

Geometric Estimation and Reconstruction From Tomographic Data

by

Peyman Milanfar

B.S., University of California at Berkeley (1988)

S.M., E.E., Massachusetts Institute of Technology (1990), (1992)

Submitted to the Department of Electrical Engineering and
Computer Science

in partial fulfillment of the requirements for the degree of

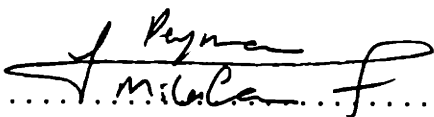
Doctor of Philosophy

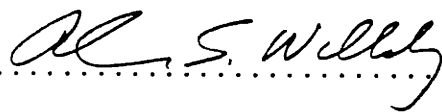
at the

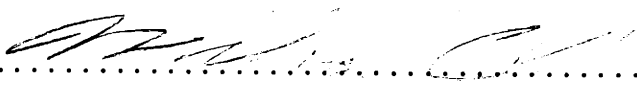
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 1993

©Massachusetts Institute of Technology, 1993. All rights reserved.

Author 
Department of Electrical Engineering and Computer Science
June 18, 1993

Certified by 
Alan S. Willsky
Professor of Electrical Engineering
Thesis Supervisor

Certified by 
William C. Karl
Research Scientist
Thesis Supervisor

Accepted by
Campbell L. Searle
Chairman, Departmental Committee on Graduate Students

Geometric Estimation and Reconstruction From Tomographic Data

by

Peyman Milanfar

Submitted to the Department of Electrical Engineering and Computer Science
on June 18, 1993, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

Tomographic signal processing encompasses a wide variety of applications in such diverse areas as medical imaging, geophysical and oceanographic signal processing, machine vision, and astrophysics. All such applications, however, share a common underlying structure. An object is to be reconstructed from its noisy, and possibly sparse, line-integral projections. Mathematically, the problem is that of inversion of the Radon-transform of a function from noisy data. Classical techniques such as the algebraic and transform techniques, along with their variants have constituted the main avenues of study in tomography in the past 2 decades. Many variations on these themes have been proposed and implemented with varying degrees of success. Most research efforts in tomography have ignored the issue of extracting explicit geometric information from the projection data. At best, these algorithms have allowed the incorporation of some geometric constraints in the reconstruction process. Until recently, however, the data itself has seldom been considered as directly containing encoded geometric information about the objects we seek to reconstruct. In this thesis, we present a framework for the direct extraction of geometric information from tomographic data and for image reconstruction based on this estimated information. In particular, we exploit the linear relationship between the moments of an image and the moments of its Radon transform to estimate the moments of the image from noisy projection measurements. These estimated moments, and the original data, then form the basis of our novel reconstruction algorithms. For finitely parameterized images, we study Maximum Likelihood, Minimum Description Length criteria, and array processing algorithms for the estimation of the image parameters. For general image reconstruction from estimated moments, we study iterative variational formulations based on divergence minimization criteria. We demonstrate that with appropriate initialization based on the available data, our methods show superior performance when compared to classical reconstruction algorithms.

Thesis Supervisor: Alan S. Willsky
Title: Professor of Electrical Engineering

Thesis Supervisor: William C. Karl
Title: Research Scientist

Acknowledgements

To say that Alan Willsky has had a profound influence on my education would be stating the obvious. I have often told friends and family that the most interesting part of my graduate studies has been the experience of working with Alan.

The help of Professor George Verghese has been instrumental in arriving at the results in Chapter 6 of this thesis. Aside from this, he has been a constant source of help and encouragement as my academic advisor for the past few years.

Thanks to Dr. Clem Karl who has always been there for help and conversation. It has certainly been a privilege (and a pleasure) to have someone sharp who is willing to listen to ideas, no matter how off the wall they may be. Also, Clem's meticulous reading of the drafts of this thesis has no doubt made it more readable.

I have had many officemates throughout the last 4.5 years, but without a doubt Mickey Bhatia is the funniest and easiest going dude I've ever had the pleasure of sharing an office with. Mark Luetggen and Eric "You Scared Me" Miller have always been there to chit chat and talk about technical problems and generally provide comic relief. Without the company of Ali Tassoudji, the afternoon tea would not have tasted as good. Through him I have learned to laugh and let the world laugh with me. These guys made the office a place you actually wanted to go to and that is rare.

It is really unfortunate that MIT does not give out honorary PhD's. My "cohabitant" and better half Sheila deserves an honorary PhD in both TLC and patience for putting up with my Monty-Python-esque life as an MIT graduate student. How do I thank thee.....

I thank my parents, my brother, and the rest of my family in Iran for their unfailing support both emotional and material. I will always be grateful to my parents for constantly reminding me that I could achieve whatever I wanted. I wish all parents would remind their children of the same and give them the same kind of love and support that my brother and I have been lucky enough to have.

Finally, I wish to acknowledge the generous support of the Clement Vaturi fellowship which I have held for the past year and a half.

I dedicate this thesis to Smoky and Bonita. The only cats in the world with working knowledge of tomographic reconstruction.

“I fully realize that I have not succeeded in answering all of your questions. ...Indeed, I feel I have not answered any of them completely. The answers I have found only serve to raise a whole new set of questions, which only lead to more problems, some of which we weren't even aware were problems.

To sum it all up...in some ways I feel we are confused as ever, but I believe we are confused on a higher level, and about more important things.”

—Anonymous truth-sayer

Contents

1	Introduction	15
1.1	Classical Reconstruction Techniques	17
1.2	Geometric Reconstruction	19
1.3	Contributions	22
1.4	Organization	23
2	Preliminaries	25
2.1	The Radon Transform and Its Properties	25
2.2	Geometry of the Projection Space	31
2.2.1	The Classical Radon Transform Constraints	31
2.2.2	Number of Projections and Reconstructability	32
3	Reconstruction of Finitely Parameterized Objects	40
3.1	Introduction	40
3.2	The Reconstruction Problem	42
3.2.1	Maximum Likelihood Approach	43
3.2.2	Minimum Description Length	45
3.2.3	Algorithmic Aspects – Computing A Good Initial Guess	46
3.3	Experimental Results	58
3.3.1	ML Based Reconstruction	60
3.3.2	MDL Reconstructions	68
3.3.3	Polygonal Reconstruction of Non-polygonal Objects	68
3.3.4	Initial Guess Algorithm	71

