2-7-2011

Orthogonality and convergence of discrete Zernike polynomials

Joseph Allen

Joseph Allen
Candidate

Mathematics
Department

This thesis is approved, and it is acceptable in quality and form for publication:

Approved by the Thesis Committee:

Peter T Ewbank, Chairperson

M Cristina Pereyra

Hugh Denham

Orthogonality and Convergence of Discrete Zernike Polynomials

by

Joseph Herbert Allen

B.S. Mathematics, NM Institute of Mining and Technology, 2003
B.S. Physics, NM Institute of Mining and Technology, 2003

THESIS

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Master of Science
Mathematics

The University of New Mexico

Albuquerque, New Mexico

December 2010

Dedication

To Rebekah, Evelyn, and Aurelia:
my inspiration.

Acknowledgments

To my advisor, Dr. Embid, and my family: Rebekah, Nicole, Donald and Agatha, Matthew and Deborah. I am endlessly grateful.

To everyone who made this possible, thank you.

Orthogonality and Convergence of Discrete Zernike Polynomials

by

Joseph Herbert Allen

ABSTRACT OF THESIS

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Master of Science
Mathematics

The University of New Mexico

Albuquerque, New Mexico

December 2010

Orthogonality and Convergence of Discrete Zernike Polynomials

by

Joseph Herbert Allen

B.S. Mathematics, NM Institute of Mining and Technology, 2003
B.S. Physics, NM Institute of Mining and Technology, 2003

M.S., Mathematics, University of New Mexico, 2010

Abstract

The Zernike polynomials are an infinite set of orthogonal polynomials over the unit disk, which are rotationally invariant. They are frequently utilized in optics, opthalmology, and image recognition, among many other applications, to describe spherical aberrations and image features. Discretizing the continuous polynomials, however, introduces errors that corrupt the orthogonality. Minimizing these errors requires numerical considerations which have not been addressed. This work examines the orthonormal polynomials visually with the Gram matrix and computationally with the rank and condition number. The convergence of the Fourier-Zernike coefficients and the Fourier-Zernike series are also examined using various measures of error. The orthogonality and convergence are studied over six grid types and resolutions, polynomial truncation order, and function smoothness. The analysis concludes with design criteria for computing an accurate analysis with the discrete Zernike polynomials.

Contents

List of Figures

List of Tables

Glossary

DZP	Discrete Zernike polynomial
FZC	Fourier-Zernike coefficients
FZS	Fourier-Zernike series
GLPF	Gaussian, low-pass filter
PSNR	Peak, signal-to-noise ratio
QR	Orthogonal triangular decomposition
RMS	Root-mean-squared
SVD	Singular value decomposition
ZP	Zernike polynomial
$V_n^{\pm m}$	Traditional Zernike polynomial, normalized amplitude at $\rho = 1$ and $\theta = 0$
$\Lambda_n^{\pm m}$	Modified Zernike polynomial, normalized for an orthonormal basis

Chapter 1

Introduction

1.1 Overview

The preferred tool for advanced geometric analysis of optical aberrations to describe complex distributions in a circular domain are the Zernike polynomials (ZP). The ZP are a complete, orthogonal basis which are rotationally invariant (Fig. 1.1 illustrates a few). They also describe optical aberrations observed in physical systems in a convenient and concise manner [34]. Application of the ZP to digital images, however, requires sophisticated techniques to preserve their accuracy, orthogonality, and convergence. This work critically examines the orthogonality of the discrete Zernike polynomials (DZP) and the convergence of the Fourier-Zernike series (FZS) and Fourier-Zernike coefficients (FZC).

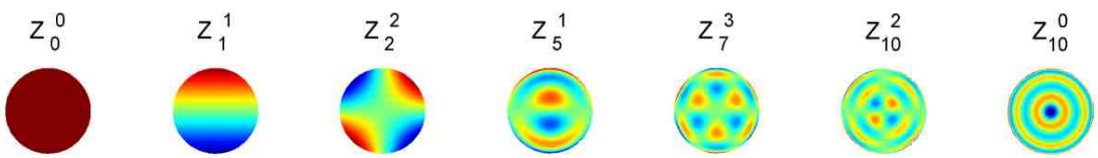


Figure 1.1: Seven examples of the Zernike polynomials.

1.2 Problem Description

The continuous ZP are a complete, orthogonal polynomial basis over the unit disk which is rotationally invariant [4]. The DZP, however, are not. Straightforward applications typically compute the ZP without considering how the grid type, grid resolution, series truncation, or the smoothness affect the accuracy.

The DZP sampled over a discrete grid does not preserve the orthogonality and most applications do not realize the potential errors. There are currently no guidelines or criteria for choosing a grid type or the number of grid points. Furthermore, there are few studies on how the grid affects the accuracy of the solutions [35]. A feasible number of polynomials available for an analysis is also limited by the computational resources, and certainly cannot be infinite. Most studies utilize low-order approximations by convention, not by accuracy considerations [2, 3, 11, 14, 35]. An additional source of error that appears in a finite series approximation is the Gibbs-Wilbraham phenomenon. This artifact appears when the original function has discontinuities or insufficient smoothness for uniform convergence of the partial sum. The characteristic ringing has been observed in the image reconstruction from the ZP, but not addressed [35].

1.3 Solution Approach

The orthogonality of the DZP and convergence of the FZS and FZC require numerical examination. The orthogonality may be observed and quantified with tools from linear algebra: Gram matrix, rank, and condition number. Since the convergence depends on many factors, the errors require numerical exploration as a function of each parameter.

The Gram matrix is a valuable tool to visually inspect which elements in a basis overlap, or have a non-zero inner product. The Gram matrix computes the inner product between every element in a set [19]. It not only shows overlap between the elements, but it also shows which elements have unit length. In addition to the qualitative review, the rank and condition number deliver critical, quantitative measures of the orthogonality and solution stability. The rank of a system measures the maximal number of linearly independent elements, while the condition number estimates the amount of accuracy which may be lost when computing a solution with the system [19, 26, 8].

In this analysis, the DZP are computed over six grid types at various resolutions. The orthogonality is examined using the Gram matrix for a few combinations. The rank and condition number are then computed for large array of grid combinations and examined for conditions which produce DZP systems of full-rank and well-conditioned. Several measures of error and basic image normalization are introduced. The convergence of the solutions for the FZC and FZS are then examined over the various DZP systems. Smoothing an image prior to computing the FZC or FZS reduces the Gibbs-Wilbraham phenomenon, and is examined as a function of the smoothing amount. The orthogonality analysis and convergence observations should yield valuable guidelines for selecting parameters to perform an analysis with the DZP.

1.4 Thesis Outline

This work critically examines the orthogonality and convergence of the DZP over various grids and resolutions, polynomial truncation order, and function smoothness using a smoothing filter.

Ch. 2 reviews the continuous Zernike polynomials. It presents previous applications and a summary of the historical development. It also introduces the definitions, properties, and illustrations of the ZP. It begins with the classical derivation by Born and Wolf and presents an alternative formulation using normalization of the orthogonality constant. It then elaborates on the orthogonalization process that Zernike originally employed and suggests that it may be used numerically to re-enforce the orthogonality of the discrete system. Using the ZP as a basis for the Fourier-Zernike series expansion of a function in terms of the Fourier-Zernike coefficients is also discussed.

Ch. 3 presents the discrete Zernike polynomials. It examines the various grids and techniques utilized in the calculation of the DZP. Various grids are introduced including some from the literature and new ones for comparison. The Fourier-Zernike coefficients are computed simultaneously from a matrix equation, not from individual calculations of each coefficient. The orthogonality of the system is then examined using the Gram matrix, rank, and condition number. Ch. 3 also introduces several measures of error. The convergence of the Fourier-Zernike coefficients and series is then examined as a function of various parameters.

Ch. 4 concludes this work, summarizing the techniques and criteria to observe for an accurate analysis using the Zernike polynomials.

Chapter 2

Continuous Zernike Polynomials

2.1 Overview

The ZP were originally derived by Frederick (Frits) Zernike in 1930's during his development of phase-contrast microscopy [36]. They were later examined further, derived rigorously and shown to be a unique set of polynomials which is orthogonal over the unit disk, contains only polynomials that are invariant to rotations about the origin, and contains only one polynomial for each pair of permissible values of the order and repetition pair (n, m) [4]. Unlike previous tools outlined by Seidel, they offered the ability to describe separate and higher-order spherical aberrations. While the low-order ZP are similar to the aberrations described by Seidel, they are not equivalent, and require different considerations upon their applications [34]. The ZP are typically utilized in applications requiring a descriptive function over a circular aperture, particularly in microscopy, ophthalmology, and image processing [2, 14, 9]. They have also been utilized to describe spherical aberrations, wavefront analysis, atmospheric turbulence, engineering defects, corneal topology, and image discrimination [7, 12, 22, 11, 27, 1].

The ZP possess the necessary properties to build a foundation for image analysis and reconstruction and additional properties for practical applications. The ZP are an infinite set of 2-dimensional polynomials that form a complete, orthogonal basis over the unit disk [5]. This ensures the existence of a unique description in terms of ZP of any function defined over the unit disk. The ZP are rotationally invariant and possesses many other derived invariants [25, 9]. Furthermore, they behave well under rotation, preserving the magnitude of the modes and only changing their phase. The selection of the ZP over any other basis is primarily due to their physical interpretation of wavefront aberrations. The ZP describe the same aberrations (e.g., astigmatism, coma, trefoil, etc.) observed in optical systems to provide a convenient and concise description [34]. The ZP are also a suitable basis for pattern recognition in images with arbitrary rotation.

Sec. 2.2 outlines the derivation of the ZP and how the rotational invariance, orthogonality, and normalization define this unique set of polynomials. Sec. 2.3 presents the generalized formulation of the ZP with both the traditional normalization and the modified normalization constants. Then Sec. 2.4 discusses the use of the ZP as a basis for expanding a function in terms of the Fourier-Zernike series and reconstructing the function from the representation in Fourier-Zernike coefficients. Sec. 2.5 concludes the discussion of the continuous ZP.

2.2 Derivation

A summary of the important results and properties is presented here, while a complete derivation is available in *Principles of Optics*, appx. VII, [5]. Infinitely many complete sets of polynomials exist which satisfy the following orthogonality condition over the closed unit disk, $\Omega = \overline{\mathbb{D}}$.

$$\iint_{\Omega} V_{\beta}^{*}(x,y) V_{\alpha}(x,y) d\Omega = A_{\alpha\beta}\delta_{\alpha\beta}, \tag{2.1}$$

where V_α and V_β are functions from the set, δ is the Kronecker delta, $A_{\alpha\beta}$ is the normalization constant, and * denotes the complex conjugate. The first major property that distinguishes the ZP from other sets is the rotational invariance, i.e. rotations about the origin by some angle as given below.

$$\begin{aligned} \tilde{x} &= +x\cos\theta + y\sin\theta, \\ \tilde{y} &= -x\sin\theta + y\cos\theta. \end{aligned} \tag{2.2}$$

Since each polynomial $V(x,y)$ is transformed into a polynomial of the same form, we obtain the following relationship:

$$V(\tilde{x},\tilde{y}) = G(\theta)V(x,y). \tag{2.3}$$

From this result, explicit properties on the continuous function, G, follow:

$$G(0) = 1, \tag{2.4a}$$

$$G(\theta + 2\pi) = G(\theta), \tag{2.4b}$$

$$G(\theta_1 + \theta_2) = G(\theta_1)G(\theta_2), \tag{2.4c}$$

The function which uniquely satisfies those requirements is the natural, exponential function for any integer, ℓ.

$$G(\theta) = e^{i\ell\theta}. \tag{2.5}$$

It has been shown that $V(x, y)$ is rotational invariant about the origin if and only if it has the following form [4]:

$$V(\rho\cos\theta, \rho\sin\theta) = R(\rho)e^{i\ell\theta}. \tag{2.6}$$

It can be seen that $R(\rho) = V(\rho, 0)$ is only a function of ρ. Expanding in powers of $\cos\theta$ and $\sin\theta$, it follows that V is a polynomial in x and y of degree n if and only if R is a polynomial in ρ of degree n when n is greater than or equal to $|\ell|$ and n is even or odd according to ℓ being even or odd. The second major property that distinguishes the ZP is that they contain a single polynomial for each pair of permissible values of the degree, n, and the angular dependence, ℓ (i.e. integers n and ℓ such that $n \geq 0, n \geq |\ell|$ and $n - |\ell|$ is even).

The ZP have the following separable form for the radial and azimuthal components,

$$V_n^\ell(\rho,\theta) = R_n^\ell(\rho)e^{i\ell\theta}. \tag{2.7}$$

Using the orthogonality property from Eq. 2.1 and substituting in the form of the ZP, Eq. 2.7, the orthogonality condition on the radial function R follows,

$$\int_0^1 R_{n_1}^\ell(\rho)R_{n_2}^\ell(\rho)\rho d\rho = \frac{A_{n_1n_2}^\ell}{2\pi}\delta_{n_1n_2}, \tag{2.8}$$

with the weighted inner product,

$$\langle f(\rho), g(\rho)\rangle = \int_0^1 f(\rho)g(\rho)\rho d\rho. \tag{2.9}$$

The radial polynomial is obtained by taking the restrictions on n and ℓ and orthogonalizing the standard polynomial basis with the weighting factor ρ. An appropriate normalization constant is traditionally chosen such that $R_n^{-\ell} = R_n^\ell$. To obtain the following, real ZP, an additional variable, m, is used to separate the complex exponential, where $m = |\ell|$ is a strictly, non-negative integer.

$$\begin{aligned} V_n^{+m}(\rho,\theta) &= \cos(m\theta)R_n^m(\rho), \\ V_n^{-m}(\rho,\theta) &= \sin(m\theta)R_n^m(\rho). \end{aligned} \tag{2.10}$$

n	$m=0$	$m=1$	$m=2$
0	1		
1		ρ	
2	$2\rho^2-1$		ρ^2
3		$3\rho^3-2\rho$	
4	$6\rho^4-6\rho^2+1$		$4\rho^4-3\rho^2$
5		$10\rho^5-12\rho^3+3\rho$	
6	$20\rho^6-30\rho^4+12\rho^2-1$		$15\rho^6-20\rho^4+6\rho^2$
7		$35\rho^7-60\rho^5+30\rho^3-4\rho$	
8	$70\rho^8-140\rho^6+90\rho^4-20\rho^2+1$		$56\rho^8-105\rho^6+60\rho^4-10\rho^2$
9		$126\rho^9-280\rho^7+210\rho^5-60\rho^3+5\rho$	

Table 2.1: Orthogonalizing the standard polynomial basis $\{\rho^0, \rho^1, \rho^2, ...\}$ with the weighted inner product yields the radial ZP, $R_n^m(\rho)$. They increase order in powers of 2 due to the weighted inner product.

The first few radial ZP are shown explicitly in Tab. 2.1 for $n \leq 9, m \leq 2$. Since the number of ZP grow in a triangular fashion, they contain $K = \frac{1}{2}(n_{\max}+1)(n_{\max}+2)$ linearly independent polynomials in x^p, y^q, and $x^r y^s$ for $0 \leq p, q, r+s \leq n_{\max}$, where $n_{\max}$ is the maximum degree. Therefore every polynomial in x, y and xy may be expressed as a linear combination of a finite number of ZP, $V_n^{\pm m}$, and the set is complete by the Weierstrass approximation theorem [5]. Furthermore, the ZP are the unique set which is orthogonal over $\overline{\mathbb{D}}$, contains only polynomials that are invariant to rotations about the origin, and contains only one polynomial for each pair of permissible values of the order and repetition pair (n, m) [4].

This section presented the required form of the ZP for rotational invariance, the orthogonality condition on the radial polynomials, the orthogonalization procedure to generate the radial polynomials, and the unique enumeration of the polynomials by their order, n, and repetition number, m. The next section presents formulas and illustrations of the ZP.

2.3 Formulation

The normalization factor has traditionally been obtained by normalizing the polynomial such that $V_n^{\pm m}(1,0) = 1$, as Zernike performed. Also note that the symbol V appears frequently in literature discussing the traditional Zernike polynomials. Utilizing properties of the hyper-geometric polynomials, particularly their orthogonality and normalization, the generalized formulation is given by [5]:

$$\begin{aligned} V_n^{+m}(\rho,\theta) &= \cos(m\theta)\sum_{s=0}^{\frac{n-m}{2}} r_n^m(s)\rho^{n-2s}, \\ V_n^{-m}(\rho,\theta) &= \sin(m\theta)\sum_{s=0}^{\frac{n-m}{2}} r_n^m(s)\rho^{n-2s}, \\ r_n^m(s) &= \frac{(-1)^s(n-s)!}{s!\left(\frac{n+m}{2}-s\right)!\left(\frac{n-m}{2}-s\right)!}. \end{aligned} \tag{2.11}$$

The orthogonality condition is

$$\int_0^{2\pi}\int_0^1 \left(V_{n_2}^{\pm m_2}(\rho,\theta)\right)^* \left(V_{n_1}^{\pm m_1}(\rho,\theta)\right)\rho d\rho d\theta = \frac{\pi}{n+1}\delta_{m_1 m_2}\delta_{n_1 n_2}. \tag{2.12}$$

The ZP are shown in Figs. 2.1 and 2.2 up to the 10^{th} degree. Note that in these figures $Z_n^{\pm m} = V_n^{\pm m}$. The order, n, increases along the diagonal from the top-left to the bottom-right. The positive repetition numbers, m, are above the diagonal while the negative values are below. The subsets of polynomials with identical repetition numbers are grouped along the diagonals. The ZP increase in the radial oscillations for increasing orders (compare Z_0^0 through Z_{10}^0) while they increase in the angular oscillation for increasing repetition numbers (compare Z_0^0 through Z_{10}^{10}). The polynomials with negative repetition numbers are identical to the polynomials with the positive repetition number rotated by $\frac{\pi}{2m}$ radians.

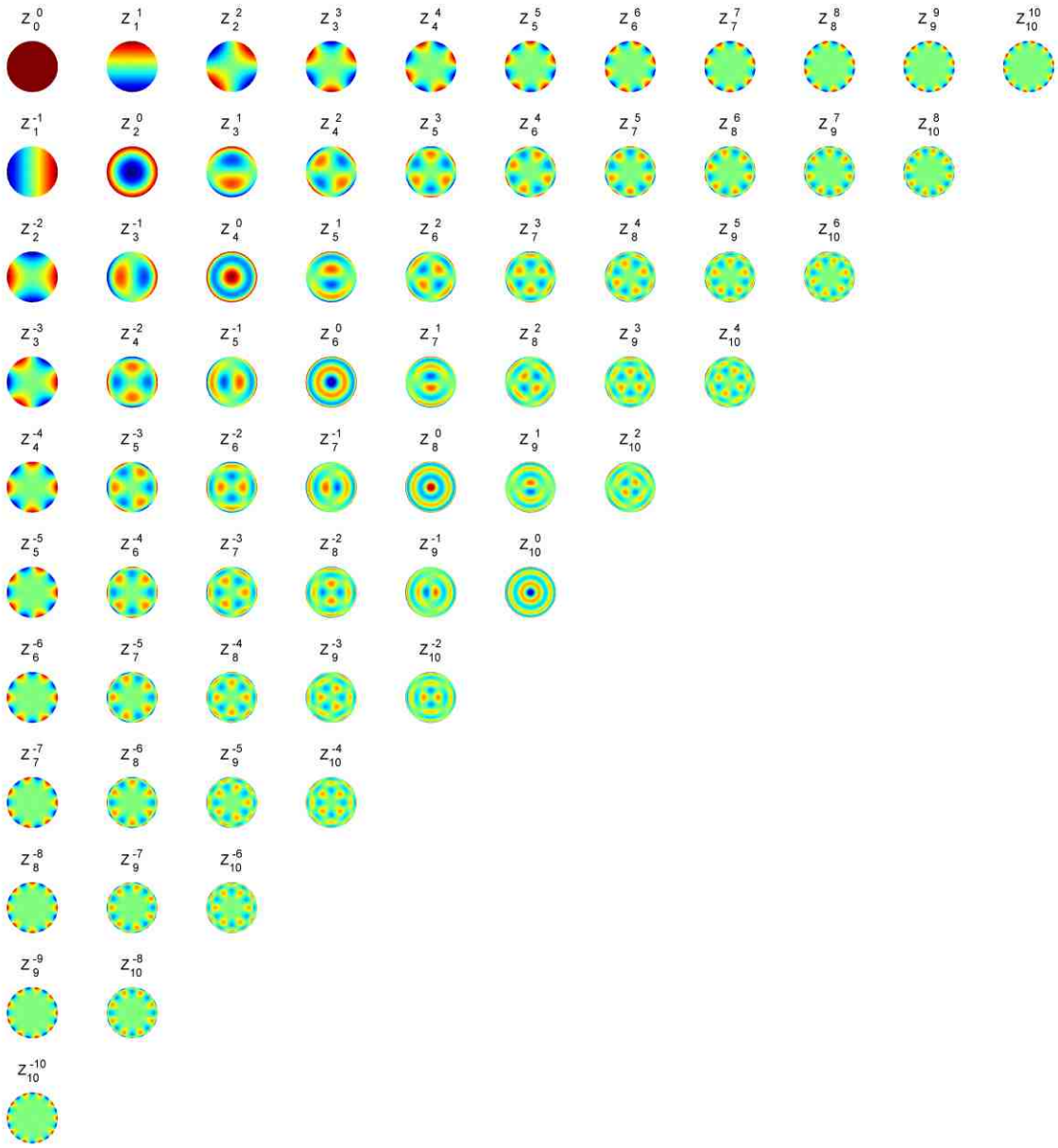


Figure 2.1: The continuous Zernike polynomials plotted in 2 dimensions up to the 10^{th} degree. The radial oscillations increase for increasing orders (subscript) while the angular oscillations increase for increasing repetition numbers (superscript). In this pseudo-color scale, green represents zero, red represents positive values, and blue represents negative values.

Figure 2.2: The continuous Zernike polynomials plotted in 3 dimensions up to the 10^{th} degree.

For this analysis with the ZP, an alternative normalization is utilized to simplify the orthogonality relationship. This is achieved by absorbing the normalization factor into the definition of the polynomials. These modified polynomials are represented by the symbol Λ, which looks like an upside-down V, to emphasize their subtle difference from the traditional Zernike polynomials. The orthonormal Zernike polynomials are then,

$$\begin{aligned}
\Lambda_n^{+m}(\rho,\theta) &= \sqrt{\frac{n+1}{\pi}}\cos(m\theta)\sum_{s=0}^{\frac{n-m}{2}} r_n^m(s)\rho^{n-2s}, \\
\Lambda_n^{-m}(\rho,\theta) &= \sqrt{\frac{n+1}{\pi}}\sin(m\theta)\sum_{s=0}^{\frac{n-m}{2}} r_n^m(s)\rho^{n-2s}, \\
r_n^m(s) &= \frac{(-1)^s(n-s)!}{s!\left(\frac{n+m}{2}-s\right)!\left(\frac{n-m}{2}-s\right)!}.
\end{aligned} \tag{2.13}$$

The orthogonality condition becomes,

$$\int_0^{2\pi}\int_0^1 \left(\Lambda_{n_2}^{\pm m_2}(\rho,\theta)\right)^* \left(\Lambda_{n_1}^{\pm m_1}(\rho,\theta)\right)\rho d\rho d\theta = \delta_{m_1 m_2}\delta_{n_1 n_2}. \tag{2.14}$$

This section presented the traditional ZP, normalized for unity amplitude, and the modified ZP, normalized to produce an orthonormal set. The next section utilizes the ZP as a basis to expand a function in a series.

2.4 Fourier-Zernike Series and Coefficients

Since the ZP are a complete and orthogonal set of unique polynomials, we can describe a function, f, using the ZP as a basis. This decomposition will be referred to as the Fourier-Zernike series (FZS). The FZS is composed of the addition of the ZP, Λ_n^m, and the magnitude of each ZP, referred to as the Fourier-Zernike coefficients (FZC), λ_n^m. The symbol to represent the coefficients, λ, was selected to parallel the symbol selected for the modified polynomials. Each coefficient is the projection of the function onto each ZP given by the following equation:

$$\lambda_n^m = \int_0^{2\pi} \int_0^1 \Lambda_n^m(\rho,\theta) f(\rho,\theta) \rho d\rho d\theta. \tag{2.15}$$

The FZS is assembled by adding the contribution from all polynomials, summing over both indices,

$$f(\rho,\theta) = \sum_{n=0}^{\infty} \sum_{\substack{m=-n \\ m-n \text{ is even}}}^{n} \lambda_n^m \Lambda_n^m(\rho,\theta). \tag{2.16}$$

The next section summarizes the properties and formulation of the ZP presented in this chapter.

2.5 Summary

This chapter introduced the Zernike polynomials. The rotation invariance required the polynomials to be of the form, $R(\rho)e^{im\theta}$. To find the radial polynomials, R_n^m, we then utilized the weighted, inner product (Eq. 2.9), to orthogonalize the standard polynomial basis, $\{\rho^0, \rho^1, ...\}$. After computing the explicit formulation of a few polynomials (Tab. 2.1), the generalized Zernike polynomials were obtained for any order, n, and repetition number, m, and rendered them up to the 10^{th} order (Fig. 2.1). The traditional formulation of the ZP were then modified the to obtain the orthonormal, ZP, Λ_n^m (Eq. 2.13). Finally, the Fourier-Zernike series (Eq. 2.16) and coefficients (Eq. 2.15) were presented which serve as the theoretical foundation to develop the discrete tools in the next chapter.

Chapter 3

Discrete Zernike Polynomials

3.1 Overview

Application of the ZP to image analysis requires computation on modern computers with floating point arithmetic. This requires discretization of both the domain and the polynomials. Furthermore, since it is not possible to compute an infinite number of terms in the FZS, we can only compute the ZP up to some maximum order, $n_{\max}$.

All approaches to compute the discrete Zernike polynomials (DZP) necessarily introduce numerical errors due to grid selection, quantization, and round-off errors. The goal is to examine which techniques minimize the errors to maximize the orthogonality and convergence. This chapter explores the techniques necessary to compute the representation of an image in terms of the FZC, and how to reconstruct the image from the FZS.

Sec. 3.2 introduces various grids that partition the domain into subdomains. The grid algorithms are discussed including the generation of grids with similar numbers of subdomains. Sec. 3.3 discusses the formulation of the FZC as a matrix equation. In this formulation, tools from linear algebra are employed to analyze the system. Sec. 3.4 examines the orthogonality of the system evaluated over each grid using the Gram matrix. This provides a visual inspection tool of the linear independence between the ZP as a basis. From there, the condition number and rank of the systems are computed in Sec. 3.5. This provides a quantitative measure to compare the orthogonality. The performance of the various grids is compared in an exploratory study of the dependence of the orthogonality on the grid type and resolution.

Sec. 3.6 presents several formulas to quantify the error (e.g., root-mean squared, peak signal-to-noise ratio, etc.) between two images on the same grid. These error formulas will be utilized extensively in the remainder of the analysis. Sec. 3.7 discusses basic techniques to resample digital images in the standard Cartesian grid onto the various other grids. Moreover, it examines the errors introduced by resampling.

Sec. 3.8 then examines the convergence of FZC. Rotational invariance is also briefly examined. Sec. 3.9 utilizes the FZC to construct the FZS and reconstruct the original image. The errors between the reconstructed image and the original image are analyzed for the dependence on various parameters.

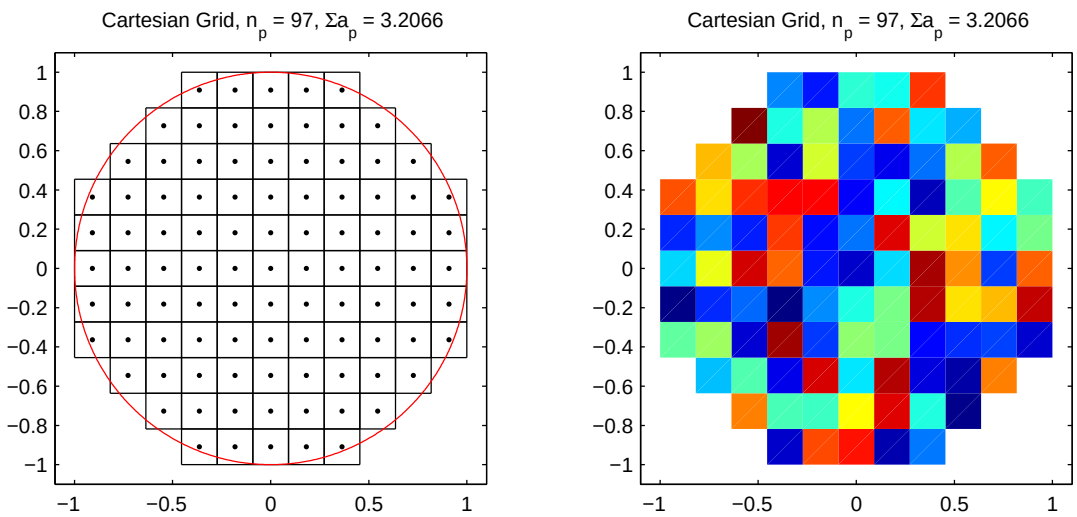
3.2 Grid Partitioning

A grid is a collection of vertices connected in a simple pattern that partitions the entire domain, or picture, into a finite set of subdomains, or picture elements (pixels). For this analysis, six tessellation techniques were selected to offer a comparison of which technique is best suited for computation of the ZP. The grid types are: (1) Cartesian, (2) simple polar (polar1), (3) improved polar (polar2), (4) advanced polar (polar3), (5) triangular, and (6) random. The Cartesian grid (including one converted to polar coordinates) is commonly utilized in literature in performing calculations on data from an image [1, 2, 3, 11, 25, 27, 28]. It has been widely employed in many applications, despite the incomplete partitioning of the unit disk and offers questionable performance [35]. The improved polar grid has also been utilized in literature, but facilitates more accurate calculations in the polar coordinate system, showing a notable increase in computational stability [9, 35]. The simple polar and advanced polar grids are introduced here for calculations also in the polar coordinate system, but have not been observed in any applications. The triangular grid is used commonly in finite element analysis, but is introduced to this application for exploration. The random grid is generally not utilized in any applications, but is utilized here as a baseline; to characterize the performance of the designed grids over a totally arbitrary selection of points.

