In this dissertation, we introduce analytical tools and numerical machinery for computing with certain classes of Generalized Prolate Spheroidal Functions (GPSFs). Classical Prolate Spheroidal Wave Functions (PSWFs) are a natural and effective tool for computing with bandlimited functions defined on the interval. Slepian et al. demonstrated that GPSFs extend this apparatus to higher dimensions. While the analytical and numerical machinery in one dimension is fairly complete, the situation in higher dimensions is less satisfactory. In order to adequately address the challenges of computing with GPSFs, it is first necessary to efficiently and stably evaluate a certain class of Zernike polynomials, a natural basis for representing smooth functions on the unit ball. Thus, we start with developing the requisite preliminary analytical and numerical apparatus before constructing algorithms and analytical tools for computing with bandlimited functions and GPSFs in higher dimensions. We present the theory and numerical schemes constructed for both Zernike polynomials and GPSFs. In the process of developing these tools, we observed that similar techniques could be used for the evaluation of another family of special functions, Incomplete Gamma Functions. The resulting algorithms (see [14]) are also included in this work.

On generalized prolate spheroidal functions

Philip Greengard[†][⋆], Technical Report YALEU/DCS/TR-1547
April 24, 2019

[⋆] This author's work was supported in part under ONR N00014-14-1-0797, AFOSR FA9550-16-0175, and NIH 1R01HG008383.

[†] Dept. of Mathematics, Yale University, New Haven, CT 06511

Keywords: *Generalized Prolate Spheroidal Functions, Zernike polynomials, Special functions, Function Approximation, Incomplete Gamma Function*

Abstract

On Generalized Prolate Spheroidal Functions

Philip Greengard

2019

In this dissertation, we introduce analytical tools and numerical machinery for computing with certain classes of Generalized Prolate Spheroidal Functions (GPSFs). Classical Prolate Spheroidal Wave Functions (PSWFs) are a natural and effective tool for computing with bandlimited functions defined on the interval. Slepian et al. demonstrated that GPSFs extend this apparatus to higher dimensions. While the analytical and numerical machinery in one dimension is fairly complete, the situation in higher dimensions is less satisfactory. In order to adequately address the challenges of computing with GPSFs, it is first necessary to efficiently and stably evaluate a certain class of Zernike polynomials, a natural basis for representing smooth functions on the unit ball. Thus, we start with developing the requisite preliminary analytical and numerical apparatus before constructing algorithms and analytical tools for computing with bandlimited functions and GPSFs in higher dimensions. We present the theory and numerical schemes constructed for both Zernike polynomials and GPSFs. In the process of developing these tools, we observed that similar techniques could be used for the evaluation of another family of special functions, Incomplete Gamma Functions. The resulting algorithms (see [14]) are also included in this work.

On Generalized Prolate Spheroidal Functions

A Dissertation
Presented to the Faculty of the Graduate School
of
Yale University
in Candidacy for the Degree of
Doctor of Philosophy

by

Philip Greengard

Dissertation Director: Vladimir Rokhlin

May 2019

Contents

1 Introduction 16

2 Zernike Polynomials: Evaluation, Quadrature, and Interpolation 18

2.1 Background 18

2.2 Mathematical Preliminaries 19

2.2.1 Jacobi Polynomials 21

2.2.2 Gaussian Quadratures 24

2.2.3 Zernike Polynomials 25

2.3 Numerical Evaluation of Zernike Polynomials 27

2.4 Quadrature for Zernike Polynomials 28

2.5 Approximation of Zernike Polynomials 33

2.6 Numerical Experiments 35

2.7 Appendix A: Mathematical Properties of
Zernike Polynomials 42

2.7.1 Special Values 44

2.7.2 Hypergeometric Function 45

2.7.3 Interrelations 45

2.7.4 Limit Relations 46

2.7.5 Zeros 46

2.7.6 Inequalities 46

2.7.7 Integrals 47

2.7.8 Generating Function 47

2.7.9 Differential Equation 47

2.7.10 Recurrence Relations 48

2.7.11 Differential Relations 49

2.8 Appendix B: Numerical Evaluation of Zernike Polynomials in $\mathbb{R}^{p+2}$ 49

2.9 Appendix C: Spherical Harmonics in $\mathbb{R}^{p+2}$ 51

2.10 Appendix D: The Shifted Jacobi Polynomials $P_n^{(k,0)}(2x-1)$ 53

2.10.1 Numerical Evaluation of the Shifted Jacobi Polynomials . . 54

2.10.2 Prüfer Transform 55

2.10.3 Roots of the Shifted Jacobi Polynomials 58

2.11 Appendix E: Notational Conventions for

Zernike Polynomials 61

2.11.1 Zernike Fringe Polynomials 62

2.11.2 ANSI Standard Zernike Polynomials 62

2.11.3 Wyant and Creath Notation 63

3 GPSFs 68

3.1 Background 68

3.2 Mathematical and Numerical Preliminaries 69

3.2.1 Jacobi Polynomials 70

3.2.2 Zernike Polynomials 73

3.2.3 Numerical Evaluation of Zernike Polynomials 75

3.2.4 Modified Zernike polynomials, $\overline{T}_{N,n}$ 76

3.2.5 Prüfer Transform 81

3.2.6 Miscellaneous Analytical Facts 84

3.2.7 Spherical Harmonics in $\mathbb{R}^{p+2}$ 89

3.2.8 Generalized Prolate Spheroidal Functions 92

3.3 Analytical Apparatus 100

3.3.1 Properties of GPSFs 100

3.3.2 Decay of the Expansion Coefficients of GPSFs in Zernike Polynomials 101

3.3.3 Tridiagonal Nature of $L_{N,c}$ 103

3.4 Numerical Evaluation of GPSFs 106

3.4.1 Numerical Evaluation of the Single Eigenvalue $\beta_{N,i}$ 107

3.4.2 Numerical Evaluation of Eigenvalues $\beta_{N,0}, \beta_{N,1}, ..., \beta_{N,k}$ . . . 110

3.5 Quadratures for Band-limited Functions 115

3.5.1 Roots of $\Phi_{0,n}$ 121

3.5.2 Gaussian Quadratures for $\Phi_{0,n}$ 125

3.6 Interpolation via GPSFs 128

3.6.1 Dimension of the Class of Bandlimited Functions 133

3.7 Numerical Experiments 137

3.8 Miscellaneous Properties of GPSFs 145

3.8.1 Properties of the Derivatives of GPSFs 145

3.8.2 Derivatives of GPSFs and Corresponding Eigenvalues With Respect to c 150

3.9 Appendix A 150

3.9.1 Derivation of the Integral Operator Q_c 150

4 An Algorithm for the Evaluation of the Incomplete Gamma Function 152

4.1 Background 152

4.2 Preliminaries 153

4.3 Mathematical Apparatus 157

4.4 $P(m+1,x)$ for Small x 164

4.5 $P(m+1,x)$ for Large x 165

4.6 Evaluation of $P(m+1,x)$ for Small m and Intermediate x via Summation 169

4.7 Evaluation of $P(m+1,x)$ for Large m and Intermediate x via Asymptotic Expansion 171

4.8 Description of Algorithm 178

4.8.1 Regime 1: $-1 < m \leq 10,000$ 179

4.8.2 Regime 2: $m > 10,000$ 179

4.9 Numerical Experiments 182

List of Figures

2.1 An illustration of locations of Zernike polynomial quadrature nodes with 20 radial nodes and 40 angular nodes. 31

2.2 Magnitudes of coefficients of interpolation of $f_1(x,y) = (1+25(x^2+y^2))^{-1}$ for $N = 0$ 40

2.3 Coefficients of the Zernike expansion for $N = 10$ of $J_{10}(10r)\cos(10\theta)$ using Chebyshev and Zernike interpolation in the radial direction with 81 points in the angular direction and 41 points in the radial direction. 41

2.4 The Wyant and Creath polynomials $Q^{|m|}_{(n+|m|)/2}$ 63

2.5 Zernike polynomials $\overset{*}{Z}{}^m_n$ in standard double index notation 64

2.6 The polynomials $R_{|m|,(n-|m|)/2}$ 64

2.7 Fringe Zernike Polynomial Ordering 65

2.8 ANSI Standard Zernike Polynomial Ordering 65

3.1 Eigenvalues of F_c (see (3.97)) for $c = 100$ and $p = 0$ 139

3.2 Eigenvalues of F_c (see (3.97)) for $c = 50$ and $p = 1$ 139

3.3 Plots of GPSFs $\Phi_{0,n}$ (see (3.111)) with $c = 50$ and $p = 1$ 140

3.4 Plots of GPSFs $\Phi_{0,n}$ (see (3.111)) with $c = 100$ and $p = 0$ 140

3.5 Plots of GPSFs $\Phi_{10,n}$ (see (3.111)) with $c = 50$ and $p = 1$ 141

3.6 Plots of GPSFs $\Phi_{25,n}$ (see (3.111)) with $c = 100$ and $p = 0$ 141

4.1 An illustrative domain of S_m, the $m^{2/3}$-neighborhood of the interval on the real line $(-m^{2/3}, m^{2/3})$. 173

4.2 $\log_{10}$ of relative errors for evaluation of $P(m+2,x)$ via k-term asymptotic expansion (4.105) for different x and for $k=10$, $k=15$, and $k=20$ 185

4.3 CPU times for evaluation of $P(m+1,x)$ via k-term asymptotic expansion (4.105) for different k and $m+1=x=10^7$ 186

4.4 $\log_{10}$ of the magnitude of the k^{th} term of asymptotic expansion (4.105) for different m 187

4.5 CPU times for evaluation of $P(m+1,x)$ by direct summation (4.83) and by asymptotic expansion (4.105) 188

4.6 CPU times of evaluation of $P(m+1,x)$ via direct summation (4.83) for different x 189

List of Tables

2.1 Locations in the radial and angular directions of Zernike polynomial quadrature nodes with 40 angular nodes and 20 radial nodes. . . . 32

2.2 Quadratures for $f_1(x, y) = (1 + 25(x^2 + y^2))^{-1}$ over the unit disk several different numbers of nodes 38

2.3 Quadratures for $f_2(r, \theta) = J_{100}(150r)\cos(100\theta)$ using several different numbers of nodes 39

2.4 Quadratures for $f_3(x, y) = P_8(x)P_{12}(y)$ (see (2.58)) using several different numbers of nodes 40

2.5 Coefficients of the interpolation of the function $f_4(x, y) = P_2(x)P_4(y)$ into Zernike polynomials of degree at most 8. The entry corresponding to N, n is the coefficient of $R_{N,n}\cos(N\theta)$. 41

2.6 Zernike Fringe Polynomials. This table lists the first 24 Zernike polynomials in what is sometimes called the "Fringe", "Air Force", or "University of Arizona" ordering (see, for example [23], p. 198, or [30], p. 31). They are often also denoted by $Z_\ell(\rho, \theta)$, where ℓ is the index. $\diamond$ See formulas (2.27) and (2.30). † See, for example, [15]. More complex aberrations are usually not named. 66

2.7 ANSI Standard Zernike Polynomials. This table lists the first 24 Zernike polynomials in the ANSI Standard ordering, also referred to as the "OSA Standard" or "Noll" ordering (see, for example [23], p. 201, or [20]). They are often also denoted by $Z_\ell(\rho, \theta)$, where ℓ is the index. ◇ See formulas (2.27) and (2.30). † See, for example, [15]. More complex aberrations are usually not named. 67

3.1 Quadratures for $e^{ic\langle x,t\rangle}$ where $x = (0.9, 0.2)$ over the unit disk using several different numbers of radial nodes for $c = 20$. Chebyshev quadratures are used in the radial direction. 139

3.2 Quadratures for $e^{ic\langle x,t\rangle}$ where $x = (0.9, 0.2)$ over the unit disk using several different numbers of angular nodes for $c = 20$. Chebyshev quadratures are used in the radial direction. 140

3.3 Quadratures for $e^{ic\langle x,t\rangle}$ where $x = (0.9, 0.2)$ over the unit disk using several different numbers of radial nodes for $c = 20$. Gaussian quadratures generated via Algorithm 3.5.2 are used in the radial direction. 141

3.4 Quadratures for $e^{ic\langle x,t\rangle}$ where $x = (0.9, 0.2)$ over the unit disk using several different numbers of radial nodes for $c = 100$. Chebyshev quadratures are used in the radial direction. 142

3.5 Quadratures for $e^{ic\langle x,t\rangle}$ where $x = (0.9, 0.2)$ over the unit disk using several different numbers of angular nodes for $c = 100$. Chebyshev quadratures are used in the radial direction. 142

3.6 Quadratures for $e^{ic\langle x,t\rangle}$ where $x = (0.9, 0.2)$ over the unit disk using several different numbers of radial nodes for $c = 100$. Gaussian quadratures generated using Algorithm 3.5.2 are used in the radial direction. 142

3.7 Coefficients, obtained via formula (3.245), of the GPSF expansion of the function on the unit disk $e^{ic\langle x,t\rangle}$ where $x = (0.3, 0.4)$. 143

3.8 Coefficients, obtained via formula (3.245), of the GPSF expansion of the function on the unit disk $e^{ic\langle x,t\rangle}$ where $x = (0.3, 0.4)$. 144

3.9 Coefficients, obtained via formula (3.245), of the GPSF expansion of the function on the unit disk $e^{ic\langle x,t\rangle}$ where $x = (0.3, 0.4)$. 145

4.1 Bounds on the error of evaluating $P(m+1,x)$ via sum (4.83) using k terms for different m, x, k. 172

4.2 Relative errors for the evaluation of $P(m+1,x)$ via 10-term asymptotic expansion (4.105) for different x 185

4.3 CPU times for the evaluation of $P(m+1,x)$ via k-term asymptotic expansion (4.105) for different k 186

4.4 Numerical results for $\log_{10}$ of the magnitude of the k^{th} term of asymptotic expansion (4.105) and errors of the k-term expansion . . 187

4.5 CPU times and errors for the evaluation of $P(m,m)$ by direct summation (4.83) for different m 188

4.6 CPU times for the evaluation of $P(m+1,x)$ by direct summation (4.83) for different x 189

4.7 Absolute errors for the evaluation of $P(m+1,x)$ by direct summation (4.83) and asymptotic expansion (4.105) in extended precision 190

Acknowledgements

I am immensely grateful to my dissertation advisor, Vladimir Rokhlin. He has taught me a huge amount about mathematics, applied mathematics, computing, programming, the research process, pedagogy, and many other subjects. He is a phenomenally good teacher: clear, concise, and kind.

Mark Tygert was highly enjoyable to work with. Over daily conversations, he explained to me a vast amount about a wide range of subjects. He has provided me with great advice and was an incredibly supportive manager.

Manas Rachh has continually taught me useful things from the day I arrived at Yale. Kirill Serkh was a great collaborator and explained a lot to me about programming and numerical analysis.

Thank you to my family for all of their help, both technical and otherwise.

Special thanks are due to Leslie Greengard, Jeremy Hoskins, and Jonathan Goodman.

Finally, thank you to Ronald Coifman and Mike O'Neil for being readers of this dissertation.

Chapter 1

Introduction

Classical Prolate spheroidal wave functions (PSWFs) provide a natural and effective tool for computing with bandlimited functions defined on an interval (see [25]). As demonstrated by Slepian et al. in [26], certain classes of generalized prolate spheroidal functions (GPSFs) extend this apparatus to higher dimensions. While the analytical and numerical apparatus in one dimension is fairly complete (see, for example, [31] and [22]), the situation in higher dimensions is less satisfactory. In order to adequately address the challenges of computing with bandlimited functions in higher dimensions, we first constructed numerical tools for stably and efficiently evaluating Zernike polynomials, a basis used to represent smooth functions on the unit ball. In Chapter 2, we describe algorithms for evaluating Zernike polynomials as well as analysis and numerical tools for integrating and interpolating with Zernike polynomials in two and higher dimensions.

Using the theory and numerical schemes involving Zernike polynomials from Chapter 2, we introduce theory and numerical tools for computing with GPSFs and bandlimited functions in Chapter 3. We present algorithms for evaluating GPSFs and their corresponding eigenvalues. We also introduce numerical schemes for integrating and interpolating bandlimited functions defined on the unit ball in higher dimensions.

In constructing tools for Zernike polynomials and GPSFs, we observed that similar techniques could be used for the evaluation of Incomplete Gamma Functions, $P(m, x)$ for $m, x > 0$. In the fourth and final chapter, we introduce these algorithms. The number of operations required for evaluation is $O(1)$ for all x and m. Nearly full double and extended precision accuracies are achieved in their respective environments and the performance of the scheme is illustrated via several numerical examples.

Chapter 2

Zernike Polynomials: Evaluation, Quadrature, and Interpolation

2.1 Background

Zernike polynomials are a family of orthogonal polynomials that are a natural basis for the approximation of smooth functions on the unit disk. Among other applications, they are widely used in optics and atmospheric sciences and are the natural basis for representing Generalized Prolate Spheroidal Functions (see [26]).

In this chapter, we provide a self-contained reference on Zernike polynomials, including tables of properties, an algorithm for their evaluation, and what appear to be new numerical schemes for quadrature and interpolation. We also introduce properties of Zernike polynomials in higher dimensions and several classes of numerical algorithms for Zernike polynomial discretization in $\mathbb{R}^n$. The quadrature and interpolation schemes provided use a tensor product of equispaced nodes in the angular direction and roots of certain Jacobi polynomials in the radial direction. An algorithm for the evaluation of these roots is also introduced.

The structure of this chapter is as follows. In Section 2.2 we introduce several technical lemmas and provide basic mathematical background that will be used in subsequent sections. In Section 2.3 we provide a recurrence relation for the evaluation of Zernike polynomials. Section 2.4 describes a scheme for integrating Zernike polynomials over the unit disk. Section 2.5 contains an algorithm for the interpolation of Zernike polynomials. In Section 2.6 we give results of numerical experiments with the quadrature and interpolation schemes introduced in the preceding sections. In Appendix A, we describe properies of Zernike polynomials in $\mathbb{R}^n$. Appendix B contains a description of an algorithm for the evaluation of Zernike polynomials in $\mathbb{R}^n$. Appendix C includes an description of Spherical Harmonics in higher dimensions. In Appendix D, an overview is provided of the family of Jacobi polynomials whose roots are used in numerical algorithms for high-dimensional Zernike polynomial discretization. Appendix D also includes a description of an algorithm for computing their roots. Appendix E contains notational conventions for Zernike polynomials.

2.2 Mathematical Preliminaries

In this section, we introduce notation and several technical lemmas that will be used in subsequent sections.

For notational convenience and ease of generalizing to higher dimensions, we will be denoting by $S_N^\ell(\theta) : \mathbb{R} \to \mathbb{R}$, the function defined by the formula

$$S_N^\ell(\theta) = \begin{cases} (2\pi)^{-1/2} & \text{if } N = 0, \\ \sin(N\theta)/\sqrt{\pi} & \text{if } \ell = 0,\ N > 0, \\ \cos(N\theta)/\sqrt{\pi} & \text{if } \ell = 1,\ N > 0. \end{cases} \tag{2.1}$$

where $\ell \in \{0, 1\}$, and N is a non-negative integer. In accordance with standard practice, we will denoting by $\delta_{i,j}$ the function defined by the formula

$$\delta_{i,j} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases} \tag{2.2}$$

The following lemma is a classical fact from elementary calculus.

Lemma 2.2.1 *For all $n \in \{1, 2, ...\}$ and for any integer $k \geq n + 1$,*

$$\frac{1}{k}\sum_{i=1}^{k} \sin(n\theta_i) = \int_0^{2\pi} \sin(n\theta) d\theta = 0 \tag{2.3}$$

and

$$\frac{1}{k}\sum_{i=1}^{k} \cos(n\theta_i) = \int_0^{2\pi} \cos(n\theta) d\theta = 0 \tag{2.4}$$

where

$$\theta_i = i\frac{2\pi}{k} \tag{2.5}$$

for $i = 1, 2, ..., k$.

The following technical lemma will be used in Section 2.4.

Lemma 2.2.2 *For all $m \in \{0, 1, 2, ...\}$, the set of all points $(N, n, \ell) \in \mathbb{R}^3$ such that $\ell \in \{0, 1\}$, N, n are non-negative integers, and $N + 2n \leq 2m - 1$ contains exactly $2m^2 + 2m$ elements.*

Proof. Lemma 2.2.2 follows immediately from the fact that the set of all pairs of non-negative integers (N, n) satisfying $N + 2n \leq 2m - 1$ has $m^2 + m$ elements where m is a non-negative integer. ■

The following is a classical fact from elementary functional analysis. A proof can be found in, for example, [27].

Lemma 2.2.3 *Let* $f_1, ..., f_{2n-1} : [a, b] \to \mathbb{R}$ *be a set of orthonormal functions such that for all* $k \in \{1, 2, ..., 2n-1\}$,

$$\int_a^b f_k(x)dx = \sum_{i=1}^{n} f_k(x_i)\omega_i dx \tag{2.6}$$

where $x_i \in [a, b]$ *and* $\omega_i \in \mathbb{R}$. *Let* $\phi : [a, b] \to \mathbb{R}$ *be defined by the formula*

$$\phi(x) = a_1 f_1(x) + ... + a_{n-1} f_{n-1}(x). \tag{2.7}$$

Then,

$$a_k = \int_a^b \phi(x) f_k(x)dx = \sum_{i=1}^{n} \phi(x_i) f_k(x_i)\omega_i. \tag{2.8}$$

for all $k \in \{1, 2, ..., n-1\}$.

2.2.1 Jacobi Polynomials

In this section, we define Jacobi polynomials and summarize some of their properties.

Jacobi Polynomials, denoted $P_n^{(\alpha,\beta)}$, are orthogonal polynomials on the interval $(-1, 1)$ with respect to weight function

$$w(x) = (1-x)^{\alpha}(1+x)^{\beta}. \tag{2.9}$$

Specifically, for all non-negative integers n, m with $n \neq m$ and real numbers $\alpha, \beta > -1$,

$$\int_{-1}^{1} P_n^{(\alpha,\beta)}(x) P_m^{(\alpha,\beta)}(x)(1-x)^{\alpha}(1+x)^{\beta} dx = 0 \tag{2.10}$$

The following lemma, provides a stable recurrence relation that can be used to evaluate a particular class of Jacobi Polynomials (see, for example, [1]).

Lemma 2.2.4 *For any integer $n \geq 1$ and $N \geq 0$,*

$$\begin{aligned} P_{n+1}^{(N,0)}(x) = {} & \frac{(2n+N+1)N^2 + (2n+N)(2n+N+1)(2n+N+2)x}{2(n+1)(n+N+1)(2n+N)} P_n^{(N,0)}(x) \\ & - \frac{2(n+N)(n)(2n+N+2)}{2(n+1)(n+N+1)(2n+N)} P_{n-1}^{(N,0)}(x), \end{aligned} \tag{2.11}$$

where

$$P_0^{(N,0)}(x) = 1 \tag{2.12}$$

and

$$P_1^{(N,0)}(x) = \frac{N + (N+2)x}{2}. \tag{2.13}$$

The Jacobi Polynomial $P_n^{(N,0)}$ is defined in (2.10).

The following lemma provides a stable recurrence relation that can be used to evaluate derivatives of a certain class of Jacobi Polynomials. It is readily obtained by differentiating (2.11) with respect to x,

Lemma 2.2.5 *For any integer $n \geq 1$ and $N \geq 0$,*

$$\begin{aligned} P_{n+1}^{(N,0)\prime}(x) = {} & \frac{(2n+N+1)N^2 + (2n+N)(2n+N+1)(2n+N+2)x}{2(n+1)(n+N+1)(2n+N)} P_n^{(N,0)\prime}(x) \\ & - \frac{2(n+N)(n)(2n+N+2)}{2(n+1)(n+N+1)(2n+N)} P_{n-1}^{(N,0)\prime}(x) \\ & + \frac{(2n+N)(2n+N+1)(2n+N+2)}{2(n+1)(n+N+1)(2n+N)} P_n^{(N,0)}(x), \end{aligned} \tag{2.14}$$

where

$$P_0^{(N,0)\prime}(x) = 0 \tag{2.15}$$

and

$$P_1^{(N,0)\prime}(x) = \frac{(N+2)}{2}. \tag{2.16}$$

The Jacobi Polynomial $P_n^{(N,0)}$ is defined in (2.10) and $P_n^{(N,0)\prime}(x)$ denotes the derivative of $P_n^{(N,0)}(x)$ with respect to x.

The following lemma, which provides a differential equation for Jacobi polynomials, can be found in [1]

Lemma 2.2.6 *For any integer n,*

$$(1-x^2)P_n^{(k,0)\prime\prime}(x) + (-k-(k+2)x)P_n^{(k,0)\prime}(x) + n(n+k+1)P_n^{(k,0)}(x) = 0 \tag{2.17}$$

for all $x \in [0,1]$ where $P_n^{(N,0)}$ is defined in (2.10).

Remark 2.2.1 *We will be denoting by $\widetilde{P}_n : [0,1] \to \mathbb{R}$ the shifted Jacobi polynomial defined for any non-negative integer n by the formula*

$$\widetilde{P}_n(x) = \sqrt{2n+2}P_n^{(1,0)}(1-2x) \tag{2.18}$$

where $P_n^{(1,0)}$ is defined in (2.10). The roots of $\widetilde{P}_n$ will be used in Section 2.4 and Section 2.5 in the design of quadrature and interpolation schemes for Zernike polynomials.

It follows immediately from the combination of (2.10) and (2.18) that the polynomials $\widetilde{P}_n$ are orthogonal on $[0, 1]$ with respect to weight function

$$w(x) = x. \tag{2.19}$$

That is, for any non-negative integers i, j,

$$\int_0^1 \widetilde{P}_i(r)\widetilde{P}_j(r)rdr = \delta_{i,j}. \tag{2.20}$$

2.2.2 Gaussian Quadratures

In this section, we introduce Gaussian Quadratures.

Definition 2.2.1 *A Gaussian Quadrature with respect to a set of functions*

$$f_1, ..., f_{2n-1} : [a, b] \to \mathbb{R} \tag{2.21}$$

and non-negative weight function $w : [a, b] \to \mathbb{R}$ is a set of n nodes, $x_1, ..., x_n \in [a, b]$, and n weights, $\omega_1, ..., \omega_n \in \mathbb{R}$, such that, for any integer $j \leq 2n - 1$,

$$\int_a^b f_j(x)w(x)dx = \sum_{i=0}^{n} \omega_i f_j(x_i). \tag{2.22}$$

The following is a well-known lemma from numerical analysis. A proof can be found in, for example, [27].

Lemma 2.2.7 *Suppose that $p_0, p_1, ... : [a, b] \to \mathbb{R}$ is a set of orthonormal polynomials with respect to some non-negative weight function $w : [a, b] \to \mathbb{R}$ such that* *polynomial p_i is of degree i. Then,*

i) Polynomial p_i has exactly i roots on $[a, b]$.

ii) For any non-negative integer n and for $i = 0, 1, ..., 2n - 1$, we have

$$\int_a^b p_i(x)w(x)dx = \sum_{k=1}^{n} \omega_k p_i(x_k) \tag{2.23}$$

where $x_1, ..., x_n \in [a, b]$ are the n roots of p_n and where weights $\omega_1, ..., \omega_n \in \mathbb{R}$ solve the $n \times n$ system of linear equations

$$\sum_{k=1}^{n} \omega_k p_j(x_k) = \int_a^b w(x)p_j(x)dx \tag{2.24}$$

with $j = 0, 1, ..., n - 1$.

iii) The weights, ω_i, satisfy the identity,

$$\omega_i = \left(\sum_{k=0}^{n-1} p_k(x_i)^2 \right)^{-1} \tag{2.25}$$

for $i = 1, 2, ..., n$.

2.2.3 Zernike Polynomials

In this section, we define Zernike Polynomials and describe some of their basic properties.

Zernike polynomials are a family of orthogonal polynomials defined on the unit ball in $\mathbb{R}^n$. In this chapter, we primarily discuss Zernike polynomials in $\mathbb{R}^2$, however nearly all of the theory and numerical machinery in two dimensions generalizes naturally to higher dimensions. The mathematical properties of Zernike polynomials in $\mathbb{R}^n$ are included in Appendix A.

Zernike Polynomials are defined via the formula

$$Z^{\ell}_{N,n}(x) = R_{N,n}(r) S^{\ell}_N(\theta) \tag{2.26}$$

for all $x \in \mathbb{R}^2$ such that $\|x\| \le 1$, (r,θ) is the representation of x in polar coordinates, N, n are non- negative integers, S^{ℓ}_N is defined in (2.1), and $R_{N,n}$ are polynomials of degree $N + 2n$ defined by the formula

$$R_{N,n}(x) = x^N \sum_{k=0}^{n} (-1)^k \binom{n+N+\frac{p}{2}}{k} \binom{n}{k} (x^2)^{n-k} (1-x^2)^k, \tag{2.27}$$

for all $0 \le x \le 1$. Furthermore, for any non-negative integers N, n, m,

$$\int_0^1 R_{N,n}(x) R_{N,m}(x) x \, dx = \frac{\delta_{n,m}}{2(2n+N+1)} \tag{2.28}$$

and

$$R_{N,n}(1) = 1. \tag{2.29}$$

We define the normalized polynomials $\overline{R}_{N,n}$ via the formula

$$\overline{R}_{N,n}(x) = \sqrt{2(2n+N+1)} R_{N,n}(x), \tag{2.30}$$

so that

$$\int_0^1 \left(\overline{R}_{N,n}(x)\right)^2 x \, dx = 1, \tag{2.31}$$

where N and n are non-negative integers. We define the normalized Zernike polynomial, $\overline{Z}^{\ell}_{N,n}$, by the formula

$$\overline{Z}_{N,n}(x) = \overline{R}_{N,n}(r) S^{\ell}_{N}(\theta) \tag{2.32}$$

where $x \in \mathbb{R}^2$ satisfies $\|x\| \leq 1$, and N, n are non-negative integers. We observe that $\overline{Z}^{\ell}_{N,n}$ has L^2 norm of 1 on the unit disk.

In an abuse of notation, we use $Z^{\ell}_{N,n}(x)$ and $Z^{\ell}_{N,n}(r, \theta)$ interchangeably where (r, θ) is the polar coordinate representation of $x \in \mathbb{R}^2$.

2.3 Numerical Evaluation of Zernike Polynomials

In this section, we provide a stable recurrence relation (see Lemma 2.3.1) that can be used to evaluate Zernike Polynomials.

Lemma 2.3.1 *The polynomials $R_{N,n}$, defined in (2.27) satisfy the recurrence relation*

$$\begin{aligned} R_{N,n+1}(x) = \\ &- \frac{((2n+N+1)N^2 + (2n+N)(2n+N+1)(2n+N+2)(1-2x^2))}{2(n+1)(n+N+1)(2n+N)} R_{N,n}(x) \\ &- \frac{2(n+N)(n)(2n+N+2)}{2(n+1)(n+N+1)(2n+N)} R_{N,n-1}(x) \end{aligned} \tag{2.33}$$

where $0 \leq x \leq 1$, N is a non-negative integer, n is a positive integer, and

$$R_{N,0}(x) = x^N \tag{2.34}$$

and

$$R_{N,1}(x) = -x^N \frac{N + (N+2)(1-2x^2)}{2}. \tag{2.35}$$

Proof. According to [1], for any non-negative integers n and N,

$$R_{N,n}(x) = (-1)^n x^N P_n^{(N,0)}(1-2x^2), \tag{2.36}$$

where $0 \le x \le 1$, N and n are nonnegative integers, and $P_n^{(N,0)}$ denotes a Jacobi polynomial (see (2.10)).

Identity (2.33) follows immediately from the combination of (2.36) and (2.11).

■

Remark 2.3.1 *The algorithm for evaluating Zernike polynomials using the recurrence relation in Lemma 2.3.1 is known as Kintner's method (see [16] and, for example, [6]).*

2.4 Quadrature for Zernike Polynomials

In this section, we provide a quadrature rule for Zernike Polynomials.

The following lemma follows immediately from applying Lemma 2.2.7 to the polynomials $\widetilde{P}_n$ defined in (2.18).

Lemma 2.4.1 *Let* $\{r_1, ..., r_m\}$ *be the* m *roots of* $\widetilde{P}_m$ *(see (2.18)) and* $\{\omega_1, ..., \omega_m\}$ *the* m *weights of the Gaussian quadrature (see (2.22)) for the polynomials*

$$\widetilde{P}_0, \widetilde{P}_1, ..., \widetilde{P}_{2m-1} \tag{2.37}$$

where $\widetilde{P}_n$ is defined in (2.18). Then, for any polynomial q of degree at most $2m-1$,

$$\int_0^1 q(x)xdx = \sum_{i=1}^{m} q(r_i)\omega_i. \tag{2.38}$$

The following theorem provides a quadrature rule for Zernike Polynomials.

Theorem 2.4.2 *Let $\{r_1, ..., r_m\}$ be the m roots of $\widetilde{P}_m$ (see (2.18)) and $\{\omega_1, ..., \omega_m\}$ the m weights of the Gaussian quadrature (see (2.22)) for the polynomials*

$$\widetilde{P}_0, \widetilde{P}_1, ..., \widetilde{P}_{2m-2}. \tag{2.39}$$

Then, for all $\ell \in \{0, 1\}$ and for all $N, n \in \{0, 1, ...\}$ such that $N + 2n \leq 2m - 1$,

$$\int_D Z^{\ell}_{N,n}(x)dx = \sum_{i=1}^{m} R_{N,n}(r_i)\omega_i \sum_{j=1}^{2m} \frac{2\pi}{2m} S^{\ell}_N(\theta_j) \tag{2.40}$$

where $R_{N,n}$ is defined in (2.27), θ_j is defined by the formula

$$\theta_j = j\frac{2\pi}{2m} \tag{2.41}$$

for $j \in \{1, 2, ..., 2m\}$, and $D \subseteq \mathbb{R}^2$ denotes the unit disk. Furthermore, there are exactly $2m^2 + m$ Zernike Polynomials of degree at most $2m - 1$.

Proof. Applying a change of variables,

$$\int_D Z^{\ell}_{N,n}(x)dx = \int_0^1 \int_0^{2\pi} R_{N,n}(r)S^{\ell}_N(\theta)rdrd\theta, \tag{2.42}$$

where $Z^{\ell}_{N,n}$ is a Zernike polynomial (see (2.26)) and where $R_{N,n}$ is defined in (2.28). Changing the order of integration of (2.42), we obtain

$$\int_D Z^{\ell}_{N,n}(x)dx = \int_0^1 rR_{N,n}(r)dr \int_0^{2\pi} S^{\ell}_N(\theta)d\theta. \tag{2.43}$$

Applying Lemma 2.2.1 and Lemma 2.4.1 to (2.43), we obtain

$$\int_D Z_{N,n}^\ell(x)dx = \sum_{i=1}^{m} R_{N,n}(r_i)\omega_i \sum_{j=1}^{2m} \frac{2\pi}{2m} S_N^\ell(\theta_j) \tag{2.44}$$

for $N+2n \leq 2m-1$. The fact that there are exactly $2m^2+m$ Zernike polynomials of degree at most $2m-1$ follows immediately from the combination of Lemma 2.2.2 with the fact that there are exactly m Zernike polynomials of degree at most $2m-1$ that are of the form $Z_{0,n}^\ell$. ■

Remark 2.4.1 *It follows immediately from Lemma 2.4.2 that for all $m \in \{1, 2, ...\}$, placing m nodes in the radial direction and $2m$ nodes in the angular direction (as described in Lemma 2.4.2), integrates exactly the $2m^2+m$ Zernike polynomials on the disk of degree at most $2m-1$.*

Remark 2.4.2 *The n roots of $\widetilde{P}_n$ (see 2.20) can be found by using, for example, the algorithm described in Section 2.10.3.*

Remark 2.4.3 *For Zernike polynomial discretization in $\mathbb{R}^{k+1}$, roots of the polynomials $\widetilde{P}_n^k$ are used, where $\widetilde{P}_n^k$ is defined by the formula*

$$\widetilde{P}_n^k(x) = \sqrt{k+2n+1} P_n^{(k,0)}(1-2x). \tag{2.45}$$

Properties of this class of Jacobi polynomials are provided in Appendix D in addition to an algorithm for finding their roots.

The following remark illustrates that the advantage of quadrature rule (2.40) is especially noticeable in higher dimensions.

Remark 2.4.4 *Quadrature rule (2.40) integrates all Zernike polynomials up to order $2m-1$ using the m roots of $\widetilde{P}_m$ (see (2.20)) as nodes in the radial direction.*

Using Guass-Legendre nodes instead of roots of $\widetilde{P}_m$ would require using $m+1$ nodes in the radial direction.

The equivalent of quadrature rule (2.40) in $p+2$ dimensions uses the roots of $\widetilde{P}_m^{p+1}$ (see (2.111)) as nodes in the radial direction. Using Gauss-Legendre nodes instead of these nodes would require using an extra $p+1$ nodes in the radial direction or approximately $(p+1)m^{p+1}$ extra nodes total.

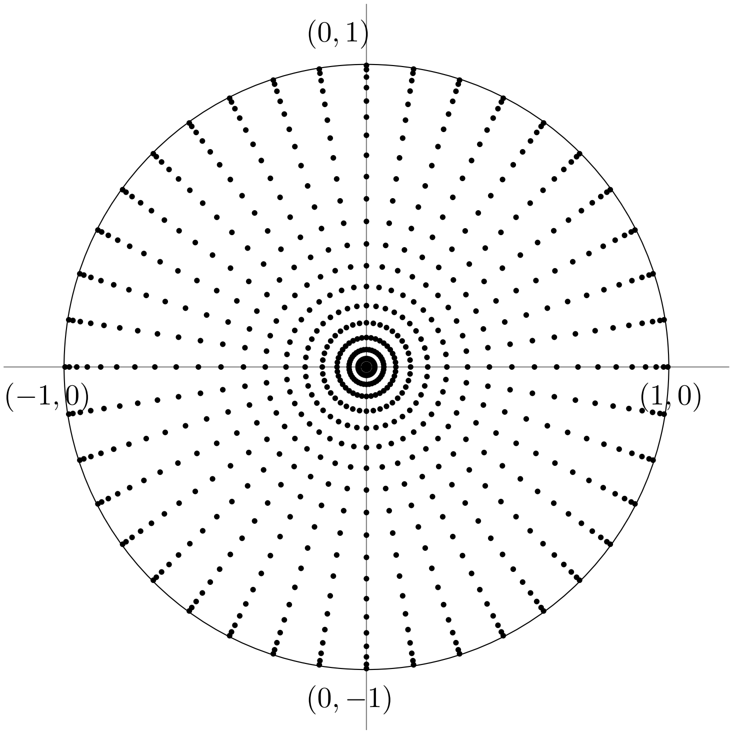

Figure 2.1: An illustration of locations of Zernike polynomial quadrature nodes with 20 radial nodes and 40 angular nodes.

The following remark shows that we can reduce the total number of nodes in quadrature rule (2.40) while still integrating the same number of functions.

Remark 2.4.5 *Quadrature rule (2.40) integrates all Zernike polynomials of order up to $2m-1$ using a tensor product of $2m$ equispaced nodes in the angular direction and the m roots of $\widetilde{P}_m$ (see 2.18) in the radial direction. However, for large enough N and small enough j, $Z_{N,n}(r_j)$ is of magnitude smaller than machine precision,*

node	θ
1	0.0000000000000000
2	0.1570796326794897
3	0.3141592653589793
4	0.4712388980384690
5	0.6283185307179586
6	0.7853981633974483
7	0.9424777960769379
8	1.0995574287564280
9	1.2566370614359170
10	1.4137166941154070
11	1.5707963267948970
12	1.7278759594743860
13	1.8849555921538760
14	2.0420352248333660
15	2.1991148575128550
16	2.3561944901923450
17	2.5132741228718340
18	2.6703537555513240
19	2.8274333882308140
20	2.9845130209103030
21	3.1415926535897930
22	3.2986722862692830
23	3.4557519189487720
24	3.6128315516282620
25	3.7699111843077520
26	3.9269908169872410
27	4.0840704496667310
28	4.2411500823462210
29	4.3982297150257100
30	4.5553093477052000
31	4.7123889803846900
32	4.8694686130641790
33	5.0265482457436690
34	5.1836278784231590
35	5.3407075111026480
36	5.4977871437821380
37	5.6548667764616280
38	5.8119464091411170
39	5.9690260418206070
40	6.1261056745000970

node	r
1	0.0083000442070672
2	0.0276430533525631
3	0.0575344576368137
4	0.0973041282065463
5	0.1460632469641095
6	0.2027224916634053
7	0.2660161417643405
8	0.3345303010944863
9	0.4067344665164935
10	0.4810157112964263
11	0.5557147130369888
12	0.6291628194156031
13	0.6997193231640498
14	0.7658081136864078
15	0.8259528873644578
16	0.8788101326763239
17	0.9231991629103781
18	0.9581285688822349
19	0.9828187818547442
20	0.9967238933309499

Table 2.1: Locations in the radial and angular directions of Zernike polynomial quadrature nodes with 40 angular nodes and 20 radial nodes.

where r_j *denotes the* j^{th} *smallest root of* $\widetilde{P}_m$*. As a result, in order to integrate exactly* $Z_{N,n}$ *for large* N*, we can use fewer equispaced nodes in the angular direction* *at radius r_j.*

2.5 Approximation of Zernike Polynomials

In this section, we describe an interpolation scheme for Zernike Polynomials.

We will denote by $r_1, ..., r_M$ the M roots of $\widetilde{P}_M$ (see 2.18).

Theorem 2.5.1 *Let M be a positive integer and $f : D \to \mathbb{R}$ be a linear combination of Zernike polynomials of degree at most $M-1$. That is,*

$$f(r,\theta) = \sum_{i,j} \alpha_{i,j}^{\ell} \overline{Z}_{i,j}^{\ell}(r,\theta) \tag{2.46}$$

where i, j are non-negative integers satisfying

$$i + 2j \leq M - 1 \tag{2.47}$$

and where $\overline{Z}_{i,j}^{\ell}(r,\theta)$ is defined by (2.32) and S_i^{ℓ} is defined by (2.1). Then,

$$\alpha_{i,j}^{\ell} = \sum_{k=1}^{M} \left[\overline{R}_{i,j}(r_k)\omega_k \sum_{l=1}^{2M-1} \frac{2\pi}{2M-1} f(r_k,\theta_l) S_i^{\ell}(\theta_l) \right] \tag{2.48}$$

where $r_1, ..., r_M$ denote the M roots of $\widetilde{P}_M$ (see 2.18) and θ_l is defined by the formula

$$\theta_l = l \frac{2\pi}{2M-1} \tag{2.49}$$

for $l = 1, 2, ..., 2M-1$.