3.3.5	Conclusions from Experiments	75
3.4	Conclusions	77
3-A	Alternate Parameterizations	78
3-A.1	The Affine Representation of Triangles	79
3-B	Theoretical Results on the Initial Guess Algorithm	81
4	Moments and the Radon Transform	89
4.1	Introduction	89
4.2	Basic Definitions	91
4.3	Basic Results	95
4.3.1	Relating the Moments of f to the Moments of $g = \mathfrak{R}f$	95
4.3.2	Analytic Constraints on the Number of Uniquely Computable Moments	99
4.3.3	Specific Results for Binary Polygonal Objects	104
4.4	New Interpretations of Classical Reconstruction Algorithms	106
4.5	Optimal Estimation of Moments From Noisy Projections	108
4.5.1	Measurement and Noise Model	108
4.5.2	Maximum Likelihood (ML) Moment Estimation	110
4.5.3	Statistics of the Estimated Moments	111
4.6	Other Results on Moments and the Radon Transform	114
4.6.1	Consistent Estimation Using Moments	115
4.6.2	Moments as Sufficient Statistics	125
4.6.3	Better Sampling Strategies	129
4.7	Some Examples and Experiments	132
4.7.1	Consistent Estimation with Constraints	132
4.7.2	An Approximation of the Radon Transform	134
4.7.3	ML Moment Estimation Example and Performance	135
4-A	Proof of Lemma 2	139
4-B	Proof of Result 6	141
4-C	Proof of Result 7	142

4-D Proof of Lemma 3	142
5 Variational Formulations for Reconstruction from Moments	144
5.1 Introduction	144
5.2 Regularization	146
5.2.1 I-Divergence Regularization	146
5.2.2 Quadratic Regularization	150
5.3 Solution of Regularized Problems	151
5.3.1 I-Divergence Solution	152
5.3.2 Quadratic Solution	155
5.3.3 Properties of the Solutions	156
5.4 Recursive Regularization	160
5.4.1 Recursive I-divergence Regularization (RIDR)	161
5.4.2 Recursive Quadratic Regularization (RQR)	163
5.4.3 Recursive Regularization and Equality-Constrained Minimization	164
5.5 Numerical Examples	167
5.5.1 Example set I	169
5.5.2 Example set II	185
5.5.3 Example set III: The MIT Ellipse	191
5.6 Conclusion	192
6 Reconstruction of Polygons from Moments: Connections to Array Processing	198
6.1 Introduction	198
6.2 Mathematical Background	199
6.2.1 Vertices from Moments via Prony's Method	204
6.2.2 Remarks	207
6.3 Statistics of Estimated Complex Moments	211
6.4 Connections to Array Processing	214
6.5 Reconstruction Algorithms	215
6.5.1 Least Squares Prony Technique	216

6.5.2	Matrix Pencil Technique	219
6.6	Numerical Examples	221
6.7	Conclusion	230
7	Contributions, Conclusions, and Future Research	232
7.1	Contributions	232
7.1.1	Parametric Model-Based Reconstruction	232
7.1.2	Direct Extraction of Geometric Information from Tomographic Data	233
7.1.3	Variational Formulations for Reconstruction from Moments . .	234
7.1.4	Array Processing Methods for Reconstruction from Moments .	236
7.2	Future Research Directions	236
7.2.1	Geometric Constraints and Projection Data	237
7.2.2	Alternative Variational Formulations	239
7.2.3	Applications to Pattern Recognition and Data Compression .	241
7.2.4	Generalizations of the Motzkin-Schoenberg Formula	242
7.2.5	Other Array Processing Algorithms and Their Performance . .	243
7.3	Conclusions	244

List of Figures

2-1	The Radon transform	26
2-2	Projections of projections	34
2-3	Two objects having identical vertical and horizontal projections . . .	36
2-4	An additive vs. a non-additive ellipse with respect to the coordinate axes	38
3-1	A binary, polygonal object	47
3-2	A projection of a binary, polygonal object	47
3-3	Illustration of Result 2	57
3-4	Sample reconstruction of a triangle at 0 SNR 50 views, 20 samples/view: True(-), Reconstruction(- -)	61
3-5	Sample reconstruction of a hexagon at 0 SNR 50 views, 20 samples/view : True(-), Reconstruction(- -)	62
3-6	From Top to Bottom: Sinograms with 50 projections and 20 samples per projection of I) Noiseless Hexagon, II) Noisy data at SNR of 0 dB III) Reconstructed Hexagon. In each of these images, the horizontal axis is θ , the vertical axis is t , and the intensity values are the values of the corresponding projections mapped to the grayscale range of $[0, 255]$	63
3-7	Sample reconstruction of a hexagon at 0 SNR using FBP: 64 views, 64 samples per view	64
3-8	Sample reconstruction of a Hexagon at 0 SNR 64 views, 64 samples/view: True(-), Reconstruction(- -)	64

3-9	Mean performance curves for ML reconstructions of a triangle and a hexagon	65
3-10	Performance as a function of Number of Views	66
3-11	Performance as a Function of Number of Samples per View	67
3-12	Cost vs number of sides for the hexagon in Figure 3-5	69
3-13	Minimum MDL costs and Sample Reconstructions for an Ellipse . . .	70
3-14	True object (-), reconstruction (-), initial guess (o) picked by eye . .	71
3-15	True object (-), reconstruction (-), initial guess (o) picked using Initial Guess Algorithm	72
3-16	FBP Reconstruction of Non-polygonal, Non convex Object: 3rd order Butterworth filter with 0.15 normalized cutoff frequency	72
3-17	A sample path of the reconstruction error at SNR=0	73
3-18	Percent error for Hexagon vs SNR after outlier removal	75
3-19	A typical reconstruction at a local minimum with SNR=0	76
3-20	A typical reconstruction at a local minimum with SNR=10	76
3-21	Relative error in matching second order moments using the Initial Guess Algorithm	88
4-1	A diagram of the decomposition of the Radon transform	98
4-2	A diagram for the proof of Result 8	118
4-3	Condition Number of $\mathbf{D}_N^T \mathbf{B}_N^T \mathbf{B}_N \mathbf{D}_N$ vs the number of views for various N	132
4-4	Condition Number of $\mathbf{D}_N^T \mathbf{B}_N^T \mathbf{B}_N \mathbf{D}_N$ vs decreasing angular coverage for various N	133
4-5	The functions $g_N(t, 0)$ for $N = 5, 10, 15,$ and 20	136
4-6	The MIT ellipse phantom	137
4-7	Computed (-) and ML Estimated (o) moments up to order 10 of the MIT Ellipse image versus their index	138
4-8	The difference values between computed and ML Estimated moments up to order 10 of the MIT Ellipse image versus their index	138

