Encyclopedia of Mathematics and Its Applications 181

ORTHOGONAL POLYNOMIALS IN THE SPECTRAL ANALYSIS OF MARKOV PROCESSES

Birth–Death Models and Diffusion

Manuel Domínguez de la Iglesia

ORTHOGONAL POLYNOMIALS IN THE SPECTRAL ANALYSIS OF MARKOV PROCESSES

In pioneering work in the 1950s, S. Karlin and J. McGregor showed that the probabilistic aspects of certain Markov processes can be studied by analyzing the orthogonal eigenfunctions of associated operators. In the decades since, many authors have extended and deepened this surprising connection between orthogonal polynomials and stochastic processes.

This book gives a comprehensive analysis of the spectral representation of the most important one-dimensional Markov processes, namely discrete-time birth–death chains, birth–death processes and diffusion processes, and brings together all the main results from the extensive literature on the topic with detailed examples and applications. It also features an introduction to the basic theory of orthogonal polynomials and has a selection of exercises at the end of each chapter. The book is suitable for graduate students with a solid background in stochastic processes as well as researchers in orthogonal polynomials and special functions who want to learn about applications of their work to probability.

Encyclopedia of Mathematics and Its Applications

This series is devoted to significant topics or themes that have wide application in mathematics or mathematical science and for which a detailed development of the abstract theory is less important than a thorough and concrete exploration of the implications and applications.

Books in the **Encyclopedia of Mathematics and Its Applications** cover their subjects comprehensively. Less important results may be summarized as exercises at the ends of chapters. For technicalities, readers can be referred to the bibliography, which is expected to be comprehensive. As a result, volumes are encyclopedic references or manageable guides to major subjects.

All the titles listed below can be obtained from good booksellers or from Cambridge University Press.
For a complete series listing visit www.cambridge.org/mathematics.

131 H. Morimoto *Stochastic Control and Mathematical Modeling*
132 G. Schmidt *Relational Mathematics*
133 P. Kornerup and D. W. Matula *Finite Precision Number Systems and Arithmetic*
134 Y. Crama and P. L. Hammer (eds.) *Boolean Models and Methods in Mathematics, Computer Science, and Engineering*
135 V. Berthé and M. Rigo (eds.) *Combinatorics, Automata and Number Theory*
136 A. Kristály, V. D. Rădulescu and C. Varga *Variational Principles in Mathematical Physics, Geometry, and Economics*
137 J. Berstel and C. Reutenauer *Noncommutative Rational Series with Applications*
138 B. Courcelle and J. Engelfriet *Graph Structure and Monadic Second-Order Logic*
139 M. Fiedler *Matrices and Graphs in Geometry*
140 N. Vakil *Real Analysis through Modern Infinitesimals*
141 R. B. Paris *Hadamard Expansions and Hyperasymptotic Evaluation*
142 Y. Crama and P. L. Hammer *Boolean Functions*
143 A. Arapostathis, V. S. Borkar and M. K. Ghosh *Ergodic Control of Diffusion Processes*
144 N. Caspard, B. Leclerc and B. Monjardet *Finite Ordered Sets*
145 D. Z. Arov and H. Dym *Bitangential Direct and Inverse Problems for Systems of Integral and Differential Equations*
146 G. Dassios *Ellipsoidal Harmonics*
147 L. W. Beineke and R. J. Wilson (eds.) with O. R. Oellermann *Topics in Structural Graph Theory*
148 L. Berlyand, A. G. Kolpakov and A. Novikov *Introduction to the Network Approximation Method for Materials Modeling*
149 M. Baake and U. Grimm *Aperiodic Order I: A Mathematical Invitation*
150 J. Borwein et al. *Lattice Sums Then and Now*
151 R. Schneider *Convex Bodies: The Brunn–Minkowski Theory (Second Edition)*
152 G. Da Prato and J. Zabczyk *Stochastic Equations in Infinite Dimensions (Second Edition)*
153 D. Hofmann, G. J. Seal and W. Tholen (eds.) *Monoidal Topology*
154 M. Cabrera García and Á. Rodríguez Palacios *Non-Associative Normed Algebras I: The Vidav–Palmer and Gelfand–Naimark Theorems*
155 C. F. Dunkl and Y. Xu *Orthogonal Polynomials of Several Variables (Second Edition)*
156 L. W. Beineke and R. J. Wilson (eds.) with B. Toft *Topics in Chromatic Graph Theory*
157 T. Mora *Solving Polynomial Equation Systems III: Algebraic Solving*
158 T. Mora *Solving Polynomial Equation Systems IV: Buchberger Theory and Beyond*
159 V. Berthé and M. Rigo (eds.) *Combinatorics, Words and Symbolic Dynamics*
160 B. Rubin *Introduction to Radon Transforms: With Elements of Fractional Calculus and Harmonic Analysis*
161 M. Ghergu and S. D. Taliaferro *Isolated Singularities in Partial Differential Inequalities*
162 G. Molica Bisci, V. D. Radulescu and R. Servadei *Variational Methods for Nonlocal Fractional Problems*
163 S. Wagon *The Banach–Tarski Paradox (Second Edition)*
164 K. Broughan *Equivalents of the Riemann Hypothesis I: Arithmetic Equivalents*
165 K. Broughan *Equivalents of the Riemann Hypothesis II: Analytic Equivalents*
166 M. Baake and U. Grimm (eds.) *Aperiodic Order II: Crystallography and Almost Periodicity*
167 M. Cabrera García and Á. Rodríguez Palacios *Non-Associative Normed Algebras II: Representation Theory and the Zel'manov Approach*
168 A. Yu. Khrennikov, S. V. Kozyrev and W. A. Zúñiga-Galindo *Ultrametric Pseudodifferential Equations and Applications*
169 S. R. Finch *Mathematical Constants II*
170 J. Krajíček *Proof Complexity*
171 D. Bulacu, S. Caenepeel, F. Panaite and F. Van Oystaeyen *Quasi-Hopf Algebras*
172 P. McMullen *Geometric Regular Polytopes*
173 M. Aguiar and S. Mahajan *Bimonoids for Hyperplane Arrangements*
174 M. Barski and J. Zabczyk *Mathematics of the Bond Market: A Lévy Processes Approach*
175 T. R. Bielecki, J. Jakubowski and M. Niewęgłowski *Structured Dependence between Stochastic Processes*
176 A. A. Borovkov, V. V. Ulyanov and Mikhail Zhitlukhin *Asymptotic Analysis of Random Walks: Light-Tailed Distributions*
177 Y.-K. Chan *Foundations of Constructive Probability Theory*
178 L. W. Beineke, M. C. Golumbic and R. J. Wilson (eds.) *Topics in Algorithmic Graph Theory*
179 H.-L. Gau and P. Y. Wu *Numerical Ranges of Hilbert Space Operators*
180 P. A. Martin *Time-Domain Scattering*

ENCYCLOPEDIA OF MATHEMATICS AND ITS APPLICATIONS

Orthogonal Polynomials in the Spectral Analysis of Markov Processes

Birth–Death Models and Diffusion

MANUEL DOMÍNGUEZ DE LA IGLESIA
Universidad Nacional Autónoma de México

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi – 110025, India

103 Penang Road, #05–06/07, Visioncrest Commercial, Singapore 238467

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781316516553
DOI: 10.1017/9781009030540

First published 2022

A catalogue record for this publication is available from the British Library.

Library of Congress Cataloging-in-Publication Data
Names: Iglesia, Manuel Domínguez de la, author.
Title: Orthogonal polynomials in the spectral analysis of Markov processes : birth-death models and diffusion / Manuel Domínguez de la Iglesia.
Description: Cambridge ; New York, NY : Cambridge University Press, 2022. |
Series: Encyclopedia of mathematics and its applications |
Includes bibliographical references and index.
Identifiers: LCCN 2021029856 (print) | LCCN 2021029857 (ebook) |
ISBN 9781316516553 (hardback) | ISBN 9781009030540 (epub)
Subjects: LCSH: Birth and death processes (Stochastic processes) | Orthogonal polynomials. |
Spectral theory (Mathematics) | BISAC: MATHEMATICS / Mathematical Analysis
Classification: LCC QA274.75 .I42 2022 (print) | LCC QA274.75 (ebook) |
DDC 519.2/33–dc23
LC record available at https://lccn.loc.gov/2021029856
LC ebook record available at https://lccn.loc.gov/2021029857

ISBN 978-1-316-51655-3 Hardback

To my wife Diana and my son Jorge

Contents

Preface

The connection between stochastic processes, special functions and orthogonal polynomials has a long history. From the 1930s N. Wiener and later K. Itô knew about the connection between Hermite polynomials and integration theory with respect to Brownian motion. Around the 1950s many authors like M. Kac [80], W. Feller [53]–[56], E. Hille [71], W. Ledermann and G. E. Reuter [111], J. F. Barrett and D. G. Lampard [6], S. Karlin and J. McGregor [82]–[89], H. P. McKean [116] and D. G. Kendall [103] established an important connection between the transition probability functions of diffusion processes, continuous-time birth–death processes and discrete-time birth–death chains (in this order) by means of a spectral representation. This spectral representation is based on the spectral analysis of the infinitesimal operator associated with these special types of Markov processes and many probabilistic aspects can be analyzed in terms of the corresponding orthogonal eigenfunctions and eigenvalues. In the following years these relationships were developed further, finding connections with other stochastic processes like random matrices, Sheffer systems, Lévy processes, stochastic integration theory or Stein's method. For a brief account of all these relations see [129].

