V.V.

USCIPI Report 730

THE DEGREES OF FREEDOM OF SAMPLED IMAGES

by

Dennis Grant McCaughey

June 1977

Image Processing Institute
University of Southern California
University Park
Los Angeles, California 90007

This research was supported by the Advanced Research Projects Agency of the Department of Defense and was monitored by the Wright Patterson Air Force Base under Contract No. F-33615-76-C-1203, ARPA Order No. 3119 and by the National Institutes of Health in the form of a training grant Contract No. GM01734.

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)

Image Processing Institute
University of Southern California, University Park
Los Angeles, California 90007

2a. REPORT SECURITY CLASSIFICATION: UNCLASSIFIED

2b. GROUP

3. REPORT TITLE

THE DEGREES OF FREEDOM OF SAMPLED IMAGES

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Technical Report, June 1977

5. AUTHOR(S) (First name, middle initial, last name)

Dennis Grant McCaughey

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
June 1977	167	48

8a. CONTRACT OR GRANT NO.: F33615-76-C-1203

b. PROJECT NO.: ARPA Order No. 3119

c.

d.

9a. ORIGINATOR'S REPORT NUMBER(S): USCIPI Report 730

9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. DISTRIBUTION STATEMENT

Approved for release: distribution unlimited

11. SUPPLEMENTARY NOTES

12. SPONSORING MILITARY ACTIVITY

Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, Virginia 22209

13. ABSTRACT

This dissertation presents a degree of freedom or information content analysis of images and imaging systems in the context of digital image processing. As such it represents an attempt to quantify the number of truly independent samples one gathers with imaging devices.

In quantifying the degrees of freedom of an imaging system it is necessary to develop an appropriate model. In this work the imaging system is modeled as a linear system through the continuous-discrete imaging equation. The associated gram matrix is then employed as an aid in defining the system degrees of freedom. The gram matrix eigenvalues are shown to be related to those of the associated continuous-continuous model and can be used to predict the discretized system performance. These ideas are then applied to the tomographic or projection imaging system; and result in the ability to predict the performance of this system by indicating where redundant data is achieved, and the best ways of increasing the degrees of freedom with a minimum sample increase.

The degrees of freedom of a sampled image itself are developed as an approximation problem. Here bicubic splines with variable knots are employed in an attempt to answer the question as to what extent images are finitely representable in the context of a digital computer.

Relatively simple algorithms for good knot placement are given, and result

DD FORM 1 NOV 65 1473

UNCLASSIFIED

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
in spline approximations that achieve significant parameter reductions at acceptable error levels. The knots themselves are shown to be useful as an indicator of image activity, and have potential as an image segmentation device.						
* *						
Key Words: Digital Image Restoration, Splines, Image Processing, Grammian, Ill Conditioning, Continuous Discrete Model, Degrees of Freedom.						

ABSTRACT

This dissertation presents a degree of freedom or information content analysis of images and imaging systems in the context of digital image processing. As such it represents an attempt to quantify the number of truly independent samples one gathers with imaging devices.

In quantifying the degrees of freedom of an imaging system it is necessary to develop an appropriate model. In this work the imaging system is modeled as a linear system through the continuous-discrete imaging equation. The associated gram matrix is then employed as an aid in defining the system degrees of freedom. The gram matrix eigenvalues are shown to be related to those of the associated continuous-continuous model and can be used to predict the discretized system performance. These ideas are then applied to the tomographic or projection imaging system, and result in the ability to predict the performance of this system by indicating where redundant data is achieved, and the best ways of increasing the degrees of freedom with a minimum sample increase.

The degrees of freedom of a sampled image itself are developed as an approximation problem. Here bicubic splines with variable knots are employed in an attempt to answer the question as to what extent images are finitely representable in the context of a digital

computer.

Relatively simple algorithms for good knot placement are given, and result in spline approximations that achieve significant parameter reductions at acceptable error levels. The knots themselves are shown to be useful as an indicator of image activity, and have potential as an image segmentation device.

ACKNOWLEDGEMENT

I would like to express my sincere appreciation to the members of my dissertation committee whose suggestions and support helped make this research possible. I would especially like to express my gratitude to Professor Harry Andrews who served as committee chairman, and in that capacity provided the necessary guidance without which the perspective in this work would have been lost. I would also like to thank Professor T. Harris for pointing out the similarity between the continuous-discrete and regression models, and to Professor G. Bekey for his helpful suggestions concerning the biomedical applications of this work.

Dr. Z. H. Cho of the Laboratory of Nuclear Medicine and Radiation Biology at U.C.L.A. deserves thanks for providing the experimental data for the tomographic results.

This research was sponsored in part by the National Institutes of Health in the form of a training grant under contract number GM01734; additional support was provided by the Advanced Research Projects Agency of the Department of Defense and was monitored by Wright-Patterson Air Force Base under contract number F-33615-76-C-1203 ARPA order #3119.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

Figure

LIST OF TABLES

ABBREVIATIONS

APC	Armored Personnel Carrier
ART	Algebraic Reconstruction Technique
CCD	Charge Coupled Device
DOF	Degrees of Freedom
LAX	Los Angeles International Airport
PSF	Point Spread Function
SAW	Surface Acoustic Wave
SIPSF	Space-Invariant Point Spread Function
SVD	Singular Value Decomposition
SVPSF	Space-Variant Point Spread Function

Chapter 1

INTRODUCTION

This dissertation is concerned with the concept of degrees of freedom (DOF) or information content of images and imaging systems arising in digital image processing. This concept is important, as it is fundamental to problems such as image coding and image restoration. In coding problems, one is interested in the transmission of that information relevant to the users' needs and in the elimination of irrelevant data; while in the restoration problems, the object is to be restored from samples of a corrupted image.

These two examples are given to illustrate that the subject could be approached from two points of view. Namely, the image could be treated as the output of an imaging system whose characteristics are known and thus dependent on the DOF of the imaging system; or the subject could be considered by itself as in the coding problem where the imaging system characteristics are either secondary in importance or considered ideal.

Fundamentally, this concept of degrees of freedom can be viewed as an attempt to quantify the number of truly independent samples of data one gathers with photographic or other imaging devices. As image sensor technology grows, the quantity of data gathered

increases, and it becomes reasonable to ask what the true increase in information content is as one increases image samples. This is especially important in medical imaging applications where an increase in the quantity of data gathered, while not producing a corresponding increase in image information, subjects the patient to an unnecessary increase in radiation exposure. Thus in designing imaging systems for medical applications it is extremely important that the information content of the imaging system be quantized.

Since half the thrust of this dissertation will be towards information content in imaging systems in general, and to the tomographic, or projection, imaging system in particular, a mathematical model for imaging will be necessary.

1.1 Mathematical Imaging Models

In modelling imaging systems an assumption often made is that of a linear system. While it is not true that every imaging system is in fact linear, this assumption is useful in that it makes the analysis tractable and provides reasonable results. Even if untrue, the system can often be considered linear if the region of observation is kept small enough. Thus imaging systems can be analyzed by considering them in terms of two-dimensional linear system theory. In applying these methods to imaging systems the assumption is made that an image, g, is related to the original object, f, by a

superposition integral as follows:

$$g(x, y) = \iint_R h(x, y;\xi, \eta) f(\xi, \eta) d\xi d\eta . \qquad (1\text{-}1)$$

Here $h(x, y;\xi, \eta)$ represents the weighting function of the imaging system and R is the region of integration over the input coordinate system. In this dissertation the assumptions will be made that $h(x, y;\xi, \eta)$ is continuous in x and y, bounded, and such that

$$\iiiint |h(x, y;\xi, \eta)|^2 \, dxdyd\xi d\eta < \infty$$

Furthermore, it will be assumed that $f(\xi, \eta)$ will also be bounded and at least piecewise continuous. These assumptions will imply that g(x, y) will be continuous in x and y. This assumption on the part of g(x, y) is necessary for digital processing since only a sampled version of g will be dealt with. Thus if we define g_i to be g(x, y) sampled in the image coordinate plane (x, y) and similarly for h, then the vector form of equation (1-1), the continuous-discrete imaging model, is obtained [1-1, 1-2]

$$\underline{g} = \iint_R \underline{h}(\xi, \eta) f(\xi, \eta) d\xi d\eta + \underline{n} \qquad (1\text{-}2)$$

where $\underline{n}$ is an error term and all vectors are Nx1 columns. Here the aim, as in any reconstruction process, is to recover f as best as possible, in some sense, knowing $\underline{h}$ and $\underline{g}$. This model is depicted in Fig. (1.1).

Intuitively, the concept of degrees of freedom (DOF) fits nicely into this model as can be seen by noting that $f(\xi, \eta)$ is defined on a continuum, namely R, and as such represents a noncountably infinite number of DOF. The image g_i is described by a finite number of samples N, and at best represents N DOF. However, the imaging integral equation causes reduction of this number due to the point spread function (PSF) blur provided by $\underline{h}(\xi, \eta)$. Here the PSF is described by an N-vector whose elements are continuous functions of the object coordinate system (ξ, η). One's intuition might serve here in the concept that the greater the blur, the fewer the DOF. Thus, if $h_i(\xi, \eta)$ is a "narrow" function the DOF would be greater than if $h_i(\xi, \eta)$ were a "broad" function. Because $h_i(\xi, \eta)$ can represent a space variant point spread function (SVPSF), we can further let our intuition suggest that in regions in which the object is in "better focus" (i.e., narrower PSF) we would be obtaining greater DOF than in regions of "poorer focus" (i.e., broader or greater blur PSF).

The rationale behind the separation of the problem into two classes should be a little clearer now. If the system is taken to be ideal, then $\underline{g}$ is simply $f(\xi, \eta)$ sampled in the output coordinate plane and we are confronted with the problem of relating the DOF of a sampled image to its original unsampled version. It must also be considered that to sensibly discuss sampled images, one must be

dealing with ones that are band limited. If $f(\xi, \eta)$ is not band limited, which is often the case, we can assume that its sampled version $\underline{f}$ is obtained from some other function whose band limited version coincides with f at the sample points. In this case, h can be taken as an ideal low pass filter.

1.2 Research Objectives

The objectives of this research will be to develop a measure of the DOF of imaging systems and to apply this to two areas of study. The first is that of the tomographic or projection imaging system. The second area will be that of developing "smart sensors" by variably adapting to the DOF of the sensed image. With respect to the former projection imaging system [1-3], it will be necessary to develop a weighting function for the tomographic imaging system and its associated continuous-discrete model.

By quantifying the degrees of freedom of the tomographic imaging system we will, for the first time, be able to predict the resolution capability of the system for large numbers of samples. It will also be shown that the resolution limits in tomography are not a function of the particular reconstruction algorithm employed, but are fundamental to the process itself.

While the main intent in the first area of research is to investigate the degrees of freedom and information content in the projection

imaging process, structure will be shown to exist that makes a linear algebraic solution to the reconstruction problem feasible for images of a rather large dimension. This linear algebraic algorithm will be developed, and the results presented in this dissertation concerning information content in projection imaging will be obtained from this model.

It will also be shown that the difficulty in image restoration arises not so much from the sampling scheme used, but is innate to the original continuous-continuous model of linear imaging systems in general. This will be accomplished by relating the eigenvalues of the gram matrix of the kernel vector $\underline{h}$ with the singular values of the original kernel $h(x, y; \xi, \eta)$. The gram matrix will be discussed in Ch. 3. The situation where $h(x, y; \xi, \eta)$ is unknown will also be treated. Here the problem will be considered as a two dimensional approximation problem and the concept of an "epsilon degrees of freedom" will be developed. By this it is meant that the degrees of freedom of an image at a level epsilon will be the minimum number of functions needed to approximate $f(\xi, \eta)$ within an accuracy of epsilon assuming a particular metric.

From a "smart sensor" viewpoint by way of motivation, if we consider a sampled image consisting of N^2 samples that could be approximated to an acceptable error by a least squares polynomial of

M^2 variables with $M^2 \ll N^2$, it would be reasonable to say that this image had less than N^2 DOF in a least squares sense using a polynomial as the approximation technique. This approach is taken to circumvent the difficulty of associating a finite DOF to an image defined on a continuum which obviously has an uncountably infinite DOF if we desire to specify that image exactly. However, if we are willing to accept an approximation with a small but nonzero error, then the possibility exists in quantifying the DOF in this manner.

In effect this represents an attempt to "bridge the gap" between the continuous domain upon which images are defined and the discrete point sets involved in digital computations. Shannon's sampling theory represents one method for reconstructing uniformly sampled band limited images, but by taking a more general approximation theoretic approach other sampling and reconstruction techniques may be used. For example, the desirability of nonuniform or adaptive sampling can be illustrated by considering an image that contains high frequency information in a small region in its domain of definition with the remainder containing only low frequency components. If it is desired to reconstruct this sampled image using Shannon sampling, then it must be sampled at the Nyquist rate defined by that small high frequency zone. Intuitively, it would seem then that the regions of low frequency content are over sampled.

Realizing that we can never collect an infinite number of samples for application on a digital computer and that some error always results from this technique, it might be possible to achieve a reasonable error by a finite "Shannon interpolation" at the respective Nyquist rates in each of those regions. More formally, in one dimension if B_T is the bandwidth of the function and T is the length of the interval of observation then there are approximately $2\,B_T T$ independent samples from this function in the interval $[0, T]$. If we wish to say that in the interval $[o, a]$, $a < T$ the highest "effective" frequency component is B_T and in $[a, T]$ the highest "effective" frequency component is B_1, $(B_1 < B_T)$, then the number of independent samples in $[0, T]$ becomes $2\lceil B_T a + B_1(T-a)\rceil$. Since $B_1 < B_T$ it follows that $2\lceil B_T a + B_1(T-a)\rceil < 2B_T T$. Depending on the ratio of B_T to B_1 it might be possible to reconstruct the function to reasonable error by this approach with far fewer samples.

To illustrate the applicability of adaptive processing consider that advances in charge coupled device (CCD) sensors are such that some preprocessing within the sensor itself is not so unrealistic. This preprocessing could involve some evaluation as to what data constitutes information to the user and transmits only that data relevant to the users' needs. Surface acoustic wave devices (SAW) are becoming available that can provide a Fourier transform of an

image at video data rates so that it is possible to obtain a sensor that provides a transform of the image as its output. This increase in sensor sophistication coupled with the ability to gather large quantities of data, the ability to do adaptive sampling or some other more exotic processing to get at the real information content in the data, may provide fruitful results.

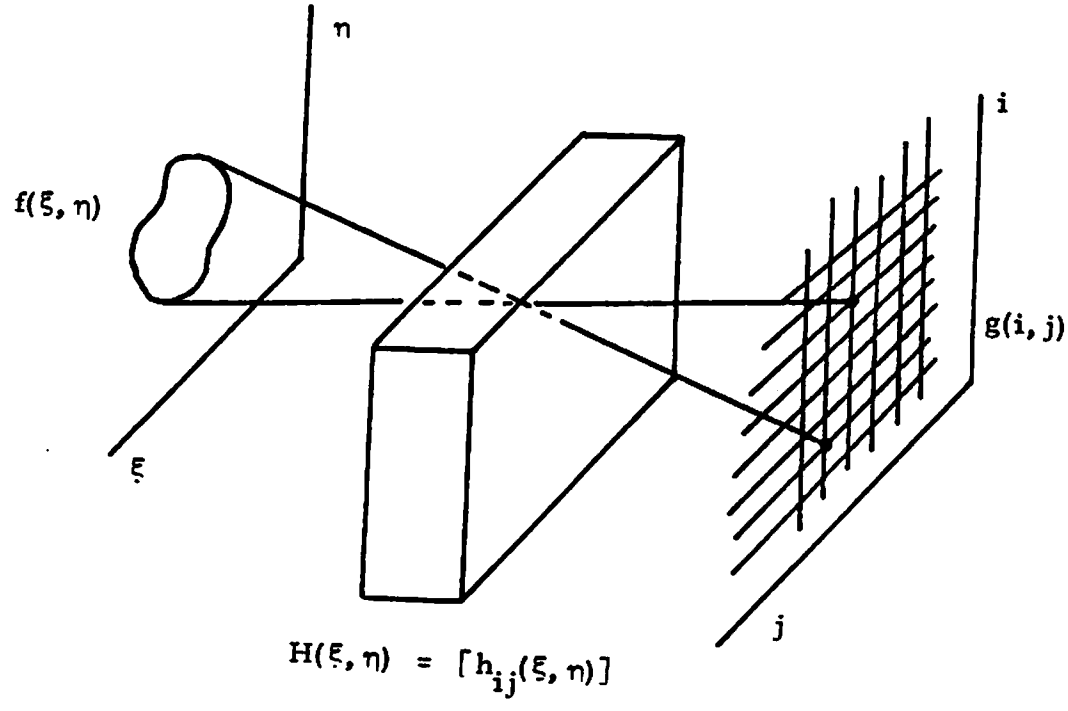


Figure 1.1. Continuous discrete model.

Chapter 2

REVIEW OF THE STATE OF THE ART

2.1 Tomographic Literature Review

The concept of degrees of freedom (DOF) has arisen in the imaging literature in an attempt to quantify the number of truly independent samples of data one gathers with photographic or other image sensing devices.

In obtaining an estimate, $\hat{f}$, for f in (1-2), consideration must be taken of the fact that the PSF kernel vector, $\underline{h}$, is comprised of kernel functions that are not necessarily independent. Neglecting the error term for the moment, this dependency of kernel functions implies that some of the output samples can be predicted by linear combinations of the others, are thus superfluous, and serve to reduce the DOF of the imaging system. The error term aggravates this situation in so far as, if an output element can be predicted by a linear combination of the others to within an accuracy better than the measurement error, that measurement adds no new information to what is already known. For one dimensional systems it has been suggested that the Gramian formed from the PSF N-vector could provide a quantitative measure of the DOF available in an imaging system. The entries in this matrix represent the correlation or

overlap of each PSF with its neighbors. If we consider the following one dimensional continuous-discrete model

$$\underline{g} = \int_R \underline{h}(\xi) f(\xi) d\xi$$

The gram matrix $[\Gamma]$ is given by

$$[\Gamma] = \int_R \underline{h}(\xi) \underline{h}^*(\xi) d\xi$$

where all vectors are considered as Nx1 column vectors and * indicates vector conjugate transpose. The rank and/or eigenvalue map could be used in the definition of the DOF [2-1, 2-2, 2-3]. In this dissertation we pursue this development in two dimensions and apply the analysis to the tomographic or projection imaging system as an imaging system example.

Projection imaging or three-dimensional reconstruction has been the subject of much research in recent years. This is the result of important applications such as electron microscopy [2-4, 2-5], radio astronomy [2-6, 2-7], and trans-axial tomography [2-8, 2-9, 2-10], to name a few. There has been a somewhat intense effort in this last application in the medical community with the proliferation of articles, reports, conferences, and even the manufacture of equipment. While the mathematical basis upon which transaxial tomography is founded is quite sound, many of the practioners of the technique have developed a variety of image reconstruction

approaches which are seemingly unrelated and controversial.

Basically the technique requires computational reconstruction of an image and researchers have developed algorithms which can be roughly classified as follows: convolutional [2-6, 2-7], Fourier transform [2-4], algebraic, in the sense of ART [2-9], and linear algebraic as described by Kashyap [2-10]. The convolutional and Fourier transform algorithms can be considered as closed-form algorithms in the sense that a continuous solution is first obtained, and then discretized for implementation on a digital computer. While mathematically elegant, this sequence gives little insight into the errors incurred due to discretization and the presence of noise or measurement errors in the system. Linear algebraic approaches somewhat circumvent this situation, but have not received as much attention, most likely, as large matrices are involved for images of even a moderate dimension. Nevertheless, the information content in a particular sampling geometry can be determined by considering the spectrum of the point spread function matrix describing the imaging system. This is not to say that linear algebraic methods are the only way to approach the information content or error analysis in the projection imaging system. Klug and Crowther [2-11] have formulated the image reconstruction problem as an eigenvalue problem in the continuous domain, and

have drawn analogies to optical data processing systems in analyzing the effect of a finite number of projections. In addition to this effort, Smith, Peters and Bates [2-12] have considered the effect of a finite number of projections in the reconstruction process.

2.2 Image Approximation Literature Review

The concept of the degrees of freedom of a sampled image as an approximation problem arises quite naturally in the context of image coding by transform methods. In transform image coding an orthogonal transformation is performed on a sampled image matrix and a bandwidth reduction is obtained by transmitting only those transform coefficients above a certain threshold whose level is consistent with the desired error [2-13]. The large bandwidth reductions reported are due, in part, to the compacting-of-image-energy property of the orthogonal transforms employed. However, any compacting in the transform domain is at the expense of an increased dynamic range in the transform coefficients because of the conservation of energy inherent to all orthogonal transformations.

Another situation is the application of the singular value decomposition (SVD) algorithm [2-14] to the sampled image matrix where upon the number of degrees of freedom can be equated with the number of effectively nonzero singular values, with the remaining parameters describing the orthogonal singular vectors.

In both of these situations we are dealing with a sampled version of the image only, and as a result the degrees of freedom could be affected by the sampling method used and thus could lead to misleading results. The point to be made here is that the degrees of freedom should be a characteristic of the original image and reflected in the sampled image only by our inability to collect an uncountably infinite number of samples for application on a computer.

It is to this end that approximation theory is directed. In fact the Remes algorithm represents a method for finding the best approximating polynomial (in the uniform norm) of degree $\leq n$ to a continuous function in the interval $[a, b]$ in terms of a sequence of solutions involving finite sample point sets with $n+2$ elements each [2-15]. Unfortunately, no equivalent algorithm is available in two dimensions but it should be noted that the Stone-Weinstrass theorem assures the ability to uniformly approximate a continuous function of 2 variables by a bivariate polynomial with arbitrary accuracy [2-16, 2-17].

Recently Hou [2-18] and Peyrovian [2-19] have employed spline functions to image restoration problems with considerable success. However, in both of these works the knots were both fixed and uniformly spaced with no attempt being made to improve the approximation by adjusting the knot placements. In one dimension

the existence of best approximating splines with both fixed and free knots have been shown to exist [2-20], and that the approximation capabilities of splines are greatly increased with the allowance of free knots. Recently Schultz has developed error bounds for multivariate spline approximations, in both L_∞ and L_2 norms in terms of the maximum mesh width of the knots and the moduli of continuity [2-21, 2-22].

While Schultz' results are mainly concerned with error bounds and not the best approximating properties of splines, it is one purpose of this dissertation to demonstrate experimentally that a sensible placement of the knots adds considerable power to the image approximating capability of bivariate cubic splines. Furthermore, since the mapping of the sampled image into its spline coefficients represents a nonenergy conserving transformation, a bandwidth reduction (manifested as a reduction in the number of spline coefficients needed to describe the image) that is not necessarily at the expense of an increased dynamic range. Thus the possibility exists of more efficiently quantizing and encoding those coefficients than those of the more widely used orthogonal transforms.

2.3 Overview of the Dissertation

The main features of this dissertation are: a) the adaptation of a continuous-discrete model for image restoration in general and its

application to the tomographic imaging system in particular; b) the use of the eigenvalues of the gram matrix in predicting the tomographic imaging system performance; c) the adoption of an approximation theoretic approach to relating the degrees of freedom of a sampled image to its original unsampled version when a system weighting function is either unknown or of secondary importance to the particular image processing task; and d) the application of that approximation theory to the design of "smart sensor" imaging arrays.

The following is a brief summary of the dissertation on a chapter by chapter basis.

Chapter 1 serves as an introduction to the notion of image degrees of freedom and as such outlines the difficulties involved. Models for imaging systems are introduced along with their appropriate constraints that will be used throughout the work.

Chapter 2 serves as a review of the research efforts towards characterizing the degrees of freedom of imaging systems with a known weighting function. The recent efforts in algorithm development and information content in the tomographic imaging system are also covered. The main intent is to review known systems and the approximation theoretic approach will be reviewed here and expanded upon in Chapter 5 after it has been properly motivated.

Chapter 3 contains the development of the tools necessary for the investigation of information content in imaging systems. Here the

relation between the gram matrix eigenvalues and system degree of freedom is developed along with algorithms for estimating $\hat{f}$ using the gram matrix of the continuous-discrete model. This is done for both the case where the system of equations is invertible, and where the system is singular and a constrained least squares approach is necessary. The separable gramian is also discussed and shown to result in a significant computational reduction. Finally, the relation of the gram matrix eigenvalues to the continuous-continuous model eigenvalues is explored and bounds are given for the separable gramian.

Chapter 4 is the application of the results of Chapter 3 to the tomographic imaging system. A continuous-discrete model along with its associated gramian is developed. Structure in the gramian is shown to exist making a linear algebraic solution possible. The degrees of freedom of the tomographic system are obtained using the gramian and are shown to be in excellent agreement with the general type of system tomographic imaging describes. Numerical results include some excellent reconstructions for experimental projection data.

Chapter 5 is devoted to the information content of a sampled image. In this chapter the problem is shown to be an approximation problem related to the prior methods (where the approximating functions are taken to be linear combination of the PSF kernels in the

continuous-discrete model). However, in this case there is considerably more flexibility. The concept of an "epsilon degrees of freedom" is defined and related to sampled images using splines with free knots.

Chapter 6 is devoted to the experimental results of the "free" or "variable knot" splines of Chapter 5.

Chapter 7 contains a summary of the dissertation along with some conclusions and possible future work.

Appendix A deals with the relevant computational properties of normalized B-splines.

Considering the dual nature of this work a flow diagram of the chapters is given in Fig. 2.1. It should be useful in visualizing the directions taken through the dissertation.

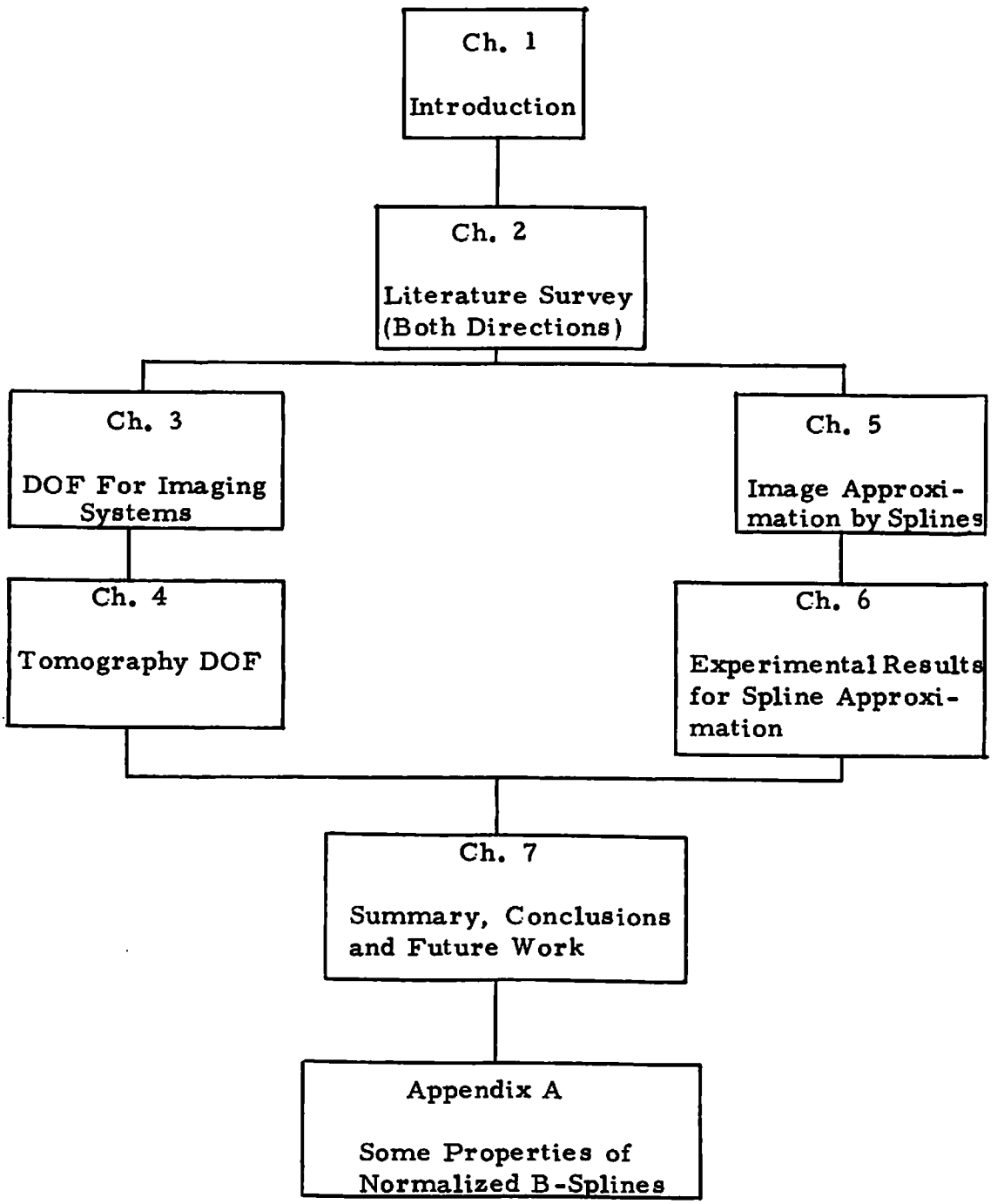


Figure 2.1. Organization of the Dissertation.

Chapter 3

THE CONTINUOUS-DISCRETE MODEL AND THE GRAMIAN

3.1 General Development

In this section we relate the DOF of the imaging system (manifested as dependent data) to the linear dependence amongst the PSF kernels of Eq. (1-2). In doing this we will select an arbitrary unit length vector $\underline{\alpha}$ which when dotted with the imaging equation yields

$$\langle \underline{\alpha}, \underline{g} \rangle = \langle \underline{\alpha}, \iint_R \underline{h}(\xi, \eta) f(\xi, \eta) d\xi\, d\eta \rangle + \langle \underline{\alpha}, \underline{n} \rangle \quad . \tag{3-1}$$

If the first term on the right is much smaller than the second, then its contribution can be considered minimal. Taking this with the fact that $\|\underline{\alpha}\| = 1$, we must reduce the degree of independence or number of DOF by one for every independent $\underline{\alpha}$ such that

$$\left| \langle \underline{\alpha}, \iint_R \underline{h}(\xi, \eta) f(\xi, \eta) d\xi d\eta \rangle \right|^2 \leq \|\underline{n}\|^2 \tag{3-2}$$

where $\|\underline{n}\|^2$ can be considered the sensor noise which contributes to the definition of the signal to noise ratio of the imaging system. We can further investigate the implications of equation (3-2) by noting that the vector inner product on the left can be related to the continuous inner product of f with $\underline{h}$ by

$$\left|\left\langle \underline{\alpha}, \iint_R \underline{h}(\xi,\eta) f(\xi,\eta) d\xi d\eta \right\rangle\right|^2 = \left| \iint_R f(\xi,\eta) \left\langle \underline{h}(\xi,\eta), \underline{\alpha} \right\rangle d\xi d\eta \right|^2$$

and applying the Schwartz Inequality

$$\left| \iint_R f(\xi,\eta) \left\langle \underline{h}(\xi,\eta), \underline{\alpha} \right\rangle d\xi d\eta \right|^2 \leq \iint_R |f(\xi,\eta)|^2 d\xi d\eta \iint_R \left|\left\langle \underline{h}(\xi,\eta), \underline{\alpha} \right\rangle\right|^2 d\xi d\eta$$

$$= \| f(\xi,\eta) \|_2^2 \, \underline{\alpha}^* [\Gamma] \underline{\alpha}$$

Here $[\Gamma]$ is the gramian of the PSF kernels and is defined as

$$[\Gamma] = \iint_R \underline{h}(\xi,\eta) \underline{h}^*(\xi,\eta) d\xi d\eta$$

or

$$\gamma_{ij} = \iint_R h_i(\xi,\eta) h_j^*(\xi,\eta) d\xi d\eta$$

which is Hermitian having real eigenvalues only. Because $f(\xi,\eta)$ can be considered bounded, $|f(\xi,\eta)| \leq E$, we will assume in conjunction with equation (3-2) that

$$E^2 \underline{\alpha}^* [\Gamma] \underline{\alpha} \leq \| \underline{n} \|^2$$

Expanding $[\Gamma]$ into its eigenvector decomposition $[\Gamma] = [U][\Lambda][U]^*$ and let $\underline{\beta} = [U]^* \underline{\alpha}$, another unit vector, we have

$$E^2 \underline{\beta}^* [\Lambda] \underline{\beta} \leq \| \underline{n} \|^2 .$$

Clearly $\underline{\beta}^* [\Lambda] \underline{\beta}$ is minimized by the allowing of $\underline{\beta}$ to be the eigenvector associated with the smallest eigenvalue of $[\Gamma]$, which leads to the conclusion, that for every eigenvalue λ such that:

$$E^2\lambda_i < \|\underline{n}\|^2$$

one degree of freedom is lost.

We can now turn our attention to the problem of obtaining an estimate of the object $\hat{f}$, and demonstrate the role played by the gramian. In doing this we will determine a minimum normed least squares estimator with the constraint that the norm of the error be equal to the norm of the noise vector. The least squares approach is taken as it leads to a generalized inverse solution in the discrete-discrete model which Rao [3-1, 3-2] has shown to be a minimum variance unbiased estimate of the original object, f. The discrete-discrete case is developed first and serves to motivate a minimum normed least squares solution to the continuous-discrete problem. The discrete-discrete problem is nothing more than a regression problem wherein the image vector $\underline{g}$ is related to the object vector $\underline{f}$ through a blur matrix $[H]$ as follows

$$\underline{g} = [H]\underline{f} + \underline{n} \qquad (3\text{-}3)$$

Here all vectors are Mx1 column vectors and $[H]$ is an MxN matrix with M≠N in general. $\underline{n}$ is a noise vector whose elements will be taken to be identically distributed independent random variables so that $E\{\underline{n}\,\underline{n}^T\} = [R_n] = [I]$ [3-2]. An estimate $\hat{f}$ can be obtained from the following:

$$\text{minimize} \quad \|\underline{f}\|_2^2$$

subject to

$$\| [H]\underline{f} - \underline{g} \|_2^2 = \| \underline{n} \|_2^2$$

which results in the minimization of

$$w(\underline{f}) = \underline{f}^T \underline{f} + \gamma \{ ([H]\underline{f} - \underline{g})^* ([H]\underline{f} - \underline{g}) + \| \underline{n} \|_2^2 \}$$

This leads to

$$[[I] + \gamma [H]^* [H]] \hat{\underline{f}} = \gamma [H]^* \underline{g}$$

resulting in

$$\hat{\underline{f}} = \left[\frac{1}{\gamma} [I] + [H]^* [H] \right]^{-1} [H]^* \underline{g}$$

which can also be expressed as [3-3]

$$\hat{\underline{f}} = [H]^* \left[\frac{1}{\gamma} [I] + [H] [H]^* \right]^{-1} \underline{g} \qquad (3\text{-}4)$$

The Moore-Penrose pseudo inverse, $[H]^+$, is given by [3-3, 3-4]

$$[H]^+ = \lim_{\gamma \to \infty} [H]^* \left[\frac{1}{\gamma} [I] + [H][H]^* \right]^{-1} \qquad (3\text{-}5)$$

and taking $\hat{\underline{f}} = [H]^+ \underline{g}$ results in a minimum normed least squares solution to Eq. (3-3). Also note that $[H]^+$ involves the matrix $[H][H]^*$ which is a discrete form of the gramian.

The continuous-discrete model can be treated in a similar manner by stating the problem as

$$\text{minimize} \quad \int f^2(\theta) d\theta$$

$$\text{subject to} \quad \left\| \int \underline{h}(\theta) f(\theta) d\theta - \underline{g} \right\|_2^2 = \left\| \underline{n} \right\|_2^2$$

where $\theta = (\xi, \eta)$ for notational simplicity. The criterion, $\omega(f(\theta))$, to be minimized is given by

$$\omega(\hat{f}(\theta)) = \int \hat{f}^2(\theta) d\theta + \gamma \left\{ \left\| \int \underline{h}(\theta) \hat{f}(\theta) d\theta - \underline{g} \right\|_2^2 + \left\| \underline{n} \right\|_2^2 \right\} \qquad (3\text{-}6)$$

Minimizing $\omega(\hat{f}(\theta))$ necessitates the use of the calculus of variations wherein we let $\hat{f}(\theta) = \hat{f}_0(\theta) + \epsilon\Delta(\theta)$. $\hat{f}_0(\theta)$ is the estimate minimizing Eq. (3-6) and $\Delta(\theta)$ is any function of θ. Thus we have that:

1) $\omega(\hat{f}(\theta))\Big|_{\epsilon=0}$ = minimum value

achieved at the $\hat{f}(\theta)$ such that

2) $\dfrac{\partial}{\partial \epsilon} \omega(\hat{f}(\theta))\Big|_{\epsilon=0} = 0$

and

3) $\dfrac{\partial^2}{\partial \epsilon^2} \omega(\hat{f}(\theta))\Big|_{\epsilon=0} \geq 0$

to guarantee a minimum. Substituting $\hat{f}(\theta) = \hat{f}_0(\theta) + \epsilon\Delta(\theta)$ into Eq. (3-6) we obtain

$$\omega(\hat{f}(\theta)) = \int (\hat{f}_0(\theta) + \epsilon\Delta(\theta))^2 d\theta + \gamma \underline{g}^T \underline{g} - 2\gamma \underline{g}^T \int \underline{h}(\theta)(\hat{f}_0(\theta) + \epsilon\Delta(\theta)) d\theta$$

$$+ \gamma \iint \underline{h}^*(\theta) \underline{h}(\eta)(\hat{f}_0(\theta) + \epsilon\Delta(\theta)) \cdot (\hat{f}_0(\eta) + \epsilon\Delta(\eta)) d\theta d\eta + \gamma \left\| \underline{n} \right\|_2^2 \qquad (3\text{-}7)$$

The first derivative with respect to ϵ results in

$$\frac{\partial w}{\partial \epsilon} w(\hat{\underline{f}}(\theta)) = 2\int (\hat{f}_0(\theta) + \epsilon\Delta(\theta))\Delta(\theta)d\theta - 2\gamma\, \underline{g}^T \int \underline{h}(\theta)\Delta(\theta)d\theta$$

$$+ 2\gamma \int h^*(\eta)(\hat{f}_0(\eta) + \epsilon\Delta(\eta))d\eta \int \underline{h}(\theta)\Delta(\theta)d\theta \qquad (3\text{-}8)$$

$$\frac{\partial^2}{\partial \epsilon^2} w(\hat{f}(\theta)) = 2\int \Delta^2(\theta)d\theta + 2\gamma \iint \underline{h}^*(\theta)\underline{h}(\eta)\Delta(\theta)\Delta(\eta)d\theta\, d\eta \geq 0$$

If $\gamma \geq 0$ we will have a minimum. Setting Eq. (3-8) equal to zero and evaluating at $\epsilon = 0$ we obtain a minimum at $\hat{f}_0(\theta)$ satisfying

$$\int \Delta(\theta) \{\hat{f}_0(\theta) - \gamma \underline{h}^*(\theta)\underline{g}(\theta) + \gamma \underline{h}^*(\theta)\int \underline{h}(\eta)\hat{f}_0(\eta)d\eta\}d\theta = 0$$

which will be satisfied if

$$\hat{f}_0(\theta) + \gamma \underline{h}^*(\theta)\int \underline{h}(\eta)\hat{f}_0(\eta)d\eta = \gamma \underline{h}^*(\theta)\underline{g} \qquad (3\text{-}9)$$

Notice that this expression is of the algebraic form

$$(I + \gamma \mathcal{R})\hat{f}_0(\eta) = \gamma \underline{h}^*(\theta)\underline{g}$$

where

$$\mathcal{R}\hat{f}_0(\eta) \triangleq \int \underline{h}^*(\theta)\underline{h}(\eta)\hat{f}_0(\eta)d\eta$$

which can also be written

$$\mathcal{R}\hat{f}_0(\eta) = \int K(\theta, \eta)\hat{f}_0(\eta)d\eta$$

where

$$K(\theta, \eta) = \underline{h}^*(\theta)\underline{h}(\eta)$$

so that we recognize Eq. (3-9) as a Fredholm integral equation of the second kind which can be solved by iterated kernels giving

$$\hat{f}_0(\eta) = \sum_{i=0}^{\infty} (-\gamma \mathcal{K})^i \delta \underline{h}^*(\theta)\underline{g}$$

since

$$\sum_{i=0}^{\infty} (-\gamma \mathcal{K})^i = (I + \gamma \mathcal{K})^{-1} \tag{3-10}$$

if $(I+\gamma\mathcal{K})^{-1}$ exists and $\|\gamma \mathcal{K}\| \leq 1$.