4-9	Trace of covariance matrix versus moment order up to order 10	139
5-1	Taking uncertainty into account when picking prior	159
5-2	Phantom for Example set I	169
5-3	Left to Right: Sinogram of Phantom, Noisy Sinogram at SNR of 10 .	169
5-4	Filtered Backprojection with 3rd order Butterworth filter with 0.25 normalized cutoff frequency. SNR=10	170
5-5	Counter-clockwise from upper left: Phantom, computed FBP initial estimate (% MSE=55.3, RQR result after 2 iterations (% MSE=17.8), RQR result after 8 iterations (% MSE=13.4). Data: 64 projections w/ 64 samples per projection at SNR= 10; moments up to order 8 were used.	172
5-6	Counter-clockwise from upper left: Phantom, uniform initial estimate (% MSE=65.7), IDR solution (% MSE=55.9), RIDR solution (% MSE=15.8). Data: 64 projections w/ 64 samples per projection at SNR= 10; mo- ments up to order 8 were used.	173
5-7	Counter-clockwise from upper left: Phantom, Initialization computed using (5.83) without support information (% MSE=39.8), IDR solu- tion (% MSE=31), RIDR solution (% MSE=10.3). Data: 64 projec- tions, 64 samples per projection with SNR=10; moments up to order 8 used.	175
5-8	Counter-clockwise from upper left: Phantom, f_0 based on FBP (% MSE=55.3), RIDR solution after 3 iterations (% MSE=38.1), RIDR solution after 10 iterations (% MSE=11.1). Data: 64 projections w 64 samples per projection at SNR= 10; moments up to order 8 used. . .	177
5-9	MSE versus number of moments used in reconstructing the phantom of Figure 5-2	178
5-10	Counter-clockwise from upper left: reconstructions using moments up to order 2, 5, 8, and 11. Data: 64 projections w/64 samples per pro- jection at SNR= 10. Initial guess was based on FBP in every case. . .	178

5-11	MSE versus SNR in reconstructing the phantom of Figure 5-2 : Moments up to order 8 used.	179
5-12	Counter-clockwise from upper left: Phantom, Initialization computed using (5.83) with support information (% MSE=37.9), IDR solution (% MSE=28), RIDR solution (% MSE=8.9). Data: 64 projections, 64 samples per projection with SNR=10; moments up to order 8 used. .	181
5-13	Counter-clockwise from upper left: Phantom, Initialization computed using FBP with support information (% MSE=10.95), IDR solution (% MSE=10.7), RIDR solution (% MSE=6.2). Data: 64 projections, 64 samples per projection with SNR=10; moments up to order 8 used.	182
5-14	MDL cost for polygonal fit vs number of sides	184
5-15	Counter-clockwise from upper left: Phantom, FBP-based initial estimate with estimated support (% MSE=24.3), RIDR result after 3 iterations (% MSE=15.3), RIDR result after 10 iterations (% MSE=9.4). Data: 64 views, 64 samples per view at SNR=10; moments up to order 8 used.	184
5-16	MSE of Initial Estimates vs. MSE of final RIDR reconstructions for Example set I: Moments up to order 8 used	185
5-17	Phantom for Example set II	187
5-18	Counter-clockwise from top left: Phantom, Initial Estimate from FBP, IDR reconstruction, Final RIDR reconstruction (32 views, 64 samples per view, SNR =10, moments up to order 10 used)	188
5-19	Left to right: Noiseless and noisy sinograms for phantom in Figure 5-17 : 32 views, 64 samples per view, SNR =10	188
5-20	Counter-clockwise from top left: Phantom, Initial Estimate from FBP, IDR reconstruction, Final RIDR reconstruction (64 views, 64 samples per view, SNR =10, moments up to order 10 used)	189
5-21	Left to right: Noiseless and noisy sinograms for phantom in Figure 5-17 : 64 views, 64 samples per view SNR =10	189

5-22	Counter-clockwise from top left: Phantom, Initial Estimate from FBP, IDR reconstruction, Final RIDR reconstruction (128 views, 64 samples per view, SNR =10, moments up to order 10 used)	190
5-23	Left to right: Noiseless and noisy sinograms for phantom in Figure 5-17 : 128 views, 64 samples per view, SNR =10	190
5-24	MSE versus number of views for computed FBP prior, IDR solution and RIDR solution, SNR=10, moments up to order 10 used	191
5-25	The MIT ellipse phantom	194
5-26	The noiseless sinogram of the MIT ellipse: 60 views, 81 samples/view	194
5-27	Counter-clockwise from top left:MIT ellipse sinograms at SNR = 20, 10, 0, and -10: each contains 60 views with 81 samples/view	195
5-28	Counter-clockwise from top left:FBP reconstructions of MIT ellipse at SNR = 20, 10, 0, and -10	196
5-29	Counter-clockwise from top left:IDR reconstructions of MIT ellipse at SNR = 20, 10, 0, and -10. Moments up to order 5 used.	196
5-30	Counter-clockwise from top left: RIDR reconstructions of MIT ellipse at SNR = 20, 10, 0, and -10. Moments up to order 5 used.	197
5-31	MSE versus SNR for FBP and RIDR solution with FBP-based initial estimate: Moments up to order 5 used.	197
6-1	Three distinct regions corresponding to the same vertices	208
6-2	Performance curves at SNR=55	223
6-3	Overlaid performance curves at SNR=55	224
6-4	Sample reconstructions at SNR=55 solid: actual, circles: reconstructed	224
6-5	Sample reconstructions at SNR=55 solid: actual, circles: reconstructed	225
6-6	Sample reconstructions at SNR=55 solid: actual, circles: reconstructed	225
6-7	Sample reconstructions at SNR=55 solid: actual, circles: reconstructed	226
6-8	Sample reconstructions at SNR=150 solid: actual, circles: reconstructions	228

6-9	Sample reconstructions at SNR=150 solid: actual, circles: reconstructions	229
6-10	Sample reconstructions at SNR=150 solid: actual, circles: reconstructions	229
7-1	Parallel Beam Projection Data	237
7-2	Fan-Beam projection data	239

Chapter 1

Introduction

Tomographic signal processing encompasses a wide variety of applications in such diverse areas as medical imaging [40], geophysical and oceanographic signal processing [65, 50], machine vision [80], and astronomy [16]. All such applications, however, share a common underlying structure. An object is to be reconstructed from its noisy, and possibly sparse, line-integral projections. Mathematically, the problem is that of inversion of the Radon-transform of a function given noisy observations. Classical techniques such as the Algebraic Reconstruction Technique (ART) and Filtered Backprojection (FBP) [40], along with many variants of these techniques, have been widely applied to pixel-based reconstruction of images from noisy projections. In many applications of tomography, however, the aim is to extract a rather small set of geometric features from a given set of projection data [7, 98, 92]. In such instances, a full pixel-by-pixel reconstruction of the object is a rather inefficient and non-robust approach. In addition, it is often difficult or impossible to collect a dense data set of high quality. A case in point is ocean acoustic tomography where due to the enormous size and nonstationarity of the objects (bodies of water) under study, only sparse measurements of relatively poor quality are possible [65, 50]. When the available data set is sparse and very noisy, classical reconstruction techniques fail to produce acceptable reconstructions. Two main sources can be cited for the shortcomings of these classical techniques. First, classical techniques are invariably aimed at reconstructing every pixel value of the underlying object with little regard for the