An ideal grid covers the entire domain with no overlap or gaps: $\Omega = \bigcup_{j=1}^{J} \Omega_j$, where $\Omega_{j_1} \bigcap \Omega_{j_2} = \emptyset$ when, $j_1 \neq j_2$. The perimeter of a real grid, however, may not entirely cover the unit disk and pixels can even extend outside of it. The three basic types of pixels are a triangle, square, or sector. A collection of all the pixel faces over a grid for rendering is referred to as a patch. Patches are only composed of polygons, typically triangles, so curved boundaries appear as straight lines between the vertices. To facilitate relevant comparisons between the various grids, it is necessary to employ the same number of pixels regardless of the partitioning technique. Due to necessary

variations in the grid generating techniques, however, the actual number of pixels in each grid, n_p, varies slightly from the requested number of pixels, n_{req}, but typically $< 5\%$.

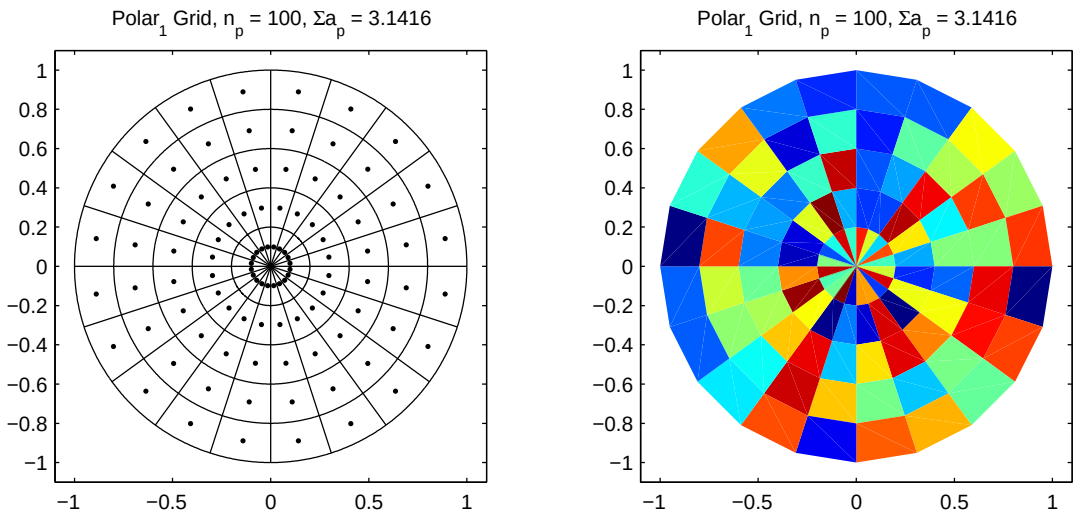
The Cartesian grid, shown in Fig. 3.1(a), is an adaption from standard image formats which utilizes equally sized, square pixels over a square domain. The pixels with center-points outside of the unit disk are discarded and only the ones inside will be used for computation. The total area is the number of pixels, n_p, times the area of a single pixel. The actual number of pixels utilized in the grid is $n_p \approx \frac{\pi}{4} n_{\text{req}}$, to the nearest integer, representing the ratio of the area of the disk to the area of a square. The Cartesian patch, shown in Fig. 3.1(b), is a color rendering of the Cartesian grid. The face of each square pixel is colored according to the value of that pixel.



(a) (b)

Figure 3.1: A Cartesian grid outlined in (a) and rendered in (b) is partitioned using square pixels over the unit disk. Only the pixels with center-points (black points) inside the unit circle (red circle) are retained for computation.

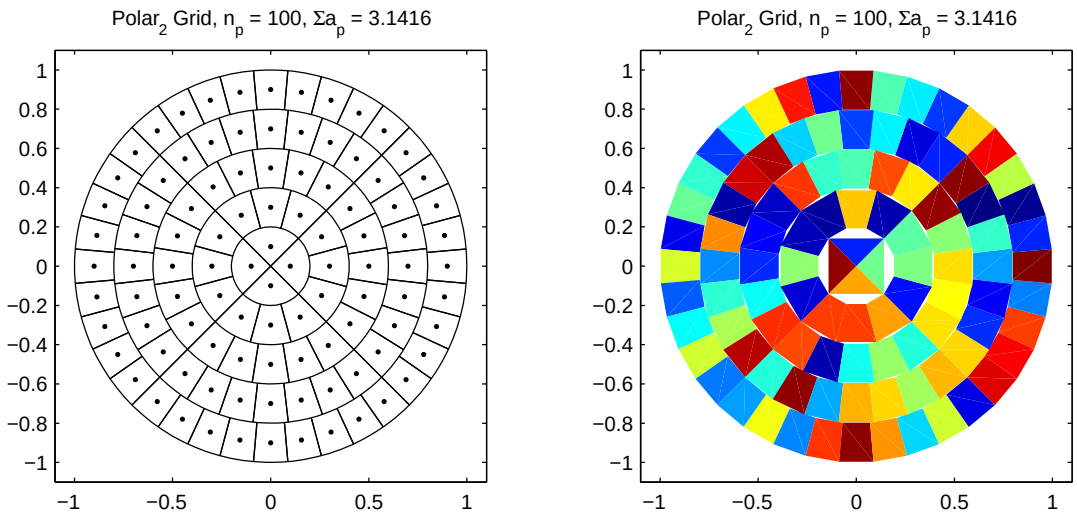
The simple polar grid, shown in Fig. 3.2(a), is generated by equally spacing the radial and angular coordinates. For simplicity, the number of angular divisions, n_θ, is set as a fixed multiple of the number of angular divisions, n_r, specifically 4. The actual number of pixels in the grid is then, $n_p = n_r n_\theta = 4n_r^2$. The number of radial divisions is approximately $n_r \approx \sqrt{\frac{n_{\text{req}}}{4}}$, to the nearest integer, where n_{req} is the requested number of points. The resulting pixels are sectors that cover the entire, closed unit disk with no overlaps or gaps. The area of each sector depends on the radius to the center of the j^{th} pixel, $[a_p]_j = \frac{2\pi}{n_r n_\theta}[r_p]_j$. The patch over the simple polar grid is shown in Fig. 3.2(b).



(a) (b)

Figure 3.2: A simple polar grid (a) partitioned from sectors of uniform radial and angular spacings. The center-point of each pixel (black dot) is spaced evenly between the divisions, and does not represent the center of mass for a given pixel. The color rendering in (b) uses triangles and cannot accurately represent the curved boundaries.

An improved polar grid [35] of pixels with equal area is shown in Fig. 3.3(a). Similar to the simple polar grid, the improved polar grid is a function of two variables: n_r specifies the number of radial divisions and n_θ for the angular divisions. The angular divisions in this case, however, increases as the radius increases, so n_θ only specifies the initial number of angular divisions at the center. The total number of pixels in this grid is $n_r^2 n_\theta$, so the area of each pixel is $\frac{\pi}{n_r^2 n_\theta}$. To reduce the number of parameters for creating this grid, the number of angular divisions is set to a constant, $n_\theta = 4$. The pixels in this grid are also sectors, similar to the simple polar grid, and cover the entire, closed unit disk with no overlap or gaps. The patch over the improved polar grid is shown in Fig. 3.3(b). The rendered patch appears to have gaps between the pixels (prominent in the center), but is merely an artifact of the triangles used to connect the vertices.



(a) (b)

Figure 3.3: An improved polar grid (a) partitioned from sectors of uniform radial and adaptive angular spacings. The center-point of each pixel (black dot) is spaced evenly between the divisions. The apparent gaps between the pixels in the color rendering (b) are not real, but artifacts from the triangles used to connect the vertices.

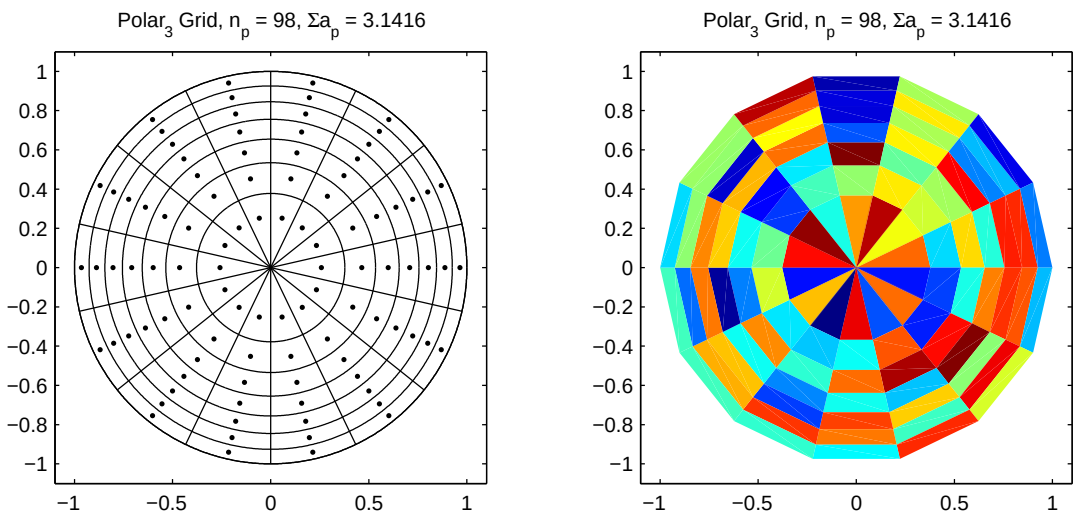
The advanced polar grid, shown in Fig. 3.4(a), is generated in the same fashion as the simple polar grid, but with varying spacing in the radial coordinate. The vertices were first generated as the square root of the radial coordinate, $[r_v]_j = \sqrt{r_j}$, where r_j are linearly spaced points between and including [0, 1]. The point inside each pixel, referred to as the central-point to distinguish it from the center-point (i.e. midpoint) and the centroid (i.e. center of mass), was calculated from the following formula:

$$\begin{aligned} r_p &= \frac{2}{3}\frac{r_{v_1}^2 + r_{v_1} r_{v_2} + r_{v_2}^2}{r_{v_1} + r_{v_2}}, \\ \theta_p &= \theta_{v_2} - \theta_{v_1} \end{aligned} \tag{3.1}$$

which improves the convergence of the following integral,

$$\varepsilon = \frac{1}{r_2 - r_1}\int_{r_1}^{r_2} \left(f(r) - f(r_p)\right) r dr, \tag{3.2}$$

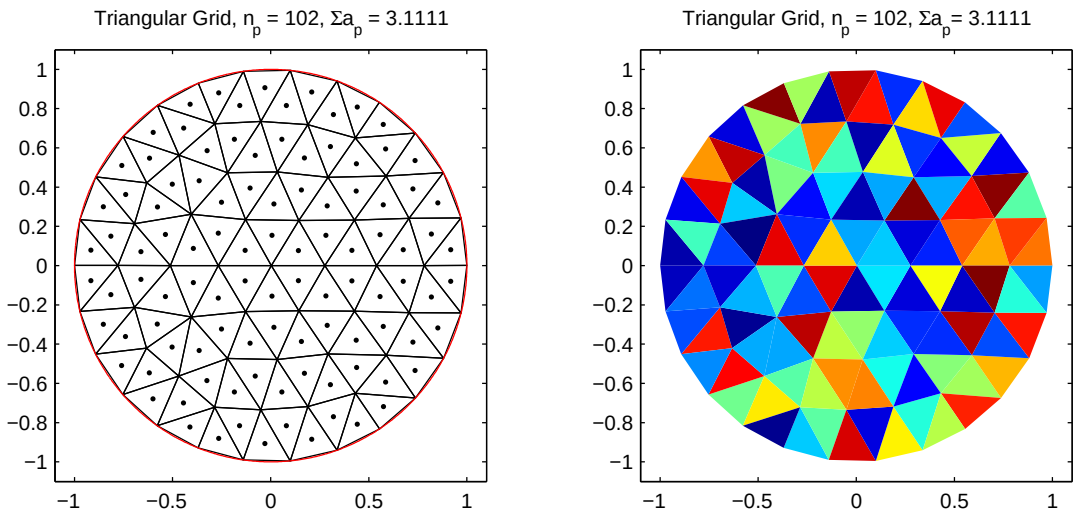
The number of angular divisions, n_θ, is set as a fixed multiple of the number of angular divisions, n_r, specifically 2. This is chosen to avoid oversampling in either coordinate. The actual number of pixels in the grid is then, $n_p = n_r n_\theta = 2n_r^2$. The number of radial divisions is approximately $n_r \approx \sqrt{\frac{n_{\text{req}}}{2}}$, to the nearest integer, where n_{req} is the requested number of points. The resulting pixels are sectors that cover the entire, closed unit disk with no overlaps or gaps. The area of each sector depends on the number of angular divisions and the radial boundaries, $a_p = \frac{\pi}{n_\theta}(r_{v_2}^2 - r_{v_1}^2)$. The patch over the advanced polar grid is shown in Fig. 3.4(b).



(a) (b)

Figure 3.4: The advanced polar grid (a) is a variation of the simple polar grid with finer spacing toward the perimeter. This feature was added from the observation that the radial ZP oscillate frequently there. The rendering of the grid (b) again shows how the pixels are represented with finer, radial resolution at the perimeter but have constant, angular resolution.

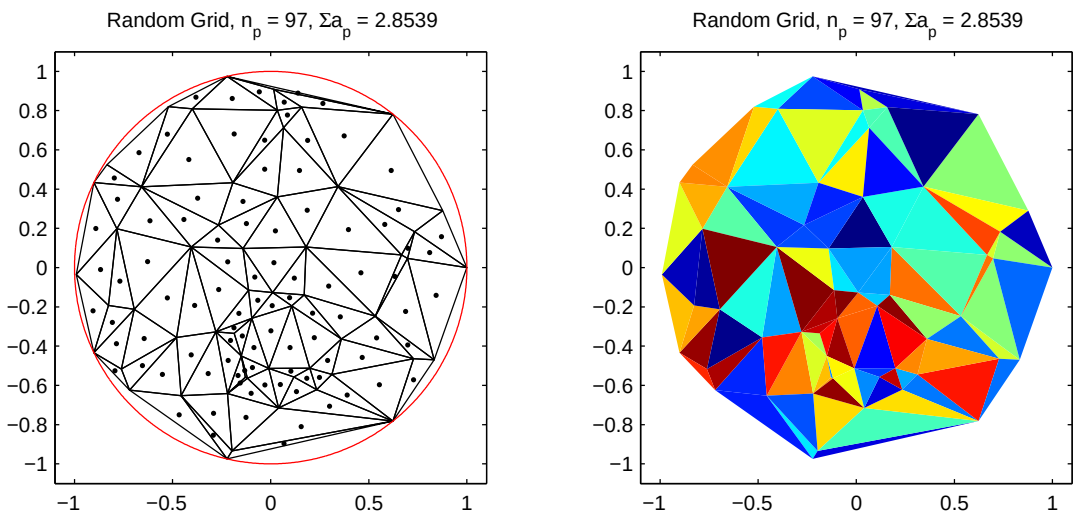
A triangular grid of pixels with approximately equal area is shown in Fig. 3.5(a). A grid generator [24] attempts to evenly space the vertices over the domain by a forcing-based smoothing procedure. The result produces near, equilateral triangles with similar area. Since the procedure optimizes the spacing between the vertices, the exact number of pixels in a grid only approximately equals the requested number of pixels. The empirical relationship between the desired vertex spacing, h, to the requested number of pixels was approximated by $h = 10^{\left(-0.02262x^2-0.34003x+0.14158\right)}$ where $x = \log_{10} n_{\text{req}}$. The points inside of these triangular pixels is calculated from the centroid (center-of-mass) of their vertices. Since this grid is composed of triangles, there is some gap between the boundary of the triangles and the unit circle. The patch rendering the colors over the triangular grid is shown in Fig. 3.5(b).



(a) (b)

Figure 3.5: A triangular grid (a) is generated by a forced smoothing of the distance between the vertices. After the distances between adjacent vertices reaches an optimum value, each center-point is calculated as the center of mass of the vertices. The rendering of the triangular grid (b) is composed of triangles with similar, but not equal, area.

A random grid is shown in Fig. 3.6(a). The location of the vertices are randomly generated in Cartesian coordinates using a uniform distribution. Any vertices outside of the unit circle are discarded. Additional vertices are then placed on the perimeter at equally spaced angles. Next, the vertices are connected to form triangles using the Delaunay tessellation algorithm. The point inside each pixel is calculated as the centroid of their vertices. The number of vertices left inside the unit circle after elimination is $n_{\text{in}} \approx \left(\frac{\pi}{4}\right)\left(\frac{2n_{\text{req}}}{\pi}\right)$. The number of vertices added to the perimeter is $n_{\text{per}} \approx \sqrt{n_{\text{in}}}$. After the Delaunay tessellation, the number of pixels, n_p, is roughly equal to the number of requested points, n_{req}, on average. The patch over the random grid is shown in Fig. 3.6(b). This technique frequently produces acute and near degenerate triangles. The perimeter of the tessellation also does not meet the unit circle, resulting in gaps.

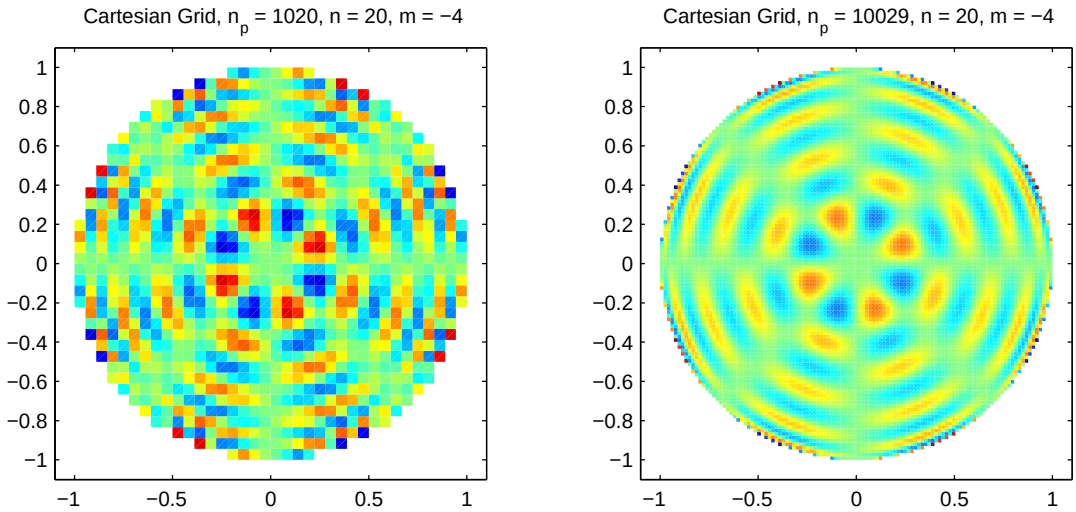


(a) (b)

Figure 3.6: A random grid (a) generated from a uniform distribution over Cartesian coordinates. Vertices outside the unit circle are discarded and additional vertices on the unit circle are added and connected using a Delaunay tessellation. The rendering of the grid (b) shows high variation in the size and orientation of the triangles. Near degenerate triangles occur often.

Using a single ZP for comparison, each of the previously described grids were used to sample the polynomial at approximately 1,000 and 10,000 pixels. The ZP Λ_{20}^{-4} is chosen for a relatively high degree of variation in both the radial and angular coordinates, which represents most of the remaining ZP of equal order. The low-resolution grids of 1,000 points are selected to demonstrate particularly how each grid loses information about the polynomial when under-sampling. The high-resolution grids of 10,000 points are chosen to demonstrate how well each grid can render the polynomial. The results are shown in Figs. 3.7 through 3.12. It is important to qualitatively examine the smoothness, symmetry, and convergence.

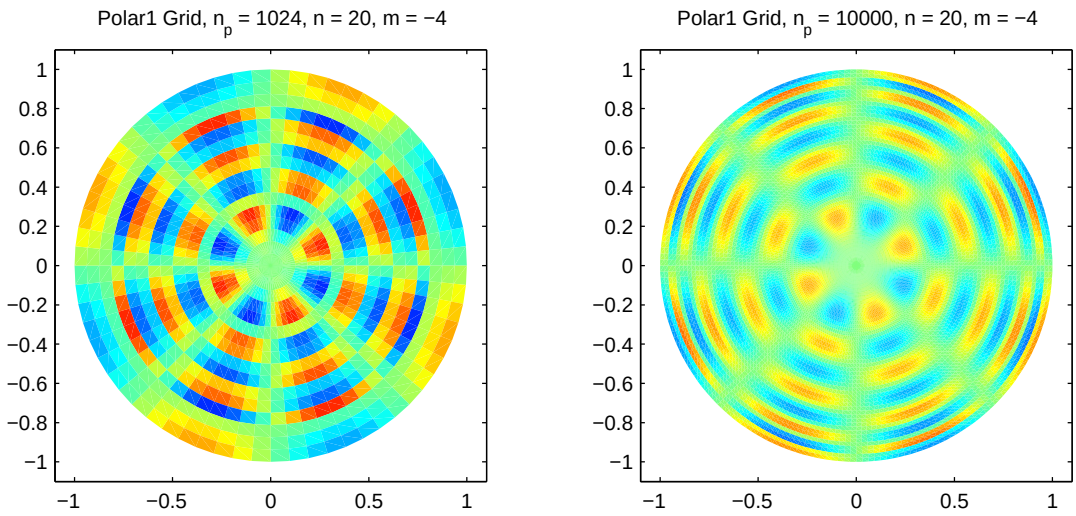
Since the ZP are smooth polynomials in the continuous domain, they should also show smooth transitions between values of adjacent pixels. This polynomial should also have symmetry with respect to rotation ($\theta' = \theta + \frac{j\pi}{4}, j = 1, 2, ..., 7$) and mirror reflections (e.g., about $\theta = \frac{j\pi}{4}, j = 0, 1, 2, 3$). Lastly, the sampled polynomial should converge uniformly over the domain for increasing grid resolutions. The following examination is not intended to be comprehensive, but rather a preemptive breakdown of the strengths and weaknesses that each grid offers. To facilitate a rough comparison, the properties of each grid will be judged by the following subjective descriptors: excellent, good, fair, poor. The comparison is summarized at the end of this section in Tab. 3.1.



(a) (b)

Figure 3.7: Λ_{20}^{-4} sampled over the Cartesian grid with (a) 1,020 points and (b) 10,029 points. The polynomial in (a) has poor smoothness or symmetry, whereas, in (b) it is better, but only on the interior. The convergence is fair since the values around the perimeter remain irregular even while the central region appears good.

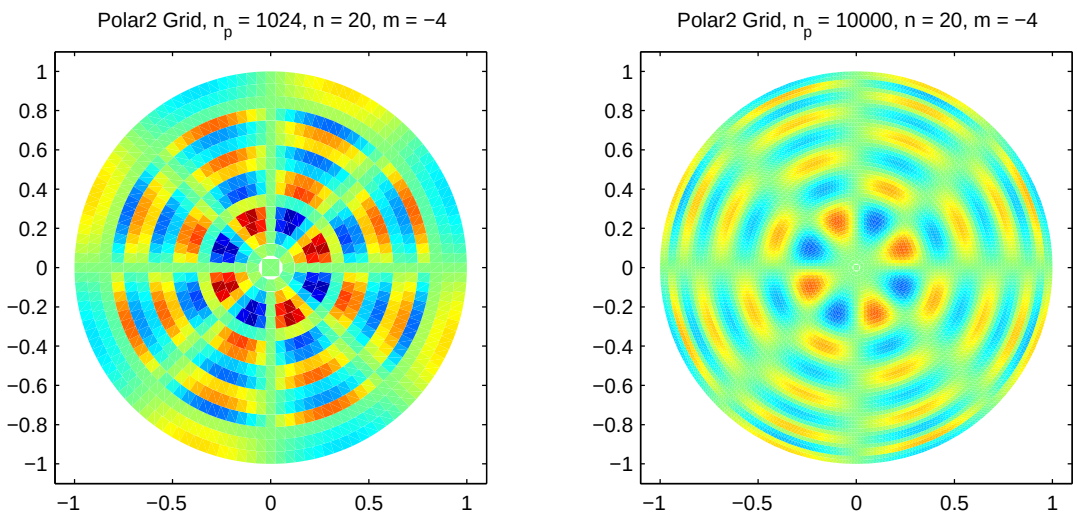
The Λ_{20}^{-4} polynomial sampled at low resolution 3.7(a) and high resolution 3.7(b) over the Cartesian grid shows fair representation overall. The low-resolution grid introduces significant variability between adjacent pixels and virtually no symmetry. While the smoothness does improve at the higher resolution grid on the interior, it remains poor around the perimeter values. The symmetry also improves, but not greatly, again due to the variability at the perimeter. The convergence is rated as fair mostly since the interior appears to improve but the perimeter remains highly variable.



(a) (b)

Figure 3.8: Λ_{20}^{-4} sampled over the simple polar grid with (a) 1,024 points and (b) 10,000 points. The polynomial has excellent smoothness and excellent symmetry at both resolutions. The convergence is good over most of the domain, but, the pixels around the perimeter appear to omit finer details in the radial coordinate.

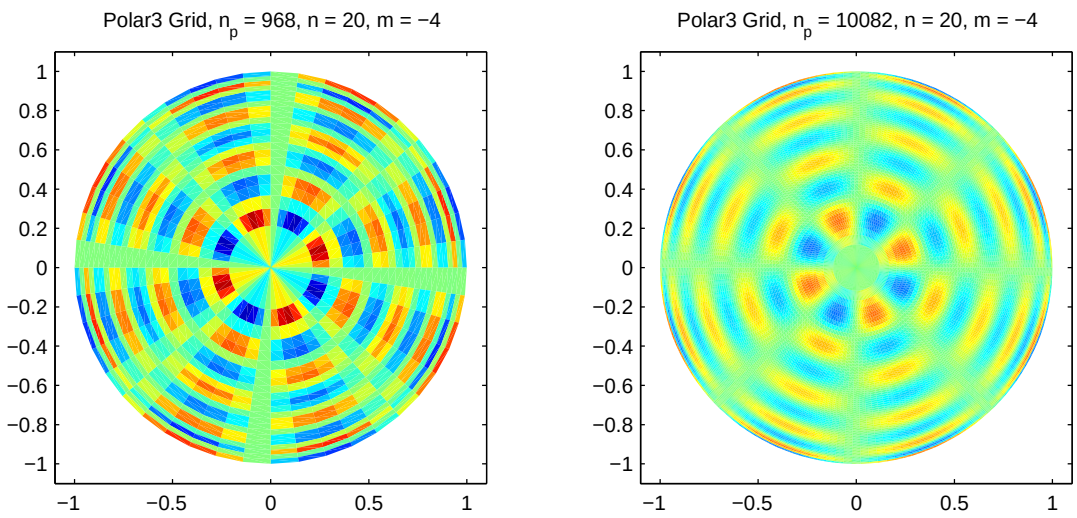
The Λ_{20}^{-4} polynomial sampled at low resolution 3.8(a) and high resolution 3.8(b) over the simple polar (polar 1) grid shows good representation overall. The pixel values transition smoothly azimuthally at both resolutions, while they transition rapidly, but as required, radially. Thus the smoothness for the simple polar grid is rated as good. The polynomial appears to have perfect symmetry at both resolutions and receives an excellent rating. The convergence was rated as good since the interior appears to improve very well but low-resolution version loses detail around the perimeter.



(a) (b)

Figure 3.9: Λ_{20}^{-4} sampled over the improved polar grid with (a) 1,024 points and (b) 10,000 points. The polynomial has excellent smoothness and excellent symmetry at both resolutions. The convergence is only rated as good, since the pixels around the perimeter obscure significant details in the radial coordinate.

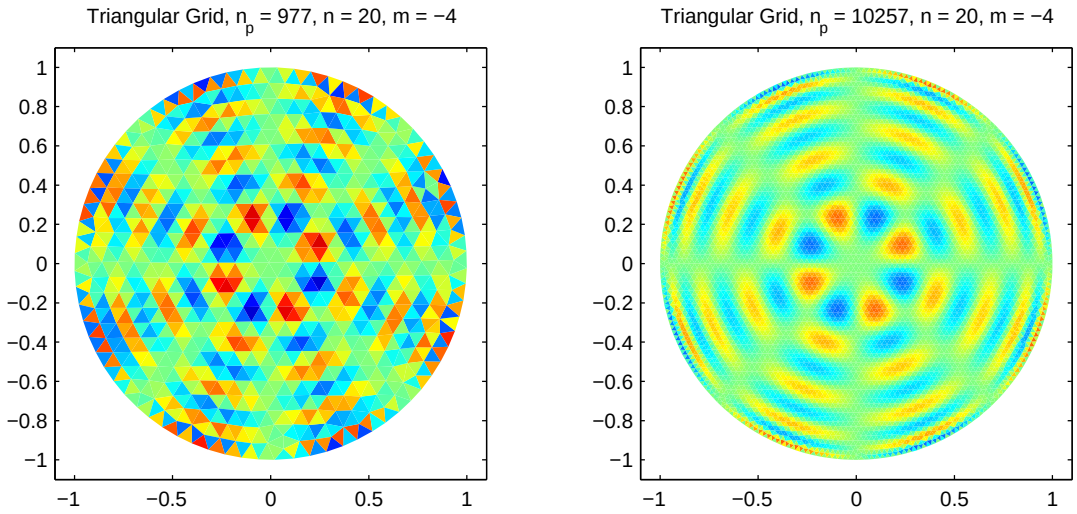
The Λ_{20}^{-4} polynomial sampled at low resolution 3.9(a) and high resolution 3.9(b) over the improved polar (polar 2) grid shows almost identical behavior as the simple polar grid. The convergence, however, is only as fair since it appears that the center of the polynomial is oversampled at the expense of under-sampling the perimeter. It appears that this grid significantly over-samples the azimuthal coordinate at all radial segments.



(a) (b)

Figure 3.10: Λ_{20}^{-4} sampled over the advanced polar grid with (a) 968 points and (b) 10,082 points. The polynomial has excellent smoothness and excellent symmetry. The convergence is rated as excellent; it appears that the polynomials is represented very well at both resolutions over the entire domain. The high-resolution version appears as a refinement of the low-resolution one over the entire domain. Also notice that the polynomial in (b) renders the most detail around the perimeter than any other grid.