To obtain an explicit form for the solution and conditions for convergence note that

$$\mathcal{K}^2 \hat{f}_0(\eta) = \iint \underline{h}^*(\theta)\underline{h}(\xi)\underline{h}^*(\xi)\underline{h}(\eta)\hat{f}_0(\eta)d\eta d\xi$$

$$= \int h^*(\theta)[\Gamma]\underline{h}(\eta)\hat{f}_0(\eta)d\eta$$

implying that the kernel of this transformation $K^2(\theta, \eta)$ is given by

$$K^2(\theta, \eta) = \underline{h}^*(\theta)[\Gamma]\underline{h}(\eta)$$

and that $\mathcal{K}^2\hat{f}(\eta)$ can be written

$$\mathcal{K}^n \hat{f}_0(\eta) = \int \underline{h}^*(\theta)[\Gamma]^{n-1}\underline{h}(\eta)\hat{f}_0(\eta)d\eta$$

and that $K^n(\theta, \eta) = \underline{h}^*(\theta)[\Gamma]^{n-1}\underline{h}(\eta)$. Thus

$$\hat{f}_0(\theta) = \gamma \lim_{n\to\infty} \left\{ \sum_{i=0}^{n} (-\gamma \mathcal{K})^i \right\} \underline{h}^*(\eta)\underline{g}$$

$$= \gamma \lim_{n\to\infty} \left\{ \sum_{i=0}^{n} (-\gamma)^i \int \underline{h}^*(\theta)[\Gamma]^{i-1}\underline{h}(\eta)\underline{h}^*(\eta)d\eta \right\} \underline{g}$$

$$= \gamma \lim_{n\to\infty} \left\{ \sum_{i=0}^{n} (-\gamma^i)\, \underline{h}^*(\theta)[\Gamma]^i \right\} \underline{g}$$

$$\hat{f}_0(\theta) = \gamma\, \underline{h}^*(\theta)[[I] + \gamma[\Gamma]]^{-1}\, \underline{g} \qquad (3\text{-}11)$$

if the limit exists. However since $[I]$ and $[\Gamma]$ are both matrices and $[\Gamma]$ has nonnegative eigenvalues only, every eigenvalue of $[I] + \gamma[\Gamma]$ will be nonzero for $\gamma > 0$ implying that $[[I] + \gamma[\Gamma]]^{-1}$ exists. Thus we can rearrange the right side of eq. (3-11) and obtain the estimate, $\hat{f}(\xi, \eta) = \hat{f}_0(\theta)$

$$\hat{f}(\xi, \eta) = \underline{h}^*(\xi, \eta)\left[\frac{1}{\gamma}[I] + [\Gamma]\right]^{-1} \underline{g} \qquad (3\text{-}12)$$

Note the similarity between (3-11) and (3-9), and that a Moore-Penrose pseudo inverse will be obtained by letting $\gamma \to \infty$ so that

$$\hat{f}(\xi, \eta) = \underline{h}^*(\xi, \eta)[\Gamma]^{+}\underline{g}$$

which will be

$$\hat{f}(\xi, \eta) = h^*(\xi, \eta)[\Gamma]^{-1}\underline{g}$$

if $[\Gamma]$ is nonsingular.

A criterion for the convergence of the series in eq. (3-10) is that $\|\gamma\mathcal{H}\| = \|\gamma\|\,\|\mathcal{H}\| < 1$. This has physical significance for passive imaging systems since the output image energy will be less than the input image energy for such systems. Thus

$$\| g\|_2^2 \leq \| f \|_2^2$$

so that

$$\frac{\|\mathcal{K}f\|_2}{\|f\|_2} \leq 1$$

resulting in

$$\|\mathcal{K}\|_2 = \sup_{\|f\| \neq 0} \frac{\|\mathcal{K}f\|_2}{\|f\|_2} \leq 1$$

At this point it is interesting to investigate the relative error in the estimate $\hat{f}(\xi, \eta)$ is of the form

$$\hat{f}(\xi, \eta) = \langle h^*(\xi, \eta), \alpha \rangle$$

where

$$\underline{\alpha} = \left[\frac{1}{\gamma}[I] + [\Gamma]\right]^{-1} \underline{g}$$

If we let $\underline{\alpha}' = [U]^{-1}\underline{\alpha}$ so that

$$\hat{f}(\xi, \eta) = \underline{\varphi}^*(\xi, \eta), \underline{\alpha}'$$

where the $\varphi_i^*(\xi, \eta)$ are orthogonal and $\underline{\varphi}^*(\xi, \eta) = \underline{h}^*(\xi, \eta)[U]$, then the relative error in $\hat{f}(\xi, \eta)$ will be given by $\frac{\|\delta \underline{\alpha}'\|}{\|\underline{\alpha}'\|}$. This can be accomplished by expanding in the eigenspace of $[\Gamma]$ by letting $[U]$ be such that

$$[U]^*[\Gamma][U] = [\Lambda]$$

where $[\Lambda]$ is the diagonal matrix of eigenvalues of $[\Gamma]$, and $[U]^*[U] = [I]$. Thus

$$\underline{\varphi}(\xi, \eta) = [U]^* \underline{h}(\xi, \eta)$$

and the solution of the continuous-discrete imaging equation will be a solution up to the degrees of freedom of $[\Gamma]$. Then we note that

$$\iint_R \underline{\varphi}(\xi, \eta)\underline{\varphi}^*(\xi, \eta)d\xi d\eta = \iint_R [U]^* \underline{h}(\xi, \eta)\underline{h}^*(\xi, \eta)[U]d\xi d\eta$$

$$= [U]^*[\Gamma][U]$$

$$= [\Lambda],$$

i.e. the set $\{\varphi(\xi, \eta)\}$ is orthogonal. Then $\underline{\alpha}'$ is as follows

$$\underline{\alpha}' = [U]^* \left[\frac{1}{\gamma}[I] + [\Gamma]\right]^{-1} \underline{g}$$

$$= \left[\frac{1}{\gamma}[U] + [U][\Lambda][U]^*[U]\right]^{-1} \underline{g}$$

$$= \left[\frac{1}{\gamma}[U] + [U][\Lambda]\right]^{-1} \underline{g}$$

If $[\Gamma]$ is nonsingular the pseudo inverse solution is obtained by letting $\gamma \to \infty$ so that

$$\underline{\alpha}' = [\Lambda]^{-1}[U]^* \underline{g}$$

It is easily verified that in terms of $\underline{\varphi}(\xi, \eta)$

$$\hat{f}(\xi, \eta) = \langle \underline{\varphi}^*(\xi, \eta), \underline{\alpha}' \rangle = \underline{h}^*(\xi, \eta)[\Gamma]^{-1} \underline{g} \qquad (3\text{-}13)$$

The results of equation (3-13) indicate an ill-conditioning problem in which an upper bound on the relative error $\|\delta \underline{a}'\| / \|\underline{a}'\|$ can be found to be

$$\frac{\|\delta \underline{a}'\|}{\|\underline{a}'\|} \leq \frac{|\lambda_{max}|}{|\lambda_{min}|} \frac{\|\underline{n}\|}{\|\underline{g}\|} .$$

In general then we can obtain our estimates for $\hat{f}(\xi, \eta)$ in the eigenspace of the point spread functions with the same results, and can generalize to other quadratic constraints by letting $[C]$ be a nonsingular constraint matrix and postulate the problem as follows.

$$\text{minimize} \quad \underline{a}^* [C] \underline{a}$$

$$\text{subject to} \quad \|\underline{g} - [U][\Lambda]\underline{a}\|^2 = 0 .$$

The criterion to be minimized becomes

$$\underline{a}^* [C] \underline{a} + \gamma \|\underline{g} - [U][\Lambda]\underline{a}\|^2$$

when γ is the usual Lagrange multiplier. Performing this minimization results in a constrained least squares solution of

$$\hat{\underline{a}} = (\gamma^{-1}[C] + [\Lambda]^2)^{-1}[\Lambda][U]^* \underline{g} \qquad (3\text{-}14a)$$

and

$$\hat{f}(\xi, \eta) = \hat{\underline{a}}^* [U]^* \underline{h}(\xi, \eta) . \qquad (3\text{-}14b)$$

Proper conditions on $[C]$ and γ result in the pseudoinverse reconstruction and are developed elsewhere [3-5, 3-6].

3.2 Separable Kernels

In image processing we are often confronted with very large matrices that make implementation of the algorithms of section 3.1 difficult if not impossible. If it is noted that if a sampled image data

point is obtained at each coordinate tuple (x_i, y_i) and that there are N_x and N_y samples in the x and y directions respectively, then the image vector $\underline{g}$ is an $N_x \cdot N_y \times 1$ column vector and $[\Gamma]$ will be a $N_x N_y \times N_x N_y$ matrix. Thus for images of even a moderate dimension, say 100 by 100, $[\Gamma]$ will be a matrix whose dimension is $10^4 \times 10^4$ containing 10^8 elements. Clearly some structure in $[\Gamma]$ must be found that allows a simplification.

The structure investigated in this section will be that of a separable kernel. By a separable kernel it is meant that $h_i(\xi, \eta)$ can be written as the product of two functions as follows:

$$h_i(\xi, \eta) = h_i^{(1)}(\xi) h_i^{(2)}(\eta) \qquad \forall i = 1, 2, \ldots, N_x \cdot N_y \tag{3-15}$$

This subject is of interest because it often occurs in practice and that separability allows a significant computation reduction over the general case.

From (3-15) the kernel vector $\underline{h}(\xi, \eta)$ of (1-2) becomes

$$\underline{h}(\xi, \eta) = \underline{h}^{(1)}(\xi) \otimes \underline{h}^{(2)}(\eta)$$

where $\otimes$ denotes the Kronecker, or direct product between $\underline{h}^{(1)}$ and $\underline{h}^{(2)}$ thus:

$$\underline{h}(\xi,\eta) = \begin{bmatrix} h_1^{(1)}(\xi)\cdot\underline{h}^{(2)}(\eta) \\ \hline h_2^{(1)}(\xi)\cdot\underline{h}^{(2)}(\eta) \\ \hline \vdots \\ h_{N_x}^{(1)}(\xi)\cdot\underline{h}^{(2)}(\eta) \end{bmatrix}_{N_x\cdot N_y x 1}$$

The computational savings provided by kernel separability lies in the fact that $[\Gamma]$ can be shown to be the Kronecker product of the two gramians $[\Gamma]^{(1)}$ and $[\Gamma]^{(2)}$ as follows

$$[\Gamma] = [\Gamma]^{(1)} \otimes [\Gamma]^{(2)}$$

where

$$[\Gamma]^{(1)} = \int_{R_\xi} \underline{h}^{(1)}(\xi)\underline{h}^{(1)*}(\xi)d\xi \qquad (3\text{-}16a)$$

$$[\Gamma]^{(2)} = \int_{R_\eta} \underline{h}^{(2)}(\eta)\underline{h}^{(2)*}(\eta)d\eta \qquad (3\text{-}16b)$$

To show this consider,

$$[\Gamma] = \iint_R \underline{h}(\xi,\eta)\underline{h}^*(\xi,\eta)d\xi d\eta$$

which for separable kernels becomes

$$[\Gamma] = \iint_R \{\underline{h}^{(1)}(\xi)\otimes\underline{h}^{(2)}(\eta)\}\{\underline{h}^{(1)}(\xi)\otimes\underline{h}^{(2)}(\eta)\}^* d\xi d\eta$$

From Bellman [3-7] we know that

$$\{\underline{h}^{(1)}(\xi)\otimes\underline{h}^{(2)}(\eta)\}^* = \underline{h}^{(1)}(\xi)^*\otimes\underline{h}^{(2)}(\eta)^*$$

thus

$$[\Gamma] = \iint_R \{\underline{h}^{(1)}(\xi) \otimes \underline{h}^{(2)}(\eta)\} \{\underline{h}^{(1)*}(\xi) \otimes \underline{h}^{(2)*}(\eta)\} \, d\xi d\eta$$

But by the algebra of Kroneckers [3-7]

$$\{\underline{h}^{(1)}(\xi) \otimes \underline{h}^{(2)}(\eta)\} \{\underline{h}^{(1)*}(\xi) \otimes \underline{h}^{(2)*}(\eta)\} = \{\underline{h}^{(1)}(\xi)\underline{h}^{(1)*}(\xi)\} \otimes \{\underline{h}^{2}(\eta)\underline{h}^{(2)*}(\eta)\}$$

so that

$$[\Gamma] = \int_{R_\xi} \underline{h}^{(1)}(\xi)\underline{h}^{(1)*}(\xi)d\xi \otimes \int_{R_\eta} \underline{h}^{(2)}(\eta)\underline{h}^{(2)*}(\eta)d\eta$$

$$[\Gamma] = [\Gamma]^{(1)} \otimes [\Gamma]^{(2)} \tag{3-17}$$

Clearly it is a significantly simpler task to diagnoalize two matrices of dimension N_x and N_y each than one $N_xN_y \times N_xN_y$ matrix. By way of illustration, since $[\Gamma]^{(1)}$ and $[\Gamma]^{(2)}$ are Hermitian, they are diagonalizable as follows

$$[\Lambda]^{(1)} = [U]^{(1)*}[\Gamma]^{(1)}[U]^{(1)}$$

$$[\Lambda]^{(2)} = [U]^{(2)*}[\Gamma]^{(2)}[U]^{(2)}$$

then from the algebra of Kroneckers [3-7]

$$[\Lambda] = [\Lambda]^{(1)} \otimes [\Lambda]^{(2)} = [U]^{(1)*}[\Gamma]^{(1)}[U]^{(1)} \otimes [U]^{(2)*}[\Gamma]^{(2)}[U]^{(2)}$$

$$= [U]^{(1)*} \otimes [U]^{(2)*}[\Gamma]^{(1)} \otimes [\Gamma]^{(2)}[U]^{(1)} \otimes [U]^{(2)}$$

and finally from Bellman [3-7] we have that

$$[[\Gamma]^{(1)} \otimes [\Gamma]^{(2)}]^{-1} = [\Gamma]^{(1)^{-1}} \otimes [\Gamma]^{(2)^{-1}}$$

Consequently, the coefficients for the estimate of f become

$$\underline{\alpha} = [\Lambda]^{(1)^{-1}} [U]^{(1)^*} \otimes [\Lambda]^{(2)^{-1}} [U]^{(2)^*} \underline{g} \tag{3-18a}$$

$$\hat{f}(\xi, \eta) = \underline{g}^* [\Gamma]^{(1)^{-1}} \underline{h}^{(1)}(\xi) \otimes [\Gamma]^{(2)^{-1}} \underline{h}^{(2)}(\eta) \tag{3-18b}$$

For the case where $[\Gamma]$ is singular and the constrained least squares solutions are used, we have the following

$$\underline{\alpha} = [\gamma^{-1}[C] + [\Lambda]^{(1)} \otimes [\Lambda]^{(2)}]^{-1} [\Lambda]^{(1)} [U]^{(1)} \otimes [\Lambda]^{(2)} [U]^{(2)} \underline{g}$$

If we let $[C]$ be separable and nonsingular as is often the case (take for example $[C] = [I]$), then

$$\hat{\underline{\alpha}} = [\gamma^{-1}[C]^{(1)} \otimes [C]^{(2)} + [[\Lambda]^{(1)} \otimes [\Lambda]^{(2)}]^2]^{-1} [\Lambda]^{(1)} [U]^{(1)} \otimes [\Lambda]^{(2)} [U]^{(2)} \underline{g}$$

$$= [C]^{(1)^{-1}} \otimes [C]^{(2)^{-1}} [\gamma^{-1}[I] + [C]^{(2)^{-1}} [\Lambda]^{(1)^2} \otimes [C]^{(2)^{-1}} [\Lambda]^{(2)^2}]^{-1}$$

$$\cdot [\Lambda]^{(1)} [U]^{(1)} \otimes [\Lambda]^{(2)} [U]^{(2)} \underline{g}$$

$$\hat{\underline{f}} = \hat{\underline{\alpha}}^* [U]^{(1)^*} \otimes [U]^{(2)^*} \underline{h}_1(\xi) \otimes \underline{h}_2(\eta)$$

Thus we see that the constrained least squares solution to a problem with separable $[\Gamma]$ and $[C]$ matrices involves the diagonalization and inversion of far smaller dimensioned matrices. The discussion of another attractive structure is deferred until the section dealing with the tomography gramian.

3.3 Gramian Eigenvalues and System Eigenvalues

In the preceding section the degrees of freedom of the continuous-discrete model of a linear imaging system has been identified with the number of effectively nonzero eigenvalues of its gram-matrix. This prompts the question of exactly what is the relationship between the gramian eigenvalues and the original continuous system eigenvalues, for it would certainly seem reasonable that they exhibit a similarity.

In this section we will apply a result of Keller [3-8] that will allow us to relate the eigenvalues of $[\Gamma]$ to the singular values of a four continuous variable kernel $h(x, y;\xi, \eta)$ which is not necessarily hermitian. To do this we form the auxiliary kernels $h'(x, y;\xi, \eta)$ and $h''(x, y;\xi, \eta)$ and formulate an outer product expansion for $h(x, y;\xi, \eta)$. Under the original assumption, namely, that $h(x, y;\xi, \eta)$ is square integrable, i.e.,

$$\iiiint_R |h(x, y;\xi, \eta)|^2 \, dxdyd\xi d\eta < \infty$$

and defining the adjoint kernel $h^*(\xi, \eta;x, y)$ as

$$h^*(\xi, \eta;x, y) = h(x, y;\xi, \eta)$$

then the auxiliary kernels are:

$$\begin{aligned} h'(x, y;\xi, \eta) &= \iint h(x, y;\rho, \theta)h^*(\rho, \theta;\xi, \eta)d\rho d\theta \\ &= \iint h(x, y;\rho, \theta)h(\xi, \eta;\rho, \theta)d\rho d\theta \end{aligned} \tag{3-19}$$

$$h''(x, y;\xi, \eta) = \iint h^*(x, y;\rho, \theta)h(\rho, \theta;\xi, \eta)d\rho d\theta$$

$$= \iint h(\rho, \theta;x, y)h(\rho, \theta;\xi, \eta)d\rho d\theta \qquad (3\text{-}20)$$

and

$$\mu_j\varphi_j(x, y) = \iint h'(x, y;\xi, \eta)\varphi_j(\xi, \eta)d\xi d\eta \qquad (3\text{-}21)$$

$$\mu_j\psi_j(x, y) = \iint h'(x, y;\xi, \eta)\psi_j(\xi, \eta)d\xi d\eta \qquad (3\text{-}22)$$

$$|\mu_j|^{\frac{1}{2}}\varphi_j(x, y) = \iint h(x, y;\xi, \eta)\psi_j(\xi, \eta)d\xi d\eta \qquad (3\text{-}23)$$

$$|\mu_j|^{\frac{1}{2}}\psi_j(x, y) = \iint h^*(x, y;\xi, \eta)\varphi_j(\xi, \eta)d\xi d\eta \qquad (3\text{-}24)$$

Then $h(x, y;\xi, \eta)$ admits of the following expansion

$$h(x, y;\xi, \eta) = \sum_{k=1}^{\infty} |\mu_k|^{\frac{1}{2}}\varphi_k(x, y)\psi_k^*(\xi, \eta) \qquad (3\text{-}25)$$

If we sample $h'(x, y;\xi, \eta)$ at the coordinate tuples $\underline{x}^{(i)} = (x_i, y_i)$ and $\underline{x}^{(j)} = (x_j, y_j)$, $i = 1, \ldots, N_xN_y$, $j = 1, \ldots, N_xN_y$, it should be clear that

$$h'(\underline{x}^{(i)}, \underline{x}^{(j)}) = \gamma_{ij}$$

Let $\mathscr{L}_{Nf}$ be a two dimensional quadrature rule approximating $\iint_R f(x,y)dxdy$ given by:

$$\mathscr{L}_{N_x \cdot N_y}f = \sum_{i=1}^{N_x}\sum_{j=1}^{N_y} f(x_i y_j)w_{ij}, \quad \sum_{i}^{N_x}\sum_{i}^{N_y} w_{ij} = 1, \quad N = N_x \cdot N_y$$

$$\approx \iint_R f(x, y)dxdy$$

Sampling f(x, y) on a cartesian product grid we can separate w_{ij} into a product of two weights $w_i^{(x)}, w_j^{(y)}$ [3-9] where these are the respective quadrature weights for one dimensional integrals along the x and y directions. The quadrature rule then becomes

$$\underset{N_x X N_y}{\mathcal{I}f} = \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} f(x_i, y_j) w_i^{(x)} w_j^{(y)}$$

To simplify the quadrature notation let $\underline{x} = (x, y)$, $\underline{\theta} = (\xi, \eta)$ then the samples become:

$$\underline{x}^{(i)} = (x_i, y_j) \qquad i = 1, 2, \ldots, N_x N_y \qquad \underline{x}^{(i)} \epsilon R^2$$

$$\underline{\theta}^{(j)} = (\xi_j, \eta_j) \qquad j = 1, 2, \ldots, N_x N_y \qquad \underline{\theta}^{(j)} \epsilon R^2$$

$$w^{(j)} = w_j^{(x)} \cdot w_j^{(y)} \qquad j = 1, 2, \ldots, N_x N_y$$

Then the quadrature approximation to (3-21) becomes

$$\mu_k \varphi_k(\underline{x}^{(i)}) \approx \sum_{\underline{\theta}^{(j)}} h'(\underline{x}^{(i)}; \underline{\theta}^{(j)}) w^{(j)} \varphi_k(\underline{\theta}^{(j)}) \qquad (3\text{-}26)$$

or in vector form

$$\mu_k \underline{\varphi}_k \approx [\Gamma][w] \underline{\varphi}_k$$

where $[w]$ is a diagonal matrix of quadrature weights. If we let λ_k be the k^{th} eigenvalue of $[\Gamma][w]$ and $\underline{U}_k$ its associated eigenvector, then

$$\lambda_k \underline{U}_k = [\Gamma][w] \underline{U}_k$$

considering those quadrature rules with weights greater than zero, and such that

$$\lim_{\substack{N_x \to \infty \\ N_y \to \infty}} \; \mathop{\not\int\!\!\int}_{N_x \times N_y} f = \iint f(x, y)\,dx\,dy$$

Keller has shown that for every eigenvalue λ_k of $[\Gamma][w]$ there exists an eigenvalue μ_k in (3-21) such that

$$|\lambda_k - \mu_k| \leq [A_R + o(1)]^{\frac{1}{2}} \max_{\underline{x}} |\epsilon(\underline{x})| \tag{3-27}$$

where

$$|\epsilon(\underline{x})| = \left| \iint_R h'(\underline{x}; \xi, \eta)\, \varphi_k(\xi, \eta)\, d\xi\, d\eta - \mathop{\not\int\!\!\int}_{N_x N_y} h' \varphi_k \right|$$

and A_R is the area of R. By letting R equal the unit rectangle and employing a rectangular integration rule with evenly spaced samples such that $w^{(j)} = \frac{1}{N_x \cdot N_y}$, then the eigenvalues of $[\Gamma]$ converge to a constant, (dependent on the number of samples), times some eigenvalue in eq. (3-21).

Since the error in the estimate $\hat{f}$ in eq. (3-6) is upper bounded by the condition number of $[\Gamma]$, $\left|\frac{\lambda_{max}}{\lambda_{min}}\right|$, the actual value of the constant is unimportant. What is important is that the shape of the eigenvalue map of $[\Gamma][w]$ will be in good agreement with the spectrum of the kernel in the continuous-continuous model. These results serve to put on a firm footing what should be our intuitive feeling towards

the image restoration problem, namely, that the difficulty in restoration is not so much a function of the sampling method used (for any one consistent with a quadrature rule $\mathcal{J}_n$ such that $\mathcal{J}_n f \to \int f d\underline{x}$ could be used), but is innate to the original continuous-continuous formulation of the image restoration problem.

3.4 Eigenvalue Error Bounds for Separable Kernels

For separable kernels actual numerical bounds on the error $|\lambda_k - \mu_k|$ can be obtained. Wielandt [3-10] investigated and developed bounds for the errors in estimating the eigenvalues of a hermitian operator in one dimension using various quadrature rules. Since the two dimensional separable problem reduces to two one dimensional problems, it is relatively straightforward to extend those results to the two dimensional separable problem. In the previous section the eigenvalues of $[\Gamma][w]$ were investigated where $w_i = \frac{1}{N_x N_y}$ for all i. This presents no difficulties in obtaining $\hat{f}$ in Eq. (3-5) since we simply scale both sides by $\frac{1}{N_x N_y}$. The results of the previous section are also valid for other quadrature rules where neither the sampling subintervals nor the quadrature weights are constant. To deal with this we must multiply both sides of Eq. (3-13) by the appropriate quadrature weight matrix $[w]$ and obtain

$$[w]\underline{g} = [w][\Gamma][U]\underline{\alpha} \qquad (3\text{-}28)$$

If we restrict our rules to those with the property

$$\frac{\max w_i}{\min w_i} < M \quad \text{as } N_x N_y \to \infty$$

then the invertibility of (3-13) is dependent on $[\Gamma]$ and not $[w]$. Now $[w][\Gamma]$ is similar to $[\Gamma][w]$ which is itself similar to the hermitian matrix $[w]^{\frac{1}{2}}[\Gamma][w]^{\frac{1}{2}}$ implying that these matrices all have the same eigenvalues. Thus we can analyze the separable gramian eigenvalues using Wielandt's results applied to $[\Gamma][w]$.

As before, we form the auxiliary kernel $h'(x, y;\xi, \eta)$ as follows

$$\begin{aligned} h'(x, y;\xi, \eta) &= \iint_R h(x, y;\rho, \theta)h(\xi, \eta;\rho, \theta)d\rho d\theta \\ &= \iint_R h^{(1)}(x, \rho)h^{(2)}(y, \theta)h^{(1)}(\xi, \rho)h^{(2)}(\eta, \theta)d\rho d\theta \\ &= \iint_{R_x} h^{(1)}(x, \rho)h^{(1)}(\xi, \rho)d\rho \int_{R_y} h^{(2)}(y, \theta)h^{(2)}(\eta, \theta)d\theta \end{aligned}$$

Then

$$h'(x, y;\xi, \eta) = h^{(1)'}(x, \xi)h^{(2)'}(y, \eta) \tag{3-29}$$

For a separable kernel Eq. (3-21) becomes

$$\mu_j^{(1)}\mu_k^{(2)}\varphi_j^{(1)}(x)\varphi_k^{(2)}(y) = \int_{R_x} h^{(1)'}(x, \xi)\varphi_j^{(1)}(\xi)d\xi \int_{R_y} h^{(2)'}(y, \eta)\varphi_k^{(2)}(\eta)d\eta \tag{3-30}$$

where

$$\mu_i^{(1)}\varphi_i^{(1)}(x) = \int_{R_x} h^{(1)'}(x, \xi)\varphi_i^{(1)}(\xi)d\xi \tag{3-31}$$

$$\mu_j^{(2)}\varphi_j^{(2)}(y) = \int_{R_y} h^{(2)'}(y,\eta)\varphi_j^{(2)}(\eta)d\eta \qquad (3-32)$$

The appropriate quadrature rules with x and y sampled on a cartesian produce grid yield

$$\lambda_i^{(1)}\underline{U}_i^{(1)} = [\Gamma]^{(1)}[w]^{(1)}\underline{U}_i^{(1)} \qquad (3-33)$$

$$\lambda_j^{(2)}\underline{U}_j^{(2)} = [\Gamma]^{(2)}[w]^{(2)}\underline{U}_j^{(2)} \qquad (3-34)$$

which implies

$$\lambda_i^{(1)}\lambda_j^{(2)} = \underline{U}_i^{(1)}\otimes\underline{U}_j^{(2)}\cdot[\Gamma]^{(1)}\otimes[\Gamma]^{(2)}[w]^{(1)}\otimes[w]^{(2)}\cdot\underline{U}_i^{(1)}\otimes\underline{U}_j^{(2)}$$

The desired result is to bound

$$|\lambda_i^{(1)}\lambda_j^{(2)}-\mu_i^{(1)}\mu_j^{(2)}| \ .$$

To do this consider that

$$\lambda_i^{(1)}\lambda_j^{(2)}-\mu_i^{(1)}\mu_j^{(2)} = (\lambda_i^{(1)}-\mu_i^{(1)})(\lambda_j^{(2)}-\mu_j^{(2)}) + \mu_i^{(1)}(\lambda_j^{(2)}-\mu_j^{(2)}) + \mu_j^{(2)}(\lambda_i^{(1)}-\mu_i^{(1)})$$

Since $\mu_j^{(2)} \geq 0$ $\forall j$ and $\mu_i^{(1)} \geq 0$ $\forall i$, taking the triangle inequality implies that

$$|\lambda_i^{(1)}\lambda_j^{(2)}-\mu_i^{(1)}\mu_j^{(2)}| \leq |\lambda_i^{(1)}-\mu_i^{(2)}|\,|\lambda_j^{(2)}-\mu_j^{(2)}| + \mu_i^{(1)}|\lambda_j^{2}-\mu_j^{(2)}| + \mu_j^{(2)}|\lambda_i^{(1)}-\mu_i^{(1)}|$$

Since we can consider that $\mu_i^{(1)} \leq 1$ and $\mu_j^{(2)} \leq 1$ we obtain

$$|\lambda_i^{(1)}\lambda_j^{(2)}-\mu_i^{(1)}\mu_j^{(2)}| \le |\lambda_i^{(1)}-\mu_i^{(1)}||\lambda_j^{(2)}-\mu_j^{(2)}| + |\lambda_j^{(2)}-\mu_j^{(2)}| + |\lambda_i^{(1)}-\mu_i^{(1)}| \tag{3-35}$$

Now Wielandt's result applied to $h^{(1)'}(x, \xi)$ for rectangular integration with N_x points from $[0, 1]$ equally spaced such that $w_i = \frac{1}{N_x-1}$, for all i is such that if $h^{(1)'}(x, \xi)$ is a hermitian kernel such that $|\frac{\partial}{\partial x} h^{(1)'}(x, \varepsilon)| \le L^{(1)}$ for all $\varepsilon \in [0, 1]$ then for every $\lambda_i^{(1)}$, an eigenvalue of $[\Gamma]^{(1)} \frac{1}{N_x}$, there exists a $\mu_i^{(1)}$, an eigenvalue of (3-27), such that

$$|\lambda_i^{(1)}-\mu_i^{(1)}| \le \frac{1.08L^{(1)}}{N_x}$$

The results for $[\Gamma]^{(2)}$ are identical. Applying this to (3-35) we have for rectangular integration with N_x and N_y evenly spaced points in the x and y directions that

$$|\lambda_i^{(1)}\lambda_j^{(2)}-\mu_i^{(1)}\mu_j^{(2)}| \le \frac{(1.08)^2}{N_x N_y} L^{(1)}L^{(2)} + \frac{1.08L^{(1)}}{N_x} + \frac{1.08L^{(2)}}{N_y} \tag{3-36}$$

where

$$\left|\frac{\partial}{\partial x} h^{(1)'}(x, \xi)\right| \le L^{(1)} \quad \forall \varepsilon \in [0, 1]$$

and

$$\left|\frac{\partial}{\partial y} h^{(2)'}(y, \eta)\right| \le L^{(2)} \quad \forall \eta \in [0, 1]$$

If the kernels $h^{(1)'}(x, \varepsilon)$ and $h^{(2)'}(y, \eta)$ are twice differentiable such that

$$\left|\frac{\partial^2}{\partial x^2} h^{(1)'}(x, \varepsilon)\right| \le L^{(1)} \quad \forall \varepsilon \in [0, 1]$$

and

$$\left|\frac{\partial^2}{\partial y^2}\, h^{(2)'}(y,\eta)\right| \le L^{(2)} \quad \forall \eta \in [0,1]$$

we can apply Wielandt's result for Simpson's rule to (3-35) and obtain

$$|\lambda_i^{(1)}\lambda_j^{(2)} - \mu_i^{(1)}\mu_j^{(2)}| \le \frac{(.75)^2 L^{(1)} L^{(2)}}{(N_x-1)^2(N_y-1)^2} + \frac{.75L^{(1)}}{(N_x-1)^2} + \frac{.75L^{(2)}}{(N_y-1)^2} \qquad (3\text{-}37)$$

For Gaussian quadrature over a cartesian product grid with N_x and N_y points along the x and y directions respectively, if the auxiliary kernels $h^{(1)'}(x,\xi)$ and $h^{(2)'}(y,\eta)$ possess continuous p^{th} partials

$$\frac{\partial^{p1}}{\partial x^{p1}}\, h^{(1)'}(x,\xi) \quad \text{and} \quad \frac{\partial^{p2}}{\partial y^{p2}}\, h^{(2)'}(y,\eta)$$

then for $N_x \ge p_1$ and $N_y \ge p_2$ using Wielandt's results for Gaussian quadrature it can be shown

$$\begin{aligned}|\lambda_i^{(1)}\lambda_j^{(2)} - \mu_i^{(1)}\mu_j^{(2)}| \le\ & 16\left(\frac{d_1}{N_x-1}\right)^{p1}\left(\frac{d_2}{N_y-1}\right)^{p2} \max_{x,\xi}\left|\frac{\partial^{p1}}{\partial x^{p1}}\, h^{(1)'}(x,\xi)\right| \\ & \cdot \max_{y,\eta}\left|\frac{\partial^{p2}}{\partial y^{p2}}\, h^{(2)'}(y,\eta)\right| \\ & + 4\left(\frac{d_1}{N_x-1}\right)^{p1} \max_{x,\xi}\left|\frac{\partial^{p1}}{\partial x^{p1}}\, h^{(1)'}(x,\xi)\right| \\ & + 4\left(\frac{d_2}{N_y-1}\right)^{p2} \max_{y,\eta}\left|\frac{\partial^{p2}}{\partial y^{p2}}\, h^{(2)'}(y,\eta)\right|\end{aligned}$$

where

$$d_1, d_2 = \begin{cases} 1.92 & \text{if } p_1, p_2 \leq 6 \\ 4.0 & \text{if } p_1, p_2 \geq 7 \end{cases}$$

These results are summarized in Table 3.1 for each of the aforementioned quadrature rules in both one and two dimensions.

3.5 Conclusions

The concept of the gramian eigenvalues in determining the degrees of freedom of the imaging system have been developed. The gramian eigenvalues have been related to the system singular values by way of auxiliary kernels and have been shown to be closely related to those singular values. Since for every L_2 kernel these singular values tend to zero, the numerical instability in digital image restoration arises from the original continuous-continuous model for imaging systems. For imaging system kernels that are separable, actual numerical bounds for the distance between the gramain eigenvalues and the square of the system singular values were found. This should be quite useful in obtaining upper bounds for the number of effectively independent samples one can obtain for separable imaging systems.

Table 3.1. Summary of Eigenvalue Error Bounds for Various Quadrature Rules.

INTEGRATION METHOD	1 DIMENSIONAL BOUND $\lvert\lambda_i-\mu_i\rvert \le$	CORRESPONDING 2 DIMENSIONAL SEPARABLE BOUND $\lvert\lambda_i^{(1)}\lambda_i^{(2)}-\mu_i^{(1)}\mu_i^{(2)}\rvert \le$
Rectangular Integration	$\frac{1.08L}{N}$	$\frac{(1.08)^2 L^{(1)}L^{(2)}}{N_x N_y} + 1.08\left[\frac{L^{(1)}}{N_x} + \frac{L^{(2)}}{N_y}\right]$
Simpson's Rule	$\frac{.75L}{(N-1)^2}$	$\frac{(.75)^2 L^{(1)}L^{(2)}}{(N_x-1)^2(N_y-1)^2} + .75\left[\frac{L^{(1)}}{(N_x-1)^2} + \frac{L^{(2)}}{(N_y-1)^2}\right]$
Gaussian Quadrature	$4\left(\frac{d}{N-1}\right)^P \max_{x,\xi}\left\lvert\frac{\partial^P}{\partial x^P} h(x,\xi)\right\rvert$ $d = \begin{cases} 1.92 & \text{if } P \le 6 \\ 4.0 & \text{if } P \ge 7 \end{cases}$	$16\left(\frac{d_1}{N_x-1}\right)^{P_1}\left(\frac{d_2}{N_y-1}\right)^{P_2} \max_{x,\xi}\left\lvert\frac{\partial^{P_1}}{\partial x^{P_1}} h^{(1)'}(x,\xi)\right\rvert$ $\cdot \max_{y,\eta}\left\lvert\frac{\partial^{P_2}}{\partial y^{P_2}} h^{(2)'}(y,\eta)\right\rvert$ $+ 4\left(\frac{d_1}{N_x-1}\right)^{P_1} \max_{x,\xi}\left\lvert\frac{\partial^{P_1}}{\partial x^{P_1}} h^{(1)'}(x,\xi)\right\rvert$ $+ 4\left(\frac{d_2}{N_y-1}\right)^{P_2} \max_{y,\eta}\left\lvert\frac{\partial^{P_2}}{\partial y^{P_2}} h^{(2)'}(y,\eta)\right\rvert$

Chapter 4

PROJECTION IMAGING AND THE GRAMIAN[1]

4.1 Continuous-Discrete Model

We are now in a position to apply these analytical methods obtained through the continuous-discrete imaging model to the transaxial tomographic projection imaging system. In the ideal case the output of such a system, as modeled in Figure 4.1, is a projection, $\rho(r, \theta)$, related to the original object, $f(\xi, \eta)$, as follows:

$$\rho(r, \theta) = \iint_R f(\xi, \eta) h(\theta, r; \xi, \eta) d\xi d\eta . \qquad (4\text{-}1)$$

In any real system finite detector width, beam spread, scattering and other phenomena cause the line projection to be distorted. Thus it is appropriate to consider the distorted projection, $\rho(r, \theta)$ to be related to $f(\xi, \eta)$ through the general blur $h(\theta, r; \xi, \eta)$. In the important case where the blur is space invariant along r, and independent of θ, the projection becomes

$$\rho(r, \theta) = \iint_R f(\xi, \eta) b(\xi \cos \theta + \eta \sin \theta - r) d\xi d\eta . \qquad (4\text{-}2a)$$

[1]This and portions of the proceeding chapter are summarized in McCaughey, D. and H. C. Andrews, "Degrees of Freedom in Projection Imaging," IEEE Transactions on Acoustics, Speech and Signal Processing, vol. 25, no. 1, February, 1977.

where $b(\xi \cos\theta + \eta \sin\theta - \eta)$ is $b(r)$ evaluated along the line $(\xi \cos\theta + \eta \sin\theta - r)$. For an infinitely narrow blur we have the ideal imaging system, and the PSF kernel would be a line mass such that

$$b(\theta r;\xi,\eta) = \delta(\xi \cos\theta + \eta \sin\theta - r)$$

along the line $(\xi \cos\theta + \eta \sin\theta - r)$. Hereafter R will be taken to be the unit circle.