quality and quantity of the available data. To put it differently, there is no implicit or explicit mechanism to control greed and focus information, thus preventing one from attempting to extract more information from a data set than it actually contains. The second type of shortcoming results from the fact that if we assume that the projection data are corrupted by Gaussian white noise, the process of reconstruction will have the net effect of “coloring” this noise. This effect manifests itself in the object domain in the form of spurious features which will complicate the detection of geometric features. This observation points out the importance of working directly with the projection data when the final goal is the extraction of geometric information. In this thesis, we present a statistically optimal framework for the estimation and reconstruction of objects based on the direct extraction of geometric information from noisy projection measurements. In particular, we develop optimal algorithms for the estimation of the moments of an image directly from its noisy projection measurements. Given these estimated moments, and the original projections, we develop efficient optimal algorithms for the reconstruction of images based on this geometric information. Several advantages are inherent to our approach. The most important advantage of our approach is that it provides a sound framework for image reconstruction with controlled degrees of freedom. This level of control not only manifests itself in terms of better reconstructions, but it also translates into significant computational savings. Another advantage is that in our framework, in contrast to most classical algorithms, prior geometric information such as spatial support, and other information such as positivity of the image values can be directly incorporated into the reconstruction process. Although we have not addressed the issue of recognition and classification of images from estimated moments, such application of moment estimation from noisy projections is clearly a possibility. In contrast to classical approaches to diagnostic tomography which are based on direct image reconstruction, given the estimated moments of the underlying image and the respective covariances of these estimates, the problem of classification of the underlying image based on measured projections can be addressed in a statistically optimal fashion *without* resorting to reconstruction. This application of moment estimation from projections

would, at least in theory, eliminate the need for reconstruction if the end objective is classification and recognition of particular features of images.

In Section 1.1 we will provide a brief summary of classical tomographic reconstruction techniques. Following this, in Section 1.2, we summarize the literature on geometrically based reconstruction techniques, such as ours, which have explicitly studied the direct extraction of geometric information from projection data and on reconstruction algorithms based on such information. In particular, we highlight those research efforts that have led to the work presented in this thesis. In Section 1.3 we describe the major contributions of this thesis. Section 1.4 will describe the organizational layout of this thesis. In the remainder of this thesis, the values of the Radon transform, whose formal definition is stated in Section 2.1 are interchangeably referred to as projections. In addition, depending on the context, the term *image* is occasionally interchanged with *function*. The term *object* is used instead of image if it is assumed that what is to be estimated is a single geometric entity such as a binary polygon.

1.1 Classical Reconstruction Techniques

Perhaps the most widely used, and arguably the most influential set of classical reconstruction techniques, are the *transform techniques* such as Filtered Backprojection (FBP) and the *iterative techniques* such as the Algebraic Reconstruction Technique (ART) [40]. The transform methods seeks to construct effective approximate inversion algorithms for the Radon transform and do not perform well in the absence of full angular coverage and high signal to noise ratios. In iterative techniques, sometimes referred to as *Finite Series Expansion Methods* the function f representing the image is approximated by a finite sum of weighted basis functions $\{\psi_i\}$. i.e.

$$f(x, y) \approx \sum_i f_i \psi_i(x, y) \quad (1.1)$$

The reconstruction process now consists of estimating the weighting factors f_i . The noisy measurements here are in essence given as the noisy inner product of f with measurement basis functions $\{\phi_i\}$.

$$d_i = \iint_{\mathcal{O}} f(x, y) \phi_i(x, y) dx dy + e_i \quad (1.2)$$

The estimation of the vector $\bar{f} = [f_1, f_2, \dots, f_n]^T$ can now be formulated as a linear estimation problem.

$$d = H\bar{f} + e \quad (1.3)$$

where

$$d = [d_1, d_2, \dots, d_n]^T \quad (1.4)$$

$$e = [e_1, e_2, \dots, e_n]^T \quad (1.5)$$

$$H_{ij} = \iint \phi_i(x, y) \psi_j(x, y) dx dy. \quad (1.6)$$

The matrix H is commonly referred to as the projection matrix. Many variants of the above formulation have been studied in the literature [40] with each corresponding to essentially a different choice of the basis functions ϕ_i and ψ_i . Although the formulation (1.3) suggest a straightforward *linear* estimation framework, the typically large dimensionality of H make these approaches computationally difficult even though efficient iterative algorithms are usually applied to solve such linear estimation problems. Recently, however, some researchers have made use of particular families of functions $\psi_i(x, y)$, [11] (natural pixels), [12] (polar pixels), and families of overlapping functions [37], to produce more computationally efficient and robust algorithms.

As discussed earlier in this chapter, in many applications of tomography such as ocean acoustic imaging, data is collected sparsely either due to the geometry of the object or the imaging apparatus, or due to cost or damage incurred as a result of imaging (such as when large doses of radiation may harm the subject under study). In such instances, data is typically sparse both in terms of the number of views and the number of samples per view. Early on, the standard Algebraic Reconstruction

Techniques (ART) were applied to solve the problem with no particular provisions made for the missing angles. These resulted in reconstructions with streaking effects [13]. Several modifications of the series expansion techniques have since been proposed to deal with the case of sparse angles. Later work by Wood, Mackovski, and Morf [103] showed that with the addition of an object prior covariance, the linear minimum variance estimate gave better reconstructions.

Variations of transform techniques have also been proposed for the limited angle reconstruction problem. Louis [58], for instance, has used the Radon transform consistency conditions to “estimate” the missing projections. He then uses FBP to recover the object. Some of our work in the later chapters will be closely related to this idea. More recently, Reed and Shepp [73] have devised an algorithm called “squashing” which, aside from its computational savings, is effective in relieving the major artifacts caused by the lack of angular measurement. The drawback of this technique is that it does not provide any tools for the incorporation of prior knowledge, whether geometric or probabilistic, into the reconstruction process.

Constraints given by prior knowledge and consistency have also been used in iterative reconstruction algorithms known as Projection onto Convex Sets (POCS) [68]. These algorithms iteratively reconstruct the object by projecting the data between convex constraint sets in a given function space. Convergence is guaranteed due to the convexity of the constraint sets.

1.2 Geometric Reconstruction

In light of the success of the classical reconstruction techniques in the full-view, high SNR cases, little attention has been paid to geometric/parametric modeling in tomographic reconstruction. The approach of direct geometric modeling is to parameterize the class of objects of interest so that tomographic reconstruction is in essence reduced to the optimal estimation of finitely many *geometric* features, or parameters. This is in a sense similar to the finite series expansion methods, with the fundamental difference being that in geometric reconstruction, parameters represent geometric detail through

an appropriate choice of basis functions that describe the image. Tomographic reconstruction is often a geometric problem since in many cases, the end goal is to obtain some geometric information about the underlying image. Hence the formulation of geometric models appears to be a natural framework for its solution. Given the success of classical techniques in high SNR, full view cases, geometric reconstruction will perhaps find its best applications in the low SNR, limited angle situations where the explicit modeling of the underlying image as a finitely parameterized function can be exploited to obtain accurate estimates of the unknown parameters. Recall that the general algebraic formulation of the pixel-based tomographic reconstruction problem (1.3) can be viewed as a high-dimensional linear estimation problem. The fundamental contribution of geometric/parametric analysis of tomographic reconstruction algorithms is that by limiting the number of parameters to be estimated, and carefully choosing the parameterization itself, the information contained in the measured projections is *focused* to where it is ultimately most valuable: in describing a few important geometric features of the underlying image. In essence, geometric reconstruction algorithms provide a framework in which it is possible to control the degrees of freedom of the reconstruction and still obtain good results. This is the idea of controlling the *greed factor*. By focusing the information contained in the measured projections, one is prevented from attempting to extract more information from the data than it actually contains. Classical reconstruction algorithms are, in this sense, greedy in that regardless of the quality and quantity of the available data, the aim is always to reconstruct the pixel-by-pixel values of the underlying image.