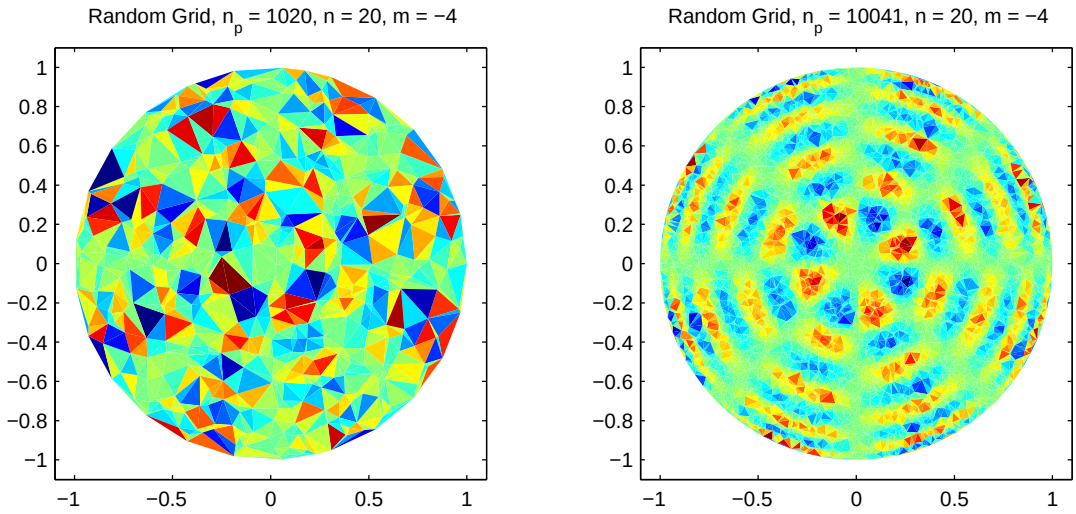
The Λ_{20}^{-4} polynomial sampled at low resolution 3.10(a) and high resolution 3.10(b) over the advanced polar (polar 3) grid shows improved representation at both resolutions. The smoothness over the high-resolution sampling appears superb over the entire domain. While the smoothness of the low-resolution version appears somewhat abrupt, this is, however, tolerable since the transitions represent the polynomial's transitions. This grid, like the previous grids in the polar domain, also has excellent symmetry. Furthermore, the convergence of the polynomial on this grid appears substantially improved; most notably around the perimeter. Not only does the low-resolution grid appear to render the polynomial very well, but the high-resolution also reveals more details. Overall, the advanced polar grid rates excellent in all areas.



(a) (b)

Figure 3.11: Λ_{20}^{-4} sampled over the triangular grid with (a) 977 points and (b) 10,257 points. The triangular grid is similar to the Cartesian and improved polar grids since it is composed of pixels with similar area. This grid renders the perimeter better than the crude Cartesian approximation, but not as well as any of the polar-type grids.

Λ_{20}^{-4} sampled at low resolution 3.11(a) and high resolution 3.11(b) over the triangular grid shows fair representation of the polynomial overall. The smoothness is poor at the lower resolution but is good at higher resolution. The symmetry is also poor at the lower resolution and good at the higher resolution. The convergence, like that of the Cartesian grid, rates only fair since significantly more points are required to render a better approximation of the polynomial, yet still contains irregularities around the perimeter.



(a) (b)

Figure 3.12: Λ_{20}^{-4} sampled over the random grid with (a) 1,020 points and (b) 10,041 points. The random grid has poor smoothness, symmetry, and convergence at both resolutions. As expected, these qualities suffer from the the haphazard location and size of the pixels.

The low-resolution 3.12(a) and high-resolution 3.12(b) random grids both poorly represent the Λ_{20}^{-4} polynomial overall. The polynomial rendered at either resolution shows virtually no smoothness or symmetry. Increasing the number of samples does not appear to improve the representation by any appreciable amount.

	Cartesian	Polar 1	Polar 2	Polar 3	Triangular	Random
Smoothness	fair	excellent	excellent	excellent	fair	poor
Symmetry	fair	excellent	excellent	excellent	fair	poor
Convergence	fair	good	fair	excellent	fair	poor

Table 3.1: The qualitative assessment of the grids representing the ZP Λ_n^m shows better marks for the grids in the polar coordinate system.

The preliminary examination of the various grids shows that using rectangular or triangular elements (as opposed to sectors) to represent the ZP Λ_n^m have serious limitations. Tab. 3.1 summarizes the initial assessment of how well each grid ranks with smoothness, symmetry, consistency, and convergence. Sectors, in the same polar coordinate system as the polynomials, have a clear advantage qualitatively. All the grids in polar coordinates appear generally represent the polynomial fairly well. The uniform pixel area in the improved polar grid (polar 2), however, over-samples around the perimeter azimuthally at the expense of under-sampling there radially. The pixels at the perimeter in most cases had marginal convergence and consistency. The center of Λ_n^m, and similarly for the other ZP not shown, showed fewer oscillations in the radial coordinate toward the center and more oscillations toward the perimeter. The oscillations in the azimuthal coordinate, however, were independent of the radial location; not surprisingly, when considering the regularity of the sine and cosine functions. After this preliminary, qualitative examination, the advanced polar (polar 3) grid, with more samples toward the perimeter, seems to be better suited to represent the ZP. The next section utilizes these grids to sample the entire set of ZP as a system of matrix equations using these grids to develop the DZP.

3.3 Discrete Zernike Polynomial System

To explore the suitability of the DZP as a basis for digital image analysis, it is critical to examine the orthogonality of the polynomials and reconstruction accuracy of the set. Prior to that task, the ZP first described in Eq. 2.13 on page 14 must be recast into the discrete domain, i.e., into a matrix equation.

First, a finite set of the order and repetition pairs are generated up to some maximal order, $n_{\max}$: $\mathbf{n} = [n_1, n_2, ..., n_K]$ and $\mathbf{m} = [m_1, m_2, ..., m_K]$, where $K = \frac{1}{2}(n_{\max} + 1)(n_{\max} + 2)$. A single variable, k, indexes all of the DZP. The order in which the (n, m) pairs are sorted is discussed later near Tab. 3.2. The ZP and function are then approximated with piecewise-constant approximations by sampling at the center-point of each pixel over a grid with J pixels, as in the following:

$$\begin{aligned} \mathbf{\Lambda} &= [\Lambda_{jk}], \\ \Lambda_{jk} &= \Lambda_{n_k}^{m_k}(x_j, y_j), \end{aligned} \tag{3.3}$$

and,

$$\begin{aligned} \mathbf{f} &= [f_j], \\ f_j &= f(x_j, y_j). \end{aligned} \tag{3.4}$$

From Eq. 2.15, the constant terms move out of the integration, the integration becomes a summation, to obtain the approximated FZC of the function, f, up to order $n_{\max}$:

$$\tilde{\lambda}_k = \sum_{j=1}^{J} \omega_j f_j \Lambda_{jk}, \tag{3.5}$$

where,

$$\omega_j = \int_{\Omega_j} d\Omega \tag{3.6}$$

represents the area of the j^{th} pixel.

To recover the approximated function, the DZP are multiplied by their respective FZC and summed over all polynomials. Similar to Eq. 2.16 for the FZS, we have the following, truncated approximation to the function at each of the pixels, for $j = 1, 2, ..., J$.

$$\tilde{f}_j = \frac{1}{\omega_j} \sum_{k=1}^{K} \tilde{\lambda}_k \Lambda_{jk} \tag{3.7}$$

Modifying Eqs. 3.3 and 3.4 to implicitly include the weight terms, $\hat{\Lambda}_{jk} = \sqrt{\omega_j}\Lambda_{jk}$ and $\hat{f}_j = \sqrt{\omega_j} f_j$, leads to the following system,

$$\hat{\mathbf{\Lambda}}\tilde{\boldsymbol{\lambda}} = \hat{\mathbf{f}} \tag{3.8}$$

where $\hat{\mathbf{\Lambda}}$ is $J \times K$, $\tilde{\boldsymbol{\lambda}}$ is $K \times 1$, and $\hat{\mathbf{f}}$ is $J \times 1$.

Again, J is the total number of pixels in the grid and K is the total number of polynomials up to order $n_{\max}$. This change allows use of the standard, inner product for later computations instead of the weighted inner product.

As mentioned earlier, it is important to discuss the order in which the DZP are collected. Three ordering schemes are presented in Tab. 3.2. For this analysis, it was found that sorting the DZP by their repetition number then by their order (the 3$^{\text{rd}}$ scheme in Tab. 3.2) reduced the condition number of the Gram matrix (discussed in the following sections). This sorting scheme is standard for the remainder of this analysis unless stated otherwise. The second ordering scheme is convenient for initial computation of the polynomials, but then they are re-sorted to third scheme for analysis of the system.

	1		2		3	
k	n	m	n	m	n	m
1	0	0	0	0	4	−4
2	1	−1	1	+1	3	−3
3	1	+1	1	−1	2	−2
4	2	−2	2	+2	4	−2
5	2	0	2	−2	1	−1
6	2	+2	2	0	3	−1
7	3	−3	3	+3	0	0
8	3	−1	3	−3	2	0
9	3	+1	3	+1	4	0
10	3	+3	3	−1	1	+1
11	4	−4	4	+4	3	+1
12	4	−2	4	−4	2	+2
13	4	0	4	+2	4	+2
14	4	+2	4	−2	3	+3
15	4	+4	4	0	4	+4

Table 3.2: Three sorting schemes of the order and repetition numbers, up to $n_{\max} = 4 \Rightarrow K = 15$. The first scheme sorts by n then by m. The second scheme sorts ascending by n, then descending by the magnitude of m. The third scheme sorts by m then by n.

This section developed the formulas for computing the approximate FZC and FZS as a matrix equation. The $J \times K$ matrix $\hat{\mathbf{\Lambda}}$ contains all of the DZP up to the maximum order $n_{\max}$ which utilizes J pixels from one of the previously described grids, and K polynomials where $K = \frac{1}{2}(n_{\max} + 1)(n_{\max} + 2)$. The next section introduces the Gram matrix for examining the orthogonality of the DZP.

3.4 Gram Matrix

The Gram matrix, also known as the Gramian, is the compilation of a set of vectors, $\{\mathbf{\Lambda}_{*k}\}$ for $k = 1, 2, ..., K$, in an inner product space, which is Hermitian and positive semidefinite, with the entries given by [29]:

$$G_{k_1 k_2} = \langle \mathbf{\Lambda}_{*k_1}, \mathbf{\Lambda}_{*k_2} \rangle \,, \tag{3.9}$$

The assembled matrix for the DZP appears as,

$$\mathbf{G} = \begin{pmatrix} \langle \mathbf{\Lambda}_{*1}, \mathbf{\Lambda}_{*1} \rangle & \langle \mathbf{\Lambda}_{*1}, \mathbf{\Lambda}_{*2} \rangle & \cdots & \langle \mathbf{\Lambda}_{*1}, \mathbf{\Lambda}_{*K} \rangle \\ \langle \mathbf{\Lambda}_{*2}, \mathbf{\Lambda}_{*1} \rangle & \langle \mathbf{\Lambda}_{*2}, \mathbf{\Lambda}_{*2} \rangle & \cdots & \langle \mathbf{\Lambda}_{*2}, \mathbf{\Lambda}_{*K} \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \mathbf{\Lambda}_{*K}, \mathbf{\Lambda}_{*1} \rangle & \langle \mathbf{\Lambda}_{*K}, \mathbf{\Lambda}_{*2} \rangle & \cdots & \langle \mathbf{\Lambda}_{*K}, \mathbf{\Lambda}_{*K} \rangle \end{pmatrix} . \tag{3.10}$$

When the set is orthonormal, the Gram matrix reduces to the identity matrix. Non-zero values in any of the off-diagonal locations indicates that those vectors overlap, i.e., are not orthogonal. The Gram matrix allows us to examine the orthogonality of the DZP over a given grid and up to some maximum order.

The Gram matrix for this analysis appears in the following associated system of normal equations:

$$\hat{\mathbf{\Lambda}}^{\mathrm{T}} \hat{\mathbf{\Lambda}} \tilde{\boldsymbol{\lambda}} = \hat{\mathbf{\Lambda}}^{\mathrm{T}} \mathbf{f} \tag{3.11}$$

The new matrix, $\mathbf{G} = \hat{\mathbf{\Lambda}}^{\mathrm{T}}\hat{\mathbf{\Lambda}}$, is the Gram matrix of the system associated with $\hat{\mathbf{\Lambda}}$. When assembling the DZP matrix, $\hat{\mathbf{\Lambda}}$, each polynomial is explicitly normalized numerically. This ensures that all of the diagonal elements of the Gram matrix precisely equal 1, as in the following,

$$\hat{\mathbf{\Lambda}}_{*k} \rightarrow \frac{\hat{\mathbf{\Lambda}}_{*k}}{\|\hat{\mathbf{\Lambda}}_{*k}\|_2} \Rightarrow \left\langle \hat{\mathbf{\Lambda}}_{*k}, \hat{\mathbf{\Lambda}}_{*k} \right\rangle = 1 \tag{3.12}$$

Ideally, the off diagonal elements in $\mathbf{G}$ would be zero or close to zero given machine precision; $\varepsilon \approx 2^{-52} \approx 10^{-16}$ for double precision arithmetic. The DZP are sorted first by the repetition number, m, then by the order, n. In this fashion, the orthogonality of the subsets of polynomials with equal m appear in the Gram matrix as blocks on the diagonal. The block in the center is the largest since this subset of purely radial polynomials (i.e., $m = 0$) has the most members. The orthogonality of the subsets of polynomials grouped with equal n appear spotted throughout the entire matrix. The off-diagonal, non-zero blocks indicate interference between polynomials of various orders, n, while the on-diagonal, non-zero blocks indicate interference between polynomials of similar repetition, m. Figs. 3.13 through 3.18 illustrate the Gram matrix associated with the 6 grids at 3 different resolutions: 20, 200, and 2,000 pixels. The maximum polynomial order is constant, $n_{\max} = 10 \Rightarrow K = 66$.

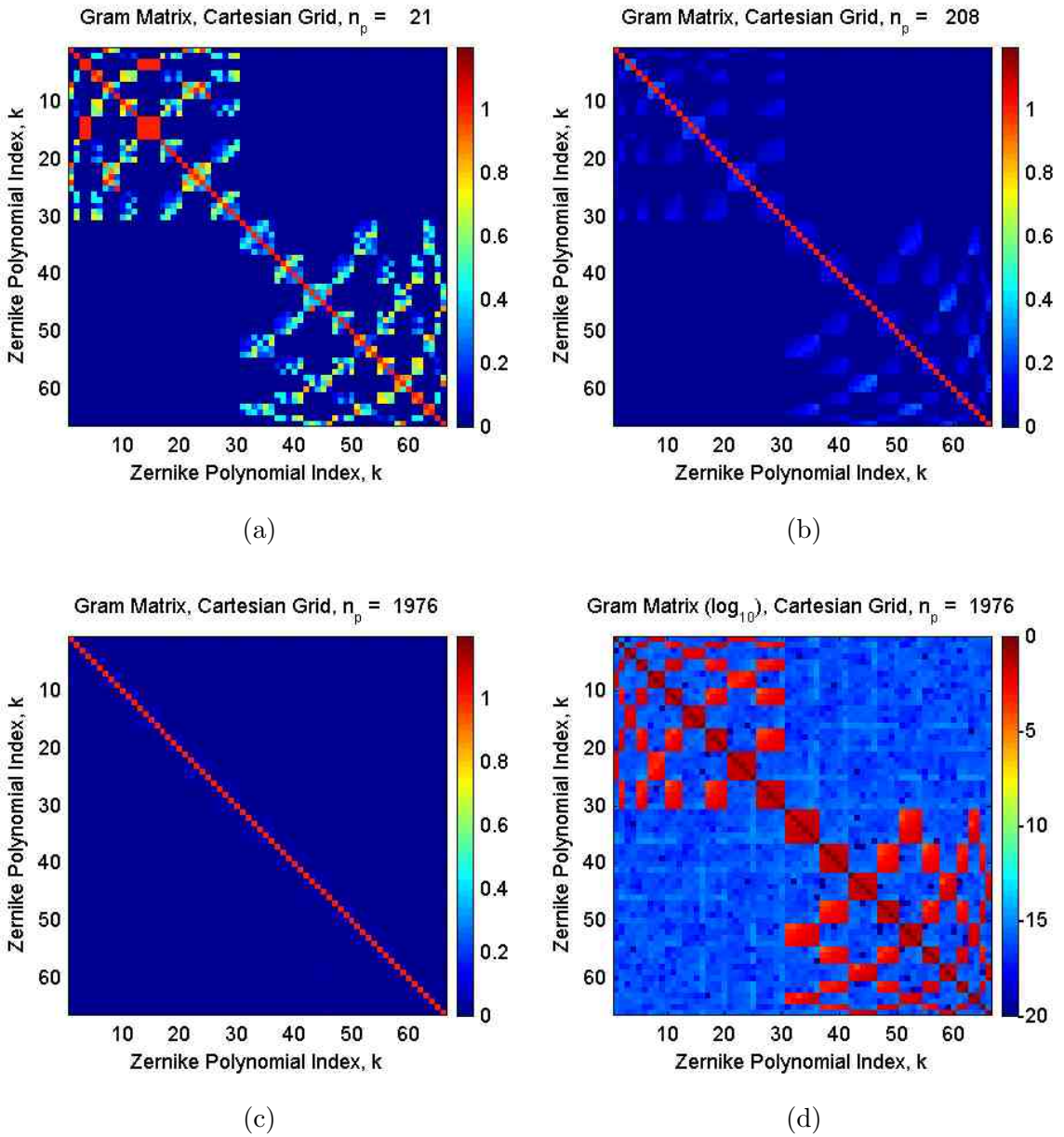


Figure 3.13: The Gram matrix **G** is shown in a linear colorscale for the DZP up to the 10^{th} degree over the Cartesian grid for: (a) 21, (b) 208, and (c) 1,976 points. The image in (d) is the same Gram matrix as in (c) but uses a logarithmic colorscale. The off-diagonal entries tend toward zero for increasing resolution, n_p, but are still quite evident as in (d). This indicates that the system does not improve orthogonality rapidly.

The images in Fig. 3.13 illustrates the the Gram matrix for the DZP over the Cartesian grid at several resolutions: 21 points in (a), 208 in (b), and 1,976 in both (c) and (d). Fig. 3.13(a) shows many non-zero entries off of the diagonal indicating a significant amout of aliasing between the polynomials. Essentially, the set of polynomials is not adequately represented and the system is highly dependent. The pattern indicates that polynomials which are even/odd with $\text{sgn}(m_{k_1}) = \text{sgn}(m_{k_2})$ generally only interfere with each other for any given order. This suggests that the interference pattern is caused by the symmetry in the grid combined with the alignment of the polynomials. Fig. 3.13(b) shows similar behavior as in (a) but with a reduced effect. The effect almost disappears in Fig. 3.13(c) except for some residual interference along the block-diagonal elements. This indicates a strong interference within the subsets of polynomials with the same repetition number, m. The Gram matrix in Fig. 3.13(d) is identical to that in (c) but rendered in a logarithmic colorscale. The even/odd interference pattern re-emerges and indicates that the orthogonality of the polynomials converges slowly.

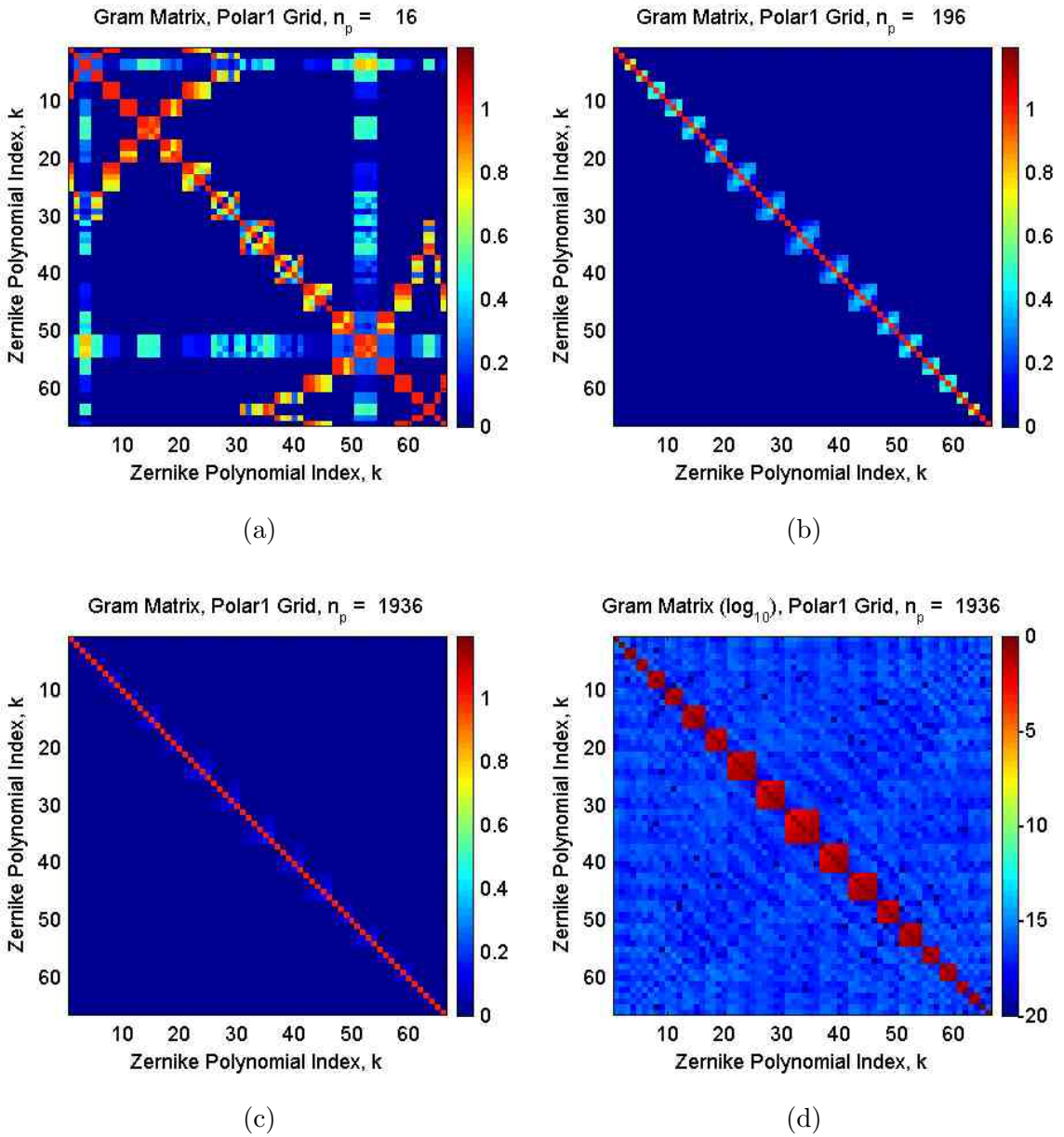


Figure 3.14: The Gram matrix $\mathbf{G}$ is shown in a linear colorscale for the DZP up to the 10^{th} degree over the simple polar grid for: (a) 16, (b) 196, and (c) 1,936 points. The image in (d) is the same Gram matrix as in (c) but uses a logarithmic colorscale. The far, off-diagonal entries are $\approx 10^{-16}$, effectively zero; indicating ideal orthogonality. The block-diagonal elements, however, still show significant interference for all grid resolutions.

The images in Fig. 3.14 illustrates the the Gram matrix for the DZP over the simple polar grid at several resolutions: 16 points in (a), 196 in (b), and 1,936 in both (c) and (d). Fig. 3.14(a) shows non-zero entries off of the diagonal across horizontal and vertical strips and on the block diagonals. The strips indicate interference from a particular repetition number with all of the other polynomials, while the block diagonal elements indicate interference within each of the families of repetition number. The interference along the strips is due to under-sampling in the azimuthal coordinate and quickly vanishes at higher resolutions as in Fig. 3.14(b) and (c). The interference within each of the subsets of repetition numbers remains significant even at the higher resolution as seen in (c). Under the greater scrutiny of the logarithmic colorscale as in (d), we can see that all of the aliasing, except for the block-diagonal elements, disappears, i.e. $\approx 10^{-16}$. This a direct result of sampling the DZP in polar coordinates.

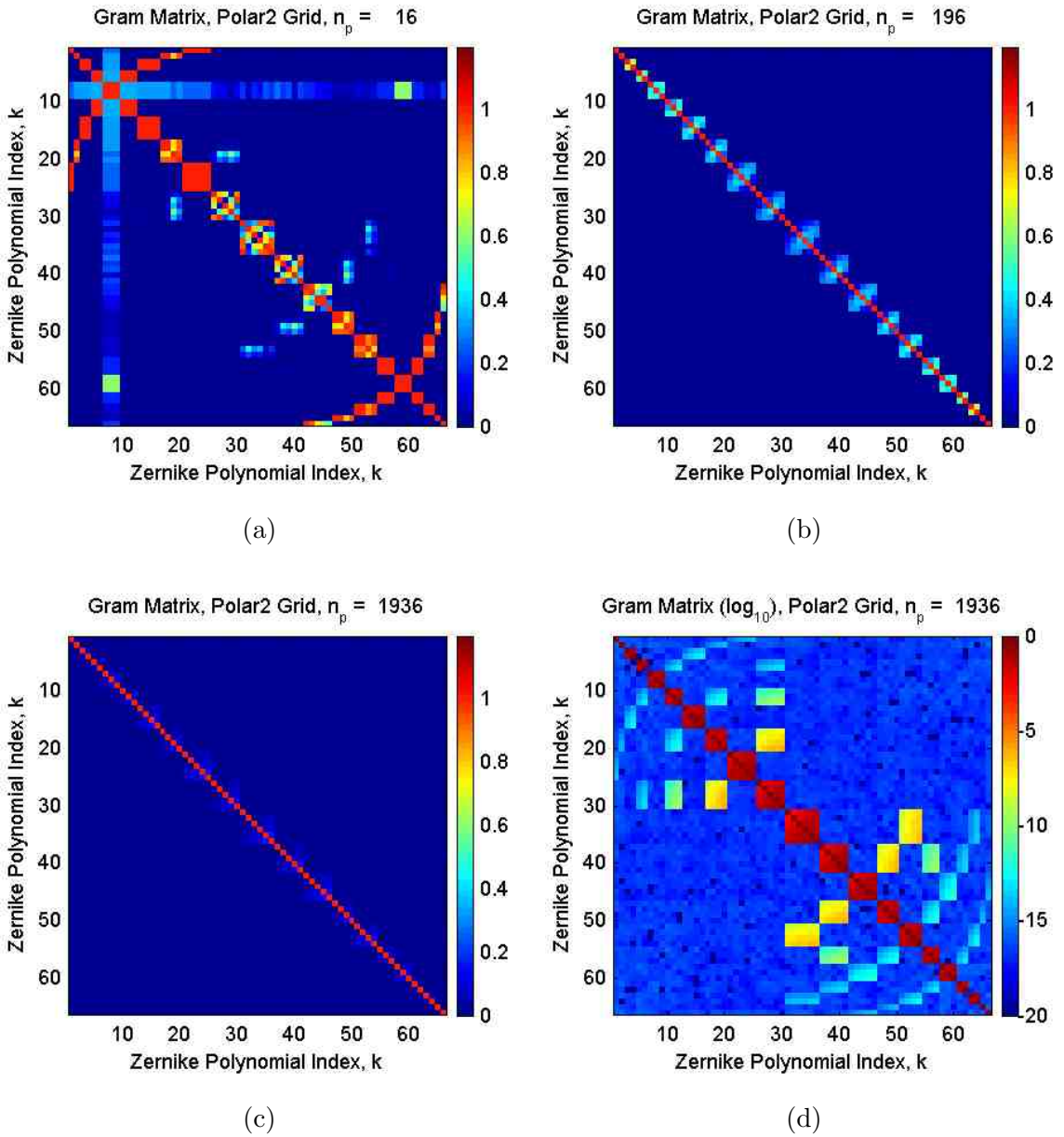
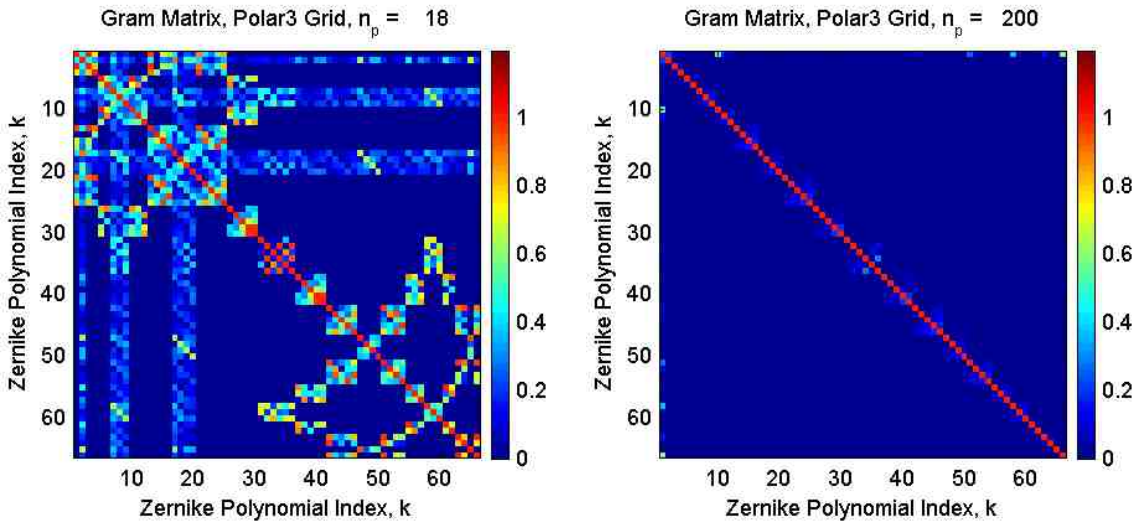


Figure 3.15: The Gram matrix $\mathbf{G}$ is shown in a linear colorscale for the DZP up to the 10^{th} degree over the improved polar grid for: (a) 16, (b) 196, and (c) 1,936 points. The image in (d) is the same Gram matrix as in (c) but uses a logarithmic colorscale. Most of the far, off-diagonal entries are effectively zero ($\approx 10^{-16}$), but there is significantly more interference than in the previous simple polar grid. The block-diagonal elements remain prominent as in the previous grid.

The images in Fig. 3.15 shows the the Gram matrix for the DZP over the improved polar grid at several resolutions: 16 points in (a), 196 in (b), and 1,936 in both (c) and (d). The image in Fig. 3.15(a) shows interference between the polynomials, but not at any regular or repeatable occurrence. At such low resolutions on the grid, the alias may be due to a combination of under-sampling both the radial and azimuthal coordinates. At higher resolutions, as in (b) and (c), the only remaining interference is between the subsets of polynomials with identical repetition numbers. The image in (d), however, shows that in greater detail, an interference pattern is still present. This patter is similar to the one originally observed with the Cartesian grid; the even/odd polynomials tend to only interfere with the even/odd polynomials. This is likely due to the similar symmetry that this grid shares with the Cartesian grid: both grids have perfect symmetry about the horizontal and vertical axes through the origin. This increases the ability of polynomials with similar symmetry to the grid to cause interference and reduce the convergence to an orthogonal system.