The main goal of this monograph is to give a comprehensive analysis of the main results related to the spectral representation of the most important one-dimensional Markov processes, namely discrete-time birth–death chains (also called random walks in some references, see [87]), birth–death processes and diffusion processes. Since the pioneering work of S. Karlin and J. McGregor in the 1950s, many authors have contributed to finding more applications of the spectral representation of the transition probability functions of these processes. This monograph tries to gather all the important results that appear in many publications over the last 60 years in one common text. The contents of this monograph served as a one-semester graduate advanced course taught at the Instituto de Matemáticas of the Universidad Nacional Autónoma de México in Fall 2018. The interested audience can be divided into two categories. On the one hand, it is intended for graduate students who have a solid background in the field of stochastic processes but are not so familiar with the theory of special functions and orthogonal polynomials. This monograph will give them alternative methods and ways of studying basic Markov processes by spectral methods. On the other hand, the book may also serve for students or researchers who are familiar with the theory of special functions and orthogonal polynomials but want to learn more about the connection between basic Markov processes and orthogonal polynomials.

In the experience of the author, graduate students are typically more familiar with probability theory and stochastic processes. This is the reason why an introduction to orthogonal polynomials is included in Chapter 1. This chapter also includes the concept of the *Stieltjes transform* and some of its properties, which will play a very important role in the spectral analysis of discrete-time birth–death chains and birth–death processes. A section about the spectral theorem for orthogonal polynomials (or *Favard's theorem*) will give insights about the relation between tridiagonal Jacobi matrices and spectral probability measures. We will focus then on the *classical families of orthogonal polynomials*, both of a continuous and a discrete variable. These families are characterized by the fact that they are eigenfunctions of a second-order differential or difference operator of hypergeometric type solving certain *Sturm–Liouville problems*. These classical families are part of the so-called *Askey scheme*.

In Chapter 2 we will perform the spectral analysis of discrete-time birth–death chains on $\mathbb{N}_0$, which are the most basic and important discrete-time Markov chains. These chains are characterized by a tridiagonal one-step transition probability matrix. We will obtain the so-called *Karlin–McGregor integral representation formula* of the n-step transition probability matrix in terms of orthogonal polynomials with respect to a probability measure ψ with support inside the interval $[-1,1]$. We will give an extensive collection of examples related to orthogonal polynomials, including gambler's ruin, the Ehrenfest model, the Bernoulli–Laplace model and the Jacobi urn model. The chapter ends with applications of the Karlin–McGregor formula to probabilistic aspects of discrete-time birth–death chains, such as recurrence, absorption, the strong ratio limit property and the limiting conditional distribution. Finally we will apply spectral methods to discrete-time birth–death chains on $\mathbb{Z}$, which are not so much studied in the literature.

In Chapter 3 we will perform the spectral analysis of birth–death processes on $\mathbb{N}_0$, which are the most basic and important continuous-time Markov chains. In this case, these processes will be characterized by an infinitesimal operator, which is a tridiagonal matrix whose spectrum is inside the interval $(-\infty,0]$. Again, we will obtain the *Karlin–McGregor integral representation formula* of the transition probability functions of the process in terms of orthogonal polynomials with respect to a probability measure ψ with support inside the interval $[0,\infty)$. Although many of the results are similar or equivalent to those of discrete-time birth–death chains, the methods and techniques are quite different. For instance, in this chapter, we will have to prove that the Karlin–McGregor representation formula is in fact a transition probability function of a birth–death process, something that was not necessary for the case of discrete-time birth–death chains. We will also provide an extensive collection of examples related to orthogonal polynomials, including the $M/M/k$ queue with $1 \leq k \leq \infty$ servers, the continuous-time Ehrenfest and Bernoulli–Laplace urn models, a genetics model of Moran and linear birth–death processes. As in the case of discrete-time birth–death chains, we will apply the Karlin–McGregor formula to probabilistic aspects of birth–death processes, such as processes with killing, recurrence, absorption, the strong ratio limit property, the limiting conditional distribution, the decay parameter, quasi-stationary distributions and bilateral birth–death processes on $\mathbb{Z}$.

In Chapter 4 we will perform the spectral analysis of one-dimensional diffusion processes, which are the most basic and important continuous-time Markov processes where now the state space is a continuous interval contained in $\mathbb{R}$. Diffusion processes are characterized by an infinitesimal operator, which is a second-order differential operator with a drift and a diffusion coefficient. We will obtain a spectral representation of the transition probability density of the process in terms of the orthogonal eigenfunctions of the corresponding infinitesimal operator, for which we will have to solve a *Sturm–Liouville problem* with certain boundary conditions. An analysis of the behavior of these boundary points will also be made. We will also give an extensive collection of examples related to special functions and orthogonal polynomials, including Brownian motion with drift and scaling, the Orstein–Uhlenbeck process, a population growth model, the Wright–Fisher model, the Jacobi diffusion model and the Bessel process, among others. Finally, we will study the concept of quasi-stationary distributions, for which the spectral representation will play an important role.

I would like to thank F. Alberto Grünbaum for introducing me to the fascinating connection between orthogonal polynomials and Markov processes. Back in 2009 I was visiting him as an undergraduate student at the University of California, Berkeley and we were studying one example of matrix-valued orthogonal polynomials coming from group representation theory which had a nice interpretation in terms of two-dimensional Markov chains. This was my first connection to the subject that brought me to write this monograph. I would also like to thank Eric A. van Doorn for reading the manuscript and providing an important list of corrections and additional material to include in the book. Unfortunately he tragically passed away before being

able to read the final version of this book. In closing I would like to thank the staff at Cambridge University Press for their support and cooperation during the preparation of this book.

Manuel Domínguez de la Iglesia[1]
Ciudad de México

[1] Partially supported by PAPIIT-DGAPA-UNAM grants IA102617 and IN104219 (México) and CONACYT grant A1-S-16202 (México).

1

Orthogonal Polynomials

In this chapter we introduce some basic definitions and properties about the theory of special functions and orthogonal polynomials on the real line. In the first section we will introduce some basic special functions and the concept of the *Stieltjes transform*, which will be used frequently in the text. In Section 1.2 we will give some properties of the general theory of orthogonal polynomials. Section 1.3 is devoted to the *spectral theorem* and in particular applied to orthogonal polynomials, in which case it is usually called *Favard's theorem*. In Sections 1.4 and 1.5 we will focus on the so-called *classical orthogonal polynomials*, both of a continuous and a discrete variable. These special families, apart from being orthogonal, are characterized by the fact that they are eigenfunctions of a second-order differential operator (in the continuous variable) or a second-order difference operator (in the discrete variable) of the *Sturm–Liouville* type. Finally, in Section 1.6, we describe the *Askey scheme*, which is a way of organizing orthogonal polynomials of hypergeometric type into a hierarchy, where the classical orthogonal polynomials are included. This chapter is based on references [3, 9, 16, 74, 135, 137, 142].

1.1 Some Special Functions and the Stieltjes Transform

The *Gamma function* is a complex-valued function that extends the domain of the factorial function of a nonnegative integer $n!$. It was introduced by Euler in 1789 and it is defined by its integral representation

$$\Gamma(z) = \int_0^\infty e^{-t} t^{z-1} dt, \quad \operatorname{Re} z > 0. \tag{1.1}$$

Integrating by parts we obtain the functional equation

$$z\Gamma(z) = \Gamma(z+1), \quad \operatorname{Re} z > 0.$$

The formula above can also be written as

$$(z)_n \Gamma(z) = \Gamma(z+n), \quad n \geq 0,$$

where $(z)_n$ is the *Pochhammer symbol*

$$(z)_n = \begin{cases} 1, & \text{if } n = 0, \\ z(z+1)\cdots(z+n-1), & \text{if } n \geq 1. \end{cases} \tag{1.2}$$

From here we also observe that if n is a nonnegative integer, then $\Gamma(n+1) = n!$.

The *Beta function* is defined by the integral

$$\mathrm{B}(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}dt, \quad \operatorname{Re} x,\ \operatorname{Re} y > 0. \tag{1.3}$$

It is symmetric, i.e. $\mathrm{B}(x,y) = \mathrm{B}(y,x)$, and it is related to the Gamma function by the well-known formula

$$\mathrm{B}(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}.$$

A *hypergeometric series* $\sum_{n=0}^{\infty} c_n$ is a series for which $c_0 = 1$ and the ratio of consecutive terms is a rational function of the summation index n, i.e. one for which

$$\frac{c_{n+1}}{c_n} = \frac{P(n)}{Q(n)},$$

where $P(n)$ and $Q(n)$ are polynomials. In this case, c_n is called a hypergeometric term. If the polynomials are completely factored, the ratio of successive terms can be written as

$$\frac{c_{n+1}}{c_n} = \frac{P(n)}{Q(n)} = \frac{(n+a_1)(n+a_2)\cdots(n+a_p)}{(n+b_1)(n+b_2)\cdots(n+b_q)(n+1)},$$

where the factor $n+1$ in the denominator is present for historical reasons of notation. From here we define the *generalized hypergeometric function* as

$${}_pF_q\left(\begin{matrix} a_1, \ldots, a_p \\ b_1, \ldots, b_q \end{matrix}; x\right) = \sum_{n=0}^{\infty} c_n x^n = \sum_{n=0}^{\infty} \frac{(a_1)_n \cdots (a_p)_n}{(b_1)_n \cdots (b_q)_n} \frac{x^n}{n!}. \tag{1.4}$$

We can also use the following notation for generalized hypergeometric functions:

$${}_pF_q(a_1, \ldots, a_p; b_1, \ldots, b_q; x).$$

This series is absolutely convergent for all x if $p \leq q$ and for $|x| < 1$ if $p = q+1$. It is divergent for all $x \neq 0$ if $p > q+1$, as long as the series is not finite. Observe that when one of the parameters of the numerator $a_i, i = 1, \ldots, p,$ is a negative integer, then the generalized hypergeometric function is a polynomial.