On the other hand, the explicit reconstruction algorithms in parametric frameworks, are typically highly nonlinear optimization problems. This, however, does not necessarily mean that the resulting algorithms are more computationally intensive than other techniques. In fact, as we will demonstrate in this thesis, by making good use of the available data to produce reasonable initial estimates, these nonlinear optimization problems may be solved quite efficiently.

The work of Rossi and Willsky [76] and Prince and Willsky [72] has served as the starting point of this research effort. In the work of Rossi, the object was represented

by a known profile, with only three geometric parameters; namely size, location, and eccentricity. These parameters were then estimated from projection data using the Maximum Likelihood (ML) formulation. In their approach, the number of unknown parameters was fixed and the main focus of their work was on performance analysis. Prince, working in the projection space, used a priori information such as prior probabilities on sinograms and consistency conditions to compute Maximum A Posteriori (MAP) estimates of the sinogram and then used CBP to obtain a reconstruction. He made use of prior assumptions about shape, such as convexity, to reconstruct convex objects from support samples which were extracted from noisy projections through optimal filtering techniques. The approach of Prince provided an explicit method of integrating geometric information into the reconstruction process but was in essence still a pixel-by-pixel reconstruction. Extending these ideas, Lele, Kulkarni, and Willsky [54] made use of only support information to produce polygonal reconstructions. Karl [47] also has studied the reconstruction of 3-D objects from two-dimensional silhouette projections.

The geometric modeling approach of Rossi and Willsky was expanded upon to include a more general set of objects by Bresler and Mackovski [10] and Fessler and Mackovski [24]. The former work chose sequences of 3-D cylinders with unknown radius, position, and orientation to model blood vessels being tomographically imaged in 3-D. The latter work used ellipsoids with unknown parameters to reconstruct 3-D arterial trees from a few magnetic resonance angiograms. Recently, Thirion [98] has introduced a technique to extract boundaries of objects from raw tomographic data through edge detection and in the sinogram. Other work in geometric reconstruction by Chang [14] and more recently Kuba, Volcic, Gardner and Fishburn, [51, 101, 25, 27] has been concerned with the reconstruction of binary objects from only two noise-free projections.

In this thesis, we provide a framework for the optimal reconstruction of an image from its noisy Radon transform samples by first estimating the moments of the underlying image from the noisy raw data. Having done so, we present efficient algorithms for the optimal reconstruction of the underlying image using the estimated moments

and, in some cases, the original projection data. The problem of estimating moments from noisy projections turns out to be a simple linear estimation problem, while the problem of reconstructing the image from its estimated moments is a highly ill-posed inverse problem. We present stable and efficient algorithms for reconstructing images from their estimated moments.

1.3 Contributions

The contributions of this thesis can be group into several categories.

1. With regard to geometric model-based reconstruction, our work serves to continue the works of Rossi, Prince, and Bresler [10, 24, 72, 76] to:
 - cover a much less restrictive object class, namely binary polygonal objects rather than convex, ellipsoidal or cylindrical shapes,
 - provide an automatic mechanism for identifying the statistically optimal number of reconstruction parameters
2. We have explicitly used the linear relationship between the moments of a function and those of its Radon transform in an estimation theoretic framework. Although this relationship has been known to mathematicians for some time (since at least 1966 to be exact [30]), the engineering community has not made much use of it for reconstruction purposes until now. The explicit use of this relationship has many apparent advantages. We have studied the following aspects:
 - It is possible, and quite simple, to extract geometric information about an object in the form of moments directly from (the moments of) its noisy Radon transform data and without reconstruction of the underlying image.
 - The (unique) computation of the moments of a function from its Radon transform requires a minimum number of views. This, in essence provides one with a theoretical upper bound on the amount of uniquely identifiable geometric information that is contained in a limited data set.

- The linear nature of the dependence of the moments of an image on those of its Radon transform make the explicit computation of estimation error covariances possible. This allows us to specify not only the geometric features of an image directly from projections, but also the degree of confidence which we place on these estimates.
 - The moments of the projections are shown to be sufficient statistics for the estimation of a function f from its Radon transform projections.
 - The Radon transform operator \mathfrak{R} as an operator from one Hilbert space to another, can be decomposed in terms of operators which map a function and its Radon transform to their respective moments. This leads to a new set of interpretations of classical reconstruction algorithms.
 - The problem of estimation of moments of an image from raw projection data is numerically more stable for a particular set of sampling strategies in the projection domain. Hence, we can use the analysis of the stability of this estimation problem to arrive at optimal sampling strategies.
3. Given the estimated moments of an image from projections, we have devised novel reconstruction algorithms for estimating the image. In particular, we have investigated algorithms for:
- direct reconstruction of binary polygonal objects using a finite number of their moments and the raw data, via Maximum Likelihood estimation
 - reconstruction of binary polygonal objects using a finite number of their moments via array processing algorithms
 - variational formulations with moment constraints based on divergence and quadratic cost functionals for reconstruction of general images.

1.4 Organization

The organization of this thesis is as follows. Chapter 2 contains a brief review of essential concepts and definitions which we will refer to throughout the thesis. Chap-

ter 3 describes our contributions to the study of the reconstruction of parameterized objects from projections. Chapter 4 contains the contributions we have made in making explicit use of the connection between the moments of a function and those of its Radon transform. In this chapter we describe the optimal estimation of moments of a function from noisy Radon transform projection data. In Chapter 5 we discuss novel moment-based variational reconstruction algorithms that are based on divergence and quadratic minimization criteria. In Chapter 6 we describe reconstruction algorithms for binary polygonal objects based on array processing techniques. Finally, in Chapter 7 we summarize the contributions of this thesis, propose several directions of future research and state our concluding remarks.

Chapter 2

Preliminaries

In this chapter, we present some preliminary concepts and definitions about the properties of the Radon Transform, its range, and support. In this context, we will describe the role of basic geometric constraints in tomographic reconstruction. The facts, definitions and results presented in this chapter will serve not only to highlight the geometric aspects of the properties of the Radon transform, but also as a collection of concepts and definitions to which we will refer throughout the thesis.