For this ideal imaging system the line mass projection becomes

$$\rho_L(r,\theta) = \iint_R f(\xi,\eta)\delta(\xi \cos\theta + \eta \sin\theta - r)d\xi d\eta \qquad (4\text{-}2b)$$

and it can be shown that $\rho(r,\theta) = \rho_L(r,\theta) * b(r)$. If the Fourier transform of $\rho(r,\theta)$ is denoted by $\mathcal{F}\{\rho(r,\theta)\}$, then

$$\mathcal{F}\{p(r,\theta)\} = \mathcal{F}\{p_L(r,\theta)\}\mathcal{F}\{b(r)\} \qquad (4\text{-}3)$$

where

$$\mathcal{F}\{p(r,\theta)\} = \int_{-\infty}^{\infty} p(r,\theta)e^{-2\pi j u}dr.$$

From the projection slice theorem [4-1] we can relate the Fourier transform of the projection to the central section of two-dimensional Fourier transform of $f(\xi,\eta)$ as follows:

$$\mathcal{F}\{p_L(r,\theta)\} = \iint_R f(\xi,\eta)e^{-j2\pi u(\xi \cos\theta + \eta \sin\theta)}d\xi d\eta.$$

Denoting $\mathcal{F}\{b(r)\}$ as $B(u)$ and letting $g(u,\theta) = \mathcal{F}\{p(r,\theta)\}$ we have

$$g(u, \theta) = B(u) \iint_R f(\xi, \eta) e^{-j2\pi u(\xi \cos \theta + \eta \sin \theta)} d\xi d\eta . \qquad (4\text{-}4)$$

Finally letting $g_{k,i}$ be $g(u, \theta)$ evaluated at the coordinate tuple (u_k, θ_i), we have the following continuous-discrete characterization of the projection imaging process:

$$g_{k,i} = B(u_k) \iint_R f(\xi, \eta) e^{-j2\pi u_k(\xi \cos \theta_i + \eta \sin \theta_i)} d\xi d\eta \qquad (4\text{-}5)$$

$$i = 1, 2, \ldots, M_\theta$$

$$k = 1, 2, \ldots, M_r$$

where M_θ and M_r are the number of samples taken around θ and along r, respectively. In the notation of the previous section, then, the total number of samples N must equal $M_\theta M_r = N$.

By lexicographically ordering the indices (i, k) to form the image vector $\underline{g}$, the above imaging system has a gramian consisting of the following blocked matrix

$$[\Gamma] = \begin{bmatrix} [\Gamma]^{(1,1)} & [\Gamma]^{(1,2)} & \cdots & [\Gamma]^{(1, M_\theta)} \\ [\Gamma]^{(2,1)} & [\Gamma]^{(2,2)} & \cdots & \\ \vdots & & & \\ [\Gamma]^{(M_\theta, 1)} & \cdots & & [\Gamma]^{(M_\theta, M_\theta)} \end{bmatrix}$$

where each $[\Gamma]^{(i,m)}$ is a symmetric $M_r \times M_r$ matrix whose k, ℓ entry is given by

$$\gamma_{(k,\ell)}^{(i,m)} = B(u_k)B^*(u_\ell) \iint_R e^{-j2\pi u_k(\xi\cos\theta_i+\eta\sin\theta_i)} e^{+j2\pi u_\ell(\xi\cos\theta_m+\eta\sin\theta_m)} \cdot d\xi d\eta .$$

Letting

$$\xi = r\cos\phi$$

$$\eta = r\sin\phi$$

$$dxdy = rdrd\phi$$

$$\gamma_{(k,\ell)}^{(\ell,m)} = B(u_k)B^*(u_\ell)\int_0^1\int_0^{2\pi} e^{-j2\pi u_k r\cos(\phi-\theta_i)} e^{j2\pi u_\ell r\cos(\phi-\theta_m)} rdrd\phi$$

Letting $\psi = \phi-\theta_i$, $d\psi = d\phi$, $-\theta_i \le \psi \le 2\pi - \theta_i$ then

$$\gamma_{(k,\ell)}^{(\ell,m)} = B(u_k)B^*(u_\ell)\int_0^1\int_{\theta_i}^{2\pi-\theta_i} e^{-j2\pi u_k r\cos(\psi)} e^{j2\pi u_\ell r\cos(\psi+\theta_i-\theta_m)} rdrd\psi$$

Temporarily let $\Delta\theta = \theta_i-\theta_m$

$$\gamma_{(k,\ell)}^{(\ell,m)} = B(u_k)B^*(u_\ell)\int_0^1\int_{-\theta_i}^{2\pi-\theta_i} e^{-j2\pi u_k r\cos(\psi)} e^{j2\pi u_\ell r[\cos(\psi)\cos(\Delta\theta) - \sin(\psi)\sin(\Delta\theta)]} d\psi rdr$$

$$= B(u_k)B^*(u_\ell)\int_0^1\int_{-\theta_i}^{2\pi-\theta_i} e^{-j2\pi r[\cos(\psi)(u_\ell\cos(\Delta\theta)-u_k)-\sin(\psi)u_\ell\sin(\Delta\theta)]} d\psi rdr$$

Define:

$$e_c \triangleq u_\ell\cos(\Delta\theta) - u_k \qquad e_s \triangleq u_\ell\sin\Delta\theta$$

$$z = \sqrt{e_c^2+e_s^2} \qquad e_c = z\cos(\alpha),\ e_s = z\sin(\alpha)$$

$$\alpha = \tan^{-1}(e_s/e_c)$$

We have that

$$\gamma^{(\ell, m)}_{(k, \ell)} = B(u_k)B^*(u_\ell)\int_0^1 \int_{-\theta_i}^{2\pi-\theta_i} e^{j2\pi r[z\cos(\psi)\cos(\alpha)-z\sin(\psi)\sin(\alpha)]} d\psi\, r dr$$

$$= B(u_k)B^*(u_\ell)\int_0^1 \int_{-\theta_i}^{2\pi-\theta_i} e^{j2\pi rz\cos(\psi+\alpha)} d\psi\, r dr$$

$$= B(u_k)B^*(u_\ell)2\pi\int_0^1 J_0(2\pi z) r dr$$

$$= B(u_k)B^*(u_\ell)\frac{J_1(2\pi z)}{z}$$

By resubstituting for z we have the desired result

$$\gamma^{(\ell, m)}_{(k, \ell)} = B(u_k)B^*(u_\ell)\pi\left\{\frac{2J_1\left[2\pi(u_k^2 - 2u_k u_\ell \cos(\theta_i-\theta_m) + u_\ell^2)^{\frac{1}{2}}\right]}{2\pi(u_k^2 - 2u_k u_\ell \cos(\theta_i-\theta_m) + u_\ell^2)^{\frac{1}{2}}}\right\} \qquad (4\text{-}6)$$

where $J_1(\cdot)$ is the first order Bessel Function of the first kind. This functional form for the entries in $[\Gamma]^{(i, m)}$ combined with the block structure resulting from the lexicographic ordering of the k, i indices from equation (4-5) provide a structure for $[\Gamma]$ which facilitates its generation, diagonalization, and inversion. If we note that θ_i and θ_m influence $[\Gamma]$ only through the cosine of their difference, by sampling θ uniformly over $[0, 2\pi)$ and u independent of θ, and noting that the cosine property of $\cos(\theta_i-\theta_m)$ is even symmetric about M_θ, $[\Gamma]$ will have the following form which is even symmetric circulant only in the indices i, m:

$$[\Gamma] = \begin{bmatrix} [\Gamma]^{(1,1)} & [\Gamma]^{(1,2)} & [\Gamma]^{(1,3)} \cdots & [\Gamma]^{(1,3)} & [\Gamma]^{(1,2)} \\ [\Gamma]^{(1,2)} & [\Gamma]^{(1,1)} & [\Gamma]^{(1,2)} \cdots & [\Gamma]^{(1,4)} & [\Gamma]^{(1,3)} \\ \vdots & & & & \vdots \\ [\Gamma]^{(1,2)} & [\Gamma]^{(1,3)} & [\Gamma]^{(1,4)} \cdots & [\Gamma]^{(1,2)} & [\Gamma]^{(1,1)} \end{bmatrix} \qquad (4\text{-}7)$$

From this it is seen that each row of blocks is a right circular shift of the preceding row. Thus for simplicity we can let $[\Gamma]^{(i,m)} = [\Gamma]^{(i)}$ hereon. In the special case where each $[\Gamma]^{(i)}$ is itself a circulant, $[\Gamma]$ becomes a block circulant matrix. Hunt [4-2] has shown that block circulant matrices are diagonalized by an orthogonal matrix that is the kronecker product of two matrices each of which represents a discrete Fourier transform. However, for our projection imaging case $[\Gamma]^{(i)}$ is only symmetric and not circulant so a more general result than Hunt's is required if we are to effect a computational reduction in diagonalizing $[\Gamma]$.

In generalizing Hunt's result to include the pseudo-circulant[2] type matrix $[\Gamma]$ we are confronted with, we will consider a matrix $[\Gamma]$, whose elements are taken from a complex vector space V. This will allow a far more general result applicable to circulant

[2] The term pseudo-circulant is introduced here to emphasize that $[\Gamma]$ is neither circulant nor block circulant, but simply has two of its four indices operating as in circulants.

matrices of rather arbitrary elements. By a complex vector space is meant a vector space whose source of scalars is the field of complex numbers, $\mathcal{C}$. Then $[\Gamma]$ is given by

$$[\Gamma] = \begin{bmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(M_\theta-1) \\ \gamma(M_\theta-1) & \gamma(0) & \cdots & \gamma(M_\theta-2) \\ \vdots & & & \\ \gamma(1) & & & \gamma(0) \end{bmatrix}$$

where $\gamma(j)$ are elements of V and may be scalars, vectors, matrices, or tensors.

Letting $\omega(k)$ be one of M_θ roots of unity the closure of the complex vector space V implies that

$$\omega(k)\gamma(i)\in V \quad \forall\, \gamma(i)\in V, \omega(k)\in\mathcal{C}.$$

Following Hunt, we have in vector form

$$\underline{\omega}(k) = \begin{bmatrix} e^{j2\pi k/M_\theta} \\ \vdots \\ e^{j2\pi ki/M_\theta} \\ \vdots \end{bmatrix}$$

and

$$[\omega_{M_\theta}] = \frac{1}{\sqrt{M_\theta}} \left[\underline{\omega}(0) \,\middle|\, \underline{\omega}(1) \,\middle|\, \cdots \,\middle|\, \underline{\omega}(M_\theta-1)\right]. \tag{4-8}$$

From the properties of the complex exponentials the matrix $[\omega_{M_\theta}]$ is orthogonal and

$$[\Lambda_V] = [\omega_{M_\theta}][\Gamma][\omega_{M_\theta}].$$

The matrix $[\Lambda_V]$ is a diagonal matrix whose diagonal elements are elements of the complex vector space V which may in fact be matrices. The case of interest here is when the vectors $\gamma(i)$ are taken to be matrices comprised of complex numbers. The set of all such matrices clearly forms a complex vector space, and the above results apply. Using Kronecker products [4-3] it follows that the matrix $[\Gamma]$ can be reduced to block diagonal form by the matrix $\frac{1}{\sqrt{M_\theta}}[\omega_{M_\theta}] \otimes [I]$ and

$$\begin{bmatrix} [\Lambda]^{(1)} & & & \\ & [\Lambda]^{(2)} & & \\ & & \ddots & \\ & & & [\Lambda]^{(M_\theta)} \end{bmatrix} = \frac{1}{M_\theta}[[\omega_{M_\theta}]^* \otimes [I]][\Gamma][[\omega_{M_\theta}] \otimes [I]] \tag{4-9}$$

where $[I]$ and $[\Lambda]^{(i)}$ are $M_r \times M_r$ matrices, $[I]$ being a diagonal, $[\Lambda]^{(i)}$ being full and $[\omega_{M_\theta}]$ being defined by equation (4-8). The block diagonal matrix on the left side of (4-9) contains diagonal matrices that are symmetric if $[\Gamma]^{(i)}$ is symmetric $\forall i$ and thus diagonalizable. In addition the orthogonal matrix reducing the left side of (4-9) to diagonal form is itself block diagonal, the blocks being the orthogonal matrices diagonalizing $[\Lambda]^{(i)}$. We thus have the result that the complete diagonalization of $[\Gamma]$ becomes

$$[\Lambda] = \begin{bmatrix} [U]^{(1)*} & & \bigcirc \\ & [U]^{(2)*} & \\ \bigcirc & & \ddots \\ & & & [U]^{(M_\theta)*} \end{bmatrix} \begin{bmatrix} [\Lambda]^{(1)} & & \bigcirc \\ & [\Lambda]^{(2)} & \\ \bigcirc & & \ddots \\ & & & [\Lambda]^{(M_\theta)} \end{bmatrix} \begin{bmatrix} [U]^{(1)} & & \bigcirc \\ & [U]^{(2)} & \\ \bigcirc & & \ddots \\ & & & [U]^{(M_\theta)} \end{bmatrix} \quad (4\text{-}10)$$

where $[U]^{(i)*}[\Lambda]^{(i)}[U]^{(i)} = [D]^{(i)}$ for all i, $[D]^{(i)}$ being a diagonal matrix of complex scalar elements.

If $\forall i$, $[\Gamma]^{(i)}$, is a circulant matrix, Hunt's result for block circulant matrices can be obtained as a special case of (4-10) by noting that the set of all circulant matrices whose elements are complex numbers comprises a complex vector space.

The results of equation (4-10) can be summarized as requiring a general discrete Fourier transform to reduce the block diagonal form followed by individual subblock diagonalizations provided by a singular value decomposition (SVD) routine developed by Golub [4-4]. By not performing the multiplications involving zero, a significant computational reduction is achieved in diagonalizing the transaxial projection imaging gramian. Furthermore, the nature of $\cos(\theta_i - \theta_m)$ over $[0, 2\pi)$ implies a further computational reduction in that only

the first $M_\theta/2+1$ blocks of $[\Gamma]$ need be calculated for M_θ an even number since the first row becomes

$$[\Gamma]^{(1)}[\Gamma]^{(2)}\ldots[\Gamma]^{(M_\theta/2)}[\Gamma]^{(M_\theta/2+1)}[\Gamma]^{(M_\theta/2)}\ldots[\Gamma]^{(2)} .$$

Thus once $[\Gamma]^{(i)}$ $i = 1, 2, \ldots, M_\theta/2+1$, is determined, $[\Gamma]$ is known in its entirety. This also applies to the reduction from block diagonal form to diagonal form since the discrete Fourier transform is conjugate symmetric about the folding frequency, and only $[\Lambda]^{(i)}$ $i = 1, \ldots, M_\theta/2+1$ need actually be diagonalized by the SVD.

To illustrate the significance of the above simplifications consider the case where 64 samples are taken for M_r and 512 samples for M_θ. $[\Gamma]$ is then a 32,768 dimensioned matrix containing over 10^9 elements. Without the above computational savings it would be unfeasible to even calculate $[\Gamma]$, let alone diagonalize it. However, by calculating only the first $M_\theta/2+1$ blocks of $[\Gamma]$, performing the Fourier transform indicated in (4.9) and employing the SVD algorithm of Golub [4-4], this particular $[\Gamma]$ was calculated, diagonalized and inverted in under 2 hour cpu time on a PDP-10 computer.

The gramian developed in the above transaxial tomographic form has been based on the even sampling of u (or r) and θ resulting in $M_r \times M_\theta$ data values. While this even sampling of u and θ has resulted in a significant computational reduction, it does present some difficulty. In discussing this and for the remainder of the chapter we

consider the projection imaging system to be ideal; namely, $\rho(r, \theta) = \rho_L(r, \theta)$ and $B(u) = 1$ in equation (4-4). Looking at (4-5) it can be seen that if $u_k = 0$, for some k, then the k-th row of $[\Gamma]^{(i)}$ will be the same for each $i = 1, \ldots, M_\theta$. Therefore the rank of $[\Gamma]$ must be no greater than $M_\theta(M_r - 1) + 1$ and consequently the gramian is singular. Physically the $u_k = 0$ term represents the d.c. or average value of $f(\xi, \eta)$ over R, which of course is rotation invariant, and thus no new information is gained concerning the d.c. value after the first projection is obtained. Naturally this will result if a discrete Fourier transform is applied to the projection data to obtain $\underline{g}$ and the u = 0 term is retained in each projection. Because of this inherent singularity in $[\Gamma]$, the estimate of $\hat{f}(\xi, \eta)$ must necessarily result from either a pseudo-inverse or a constrained least squares solution as developed in equation (3-14). For a minimum norm estimate of $\hat{f}(\xi, \eta)$ we would set the constraint matrix such that $[C] = [I]$. Utilizing a discrete Fourier transform on the projection data to obtain $\underline{g}$ we have

$$\underline{g} = \begin{bmatrix} \underline{g}^{(1)} \\ \underline{g}^{(2)} \\ \vdots \\ \underline{g}^{(M_\theta)} \end{bmatrix}$$

where $\underline{g}^{(i)}$ is an $M_r \times 1$ column vector whose k-th element, $g_k^{(i)}$, is the k-th harmonic of the discrete Fourier transform of the i-th projection.

Utilizing this result, the following matrix representation of the reconstruction algorithm is obtained.

$$\hat{f}(\xi,\eta) = \frac{1}{M_\theta}\, \underline{h}^*(\xi,\eta)[[\omega_{M_\theta}] \otimes [I]] \begin{bmatrix} [U]^{(1)} & & \\ & [U]^{(2)} & \\ & & \ddots \\ & & & [U]^{(M_\theta)} \end{bmatrix}$$

$$[\gamma^{-1}[I]+[\Lambda]^2]^{-1}[\Lambda] \begin{bmatrix} [U]^{(1)*} & & \\ & [U]^{(2)*} & \\ & & \ddots \\ & & & [U]^{(M_\theta)*} \end{bmatrix} [[\omega_{M_\theta}] \otimes [I]] \begin{bmatrix} \underline{g}^{(1)} \\ \underline{g}^{(2)} \\ \vdots \\ \underline{g}^{(M_\theta)} \end{bmatrix} \tag{4-11}$$

It is the above equation which is utilized as the reconstruction algorithm for the images in subsequent sections of this chapter, and a few comments are in order.

Clearly the matrix $[\Gamma]$ must be diagonalized and a constrained least squares inverse precalculated only when the geometry is changed. Recalling that

$$\hat{f}(\xi,\eta) = \underline{h}^*(\xi,\eta)\underline{\alpha}$$

and that the $[[\omega_\theta] \otimes [I]]$ can be implemented using fast Fourier techniques, the bulk of the computations in calculating $\underline{\alpha}$ lies in the multiplication by the constrained least squares matrix. Since this matrix is block diagonal, by not performing the calculations involving zeros, it should be evident that this step involves the same

number of operations as the convolutional portion of a convolution algorithm for the same projection data format. However, the convolution algorithm involves only a single sum in the back projection step [4-1] while this algorithm involves a sum over both the u_k and θ_i indices where $k = 1, \ldots, M_r$ and $i = 1, \ldots, M_\theta$. Thus the convolution algorithm should be more efficient computationally. One advantage of the continuous-discrete formulation described herein is that it provides as an output, a function which can be evaluated at the points of interest without an interpolation step in the algorithm. This eliminates one source of error while providing flexibility in the output format. Reconstructing a 64 x 64 output array, approximately 8 min of CPU time was required for $M_r = 32$ and $M_\theta = 128$. This is somewhat slower than the corresponding 4 min for the convolutional algorithm, and for larger arrays the computational advantage should favor the convolution algorithm.

4.2 Experimentally Determined Degrees of Freedom

To further emphasize the usefulness of the gramian in estimating the degrees of freedom of an imaging system, some computational examples are developed below. The resulting $[\Gamma]$ was determined and diagonalized by the procedure described in the preceding section for nine different combinations of M_r and M_θ. Three values of M_r equaling 16, 32, and 64 were selected and for each of these, three values of M_θ were selected as $2M_r$, $4M_r$, and $8M_r$. Thus the

resulting gramians ranged in dimension from 512 x 512 up to 32,768 x 32,768. Table (4.1) lists the appropriate combinations.[3]

The gramian associated with the nine combinations of M_r and M_θ listed in Table (4.1), were formed and the procedure for reducing the full $[\Gamma]$ to block diagonal form using Fourier methods ($[\omega_{M\theta}] \otimes [I]$) was employed. The block diagonal form was further reduced utilizing the SVD routines of Golub [4-4], yielding a completely diagonal $M_r M_\theta \times M_r M_\theta$ matrix with eigenvalues on the diagonal. The eigenvalues so computed are plotted in figures 4.2, 4.3, 4.4 for M_r = 16, 32, and 64 respectively. In each figure results are shown for M_θ equaling $2M_r$, $4M_r$, $8M_r$. Since nonsingular permutation matrices can be included in the eigenvector matrices of $[\Gamma]$, without loss of generality the eigenvalues are plotted in decreasing order. Note that due to the large dynamic ranges involved, a logarithmic scale is used and that λ_{MAX} is always unity. This is taken to be the case since $[\Gamma]$ can always be normalized by dividing by λ_{MAX}. Figures 4.2, 4.3 and 4.4 display a characteristic behavior in the

[3] A parenthetical comment is in order here for those readers familiar with tomographic scanners. In this chapter we have gathered M_r sample points along each radius from (0, 1) and have gathered M_θ projections at angles from $(0, 2\pi)$. In practice projection data is often taken radially from (-1, 1) and at angles from $(0, \pi)$ (see fig.4.1). Clearly the same region is covered, the only difference being a permutation of indices.

spectrum of $[\Gamma]$. In each plot, after an initial decline, λ_n remains relatively constant with increasing n until a critical point, called N_{crit}, is reached; whereupon an abrupt drop of 4 orders of magnitude in λ_n occurs. Subsequently λ_n decreases more slowly with increasing n. However, the eigenvalues were calculated on a DEC PDP-10 digital computer in single precision arithmetic and since the SVD algorithm calculates eigenvalues in decreasing order, little confidence should be attributed to values of λ less than 10^{-6}. Consequently we can attribute the tail of the eigenvalues spectrum to be due to computer roundoff and computational noise. This coupled with the abrupt drop in λ at N_{crit}, will allow us to roughly estimate the degrees of freedom in the projection imaging system to be N_{crit}. Note that this abrupt drop allows us to set a threshold of approximately $\lambda = 10^{-4}$ and achieve a consistent mid point of this step drop. Table 4.1 includes the number of eigenvalues above 10^{-4}.

This behavior of relatively constant λ out to a point followed by an abrupt drop has been shown to be characteristic of ideal circularly symmetric systems that are bandlimited and observed over a circle of finite radius [4-5]. In considering the projection slice theorem, we note that the Fourier transform of the output projection is the central section of the Fourier transform of the original cross section. Therefore even ideal projection imaging (i.e. no blur) is "bandlimited" by inclusion of only M_r harmonics, and thus this eigenvalue

behavior should come as no surprise. More will be said about this bandlimiting process later.

In figures 4.2, 4.3, and 4.4 it is evident that increasing M_θ from $2M_r$ to $4M_r$ results in a significant increase, but not a doubling, in N_{crit}. This behavior is born out in table 4.1. Also in table 4.1 the value $\pi^2(M_r-1)^2/4$ is included with other parameters of interest. Note that the number $\pi^2(M_r-1)^2/4$ is in relatively good agreement with the value at which N_{crit} does not significantly increase with increasing M_θ, and that N_{crit} is always less than the maximum possible rank of $[\Gamma]$. This agreement is related to the bandlimiting mentioned before and can be explained by noting that the continuous-continuous ideal projection imaging system described in section 4.1 when band-limited with a one sided bandwidth B and constrained to a circular input region of radius R, is a circularly symmetric ideal imaging system that is bandlimited and observed over a finite area. Gori and Guattari [4-6] have shown that the degrees of freedom of such a continuous-continuous system are in reasonable agreement with the Shannon number, which for a circularly symmetric system of radius R and bandwidth B has been shown to be $\pi^2R^2B^2$. In our case if we consider the pupil to have Fourier harmonics no higher than $(M_r-1)/2$ and take R equal to unity we have the following estimate of the degrees of freedom or N_{DOF}:

$$N_{DOF} \doteq \frac{\pi^2 (M_r - 1)^2}{4} \quad .$$

Note that this estimate of the maximum number of effectively nonzero eigenvalues of $[\Gamma]$ is a function of M_r only and not M_θ. Intuitively this makes sense since increasing samples in θ more densely fills the circle of nonzero frequency components of radius $(M_r - 1)/2$ in frequency space and does not increase its possible area. This is exactly the situation in the continuous case for the circularly symmetric system where the accessible Fourier area is a function of the radial bandwidth B. More rigorously, the results of Section 3 have shown that for sufficiently large M_θ, the rectangular integration rule implicit in even sampling provides a Gramian whose spectrum approaches that of the continuous-continuous model of the ideal projection imaging system whose aforementioned degrees of freedom are independent of θ.

If M_r is specified, considering that $N_{crit} \leq M_r \cdot M_\theta$ we should certainly not expect an increase in reconstruction performance if M_θ is taken much greater than N_{DOF}/M_r. This results in an estimate for the number of projections once M_r is specified as follows:

$$M_\theta \leq \frac{\pi^2}{4} \frac{(M_r - 1)^2}{M_r}$$

which for M_r sufficiently large reduces to

$$M_\theta \leq \frac{\pi^2}{4} M_r$$

4.3 Experimentally Determined Projections

The previous section was devoted to investigating the degrees of freedom structure provided by the eigenvalue spectrum of the gramian of the imaging system. In this section we turn our attention to the actual reconstruction algorithm to estimate the object $\hat{f}(\xi, \eta)$.

Shown in figures 4.5, 4.6, and 4.7 are perspective plots of a reconstructed phantom for M_r equaling 16, 32, and 64. As before multiple values of M_θ are taken for each M_r. The phantom consists of two circular regions of density of 0.1 and 0.05 superimposed on a circular background of density 0.1. These results we obtained with the algorithm of equation (4.11) where γ is such that the dynamic range of the elements of the diagonal matrix $[\gamma^{-1}[I] + [\Lambda]^2]$ is 10^4. As expected from the eigenvalue plot for $M_r = 16$ an increase from $M_\theta = 2M_r$ to $M_\theta = 4M_r$ produces an improvement in the reconstruction while an increase from $M_\theta = 4M_r$ to $M_\theta = 8M_r$ has a far lesser effect as shown in Fig. 4.5. However, for the cases of M_r equaling 32 and 64 the respective increases from $M_\theta = 2M_r$ to $M_\theta = 4M_r$ produce little noticeable improvement. This behavior is not predicted by eigenvalue plots for M_r equaling 32, and 64, nor should it be. These plots in reality predict an upper limit to the improvement in the reconstruction as a function of an increased number of

projections given the number of sample points per projection. For the cases of M_r equaling 32 and 64 the full eigenvalue spectrum of the matrix $[\Gamma]$ is probably not necessary to achieve a reasonable reconstruction of this relatively simple image. This fact is somewhat emphasized by the results of Figure 4.8 where for the case of M_r equaling 64 a very reasonable reconstruction is obtained for M_θ equaling 64. In analyzing this it should be considered that in these experiments M_r was specified a priori at 3 different levels and for this simple phantom M_r equaling 32 and 64 probably indicates radial oversampling. This would seem to indicate that the radial sampling rate should be set to reflect the highest spatial frequency believed present in the data. Thus if M_r is determined in this manner a more substantial improvement in the reconstruction should be evident in increasing M_θ from $2M_r$ to $4M_r$, but in no case will a comparable improvement be exhibited by increasing M_θ to $4M_r$ to $8M_r$.

Figure 4.9 shows the reconstruction of a monkey's head which was irradiated by a G_d^{153} radioisotope source. Hence $M_r = 32$ and projections were taken at 1 degree increments from 0 to 360 degrees. Subsets of these projections were taken to obtain M_θ equaling $2M_r$, $4M_r$, and $8M_r$. In this case increasing M_θ from $2M_r$ to $4M_r$ produces an improvement in the reconstruction while an increase from $4M_r$ to

$8M_r$ produces an improvement of a comparably lesser degree. Obtaining the projections by the method described probably introduced some angular error and the improvement of Fig. 4.9c over Fig.4.9b is due more to these errors being averaged than any significant increase in information content.

4.4 Conclusions

This chapter has attempted to present a generalized degree of freedom analysis for imaging systems by utilizing the gramian of the point spread function kernel. The gramian's eigenvalue spectrum provides an indication of the DOF, and various sensor signal to noise ratios and computational noise considerations become useful parameters in the eigenvalue space. The general gramian approach to imaging was then directed to the specific imaging geometry of transaxial tomographic projections. The associated gramian was shown to have considerable structure allowing computational savings in calculation of the eigenvalue spectrum. These computational savings utilized a fast Fourier routine to reduce the gramian to a block diagonal form. This block diagonal form was then further reduced by performing a series of lower dimensional singular value decompositions resulting in diagonalization procedures to handle extremely large sized gramians. This linear algebraic approach also provided a useful reconstruction algorithm for obtaining

estimates of the original object from the projection data.

Experimental verification of the analytic work was then developed by computational procedures on nine different sized gramians for different projection imaging sampling geometries. The computational eigenvalue spectra so obtained agreed quite well with the theoretical predictions and also demonstrated some additional structure in the imaging system inherent in circular imaging. The degrees of freedom were quite easily obtained from the eigenvalues. A critical number of eigenvalues was consistently observed and in retrospect this number agreed quite well with the inherent bandlimit of such systems when analyzed from a continuous-continuous imaging model.

Finally, the reconstruction algorithm, which inherently makes use of the gramian data, was exercised for obtaining estimates of the original objects from their projection data. Again experimental consistency provided added confirmation of the results from the gramian analysis. Essentially the visual quality of the pictorial reconstructions agreed quite well with the predicted behavior based upon the degree of freedom analysis.

M_r	M_θ	Gramian Dimension $N=M_r M_\theta$	Maximum Rank $M_\theta(M_r-1)+1$	N_{crit}	N_{DOF} $\pi^2(M_r-1)^2/4$	# Eigenvalues > 10^{-4}
	32	512	481	381		401
16	64	1024	961	549	555	612
	128	2048	1921	578		688
	64	2048	1984	1543		1588
32	128	4096	3969	2287	2371	2390
	256	8192	7937	2242		2636
	128	8192	8075	6245		6320
64	256	16384	16124	9237	9795	9434
	512	32768	32257	9928		10312

Table 4.1. Rank and Degrees of Freedom Results for Various Projection Imaging Gramians.

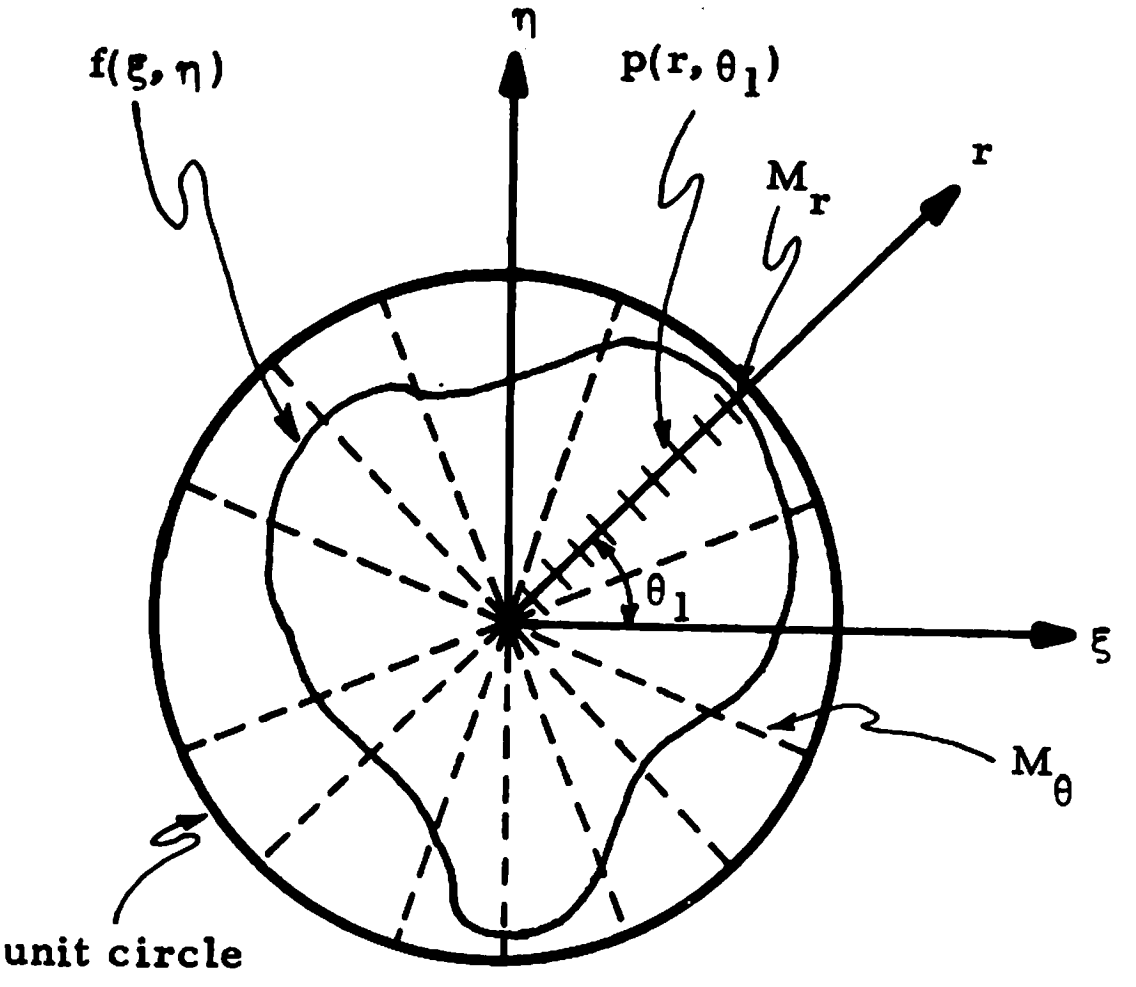


Figure 4.1. Projection imaging geometry.

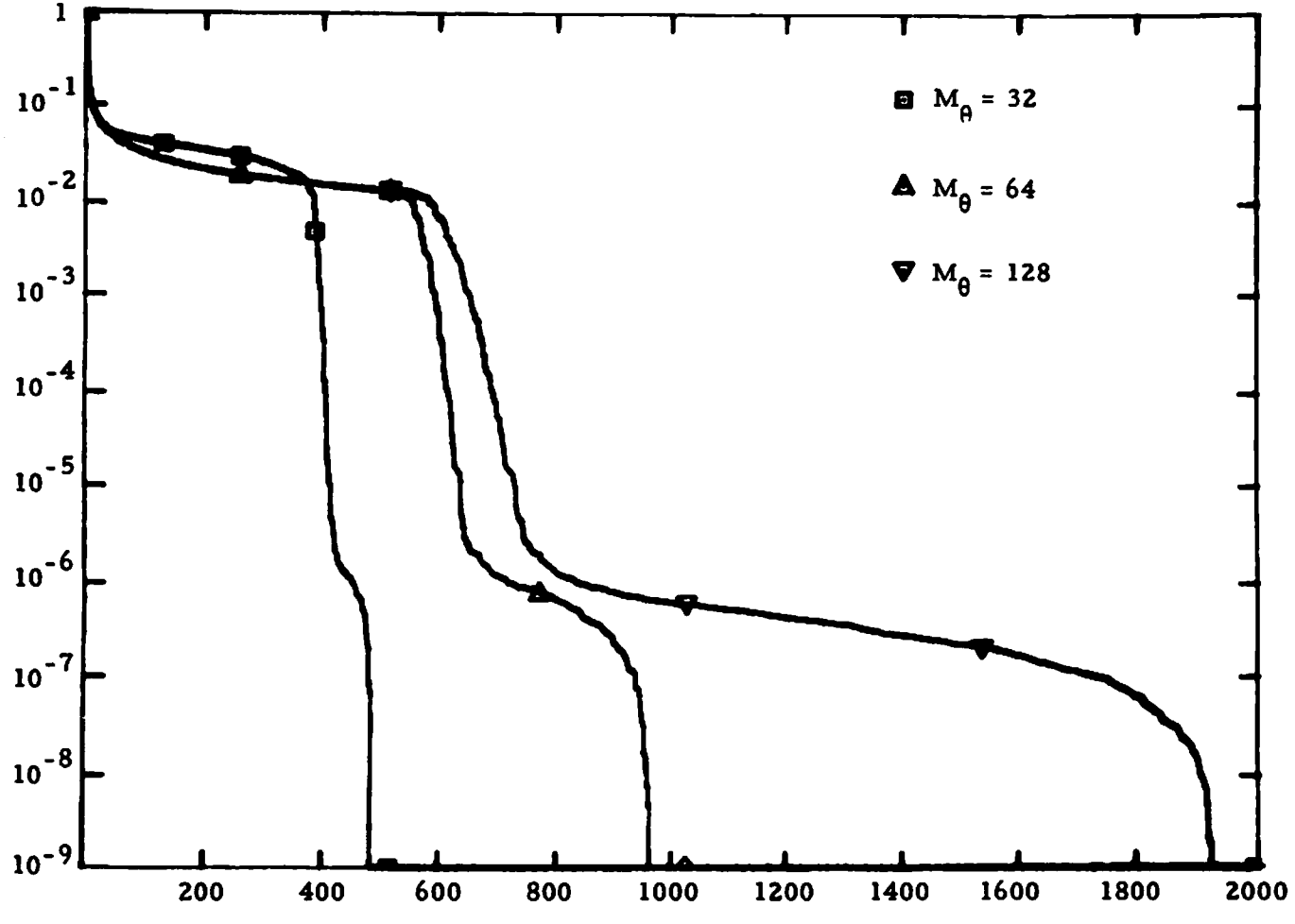


Figure 4.2. Gramian Eigenvalues for $M_r = 16$.