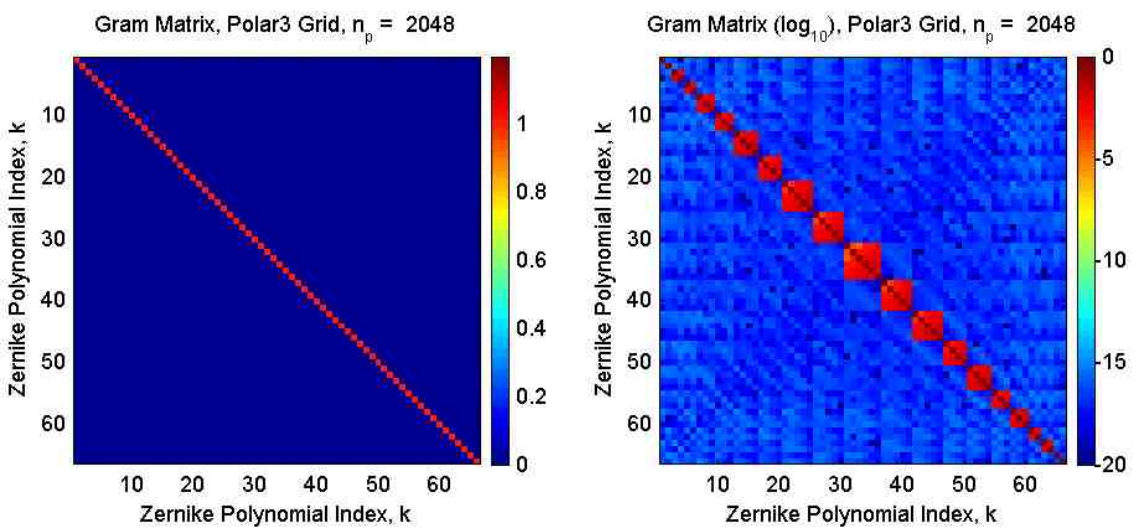
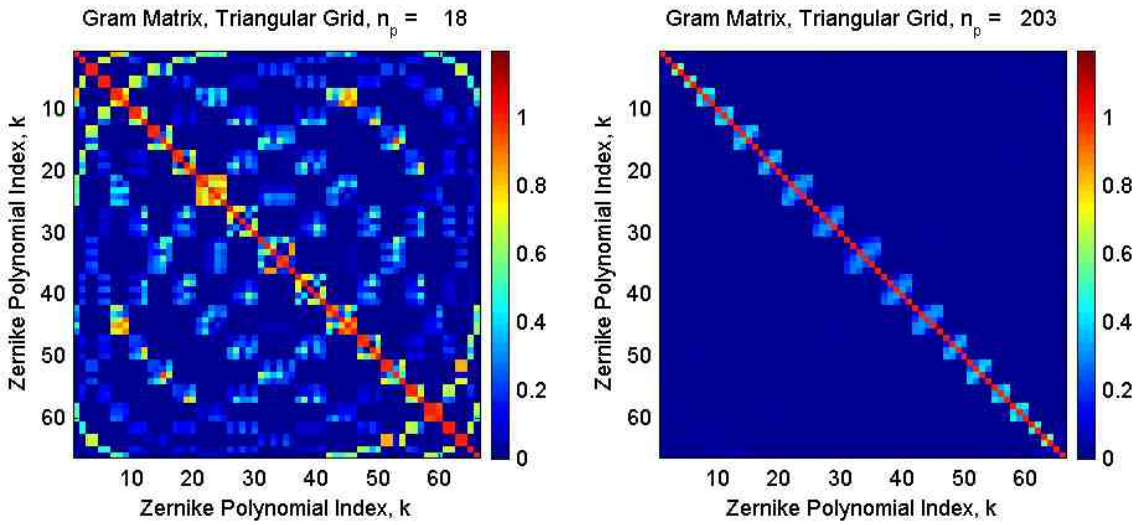


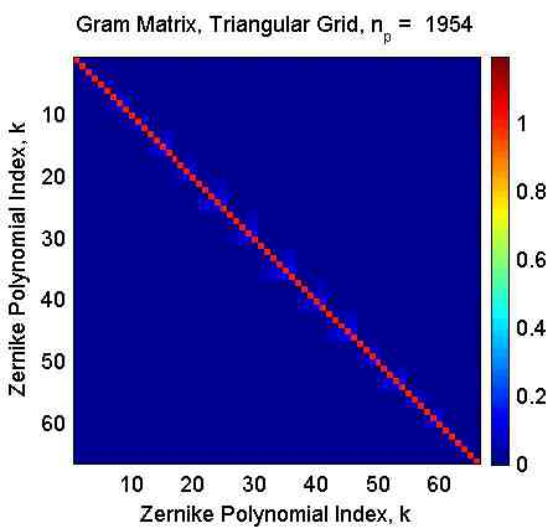
Figure 3.16: The Gram matrix $\mathbf{G}$ is shown in a linear colorscale for the DZP up to the 10^{th} degree over the advanced polar grid for: (a) 18, (b) 200, and (c) 2,048 points. The image in (d) is the same Gram matrix as in (c) but uses a logarithmic colorscale. Similar to the orthogonality of the simple polar grid, the far, off-diagonal entries are $\approx 10^{-16}$, effectively zero which indicates ideal orthogonality. The block-diagonal elements still show interference at all grid resolutions, but much less than any other grid.

The images in Fig. 3.16 shows the the Gram matrix for the DZP over the advanced polar grid at several resolutions: 18 points in (a), 200 in (b), and 2,048 in both (c) and (d). The image in Fig. 3.16(a) shows heavy interference which quickly vanishes as in (b) and (c). The non-zero elements along the diagonal blocks in (d) indicates that the polynomials within the subsets of identical repetition numbers still interfere with each other. Similar to the results observed with the simple polar grid in Fig. 3.14(d), the polynomials with different repetition numbers, m, are effectively orthogonal. This is evident from the trivial values, $\approx 10^{-16}$, off the diagonal and off of the diagonal blocks. The values in the diagonal blocks for the advanced polar grid, however, are typically at least an order of magnitude smaller than those in the simple polar grid.

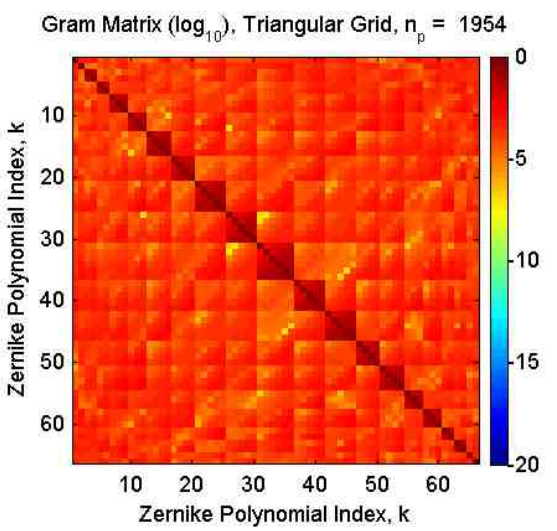


(a)

(b)



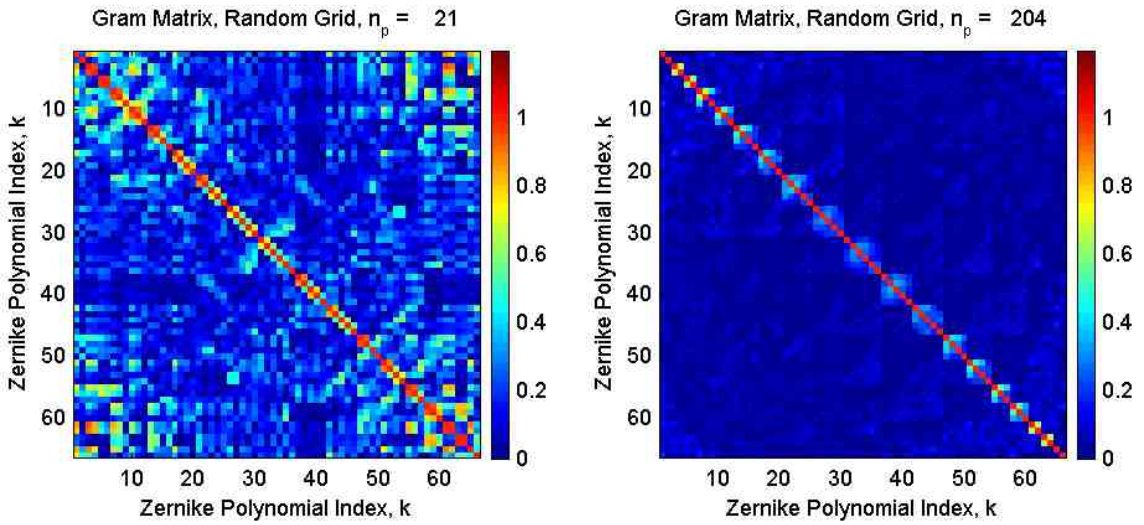
(c)



(d)

Figure 3.17: The Gram matrix $\mathbf{G}$ is shown in a linear colorscale for the DZP up to the 10^{th} degree over the triangular grid for: (a) 18, (b) 203, and (c) 1,954 points. The image in (d) is the same Gram matrix as in (c) but uses a logarithmic colorscale. At low resolution (a), the under-sampled DZP show significant interference with many other of the polynomials in an interesting pattern. At the higher resolutions, as in (b) and (c), the interference appears to decrease, except for the block diagonal elements. As we can see in (d), however, the DZP still show non-trivial interference between all polynomials where the elements are $\approx 10^{-5}$.

Fig. 3.17 shows the the Gram matrix for the DZP over the triangular grid at several resolutions: 18 points in (a), 203 in (b), and 1,954 in both (c) and (d). The image in Fig. 3.17(a) shows interference between the majority of the polynomials. The interference is spread over polynomials of similar order: even/odd polynomials interfere with even/odd polynomials, respectively. In this case, it does not depend on the sign of the repetition number, as in the case with the Cartesian and improved polar grids. This aliasing is likely due to under-sampling in the azimuthal coordinate. Also present is the interference between polynomials within the families of identical repetition numbers. This is again seen from the non-zero elements in the blocks on the diagonal, as in (b) and less pronounced in (c). The image in Fig. 3.17(d) shows tremendous interference between all polynomials regardless of order or repetition. This is likely due to inadequate resolution near the perimeter, as observed in the previous section for this grid.



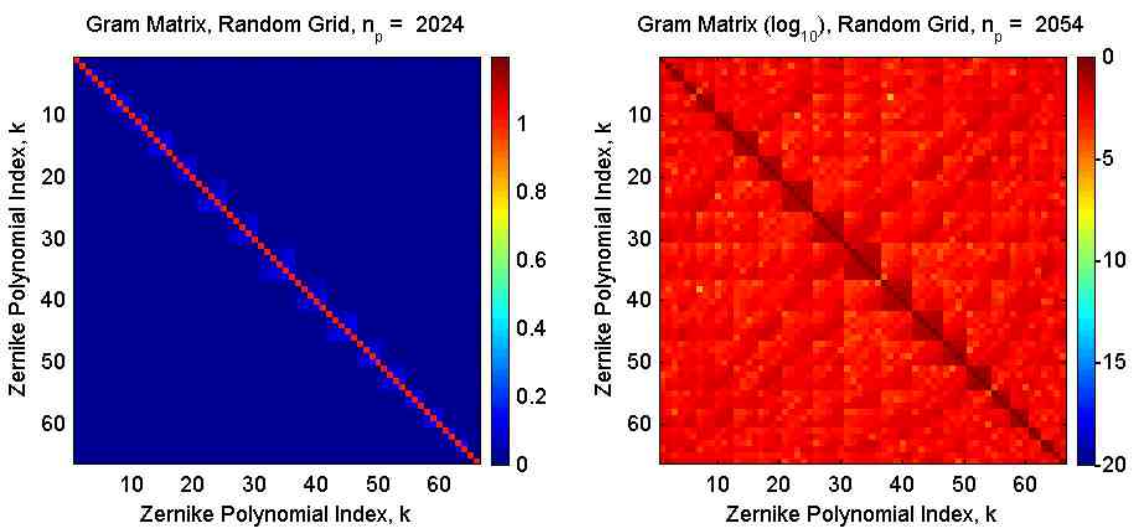


Figure 3.18: The Gram matrix $\mathbf{G}$ is shown in a linear colorscale for the DZP up to the 10^{th} degree over the random grid for: (a) 21, (b) 204, and (c) 2,024 points. The image in (d) is the same Gram matrix as in (c) but uses a logarithmic colorscale. At low resolution (a), the under-sampled DZP show tremendous interference between all polynomials. At the higher resolutions, as in (b) and (c), the interference decrease, except for the block diagonal elements similar to the behavior in the triangular grid. As we can see in (d), however, the DZP show even more interference between all polynomials where the elements are $\approx 10^{-4}$ or more.

Lastly, the images in Fig. 3.18 shows the the Gram matrix for the DZP over the random grid at several resolutions: 21 points in (a), 204 in (b), 2,024 in (c), and 2,054 in (d). The image in Fig. 3.18(a) shows substantial interference between the all of the polynomials, especially toward the corners which represents the high-order DZP sensitive to misrepresentation at the perimeter. Fig. 3.18(b) shows lower values, but, the interference is still spread over all polynomials. The interference further decreases, as in (c), but is still much higher ($\approx 10^{-4}$) than desirable, as in (d).

The Gram matrix computed at several resolutions for each grid type provides a valuable inspection tool to examine the orthogonality of DZP. In all cases, it is apparent that the polynomials within the subsets of identical repetition number, m, typically interfere with each other. While this aliasing was present for all systems, it was, however, less pronounced in the advanced polar grid. It is also common to observe polynomials of interfere with each other with similar parity; even/odd polynomials interfered with even/odd polynomials, respectively, as seen in the Cartesian, improved polar, and triangular grids. Based on the results illustrated here, the random, triangular and improved polar grids appear to have the least orthogonality, while the advanced polar, simple polar and the Cartesian have the best. The Gram matrix shows even more insight to the orthogonality of the DZP, but what is needed is a quantitative metric to determine what resolution, n_p, is best for a given order, $n_{\max}$. The next section computes the rank and condition number to accomplish this.

3.5 Rank and Condition

In this section, the condition number and rank are computed to numerically explore which systems of the DZP are well-conditioned and full-rank. After a brief review of the necessary definitions, the condition number and rank are then be computed for each of the 6 grids previously described at various resolutions, n_p, and maximum polynomial order, $n_{\max}$.

The rank of a matrix, $R(\hat{\mathbf{\Lambda}}) = r$, can be defined as the number of independent columns, r [19]. The $J \times K$ matrix, $\hat{\mathbf{\Lambda}}$, is said to have full-rank, or linearly independent, when the rank is equal to the number of columns. The system is rank-deficient, or linearly dependent, otherwise, $r < K$. To facilitate a comparison of systems with various sizes, the rank will be normalized by the number of columns, such that the rank is bound in the interval $[0, 1]$,

$$\frac{R(\hat{\mathbf{\Lambda}})}{K}. \tag{3.13}$$

The condition number of a square matrix, $\mathbf{A}$, is commonly defined as [19, 26, 8]:

$$\kappa(\mathbf{A}) = \|\mathbf{A}\|\|\mathbf{A^{-1}}\|, \tag{3.14}$$

and in the case of a rectangular, or least-squares system, it generalizes to,

$$\kappa(\hat{\mathbf{\Lambda}}) = \|\hat{\mathbf{\Lambda}}\|\|\hat{\mathbf{\Lambda}}^{+}\|, \tag{3.15}$$

where $^{+}$ represents the pseudo-inverse or the Moore-Penrose inverse of $\hat{\mathbf{\Lambda}}$.

The condition number varies widely over many orders of magnitude, $1 \leq \kappa(\hat{\mathbf{\Lambda}}) \leq \infty$. To facilitate a comparison of various orders, the order of the condition number will be taken from the common logarithm. This is also a convenient estimation for the degree of accuracy we can attain a solution; one can generally expect to lose $\log_{10} \kappa(\hat{\mathbf{\Lambda}})$ digits of accuracy computing a solution using $\hat{\mathbf{\Lambda}}$.

The rank and condition number are both computed numerically from the singular value decomposition (SVD). The SVD translates a matrix into a diagonal form, provided the proper orthogonal coordinates for the domain and range. The SVD of an $J \times K$ matrix ($J \geq K$), $\hat{\mathbf{\Lambda}}$, is the product of a diagonal matrix of singular values, Σ, the left singular vectors, U, and the right singular vectors, V, which has the following form [8]:

$$\begin{aligned} \hat{\mathbf{\Lambda}} &= \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*, \\ \mathbf{\Sigma} &= \mathrm{diag}(\sigma_1, \sigma_2, ..., \sigma_K), \\ \mathbf{U}^*\mathbf{U} &= \mathbf{I}, \\ \mathbf{V}^*\mathbf{V} &= \mathbf{I}, \end{aligned} \tag{3.16}$$

where the singular values are ordered largest to smallest.

The induced matrix norm in the condition number, $\|\cdot\|$, is typically taken as the 2-norm. In this case the condition number is given by the ratio of the largest singular value, σ_1 to the smallest singular value, σ_K,

$$\kappa(\hat{\mathbf{\Lambda}}) = \|\hat{\mathbf{\Lambda}}\|_2 \|\hat{\mathbf{\Lambda}}^+\|_2 = \frac{\sigma_1}{\sigma_K}. \tag{3.17}$$

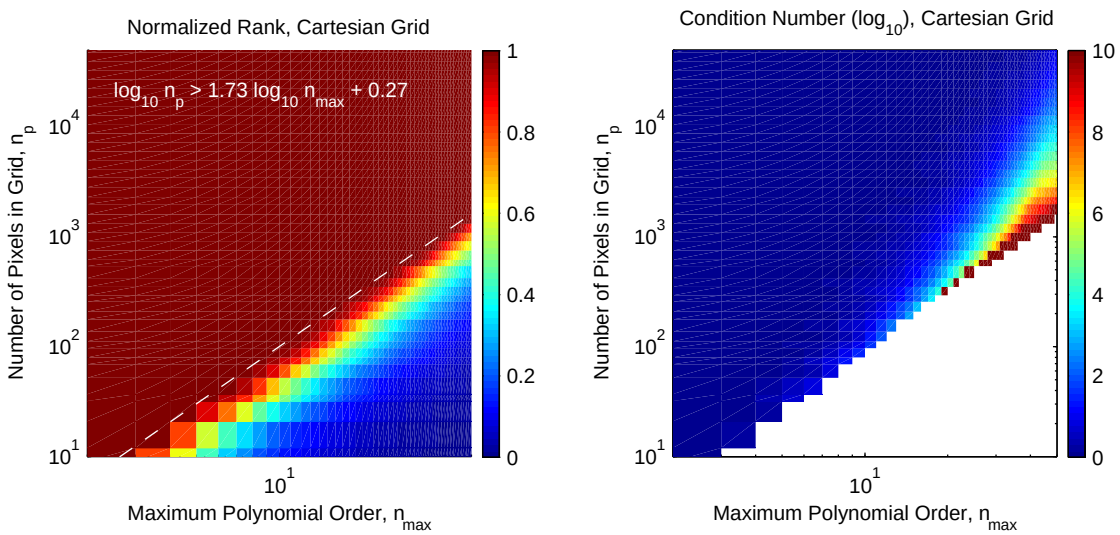
The SVD is also utilized to estimate the rank. For a linearly dependent system, the singular values tend to decrease by many orders of magnitude after the principal singular values. The rank is calculated by counting the number on non-trivial singular values above a threshold [17],

$$\tau = K\varepsilon(\sigma_1) \tag{3.18a}$$

$$R(\hat{\mathbf{\Lambda}}) = \sum_{k=1}^{K} \sigma_k > \tau \tag{3.18b}$$

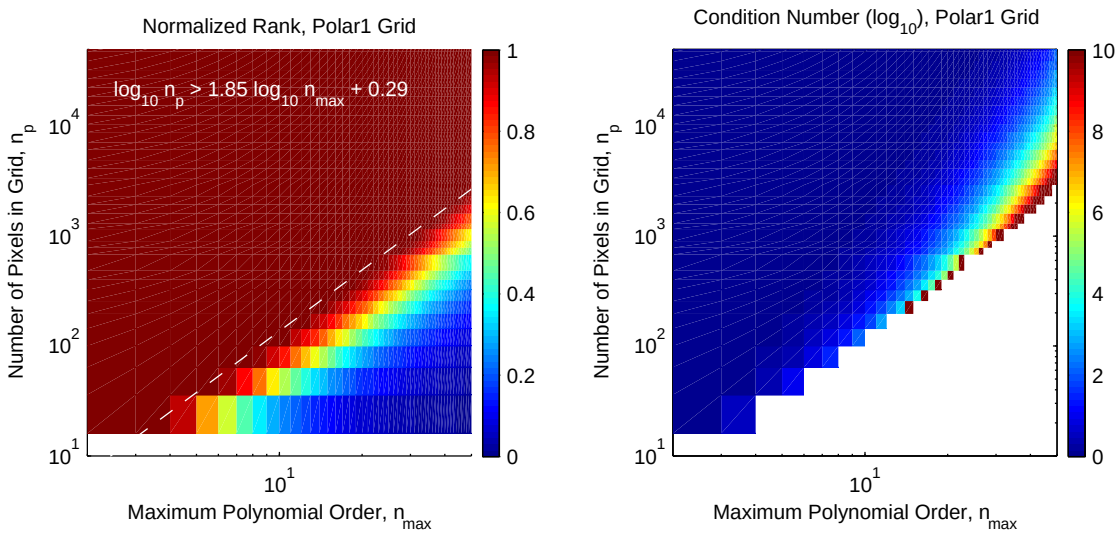
where $\varepsilon(\cdot)$ is the machine precision to the next, larger floating point number ($\varepsilon \approx 2^{-52} \approx 10^{-16}$ for double precision variables) [17].

A numerical exploration of the rank and condition number was performed over each of the 6 grids at various resolutions and maximum order of the DZP. The number of points, n_p, in each grid ranged logarithmically from 10^1 to 10^5. The maximum order of the DZP, $n_{\max}$, range linearly from 2 to 50. At each combination of $(n_{\max}, n_p)$, the rank and condition number are computed from the economic SVD. When the normalized rank was near unity, $\frac{1}{K}R(\hat{\mathbf{\Lambda}}) > 0.99$, the system is considered full-rank and the order of the condition number is also computed, $\log_{10}\kappa(\hat{\mathbf{\Lambda}})$. The results are plotted over logarithmic coordinates to emphasize the the change from full-rank to rank-deficient over the apparent minimum-resolution barrier; the rank is unity above some line in each image, but deteriorates rapidly below it. The boundary between full-rank and the rank-deficient domains is estimated. The minimum-resolution criteria is then computed using a linear fit.



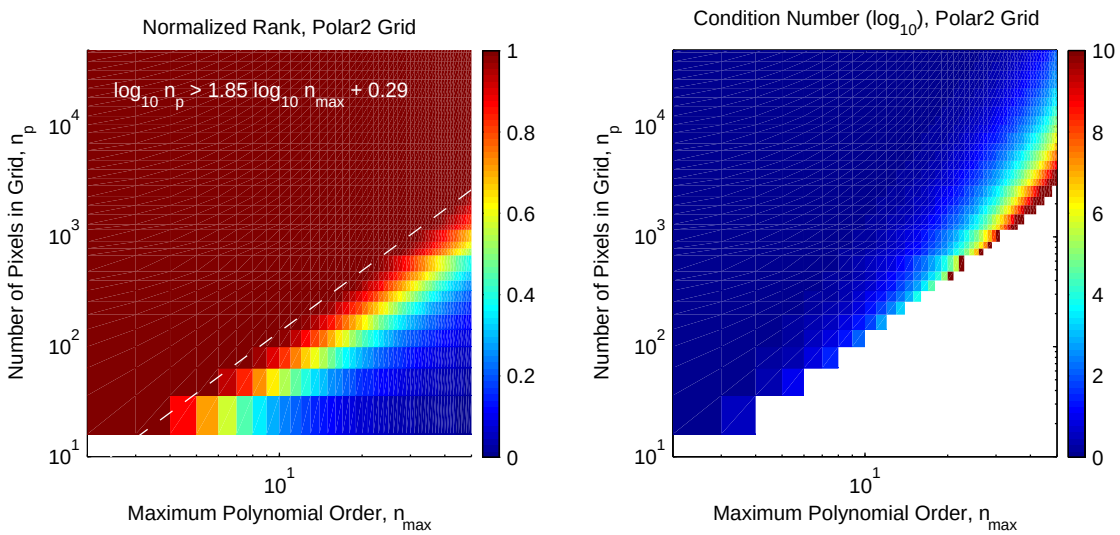
(a) (b)

Figure 3.19: The normalized rank (a) and the order of the condition number (b) for the Cartesian grid. The condition number rises at higher polynomial orders faster than the minimum resolution criteria.



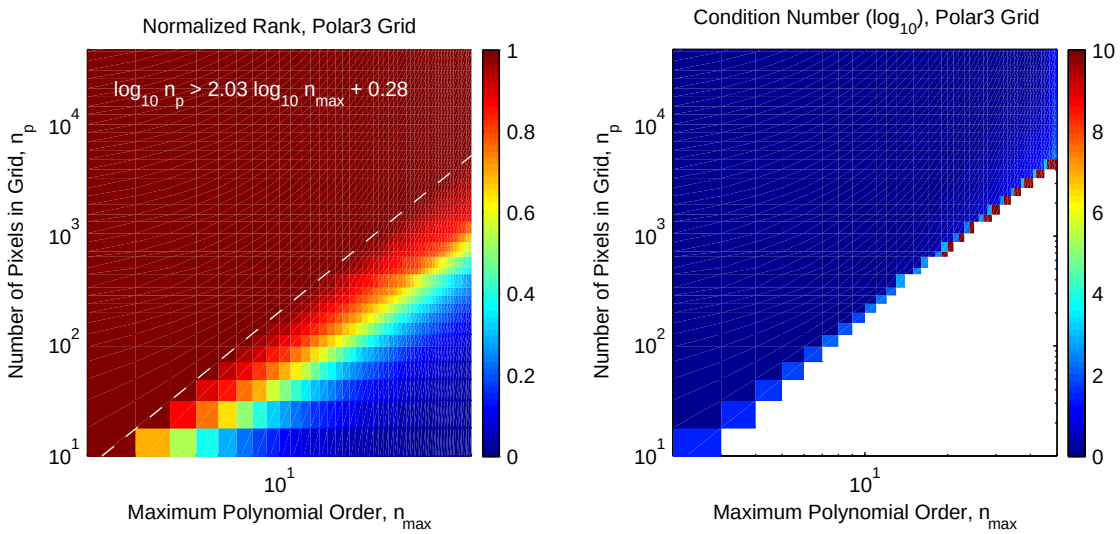
(a) (b)

Figure 3.20: The normalized rank (a) and the order of the condition number (b) for the simple polar grid. The condition number is very high at the higher polynomial orders.



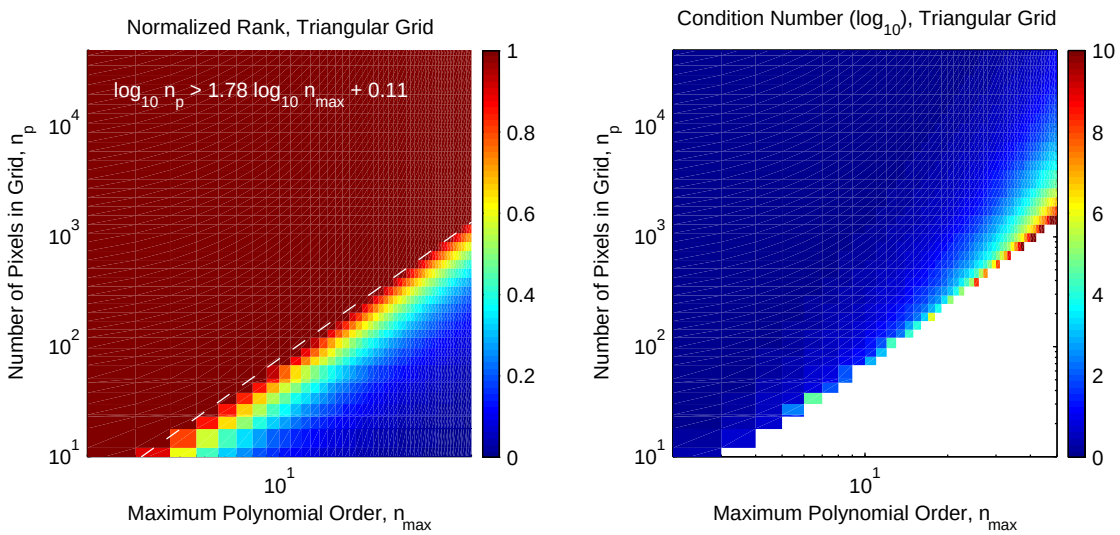
(a) (b)

Figure 3.21: The normalized rank (a) and the order of the condition number (b) for the improved polar grid. The condition number is also very high at high polynomial orders.



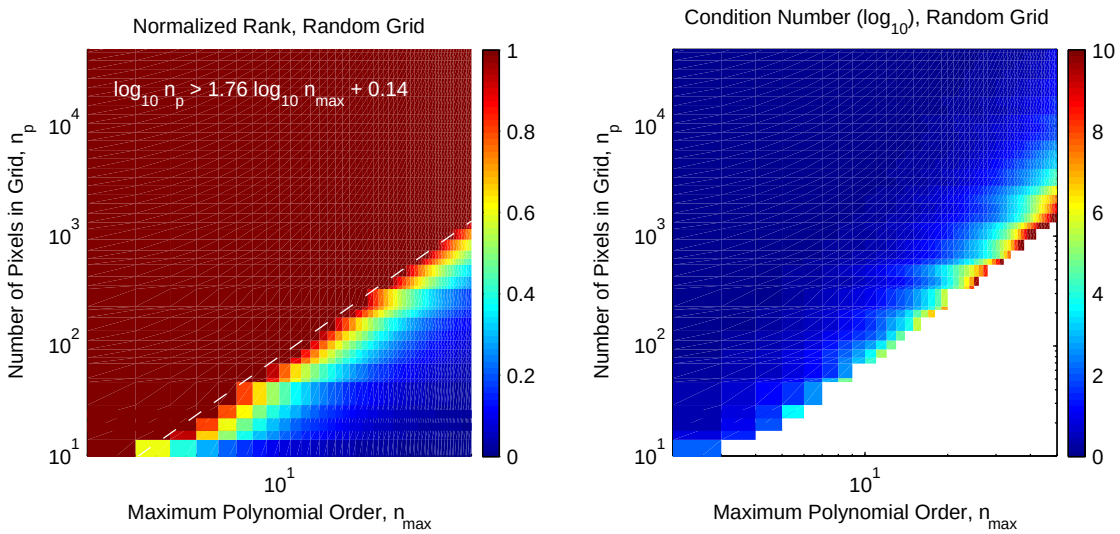
(a) (b)

Figure 3.22: The normalized rank (a) and the order of the condition number (b) for the advanced polar grid. The condition number appears much lower, even at high polynomial orders.