2.1 The Radon Transform and Its Properties

The Radon Transform [39, 40, 59] of a function $f(x, y)$, denoted as $g(t, \theta)$, is defined as its line integrals along lines forming an angle of $\theta + \frac{\pi}{2}$ with the x-axis and a radial distance t from the origin. More precisely, if $f(x, y)$ has support in some compact region of the plane \mathcal{O} , its Radon Transform is given by

$$g(t, \theta) \equiv \mathfrak{R}f = \int \int_{\mathcal{O}} f(x, y) \delta(t - \mathbf{x}^T \boldsymbol{\omega}) dx dy, \quad (2.1)$$

where

$$\mathbf{x} = [x, y]^T, \quad (2.2)$$

$$\boldsymbol{\omega} = [\cos(\theta), \sin(\theta)]^T. \quad (2.3)$$

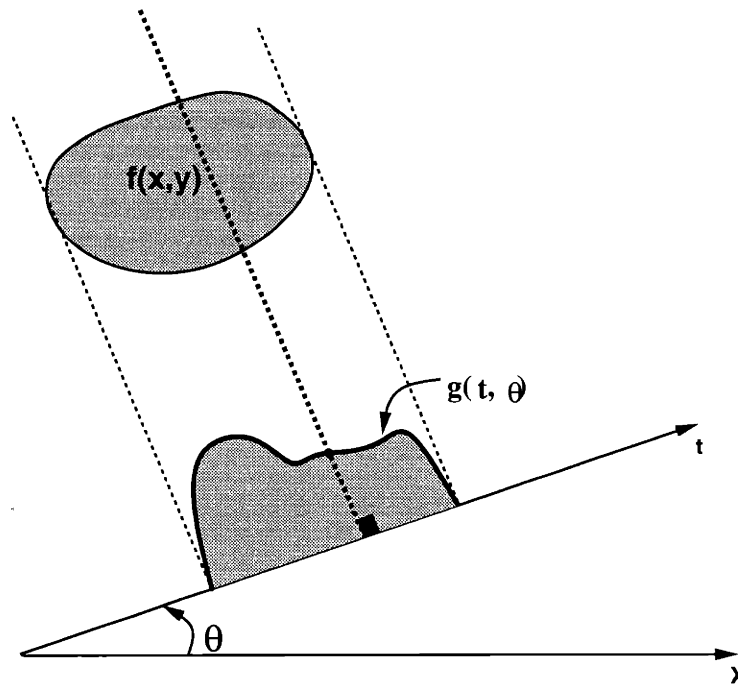


Figure 2-1: The Radon transform

and $\delta(\cdot)$ denotes the Dirac delta function. See Figure 2-1.

Note that the vector ω is a unit direction vector at angle θ with the x-axis. In what follows we will assume that the function $f(x, y)$ vanishes outside of the closed unit disk D so that $\mathcal{O} \subset D$. We shall also assume that $f(x, y)$ is a non-negative function, and that it is square-integrable so that (2.1) is well defined. The assumption of positivity is simply a consequence of practical considerations about the types of applications we have in mind. It is also sufficient to consider the values of $\theta \in [0, \pi)$ due to the symmetry properties of the Radon Transform which we will discuss later. The function $g(t, \theta)$ has its support on the cylinder given by

$$\mathcal{S} = \{(t, \theta) : -1 \leq t \leq 1, 0 \leq \theta < \pi\} \quad (2.4)$$

The Radon Transform maps the object domain (or spatial domain) (x, y) to the projection domain (t, θ) . Each point (t, θ) in the projection domain corresponds to a

line in the object domain through the following equation:

$$x \cos(\theta) + y \sin(\theta) = t \quad (2.5)$$

Correspondingly, a point $[x_0, y_0]^T$ in the object domain maps to a sinusoid in the projection domain through (2.5).

The Radon Transform is a linear operator and has several useful properties which we describe next. The proofs of the following properties, if not explicitly discussed, may be found in [30, 46] among other places.

Property 1 (Linearity) *If a and b are real numbers,*

$$\mathfrak{R}(af_1(x, y) + bf_2(x, y)) = ag_1(t, \theta) + bg_2(t, \theta) \quad (2.6)$$

Property 2 (Bounded Support) *If $f(x, y) = 0$ for $|x|, |y| > \frac{d}{2}$, we have that $g(t, \theta) = 0$ for $|t| > \frac{\sqrt{2}d}{2}$.*

Property 3 (Symmetry) *The Radon Transform is symmetric as follows:*

$$g(-t, \theta + \pi) = g(t, \theta) \quad (2.7)$$

Property 4 (Periodicity) *The Radon Transform is periodic in the variable θ with period 2π . i.e. for every integer k , we have:*

$$g(t, \theta + 2k\pi) = g(t, \theta) \quad (2.8)$$

Property 5 (Shift) *Given the coordinates of the point $\mathbf{p}_0 = [x_0, y_0]^T$, we have*

$$\mathfrak{R}f(x - x_0, y - y_0) = g(t - \omega^T \mathbf{p}_0, \theta) \quad (2.9)$$

where ω is the unit direction vector at angle θ .

Property 6 (Rotation) *Let $f(r, \phi)$ denote the polar representation of the function*

$f(x, y)$ through the transformations

$$x = r \cos(\phi) \quad (2.10)$$

$$y = r \sin(\phi). \quad (2.11)$$

The Radon transform of f then satisfies:

$$\mathfrak{R}f(r, \phi + \theta_0) = g(t, \theta + \theta_0) \quad (2.12)$$

Property 7 (Scaling) *If a is a nonzero real number, we have*

$$\mathfrak{R}f(ax, ay) = \frac{1}{|a|}g(at, \theta) \quad (2.13)$$

Property 8 (Convolution) *Let $f(\mathbf{x})$, with $\mathbf{x} = [x, y]^T$, be given by the 2-d convolution of two functions f_1 and f_2 which have Radon transforms g_1 and g_2 , respectively. i.e.*

$$f(\mathbf{x}) = \int \int_{\mathcal{R}^2} f_1(\mathbf{s})f_2(\mathbf{x} - \mathbf{s}) ds. \quad (2.14)$$

Then we have:

$$\mathfrak{R}f(\mathbf{x}) = \int_{\mathcal{R}} g_1(s, \theta)g_2(t - s, \theta) ds \quad (2.15)$$

Property 9 (Differentiation in t) *Suppose that $f(x, y)$ is differentiable in x and y . Then, given real numbers a and b , we have:*

$$\mathfrak{R}\left(a\frac{\partial f}{\partial x} + b\frac{\partial f}{\partial y}\right) = (a \cos(\theta) + b \sin(\theta))\frac{\partial g(t, \theta)}{\partial t} \quad (2.16)$$

From this, it then follows that if $P(\cdot, \cdot)$ is a polynomial differential operator of order k with constant coefficients, we have:

$$\mathfrak{R}\left(P\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right)f(x, y)\right) = P(\cos(\theta), \sin(\theta))\frac{\partial^k g(t, \theta)}{\partial t^k} \quad (2.17)$$

Property 10 (Differentiation in θ) *Provided that all relevant quantities exist,*

$$\Re((ax + by)f(x, y)) = \left(\frac{a}{\sin(\theta)} - \frac{b}{\cos(\theta)} \right) \frac{\partial g(t, \theta)}{\partial \theta}. \quad (2.18)$$

The above ten properties constitute the most basic properties of the Radon Transform. Other properties of the Radon Transform can be established that relate the Fourier Transform of $f(x, y)$ to its Radon Transform. The relation between the Fourier Transform and the Radon Transform is actually only a Corollary to a more general result which we discuss next.