Many of the known special functions can be represented in terms of generalized hypergeometric functions. For example, the simplest cases of ${}_0F_0$ and ${}_1F_0$ correspond to the exponential series and the binomial series, respectively. Indeed,

$${}_0F_0(-;\,-\,;x) = \sum_{n=0}^{\infty} \frac{x^n}{n!} = e^x,$$

$$
{}_1F_0(a;\,-\,;x) = \sum_{n=0}^{\infty} \frac{(a)_n x^n}{n!} = \sum_{n=0}^{\infty} \frac{\Gamma(z+n)}{\Gamma(a)\Gamma(n+1)} x^n = \sum_{n=0}^{\infty} \binom{a+n-1}{n} x^n = (1-x)^{-a}.
$$

If $p = 2$ and $q = 1$, the function becomes what is called the *Gaussian hypergeometric function* ${}_2F_1(a,b;c;x)$ and it is related to the solutions of *Euler's hypergeometric differential equation*

$$
x(1-x)y''(x) + [c-(a+b+1)x]y'(x) - aby(x) = 0. \tag{1.5}
$$

We will see later the relation of this equation with the Jacobi polynomials. All families of orthogonal polynomials in the Askey scheme admit a representation in terms of hypergeometric series, as we will see later. For more information about generalized hypergeometric functions see [3, Chapter 2].

The *Stieltjes transform* (also known as the Cauchy transform) of a measure ψ defined on $\mathbb{R}$ is defined as the complex-valued function

$$
B(z;\psi) = \int_{\mathbb{R}} \frac{d\psi(x)}{x-z}, \quad z \in \mathbb{C}\setminus\mathbb{R}. \tag{1.6}
$$

This transform is related to the *generating function of the moments* of the measure ψ, since, formally

$$
B(z;\psi) = -\frac{1}{z}\int_{\mathbb{R}} \frac{1}{1-x/z} d\psi(x) = -\frac{1}{z}\sum_{n=0}^{\infty}\int_{\mathbb{R}} \frac{x^n}{z^n}\, d\psi(x) = -\sum_{n=0}^{\infty} \frac{\mu_n}{z^{n+1}}, \tag{1.7}
$$

where $\mu_n = \int_{\mathbb{R}} x^n d\psi(x)$ are the *moments* of the measure. In the case where $\mathrm{supp}(\psi) \subseteq [-A,A]$, then $|\mu_n| \le 2A^n$, implying that the series (1.7) is absolutely convergent for $|z| > A$. In this case, the Stieltjes transform is completely determined in terms of the moments of the measure ψ. In general, the expansion of the Stieltjes transform (1.6) has to be interpreted as an asymptotic expansion of the Stieltjes transform $B(z;\psi)$ as $|z| \to \infty$.

There is a formula which allows to calculate the measure ψ if we have information about the corresponding Stieltjes transform. This formula is known as the *Perron–Stieltjes inversion formula*. It has several versions, but the one we will use in this text is included in the following result.

Proposition 1.1 ([51, Theorem X.6.1]) *Let ψ be a probability measure with finite moments and $B(z;\psi)$ its Stieltjes transform* (1.6)*. Then*

$$
\int_a^b d\psi(x) + \frac{1}{2}\psi(\{a\}) + \frac{1}{2}\psi(\{b\}) = \frac{1}{\pi}\lim_{\varepsilon\downarrow 0}\int_a^b \mathrm{Im}B(x+i\varepsilon;\psi)\,dx, \tag{1.8}
$$

where $\psi(\{a\}) \ge 0$ is the magnitude or size of the mass at an isolated point a. If the measure is absolutely continuous at a then $\psi(\{a\}) = 0$.

Proof Observe that

$$2i\mathrm{Im}B(z;\psi) = B(z;\psi) - \overline{B(z;\psi)} = B(z;\psi) - B(\bar{z};\psi) = \int_{\mathbb{R}} \left[\frac{1}{x-z} - \frac{1}{x-\bar{z}}\right] d\psi(x)$$
$$= \int_{\mathbb{R}} \frac{z-\bar{z}}{|x-z|^2} d\psi(x) = 2i \int_{\mathbb{R}} \frac{\mathrm{Im}z}{|x-z|^2}\, d\psi(x).$$

Therefore

$$\mathrm{Im}B(x+i\varepsilon;\psi) = \int_{\mathbb{R}} \frac{\varepsilon}{|s-(x+i\varepsilon)|^2}\, d\psi(s) = \int_{\mathbb{R}} \frac{\varepsilon}{(s-x)^2+\varepsilon^2}\, d\psi(s).$$

Integrating and exchanging integrals (which is allowed since the integrand is positive) we have that

$$\int_a^b \mathrm{Im}B(x+i\varepsilon;\psi)\, dx = \int_{\mathbb{R}} \left[\int_a^b \frac{\varepsilon}{(s-x)^2+\varepsilon^2} dx\right] d\psi(s).$$

The internal integral can be calculated explicitly by making the change of variables $y = (x-s)/\varepsilon$:

$$\chi_\varepsilon(s) = \int_a^b \frac{\varepsilon}{(s-x)^2+\varepsilon^2}\, dx = \int_{(a-s)/\varepsilon}^{(b-s)/\varepsilon} \frac{1}{1+y^2}\, dy = \arctan y \Big|_{y=(a-s)/\varepsilon}^{y=(b-s)/\varepsilon}.$$

We have that $0 \le \chi_\varepsilon(s) \le \pi$ and when we take the limit (which is also allowed using the Lebesgue dominated convergence theorem since ψ is a probability measure and $\chi_\varepsilon(s)$ is bounded and positive) we have that

$$\lim_{\varepsilon\downarrow 0} \chi_\varepsilon(s) = \begin{cases} \pi, & \text{if} \quad a < s < b, \\ \frac{\pi}{2}, & \text{if} \quad s = a \text{ or } s = b. \end{cases} \qquad \square$$

As a consequence of the previous proposition we also have the formula

$$\int_a^b d\psi(x) = \frac{1}{\pi} \lim_{\varepsilon\downarrow 0} \lim_{\eta\downarrow 0} \int_{a+\eta}^{b-\eta} \mathrm{Im}B(x+i\varepsilon;\psi)\, dx. \tag{1.9}$$

When the measure is absolutely continuous with respect to the Lebesgue measure, i.e. $d\psi(x) = \psi(x)\, dx$ (abusing the notation), we have

$$\psi(x) = \frac{1}{\pi} \lim_{\varepsilon\downarrow 0} \mathrm{Im}B(x+i\varepsilon;\psi) = \lim_{\varepsilon\downarrow 0} \frac{B(x+i\varepsilon;\psi) - B(x-i\varepsilon;\psi)}{2\pi i}. \tag{1.10}$$

Finally, for measures that have an absolutely continuous part and a discrete part, there is a direct way to calculate the size of the jump. Indeed, assume that $\psi = \widehat{\psi} + \psi(\{a\})\delta_a$, where $\delta_a(x) = \delta(x-a)$ is the Dirac delta distribution which is defined, as usual, by $\int_{\mathbb{R}} f(x)\delta(x-a)\, dx = f(a)$. Then, since the Stieltjes transform is linear, we have

$$B(z;\psi) = B(z;\widehat{\psi}) + \frac{\psi(\{a\})}{a-z}.$$

Evaluating at $z = a + i\varepsilon$ and taking imaginary parts, we have

$$\mathrm{Im}B(a+i\varepsilon;\psi) = \mathrm{Im}B(a+i\varepsilon;\widehat{\psi}) + \mathrm{Im}\frac{\psi(\{a\})}{-i\varepsilon} = \mathrm{Im}B(a+i\varepsilon;\widehat{\psi}) + \frac{\psi(\{a\})}{\varepsilon}.$$

Therefore we get

$$\psi(\{a\}) = \varepsilon\mathrm{Im}B(a+i\varepsilon;\psi) - \varepsilon\mathrm{Im}B(a+i\varepsilon;\widehat{\psi}). \tag{1.11}$$

Taking limits as $\varepsilon \downarrow 0$ we observe that $B(a+i\varepsilon;\widehat{\psi})$ is bounded since $\widehat{\psi}$ is absolutely continuous. Therefore the meaningful isolated points (where $\psi(\{a\}) > 0$) must be those satisfying

$$\lim_{\varepsilon\downarrow 0}\mathrm{Im}B(a+i\varepsilon;\psi) = \infty,$$

while the size of the jump at $x = a$ is given by

$$\psi(\{a\}) = \lim_{\varepsilon\downarrow 0}\varepsilon\mathrm{Im}B(a+i\varepsilon;\psi) \geq 0. \tag{1.12}$$

Example 1.2 Let $B(z;\psi)$ be given by

$$B(z;\psi) = \frac{1}{1-z}, \quad z \in \mathbb{C}\setminus\{1\}.$$

According to (1.11) there will be a pole at $z = 1$, so it is a candidate for a singular part of the measure. Assume that $\psi = \widehat{\psi} + \psi(\{1\})\delta_1$, where $\widehat{\psi}$ is the absolutely continuous part. Then, by (1.10), we have

$$\widehat{\psi}(x) = \frac{1}{\pi}\lim_{\varepsilon\downarrow 0}\mathrm{Im}\frac{1}{1-x-i\varepsilon} = \frac{1}{\pi}\lim_{\varepsilon\downarrow 0}\mathrm{Im}\left(\frac{1-x+i\varepsilon}{(1-x)^2+\varepsilon^2}\right) = \frac{1}{\pi}\lim_{\varepsilon\downarrow 0}\frac{\varepsilon}{(1-x)^2+\varepsilon^2}.$$

We observe that if $x \neq 1$, then $\widehat{\psi}(x) = 0$. Therefore the measure ψ consists only of a singular part at $x = 1$. The value of $\psi(\{1\})$ is given by (1.12) and it is easy to see that

$$\psi(\{1\}) = \lim_{\varepsilon\downarrow 0}\varepsilon\mathrm{Im}B(1+i\varepsilon;\psi) = \lim_{\varepsilon\downarrow 0}\varepsilon\frac{\varepsilon}{\varepsilon^2} = 1.$$

Therefore $\psi(x) = \delta_1(x)$. ◊

Example 1.3 Consider the Stieltjes transform given by

$$B(z;\psi) = -2z + 2\sqrt{z^2-1}, \quad z \in \mathbb{C}\setminus[-1,1],$$

where the branch of the square root is determined by analytic continuation from positive values for real $z > 1$. We observe that there are no singular points, so the measure will consist only of an absolutely continuous part. From (1.10) we get

$$\begin{aligned}\psi(x) &= \frac{1}{\pi}\lim_{\varepsilon\downarrow 0}\mathrm{Im}B(x+i\varepsilon;\psi)\\ &= \frac{1}{\pi}\lim_{\varepsilon\downarrow 0}\left(-2\varepsilon + 2\mathrm{Im}\sqrt{x^2-\varepsilon^2+2ix\varepsilon-1}\right) = \frac{2}{\pi}\mathrm{Im}\sqrt{x^2-1}.\end{aligned}$$

The last part has only imaginary part when $|x| \leq 1$. Therefore

$$\psi(x) = \frac{2}{\pi}\sqrt{1-x^2}, \quad |x| < 1,$$

which is the *Wigner semicircle distribution*. ◊

In Chapters 2 and 3 we will see several examples of computation of measures using the Perron–Stieltjes inversion formula.