Proof. Clearly,

$$\alpha_{i,j}^{\ell} = \int_D f(r,\theta) \overline{Z}_{i,j}^{\ell} = \int_0^{2\pi} \int_0^1 f(r,\theta) \overline{R}_{i,j}(r) S_i^{\ell}(\theta) r dr d\theta. \tag{2.50}$$

Changing the order of integration of (2.50) and applying Lemma 2.2.1 and Lemma 2.2.3, we obtain

$$\begin{aligned} \alpha_{i,j}^{\ell} &= \int_0^1 \overline{R}_{i,j}(r) r \int_0^{2\pi} f(r,\theta) S_i^{\ell}(\theta) d\theta dr \\ &= \int_0^1 \overline{R}_{i,j}(r) r \sum_{l=1}^{2M-1} \frac{2\pi}{2M-1} f(r,\theta_l) S_i^{\ell}(\theta_l) dr. \end{aligned} \tag{2.51}$$

Applying Lemma 2.2.3 to (2.51), we obtain

$$\alpha_{i,j}^{\ell} = \sum_{k=1}^{M} \left[\overline{R}_{i,j}(r_k) \omega_k \sum_{l=1}^{2M-1} \frac{2\pi}{2M-1} f(r_k,\theta_l) S_{i,j}^{\ell}(\theta_l) \right]. \tag{2.52}$$

■

Remark 2.5.1 *Suppose that $f : D \to \mathbb{R}$ is a linear combination of Zernike polynomials of degree at most $M-1$. It follows immediately from Theorem 2.5.1 and Theorem 2.4.2 that we can recover exactly the $M^2/2 + M/2$ coefficients of the Zernike polynomial expanison of f by evaluation of f at $2M^2 - M$ points via (2.48).*

Remark 2.5.2 *Recovering the $M^2/2 + M/2$ coefficients of a Zernike expansion of degree at most $M-1$ via (2.52) requires $O(M^3)$ operations by using a FFT to compute the sum*

$$\sum_{l=1}^{2M-1} \frac{2\pi}{2M-1} f(r,\theta_l) S_{i,j}^{\ell}(\theta_l) \tag{2.53}$$

and then naively computing the sum

$$\alpha_{i,j}^{\ell} = \sum_{k=1}^{M} \overline{R}_{i,j}(r_k) \omega_k \sum_{l=1}^{2M-1} \frac{2\pi}{2M-1} f(r_k,\theta_l) S_{i,j}^{\ell}(\theta_l). \tag{2.54}$$

Remark 2.5.3 *Sum (2.54) can be computed using an FMM (see, for example, [2]) which would reduce the evaluation of sum (2.52) to a computational cost of* $O(M^2 \log(M))$.

Remark 2.5.4 *Standard interpolation schemes on the unit disk often involve representing smooth functions as expansions in non-smooth functions such as*

$$T_n(r) S_N^\ell(\theta) \tag{2.55}$$

where n *and* N *are non-negative integers,* T_n *is a Chebyshev polynomial, and* S_N^ℓ *is defined in (2.1). Such interpolation schemes are amenable to the use of an FFT in both the angular and radial directions and thus have a computational cost of only* $O(M^2 \log(M))$ *for the interpolation of an* M*-degree Zernike expansion.*

However, interpolation scheme (2.48) has three main advantages over such a scheme:

i) In order to represent a smooth function on the unit disk to full precision, a Zernike expansion requires approximately half as many terms as an expansion into functions of the form (2.55) (see Figure 2.3).

ii) Each function in the interpolated expansion is smooth on the disk.

iii) The expansion is amenable to filtering.

2.6 Numerical Experiments

The quadrature and interpolation formulas described in Sections 2.4 and 2.5 were implemented in Fortran 77. We used the Lahey/Fujitsu compiler on a 2.9 GHz Intel i7-3520M Lenovo laptop. All examples in this section were run in double precision arithmetic.

In each table in this section, the column labeled "nodes" denotes the number of nodes in both the radial and angular direction using quadrature rule (2.40). The column labeled "exact integral" denotes the true value of the integral being tested. This number is computed using adaptive gaussian quadrature in extended precision. The column labeled "integral via quadrature" denotes the integral approximation using quadrature rule (2.40).

We tested the performance of quadrature rule (2.40) in integrating three different functions over the unit disk. In Table 2.2 we approximated the integral over the unit disk of the function f_1 defined by the formula

$$f_1(x, y) = \frac{1}{1 + 25(x^2 + y^2)}. \tag{2.56}$$

In Table 2.3 we use quadrature rule (2.40) to approximate the integral over the unit disk of the function f_2 defined by the formula

$$f_2(r, \theta) = J_{100}(150r) \cos(100\theta)). \tag{2.57}$$

In Table 2.4, we use quadrature rule (2.40) to approximate the integral over the unit disk of the function f_3 defined by the formula

$$f_3(r, \theta) = P_8(x)P_{12}(y). \tag{2.58}$$

We tested the performance of interpolation scheme (2.46) on two functions defined on the unit disk.

In Figure 2.2 we plot the magnitude of the coefficients of the Zernike polynomials $R_{0,n}$ for $n = 0, 1, ..., 10$ using interpolation scheme (2.46) with 21 nodes in the radial direction and 41 in the angular direction on the function f_1 defined in (2.56). All coeficients of other terms were of magnitude smaller than 10^{-14}. In Table 2.5 we list the interpolated coefficients of the Zernike polynomial expansion of the function f_4 defined by the formula

$$f_4(x, y) = P_2(x)P_4(y) \tag{2.59}$$

where P_i is the ith degree Legendre polynomial. Listed are the coefficients using interpolation scheme (2.46) with 5 points in the radial direction and 9 points in the angular direction of Zernike polynomials

$$R_{N,n} \cos(N\theta) \tag{2.60}$$

where $N = 0, 1, ..., 8$ and $n = 0, 1, 2, 3, 4$. All other coefficients were of magnitude smaller than 10^{-14}. We interpolated the Bessel function

$$J_{10}(10r)cos(10\theta) \tag{2.61}$$

using interpolation scheme (2.46) and plot the resulting coefficients of the Zernike polynomials

$$R_{10,n} \cos(10\theta) \tag{2.62}$$

for $n = 0, ..., 16$ in Figure 2.3. All other coefficients were approximately 0 to machine precision. In Figure 2.3, we plot the coefficients of the Chebyshev expansion obtained via Chebyshev interpolation of the radial component of (2.61).

radial nodes	angular nodes	exact integral	integral via quadrature	relative error
5	10	0.4094244859413851	0.4097244673896003	0.732691×10^{-3}
10	20	0.4094244859413851	0.4094251051077367	0.151228×10^{-5}
15	30	0.4094244859413851	0.4094244870531256	0.271537×10^{-8}
20	40	0.4094244859413851	0.4094244859432513	0.455821×10^{-11}
25	50	0.4094244859413851	0.4094244859413883	0.791759×10^{-14}
30	60	0.4094244859413851	0.4094244859413848	0.630994×10^{-15}
35	70	0.4094244859413851	0.4094244859413850	0.142503×10^{-15}
40	80	0.4094244859413851	0.4094244859413858	0.181146×10^{-14}

Table 2.2: Quadratures for $f_1(x, y) = (1 + 25(x^2 + y^2))^{-1}$ over the unit disk several different numbers of nodes

radial nodes	angular nodes	exact integral	integral via quadrature
5	10	0	$0.2670074163846569 \times 10^{-1}$
10	20	0	$0.2606355680939063 \times 10^{-2}$
15	30	0	$0.3119143925398078 \times 10^{-15}$
20	40	0	$0.0000000000000000 \times 10^{0}$
25	50	0	$0.3228321977714574 \times 10^{-1}$
30	60	0	$0.4945592102178045 \times 10^{-16}$
35	70	0	$0.1147861841710902 \times 10^{-16}$
40	80	0	$0.8148891073315595 \times 10^{-16}$
45	90	0	$-0.7432759692263743 \times 10^{-16}$
50	100	0	$0.3207999037057322 \times 10^{-1}$
55	110	0	$-0.1399753743762347 \times 10^{-15}$
60	120	0	$0.3075136040459932 \times 10^{-16}$
65	130	0	$-0.9458788981593222 \times 10^{-16}$
70	140	0	$0.2045957446273746 \times 10^{-17}$
75	150	0	$0.2416178317504225 \times 10^{-16}$

Table 2.3: Quadratures for $f_2(r, \theta) = J_{100}(150r)\cos(100\theta)$ using several different numbers of nodes

radial nodes	angular nodes	integral via quadrature	exact integral	relative error
5	10	$-0.8998055487754142 \times 10^{-2}$	$-0.1527947805159123 \times 10^{-2}$	-0.830191×10^{0}
10	20	$0.1655201967553289 \times 10^{-1}$	$-0.1527947805159123 \times 10^{-2}$	-0.109231×10^{1}
15	30	$-0.1527947805159138 \times 10^{-2}$	$-0.1527947805159123 \times 10^{-2}$	$-0.979221 \times 10^{-14}$
20	40	$-0.1527947805159132 \times 10^{-2}$	$-0.1527947805159123 \times 10^{-2}$	$-0.567665 \times 10^{-14}$
25	50	$-0.1527947805159108 \times 10^{-2}$	$-0.1527947805159123 \times 10^{-2}$	0.102180×10^{-13}
30	60	$-0.1527947805159144 \times 10^{-2}$	$-0.1527947805159123 \times 10^{-2}$	$-0.134820 \times 10^{-13}$
35	70	$-0.1527947805159128 \times 10^{-2}$	$-0.1527947805159123 \times 10^{-2}$	$-0.269641 \times 10^{-14}$
40	80	$-0.1527947805159155 \times 10^{-2}$	$-0.1527947805159123 \times 10^{-2}$	$-0.210036 \times 10^{-13}$

Table 2.4: Quadratures for $f_3(x, y) = P_8(x)P_{12}(y)$ (see (2.58)) using several different numbers of nodes

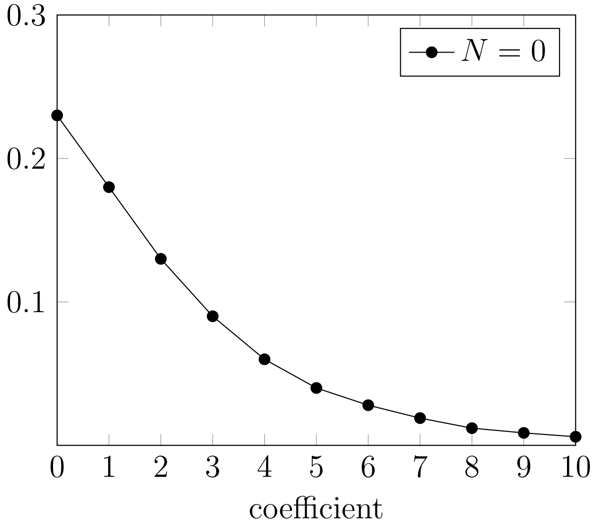

Figure 2.2: Magnitudes of coefficients of interpolation of $f_1(x, y) = (1 + 25(x^2 + y^2))^{-1}$ for $N = 0$

N	$n=0$	$n=1$	$n=2$	$n=3$	$n=4$
0	0.02942	0.03297	-0.11998	0.01373	0.53776×10^{-16}
1	-0.48788×10^{-16}	0.76567×10^{-17}	0.99670×10^{-18}	0.22059×10^{-16}	-
2	0.02967	0.11495	-0.00647	-0.90206×10^{-16}	-
3	0.58217×10^{-16}	-0.73297×10^{-16}	0.19321×10^{-17}	-	-
4	0.04926	-0.03238	-0.13010×10^{-16}	-	-
5	0.77604×10^{-16}	0.10474×10^{-15}	-	-	-
6	0.09714	-0.11102×10^{-15}	-	-	-
7	-0.18100×10^{-16}	-	-	-	-
8	0.77241×10^{-16}	-	-	-	-

Table 2.5: Coefficients of the interpolation of the function $f_4(x, y) = P_2(x)P_4(y)$ into Zernike polynomials of degree at most 8. The entry corresponding to N, n is the coefficient of $R_{N,n} \cos(N\theta)$.

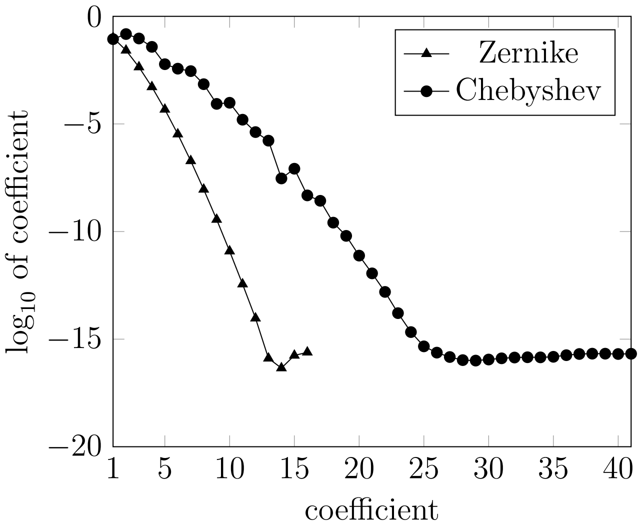

Figure 2.3: Coefficients of the Zernike expansion for $N = 10$ of $J_{10}(10r) \cos(10\theta)$ using Chebyshev and Zernike interpolation in the radial direction with 81 points in the angular direction and 41 points in the radial direction.

2.7 Appendix A: Mathematical Properties of Zernike Polynomials

In this appendix, we define Zernike polynomials in $\mathbb{R}^{p+2}$ and describe some of their basic properties. Zernike polynomials, denoted $Z_{N,n}^{\ell}$, are a sequence of orthogonal polynomials defined via the formula

$$Z_{N,n}^{\ell}(x) = R_{N,n}(\|x\|)S_N^{\ell}(x/\|x\|), \tag{2.63}$$

for all $x \in \mathbb{R}^{p+2}$ such that $\|x\| \leq 1$, where N and n are nonnegative integers, S_N^{ℓ} are the orthonormal surface harmonics of degree N (see Appendix C), and $R_{N,n}$ are polynomials of degree $2n + N$ defined via the formula

$$R_{N,n}(x) = x^N \sum_{m=0}^{n} (-1)^m \binom{n + N + \frac{p}{2}}{m} \binom{n}{m} (x^2)^{n-m}(1 - x^2)^m, \tag{2.64}$$

for all $0 \leq x \leq 1$. The polynomials $R_{N,n}$ satisfy the relation

$$R_{N,n}(1) = 1, \tag{2.65}$$

and are orthogonal with respect to the weight function $w(x) = x^{p+1}$, so that

$$\int_0^1 R_{N,n}(x)R_{N,m}(x)x^{p+1}\,dx = \frac{\delta_{n,m}}{2(2n + N + \frac{p}{2} + 1)}, \tag{2.66}$$

where

$$\delta_{n,m} = \begin{cases} 1 & \text{if } n = m, \\ 0 & \text{if } n \neq m. \end{cases} \tag{2.67}$$

We define the polynomials $\overline{R}_{N,n}$ via the formula

$$\overline{R}_{N,n}(x) = \sqrt{2(2n + N + p/2 + 1)}R_{N,n}(x), \tag{2.68}$$

so that

$$\int_0^1 \big(\overline{R}_{N,n}(x)\big)^2 x^{p+1}\, dx = 1, \tag{2.69}$$

where N and n are nonnegative integers. We define the normalized Zernike polynomial, $\overline{Z}_{N,n}^{\ell}$, by the formula

$$\overline{Z}_{N,n}(x) = \overline{R}_{N,n}(\|x\|)S_N^{\ell}(x/\|x\|) \tag{2.70}$$

for all $x \in \mathbb{R}^{p+2}$ such that $\|x\| \le 1$, where N and n are nonnegative integers, S_N^{ℓ} are the orthonormal surface harmonics of degree N (see Appendix C), and $R_{N,n}$ is defined in (2.64). We observe that $\overline{Z}_{N,n}^{\ell}$ has L^2 norm of 1 on the unit ball in $\mathbb{R}^{p+2}$.

In an abuse of notation, we refer to both the polynomials $Z_{N,n}^{\ell}$ and the radial polynomials $R_{N,n}$ as Zernike polynomials where the meaning is obvious.

Remark 2.7.1 *When $p = -1$, the Zernike polynomials take the form*

$$Z_{0,n}^1(x) = R_{0,n}(|x|) = P_{2n}(x), \tag{2.71}$$

$$Z_{1,n}^2(x) = \operatorname{sgn}(x) \cdot R_{1,n}(|x|) = P_{2n+1}(x), \tag{2.72}$$

for $-1 \le x \le 1$ and nonnegative integer n, where P_n denotes the Legendre polyno-

mial of degree n and

$$\operatorname{sgn}(x) = \begin{cases} 1 & \text{if } x > 0, \\ 0 & \text{if } x = 0, \\ -1 & \text{if } x < 0, \end{cases} \tag{2.73}$$

for all real x.

Remark 2.7.2 *When $p = 0$, the Zernike polynomials take the form*

$$Z^1_{N,n}(x_1, x_2) = R_{N,n}(r)\cos(N\theta)/\sqrt{\pi}, \tag{2.74}$$

$$Z^2_{N,n}(x_1, x_2) = R_{N,n}(r)\sin(N\theta)/\sqrt{\pi}, \tag{2.75}$$

where $x_1 = r\cos(\theta)$, $x_2 = r\sin(\theta)$, and N and n are nonnegative integers.

2.7.1 Special Values

The following formulas are valid for all nonnegative integers N and n, and for all $0 \le x \le 1$.

$$R_{N,0}(x) = x^N, \tag{2.76}$$

$$R_{N,1}(x) = x^N\big((N + p/2 + 2)x^2 - (N + p/2 + 1)\big), \tag{2.77}$$

$$R_{N,n}(1) = 1, \tag{2.78}$$

$$R^{(k)}_{N,n}(0) = 0 \quad \text{for } k = 0, 1, \ldots, N-1, \tag{2.79}$$

$$R^{(N)}_{N,n}(0) = (-1)^n N! \binom{n + N + \frac{p}{2}}{n}. \tag{2.80}$$

2.7.2 Hypergeometric Function

The polynomials $R_{N,n}$ are related to the hypergeometric function ${}_2F_1$ (see [1]) by the formula

$$R_{N,n}(x) = (-1)^n \binom{n+N+\frac{p}{2}}{n} x^N {}_2F_1\Big(-n, n+N+\frac{p}{2}+1; N+\frac{p}{2}+1; x^2\Big), \quad (2.81)$$

where $0 \le x \le 1$, and N and n are nonnegative integers.

2.7.3 Interrelations

The polynomials $R_{N,n}$ are related to the Jacobi polynomials via the formula

$$R_{N,n}(x) = (-1)^n x^N P_n^{(N+\frac{p}{2},0)}(1-2x^2), \quad (2.82)$$

where $0 \le x \le 1$, N and n are nonnegative integers, and $P_n^{(\alpha,\beta)}$, $\alpha, \beta > -1$, denotes the Jacobi polynomials of degree n (see [1]).

When $p = -1$, the polynomials $R_{N,n}$ are related to the Legendre polynomials via the formulas

$$R_{0,n}(x) = P_{2n}(x), \quad (2.83)$$

$$R_{1,n}(x) = P_{2n+1}(x), \quad (2.84)$$

where $0 \le x \le 1$, n is a nonnegative integer, and P_n denotes the Legendre polynomial of degree n (see [1]).

2.7.4 Limit Relations

The asymptotic behavior of the Zernike polynomials near 0 as the index n tends to infinity is described by the formula

$$\lim_{n\to\infty} \frac{(-1)^n R_{N,n}\left(\frac{x}{2n}\right)}{(2n)^{p/2}} = \frac{J_{N+p/2}(x)}{x^{p/2}}, \tag{2.85}$$

where $0 \leq x \leq 1$, N is a nonnegative integer, and J_ν denotes the Bessel functions of the first kind (see [1]).

2.7.5 Zeros

The asymptotic behavior of the zeros of the polynomials $R_{N,n}$ as n tends to infinity is described by the following relation. Let $x_{N,m}^{(n)}$ be the mth positive zero of $R_{N,n}$, so that $0 < x_{N,1}^{(n)} < x_{N,2}^{(n)} < \ldots$. Likewise, let $j_{\nu,m}$ be the mth positive zero of J_ν, so that $0 < j_{\nu,1} < j_{\nu,2} < \ldots$, where J_ν denotes the Bessel functions of the first kind (see [1]). Then

$$\lim_{n\to\infty} 2n x_{N,m}^{(n)} = j_{N+p/2,m}, \tag{2.86}$$

for any nonnegative integer N.

2.7.6 Inequalities

The inequality

$$|R_{N,n}(x)| \leq \binom{n+N+\frac{p}{2}}{n} \tag{2.87}$$

holds for $0 \leq x \leq 1$ and nonnegative integer N and n.

2.7.7 Integrals

The polynomials $R_{N,n}$ satify the relation

$$\int_0^1 \frac{J_{N+p/2}(xy)}{(xy)^{p/2}} R_{N,n}(y) y^{p+1}\, dy = \frac{(-1)^n J_{N+p/2+2n+1}(x)}{x^{p/2+1}}, \tag{2.88}$$

where $x \geq 0$, N and n are nonnegative integers, and J_ν denotes the Bessel functions of the first kind.

2.7.8 Generating Function

The generating function associated with the polynomials $R_{N,n}$ is given by the formula

$$\frac{\left(1 + z - \sqrt{1 + 2z(1-2x^2) + z^2}\right)^{N+p/2}}{(2zx)^{N+p/2} x^{p/2} \sqrt{1 + 2z(1-2x^2) + z^2}} = \sum_{n=0}^{\infty} R_{N,n}(x) z^n, \tag{2.89}$$

where $0 \leq x \leq 1$ is real, z is a complex number such that $|z| \leq 1$, and N is a nonnegative integer.

2.7.9 Differential Equation

The polynomials $R_{N,n}$ satisfy the differential equation

$$(1 - x^2) y''(x) - 2xy'(x) + \left(\chi_{N,n} + \frac{\frac{1}{4} - (N + \frac{p}{2})^2}{x^2}\right) y(x) = 0, \tag{2.90}$$

where

$$\chi_{N,n} = (N + \tfrac{p}{2} + 2n + \tfrac{1}{2})(N + \tfrac{p}{2} + 2n + \tfrac{3}{2}), \tag{2.91}$$

and

$$y(x) = x^{p/2+1} R_{N,n}(x), \tag{2.92}$$

for all $0 < x < 1$ and nonnegative integers N and n.

2.7.10 Recurrence Relations

The polynomials $R_{N,n}$ satisfy the recurrence relation

$$\begin{aligned} &2(n+1)(n+N+\tfrac{p}{2}+1)(2n+N+\tfrac{p}{2})R_{N,n+1}(x) \\ &\quad = -\big((2n+N+\tfrac{p}{2}+1)(N+\tfrac{p}{2})^2 + (2n+N+\tfrac{p}{2})_3(1-2x^2)\big)R_{N,n}(x) \\ &\qquad - 2n(n+N+\tfrac{p}{2})(2n+N+\tfrac{p}{2}+2)R_{N,n-1}(x), \end{aligned} \tag{2.93}$$

where $0 \le x \le 1$, N is a nonnegative integer, n is a positive integer, and $(\cdot)_n$ is defined via the formula

$$(x)_n = x(x+1)(x+2)\ldots(x+n-1), \tag{2.94}$$

for real x and nonnegative integer n. The polynomials $R_{N,n}$ also satisfy the recurrence relations

$$(2n+N+\tfrac{p}{2}+2)xR_{N+1,n}(x) = (n+N+\tfrac{p}{2}+1)R_{N,n}(x) + (n+1)R_{N,n+1}(x), \tag{2.95}$$

for nonnegative integers N and n, and

$$(2n+N+\tfrac{p}{2})xR_{N-1,n}(x) = (n+N+\tfrac{p}{2})R_{N,n}(x) + nR_{N,n-1}(x), \tag{2.96}$$

for integers $N \geq 1$ and $n \geq 0$, where $0 \leq x \leq 1$.

2.7.11 Differential Relations

The Zernike polynomials satisfy the differential relation given by the formula

$$\begin{aligned}
&(2n+N+\tfrac{p}{2})x(1-x^2)\frac{d}{dx}R_{N,n}(x) \\
&\quad = \big(N(2n+N+\tfrac{p}{2}) + 2n^2 - (2n+N)(2n+N+\tfrac{p}{2})x^2\big)R_{N,n}(x) \\
&\qquad\quad + 2n(n+N+\tfrac{p}{2})R_{N,n-1}(x), \qquad (2.97)
\end{aligned}$$

where $0 < x < 1$, N is a nonnegative integer, and n is a positive integer.

2.8 Appendix B: Numerical Evaluation of Zernike Polynomials in $\mathbb{R}^{p+2}$

The main analytical tool of this section is Lemma 2.8.1 which provides a recurrence relation that can be used for the evaluation of radial Zernike Polynomials, $R_{N,n}$.

According to [1], radial Zernike polynomials, $R_{N,n}$, are related to Jacobi polynomials via the formula

$$R_{N,n}(x) = (-1)^n x^N P_n^{(N+\frac{p}{2},0)}(1-2x^2), \qquad (2.98)$$

where $0 \leq x \leq 1$, N and n are nonnegative integers, and $P_n^{(\alpha,0)}$ is defined in (2.10).

The following lemma provides a relation that can be used to evaluate the polynomial $R_{N,n}$.

Lemma 2.8.1 *The polynomials $R_{N,n}$ satisfy the recurrence relation*

$$\begin{aligned} &2(n+1)(n+N+\tfrac{p}{2}+1)(2n+N+\tfrac{p}{2})R_{N,n+1}(x) \\ &\quad = -\big((2n+N+\tfrac{p}{2}+1)(N+\tfrac{p}{2})^2+(2n+N+\tfrac{p}{2})_3(1-2x^2)\big)R_{N,n}(x) \\ &\qquad\quad - 2n(n+N+\tfrac{p}{2})(2n+N+\tfrac{p}{2}+2)R_{N,n-1}(x), \end{aligned} \tag{2.99}$$

where $0 \le x \le 1$, N is a nonnegative integer, n is a positive integer, and $(\cdot)_n$ is defined via the formula

$$(x)_n = x(x+1)(x+2)\dots(x+n-1), \tag{2.100}$$

for real x and nonnegative integer n.

Proof. It is well known that the Jacobi polynomial $P_n^{(\alpha,0)}(x)$ satisfies the recurrence relation

$$a_{1n}P_{n+1}^{(\alpha,0)} = (a_{2n}+a_{3n}x)P_n^{(\alpha,0)}(x) - a_{4n}P_{n-1}^{(\alpha,0)}(x) \tag{2.101}$$

where

$$\begin{aligned} a_{1n} &= 2(n+1)(n+\alpha+1)(2n+\alpha) \\ a_{2n} &= (2n+\alpha+1)\alpha^2 \\ a_{3n} &= (2n+\alpha)(2n+\alpha+1)(2n+\alpha+2) \\ a_{4n} &= 2(n+\alpha)(n)(2n+\alpha+2) \end{aligned} \tag{2.102}$$

Identity (2.99) follows immediately from the combination of (2.101) and (2.102). ■

2.9 Appendix C: Spherical Harmonics in $\mathbb{R}^{p+2}$

Suppose that S^{p+1} denotes the unit sphere in $\mathbb{R}^{p+2}$. The spherical harmonics are a set of real-valued continuous functions on S^{p+1}, which are orthonormal and complete in $L^2(S^{p+1})$. The spherical harmonics of degree $N \geq 0$ are denoted by $S_N^1, S_N^2, \ldots, S_N^\ell, \ldots, S_N^{h(N)} \colon S^{p+1} \to \mathbb{R}$, where

$$h(N) = (2N+p)\frac{(N+p-1)!}{p!\,N!}, \tag{2.103}$$

for all nonnegative integers N.

The following theorem defines the spherical harmonics as the values of certain harmonic, homogeneous polynomials on the sphere (see, for example, [3]).

Theorem 2.9.1 *For each spherical harmonic S_N^ℓ, where $N \geq 0$ and $1 \leq \ell \leq h(N)$ are integers, there exists a polynomial $K_N^\ell \colon \mathbb{R}^{p+2} \to \mathbb{R}$ which is harmonic, i.e.*

$$\nabla^2 K_N^\ell(x) = 0, \tag{2.104}$$

for all $x \in \mathbb{R}^{p+2}$, and homogenous of degree N, i.e.

$$K_N^\ell(\lambda x) = \lambda^N K_N^\ell(x), \tag{2.105}$$

for all $x \in \mathbb{R}^{p+2}$ and $\lambda \in \mathbb{R}$, such that

$$S_N^\ell(\xi) = K_N^\ell(\xi), \tag{2.106}$$

for all $\xi \in S^{p+1}$.

The following theorem is proved in, for example, [3].

Theorem 2.9.2 *Suppose that N is a nonnegative integer. Then there are exactly*

$$(2N+p)\frac{(N+p-1)!}{p!\,N!} \tag{2.107}$$

linearly independent, harmonic, homogenous polynomials of degree N in $\mathbb{R}^{p+2}$.

The following theorem states that for any orthogonal matrix U, the function $S_N^\ell(U\xi)$ is expressible as a linear combination of $S_N^1(\xi), S_N^2(\xi), \dots, S_N^{h(N)}(\xi)$ (see, for example, [3]).

Theorem 2.9.3 *Suppose that N is a nonnegative integer, and that*

$$S_N^1, S_N^2, \dots, S_N^{h(N)} \colon S^{p+1} \to \mathbb{R} \tag{2.108}$$

are a complete set of orthonormal spherical harmonics of degree N. Suppose further that U is a real orthogonal matrix of dimension $p+2 \times p+2$. Then, for each integer $1 \le \ell \le h(N)$, there exists real numbers $v_{\ell,1}, v_{\ell,2}, \dots, v_{\ell,h(N)}$ such that

$$S_N^\ell(U\xi) = \sum_{k=1}^{h(N)} v_{\ell,k} S_N^k(\xi), \tag{2.109}$$

for all $\xi \in S^{p+1}$. Furthermore, if V is the $h(N) \times h(N)$ real matrix with elements $v_{i,j}$ for all $1 \le i, j \le h(N)$, then V is also orthogonal.

Remark 2.9.1 *From Theorem (2.9.3), we observe that the space of linear combinations of functions S_N^ℓ is invariant under all rotations and reflections of S^{p+1}.*

The following theorem states that if an integral operator acting on the space of functions $S^{p+1} \to \mathbb{R}$ has a kernel depending only on the inner product, then the spherical harmonics are eigenfunctions of that operator (see, for example, [3]).

Theorem 2.9.4 (Funk-Hecke) *Suppose that $F\colon [-1,1] \to \mathbb{R}$ is a continuous function, and that $S_N\colon S^{p+1} \to \mathbb{R}$ is any spherical harmonic of degree N. Then*

$$\int_\Omega F(\langle \xi, \eta \rangle) S_N(\xi)\, d\Omega(\xi) = \lambda_N S_N(\eta), \tag{2.110}$$

for all $\eta \in S^{p+1}$, where $\langle \cdot, \cdot \rangle$ denotes the inner product in $\mathbb{R}^{p+2}$, the integral is taken over the whole area of the hypersphere Ω, and λ_N depends only on the function F.

2.10 Appendix D: The Shifted Jacobi Polynomials $P_n^{(k,0)}(2x-1)$

In this section, we introduce a class of Jacobi polynomials that can be used as quadrature and interpolation nodes for Zernike polynomials in $\mathbb{R}^{p+2}$.

We define $\widetilde{P}_n^k(x)$ to be the shifted Jacobi polynomials on the interval $[0,1]$ defined by the formula

$$\widetilde{P}_n^k(x) = \sqrt{k+2n+1}\, P_n^{(k,0)}(1-2x) \tag{2.111}$$

where $k > -1$ is a real number and where $P_n^{(k,0)}$ is defined in (2.10). It follows immediately from (2.111) that $\widetilde{P}_n^k(x)$ are orthogonal with respect to weight function x^k. That is, for all non-negative integers n, the Jacobi polynomial $\widetilde{P}_n^k$ is a polynomial of degree n such that

$$\int_0^1 \widetilde{P}_i^k(x) \widetilde{P}_j^k(x) x^k dx = \delta_{i,j} \tag{2.112}$$

for all non-negative integers i, j where $k > -1$.

The following lemma, which follows immediately from the combination of Lemma

2.2.6 and (2.111), provides a differential equation satisfied by $\widetilde{P}_n^k$.

Lemma 2.10.1 *Let $k > -1$ be a real number and let n be a non-negative integer. Then, $\widetilde{P}_n^k$ satisfies the differential equation,*

$$r - r^2\widetilde{P}_n^{k\prime\prime}(r) + (k - rk + 1 - 2r)\widetilde{P}_n^{k\prime}(r) + n(n+k+1)\widetilde{P}_n^k(r) = 0. \tag{2.113}$$

for all $r \in (0, 1)$.

The following recurrence for $\widetilde{P}_n^k$ follows readily from the combination of Lemma 2.111 and (2.11).

Lemma 2.10.2 *For all non-negative integers n and for all real numbers $k > -1$,*

$$\begin{aligned}\widetilde{P}_{n+1}^k(r) = {} & \frac{(2n+N+1)N^2 + (2n+N)(2n+N+1)(2n+N+2)(1-2r)}{2(n+1)(n+N+1)(2n+N)} \\ & \cdot \frac{\sqrt{2n+k+1}}{\sqrt{2(n+1)+k+1}}\widetilde{P}_n^k(r) \\ & - \frac{2(n+N)(n)(2n+N+2)}{2(n+1)(n+N+1)(2n+N)}\frac{\sqrt{2(n-1)+k+1}}{\sqrt{2(n+1)+k+1}}\widetilde{P}_{n-1}^k(r)\end{aligned} \tag{2.114}$$

2.10.1 Numerical Evaluation of the Shifted Jacobi Polynomials

The following observations provide a way to evaluate $\widetilde{P}_n^k$ and its derivatives.

Observation 2.10.1 *Combining (2.11) with (2.111), we observe that $\widetilde{P}_n^k(x)$ can be evaluated by first evaluating $P_n^{(k,0)}(1-2x)$ via recurrence relation (2.11) and then multiplying the resulting number by*

$$\sqrt{k+2n+1}. \tag{2.115}$$

Observation 2.10.2 *Combining (2.14) with (2.111), we observe that the polynomial* $\widetilde{P}_n^{k\prime}(x)$ *(see (2.111)), can be evaluated by first evaluating* $P_n^{(k,0)\prime}(1-2x)$ *via recurrence relation (2.14) and then multiplying the resulting number by*

$$-2\sqrt{k+2n+1}. \tag{2.116}$$

2.10.2 Prüfer Transform

In this section, we describe the Prüfer Transform, which will be used in Section 2.10.3. A more detailed description of the Prüfer Transform can be found in [11].

Lemma 2.10.3 (Prüfer Transform) *Suppose that the function* $\phi : [a,b] \to \mathbb{R}$ *satisfies the differential equation*

$$\phi''(x) + \alpha(x)\phi'(x) + \beta(x)\phi(x) = 0, \tag{2.117}$$

where $\alpha, \beta : (a,b) \to \mathbb{R}$ *are differential functions. Then,*

$$\frac{d\theta}{dx} = -\sqrt{\beta(x)} - \left(\frac{\beta'(x)}{4\beta(x)} + \frac{\alpha(x)}{2}\right) sin(2\theta), \tag{2.118}$$

where the function $\theta : [a,b] \to \mathbb{R}$ *is defined by the formula,*

$$\frac{\phi'(x)}{\phi(x)} = \sqrt{\beta(x)}\tan(\theta(x)). \tag{2.119}$$

Proof. Introducing the notation

$$z(x) = \frac{\phi'(x)}{\phi(x)} \tag{2.120}$$

for all $x \in [a, b]$, and differentiating (2.120) with respect to x, we obtain the identity

$$\frac{\phi''}{\phi} = \frac{dz}{dx} + z^2(x). \tag{2.121}$$

Substituting (2.121) and (2.120) into (2.117), we obtain,

$$\frac{dz}{dx} = -(z^2(x) + \alpha(x)z(x) + \beta(x)). \tag{2.122}$$

Introducing the notation,

$$z(x) = \gamma(x)\tan(\theta(x)), \tag{2.123}$$

with θ, γ two unknown functions, we differentiate (2.123) and observe that,

$$\frac{dz}{dx} = \gamma(x)\frac{\theta'}{\cos^2(\theta)} + \gamma'(x)\tan(\theta(x)) \tag{2.124}$$

and squaring both sides of (2.123), we obtain

$$z(x)^2 = \tan^2(\theta(x))\gamma(x)^2. \tag{2.125}$$

Substituting (2.124) and (2.125) into (2.122) and choosing

$$\gamma(x) = \sqrt{\beta(x)} \tag{2.126}$$

we obtain

$$\frac{d\theta}{dx} = -\sqrt{\beta(x)} - \left(\frac{\beta'(x)}{4\beta(x)} + \frac{\alpha(x)}{2}\right) sin(2\theta). \tag{2.127}$$

■

Remark 2.10.3 *The Prüfer Transform is often used in algorithms for finding the roots of oscillatory special functions. Suppose that $\phi : [a,b] \to \mathbb{R}$ is a special function satisfying differential equation (2.117). It turns out that in most cases, coefficient*

$$\beta(x) \tag{2.128}$$

in (2.117) is significantly larger than

$$\frac{\beta'(x)}{4\beta(x)} + \frac{\alpha(x)}{2} \tag{2.129}$$

on the interval $[a,b]$, where α and β are defined in (2.117).

Under these conditions, the function θ (see (2.119)), is monotone and its derivative neither approaches infinity nor 0. Furthermore, finding the roots of ϕ is equivalent to finding $x \in [a,b]$ such that

$$\theta(x) = \pi/2 + k\pi \tag{2.130}$$

for some integer k. Consequently, we can find the roots of ϕ by solving well-behaved differential equation (2.127).

Remark 2.10.4 *If for all $x \in [a,b]$, the function $\sqrt{\beta(x)}$ satisfies*

$$\sqrt{\beta(x)} > \frac{\beta'(x)}{4\beta(x)} + \frac{\alpha(x)}{2}, \tag{2.131}$$

then, for all $x \in [a,b]$, we have $\frac{d\theta}{dx} < 0$ (see (2.118)) and we can view $x : [-\pi, \pi] \to$

$\mathbb{R}$ *as a function of θ where x satisfies the first order differential equation*

$$\frac{dx}{d\theta} = \left(-\sqrt{\beta(x)} - \left(\frac{\beta'(x)}{4\beta(x)} + \frac{\alpha(x)}{2} \right) sin(2\theta) \right)^{-1}. \tag{2.132}$$

2.10.3 Roots of the Shifted Jacobi Polynomials

The primary purpose of this section is to describe an algorithm for finding the roots of the Jacobi polynomials $\widetilde{P}_n^k$. These roots will be used in Section 2.4 for the design of quadratures for Zernike Polynomials.

The following lemma follows immediately from applying the Prufer Transform (see Lemma 2.10.3) to (2.113).

Lemma 2.10.4 *For all non-negative integers n, real $k > -1$, and $r \in (0,1)$,*

$$\frac{d\theta}{dr} = -\left(\frac{n(n+k+1)}{r-r^2} \right)^{1/2} - \left(\frac{1-2r+2k-2kr}{4(r-r^2)} \right) \sin(2\theta(r)). \tag{2.133}$$

where the function $\theta : (0,1) \to \mathbb{R}$ is defined by the formula

$$\frac{\widetilde{P}_n^k(r)}{\widetilde{P}_n^{k\prime}(r)} = \left(\frac{n(n+k+1)}{r-r^2} \right)^{1/2} \tan(\theta(r)), \tag{2.134}$$

where $\widetilde{P}_n^k$ is defined in (2.112).

Remark 2.10.5 *For any non-negative integer n,*

$$\frac{d\theta}{dr} < 0 \tag{2.135}$$

for all $r \in (0,1)$. Therefore, applying Remark 2.10.4 to (2.133), we can view r as a function of θ where r satisfies the differential equation

$$\frac{dr}{d\theta} = \left(-\left(\frac{n(n+k+1)}{r-r^2} \right)^{1/2} - \left(\frac{1-2r+2k-2kr}{4(r-r^2)} \right) \sin(2\theta(r)) \right)^{-1}. \tag{2.136}$$

Algorithm

In this section, we describe an algorithm for the evaluation of the n roots of $\widetilde{P}_n^k$. We denote the n roots of $\widetilde{P}_n^k$ by $r_1 < r_2 < ... < r_n$.

Step 1. Choose a point, $x_0 \in (0,1)$, that is greater than the largest root of $\widetilde{P}_n^k$. For example, for all $k \geq 1$, the following choice of x_0 will be sufficient:

$$x_0 = \begin{cases} 1 - 10^{-6} & \text{if } n < 10^3, \\ 1 - 10^{-8} & \text{if } 10^3 \leq n < 10^4, \\ 1 - 10^{-10} & \text{if } 10^4 \leq n < 10^5. \end{cases} \tag{2.137}$$

Step 2. Defining θ_0 by the formula

$$\theta_0 = \theta(x_0), \tag{2.138}$$

where θ is defined in (2.134), solve the ordinary differential equation $\frac{dr}{d\theta}$ (see (2.136)) on the interval $(\pi/2, \theta_0)$, with the initial condition $r(\theta_0) = x_0$. To solve the differential equation, it is sufficient to use, for example, second order Runge Kutta with 100 steps (independent of n). We denote by $\tilde{r}_n$ the approximation to $r(\pi/2)$ obtained by this process. It follows immediately from (2.130) that $\tilde{r}_n$ is an approximation to r_n, the largest root of $\widetilde{P}_n^k$.

Step 3. Use Newton's method with $\tilde{r}_n$ as an initial guess to find r_n to high precision. The polynomials $\widetilde{P}_n^k$ and $\widetilde{P}_n^{k\prime}$ can be evaluated via Observation 2.10.1 and Observation 2.10.2.

Step 4. With initial condition

$$x(\pi/2) = r_n, \tag{2.139}$$

solve differential equation $\frac{dr}{d\theta}$ (see (2.136)) on the interval

$$(-\pi/2, \pi/2) \tag{2.140}$$

using, for example, second order Runge Kuta with 100 steps. We denote by $\tilde{r}_{n-1}$ the approximation to

$$r(-\pi/2) \tag{2.141}$$

obtained by this process.

Step 5. Use Newton's method, with initial guess $\tilde{r}_{n-1}$, to find to high precision the second largest root, r_{n-1}.

Step 6. For $k = \{1, 2, ..., n-1\}$, repeat Step 4 on the interval

$$(-\pi/2 - k\pi, -\pi/2 - (k-1)\pi) \tag{2.142}$$

with intial condition

$$x(-\pi/2 - (k-1)\pi) = r_{n-k+1} \tag{2.143}$$

and repeat Step 5 on $\tilde{r}_{n-k}$.