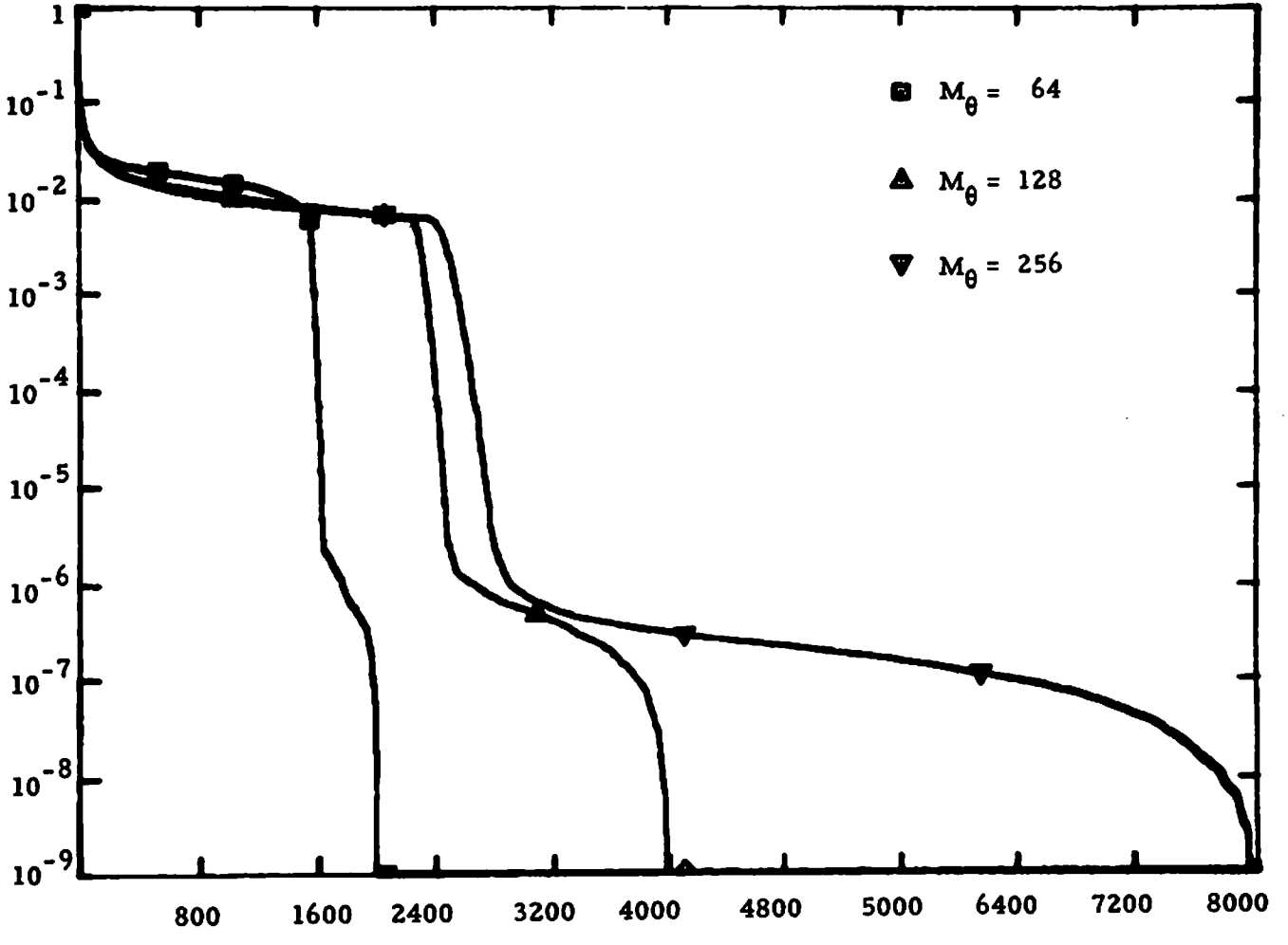


Figure 4.3. Gramian Eigenvalues for $M_r = 32$.

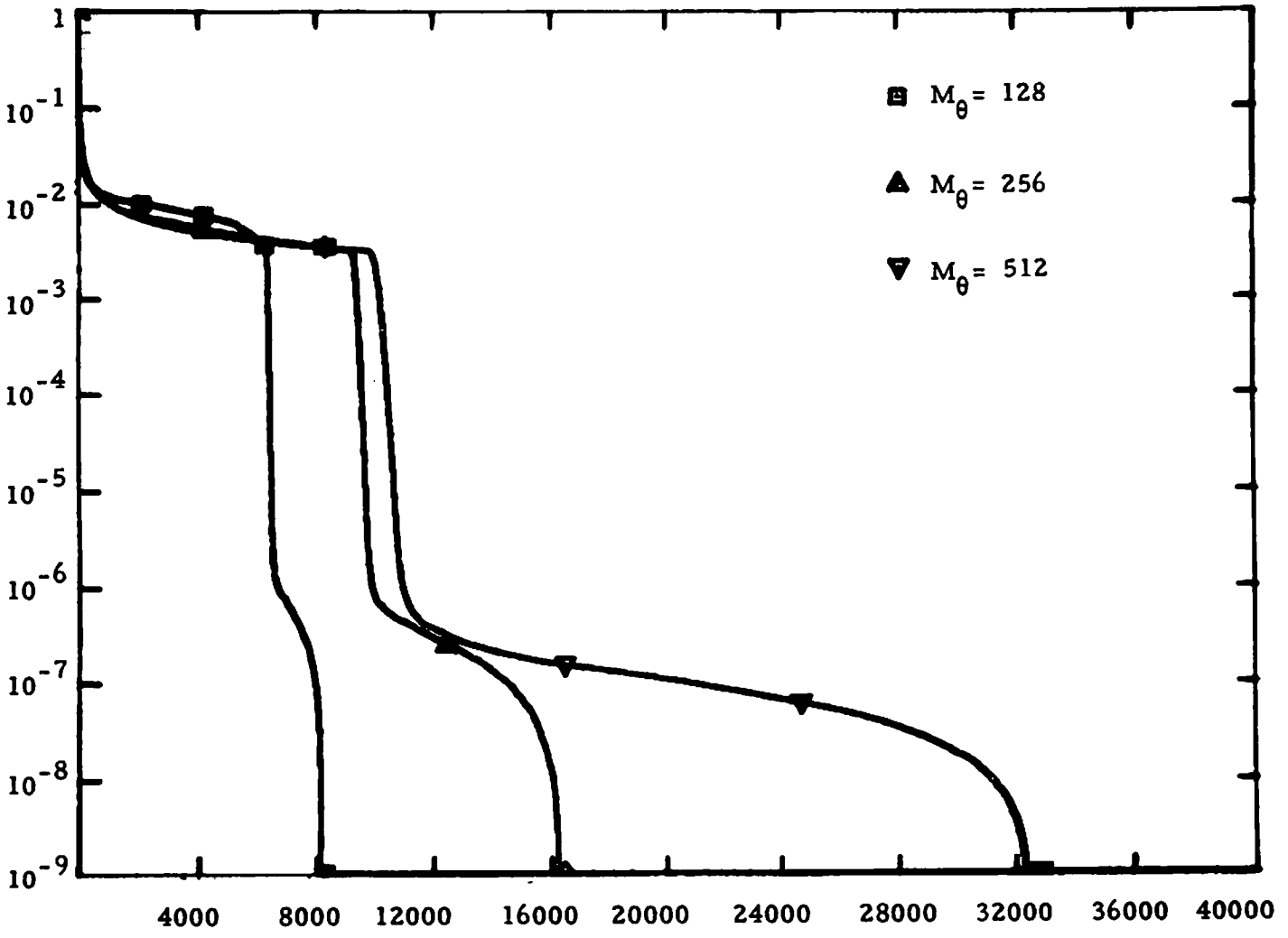


Figure 4.4. Gramian Eigenvalues for $M_r = 64$.

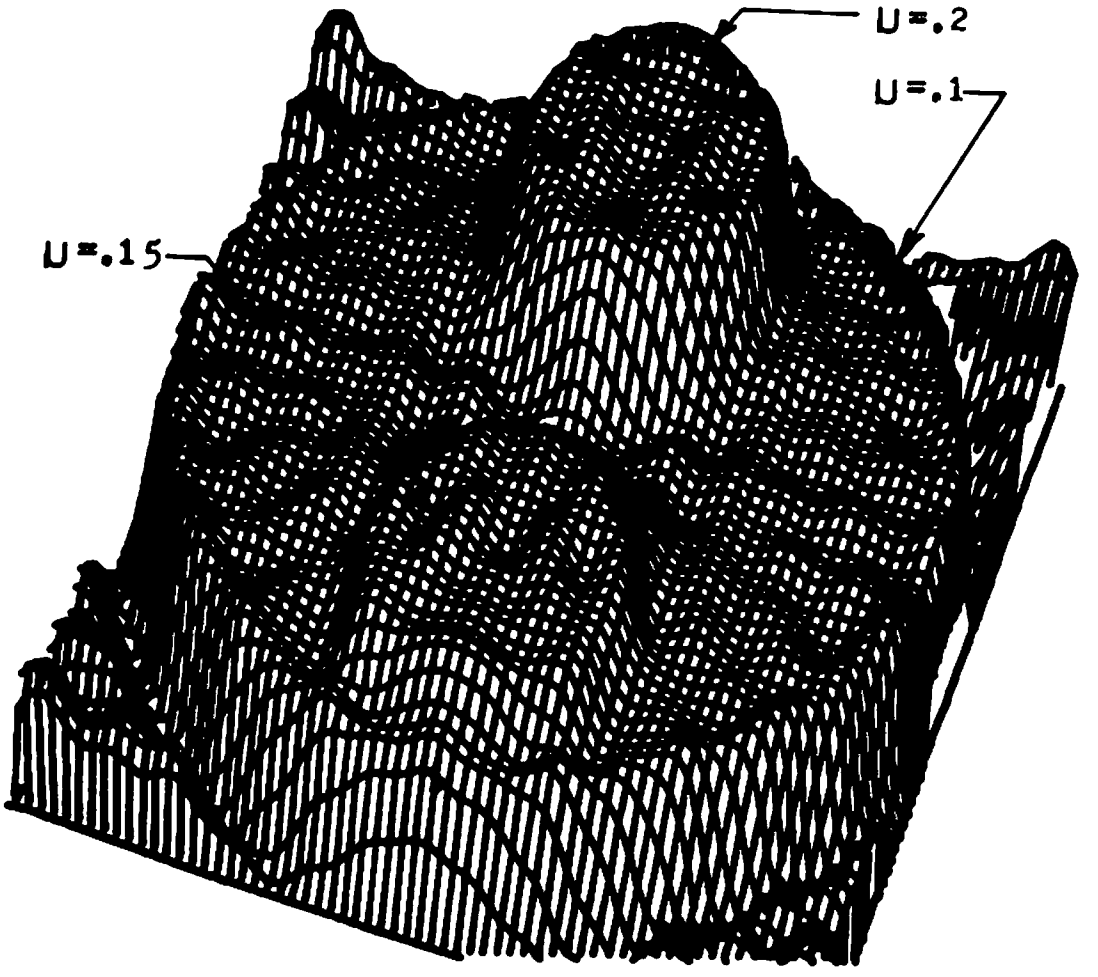


Figure 4.5(a). Phantom Reconstruction with $M_r = 16$, $M_\theta = 32$.

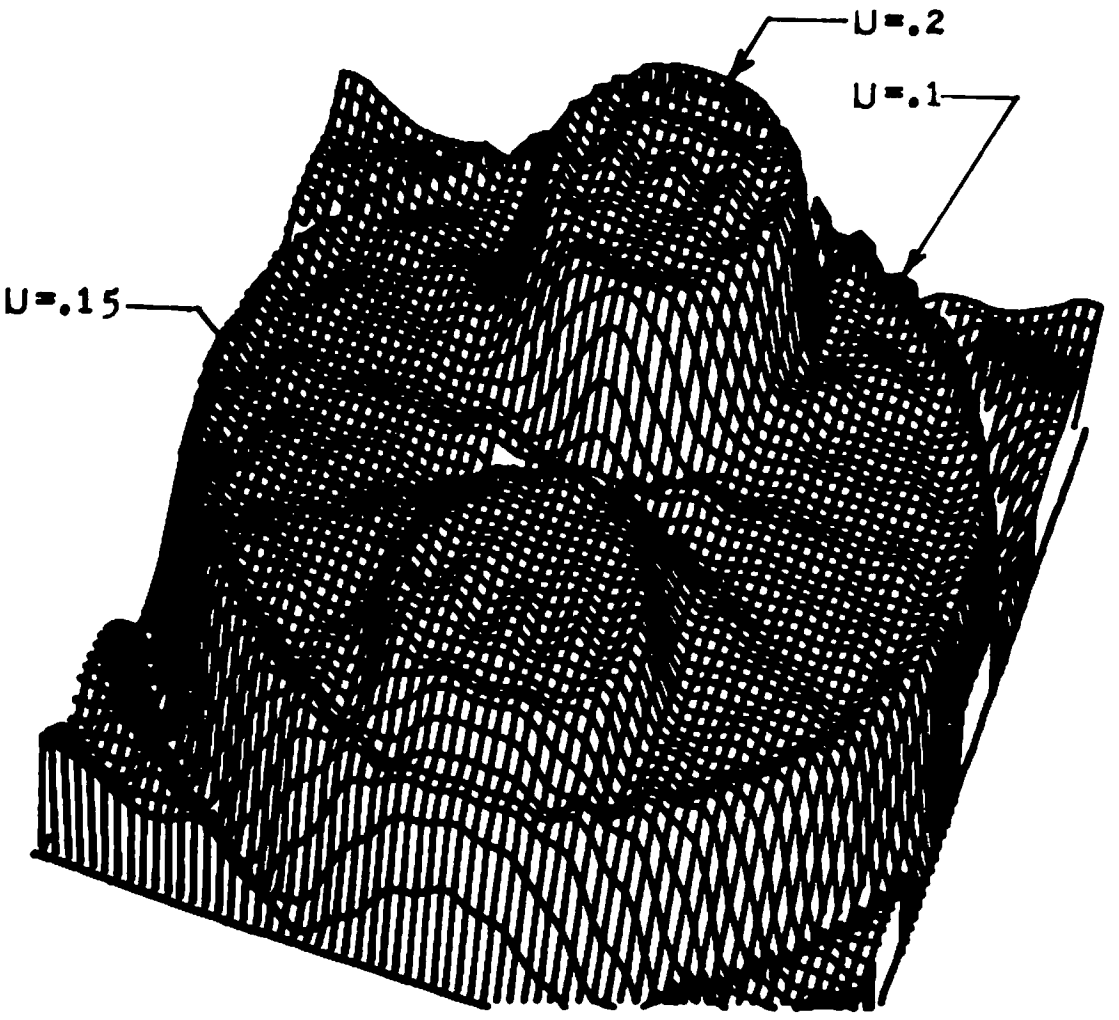


Figure 4.5(b). Phantom Reconstruction with $M_r = 16$, $M_\theta = 64$.

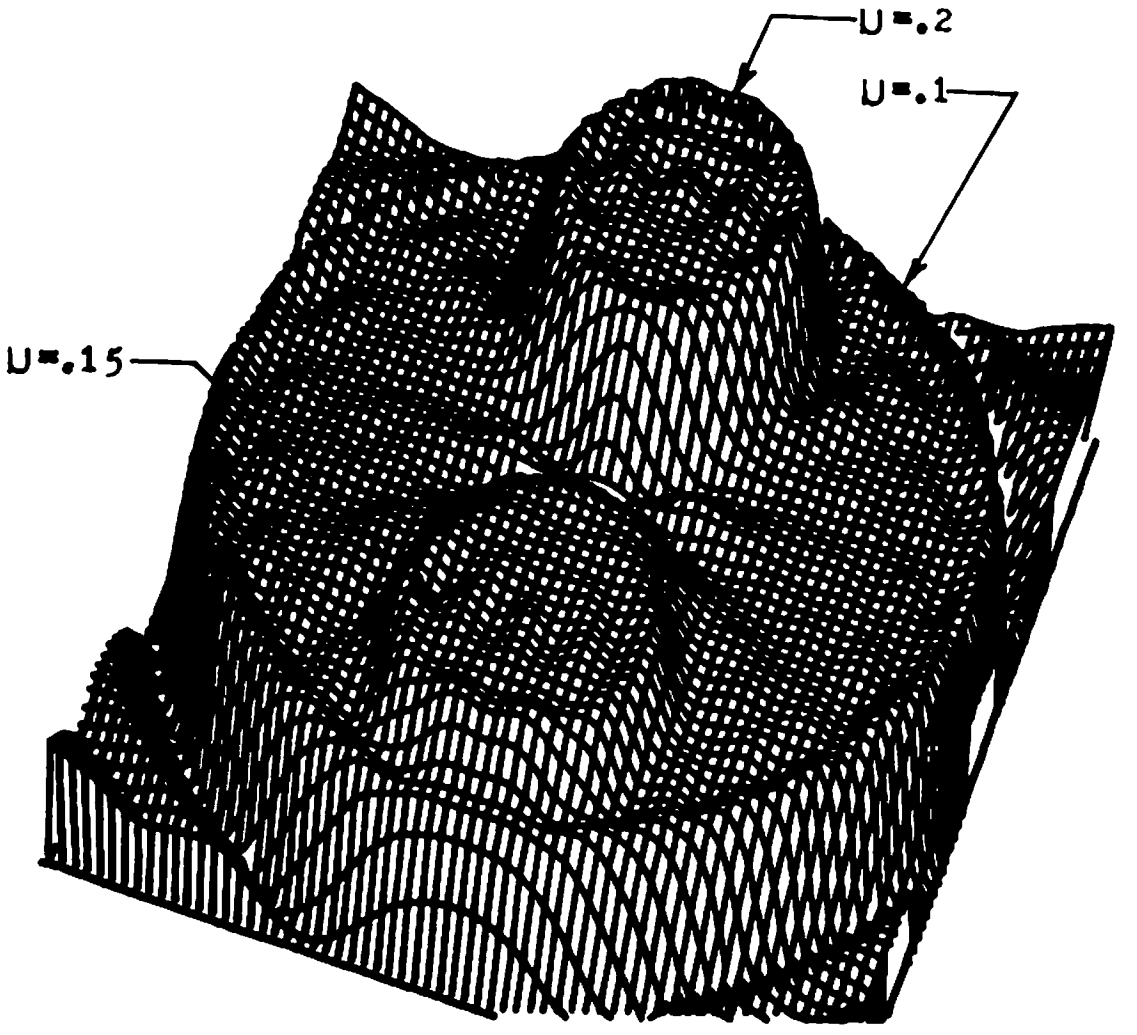


Figure 4.5(c). Phantom Reconstruction with $M_r = 16$, $M_\theta = 128$.

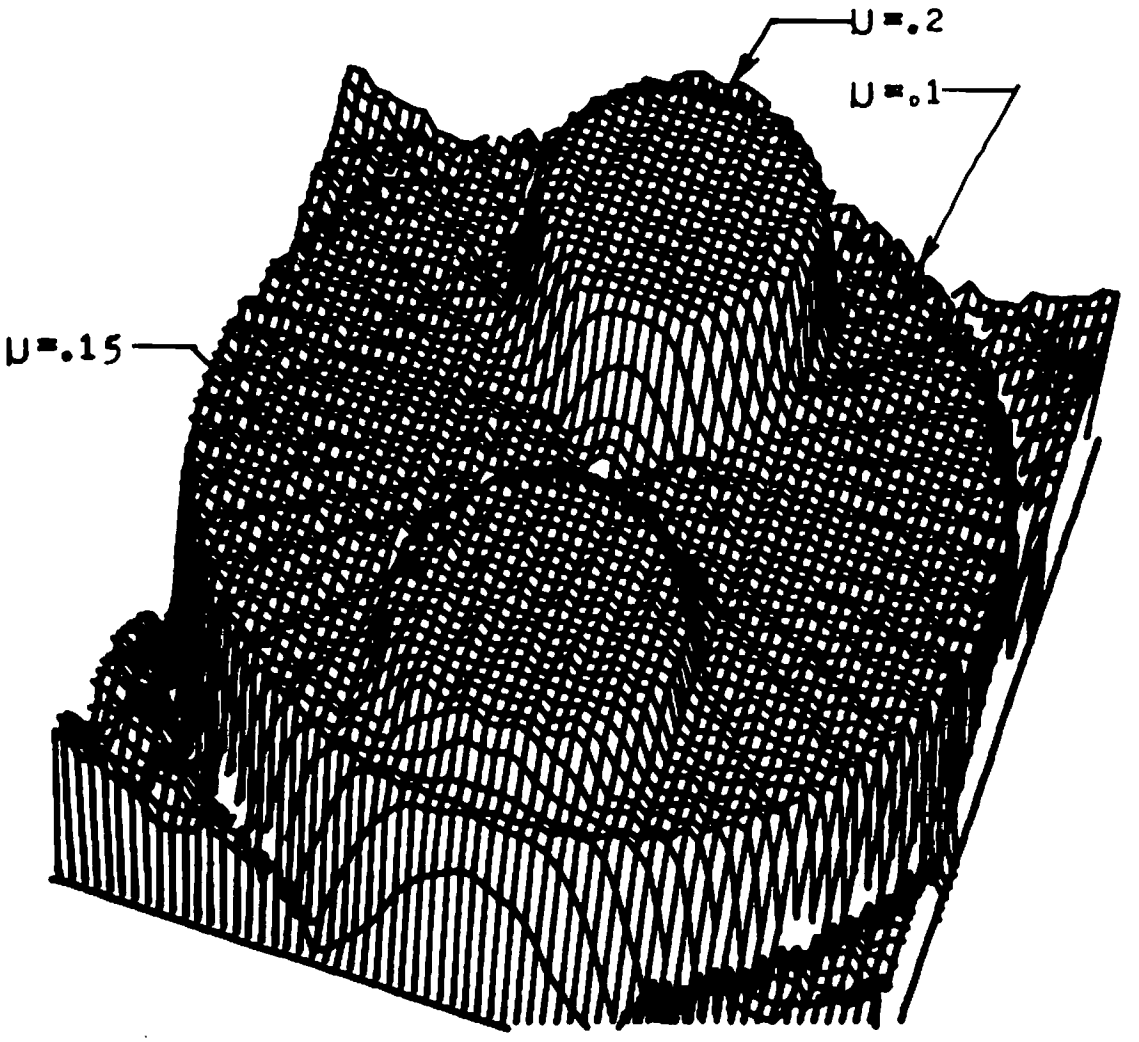


Figure 4.6(a). Phantom Reconstruction with $M_r = 32$, $M_\theta = 64$.

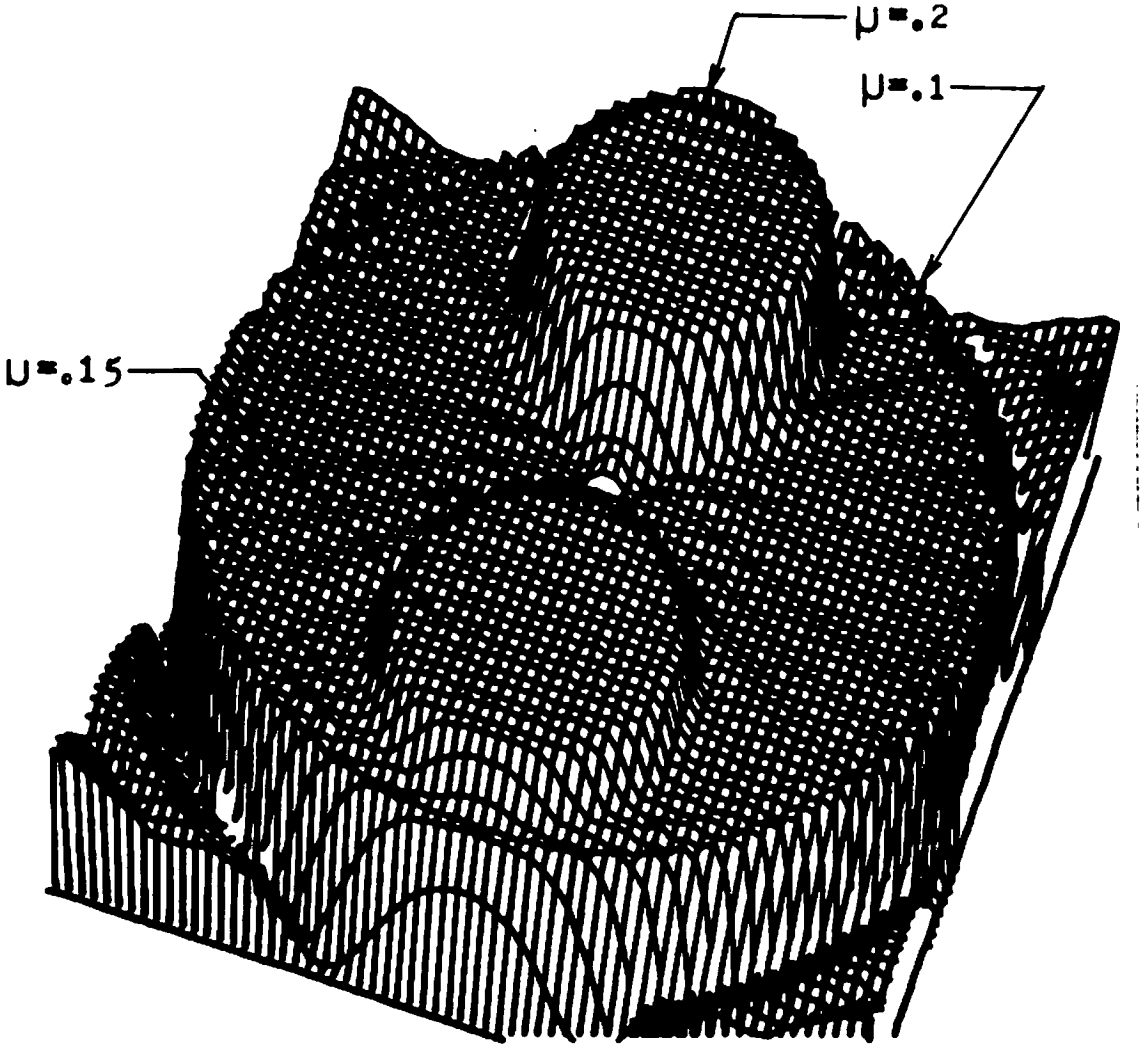


Figure 4.6(b). Phantom Reconstruction with $M_r = 32$, $M_\theta = 128$.

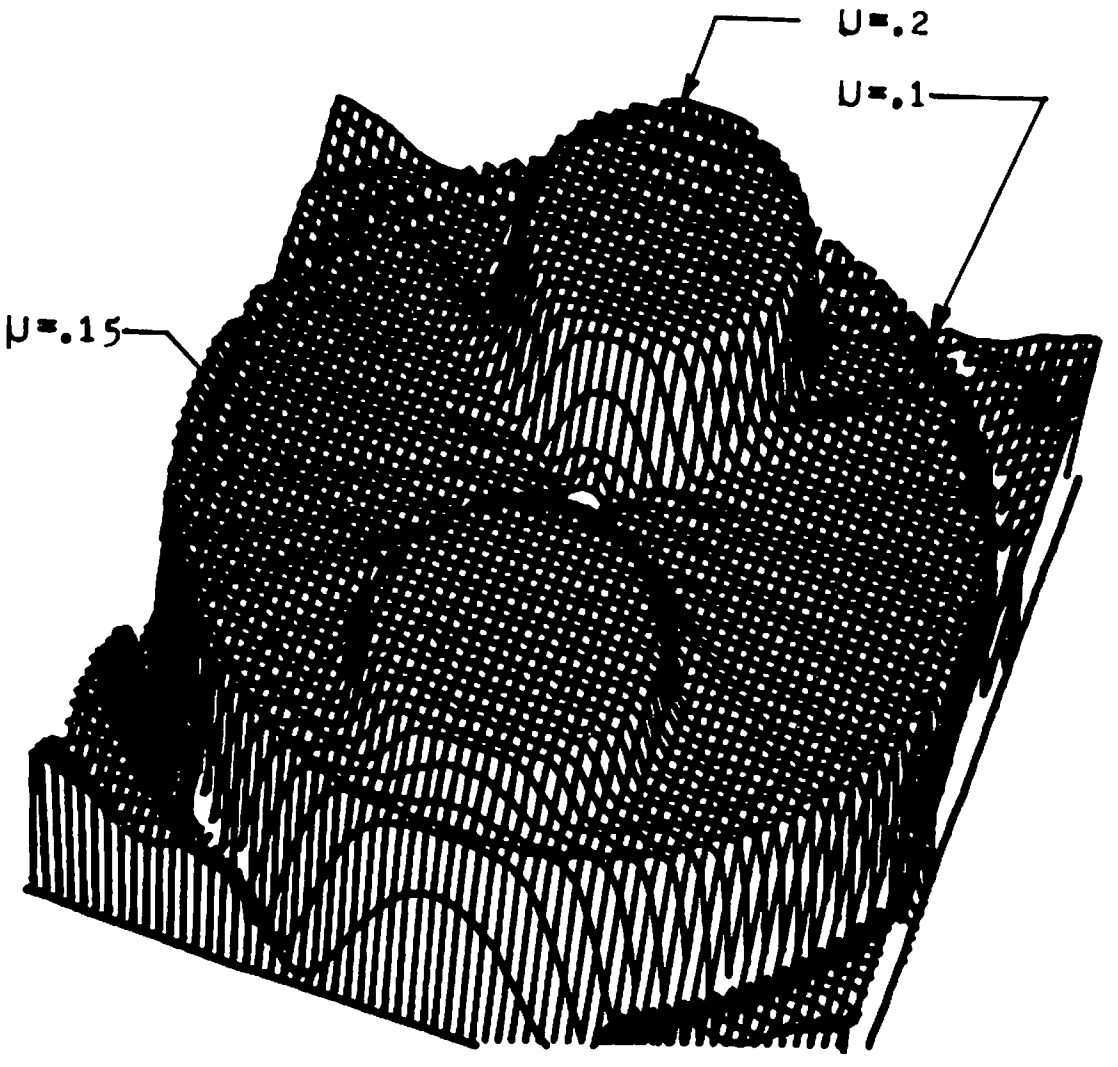


Figure 4.6(c). Phantom Reconstruction with $M_r = 32$, $M_\theta = 256$.

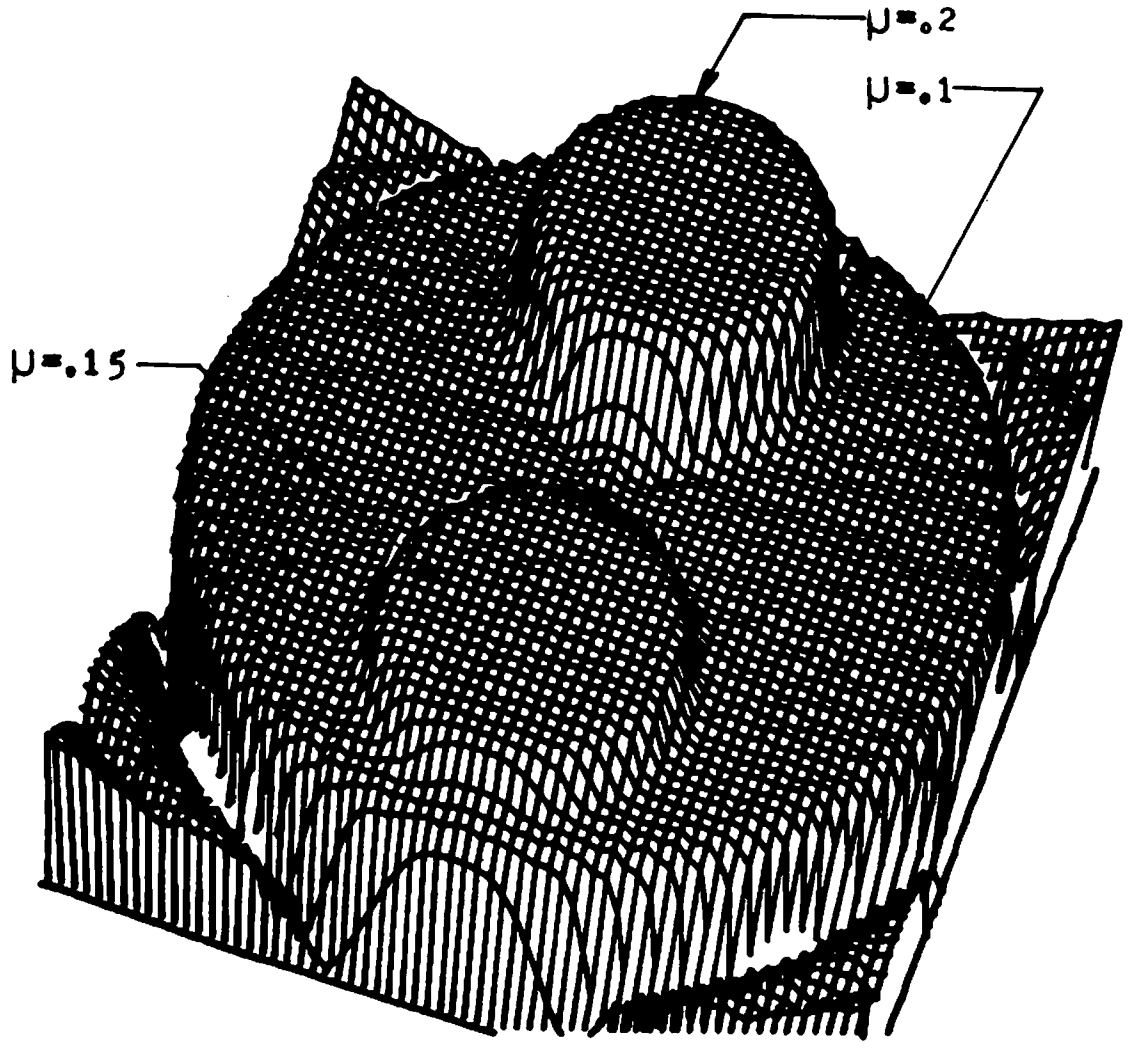


Figure 4.7(a). Phantom Reconstruction with $M_r = 64$, $M_\theta = 128$.

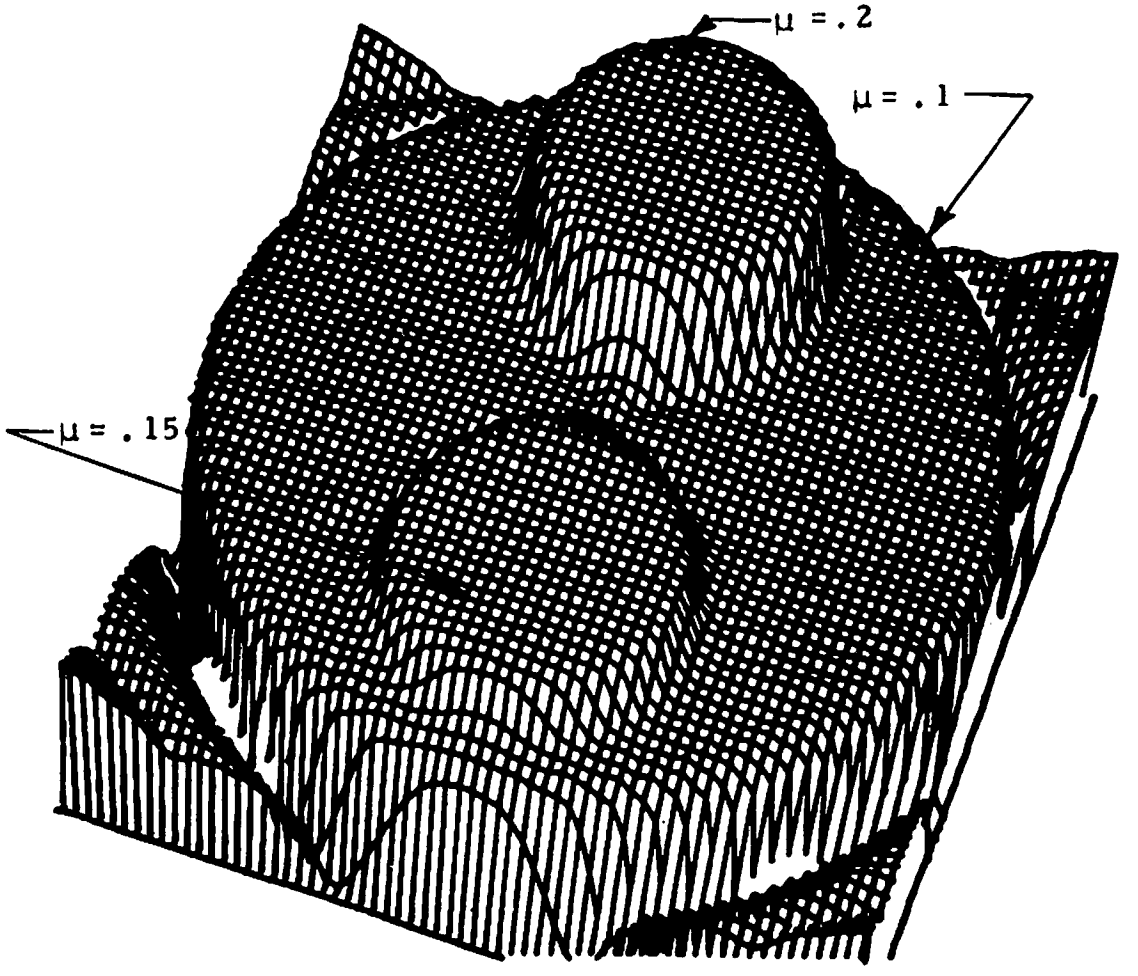


Figure 4.7(b). Phantom Reconstruction with $M_r = 64$, $M_\theta = 256$.