Remark 1.4 As we have seen in (1.7), the Stieltjes transform is related to the generating function of the moments of a probability measure ψ. This is not exactly the same as the usual *moment generating function*, which is defined as

$$M_X(t) = \mathbb{E}(e^{tX}) = \sum_{n=0}^{\infty} \mu_n \frac{t^n}{n!},$$

where X is the random variable associated with the probability measure ψ. This moment generating function is more related to the *Laplace transform*. Indeed, assume that the probability measure is absolutely continuous and supported on $[0, \infty)$. Then the Laplace transform is defined by

$$\mathcal{L}[\psi](s) = \int_0^{\infty} e^{-sx}\psi(x)\,dx.$$

Then we have $\mathcal{L}[\psi](-t) = M_X(t)$. The Stieltjes transform arises naturally as an iteration of the Laplace transform. Indeed, if we call $\phi(s) = \mathcal{L}[\psi](s)$ then, formally, we have

$$\begin{aligned}
\mathcal{L}[\phi](t) &= \int_0^{\infty} e^{-st}\phi(s)ds = \int_0^{\infty} e^{-st}\left(\int_0^{\infty} e^{-su}\psi(u)du\right) ds \\
&= \int_0^{\infty} \psi(u)\left(\int_0^{\infty} e^{-s(t+u)}ds\right) du \\
&= \int_0^{\infty} \psi(u)\left(-\frac{1}{t+u}e^{-s(t+u)}\Big|_{s=0}^{s=\infty}\right) du = \int_0^{\infty} \frac{\psi(u)}{t+u}\,du, \quad \mathrm{Re}(t) > 0.
\end{aligned}$$

Therefore $B(t;\psi) = \mathcal{L}^2[\psi](-t)$. A good reference about Stieltjes transforms in connection with the Laplace transform can be found in Chapter VIII of [146]. ◊

1.2 General Properties of Orthogonal Polynomials

Let ψ be a positive Borel measure on $\mathbb{R}$ with infinite support and let us assume that the moments

$$\mu_n = \int_{\mathbb{R}} x^n d\psi(x), \quad n \geq 0,$$

exist and are finite. We normalize the measure in such a way that $\mu_0 = 1$, so we have a probability measure. Following *Lebesgue's decomposition theorem* any Borel measure on the real line can be decomposed into three measures such that

$$\psi = \psi_c + \psi_d + \psi_{sc},$$

where ψ_c is absolutely continuous, ψ_d is discrete and ψ_{sc} is singular continuous. The absolutely continuous measure ψ_c is classified by the Radon–Nikodym theorem and can always be written (abusing the notation) as $d\psi_c(x) = \psi_c(x)dx$, with respect to the Lebesgue measure. The discrete measure ψ_d can always be written as

$$d\psi_d(x) = \sum_k \psi(\{x_k\})\delta(x - x_k)\, dx,$$

where k runs over a countable set, x_k are the mass points, $\psi(\{x_k\})$ are the sizes or magnitudes of these jumps and $\delta(x-a)$ is the Dirac delta distribution. Finally, the singular continuous measure ψ_{sc} is defined over a set of measure 0. The *Cantor measure* (the probability measure on the real line whose cumulative distribution function is the Cantor function) is an example of a singular continuous measure. In this text we consider positive Borel measures on $\mathbb{R}$ with either only an absolutely continuous part or only a discrete part (or a combination of both).

Associated with this measure ψ we can consider the Hilbert space L^2_ψ with the inner product

$$(f,g)_\psi = \int_{\mathbb{R}} f(x)g(x)\, d\psi(x), \tag{1.13}$$

of all measurable real functions f such that $(f,f)_\psi = \|f\|^2_\psi < \infty$. If the support of the measure is given by $\mathcal{S} \subseteq \mathbb{R}$, then this space will be written as $L^2_\psi(\mathcal{S})$. When $\mathcal{S}$ is a countable set, for example $\mathbb{N}_0 = \{0, 1, \ldots\}$, this space is usually denoted by $\ell^2_\psi(\mathbb{N}_0)$.

We say that $(p_n(x))_n$ is a *sequence of polynomials* if each element is a polynomial of degree exactly n in the real variable x. A sequence of polynomials is *monic* if the leading coefficient of each polynomial is exactly 1. A sequence of polynomials $(p_n)_n$ is *orthogonal* with respect to a Borel measure ψ if

$$(p_n, p_m)_\psi = \int_{\mathbb{R}} p_n(x)p_m(x)\, d\psi(x) = d_n^2\delta_{nm},$$

where $d_n^2 = \|p_n\|^2_\psi > 0$. If the norm is always identically 1, we say that the polynomial sequence is *orthonormal* and we denote it by $(P_n)_n$. When we work with the sequence of monic orthogonal polynomials, we will use the notation $(\widehat{P}_n)_n$ and its norms will be denoted by $\|\widehat{P}_n\|^2_\psi = \zeta_n$.

Given a Borel measure ψ on $\mathbb{R}$ with infinite support and finite moments, it will always be possible to build a sequence of orthogonal polynomials. A direct way is through the *Gram–Schmidt orthogonalization process* applied to the set $\{1, x, x^2, \ldots\}$.

This method builds the polynomials one by one taking into account that all the previous ones have already been calculated. Specifically

$$\begin{aligned}
\widehat{P}_0(x) &= 1, \\
\widehat{P}_1(x) &= x - \frac{(\widehat{P}_0, x)_\psi}{(\widehat{P}_0, \widehat{P}_0)_\psi}\widehat{P}_0(x), \\
&\vdots \qquad\qquad \vdots \\
\widehat{P}_k(x) &= x^k - \sum_{j=0}^{k-1} \frac{(\widehat{P}_j, x^k)_\psi}{(\widehat{P}_j, \widehat{P}_j)_\psi}\widehat{P}_j(x).
\end{aligned}$$

Once they have been computed, the monic polynomials can be normalized by dividing them by $\|\widehat{P}_k\|_\psi = \sqrt{\zeta_k}$. Observe that the monic orthogonal polynomials have always real coefficients.

Another way to define orthogonal polynomials is through determinants associated with the moments. Consider the determinant

$$\Delta_n = \begin{vmatrix} \mu_0 & \mu_1 & \cdots & \mu_n \\ \mu_1 & \mu_2 & \cdots & \mu_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_n & \mu_{n+1} & \cdots & \mu_{2n} \end{vmatrix}, \quad \Delta_{-1} = 1.$$

The quadratic form associated with the matrix of the previous determinant, which we denote by (Δ_n), is always positive definite. Indeed, for any real vector $v = (v_0, v_1, \ldots, v_n)^T$, we have that

$$v^T(\Delta_n)v = \sum_{j,k=0}^{n} \mu_{j+k} v_j v_k = \int_{\mathbb{R}} \left[\sum_{j=0}^{n} v_j x^j\right]^2 d\psi(x),$$

which is clearly positive. Thus $\Delta_n > 0, n \geq 0$. $\Delta_n, n \geq 0$ are usually called *Hankel determinants*.

It is easy to see that the sequence of polynomials defined by

$$p_n(x) = \begin{vmatrix} \mu_0 & \mu_1 & \cdots & \mu_{n-1} & 1 \\ \mu_1 & \mu_2 & \cdots & \mu_n & x \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mu_n & \mu_{n+1} & \cdots & \mu_{2n-1} & x^n \end{vmatrix}, \quad n \geq 0, \tag{1.14}$$

is orthogonal with respect to the measure ψ. To see that, we simply evaluate the inner product $(p_n, x^m)_\psi = 0, m = 0, 1, \ldots, n-1$ observing that we always have a repeated column, so the determinant is 0. Alternatively, we have $(p_n, x^n)_\psi = \Delta_n > 0$. Thus

$$p_n(x) = \Delta_{n-1}x^n + \text{lower degree terms},$$

and we have that

$$(p_n, p_n)_\psi = (p_n, \Delta_{n-1}x^n)_\psi = \Delta_{n-1}\Delta_n.$$

Therefore, the polynomials

$$P_n(x) = \frac{1}{\sqrt{\Delta_{n-1}\Delta_n}} p_n(x)$$

are orthonormal, and the leading coefficient is given by $h_n = \sqrt{\Delta_{n-1}/\Delta_n} = \zeta_n^{-1/2}$. The monic family can be written as

$$\widehat{P}_n(x) = \frac{1}{\Delta_{n-1}} p_n(x) = \sqrt{\frac{\Delta_n}{\Delta_{n-1}}} P_n(x).$$