2.11 Appendix E: Notational Conventions for Zernike Polynomials

In two dimensions, the Zernike polynomials are usually indexed by their azimuthal order and radial order. In this section, we use a slightly different indexing scheme, which leads to simpler formulas and generalizes easily to higher dimensions (see Section 2.2.3 for our definition of the Zernike polynomials $Z_{N,n}^{\ell}$ and the radial polynomials $R_{N,n}$). However, for the sake of completeness, we describe in this section the standard two dimensional indexing scheme, as well as other widely used notational conventions.

If $|m|$ denotes the azimuthal order and n the radial order, then the Zernike polynomials in standard two index notation (using asterisks to differentiate them from the polynomials $Z_{N,n}^{\ell}$ and $R_{N,n}$) are

$$\overset{*}{Z}{}_n^m(\rho,\theta) = \overset{*}{R}{}_n^{|m|}(\rho) \cdot \begin{cases} \sin(|m|\theta) & \text{if } m < 0, \\ \cos(|m|\theta) & \text{if } m > 0, \\ 1 & \text{if } m = 0, \end{cases} \tag{2.144}$$

where

$$\overset{*}{R}{}_n^{|m|}(\rho) = \sum_{k=0}^{\frac{n-|m|}{2}} \frac{(-1)^k (n-k)!}{k!\left(\frac{n+|m|}{2}-k\right)!\left(\frac{n-|m|}{2}-k\right)!} \rho^{n-2k}, \tag{2.145}$$

for all $m = 0, \pm 1, \pm 2, \dots$ and $n = |m|, |m|+2, |m|+4, \dots$ (see Figure 2.144); they are normalized so that

$$\overset{*}{R}{}_n^{|m|}(1) = 1, \tag{2.146}$$

for all $m = 0, \pm 1, \pm 2, \ldots$ and $n = |m|, |m| + 2, |m| + 4, \ldots$. We note that

$$\overset{*}{R}{}_n^{|m|}(\rho) = R_{|m|, \frac{n-|m|}{2}}(\rho), \tag{2.147}$$

for all $m = 0, \pm 1, \pm 2, \ldots$ and $n = |m|, |m| + 2, |m| + 4, \ldots$, where R is defined by (2.27) (see Figure 2.6); equivalently,

$$R_{N,n}(\rho) = \overset{*}{R}{}_{N+2n}^{N}(\rho), \tag{2.148}$$

for all nonnegative integers N and n.

Remark 2.11.1 *The quantity $n + |m|$ is sometimes referred to as the "spacial frequency" of the Zernike polynomial $\overset{*}{Z}{}_n^m(\rho, \theta)$. It roughly corresponds to the frequency of the polynomial on the disc, as opposed to the azimuthal frequency $|m|$ or the order of the polynomial n.*

2.11.1 Zernike Fringe Polynomials

The Zernike Fringe Polynomials are the standard Zernike polynomials, normalized to have L^2 norm equal to π on the unit disc and ordered by their spacial frequency $n + |m|$ (see Table 2.6 and Figure 2.7). This ordering is sometimes called the "Air Force" or "University of Arizona" ordering.

2.11.2 ANSI Standard Zernike Polynomials

The ANSI Standard Zernike polynomials, also referred to as OSA Standard Zernike polynomials or Noll Zernike polynomials, are the standard Zernike polynomials, normalized to have L^2 norm π on the unit disc and ordered by n (the order of the polynomial on the disc; see Table 2.7 and Figure 2.8).

2.11.3 Wyant and Creath Notation

In [30], James Wyant and Katherine Creath observe that it is sometimes conviement to factor the radial polynomial $\overset{*}{R}{}^{|m|}_{2n-|m|}$ into

$$\overset{*}{R}{}^{|m|}_{2n-|m|}(\rho) = Q^{|m|}_n(\rho)\rho^{|m|}, \tag{2.149}$$

for all $m = 0, \pm 1, \pm 2, \ldots$ and $n = |m|, |m| + 1, |m| + 2, \ldots$, where the polynomial $Q^{|m|}_n$ is of order $2(n - |m|)$ (see Figure 2.4). Equivalently, the factorization can be written as

$$\overset{*}{R}{}^{|m|}_n(\rho) = Q^{|m|}_{\frac{n+|m|}{2}}(\rho)\rho^{|m|}, \tag{2.150}$$

for all $m = 0, \pm 1, \pm 2, \ldots$ and $n = |m|, |m| + 2, |m| + 4, \ldots$.

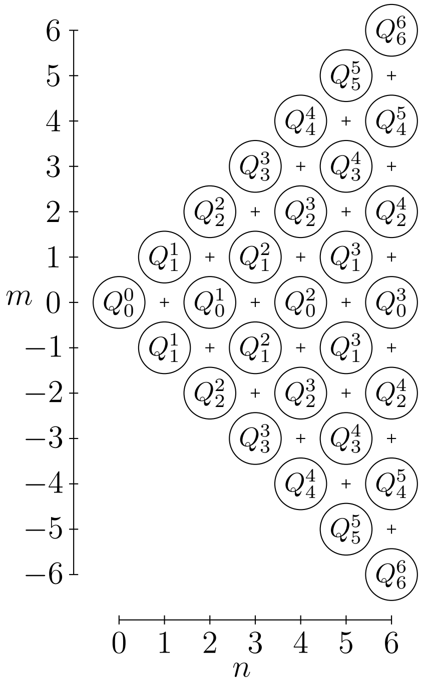

Figure 2.4: The Wyant and Creath polynomials $Q^{|m|}_{(n+|m|)/2}$

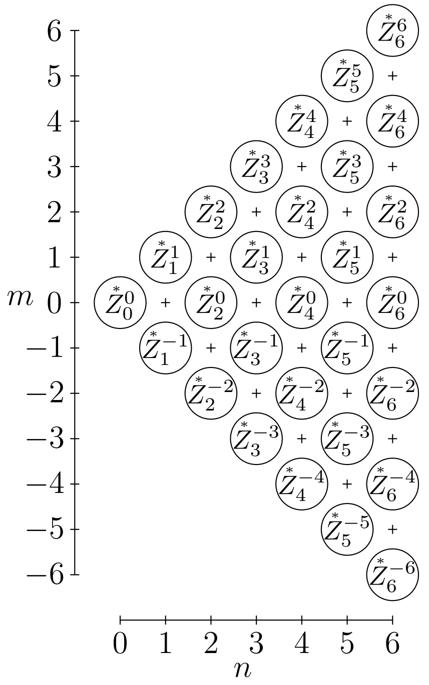
6 $\overset{*}{Z}_6^6$
5 $\overset{*}{Z}_5^5$ +
4 $\overset{*}{Z}_4^4$ + $\overset{*}{Z}_6^4$
3 $\overset{*}{Z}_3^3$ + $\overset{*}{Z}_5^3$ +
2 $\overset{*}{Z}_2^2$ + $\overset{*}{Z}_4^2$ + $\overset{*}{Z}_6^2$
1 $\overset{*}{Z}_1^1$ + $\overset{*}{Z}_3^1$ + $\overset{*}{Z}_5^1$ +
m 0 $\overset{*}{Z}_0^0$ + $\overset{*}{Z}_2^0$ + $\overset{*}{Z}_4^0$ + $\overset{*}{Z}_6^0$
−1 $\overset{*}{Z}_1^{-1}$ + $\overset{*}{Z}_3^{-1}$ + $\overset{*}{Z}_5^{-1}$ +
−2 $\overset{*}{Z}_2^{-2}$ + $\overset{*}{Z}_4^{-2}$ + $\overset{*}{Z}_6^{-2}$
−3 $\overset{*}{Z}_3^{-3}$ + $\overset{*}{Z}_5^{-3}$ +
−4 $\overset{*}{Z}_4^{-4}$ + $\overset{*}{Z}_6^{-4}$
−5 $\overset{*}{Z}_5^{-5}$ +
−6 $\overset{*}{Z}_6^{-6}$
0 1 2 3 4 5 6
n

Figure 2.5: Zernike polynomials $\overset{*}{Z}_n^m$ in standard double index notation

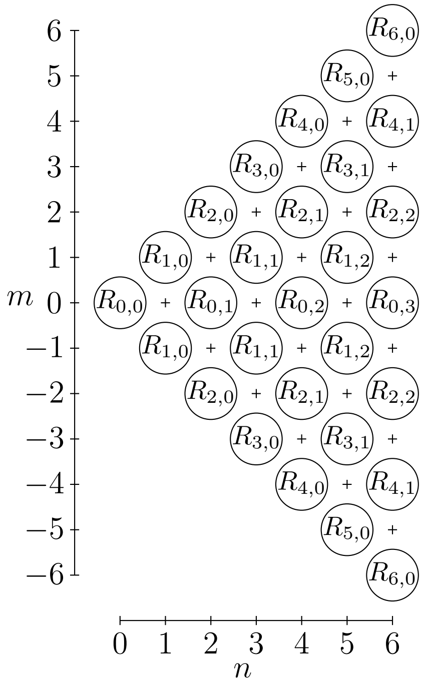
6 $R_{6,0}$
5 $R_{5,0}$ +
4 $R_{4,0}$ + $R_{4,1}$
3 $R_{3,0}$ + $R_{3,1}$ +
2 $R_{2,0}$ + $R_{2,1}$ + $R_{2,2}$
1 $R_{1,0}$ + $R_{1,1}$ + $R_{1,2}$ +
m 0 $R_{0,0}$ + $R_{0,1}$ + $R_{0,2}$ + $R_{0,3}$
−1 $R_{1,0}$ + $R_{1,1}$ + $R_{1,2}$ +
−2 $R_{2,0}$ + $R_{2,1}$ + $R_{2,2}$
−3 $R_{3,0}$ + $R_{3,1}$ +
−4 $R_{4,0}$ + $R_{4,1}$
−5 $R_{5,0}$ +
−6 $R_{6,0}$
0 1 2 3 4 5 6
n

Figure 2.6: The polynomials $R_{|m|,(n-|m|)/2}$

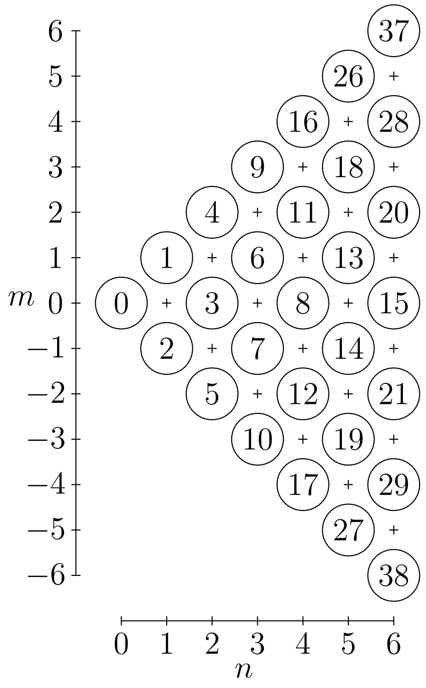

Figure 2.7: Fringe Zernike Polynomial Ordering

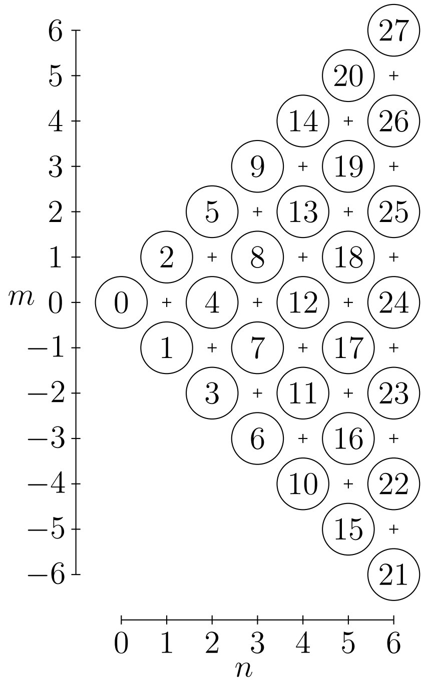

Figure 2.8: ANSI Standard Zernike Polynomial Ordering

index	n	m	spacial frequency	polynomial$^\diamond$	name†
0	0	0	0	1	piston
1	1	1	2	$\overline{R}_{1,0}(\rho)\cos(\theta)$	tilt in x-direction
2	1	-1	2	$\overline{R}_{1,0}(\rho)\sin(\theta)$	tilt in y-direction
3	2	0	2	$\overline{R}_{0,1}(\rho)$	defocus (power)
4	2	2	4	$\overline{R}_{2,0}(\rho)\cos(2\theta)$	defocus + astigmatism 45°/135°
5	2	-2	4	$\overline{R}_{2,0}(\rho)\sin(2\theta)$	defocus + astigmatism 90°/180°
6	3	1	4	$\overline{R}_{1,1}(\rho)\cos(\theta)$	tilt + horiz. coma along x-axis
7	3	-1	4	$\overline{R}_{1,1}(\rho)\sin(\theta)$	tilt + vert. coma along y-axis
8	4	0	4	$\overline{R}_{0,2}(\rho)$	defocus + spherical aberration
9	3	3	6	$\overline{R}_{3,0}(\rho)\cos(3\theta)$	trefoil in x-direction
10	3	-3	6	$\overline{R}_{3,0}(\rho)\sin(3\theta)$	trefoil in y-direction
11	4	2	6	$\overline{R}_{2,1}(\rho)\cos(2\theta)$	
12	4	-2	6	$\overline{R}_{2,1}(\rho)\sin(2\theta)$	
13	5	1	6	$\overline{R}_{1,2}(\rho)\cos(\theta)$	
14	5	-1	6	$\overline{R}_{1,2}(\rho)\sin(\theta)$	
15	6	0	6	$\overline{R}_{0,3}(\rho)$	
16	4	4	8	$\overline{R}_{4,0}(\rho)\cos(4\theta)$	
17	4	-4	8	$\overline{R}_{4,0}(\rho)\sin(4\theta)$	
18	5	3	8	$\overline{R}_{3,1}(\rho)\cos(3\theta)$	
19	5	-3	8	$\overline{R}_{3,1}(\rho)\sin(3\theta)$	
20	6	2	8	$\overline{R}_{2,2}(\rho)\cos(2\theta)$	
21	6	-2	8	$\overline{R}_{2,2}(\rho)\sin(2\theta)$	
22	7	1	8	$\overline{R}_{1,3}(\rho)\cos(\theta)$	
23	7	-1	8	$\overline{R}_{1,3}(\rho)\sin(\theta)$	
24	8	0	8	$\overline{R}_{0,4}(\rho)$	

Table 2.6: Zernike Fringe Polynomials. This table lists the first 24 Zernike polynomials in what is sometimes called the “Fringe”, “Air Force”, or “University of Arizona” ordering (see, for example [23], p. 198, or [30], p. 31). They are often also denoted by $Z_\ell(\rho,\theta)$, where ℓ is the index.
$\diamond$ See formulas (2.27) and (2.30).
† See, for example, [15]. More complex aberrations are usually not named.

index	n	m	spacial frequency	polynomial$^\diamond$	name†
0	0	0	0	1	piston
1	1	-1	2	$\overline{R}_{1,0}(\rho)\sin(\theta)$	tilt in y-direction
2	1	1	2	$\overline{R}_{1,0}(\rho)\cos(\theta)$	tilt in x-direction
3	2	-2	4	$\overline{R}_{2,0}(\rho)\sin(2\theta)$	defocus + astigmatism 90°/180°
4	2	0	2	$\overline{R}_{0,1}(\rho)$	defocus (power)
5	2	2	4	$\overline{R}_{2,0}(\rho)\cos(2\theta)$	defocus + astigmatism 45°/135°
6	3	-3	6	$\overline{R}_{3,0}(\rho)\sin(3\theta)$	trefoil in y-direction
7	3	-1	4	$\overline{R}_{1,1}(\rho)\sin(\theta)$	tilt + vert. coma along y-axis
8	3	1	4	$\overline{R}_{1,1}(\rho)\cos(\theta)$	tilt + horiz. coma along x-axis
9	3	3	6	$\overline{R}_{3,0}(\rho)\cos(3\theta)$	trefoil in x-direction
10	4	-4	8	$\overline{R}_{4,0}(\rho)\sin(4\theta)$	
11	4	-2	6	$\overline{R}_{2,1}(\rho)\sin(2\theta)$	
12	4	0	4	$\overline{R}_{0,2}(\rho)$	
13	4	2	6	$\overline{R}_{2,1}(\rho)\cos(2\theta)$	
14	4	4	8	$\overline{R}_{4,0}(\rho)\cos(4\theta)$	
15	5	-5	10	$\overline{R}_{5,0}(\rho)\sin(5\theta)$	
16	5	-3	8	$\overline{R}_{3,1}(\rho)\sin(3\theta)$	
17	5	-1	6	$\overline{R}_{1,2}(\rho)\sin(\theta)$	
18	5	1	6	$\overline{R}_{1,2}(\rho)\cos(\theta)$	
19	5	3	8	$\overline{R}_{3,1}(\rho)\cos(3\theta)$	
20	5	5	10	$\overline{R}_{5,0}(\rho)\cos(5\theta)$	
21	6	-6	12	$\overline{R}_{6,0}(\rho)\sin(6\theta)$	
22	6	-4	10	$\overline{R}_{4,1}(\rho)\sin(4\theta)$	
23	6	-2	8	$\overline{R}_{2,2}(\rho)\sin(2\theta)$	
24	6	0	6	$\overline{R}_{0,3}(\rho)$	

Table 2.7: ANSI Standard Zernike Polynomials. This table lists the first 24 Zernike polynomials in the ANSI Standard ordering, also referred to as the "OSA Standard" or "Noll" ordering (see, for example [23], p. 201, or [20]). They are often also denoted by $Z_\ell(\rho,\theta)$, where ℓ is the index.

$\diamond$ See formulas (2.27) and (2.30).

† See, for example, [15]. More complex aberrations are usually not named.

Chapter 3

GPSFs

3.1 Background

Generalized Prolate Spheroidal Functions (GPSFs) are functions $\psi_j : \mathbb{R}^n \to \mathbb{C}$ satisfying

$$\lambda_j \psi_j(x) = \int_B \psi_j(t) e^{ic\langle x,t\rangle} dt \tag{3.1}$$

for some $\lambda_j \in \mathbb{C}$ where B denotes the unit ball in $\mathbb{R}^n$. A function $f : \mathbb{R}^n \to \mathbb{C}$ is referred to as bandlimited with bandlimit $c > 0$ if

$$f(x) = \int_B \sigma(t) e^{ic\langle x,t\rangle} dt \tag{3.2}$$

where B denotes the unit ball in $\mathbb{R}^n$ and σ is a square-integrable function defined on B. Bandlimited functions are encountered in a variety of applications including in signal processing, antenna design, radar, etc.

Much of the theory and numerical machinery of GPSFs in two dimensions is described in [24]. In this chapter, we provide analytical and numerical tools for GPSFs in $\mathbb{R}^n$ for all $n > 0$. We introduce algorithms for evaluating GPSFs, quadrature rules for integrating bandlimited functions, and numerical interpolation schemes for representing bandlimited functions in GPSF expansions. We also provide numerical machinery for efficient evaluation of eigenvalues λ_j (see (3.1)).

The structure of this chapter is as follows. In Section 3.2 we provide basic mathematical background that will be used throughout the remainder of the chapter. Section 3.3 contains analytical facts related to the numerical evaluation of GPSFs that will be used in subsequent sections. In Section 3.4, we describe a numerical scheme for evaluating GPSFs. Section 3.5 contains a quadrature rule for integrating bandlimited functions. Section 3.6 includes a numerical scheme for expanding bandlimited functions into GPSFs. In Section 3.7, we provide the numerical results of implementing the quadrature and interpolation schemes as well as plots of GPSFs and their eigenvalues. In Section 3.8 we provide certain miscellaneous properties of GPSFs.

3.2 Mathematical and Numerical Preliminaries

In this section, we introduce notation and elementary mathematical and numerical facts which will be used in subsequent sections.

In accordance with standard practice, we define the Gamma function, $\Gamma(x)$, by the formula

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt \tag{3.3}$$

where e will denote the base of the natural logarithm. We will be denoting by $\delta_{i,j}$

the function defined by the formula

$$\delta_{i,j} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases} \tag{3.4}$$

The following is a well-known technical lemma that will be used in Section 3.3.2.

Lemma 3.2.1 *For any real number $a > 0$ and for any integer $n > ae$,*

$$\frac{a^n \sqrt{n}}{\Gamma(n+1)} < 1 \tag{3.5}$$

where $\Gamma(n)$ is defined in (3.3).

The following lemma follows immediately from Formula 9.1.10 in [1].

Lemma 3.2.2 *For all real numbers $x \in [0,1]$, and for all real numbers $\nu \geq -1/2$,*

$$|J_\nu(x)| \leq \frac{|x/2|^\nu}{\Gamma(\nu+1)} \tag{3.6}$$

where J_ν is a Bessel function of the first kind and $\Gamma(\nu)$ is defined in (3.3).

3.2.1 Jacobi Polynomials

In this section, we summarize some properties Jacobi polynomials.

Jacobi Polynomials, denoted $P_n^{(\alpha,\beta)}$, are orthogonal polynomials on the interval $(-1,1)$ with respect to weight function

$$w(x) = (1-x)^\alpha (1+x)^\beta. \tag{3.7}$$

Specifically, for all non-negative integers n, m with $n \neq m$ and real numbers $\alpha, \beta > -1$,

$$\int_{-1}^{1} P_n^{(\alpha,\beta)}(x) P_m^{(\alpha,\beta)}(x)(1-x)^{\alpha}(1+x)^{\beta} dx = 0 \tag{3.8}$$

The following lemma, provides a stable recurrence relation that can be used to evaluate a particular class of Jacobi Polynomials (see, for example, [1]).

Lemma 3.2.3 *For any integer $n \geq 1$ and $\alpha > -1$,*

$$\begin{aligned} P_{n+1}^{(\alpha,0)}(x) = {} & \frac{(2n+\alpha+1)\alpha^2 + (2n+\alpha)(2n+\alpha+1)(2n+\alpha+2)x}{2(n+1)(n+\alpha+1)(2n+\alpha)} P_n^{(\alpha,0)}(x) \\ & - \frac{2(n+\alpha)(n)(2n+\alpha+2)}{2(n+1)(n+\alpha+1)(2n+\alpha)} P_{n-1}^{(\alpha,0)}(x), \end{aligned} \tag{3.9}$$

where

$$P_0^{(\alpha,0)}(x) = 1 \tag{3.10}$$

and

$$P_1^{(\alpha,0)}(x) = \frac{\alpha + (\alpha+2)x}{2}. \tag{3.11}$$

The Jacobi Polynomial $P_n^{(\alpha,0)}$ is defined in (3.8).

The following lemma provides a stable recurrence relation that can be used to evaluate derivatives of a certain class of Jacobi Polynomials. It is readily obtained by differentiating (3.9) with respect to x,

Lemma 3.2.4 *For any integer $n \geq 1$ and $\alpha > -1$,*

$$\begin{aligned} P_{n+1}^{(\alpha,0)\prime}(x) &= \frac{(2n+\alpha+1)\alpha^2 + (2n+\alpha)(2n+\alpha+1)(2n+\alpha+2)x}{2(n+1)(n+\alpha+1)(2n+\alpha)} P_n^{(\alpha,0)\prime}(x) \\ &- \frac{2(n+\alpha)(n)(2n+\alpha+2)}{2(n+1)(n+\alpha+1)(2n+\alpha)} P_{n-1}^{(\alpha,0)\prime}(x) \\ &+ \frac{(2n+\alpha)(2n+\alpha+1)(2n+\alpha+2)}{2(n+1)(n+\alpha+1)(2n+\alpha)} P_n^{(\alpha,0)}(x), \end{aligned} \tag{3.12}$$

where

$$P_0^{(\alpha,0)\prime}(x) = 0 \tag{3.13}$$

and

$$P_1^{(\alpha,0)\prime}(x) = \frac{\alpha+2}{2}. \tag{3.14}$$

The Jacobi Polynomial $P_n^{(\alpha,0)}$ is defined in (3.8) and $P_n^{(\alpha,0)\prime}(x)$ denotes the derivative of $P_n^{(\alpha,0)}(x)$ with respect to x.

The following two lemmas, which provide a differential equation and a recurrence relation for Jacobi polynomials, can be found in, for example, [1].

Lemma 3.2.5 *For any integer $n \geq 2$ and $\alpha > -1$,*

$$(1-x^2)P_n^{(\alpha,0)\prime\prime}(x) + (-\alpha-(\alpha+2)x)P_n^{(\alpha,0)\prime}(x) + n(n+\alpha+1)P_n^{(\alpha,0)}(x) = 0 \tag{3.15}$$

for all $x \in [0,1]$ where $P_n^{(\alpha,0)}$ is defined in (3.8).

Lemma 3.2.6 *For all $\alpha > -1$, $x \in (0,1)$, and any integer $n \geq 2$,*

$$a_{1n}P_{n+1}^{(\alpha,0)} = (a_{2n} + a_{3n}x)P_n^{(\alpha,0)}(x) - a_{4n}P_{n-1}^{(\alpha,0)}(x) \tag{3.16}$$

where

$$\begin{aligned} a_{1n} &= 2(n+1)(n+\alpha+1)(2n+\alpha) \\ a_{2n} &= (2n+\alpha+1)\alpha^2 \\ a_{3n} &= (2n+\alpha)(2n+\alpha+1)(2n+\alpha+2) \\ a_{4n} &= 2(n+\alpha)(n)(2n+\alpha+2) \end{aligned} \tag{3.17}$$

and

$$\begin{aligned} P_0^{(\alpha,0)}(x) &= 1 \\ P_1^{(\alpha,0)}(x) &= \frac{\alpha+(\alpha+2)x}{2}. \end{aligned} \tag{3.18}$$

3.2.2 Zernike Polynomials

In this section, we describe properties of Zernike polynomials, which are a family of orthogonal polynomials on the unit ball in $\mathbb{R}^{p+2}$. They are the natural basis for representing GPFS.

Zernike polynomials are defined via the formula

$$Z_{N,n}^{\ell}(x) = R_{N,n}(\|x\|)S_N^{\ell}(x/\|x\|), \tag{3.19}$$

for all $x \in \mathbb{R}^{p+2}$ such that $\|x\| \leq 1$, where N and n are nonnegative integers, S_N^{ℓ} are the orthonormal surface harmonics of degree N (see Section 3.2.7), and $R_{N,n}$ are polynomials of degree $2n+N$ defined via the formula

$$R_{N,n}(x) = x^N \sum_{m=0}^{n} (-1)^m \binom{n+N+\frac{p}{2}}{m} \binom{n}{m} (x^2)^{n-m}(1-x^2)^m, \tag{3.20}$$

for all $0 \leq x \leq 1$. The polynomials $R_{N,n}$ satisfy the relation

$$R_{N,n}(1) = 1, \tag{3.21}$$

and are orthogonal with respect to the weight function $w(x) = x^{p+1}$, so that

$$\int_0^1 R_{N,n}(x) R_{N,m}(x) x^{p+1} \, dx = \frac{\delta_{n,m}}{2(2n + N + \frac{p}{2} + 1)}. \tag{3.22}$$

We define the polynomials $\overline{R}_{N,n}$ via the formula

$$\overline{R}_{N,n}(x) = \sqrt{2(2n + N + p/2 + 1)} R_{N,n}(x), \tag{3.23}$$

so that

$$\int_0^1 \big(\overline{R}_{N,n}(x)\big)^2 x^{p+1} \, dx = 1, \tag{3.24}$$

where N and n are nonnegative integers. In an abuse of notation, we refer to both the polynomials $Z^{\ell}_{N,n}$ and the radial polynomials $R_{N,n}$ as Zernike polynomials where the meaning is obvious.

Remark 3.2.1 *When $p = -1$, Zernike polynomials take the form*

$$\begin{aligned} Z^1_{0,n}(x) &= R_{0,n}(|x|) = P_{2n}(x), \\ Z^2_{1,n}(x) &= \operatorname{sgn}(x) \cdot R_{1,n}(|x|) = P_{2n+1}(x), \end{aligned} \tag{3.25}$$

for $-1 \leq x \leq 1$ and nonnegative integer n, where P_n denotes the Legendre polyno-

mial of degree n and

$$\operatorname{sgn}(x) = \begin{cases} 1 & \text{if } x > 0, \\ 0 & \text{if } x = 0, \\ -1 & \text{if } x < 0, \end{cases} \tag{3.26}$$

for all real x.

Remark 3.2.2 *When $p = 0$, Zernike polynomials take the form*

$$Z^1_{N,n}(x_1, x_2) = R_{N,n}(r)\cos(N\theta), \tag{3.27}$$

$$Z^2_{N,n}(x_1, x_2) = R_{N,n}(r)\sin(N\theta), \tag{3.28}$$

where $x_1 = r\cos(\theta)$, $x_2 = r\sin(\theta)$, and N and n are nonnegative integers.

The following lemma, which can be found in, for example, [1], shows how Zernike polynomials are related to Jacobi polynomials.

Lemma 3.2.7 *For all non-negative integers N, n,*

$$R_{N,n}(x) = (-1)^n x^N P_n^{(N+\frac{p}{2},0)}(1 - 2x^2), \tag{3.29}$$

where $0 \le x \le 1$, and $P_n^{(\alpha,0)}$, $\alpha > -1$, is defined in (3.20).

3.2.3 Numerical Evaluation of Zernike Polynomials

In this section, we provide a stable recurrence relation (see Lemma 3.2.8) that can be used to evaluate Zernike Polynomials.

The following lemma follows immediately from applying Lemma 3.2.7 to (3.16).

Lemma 3.2.8 *The polynomials $R_{N,n}$, defined in (3.20) satisfy the recurrence relation*

$$\begin{aligned} &R_{N,n+1}(x) = \\ &-\frac{((2n+N+1)N^2+(2n+N)(2n+N+1)(2n+N+2)(1-2x^2))}{2(n+1)(n+N+1)(2n+N)}R_{N,n}(x) \\ &-\frac{2(n+N)(n)(2n+N+2)}{2(n+1)(n+N+1)(2n+N)}R_{N,n-1}(x) \end{aligned} \tag{3.30}$$

where $0 \leq x \leq 1$, N is a non-negative integer, n is a positive integer, and

$$R_{N,0}(x) = x^N \tag{3.31}$$

and

$$R_{N,1}(x) = -x^N \frac{N+(N+2)(1-2x^2)}{2}. \tag{3.32}$$

Remark 3.2.3 *The algorithm for evaluating Zernike polynomials using the recurrence relation in Lemma 3.2.8 is known as Kintner's method (see [16] and, for example, [6]).*

3.2.4 Modified Zernike polynomials, $\overline{T}_{N,n}$

In this section, we define the modified Zernike polynomials, $\overline{T}_{N,n}$ and provide some of their properties. This family of functions will be used in Section 3.4 for the numerical evaluation of GPSFs.

We define the function $T_{N,n}$ by the formula

$$T_{N,n}(r) = r^{\frac{p+1}{2}} R_{N,n}(r) \tag{3.33}$$

where N, n are non-negative integers. We define $\overline{T}_{N,n} : [0,1] \to \mathbb{R}$ by the formula,

$$\overline{T}_{N,n}(r) = r^{\frac{p+1}{2}} \overline{R}_{N,n}(r) \tag{3.34}$$

where N, n are non-negative integers and $\overline{R}_{N,n}$ is a normalized Zernike polynomial defined in (3.23), so that

$$\int_0^1 (\overline{T}_{N,n}(r))^2 dr = 1. \tag{3.35}$$

Lemma 3.2.9 *The functions $\overline{T}_{N,n}$ are orthonormal on the interval $(0,1)$ with respect to weight function $w(x) = 1$. That is,*

$$\int_0^1 \overline{T}_{N,n}(r) \overline{T}_{N,m}(r) dr = \delta_{n,m}. \tag{3.36}$$

Proof. Using (3.34), (3.22) and (3.24), for all non-negative integers N, n, m,

$$\begin{aligned}\int_0^1 \overline{T}_{N,n}(r) \overline{T}_{N,m}(r) dr &= \int_0^1 r^{\frac{p+1}{2}} \overline{R}_{N,n}(r) r^{\frac{p+1}{2}} \overline{R}_{N,m}(r) dr \\ &= \int_0^1 \overline{R}_{N,n}(r) \overline{R}_{N,m}(r) r^{p+1} dr \\ &= \delta_{n,m}\end{aligned} \tag{3.37}$$

■

The following identity follows immediately from the combination of (3.34),(3.29), and (3.23).

Lemma 3.2.10 *For all non-negative integers N, n,*

$$\overline{T}_{N,n}(r) = P_n^{(N+p/2,0)}(1-2r^2)(-1)^n \sqrt{2(2n+N+p/2+1)}\, r^{\frac{p+1}{2}} \tag{3.38}$$

where $\overline{T}_{N,n}$ is defined in (3.34) and $P_n^{(N+p/2,0)}$ is a Zernike polynomial defined in

(3.8).

The following lemma, which provides a differential equation for $\overline{T}_{N,n}$, follows immediately from substituting (3.38) into Lemma 3.40.

Lemma 3.2.11 *For all $r \in [0,1]$, non-negative integers N, n and real $p \geq -1$,*

$$(1-r^2)\overline{T}''_{N,n}(r) - 2r\overline{T}'_{N,n}(r) + \left(\chi_{N,n} + \frac{\frac{1}{4} - (N+\frac{p}{2})^2}{r^2}\right)\overline{T}_{N,n}(r) = 0 \tag{3.39}$$

where $\chi_{N,n}$ is defined by the formula

$$\chi_{N,n} = (N + p/2 + 2n + 1/2)(N + p/2 + 2n + 3/2). \tag{3.40}$$

The following lemma provides a recurrence relation satisfied by $\overline{T}_{N,n}$. It follows immediately from the combination of Lemma 3.2.10 and (3.9).

Lemma 3.2.12 *For any non-negative integers N, n and for all $r \in [0,1]$,*

$$\begin{aligned} r^2\overline{T}_{N,n}(r) =& \frac{\sqrt{2(2n+N+p/2+1)}}{\sqrt{2(2(n-1)+N+p/2+1)}}\frac{a_{4n}}{2a_{3n}}\overline{T}_{N,n-1}(r) \\ &+ \frac{a_{2n}+a_{3n}}{2a_{3n}}\overline{T}_{N,n}(r) \\ &+ \frac{\sqrt{2(2n+N+p/2+1)}}{\sqrt{2(2(n+1)+N+p/2+1)}}\frac{a_{1n}}{2a_{3n}}\overline{T}_{N,n+1}(r) \end{aligned} \tag{3.41}$$

where $\overline{T}_{N,n}$ is defined in (3.34) and

$$\begin{aligned} a_{1n} &= 2(n+1)(n+N+p/2+1)(2n+N+p/2) \\ a_{2n} &= (2n+N+p/2+1)N+p/2^2 \\ a_{3n} &= (2n+N+p/2)(2n+N+p/2+1)(2n+N+p/2+2) \\ a_{4n} &= 2(n+N+p/2)(n)(2n+N+p/2+2). \end{aligned} \tag{3.42}$$

Proof. Applying the change of variables $1 - 2r^2 = x$ to (3.16) and setting $\alpha = N + p/2$, we obtain

$$\begin{aligned} r^2 P_n^{(N+p/2,0)}(1-2r^2) =& \frac{a_{2n}}{2a_3 n} P_n^{(N+p/2,0)}(1-2r^2) + \frac{1}{2} P_n^{(N+p/2,0)}(1-2r^2) \\ & - \frac{a_{4n}}{2a_3 n} P_{n-1}^{(N+p/2,0)}(1-2r^2) - \frac{a_{1n}}{2a_3 n} P_{n+1}^{(N+p/2,0)}(1-2r^2). \end{aligned} \tag{3.43}$$

Identity (3.41) follows immediately m the combination of (3.43) with Lemma 3.2.10. ■

The following observation provides a scheme for computing $\overline{T}_{N,n}$.

Observation 3.2.4 *Combining (3.34), Lemma 3.2.8, and (3.23), we observe that the modified Zernike polynomial $\overline{T}_{N,n}(r)$ can be evaluated by first computing*

$$P_n^{(N+p/2,0)}(1-2r^2) \tag{3.44}$$

via recurrence relation (3.16) and then multiplying the resulting number by

$$r^N(-1)^n \sqrt{2(2n+N+p/2+1)} r^{\frac{p+1}{2}}. \tag{3.45}$$

We define the function $\overline{T}^*_{N,n}$ by the formula

$$\overline{T}^*_{N,n}(r) = \frac{\overline{T}_{N,n}(r)}{r^{N+\frac{p+1}{2}}}. \tag{3.46}$$

where N, n are non-negative integers and $r \in (0,1)$. The following technical lemma involving $\overline{T}^*_{N,n}$ will be used in Section 3.3.3.

Lemma 3.2.13 *For all non-negative integers N, n,*

$$\overline{T}^*_{N,n}(0) = \sqrt{2(2n+N+p/2+1)}(-1)^n \binom{n+N+p/2}{n}. \tag{3.47}$$

Proof. Combining (3.34) and (3.20), we observe that

$$\overline{T}_{N,n}(r) = \sum_{k=0}^{n} a_{N+k} r^{N+\frac{p+1}{2}+2k} \tag{3.48}$$

where a_{N+k} is some real number for $k = 0, 1, ..., n$. In particular, using (3.20),

$$a_N = \sqrt{2(2n+N+p/2+1)}(-1)^n \binom{n+N+p/2}{n}. \tag{3.49}$$

Combining (3.46) and (3.49), we obtain (3.47). ■

The following lemma provides a relation that will be used in section 3.4.1 for the evaluation of certain eigenvalues.

Lemma 3.2.14 *Suppose that N is a nonnegative integer and that $n \geq 1$ is an integer. Then*

$$\begin{aligned} &\tilde{a}_n x T'_{N,n+1}(x) - \tilde{b}_n x T_{N,n+1}(x) + \tilde{c}_n x T'_{N,n-1}(x) \\ &= a_n T_{N,n+1}(x) - b_n T_{N,n}(x) + c_n T_{N,n-1}(x), \end{aligned} \tag{3.50}$$

for all $0 \leq x \leq 1$, *where*

$$\begin{aligned}
\tilde{a}_n &= 2(n+N+1)(2n+N), \\
\tilde{b}_n &= 2N(2n+N+1), \\
\tilde{c}_n &= -2n(2n+N+2), \\
a_n &= (2N+4n+5)(n+N+1)(2n+N), \\
b_n &= N(2n+N+1) - 2(2n+N)_3, \\
c_n &= n(2N+4n-1)(2n+N+2),
\end{aligned} \tag{3.51}$$

with $(\cdot)_k$ *denoting the Pochhammer symbol or rising factorial.*

3.2.5 Prüfer Transform

In this section, we describe the Prüfer Transform, which will be used in Section 3.5.1 in an algorithm for finding the roots of GPSFs. A more detailed description of the Prüfer Transform can be found in [11].

Lemma 3.2.15 (Prüfer Transform) *Suppose that the function* $\phi : [a, b] \to \mathbb{R}$ *satisfies the differential equation*

$$\phi''(x) + \alpha(x)\phi'(x) + \beta(x)\phi(x) = 0, \tag{3.52}$$

where $\alpha, \beta : (a, b) \to \mathbb{R}$ *are differential functions. Then,*

$$\frac{d\theta}{dx} = -\sqrt{\beta(x)} - \left(\frac{\beta'(x)}{4\beta(x)} + \frac{\alpha(x)}{2}\right) sin(2\theta), \tag{3.53}$$

where the function $\theta : [a, b] \to \mathbb{R}$ *is defined by the formula,*

$$\frac{\phi'(x)}{\phi(x)} = \sqrt{\beta(x)} \tan(\theta(x)). \tag{3.54}$$

Proof. Introducing the notation

$$z(x) = \frac{\phi'(x)}{\phi(x)} \tag{3.55}$$

for all $x \in [a, b]$, and differentiating (3.55) with respect to x, we obtain the identity

$$\frac{\phi''}{\phi} = \frac{dz}{dx} + z^2(x). \tag{3.56}$$

Substituting (3.56) and (3.55) into (3.52), we obtain,

$$\frac{dz}{dx} = -(z^2(x) + \alpha(x)z(x) + \beta(x)). \tag{3.57}$$

Introducing the notation,

$$z(x) = \gamma(x)\tan(\theta(x)), \tag{3.58}$$

with θ, γ two unknown functions, we differentiate (3.58) and observe that,

$$\frac{dz}{dx} = \gamma(x)\frac{\theta'}{\cos^2(\theta)} + \gamma'(x)\tan(\theta(x)) \tag{3.59}$$

and squaring both sides of (3.58), we obtain

$$z(x)^2 = \tan^2(\theta(x))\gamma(x)^2. \tag{3.60}$$

Substituting (3.59) and (3.60) into (3.57) and choosing

$$\gamma(x) = \sqrt{\beta(x)} \tag{3.61}$$

we obtain

$$\frac{d\theta}{dx} = -\sqrt{\beta(x)} - \left(\frac{\beta'(x)}{4\beta(x)} + \frac{\alpha(x)}{2}\right) sin(2\theta). \tag{3.62}$$

■

Remark 3.2.5 *The Prüfer Transform is often used in algorithms for finding the roots of oscillatory special functions. Suppose that $\phi : [a,b] \to \mathbb{R}$ is a special function satisfying differential equation (3.34). It turns out that in most cases, coefficient*

$$\beta(x) \tag{3.63}$$

in (3.52) is significantly larger than

$$\frac{\beta'(x)}{4\beta(x)} + \frac{\alpha(x)}{2} \tag{3.64}$$

on the interval $[a,b]$, where α and β are defined in (3.52).

Under these conditions, the function θ (see (3.54)), is monotone and its derivative neither approaches infinity nor 0. Furthermore, finding the roots of ϕ is equivalent to finding $x \in [a,b]$ such that

$$\theta(x) = \pi/2 + k\pi \tag{3.65}$$

for some integer k. Consequently, we can find the roots of φ by solving (3.62), a well-behaved differential equation.

Remark 3.2.6 *If for all $x \in [a,b]$, the function $\sqrt{\beta(x)}$ satisfies*

$$\sqrt{\beta(x)} > \frac{\beta'(x)}{4\beta(x)} + \frac{\alpha(x)}{2}, \tag{3.66}$$

then, for all $x \in [a,b]$, we have $\frac{d\theta}{dx} < 0$ (see (3.53)) and we can view $x : [-\pi,\pi] \to \mathbb{R}$ as a function of θ where x satisfies the first order differential equation

$$\frac{dx}{d\theta} = \left(-\sqrt{\beta(x)} - \left(\frac{\beta'(x)}{4\beta(x)} + \frac{\alpha(x)}{2}\right) sin(2\theta)\right)^{-1}. \tag{3.67}$$

3.2.6 Miscellaneous Analytical Facts

In this section, we provide several facts from analysis that will by used in subsequent sections.