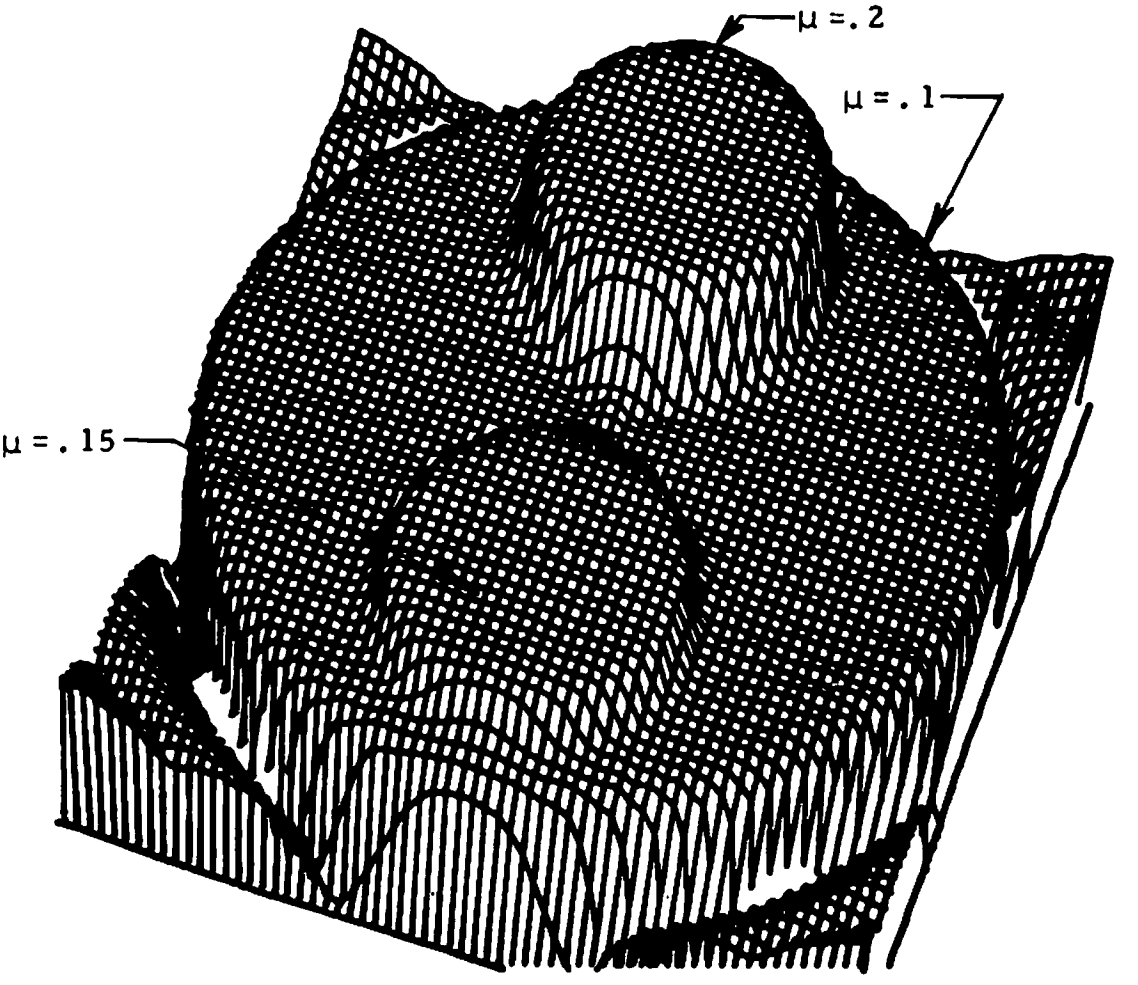


Figure 4.7(c). Phantom Reconstruction with $M_r = 64$, $M_\theta = 512$.

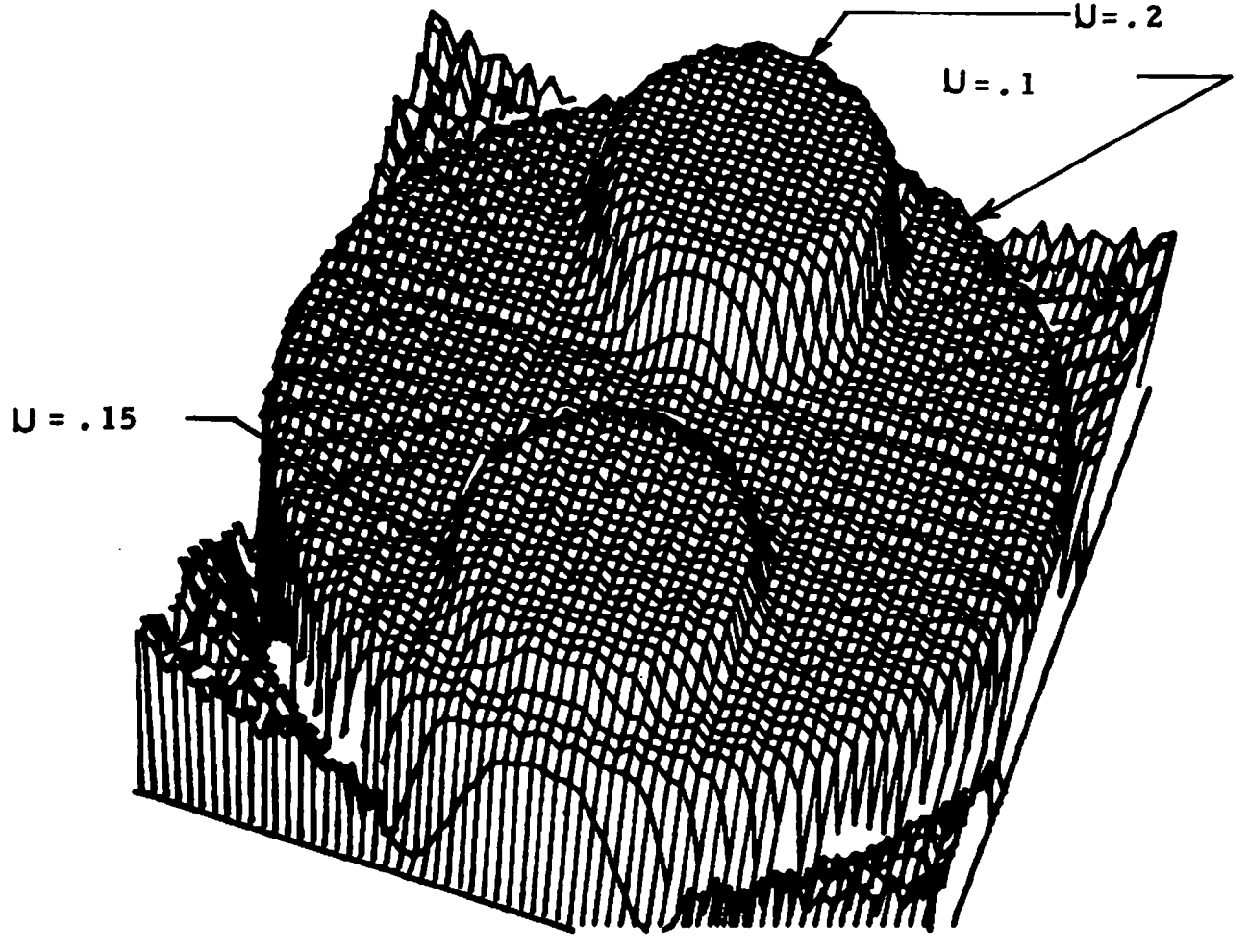


Figure 4.8. Phantom Reconstruction with $M_r = 64$, $M_\theta = 64$.

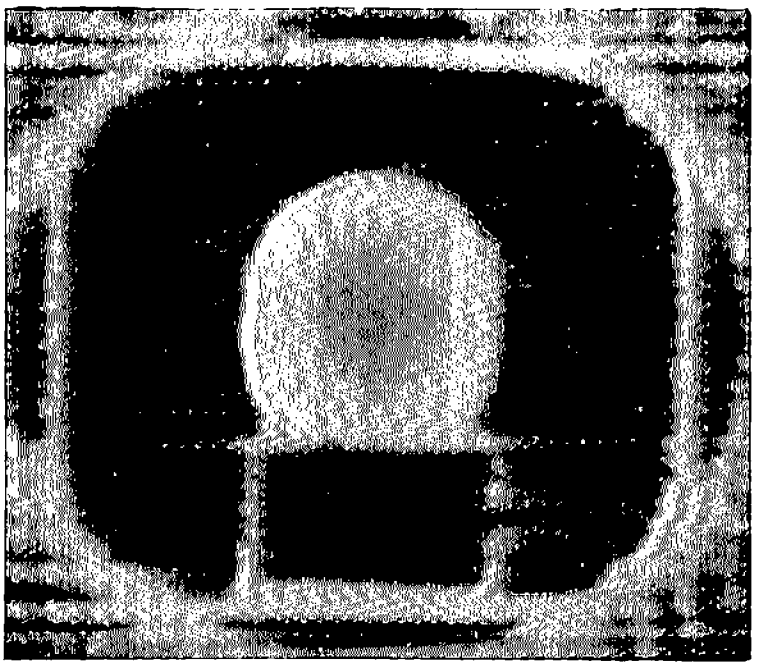


Figure 4.9(a). Monkey's Head Reconstruction with $M_r = 32$, $M_\theta = 64$.

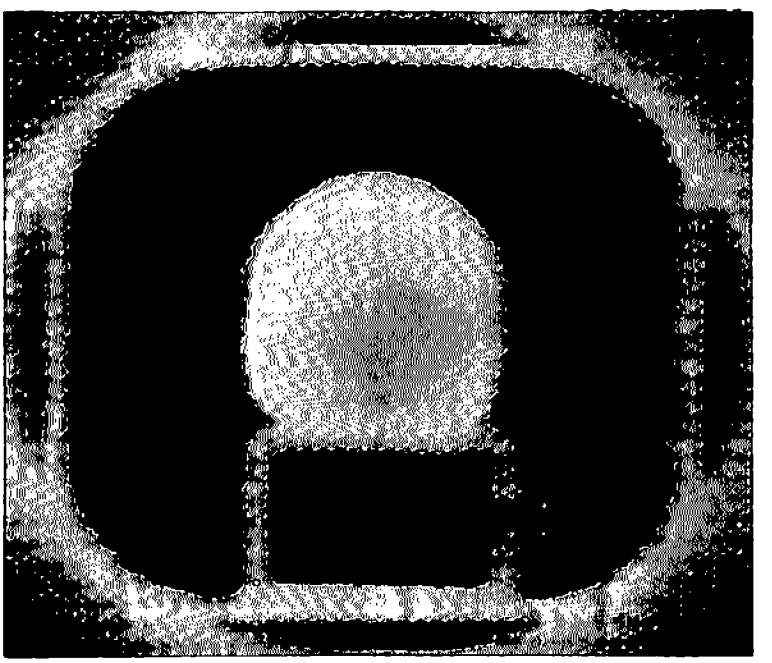


Figure 4.9(b). Monkey's Head Reconstruction with $M_r = 32, M_\theta = 128$.

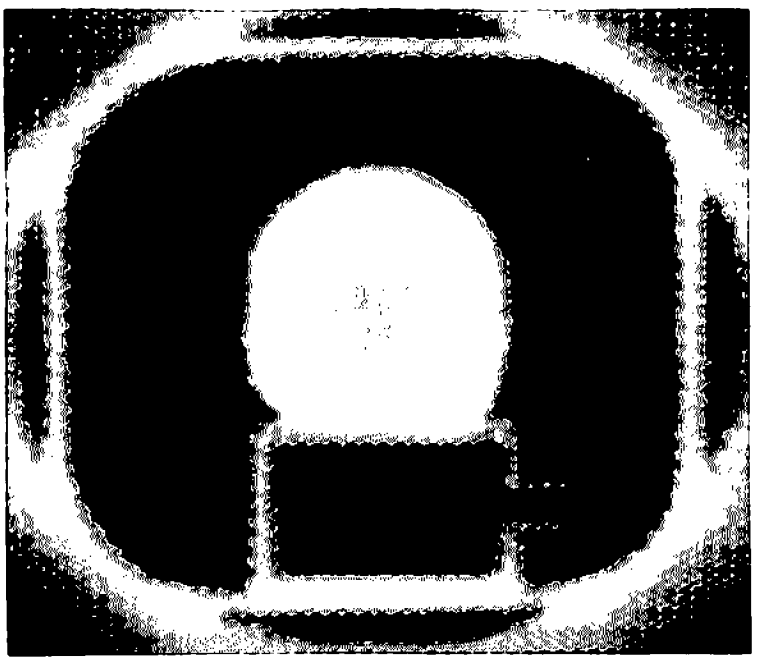


Figure 4.9(c). Monkey's Head Reconstruction with $M_r = 32$, $M_\theta = 256$.

Chapter 5

THE DEGREES OF FREEDOM OF SAMPLED IMAGE

5.1 Introduction

In the preceding chapter the number of effectively nonzero eigenvalues of the gram matrix has been used to characterize the degrees of freedom of a sampled image when viewed as the output of a linear imaging system. However, difficulties arise if the image is considered as an entity to itself as in the coding problem. While it is reasonable in the discussion of sampled images to assume that we are dealing with bandlimited scenes sampled at least at the respective Nyquist rates along the x and y directions in the output palne, we will thus assume that every image is the output of a linear imaging system, namely the ideal low pass spatial filter whose Fourier transform H(u, v) is as follows:

$$H(u, v) = \text{Rect}\left(\frac{u}{2B_x}\right)\text{Rect}\left(\frac{v}{2B_y}\right)$$

where

$$\text{Rect}(x) = \begin{cases} 1 & |x| \leq \frac{1}{2} \\ 0 & \text{otherwise} \end{cases}$$

B_x and B_y are the one-sided bandwidths in the x and y directions. The system weighting function is then given by

$$h(x, y;\xi, \eta) = 2B_x \ \mathrm{Sinc}(2B_x(x-\xi))2B_y \ \mathrm{sinc}(2B_y(y-\eta))$$

where

$$\mathrm{sinc}(x) = \frac{\sin(\pi x)}{\pi x}$$

While this assumption is certainly reasonable it leads to trivial results as will now be shown.

We can consider the bandlimited image f(x, y) under consideration to be related to some other non-bandlimited image $f'(\xi, \eta)$ through the ideal bandpass filter $h(x, y;\xi, \eta)$ as follows

$$f(x, y) = \int_{-\infty}^{\infty}\int 2B_x \ \mathrm{sinc}(2B_x(x-\xi))2B_y \ \mathrm{sinc}(2B_y(y-\eta))f'(\xi, \eta)d\xi d\eta \qquad (5\text{-}1)$$

which results in the following continuous-discrete representation

$$f(x_i, y_j) = \int_{-\infty}^{\infty}\int 2B_x \ \mathrm{sinc}(2B_x(x_i-\xi))2B_y \ \mathrm{sinc}(2B_y(y_j-\eta))f'(\xi, \eta)d\xi d\eta$$

Note that the kernel of Eq. (5-1) separates implying that $[\Gamma] = [\Gamma]^{(1)} \otimes [\Gamma]^{(2)}$ defining

$$[\Gamma]^{(1)} = [\gamma^{(1)}_{i,j}]$$

$$[\Gamma]^{(2)} = [\gamma^{(2)}_{m,n}]$$

$\gamma^{(1)}_{i,j}$ and $\gamma^{(2)}_{m,n}$ are as follows

$$\gamma^{(1)}_{i,j} = 4B_x^2 \int_{-\infty}^{\infty} \mathrm{sinc}(2B_x(x_i-\xi))\mathrm{sinc}(2B_x(x_j-\xi))d\xi \qquad i, j = 1, \ldots, N_x$$

$$\gamma^{(2)}_{m,n} = 4B_y^2 \int_{-\infty}^{\infty} \mathrm{sinc}(2B_y(y_m-\eta))\mathrm{sinc}(2B_y(y_n-\eta))d\eta \quad m, n = 1, 2, \ldots, N_y$$

Working with $[\Gamma]^{(1)}$ if we let $\theta = x_i - \xi$, it can be shown that $\gamma^{(1)}_{i,j}$ is given by the following convolution:

$$\gamma^{(1)}_{i,j} = 4B_x^2 \int_{-\infty}^{\infty} \text{sinc}(2B_x(x_i - x_j - \theta))\text{sinc}(2B_x\theta)d\theta$$

$$= 2B_x \text{ sinc}(2B_x(x_i - x_j))$$

Note that

$$\gamma^{(1)}_{i,j} = \begin{cases} 2B_x & \text{if } x_i = x_j \\ 0 & \text{if } \frac{x_i - x_j}{2B_x} = 1, 2, \ldots \end{cases}$$

and

$$\gamma^{(2)}_{m,n} = \begin{cases} 2B_y & \text{if } y_m = y_n \\ 0 & \text{if } \frac{y_m - y_n}{2B_y} = 1, 2, \ldots \end{cases}$$

Thus we see that sampling at the respective Nyquist rates along x and y produces a gramian that is an $N_x \cdot N_y$ square matrix that is diagonal as follows

$$[\Gamma] = 4B_xB_y \cdot [I]_{N_x \times N_x} \otimes [I]_{N_y \times N_y}$$

$$[\Gamma] = 4B_xB_y [I]_{N_x \cdot N_y \times N_x \cdot N_y}$$

where $[I]_{N \times N}$ is an NxN identity matrix.

This is a little less than satisfactory for another reason, namely that to obtain this full rank diagonal gram matrix we must

sample at the respective Nyquist rates everywhere along x and y. Since images are rarely stationary, the frequency content of the image is not constant over the entire region upon which the image is defined which would intuitively suggest that we are oversampling unless the frequency content of the image is spatially invariant.

Thus to define the degrees of freedom for an image of this type another approach is necessary. One such approach is to apply the singular value decomposition (SVD) algorithm [5-1] to the sampled image matrix whereupon the number of degrees of freedom can be equated with the number of effectively non-zero singular values with additional parameters needed to describe the singular vectors.

This method deals with a sampled version of the image only and in such a way that the degrees of freedom are affected by the sampling method used and can result in misleading results. This is readily apparent by considering an image $f(x, y)$ that can be written as the product of two functions $f_1(x)$ and $f_2(x)$ as $f(x, y) = f_1(x) \cdot f_2(y)$. If this image is sampled on a Cartesian grid (x_i, y_k), $i = 1, 2, \ldots, N, k = 1, 2, \ldots, N$, then the image matrix can be written as the outer product of the two vectors $[f_1(x_1) \ldots f_1(x_N)]$ and $[f_2(y_1) \ldots f_2(y_N)]$ and will be a rank one matrix for all such separable images f. This represents, at most, 2N degrees of freedom. Again the point to be made is that the degrees of freedom should be a characteristic of the original

image and reflected in the sampled image only by our inability to collect an uncountably infinite number of samples for application on a computer.

This brings up some conceptual difficulties since the number of degrees of freedom of a function defined on a continuum is countably infinite at best, viz., the space of all square integrable functions on $[0, 1]$ wherein any function can be expressed in a L_2 sense by a countably infinite expansion of any orthonormal basis functions that span the space. Thus we need to determine in what sense a class of functions defined on the unit rectangle, $\mathcal{S} = \{x, y; -1 \leq x, y \leq 1\}$ are finitely representable. It is to this end to which approximation theory is directed.

5.2 Degrees of Freedom and Shannon's Sampling Formula

Before proceeding further a few words considering bandlimited functions are in order since Shannon's sampling theorem provides a method to relate the original function to its sampled version and, to some extent, the ability to finitely represent the image when it is available only over the rectangle $\mathcal{R} = \{x, y: |x| \leq X/2, |y| \leq Y/2\}$. In this case the image f(x, y) is given in terms of its samples taken at the respective Nyquist rates along x and y as follows:

$$f(x,y) = \lim_{\substack{N_x \to \infty \\ N_y \to \infty}} \sum_{i=-N_x}^{N_x} \sum_{y=-N_y}^{N_y} f\left(\frac{i}{2B_x}, \frac{j}{2B_y}\right) \operatorname{sinc}\left(2B_x\left(x - \frac{i}{2B_x}\right)\right)$$

$$\cdot \operatorname{sinc}\left(2B_y\left(y - \frac{j}{2B_y}\right)\right) \qquad (5\text{-}2)$$

Here B_x and B_y are the respective bandwidths in the x and y directions. Equation (5-2) involves infinite sums and in any practical situation an image f(x, y) is taken to be non-zero only over the rectangle, $\mathcal{R}$, so that $N_x = B_x X$ and $N_y = B_y Y$ results in $4B_x X B_y Y$ samples of f(x, y) for application in (5-2). Then finite term Shannon interpolation becomes an approximation problem with the L_2 error given by:

$$\left[\sum_{|i|>N_x} \sum_{|j|>N_y} \left|f\left(\frac{i}{2B_x}, \frac{j}{2B_y}\right)\right|^2\right]^{\frac{1}{2}}$$

which is dependent upon the rate at which $\left|f\left(\frac{i}{2B_x}, \frac{j}{2B_y}\right)\right|^2 \to 0$ outside $\mathcal{R}$.

The results of an earlier portion of this section concerning the gram matrix of the bandlimiting operator are interesting in this context, in that they indicate that sampling at the respective Nyquist rates along the x and y directions will produce a diagonal gramian of constant entries whose dimension will be $4B_x X B_y Y$. Increasing the number of samples beyond this point will force the gram matrix to contain non-zero off diagonal terms implying an increased

correlation between samples. Since this gram matrix is that of a bandlimited operator observed over a finite area, increasing the number of samples sufficiently will produce a gramian eigenvalue map similar to that of the continuous-continuous operator--namely one that is constant out to a point, not significantly greater than $4B_xXB_yY$, with an abrupt drop beyond that point. Thus we can consider that there could be approximately $4B_xXB_yY$ independent samples of f(x, y) in $\mathcal{R}$.

In reality Shannon's sampling theorem provides a method of bridging the gap between the continuous and discrete for bandlimited images by providing a reconstruction method involving uniformly spaced samples over $\mathcal{R}$. However, by considering the problem in a more general approach it may be possible to define other adaptive and possibly more efficient approximation methods.

In this light consider that the spatial frequency content of most images is not constant over the entire image and the possibility exists that if the rectangle $\mathcal{R}$ is broken up into 2 sets $\mathcal{R}_1$ and $\mathcal{R}_2$ such that $\mathcal{R} = \mathcal{R}_1 \cup \mathcal{R}_2$ then the portions of f(x, y) defined over $\mathcal{R}_1$ and $\mathcal{R}_2$ can be considered sections of a bandlimited image with x and y bandwidths $B_x^{(1)}$, $B_x^{(2)}$ and $B_y^{(1)}$, $B_y^{(2)}$ in each of these regions. By way of example consider $\mathcal{R}_1 = \{x, y: -X/2 \le x \le 0, |y| < Y/2\}$, $\mathcal{R}_2 = \{x, y: 0 \le x \le X/2, |y| \le Y/2\}$. Then f(x, y) need be sampled only at the respective Nyquist rates in each of these regions.

For $\mathcal{R}_1$ and $\mathcal{R}_2$ respectively we obtain $N_x^{(1)}N_y = 2B_x^{(1)}XB_yY$ and $N_x^{(2)}N_y = 2B_x^{(2)}XB_yY$ so that the approximate number of independent samples is $2XYB_y(B_x^{(1)}+B_x^{(2)})$. Now either $B_x^{(1)} \leq B_x$ or $B_x^{(2)} \leq B_x$ so that it follows that the number of independent samples is less than or equal to $4B_xXB_yY$. This heuristic argument is presented to motivate the question of why should an image be sampled over its entire domain of definition at a Nyquist rate dependent on a bandwidth that may be relevant to only a small region. A similar idea will be presented in the next chapter for bicubic splines.

In summary then the eigenvalue map of the bandlimited gram matrix gives an upper bound to the number of independent samples available and a subsectioning of the image represents a low order attempt at reducing this upper bound--hopefully at an acceptable error.

5.3 The Degrees of Freedom Viewed as an Approximation Problem

In characterizing the degrees of freedom of an image as an approximation problem we are confronted with two questions, namely: 1) to what extent is f(x, y) finitely representable, and 2) the determination of the finite representation from a sampled version of f(x, y).

In dealing with these questions we will take f(x, y) to be an element of a metric space W with metric d_W. For example, this could be the space $L^2(\mathcal{S})$ where $\mathcal{S} = \{x, y: -1 \leq x, y \leq 1\}$, if we do not distinguish functions that differ only on a set of measure zero. Consider

then the following generalized bivariate polynomial approximation scheme $\mathcal{L}_N^o$:

$$\hat{f}(x,y) = \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} a_{ij}^o \ell_i(x)\ell_j(x) \triangleq \mathcal{L}_N^o \quad N = N_x \cdot N_y$$

with the property that if we have any other approximation scheme $\mathcal{L}_N$ where

$$\mathcal{L}_N = \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} a_{ij} \ell_i(x)\ell_j(y)$$

that

$$d_W(f, \mathcal{L}_N^o) \leq d_W(f, \mathcal{L}_N) .$$

Then $\mathcal{L}_N^o$ is a best approximating generalized polynomial in the metric d_W. We also require that $\mathcal{L}_N^o \to f(x,y)$ as $N \to \infty$ so that for every $\epsilon > 0$, there exists an $N(\epsilon, \ell, d_W)$ (in general dependent on the set $\{\ell\}$ and the metric d_W) such that $N_x N_y > N(\epsilon, \ell, d_W)$ implies that $d_W(f, \mathcal{L}_W^o) < \epsilon$.

Thus we could define the degrees of freedom of f at level epsilon, or more succinctly the epsilon degrees of freedom, DOF (d_W, ϵ), as

$$DOF(d_W, \epsilon) = \inf_{\{\ell\} \neq f} \{N(\epsilon, \ell, d_W): d_W(f, \mathcal{L}_N^o) < \epsilon\} .$$

Here the infimum is taken over all approximating functions not equal to f to avoid the trivial case of $DOF(d_W, \epsilon)$ equaling 1, (the approximating function being f itself).

In general, this is a difficult problem since for each set of functions $\{\ell_1(x)\ \ell_1(y)\ldots\ell_{N_x}(x)\ell_{N_y}(y)\}$ we must find a best approximating bivariate generalized polynomial (if it exists) and then determine the set of functions minimizing N_xN_y. Thus we are concerned with the existence of a best approximation in addition to its determination.

However, if we consider a one dimensional case where N is taken to be the space of bandlimited L_2 functions observed over a finite interval of the real line, the distance d_W being the L_2 metric, the result is known. Here for every $\epsilon>0$ the inf $N(\epsilon, \ell, d_W)$ was found by Landau and Pollak [5-2] to be achieved by the functions $\{\phi\}$ which are related to the prolate spheroidal waveforms, as first described by Slepian and Pollak [5-3]. While this is known, the determination and utilization of these waveforms for large space bandwidth products has met with little success. Most likely then to arrive at any significant results in two dimensions, we will have to restrict the search to those sets of functions that are computationally feasible and possess the approximating properties required.

In the context of a digital computer the actual functions available are those that can be obtained by finite sums and products. Thus ultimately all functions must be reduced to those that can either be generated by recursions or by truncated and shifted polynomials. So

a restriction of the search of functions to polynomial splines would seem to be reasonable.

Polynomial splines are chosen due to their approximation properties and the fact that they possess a basis namely, the normalized B-spline basis, that provides a local basis property thus allowing a rapid generation with the matrices involved in generating a B-spline fit to a function f being well conditioned. With DeBoor's algorithm for computations using normalized B-splines [5-4] no difficulties are encountered in handling multiple order knots. Hereafter we will consider a spline $S_{k, N_x N_y}$ of the order k with (the degree being equal to k-1) N_x and N_y knots in the x and y directions respectively, to be of the following form:

$$S_{k, N_x N_y}(x, y) \triangleq \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} S_{i,j} N_{i,k}(\underline{\xi}; x) N_{j,k}(\underline{\eta}; y) = \hat{f}(x, y)$$

where $N_{i,k}(\cdot)$ are the normalized B splines of order k satisfying the following recursion relationship over the knot sets $\{\xi_1, \xi_2, \ldots, \xi_{N_x}\}$ and $\{\eta_1, \eta_2, \ldots, \eta_{N_y}\}$ [5-4].

$$N_{i,k}(\underline{\xi}; x) = \frac{x-\xi_i}{\xi_{i+k-1}-\xi_i} N_{i,k-1}(\underline{\xi}; x) + \frac{\xi_{i+k}-x}{\xi_{i+k}-\xi_{i+1}} N_{i+1,k-1}(\underline{\xi}; x)$$

$$N_{i,1}(\underline{\xi}; x) = \begin{cases} 1 & x \in [\xi_i, \xi_{i+1}) \\ 0 & \text{otherwise} \end{cases}$$

and

$$\sum_{i=1}^{N_x} N_{i,k}(\xi;x) = 1 \qquad \forall x \epsilon [\xi_1, \ldots, \xi_{N_x}]$$

These functions are discussed in Appendix A.

For the duration of this chapter we will consider $f(x, y) \epsilon L_2(\mathcal{S})$ with the distance being the L_2 metric. Thus for any x and y knots sets $\{\xi_1 \ldots \xi_{N_x}\}$ and $\{\eta_1, \ldots, \eta_{N_y}\}$ the matrix $[S_{i,j}]$ of coefficients minimizing the L_2 error for a particular set of x and y knots is given by the solution to the following matrix representation of the normal equations:

$$[\Gamma]^{(x)}_{N_x \times N_y} [S_{ij}]_{N_x \times N_y} [\Gamma]^{(y)}_{N_x \times N_y} = \iint_E \underline{N}_k(\underline{\xi};x)\underline{N}_k^T(\underline{\eta};y) f(x, y)dxdy \tag{5-3}$$

where $\underline{N}_k(\underline{\xi};x)$ and $\underline{N}_k(\underline{\eta};y)$ are respectively the N_x and N_y column vectors of the normalized B-spline basic functions and

$$[\Gamma]^{(x)} = [\gamma^{(x)}_{ij}],$$

$$[\Gamma]^{(y)} = [\gamma^{(y)}_{ij}]$$

where

$$\gamma^{(x)}_{ij} = \int_{-1}^{1} N_{i,k}(\underline{\xi};x)N_{j,k}(\underline{\xi};x)dx$$

$$\gamma^{(y)}_{ij} = \int_{-1}^{1} N_{i,k}(\underline{\eta};y)N_{j,k}(\underline{\eta};y)dy$$

Thus finding the best approximation to f(x, y) over the x and y knots sets can be stated as follows:

Minimize: $$\left[\iint_{\mathcal{S}} |f(x,y) - S_{k,N_x N_y}(x,y)|^2 dxdy \right]^{\frac{1}{2}}$$

over all possible x and y knot vectors $\{\xi_1, \ldots, \xi_{N_x}\}$, $\{\eta_1, \ldots, \eta_{N_y}\}$ subject to

$$|\xi_i| \leq 1 \qquad \forall i = 1, \ldots, N_x$$

$$|\eta_j| \leq 1 \qquad \forall j = 1, \ldots, N_y$$

This is nothing more than nonlinear minimization over the possible knot sets in $\mathcal{S}$, the solution of which has been shown to exist by Rice [5-5].

Thus specifying an error ϵ, we can find the epsilon degrees of freedom by a sequence of minimizations decreasing the number of knots in the x and y directions until we reach a point at which the error will be exceeded with any further restriction in the number of knots.

For this to make sense we must be assured that for every $\epsilon > 0$ there exists N_x and N_y such that

$$\| f(x,y) - S_{k,N_x N_y}(x,y) \|_2 < \epsilon$$

and that a minimum over the knot sets exists. Addressing this convergence problem Schultz [5-6] has shown for k = 4 that

$$\|f(x,y) - S_{N_x N_y}(x,y)\|_2 \le \left(\frac{\bar{\rho}}{\pi}\right)^4 \left\{ \left\| \frac{\partial^4}{\partial x^4} f(x,y) \right\|_2 + \left\| \frac{\partial^2}{\partial x^2} \frac{\partial^2}{\partial y^2} f(x,y) \right\|_2 + \left\| \frac{\partial^4}{\partial y^4} f(x,y) \right\|_2 \right\}$$

where $\bar{\rho} = \max\{\max_i(\xi_{i+1}-\xi_i),\ \max_j(\eta_{j+1}-\eta_j)\}$. Thus taking $N_x \ge \frac{2}{\bar{\rho}}$, $N_y \ge \frac{2}{\bar{\rho}}$ we have as $N_x, N_y \to \infty$ such that $\bar{\rho} \to 0$ the bicubic spline approximation will converge to $f(x,y)$ in an L_2 sense over the unit rectangle $\mathcal{S}$.

5.4 Computational Considerations and Conclusions

Heretofore the concern has been the existence of a solution to the problem of finding $DOF(d_W, \varepsilon)$ of an image on the unit rectangle. While a solution exists in the form of a sequence of best approximating splines with a successively decreasing number of knots, this requires the solution of a nonlinear minimization at each step. For approximations of the form of (5-3) involving even a modest number of knots this is computationally infeasible. Furthermore since we are dealing with a computer, the right side of (5-3) must be calculated using numerical quadrature and this involves only an approximation. Thus we must settle for a procedure that provides an approximation with a good, if not optimal, knot placement. In the next chapter some numerical results involving bicubic spline approximations to simulated and actual images will be given. It will

be shown that there exist relatively simple methods for placing the knots that provide much better approximations than a bicubic spline with uniformly placed knots.

Chapter 6

EXPERIMENTAL RESULTS FOR SPLINE APPROXIMATION

6.1 Introduction

The purpose of this chapter is to present some numerical results concerning the "epsilon degrees of freedom" concepts developed in the preceding chapter. This was developed as an approximation problem whose solution was seen to involve the determination of a sequence of best approximating (in an L_2 sense) B-splines with variable knots. While the determination at each step of such a best approximating spline is simply a nonlinear minimization problem over the knots defining the spline, it is computationally infeasible. Thus we must follow DeBoor [6-1] and settle for spline approximations with good if not optimal knot placements. In what follows, two easily implemented methods for placing the knots will be given that can result in a significant error reduction over the uniform knot case for the proper class of images.

The results in this chapter will be developed using cubic splines giving the following fourth order spline approximation $\hat{f}(x, y)$

$$\hat{f}(x, y) = \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} S_{ij} N_{i,4}(\underline{\xi}; x) N_{j,4}(\underline{\eta}; y) \triangleq S_{4, N_x, N_y}(x, y) \quad (6\text{-}1)$$

where $\underline{\xi}$ and $\underline{\eta}$ are the knot vectors in the x and y directions respectively. The spline coefficients, S_{ij} were obtained by solving the following system of equations:

$$\hat{f}(x_\ell, y_m) = \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} S_{ij} N_{i,4}(\underline{\xi}; x_\ell) N_{j,4}(\underline{\eta}; y_m)$$

for $\ell = 1, 2, \ldots, N$ and $m = 1, 2, \ldots, N$. In matrix notation this becomes

$$[f(x_\ell, y_m)]_{N \times N} = [N_{i,4}(\underline{\xi}, x_\ell)]^T_{N \times N_y} [S_{ij}]_{N_x \times N_y} [N_{j,4}(\underline{\eta}, y_m)]_{N_y \times N} \tag{6-2}$$

where $[\ ]^T$ indicates matrix transpose. To simplify notation let

$$[f(x_\ell, y_m)] = [F]$$

$$[N]^{(\underline{\xi})} = [N_{i,4}(\underline{\xi}, x_\ell)]$$

$$[N]^{(\underline{\eta})} = [N_{j,4}(\underline{\eta}; y_m)]$$

equation (6-2) becomes

$$[F] = [N]^{(\underline{\xi})^T} [S_{ij}] [N]^{(\underline{\eta})}$$

Since $N > N_x$ and $N > N_y$ in general equation (6-2) cannot be solved exactly. However the spline coefficients minimizing the normalized least squares error ϵ given by

$$\epsilon = \frac{\sum_{\ell=1}^{N}\sum_{m=1}^{N} |f(x_\ell, y_m) - \hat{f}(x_\ell, y_m)|^2}{\sum_{\ell=1}^{N}\sum_{m=1}^{N} |f(x_\ell, y_m)|^2} \tag{6-3}$$

can be obtained by taking $[S_{ij}]$ to be

$$[S_{ij}] = [[N]^{(\underline{\xi})^T}[N]^{(\underline{\xi})}]^{-1}[N]^{(\underline{\xi})^T}[F][N]^{(\underline{\eta})^T}[[N]^{(\underline{\eta})^T}[N]^{(\underline{\eta})}]^{-1} \tag{6-4}$$

6.2 Knot Placement from Projections

In this section we investigated the possibility of placing the knots for the x and y knot vectors from the projections of f(x, y) along the y and x directions respectively. To do this we borrowed an idea suggested by DeBoor [6-1], wherein he suggested placing the $i+1^{\underline{st}}$ knot with respect to the $i^{\underline{th}}$ knot for a $k^{\underline{th}}$ order spline according to the following

$$\int_{\xi_i}^{\xi_{i+1}} |f^{(k)}(x)|^{1/k} dx = \text{constant}$$

where $f^{(k)}(x)$ indicates the k^{th} derivative of f(x). To determine the x knot set for a bicubic spline fit the knots were such that

$$\int_{\xi_i}^{\xi_{i+1}} \left| \frac{\partial^4}{\partial x^4} \int_{-1}^{1} f(x, y)dy \right|^{\frac{1}{4}} dx = \text{constant} \tag{6-5}$$

The y knot set is obtained in a similar manner. Note that if $f(x, y)$ can be written as $f(x, y) = f_1(x) \cdot f_2(y)$ then equation (6-4) can be written as

$$\int_{\xi_i}^{\xi_{i+1}} \left| \frac{\partial^4}{\partial x^4} f_1(x) \right|^{\frac{1}{4}} \left| \int_{-1}^{1} f_2(y) dy \right|^{\frac{1}{4}} dx = \text{constant}$$

and we have that

$$\int_{\xi_i}^{\xi_{i+1}} \left| \frac{\partial^4}{\partial x^4} f_1(x) \right|^{\frac{1}{4}} dx = \frac{\text{constant}}{\left| \int_{-1}^{1} f_2(y) dy \right|^{\frac{1}{4}}}$$

It is reasonable to expect that this method of placing the knots would be quite effective for images that are either separable or exhibit a high degree of separability, and not so effective for images that do not. This point is illustrated for the following experiments involving two images--one an analytic image consisting of a Gaussian pulse with standard deviation σ^2 equaling 0.1 and the other consisting of an actual image of an armored personnel carrier (APC). For each experiment N was taken to be 128 so that the effects of the quadrature error implicit in the residual of equation (6-3) and the computations in equation (6-4) should not be a significant factor in evaluating the results.

The results for a bicubic spline fit to the Gaussian pulse with 10 knots placed by the projection method are shown in figure 6.1.

Also shown are the results obtained with 10 and 20 knots placed uniformly, the original image, and the knot placements for 10 knots placed uniformly and by the projection method. The corresponding mean square errors are tabulated in table 6.1. Note that visually the results of 10 knots placed according to this projection method are as good as of 20 knots placed uniformly and are far better than those obtained by placing 10 knots uniformly. From the results of table 6.1 it is clear that placing 10 knots in the x and y directions by the projection method is better than placing them uniformly even up to the situation employing 20 knots.