Theorem 1 ([90]) *Let the square integrable function $f(x, y)$, with support $\mathcal{O} \subset D$, and its Radon Transform $g(t, \theta)$ be given. Then for any square-integrable function $F(t)$ defined over $[-1, 1]$ we have*

$$\int_{-1}^1 g(t, \theta) F(t) dt = \int \int_{\mathcal{O}} f(x, y) F(x \cos(\theta) + y \sin(\theta)) dx dy \quad (2.19)$$

This result in essence follows from the definition of the Radon Transform.

$$\begin{aligned} \int_{-1}^1 g(t, \theta) F(t) dt &= \int_{-1}^1 \int \int_{\mathcal{O}} f(x, y) \delta(t - x \cos(\theta) - y \sin(\theta)) F(t) dx dy dt \\ &= \int \int_{\mathcal{O}} \int_{-1}^1 f(x, y) \delta(t - x \cos(\theta) - y \sin(\theta)) F(t) dt dx dy \\ &= \int \int_{\mathcal{O}} f(x, y) F(x \cos(\theta) + y \sin(\theta)) dx dy \end{aligned}$$

As a corollary to Theorem 1, the celebrated *Projection-Slice Theorem* [55] is obtained by letting $F(t) = e^{-it}$. This result relates the Fourier Transform of f to its Radon Transform.

Corollary 1 (Projection-Slice Theorem) *The Fourier and Radon Transforms of $f(x, y)$ are related by the following identity.*

$$\int_{-1}^1 g(t, \theta) e^{-it} dt = \int \int_{\mathcal{O}} f(x, y) e^{-i(x \cos(\theta) + y \sin(\theta))} dx dy \quad (2.20)$$

Recall that the Fourier Transform of the function f is defined by

$$F(u, v) = \int \int_{\mathcal{O}} f(x, y) e^{-i(ux+vy)} dx dy \quad (2.21)$$

Hence, the Projection Slice Theorem states that for any fixed angle θ , the 1-D Fourier Transform of $g(t, \theta)$ is equal to a central slice at an angle θ through the 2-D Fourier Transform of the function $f(x, y)$. This result forms the basis of a set of important classical reconstruction techniques known as the *Transform Methods* [40].

In the next result, we have established a relationship between the Radon Transform of $f(x, y)$ and the support function [9] of its domain \mathcal{O} . The support function $h(\theta)$ of the compact set \mathcal{O} is defined as:

$$h(\theta) = \sup_{\mathbf{x} \in \mathcal{O}} \mathbf{x}^T \boldsymbol{\omega}, \quad (2.22)$$

where $\boldsymbol{\omega}$ is defined in (2.3). Given this definition we have

Result 1 *Let $f(x, y)$ be a positive function, with support in a compact subset \mathcal{O} of the closed unit disk. Denote by $g(t, \theta)$ the Radon Transform of $f(x, y)$, and by $h(\theta)$ the support function of \mathcal{O} . Then, for every positive integer m ,*

$$\left| \int_{-1}^1 g(t, \theta) t^m dt \right| \leq \mu_{00} |h(\theta)|^m \quad (2.23)$$

where

$$\mu_{00} = \int \int_{\mathcal{O}} f(x, y) dx dy \quad (2.24)$$

Proof: By definition, for all $\mathbf{x} \in \mathcal{O}$, we have

$$\mathbf{x}^T \boldsymbol{\omega} \leq h(\theta). \quad (2.25)$$

Applying Theorem 1, the result follows immediately. \square

Note that (2.23) actually relates the moments of $g(t, \theta)$ (the left hand side of the inequality) to the support function of the set \mathcal{O} . Since we have assumed that $\mathcal{O} \subset D$, it follows that the magnitude of the support function $h(\theta)$ is bounded above by unity.

Hence, (2.23) shows that all the moments of $g(t, \theta)$ must be bounded in magnitude by the zeroth order moment μ_{00} .

2.2 Geometry of the Projection Space

In this section we discuss some properties of the range and null-space of the Radon transform. In Section 2.2.1 we discuss the classical consistency results which define the range of the Radon transform, while Section 2.2.2 contains a review of results from the literature on reconstruction of binary planar sets from few projections in the absence of noise. These results provide insights into the structure of the null-space of the Radon transform.

2.2.1 The Classical Radon Transform Constraints

Not all functions $g(t, \theta)$ are Radon Transforms of some $f(x, y)$. For instance, a necessary condition for $g(t, \theta)$ to be the Radon transform of a function $f(x, y) \geq 0$ is that $g(t, \theta) \geq 0$. Several well-known conditions, known as the consistency relations, specify necessary and sufficient conditions for a function $g(t, \theta)$ to be a valid Radon transform. The Radon transform of a function $f(x, y)$ is constrained to lie in a particular functional subspace of the space of all real-valued functions $g(t, \theta) : \mathcal{R} \times \mathbf{S}^1 \rightarrow \mathcal{R}$, where \mathbf{S}^1 is the unit circle. This subspace is characterized by the fact that g must be an even function of t , and that certain coefficients of the Fourier expansion of g must be zero. For the sake of simplicity, we state the corresponding classical theorem in 2 dimensions and for square-integrable functions with bounded support. General versions of this result hold true not only in higher dimensions, but more generally over certain Banach spaces.

Theorem 2 ([39, 59]) *Let $L^2(D)$ denote the space of all square-integrable functions $f : D \rightarrow \mathcal{R}$ with support inside the closed unit disk D . In order for $g(t, \theta)$ to be the Radon transform of some function $f \in L^2(D)$, it is necessary and sufficient that*

1. $g \in L^2([-1, 1] \times [0, 2\pi])$,

2. $g(t, \theta + \pi) = g(-t, \theta)$, and

3. the integral

$$\int_{-1}^1 g(t, \theta) t^k dt \quad (2.26)$$

be a homogeneous polynomial of degree k in $\cos(\theta)$ and $\sin(\theta)$ for every positive integer k .

Let us make some observations regarding the third condition in the above theorem. First, note that the expression (2.26), for a fixed $\theta = \theta_0$, represents the **moments** of the function $g(t, \theta_0)$. So this condition states that the k^{th} order moments of the projection function g in any fixed direction θ_0 are homogeneous polynomials of order k in $\cos(\theta_0)$ and $\sin(\theta_0)$. Furthermore, the following identity is easily established as a corollary to Theorem 1.

Corollary 2 *Given a square-integrable function $f(x, y)$ with compact support $\mathcal{O} \subset D$, we have*

$$\int_{-1}^1 g(t, \theta) t^k dt = \iint_{\mathcal{O}} f(x, y) (x \cos(\theta) + y \sin(\theta))^k dx dy \quad (2.27)$$

for all positive integers k .

The equation (2.27) is the explicit connection between the moments of f and those of its Radon transform g , as we will show in Chapter 4.