The following theorem is an identity involving the incomplete beta function.

Theorem 3.2.16 *Suppose that $a, b > 0$ are real numbers and n is a nonnegative integer. Then*

$$B_x(a+n,b) = \frac{\Gamma(a+n)}{\Gamma(a+b+n)}\left(\frac{\Gamma(a+b)}{\Gamma(a)}B_x(a,b) - (1-x)^b \sum_{k=1}^{n} \frac{\Gamma(a+b+k-1)}{\Gamma(a+k)} x^{a+k-1}\right) \tag{3.68}$$

for all $0 \le x \le 1$, where $B_x(a,b)$ denotes the incomplete beta function.

The following lemma is an identity involving the gamma function.

Lemma 3.2.17 *Suppose that n is a nonnegative integer. Then*

$$\sqrt{\pi} + \sum_{k=1}^{n} \frac{\Gamma(k+\frac{1}{2})}{\Gamma(k+1)} = \frac{2\Gamma(n+\frac{3}{2})}{\Gamma(n+1)}. \tag{3.69}$$

The following two lemmas are identities involving the incomplete beta function.

Lemma 3.2.18 *Suppose that* $0 \le r \le 1$. *Then*

$$B_{1-r^2}(1, \tfrac{1}{2}) = 2(1-r). \tag{3.70}$$

Lemma 3.2.19 *Suppose that* $0 \le r \le 1$. *Then*

$$B_{1-r^2}(\tfrac{1}{2}, \tfrac{1}{2}) = 2\arccos(r). \tag{3.71}$$

Bessel Functions

The primary analytical tool of this subsection is Theorem 3.2.25.

The following lemmas 3.2.20, 3.2.21, 3.2.22, 3.2.23, 3.2.24 describe the limiting behavior of certain integrals involving Bessel functions.

Lemma 3.2.20 *Suppose that* $\nu > 0$. *Then*

$$\int_0^1 (J_\nu(2cr))^2 \frac{1}{r}\, dr = \frac{1}{2\nu} + O\Big(\frac{1}{c}\Big), \tag{3.72}$$

as $c \to \infty$.

Lemma 3.2.21 *Suppose that* $\nu > 0$. *Then*

$$\int_0^1 (J_\nu(2cr))^2\, dr = \frac{1}{2\pi}\frac{\log(c)}{c} + o\Big(\frac{\log(c)}{c}\Big), \tag{3.73}$$

as $c \to \infty$.

Lemma 3.2.22 *Suppose that* $\nu > 0$ *is real and* k *is a positive integer. Then*

$$\int_0^1 (J_\nu(2cr))^2 r^k\, dr = O\Big(\frac{1}{c}\Big), \tag{3.74}$$

as $c \to \infty$.

Lemma 3.2.23 *Suppose that n is a positive integer. Then*

$$\int_0^1 \frac{(J_n(2cr))^2}{r} \arccos(r)\, dr = \frac{\pi}{4n} - \frac{1}{2\pi}\frac{\log(c)}{c} + o\Big(\frac{\log(c)}{c}\Big), \tag{3.75}$$

as $c \to \infty$.

Lemma 3.2.24 *Suppose that n and k are positive integers. Then*

$$\int_0^1 (J_n(2cr))^2(1-r^2)^{k-\frac{1}{2}}\, dr = \frac{1}{2\pi}\frac{\log(c)}{c} + o\Big(\frac{\log(c)}{c}\Big), \tag{3.76}$$

as $c \to \infty$.

The following theorem describes the limiting behavior of a certain integral involving a Bessel function and the incomplete beta function.

Theorem 3.2.25 *Suppose that $p \geq -1$ is an integer. Then*

$$\int_0^1 \frac{(J_{p/2+1}(2cr))^2}{r} B_{1-r^2}(\tfrac{p}{2}+\tfrac{3}{2},\tfrac{1}{2})\, dr = \frac{\sqrt{\pi}\,\Gamma(\frac{p}{2}+\frac{3}{2})}{(p+2)\Gamma(\frac{p}{2}+2)} - \frac{1}{\pi}\frac{\log(c)}{c} + o\Big(\frac{\log(c)}{c}\Big) \tag{3.77}$$

as $c \to \infty$, where $B_x(a,b)$ denotes the incomplete beta function.

Proof. Suppose that $p \geq -1$ is an odd integer, and let $n = \frac{p}{2}+\frac{1}{2}$. Then

$$\int_0^1 \frac{\big(J_{p/2+1}(2cr)\big)^2}{r} B_{1-r^2}(\tfrac{p}{2}+\tfrac{3}{2},\tfrac{1}{2})\, dr = \int_0^1 \frac{\big(J_{n+1/2}(2cr)\big)^2}{r} B_{1-r^2}(1+n,\tfrac{1}{2})\, dr. \tag{3.78}$$

By Theorem 3.2.16 and Lemma 3.2.18, we observe that

$$\begin{aligned}
&\int_0^1 \frac{\big(J_{n+1/2}(2cr)\big)^2}{r} B_{1-r^2}(1+n,\tfrac{1}{2})\,dr \\
&\quad = \frac{\Gamma(n+1)}{\Gamma(n+\frac{3}{2})} \int_0^1 \frac{\big(J_{n+1/2}(2cr)\big)^2}{r} \Big(\frac{\sqrt{\pi}}{2} B_{1-r^2}(1,\tfrac{1}{2}) - r\sum_{k=1}^{n} \frac{\Gamma(k+\frac{1}{2})}{\Gamma(k+1)}(1-r^2)^k\Big)\,dr \\
&\quad = \frac{\Gamma(n+1)}{\Gamma(n+\frac{3}{2})} \int_0^1 \frac{\big(J_{n+1/2}(2cr)\big)^2}{r} \Big(\sqrt{\pi}(1-r) - r\sum_{k=1}^{n} \frac{\Gamma(k+\frac{1}{2})}{\Gamma(k+1)}(1-r^2)^k\Big)\,dr,
\end{aligned} \tag{3.79}$$

where $0 \le r \le 1$ and n is a nonnegative integer. By lemmas 3.2.20, 3.2.21, and 3.2.22, it follows that

$$\begin{aligned}
&\int_0^1 \frac{\big(J_{n+1/2}(2cr)\big)^2}{r} B_{1-r^2}(1+n,\tfrac{1}{2})\,dr \\
&\quad = \frac{\Gamma(n+1)}{\Gamma(n+\frac{3}{2})} \Big(\frac{\sqrt{\pi}}{2n+1} - \frac{1}{2\pi}\Big(\sqrt{\pi} + \sum_{k=1}^{n} \frac{\Gamma(k+\frac{1}{2})}{\Gamma(k+1)}\Big)\frac{\log(c)}{c}\Big) + o\Big(\frac{\log(c)}{c}\Big),
\end{aligned} \tag{3.80}$$

as $c \to \infty$, where $0 \le r \le 1$ and n is a nonnegative integer. Applying Lemma 3.2.17,

$$\begin{aligned}
&\int_0^1 (J_{n+1/2}(2cr))^2 B_{1-r^2}(1+n,\tfrac{1}{2})\,dr \\
&\quad = \frac{\Gamma(n+1)}{\Gamma(n+\frac{3}{2})} \Big(\frac{\sqrt{\pi}}{2n+1} - \frac{1}{\pi}\frac{\Gamma(n+\frac{3}{2})}{\Gamma(n+1)}\frac{\log(c)}{c}\Big) + o\Big(\frac{\log(c)}{c}\Big) \\
&\quad = \frac{\sqrt{\pi}\,\Gamma(n+1)}{2(n+\frac{1}{2})\Gamma(n+\frac{3}{2})} - \frac{1}{\pi}\frac{\log(c)}{c} + o\Big(\frac{\log(c)}{c}\Big),
\end{aligned} \tag{3.81}$$

as $c \to \infty$, where $0 \le r \le 1$ and n is a nonnegative integer. Therefore,

$$\int_0^1 \frac{(J_{p/2+1}(2cr))^2}{r} B_{1-r^2}(\tfrac{p}{2}+\tfrac{3}{2},\tfrac{1}{2})\,dr = \frac{\sqrt{\pi}\,\Gamma(\frac{p}{2}+\frac{3}{2})}{(p+2)\Gamma(\frac{p}{2}+2)} - \frac{1}{\pi}\frac{\log(c)}{c} + o\Big(\frac{\log(c)}{c}\Big), \tag{3.82}$$

as $c \to \infty$, for all $0 \leq r \leq 1$ and odd integers $p \geq -1$.

The proof in the case when $p \geq 0$ is an even integer is essentially identical.

■

The Area and Volume of a Hypersphere

The following theorem provides well-known formulas for the volume and area of a $(p+2)$-dimensional hypersphere. The formulas can be found in, for example, [17].

Theorem 3.2.26 *Suppose that $S^{p+2}(r) = \{x \in \mathbb{R}^{p+2} : \|x\| = r\}$ denotes the $(p+2)$-dimensional hypersphere of radius $r > 0$. Suppose further that $A_{p+2}(r)$ denotes the area of $S^{p+2}(r)$ and $V_{p+2}(r)$ denotes the volume enclosed by $S^{p+2}(r)$. Then*

$$A_{p+2}(r) = \frac{2\pi^{p/2+1}}{\Gamma(\frac{p}{2}+1)} r^{p+1}, \tag{3.83}$$

and

$$V_{p+2}(r) = \frac{\pi^{p/2+1}}{\Gamma(\frac{p}{2}+2)} r^{p+2}. \tag{3.84}$$

The following theorem provides a formula for the volume of the intersection of two $(p+2)$-dimensional hyperspheres (see, for example, [17]).

Theorem 3.2.27 *Suppose that $p \geq -1$ is an integer, let B denote the closed unit ball in $\mathbb{R}^{p+2}$, and let $B(c)$ denote the set $\{x \in \mathbb{R}^{p+2} : \|x\| \leq c\}$, where $c > 0$. Then*

$$\int_{\mathbb{R}^D} \mathbb{1}_B(u-t)\mathbb{1}_B(t)\, dt = V_{p+2}(1) \frac{B_{1-\|u\|^2/4}(\frac{p}{2}+\frac{3}{2},\frac{1}{2})}{B(\frac{p}{2}+\frac{3}{2},\frac{1}{2})}, \tag{3.85}$$

for all $u \in B(2)$, where $B(a,b)$ denotes the beta function, $B_x(a,b)$ denotes the incomplete beta function, V_{p+2} is defined by (3.84), and $\mathbb{1}_A$ is defined via the formula

$$\mathbb{1}_A(x) = \begin{cases} 1 & \textit{if } x \in A, \\ 0 & \textit{if } x \notin A. \end{cases} \tag{3.86}$$

3.2.7 Spherical Harmonics in $\mathbb{R}^{p+2}$

Suppose that S^{p+1} denotes the unit sphere in $\mathbb{R}^{p+2}$. The spherical harmonics are a set of real-valued continuous functions on S^{p+1}, which are orthonormal and complete in $L^2(S^{p+1})$. The spherical harmonics of degree $N \geq 0$ are denoted by $S_N^1, S_N^2, \ldots, S_N^\ell, \ldots, S_N^{h(N)} \colon S^{p+1} \to \mathbb{R}$, where

$$h(N) = (2N + p)\frac{(N + p - 1)!}{p!\, N!}, \tag{3.87}$$

for all nonnegative integers N.

The following theorem defines the spherical harmonics as the values of certain harmonic, homogeneous polynomials on the sphere (see, for example, [3]).

Theorem 3.2.28 *For each spherical harmonic S_N^ℓ, where $N \geq 0$ and $1 \leq \ell \leq h(N)$ are integers, there exists a polynomial $K_N^\ell \colon \mathbb{R}^{p+2} \to \mathbb{R}$ which is harmonic, i.e.*

$$\nabla^2 K_N^\ell(x) = 0, \tag{3.88}$$

for all $x \in \mathbb{R}^{p+2}$, and homogenous of degree N, i.e.

$$K_N^\ell(\lambda x) = \lambda^N K_N^\ell(x), \tag{3.89}$$

for all $x \in \mathbb{R}^{p+2}$ and $\lambda \in \mathbb{R}$, such that

$$S_N^\ell(\xi) = K_N^\ell(\xi), \tag{3.90}$$

for all $\xi \in S^{p+1}$.

The following lemma follows immediately from the orthonormality of spherical harmonics and Theorem 3.2.28.

Lemma 3.2.29 *For all $N > 0$ and for all $1 \leq \ell \leq h(N)$,*

$$\int_{S^{p+1}} S_N^\ell(x)dx = 0. \tag{3.91}$$

For $N = 0$ and $\ell = 1$, S_N^ℓ is the constant function defined by the formula

$$S_0^1(x) = A_{p+2}(1)^{(-1/2)} \tag{3.92}$$

where A_{p+2} is defined in (3.83).

The following theorem is proved in, for example, [3].

Theorem 3.2.30 *Suppose that N is a nonnegative integer. Then there are exactly*

$$(2N+p)\frac{(N+p-1)!}{p!\,N!} \tag{3.93}$$

linearly independent, harmonic, homogenous polynomials of degree N in $\mathbb{R}^{p+2}$.

The following theorem states that for any orthogonal matrix U, the function $S_N^\ell(U\xi)$ is expressible as a linear combination of $S_N^1(\xi), S_N^2(\xi), \ldots, S_N^{h(N)}(\xi)$ (see, for example, [3]).

Theorem 3.2.31 *Suppose that N is a nonnegative integer, and that*

$$S_N^1, S_N^2, \dots, S_N^{h(N)} \colon S^{p+1} \to \mathbb{R} \tag{3.94}$$

are a complete set of orthonormal spherical harmonics of degree N. Suppose further that U is a real orthogonal matrix of dimension $p+2 \times p+2$. Then, for each integer $1 \leq \ell \leq h(N)$, there exist real numbers $v_{\ell,1}, v_{\ell,2}, \dots, v_{\ell,h(N)}$ such that

$$S_N^\ell(U\xi) = \sum_{k=1}^{h(N)} v_{\ell,k} S_N^k(\xi), \tag{3.95}$$

for all $\xi \in S^{p+1}$. Furthermore, if V is the $h(N) \times h(N)$ real matrix with elements $v_{i,j}$ for all $1 \leq i,j \leq h(N)$, then V is also orthogonal.

Remark 3.2.7 *From Theorem (3.2.31), we observe that the space of linear combinations of functions S_N^ℓ is invariant under all rotations and reflections of S^{p+1}.*

The following theorem states that if an integral operator acting on the space of functions $S^{p+1} \to \mathbb{R}$ has a kernel depending only on the inner product, then the spherical harmonics are eigenfunctions of that operator (see, for example, [3]).

Theorem 3.2.32 (Funk-Hecke) *Suppose that $F \colon [-1,1] \to \mathbb{R}$ is a continuous function, and that $S_N \colon S^{p+1} \to \mathbb{R}$ is any spherical harmonic of degree N. Then*

$$\int_\Omega F(\langle \xi, \eta \rangle) S_N(\xi)\, d\Omega(\xi) = \lambda_N S_N(\eta), \tag{3.96}$$

for all $\eta \in S^{p+1}$, where $\langle \cdot, \cdot \rangle$ denotes the inner product in $\mathbb{R}^{p+2}$, the integral is taken over the whole area of the hypersphere Ω, and λ_N depends only on the function F.

3.2.8 Generalized Prolate Spheroidal Functions

Basic Facts

In this section, we summarize several facts about generalized prolate spheroidal functions (GPSFs). Let B denote the closed unit ball in $\mathbb{R}^{p+2}$. Given a real number $c > 0$, we define the operator $F_c \colon L^2(B) \to L^2(B)$ via the formula

$$F_c[\psi](x) = \int_B \psi(t) e^{ic\langle x,t\rangle} \, dt, \tag{3.97}$$

for all $x \in B$, where $\langle \cdot, \cdot \rangle$ denotes the inner product on $\mathbb{R}^{p+2}$. Clearly, F_c is compact. Obviously, F_c is also normal, but not self-adjoint. We denote the eigenvalues of F_c by $\lambda_0, \lambda_1, \dots, \lambda_n, \dots$, and assume that $|\lambda_j| \geq |\lambda_{j+1}|$ for each non-negative integer j. For each non-negative integer j, we denote by ψ_j the eigenfunction corresponding to λ_j, so that

$$\lambda_j \psi_j(x) = \int_B \psi_j(t) e^{ic\langle x,t\rangle} \, dt, \tag{3.98}$$

for all $x \in B$. We assume that $\|\psi_j\|_{L^2(B)} = 1$ for each j. The following theorem is proved in [26] and describes the eigenfunctions and eigenvalues of F_c.

Theorem 3.2.33 *Suppose that $c > 0$ is a real number and that F_c is defined by* (3.97)*. Then the eigenfunctions $\psi_0, \psi_1, \dots, \psi_n, \dots$ of F_c are real, orthonormal, and complete in $L^2(B)$. For each j, the eigenfunction ψ_j is either even, in the sense that $\psi_j(-x) = \psi_j(x)$ for all $x \in B$, or odd, in the sense that $\psi_j(-x) = -\psi_j(x)$ for all $x \in B$. The eigenvalues corresponding to even eigenfunctions are real, and the eigenvalues corresponding to odd eigenfunctions are purely imaginary. The domain on which the eigenfunctions are defined can be extended from B to $\mathbb{R}^{p+2}$ by requiring that* (3.98) *hold for all $x \in \mathbb{R}^{p+2}$; the eigenfunctions will then be orthogonal on*

$\mathbb{R}^{p+2}$ *and complete in the class of band-limited functions with bandlimit* c.

We define the self-adjoint operator $Q_c \colon L^2(B) \to L^2(B)$ via the formula

$$Q_c = \left(\frac{c}{2\pi}\right)^{p+2} F_c^* \cdot F_c. \tag{3.99}$$

Since F_c is normal, it follows that Q_c has the same eigenfunctions as F_c, and that the jth eigenvalue μ_j of Q_c is connected to λ_j via the formula

$$\mu_j = \left(\frac{c}{2\pi}\right)^{p+2} |\lambda_j|^2. \tag{3.100}$$

We also observe that

$$Q_c[\psi](x) = \left(\frac{c}{2\pi}\right)^{p/2+1} \int_B \frac{J_{p/2+1}\big(c\|x-t\|\big)}{\|x-t\|^{p/2+1}} \psi(t)\, dt, \tag{3.101}$$

for all $x \in \mathbb{R}^{p+2}$, where J_ν denotes the Bessel functions of the first kind and $\|\cdot\|$ denotes Euclidean distance in $\mathbb{R}^{p+2}$ (see Appendix A for a proof).

We observe that

$$Q_c[\psi](x) = \mathbb{1}_B(x) \cdot \mathcal{F}^{-1}\big[\mathbb{1}_{B(c)}(t) \cdot \mathcal{F}[\psi](t)\big](x), \tag{3.102}$$

where $\mathcal{F} \colon L^2(\mathbb{R}^{p+2}) \to L^2(\mathbb{R}^{p+2})$ is the $(p+2)$-dimensional Fourier transform, $B(c)$ denotes the set $\{\, x \in \mathbb{R}^{p+2} : \|x\| \le c \,\}$, and $\mathbb{1}_A$ is defined via the formula

$$\mathbb{1}_A(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A. \end{cases} \tag{3.103}$$

From (3.102) it follows that $\mu_j < 1$ for all j.

We observe further that Q_c is closely related to the operator $P_c \colon L^2(\mathbb{R}^{p+2}) \to$

$L^2(\mathbb{R}^{p+2})$, defined via the formula

$$P_c[\psi](x) = \Big(\frac{c}{2\pi}\Big)^{p/2+1} \int_{\mathbb{R}^{p+2}} \frac{J_{p/2+1}\big(c\|x-t\|\big)}{\|x-t\|^{p/2+1}} \psi(t)\, dt, \tag{3.104}$$

which is the orthogonal projection onto the space of bandlimited functions on $\mathbb{R}^{p+2}$ with bandlimit $c > 0$.

Eigenfunctions and Eigenvalues of F_c

In this section we describe the eigenvectors and eigenvalues of the operator F_c, defined in (3.97). Suppose that ψ is some eigenfunction of the integral operator F_c, with corresponding complex eigenvalue λ, so that

$$\lambda\psi(x) = \int_B \psi(t) e^{ic\langle x,t\rangle}\, dt, \tag{3.105}$$

for all $x \in B$ (see Theorem 3.2.33).

Observation 3.2.8 *The operator F_c, defined by* (3.97)*, is spherically symmetric in the sense that, for any $(p+2)\times(p+2)$ orthogonal matrix U, F_c commutes with the operator $\hat{U}\colon L^2(B) \to L^2(B)$, defined via the formula*

$$\hat{U}[\psi](x) = \psi(Ux), \tag{3.106}$$

for all $x \in B$. Hence, the problem of finding the eigenfunctions and eigenvalues of F_c is amenable to the separation of variables.

Suppose that

$$\psi(x) = \Phi_N^\ell(\|x\|) S_N^\ell(x/\|x\|), \tag{3.107}$$

where S_N^ℓ, $\ell = 0, 1, \ldots, h(N,p)$ denotes the spherical harmonics of degree N (see Section 3.2.7), and $\Phi_N^\ell(r)$ is a real-valued function defined on the interval $[0,1]$. We observe that

$$e^{ic\langle x,t\rangle} = \sum_{N=0}^{\infty} \sum_{\ell=1}^{h(N,p)} i^N (2\pi)^{p/2+1} \frac{J_{N+p/2}(c\|x\|\|t\|)}{(c\|x\|\|t\|)^{p/2}} S_N^\ell(x/\|x\|) S_N^\ell(t/\|t\|), \tag{3.108}$$

where $x, t \in B$, and where J_ν denotes the Bessel functions of the first kind (see Section VII of [26] for a proof). Substituting (3.107) and (3.108) into (3.105), we find that

$$\lambda \Phi_N^\ell(r) = i^N (2\pi)^{p/2+1} \int_0^1 \frac{J_{N+p/2}(cr\rho)}{(cr\rho)^{p/2}} \Phi_N^\ell(\rho) \rho^{p+1} \, d\rho, \tag{3.109}$$

for all $0 \le r \le 1$. We define the operator $H_{N,c} \colon L^2\big([0,1], \rho^{p+1}\, d\rho\big) \to L^2\big([0,1], \rho^{p+1}\, d\rho\big)$ via the formula

$$H_{N,c}[\Phi](r) = \int_0^1 \frac{J_{N+p/2}(cr\rho)}{(cr\rho)^{p/2}} \Phi(\rho) \rho^{p+1} \, d\rho, \tag{3.110}$$

where $0 \le r \le 1$, and observe that $H_{N,c}$ is clearly compact and self-adjoint, and does not depend on ℓ. Dropping the index ℓ, we denote by $\beta_{N,0}, \beta_{N,1}, \ldots, \beta_{N,n}, \ldots$ the eigenvalues of $H_{N,c}$, and assume that $|\beta_{N,n}| \ge |\beta_{N,n+1}|$ for each nonnegative integer n. For each nonnegative integer n, we let $\Phi_{N,n}$ denote the eigenvector corresponding to eigenvalue $\beta_{N,n}$, so that

$$\beta_{N,n} \Phi_{N,n}(r) = \int_0^1 \frac{J_{N+p/2}(cr\rho)}{(cr\rho)^{p/2}} \Phi_{N,n}(\rho) \rho^{p+1} \, d\rho, \tag{3.111}$$

for all $0 \le r \le 1$. Clearly, the eigenfunctions $\Phi_{N,n}$ are purely real. We assume that $\|\Phi_{N,n}\|_{L^2([0,1],\rho^{p+1}\, d\rho)} = 1$ and that $\Phi_{N,n}(1) > 0$ for each nonnegative integer N and n (see Theorem 3.8.6). It follows from (3.111) and (3.109) that the eigenvectors and eigenvalues of F_c are given by the formulas

$$\psi^{\ell}_{N,n}(x) = \Phi_{N,n}(\|x\|)S^{\ell}_{N}(x/\|x\|), \tag{3.112}$$

and

$$\lambda^{\ell}_{N,n} = i^N(2\pi)^{p/2+1}\beta_{N,n}, \tag{3.113}$$

respectively, where $x \in B$, N and n are nonnegative integers, and ℓ is an integer so that $1 \le \ell \le h(N,p)$ (see Section 3.2.7). We note in formula (3.113) the expected degeneracy of eigenvalues due to the spherical symmetry of the integral operator F_c (see Observation 3.2.8); we denote $\lambda^{\ell}_{N,n}$ by $\lambda_{N,n}$ where the meaning is clear. We also make the following observation.

Observation 3.2.9 *The domain on which the functions $\Phi_{N,n}$ are defined may be extended from the interval $[0,1]$ to the complex plane $\mathbb{C}$ by requiring that* (3.105) *hold for all $r \in \mathbb{C}$. Moreover, the functions $\Phi_{N,n}$, extended in this way, are entire.*

The Dual Nature of GPSFs

In this section, we observe that the eigenfunctions $\Phi_{N,0}, \Phi_{N,1}, \dots, \Phi_{N,n}, \dots$ of the integral operator $H_{N,c}$, defined in (3.110), are also the eigenfunctions of a certain differential operator.

Let $\beta_{N,n}$ denote the eigenvalue corresponding to the eigenfunction $\Phi_{N,n}$, for all nonnegative integers N and n, so that

$$\beta_{N,n}\Phi_{N,n}(r) = \int_0^1 \frac{J_{N+p/2}(cr\rho)}{(cr\rho)^{p/2}}\Phi_{N,n}(\rho)\rho^{p+1}\, d\rho, \tag{3.114}$$

where $0 \le r \le 1$, N and n are nonnegative integers, and J_ν denotes the Bessel functions of the first kind (see (3.111)). Making the substitutions

$$\varphi_{N,n}(r) = r^{(p+1)/2}\Phi_{N,n}(r), \tag{3.115}$$

and

$$\gamma_{N,n} = c^{(p+1)/2}\beta_{N,n}, \tag{3.116}$$

we observe that

$$\gamma_{N,n}\varphi_{N,n}(r) = \int_0^1 J_{N+p/2}(cr\rho)\sqrt{cr\rho}\,\varphi_{N,n}(\rho)\,d\rho, \tag{3.117}$$

where $0 \le r \le 1$, and N and n are arbitrary nonnegative integers. We define the operator $M_{N,c}\colon L^2([0,1]) \to L^2([0,1])$ via the formula

$$M_{N,c}[\varphi](r) = \int_0^1 J_{N+p/2}(cr\rho)\sqrt{cr\rho}\,\varphi(\rho)\,d\rho, \tag{3.118}$$

where $0 \le r \le 1$, and N is an arbitrary nonnegative integer. Obviously, $M_{N,c}$ is compact and self-adjoint. Clearly, the eigenvalues of $M_{N,c}$ are $\gamma_{N,0}, \gamma_{N,1}, \dots, \gamma_{N,n}, \dots$, and $\varphi_{N,n}$ is the eigenfunction corresponding to eigenvalue $\gamma_{N,n}$, for each nonnegative integer n.

We define the differential operator $L_{N,c}$ via the formula

$$L_{N,c}[\varphi](x) = \frac{d}{dx}\left((1-x^2)\frac{d\varphi}{dx}(x)\right) + \left(\frac{\frac{1}{4} - (N+\frac{p}{2})^2}{x^2} - c^2x^2\right)\varphi(x), \tag{3.119}$$

where $0 < x < 1$, N is a nonnegative integer, and φ is twice continuously differentiable. Let C be the class of functions φ which are bounded and twice continuously differentiable on the interval $(0,1)$, such that $\varphi'(0) = 0$ if $p = -1$ and $N = 0$, and

$\varphi(0) = 0$ otherwise. Then it is easy to show that, operating on functions in class C, $L_{N,c}$ is self-adjoint. From Sturmian theory we obtain the following theorem (see [26]).

Theorem 3.2.34 *Suppose that $c > 0$, N is a nonnegative integer, and $L_{N,c}$ is defined via* (3.119)*. Then there exists a strictly increasing unbounded sequence of positive numbers $\chi_{N,0} < \chi_{N,1} < \ldots$ such that for each nonnegative integer n, the differential equation*

$$L_{N,c}[\varphi](x) + \chi_{N,n}\varphi(x) = 0 \tag{3.120}$$

has a solution which is bounded and twice continuously differentiable on the interval $(0,1)$, so that $\varphi'(0) = 0$ if $p = -1$ and $N = 0$, and $\varphi(0) = 0$ otherwise.

The following theorem is proved in [26].

Theorem 3.2.35 *Suppose that $c > 0$, N is a nonnegative integer, and the operators $M_{N,c}$ and $L_{N,c}$ are defined via* (3.118) *and* (3.119) *respectively. Suppose also that $\varphi\colon (0,1) \to \mathbb{R}$ is in $L^2([0,1])$, is twice differentiable, and that $\varphi'(0) = 0$ if $p = -1$ and $N = 0$, and $\varphi(0) = 0$ otherwise. Then*

$$L_{N,c}\big[M_{N,c}[\varphi]\big](x) = M_{N,c}\big[L_{N,c}[\varphi]\big](x), \tag{3.121}$$

for all $0 < x < 1$.

Remark 3.2.10 *Since Theorem 3.2.34 shows that the eigenvalues of $L_{N,c}$ are not degenerate, Theorem 3.2.35 implies that $L_{N,c}$ and $M_{N,c}$ have the same eigenfunctions.*

Zernike Polynomials and GPSFs

In this section we describe the relationship between Zernike polynomials and GPSFs. We use $\varphi^c_{N,n}$, where $c > 0$ and N and n are arbitrary nonnegative integers, to denote the nth eigenfunction of $L_{N,c}$, defined in (3.119); we denote by $\chi_{N,n}(c)$ the eigenvalue corresponding to eigenfunction $\varphi^c_{N,n}$.

For $c = 0$, the eigenfunctions and eigenvalues of the differential operator $L_{N,c}$, defined in (3.119), are

$$\overline{T}_{N,n}(x) \tag{3.122}$$

and

$$\chi_{N,n}(0) = (N + \tfrac{p}{2} + 2n + \tfrac{1}{2})(N + \tfrac{p}{2} + 2n + \tfrac{3}{2}), \tag{3.123}$$

respectively, where $0 \leq x \leq 1$, N and n are arbitrary nonnegative integers, and $\overline{T}_{N,n}$ is defined in (3.34).

For small $c > 0$, the connection between Zernike polynomials and GPSFs is given by the formulas

$$\varphi^c_{N,n}(x) = \overline{T}_{N,n}(x) + o(c^2), \tag{3.124}$$

and

$$\chi_{N,n}(c) = \chi_{N,n}(0) + o(c^2), \tag{3.125}$$

as $c \to 0$, where $0 \leq x \leq 1$ and N and n are arbitrary nonnegative integers (see [26]).

For $c > 0$, the functions $T_{N,n}$ are also related to the integral operator $M_{N,c}$, defined in (3.118), via the formula

$$M_{N,c}\big[T_{N,n}\big](x) = \int_0^1 J_{N+p/2}(cxy)\sqrt{cxy}\,T_{N,n}(y)\,dy = \frac{(-1)^n J_{N+p/2+2n+1}(cx)}{\sqrt{cx}}, \tag{3.126}$$

where $x \geq 0$ and N and n are arbitrary nonnegative integers (see Equation (85) in [13]).

3.3 Analytical Apparatus

In this section, we provide analytical apparatus relating to GPSFs that will be used in numerical schemes in subsequent sections.

3.3.1 Properties of GPSFs

The following theorem provides a formula for ratios of eigenvalues $\beta_{N,n}$ (see (3.111)), and finds use in the numerical evaluation of $\beta_{N,n}$. A proof follows immediately from Theorem 7.1 of [22].

Theorem 3.3.1 *Suppose that N is a nonnegative integer. Then*

$$\frac{\beta_{N,m}}{\beta_{N,n}} = \frac{\int_0^1 x\Phi'_{N,n}(x)\Phi_{N,m}(x)x^{p+1}\,dx}{\int_0^1 x\Phi'_{N,m}(x)\Phi_{N,n}(x)x^{p+1}\,dx}, \tag{3.127}$$

for each nonnegative integers n and m.

3.3.2 Decay of the Expansion Coefficients of GPSFs in Zernike Polynomials

Since the functions $\Phi_{N,n}$ are analytic on $\mathbb{C}$ for all nonnegative integers N and n (see Observation 3.2.9), and $\Phi^{(k)}_{N,n}(0) = 0$ for $k = 0, 1, \ldots, N-1$ (see Theorem 3.8.5), the functions $\Phi_{N,n}$ are representable by a series of Zernike polynomials of the form

$$\Phi_{N,n}(r) = \sum_{k=0}^{\infty} a_{n,k} \overline{R}_{N,k}(r), \tag{3.128}$$

for all $0 \leq r \leq 1$, where $a_{n,0}, a_{n,1}, \ldots$ satisfy

$$a_{n,k} = \int_0^1 \overline{R}_{N,k}(r)\Phi_{N,n}(r)dr \tag{3.129}$$

where $\overline{R}_{N,n}$ is defined in (3.23). The following technical lemma will be used in the proof of Theorem 3.3.3.

Lemma 3.3.2 *For any integer $p \geq -1$, for all $c > 0$, and for all $\rho \in [0,1]$,*

$$\left| \int_0^1 \frac{J_{N+p/2}(cr\rho)}{(cr\rho)^{p/2}} \overline{R}_{N,k}(r) r^{p+1} dr \right| < \left(\frac{1}{2}\right)^{N+p/2+2k+1} \tag{3.130}$$

for any non-negative integers N, k such that $N + 2k \geq ec$ where $\overline{R}_{N,n}$ is defined in (3.23) and $J_{N+p/2}$ is a Bessel function of the first kind.

Proof. According to equation (85) in [13],

$$\int_0^1 \frac{J_{N+p/2}(cr\rho)}{(cr\rho)^{p/2}} R_{N,k}(r) r^{p+1} dr = \frac{(-1)^n J_{N+p/2+2k+1}(c\rho)}{(c\rho)^{p/2+1}}, \tag{3.131}$$

where $J_{N+p/2}$ is a Bessel function of the first kind. Applying Lemma 3.2.2 to (3.131), we obtain

$$\left|\int_0^1 \frac{J_{N+p/2}(cr\rho)}{(cr\rho)^{p/2}}\overline{R}_{N,k}(r)r^{p+1}dr\right| \leq \frac{(c\rho/2)^{N+p/2+2k+1}}{(c\rho)^{p/2+1}}\frac{\sqrt{2(N+p/2+2k+1)}}{\Gamma(N+p/2+2k+2)}. \tag{3.132}$$

Combining Lemma 3.2.1 and (3.132), we have

$$\begin{aligned}\left|\int_0^1 \frac{J_{N+p/2}(cr\rho)}{(cr\rho)^{p/2}}\overline{R}_{N,k}(r)r^{p+1}dr\right| &\leq \left(\frac{1}{2}\right)^{N+p/2+2k+1}(c\rho)^{N+2k}\frac{\sqrt{2(2k+N)}}{\Gamma(2k+N+1)}\\ &\leq \left(\frac{1}{2}\right)^{N+p/2+2k+1}\end{aligned} \tag{3.133}$$

for $N+2k \geq ec$. ■

The following theorem shows that the coefficients $a_{N,k}$ of GPSFs in a Zernike polynomial basis decay exponentially and establishes a bound for the decay rate.

Theorem 3.3.3 *For all non-negative integers N, n, k and for all $c > 0$,*

$$\int_0^1 \Phi_{N,n}(r)\overline{R}_{N,k}r^{p+1}dr < (p+2)^{-1/2}(\beta_{N,n})^{-1}\left(\frac{1}{2}\right)^{N+p/2+2k+1} \tag{3.134}$$

where $N+2k \geq ec$.

Proof. Combining (3.114) and (3.129),

$$\begin{aligned}&\int_0^1 \Phi_{N,n}(r)\overline{R}_{N,k}r^{p+1}\\ &\qquad= \int_0^1 (\beta_{N,n})^{-1}\left(\int_0^1 \frac{J_{N+p/2}(cr\rho)}{(cr\rho)^{p/2}}\Phi_{N,n}(\rho)\rho^{p+1}d\rho\right)\overline{R}_{N,k}(r)r^{p+1}dr.\end{aligned} \tag{3.135}$$

Changing the order of integration of (3.135),

$$\begin{aligned}\int_0^1 &\Phi_{N,n}(r)\overline{R}_{N,k}r^{p+1} \\ &= (\beta_{N,n})^{-1}\int_0^1 \Phi_{N,n}(\rho)\rho^{p+1}\int_0^1 \frac{J_{N+p/2}(cr\rho)}{(cr\rho)^{p/2}}\overline{R}_{N,k}(r)r^{p+1}drd\rho.\end{aligned} \tag{3.136}$$

Applying Lemma 3.3.2 to (3.136) and applying Cauchy-Schwarz, we obtain

$$\begin{aligned}\int_0^1 \Phi_{N,n}(r)\overline{R}_{N,k}r^{p+1} &\le (\beta_{N,n})^{-1}\left(\frac{1}{2}\right)^{N+p/2+2k+1}\int_0^1 \Phi_{N,n}(\rho)\rho^{p+1}d\rho \\ &\le (p+2)^{-1/2}(\beta_{N,n})^{-1}\left(\frac{1}{2}\right)^{N+p/2+2k+1}.\end{aligned} \tag{3.137}$$

for $N + 2k \geq ec$. ■

3.3.3 Tridiagonal Nature of $L_{N,c}$

In this section, we show that in the basis of $\overline{T}_{N,n}$ (see (3.34)), the matrix representing differential operator $L_{N,c}$ (see (3.119)) is symmetric and tridiagonal.

The following lemma provides an identity relating the differential operator $L_{N,c}$ to $\overline{T}_{N,n}$.

Lemma 3.3.4 *For all non-negative integers N, n and real numbers $c > 0$*

$$L_{N,c}[\overline{T}_{N,n}] = -\chi_{N,n}\overline{T}_{N,n}(x) - c^2x^2\overline{T}_{N,n}(x) \tag{3.138}$$

for all $x \in [0,1]$ where $\chi_{N,n}$ is defined in (3.40) and $L_{N,c}$ is defined in (3.119).

Proof. Applying $L_{N,c}$ to $\overline{T}_{N,n}$, we obtain

$$L_{N,c}[\overline{T}_{N,n}](x) = (1-x^2)\overline{T}''_{N,n}(x) - 2x\overline{T}_{N,n}(x) + \left(\frac{\frac{1}{4}-(N+\frac{p}{2})^2}{x^2} - c^2x^2\right)\overline{T}_{N,n}(x).$$

$$\tag{3.139}$$

Identity (3.138) follows immediately from the combination of (3.39) and (3.139). ■

The following theorem follows readily from the combination of Lemma 3.3.4 and Lemma 3.2.12.

Theorem 3.3.5 *For any non-negative integer N, any integer $n \geq 1$, and for all $r \in (0,1)$,*

$$L_{N,c}[\overline{T}_{N,n}] = a_n \overline{T}_{N,n-1}(r) + b_n \overline{T}_{N,n}(r) + c_n \overline{T}_{N,n+1}(r) \tag{3.140}$$

where

$$\begin{aligned} a_n =& \frac{-c^2(n+N+p/2)n}{(2n+N+p/2)\sqrt{2n+N+p/2+1}\sqrt{2n+N+p/2-1}} \\ b_n =& \frac{-c^2(N+p/2)^2}{2(2n+N+p/2)(2n+N+p/2+2)} - \frac{c^2}{2} + \chi_{N,n} \\ c_n =& \frac{-c^2(n+1+N+p/2)(n+1)}{(2n+N+p/2+2)\sqrt{2n+N+p/2+3}\sqrt{2n+N+p/2+1}} \end{aligned} \tag{3.141}$$

and $\chi_{N,n}$ is defined in (3.40).

Observation 3.3.1 *It follows immediately from Theorem 3.3.5 that the matrix corresponding to the differential operator $L_{N,c}$ acting on the $\overline{T}_{N,n}$ basis is symmetric* *and tridiagonal. Specifically, for any positive integer n and for all $r \in (0,1)$,*

$$\begin{bmatrix} b_0 & c_0 & & & & 0 \\ c_0 & b_1 & c_1 & & & \\ & c_1 & b_2 & c_2 & & \\ & & \ddots & \ddots & \ddots & \\ & & & c_{n-2} & b_{n-1} & c_{n-1} \\ 0 & & & & c_{n-1} & b_n \end{bmatrix} \begin{bmatrix} \overline{T}_{N,0}(r) \\ \\ \vdots \\ \\ \\ \overline{T}_{N,n}(r) \end{bmatrix} + \begin{bmatrix} 0 \\ \\ \vdots \\ \\ 0 \\ c_n \overline{T}_{N,n+1}(r) \end{bmatrix} = \begin{bmatrix} \overline{T}_{N,0}(r) \\ \\ \vdots \\ \\ \\ \overline{T}_{N,n}(r) \end{bmatrix} \tag{3.142}$$

where b_k and c_k are defined in (3.141) and $\overline{T}_{N,k}$ is defined in (3.34).

Observation 3.3.2 *Let A be the infinite symmetric tridiagonal matrix satisfying $A_{1,1} = b_0$, $A_{1,2} = c_0$ and for all integers $k \geq 2$,*

$$\begin{aligned} A_{k,k-1} &= c_{k-1} \\ A_{k,k} &= b_k \\ A_{k,k+1} &= c_k, \end{aligned} \tag{3.143}$$

where b_k, c_k are defined in (3.141). That is,

$$A = \begin{bmatrix} b_0 & c_0 & & & \\ c_0 & b_1 & c_1 & & \\ & c_1 & b_2 & c_2 & \\ & & \ddots & \ddots & \ddots \end{bmatrix}. \tag{3.144}$$

Suppose further that we define the infinite vector α_n by the equation

$$a_n = (a_{n,0}, a_{n,1}, ...)^T, \tag{3.145}$$

where $a_{n,k}$ is defined in (3.129). By the combination of Theorem 3.2.35 and Remark

3.23, we know that $\varphi_{N,n}$ is the eigenfunction corresponding to $\chi_{N,n}(c)$, the nth smallest eigenvalue of differential operator $L_{N,c}$. Therefore,

$$A\alpha_n = \chi_{N,n}(c)\alpha_n. \tag{3.146}$$

Furthermore, the $a_{n,k}$ decay exponentially fast in k (see Theorem 3.3.3).