Figure 6.2 shows the results of a bicubic spline approximation to an actual image of an APC for 40 uniform knots in each direction and for 40 knots placed by the projection method. Also shown are the corresponding knot placements in each direction for both cases.

Figure 6.3 is the original. Here the results indicate that the projection method is not so good as the uniform method. This is evident in table 6.2 where the projection method results in an error that is an order of magnitude greater than that obtained in the uniform knot case.

While these results show that the proper placement of the knots can result in a significant error improvement in the approximation of separable images, another method must be employed for those images that are not separable.

6.3 Spline Approximations by Subsectioning

In this section the possibility of subsectioning the image and using different knot densities in each of the subsections is investigated. This method might provide fruitful results when one considers an L_2 error bound given by Schultz [6-2] for cubic splines. Recalling that the error is given by the L_2 norm, $\|\cdot\|_2$ of the difference between the function and its approximation, this bound is given by

$$\|f-\hat{f}\|_2 \leq 4\left(\frac{\bar{\rho}}{\pi}\right)^{-4}\left\{\left\|\frac{\partial^4}{\partial x^4} f(x,y)\right\|_2 + \left\|\frac{\partial^2}{\partial x^2}\frac{\partial^2}{\partial y^2} f(x,y)\right\|_2 + \left\|\frac{\partial^4}{\partial y^4} f(x,y)\right\|_2\right\} \tag{6-6}$$

where

$$\bar{\rho} = \max\{\max_i(\xi_{i+1}-\xi_i), \max_j(\eta_{j+1}-\eta_j)\} \tag{6-7}$$

From the discussion of Chapter 5 concerning local bandwidth and Shannon sampling it follows that if the image derivative energy is large only over a small region, then using a uniform knot bicubic spline with knot mesh width equaling $\bar{\rho}$ given by equation (6-7) should result in an overly good approximation of the image in those regions where the image derivative energy is low. Thus we should be able to obtain reasonable results by employing a different bicubic spline with uniformly spaced knots in each subsection, the knot density in each subsection being proportional to the value of

$$\left\| \frac{\partial^4}{\partial x^4} f(x, y) \right\|_2 + \left\| \frac{\partial^2}{\partial x^2} \frac{\partial^2}{\partial y^2} f(x, y) \right\|_2 + \left\| \frac{\partial^4}{\partial y^4} f(x, y) \right\|_2$$

in that subsection.

This approach was taken for the APC of figure 6.3. The image was partitioned into 64 square sections and the derivative energy in each section was calculated by central differencing. Figure 6.4 shows the results for this method along with the associated knot density. It can be seen that this associated knot density is highest in the region containing the insignia and lowest in the sandy regions surrounding the vehicle. The bright lines at the subsection boundaries indicate a fourth order knot at the boundary, and note that the regions containing the minimum number of knots have x and y knot vectors consisting of two fourth order knots each at the subsection boundary. As noted in Appendix A this corresponds to a bivariate cubic polynomial approximation in each such subsection.

Table 6.3 lists the number of parameters necessary to reconstruct the image, the reconstruction error, the data reduction ratio for the uniform knot case, the projection placement case, and the subpartitioning case. Note that the error is lowest for the subpartitioning case and that it provides the best reconstruction.

To further explore the subpartitioning method a series of bicubic approximations involving image subsections of different sizes

was run on the APC image, an aerial reconnaisance image and an image of Los Angeles International Airport (LAX). For this series the image dimension N was taken to be 256 and three subsection sizes of 32 x 32, 16 x 16, and 8 x 8 pixels were used for both images. For each subpartition size three knot density ranges were employed. In all cases the maximum knot density is taken to be such that the matrices of normalized B-splines in equation (6-4) are nonsingular thus resulting in the image being interpolated in at least one subsection. The lowest knot density in each subpartition sequence was taken to be that corresponding to a fourth order knot placed at each of the subregion boundaries for the x and y knot vectors respectively. The number of knots was then increased by raising the minimum number of knots employed. The results for the APC image along with the associated knot densities for subpartitions of size 32 x 32, 16 x 16 and 8 x 8 are shown in figures 6.5, 6.6, and 6.7 respectively. Figure 6.8 is the original. Here the fourth order knots at the subpartition boundaries are not displayed for aesthetic purposes. Note that all of the approximations are quite good and that the knot densities are quite adaptive for each subpartition size. The corresponding error, data reduction ratio, and number of parameters necessary for the bicubic spline approximation are listed in table 6.4. Note that the error for the 32 x 32 case corresponding to

the data reduction ratio of 5.62:1 is lower than the error for the 16 x 16 case for the 5.31:1 data reduction ratio. This would seem to indicate that for low error levels the 32 x 32 partition size is better than the 16 x 16 sized partition which itself is better than the 8 x 8 case.

This same sequence was obtained for the reconnaisance image and the results along with the corresponding knot density patterns are shown in figures 6.9, 6.10, and 6.11 for the 32 x 32, 16 x 16, and 8 x 8 cases respectively. Figure 6.12 is the original. Table 6.5 lists the relevant errors and corresponding parameters as for the APC. Note again that the best reconstruction for the 32 x 32 case is at a lower error than the 16 x 16 case, and at a higher corresponding data reduction ratio. This would again seem to indicate that for low error levels the 32 x 32 sized partition is better than the 16 x 16 partition.

The sequence for the image of LAX is contained in figures 6.13, 6.14, and 6.15. Figure 6.16 is the original. Note here that while the errors are relatively high the results are quite good visually and that the knots serve fairly well in locating the areas containing the aircraft. Note also that the best results occur for the 8 x 8 subpartitioning case. This is most likely due to the relatively small size of the items of interest - in this case the aircraft.

6.4 Summary and Conclusions

In this chapter the attempt has been to demonstrate the utility of variable knot splines in achieving a data reduction and quantifying the degrees of freedom of sample images by the number of variable knot bicubic splines necessary to approximate a particular image at an error level epsilon. Considerable success was achieved for an analytical image consisting of a Gaussian pulse with $\sigma = 0.1$. Here 10 knots, each in the x and y directions were placed by an algorithm dependent on the 4^{th} partials of the projections and provided an error reduction of three orders of magnitude over the situation where each of the knot sets were uniformly spaced in the x and y directions. The error was such that 10 knots placed in the above manner provided an error lower than that achieved by placing 20 knots uniformly in the x and y directions. This method was not very successful for the APC image due to the lack of any separability characteristics. However by subsectioning the image and employing a different spline approximation in each subsection whose knot density was dependent on the derivative energy in that region good results were obtained. A high degree of adaptability was in evidence through the knot density patterns with acceptable errors being obtained at reasonable data reduction ratios.

NUMBER OF KNOTS	PLACEMENT MODE	MEAN SQUARE ERROR
10	uniform	2.158×10^{-3}
10	4th differences	1.29×10^{-6}
20	uniform	3.24×10^{-6}

Table 6.1. Mean Square Error for Knot Placement on Gaussian Pulse $\sigma^2 = 0.1$.

NUMBER OF KNOTS	PLACEMENT MODE	MEAN SQUARE ERROR
40	uniform	6.413×10^{-3}
40	4th differences	6.308×10^{-2}

Table 6.2. Mean Square Error for Knot Placement on APC Image.

PLACEMENT MODE	NUMBER OF PARAMETERS	DATA REDUCTION RATIO	MEAN SQUARE ERROR
uniform	1936	8.46:1	6.413×10^{-3}
4th differences on projections	2016	8.12:1	6.308×10^{-2}
subpartitioning	1885	8.69:1	5.578×10^{-3}

Table 6.3. Data Reduction and Errors for a 128 x 128 Bicubic Spline Reconstruction of the APC Image.

SUBSECTION SIZE	NUMBER OF PARAMETERS NECESSARY TO DEFINE BICUBIC SPLINE	DATA REDUCTION RATIO	MEAN SQUARE ERROR
32 x 32	6,889	9.5 :1	.38%
	8,465	7.74:1	.32%
	11,649	5.62:1	.25%
16 x 16	7,071	9.26:1	.37%
	9,379	6.88:1	.31%
	12,334	5.31:1	.26%
8 x 8	18,070	3.62:1	.24%
	27,070	2.42:1	.17%
	38,021	1.72:1	.1 %

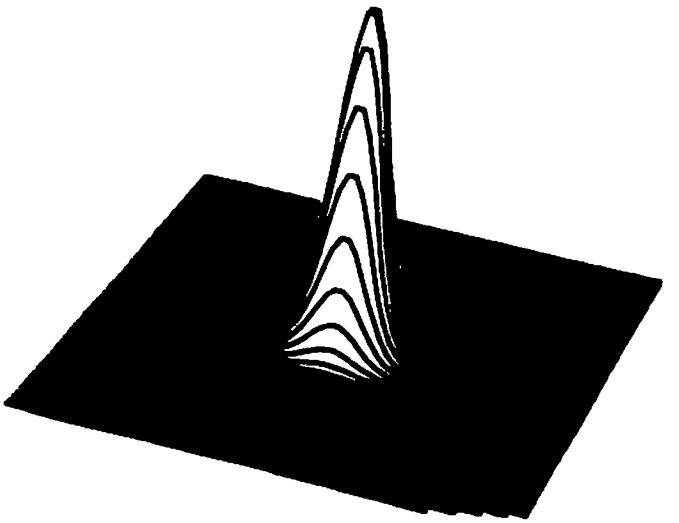
Table 6.4. Mean Square Error, Data Reduction and Number of Parameters for Subsectioned Bicubic Spline Approximation of a 256 x 256 APC Image.

SUBSECTION SIZE	NUMBER OF PARAMETERS NECESSARY TO DEFINE BICUBIC SPLINE	DATA REDUCTION RATIO	MEAN SQUARE ERROR
32 x 32	13,995	4.69:1	1.4 %
	18,910	3.46:1	.63%
	25,572	2.56:1	.38%
16 x 16	12,375	5.29:1	1.2 %
	17,465	3.71:1	.70%
	27,658	2.36:1	.39%
8 x 8	19,767	3.31:1	.77%
	27,809	2.35:1	.49%
	38,229	1.71:1	.31%

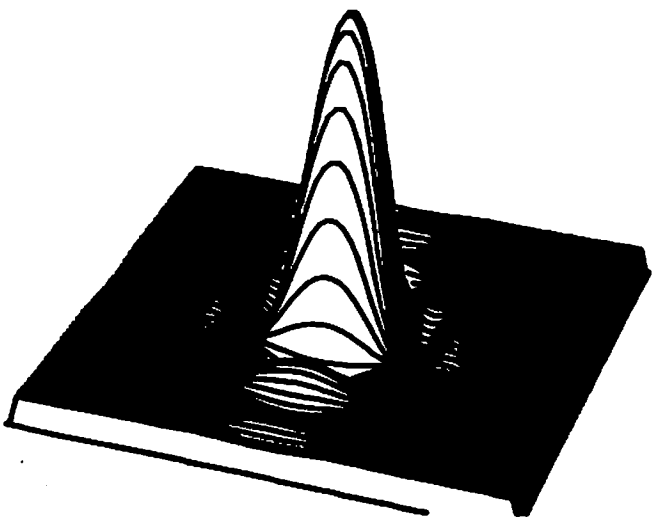
Table 6.5. Mean Square Error, Data Reduction Ratio and Number of Parameters for Subsectioned Bicubic Spline Approximation of a 256 x 256 Reconnaissance Image.

SUBSECTION SIZE	NUMBER OF PARAMETERS NECESSARY TO DEFINE BICUBIC SPLINE	DATA REDUCTION RATIO	MEAN SQUARE ERROR
32 x 32	7,129	9.19:1	2.8%
	11,954	5.48:1	1.9%
	19,245	3.41:1	1.3%
16 x 16	7,135	9.19:1	2.5%
	12,248	5.35:1	1.8%
	23,177	2.82:1	1.1%
8 x 8	18,449	3.55:1	1.5%
	27,361	2.39:1	1.0%
	38,216	1.71:1	.66%

Table 6.6 Mean Square Error, Data Reduction Ratio and Number of Parameters for Subsectioned Bicubic Spline Approximation of a 256 x 256 Image of LAX.

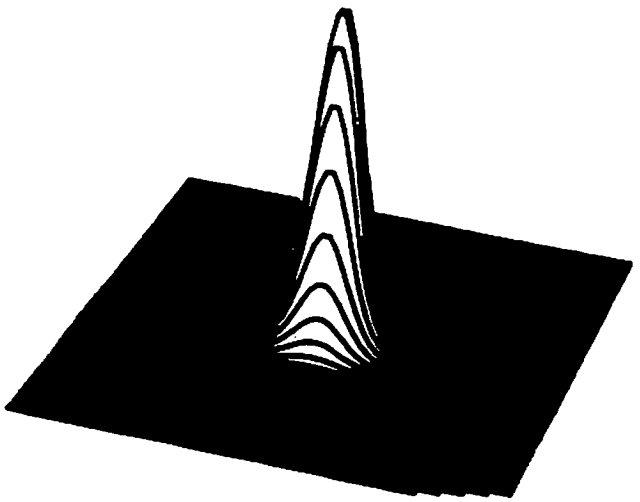


Bi-Cubic Spline Fit with 10 Knots in X and Y Direction Determined from 4th Partials of X and Y Projections.

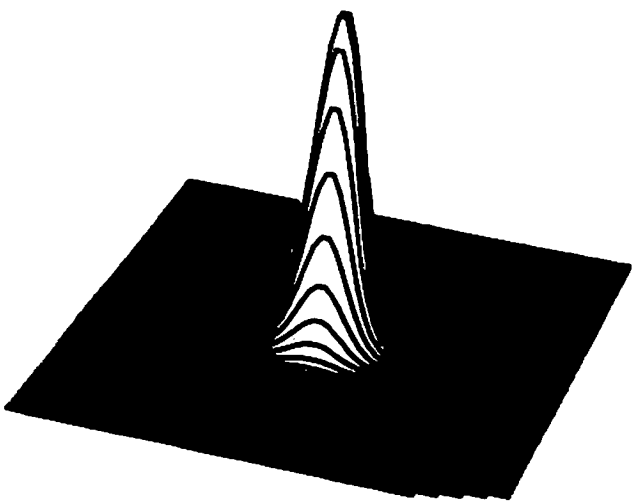


Bi-Cubic Spline Fit with 10 Uniform Knots in X and Y Directions.

Figure 6.1 Results For A Bi-Cubic Spline Fit To A Gaussian Pulse With $\sigma^2 = .1$

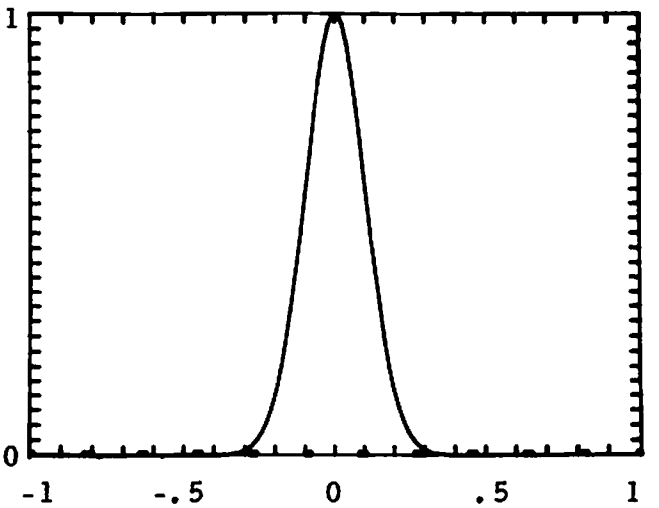


Bi-Cubic Spline Fit with 20 Uniform Knots in X and Y Directions

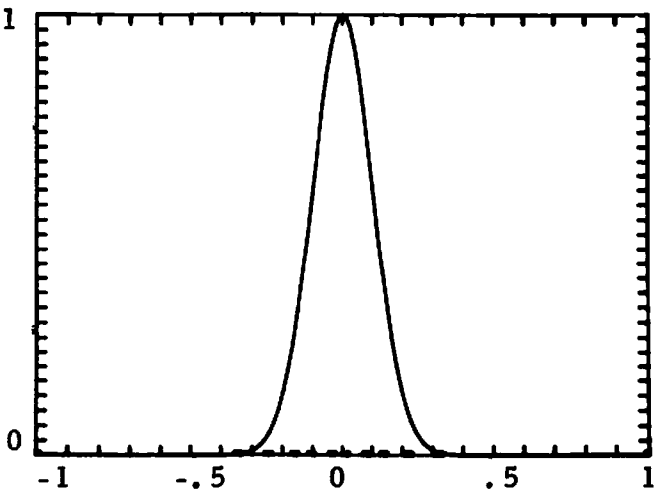


Original

Figure 6.1 (continued). Results For A Bi-Cubic Spline Fit To A Gaussian Pulse With $\sigma^2 = 1$

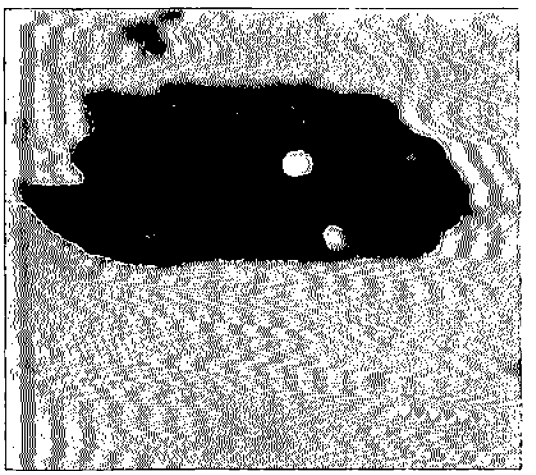


Y Projection and X-Knot Placement for 10 Uniform Knots (Y-Knot Placement Identical)

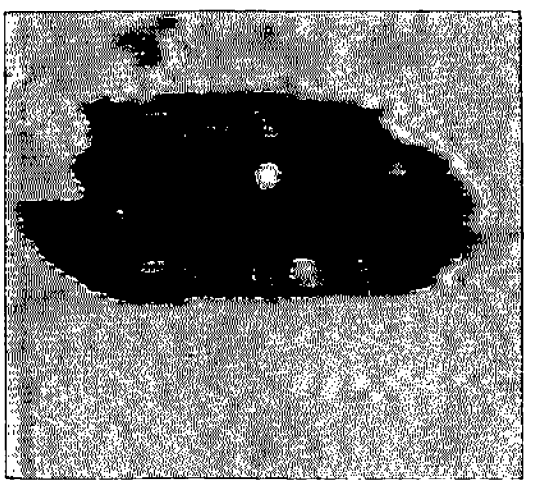


X Projection of f(x, y) and Y-Knot Placement from 4th Partial Algorithm for 10 Knots (X-Knot Placement Identical)

Figure 6.1 (continued). Results For A Bi-Cubic Spline Fit To A Gaussian Pulse With $\sigma^2 = .1$

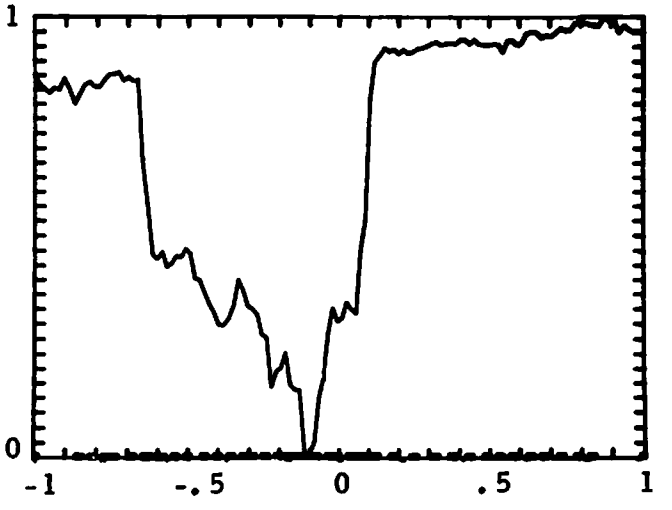


40 Knots Uniformly Placed in X and Y Directon

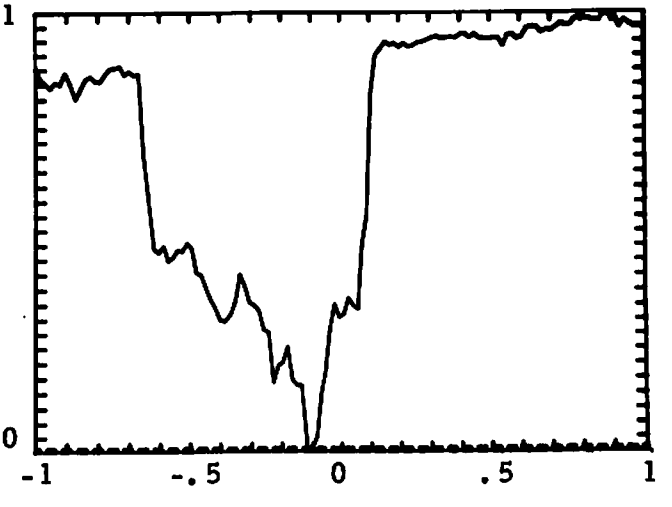


40 Knots in X and Y Directions placed by 4th Differences on X and Y Projections

Figure 6.2 Bi-Cubic Spline Approximation For APC With 40 Knots

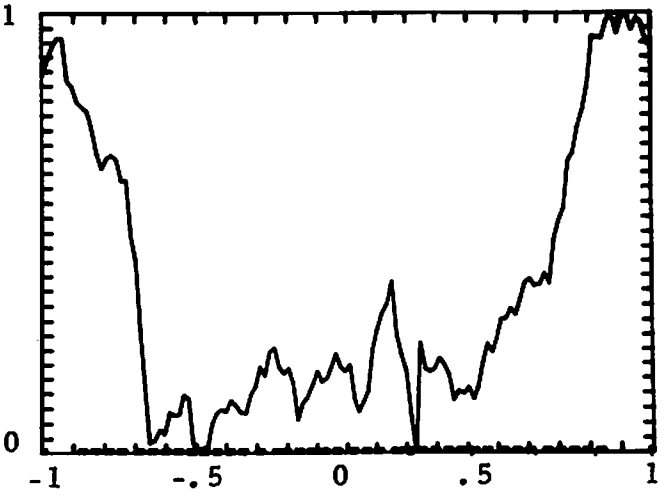


Knot Placement from 4th Differences
on Row Projections, 40 Knots

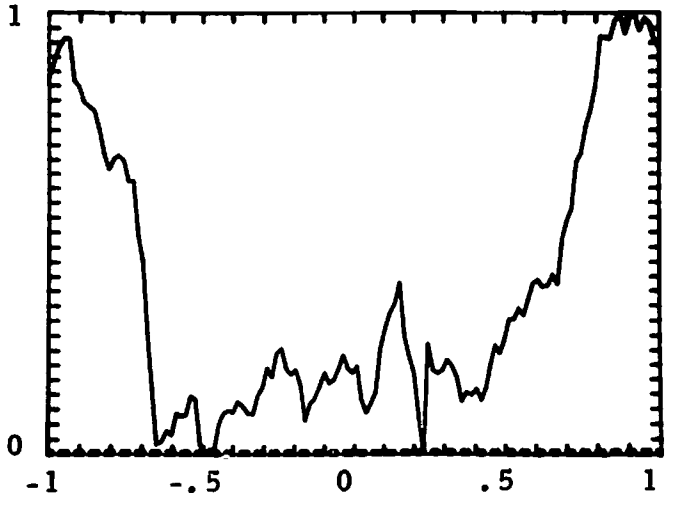


40 Knots Placed Uniformly

Figure 6.2 (continued) Column Knot
Placements For The APC Image



Knot Placement from 4th Differences on Column Projections, 40 Knots



40 Knots Placed Uniformly

Figure 6.2 (continued) Row Knot Placements For The APC Image

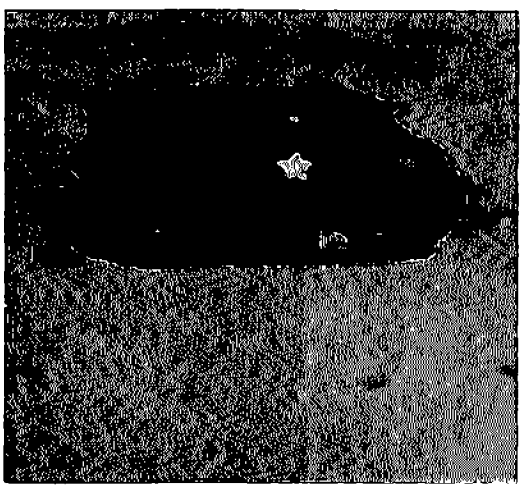
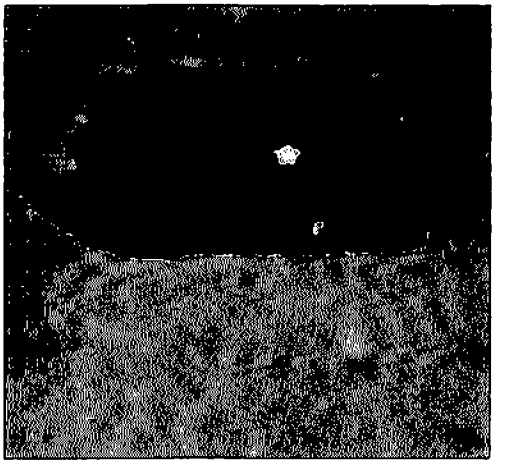
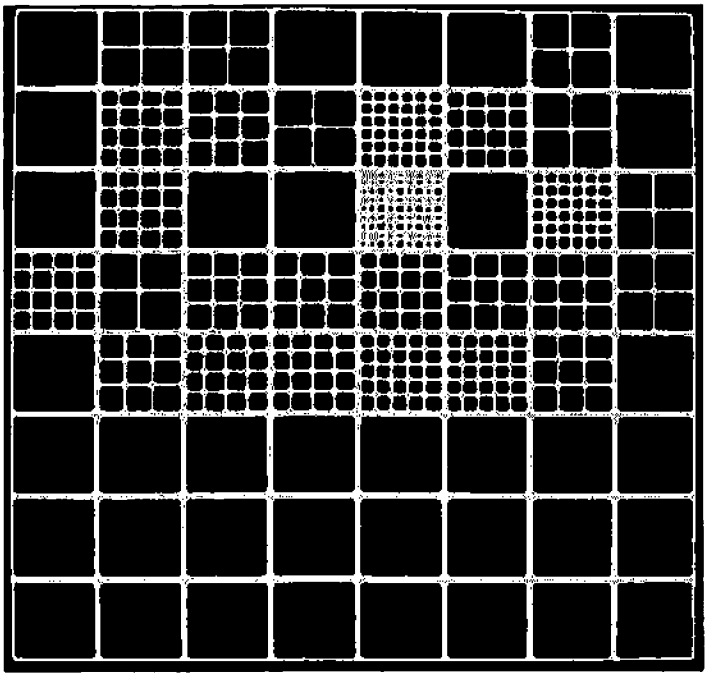


Figure 6.3 Original 128 x 128 Pixel APC Image

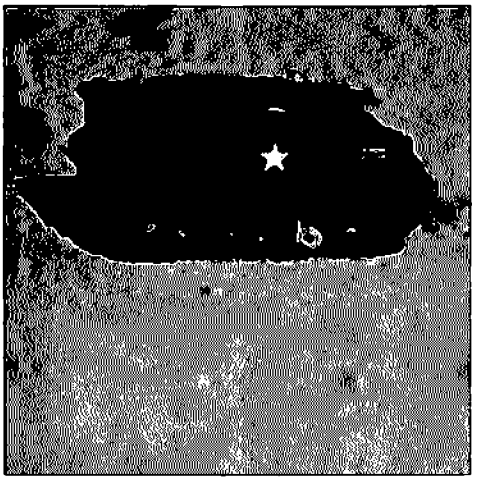


Bi-Cubic Spline Approximation
by Subpartitioning

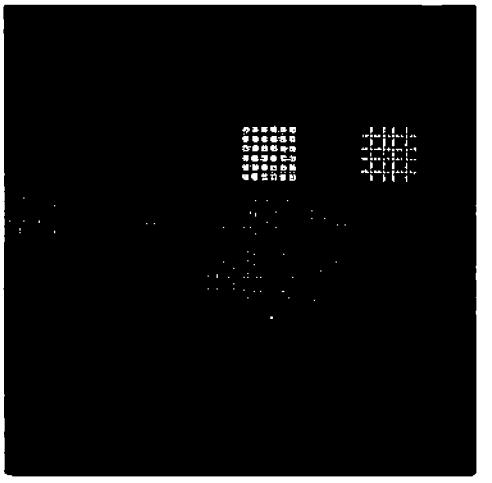


Associated Knot Density

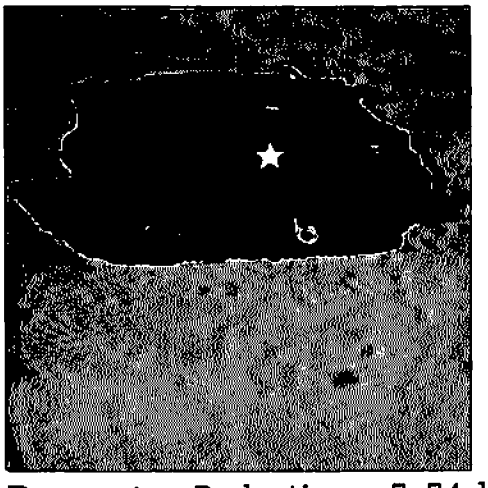
Figure 6.4 Bi-Cubic Spline Approximation
by Subsectioning



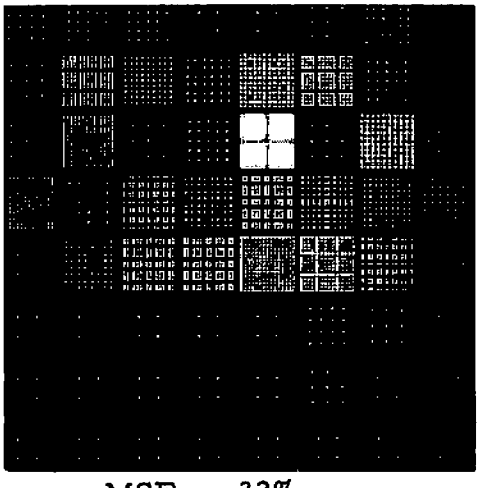
Parameter Reduction = 9.5:1



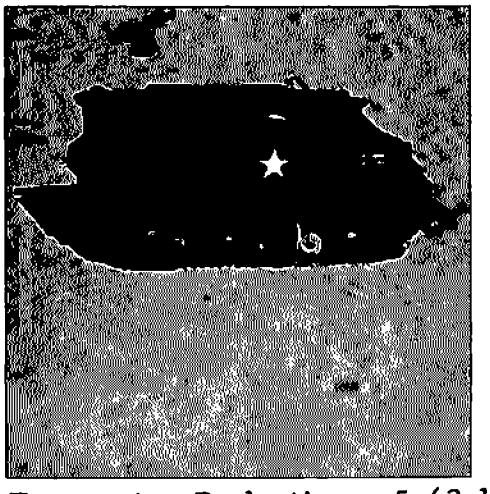
MSE = .38%



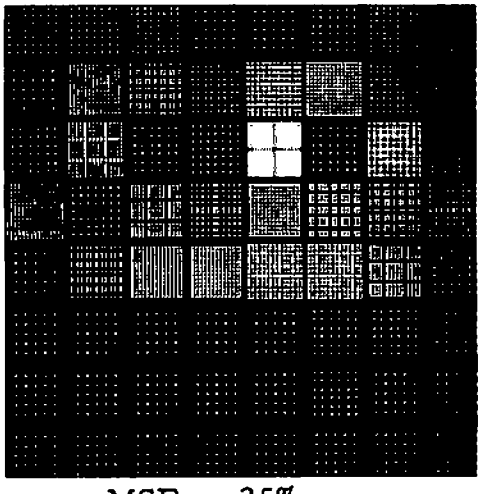
Parameter Reduction = 7.74:1



MSE = .32%

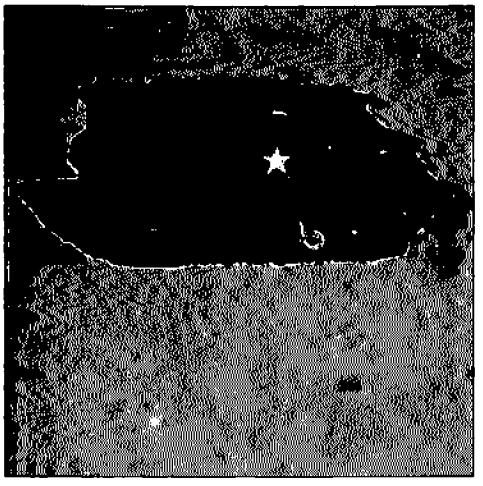


Parameter Reduction = 5.62:1

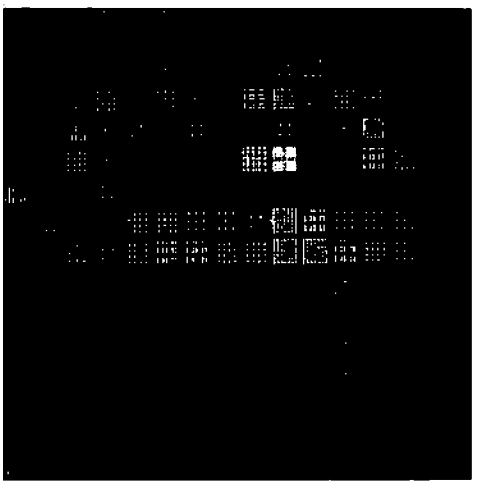


MSE = .25%

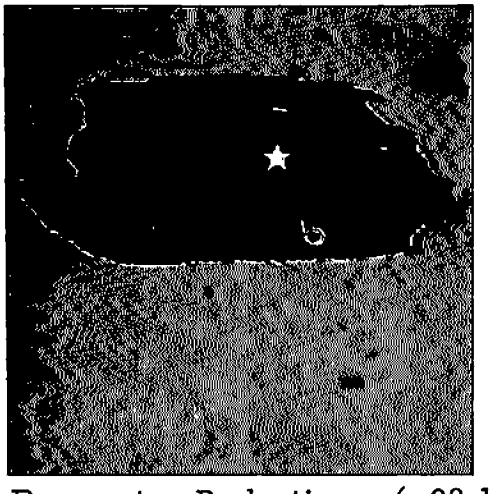
Figure 6.5 Bicubic Spline Reconstructions and Associated Knot Densities for an APC Photograph Using Subregions of Size 32 by 32.



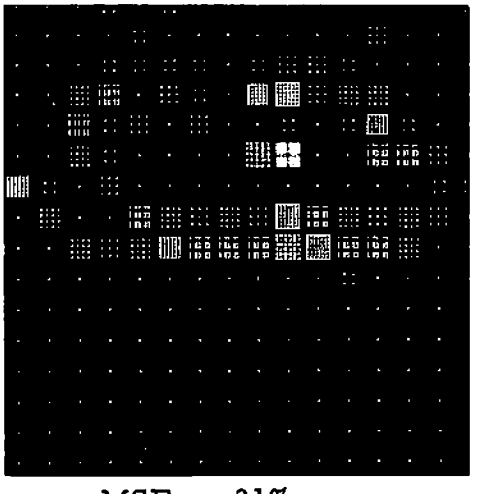
Parameter Reduction = 9.26:1



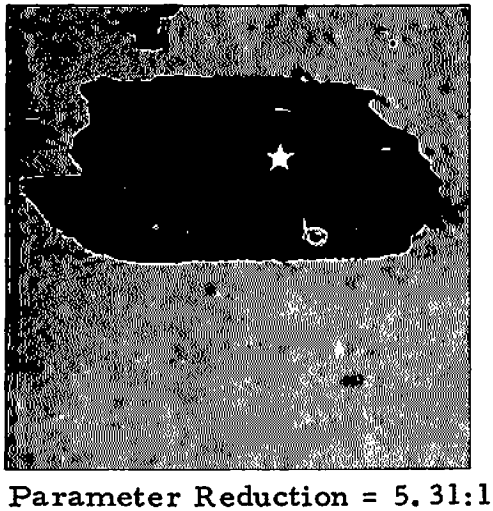
MSE = .37%



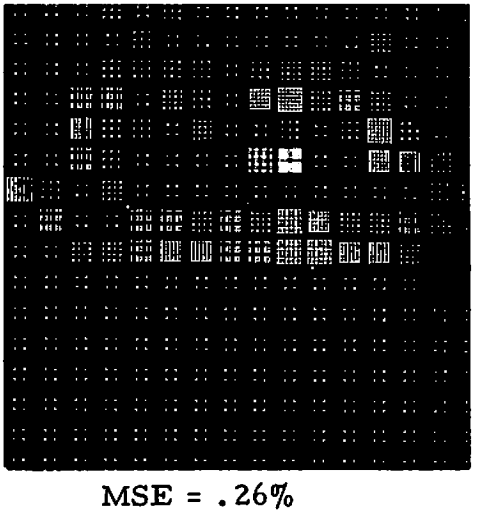
Parameter Reduction = 6.98:1



MSE = .31%

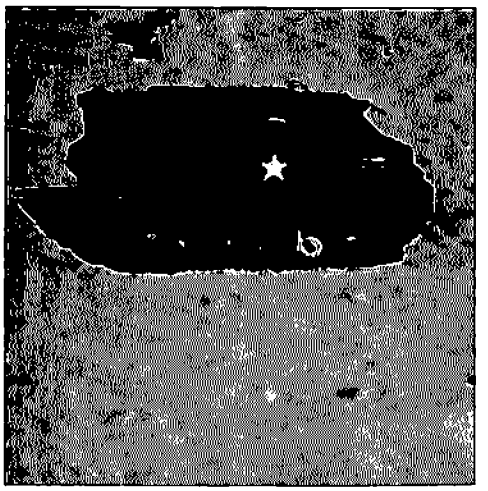


Parameter Reduction = 5.31:1

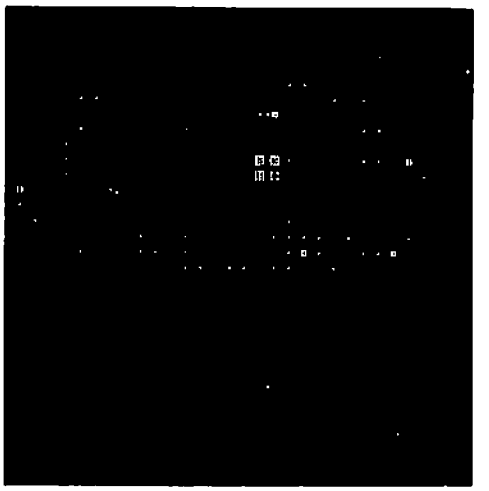


MSE = .26%

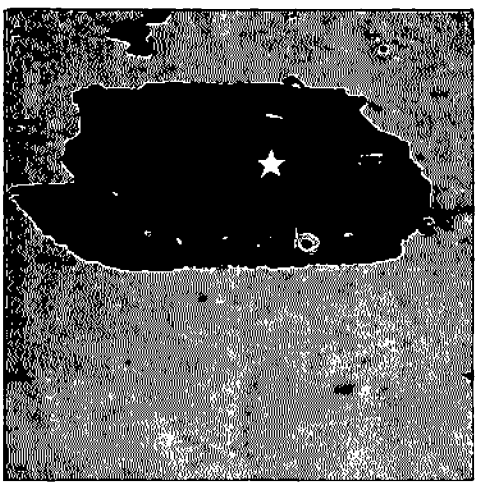
Figure 6.6 Bicubic Spline Reconstructions and Associated Knot Densities for an APC Photograph Using Subregions of Size 16 by 16.