(a) (b)

Figure 3.23: The normalized rank (a) and the order of the condition number (b) for the triangular grid. The condition number is moderate at the higher polynomial orders.



(a) (b)

Figure 3.24: The normalized rank (a) and the order of the condition number (b) for the random grid. The condition number is also moderately high at the high polynomial orders.

Fig. 3.19(a) shows the normalized rank of the DZP system computed on the Cartesian grid. For this system, the transition region between rank-deficient and full-rank appears rather sharp. Fig. 3.19(b) reveals that even when the system has full rank, the condition number increases faster for increasing, $n_{\max}$. The rank and condition number of the polar1 and polar2 grids show similar behavior in Fig. 3.20 and Fig. 3.21. The minimum resolution criteria of the polar3 grid, however, is noticeably higher in Fig. 3.22(a). Despite this higher and broader transition region, the condition number of the polar3 grid, Fig. 3.22(b), is dramatically lower compared to similar locations on all of the other graphs. This suggests that the DZP over the polar3 grid have better orthogonality. The rank of the triangular and random grids, Fig. 3.23(a) and Fig. 3.24(a), show low and sharp transition regions, similar to that of the Cartesian. The condition number of the triangular and random grids, Fig. 3.23(b) and Fig. 3.24(b), however, are noticeably lower.

Grid Type	Minimum Resolution Criteria
Cartesian	$\log_{10} n_p > 1.73 \log_{10} n_{\max} + 0.27$
Polar 1	$\log_{10} n_p > 1.85 \log_{10} n_{\max} + 0.29$
Polar 2	$\log_{10} n_p > 1.85 \log_{10} n_{\max} + 0.29$
Polar 3	$\log_{10} n_p > 2.03 \log_{10} n_{\max} + 0.28$
Triangular	$\log_{10} n_p > 1.78 \log_{10} n_{\max} + 0.11$
Random	$\log_{10} n_p > 1.76 \log_{10} n_{\max} + 0.14$

Table 3.3: The minimum number of pixels, n_p, for the discrete system to be full-rank depends on the maximum order, $n_{\max}$, of the DZP system, and varies by grid type. The minimum resolution criteria for the various grids were similar, except for the advanced polar grid (polar 3) which is slightly higher.

Tab. 3.3 summarizes the minimum-resolution criteria necessary for full-rank DZP systems, but is not sufficient for stability and accuracy. These formulas serve as guideline for selecting the minimum number of points when creating a grid as a function of the maximum polynomial order. The formula for the advanced polar grid is slightly higher and serves as a conservative, single guideline.

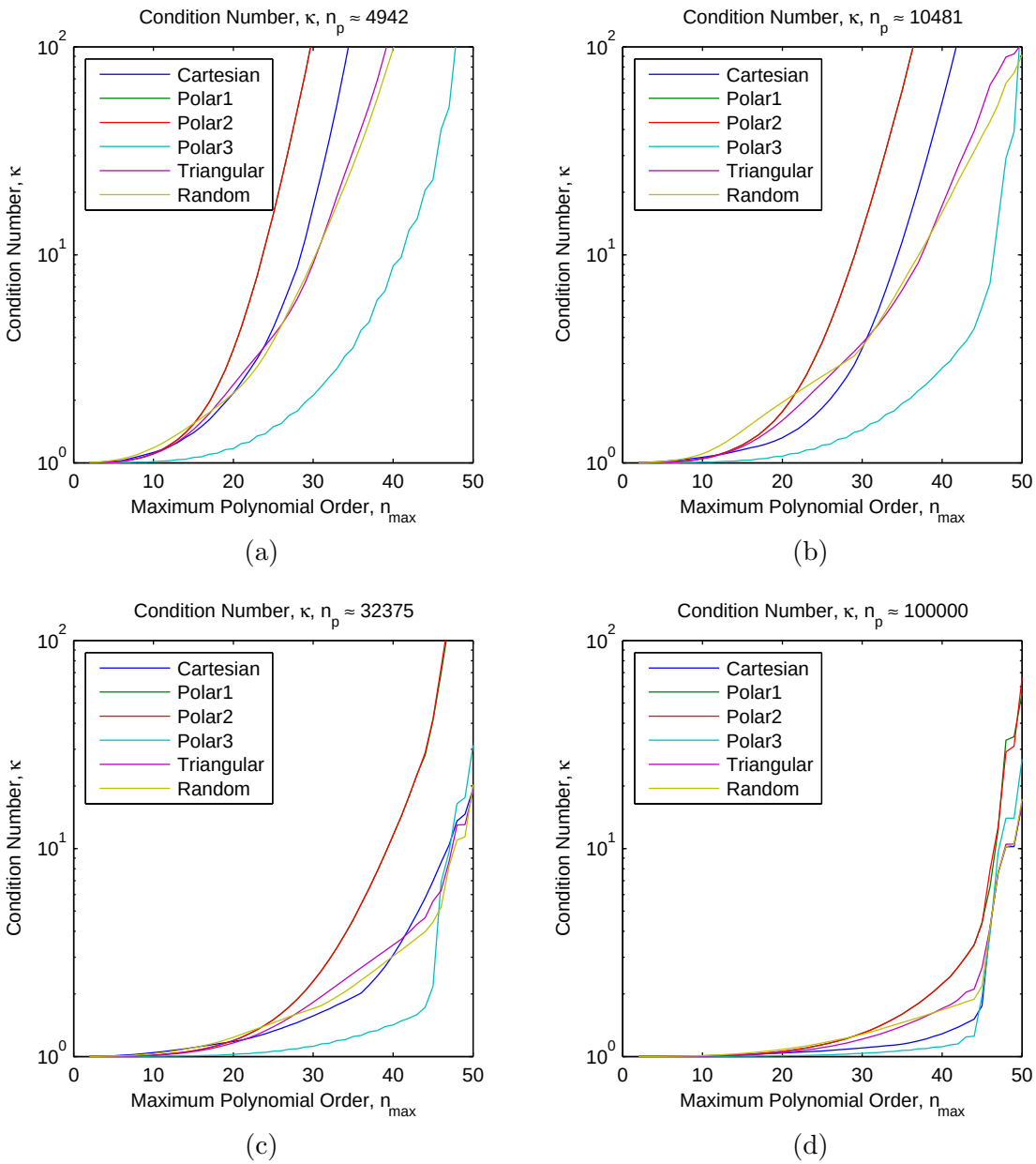
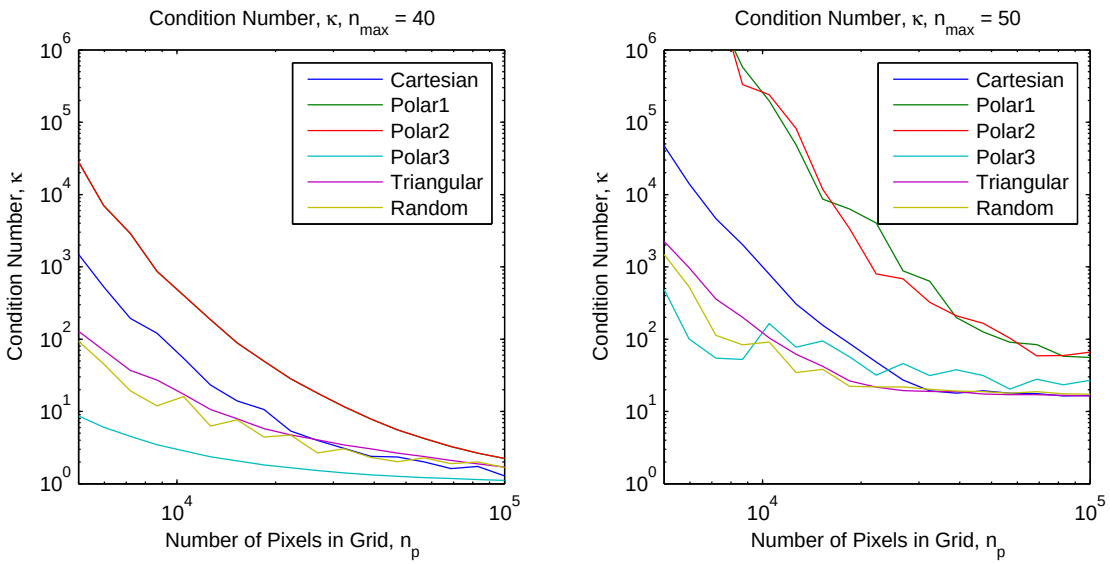


Figure 3.25: The condition number versus the polynomial order for each grid with (a) $\approx 5,000$, (b) $\approx 10,000$, (c) $\approx 33,000$, and (d) $\approx 100,000$, pixels for various orders of DZP. The condition number for the system over the polar 3 grid is typically much lower than the other grids. The results in (d) reveal the barrier in double-precision arithmetic for $n_{\max} > 40$.

The condition number in Fig. 3.25 shows that additional resolution is required for a high quality solution. Utilizing the results from the previous, exploratory computations and selecting a few resolutions, the condition number as a function of the polynomial order is shown for grids with approximately 5,000, 10,000, 33,000, and 100,000 pixels. These resolutions are the total number of pixels over the circular domain. Translating this to common resolutions for an image on a square domain is made by $\sqrt{\frac{4}{\pi}n_p}$ as a rough estimate. This suggests that we would require images with resolutions of 80×80, 116×116, 200×200, and 356×356, which are all modest resolutions by current computing standards. The condition number generally improves for increasing resolutions, as shown in Fig. 3.26. Fig. 3.26(a) also shows that the advanced polar grid is a well-conditioned system for all resolutions, but the Cartesian grid closes the gap at higher resolutions. Fig. 3.26(b) shows that increasing the order beyond $n_{\max} > 40$ all systems loose conditioning. This is a computational barrier imposed by double-precision arithmetic.



(a) (b)

Figure 3.26: The condition number of the DZP systems for $n_{\max}$ = (a) 40, and (b) 50. The condition number decreases as expected for increasing grid resolutions as in (a). The advanced polar grid has the smallest condition number for all these grid resolutions above the minimum-resolution criteria. The condition numbers in (b) are not only significantly higher, but also encounters a computational barrier.

In this section, the rank and condition number of the DZP matrix, $\hat{\mathbf{\Lambda}}$, are computed at various grid resolutions and polynomial orders for each of the grids. An exploratory computation yielded valuable information where the systems had full-rank and are well-conditioned. A minimum-resolution criteria is estimated for all of the grids which established an absolute lower bound on the grid resolution required to solve for the DZP at a selected order. The condition number showed that additional resolution is required for stability and accuracy, and that these resolutions are attainable by current image processing standards. Furthermore, computation of DZP above the 40^{th} order is not recommended. The next section presents several normalized measures to discuss the error convergence in the subsequent sections.

3.6 Error

There are various measures to quantify the error between two images: some are commonly used in the field of image analysis [33, 30, 32, 13], some are used in matrix analysis [19, 26], while others are introduced in this section. The two images are referred to as the reference image $\mathbf{I}^{\text{ref.}}$, and the test image, $\mathbf{I}^{\text{test}}$. The errors are non-dimensional either by their typically formulation or use some property of the reference image to induce the normalization. The images must have the same number of pixels, J, and corresponding pixels, $I_j^{\text{test}} \approx I_j^{\text{ref.}}$.

The normalized, mean absolute error measures the average residual [30], normalized by the range of observed values in the reference image,

$$\varepsilon_{\text{mean}}(\mathbf{I}^{\text{ref.}}, \mathbf{I}^{\text{test}}) = \frac{\frac{1}{J}\sum_{j=1}^{J}\left|I_j^{\text{ref.}} - I_j^{\text{test}}\right|}{\max_j I_j^{\text{ref.}} - \min_j I_j^{\text{ref.}}}. \tag{3.19}$$

The normalized, maximum absolute error measures the maximum residual, normalized by the range of observed values in the reference image,

$$\varepsilon_{\text{max}}(\mathbf{I}^{\text{ref.}}, \mathbf{I}^{\text{test}}) = \frac{\max_j\left|I_j^{\text{ref.}} - I_j^{\text{test}}\right|}{\max_j I_j^{\text{ref.}} - \min_j I_j^{\text{ref.}}}. \tag{3.20}$$

The normalized, root-mean-squared error measures the variation of the residuals [33], normalized by the range of observed, reference values,

$$\varepsilon_{\text{rms}}(\mathbf{I}^{\text{ref.}}, \mathbf{I}^{\text{test}}) = \sqrt{\frac{\sum_{j=1}^{J}\left(I_j^{\text{ref.}} - I_j^{\text{test}}\right)^2}{\sum_{j=1}^{J}\left(I_j^{\text{ref.}}\right)^2}}. \tag{3.21}$$

The peak, signal-to-noise ratio (PSNR) measures the power of a signal versus the power of the corrupting noise (in dB) [32]. PSNR is also used frequently as a metric in analysis of lossy compression algorithms [13], and is given by,

$$\varepsilon_{\text{psnr}}(\mathbf{I}^{\text{ref.}}, \mathbf{I}^{\text{test}}) = 20 \log_{10} \left(\frac{\max_b \mathbf{I}^{\text{ref.}}}{\sqrt{\frac{1}{J} \sum_{j=1}^{J} \left(I_j^{\text{ref.}} - I_j^{\text{test}} \right)^2}} \right). \tag{3.22}$$

where, $\max_b \mathbf{I}^{\text{ref.}}$ is the maximum value for the bit-depth of the reference image (e.g., $255 = 2^b - 1$ for an 8-bit image). When dealing with normalized images, however, this is replaced by the maximum value obtained from the referenced image.

The angle which measures the closeness of fit is typically used in analysis of the solutions of least-squares problems [26]. Given the matrix problem $\mathbf{Ax} = \mathbf{b}$ with the solutions $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$ and $\mathbf{y} = \mathbf{Ax}$, the closeness of fit is given by $\theta = \cos^{-1} \frac{\|\mathbf{y}\|}{\|\mathbf{b}\|}$, where $0 \leq \theta \leq \frac{\pi}{2}$ and all norms are in the 2-norm. In this image analysis, we have,

$$\varepsilon_\theta(\mathbf{I}^{\text{ref.}}, \mathbf{I}^{\text{test}}) = \cos^{-1} \frac{\|\mathbf{I}^{\text{test}}\|}{\|\mathbf{I}^{\text{ref.}}\|} = \sin^{-1} \frac{\|\mathbf{I}^{\text{ref.}} - \mathbf{I}^{\text{test}}\|}{\|\mathbf{I}^{\text{ref.}}\|}. \tag{3.23}$$

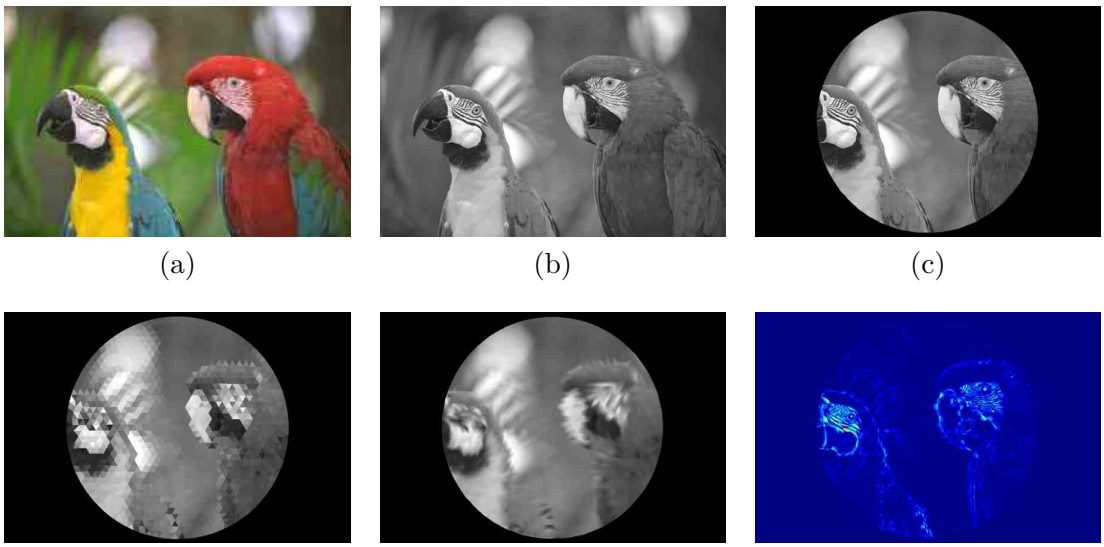
An additional measure is how a solution $\|\mathbf{y}\|$ falls short of the maximum possible value [26], $\eta = \frac{\|\mathbf{A}\|\|\mathbf{x}\|}{\|\mathbf{y}\|}$, where $0 \leq \eta \leq \kappa(\mathbf{A})$. In this particular analysis, it becomes,

$$\varepsilon_\eta(\mathbf{I}^{\text{ref.}}, \mathbf{I}^{\text{test}}) = \frac{\|\hat{\mathbf{\Lambda}}\|\|\tilde{\boldsymbol{\lambda}}\|}{\|\mathbf{I}^{\text{test}}\|}. \tag{3.24}$$

In this section, various error measures have been presented for later use in the analysis of the DZP on an associated grid. The error types are: mean, max, RMS, PSNR, θ, and η. Each of these error measures different qualities about the residuals between a test image and the reference image. These error formulations are utilized in the following sections to examine the convergence of the resampled images on the new grids and the reconstructed images from the FZS solutions.

3.7 Image Resampling

Introducing a grid with a new set of pixels requires interpolation of the pixel values in the original or source image. It has already been shown that the convergence rate of the re-sampled image is proportional to the size of the new pixels to the order of the interpolant up to some constant which depends on the smoothness of the original image [35]. The goal of this section is to quantify the errors, as a function of the grid resolution n_p, introduced by re-sampling onto a grid, and subsequently re-sampling back to the original image domain all using linear interpolation. Also of interest is how the use of a smoothing pre-filter can improve the re-sampling process, when fine details from the image are not required.



(a) (b) (c)

(d) (e) (f)

Figure 3.27: The re-sampling process to produce the approximated image. The original image in (a) is: (b) converted to grayscale, (c) cropped, (d) re-sampled onto the triangular grid, and (e) re-sampled back to the original pixels. The absolute value of the residual between (c) and (e) shows in (f) that regions of high-contrast and high-detail lose the most information, as expected.

Fig. 3.27 illustrates the steps necessary to examine the re-sampling process. The original color image in (a) is first converted to the grayscale image in (b) using the formula, $I_j = 0.2989R_j + 0.5870G_j + 0.1140B_j$, for the j^{th} pixel, where R, G, B are the red, green, and blue channels [18]. The image in (b) is circularly cropped within the radius $\rho_0 = 250$, to obtain the image in (c), retaining only 196,231 pixels of the original 393,216, or 50%. Using linear interpolation, the image in (d) is computed from the pixel values in (c); in this case, a triangular grid with only $n_p = 2,000$ pixels, or 1% of the 196,231. Those values are then used to produce the approximated image in (e). The residual image in (f) is the absolute value of the difference between (c) and (e) rendered in pseudo-color. The errors are then computed for a series of standard images utilizing this technique. Fig. 3.28 shows the 24 images utilized in this analysis [15].

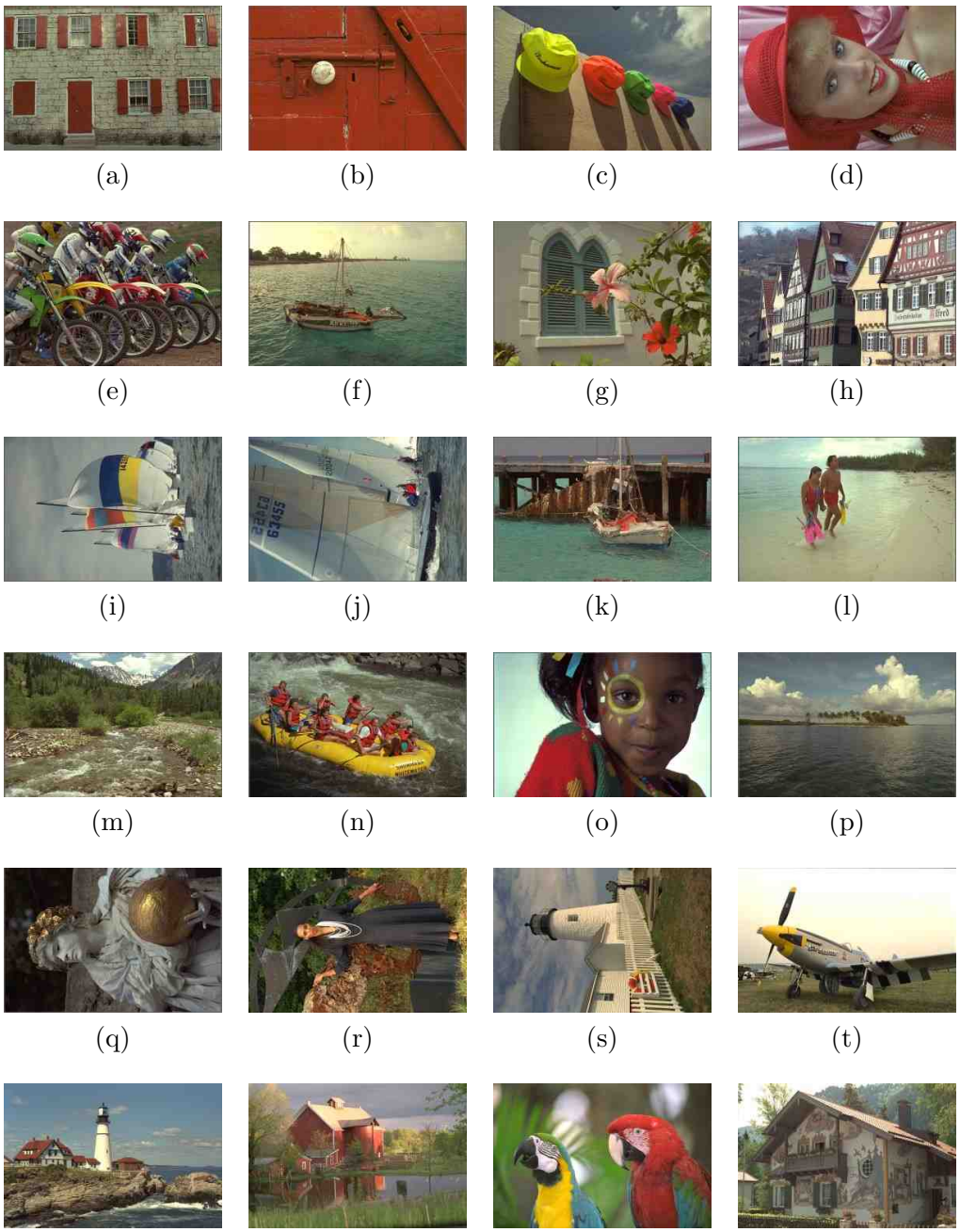


Figure 3.28: Standard test images used for the re-sampling analysis. The original images are 768×512, 0.39 Mpixel, 24-bit color, with lossless compression.

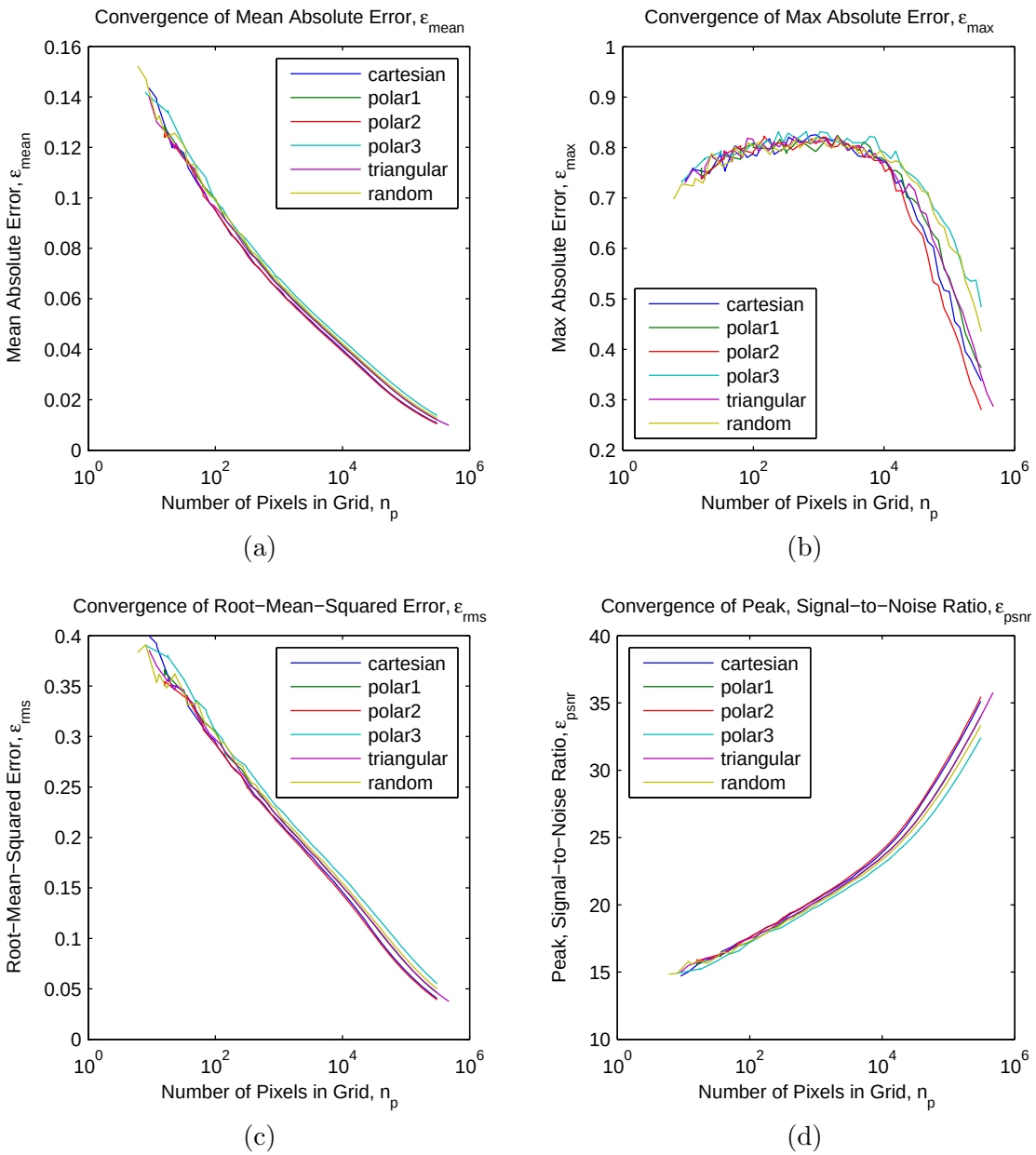


Figure 3.29: The errors associated with re-sampling onto a grid and back decrease for the mean and RMS errors (a) and (c) with increasing grid resolutions, as expected, but do not vanish. The max error and PSNR (b) and (d) show sudden improvements near $n_p \approx 10^4$.

For increasing grid resolutions, the errors over the entire image decreases, as expected. Fig. 3.29 shows the results of the resampling errors. The results are averaged over the entire set of test images for each of the grid types. The grids with uniform pixel areas (i.e., Cartesian, polar 2, and triangular) perform slightly better than the others. This is presumably because constant pixel area preserves the original resolution evenly over the entire region, while the larger and/or odd shaped pixels of the other grids tend to lose detail. The errors are measured up to $n_p \approx 316,000$, higher than the original $\approx 200,000$ pixels in the original images. This demonstrates that even at oversampled grid resolutions, the process of re-sampling introduces non-trivial errors. The results also show that the different measures of error are relatively consistent between grid types.

(a) (b)

Figure 3.30: An original image in grayscale (a), and the smoothed version in (b) using a Gaussian, low-pass filter with standard deviation of 4 pixels (1.6% of ρ_0).

All the errors associated with the re-sampling process rapidly decrease by discarding the details of the original image with the use of a pre-filter. A common tool for smoothing is the 2-dimensional convolution of the image with a Gaussian, low-pass filter (GLPF), as shown in Fig. 3.30 [10]. This particular image is smoothed with a discrete Gaussian kernel of standard deviation, $\sigma = 4$ pixels, representing only 1.6% ($\frac{\sigma}{\rho_0} = 0.016$) of the aperture radius, $\rho_0 = 250$ pixels.