Remark 3.3.3 *The eigenvalues $\chi_{N,n}$ of differential operator $L_{N,c}$ and the coefficients in the Zernike expansion of the eigenfunctions $\Phi_{N,n}$ can be computed numerically to high relative precision by the following process. First, we reduce the infinite dimensional matrix A (see (3.144)) to A_K, its upperleft $K \times K$ submatrix where K is chosen, using Theorem 3.3.3, so that $a_{n,K-1}$ is smaller than machine precision and is in the regime of exponential decay. We then use standard algorithms to find the eigenvalues and eigenvectors of matrix A_K. See Algorithm 3.4.1 for a more detailed description of the algorithm.*

3.4 Numerical Evaluation of GPSFs

In this section, we describe an algorithm (Algorithm 3.4.1) for the evaluation of $\Phi_{N,n}(r)$ (see (3.111)) for all $r \in [0, 1]$.

Algorithm 3.4.1

Step 1. Use Theorem 3.3.3 to determine how many terms are needed in a Zernike expansion of $\Phi_{N,n}$. We assume that we want a K term expansion.

Step 2. Generate A_K, the symmetric, tri-diagonal, upper-left $K \times K$ sub-matrix of A (see (3.144)).

Step 3. Use an eigensolver to find the eigenvector, $\tilde{a}_n$, corresponding to the $n+1^{\text{th}}$ largest eigenvalue, $\tilde{\chi}_{N,n}$. That is, find $\tilde{a}_n$ and $\tilde{\chi}_{N,n}$ such that

$$A_K \tilde{a}_n = \tilde{\chi}_{N,n} \tilde{a}_n \tag{3.147}$$

where we define the components of $\tilde{a}_{N,n}$ by the formula,

$$\tilde{a}_n = (a_{n,0}, a_{n,1}, ..., a_{n,K-1}). \tag{3.148}$$

Step 4. Evaluate $\Phi_{N,n}(r)$ by the expansion

$$\Phi_{N,n}(r) = \sum_{i=0}^{k} a_{n,i} \overline{R}_{N,i}(r) \tag{3.149}$$

where, $\overline{R}_{N,i}$ is evaluated via Lemma 3.2.8 and $a_{n,i}$ are the components of eigenvector (3.148) recovered in Step 3.

Remark 3.4.1 *It turns out that because of the structure of A_K, standard numerical algorithms will compute the components of eigenvector $\tilde{a}_n$ (see 3.148)), and thus the coefficients of a GPSF in a Zernike expansion, to high relative precision. In particular, the components of $\tilde{a}_n$ that are of magnitude far less than machine precision, are computed to high relative precision. For example, when using double-precision arithmetic, a component of $\tilde{a}_n$ of magnitude 10^{-100} will be computed in absolute precision to 116 digits. This fact is proved in a more general setting in [21].*

3.4.1 Numerical Evaluation of the Single Eigenvalue $\beta_{N,i}$

In this section, we describe a sum that can be used to evaluate the eigenvalue $\beta_{N,n}$ (see Theorem 3.111) for fixed n to high relative precision.

The following is a technical lemma will be used in the proof of Theorem 3.4.3.

Lemma 3.4.1 *For all non-negative integers N, k,*

$$\int_0^1 \rho^N \varphi_{N,k}(\rho) \rho^{\frac{p+1}{2}} d\rho = \frac{a_{k,0}}{\sqrt{2N+p+2}} \tag{3.150}$$

where $\varphi_{N,k}$ is defined in (3.115) and $p \geq -1$ is an integer.

Proof. Using (3.20),

$$\int_0^1 \rho^N \varphi_{N,k}(\rho) \rho^{\frac{p+1}{2}} d\rho = \int_0^1 R_{N,0}(\rho) \varphi_{N,k}(\rho) \rho^{\frac{p+1}{2}} d\rho \tag{3.151}$$

Applying (3.129) and (3.34) to (3.151), we obtain

$$\int_0^1 \rho^N \varphi_{N,k}(\rho) \rho^{\frac{p+1}{2}} d\rho = \frac{1}{\sqrt{2N+p+2}} \int_0^1 \overline{T}_{N,0}(\rho) \varphi_{N,k}(\rho) d\rho = \frac{a_{k,0}}{\sqrt{2N+p+2}}. \tag{3.152}$$

■

We will denote by $\varphi^*_{N,n}(r)$ the function on $[0,1]$ defined by the formula

$$\varphi^*_{N,n}(r) = \frac{\varphi_{N,n}(r)}{r^{N+\frac{p}{2}}} \tag{3.153}$$

where N, n are non-negative integers.

The following identity will be used in the proof of Theorem 3.4.3.

Lemma 3.4.2 *For all non-negative integers N, k,*

$$\varphi^*_{N,k}(0) = \sum_{i=0}^{\infty} a_{k,i} \sqrt{2(2i+N+p/2+1)} (-1)^i \binom{i+N+p/2}{i}. \tag{3.154}$$

where $\varphi^*_{N,k}$ is defined in (3.153) and $a_{k,i}$ is defined in (3.129).

Proof. Combining (3.153) and (3.46), we have

$$\varphi^*_{N,k}(r) = \frac{\varphi_{N,k}(r)}{r^{N+\frac{p}{2}}} = \sum_{i=0}^{\infty} a_{k,i} \frac{\overline{T}_{N,i}(r)}{r^{N+\frac{p}{2}}} = \sum_{i=0}^{\infty} a_{k,i} \overline{T}^*_{N,i}(r) \tag{3.155}$$

where $\overline{T}^*_{N,n}$ is defined in (3.46) and $\overline{T}_{N,n}$ is defined in (3.34). Identity (3.154) follows immediately from applying Lemma 3.2.13 to (3.155) and setting $r = 0$. ■

The following theorem provides a formula that can be used to compute $\beta_{N,n}$ (see (3.116)), an eigenvalue of integral operator $H_{N,c}$ (see (3.110).

Theorem 3.4.3 *For all non-negative integers N, k,*

$$\beta_{N,k} = \frac{a_{k,0} c^N (2^{N+p/2} \Gamma(N+p/2+1)\sqrt{2N+p+2})^{-1}}{\sum\limits_{i=0}^{\infty} a_{k,i} \sqrt{2(2i+N+p/2+1)}(-1)^i \binom{i+N+p/2}{i}} \tag{3.156}$$

where $\beta_{N,k}$ is defined in (3.114) and $a_{k,i}$ are defined in (3.129).

Proof. It is well known that $J_{N+p/2}$, a Bessel Function of the first kind, satisfies the identity

$$J_{N+p/2} = \left(\frac{cr\rho}{2}\right)^{N+p/2} \sum_{k=0}^{\infty} \frac{(-(cr\rho)^2/4)^k}{k!\Gamma(N+p/2+k+1)} \tag{3.157}$$

where $\Gamma(n)$ is the gamma function. Dividing both sides of (3.117) by $r^{N+\frac{(p+1)}{2}}$, we obtain the equation

$$\gamma_{N,k} \varphi^*_{N,k}(r) = \int_0^1 \frac{J_{N+p/2}(cr\rho)}{r^{N+\frac{p}{2}}} \sqrt{c\rho}\, \varphi_{N,k}(\rho) d\rho \tag{3.158}$$

where $\varphi^*_{N,k}$ is defined in (3.153). Setting $r = 0$, in (3.158) and subsituting in

(3.154) and (3.157), we obtain

$$\gamma_{N,k} = \int_0^1 \left(\frac{c\rho}{2}\right)^{N+p/2} \frac{(c\rho)^{1/2}}{\Gamma(N+p/2+1)} \varphi_{N,k}(\rho) d\rho \\ \left(\sum_{i=0}^{\infty} a_{k,i} \sqrt{2(2i+N+p/2+1)} (-1)^i \binom{i+N+p/2}{i}\right)^{-1}. \tag{3.159}$$

Equation (3.156) follows immediately from applying Lemma 3.4.1 and (3.116) to (3.159). ■

Remark 3.4.2 *For any non-negative integers N, k, the eigenvalue $\beta_{N,k}$ can be evaluated stably by first using Algorithm 3.4.1 to compute the eigenvector $\tilde{a}_k$ (see (3.148)), and then evaluating $\beta_{N,k}$ via sum (3.156) where $\tilde{a}_k$ are approximations to a_k. In (3.156), the sum*

$$\sum_{i=0}^{\infty} a_{k,i} \sqrt{2(2i+N+p/2+1)} (-1)^i \binom{i+N+p/2}{i} \tag{3.160}$$

can be computed to high relative precision by truncating the sum at a point when the partial sum up to that point is a factor of machine precision larger than the next term.

3.4.2 Numerical Evaluation of Eigenvalues $\beta_{N,0}, \beta_{N,1}, ..., \beta_{N,k}$

In this section, we describe an algorithm for numerically evaluating the eigenvalues $\beta_{N,0}, \beta_{N,1}, ..., \beta_{N,k}$ (see (3.111)) for any non-negative integers N, k (see Algorithm 3.4.2).

In Observation 3.4.3, we describe a stable numerical scheme for converting an expansion of the form

$$\sum_{i=0}^{K} x_i r \overline{T}'_{N,i}(r), \tag{3.161}$$

where $x_0, ..., x_K$ are real numbers, into an expansion of the form

$$\sum_{i=0}^{K} \alpha_i \overline{T}_{N,i}(r) \tag{3.162}$$

where $\alpha_0, ..., \alpha_K$ are real numbers, $\overline{T}_{N,n}(r)$ is defined in (3.34), and $\overline{T}'_{N,n}(r)$ denotes the derivative of $\overline{T}_{N,n}(r)$ with respect to r.

Observation 3.4.3 *Fix $\epsilon > 0$ and let $x_0, ..., x_K$ be a sequence of real numbers such that*

$$\sum_{i=K_1+1}^{K} |x_k| < \epsilon \tag{3.163}$$

where $0 \le K_1 \le K$. Using (3.34), we have

$$\sum_{i=0}^{K} x_i \overline{T}_{N,n}(r) = \sum_{i=0}^{K} \alpha_i T_{N,n}(r) \tag{3.164}$$

where $x_0, ..., x_K$ are real numbers and α_i is defined by the formula

$$\alpha_i = x_i \sqrt{2(2i + N + p/2 + 1)}. \tag{3.165}$$

Scaling both sides of (3.50), we obtain

$$\begin{aligned} &\alpha_0 r T'_{N,0}(r) - \frac{\alpha_0 \tilde{b}_1}{\tilde{a}_1} r T'_{N,1}(r) + \frac{\alpha_0 \tilde{c}_1}{\tilde{a}_1} r T'_{N,2}(r) \\ &= \frac{\alpha_0 a_1}{\tilde{a}_1} T_{N,0}(r) - \frac{\alpha_0 b_1}{\tilde{a}_1} T_{N,1}(r) + \frac{\alpha_0 c_1}{\tilde{a}_1} T_{N,2}(r). \end{aligned} \tag{3.166}$$

where $a_i, b_i, c_i, \tilde{a}_i, \tilde{b}_i, \tilde{c}_i$ are defined in Lemma 3.2.14. Scaling (3.50) and adding the resulting equation to (3.166), we obtain

$$\begin{aligned}
&\alpha_0 rT'_{N,0}(r) - \frac{\alpha_0\tilde{b}_1}{\tilde{a}_1} rT'_{N,1}(r) + \frac{\alpha_0\tilde{c}_1}{\tilde{a}_1} rT'_{N,2}(r) \\
&+ \left(\left(\frac{\alpha_0\tilde{b}_1}{\tilde{a}_1} + \alpha_1\right)\tilde{a}_2^{-1}\right)\left(\tilde{a}_2 rT'_{N,1}(r) - \tilde{b}_2 rT'_{N,2}(r) + \tilde{c}_2 rT'_{N,3}(r)\right) \\
&= \frac{\alpha_0 a_1}{\tilde{a}_1} T_{N,0}(r) - \frac{\alpha_0 b_1}{\tilde{a}_1} T_{N,1}(r) + \frac{\alpha_0 c_1}{\tilde{a}_1} T_{N,2}(r) \\
&+ \left(\left(\frac{\alpha_0\tilde{b}_1}{\tilde{a}_1} + \alpha_1\right)\tilde{a}_2^{-1}\right)\left(a_2 T_{N,1}(r) - b_2 T_{N,2}(r) + c_2 T_{N,3}(r)\right).
\end{aligned} \tag{3.167}$$

Simplying the left hand side of (3.167), we have

$$\begin{aligned}
&\alpha_0 rT'_{N,0}(r) + \alpha_1 rT'_{N,1}(r) + \left(\frac{\alpha_0\tilde{c}_1}{\tilde{a}_1} - \frac{\tilde{b}_2}{\tilde{a}_2}\left(\frac{\alpha_0\tilde{b}_1}{\tilde{a}_1} + \alpha_1\right)\right) rT'_{N,2}(r) \\
&+ \left(\left(\frac{\alpha_0\tilde{b}_1}{\tilde{a}_1} + \alpha_1\right)\tilde{a}_2^{-1}\right)\left(\tilde{c}_2 rT'_{N,3}(r)\right).
\end{aligned} \tag{3.168}$$

We continue by adding scaled versions of (3.50) to (3.167) until the expansion on the left hand side of (3.167) approximates (3.164). After $K_1 + 1$ steps, the new expansion will be accurate to approximately ϵ precision. Specifically, at the start of step k, for $2 \le k \le K_1 + 1$, we have

$$\begin{aligned}
&\alpha_0 rT'_{N,0}(r) + \alpha_1 rT'_{N,1}(r) + ... + \alpha_{k-2} rT'_{N,k-2}(r) + x_{k-1} rT'_{N,k-1}(r) + x_k rT'_{N,k}(r) \\
&= y_0 T_{N,0} + y_1 T_{N,1} + ... + y_k T_{N,k}
\end{aligned} \tag{3.169}$$

where $x_{k-1}, x_k, x_{k+1}, y_0, y_1, ..., y_k$ are some real numbers. Scaling both sides of

(3.50) and adding the resulting equation to (3.169), we obtain

$$\begin{aligned}
&\alpha_0 rT'_{N,0}(r) + \alpha_1 rT'_{N,1}(r) + ... + \alpha_{k-2} rT'_{N,k-2}(r) + x_{k-1} rT'_{N,k-1}(r) + x_k rT'_{N,k}(r) \\
&+ \left(\frac{-x_{k-1} + \alpha_{k-1}}{\tilde{a}_k} \right) \left(\tilde{a}_k rT'_{N,k-1}(r) - \tilde{b}_k rT'_{N,k}(r) + \tilde{c}_k rT'_{N,k+1}(r) \right) \\
&= y_0 T_{N,0} + y_1 T_{N,1} + ... + y_k T_{N,k} \\
&+ \left(\frac{-x_{k-1} + \alpha_{k-1}}{\tilde{a}_k} \right) \left(a_k T_{N,k-1}(r) - b_k T_{N,k}(r) + c_k T_{N,k+1}(r) \right).
\end{aligned} \tag{3.170}$$

Simplifying both sides of (3.170), we have

$$\begin{aligned}
&\alpha_0 rT'_{N,0}(r) + \alpha_1 rT'_{N,1}(r) + ... + \alpha_{k-2} rT'_{N,k-2}(r) + \alpha_{k-1} rT'_{N,k-1}(r) \\
&+ \left(\frac{-x_{k-1} + \alpha_{k-1}}{\tilde{a}_k} (-\tilde{b}_k) + x_k \right) rT'_{N,k}(r) + \left(\frac{-x_{k-1} + \alpha_{k-1}}{\tilde{a}_k} \tilde{c}_k \right) rT'_{N,k+1}(r) \\
&= y_0 T_{N,0} + y_1 T_{N,1} + ... + \left(\frac{-x_{k-1} + \alpha_{k-1}}{\tilde{a}_k} a_k + y_{k-1} \right) T_{N,k-1}(r) \\
&+ \left(\frac{-x_{k-1} + \alpha_{k-1}}{\tilde{a}_k} (-b_k) + y_k \right) T_{N,k}(r) + \left(\frac{-x_{k-1} + \alpha_{k-1}}{\tilde{a}_k} c_k \right) T_{N,k+1}(r).
\end{aligned} \tag{3.171}$$

We then scale back each term in the new expansion in $T_{N,n}$ to get an expansion in $\overline{T}_{N,n}$. That is, we scale the i^{th} term in the new expansion by

$$(2(2i + N + p/2 + 1))^{-(1/2)}. \tag{3.172}$$

The following observation, when combined with Observation 3.4.3, provides a numerical scheme for evaluating integrals of the form

$$\int_0^1 r\Phi'_{N,n}(r)\Phi_{N,m}(r) r^{p+1} dr. \tag{3.173}$$

This scheme will be used in Algorithm 3.4.2.

Observation 3.4.4 *Suppose that*

$$r\Phi'_{N,n}(r) = \sum_{i=0}^{K} x_i \overline{R}_{N,i}(r) \tag{3.174}$$

and

$$\Phi_{N,m}(r) = \sum_{i=0}^{K} y_i \overline{R}_{N,i}(r). \tag{3.175}$$

where x_i, y_i are real numbers. Then, combining (3.22) with (3.24), we have,

$$\int_0^1 r\Phi'_{N,n}(r)\Phi_{N,m}(r) r^{p+1} dr = \int_0^1 \sum_{i=0}^{K} x_i \overline{R}_{N,i}(r) \sum_{i=0}^{K} y_i \overline{R}_{N,i}(r) r^{p+1} dr = \sum_{i=0}^{K} x_i y_i. \tag{3.176}$$

We now describe an algorithm for evaluating the eigenvalues $\beta_{N,0}, \beta_{N,1}, ..., \beta_{N,k}$ for any non-negative integers N, k.

Algorithm 3.4.2

Step 1. Use Algorithm 3.4.1 to recover the Zernike expansions of the GPSFs

$$\Phi_{N,0}, \Phi_{N,1}, ..., \Phi_{N,n}. \tag{3.177}$$

Step 2. Compute the eigenvalue $\beta_{N,0}$ (see (3.111)) using Remark 3.4.2.

Step 3. Use Observation 3.4.3 to evaluate the $\overline{R}_{N,n}$ expansion of $r\Phi'_{N,0}$ and $r\Phi'_{N,1}$.

Step 4. Use Observation 3.4.4 to compute the integrals

$$\int_0^1 r\Phi'_{N,1}(r)\Phi_{N,0}(r) r^{p+1} dr \tag{3.178}$$

and

$$\int_0^1 r\Phi'_{N,0}(r)\Phi_{N,1}(r)r^{p+1}dr \tag{3.179}$$

where the Zernike expansions of $\Phi_{N,0}(r), \Phi_{N,1}(r)$ were computed in Step 1 and the Zernike expansions of $r\Phi'_{N,0}(r), \Phi'_{N,1}(r)$ were computed in Step 3.

Step 5. Using Theorem 3.3.1, evaluate $\beta_{N,1}$ using the formula

$$\beta_{N,1} = \beta_{N,0}\frac{\int_0^1 r\Phi'_{N,1}(r)\Phi_{N,0}(r)r^{p+1}\,dr}{\int_0^1 r\Phi'_{N,0}(r)\Phi_{N,1}(r)r^{p+1}\,dr}. \tag{3.180}$$

where $\beta_{N,0}$ was obtained in Step 2 and the numerator and denominator of (3.180) were evaluated in Step 4.

Step 6. Repeat Steps 3-5 k times, each time computing the next eigenvalue, $\beta_{N,i+1}$ via the formula

$$\beta_{N,i+1} = \beta_{N,i}\frac{\int_0^1 r\Phi'_{N,i+1}(r)\Phi_{N,i}(r)r^{p+1}\,dr}{\int_0^1 r\Phi'_{N,i}(r)\Phi_{N,i+1}(r)r^{p+1}\,dr}. \tag{3.181}$$

3.5 Quadratures for Band-limited Functions

In this section, we describe a quadrature scheme for bandlimited functions using nodes that are a tensor product of roots of GPSFs in the radial direction and nodes that integrate spherical harmonics in the angular direction.

The following lemma shows that a quadrature rule that accurately integrates complex exponentials, also integrates bandlimited functions accurately.

Lemma 3.5.1 *Let $\xi_1, ..., \xi_n \in B$ and $w_1, ..., w_n \in \mathbb{R}$ weights such that*

$$\left| \int_B e^{ic\langle x,t\rangle} dt - \sum_{i=1}^{n} w_i e^{ic\langle x,\xi_i\rangle} \right| < \epsilon \tag{3.182}$$

for all $x \in B$ where B denotes the unit ball in $\mathbb{R}^n$ for any non-negative integer n and $\epsilon > 0$ is fixed. Then, for all $f : B \to \mathbb{C}$ such that

$$f(x) = \int_B \sigma(t) e^{ic\langle x,t\rangle} dt \tag{3.183}$$

where $\sigma \in L^2(B)$, we have

$$\left| \int_B f(x) dx - \sum_{i=1}^{n} w_i f(\xi_i) \right| < \epsilon \int_B |\sigma(t)| dt \tag{3.184}$$

Proof. Clearly,

$$\begin{aligned} \left| \int_B f(t) dt - \sum_{i=1}^{n} w_i f(\xi_i) \right| &= \left| \int_B \int_B \sigma(t) e^{ic\langle x,t\rangle} dt dx - \sum_{i=0}^{n} w_i \int_B \sigma(t) e^{ic\langle \xi_i,t\rangle} dt \right| \\ &= \left| \int_B \sigma(t) \left(\int_B e^{ic\langle x,t\rangle} dx - \sum_{i=0}^{n} w_i e^{ic\langle \xi_i,t\rangle} \right) dt \right|. \end{aligned} \tag{3.185}$$

Applying (3.182) to (3.185), we obtain

$$\begin{aligned} \left| \int_B f(t) dt - \sum_{i=1}^{n} w_i f(\xi_i) \right| &\leq \int_B |\sigma(t)| \left| \int_B e^{ic\langle x,t\rangle} dx - \sum_{i=0}^{n} w_i e^{ic\langle \xi_i,t\rangle} \right| dt \\ &< \epsilon \int_B |\sigma(t)| dt. \end{aligned} \tag{3.186}$$

■

The following technical lemma will be used in the construction of quadratures for bandlimited functions.

Lemma 3.5.2 *For any positive integer K and any integer $p \geq -1$,*

$$\left| \int_B e^{ic\langle x,t\rangle} dt - \int_B \sum_{N=0}^{K} \sum_{\ell=1}^{h(N,p)} i^N (2\pi)^{p/2+1} \frac{J_{N+p/2}(c\|x\|\|t\|)}{(c\|x\|\|t\|)^{p/2}} S_N^\ell(x/\|x\|) S_N^\ell(t/\|t\|) dt \right| \\ \leq (2\pi)^{p/2+1} \sum_{N=K+1}^{\infty} \frac{c^{2N}(1/2)^{2N+p}}{\Gamma(N+p/2+1)^2} \frac{\pi^{p/2+1}}{\Gamma(p/2+2)} \left(\sum_{\ell=1}^{h(N,p)} |S_N^\ell(x/\|x\|)| \right) \tag{3.187}$$

for all $x \in B$ and $c > 0$.

Proof. It follows immediately from (3.108) that for any integer $p \geq -1$ and for all $x \in \mathbb{R}^{p+2}$,

$$\left| \int_B e^{ic\langle x,t\rangle} dt - \int_B \sum_{N=0}^{K} \sum_{\ell=1}^{h(N,p)} i^N (2\pi)^{p/2+1} \frac{J_{N+p/2}(c\|x\|\|t\|)}{(c\|x\|\|t\|)^{p/2}} S_N^\ell(x/\|x\|) S_N^\ell(t/\|t\|) dt \right| \\ \leq (2\pi)^{p/2+1} \sum_{N=K+1}^{\infty} \sum_{\ell=1}^{h(N,p)} \left| S_N^\ell(x/\|x\|) \right| \int_B \left| \frac{J_{N+p/2}(c\|x\|\|t\|)}{(c\|x\|\|t\|)^{p/2}} S_N^\ell(t/\|t\|) \right| dt \tag{3.188}$$

where $r = \|x\|$, B denotes the unit ball in $\mathbb{R}^{p+1}$, and S_N^ℓ is defined in (3.90). Applying Cauchy-Schwarz and Lemma 3.2.2 to (3.188) and using the fact that Spherical Harmonics have L^2 norm of 1, we obtain,

$$\left| \int_B e^{ic\langle x,t\rangle} dt - \int_B \sum_{N=0}^{K} \sum_{\ell=1}^{h(N,p)} i^N (2\pi)^{p/2+1} \frac{J_{N+p/2}(c\|x\|\|t\|)}{(c\|x\|\|t\|)^{p/2}} S_N^\ell(x/\|x\|) S_N^\ell(t/\|t\|) dt \right| \\ \leq (2\pi)^{p/2+1} \sum_{N=K+1}^{\infty} \sum_{\ell=1}^{h(N,p)} \left| S_N^\ell(x/\|x\|) \right| \int_B \left| \frac{(c\|x\|\|t\|)^N (1/2)^{N+p/2}}{\Gamma(N+p/2+1)} \right|^2 dt. \tag{3.189}$$

Equation (3.187) follows immediately from applying (3.84) and (3.87) to (3.189). ■

Remark 3.5.1 *Lemma 3.5.2 shows that a complex exponential on the unit ball is well approximated by a function of the form*

$$\sum_{N=0}^{K} \sum_{\ell=1}^{h(N,p)} i^N (2\pi)^{p/2+1} \frac{J_{N+p/2}(c\|x\|\|t\|)}{(c\|x\|\|t\|)^{p/2}} S_N^\ell(x/\|x\|) S_N^\ell(t/\|t\|) dt \tag{3.190}$$

where the error of the approximation decays super-exponentially in K. Furthermore, the spherical harmonics S_N^ℓ integrate to 0 for all $N \geq 1$ (see Lemma 3.5.3). Combining these facts, we observe that in order to integrate a complex exponential on the unit ball, it is sufficient to use a quadrature rule that is the tensor product of nodes in the angular direction that integrate all spherical harmonics S_N^ℓ for sufficiently large N and nodes in the radial direction that integrate functions of the form

$$\frac{J_{p/2}(cr\rho)}{(cr\rho)^{p/2}} \rho^{p+1}. \tag{3.191}$$

We will show in Remark 3.5.2 that accurately computing functions of the form of (3.191) can be done using a quadrature rule that integrates GPSFs.

The following lemma shows that (3.191) is well represented by an expansion in GPSFs. This lemma will be used to construct quadrature nodes for bandlimited functions.

Lemma 3.5.3 *For all real numbers $r, \rho \in (0, 1)$,*

$$\frac{J_{p/2}(cr\rho)}{(cr\rho)^{p/2}} \rho^{p+1} = \sum_{i=0}^{\infty} \beta_{0,i} \Phi_{0,i}(r) \Phi_{0,i}(\rho) \tag{3.192}$$

where $J_{p/2}$ is a Bessel function, $\Phi_{0,n}$ is defined in (3.111) and $\beta_{0,i}$ is defined in (3.114).

Proof. Since $\Phi_{0,i}$ are complete in $L^2[0,1]_{r^{p+1}}$,

$$\frac{J_{p/2}(cr\rho)}{(cr\rho)^{p/2}}\rho^{p+1} = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \alpha_{i,j}\Phi_{0,i}(r)\Phi_{0,j}(\rho) \tag{3.193}$$

where $\alpha_{i,j}$ is defined by the formula

$$\alpha_{i,j} = \int_0^1 \int_0^1 \frac{J_{p/2}(cr\rho)}{(cr\rho)^{p/2}} r^{p+1}\Phi_{0,i}(r)\Phi_{0,j}(\rho) dr \rho^{p+1} d\rho. \tag{3.194}$$

Changing the order of integration of (3.194) and substituting in (3.114), we obtain,

$$\begin{aligned}\alpha_{i,j} &= \int_0^1 \Phi_{0,j}(r) \int_0^1 \frac{J_{p/2}(cr\rho)}{(cr\rho)^{p/2}} \rho^{p+1}\Phi_{0,i}(\rho) d\rho r^{p+1} dr \\ &= \beta_{0,i}\int_0^1 \Phi_{0,j}(r)\Phi_{0,i}(r) r^{p+1} dr \\ &= \delta_{i,j}\beta_{0,i}\end{aligned} \tag{3.195}$$

where $\beta_{0,i}$ is defined in (3.114). Identity (3.192) follows immediately from the combination of (3.193) and (3.195). ■

The following remark shows that a quadrature rule that correctly integrates certain GPSFs also integrates certain Bessel functions.

Remark 3.5.2 *Let $\rho_1, ..., \rho_n$ be the n roots of $\Phi_{0,n}$ and $w_1, ..., w_n \in \mathbb{R}$ the n weights such that*

$$\int_0^1 \Phi_{0,k}(r) r^{p+1} dr = \sum_{i=0}^{n} \Phi_{0,k}(\rho_i) w_i \tag{3.196}$$

for $k = 0, 1, ..., K$. By Lemma 3.5.3,

$$\begin{aligned}&\left|\int_0^1 \frac{J_{p/2}(cr\rho)}{(cr\rho)^{p/2}}\rho^{p+1}d\rho - \sum_{i=1}^{n} \frac{J_{p/2}(cr\rho_i)}{(cr\rho_i)^{p/2}}w_i\right| \\ &= \left|\int_0^1 \left(\sum_{j=0}^{\infty} \beta_{0,j}\Phi_{0,j}(r)\Phi_{0,j}(\rho)\right) d\rho - \sum_{i=1}^{n} w_i \left(\sum_{j=0}^{\infty} \beta_{0,j}\Phi_{0,j}(r)\Phi_{0,j}(\rho_i)\right)\right|\end{aligned} \tag{3.197}$$

where $\beta_{0,j}$ is defined in (3.114). Applying (3.196) to (3.197), we obtain

$$\begin{aligned}&\left|\int_0^1 \frac{J_{p/2}(cr\rho)}{(cr\rho)^{p/2}}\rho^{p+1}d\rho - \sum_{i=1}^{n} \frac{J_{p/2}(cr\rho_i)}{(cr\rho_i)^{p/2}}\rho_i^{p+1}w_i\right| \\ &= \left|\int_0^1 \left(\sum_{j=K+1}^{\infty} \beta_{0,j}\Phi_{0,j}(r)\Phi_{0,j}(\rho)\right) d\rho - \sum_{i=1}^{n} w_i \left(\sum_{j=K+1}^{\infty} \beta_{0,j}\Phi_{0,j}(r)\Phi_{0,j}(\rho_i)\right)\right|.\end{aligned} \tag{3.198}$$

Clearly, as long as $\beta_{0,K+1}$ is in the regime of exponential decay, (3.198) is of magnitude approximately $\beta_{0,K+1}$.

We now describe a quadrature rule that correctly integrates functions of the form of (3.190). This quadrature rule uses nodes that are a tensor product of roots of $\Phi_{0,n}$ in the radial direction and nodes that integrate spherical harmonics in the angular direction.

Observation 3.5.3 *Suppose that $r_1, ..., r_n \in (0, 1)$ and weights $w_1, ..., w_n \in \mathbb{R}$ satisfy*

$$\int_0^1 \Phi_{0,k}(r)r^{p+1}dr = \sum_{i=1}^{n} w_i\Phi_{0,k}(r_i) \tag{3.199}$$

for $k = 0, 1, ..., K_1$. Suppose further that $x_1, ..., x_m \in S^{p+1}$ are nodes and $v_1, ..., v_m \in$

$\mathbb{R}$ *are weights such that*

$$\int_{S^{p+1}} S_N^\ell(x)dx = \sum_{i=1}^{m} v_i S_N^\ell(x_i) \tag{3.200}$$

for all $N \leq K_2$ and for all $\ell \in \{1, 2, ..., h(N, p)\}$. Then it follows immediately from Remark 3.5.1 and Remark 3.5.2 that

$$\left| \int_B e^{ic\langle x,t\rangle} dt - \sum_{i=0}^{m} v_i \sum_{j=1}^{n} w_j e^{ic\langle x, r_j x_i\rangle} \right| \tag{3.201}$$

will be exponentially small for large enough n, m. Lemma 3.5.1 shows that quadrature (3.201) will also accurately integrate functions of the form

$$\int_B \sigma(t) e^{ic\langle x,t\rangle} dt \tag{3.202}$$

where σ is in $L^2(B)$.

Remark 3.5.4 *A Chebyshev quadrature of the form (3.5.1) can be generated by first computing the n roots of $\Phi_{0,n}$ (see Section 3.5.1) and then solving the $n \times n$ linear system of equations*

$$\int_0^1 \Phi_{0,k}(r) r^{p+1} dr = \sum_{i=1}^{n} w_i \Phi_{0,k}(r_i) \tag{3.203}$$

for $w_1, ..., w_n$ where $r_1, ..., r_n$ are the n roots of $\Phi_{0,n}$. Section 3.5.2 contains a description of an algorithm for generating Gaussian quadratures for GPSFs.

3.5.1 Roots of $\Phi_{0,n}$

In this section, we describe an algorithm for finding the roots of $\Phi_{N,n}$. These roots will be used in the design of quadratures for GPSFs.

The following lemma provides a differential equation satisfied by $\varphi_{0,n}$. It will be used in the evaluation of roots of $\varphi_{0,n}$ later in this section.

Lemma 3.5.4 *For all non-negative integers n,*

$$\varphi''_{0,n}(r) + \alpha(r)\varphi'_{0,n}(r) + \beta(r)\varphi_{0,n}(r) = 0, \tag{3.204}$$

where

$$\alpha(r) = \frac{-2r}{1-r^2} \tag{3.205}$$

and

$$\beta(r) = \frac{1/4 - (N+p/2)^2}{(1-r^2)r^2} - \frac{c^2r^2 + \chi_{N,n}}{1-r^2} \tag{3.206}$$

where $\varphi_{0,n}$ is defined in (3.115) and $\chi_{N,n}$ is defined in (3.120).

The following lemma is obtained by applying the Prufer Transform (see Lemma 3.2.15) to (3.204).

Lemma 3.5.5 *For all non-negative integers n, real numbers $k > -1$, and $r \in (0,1)$,*

$$\frac{d\theta}{dr} = -\sqrt{\beta(r)} - \left(\frac{\beta'(r)}{4\beta(r)} + \frac{\alpha(r)}{2}\right)\sin(2\theta(r)), \tag{3.207}$$

where the function $\theta : (0,1) \to \mathbb{R}$ is defined by the formula

$$\frac{\varphi_{N,n}(r)}{\varphi'_{N,n}(r)} = \sqrt{\beta(r)}\tan(\theta(r)), \tag{3.208}$$

and $\beta'(r)$, the derivative of $\beta(r)$ with respect to r, is defined by the formula

$$\beta'(r) = \frac{-2(1/4-(N+p/2)^2)(1-2r^2)}{(1-r^2)r^3} + \frac{-2rc^2(1-r^2)+2r(-c^2r^2-\chi_{N,n})}{(1-r^2)^2} \tag{3.209}$$

and where $\alpha(r)$, $\beta(r)$ are defined in (3.205) and (3.206), $\varphi_{N,n}$ is defined in (3.115) and $\chi_{N,n}$ is defined in (3.120).

Remark 3.5.5 *For any non-negative integer n,*

$$\frac{d\theta}{dr} < 0 \tag{3.210}$$

for all $r \in (r_1, r_n)$ where r_1 and r_n are the smallest and largest roots of $\varphi_{N,n}$ respectively. Therefore, applying Remark 3.2.6 to (3.207), we can view r as a function of θ where r satisfies the differential equation

$$\frac{dr}{d\theta} = \left(-\sqrt{\beta(r)} - \left(\frac{\beta'(r)}{4\beta(r)} + \frac{\alpha(r)}{2}\right)\sin(2\theta(r))\right)^{-1} \tag{3.211}$$

where α, β, and β' are defined in (3.205), (3.206) and (3.209).

The following is a description of an algorithm for the evaluation of the n roots of $\Phi_{N,n}$. We denote the n roots of $\Phi_{N,n}$ by $r_1 < r_2 < ... < r_n$.

Algorithm 3.5.1

Step 0. Compute the $\overline{T}_{N,n}$ expansion of $\varphi_{N,n}$ using Algorithm 3.4.1.

Step 1. Use bisection to find the root in $x_0 \in (0,1)$ of $\beta(r)$ where $\beta(r)$ is defined in (3.206). If β has no root on $(0,1)$, then set $x_0 = 1$.

Step 2. If $\chi_{0,n}(c) > 1/\sqrt{c}$, place Chebyshev nodes (order $5n$, for example) on the interval $(0, x_0)$ and check, starting at x_0 and moving in the negative direction, for a sign change. Once a sign change has occured, use Newton to find an accurate approximation to the root.

If $\chi_{0,n}(c) \leq 1/\sqrt{c}$, then use three steps of Mueller's method starting at x_0, using the first and second derivatives of $\varphi_{0,n}$ followed by Newton's method.

Step 3. Defining θ_0 by the formula

$$\theta_0 = \theta(x_0), \tag{3.212}$$

where θ is defined in (3.139), solve the ordinary differential equation $\frac{dr}{d\theta}$ (see (3.211)) on the interval $(\pi/2, \theta_0)$, with the initial condition $r(\theta_0) = x_0$. To solve the differential equation, it is sufficient to use, for example, second order Runge Kutta with 100 steps (independent of n). We denote by $\tilde{r}_n$ the approximation to $r(\pi/2)$ obtained by this process. It follows immediately from (3.65) that $\tilde{r}_n$ is an approximation to r_n, the largest root of $\varphi_{N,n}$.

Step 4. Use Newton's method with $\tilde{r}_n$ as an initial guess to find r_n to high precision. The GPSF $\varphi_{N,n}$ and its derivative $\varphi'_{N,n}$ can be evaluated using the expansion evaluated in Step 0.

Step 5. With initial condition

$$x(\pi/2) = r_n, \tag{3.213}$$

solve differential equation $\frac{dr}{d\theta}$ (see (3.211)) on the interval

$$(-\pi/2, \pi/2) \tag{3.214}$$

using, for example, second order Runge Kuta with 100 steps. We denote by $\tilde{r}_{n-1}$ the approximation to

$$r(-\pi/2) \tag{3.215}$$

obtained by this process.

Step 6. Use Newton's method, with initial guess $\tilde{r}_{n-1}$, to find to high precision the second largest root, r_{n-1}.

Step 7. For $k = \{1, 2, ..., n-1\}$, repeat Step 4 on the interval

$$(-\pi/2 - k\pi, -\pi/2 - (k-1)\pi) \tag{3.216}$$

with intial condition

$$x(-\pi/2 - (k-1)\pi) = r_{n-k+1} \tag{3.217}$$

and repeat Step 5 on $\tilde{r}_{n-k}$.

3.5.2 Gaussian Quadratures for $\Phi_{0,n}$

In this section, we describe an algorithm for generating Gaussian quadratures for the GPSFs $\Phi_{0,0}, \Phi_{0,1}, ..., \Phi_{0,n}$.

Definition 3.5.1 *A Gaussian Quadrature with respect to a set of functions*

$$f_1, ..., f_{2n-1} : [a, b] \to \mathbb{R} \tag{3.218}$$

and non-negative weight function $w : [a, b] \to \mathbb{R}$ *is a set of* n *nodes,* $x_1, ..., x_n \in [a, b]$*, and* n *weights,* $\omega_1, ..., \omega_n \in \mathbb{R}$*, such that, for any integer* $j \leq 2n - 1$*,*

$$\int_a^b f_j(x)w(x)dx = \sum_{i=0}^{n} \omega_i f_j(x_i). \tag{3.219}$$

Remark 3.5.6 *In order to generate a Gaussian quadrature for GPSFs with bandlimit* $c > 0$*, we first generate a Chebyshev quadrature for GPSFs with bandlimit* $c/2$ *and then, using those nodes and weights as a starting point, we use Newton's method with step-length control to find nodes and weights that form a Gaussian quadrature for GPSFs with bandlimit* c*.*

The following is a description of an algorithm for generating Gaussian quadratures for the GPSFs

$$\Phi^c_{0,0}, ..., \Phi^c_{0,2n-1}. \tag{3.220}$$

Algorithm 3.5.2

Step 1. Using Algorithm 3.5.1, generate a Chebyshev quadrature for the functions

$$\Phi^{c/2}_{0,0}, ..., \Phi^{c/2}_{0,n-1}. \tag{3.221}$$

That is, find, $r_1, ..., r_n$, the n roots of $\Phi_{0,n}$ and weights $w_1, ..., w_n$ such that

$$\int_0^1 \Phi^{c/2}_{0,k}(r)dr = \sum_{i=1}^{n} w_i \Phi^{c/2}_{0,k}(r_i) \tag{3.222}$$

for $k = 0, 1, ..., n - 1$.

Step 2. Evaluate the vector $d = (d_0, d_1, ..., d_{2n-1})$ of discrepencies where d_k is defined by the formula

$$d_k = \int_0^1 \Phi^c_{0,k}(r)dr - \sum_{i=1}^{n} w_i \Phi^c_{0,k}(r_i) \tag{3.223}$$

for $k = 0, 1, ..., 2n - 1$.

Step 3. Generate A, the $2n \times 2n$ matrix of partial derivatives of d the n nodes and and n weights. Specifically, for $i = 1, ..., 2n$, the matrix A is defined by the formula

$$A_{i,j} = \begin{cases} \Phi^c_{0,j}(r_i) & \text{for } i = 1, ..., n, \\ \\ w_i \Phi^{c\prime}_{0,j}(r_i) & \text{for } i = n+1, ..., 2n. \end{cases} \tag{3.224}$$

where $\Phi^{c\prime}_{0,k}(r)$ denotes the derivative of $\Phi^c_{0,k}(r)$ with respect to r.

Step 4. Solve for $x \in \mathbb{R}^{2n}$ the $2n \times 2n$ linear system of equations

$$Ax = -d \tag{3.225}$$

where A is defined in (3.224) and d is defined in (3.223).

Step 5. Update nodes and weights correspondingly. That is, defining $r \in \mathbb{R}^{2n}$ to be the vector of nodes and weights

$$r = (r_1, r_2, ..., r_n, w_1, w_2, ..., w_n)^T, \tag{3.226}$$

we construct the updated vector of nodes and weights $\tilde{r}$ so that

$$\tilde{r} = r + \langle r, x \rangle r \tag{3.227}$$

Step 6. Check that the l^2 norm of $\tilde{r}$ is less than the l^2 norm of r. If now, then go back to Step 5 and divide the steplength by 2. That is, define $\tilde{r}$ by the formula,

$$\tilde{r} = r + \frac{1}{2}\langle r, x \rangle r. \tag{3.228}$$

Continue dividing the steplength by 2 until $\|\tilde{r}\|_2 < \|r\|_2$.

Step 7. Repeat steps 2-6 until the discrepencies, d_i for $i = 0, 1, ..., 2n-1$ (see (3.223)) are approximately machine precision.

3.6 Interpolation via GPSFs

In this section, we describe a numerical scheme for representing a bandlimited function as an expansion in GPSFs.