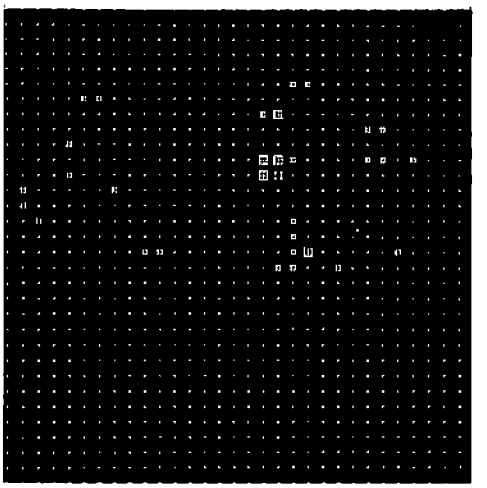
Parameter Reduction = 3.62:1



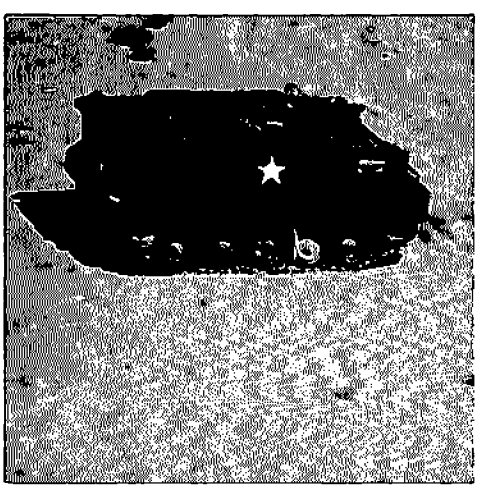
MSE = .24%



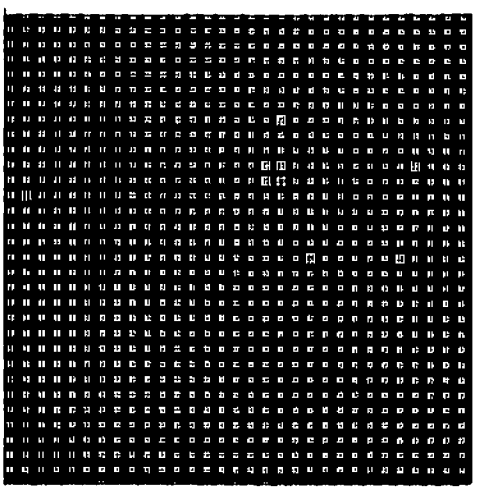
Parameter Reduction = 2.42:1



MSE = .17%



Parameter Reduction = 1.72:1



MSE = .1%

Figure 6.7 Bicubic Spline Reconstructions and Associated Knot Densities for an APC Photograph Using Subregions of Size 8 by 8.

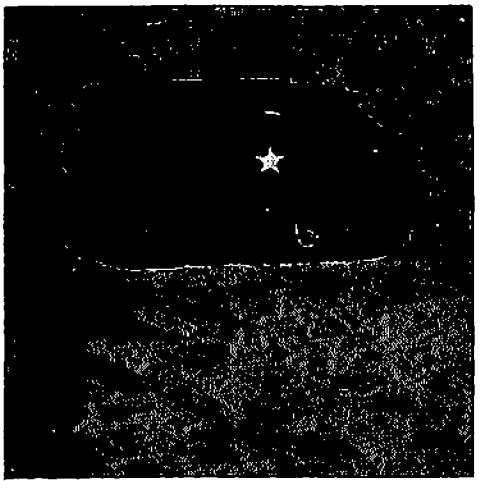
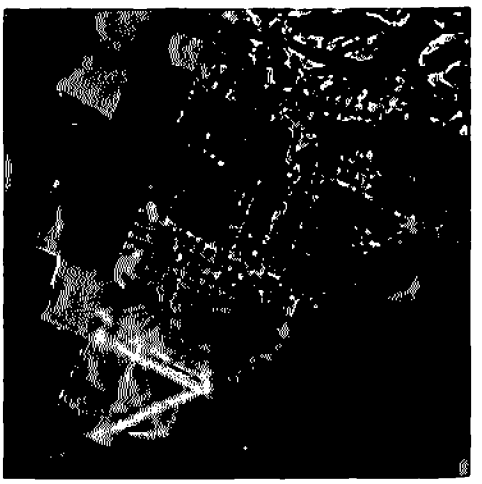
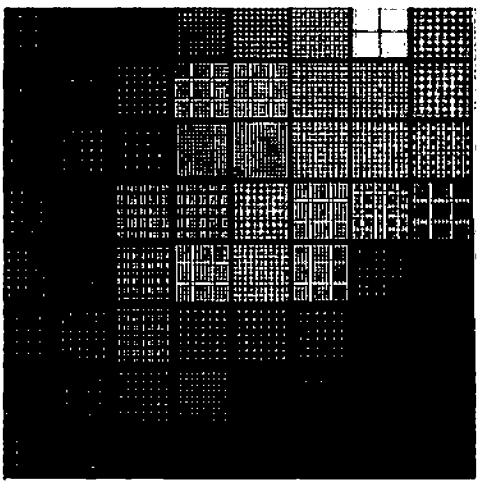


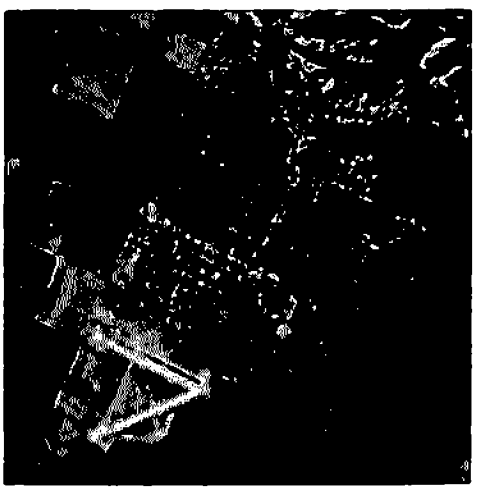
Figure 6.8 Original 256 by 256 APC Image.



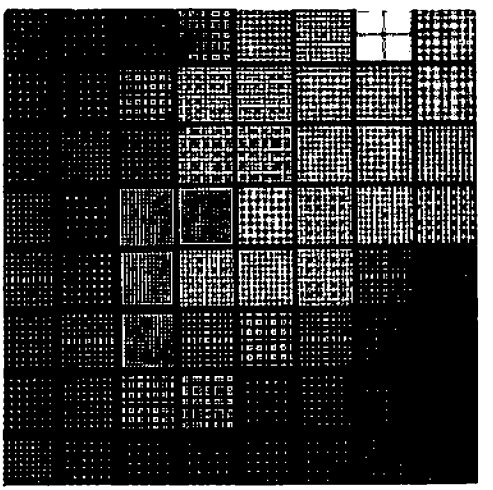
Parameter Reduction = 4.69:1



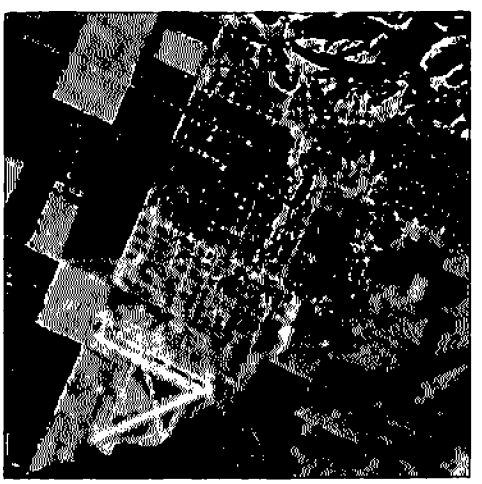
MSE = 1.4%



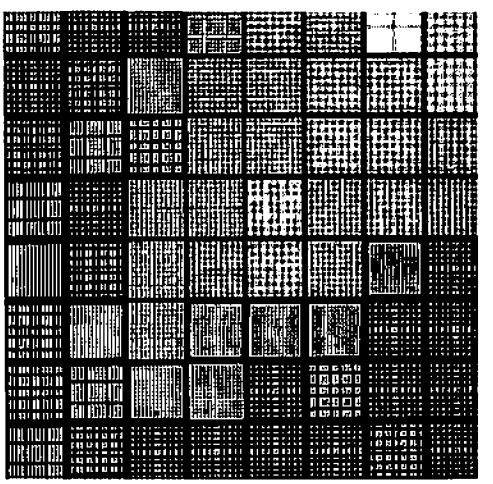
Parameter Reduction = 3.46:1



MSE = .63%

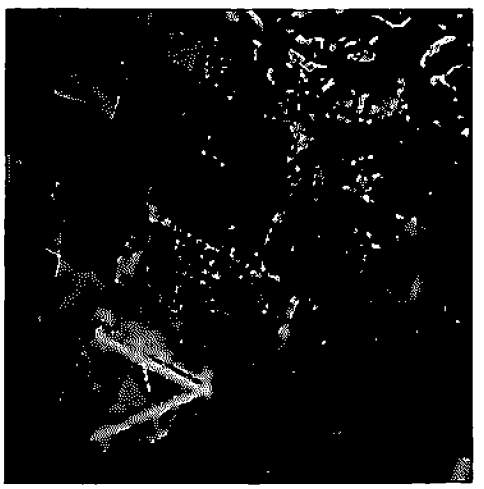


Parameter Reduction = 2.56:1

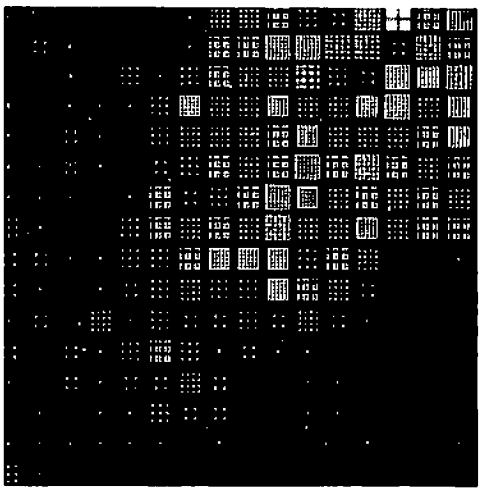


MSE = .38%

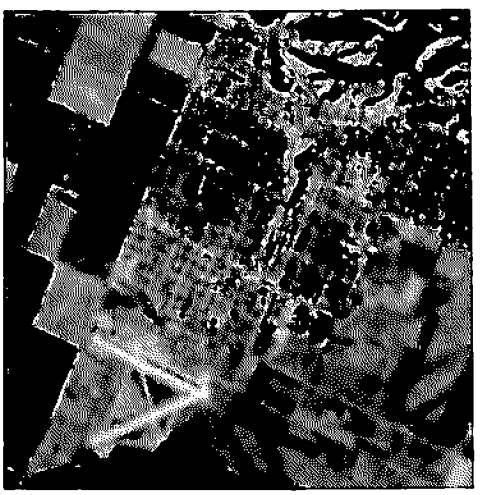
Figure 6.9 Bicubic Spline Reconstructions and Associated Knot Densities for a Reconnaissance Photograph Using Subregions of Size 32 by 32.



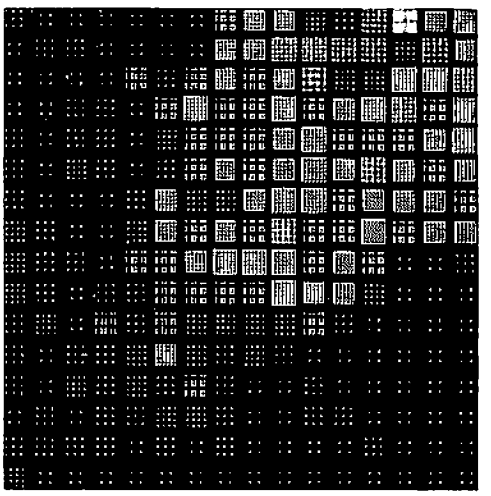
Parameter Reduction = 5.24:1



MSE = 1.2%

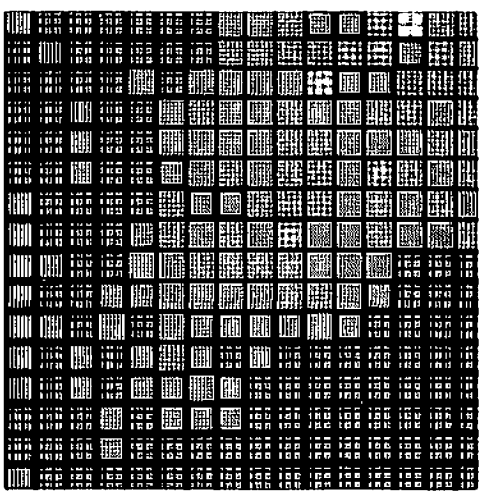


Parameter Reduction = 3.71:1



MSE = .70%

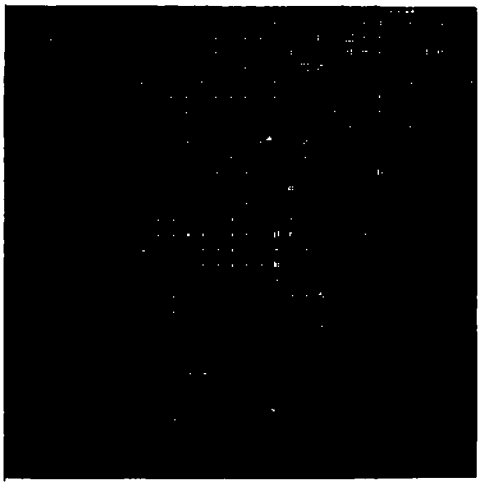
Parameter Reduction = 2.36:1



MSE = .39%

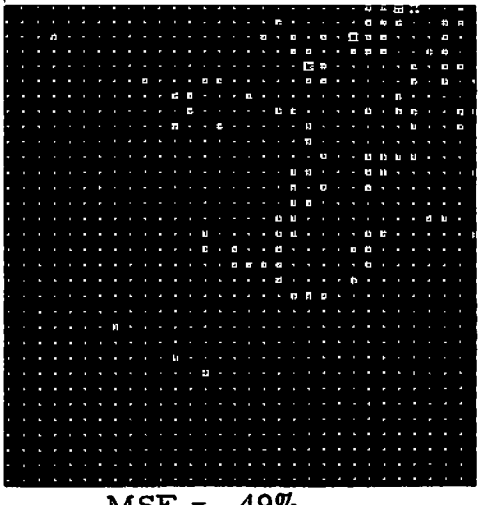
Figure 6.10 Bicubic Spline Reconstructions and Associated Knot Densities for a Reconnaissance Photograph Using Subregions of Size 16 by 16.

Parameter Reduction = 3.31:1

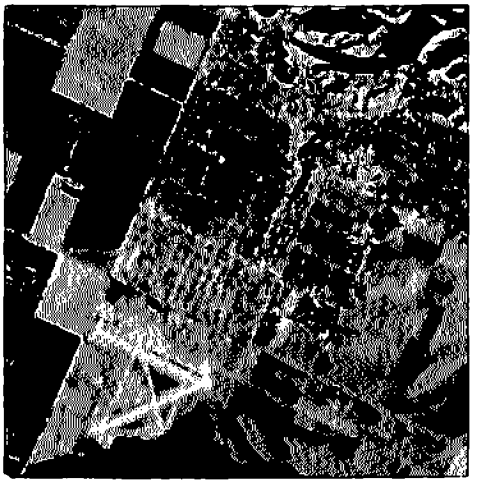


MSE = .77%

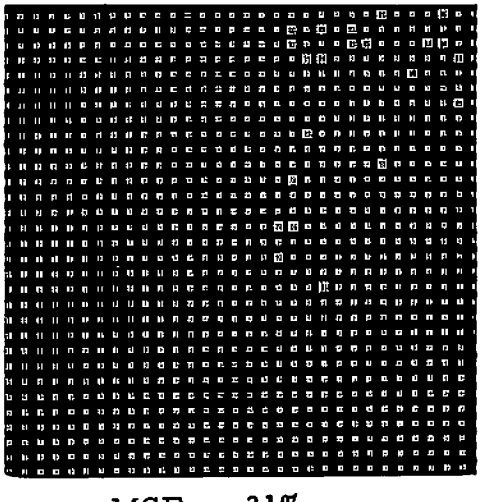
Parameter Reduction = 2.35:1



MSE = .49%



Parameter Reduction = 1.71:1



MSE = .31%

Figure 6.11 Bicubic Spline Reconstructions and Associated Knot Densities for a Reconnaissance Photograph Using Subregions of Size 8 by 8.

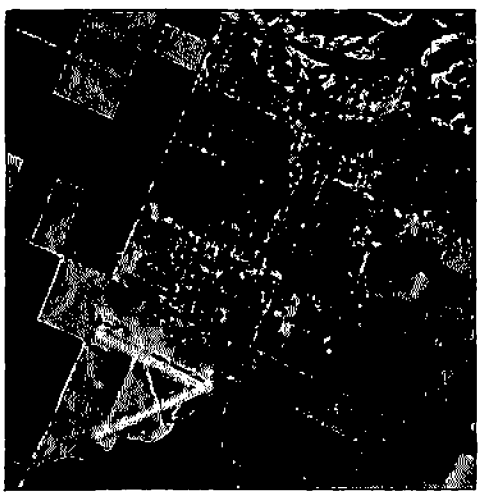
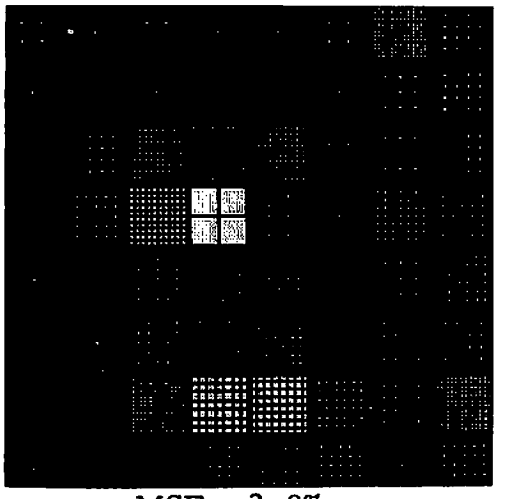
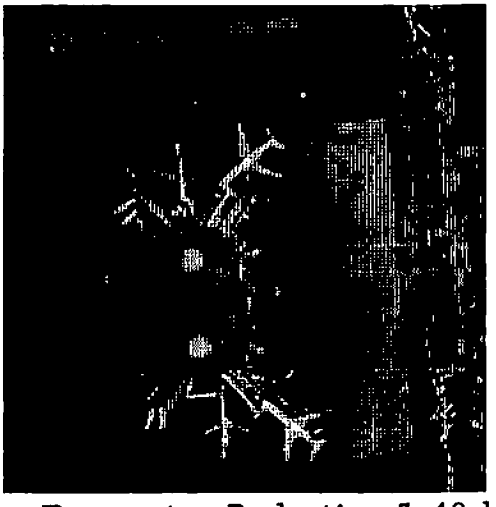


Figure 6.12 Original 256 by 256 Reconnaissance Image.

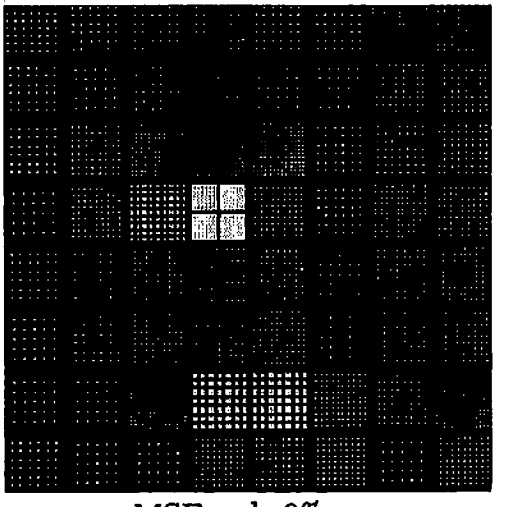
Parameter Reduction 9.19:1



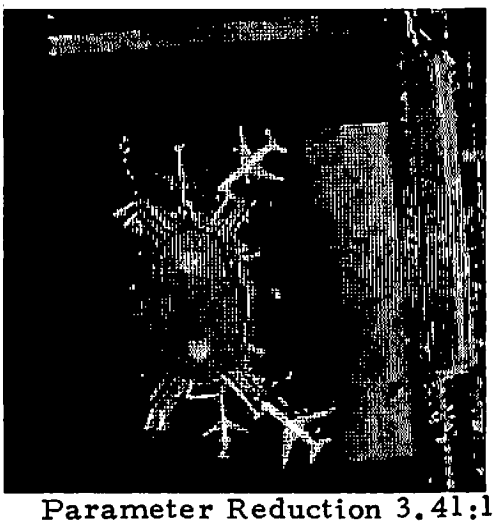
MSE = 2.8%



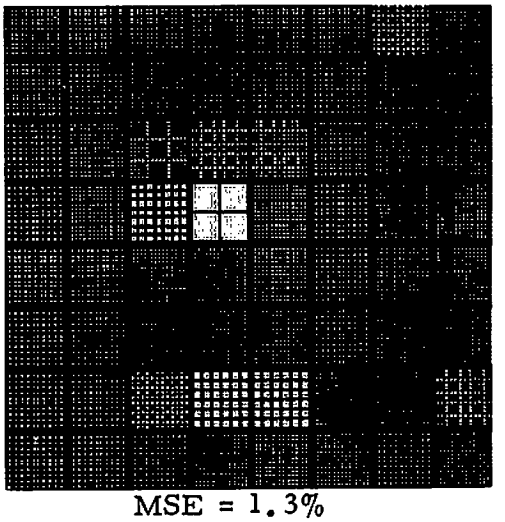
Parameter Reduction 5.48:1



MSE = 1.9%

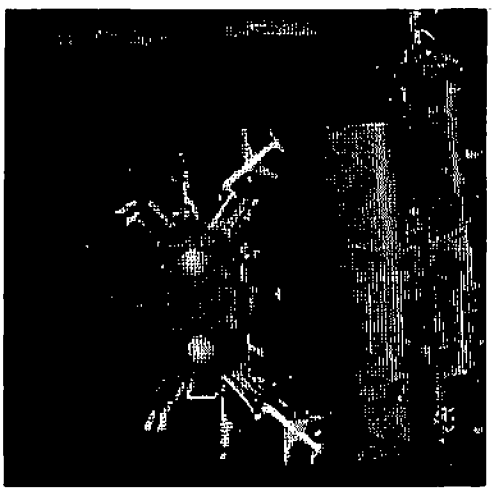


Parameter Reduction 3.41:1

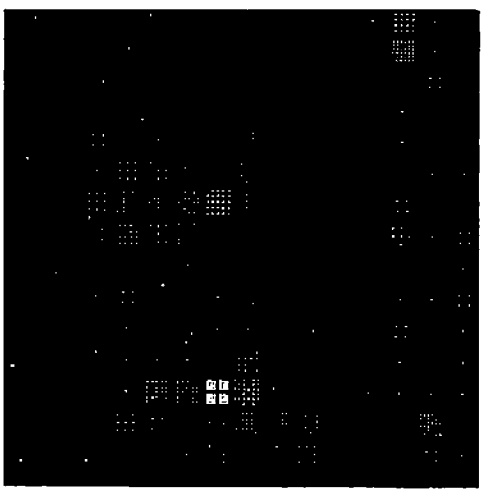


MSE = 1.3%

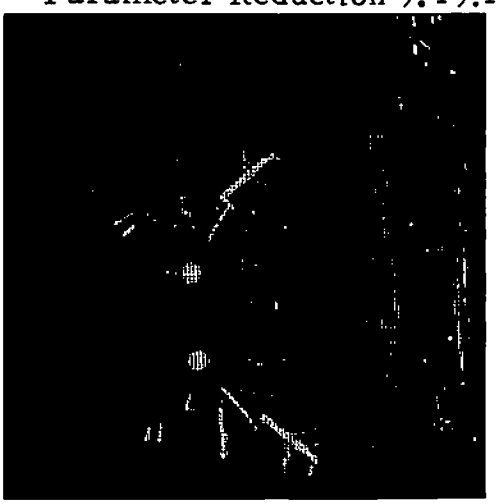
Figure 6.13. Bicubic Spline Reconstructions and Associated Knot Densities for LAX Photograph Using Subregions of Size 32 x 32.



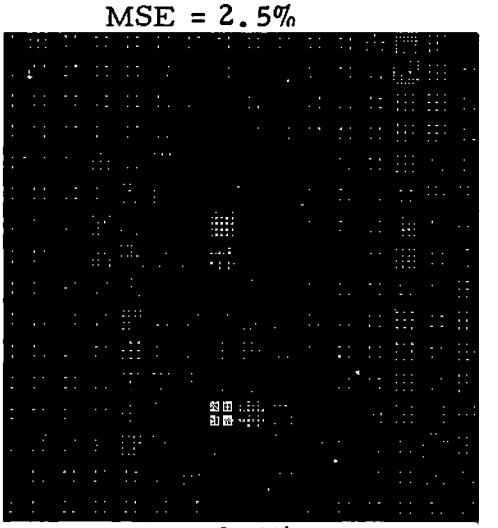
Parameter Reduction 9.19:1



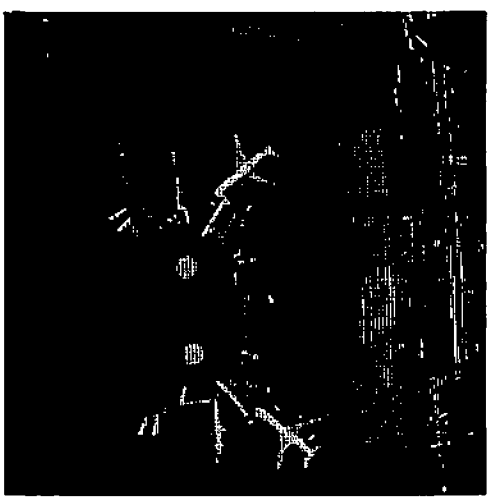
MSE = 2.5%



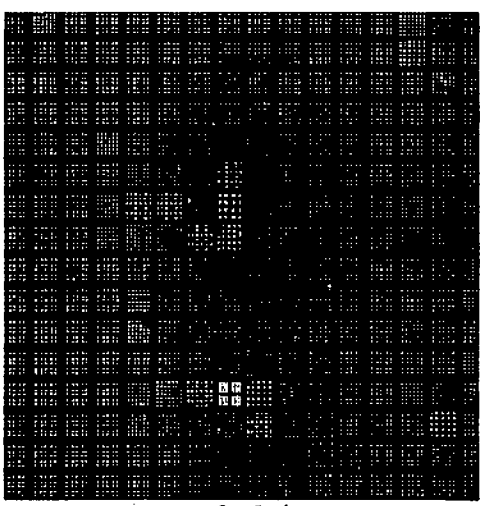
Parameter Reduction 5.35:1



MSE = 1.8%

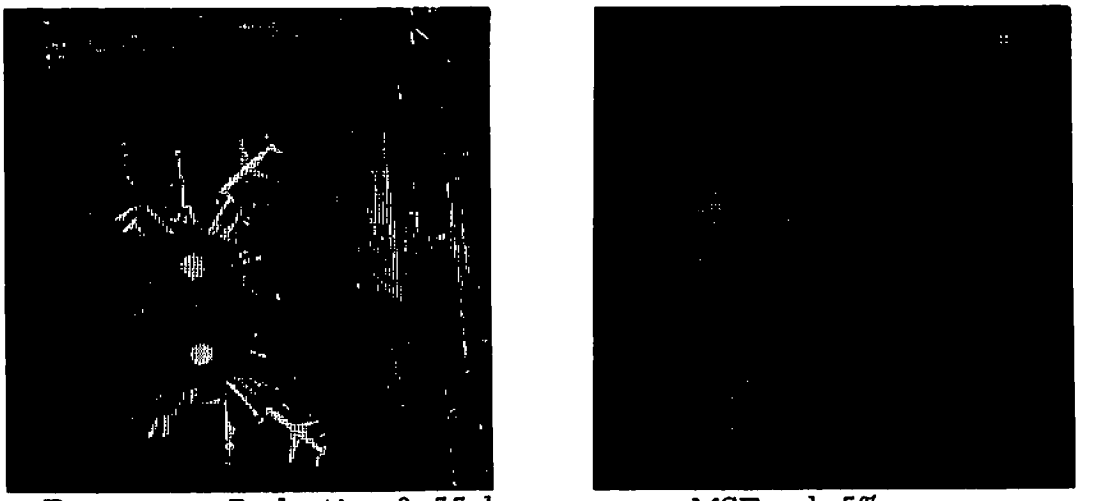


Parameter Reduction 2.82:1

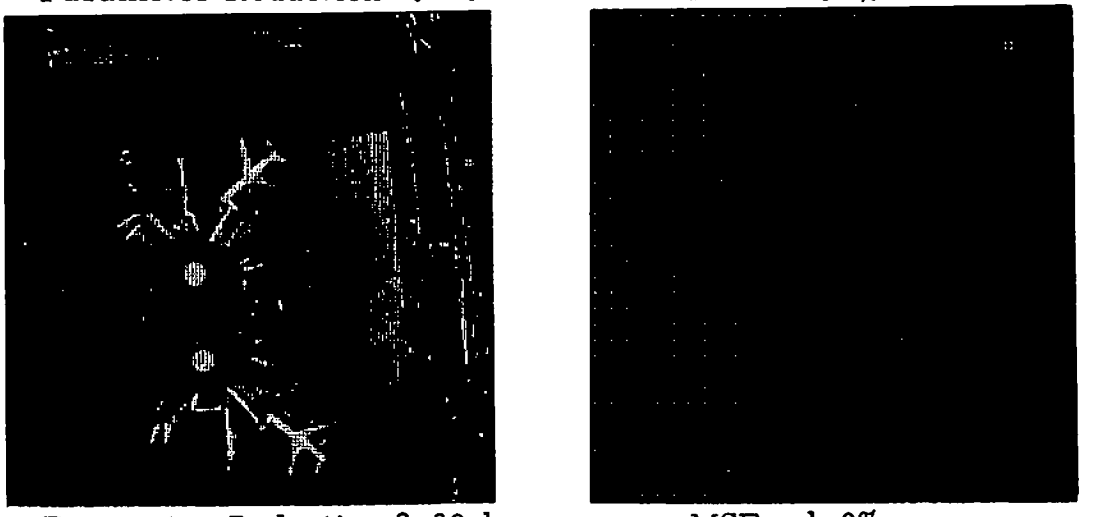


MSE = 1.1%

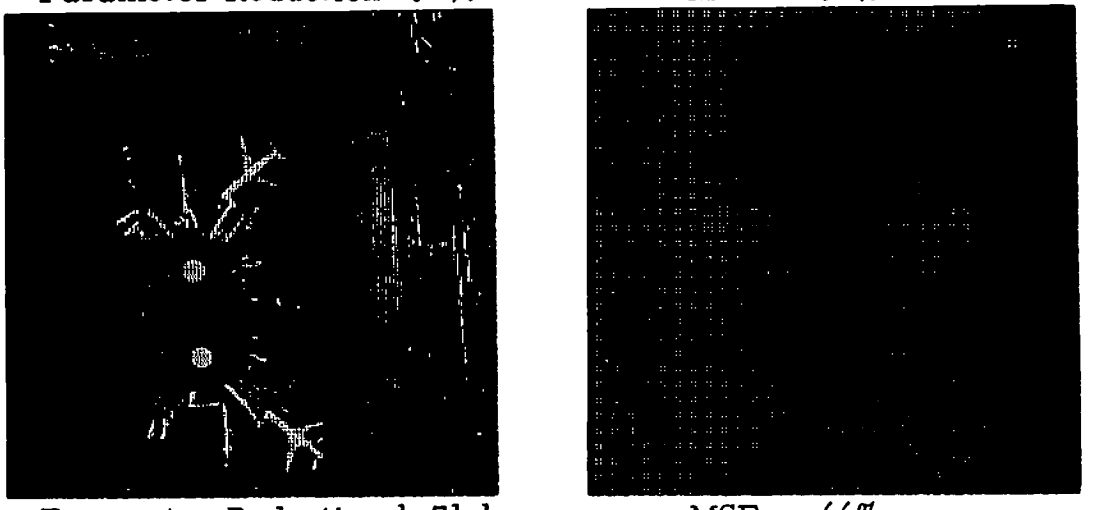
Figure 6.14. Bicubic Spline Reconstructions and Associated Knot Densities for LAX Photograph Using Subregions of Size 16 x 16.



Parameter Reduction 3.55:1 MSE = 1.5%



Parameter Reduction 2.39:1 MSE = 1.0%



Parameter Reduction 1.71:1 MSE = .66%

Figure 6.15. Bicubic Spline Reconstructions and Associated Knot Densities for LAX Photograph Using Subregions of Size 8 x 8.

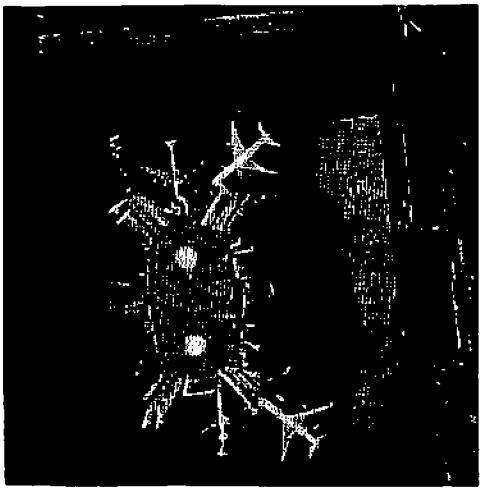


Figure 6.16. Original 256 x 256 LAX Image.

Chapter 7

SUMMARY, CONCLUSIONS AND FUTURE WORK

7.1 Summary and Conclusions

The dissertation has presented a degrees of freedom analysis of images and imaging systems in digital image processing. The analysis has been applied to image restoration when the image is viewed as the output of a linear imaging system, the problem being to restore the original image up to the degrees of freedom of the system, once this has been quantified; and to sampled images directly where the desire is to relate the sampled image to the original unsampled and unblurred version. In reality both situations are approximation problems since the restoration of an image through the continuous-discrete model involves the determination of an estimate for the original image in terms of a linear combination of the point spread function (PSF) kernels of the model, while in the latter case the researcher is free to find an approximation over a wide class of approximating functions. In modeling the restoration problem the continuous-discrete model was adopted since it most closely represents the imaging--data gathering--reconstruction sequence, eliminating any quadrature approximation to the superposition integral implicit in the linear system assumption. Further-

more, the eigenvalues of the gram matrix of the PSF kernels were shown to determine to what extent restoration of the image is possible.

The assumption that imaging systems are linear and representable by a two dimensional Fredholm integral equation of the first kind presents difficulties in the restoration process since restoration is no more than an attempt to numerically invert the associated integral equation. In this dissertation we have bounded the error between the eigenvalues of the system gram matrix and the singular values of the original continuous-continuous model kernel. It was shown for a wide class of quadrature rules with their corresponding sampling schemes that the eigenvalues of the gramian converge to the square of the singular values of the continuous-continuous kernel. Since this singular value sequence possesses zero as its only limit point, it is obvious as to why the image restoration process becomes ill-conditioned for large data ensembles.

More important is the fact that if the eigenvalue behavior of the original kernel is known, it is then possible to predict the number of effectively independent samples that can be obtained from an imaging system.

The continuous discrete model was developed for imaging systems with separable kernels and the associated gram matrix was

shown to be the direct product of two smaller gramians with a resultant computational reduction. Actual numerical bounds were obtained for the difference between the gramian eigenvalues and the continuous-continuous model singular-values for several quadrature rules.

These concepts were applied to the tomographic imaging system with excellent results. The continuous-discrete model was shown to possess a structure that made diagonalization of the tomographic gramian possible, and a linear algebraic solution feasible. The tomographic imaging system was shown to be a circularly symmetric imaging system that is bandlimited and observed over a finite radius in the output plane and whose degrees of freedom and eigenvalue behavior are known. The gramian eigenvalue behavior was exactly as expected--exhibiting a rather constant behavior out to a predicted point followed with a large drop. The reconstruction algorithm developed from the continuous-discrete model provided excellent reconstructions from real projection data.

In dealing with the image itself the degrees of freedom was approached as an approximation problem where the degrees of freedom at a level epsilon was taken to be the minimum number of functions needed to approximate the image with an error epsilon. Since this minimum is difficult to find, the functions used were cubic

splines with variable knots. By dividing the image into subregions a significant data reduction was achieved with reasonable errors. It was found that the number of knots and thus the degrees of freedom was higher in regions of higher image derivative energy than in those regions where the image was relatively constant. It was also found in one case that while the mean squared errors in this scheme were of the same order as obtained with a uniform knot spline approximation with a comparable number of parameters this method produced reconstructions with far more visible detail.

Finally it should be said that in effect this represents an attempt to bridge the gap between the continuous domain upon which images are defined and the discrete grids upon which they are sampled and defined for analysis by digital techniques.

7.2 Future Work

There are several topics concerning the spline approximations presented in this dissertation that deserve further study. This is not to say that there are no other interesting questions concerning the gram matrix, for its application to Fredholm integral equations of the second kind is certainly one example. Nevertheless the topics discussed here will mainly be concerned with spline approximations as an image coding device.

In this light we have seen that a sampled image can be approximated with a tolerable error by a bicubic spline with fewer

coefficients than pixels in the original image, and as such it represents a data reduction scheme. However this is a reduction in the number of computer words and not bits necessary to represent the image, and in order for these results to be applicable to digital communication systems they must be in terms of bit reductions. This would indicate that an efficient quantization method for the spline coefficients must be developed. The conjecture that the dynamic range of these coefficients might not be significantly greater than that of the original image makes this a particularly interesting topic for further study. Furthermore the normalized B-spline basis functions exhibit a local basis property which might transfer some of the image spatial correlation properties to the spline coefficients thus making a further de-correlation possible with some post-processing.

In this work the image was subsectioned into square regions only. In each of these regions the knot density was quite adaptive to the image derivative characteristics and it would be quite interesting to determine if this adaptivity could be employed in determining the subregion boundaries. If this is possible then variable knots splines could become a useful image segmentation tool.