Finally, let us see another way to generate the orthogonal polynomials recurrently. Assume that we have a sequence of orthogonal polynomials $(p_n)_n$. The polynomial $xp_n(x)$ has degree $n+1$ and can be expressed as a linear combination of the $n+1$ first polynomials, i.e.

$$xp_n(x) = \sum_{j=0}^{n+1} d_{n,j} p_j(x).$$

Now, multiplying by $p_k(x)$ and evaluating the inner product, it is easy to see, using the orthogonal relations, that the coefficients $d_{n,j} = 0, j = 0, 1, \ldots, n-2$. Therefore, only the last three coefficients remain and every family of orthogonal polynomials satisfies a *three-term recurrence relation* of the form

$$xp_n(x) = a_n p_{n+1}(x) + b_n p_n(x) + c_n p_{n-1}(x), \quad n \geq 0, \quad p_{-1} = 0, \tag{1.15}$$

where

$$a_n = \frac{(xp_n, p_{n+1})_\psi}{(p_{n+1}, p_{n+1})_\psi}, \quad b_n = \frac{(xp_n, p_n)_\psi}{(p_n, p_n)_\psi}, \quad c_n = \frac{(xp_n, p_{n-1})_\psi}{(p_{n-1}, p_{n-1})_\psi}.$$

We observe that the coefficient b_n is always real. Moreover, for the orthonormal family $P_n(x)$ we have, comparing the coefficients of x^{n+1} in (1.15), that $a_n = h_n/h_{n+1} = \sqrt{\zeta_{n+1}/\zeta_n} > 0$, and that $c_n = (xP_n, P_{n-1})_\psi = (P_n, xP_{n-1})_\psi = a_{n-1}$. Therefore the sequence of orthonormal polynomials $(P_n)_n$ satisfies a three-term recurrence relation of the form

$$xP_n(x) = a_n P_{n+1}(x) + b_n P_n(x) + a_{n-1} P_{n-1}(x), \quad a_n > 0, \quad b_n \in \mathbb{R}. \tag{1.16}$$

For the monic family $(\widehat{P}_n)_n$ the three-term recurrence relation will be given by

$$x\widehat{P}_n(x) = \widehat{P}_{n+1}(x) + \alpha_n \widehat{P}_n(x) + \beta_n \widehat{P}_{n-1}(x), \quad \widehat{P}_0(x) = 1, \quad \widehat{P}_1(x) = x - \alpha_0, \tag{1.17}$$

where $\alpha_{n-1} \in \mathbb{R}, \beta_n > 0$ for $n \geq 1$. The relations between these coefficients and the coefficients of the orthonormal family are given by

$$a_n = \sqrt{\frac{\zeta_{n+1}}{\zeta_n}}, \quad \alpha_n = b_n, \quad \beta_n = \frac{\zeta_n}{\zeta_{n-1}}.$$

Observe that $\zeta_n = \beta_n \cdots \beta_1$.

Another way of writing this recurrence relation is in matrix form. Denoting the column vector of orthonormal polynomials by $P(x) = (P_0(x), P_1(x), \ldots)^T$, we have that $xP(x) = JP(x)$, where J is the tridiagonal symmetric matrix

$$J = \begin{pmatrix} b_0 & a_0 & 0 & 0 & 0 & \cdots \\ a_0 & b_1 & a_1 & 0 & 0 & \cdots \\ 0 & a_1 & b_2 & a_2 & 0 & \cdots \\ 0 & 0 & a_2 & b_3 & a_3 & \cdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots \end{pmatrix}. \tag{1.18}$$

This matrix plays a very important role and it is called a *Jacobi matrix*. In particular, we will find this kind of matrix in the one-step transition probability matrix of a one-dimensional discrete-time birth–death chain and in the infinitesimal operator of a birth–death process, as we will see in the next two chapters. The inverse result, i.e. for a family of polynomials defined by (1.16), where there exists a positive measure for which they are orthogonal, is known as Favard's theorem or the spectral theorem for orthogonal polynomials. We will see more details in Section 1.3.

The powers of J can be computed formally using orthogonality properties. Observe that the relation $xP(x) = JP(x)$ implies that $x^nP(x) = J^nP(x)$. Therefore, multiplying by $P^T(x)$, integrating with respect to the measure ψ and looking at the (i,j) entry, we obtain

$$\int_{\mathbb{R}} x^n P_i(x)P_j(x)d\psi(x) = \sum_{k\geq 0} \int_{\mathbb{R}} J^n_{ik} P_k(x)P_j(x)d\psi(x) = J^n_{ij}. \tag{1.19}$$

From here we observe that the moments $(\mu_n)_n$ of the measure ψ can be computed from J^n_{00}. In general, the diagonal coefficients J^n_{ii} are the moments of the measure $d\psi_i(x) = P_i^2(x)d\psi(x)$.

The identity (1.19) can be extended to any analytic function defined on $\operatorname{supp}(\psi)$ of the form $f(x) = \sum_{n\geq 0} c_n x^n$ as

$$\int_{\mathbb{R}} f(x)P_i(x)P_j(x)\, d\psi(x) = \sum_{n\geq 0} \int_{\mathbb{R}} c_n x^n P_k(x)P_j(x)\, d\psi(x) = \sum_{n\geq 0} c_n J^n_{ij} = f(J)_{ij}.$$

For instance, the function $f(x) = (1 - zx)^{-1}$ with $z^{-1} \in \mathbb{C} \setminus \operatorname{supp}(\psi)$ gives, formally, that

$$(I - zJ)^{-1}_{00} = \int_{\mathbb{R}} \frac{P_0^2(x)}{1 - xz}\, d\psi(x) = \int_{\mathbb{R}} \frac{d\psi(x)}{1 - xz} = \sum_{n \geq 0} \mu_n z^n, \tag{1.20}$$

i.e. the generating function of the moments of ψ. In terms of the Stieltjes transform $B(z; \psi)$ defined by (1.6), we have that

$$(I - zJ)^{-1}_{00} = -\frac{1}{z} B\left(\frac{1}{z}; \psi\right). \tag{1.21}$$

Theorem 1.5 *Let J be the Jacobi matrix given by (1.18) and denote by $J^{(0)}$ the Jacobi matrix built from J by removing the first row and column. Then we have*

$$(I - zJ)^{-1}_{00} = \frac{1}{1 - b_0 z - a_0^2 z^2 (I - zJ^{(0)})^{-1}_{00}}.$$

Proof Write the Jacobi matrix J in (1.18) as

$$J = \left(\begin{array}{c|cc} b_0 & a_0 & \cdots \\ \hline a_0 & & \\ 0 & & J^{(0)} \\ \vdots & & \end{array}\right).$$

Using the well-known formula for the inverse of a 2×2 block matrix

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} (A - BD^{-1}C)^{-1} & * \\ * & * \end{pmatrix},$$

applied to the matrix $I - zJ$, we get

$$(I - zJ)^{-1}_{00} = \left[1 - zb_0 - a_0^2 z^2 e_0^T (I - zJ^{(0)})^{-1} e_0\right]^{-1} = \frac{1}{1 - zb_0 - a_0^2 z^2 (I - zJ^{(0)})^{-1}_{00}},$$

where e_0 is the canonical vector $e_0 = (1, 0, \ldots)^T$. □

Remark 1.6 If we assume that associated with the Jacobi matrices J and $J^{(0)}$ there exist positive measures ψ and $\psi^{(0)}$, respectively, then we have, using (1.20), that

$$\int_{-1}^{1} \frac{d\psi(x)}{1 - xz} = \frac{1}{1 - b_0 z - a_0^2 z^2 \displaystyle\int_{-1}^{1} \frac{d\psi^{(0)}(x)}{1 - xz}}.$$

This formula relates the generating functions of the moments of the measures ψ and $\psi^{(0)}$. In terms of Stieltjes transforms, relation (1.21) gives

$$B(z;\psi) = -\frac{1}{z - b_0 + a_0^2 B(z;\psi^{(0)})}. \tag{1.22}$$

$\Diamond$

Example 1.7 Consider the Jacobi matrix given by

$$J = \left(\begin{array}{c|cccc} 0 & 1/2 & 0 & 0 & \cdots \\ \hline 1/2 & 0 & 1/2 & 0 & \cdots \\ 0 & 1/2 & 0 & 1/2 & \cdots \\ \vdots & \vdots & \ddots & \ddots & \ddots \end{array}\right).$$

Then we have that $J^{(0)} = J$ and consequently $\psi = \psi^{(0)}$. From (1.22) we obtain a quadratic equation for $B(z;\psi) = B(z)$, given by

$$B^2(z) + 4zB(z) + 4 = 0.$$

Therefore

$$B(z) = -2z \pm 2\sqrt{z^2 - 1}.$$

On the one hand, we can discard the negative solution of $B(z)$ since as $z \to \infty$ the Stieltjes transform should vanish. On the other hand, the function $B(z)$ is well defined as a single-valued function in the complex plane from which we have removed the interval $[-1,1]$. If we approach the cut from above, $B(z)$ has a nontrivial imaginary part coming from the square root. This square root has positive values for Re $z > 1$ and negative values for Re $z > 1$. Therefore we have

$$B(z;\psi) = -2z + 2\sqrt{z^2 - 1}, \quad z \in \mathbb{C} \setminus [-1,1],$$

and the branch of the square root is determined by analytic continuation from positive values of Re $z > 1$. This example is precisely the one studied in Example 1.3 and the spectral measure is given by the Wigner semicircle distribution using the Perron–Stieltjes inversion formula. $\Diamond$

