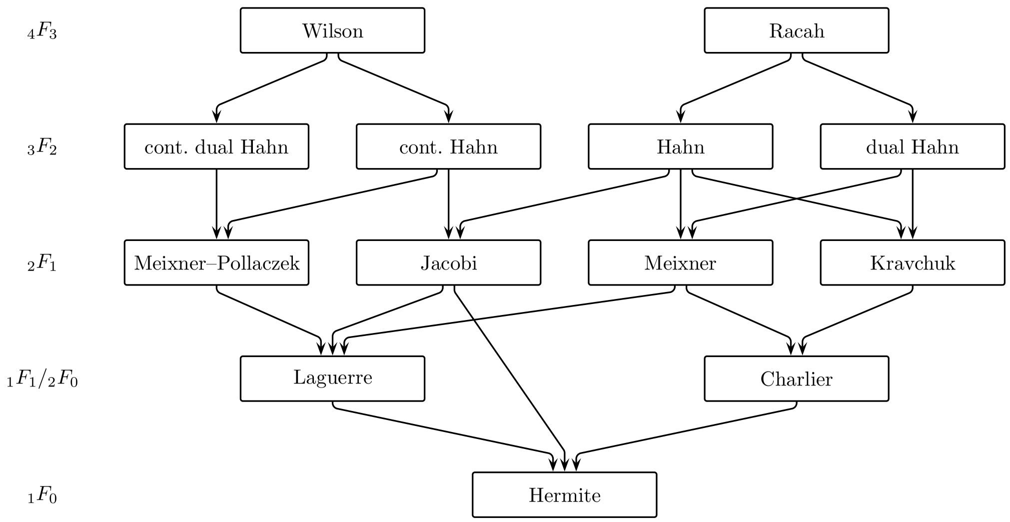


Figure 8.4: The hierarchy of hypergeometric orthogonal polynomials in the Askey scheme. The number of free real parameters is zero for the Hermite polynomials, and increases by one for each row ascended in the scheme, culminating with four free real parameters for the Wilson and Racah polynomials, with the convention that the real and imaginary parts of the parameters are counted separately in the case of the continuous Hahn polynomials. Each arrow indicates a limit relation in which the polynomials at the head of the arrow are obtained as a parametric limit of polynomials at the tail of the arrow: for example, the Laguerre and Jacobi polynomials satisfy $L_n^{(\alpha)}(x) = \lim_{\beta\to\infty} P_n^{(\alpha,\beta)}\big(1 - \frac{2x}{\beta}\big)$.