APPENDIX A

SOME PROPERTIES OF NORMALIZED B-SPLINES

In this appendix the properties of the normalized B-splines of order k are discussed. The discussion will be limited to the one dimensional case as the extension to a direct product of splines for two dimensional approximations is immediate.

For the one dimensional case a spline $S_{k, N_x}(x)$ of order k with N_x knots approximating f(x) is given by

$$S_{k, N_x}(x) \triangleq \sum_{i=1}^{N_x} s_i N_{i, k}(\underline{\xi}; x) = \hat{f}(x) \quad x \in [0, 1]$$

where $N_{i, k}(\xi, x)$ are the normalized B-splines of order k satisfying the following recursion relationship [A-1] over the knot vector $\underline{\xi}$, where

$$\underline{\xi} = (\xi_1, \xi_2, \ldots, \xi_{N_x}, \ldots, \xi_{N_x+k})$$

$$\xi_k \leq x \leq \xi_{N_x}$$

$$N_{i, k}(\underline{\xi}; x) = \begin{cases} \dfrac{x-\xi_i}{\xi_{i+k-1}-\xi_i} N_{i,k-1}(\underline{\xi}; x) + \dfrac{\xi_{i+k}-x}{\xi_{i+k} - \xi_{i+1}} N_{i+1, k-1}(\xi; x) & \\ 0 \qquad \text{otherwise} & x \in [\xi_i, \xi_{i+k}] \end{cases} \tag{A-1}$$

$$\underline{a} = [T]\underline{S}$$

where $[T]$ is as follows

$$[T] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -3 & 3 & 0 & 0 \\ 3 & -6 & 3 & 0 \\ -1 & 3 & -3 & 1 \end{bmatrix}$$

With these properties we can now demonstrate an interesting property of k^{th} order least squares splines. A k^{th} order least squares spline with N_x knots for $f(x_i)$ $i = 1, 2, \ldots, N (N \geq N_x)$ is given by the solution to the following normal equations:

$$\underset{N_x \times 1}{\underline{S}} = \left[[N]^{(\xi)^T} [N]^{(\xi)} \right]^{-1}_{N_x \times N_y} [N]^{(\xi)}_{N_x \times N} \underset{N \times 1}{\underline{f}} \qquad \text{(A-2)}$$

where $\underline{f}_{N \times 1} = [f(x_1), \ldots, f(x_N)]^T$

$$[N]^{(\xi)} = [N_{i,k}(\underline{\xi}; x_j)]_{N_x \times N}$$

If we take $N_x = k$ and the knot vector $\underline{\xi}$ to consist of two k^{th} order knots at 0 and 1 then this is equivalent to determining a k-1 degree least squares polynomial for $f(x_i)$ $i = 1, \ldots, N$. Alternately if we take the knot vector $\underline{\xi}$ to contain the two k^{th} order knots at 0 and 1 and a sufficient number of internal knots so that $N_x = N$ then the matrix $[N]^{(\xi)}_{N_x \times N_x}$ will be square and nonsingular so that eq. (A-2) becomes

$$N_{i,1}(\xi;x) = \begin{cases} 1 & x \in [\xi_i, \xi_{i+1}) \\ 0 & \text{otherwise} \end{cases}$$

and

$$\sum_{i=1}^{N_x} N_{i,k}(\underline{\xi};x) = 1$$

Note that eq. (A-1) indicates that N_x+k knots are required to generate N_x normalized B-splines of order k, and that $N_{i,k}(\underline{\xi};x)$ is nonzero only over the interval $[\xi_i, \xi_{i+k})$. Also a knot may have multiplicity p, up to k in which the multiplicity indicates a discontinuity in the (k-(p+1)) derivative at that knot. If we follow Rice [A-2] and adopt the convention that the spline is differentiable of order 0 or -1 at the knot ξ_0 if the spline is continuous or has a simple jump at ξ_0 respectively. a fourth order order knot at ξ_0 for a cubic or 4^{th} order spline indicates a simple jump in the spline at ξ. Figures A.1 and A.2 illustrate the normalized B-splines of order 4 for the knot vectors

$$\underline{\xi}_1 = [0, 0, 0, 0, .25, .5, .75, 1, 1, 1, 1]^T$$

and

$$\underline{\xi}_2 = [0, 0, 0, 0, .7, .8, .9, 1, 1, 1, 1]^T$$

respectively. $\underline{\xi}_1$ corresponds to the uniform knot case while $\underline{\xi}_2$ illustrates the effect of shifting the internal knots towards 1.

Table A.1 lists the knots over which $N_{i,4}(\underline{\xi};x)$ is non-zero for i =

1, 2, . . . , 7. Note also that there are a total of 11 knots to define these 7 nonzero normalized B-splines.

Table A.1 and figure A.1 serve to illustrate the relationship between multiple knots and the differentiability of the normalized B-splines. Note that $N_{1,4}(\underline{\xi}_1;x)$ involves a fourth order knot at x =0 so that k-(p+1) is equal to -1 and $N_{1,4}(\underline{\xi}_1;x)$ possesses a simple jump at x = 0. The third order knot at x = 0 for $N_{2,4}(\underline{\xi}_1;x)$ results in k-(p+1) equaling 0 and from figure A.1 it is clear that $N_{2,4}(\underline{\xi}_1;x)$ is merely continuous at x = 0. The second and first order knots at x = 0 for $N_{3,4}(\underline{\xi}_1;x)$ and $N_{4,4}(\underline{\xi}_1;x)$ respectively result in $N_{3,4}(\underline{\xi}_1;x)$ being once continuously differentiable and $N_{4,4}(\underline{\xi}_1;x)$ being twice continuously differentiable at x=0. The same sequence of events is true for $N_{5,4}(\underline{\xi}_1;x)$, $N_{6,4}(\underline{\xi}_1;x)$ and $N_{7,4}(\underline{\xi}_1;x)$ at x = 1.

Another interesting property of the k^{th} order normalized B-splines arises in the case when the knot vector consists of solely two k^{th} order knots at x = 0 and x = 1. In this case the k^{th} order spline with N_x = k knots can be shown to be equivalent to a k-1 degree polynomial. For k = 4 the four nonzero normalized B-splines are listed in Table A.2 for the knot vector $\underline{\xi}$ = (0, 0, 0, 0, 1, 1, 1, 1). In this case if $S_{4,4}(x) = \sum_{i=1}^{4} S_i N_{i,4}(\underline{\xi};x)$ is the fourth order spline with four knots and $\sum_{k=1}^{4} a_i x^{i-1}$ is a cubic polynomial it is simple to show that $\underline{a} = (a_1, a_2, a_3, a_4)^T$ and $S = (S_1, S_2, S_3, S_4)^T$ are related through the matrix $[T]$ by

$$\underline{S} = [N]^{(\xi)^{-1}} \underline{f}$$

corresponding to interpolation. Thus by varying the number of internal knots from 0 to N-k so that $N_x = N-k+k = N$, we can determine a least squares spline that goes from a least squares polynomial of degree k-1 to an interpolating spline of order k.

In summary this appendix has been presented to illustrate some of the computational properties of the normalized B-splines. This has been done in the hope that the interested reader might be better able to appreciate them from a computational standpoint and to more effectively apply them to approximation problems. It is also felt that these facts coupled with the recursion formula of eq. (A-1) should give the reader enough information to implement his own spline subroutines with variable knots without an excessive amount of difficulty.

	$N_{1,4}(\underline{\xi};x)$	$N_{2,4}(\underline{\xi};x)$	$N_{3,4}(\underline{\xi};x)$	$N_{4,4}(\underline{\xi};x)$	$N_{5,4}(\underline{\xi};x)$	$N_{6,4}(\underline{\xi};x)$	$N_{7,4}(\underline{\xi};x)$
$\underline{\xi}=\underline{\xi}_1=[0,0,0,0,.25,.5,.75,1,1,1,1]^T$	$\begin{bmatrix}0\\0\\0\\0\\.25\end{bmatrix}$	$\begin{bmatrix}0\\0\\0\\.25\\.5\end{bmatrix}$	$\begin{bmatrix}0\\0\\.25\\.5\\.25\end{bmatrix}$	$\begin{bmatrix}0\\.25\\.5\\.75\\1\end{bmatrix}$	$\begin{bmatrix}.25\\.5\\.75\\1\\1\end{bmatrix}$	$\begin{bmatrix}.5\\.75\\1\\1\\1\end{bmatrix}$	$\begin{bmatrix}.75\\1\\1\\1\\1\end{bmatrix}$
$\underline{\xi}=\underline{\xi}_2=[0,0,0,0,.7,.8,.9,1,1,1,1]^T$	$\begin{bmatrix}0\\0\\0\\0\\.7\end{bmatrix}$	$\begin{bmatrix}0\\0\\0\\.7\\.8\end{bmatrix}$	$\begin{bmatrix}0\\0\\.7\\.8\\.9\end{bmatrix}$	$\begin{bmatrix}0\\.7\\.8\\.9\\1\end{bmatrix}$	$\begin{bmatrix}.7\\.8\\.9\\1\\1\end{bmatrix}$	$\begin{bmatrix}.8\\.9\\1\\1\\1\end{bmatrix}$	$\begin{bmatrix}.9\\1\\1\\1\\1\end{bmatrix}$

Table A.1. Knot Vectors Over Which $N_{i,4}(\underline{\xi};x)$ is Nonzero.

NORMALIZED B-SPLINE	FUNCTIONAL FORM
$N_{1,4}(\underline{\xi};x)$	$(1-t)^3$
$N_{2,4}(\underline{\xi};x)$	$3t(1-t)^2$
$N_{3,4}(\underline{\xi};x)$	$3t^2(1-t)$
$N_{4,4}(\underline{\xi};x)$	t^3

Table A.2. Normalized B-Splines for $\underline{\xi} = (0, 0, 0, 0, 1, 1, 1, 1)^T$.

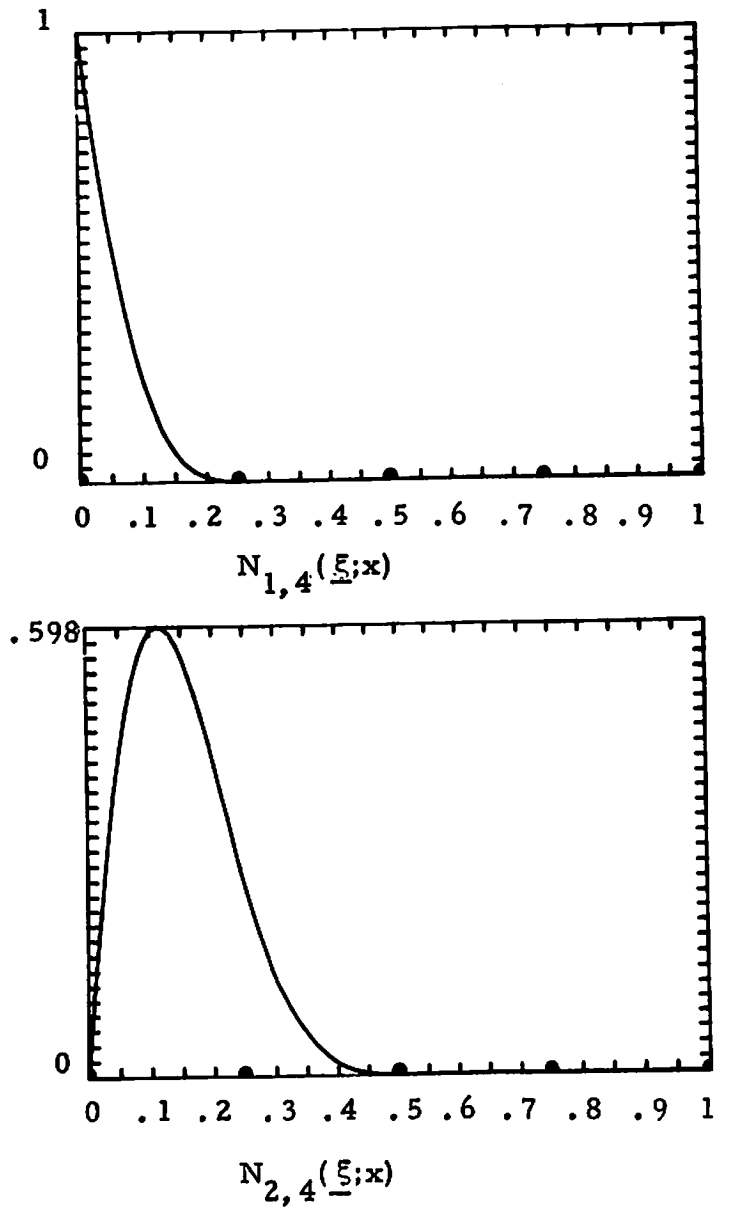
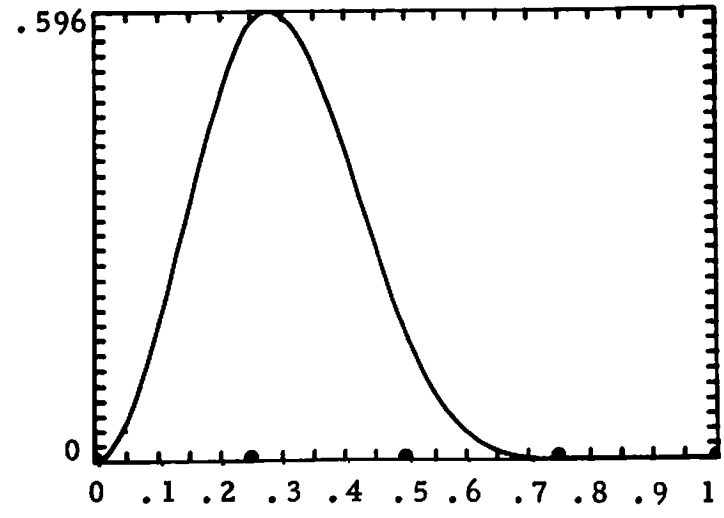


Figure A.1 Normalized 4th Order B-Splines For Knot Vector $\underline{\xi}$ = (0, 0, 0, 0, .25, .5, .75, 1, 1, 1, 1)



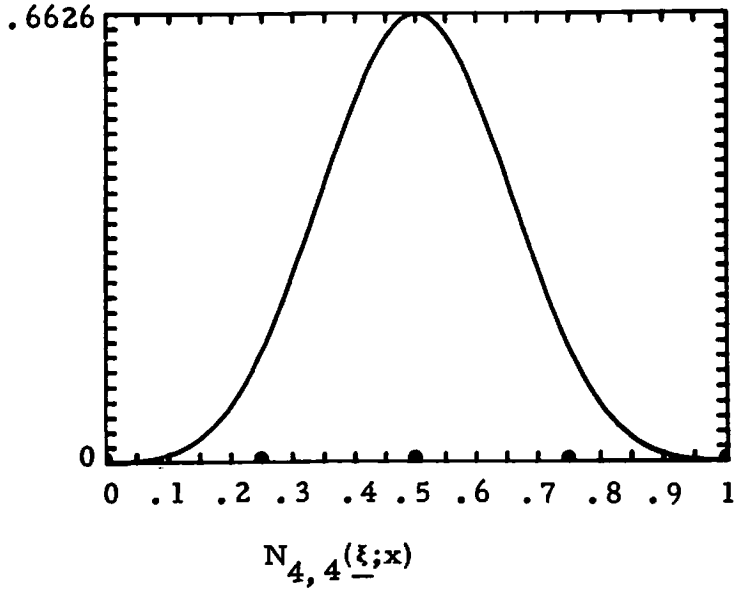
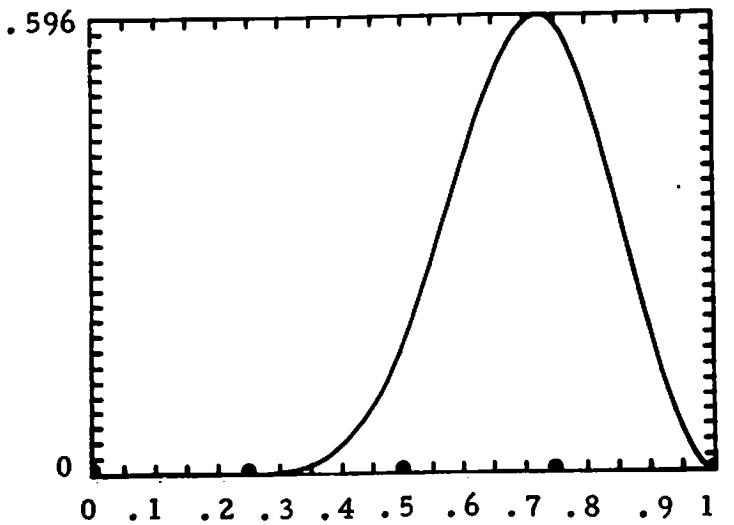


Figure A. 1 (continued) Normalized 4th Order B-Splines For Knot Vector $\underline{\xi}$=(0, 0, 0, 0, .25, .5, .75, 1, 1, 1, 1)



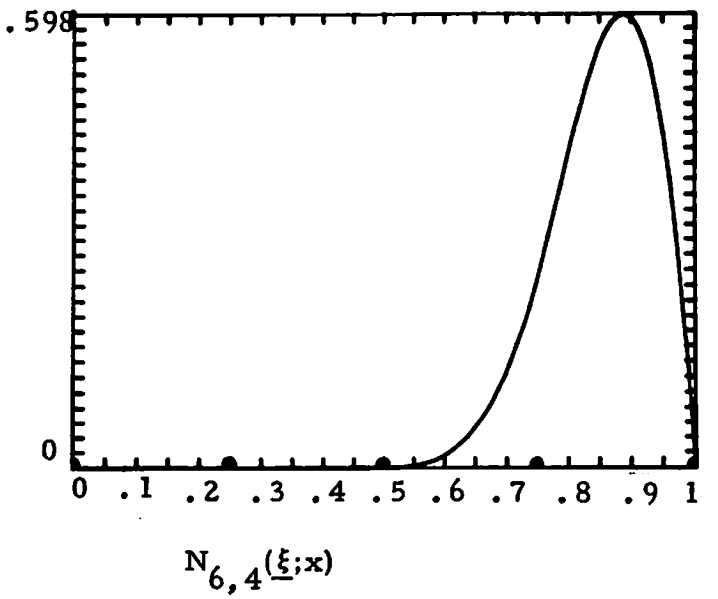


Figure A.1 (continued) Normalized 4th Order B-Splines For Knot Vector $\xi = (0, 0, 0, 0, .25, .5, .75, 1, 1, 1, 1)$

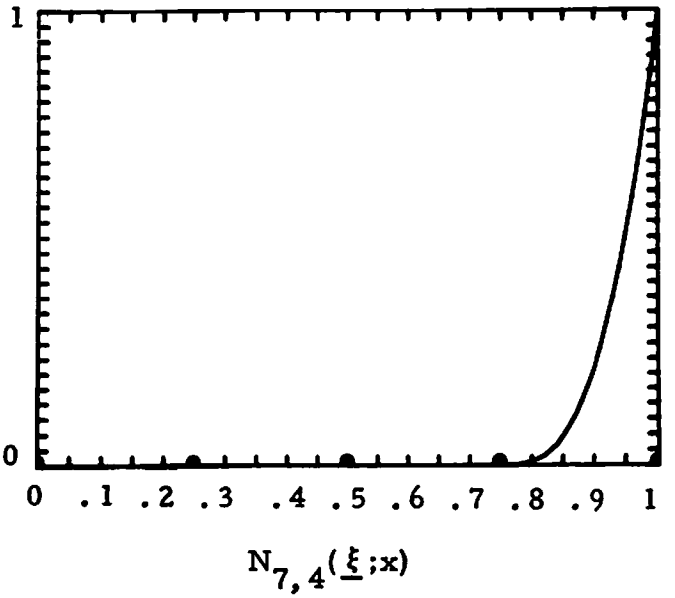


Figure A. 1 (continued) Normalized 4th Order B-Splines For Knot Vector = (0, 0, 0, 0, .25, .5, .75, 1, 1, 1, 1)

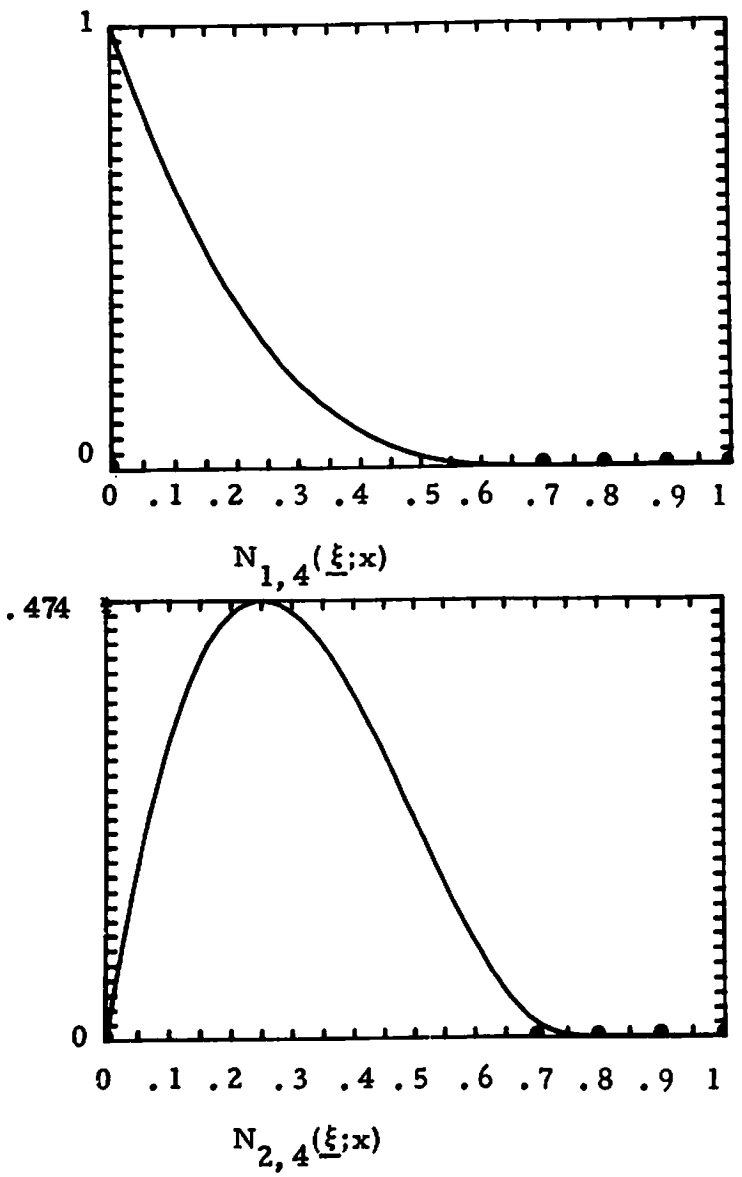
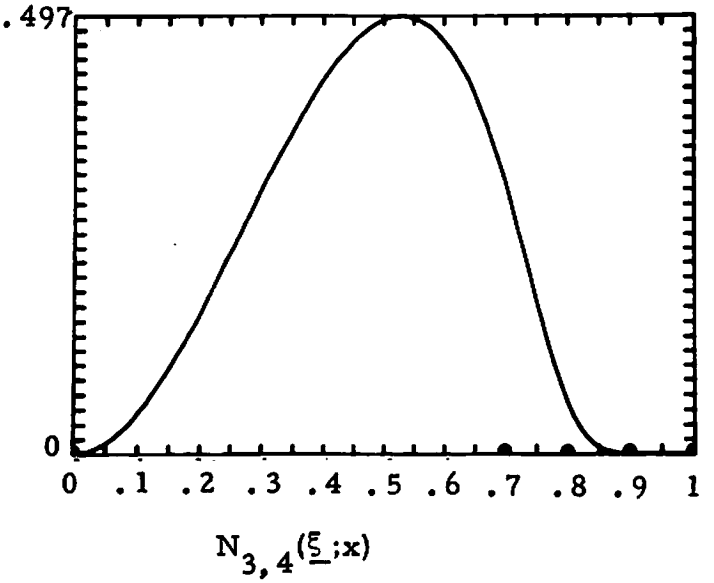


Figure A.2 Normalized 4th Order B-Splines For Knot Vector $\underline{\xi}$ = (0, 0, 0, 0, .7, .8, .9, 1, 1, 1, 1)



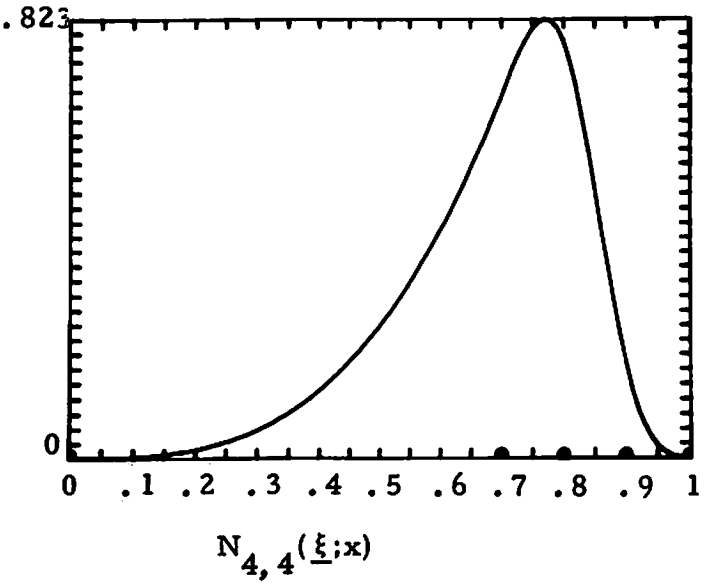
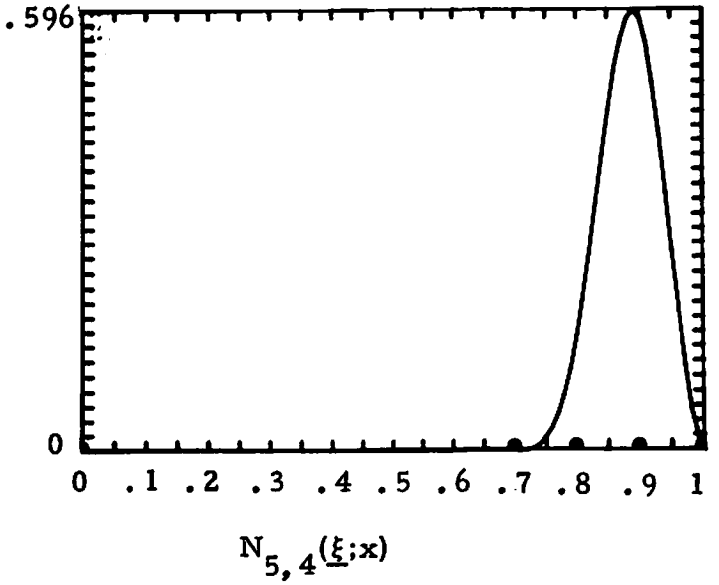


Figure A.2 (continued) Normalized 4th Order B-Splines
For Knot Vector $\underline{\xi} = (0,0,0,0,.7,.8,.9,1,1,1,1)$



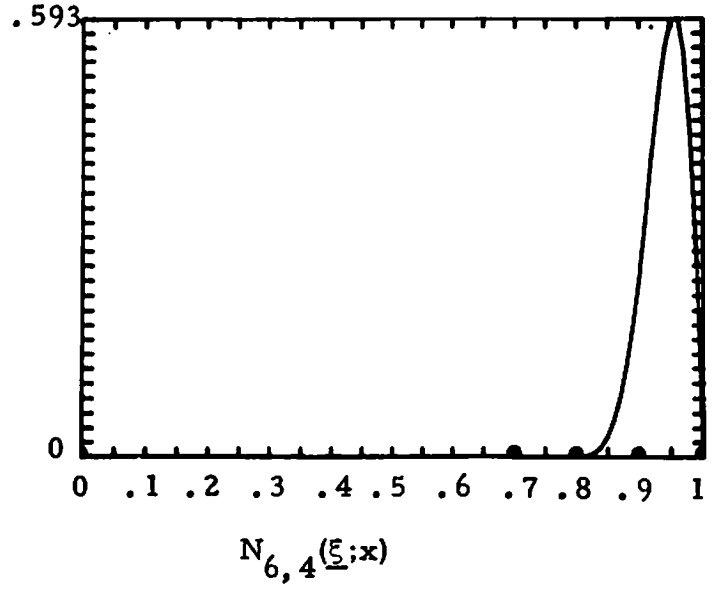


Figure A. 2 (continued) Normalized 4th Order B-Splines For Knot Vector $\underline{\xi}$ = (0, 0, 0, 0, . 7, . 8, . 9, 1, 1, 1, 1)

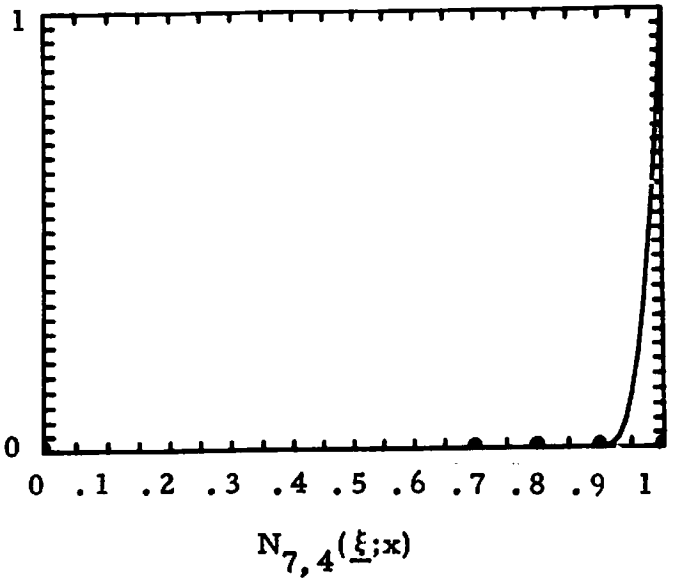


Figure A. 2 (continued) Normalized 4th Order B-Splines
For Knot Vector $\underline{\xi}$ = (0, 0, 0, 0, .7, .8, .9, 1, 1, 1, 1)

REFERENCES

[1-1] Andrews, H. C. and C. L. Patterson, "Singular Value Decompositions and Digital Image Processing," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. ASSP-24, no. 1, pp. 26-53, (February 1976).

[1-2] Hou, H. S., "Least Squares Image Restoration Using Spline Interpolation," Ph.D. dissertation, USCIPI Report No. 650, University of Southern California, Los Angeles, California (March 1976).

[1-3] Hounsfeld, G. N., "Computerized Transverse Axial Scanning (Tomography): Part I. Description of System," Brit. J. Radiol., vol. 46, pp. 1016-1022 (1973).

[2-1] Twomey, S., "The Application of Numerical Filtering to the Solution of Integral Equations Encountered in Indirect Sensing Measurements," Journal of the Franklin Institute, vol. 279, pp. 95-109 (1965).

[2-2] Twomey, S., "Information Content in Remote Sensing," Applied Optics, vol. 13, no. 4, pp. 942-945 (1976).

[2-3] Hou, H. S., and H. C. Andrews, "Fundamental Limits and Degrees of Freedom of Imaging Systems," Proc. of the OSA Topical Meeting on Image Processing, Asilomar, California (February 1976).

[2-4] Crowther, R. A., DeRosier, D. J., and A. Klug, "The Reconstruction of a Three-Dimensional Structure from Projections and its Application to Electron Microscopy," Proc. Roy. Soc. London, vol. A 317, pp. 319-340 (1970).

[2-5] Ramachandran, G. N. and A. V. Lakshminarayanan, "Three Dimensional Reconstruction from Radiographs and Electron Micrographs: Application of Convolutions Instead of Fourier Transforms," Proc. Nat. Acad. Sci., USA, vol. 68, no. 9, pp. 2236-2240 (1971).

[2-6] Bracewell, R. N., "Strip Integration in Radio Astronomy," Australian J. Phys., vol. 9, p. 198 (1956).

[2-7] Bracewell, R. N. and A. C. Riddle, "Inversion of Fan Beam Scans in Radio Astronomy," The Astrophysical Journal, vol. 150, pp. 427-434 (1967).

[2-8] Hounsfeld,G. N., "Computerized Transverse Axial Scanning (Tomography): Part I. Description of System," Brit. J. Radiol., vol. 46, pp. 1016-1022 (1973).

[2-9] Gordon, R., Bender, R., and G. T. Herman, "Algebraic Reconstruction Techniques (ART) for Three Dimensional Electron Microscopy and X-Ray Photography," J. Theor. Biol., vol. 29, pp. 471-481 (1970).

[2-10] Kashyap, R. L., and M. C. Mittal, "Picture Reconstruction from Projections," Proc. First Int'l Conf. on Pattern Recognition, Washington, D. C. (1973).

[2-11] Klug, A. and Crowther, R. A., "Three Dimensional Image Reconstruction from the Viewpoint of Information Theory," Nature, vol. 238, pp. 435-440 (1972).

[2-12] Smith, P. R., Peters, T. M., and Bates, R. H. T., "Image Reconstruction from Finite Numbers of Projections," J. Phys. A: Math., Nucl. Gen., vol. 6, pp. 361-381 (1973).

[2-13] Andrews, H. C., Computer Techniques in Image Processing, Academic Press, New York (1970).

[2-14] Golub, G. and C. Reinsch, "Singular Value Decomposition and Least Squares Solutions," Numer. Math., vol. 14, pp. 403-420 (1970).

[2-15] Blum, E. K., Numerical Analysis and Computation, Addison-Wesley, Reading, Massachusetts (1972).

[2-16] Stone, M. H., "The Generalized Weirstrass Approximation Theorem," Mathematics Magazine, vol. 21, pp. 167-187 (1948).

[2-17] _____, "The Generalized Weirstrass Approximation Theorem," Mathematics Magazine, vol. 21, pp. 237-254 (1948).

[2-18] Hou, H. S., "Least Squares Image Restoration Using Spline Interpolation," Ph.D. Dissertation, USCIPI Report No. 650, Univ. of So. Calif., Los Angeles, Calif., (March 1976).

[2-19] Peyrovian, M. J., "Image Restoration by Spline Functions," USCIPI Report No. 680, University of Southern California, Los Angeles, California (August 1976).

[2-20] Rice, J. R., The Approximation of Functions, Vol. II, Addison-Wesley, Reading, Massachusetts (1969).

[2-21] Schultz, M. H., "L^{∞}-Multivariate Approximations Theory," SIAM J. Numer. Anal., vol. 6, No. 2, pp. 161-183 (1964).

[2-22] _____, "L^{2}-Multivariate Approximation Theory," SIAM J. Numer. Anal., vol. 6, no. 2, pp. 184-204 (1969).

[3-1] Rao, C.R., Linear Statistical Interference and Its Applications, Second Edition, John Wiley & Sons, New York (1973).

[3-2] Rao, C.R., Mitra, S. K., Generalized Inverse Matrices and Its Applications, John Wiley & Sons, New York (1971).

[3-3] Albert, A., Regression and the Moore-Penrose Pseudo Inverse, Academic Press, New York (1972).

[3-4] Penrose, N., "A Generalized Inverse for Matrices," Proc. Cambridge Philos. Soc., Vol. 51, pp. 406-413 (1955).

[3-5] Hunt, B. R., "The Application of Constrained Least Square Estimation to Image Restoration by Digital Computer," IEEE Trans. on Computers, Vol. C-22, No. 9, pp. 805-812 (1973).

[3-6] Golub, G. and Reinsch, C., "Singular Value Decomposition and Least Squares Solutions," Numer. Math., vol. 14, pp. 403-420 (1970).

[3-7] Bellman, R., Introduction to Matrix Analysis, Second Edition, McGraw-Hill, New York (1970).

[3-8] Keller, H. B., "On the Accuracy of Finite Difference Approximations to the Eigenvalues of Differential and Integral Operators," Numer. Math., vol. 7, pp. 412-419 (1965).

[3-9] Blum, E. K., Numerical Analysis and Computations, Addison-Wesley, Reading, Massachusetts (1972).

[3-10] Wielandt, H., "Error Bounds for Eigenvalues of Symmetric Integral Equations," Proc. Symp. Appl. Math., vol. VI, pp. 261-282 (1956).

[4-1] Bracewell, R. N., "Strip Integration in Radio Astronomy," Australian J. Phys., Vol. 9, p. 198 (1956).

[4-2] Hunt, B. R., "A Matrix Theory Proof of the Discrete Convolution Theorem," IEEE Trans. Audio Electro Acoust., Vol. AV-19, pp. 285-288 (1971).

[4-3] Bellman, R., Introduction to Matrix Analysis, Second Edition, McGraw-Hill, New York (1970).

[4-4] Golub, G. and C. Reinsch, "Singular Value Decomposition and Least Squares Solutions," Numer. Math., Vol. 14, pp. 403-420 (1970).

[4-5] Slepian, D. L., "Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty-IV: Extensions to Many Dimensions; Generalized Prolate Spheroidal Functions," Bell System Tech. J., Vol. 43, pp. 3009-3057 (1964).

[4-6] Gori, F. and G. Guattari, "Shannon Number and Degrees of Freedom of an Image," Opt. Comm., Vol. 7, No. 2, pp. 163-165 (1973).

[5-1] Golub, G. and C. Reinsch, "Singular Value Decomposition and Least Squares Solutions," Numer. Math., Vol. 14, pp. 403-420 (1970).

[5-2] Landau, H. J. and Pollak, H. O., "Prolate Spheroidal Waveforms, Fourier Analysis, and Uncertainty-III: The Dimension of the Space of Essentially Time-and-Band-Limited Signals," Bell Systems Technical Journal, Vol. 41, No. 4, pp. 1295-1336 (1962).

[5-3] Slepian, D. and Pollak, H. O., "Prolate Spheroidal Waveforms, Fourier Analysis, and Uncertainty-I, Bell Systems Technical Journal, Vol. 40, No. 1, pp. 43-64 (1961).

[5-4] DeBoor, C., "On Calculating with B-Splines," Journal of Approximation Theory, Vol. 6, pp. 50-62 (1972).

[5-5] Rice, J. R., The Approximation of Functions - Vol. II, Addison-Wesley, Reading, Massachusetts (1969).

[5-6] Schultz, M. H., Spline Analysis, Prentice-Hall, Englewood Cliffs, New Jersey (1973).

[6-1] DeBoor, C., "Good Approximations by Splines with Variable Knots," from Spline Functions and Approximation Theory, A. Meir and A. Sharma (eds.), Birkhauser Verlag Basel und Stüttgart (1973).

[6-2] Schultz, M. H., Spline Analysis, Prentice-Hall, Englewood Cliffs, New Jersey (1973).

[A-1] DeBoor, C., "On Calculating with B-Splines," Journal of Approximation Theory, Vol. 6, pp. 50-62 (1972).

[A-2] Rice, J. R., The Approximation of Functions - Vol. II, Addison-Wesley, Reading, Massachusetts (1969).