Multiplying (1.16) by $P_n(y)$ and (1.16) (for $x = y$) by $P_n(x)$, and subtracting both formulas, we get the telescopic relation

$$\begin{aligned}(x-y)P_n(x)P_n(y) = a_n[&P_{n+1}(x)P_n(y) - P_n(x)P_{n+1}(y)] \\ &- a_{n-1}[P_n(x)P_{n-1}(y) - P_{n-1}(x)P_n(y)],\end{aligned}$$

which iterating, adding and dividing by $x - y$ gives the *Christoffel–Darboux formula*

$$K_n(x,y) \doteq \sum_{j=0}^{n} P_j(x)P_j(y) = a_n\left[\frac{P_{n+1}(x)P_n(y) - P_n(x)P_{n+1}(y)}{x-y}\right]. \tag{1.23}$$

Taking $y \to x$ we get the *confluent Christoffel–Darboux formula*

$$\sum_{j=0}^{n} P_j^2(x) = a_n[P'_{n+1}(x)P_n(x) - P_{n+1}(x)P'_n(x)]. \tag{1.24}$$

The kernel $K_n(x,y)$ generated by the Christoffel–Darboux formula has the *reproducing kernel property*, i.e for every polynomial p of degree n, we have that

$$\int_{\mathbb{R}} p(x)K_n(x,y)d\psi(x) = p(y).$$

In terms of the monic family $(\widehat{P}_n)_n$ the Christoffel–Darboux formula can be written as

$$\sum_{j=0}^{n} \frac{\widehat{P}_j(x)\widehat{P}_j(y)}{\zeta_j} = \frac{\widehat{P}_{n+1}(x)\widehat{P}_n(y) - \widehat{P}_n(x)\widehat{P}_{n+1}(y)}{\zeta_n(x-y)} \tag{1.25}$$

and the confluent formula as

$$\sum_{j=0}^{n} \frac{\widehat{P}_j^2(x)}{\zeta_j} = \frac{\widehat{P}'_{n+1}(x)\widehat{P}_n(x) - \widehat{P}_{n+1}(x)\widehat{P}'_n(x)}{\zeta_n}. \tag{1.26}$$

Observe that the Christoffel–Darboux formula is a property that holds for every (monic) sequence of polynomials generated by the three-term recurrence relation (1.17), no matter if they are orthogonal or not with respect to some measure. The sequence ζ_n is generated by $\zeta_n = \beta_n \cdots \beta_1$. In the following result we will prove certain properties of the zeros of these polynomials.

Proposition 1.8 *The zeros or roots of the monic polynomials $\widehat{P}_n$ generated by the three-term recurrence relation* (1.16) *are all real and simple. Moreover the zeros of $\widehat{P}_{n+1}$ and $\widehat{P}_n$ interlace. If the polynomials are orthogonal with respect to some measure ψ, then these zeros lie in the smallest closed interval containing supp(ψ) for all $n \geq 1$.*

Proof Let u be a complex zero of $\widehat{P}_n$. Since the coefficients of $\widehat{P}_n$ are all real then $\bar{u}$ is also a complex zero of $\widehat{P}_n$. Taking $x = u$ and $y = \bar{u}$ in the Christoffel–Darboux formula (1.25) we get a contradiction since the right-hand part should be 0 while the left-hand side is > 1 (sum of positive absolute values of complex numbers). Then all zeros must be real. Alternatively, if we had a multiple zero then the confluent Christoffel–Darboux formula (1.26) will give the same contradiction.

If $\widehat{P}_{n+1}$ and $\widehat{P}_n$ have a zero in common, then by the recursion formula (1.17), it is also a zero of $\widehat{P}_{n-1}$. Following this reasoning this zero is a zero of $\widehat{P}_0 = 1$, but this is a contradiction. Regarding the interlacing property, for $n < 2$ there is nothing to prove. For $n \geq 2$, (1.26) implies that $\widehat{P}'_{n+1}(x)\widehat{P}_n(x) - \widehat{P}_{n+1}(x)\widehat{P}'_n(x) > 0$. Assume that $y_1 < y_2$ are two consecutive zeros of $\widehat{P}_{n+1}$. Then the above inequality implies that $\widehat{P}'_{n+1}(y_j)\widehat{P}_n(y_j) > 0, j = 1, 2$. Since $\widehat{P}'_{n+1}(y_1)$ and $\widehat{P}'_{n+1}(y_2)$ must have different signs, since they are simple, it follows from the previous inequality that $\widehat{P}_n(y_j), j = 1, 2$, has different signs. Thus, $\widehat{P}_n$ has a zero in the range (y_1, y_2) by Bolzano's theorem.

Finally let $[a,b]$ the smallest closed interval containing $\operatorname{supp}(\psi)$ and $c_1, \dots, c_j$ the zeros of $\widehat{P}_n$ contained in $[a,b]$. If $j < n$ then the orthogonality implies that $\int_{\mathbb{R}} \widehat{P}_n(x) \prod_{k=1}^{j} (x - c_k) d\psi(x) = 0$. But this is a contradiction because the integrand does not change signs on $[a,b]$. Therefore $j = n$. □

For a fixed n, let $x_{n,j}$, $j = 1, \dots, n$ denote the zeros of $\widehat{P}_n$ arranged in the following form:

$$x_{n,1} < x_{n,2} < \cdots < x_{n,n}. \tag{1.27}$$

The interlacing property says that each sequence $(x_{n,i})_n$ is monotone, therefore the limits exist. Define them as

$$\xi_i = \lim_{n\to\infty} x_{n,i} \quad \text{and} \quad \eta_j = \lim_{n\to\infty} x_{n,n-j+1}, \quad i,j \geq 1. \tag{1.28}$$

We have that

$$-\infty \leq \xi_i \leq \xi_{i+1} < \eta_{j+1} \leq \eta_j \leq \infty, \quad i,j \geq 1.$$

The interval $[\xi_1, \eta_1]$ is usually called the *true interval of orthogonality* and it is the smallest closed interval containing $\operatorname{supp}(\psi)$. Therefore $\xi_1 = \inf \operatorname{supp}(\psi)$ and $\eta_1 = \sup \operatorname{supp}(\psi)$. If we call

$$\sigma = \lim_{i\to\infty} \xi_i \quad \text{and} \quad \tau = \lim_{j\to\infty} \eta_j,$$

then we have

$$-\infty \leq \xi_i \leq \sigma \leq \tau \leq \eta_j \leq \infty, \quad i,j \geq 1.$$

Therefore, defining the (possible finite) sets

$$\Xi = \{\xi_1, \xi_2, \dots\} \quad \text{and} \quad H = \{\eta_1, \eta_2, \dots\},$$

we have that (see [16, II.4.2])

$$\operatorname{supp}(\psi) = \bar{\Xi} \cup S \cup \bar{H},$$

where the bar denotes closure and $S \subset (\sigma, \tau)$. Also σ is the smallest and τ is the largest limit point of $\operatorname{supp}(\psi)$.

Another way to prove Proposition 1.8 is to write the monic polynomials $\widehat{P}_n$ as the characteristic polynomial of the truncated Jacobi matrix of dimension $n \times n$ in the following form:

$$\widehat{P}_n(x) = \begin{vmatrix} x-\alpha_0 & -1 & 0 & \cdots & 0 & 0 & 0 \\ -\beta_1 & x-\alpha_1 & -1 & \cdots & 0 & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & & -\beta_{n-2} & x-\alpha_{n-2} & -1 \\ 0 & 0 & 0 & \cdots & 0 & -\beta_{n-1} & x-\alpha_{n-1} \end{vmatrix}. \tag{1.29}$$

The zeros of the orthogonal polynomials play an important role in the approximation of integrals of the form $\int_{\mathbb{R}} f(x)d\psi(x)$ by *Gaussian quadrature formulas*. For a fixed n, let $x_{n,j}, j = 1, \dots, n$ denote the zeros of $\widehat{P}_n$ arranged in the form (1.27).

Theorem 1.9 *For any n, there exist positive numbers $\lambda_1, \dots, \lambda_n$, such that*

$$\int_{\mathbb{R}} p(x)d\psi(x) = \sum_{k=1}^{n} \lambda_k p(x_{n,k}) \tag{1.30}$$

for all polynomials p of degree at most $2n - 1$. The values λ_k admit the following representation:

$$\lambda_k = \int_{\mathbb{R}} \frac{\widehat{P}_n(x)}{\widehat{P}'_n(x_{n,k})(x - x_{n,k})} d\psi(x) = \int_{\mathbb{R}} \left[\frac{\widehat{P}_n(x)}{\widehat{P}'_n(x_{n,k})(x - x_{n,k})} \right]^2 d\psi(x). \tag{1.31}$$

Proof A proof can be found in [74, Theorem 2.4.1]. □

The numbers $\lambda_1, \dots, \lambda_n$ are usually called *Christoffel numbers*. Observe from (1.31) that they are always positive.