In general, the interpolation problem is formulated as follows. Suppose that f is defined by the formula

$$f(x) = a_1 g_1(x) + a_2 g_2(x) + ... + a_n g_n(x) \tag{3.229}$$

where $g_1, ..., g_n$ are some fixed basis functions. The interpolation problem is to recover the coefficients $a_1, ..., a_n$. This is generally done by solving the $n \times n$ linear system of equations obtained from evaluating f at certain interpolation nodes. As long as f is well-represented by the interpolation nodes, then the procedure is accurate.

As shown in the context of quadrature (see Section 3.5), GPSFs are a natural basis for representing bandlimited functions. We formulate the interpolation problem for GPSFs as recovering the coefficients of a bandlimited function f in its GPSF expansion. That is, suppose that f is of the form

$$f(x) = \int_B \sigma(t) e^{ic\langle x,t\rangle} dt. \tag{3.230}$$

where $\sigma \in L^2(B)$. Then, f is representable in the form

$$f(x) = \sum_{i=1}^{N} a_i \psi_i(x) \tag{3.231}$$

where $\psi_j(x)$ is a GPSF defined in (3.98) and a_i satisfies

$$a_i = \int_B \psi_i(t) f(t) dt. \tag{3.232}$$

The problem is the recover the coefficients a_j.

The following lemma shows that a quadrature rule that recovers the coefficients of the expansion in GPSFs of a complex exponential will also recover the coefficients in a GPSF expansion of a bandlimited function.

Lemma 3.6.1 *Suppose that for all $t \in B$,*

$$\left| \int_B \psi_j(x) e^{ic\langle x,t\rangle} dx - \sum_{k=1} w_k \psi_j(x_k) e^{ic\langle x_k,t\rangle} \right| < \epsilon \tag{3.233}$$

where B denotes the unit ball in $\mathbb{R}^{p+2}$ and ψ_j is defined in (3.98). Then,

$$\left| \int_B \psi_j(x) f(x) dx - \sum_{k=1} w_k \psi_j(x_k) f(x_k) \right| < \epsilon \tag{3.234}$$

where

$$f(x) = \int_B \sigma(t) e^{ic\langle x,t\rangle} dt. \tag{3.235}$$

The following lemma shows that the product of a complex exponential with a GPSF of bandlimit $c > 0$ is a bandlimited function with bandlimit $2c$. The proof is a slight modification of Lemma 5.3 in [24].

Lemma 3.6.2 *For all $x \in B$ where B denotes the unit ball in $\mathbb{R}^{p+2}$ and for all $c > 0$,*

$$e^{ic\langle\omega,x\rangle}\psi_j(x) = \int_B \sigma(\xi)e^{i2c\langle\xi,x\rangle}d\xi \tag{3.236}$$

where ψ_j is defined in (3.98) and σ satisfies

$$\left|\int_B \sigma(t)^2\right| \leq 4/|\lambda_j|^2. \tag{3.237}$$

where λ_j is defined in (3.98).

Proof. Using (3.98),

$$\psi_j(x)e^{ic\langle\omega,x\rangle} = \frac{1}{\lambda_j}\int_B e^{ic\langle\omega+t,x\rangle}\psi_j(t)dt. \tag{3.238}$$

Applying the change of variables $\xi = (t+\omega)/2$ to (3.238), we obtain

$$\psi_j(x)e^{ic\langle\omega,x\rangle} = \frac{1}{\lambda_j}\int_{B_\omega} e^{i2c\langle\xi,x\rangle}2\psi_j(2\xi-\omega)d\xi \tag{3.239}$$

where B_ω is the ball of radius $1/2$ centered at $\omega/2$. It follows immediately from (3.239) that

$$\psi_j(x)e^{ic\langle\omega,x\rangle} = \frac{1}{\lambda_j}\int_{B_\omega} e^{i2c\langle\xi,x\rangle}\mu(\xi)d\xi. \tag{3.240}$$

where

$$\mu(\xi) = \begin{cases} \frac{2\psi_j(2\xi-\omega)}{\lambda_j} & \text{if } \xi \in B_\omega, \\ 0 & \text{otherwise.} \end{cases} \tag{3.241}$$

Inequality (3.237) follows immediately from the combination of (3.241) with the fact that ψ_j is L^2 normalized. ■

The following observation describes a numerical scheme for recovering the coefficients in a GPSF expansion of a bandlimited function.

Observation 3.6.1 *Suppose that f is defined by the formula*

$$f(x) = \int_B \sigma(t) e^{ic\langle x,t\rangle} dt \tag{3.242}$$

where σ is some function in $L^2(B)$. Then, f is representable in the form

$$f(x) = \sum_{k=1}^{\infty} a_k \psi_k(x) \tag{3.243}$$

where

$$a_k = \int_B f(x)\psi_k(x)dx. \tag{3.244}$$

It follows immediately from the combination of Lemma 3.6.2 and Lemma 3.6.1 that using quadrature rule (3.201) with bandlimit $2c$ will integrate a_k accurately. That is, following the notation of Observation 3.5.1,

$$\left| a_k - \sum_{i=0}^{n} w_i \sum_{j=1}^{m} v_j f(r_i x_j)\psi_k(r_i x_j) \right| \tag{3.245}$$

is exponentially small for large enough m, n.

Remark 3.6.2 *When integrating a function of the form of (3.243) on the unit disk in R^2, the v_j in (3.245) are defined by the formula*

$$v_j = j\frac{2\pi}{2m-1} \tag{3.246}$$

for $j = 1, 2, ..., 2m-1$ and the sums

$$\sum_{j=1}^{m} v_j f(r_i x_j)\psi_k(r_i x_j) \tag{3.247}$$

for each i can be computed using an FFT.

The following lemma bounds the magnitudes of the coefficients of the GPSF expansion of a bandlimited function.

Lemma 3.6.3 *Suppose that f is defined by the formula*

$$f(x) = \int_B \sigma(t)e^{ic\langle x,t\rangle}dt. \tag{3.248}$$

Then,

$$f(x) = \sum_{i=1}^{N} a_i\psi_i(x) \tag{3.249}$$

where $\psi_j(x)$ is a GPSF defined in (3.98) and a_i satisfies

$$|a_i| \le |\lambda_i| \int_B |\sigma(t)|^2 dt \tag{3.250}$$

where λ_i is defined in (3.98).

Proof. Since ψ_j form an orthonormal basis for $L^2[B]$, f is representable in the form of (3.249) and for all positive integers i,

$$a_i = \int_B f(t)\psi_i(t)dt = \int_B \left(\int_B \sigma(\xi) e^{ic\langle t,\xi\rangle} d\xi \right) \psi_i(t)dt. \tag{3.251}$$

Combining (3.251) and (3.98) and using Cauchy-Schwarz, we obtain

$$|a_i| = |\lambda_i \int_B \sigma(t)\psi_i(t)dt| \leq |\lambda_i| \int_B |\sigma(t)|^2 dt \int_B |\psi_j(t)|^2 dt = |\lambda_i| \int_B |\sigma(t)|^2 dt \tag{3.252}$$

■

Remark 3.6.3 *Lemma 3.6.3 shows that in order to accurately represent a bandlimited function, f, it is sufficient to find the projection of f onto all GPSFs with corresponding eigenvalue above machine precision. In Section 3.6.1, we approximate the number of GPSFs with large corresponding eigenvalues.*

3.6.1 Dimension of the Class of Bandlimited Functions

In this section, we investigate the properties of the eigenvalues $\mu_0, \mu_1, \ldots, \mu_j, \ldots$ of the operator Q_c, defined via formula (3.99). We denote by λ_j the eigenvalues of operator F_c, defined via formula (3.97), and let ψ_j denote the eigenfunctions corresponding to λ_j, for each nonnegative integer j.

The following two theorems evaluate the sums $\sum_{j=0}^{\infty} \mu_j$ and $\sum_{j=0}^{\infty} \mu_j^2$ respectively.

Theorem 3.6.4 *Suppose that $c > 0$. Then*

$$\sum_{j=0}^{\infty} \mu_j = \frac{c^{p+2}}{2^{p+2}\Gamma(\frac{p}{2}+2)^2}. \tag{3.253}$$

Proof. From (3.98), we observe the identity

$$\sum_{j=0}^{\infty} \lambda_j \psi_j(x) \psi_j(t) = e^{ic\langle x,t \rangle}, \tag{3.254}$$

for all $x, t \in B$, where B is the closed unit ball in $\mathbb{R}^{p+2}$, and the sum on the left hand side converges in the sense of $L^2(B) \otimes L^2(B)$. By taking the squared $L^2(B) \otimes L^2(B)$ norm of both sides and using (3.84), we obtain the formula

$$\sum_{j=0}^{\infty} |\lambda_j|^2 = \frac{\pi^{p+2}}{\Gamma(\frac{p}{2}+2)^2}. \tag{3.255}$$

Since

$$\mu_j = \left(\frac{c}{2\pi}\right)^{p+2} |\lambda_j|^2, \tag{3.256}$$

for all nonnegative integer j (see (3.98)), it follows that

$$\sum_{j=0}^{\infty} \mu_j = \frac{c^{p+2}}{2^{p+2}\Gamma(\frac{p}{2}+2)^2}. \tag{3.257}$$

■

Theorem 3.6.5 *Suppose that $c > 0$. Then*

$$\sum_{j=0}^{\infty} \mu_j^2 = \frac{c^{p+2}}{2^{p+2}\Gamma(\frac{p}{2}+2)^2} - \frac{c^{p+1}\log(c)}{\pi^2 \Gamma(p+2)} + o\big(c^{p+1}\log(c)\big), \tag{3.258}$$

as $c \to \infty$.

Proof. By (3.101),

$$\sum_{j=0}^{\infty} \mu_j \psi_j(x) \psi_j(t) = \Big(\frac{c}{2\pi}\Big)^{p/2+1} \frac{J_{p/2+1}(c\|x-t\|)}{\|x-t\|^{p/2+1}}, \tag{3.259}$$

for all $x, t \in B$, where the sum on the left hand side converges in the sense of $L^2(B) \otimes L^2(B)$, and where J_ν denotes the Bessel functions of the first kind. Taking the squared $L^2(B) \otimes L^2(B)$ norm of both sides, we obtain the formula

$$\begin{aligned}
\sum_{j=0}^{\infty} \mu_j^2 &= \Big(\frac{c}{2\pi}\Big)^{p+2} \int_B \int_B \frac{\big(J_{p/2+1}(c\|x-t\|)\big)^2}{\|x-t\|^{p+2}} \, dx \, dt \\
&= \Big(\frac{c}{2\pi}\Big)^{p+2} \int_B \int_B \frac{\big(J_{p/2+1}(c\|x+t\|)\big)^2}{\|x+t\|^{p+2}} \, dx \, dt \\
&= \Big(\frac{c}{2\pi}\Big)^{p+2} \int_{\mathbb{R}^D} \int_{\mathbb{R}^D} \frac{\big(J_{p/2+1}(c\|x+t\|)\big)^2}{\|x+t\|^{p+2}} \mathbb{1}_B(x) \mathbb{1}_B(t) \, dx \, dt,
\end{aligned} \tag{3.260}$$

where $\mathbb{1}_A$ is defined via the formula

$$\mathbb{1}_A(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A. \end{cases} \tag{3.261}$$

Letting $u = x + t$, we observe that

$$\begin{aligned}
\sum_{j=0}^{\infty} \mu_j^2 &= \Big(\frac{c}{2\pi}\Big)^{p+2} \int_{\mathbb{R}^D} \int_{\mathbb{R}^D} \frac{\big(J_{p/2+1}(c\|u\|)\big)^2}{\|u\|^{p+2}} \mathbb{1}_B(u-t) \mathbb{1}_B(t) \, du \, dt \\
&= \Big(\frac{c}{2\pi}\Big)^{p+2} \int_{\mathbb{R}^D} \int_{\mathbb{R}^D} \frac{\big(J_{p/2+1}(c\|u\|)\big)^2}{\|u\|^{p+2}} \mathbb{1}_{B(2)}(u) \mathbb{1}_B(u-t) \mathbb{1}_B(t) \, du \, dt \\
&= \Big(\frac{c}{2\pi}\Big)^{p+2} \int_{B(2)} \frac{\big(J_{p/2+1}(c\|u\|)\big)^2}{\|u\|^{p+2}} \int_{\mathbb{R}^D} \mathbb{1}_B(u-t) \mathbb{1}_B(t) \, dt \, du.
\end{aligned} \tag{3.262}$$

Combining (3.262) and (3.85),

$$\begin{aligned}
\sum_{j=0}^{\infty} \mu_j^2 &= \Big(\frac{c}{2\pi}\Big)^{p+2} \int_{B(2)} \frac{\big(J_{p/2+1}(c\|u\|)\big)^2}{\|u\|^{p+2}} \cdot V_{p+2}(1) \frac{B_{1-\|u\|^2/4}(\frac{p}{2}+\frac{3}{2},\frac{1}{2})}{B(\frac{p}{2}+\frac{3}{2},\frac{1}{2})}\, du \\
&= \Big(\frac{c}{2\pi}\Big)^{p+2} \frac{V_{p+2}(1)}{B(\frac{p}{2}+\frac{3}{2},\frac{1}{2})} \int_{B(2)} \frac{\big(J_{p/2+1}(c\|u\|)\big)^2}{\|u\|^{p+2}} B_{1-\|u\|^2/4}(\tfrac{p}{2}+\tfrac{3}{2},\tfrac{1}{2})\, du \\
&= \Big(\frac{c}{2\pi}\Big)^{p+2} \frac{V_{p+2}(1)A_{p+2}(1)}{B(\frac{p}{2}+\frac{3}{2},\frac{1}{2})} \int_0^2 \frac{\big(J_{p/2+1}(cr)\big)^2}{r} B_{1-r^2/4}(\tfrac{p}{2}+\tfrac{3}{2},\tfrac{1}{2})\, dr \\
&= \Big(\frac{c}{2\pi}\Big)^{p+2} \frac{V_{p+2}(1)A_{p+2}(1)}{B(\frac{p}{2}+\frac{3}{2},\frac{1}{2})} \int_0^1 \frac{\big(J_{p/2+1}(2cr)\big)^2}{r} B_{1-r^2}(\tfrac{p}{2}+\tfrac{3}{2},\tfrac{1}{2})\, dr, \qquad (3.263)
\end{aligned}$$

where $V_{p+2}(1)$ denotes the volume of the unit ball in $\mathbb{R}^{p+2}$, $A_{p+2}(1)$ denotes the area of the unit sphere in $\mathbb{R}^{p+2}$, $B(a,b)$ denotes the beta function, and $B_x(a,b)$ denotes the incomplete beta function. Applying Theorem 3.2.26 to (3.263),

$$\begin{aligned}
\sum_{j=0}^{\infty} \mu_j^2 &= \frac{c^{p+2}}{2^{p+1}\sqrt{\pi}\Gamma(\frac{p}{2}+1)\Gamma(\frac{p}{2}+\frac{3}{2})} \int_0^1 \frac{\big(J_{p/2+1}(2cr)\big)^2}{r} B_{1-r^2}(\tfrac{p}{2}+\tfrac{3}{2},\tfrac{1}{2})\, dr \\
&= \frac{c^{p+2}}{\pi\Gamma(p+2)} \int_0^1 \frac{\big(J_{p/2+1}(2cr)\big)^2}{r} B_{1-r^2}(\tfrac{p}{2}+\tfrac{3}{2},\tfrac{1}{2})\, dr. \qquad (3.264)
\end{aligned}$$

Combining (3.264) and (3.77),

$$\begin{aligned}
\sum_{j=0}^{\infty} \mu_j^2 &= \frac{c^{p+2}}{\pi\Gamma(p+2)} \bigg(\frac{\sqrt{\pi}\,\Gamma(\frac{p}{2}+\frac{3}{2})}{(p+2)\Gamma(\frac{p}{2}+2)} - \frac{1}{\pi}\frac{\log(c)}{c} + o\Big(\frac{\log(c)}{c}\Big) \bigg) \\
&= \frac{c^{p+2}}{2^{p+2}\Gamma(\frac{p}{2}+2)^2} - \frac{c^{p+1}\log(c)}{\pi^2\Gamma(p+2)} + o\big(c^{p+1}\log(c)\big), \qquad (3.265)
\end{aligned}$$

as $c \to \infty$.

■

The following corollary follows immediately from theorems 3.6.4 and 3.6.5.

Corollary 3.6.6 *Suppose that $c > 0$. Then*

$$\sum_{j=0}^{\infty} \mu_j(1-\mu_j) = \frac{c^{p+1}\log(c)}{\pi^2\Gamma(p+2)} + o\big(c^{p+1}\log(c)\big), \tag{3.266}$$

as $c \to \infty$.

From (3.253) and (3.266) we observe that the spectrum of Q_c consists of three parts:

$$\frac{c^{p+2}}{2^{p+2}\Gamma(\frac{p}{2}+2)^2} \tag{3.267}$$

eigenvalues close to 1;

$$\frac{c^{p+1}\log(c)}{\pi^2\Gamma(p+2)} \tag{3.268}$$

eigenvalues in the transition region; and the rest close to 0.

3.7 Numerical Experiments

The quadrature and interpolation formulas described in Sections 3.5 and 3.6 were implemented in Fortran 77. We used the Lahey/Fujitsu compiler on a 2.9 GHz Intel i7-3520M Lenovo laptop. All examples in this section were run in double precision arithmetic.

In Figure 3.1 and Figure 3.2 we plot the eigenvalues $|\lambda_{N,n}|$ of integral operator F_c (see (3.97)) for different N and different c.

In Figures 3.3, 3.4, 3.5, and 3.6 we plot the GPSFs $\Phi_{N,n}(r)$ (see (3.111)) for different N, n, and c.

In Tables 3.1-3.6, we provide the performance of quadrature rule (3.201) in integrating the function

$$e^{ic\langle x,t\rangle} \tag{3.269}$$

over the unit disk where $x = (0.9, 0.2)$. We provide the results for $c = 20$ and $c = 100$ using both Chebyshev and Gaussian quadratures in the radial direction (see Remark 3.5.4).

In Tables 3.7, 3.8, and 3.9, we provide magnitudes of coefficients of the GPSF expansion of the function on the unit disk $e^{ic\langle x,t\rangle}$ where $x = (0.3, 0.4)$. These coefficients were obtained via interpolation scheme (3.245).

In each table in this section, the column labeled "c" denotes the value of c in (3.269). The column labeled "radial nodes" denotes the number of nodes in the radial direction. These nodes integrate GPSFs. The column labeled "angular nodes" gives the number of equispaced nodes used in the angular direction. The column labeled "N" denotes the N of $\Phi_{N,n}$ (see 3.111). The column labeled "n" denotes the n of $\Phi_{N,n}$. The column labeled "integral via quadrature" denotes value of the integral obtained via quadrature rule (3.201) The column labeled "relative error" denotes the relative error of the integral obtained via quadrature to the true value of the integral. The true value of the integral was obtained by a calculation in extended precision. In Tables 3.7, 3.8, and 3.9, the column labeled $|\alpha_{N,n}|$ denotes the coefficient of $\Phi_{N,n}(r) sin(\theta)$ in the GPSF expansion of (3.269). These coefficients were obtained via formula (3.245).

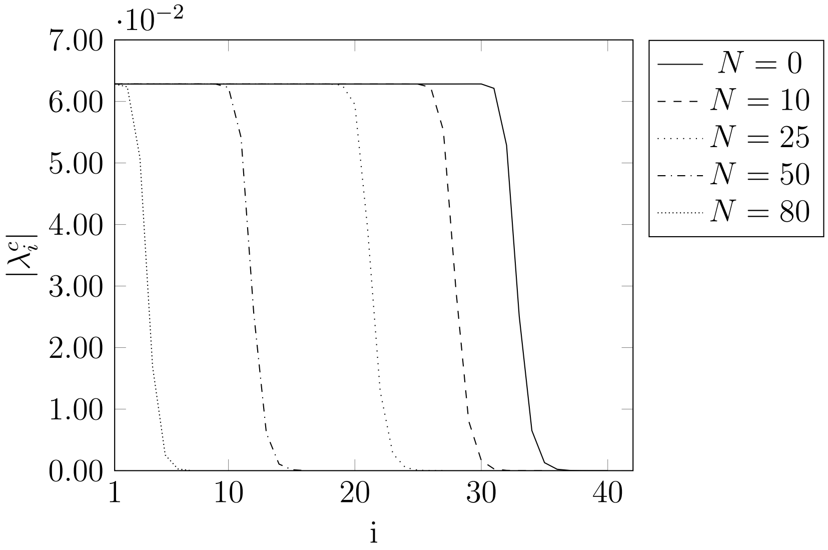

Figure 3.1: Eigenvalues of F_c (see (3.97)) for $c = 100$ and $p = 0$

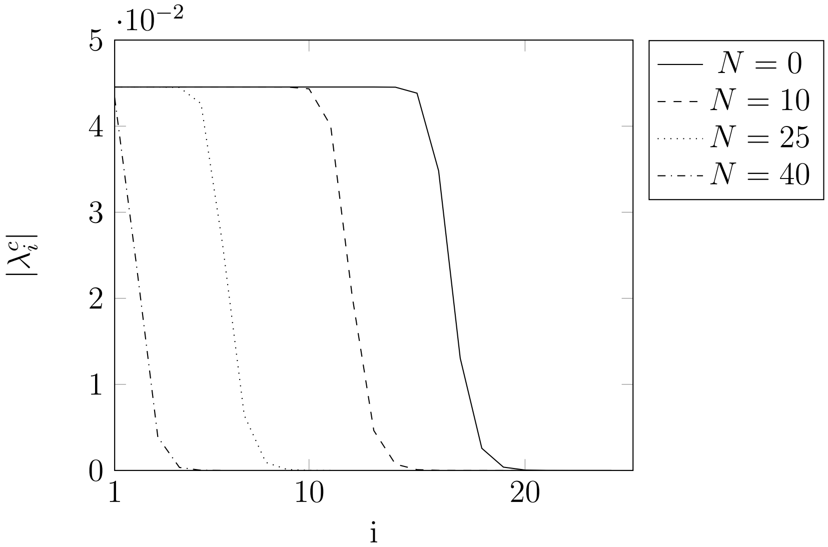

Figure 3.2: Eigenvalues of F_c (see (3.97)) for $c = 50$ and $p = 1$

c	radial nodes	angular nodes	integral via quadrature	relative error
20	6	50	$-0.1076416394449520 + i0.13791 \times 10^{-15}$	0.84109×10^{0}
20	8	50	$-0.0584248723305745 + i0.31659 \times 10^{-15}$	0.70864×10^{-3}
20	10	50	$-0.0584663050529888 + i0.26671 \times 10^{-15}$	0.15834×10^{-7}
20	12	50	$-0.0584663041272412 + i0.27929 \times 10^{-15}$	0.75601×10^{-13}
20	14	50	$-0.0584663041272372 + i0.16220 \times 10^{-15}$	0.68485×10^{-14}
20	16	50	$-0.0584663041272371 + i0.27777 \times 10^{-15}$	0.29262×10^{-14}
20	18	50	$-0.0584663041272375 + i0.23191 \times 10^{-15}$	0.75991×10^{-14}

Table 3.1: Quadratures for $e^{ic\langle x,t\rangle}$ where $x = (0.9, 0.2)$ over the unit disk using several different numbers of radial nodes for $c = 20$. Chebyshev quadratures are used in the radial direction.

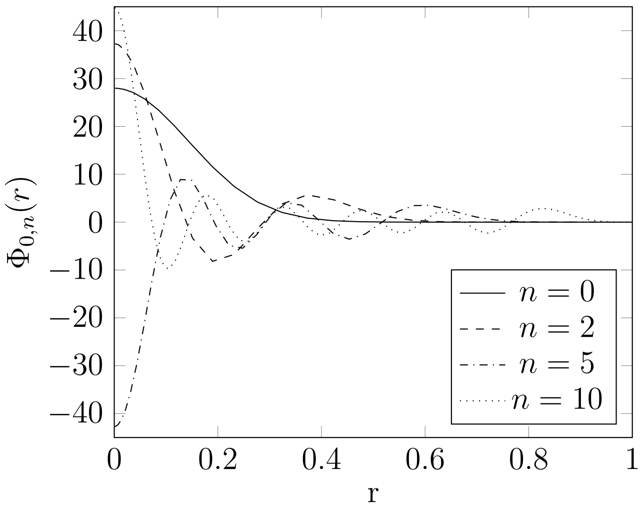

Figure 3.3: Plots of GPSFs $\Phi_{0,n}$ (see (3.111)) with $c = 50$ and $p = 1$

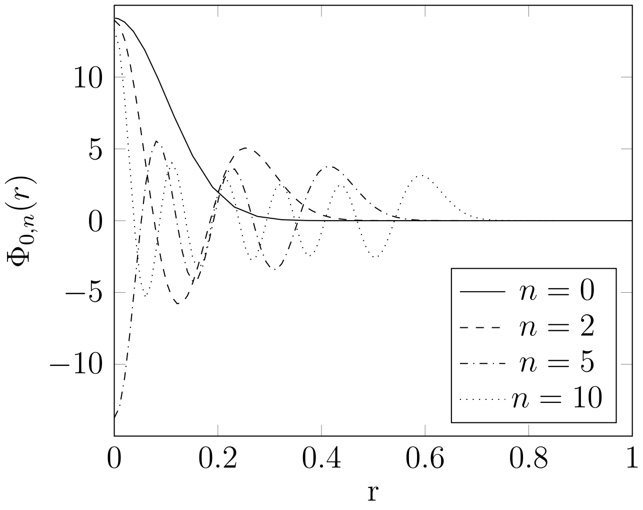

Figure 3.4: Plots of GPSFs $\Phi_{0,n}$ (see (3.111)) with $c = 100$ and $p = 0$

c	radial nodes	angular nodes	integral via quadrature	relative error
20	14	20	$-0.0856165805088149 - i0.57734 \times 10^{-16}$	0.46437×10^{0}
20	14	25	$-0.0584663041272373 + i0.10816 \times 10^{-2}$	0.18500×10^{-1}
20	14	30	$-0.0584748094426783 - i0.18258 \times 10^{-15}$	0.14547×10^{-3}
20	14	35	$-0.0584663041272371 - i0.37973 \times 10^{-8}$	0.64949×10^{-7}
20	14	40	$-0.0584663041418621 + i0.14875 \times 10^{-15}$	0.25015×10^{-9}
20	14	45	$-0.0584663041272375 - i0.94777 \times 10^{-14}$	0.16653×10^{-12}
20	14	50	$-0.0584663041272372 + i0.16220 \times 10^{-15}$	0.51483×10^{-14}
20	14	55	$-0.0584663041272368 + i0.39248 \times 10^{-15}$	0.30672×10^{-14}
20	14	60	$-0.0584663041272371 - i0.13661 \times 10^{-16}$	0.53592×10^{-14}

Table 3.2: Quadratures for $e^{ic\langle x,t\rangle}$ where $x = (0.9, 0.2)$ over the unit disk using several different numbers of angular nodes for $c = 20$. Chebyshev quadratures are used in the radial direction.

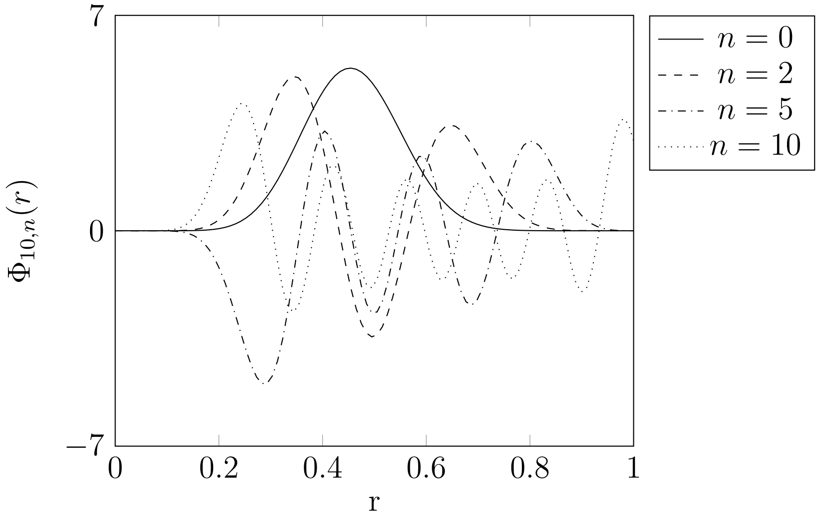

Figure 3.5: Plots of GPSFs $\Phi_{10,n}$ (see (3.111)) with $c = 50$ and $p = 1$

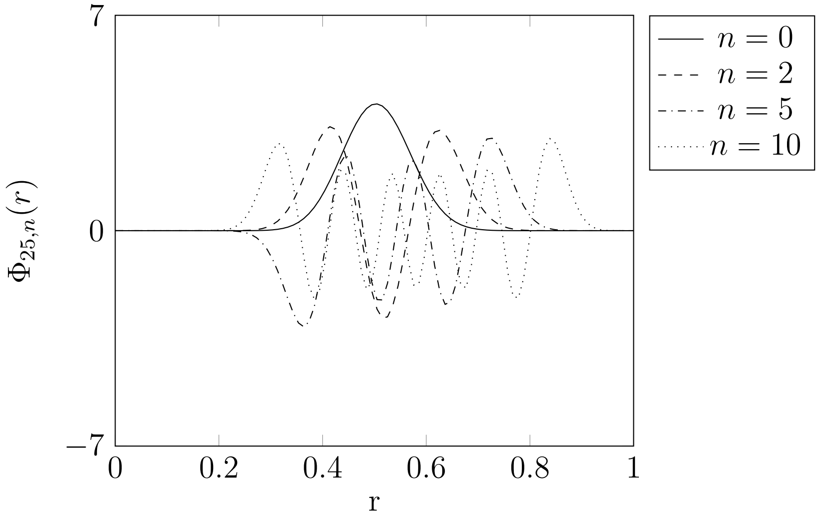

Figure 3.6: Plots of GPSFs $\Phi_{25,n}$ (see (3.111)) with $c = 100$ and $p = 0$

c	radial nodes	angular nodes	integral via quadrature	relative error
20	4	50	$-0.0510613892349747 + i0.37123 \times 10^{-15}$	0.12603×10^{0}
20	6	50	$-0.0584663254751910 + i0.11623 \times 10^{-15}$	0.36513×10^{-6}
20	8	50	$-0.0584663041272613 + i0.37340 \times 10^{-15}$	0.41931×10^{-12}
20	10	50	$-0.0584663041272369 + i0.20903 \times 10^{-15}$	0.15463×10^{-14}
20	12	50	$-0.0584663041272371 + i0.34694 \times 10^{-15}$	0.35160×10^{-14}

Table 3.3: Quadratures for $e^{ic\langle x,t\rangle}$ where $x = (0.9, 0.2)$ over the unit disk using several different numbers of radial nodes for $c = 20$. Gaussian quadratures generated via Algorithm 3.5.2 are used in the radial direction.

c	radial nodes	angular nodes	integral via quadrature	relative error
100	30	140	$0.0164989321769857 - i0.50090 \times 10^{-16}$	0.10612×10^{2}
100	32	140	$-0.0019104800874610 - i0.65264 \times 10^{-15}$	0.11305×10^{0}
100	34	140	$-0.0017165140985462 - i0.21673 \times 10^{-15}$	0.45510×10^{-4}
100	36	140	$-0.0017164370759186 - i0.27856 \times 10^{-15}$	0.63672×10^{-6}
100	38	140	$-0.0017164359820963 - i0.30840 \times 10^{-15}$	0.54009×10^{-9}
100	40	140	$-0.0017164359830235 - i0.30398 \times 10^{-15}$	0.94943×10^{-13}

Table 3.4: Quadratures for $e^{ic\langle x,t\rangle}$ where $x = (0.9, 0.2)$ over the unit disk using several different numbers of radial nodes for $c = 100$. Chebyshev quadratures are used in the radial direction.

c	radial nodes	angular nodes	integral via quadrature	relative error
100	40	115	$-0.0017164359830236 - i0.21183 \times 10^{-6}$	0.12341×10^{-3}
100	40	120	$-0.0017164338146549 - i0.44658 \times 10^{-15}$	0.12633×10^{-5}
100	40	125	$-0.0017164359830231 - i0.48252 \times 10^{-10}$	0.28112×10^{-7}
100	40	130	$-0.0017164359819925 - i0.11947 \times 10^{-14}$	0.60096×10^{-9}
100	40	135	$-0.0017164359830233 + i0.24522 \times 10^{-14}$	0.13296×10^{-11}
100	40	140	$-0.0017164359830235 - i0.30398 \times 10^{-15}$	0.94943×10^{-13}
100	40	145	$-0.0017164359830231 - i0.78770 \times 10^{-15}$	0.23749×10^{-12}
100	40	150	$-0.0017164359830231 - i0.31466 \times 10^{-15}$	0.16075×10^{-12}

Table 3.5: Quadratures for $e^{ic\langle x,t\rangle}$ where $x = (0.9, 0.2)$ over the unit disk using several different numbers of angular nodes for $c = 100$. Chebyshev quadratures are used in the radial direction.

c	radial nodes	angular nodes	integral via quadrature	relative error
100	20	150	$-0.0017164492038983 - i0.52665 \times 10^{-15}$	0.77025×10^{-5}
100	22	150	$-0.0017164359833709 - i0.60352 \times 10^{-15}$	0.20280×10^{-9}
100	24	150	$-0.0017164359830226 - i0.45244 \times 10^{-15}$	0.28465×10^{-12}
100	26	150	$-0.0017164359830229 - i0.35123 \times 10^{-15}$	0.50904×10^{-13}
100	28	150	$-0.0017164359830224 - i0.55262 \times 10^{-15}$	0.35430×10^{-12}
100	30	150	$-0.0017164359830228 - i0.63794 \times 10^{-15}$	0.39846×10^{-12}

Table 3.6: Quadratures for $e^{ic\langle x,t\rangle}$ where $x = (0.9, 0.2)$ over the unit disk using several different numbers of radial nodes for $c = 100$. Gaussian quadratures generated using Algorithm 3.5.2 are used in the radial direction.

radial nodes	angular nodes	c	N	n	$\lvert\alpha_{N,n}\rvert$
40	140	50	1	0	$0.5331000423667240 \times 10^{-2}$
40	140	50	1	1	$0.4428631717847083 \times 10^{-1}$
40	140	50	1	2	$0.1658210569373790 \times 10^{0}$
40	140	50	1	3	$0.3007289752527894 \times 10^{0}$
40	140	50	1	4	$0.1775918995268194 \times 10^{0}$
40	140	50	1	5	$0.1698366869978232 \times 10^{0}$
40	140	50	1	6	$0.1326556850627168 \times 10^{0}$
40	140	50	1	7	$0.1913962335203701 \times 10^{0}$
40	140	50	1	8	$0.1031820332780429 \times 10^{-1}$
40	140	50	1	9	$0.1525659498901890 \times 10^{0}$
40	140	50	1	10	$0.1596240985391338 \times 10^{0}$
40	140	50	1	11	$0.5077661980005956 \times 10^{-1}$
40	140	50	1	12	$0.7004482833257132 \times 10^{-1}$
40	140	50	1	13	$0.1328923889087414 \times 10^{0}$
40	140	50	1	14	$0.1238722286983581 \times 10^{0}$
40	140	50	1	15	$0.6158313809902630 \times 10^{-1}$
40	140	50	1	16	$0.9273653953916678 \times 10^{-2}$
40	140	50	1	17	$0.1222486302912020 \times 10^{-2}$
40	140	50	1	18	$0.5966018610435559 \times 10^{-3}$
40	140	50	1	19	$0.9457503976218055 \times 10^{-4}$
40	140	50	1	20	$0.7272803775518590 \times 10^{-5}$
40	140	50	1	21	$0.2471737500102828 \times 10^{-7}$
40	140	50	1	22	$0.5697214169860662 \times 10^{-7}$
40	140	50	1	23	$0.6261378248559833 \times 10^{-8}$
40	140	50	1	24	$0.2876620855784414 \times 10^{-9}$
40	140	50	1	25	$0.3487372839216281 \times 10^{-11}$
40	140	50	1	26	$0.1344784001636234 \times 10^{-11}$
40	140	50	1	27	$0.8389389113185264 \times 10^{-13}$
40	140	50	1	28	$0.3090050472181085 \times 10^{-14}$
40	140	50	1	29	$0.5438594709432636 \times 10^{-15}$

Table 3.7: Coefficients, obtained via formula (3.245), of the GPSF expansion of the function on the unit disk $e^{ic\langle x,t\rangle}$ where $x = (0.3, 0.4)$.

radial nodes	angular nodes	c	N	n	$\lvert\alpha_{N,n}\rvert$
40	140	50	10	0	$0.6083490415455435 \times 10^{-1}$
40	140	50	10	1	$0.2656230046895768 \times 10^{-1}$
40	140	50	10	2	$0.4475286860599875 \times 10^{-1}$
40	140	50	10	3	$0.4833769722091774 \times 10^{-2}$
40	140	50	10	4	$0.3152644364681537 \times 10^{-1}$
40	140	50	10	5	$0.3440099078209665 \times 10^{-1}$
40	140	50	10	6	$0.1216643028774711 \times 10^{-1}$
40	140	50	10	7	$0.1348650121618380 \times 10^{-1}$
40	140	50	10	8	$0.2761069443786074 \times 10^{-1}$
40	140	50	10	9	$0.2729520957518510 \times 10^{-1}$
40	140	50	10	10	$0.1713503999971936 \times 10^{-1}$
40	140	50	10	11	$0.4647646609621038 \times 10^{-2}$
40	140	50	10	12	$0.5498106244002701 \times 10^{-3}$
40	140	50	10	13	$0.4531628449744277 \times 10^{-3}$
40	140	50	10	14	$0.9388943333348342 \times 10^{-4}$
40	140	50	10	15	$0.1018790565231280 \times 10^{-4}$
40	140	50	10	16	$0.4628420439758330 \times 10^{-6}$
40	140	50	10	17	$0.3302969345113099 \times 10^{-7}$
40	140	50	10	18	$0.7386880328505609 \times 10^{-8}$
40	140	50	10	19	$0.5793842432833322 \times 10^{-9}$
40	140	50	10	20	$0.1808244166685658 \times 10^{-10}$
40	140	50	10	21	$0.8331243140844428 \times 10^{-12}$
40	140	50	10	22	$0.1247624356690115 \times 10^{-12}$
40	140	50	10	23	$0.6402836746674745 \times 10^{-14}$
40	140	50	10	24	$0.3219490617035674 \times 10^{-15}$
40	140	50	10	25	$0.4392156715211933 \times 10^{-16}$
40	140	50	10	26	$0.4216375878565715 \times 10^{-16}$
40	140	50	10	27	$0.1192730971046164 \times 10^{-15}$
40	140	50	10	28	$0.5964172072581517 \times 10^{-16}$
40	140	50	10	29	$0.9795188267888765 \times 10^{-16}$

Table 3.8: Coefficients, obtained via formula (3.245), of the GPSF expansion of the function on the unit disk $e^{ic\langle x,t\rangle}$ where $x = (0.3, 0.4)$.

radial nodes	angular nodes	c	N	n	$\lvert\alpha_{N,n}\rvert$
40	140	50	30	0	$0.4972797526740737 \times 10^{-3}$
40	140	50	30	1	$0.1401428942935588 \times 10^{-2}$
40	140	50	30	2	$0.2710925506457800 \times 10^{-2}$
40	140	50	30	3	$0.3545718524468668 \times 10^{-2}$
40	140	50	30	4	$0.2241476750854641 \times 10^{-2}$
40	140	50	30	5	$0.6682792235496368 \times 10^{-3}$
40	140	50	30	6	$0.1339565034261751 \times 10^{-3}$
40	140	50	30	7	$0.2092420216819930 \times 10^{-4}$
40	140	50	30	8	$0.2648137075865133 \times 10^{-5}$
40	140	50	30	9	$0.2763313112747597 \times 10^{-6}$
40	140	50	30	10	$0.2398228591769509 \times 10^{-7}$
40	140	50	30	11	$0.1734961623216772 \times 10^{-8}$
40	140	50	30	12	$0.1041121888882874 \times 10^{-9}$
40	140	50	30	13	$0.5099613478241473 \times 10^{-11}$
40	140	50	30	14	$0.1958321703329898 \times 10^{-12}$
40	140	50	30	15	$0.5129356249335817 \times 10^{-14}$
40	140	50	30	16	$0.1790936343596214 \times 10^{-15}$
40	140	50	30	17	$0.2114685083013237 \times 10^{-15}$
40	140	50	30	18	$0.1221114171480004 \times 10^{-15}$
40	140	50	30	19	$0.1775028830408308 \times 10^{-15}$
40	140	50	30	20	$0.9115790774963023 \times 10^{-16}$
40	140	50	30	21	$0.7676533284257323 \times 10^{-16}$
40	140	50	30	22	$0.1056865232130847 \times 10^{-15}$
40	140	50	30	23	$0.1282851493246300 \times 10^{-15}$
40	140	50	30	24	$0.1301036117017623 \times 10^{-15}$
40	140	50	30	25	$0.6302967734899496 \times 10^{-16}$
40	140	50	30	26	$0.7542252119317336 \times 10^{-16}$
40	140	50	30	27	$0.6734661033178358 \times 10^{-16}$
40	140	50	30	28	$0.7752009233608849 \times 10^{-16}$
40	140	50	30	29	$0.1184019207536341 \times 10^{-15}$

Table 3.9: Coefficients, obtained via formula (3.245), of the GPSF expansion of the function on the unit disk $e^{ic\langle x,t\rangle}$ where $x = (0.3, 0.4)$.

3.8 Miscellaneous Properties of GPSFs

3.8.1 Properties of the Derivatives of GPSFs

The following theorem follows immediately from (3.115) and (3.119).

Theorem 3.8.1 *Let $c > 0$. Then*

$$\begin{aligned}
&\frac{d}{dx}\Big((x^{p+1} - x^{p+3})\frac{d\Phi_{N,n}}{dx}(x)\Big) \\
&\quad + \Big(\chi_{N,n}x^{p+1} - \frac{(p+1)(p+3)}{4}x^{p+1} - N(N+p)x^{p-1} - c^2x^{p+3}\Big)\Phi_{N,n}(x) = 0,
\end{aligned} \tag{3.270}$$

where $0 < x < 1$ and N and n are arbitrary nonnegative integers.

Corollary 3.8.2 *Let $c > 0$. Then*

$$\begin{aligned}
&x^2(1 - x^2)\Phi''_{N,n}(x) + \big((p+1)x - (p+3)x^3\big)\Phi'_{N,n}(x) \\
&\quad + \Big(\chi_{N,n}x^2 - \frac{(p+1)(p+3)}{4}x^2 - N(N+p) - c^2x^4\Big)\Phi_{N,n}(x) = 0,
\end{aligned} \tag{3.271}$$

where $0 < x < 1$ and N and n are arbitrary nonnegative integers.