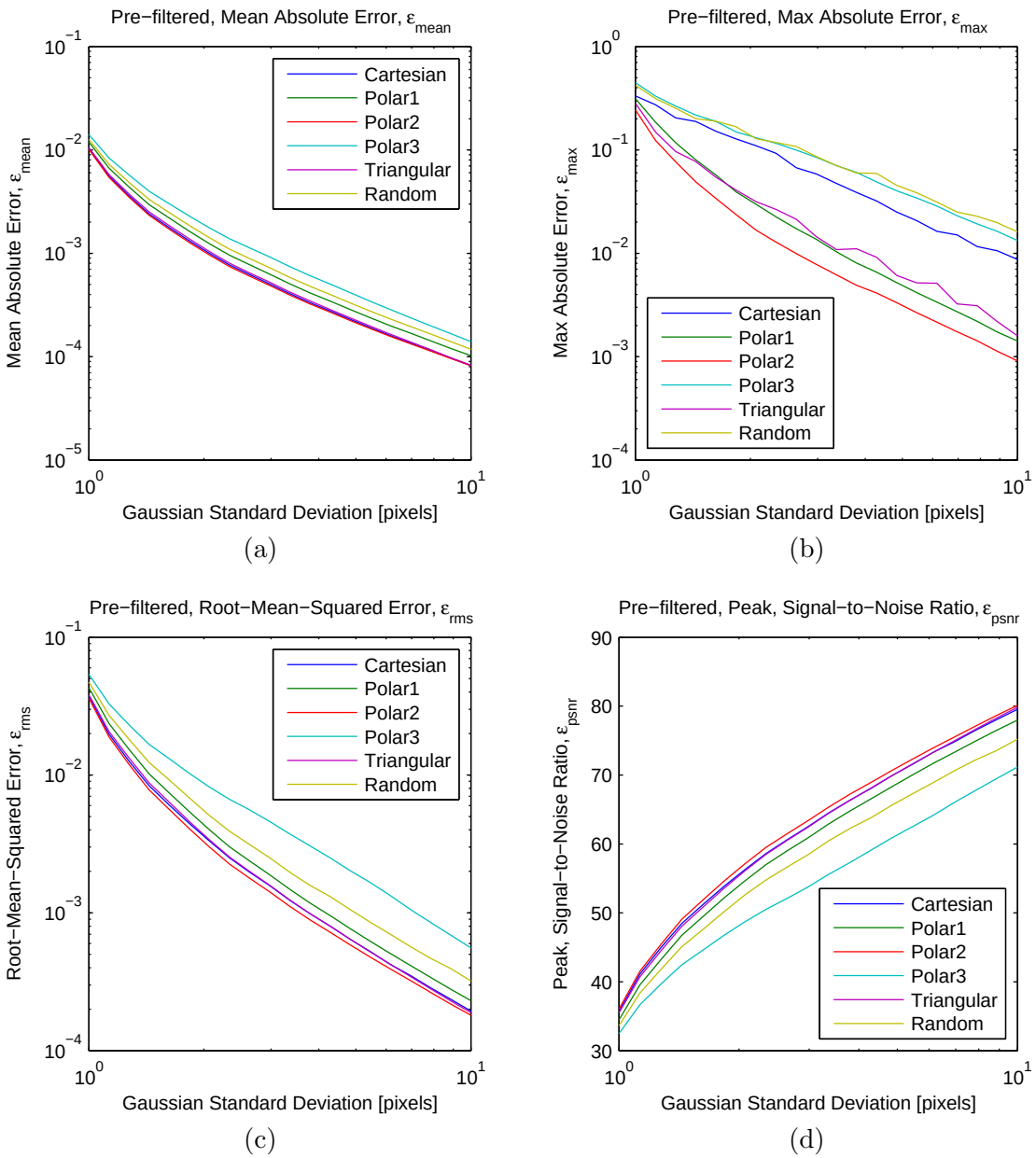


Figure 3.31: The mean, max, RMS and PSNR errors in (a) through (d) all indicate a rapid decrease in the residuals for increasing GLPF kernel size.

To examine the effect that smoothing has on the errors, the grid resolution is fixed, $n_p \approx 10^5$, while the size of the smoothing kernel varies. The results are averaged from all of the 24 test images. The size of the GLPF varies from 1 to 10 pixels or 0.4 to 4.0% of ρ_0. Fig. 3.31 shows that the errors decrease rapidly with increase smoothing. This shows the errors from resampling a smoothed image and does not include the error from smoothing the original image. Even for a slight amount of smoothing, the errors decrease by an order of magnitude. The grids with uniform-area pixels (Cartesian, polar 2, and triangular) obtain slightly better performance.

In this section, the process of re-sampling onto the grids from Sec. 3.2 was shown to introduce errors. These errors are non-trivial even for grids at higher resolution than the original image. If the image resolution is sufficient to support a well-conditioned DZP system, it is reasonable to use the original Cartesian grid from the original image and to avoid the resampling errors. Smoothing the image prior to re-sampling dramatically reduces the errors by discarding fine details before the resampling process. The next section computes the FZC and examines the convergence as a function of various parameters.

3.8 Fourier-Zernike Coefficients

The Fourier-Zernike coefficients (FZC) represent the contribution of each Zernike polynomial (ZP) in the decomposition of a function over the DZP basis. After an examination of the rotational invariance of the FZC, the goal of this section is to examine the convergence of the FZC computed using the DZP as a function of the grid type, number of pixels in the grid, the maximum polynomial order, and the smoothing pre-filter.

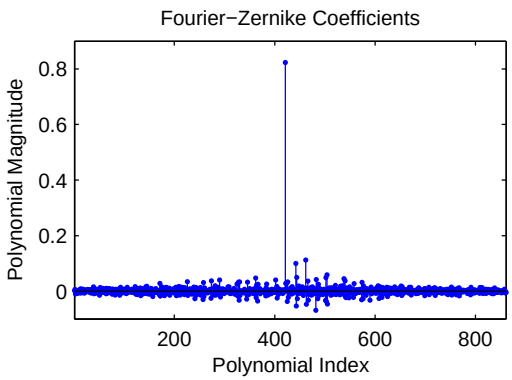
From Tab. 3.3 on p. 59, the minimum number of pixels in a grid, n_p, as a function of the maximum polynomial order, $n_{\max}$, for the DZP system, $\hat{\mathbf{\Lambda}}$, to have full-rank is given by the conservative formula, $\log_{10} n_p = 2.03 \log_{10} n_{\max} + 0.28$. If $n_{\max} = 40$ then $n_p > 3,406$. Furthermore, for most of these systems to be well-conditioned, $\log_{10} \kappa(\hat{\mathbf{\Lambda}}) < 2$, the resolution should be even higher. According to Fig. 3.26 on p. 62, the resolution should be $n_p > 10^4$. This requirement is observed for all FZC computations in this section.

The FZC is the solution to the system given in Eq. 3.8 on p. 36, $\tilde{\boldsymbol{\lambda}} = \hat{\mathbf{\Lambda}}^{-1}\hat{\mathbf{f}}$. To reiterate, the double-indexed ZP are transformed into a single-indexed system by sorting the polynomials first by the repetition number, m, then by the order, n. This essentially groups the FZC into subsets of identical repetition which are then sorted by their order. Understanding the order of the polynomials is essential for viewing how the FZC relate to a source image.

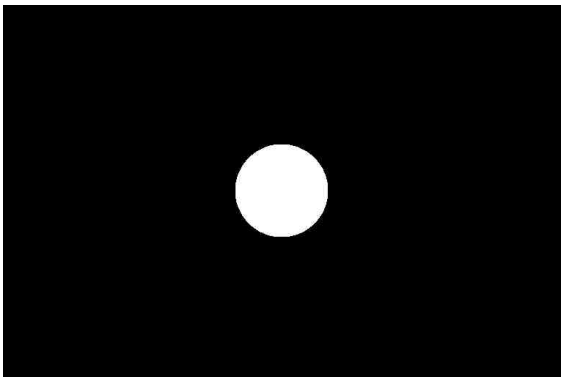
Some basic observations can be made directly from a plot of the FZC. This is illustrated in Fig. 3.32 for several images and their associated FZC representation (computed with grid type = polar3, $n_p = 10^5$, $n_{\max} = 40$). The first image in (a) shows a detailed test image and its complex representation in (b). The FZC has a very high low-order component, Λ_0^0, which is typical for most images, but also has significant high-order components elsewhere. This indicates that the source image requires significant detail, as expected; probably more than 40^{th} order. The next image of a disk in (c) requires significantly fewer FZC components (d), specifically, only the pure radial ZP grouped in the center, $m = 0$. The third image (e) shows a centered, symmetric array of lines. The FZC components in (f) appear generally only in the positive and even ZP due to the orientation of the array relative to the ZP and the 2-fold symmetry of the array. If the array were rotated, the FZC would also show components in the negative and even ZP. The FZC representation, however, is still unique from the rotational ivariance property.

The FZC should not only remain stable for rotations but should also should converge for refinements in the grid and added smoothness on the function. As the number of pixels in the grid increases, the FZC solution should tend toward higher accuracy. Additionally, as the number of polynomials increases, the previous, low-order FZC should agree with the low-order FZC computed from the larger system. These properties are particularly attractive features which allow progressive refinement of a ZP analysis and agreement with computations from various sources. To examine the convergence and stability, the measures of error described in Sec. 3.6 can also be employed here by utilizing the highest accuracy solution as the reference data. The RMS and PSNR errors, however, only provide unbiased comparisons if the RMS of the signal doesn't change for varying sample sizes.

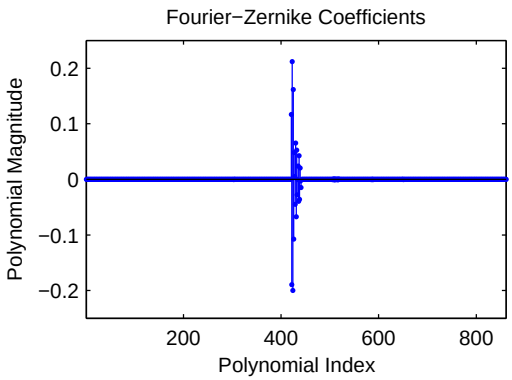
(a)



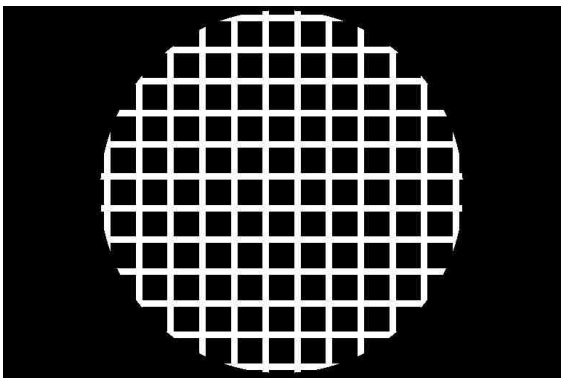
(b)



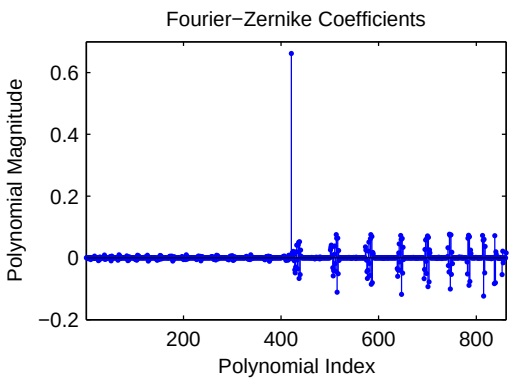
(c)



(d)



(e)



(f)

Figure 3.32: Some test images used to illustrate the associated FZC. A highly detailed image in (a) has significant FZC components in (b), especially the Λ_0^0 term. The disk in (c) has only pure radial components ($m = 0$) in (d). The symmetric grid in (e) only requires positive and even polynomials in the FZC of (f).

The DZP are invariant to rotations about the origin which is readily observed using the complex, ZP. These are related to the real, ZP as in the following,

$$\begin{aligned} \Lambda_n^{\mu} &= \Lambda_n^{+m} + i\Lambda_n^{-m}, \\ \Lambda_n^{0} &= \Lambda_n^{0}, \end{aligned} \tag{3.25}$$

where, $\mu \geq 0$ and $|\mu| = |m|$. The magnitude and phase of each polynomial is then computed as usual,

$$\begin{aligned} |\Lambda_n^{\mu}| &= \sqrt{(\Lambda_n^{+m})^2 + (\Lambda_n^{-m})^2}, \\ \angle\Lambda_n^{\mu} &= \tan^{-1}\frac{\Lambda_n^{-m}}{\Lambda_n^{+m}}. \end{aligned} \tag{3.26}$$

Using the magnitude of the complex, ZP, a comparison of the FZC can be made as the grid rotates about the source image at random angles, $0 \leq \phi \leq 2\pi$ radians. This analysis examines all grid types, a constant grid resolution $n_p = 10^5$, a single polynomial order $n_{\max} = 40$, averaged over several images with no smoothing. The errors were computed between the FZC of a reference grid, $\phi = 0$, and the FZC of the appropriate rotated grids. Fig. 3.33 shows the (a) mean error, (b) max error, (c) RMS error, and (d) the PSNR. The general trend shows that the error is not a function of the rotation angle. At a few special angles where the grids have perfect symmetry, however, the errors vanish down to working precision. Overall, the polar2 and triangular grids showed slightly less errors, indicating better rotational invariance.

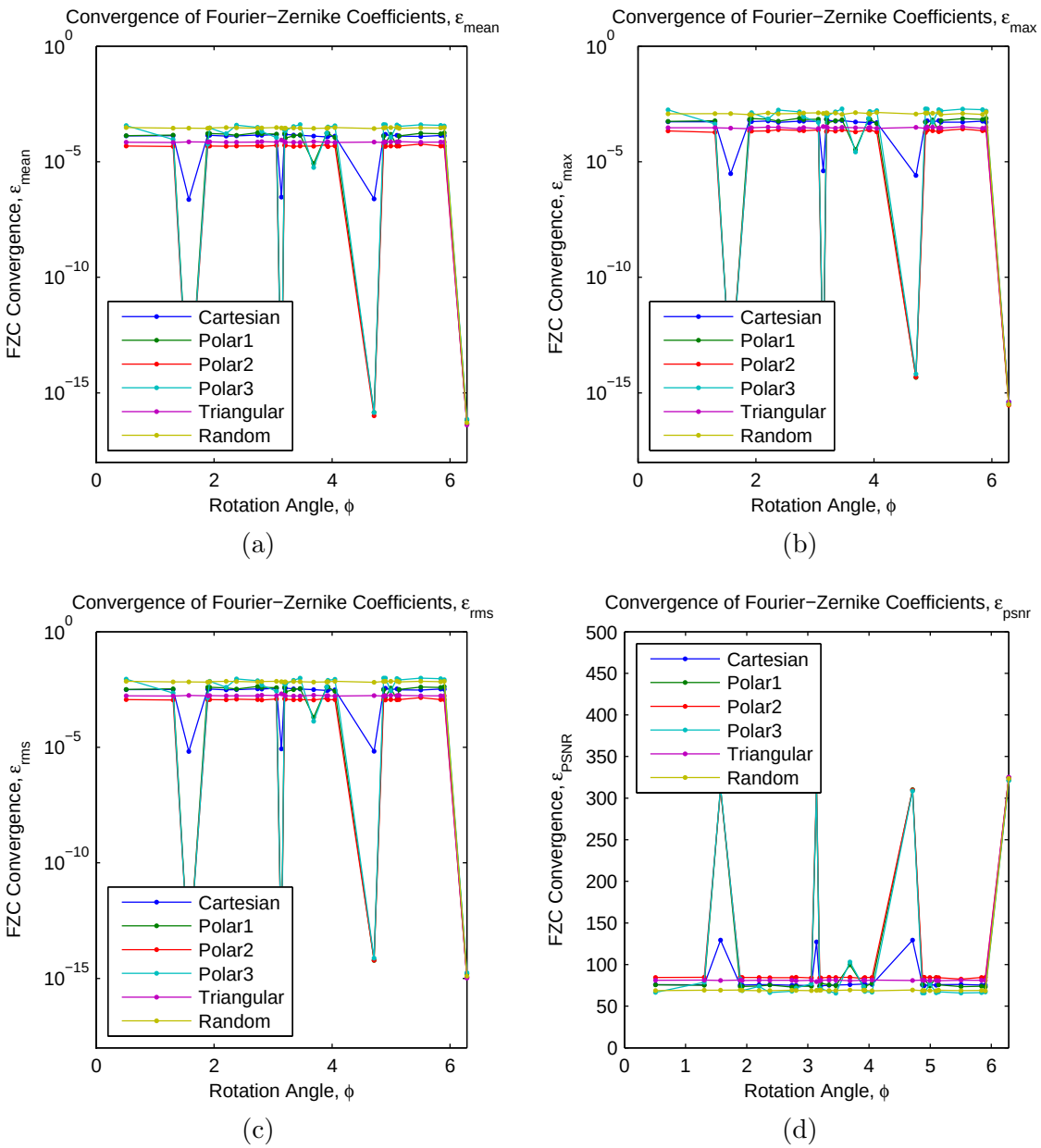


Figure 3.33: The (a) mean, (b) max, (c) RMS, and (d) PSNR errors show no dependence on the rotation angle, ϕ. The errors vanish on the polar grids for $\phi = 0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}, 2\pi$ radians. The polar2 and triangular grids had slightly less errors on average.

The FZC, $\tilde{\boldsymbol{\lambda}}$, of several images are computed at various grid resolutions, $1 \times 10^4 \leq n_p \leq 2 \times 10^5$, for each grid type. The polynomial order is fixed for all cases, $n_p = 40$, providing 861 coefficients. The error is computed between the FZC at the highest resolution as the reference data and the FZC at the other resolutions as the test data. The errors derived from each image are then averaged to improve the consistency of the error trends. Fig. 3.34 shows how the FZC converge for higher grid resolutions. The trends in all of the mean error (a), max error (b), RMS error (c), and PSNR error (d) show nearly identical behavior; the residual errors decrease, as expected. High-resolution grids show modest improvements the total accuracy of the FZC. These trends also show that while the advanced polar grid (polar3) does not improve as quickly as the other grids, it does begin at the lower resolutions with less error than any of the other grids. The FZC should not only converge with high-resolution grids, but should also be stable as the maximum polynomial order increases.

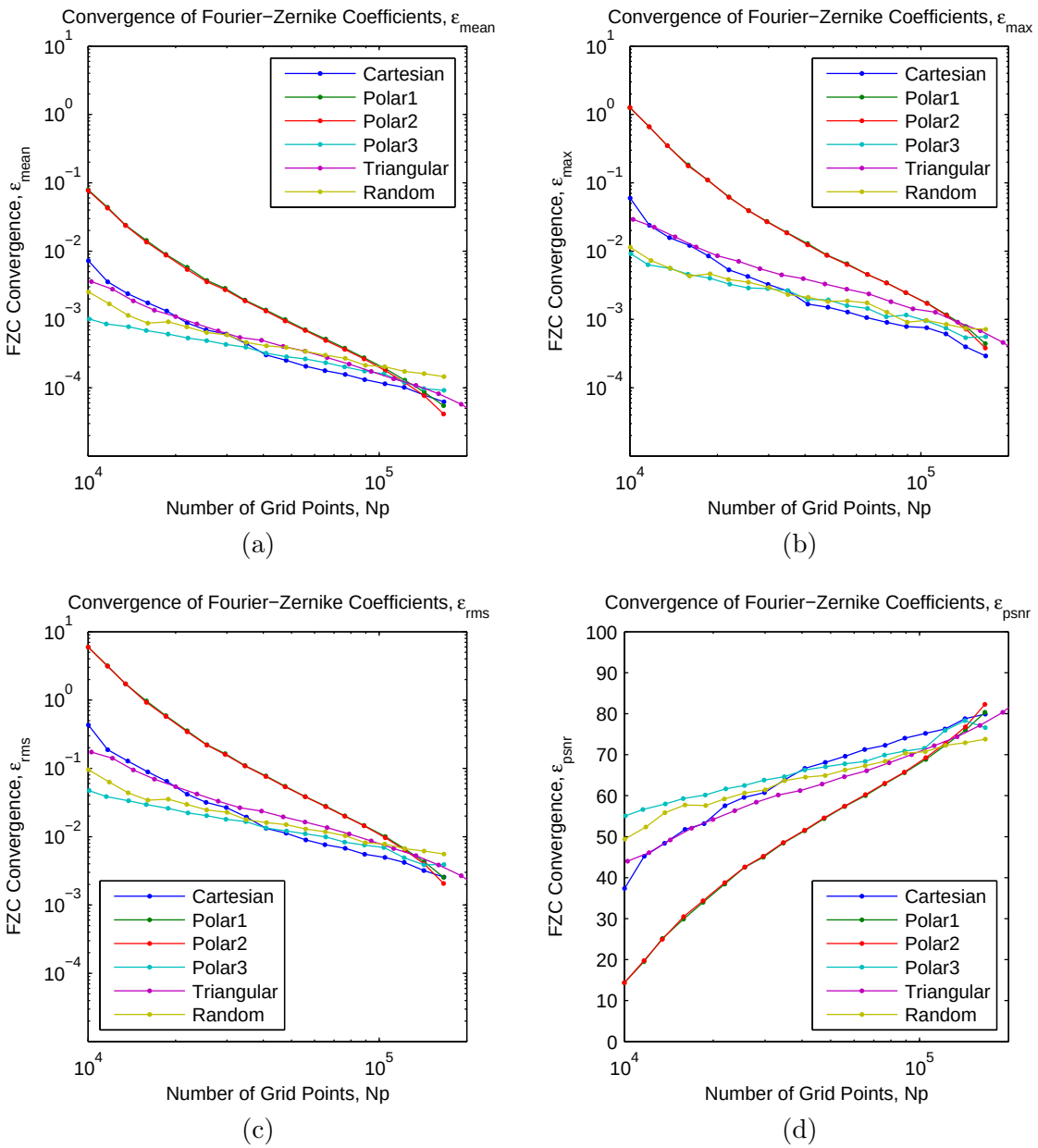


Figure 3.34: The trends in the mean error (a), max error (b), RMS error (c), and PSNR error (d) all show similar results: the FZC solutions improve for high grid resolutions and the polar3 grid has a slight advantage toward the lower resolutions.

Higher polynomial orders in an orthogonal system should increase the fidelity of the approximation progressively; adding fine details with high-order coefficients while preserving course details with the low-order coefficients previously computed. To examine this, the FZC, $\tilde{\boldsymbol{\lambda}}$, were computed for several images at various maximum polynomial orders, $2 \leq n_{\max} \leq 40$, for each grid type at a fixed resolution, $n_p \approx 1 \times 10^5$. The errors from each image are averaged together to improve the continuity of the results. Since the number of coefficients changes as a function of $n_{\max}$, the errors are carefully computed between the coefficients at a lower order and only their respective counterparts in the 861 coefficients at the maximum order, 40.

The errors are plotted in Fig. 3.35 for the mean error (a), max error (b), RMS error (c), and PSNR in (d). The general trend in the mean and max errors show that the FZC solutions are relatively stable or slightly increase for higher orders. Some increase in the errors is expected since the condition number increases for increasing orders at fixed grid resolutions, previously seen in Fig. 3.25(d) on p. 60. The polar3 grid shows the least amount of error followed by the Cartesian, random, triangular and then polar1 and polar2. This order is consistent with the the order observed in the condition numbers of the respective systems, again in 3.25(d). The RMS error increases at higher orders while the PSNR signal decreases, both indicating a loss of accuracy. The validity of these calculations is questionable, however, since both utilize RMS calculations of the residuals. Specifically, the FZC at higher orders are typically much smaller than those at low orders; effectively lowering the signal's RMS value which is subsequently used to normalize the error. This artificially increases the RMS error and lowers the PSNR for increasing polynomial orders.

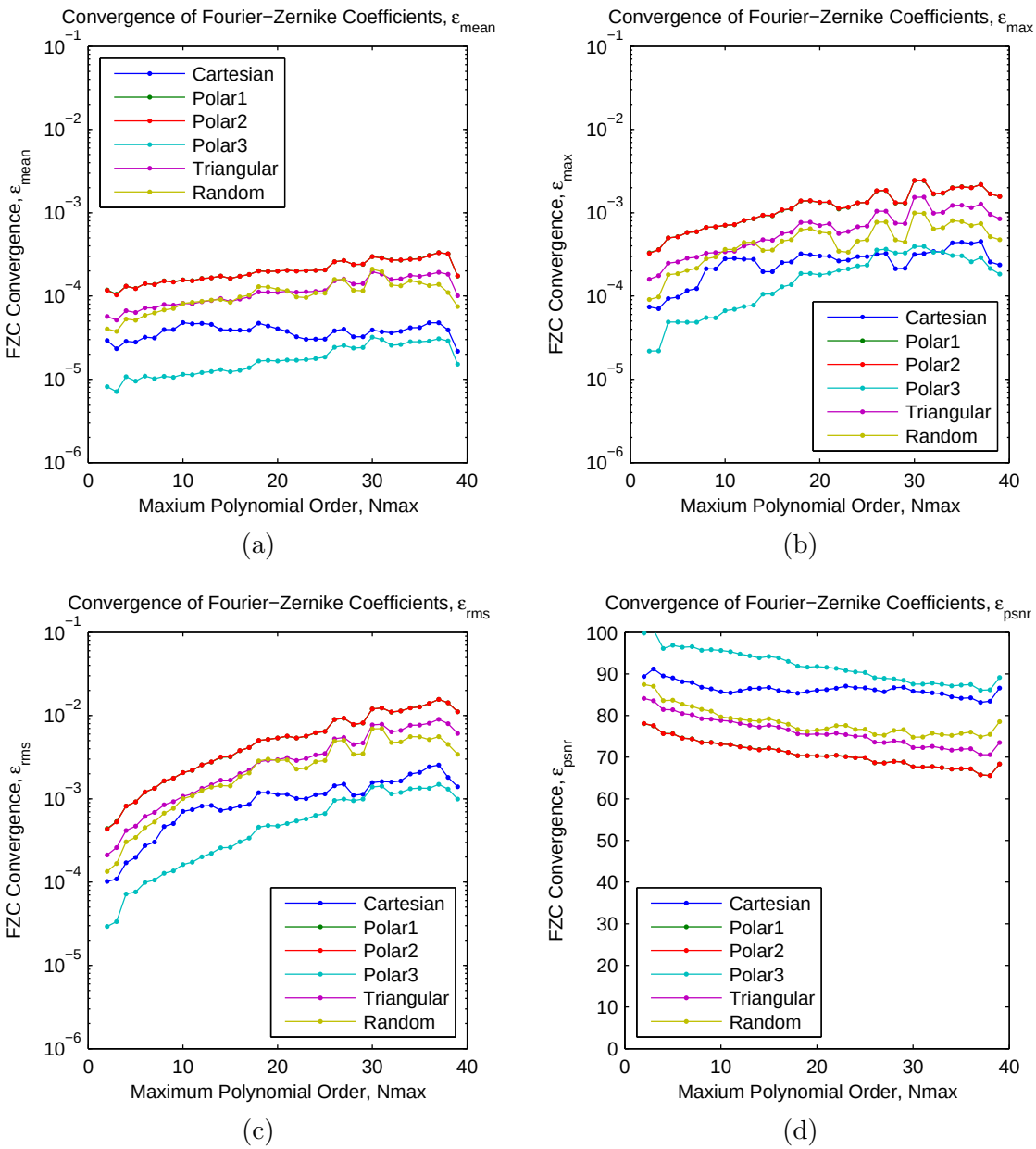


Figure 3.35: The mean error in (a) and max error in (b) of the FZC for increasing polynomial orders increases slightly, but is expected for the increasing condition numbers also as the polynomial order increase. The RMS error (c) and PSNR in (d) both indicate a loss of accuracy but are questionable calculations since the normalization factor decreases for higher orders, causing an apparent loss in accuracy.

A common technique in sampling theory is the utilization of an anti-aliasing filter. In this analysis, a Gaussian, low-pass filter smooths the image prior to sampling the signal of interest [16]. This limits the spectral bandwidth and reduces the ringing artifacts of the Gibbs-Wilbraham phenomenon. In this application, the original image is filtered, prior to both resampling of the grid and computation of the FZC. The convolution of the original image with the GLPF kernel smooths the image in 2 dimensions [10]. The single parameter which controls the amount of smoothing is the standard deviation of the Gaussian, σ. The kernel size is normalized in this application by the disk radius to provide a non-dimensional measure. The GLFP reduces the high-frequency spectral components of a signal in a Fourier analysis, and, similarly reduces the high-order FZC components in this Fourier-Zernike analysis. Fig. 3.36 shows the magnitude of each FZC (in logarithmic scale) for various levels of smoothing. The polynomials are sorted by order, n, then by repetition number, m. The low-order polynomials are to the left. The red vertical lines indicate that the FZC remain relatively constant for increasing kernel sizes. The high-order FZC remain relatively constant prior to $\frac{\sigma}{\rho_0} > 3 \times 10^{-2}$, but decrease after that.

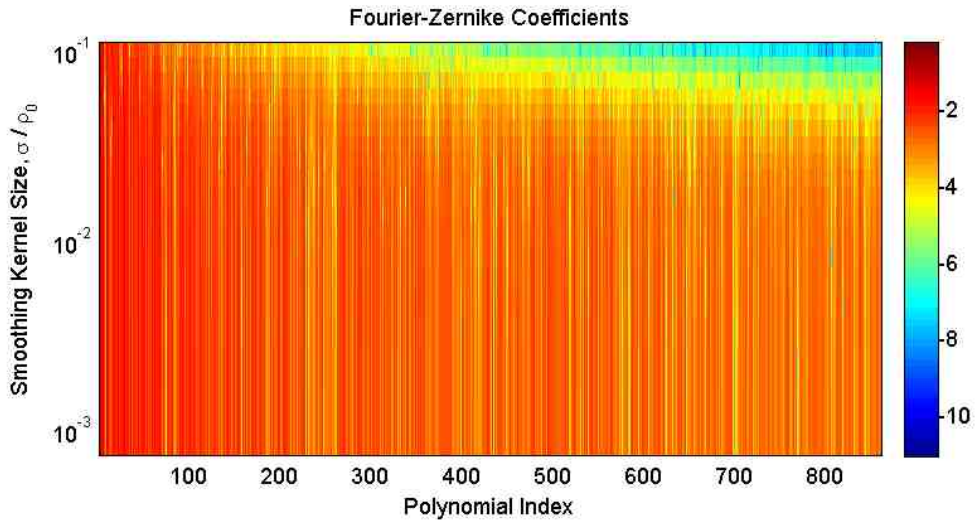


Figure 3.36: A spectrograph of an image shows the separated FZC for various amounts of smoothing. For heavy smoothing, the high-order FZC tend toward zero (blue, negative numbers in the logarithmic scale) while the low-order coefficients remain relatively stable (red vertical lines).

Examining the numerical errors introduced by smoothing reveals changes in the FZC at all levels of smoothing. The FZC are computed over a single grid type, polar3; fixed resolution, $n_p = 10^5$; various levels of smoothing, $0 \leq \frac{\sigma}{\rho_0} \leq 10^{-1}$; and averaged over several of the test images. The errors are computed between the reference FZC with no image smoothing and the FZC at various levels of smoothing. The results are averaged from the various images. Fig. 3.37 shows that the smoothing process introduces error at all levels of smoothing (the minimum smoothing is limited by the pixel size in the original image, except for the case of no smoothing). The relationship appears nearly linear, and is similar for other grid types not shown.