Proposition 1.10 *The Christoffel numbers have the following properties:*

$$1 = \sum_{k=1}^{n} \lambda_k, \tag{1.32}$$

$$\lambda_k = -\frac{\zeta_n}{\widehat{P}_{n+1}(x_{n,k})\widehat{P}'_n(x_{n,k})}, \quad k = 1, \dots, n,$$

$$\frac{1}{\lambda_k} = K_n(x_{n,k}, x_{n,k}), \quad k = 1, \dots, n.$$

Proof A proof can be found in [74, Theorem 2.4.2]. □

The three-term recurrence relation (1.16) is a second-order difference equation and therefore it must have two linearly independent solutions. One is given by $P_n(x)$, and the other can be constructed using the initial conditions

$$P_0^{(0)}(x) = 0, \quad P_1^{(0)}(x) = 1/a_0,$$

which makes $P_n^{(0)}$ a polynomial of degree $n - 1$. These polynomials are called *associated polynomials*, 0th associated polynomials, numerator polynomials or polynomials of the second kind. Multiplying the recurrence relation (1.16) by $P_n^{(0)}$ and subtracting (1.16) (for $P_n^{(0)}$) multiplied by P_n, we can see that

$$a_{n-1} \left[P_n(x)P_{n-1}^{(0)}(x) - P_n^{(0)}(x)P_{n-1}(x) \right] = a_{n-1} \begin{vmatrix} P_n(x) & P_n^{(0)}(x) \\ P_{n-1}(x) & P_{n-1}^{(0)}(x) \end{vmatrix} = -1 \neq 0.$$

Then they are linearly independent. This relation (called a Casoratian determinant) also shows that the zeros of $P_n^{(0)}(x)$ are all real, simple and interlace with the zeros of $P_n(x)$. We also have the following integral representation:

$$P_n^{(0)}(x) = \int_{\mathbb{R}} \frac{P_n(x) - P_n(y)}{x - y} d\psi(y), \quad n \geq 0. \tag{1.33}$$

Indeed, let us use $R_n(x)$ to denote the right-hand side of (1.33). Then $R_0(x) = 0$ and $R_1(x) = 1/a_0$. For x not real and $n > 0$, we have, using (1.16), that

$$\begin{aligned} a_n R_{n+1}(x) &- (x - b_n)R_n(x) + a_{n-1}R_{n-1}(x) \\ &= \int_{\mathbb{R}} \frac{-a_n P_{n+1}(y) + (x - b_n)P_n(y) - a_{n-1}P_{n-1}(y) + yP_n(y) - yP_n(y)}{x - y} d\psi(y) \\ &= \int_{\mathbb{R}} \frac{(x - y)P_n(y)}{x - y} d\psi(y) = 0, \quad n > 0. \end{aligned}$$

Another way to generate the associated polynomials is by using the Jacobi matrix $J^{(0)}$ built from the Jacobi matrix J in (1.18) by removing the first row and column (see Theorem 1.5). The Stieltjes transforms of the spectral measures associated with both Jacobi matrices are related by the formula (1.22). For more information about how to compute the spectral measure associated with the associated polynomials see [61, 136] or more recently [31].

There is an important asymptotic result that relates these two solutions of the three-term recurrence relation with the Stieltjes transform of the corresponding measure ψ.

Theorem 1.11 (Markov's theorem) *Let ψ be a positive measure defined in a bounded interval $[a,b]$ and consider the corresponding orthonormal polynomials $P_n(x)$ and the associated polynomials $P_n^{(0)}(x)$. Then we have that*

$$\lim_{n\to\infty} \frac{P_n^{(0)}(z)}{P_n(z)} = \int_a^b \frac{d\psi(x)}{z - x}, \quad z \in \mathbb{C} \setminus [a,b],$$

and the convergence is uniform on compact subsets of $\mathbb{C} \setminus [a,b]$.

Proof Details of the proof can be found in [3, Section 5.5] or [34, Chapter 3]. □

There is a nice interpretation of the previous theorem in terms of *continued fractions* (see for instance [16, Chapter IV]), given by

$$\int_a^b \frac{d\psi(x)}{z - x} = \cfrac{1}{z - b_0 - \cfrac{a_0^2}{z - b_1 - \cfrac{a_1^2}{z - b_2 - \cfrac{a_2^2}{z - b_3 - \cdots}}}}, \quad z \in \mathbb{C} \setminus [a,b]. \tag{1.34}$$

This formula can be regarded as an alternative way of computing the Stieltjes transform of a probability measure ψ and eventually computing the measure by the Perron–Stieltjes inversion formula.

Finally we give without proof the necessary conditions for the completeness of a sequence of orthogonal polynomials in the space L^2_ψ.

Theorem 1.12 *Let ψ be an absolutely continuous positive Borel measure defined on an interval (a,b) and assume that for some $c > 0$, we have*

$$\int_a^b e^{c|x|}\psi(x)dx < \infty.$$

Let $(P_n)_n$ be a sequence of orthonormal polynomials with respect to ψ. Then, for any $f \in L^2_\psi(a,b)$, we have that

$$f(x) = \sum_{n=0}^{\infty}(f,P_n)_\psi P_n(x),$$

in the sense that the partial sums of the series converge in norm in the space $L^2_\psi(a,b)$. Moreover, we have Parseval's identity

$$\|f\|^2_\psi = \sum_{n=0}^{\infty}(f,P_n)^2_\psi. \tag{1.35}$$

Proof Details of the proof, using tools from Fourier analysis, can be found in [3, Section 6.5]. □

A measure ψ satisfying Parseval's identity (1.35) is usually called *extremal*.

1.3 The Spectral Theorem for Orthogonal Polynomials

In linear algebra, when we have a linear operator acting on $\mathbb{C}^n$, we may ask ourselves under what conditions a finite-dimensional square matrix associated with the operator can be diagonalized. In finite dimensions it is enough to analyze the spectrum or eigenvalues associated with this matrix. However, when we work with infinite-dimensional vector spaces, the situation is not as simple. The spectral theorem has a broader context in the theory of linear operators on Hilbert spaces equipped with an inner product. The spectral theorem identifies a class of linear operators that can be modeled by multiplication operators. In particular, self-adjoint operators will be of special interest. In general, given a self-adjoint linear operator A defined on a Hilbert space $\mathcal{H}$, we will always be able to find a measure ψ defined on a certain measurable space $\mathcal{S}$ and a unitary operator $U\colon \mathcal{H} \to L^2_\psi(\mathcal{S})$ such that

$$\left(UAU^{-1}f\right)(x) = F(x)f(x),$$

for a certain measurable and bounded real function F defined on $\mathcal{S}$. This is a generalization of the finite-dimensional case.

In the context of orthogonal polynomials, the operator A is identified with the symmetric tridiagonal Jacobi matrix J defined by (1.18), which is obviously self-adjoint (symmetric) in the Hilbert space $\mathcal{H} = \ell^2(\mathbb{N}_0)$. We want to find a measure ψ and a complete orthonormal basis in $L^2_\psi(\mathcal{S})$ with $\mathcal{S}$ a real interval (for example, the orthonormal polynomials) in such a way that $F(x) = x$. In the literature this result is usually called *Favard's theorem*, since J. Favard proved it in 1935. However, other authors, such as A. Wintner in 1929 or M. H. Stone in 1932, proved the same theorem some years before, or almost at the same time, such as J. A. Shohat in 1936.

There are several ways to prove the spectral theorem for orthogonal polynomials and in these notes we will see two different ones. The first one builds a distribution that is located at the zeros of the monic polynomials $\widehat{P}_n$ generated by a three-term recurrence relation and then we take $n \to \infty$. This version can be found in [74, Section 2.5] or more extensively in [16, Chapter II]. A second (more general) method uses functional analysis tools and spectral theory of self-adjoint operators (see [27, Chapter 2], [109] or [133]). We keep both proofs since the first one will be used in the spectral representation of birth–death processes in Chapter 3 while the second one will be in the spectral representation of one-dimensional discrete-time birth–death chains in Chapter 2 and bilateral birth–death processes in Chapter 3.

Other methods, which will not be discussed here, are related to the theory of positive linear functionals $\mathcal{L}$ defined by $\mathcal{L}(x^n) = \mu_n$ (see [16, p. 21]); or also solving the *moment problem*, which is divided into three depending on whether the support of the measure is finite (Hausdorff), semi-infinite (Stieltjes) or $\mathbb{R}$ (Hamburger). For more information, see [2, 133, 137]. There is a close connection between these problems and the theory of *continued fractions*.

Constructive Method Using the Zeros of the Polynomials

We will prove the spectral theorem for the monic family of polynomials $(\widehat{P}_n)_n$ defined by the three-term recurrence relation (1.17). From Proposition 1.8 we know that the zeros of $\widehat{P}_n$ are all real and simple and in addition those of $\widehat{P}_n$ and $\widehat{P}_{n-1}$ interlace. For a fixed n, let $x_{n,j}, j = 1, \dots, n$ be the zeros of $\widehat{P}_n$ arranged in the form (1.27). Since the polynomial is monic, we have $\widehat{P}_n(x) > 0$ for $x > x_{n,1}$ and therefore $(-1)^{j-1}\widehat{P}'_n(x_{n,j}) > 0$. Hence, using the confluent formula (1.24), we get that $(-1)^{j-1}\widehat{P}_{n-1}(x_{n,j}) > 0$. Therefore the sequence defined by

$$\rho(x_{n,j}) = \frac{\zeta_{n-1}}{\widehat{P}'_n(x_{n,j})\widehat{P}_{n-1}(x_{n,j})}, \quad 1 \leq j \leq n, \tag{1.36}$$

takes positive values. Using the Christoffel–Darboux formula (1.25) for monic polynomials with $\zeta_j = \beta_j \cdots \beta_1$ and the confluent formula (1.26), the expression (1.36) can be rewritten, taking $x = x_{n,r}$ and $y = x_{n,s}$, as

$$\rho(x_{n,r})\sum_{k=0}^{n-1}\frac{\widehat{P}_k(x_{n,r})\widehat{P}_k(x_{n,s})}{\zeta_k}=\delta_{rs}. \tag{1.37}$$

The real matrix U defined by

$$U=(u_{r,k}),\quad 1\le r,k\le n,\quad u_{r,k}=\sqrt{\rho(x_{n,r})}\frac{\widehat{P}_{k-1}(x_{n,r})}{\sqrt{\zeta_{k-1}}},$$

satisfies $UU^T=I$. Therefore $U^TU=I$ by the uniqueness of the inverse, i.e. U is a unitary matrix. From the (r,s) entry of $U^TU=I$ it follows that

$$\sum_{r=1}^{n}\rho(x_{n,r})\widehat{P}_k(x_{n,r})\widehat{P}_j(x_{n,r})=\zeta_k\delta_{jk},\quad j,k=0,1,\ldots,n-1. \tag{1.38}$$

The previous identity shows that there exists some discrete orthogonality of the polynomials $\widehat{P}_n(x)$ when we restrict their support to the corresponding zeros. The values $\rho(x_{n,j}), j=1,\ldots,n$ are the corresponding sizes of the jumps at those zeros. Note also that for $k=j=0$, we have $\sum_{r=1}^{n}\rho(x_{n,r})=1$, and the total sum of all these quantities is exactly 1.