The following lemma connects the values of the $(k+2)$nd derivative of the function $\Phi_{N,n}$ with its derivatives of orders $k-4, k-3, \ldots, k+1$, and is obtained by repeated differentiation of (3.271).

Lemma 3.8.3 *Let $c > 0$. Then*

$$\begin{aligned}
&(x^2 - x^4)\Phi^{(k+2)}_{N,n}(x) + \big((2k+1+p)x - (4k+3+p)x^3\big)\Phi^{(k+1)}_{N,n}(x) \\
&+ \Big(k(k+p) - N(N+p) + \big[\chi_{N,n} - \tfrac{1}{4}(p+1)(p+3) \\
&\qquad - 3k(2k+1+p)\big]x^2 - c^2x^4\Big)\Phi^{(k)}_{N,n}(x) \\
&+ \Big(\big[2k\big(\chi_{N,n} - \tfrac{1}{4}(p+1)(p+3)\big) - k(k-1)(4k+1+3p)\big]x - 4kc^2x^3\Big)\Phi^{(k-1)}_{N,n}(x) \\
&+ \Big(k(k-1)\big(\chi_{N,n} - \tfrac{1}{4}(p+1)(p+3)\big) - k(k-1)(k-2)(k+p) - 6k(k-1)c^2x^2\Big)\Phi^{(k-2)}_{N,n}(x) \\
&\qquad - 4k(k-1)(k-2)c^2x\Phi^{(k-3)}_{N,n}(x) - k(k-1)(k-2)(k-3)c^2\Phi^{(k-4)}_{N,n}(x) = 0,
\end{aligned} \tag{3.272}$$

where $0 < x < 1$, N *and* n *are arbitrary nonnegative integers, and* k *is an arbitrary integer so that* $k \geq 4$. *Also,*

$$\begin{aligned}(x^2 - x^4)\Phi''_{N,n}(x) &+ \big((p+1)x - (p+3)x^3\big)\Phi'_{N,n}(x) \\ &+ \Big(-N(N+p) + \big[\chi_{N,n} - \tfrac{1}{4}(p+1)(p+2)\big]x^2 - c^2x^4\Big)\Phi_{N,n}(x) = 0, \quad (3.273)\end{aligned}$$

and

$$\begin{aligned}(x^2 - x^4)\Phi^{(3)}_{N,n}(x) &+ \big((p+3)x - (p+7)x^3\big)\Phi''_{N,n}(x) \\ &+ \Big((p+1) - N(N+p) + \big[\chi_{N,n} - \tfrac{1}{4}(p+1)(p+3) - 3(p+3)\big]x^2 - c^2x^4\Big)\Phi'_{N,n}(x) \\ &+ \Big(2\big[\chi_{N,n} - \tfrac{1}{4}(p+1)(p+3)\big]x - 4c^2x^3\Big)\Phi_{N,n}(x) = 0, \quad (3.274)\end{aligned}$$

and

$$\begin{aligned}(x^2 - x^4)\Phi^{(4)}_{N,n}(x) &+ \big((p+5)x - (p+11)x^3\big)\Phi^{(3)}_{N,n}(x) \\ &+ \Big(2(p+2) - N(N+p) + \big[\chi_{N,n} - \tfrac{1}{4}(p+1)(p+3) - 6(p+5)\big]x^2 - c^2x^4\Big)\Phi''_{N,n}(x) \\ &+ \Big(\big[4\big(\chi_{N,n} - \tfrac{1}{4}(p+1)(p+3)\big) - 6(p+3)\big]x - 8c^2x^3\Big)\Phi'_{N,n}(x) \\ &+ \Big(2\big(\chi_{N,n} - \tfrac{1}{4}(p+1)(p+3)\big) - 12c^2x^2\Big)\Phi_{N,n}(x) = 0, \quad (3.275)\end{aligned}$$

and

$$\begin{aligned}(x^2-x^4)\Phi^{(5)}_{N,n}(x)&+\big((p+7)x-(p+15)x^3\big)\Phi^{(4)}_{N,n}(x)\\ &+\Big(3(p+3)-N(N+p)+\big[\chi_{N,n}-\tfrac{1}{4}(p+1)(p+3)-9(p+7)\big]x^2-c^2x^4\Big)\Phi^{(3)}_{N,n}(x)\\ &+\Big(\big[6\big(\chi_{N,n}-\tfrac{1}{4}(p+1)(p+3)\big)-6(3p+13)\big]x-12c^2x^3\Big)\Phi''_{N,n}(x)\\ &+\Big(6\big(\chi_{N,n}-\tfrac{1}{4}(p+1)(p+3)\big)-6(p+3)-36c^2x^2\Big)\Phi'_{N,n}(x)\\ &\qquad\qquad-24c^2x\Phi_{N,n}(x)=0,\end{aligned}\tag{3.276}$$

where $0<x<1$ and N and n are arbitrary nonnegative integers.

The following corollary and theorem are obtained immediately from Lemma 3.8.3.

Corollary 3.8.4 *Let $c>0$. Then*

$$\begin{aligned}\big(k(k+p)-N(N+p)\big)\Phi^{(k)}_{N,n}(0)&\\ +\Big(k(k-1)\big(\chi_{N,n}-\tfrac{1}{4}(p+1)(p+3)\big)&-k(k-1)(k-2)(k+p)\Big)\Phi^{(k-2)}_{N,n}(0)\\ &-k(k-1)(k-2)(k-3)c^2\Phi^{(k-4)}_{N,n}(0)=0,\end{aligned}\tag{3.277}$$

where N and n are arbitrary nonnegative integers, and k is an arbitrary integer so that $k\geq 4$. Also,

$$N(N+p)\Phi_{N,n}(0)=0,\tag{3.278}$$

and

$$\big((p+1)-N(N+p)\big)\Phi'_{N,n}(0)=0,\tag{3.279}$$

and

$$\big(2(p+2) - N(N+p)\big)\Phi''_{N,n}(0) + 2\big(\chi_{N,n} - \tfrac{1}{4}(p+1)(p+3)\big)\Phi_{N,n}(0) = 0, \tag{3.280}$$

and

$$\begin{aligned}&\big(3(p+3) - N(N+p)\big)\Phi^{(3)}_{N,n}(0)\\ &\qquad + \Big(6\big(\chi_{N,n} - \tfrac{1}{4}(p+1)(p+3)\big) - 6(p+3)\Big)\Phi'_{N,n}(0) = 0,\end{aligned} \tag{3.281}$$

where N and n are arbitrary nonnegative integers.

Theorem 3.8.5 *If $N = 0$, then*

$$\Phi_{N,n}(0) \neq 0, \tag{3.282}$$

where n is an arbitrary nonnegative integer. If $N \geq 1$, then

$$\Phi^{(k)}_{N,n}(0) = 0 \quad \textit{for } k = 0, 1, \ldots, N-1, \tag{3.283}$$

and

$$\Phi^{(N)}_{N,n}(0) \neq 0, \tag{3.284}$$

where n is an arbitrary nonnegative integer.

Theorem 3.8.6 *Suppose that N and n are nonnegative integers. Then*

$$\Phi_{N,n}(1) \neq 0. \tag{3.285}$$

3.8.2 Derivatives of GPSFs and Corresponding Eigenvalues With Respect to c

The following two theorems establish formulas for the derivatives of the eigenvalues $\mu_{N,n}$ (see (3.100)) and $\beta_{N,n}$ (see (3.111)) with respect to c.

Theorem 3.8.7 *Suppose that $c > 0$ is real and that N and n are nonnegative integers. Then*

$$\frac{\partial \beta_{N,n}}{\partial c} = \beta_{N,n} \frac{(\Phi_{N,n}(1))^2 - (p+2)}{2c}, \tag{3.286}$$

and

$$\frac{\partial \mu_{N,n}}{\partial c} = \frac{\mu_{N,n}}{c}((\Phi_{N,n}(1))^2 - (p+1)). \tag{3.287}$$

3.9 Appendix A

3.9.1 Derivation of the Integral Operator Q_c

In this section we derive an explicit formula for the integral operator Q_c, defined in (3.99).

Suppose that B denotes the closed unit ball in $\mathbb{R}^{p+2}$. From (3.99),

$$Q_c[\psi](x) = \left(\frac{c}{2\pi}\right)^{p+2} \int_B \int_B e^{ic\langle x-t,u\rangle} \psi(t) \, du \, dt, \tag{3.288}$$

for all $x \in B$. We observe that

$$e^{ic\langle v,u\rangle} = \sum_{N=0}^{\infty} \sum_{\ell=1}^{h(N,p)} i^N (2\pi)^{p/2+1} \frac{J_{N+p/2}(c\|u\|\|v\|)}{(c\|u\|\|v\|)^{p/2}} S_N^\ell(u/\|u\|) S_N^\ell(v/\|v\|), \tag{3.289}$$

for all $u, v \in B$, where S_N^ℓ denotes the spherical harmonics of degree N, and J_ν denotes Bessel functions of the first kind (see Section VII of [26]). Therefore,

$$\begin{aligned}\int_B e^{ic\langle v,u\rangle}\,du &= (2\pi)^{p/2+1}\int_0^1 \frac{J_{p/2}(c\|v\|\rho)}{(c\|v\|\rho)^{p/2}}\rho^{p+1}\,d\rho \\ &= \frac{(2\pi)^{p/2+1}}{(c\|v\|)^{p/2}}\int_0^1 \rho^{p/2+1}J_{p/2}(c\|v\|\rho)\,d\rho \\ &= \Big(\frac{2\pi}{c}\Big)^{p/2+1}\frac{J_{p/2+1}(c\|v\|)}{\|v\|^{p/2+1}}, \end{aligned} \tag{3.290}$$

for all $v \in \mathbb{R}^{p+2}$, where the last equality follows from formula 6.561(5) in [12]. Combining (3.288) and (3.290),

$$Q_c[\psi](x) = \Big(\frac{c}{2\pi}\Big)^{p/2+1}\int_B \frac{J_{p/2+1}\big(c\|x-t\|\big)}{\|x-t\|^{p/2+1}}\psi(t)\,dt, \tag{3.291}$$

for all $x \in \mathbb{R}^{p+2}$.

Chapter 4

An Algorithm for the Evaluation of the Incomplete Gamma Function

4.1 Background

The evaluation of special functions is one of the most developed areas of numerical analysis. For some special functions, such as Bessel functions, the theory has been fairly complete for many decades; for others, such as Prolate Spheroidal Wave Functions, the theory is still an active area of research. In this respect, the Incomplete Gamma Function occupies an intermediate position. Its mathematical properties appear to be well understood, but the relevant numerical techniques leave much to be desired, at least in certain regimes.

In this chapter is to introduce a numerical scheme, or rather a class of numerical schemes, for the evaluation of the Incomplete Gamma Function. When calculations are performed in double precision, these numerical schemes produce (more or less) full double precision accuracy and are sufficiently fast to be compatible with standard schemes for the evaluation of other special functions. When calculations are performed in extended precision, the schemes produces roughly extended precision accuracy, though in this regime the algorithm loses much of its efficiency. The algorithm is based on the combination of an identity concerning the Incomplete Gamma Function (see [1] formula 6.5.22), with an asymptotic expansion that appears to be new (see [9], [10], [7], [28], [29]); its performance is illustrated with several numerical examples, in both double and extended precision (see Section 4.9).

The structure of this chapter is as follows. In Section 4.2, we introduce notation and summarize a number of elementary mathematical results to be used throughout the remainder of the chapter. Section 4.3 contains apparatus, consisting mainly of technical lemmas, that will be used in proofs in subsequent sections. Section 4.4 contains technical lemmas describing conditions under which, when x is much smaller than m, the function $P(m, x)$ is essentially 0. In Section 4.5, we describe conditions under which, when x is much larger than m, the function $P(m, x)$ is essentially 1. Section 4.6 describes a technique for the evaluation of $P(m, x)$ via direct summation. Section 4.7 describes an asymptotic expansion for the evaluation of $P(m, x)$. Section 4.8 contains a description of an algorithm to evaluate $P(m, x)$ for all $m, x > 0$. Section 4.9 contains the results of numerical experiments with the algorithm for $P(m, x)$ described in Section 4.8.

4.2 Preliminaries

In accordance with standard practice, we will be denoting the Incomplete Gamma Function by $\gamma(m, x)$ where

$$\gamma(m+1, x) = \int_0^x t^m e^{-t} dt, \tag{4.1}$$

for all real numbers $m > -1$ and $x > 0$. We define $\overline{\gamma}(m+1, x)$ by the formula

$$\overline{\gamma}(m+1, x) = \int_0^x \frac{t^m e^{-t}}{m^m e^{-m}} dt. \tag{4.2}$$

Let $\phi(s)$ be defined by the formula

$$\phi(s) = \log((m+s)^m e^{-(m+s)}) - \log(m^m e^{-m}) \tag{4.3}$$

for all real numbers s and we observe that

$$\overline{\gamma}(m+1, x) = \int_{-m}^{x-m} e^{\phi(s)} ds \tag{4.4}$$

We will be denoting by $P(m+1, x)$ (see [1]), the Incomplete Gamma Function scaled by the Complete Gamma Function. That is,

$$P(m+1, x) = \frac{\gamma(m+1, x)}{\Gamma(m+1)} = \frac{m^m e^{-m}}{\Gamma(m+1)} \overline{\gamma}(m+1, x). \tag{4.5}$$

Consistent with standard practice, we denote by $\Gamma(m)$ the Complete Gamma Function,

$$\Gamma(m+1) = \int_0^\infty t^m e^{-t} dt, \tag{4.6}$$

for all real numbers $m > -1$. We define $\overline{\Gamma}(m+1)$ by the formula,

$$\overline{\Gamma}(m+1) = \int_0^\infty \frac{t^m e^{-t}}{m^m e^{-m}} dt. \tag{4.7}$$

for all real numbers $m > -1$. We define f_m to be the function on $\mathbb{C}$ defined by the formula,

$$f_m(z) = \exp\big(m \log(1 + z/m) - z + z^2/2m\big) \tag{4.8}$$

and observe that

$$e^{\phi(s)} = f_m(s)e^{-s^2/2m}. \tag{4.9}$$

The following lemma, Stirling's Formula, is a classical asymptotic expansion for the Gamma Function, $\Gamma(m)$ (see (4.6)). It can be found in, for example, [12], Formula 6.1.37. Proofs for bounds on the error terms for Stirling's Formula, provided in Lemma 4.2.1, can be found in, for example, [18].

Lemma 4.2.1 *[Stirling's Formula] For all $m > -1$,*

$$\begin{aligned}\Gamma(m+1) \sim & e^{-(m+1)}(m+1)^{m+1/2}(2\pi)^{1/2} \\ & \left[1 + \frac{1}{12(m+1)} + \frac{1}{288(m+1)^2} + ... + g_k(m+1) + ...\right]\end{aligned} \tag{4.10}$$

where formulas for $g_k(m+1)$ can be found in [19]. Further, for $K > 2$, if

$$\Gamma(m+1) = e^{-(m+1)}(m+1)^{m+1/2}(2\pi)^{1/2}\left(\sum_{k=0}^{K-1} g_k(m+1) + R_K(m+1)\right) \tag{4.11}$$

then

$$|R_K(m+1)| < \frac{\Gamma(K)}{(K+1)(m+1)^K}. \tag{4.12}$$

Additionally,

$$\begin{aligned}\log(\Gamma(m+1)) \sim &(m+1/2)\log(m+1) - (m+1) + 1/2\log(2\pi) + \frac{1}{12(m+1)} \\ &- \frac{1}{360(m+1)^3} + ... + h_k(m+1) + ...\end{aligned} \tag{4.13}$$

where formulas for $h_k(m)$ can be found in, for example, [1]. Further, suppose that

$$\begin{aligned}\log(\Gamma(m+1)) =&(m+1/2)\log(m+1) - (m+1) + 1/2\log(2\pi) \\ &+ \sum_{k=0}^{K-1} h_k(m+1) + R_K(m+1).\end{aligned} \tag{4.14}$$

Then,

$$|R_K(m+1)| \leq \frac{|B_{2K}|}{(2K-1)(m+1)^{2K-1}}, \tag{4.15}$$

where B_n is n^{th} Bernoulli number (see [1]). In addition, $|R_K(m)|$ is smaller in magnitude than the first neglected term (see [18]).

The following observation will be used throughout the remainder of the chapter.

Observation 4.2.1 *Straightforward application of Stirling's Formula (see (4.10)) and its error bound (4.15) shows that for $m > 1$,*

$$\Gamma(m+1) > m^{m+1/2}e^{-m} > 0. \tag{4.16}$$

It follows immediately from (4.16) that, for $m > 1$,

$$0 < \frac{m^m e^{-m}}{\Gamma(m+1)} < m^{-1/2}. \tag{4.17}$$

The following well-known inequality will be used in the proof of Lemma 4.3.1. A proof of Lemma 4.2.2 can be found in, for example, Section VII of [8].

Lemma 4.2.2 *For all $x, \sigma > 0$,*

$$\frac{1}{(2\pi\sigma^2)^{1/2}} \int_{-\infty}^{-\sigma x} e^{-t^2/2\sigma^2} dt < \frac{1}{x} e^{-x^2/2}. \tag{4.18}$$

4.3 Mathematical Apparatus

The main analytical tools of this section are Lemma 4.3.6 and Corollary 4.3.7. They will be used in Sections 4.4 and 4.5. All other apparatus will be used in the proofs of Lemma 4.3.6 and Corollary 4.3.7 or in Sections 4.4, 4.5, 4.6, and 4.7.

We use the following lemma in the proof of Lemma 4.4.1.

Lemma 4.3.1 *For all $\alpha, m > 0$,*

$$\int_0^{m-\alpha m^{1/2}} m^{-1/2} e^{-(t-m)^2/2m+1} dt < \frac{1}{\alpha} e^{-\alpha^2/2+1}. \tag{4.19}$$

Proof. Clearly,

$$\begin{aligned} \int_0^{m-\alpha m^{1/2}} m^{-1/2} e^{-(t-m)^2/2m+1} dt &= e m^{-1/2} \int_0^{m-\alpha m^{1/2}} e^{-(t-m)^2/2m} dt \\ &< e m^{-1/2} \int_{-\infty}^{m-\alpha m^{1/2}} e^{-(t-m)^2/2m} dt \qquad (4.20) \\ &= e m^{-1/2} \int_{-\infty}^{-\alpha m^{1/2}} e^{-t^2/2m} dt. \end{aligned}$$

Applying Lemma 4.2.2 to (4.20),

$$e m^{-1/2} \int_{-\infty}^{-\alpha m^{1/2}} e^{-(t)^2/2m} dt < \frac{1}{\alpha} e^{-\alpha^2/2+1} \tag{4.21}$$

(4.19) follows immediately from the combination of (4.20) and (4.21). ■

The following elementary inequality will be used in the proof of Lemma 4.3.5.

Lemma 4.3.2 *Let $\{a_n\}_{n=1}^{\infty}$ be a non-negative, monotonically decreasing sequence in $\mathbb{R}$ such that $a_n \to 0$ as $n \to \infty$. Then,*

$$\left|\sum_{n=1}^{\infty}(-1)^{n+1}a_n\right| \leq a_1. \tag{4.22}$$

Proof. Clearly, since $\{a_n\}_{n=1}^{\infty}$ is monotonically decreasing, for all $i \in \mathbb{N}$,

$$a_i - a_{i+1} \geq 0. \tag{4.23}$$

therefore,

$$\sum_{n=1}^{\infty}(-1)^{n+1}a_n = \sum_{n=1}^{\infty}(a_{2n-1} - a_{2n}) \geq 0. \tag{4.24}$$

Combining (4.23) and (4.24) yields

$$0 \leq \sum_{n=1}^{\infty}(-1)^{n+1}a_n = a_1 - \sum_{n=1}^{\infty}(a_{2n} - a_{2n+1}) \leq a_1. \tag{4.25}$$

Inequality (4.22) follows from (4.25). ■

Lemma 4.3.3 and Lemma 4.3.4, indefinite integral identities, will be used in Section 4.8.

Lemma 4.3.3 *For all non-negative integers, n, and real numbers, $m > 0$,*

$$\begin{aligned}\int t^{2n+1}e^{-t^2/2m}dt = 2e^{-t^2/2m}(-2mt^{2n+2}-& \\ \sum_{k=1}^{n+1}(2m)^{k+1}(n+1)(n)...(n-k+2)t^{2(n+1-k)}&),\end{aligned} \tag{4.26}$$

Proof. By the change of variables $x = t^2$,

$$\int_a^b t^{2n+1} e^{-t^2/2m} dt = 2 \int_{a^2}^{b^2} x^{n+1} e^{-x/2m} dx. \tag{4.27}$$

for all $a < b$. From Formula 2.321.2 in [12], we know that for all non-negative integers n,

$$\int x^n e^{-x/2m} dx = e^{-x/2m} \left(-2mx^n - \sum_{k=1}^{n} (2m)^{k+1} n(n-1)...(n-k+1)x^{n-k} \right). \tag{4.28}$$

Combining (4.27) and (4.28) yields (4.26). ■

Lemma 4.3.4 *For all non-negative integers, n, and real numbers, m > 0,*

$$\int_0^x t^{2n+2} e^{-t^2/2m} dt = \prod_{i=0}^{n} \alpha_i (2m)^{1/2} \frac{\sqrt{\pi}}{2} erf\left(\frac{x}{\sqrt{2m}} \right) - \sum_{i=0}^{n-1} \left(\beta_i \prod_{j=0}^{n-1-i} \alpha_{n-j} \right) - \beta_n, \tag{4.29}$$

where

$$\alpha_i = (2i+1)m \tag{4.30}$$

and

$$\beta_i = mx^{2i+1} e^{-x^2/2m} \tag{4.31}$$

for $i \in \{0, 1, ...\}$, and where, in accordance with standard practice,

$$\mathit{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt. \tag{4.32}$$

Proof. Integrating by parts,

$$\int_0^x t^{2n} e^{-t^2/2m} dt = \frac{x^{2n+1}}{2n+1} e^{-x^2/2m} + \frac{1}{m(2n+1)} \int_0^x t^{2n+2} e^{-t^2/2m} dt. \tag{4.33}$$

Now, rearranging the terms of (4.33),

$$\int_0^x t^{2n+2} e^{-t^2/2m} dt = (2n+1)m \int_0^x t^{2n} e^{-t^2/2m} dt - m x^{2n+1} e^{-x^2/2m}. \tag{4.34}$$

Repeated application of identity (4.34) yields,

$$\int_0^x t^{2n+2} e^{-t^2/2m} dt = \prod_{i=0}^{n} \alpha_i \int_0^x e^{-t^2/2m} dt - \sum_{i=0}^{n-1} \left(\beta_i \prod_{j=0}^{n-1-i} \alpha_{n-j} \right) - \beta_n, \tag{4.35}$$

where

$$\alpha_i = (2i+1)m \tag{4.36}$$

and

$$\beta_i = m x^{2i+1} e^{-x^2/2m}. \tag{4.37}$$

Through a straightforward change of variables, we obtain the identity

$$\int_a^b e^{-t^2/2m} dt = (2m)^{1/2} \frac{\sqrt{\pi}}{2} \left(\mathrm{erf}\left(\frac{b}{\sqrt{2m}}\right) - \mathrm{erf}\left(\frac{a}{\sqrt{2m}}\right) \right), \tag{4.38}$$

where, in accordance with standard practice,

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt. \tag{4.39}$$

(4.29) follows directly from the combination of (4.35) and (4.38). ■

The following bound will be used in the proof of Lemma 4.3.6.

Lemma 4.3.5 *For all $m > 0$ and $\alpha \in (-m^{1/2}, m^{1/2})$,*

$$-\sum_{j=3}^{\infty} \frac{(\alpha/m^{1/2})^j}{j} \leq 1. \tag{4.40}$$

Proof. We will consider two cases.

Case 1: $\alpha \in [0, m^{1/2})$. Clearly, since $m > 0$,

$$\alpha/m^{1/2} \geq 0. \tag{4.41}$$

It follows immediately that

$$-\sum_{j=3}^{\infty} \frac{(\alpha/m^{1/2})^j}{j} \leq 0 \tag{4.42}$$

for all $\alpha \in [0, m^{1/2})$.

Case 2: $\alpha \in (-m^{1/2}, 0)$. Clearly, for all $\alpha \in (-m^{1/2}, 0)$,

$$-\sum_{j=3}^{\infty} \frac{(\alpha/m^{1/2})^j}{j} = \sum_{j=3}^{\infty} (-1)^{j+1} \frac{|\alpha/m|^{j/2}}{j}. \tag{4.43}$$

Furthermore, the sequence

$$\left\{ \frac{|\alpha/m|^{j/2}}{j} \right\}_{j=3}^{\infty} \tag{4.44}$$

is a non-negative, monotonically decreasing sequence in j. Therefore, according to Lemma 4.3.2,

$$\sum_{j=3}^{\infty}(-1)^{j+1}\frac{|\alpha/m|^{j/2}}{j} < \frac{|\alpha^3|}{3m^{3/2}} < \frac{1}{3}. \tag{4.45}$$

Combining (4.43) and (4.45) yields

$$-\sum_{j=3}^{\infty}\frac{(\alpha/m^{1/2})^j}{j} < \frac{1}{3}. \tag{4.46}$$

for all $\alpha \in (-m^{1/2}, 0)$. Combining (4.42) and (4.46) yields (4.40). ■

The following lemma will be used in the proof of Corollary 4.3.7.

Lemma 4.3.6 *For all $m > 1$ and $\alpha \in (-m^{1/2}, m^{1/2})$,*

$$\frac{(m-\alpha m^{1/2})^m e^{-(m-\alpha m^{1/2})}}{\Gamma(m+1)} < m^{-1/2}e^{-\alpha^2/2+1}, \tag{4.47}$$

where $\Gamma(m)$ is defined in equation (4.6).

Proof. It follows immediately from Observation 4.2.1 that for all $m > 1$,

$$\begin{aligned}\frac{(m-\alpha m^{1/2})^m e^{-(m-\alpha m^{1/2})}}{\Gamma(m+1)} &< \frac{(m-\alpha m^{1/2})^m e^{-(m-\alpha m^{1/2})}}{m^{m+1/2}e^{-m}}\\ &= m^{-1/2}\left(\frac{m-\alpha m^{1/2}}{m}\right)^m \frac{e^{-(m-\alpha m^{1/2})}}{e^{-m}}.\end{aligned} \tag{4.48}$$

Clearly,

$$\begin{aligned}m^{-1/2}\left(\frac{m-\alpha m^{1/2}}{m}\right)^m \frac{e^{-(m-\alpha m^{1/2})}}{e^{-m}} &= m^{-1/2}\left(1-\frac{\alpha}{m^{1/2}}\right)^m e^{\alpha m^{1/2}}\\ &= m^{-1/2}\exp\left(m\log\left(1-\frac{\alpha}{m^{1/2}}\right)\right)e^{\alpha m^{1/2}}.\end{aligned} \tag{4.49}$$

Expanding

$$\log(1 - \frac{\alpha}{m^{1/2}}) \tag{4.50}$$

into Taylor series yields

$$\begin{aligned} m^{-1/2} \exp\left(m \log\left(1 - \frac{\alpha}{m^{1/2}}\right)\right) e^{\alpha m^{1/2}} &= m^{-1/2} \exp\left(-m \sum_{j=1}^{\infty} \frac{(\alpha/m^{1/2})^j}{j}\right) e^{\alpha m^{1/2}} \\ &= m^{-1/2} e^{-\alpha^2/2} \exp\left(-\sum_{j=3}^{\infty} \frac{(\alpha/m^{1/2})^j}{j}\right) \end{aligned} \tag{4.51}$$

for $\alpha \in (-m^{1/2}, m^{1/2})$. According to Lemma 4.3.5,

$$e^{-\alpha^2/2} \exp\left(-\sum_{j=3}^{\infty} \frac{(\alpha/m^{1/2})^j}{j}\right) < e^{-\alpha^2/2+1} \tag{4.52}$$

for all $\alpha \in (-m^{1/2}, m^{1/2})$. Combining (4.49), (4.51), and (4.52) yields (4.47). ■

The following inequality provides a bound on the integrand of $P(m+1, x)$ and will be used in the proofs of Lemmas 4.4.1 and 4.5.1.

Corollary 4.3.7 *For all $m > 0$ and $t \in (0, 2m)$,*

$$\frac{t^m e^{-t}}{\Gamma(m+1)} < m^{-1/2} e^{-\frac{(t-m)^2}{2m}+1}. \tag{4.53}$$

Proof. Obviously, for all $m > 0$,

$$t = m - (\frac{-t+m}{m^{1/2}}) m^{1/2}. \tag{4.54}$$

Combining Observation 4.2.1 and (4.54) we have

$$\frac{t^m e^{-t}}{\Gamma(m+1)} = \frac{\left(m - (\frac{-t+m}{m^{1/2}})m^{1/2}\right)^m e^{-\left(m-(\frac{-t+m}{m^{1/2}})m^{1/2}\right)}}{\Gamma(m+1)}. \tag{4.55}$$

Combining Lemma 4.3.6 with (4.55) yields,

$$\begin{aligned}\frac{\left(m - (\frac{-t+m}{m^{1/2}})m^{1/2}\right)^m e^{-\left(m-(\frac{-t+m}{m^{1/2}})m^{1/2}\right)}}{\Gamma(m+1)} &< m^{-1/2} e^{-\frac{1}{2}\left(\frac{-t+m}{m^{1/2}}\right)^2+1} \\ &= m^{-1/2} e^{\frac{-(t-m)^2}{2m}+1}\end{aligned} \tag{4.56}$$

for $\frac{-t+m}{m^{1/2}} \in (-m^{1/2}, m^{1/2})$ or, equivalently, $t \in (0, 2m)$. ■

4.4 $P(m+1,x)$ for Small x

The principal purpose of this section is to introduce Lemma 4.4.1, which shows that for sufficiently small x, the function $P(m+1,x)$ is essentially 0.

Lemma 4.4.1 *For all $m > 1$ and $\alpha \in (0, m^{1/2})$,*

$$P(m+1, m - \alpha m^{1/2}) < \frac{1}{\alpha} e^{-\alpha^2/2+1} \tag{4.57}$$

with $P(m+1,x)$ defined in (4.5).

Proof. Using (4.2) and applying Corollary 4.3.7 and Lemma 4.3.1, we have

$$\begin{aligned}P(m+1, m-\alpha m^{1/2}) &= \int_0^{m-\alpha m^{1/2}} \frac{t^m e^{-t}}{\Gamma(m+1)} dt \\ &< \int_0^{m-\alpha m^{1/2}} m^{-1/2} e^{-\frac{(t-m)^2}{2m}+1} dt \\ &< \frac{1}{\alpha} e^{-\alpha^2/2+1}.\end{aligned} \tag{4.58}$$

■

Remark 4.4.1 *Suppose $m > 1$. By observing that $P(m+1,x)$ is non-negative for all $x > 0$, and applying Lemma 4.4.1 with $\alpha = m^{1/6}$, we obtain the bound*

$$|P(m+1, m - m^{2/3})| < m^{-1/6} e^{\frac{-m^{1/3}}{2}+1} \tag{4.59}$$

for all $m > 1$ where $P(m+1,x)$ is defined in (4.5).

4.5 $P(m+1, x)$ for Large x

The main purpose of this section is to introduce Lemma 4.5.3, which shows that, for sufficiently large x, the function $P(m+1,x)$ is well approximated by 1.

In the following lemma, we provide a bound to be used in the proof of Lemma 4.5.3.

Lemma 4.5.1 *For all $m > 1$ and $\alpha \in (0, m^{1/2})$,*

$$|P(m+1, 2m) - P(m+1, m + \alpha m^{1/2})| < \frac{1}{\alpha} e^{-\alpha^2/2+1}, \tag{4.60}$$

where $P(m+1,x)$ is defined in (4.5).

Proof. Clearly, by (4.5) and applying Corollary 4.3.7,

$$\begin{aligned} |P(m+1, 2m) - P(m+1, m + \alpha m^{1/2})| &= \int_{m+\alpha m^{1/2}}^{2m} \frac{t^m e^{-t}}{\Gamma(m+1)} dt \\ &< \int_{m+\alpha m^{1/2}}^{2m} m^{-1/2} e^{-\frac{(t-m)^2}{2m}+1} dt. \end{aligned} \tag{4.61}$$

Applying Lemma 4.4.1 and the change of variables $s = 2m - t$ to (4.61), we obtain

$$\int_{m+\alpha m^{1/2}}^{2m} m^{-1/2} e^{-\frac{(t-m)^2}{2m}+1} dt = \int_{0}^{m-\alpha m^{1/2}} m^{-1/2} e^{-\frac{(s-m)^2}{2m}+1} ds < \frac{1}{\alpha} e^{-\alpha^2/2+1}. \quad (4.62)$$

Combining (4.61) and (4.62) yields (4.60). ■

In the following lemma, we provide a bound to be used in the proof of Lemma 4.5.3.

Lemma 4.5.2 *For all* $m > 1$,

$$|1 - P(m+1, 2m)| < 10 m^{-1/2} e^{-m/5}, \quad (4.63)$$

where $P(m+1, x)$ *is defined in (4.5).*

Proof. Obviously, by (4.2) and (4.7),

$$\begin{aligned} |1 - P(m+1, 2m)| &= \int_{2m}^{\infty} \frac{t^m e^{-t}}{\Gamma(m+1)} dt \\ &= \int_{2m}^{\infty} \frac{t^m e^{-9t/10}}{\Gamma(m+1)} e^{-t/10} dt \\ &= \int_{2m}^{\infty} \psi_m(t) e^{-t/10} dt, \end{aligned} \quad (4.64)$$

where, by Observation 4.2.1,

$$\psi_m(t) = \frac{t^m e^{-9t/10}}{\Gamma(m+1)} < m^{-1/2} \frac{t^m e^{-9t/10}}{m^m e^{-m}}. \quad (4.65)$$

We now provide a bound for $\psi_m(t)$ on the interval $t \in (2m, \infty)$. Straightforward differentiation shows that for all $m > 1$ and $t \in (2m, \infty)$, the function $\psi_m(t)$ is decreasing as a function of t. Therefore, using (4.65), for $t \in (2m, \infty)$,

$$\psi_m(t) \le \psi_m(2m) < m^{-1/2} \frac{(2m)^m e^{-9m/5}}{m^m e^{-m}} = m^{-1/2} 2^m e^{-4m/5} < m^{-1/2}. \quad (4.66)$$

Therefore, combining (4.64) and (4.66) yields,

$$\int_{2m}^{\infty} \frac{t^m e^{-9t/10}}{\Gamma(m+1)} e^{-t/10} dt < \int_{2m}^{\infty} m^{-1/2} e^{-t/10} dt = 10 m^{-1/2} e^{-m/5}. \tag{4.67}$$

Combining (4.64) and (4.67) yields (4.63). ■

The following lemma shows that for sufficiently large x, the function $P(m+1, x)$ is well-approximated by 1.

Lemma 4.5.3 *For all $m > 1$ and $\alpha \in (0, m^{1/2})$,*

$$\left|1 - P(m+1, m + \alpha m^{1/2})\right| < \frac{1}{\alpha} e^{-\alpha^2/2+1} + 10 m^{-1/2} e^{-m/5}, \tag{4.68}$$

where $P(m+1, x)$ is defined in (4.5).

Proof. Obviously, by (4.2) and (4.5),

$$\begin{aligned} \left|1 - P(m+1, m + \alpha m^{1/2})\right| &= \int_{m+\alpha m^{1/2}}^{\infty} \frac{t^m e^{-t}}{\Gamma(m+1)} dt \\ &= \int_{m+\alpha m^{1/2}}^{2m} \frac{t^m e^{-t}}{\Gamma(m+1)} dt + \int_{2m}^{\infty} \frac{t^m e^{-t}}{\Gamma(m+1)} dt. \end{aligned} \tag{4.69}$$

According to Lemma 4.5.1,

$$\int_{m+\alpha m^{1/2}}^{2m} \frac{t^m e^{-t}}{\Gamma(m+1)} dt < m^{1/2} e^{-\alpha^2/2+1}. \tag{4.70}$$

According to Lemma 4.5.2,

$$\int_{2m}^{\infty} \frac{t^m e^{-t}}{\Gamma(m+1)} dt < 10 m^{-1/2} e^{-m/5}. \tag{4.71}$$

Combining (4.69), (4.70), and (4.71) yields (4.68). ■

The following corollary will be used in Section 4.8. It shows that for fixed m and sufficiently large x, the function $P(m+1, x)$ is well-approximated by 1.

Corollary 4.5.4 *For all $m > 1$ and $x > m$,*

$$|1 - P(m+1, x)| < \frac{m^{1/2}}{x-m} e^{-\frac{(x-m)^2}{2m}+1} + 10m^{-1/2}e^{-m/5}, \tag{4.72}$$

where $P(m+1, x)$ is defined in (4.2).

Proof. We consider two cases.

Case 1. Suppose $x \in (m, 2m)$. Obviously, by (4.2),

$$\begin{aligned} |1 - P(m+1, x)| &= \int_x^\infty \frac{t^m e^{-t}}{\Gamma(m+1)} dt \\ &= \int_x^{2m} \frac{t^m e^{-t}}{\Gamma(m+1)} dt + \int_{2m}^\infty \frac{t^m e^{-t}}{\Gamma(m+1)} dt. \end{aligned} \tag{4.73}$$

Using the identity

$$x = m + \left(\frac{x-m}{m^{1/2}} \right) m^{1/2} \tag{4.74}$$

and applying Lemma 4.5.1 to (4.73),

$$\int_x^{2m} \frac{t^m e^{-t}}{\Gamma(m+1)} dt = \int_{m+\frac{x-m}{m^{1/2}}m^{1/2}}^{2m} \frac{t^m e^{-t}}{\Gamma(m+1)} dt < \frac{m^{1/2}}{x-m} e^{-\frac{(x-m)^2}{2m}+1}. \tag{4.75}$$

Applying Lemma 4.5.2 to (4.73),

$$\int_{2m}^\infty \frac{t^m e^{-t}}{\Gamma(m+1)} dt < 10m^{-1/2}e^{-m/5}. \tag{4.76}$$

Combining (4.73), (4.75), and (4.76) yields (4.72) for all $x \in (m, 2m)$.

Case 2. Suppose $x \geq 2m$. Obviously, by (4.2),

$$|1 - P(m+1,x)| = \int_x^\infty \frac{t^m e^{-t}}{\Gamma(m+1)} dt < \int_{2m}^\infty \frac{t^m e^{-t}}{\Gamma(m+1)} dt. \tag{4.77}$$

Applying Lemma 4.5.2 to (4.77),

$$\int_{2m}^\infty \frac{t^m e^{-t}}{\Gamma(m+1)} dt < 10 m^{-1/2} e^{-m/5}. \tag{4.78}$$

Combining (4.77) and (4.78) yields (4.72) for all $x \geq 2m$. ■

4.6 Evaluation of $P(m+1,x)$ for Small m and Intermediate x via Summation

The principal purpose of this section is Lemma 4.6.2, which provides a formula for evaluating $P(m+1,x)$ (see (4.2)), for all real numbers $m > -1$ and $x > 0$, and a bound on the error of the approximation.

The following lemma provides a bound for $P(m+1,x)$ for small x and will be used in the proof of Lemma 4.6.2.

Lemma 4.6.1 *For all real numbers $m > 1$, $k \geq 0$, $x \in (0,m)$,*

$$P(m+k+1,x) < \frac{(m+k)^{1/2}}{m+k-x} \exp\left(\frac{-(k+m-x)^2}{2k+2m} + 1\right), \tag{4.79}$$

where $P(m+1,x)$ is defined in (4.5).

Proof. Clearly, since $m > 0$ and $k > 0$,

$$x = (m+k) - \alpha(m+k)^{1/2}, \tag{4.80}$$

where

$$\alpha = \frac{m+k-x}{(m+k)^{1/2}}. \tag{4.81}$$

Therefore, by (4.80), (4.81), and applying Lemma 4.4.1, we obtain,

$$\begin{aligned} P(m+k+1,x) &= P(m+k+1,(m+k)-\alpha(m+k)^{1/2}) \\ &< \frac{(m+k)^{1/2}}{m+k-x}\exp\left(\frac{-(k+m-x)^2}{2k+2m}+1\right). \end{aligned} \tag{4.82}$$

■

In the following lemma, we provide a formula for evaluating $P(m+1,x)$ and a bound on the error. In Table 4.1, we list values of the bound on the error for different values of m, x, k.

Lemma 4.6.2 *For all real numbers $m > -1$, $x > 0$, and for any positive integer $k > m+2-x$,*

$$P(m+1,x) = \sum_{i=0}^{k} \frac{x^{m+1+i}e^{-x}}{\Gamma(m+2+i)} + \rho_{k+1}(m+1,x), \tag{4.83}$$

where

$$|\rho_{k+1}(m+1,x)| < \frac{(m+1+k)^{1/2}}{m+1+k-x}\exp\left(\frac{-(k+m+1-x)^2}{2k+2m+2}+1\right). \tag{4.84}$$

Proof. By Formula 6.5.21 of [1],

$$P(m+1,x) = P(m,x) - \frac{x^m e^{-x}}{\Gamma(m+1)}. \tag{4.85}$$

Iteratively applying identity (4.85) k times yields

$$P(m+1,x) = P(m+k+2,x) + \sum_{i=0}^{k} \frac{x^{m+1+i}e^{-x}}{\Gamma(m+2+i)}. \tag{4.86}$$

According to Lemma 4.6.1,

$$0 < P(m+k+2,x) < \frac{(m+1+k)^{1/2}}{m+k+1-x} \exp\left(\frac{-(k+m+1-x)^2}{2k+2m+2}+1\right). \tag{4.87}$$

Combining (4.86) and (4.87) yields (4.84). ■

Observation 4.6.1 *For all $m > -1$, $\alpha \in (0,(m+1)^{1/2})$, and $x \in (m+1-\alpha(m+1)^{1/2}, m)$, by applying Lemma 4.6.2 with $k \geq \lambda(m+1)^{1/2}$ where $\lambda \in (1,(m+1)^{1/2})$, we obtain the bound*

$$\rho_{k+1}(m+1,x) < \frac{(m+1+\lambda(m+1)^{1/2})^{1/2}}{\lambda(m+1)^{1/2}} \exp\left(\frac{-(\lambda-\alpha)^2}{4}+1\right). \tag{4.88}$$

4.7 Evaluation of $P(m+1,x)$ for Large m and Intermediate x via Asymptotic Expansion

In this section, we introduce an asymptotic expansion for the evaluation of $P(m+1,x)$ for sufficiently large m and $x \in (-m^{2/3}, m^{2/3})$.