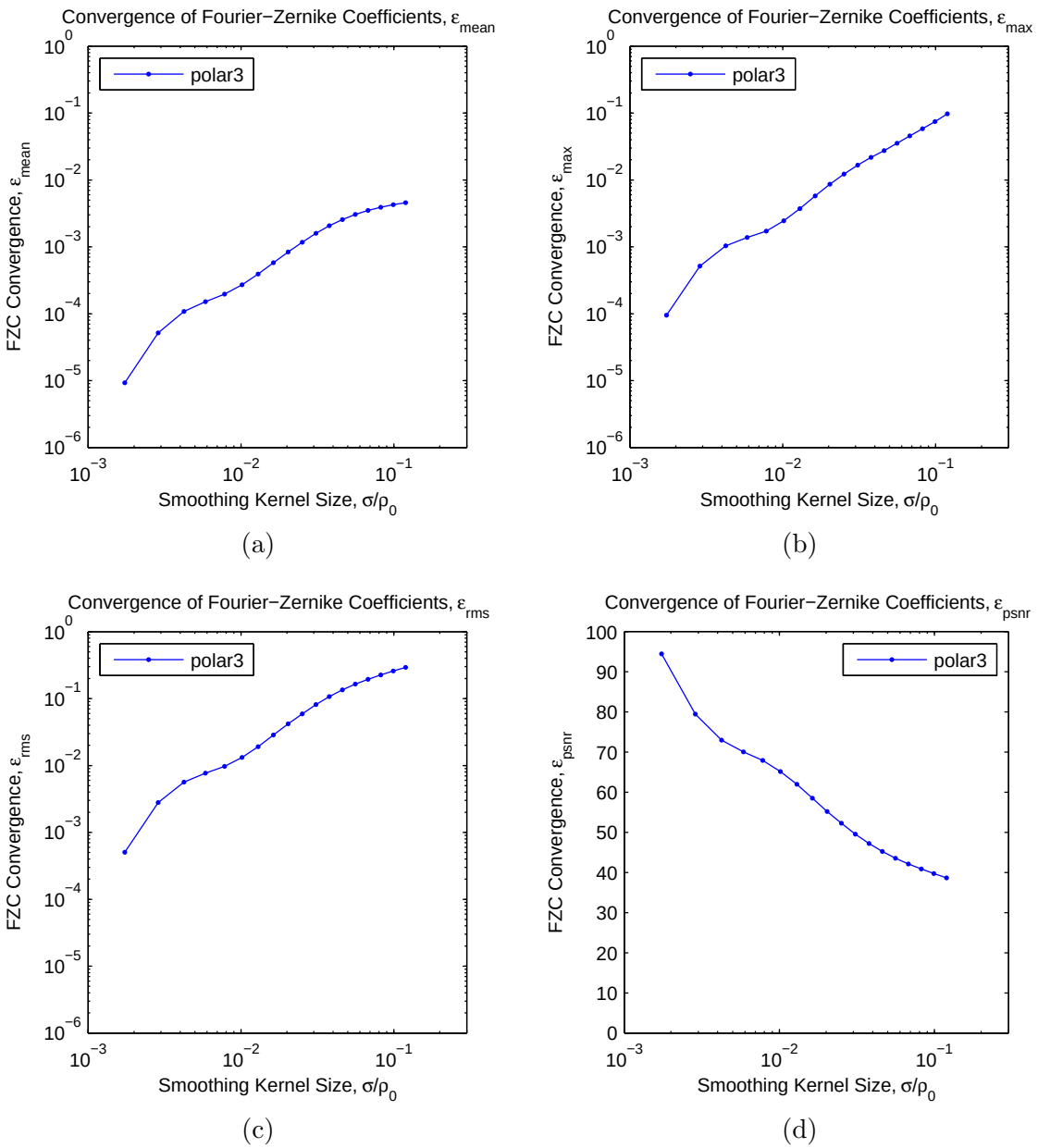


Figure 3.37: All errors indicate that smoothing an image necessary alters the FZC, as expected.

The FZC are computed for various test images and the convergence and stability are examined as a function of the grid rotation, grid type, grid resolution, polynomial order, and image smoothing. The rotational invariance of the FZC is confirmed for all grid types. The FZC converge by 1 to 2 orders of accuracy for increasing grid resolutions, $10^4 \leq n_p \leq 10^5$. This improvement is similar to the gains previously observed in the condition numbers at higher resolutions (Fig. 3.26 on p. 62). At fixed grid resolutions, the FZC are relatively stable for increasing polynomial orders; losing some accuracy, less than 1 order. This is also identical to the behavior of the condition number as previously seen in Fig. 3.25 on p. 60. Smoothing the image changes the FZC of an image, observed here as errors, and suppresses high-order FZC. An additional examination of the reconstruction errors, as in the following section, is needed to determine how these changes in the FZC effect FZS. The next section utilizes the FZC and the DZP to construct the FZS and reconstruct the images.

3.9 Fourier-Zernike Series

This section examines image reconstruction from the FZS and the associated errors as a function of the analysis parameters. The formulation for the FZS is discussed first followed by a few examples. The errors between the source image and the approximated image are then examined for both the original image resolution and the interpolated grid resolution. This provides a measure of the total error, which includes interpolation errors, and the solution error, only from the FZS. The errors are computed for all grid types and averaged over several images. The relationship is observed between the errors as a function of the grid resolution, n_p, the maximum polynomial order, $n_{\max}$, the smoothing filter size, $\frac{\sigma}{\rho 0}$, and the rotation angle, ϕ.

From Eq. 3.8, the solution of the system for the FZC is $\tilde{\boldsymbol{\lambda}} = \hat{\boldsymbol{\Lambda}}^{-1}\hat{\mathbf{f}}$. These coefficients are then used to reconstruct the original function (image), $\tilde{\mathbf{f}}$, in the least-squares sense. The approximate function is given by,

$$\hat{\boldsymbol{\Lambda}}\tilde{\boldsymbol{\lambda}} = \hat{\boldsymbol{\Lambda}}\left(\hat{\boldsymbol{\Lambda}}^{-1}\hat{\mathbf{f}}\right) = \hat{\tilde{\mathbf{f}}}. \tag{3.27}$$

This relates the value at the j^{th} pixel to the summation over all of the polynomials multiplied by the appropriate coefficient,

$$\begin{aligned} \hat{\tilde{f}}_j &= \sum_{k=1}^{K} \hat{\Lambda}_{jk}\tilde{\lambda}_k, \\ \tilde{f}_j &= \frac{\hat{\tilde{f}}_j}{\sqrt{\omega_j}}. \end{aligned} \tag{3.28}$$

The grid used to find the approximate FZC, however, does not have to be the same grid used to evaluate the FZS. This analysis does not explore mixing grid types.

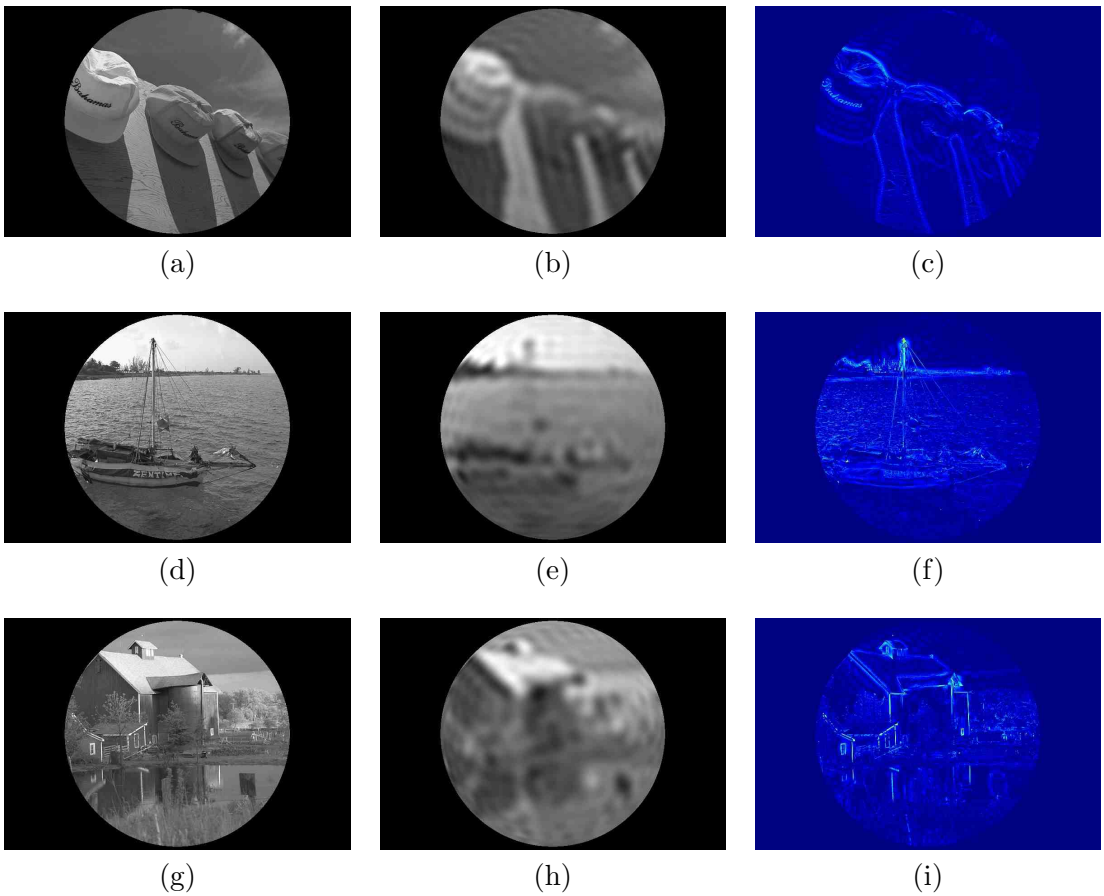


Figure 3.38: Three test images (a), (d) and (g) were approximated by their FZS (b), (e) and (h). The residuals (c), (f) and (i) show that the approximations discard fine details and introduce ringing.

Fig. 3.38 shows the results from approximating several images with the DZP. The images on the left are the source images while the images in the center are their approximated counterparts. The approximations roughly resemble the originals but show a loss of fidelity similar to blurring. The images on the right show the residuals in pseudo-color where dark-blue represents very little difference and red indicates substantial differences. Overall, the results appear reasonably accurate except for regions of high detail or sharp transitions. The ripples are the manifestation of the Gibbs-Wilbraham phenomenon.

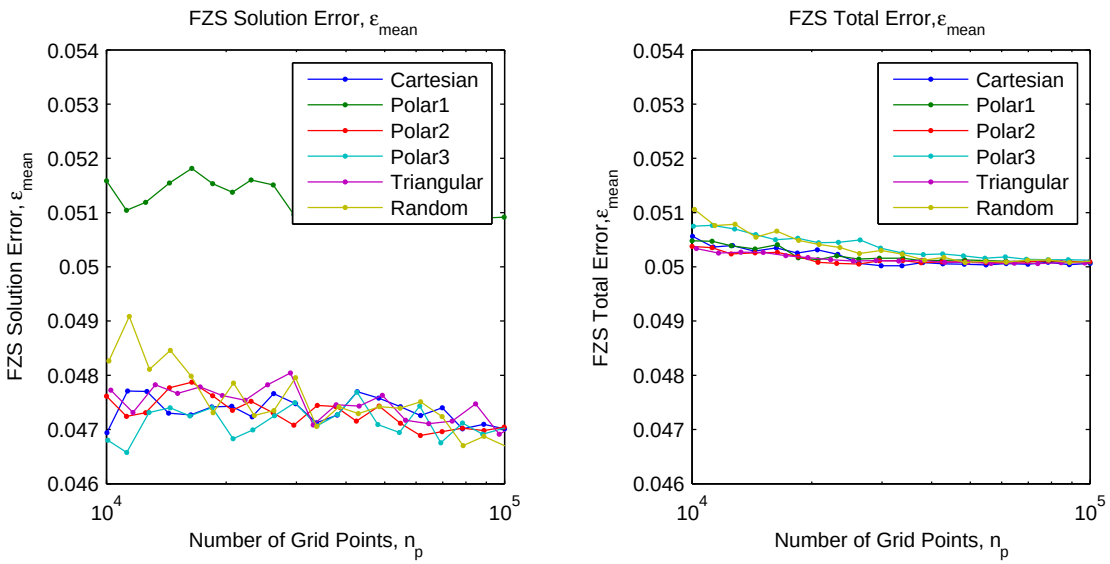
(a) (b) (c) (d) (e) (f) (g) (h)

Figure 3.39: The progressive, linear combination of the FZC from two images is preserved in the reconstruction of FZS.

An invaluable property from Eq. 3.28, is the linearity. A linear combination of FZC is preserved in the FZS. Fig. 3.39 illustrates a progressive linear combination of the FZC between two source images.

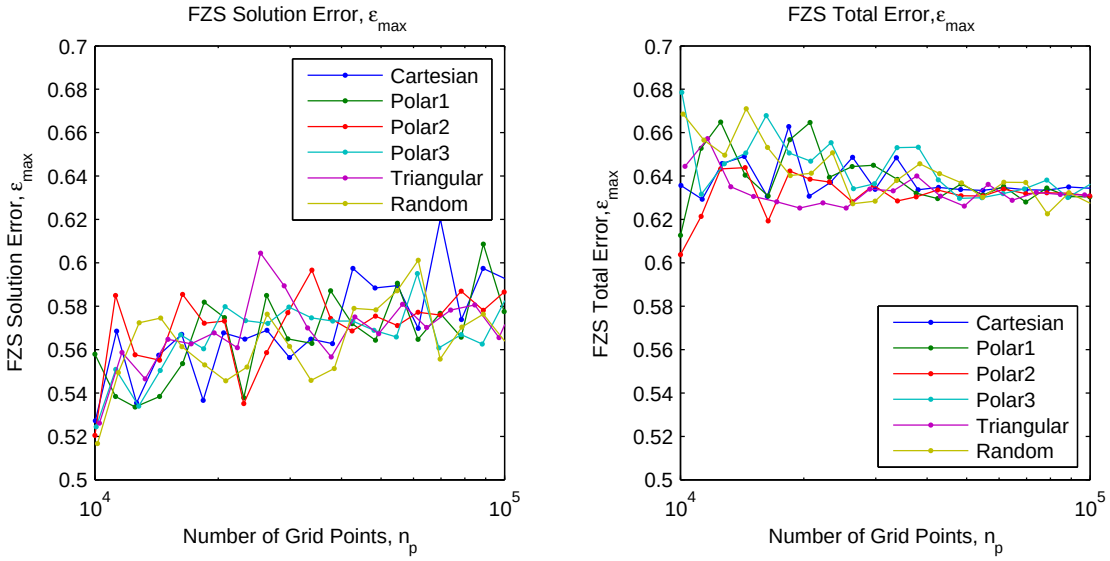
The following stages are performed to approximate an image: (1) circularly crop the image and normalize the amplitude, (2) smooth the image if required, (2) generate the desired grid, (3) interpolate the values in the grid from the values in the image, (4) calculate the DZP on the grid, (5) solve the system for the FZC, (6) evaluate the FZC and DZP to find the FZS on the grid, (7) create the approximated image by interpolating values from the grid. During these operations, there is two opportunities to examine the errors: between the values in the grid, which measures the errors in the solution, and between the values in the image, which measures the total error including interpolation errors.

Figs. 3.40-3.44 examine the errors as a function of the number of grid points, $10^4 \leq n_p \leq 10^5$. The error is calculated for each of the grid types, at the maximum polynomial order, $n_{\max} = 40$, with no smoothing, $\frac{\sigma}{\rho_0} = 0$, and no rotation angle, $\phi = 0$. The results for several images are averaged together to improve the consistency.



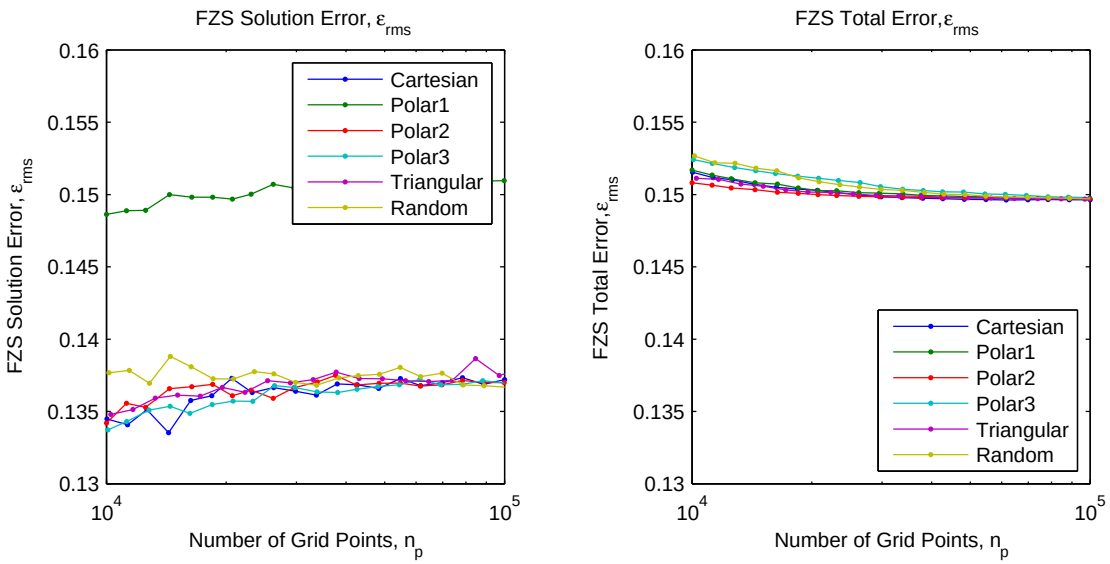
(a) (b)

Figure 3.40: The mean error of the matrix solution (a) shows virtually no dependence on the grid resolution, while the total, mean error shows very little improvement at higher resolutions.



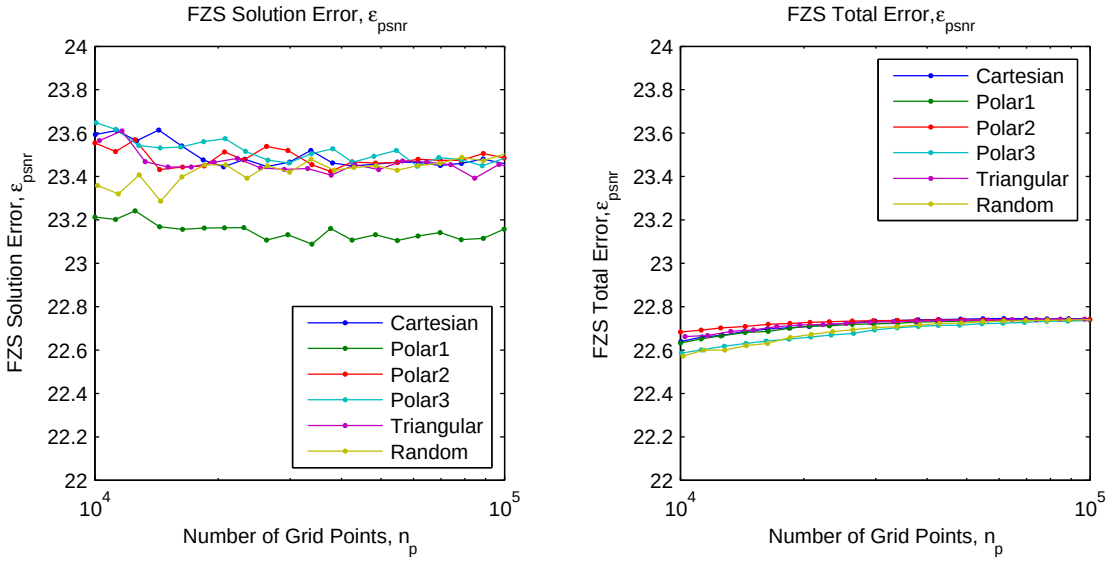
(a) (b)

Figure 3.41: The maximum error of the matrix solution (a) appears almost random, whereas, the total max error decreases variability for increasing resolution.



(a) (b)

Figure 3.42: The RMS error over the solution (a) shows a slight increase while the total error (b) shows a slight decrease.



(a) (b)

Figure 3.43: The PSNR on the grid shows no dependence on the number of grid points in the solution (a). The total PSNR between the images improves slightly at higher resolutions, but is probably due only to the grid refinements.

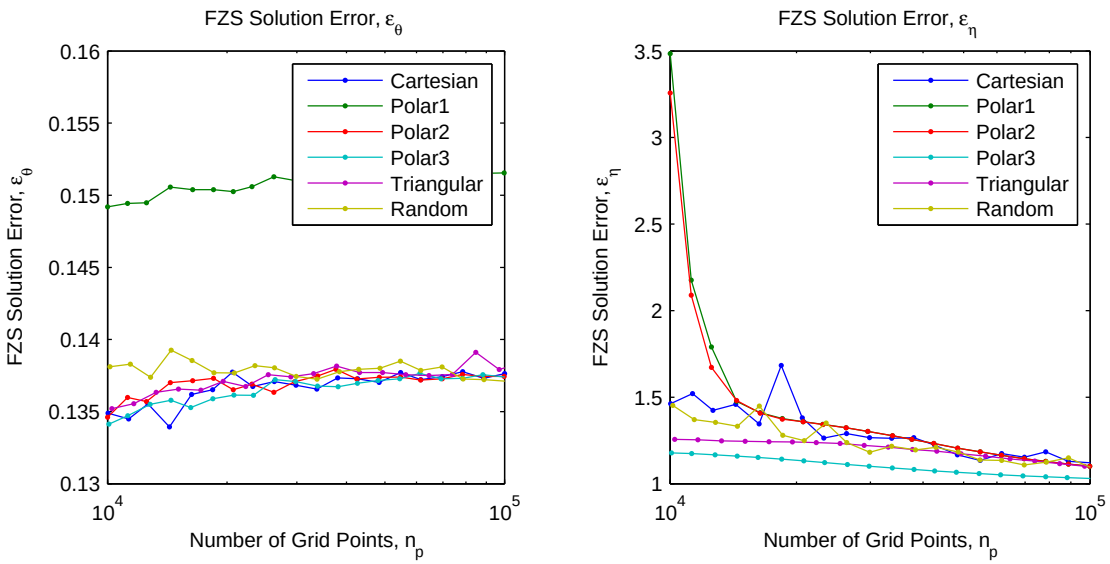
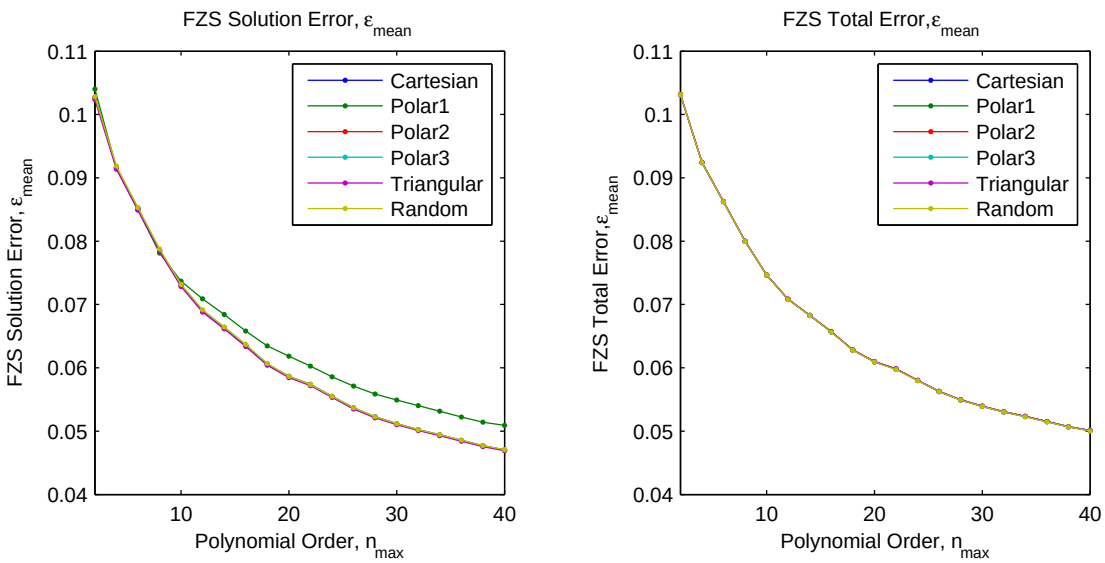


Figure 3.44: The closeness of fit of the solution (a) increases slightly. The maximum solution attained, η, follows the condition number, which indicates a well-conditioned system.

Fig. 3.40 shows the mean error does not vary significantly as a function of the grid resolution. The total, mean error is slightly higher than the solution error, as expected, but not for the polar1 grid, where it is inexplicably lower. Fig. 3.41 plots the maximum deviation between the solution (a) and the total image error (b). There is significant variation which appears to decrease in the total error at higher resolutions. Similar to the trends observed in the mean error, the RMS errors in Fig. 3.42 shows only a slight decrease in the total error. Again, the solution error for the polar1 grid seems abnormally high. The PSNR in Fig. 3.43 confirms that both the error in the solution and the total error does not significantly improve for the increasing grid resolutions. The closeness of fit, θ, in Fig. 3.44(a) does not show an improvement, but may not be an accurate measure since the size of the systems change.

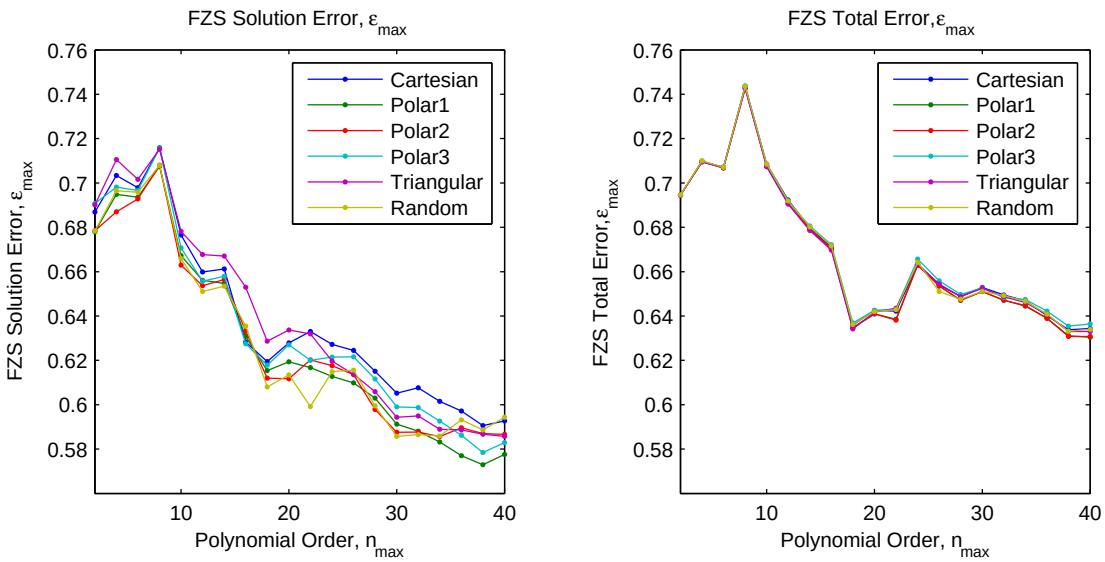
The solution approaches the maximum, $\eta \to 1$, for increasing order, which is not surprising since the condition number also approaches unity at higher resolutions. Overall, the errors do not significantly depend on the number of grid points, for these well-conditioned systems. The total errors decrease slightly, but this is probably due to the refinement in the grid's ability to resample and image with smaller pixel sizes.

Figs. 3.45-3.49 plot the error as a function of the maximum polynomial order, $2 \leq n_{\max} \leq 40$. The errors are calculated for each of the grid types, on a high-resolution grid, $n_p = 10^5$, with no smoothing, $\frac{\sigma}{\rho_0} = 0$, and no rotation angle, $\phi = 0$. The results for several images are averaged together to improve the consistency.



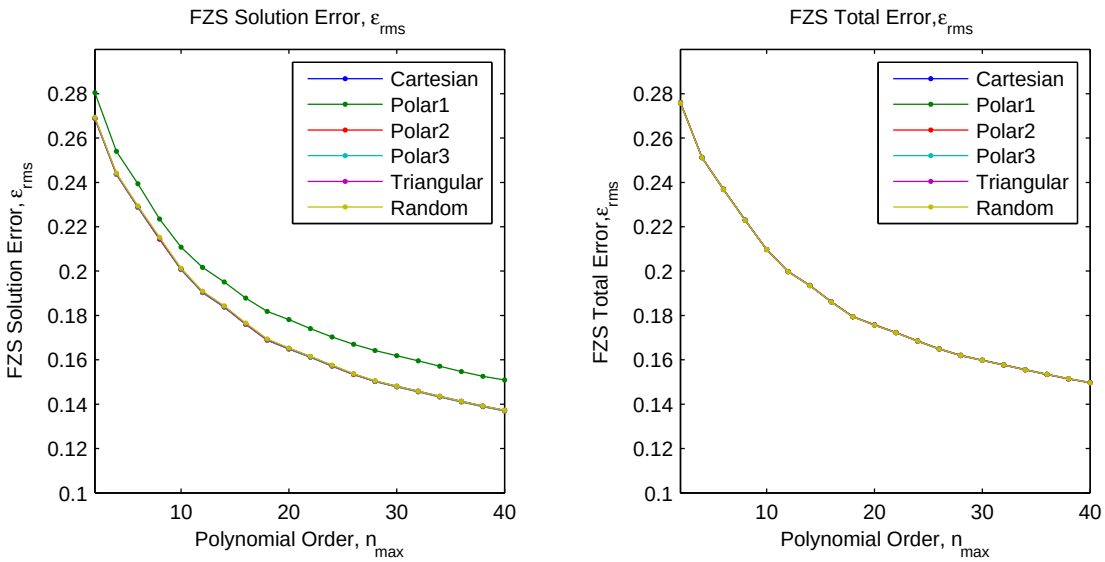
(a) (b)

Figure 3.45: The mean error between the solution residuals in (a) and the image residuals in (b) decreases for higher orders; dropping by approximately 50%.



(a) (b)

Figure 3.46: The maximum error in the solution residuals (a) and the total image residuals (b) appears to decrease overall, but with high variability.



(a) (b)

Figure 3.47: The RMS error, is higher than the mean error as expected, but shows a similar decrease in both the solution residuals (a) and the total image residuals (b).