We now introduce a sequence of distribution functions $(\psi_n)_n$ defined by

$$\psi_n(-\infty)=0,\qquad \lim_{x\downarrow x_{n,j}}\psi_n(x)-\lim_{x\uparrow x_{n,j}}\psi_n(x)=\rho(x_{n,j}). \tag{1.39}$$

Theorem 1.13 ([74, Theorem 2.5.2]) *Given a sequence of polynomials $(\widehat{P}_n)_n$ generated by the three-term recurrence relation (1.17) with $\alpha_{n-1}\in\mathbb{R}$ and $\beta_n>0$ for all $n\ge 1$, there exists a distribution function ψ such that*

$$\int_{\mathbb{R}}\widehat{P}_n(x)\widehat{P}_m(x)d\psi(x)=\zeta_n\delta_{nm}.$$

Proof From (1.38), we have that

$$1=\zeta_0=\int_{\mathbb{R}}d\psi_n(x)=\psi_n(\infty)-\psi_n(-\infty)=\sum_{r=1}^{n}\rho(x_{n,r}).$$

Therefore the functions ψ_n are uniformly bounded. Helly's selection principle (see [137, Introduction]) allows us to find a subsequence $(\psi_{n_k})_k$ of $(\psi_n)_n$ that converges to a distribution ψ, which is also non-decreasing and bounded. The same principle gives that if for all n, the moments of ψ_n exist, then the moments of ψ also exist and also the moments of the subsequence ψ_{n_k} converge to the moments of ψ. Since x^n can be written as a linear combination of the polynomials $\widehat{P}_j, j=0,1,\ldots,n$, from (1.38) we see that the moments of ψ_n exist. Therefore, taking limits in (1.38) as $n\to\infty$, we obtain the result. □

The previous result shows existence but not uniqueness. In fact there exist families of measures having all the same moments (see [74, Example 2.5.3]). Let us now see that if the coefficients α_n and β_n are bounded, then the measure is unique.

Theorem 1.14 ([74, Theorem 2.5.5]) *If the sequences $(\alpha_n)_n$ and $(\beta_n)_n$ are bounded, then the orthogonality measure ψ of Theorem 1.13 is unique.*

Proof First, if the sequences $(\alpha_n)_n$ and $(\beta_n)_n$ are bounded, then the support of ψ is also bounded. This is due to the representation (1.29) of the polynomials in terms of the truncated matrix J_n of the Jacobi matrix J. The zeros of $\widehat{P}_n$ are the eigenvalues of J_n. Therefore $|x_{n,j}| < 3M$, where M is an upper bound of both sequences $(\alpha_n)_n$ and $(\beta_n)_n$ (see [74, Theorem 1.1.1]). Therefore the support of each ψ_n is contained in the interval $(-3M, 3M)$.

Let ν be another orthogonality measure that has the same moments as the measure ψ. For any $a > 0$, we have

$$\int_{|x|\geq a} d\nu(x) \leq a^{-2n} \int_{|x|\geq a} x^{2n} d\nu(x) \leq a^{-2n} \int_{\mathbb{R}} x^{2n} d\nu(x) = a^{-2n} \int_{\mathbb{R}} x^{2n} d\psi(x).$$

Applying the quadrature formula (1.30) for $p(x) = x^{2n}$ and using (1.32), we have that

$$\int_{|x|\geq a} d\nu(x) \leq a^{-2n} \sum_{k=1}^{n+1} \lambda_k (x_{n+1,k})^{2n} \leq (A/a)^{2n} \sum_{k=1}^{n+1} \lambda_k = (A/a)^{2n},$$

where $|x_{n,j}| \leq A$ for all $n \geq 1$ and $1 \leq j \leq n$. Since a is a free parameter, in particular, if $a > A$, then $\int_{|x|\geq a} d\nu(x) = 0$ by taking $n \to \infty$ in the previous inequality. Therefore the support of ν is contained in $[-A, A]$. Let us now see that $\nu = \psi$. For $|x| \geq 2A$, we have that $\sum_{k=0}^{n} t^k x^{-k-1}$ converges to $1/(x-t)$ for all $t \in [-A, A]$ since $|t/x| < 1$ as a consequence of $|x| \geq 2A > A \geq |t|$. Thus

$$\int_{\mathbb{R}} \frac{d\psi(t)}{x-t} = \int_{\mathbb{R}} \lim_{n\to\infty} \sum_{k=0}^{n} \frac{t^k}{x^{k+1}} d\psi(t) = \lim_{n\to\infty} \int_{\mathbb{R}} \sum_{k=0}^{n} \frac{t^k}{x^{k+1}} d\psi(t) = \lim_{n\to\infty} \sum_{k=0}^{n} \frac{\mu_k}{x^{k+1}},$$

using the dominated convergence theorem, since $|t/x| \leq 1/2$. The last limit only depends on the moments and is the same for all measures that have the same moments. Then $F(x) = \int_{\mathbb{R}} \frac{d\psi(t)}{x-t}$ is uniquely determined for any x outside the circle $|x| = 2A$. Since F is analytic in $x \in \mathbb{C} \setminus [-A, A]$, F is unique. Then the theorem is a consequence of the Perron–Stieltjes inversion formula (1.8). □

For the more general case see for instance [74].

Methods from Functional Analysis and Spectral Theory

This method is based on important results on functional analysis and spectral analysis of linear operators in Hilbert spaces, which will not be proved in these notes. For more information about these results, see [125, 127].

Let $\mathcal{H}$ be a Hilbert space with an inner product $(\cdot,\cdot)$ and denote by $\mathcal{B}(\mathcal{H})$ the set of all linear operators of $\mathcal{H}$ in $\mathcal{H}$. For an operator $T \in \mathcal{B}(\mathcal{H})$, the *resolvent operator* is defined by $R(z) = (T - z)^{-1}$. The values of $z \in \mathbb{C}$ for which $R(z)$ is a bounded linear operator are called *regular values* and are denoted by $\rho(T)$. The complement of the *resolvent set* $\rho(T)$ is called the *spectrum* of T and is denoted by $\sigma(T)$. For a bounded operator T the spectrum $\sigma(T)$ is a compact subset of the disk of radius $\|T\| = \inf_{u\in\mathcal{H}}(\|Tu\|/\|u\|)$. Moreover, if T is self-adjoint, i.e. $(Tu,v) = (u,Tv)$ for all $u,v \in \mathcal{H}$, then $\sigma(T) \subset \mathbb{R}$, so that $\sigma(T) \subset [-\|T\|,\|T\|]$.

A *resolution of the identity* E of the Hilbert space $\mathcal{H}$ is a map $E\colon \mathbb{R} \to \mathcal{B}(\mathcal{H})$ such that for any Borel sets $A,B \subseteq \mathbb{R}$ we have (i) $E(A)$ is a self-adjoint projection, i.e. $E(A)^2 = E(A)$, (ii) $E(A\cap B) = E(A)E(B)$, (iii) $E(\emptyset) = 0, E(\mathbb{R}) = I_{\mathcal{H}}$, (iv) $A\cap B = \emptyset$ implies $E(A\cup B) = E(A) + E(B)$ and (v) for all $u,v \in \mathcal{H}$ the map $A \mapsto E_{u,v}(A) = (E(A)u,v)$ is a complex Borel measure. The spectral measure for orthogonal polynomials will be constructed from the map $A \mapsto E_{e_0,e_0}(A) = (E(A)e_0,e_0)$, where e_0 is the first canonical vector of the space $\ell^2(\mathbb{N}_0)$, as we will see below.

Theorem 1.15 (Spectral theorem) *Let $T\colon \mathcal{H} \to \mathcal{H}$ be a bounded self-adjoint linear operator. Then there exists a unique resolution of the identity E of $\mathcal{H}$ such that $T = \int_{\mathbb{R}} tdE(t)$, i.e.*

$$(Tu,v) = \int_{\mathbb{R}} tdE_{u,v}(t).$$

Moreover, E is supported on the spectrum $\sigma(T)$ and any of the spectral projections $E(A), A \subset \mathbb{R}$, commutes with T.

Proof See [127, Section 12.22] or also [51, 125]. □

For any continuous function f defined on the spectrum $\sigma(T)$, we can define the operator $f(T) = \int_{\mathbb{R}} f(t)dE(t)$, i.e. $(f(T)u,v) = \int_{\mathbb{R}} f(t)dE_{u,v}(t)$. Then $f(T)$ is a bounded operator with norm $\|f(T)\| = \sup_{x\in\sigma(T)} |f(x)|$. In particular this can be applied to $f(x) = 1/(x-z), z \in \rho(T)$ and $f(T)(z) = R(z)$, the resolvent operator. The spectral measure E can be obtained from the resolvent operator using the Perron–Stieltjes inversion formula. Indeed, for an open interval $(a,b) \subset \mathbb{R}$, we have that

$$E_{u,v}\left((a,b)\right) = \lim_{\delta\downarrow 0}\lim_{\varepsilon\downarrow 0} \frac{1}{2\pi i}\int_{a+\delta}^{b-\delta} \left[(R(x+i\varepsilon)u,v) - (R(x-i\varepsilon)u,v)\right] dx.$$

Compare with the Perron–Stieltjes inversion formula in (1.9). For unbounded linear operators there is also a spectral theorem, but it is a little bit more technical than the bounded case (see [27, 109, 125]).

Let J be the symmetric tridiagonal Jacobi operator (1.18) with $b_n \in \mathbb{R}, a_n > 0$, $n \geq 0$ and assume that these coefficients are bounded. J is an operator defined on the Hilbert space $\ell^2(\mathbb{N}_0)$ given by