We will denote by S_m (see Figure 4.1) the tubular $m^{2/3}$-neighborhood of the real interval $x \in (-m^{2/3}, m^{2/3})$. That is,

$$S_m = \{z \in \mathbb{C} : |z-x| \leq m^{2/3} \text{ for some } x \in (-m^{2/3}, m^{2/3})\}. \tag{4.89}$$

$m+1$	x	k	error bound (see (4.84))
100	50	60	$0.1183591 \times 10^{-16}$
100	75	95	$0.2915023 \times 10^{-16}$
100	100	130	$0.3512477 \times 10^{-16}$
100	125	165	$0.2748450 \times 10^{-16}$
100	150	200	$0.1624504 \times 10^{-16}$
1000	800	80	$0.5502590 \times 10^{-16}$
1000	900	200	$0.1624504 \times 10^{-16}$
1000	1000	310	$0.3731746 \times 10^{-16}$
1000	1100	420	$0.7018108 \times 10^{-16}$
1000	1200	540	$0.1571927 \times 10^{-16}$
10000	9200	50	$0.7854609 \times 10^{-16}$
10000	9500	380	$0.1984442 \times 10^{-16}$
10000	10000	900	$0.2302159 \times 10^{-16}$
10000	10500	1420	$0.2542989 \times 10^{-16}$
10000	10800	1730	$0.3085479 \times 10^{-16}$

Table 4.1: Bounds on the error of evaluating $P(m+1, x)$ via sum (4.83) using k terms for different m, x, k.

Observation 4.7.1 *We observe that $|z| < 2m^{2/3}$ for all $m > 0$ and $z \in S_m$. In particular, if $m > 100$, then for all $z \in S_m$, $|z| < 2m^{2/3} < m$.*

The following observations will be used in the proof of Lemma 4.7.2.

Observation 4.7.2 *Suppose we choose the branch cut for f_m to be the negative real axis with $x < -m$. Then, by (4.8), (4.89), and applying Observation 4.7.1, we observe that f_m is analytic on S_m, where f_m is defined in (4.8) and S_m is defined in (4.89).*

Observation 4.7.3 *It follows immediately from the combination of Observation 4.7.2 and Observation 4.7.1 that for all $m > 100$ and $\xi \in (-m^{2/3}, m^{2/3})$, the function f_m is analytic on the disk of radius $m^{2/3}$ centered at ξ, where f_m is defined in (4.8).*

The following lemma will be used in the proof of Lemma 4.7.2.

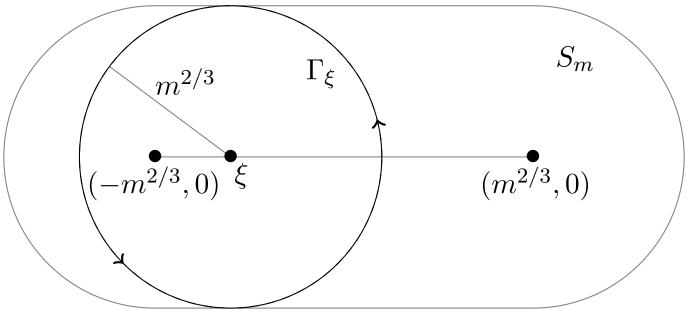

Figure 4.1: An illustrative domain of S_m, the $m^{2/3}$-neighborhood of the interval on the real line $(-m^{2/3}, m^{2/3})$.

Lemma 4.7.1 *For all $m > 0$ and $z \in S_m$,*

$$|f_m(z)| < 15, \tag{4.90}$$

where f_m is defined in (4.8) and S_m is defined in (4.89).

Proof. Obviously, by (4.8), for all $m > 0$ and $z \in S_m$,

$$\begin{aligned} |f_m(z)| &= |\exp[m\log(1+z/m) - z + z^2/2m]| \\ &= \exp[\mathrm{Re}\{m\log(1+z/m) - z + z^2/2m\}]. \end{aligned} \tag{4.91}$$

Hence, expanding $\log(1+z)$ into Taylor series, we have

$$\begin{aligned} |f_m(z)| &= \exp[\mathrm{Re}\{m\sum_{k=1}^{\infty}(-1)^{k+1}(\frac{z}{m})^k - z + z^2/2m\}] \\ &= \exp[\mathrm{Re}\{\sum_{k=3}^{\infty}(-1)^{k+1}\frac{z^k}{km^{k-1}}\}]. \end{aligned} \tag{4.92}$$

Therefore, by (4.92), and the combination of Observation 4.7.1 and Lemma 4.3.2,

$$|f_m(z)| \le \exp[\mathrm{Re}\{\sum_{k=3}^{\infty}(-1)^{k+1}\frac{(2m^{2/3})^k}{km^{k-1}}\}] \le \exp\left(\frac{(2m^{2/3})^3}{3m^2}\right) = e^{8/3} < 15. \tag{4.93}$$

■

In Lemma 4.7.2, we provide a bound to be used in the proof of Theorem 4.7.4.

Lemma 4.7.2 *For all $m > 100$ and $\xi \in (-m^{2/3}, m^{2/3})$,*

$$\left|\frac{f_m^{(k)}(\xi)}{k!}\right| < \frac{15}{m^{2k/3}}, \tag{4.94}$$

where $f_m^{(k)}$ is the k^{th} derivative of f_m defined in (4.8).

Proof. Let $m > 100$ and $\xi \in (-m^{2/3}, m^{2/3})$. Let Γ_ξ be the positively oriented circular contour of radius $m^{2/3}$ centered at ξ (see Figure 4.1). Then combining Observation 4.7.2 and the Cauchy Integral Formula and applying elementary integral transformations, we obtain

$$\begin{aligned}\left|\frac{f^{(k)}(\xi)}{k!}\right| &= \left|\frac{1}{2\pi i}\int_{\Gamma_\xi}\frac{f(z)}{(z-\xi)^{k+1}}dz\right| \\ &\le \frac{1}{2\pi}\int_{\Gamma_\xi}\frac{|f(z)|}{|(z-\xi)^{k+1}|}dz.\end{aligned} \tag{4.95}$$

By applying Lemma 4.7.1 to (4.95), we have

$$\begin{aligned}\left|\frac{f^{(k)}(\xi)}{k!}\right| &\le \frac{1}{2\pi}\int_{\Gamma_\xi}\frac{15}{|(z-\xi)^{k+1}|}dz \\ &= \frac{15}{2\pi}\int_{\Gamma_\xi}\frac{1}{m^{2(k+1)/3}}dz.\end{aligned} \tag{4.96}$$

Now, combining (4.96) with the fact that Γ_ξ is of length $2\pi m^{2/3}$ yields,

$$\left|\frac{f^{(k)}(\xi)}{k!}\right| \le 15\frac{m^{2/3}}{m^{2(k+1)/3}} = \frac{15}{m^{2k/3}}. \tag{4.97}$$

■

The following observation will be used in the proof of Lemma 4.7.3.

Observation 4.7.4 *Expanding f_m into k-order Taylor series centered at 0, and* *using (4.2) and (4.9), we obtain*

$$\begin{aligned}\overline{\gamma}(m+1,x) &= \int_{-m}^{x-m} e^{-s^2/2m} f_m(s)ds. \\ &= \int_{-m}^{x-m} e^{-s^2/2m}(1 + f_m'(0)s + ... + \frac{f_m^{(k)}(0)}{k!}s^k + R_{k+1})ds\end{aligned} \tag{4.98}$$

where $R_k(s)$ is the Taylor remainder term,

$$R_k(s) = \frac{f_m^{(k)}(\xi)s^k}{k!} \tag{4.99}$$

for some $\xi \in (0,x)$. The function $\overline{\gamma}(m+1,x)$ above is defined in (4.2), f_m is defined in (4.8), and $f_m^{(k)}$ is the k^{th} derivative of f_m.

The following lemma will be used in the proof of Theorem 4.7.4 to bound the error of asymptotic expansion (4.105).

Lemma 4.7.3 *For all $m > 100$ and $x \in (-m^{2/3}, m^{2/3})$,*

$$\left|\int_{-m^{2/3}}^{x-m} e^{-s^2/2m} R_k(s)ds\right| \le 15\frac{\Gamma(\frac{k+1}{2})2^{(k+1)/2}}{m^{k/6-1/2}}, \tag{4.100}$$

where R_k is defined in (4.99) and $\Gamma(k)$ is defined in (4.6).

Proof. Using (4.99) and applying elementary integral transformations to (4.100),

$$\begin{aligned}\left|\int_{-m^{2/3}}^{x-m} e^{-s^2/2m} R_k(s)ds\right| &< \int_{-m^{2/3}}^{x-m} \left|e^{-s^2/2m} R_k(s)\right| ds \\ &= \int_{-m^{2/3}}^{x-m} e^{-s^2/2m} \left|\frac{f^{(k)}(s)s^k}{k!}\right| ds.\end{aligned} \tag{4.101}$$

It follows immediately from applying Lemma 4.7.2 to (4.101) that

$$\begin{aligned}\left|\int_{-m^{2/3}}^{x-m} e^{-s^2/2m} R_k(s)ds\right| &< \int_{-m^{2/3}}^{x-m} e^{-s^2/2m}\left|\frac{f^{(k)}(s)s^k}{k!}\right| ds \\ &< \frac{15}{m^{2k/3}}\int_{-m^{2/3}}^{x-m} e^{-s^2/2m}|s^k|ds \\ &\leq \frac{15}{m^{2k/3}}\int_{-\infty}^{\infty} e^{-s^2/2m}|s^k|ds.\end{aligned} \tag{4.102}$$

Combining formulas 7.4.4 and 7.4.5 in [1], we obtain the identity,

$$\int_{-\infty}^{\infty} e^{-s^2/2m}|s^k|ds = \Gamma\left(\frac{k+1}{2}\right)(2m)^{(k+1)/2}, \tag{4.103}$$

and combining (4.102) and (4.103) yields,

$$\begin{aligned}\left|\int_{-m^{2/3}}^{x-m} e^{-s^2/2m} R_k(s)ds\right| &< \frac{15}{m^{2k/3}}\Gamma\left(\frac{k+1}{2}\right)(2m)^{(k+1)/2} \\ &= 15\frac{\Gamma(\frac{k+1}{2})2^{(k+1)/2}}{m^{k/6-1/2}}.\end{aligned} \tag{4.104}$$

■

The following theorem provides an asymptotic expansion for the evaluation of $P(m+1,x)$ where $P(m+1,x)$ is defined in (4.5).

Theorem 4.7.4 *For all $m > 100$ and $x \in (m - m^{2/3}, m + m^{2/3})$,*

$$P(m+1,x) \sim \frac{m^m e^{-m}}{\Gamma(m+1)}\sum_{i=0}^{\infty}\frac{f_m^{(i)}(0)}{i!}\int_{-m^{2/3}}^{x-m} e^{-s^2/2m}s^i ds, \tag{4.105}$$

where $f_m^{(k)}$ is the k^{th} derivative of f_m (see (4.8)) and $P(m+1,x)$ is defined in (4.5).

Furthermore, for all $k \in \mathbb{N}$,

$$\left| P(m+1,x) - \frac{m^m e^{-m}}{\Gamma(m+1)} \sum_{i=0}^{k-1} \left(\frac{f_m^{(i)}(0)}{i!} \int_{-m^{2/3}}^{x-m} e^{-s^2/2m} s^i ds \right) \right| < \\ 15\frac{\Gamma(\frac{k+1}{2})2^{(k+1)/2}}{m^{k/6}} + m^{-1/6} e^{\frac{-m^{1/3}}{2}+1}, \tag{4.106}$$

where $\Gamma(k)$ is defined in (4.6).

Proof. Clearly, using (4.2) and Observation 4.7.4,

$$\begin{aligned} \overline{\gamma}(m+1,x) &= \int_{-m}^{-m^{2/3}} e^{\phi(s)} ds + \int_{-m^{2/3}}^{x-m} e^{\phi(s)} ds \\ &= \int_{-m}^{-m^{2/3}} e^{\phi(s)} ds + \int_{-m^{2/3}}^{x-m} e^{-s^2/2m} (1 + f'_m(0)s + ... \\ &\qquad + \frac{f_m^{(k-1)}(0)}{(k-1)!} s^{k-1} + R_k(s)) ds, \end{aligned} \tag{4.107}$$

where R_k is the Taylor remainder term (4.99). Now, rearranging the terms of (4.107),

$$\begin{aligned} \overline{\gamma}(m+1,x) - \sum_{i=0}^{k-1} \left(\frac{f_m^{(i)}(0)}{i!} \int_{-m^{2/3}}^{x-m} e^{-s^2/2m} s^i ds \right) \\ = \int_{-m}^{-m^{2/3}} e^{\phi(s)} ds + \int_{-m^{2/3}}^{x-m} e^{-s^2/2m} R_k(s) ds. \end{aligned} \tag{4.108}$$

Using (4.5) and scaling both sides of (4.108) by

$$\frac{m^m e^{-m}}{\Gamma(m+1)} \tag{4.109}$$

we obtain,

$$
\begin{aligned}
P(m+1,x)-&\frac{m^m e^{-m}}{\Gamma(m+1)}\sum_{i=0}^{k-1}\left(\frac{f_m^{(i)}(0)}{i!}\int_{-m^{2/3}}^{x-m} e^{-s^2/2m}s^i ds\right) \\
&= \frac{m^m e^{-m}}{\Gamma(m+1)}\int_{-m}^{-m^{2/3}} e^{\phi(s)}ds + \frac{m^m e^{-m}}{\Gamma(m+1)}\int_{-m^{2/3}}^{x-m} e^{-s^2/2m}R_k(s)ds.
\end{aligned}
\tag{4.110}
$$

Combining Remark 4.4.1, (4.4), and (4.5), we obtain

$$
\frac{m^m e^{-m}}{\Gamma(m+1)}\int_{-m}^{-m^{2/3}} e^{\phi(s)}ds = P(m+1, m-m^{2/3}) < m^{-1/6}e^{\frac{-m^{1/3}}{2}+1}. \tag{4.111}
$$

Furthermore, according to Lemma 4.7.3 and Observation 4.2.1,

$$
\frac{m^m e^{-m}}{\Gamma(m+1)}\left|\int_{-m^{2/3}}^{x-m} e^{-s^2/2m}R_k(s)ds\right| < \frac{15}{m^{2k/3+1/2}}\Gamma\left(\frac{k+1}{2}\right)(2m)^{(k+1)/2}. \tag{4.112}
$$

It follows immediately from applying the triangle inequality and combining (4.110), (4.111), and (4.112) that

$$
\begin{aligned}
&\left|P(m+1,x) - \frac{m^m e^{-m}}{\Gamma(m+1)}\sum_{i=0}^{k-1}\left(\frac{f_m^{(i)}(0)}{i!}\int_{-m^{2/3}}^{x-m} e^{-s^2/2m}s^i ds\right)\right| \leq \\
&\qquad 15\frac{\Gamma(\frac{k+1}{2})2^{(k+1)/2}}{m^{k/6}} + m^{-1/6}e^{\frac{-m^{1/3}}{2}+1}.
\end{aligned}
\tag{4.113}
$$

■

4.8 Description of Algorithm

Suppose we wish to evaluate $P(m+1,x)$ for some $m > -1$ and $x > 0$. We consider the following two regimes.

4.8.1 Regime 1: $-1 < m \leq 10,000$

Numerical experiments show that in this regime, evaluation of $P(m+1, x)$ using formula (4.83) is faster than evaluation by asymptotic expansion (4.105). Hence, in this regime, we evaluate $P(m+1, x)$ directly using formula (4.83). A bound on the error of approximation (4.83) is provided in Lemma 4.6.2 and Table 4.1 includes values of the bound for different m, x, k.

When $x > m/2$, in order to evaluate sum (4.83) and without a loss of accuracy, we compute recursively the terms ω_n of (4.83) defined by the formula,

$$\omega_{m+i} = \frac{x^{m+i}e^{-x}}{\Gamma(m+i+1)} \tag{4.114}$$

by observing that

$$\omega_{k+1} = \frac{x^k e^{-x}}{\Gamma(k+1)} = \frac{x}{k}\omega_k. \tag{4.115}$$

and evaluating the initial recursive step ω_m by observing that

$$\begin{aligned} \omega_m &= \frac{x^m e^{-x}}{\Gamma(m+1)} \\ &= \exp\left(m\log(x) - x - \log(\Gamma(m+1))\right). \end{aligned} \tag{4.116}$$

We then use (4.13) to evaluate $\log(\Gamma(m+1))$.

4.8.2 Regime 2: $m > 10,000$

We first check if $x < m+1$. If so, we use Lemma 4.6.1 to determine whether $P(m+1, x)$ is sufficiently small that it is well-approximated by 0 to some user-specified accuracy. If $x > m+1$, we check if $P(m+1, x)$ is well-approximated by 1 via Corollary 4.5.4.

If $P(m+1,x)$ is neither well-approximated by 0 nor by 1, further analysis remains to show under what conditions algorithm (4.105) is computationally less expensive than (4.83). However, numerical experiments show that for most x arising in practice, evaluation of $P(m+1,x)$ by asymptotic expansion (4.105) is significantly faster than evaluation by (4.83). Hence, we evaluate $P(m+1,x)$ by asymptotic expansion (4.105). In the remainder of this section, we provide a detailed explanation of asymptotic expansion (4.105).

Remark 4.8.1 *Numerical experiments show that in this regime, evaluation of $P(m+1,x)$ via asymptotic expansion (4.105) achieves full double precision accuracy when using an expansion of* 29 *terms. That is, setting $k = 28$ in (4.106).*

Precomputation

Asymptotic expansion (4.105) includes the factors,

$$\frac{f_m^{(k)}(0)}{k!} \tag{4.117}$$

where $k \in \{0,1,2,...\}$ and $f_m^{(k)}$ is the k^{th} derivative of f_m (see (4.8)). Straightforward differentiation shows that for all k, the values $f_m^{(k)}(0)$ are defined by the formula,

$$\frac{f_m^{(k)}(0)}{k!} = \sum_{i=j_k}^{n_k} \frac{a_{k,i}}{k!m^i}, \tag{4.118}$$

for some $j_k, n_k \in \mathbb{N}$ and some $a_{k,i} \in \mathbb{R}$. The values,

$$\left\{\frac{a_{k,i}}{k!}\right\}, \tag{4.119}$$

for $i \in \{j_k, ..., n_k\}$, are computed in Mathematica and stored in Fortran DATA statements.

Evaluation

The inputs to this stage of the algorithm are $m > 10,000$ and $x > 0$.

Step 1. Given some requirement on the precision of the approximation, we use (4.106) to determine the number of terms in the expansion. For the remainder of this section, we assume that we require an expansion of K terms.

Step 2. For all $k \in \{1, ..., K\}$, compute the powers from (4.118),

$$\frac{1}{m^i}, \tag{4.120}$$

and store them for all $i \in \{j_k, ..., n_k\}$, where n_k is defined in (4.118).

Step 3. Evaluate the factors in (4.105) defined by (4.118). Specifically, for each $k \leq K$, evaluate

$$\frac{f_m^{(k)}(0)}{k!} = \sum_{i=j_k}^{n_k} \frac{a_{k,i}}{k! m^i}, \tag{4.121}$$

where $a_{k,i}$ and m^i are defined in (4.118). Note that we have already computed the quotients

$$\left\{ \frac{a_{k,i}}{k!} \right\}, \tag{4.122}$$

for $i \in \{j_k, ..., n_k\}$, in the precomputation stage, while the necessary powers of $1/m$ were computed in Step 1.

Step 4. Use Lemma 4.3.3 to evaluate the integrals of (4.105) of the form,

$$\int_{-m^{2/3}}^{x-m} e^{-s^2/2m} s^{2n+1} dt, \tag{4.123}$$

where n is a non-negative integer, and $x > 0$ and $m > 10,000$ are real numbers.

Step 5. Use Lemma 4.3.4 to evaluate the integrals of (4.105) of the form,

$$\int_{-m^{2/3}}^{x-m} e^{-s^2/2m} s^{2n} dt, \tag{4.124}$$

where n is a non-negative integer and $x > 0$ and $m > 10,000$ are real numbers.

4.9 Numerical Experiments

The algorithm of this chapter was implemented in Fortran 77. We used the Lahey/Fujitsu compiler on a 2.9 GHz Intel i7-3520M Lenovo laptop; all examples in this section were run in double precision arithmetic.

Throughout this section, we report numerical results relating to the evaluation of $P(m+1,x)$ via asymptotic expansion (4.105) and via summation (4.83) for various values of m and x. In each table in this section, the column labeled "$m+1$" denotes the value of $m+1$ in $P(m+1,x)$. The column labeled "x" denotes the value of x in $P(m+1,x)$. The column labeled "k" denotes the number of terms of expansion (4.105) used to approximate $P(m+1,x)$. The column labeled "time (μs)" denotes the time, in microseconds, required to run each evaluation. The column labeled "relative error" denotes the relative error of the approximation. The column labeled "absolute error" denotes the absolute error of the approximation. The column labeled $P(m+1,x)$ denotes the true value that is being approximated. This value was computed in extended precision using adaptive Gaussian quadrature.

In Table 4.4, the column labeled $\alpha_k(m+1,x)$ denotes $\log_{10}$ of the magnitude of the k^{th} term of asymptotic expansion (4.105). Specifically, $\alpha_k(m+1,x)$ is defined via the formula

$$\alpha_k(m+1,x) = \log_{10}\left|\frac{m^m e^{-m}}{\Gamma(m+1)}\frac{f_m^{(k)}(0)}{k!}\int_{-m^{2/3}}^{x-m} e^{-s^2/2m}s^k ds\right|. \tag{4.125}$$

In Table 4.4, the column labeled $\sigma_k(m+1,x)$ denotes the relative error of the k-term approximation (4.105). Specifically, $\sigma_k(m+1,x)$ is defined via the formula

$$\begin{aligned}&\sigma_k(m+1,x) = \\ &P(m+1,x)^{-1}\left|P(m+1,x) - \frac{m^m e^{-m}}{\Gamma(m+1)}\sum_{i=0}^{k}\left(\frac{f_m^{(i)}(0)}{i!}\int_{-m^{2/3}}^{x-m} e^{-s^2/2m}s^i ds\right)\right|.\end{aligned} \tag{4.126}$$

In Table 4.7, the column labeled "evaluator" indicates whether $P(m+1,x)$ was evaluated via sum (4.83) or asymptotic expansion (4.105).

The primary purpose of Table 4.2 and Figure 4.2 is to demonstrate that for fixed m and fixed k, evaluation of $P(m+1,x)$ via k-term asymptotic expansion (4.105) results in a smaller error for larger x.

Table 4.3 and Figure 4.3 report the numerical costs of evaluation of $P(m+1,x)$ via k-term asymptotic expansion (4.105) for different k. We report runtimes for different k with $m+1 = x = 10^7$.

The primary purpose of Table 4.4 and Figure 4.4 is to report the decrease in the magnitude of the k^{th} term of asymptotic expansion (4.105) along with the corresponding error of k-term expansion (4.105). In Table 4.4, we report these numerical results for the case $m + 1 = x = 10^4$. In Figure 4.4, we plot $\log_{10}$ of the magnitude of the k^{th} term of expansion (4.105) for $k \leq 28$. We do this for the cases $m + 1 = x = 10^4$, $m + 1 = x = 10^7$, and $m + 1 = x = 10^{10}$.

In Table 4.5 and Figure 4.5 we report the numerical costs of evaluation of $P(m+1, x)$ via summation (4.83) for different m. We report runtimes for different m with $x = m + 1$. In Figure 4.5, the horizontal line corresponds to the runtime required for evaluation of $P(m+1, x)$ via asymptotic expansion (4.105) with $k = 28$.

Table 4.6 and Figure 4.6 report the numerical costs of evaluation of summation (4.83) for fixed m and different x. We report runtimes for $m + 1 = 1000$ with various x.

Table 4.7 demonstrates that both sum (4.83) and asymptotic expansion (4.105) achieve nearly full extended precision accuracy when evaluating $P(m + 1, x)$. We demonstrate this for various values of m and x.

Observation 4.9.1 *Figure 4.5 demonstrates that for all real numbers $x > 0$ and $-1 < m < 10^3$, evaluation of $P(m + 1, x)$ via asymptotic expansion (4.105) is computationally more expensive than evaluation of $P(m + 1, x)$ via sum (4.83).*

$m+1$	$x\ (\cdot 10^3)$	k	relative error	$P(m+1,x)$
10^6	996	10	0.21472×10^{-7}	0.0000310071182110
10^6	997	10	0.15395×10^{-8}	0.0013381041673135
10^6	998	10	0.11619×10^{-9}	0.0226961140067368
10^6	999	10	0.16768×10^{-10}	0.1586552135743036
10^6	1000	10	0.53194×10^{-11}	0.5001329807608725
10^6	1001	10	0.31623×10^{-11}	0.8413447863683402
10^6	1002	10	0.27477×10^{-11}	0.9771959041012301
10^6	1003	10	0.32794×10^{-11}	0.9986382593537824
10^6	1004	10	0.46816×10^{-11}	0.9999676545526865

Table 4.2: Relative errors for the evaluation of $P(m+1,x)$ via 10-term asymptotic expansion (4.105) for different x

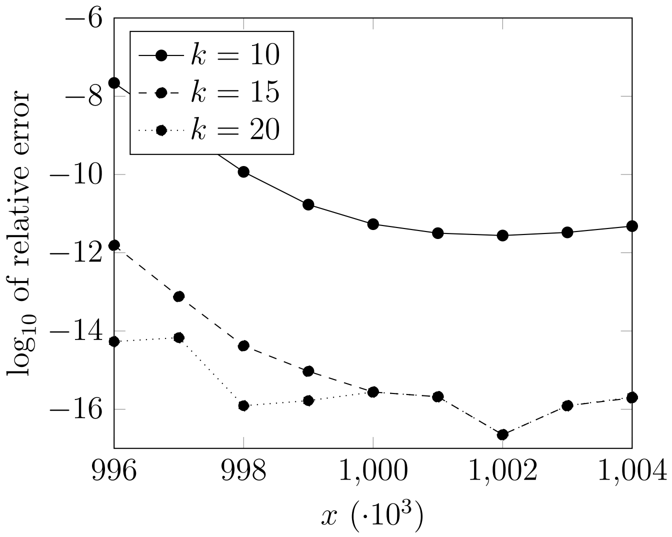

Figure 4.2: $\log_{10}$ of relative errors for evaluation of $P(m+2,x)$ via k-term asymptotic expansion (4.105) for different x and for $k = 10$, $k = 15$, and $k = 20$

$m+1$	x	k	time (μs)	relative error	$P(m+1,x)$
10^7	10^7	4	1.68	0.83327×10^{-7}	0.500042052208723698
10^7	10^7	8	2.08	0.59881×10^{-10}	0.500042052208723698
10^7	10^7	12	2.50	0.40330×10^{-15}	0.500042052208723698
10^7	10^7	16	2.92	0.60328×10^{-15}	0.500042052208723698
10^7	10^7	20	3.44	0.60328×10^{-15}	0.500042052208723698
10^7	10^7	24	3.75	0.60328×10^{-15}	0.500042052208723698
10^7	10^7	28	4.26	0.60328×10^{-15}	0.500042052208723698

Table 4.3: CPU times for the evaluation of $P(m+1,x)$ via k-term asymptotic expansion (4.105) for different k

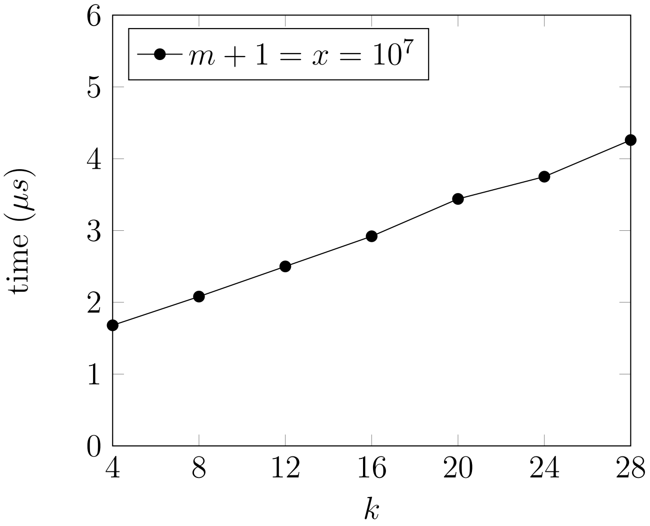

Figure 4.3: CPU times for evaluation of $P(m+1,x)$ via k-term asymptotic expansion (4.105) for different k and $m+1=x=10^7$

$m+1$	x	k	$\alpha_k(m+1,x)$	$\sigma_k(m+1,x)$	$P(m+1,x)$
10^4	10^4	0	-0.2976	0.52970×10^{-2}	0.501329808339955200
10^4	10^4	4	-4.4259	0.83144×10^{-4}	0.501329808339955200
10^4	10^4	6	-4.3803	0.13220×10^{-5}	0.501329808339955200
10^4	10^4	8	-7.2890	0.19636×10^{-5}	0.501329808339955200
10^4	10^4	10	-7.1832	0.50570×10^{-7}	0.501329808339955200
10^4	10^4	12	-7.5752	0.41020×10^{-8}	0.501329808339955200
10^4	10^4	14	-9.6979	0.19454×10^{-8}	0.501329808339955200
10^4	10^4	16	-9.8870	0.50321×10^{-10}	0.501329808339955200
10^4	10^4	18	-10.496	0.10907×10^{-10}	0.501329808339955200
10^4	10^4	20	-12.127	0.31368×10^{-11}	0.501329808339955200
10^4	10^4	22	-12.492	0.48601×10^{-13}	0.501329808339955200
10^4	10^4	24	-13.264	0.33580×10^{-13}	0.501329808339955200
10^4	10^4	26	-14.519	0.68569×10^{-14}	0.501329808339955200
10^4	10^4	28	-15.016	0.52361×10^{-16}	0.501329808339955200

Table 4.4: Numerical results for $\log_{10}$ of the magnitude of the k^{th} term of asymptotic expansion (4.105) and errors of the k-term expansion

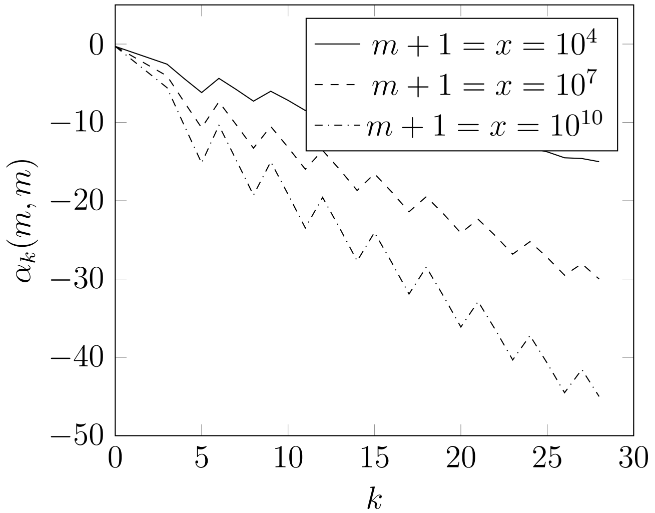

Figure 4.4: $\log_{10}$ of the magnitude of the k^{th} term of asymptotic expansion (4.105) for different m

$m+1$	x	time (μs)	relative error	$P(m+1,x)$
10^0	10^0	0.62	0.34164×10^{-16}	0.632120558828557678
10^1	10^1	0.96	0.16914×10^{-15}	0.542070285528147791
10^2	10^2	1.34	0.71081×10^{-15}	0.513298798279148664
10^3	10^3	3.00	0.41353×10^{-15}	0.504205244180215508
10^4	10^4	7.80	0.39818×10^{-15}	0.501329808339955200
10^5	10^5	20.88	0.45496×10^{-14}	0.500420522110365176
10^6	10^6	62.32	0.95799×10^{-14}	0.500132980760872591

Table 4.5: CPU times and errors for the evaluation of $P(m,m)$ by direct summation (4.83) for different m

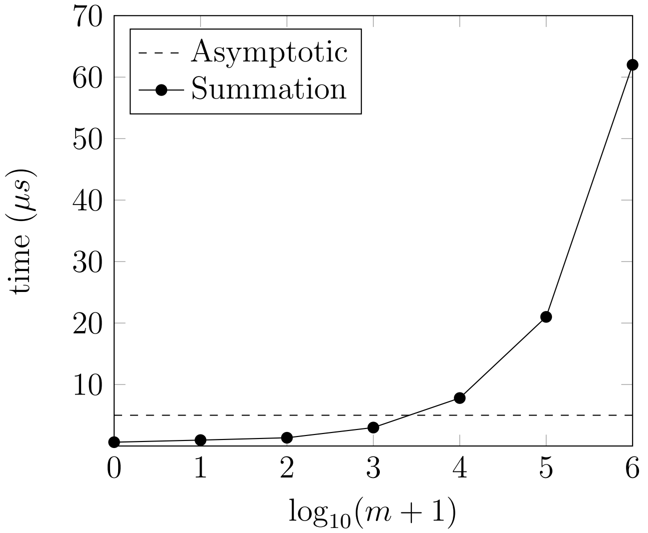

Figure 4.5: CPU times for evaluation of $P(m+1,x)$ by direct summation (4.83) and by asymptotic expansion (4.105)

$m+1$	x	time (μs)	relative error	$P(m+1,x)$
1000	900	1.93	0.50370×10^{-15}	0.000549902265711782
1000	925	2.16	0.93887×10^{-15}	0.007693713246846007
1000	950	2.28	0.99996×10^{-16}	0.055054686230738034
1000	975	2.49	0.27405×10^{-14}	0.215731105240819891
1000	1000	2.65	0.41353×10^{-15}	0.504205244180215508
1000	1025	2.90	0.62391×10^{-15}	0.786575483861807090
1000	1050	3.06	0.34303×10^{-15}	0.941328888622681922
1000	1075	3.35	0.13468×10^{-14}	0.989973597928674133
1000	1100	3.55	0.15243×10^{-14}	0.998940676746070022

Table 4.6: CPU times for the evaluation of $P(m+1,x)$ by direct summation (4.83) for different x

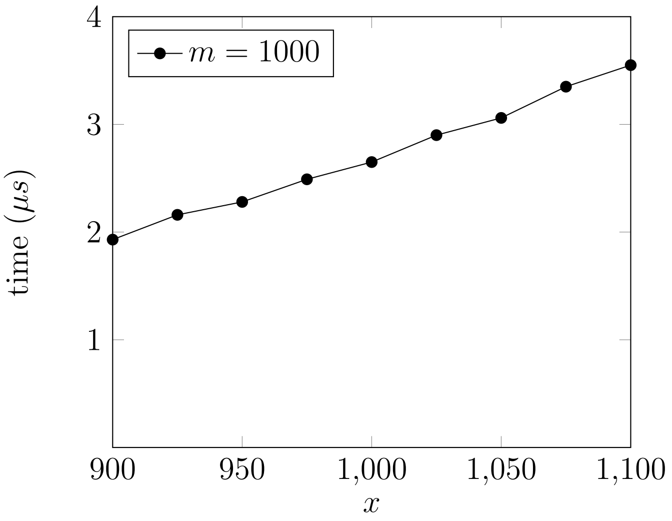

Figure 4.6: CPU times of evaluation of $P(m+1,x)$ via direct summation (4.83) for different x

$m+1$	x	evaluator	absolute error	$P(m+1,x)$
1	0.5	Sum (4.83)	0.48148×10^{-34}	0.393469340287366576396200465009
1	1	Sum (4.83)	0.10000×10^{-34}	0.632120558828557678404476229839
1	10	Sum (4.83)	0.13482×10^{-32}	0.999954600070237515148464408484
100	80	Sum (4.83)	0.10803×10^{-32}	0.017108313035133114165877307636
100	100	Sum (4.83)	0.33415×10^{-31}	0.513298798279148664857314256564
100	120	Sum (4.83)	0.80504×10^{-31}	0.972136260109479338515814832144
$10,000$	$9,000$	Sum (4.83)	0.13501×10^{-34}	0.000000000000000000000000207329
$10,000$	$10,000$	Sum (4.83)	0.49111×10^{-32}	0.501329808339955200382742251300
$10,000$	$11,000$	Sum (4.83)	0.19356×10^{-31}	0.999999999999999999999830714685
10^5	$10^5 - 10^3$	Sum (4.83)	0.31597×10^{-34}	0.000757419921174767974118465304
10^5	10^5	Sum (4.83)	0.80889×10^{-32}	0.500420522110365176693312579044
10^5	$10^5 + 10^3$	Sum (4.83)	0.45221×10^{-30}	0.999191578487074409267531226544
10^6	$10^6 - 10^3$	Sum (4.83)	0.43733×10^{-30}	0.158655213574303652463032743495
10^6	10^6	Sum (4.83)	0.51519×10^{-31}	0.500132980760872591244322817503
10^6	$10^6 + 10^3$	Sum (4.83)	0.87534×10^{-31}	0.841344786368340291627563851466
10^7	$10^7 - 10^3$	Exp. (4.105)	0.11700×10^{-30}	0.375950818831443160416162761546
10^7	10^7	Exp. (4.105)	0.24941×10^{-31}	0.500042052208723698333756164783
10^7	$10^7 + 10^3$	Exp. (4.105)	0.63748×10^{-31}	0.624121183505552339531809964939

Table 4.7: Absolute errors for the evaluation of $P(m+1,x)$ by direct summation (4.83) and asymptotic expansion (4.105) in extended precision

Bibliography

[1] Abramowitz, Milton, and Irene A. Stegun, eds. *Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables.* Washington: U.S. Govt. Print. Off., 1964.

[2] Alpert, Bradley K. and Vladimir Rokhlin. "A fast algorithm for the evaluation of Legendre expansions." *SIAM J. Sci. Stat. Comput.* 12.1 (1991): 158–179.

[3] Bateman, Harry. *Higher Transcendental Functions.* Vol. 2. Ed. Arthur Erdélyi. New York: McGraw-Hill, 1953.

[4] Ben-yu, Guo, and Wang Li-lian. "Jacobi interpolation approximations and their applications to singular differential equations." *Adv. Comput. Math.* 14 (2001): 227–276.

[5] Born, Max, and Emil Wolf. *Principles of Optics.* 6th ed. (with corrections). New York: Pergamon Press Inc., 1980.

[6] Chong, Chee-Way, P. Raveendran, and R. Mukundan. "A comparative analysis of algorithms for fast computation of Zernike moments." Pattern Recognition 36 (2003): 731–742.

[7] DiDonato, Armido and Alfred H. Morris. "Computation of the Incomplete Gamma Function Ratios and their Inverse." *ACM TOMS* 12.4 (1986): 377–393.

[8] Feller, William. *An Introduction to Probability and its Applications.* Volume 1. 3rd Ed. New York: Wiley, 1968.

[9] Gautschi, Walter. "A Computational Procedure for Incomplete Gamma Functions." *ACM TOMS* 5.4 (1979): 466-481.

[10] Gil, Amparo, Javier Seura, and Nico M. Temme. "Efficient and Accurate Algorithms for the Computation and Inversion of the Incomplete." *SIAM Journal on Scientific Computing* 34.6 (2013): A2965–A2981.

[11] Glaser, Andreas, Xiangtao Liu, and Vladimir Rokhlin. "A Fast Algorithm for the Calculation of the Roots of Special Functions."*SIAM J. Sci. Comput.* 29.4 (2007): 1420–1438.

[12] Gradshteyn, I. S., and I. M. Ryzhik. *Table of Integrals, Series, and Products.* Eds. Alan Jeffrey and Daniel Zwillinger. San Diego: Academic Press, 2000.

[13] Greengard, Philip and Kirill Serkh. "Zernike Polynomials: Evaluation, Quadrature, and Interpolation." *Yale Computer Science Technical Reports* (02/2018).

[14] Greengard, Philip and Vladimir Rokhlin. "An algorithm for the evaluation of the incomplete gamma function." *Adv Comput Math* (2018).

[15] Jagerman, Louis S. *Ophthalmologists, meet Zernike and Fourier!*. Victoria, BC, Canada: Trafford Publishing, 2007.

[16] Kintner, Eric C. "On the mathematical properties of the Zernike polynomials." *Optica Acta* 23.8 (1976): 679–680.

[17] Li, Shengqiao. "Concise Formulas for the Area and Volume of a Hyperspherical Cap." *Asian J. Math. Stat.* 4.1 (2011): 66–70.

[18] Nemes, G. “Error bounds and exponential improvement for Hermite’s asymptotic expansion for the gamma function.”*Appl. Anal. Discrete Math.* 7.1 (2013): 161–179.

[19] Nemes, G. “An explicit formula for the Coefficients in Laplace’s Method.” *Constructive Approximation* 38.3 (2013):471–487.

[20] *Ophthalmics - Corneal Topography Systems - Standard Terminology, Requirements.* ANSI Standard Z80.23-2008 (R2013).

[21] Osipov, Andrei. “Evaluation of small elements of the eigenvectors of certain symmetric tridiagonal matrices with high relative accuracy.” *Appl. Comput. Harmon. Anal.* 43 (2017): 173-211.

[22] Osipov, Andrei, Vladimir Rokhlin and Hong Xiao. *Prolate Spheroidal Wave Functions of Order Zero.* New York: Springer US, 2013.

[23] Radiant ZEMAX LLC. *ZEMAX: Optical Design Program User’s Manual.* Redmond, WA: Author, 2011.

[24] Shkolnisky, Yoel. “Prolate Spheroidal Wave Functions on a disc- Integration and approximation of two-dimensional bandlimited functions.” *Appl. Comput. Harmon. Anal.* 22 (2007): 235-256.

[25] Slepian, David and H. O. Pollak. “Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty - I.” *Bell Labs Technical Journal* 40.1 (1961): 43-63.

[26] Slepian, David. “Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty - IV: Extensions to Many Dimensions; Generalized Prolate Spheroidal Functions.” *Bell Labs Technical Journal* 43.6 (1964): 3009-3057.

[27] Stoer, Josef and Roland Bulirsch. *Introduction to Numerical Analysis*, 2nd ed., Springer-Verlag, 1992.

[28] Temme, N.M. “The Asymptotic Expansion of the Incomplete Gamma Functions.” *SIAM J. Math. Anal.* 10.4 (1979): 757–766.

[29] Temme, N.M. “On the computation of the incomplete gamma functions for large values of the parameters.” *Algorithms for Approximation.* New York: Clarendon Press (1987): 479–489.

[30] Wyant, James C., and Katherine Creath. “Basic wavefront aberration theory for optical metrology.” Applied optics and optical engineering II (1992): 28–39.

[31] Xiao, H., V. Rokhlin, N. Yarvin. “Prolate spheroidal wave functions, quadrature and interpolation.” *Inverse Problems* 17 (2001): 805-838.