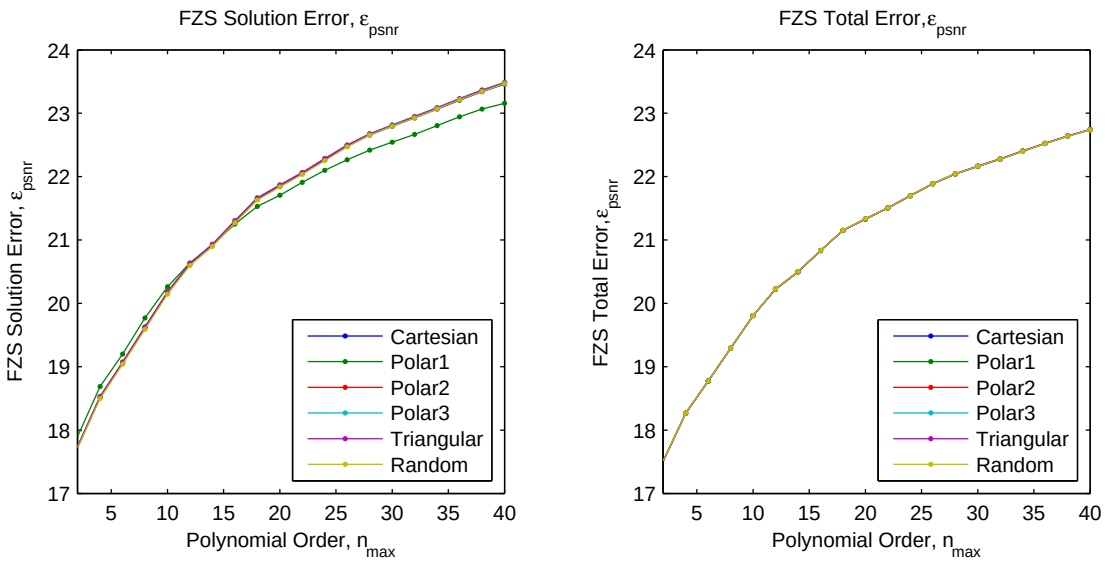


Figure 3.48: The PSNR increases in both (a) the solution residuals and (b) the total residuals.

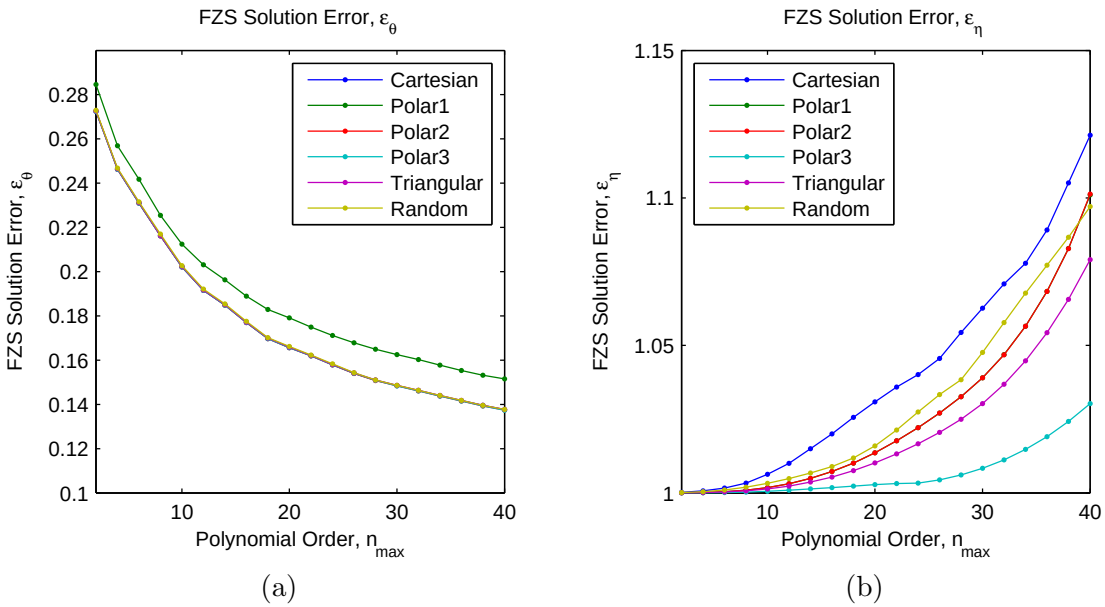
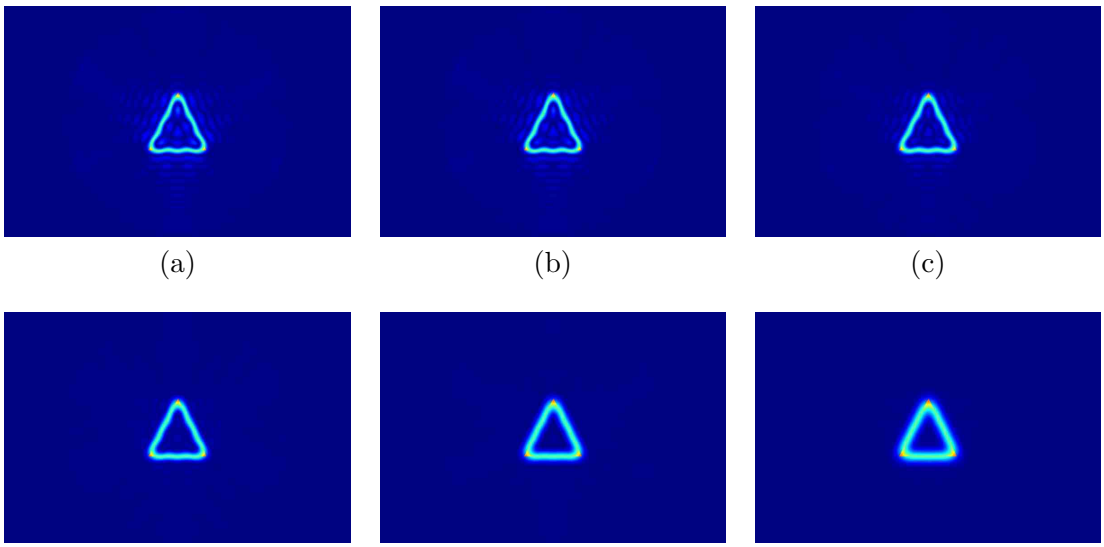


Figure 3.49: The solution angle, θ, in (a) decreases for higher orders, indicating that the least-squares solution is approaching the exact solution. The solution does not attain the maximum, $\eta > 1$, but remains well-conditioned at all polynomial orders.

Figs. 3.45-3.48 all show that the errors between the solution residuals and the total image residuals decrease for higher orders of the FZS approximations, as expected. This provides an estimate of how much accuracy can be obtained for increasing polynomial orders, hence increasing computational work. When $n_{\max} = 10, 20, 30, 40$ the number of FZC to solve for are 66, 231, 496, and 861, respectively. The corresponding errors are $\varepsilon_{\text{rms}} \approx 0.21, 0.18, 0.16, 0.15$. The accuracy of the approximation improves for higher orders of polynomials at the cost of higher computation workloads. The previous figures also show that there was neither a significant difference between the total error and the solution errors nor between the various grid types. This indicates that the polynomial order dominates the amount of accuracy which is attainable.

The appearance of the Gibbs-Wilbraham phenomenon appears when the desired features in a signal or image require higher-order approximations than what is utilized. Removing the high-frequency features and increasing the smoothness eliminates the problem but at the expense of reduced accuracy, as the following analysis shows. Fig. 3.50 illustrates the absolute value of the residuals between a source image (a solid triangle) and the approximated image from the FZS for various levels of smoothing (a)-(f), $\frac{\sigma}{\rho_0} = 0.000$, 0.016, 0.032, 0.048, 0.064, 0.080. In (a), the error around the edges appears as a cyan blue and the corners appear red for higher errors. The ringing appears as a light blue on the triangle interior and around the entire exterior. As the smoothing increases (b)-(e), the amplitude of the ringing diminishes while the errors around the triangular perimeter widens, and in (f) the ringing is not visibly present. The presence and strength of the ringing depends on the application; it may be preferable to have the higher accuracy and accept the ringing or to have reduced ringing and accept the loss of some accuracy. The following analysis quantifies some the errors associated with smoothing.

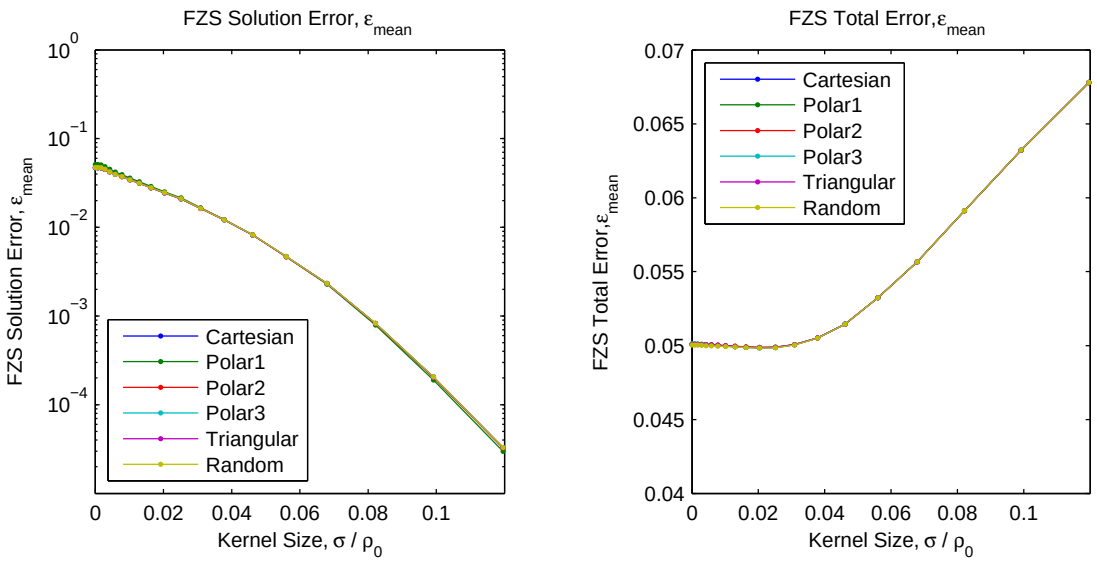


(a) (b) (c)

(d) (e) (f)

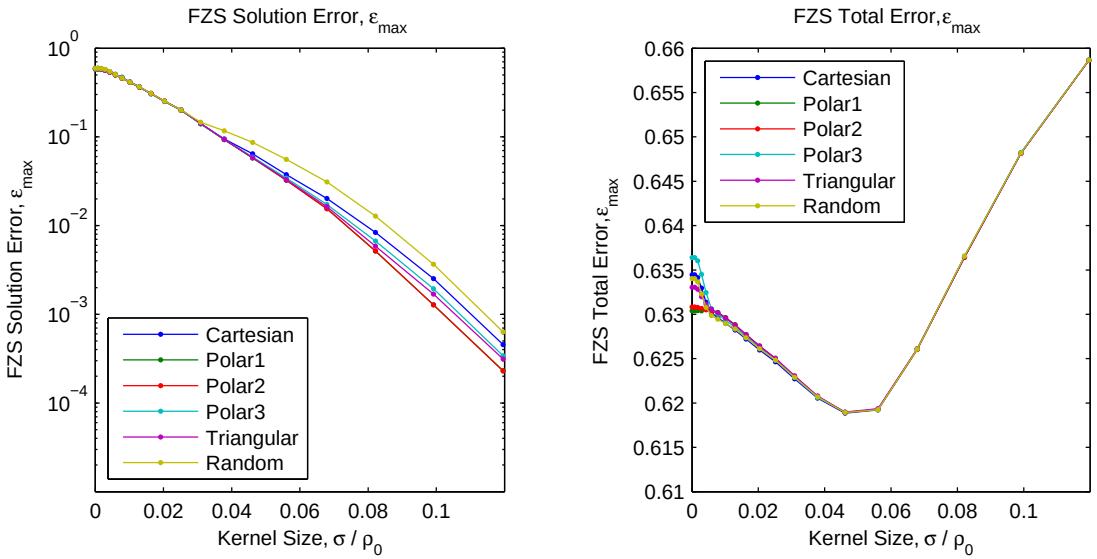
Figure 3.50: Images (a)-(f) show the absolute value of the residuals of an image of a triangle and the FZS approximations with higher levels of pre-filtering with a GLPF. The amplitude of the ringing in (a) diminishes for increasing levels of smoothing (b)-(e), and virtually disappears in (f) for $\frac{\sigma}{\rho_0} = 0.08$.

Figs. 3.51-3.55 plot the errors as a function of the smoothing prior solving for the FZC for $0 \leq \frac{\sigma}{\rho_0} \leq 0.12$. The error was calculated for each of the grid types, on high-resolution grids, $n_p = 10^5$, at maximum polynomial order, $n_{\max} = 0$, and no rotation angle, $\phi = 0$. The results for several images are averaged together to improve the consistency of the results.



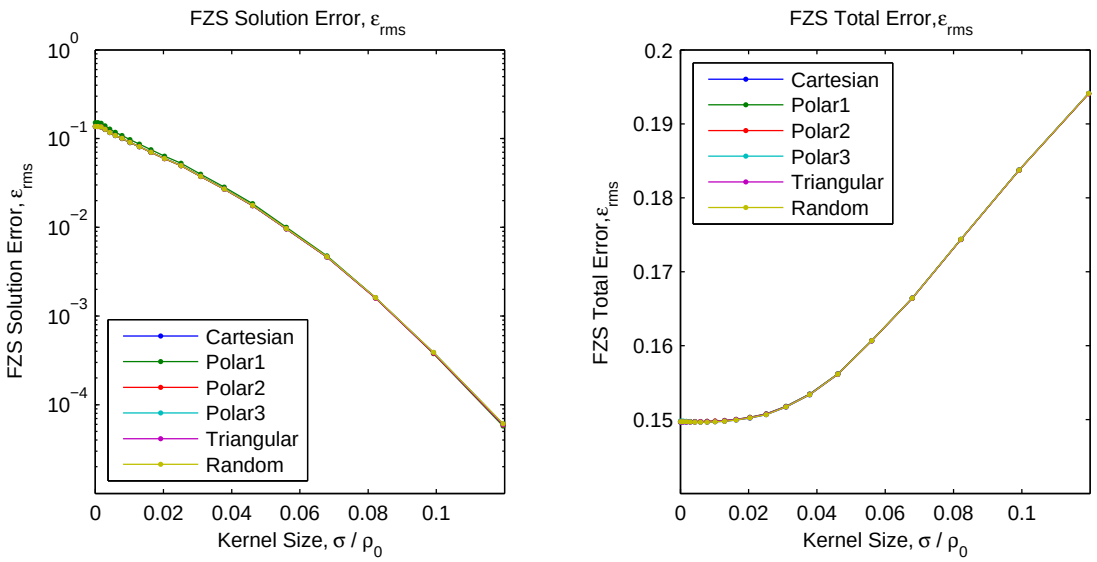
(a) (b)

Figure 3.51: The mean error of the residuals over the grid (a) dramatically decrease with smoothing. The total error of the residuals between the images (b) initially reduces slightly, but then dramatically increases.



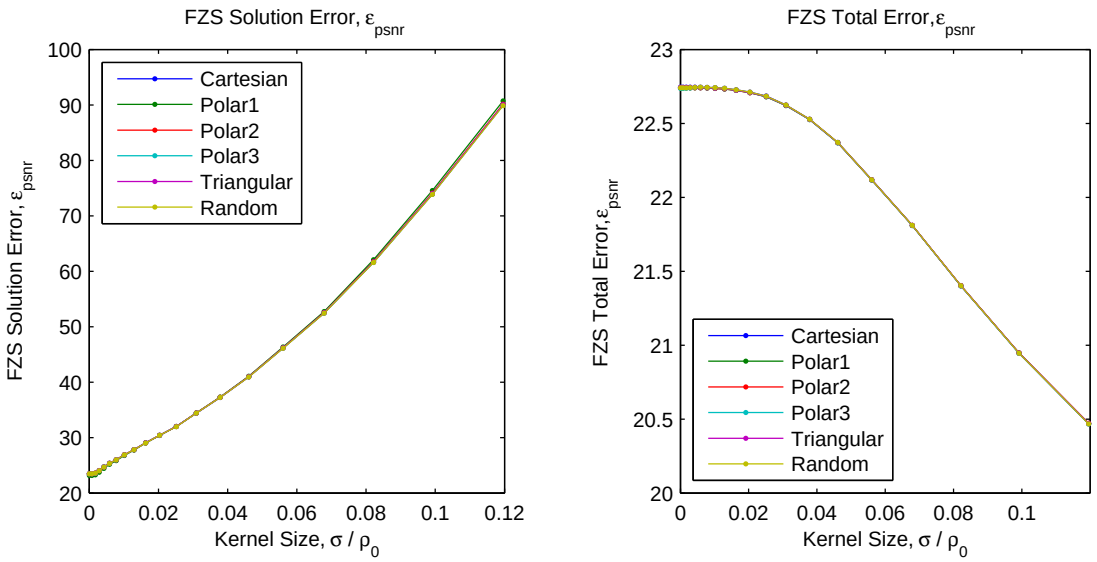
(a) (b)

Figure 3.52: The maximum error of the residuals on the grid (a) also dramatically decrease and with a good degree of stability. The total maximum error (b) shows appreciable improvement and stabilization, but then turns toward higher final errors.



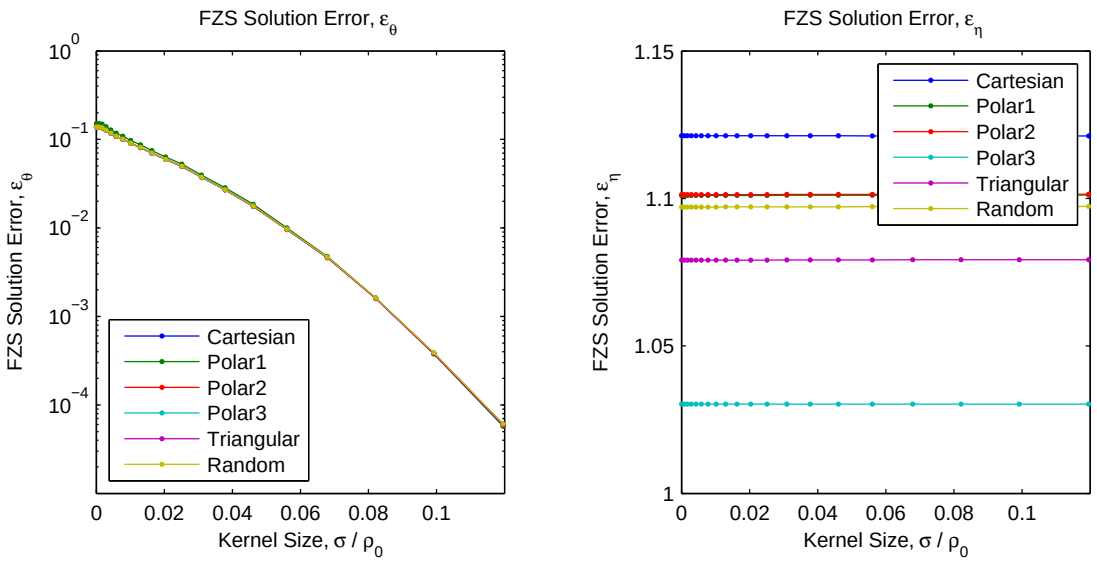
(a) (b)

Figure 3.53: The RMS error of the residuals from the grid approximation in (a) decrease rapidly for heavier smoothing. The total RMS error of the image residuals increases, but only after $\frac{\sigma}{\rho_0} > 2\%$.



(a) (b)

Figure 3.54: The PSNR increases dramatically for the grid residuals in (a), but decreases overall for the total image residuals in (b).



(a) (b)

Figure 3.55: The closeness of fit, θ, in (a) confirms that the solution to the matrix equations improves dramatically. The η for this system is unaffected by smoothing since it does not alter the matrix equations.

The errors from the residuals over the grid, which reflect the quality of the approximation after smoothing, rapidly decrease for heavier smoothing. Figs. 3.51(a)-3.54(a) all show improvements in the accuracy by several orders. It may be possible to attain perfect reconstruction (i.e., errors $\approx 10^{-16}$) over the grid with higher amounts of smoothing, but is not explored here. The total error of the residuals between the original and final images, which includes the differences introduced by smoothing, however, shows that slight smoothing does not affect the accuracy but heavy smoothing reduces accuracy. Figs. 3.51(b), 3.53(b) and 3.51(b) all confirm that the accuracy does not change appreciably for small smoothing kernel sizes, $\frac{\sigma}{\rho_0} < 2\%$. After that, the errors increase linearly with the kernel size. Fig. 3.52(b) shows that smoothing stabilizes and decreases the maximum deviation of the residuals up to $\frac{\sigma}{\rho_0} < 6\%$.

The errors in the solution and total residuals are not affected by a rotation angle. The errors (figures omitted) do not correlate to the rotation angle and show almost no variation. This confirms that the DZP are rotationally invariant, as expected.

This section utilizes the FZS to compute the approximation of an image reconstructed from the DZP. The errors between the solution residuals and the image residuals are examined as a function of the number of grid points, n_p, the maximum polynomial order, $n_{\max}$, and the amount of smoothing pre-filter, $\frac{\sigma}{\rho_0}$. These errors are compared between the grid types and averaged over several test images.

The errors do not vary significantly as a function of the number of grid points, n_p, provided that $n_p > 10^4$. They typically improve less than one order of magnitude, and the polar3 grid has the least error at low resolutions. The errors significantly depend on the maximum polynomial order, $n_{\max}$; decreasing for higher-order approximations, as expected. The amount of accuracy gained by increasing the polynomial order, however, is significantly less than the computational requirements. Furthermore, the gains diminish for higher orders. Using a GLPF to smooth the image prior to computing the FZS improves the accuracy of the solution, but decreases the total errors. The total effect was a marginal increase in accuracy for small smoothing, but a substantial decrease in accuracy for heavier smoothing. We also observe that smoothing virtually eliminates the ringing artifacts.

Chapter 4

Conclusion

The DZP provide a rotationally invariant, orthonormal basis over a circular domain. There are many ways to partition the domain to sample the continuous ZP and produce the DZP. Assembling the DZP into a matrix problem provides tools, common in linear algebra, to assess the accuracy of the FZC solutions. The Gram matrix provides a visual inspection method to find which polynomials have non-zero overlap. This inspired the design of a new grid, polar3, to improve the orthogonality of the polynomials with identical repetition number. Computing the rank of the DZP produces a minimum resolution criteria. The DZP system should not be utilized if it does not have full-rank. The condition number provides the best metric for estimating the solution's accuracy. It was also observed that the sorting order of the polynomials affects the condition number. Sorting by the repetition number, m, then by the order, n, decreases the condition number. The errors were then analyzed as a function of the grid type, grid resolution, polynomial order, smoothing, and rotation angle.

The performance of six grid types was analyzed side-by-side for the entire analysis. The polar3 grid delivers the best conditioning and the least amount of errors at significantly lower grid resolutions. The performance of the remaining grids improves at high grid resolutions, even for the random grid. In some cases, the resolution of the original image may support a well-conditioned DZP system without interpolation. This would eliminate the significant errors introduced by just the interpolation process. When interpolation is necessary, the polar3 grid delivers the best performance overall. This suggests that further improvements in the DZP are possible by grid optimization alone.

The grid resolution does not significantly affect the convergence of the FZC and the FZS, provided that the DZP system is well-conditioned. The FZC solution improved slightly for all grid types as n_p increased from 10^4 to 10^5, for $n_{\max} = 40$. The polar3 grid had lower error overall and much lower error at the lower resolutions. The FZS solution does not show significant improvements, with or without resampling. From these observations, there is no justification to employ extreme resolutions beyond what is necessary for the DZP to be well-conditioned.

Higher polynomial orders dramatically improve the accuracy of the FZS while preserving the accuracy of low-order FZC. This is an important feature for progressive refinement during an analysis. It also ensures that a comparison between a high-order approximation and a low-order approximation is valid between the polynomials of similar order (assuming proper normalization). While the accuracy increases at higher orders, the computational work and grid requirements increase faster. The accuracy improves less than first order (Fig. 3.47), yet the number of polynomials increase as second order and the minimum resolution criteria also increases as second order. This does not include the work associated with the computation of the ZP, which may be moderated by recursive or factorial-free methods [23]. Approximations above 40^{th} order also require arbitrary precision arithmetic.

Introducing the GLPF smoothing filter rapidly improve the ability of DZP to reconstruct an image over the grid. This suggests that there is a lower-limit on the smoothness of a function that the DZP can accurately represent. The solution errors of the FZS over the grid dramatically decrease for higher amounts of smoothing (e.g. Fig. 3.53(a)) and might reach numerical accuracy $\approx 10^{-16}$. The total errors (which include errors introduced by the smoothing), however, rapidly increased (e.g., Fig. 3.53(b)). This contradiction suggests that one should carefully consider the smoothness of the function to approximate with the accuracy of the FZS. The DZP cannot accurately approximate highly detailed images, as expected. The Gibbs-Wilbraham phenomenon appears as ringing in the FZS if the function does not have sufficient smoothness. Further analysis is necessary to analytically quantify the smoothness requirement as a function of the polynomial order.

Finally, the rotational invariance of the DZP was observed in the FZC and the FZS. This is readily computed from the magnitude of the complex, polynomials. The phase information is also preserved but requires additional considerations not undertaken here.

References

[1] Alireze, K., Hong, Y., H., "Invariant Image Recognition by Zernike Moments." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 12, No. 5, pp. 489, 1990. Print.

[2] Asadi, M. R., Vahedi, A., Hamidreza, A., "Leukemia Cell Recognition with Zernike Moments of Holographic Images." *IEEE Signal Processing Symposium, NORSIG*, pp. 214-217, 2006. Print.

[3] Belkasim, S., et al., "The Optimum Automatic Thresholding Using the Phase of Zernike Moments." *IEEE International Midwest Symposium on Circuits and Systems*, 47th, 2004. Print.

[4] Bhatia, A. B. and Wolf, E. "On the Circle Polynomials of Zernike and Related Orthogonal Sets." *Proc. Cambridge Phil. Soc.*, Vol. 50, No. 40, 1954. Print.

[5] Born, Max and Wolf, Emil. *Principles of Optics* — 6th ed. (with corrections), Sec. 9.2, appx. VII, Pergamon Press Inc., 1980. Print.

[6] Botev, Z. I., Grotowski, J. F., and Kroese, D. P. "Kernel Density Estimation via Diffusion." *Annals of Statistics* (submitted), 50, 40, 1990. Web.

[7] Dai, G. M., Mahajan, V. N., "Orthonormal Polynomials in Wavefront Analysis: Error Analysis." *Applied Optics*, Vol. 47, No. 19, July 2008. Print.

[8] Demmel, J. W., *Applied Numerical Linear Algebra*, SIAM, 1997. Print.

[9] Flusser, Jan, et al. *Moments and Moment Invariants in Patter Recognition*, Sec. 6.3. Wiley, 1980. Print.

[10] Gonzalez, R. C., Woods, R. E., *Digital Image Processing*, 3 ed., Pearson Prentice Hall, 2008. Print.

[11] Guotang, B., et. al., "A Defect Detection Method Based on Zernike Moments for Industrial CT Image." *International Conference on Electronic Measurement and Instruments*, 2007. Print.

[12] Hou, X., et. al., "Comparison of Annular Wavefront Interpretation with Zernike Circle Polynomials and Annular Polynomials." *Applied Optics*, Vol. 45, No. 35, December 2006. Print.

[13] Huynh-Thu, Q. and Ghanbari M., "Scope of validity of PSNR in image/video quality assessment." *IEEE Electronics Letters*, Vol. 44, No. 13, pp. 800-801, June 2008. Print.

[14] Iskander, D. R., Collins, M. J., Davis, B., "Optimal Modeling of Corneal Surfaces with Zernike Polynomials." *IEEE Transactions on Biomedical Engineering*, Vol. 48, No. 1, January 2001. Print.

[15] "Kodak Lossless True Color Image Suite." *Richard Franzen*, http://r0k.us/graphics/kodak/, Feb. 2010. Web.

[16] Lathi, B. P., *Signal Processing and Linear Systems*, Oxford University Press, 1998. Print.

[17] "Rank." *MATLAB*, http://www.mathworks.com/help/techdoc/ref/rank.html, August 2010. Web.

[18] "rgb2gray." *MATLAB*, http://www.mathworks.com/help/toolbox/ images/ref/rgb2gray.html, August 2010. Web.

[19] Meyer, C. D., *Matrix Analysis and Applied Linear Algebra*, SIAM, 2000. Print.

[20] Nijboer, B. R. A. *Thesis*. University of Groningen, 1942. Print.

[21] "Frits Zernike – Biography." *Nobelprize.org*, http://nobelprize.org/nobel_prizes/physics/laureates/1953/zernike-bio.html, 2010. Web.

[22] Noll, R. J., "Zernike Polynomials and Atmospheric Turbulence." *J. Opt. Soc. Am.*, Vol. 66, No. 3, March 1976. Print.

[23] Papakostas, G. A., et al. "Modified Factorial-Free Direct Methods for Zernike and Pseudo-Zernike Moment Computation." *IEEE Transactions on Instrumentation and Measurement*, Vol. 58, No. 7, July 2009. Print.

[24] Persson, P. O., Strang, G. "A Simple Mesh Generator in MATLAB." *SIAM Review*, Vol. 46, No. 2, pp. 329-345, June 2004. Web.

[25] Teague, M. R. "Image analysis via the General Theory of Moments." *J. Opt. Soc. Am.*, Vol. 70, No. 8, August 1980. Print.

[26] Trefethen, L. N., Bau, D., *Numerical Linear Algebra*, SIAM, 1997. Print.

[27] Turuwhenua, J., "Corneal Surface Reconstruction Algorithm Using Zernike Polynomial Representation: Improvements." *J. Opt. Soc. Am.*, Vol. 24, No. 6, June 2007. Print.

[28] Wang, L., Healey, G., "Using Zernike Moments for the Illumination and Geometry Invariant Classification of Multispectral Texture." *IEEE Transactions on Image Processing*, Vol. 7, No. 2, Feb. 1998. Print.

[29] "Gramian matrix." *Wikipedia*, http://en.wikipedia.org/wiki/gramian_matrix, July 2010. Web.

[30] "Mean absolute error." *Wikipedia*, http://en.wikipedia.org/wiki/Mean_absolute_error, July 2010. Web.

[31] "Mean squared error." *Wikipedia*, http://en.wikipedia.org/wiki/Mean_squared_error, July 2010. Web.

[32] "Peak signal-to-noise ratio." *Wikipedia*, http://en.wikipedia.org/wiki/Peak_signal-to-noise_ratio, July 2010. Web.

[33] "Root mean square deviation." *Wikipedia*, http://en.wikipedia.org/wiki/Root_mean_square_deviation, June 2010. Web.

[34] Wyant, James and Creath, Katherine. *Applied Optics and Optical Engineering.* Vol. XI, Academic Press Inc, 1992. Web.

[35] Xin, Y., et al. "Accurate Computation of Zernike Moments in Polar Coordinates." *IEEE Transactions on Image Processing*, Vol. 16, No. 2, February 2007. Print.

[36] Zernike, F. "Beugungstheorie des Schneidenverfahrens und seiner verbesserten Form, der Phasenkontrastmethode." *Physica1*, pp. 689-704, 1934. Print.

[37] Zernike, F., Brinkman, H. C. "Verh. Akad. Wet. Amst." *Proc. Sec. Sci.*, Vol. 38, No. 11, 1935. Print.