2.2.2 Number of Projections and Reconstructability

In this section we review some results and observations regarding the problem of reconstructability of a binary planar sets from a few projections in the absence of noise. These results provide insights into the structure of the null-space of the Radon transform for a particular class of objects. Although the results we present in this section are mostly of theoretical nature, with little or no reference to applications, they are useful in showing just what can be done with a very limited data set. These results are, in fact, useful from a practical point of view when few projections are available and the underlying object is binary or a close approximation thereof. Such

a situation arises in the study of cold core rings in gulf streams [65, 50]. A cold core ring is a rather homogeneous body of cold water that is found encompassed in a much larger body of warm water known as a gulf stream. In acoustic transmission tomography, measurements of these cold core rings are made by computing the attenuation of acoustic waves which are transmitted across the gulf stream. The cold core ring, due to its homogeneous nature can be adequately approximated by a binary region (in 3-D). Due to the size and continuous motion of the body of water, it is impossible to make more than a few projection measurements in a small time interval. Hence, in this scenario we are essentially concerned with the reconstruction of a binary region from very few projections. Even though the results that follow do not deal with measurement error, they provide some theoretical framework for dealing with a practical situation as the one described above.

In general, no finite set of projections will allow the exact reconstruction of an object. In what follows, we shall term those objects which are uniquely determined by few projections as *uniqueness sets* or *uniquely characterized sets*. When there are only two projection directions, the structure of uniqueness sets is well understood via the classical work of Lorentz [57], and more recent work by Kuba [51, 52], Volcic [52], and Fishburn [25]. When 3 or more projections are available, the problem of uniquely characterizable sets remains, with the exception of some recent work by Gardner [26] and Giering [31], largely unexamined. In Chapter 5 we shall present some original contributions to the theory of reconstructability of binary *polygonal* regions from few projections.

The problem of reconstructability from few projections has also been studied in the digital realm, where the reconstruction of binary digital patterns from discrete projections in 2 directions has been considered. The digital problem with 2 directions is far less complicated than its continuous counterpart, and was effectively solved by Chang [14]. In 1949, Lorentz derived necessary and sufficient conditions on two non-negative, integrable functions $g(t, \theta_1)$ and $g(t, \theta_2)$ so that these are projections of a measurable plane (binary) set [57]. He further characterized projections that determine one, more than one, or no plane set. His results, although seminal, are

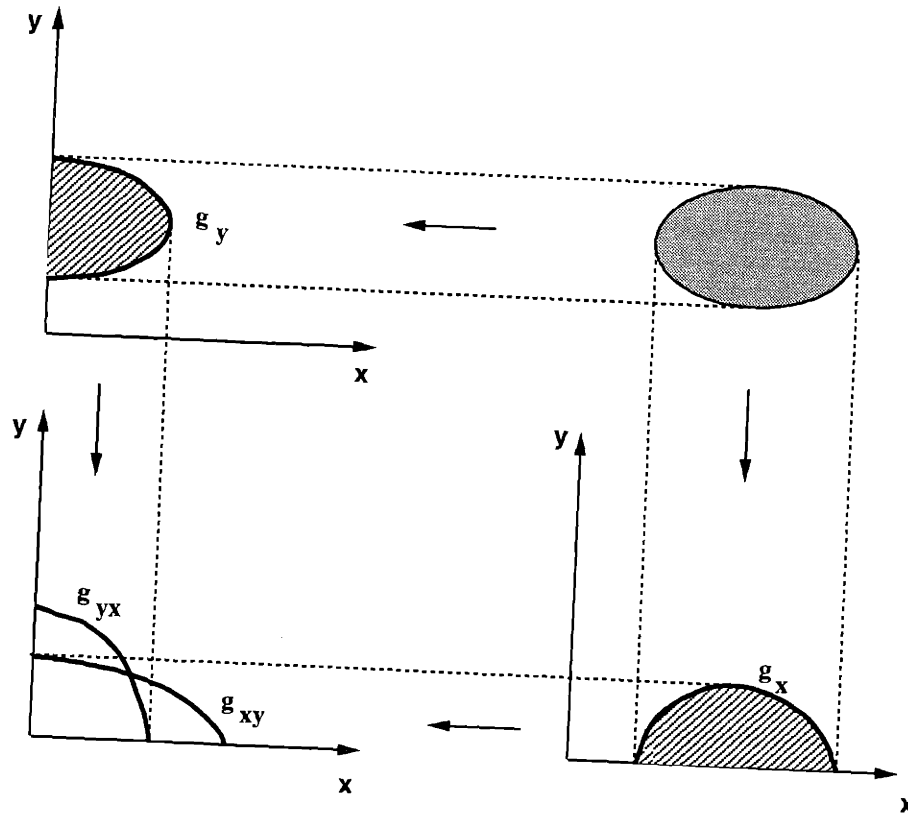


Figure 2-2: Projections of projections

mathematically quite involved and are not geometrically intuitive. The recent works of Fishburn, et al. [25], and Kuba and Volcic [52], have sought to reinterpret and generalize the work of Lorentz. In [52], Kuba and Volcic reformulate Lorentz's theorems and obtain explicit expressions for uniquely determined plane sets. To provide some insight into the null space of the Radon transform operator, we discuss some of these results in detail. These discussions will also put into perspective some of our own contributions which we will present in detail in Chapter 6.

Let $g_x(t)$ and $g_y(t)$ denote two projections of a binary object in two orthogonal directions which, without loss of generality, we take to be parallel to the coordinate axes x and y . Now consider these functions as objects in \mathcal{R}^2 and define $g_{xy}(t)$ as the projection of $g_x(t)$ in the direction of y . (See Figure 2-2). Similarly, we can define $g_{xy\dots}(t)$ for any number of indices. The following result is a reinterpretation of Lorentz's result and was stated in [52].

Theorem 3 (Kuba, Volcic) *Let $g_x(t)$ and $g_y(t)$ be non-negative, integrable functions such that*

$$\int_{-\infty}^{\infty} g_x(t) dt = \int_{-\infty}^{\infty} g_y(t) dt \quad (2.28)$$

Then

1. g_x and g_y uniquely specify a plane set iff

$$\int_0^c g_{xy}(t) dt = \int_0^c g_{yx}(t) dt \quad (2.29)$$

for all $c > 0$.

2. g_x and g_y can specify many plane sets iff

$$\int_0^c g_{xy}(t) dt \geq \int_0^c g_{yx}(t) dt \quad (2.30)$$

for all $c > 0$ and there is a $c > 0$ for which strict inequality holds.

3. g_x and g_y specify no plane sets (are inconsistent) iff

$$\int_0^c g_{xy}(t) dt < \int_0^c g_{yx}(t) dt \quad (2.31)$$

for some $c > 0$.

(It is worth noting that the functions g_{yxy} and g_{yx} look exactly alike. What distinguishes the two is that g_{yxy} is a function of the variable y while g_{yx} is a function of the variable x .)

Results with a more distinct geometric flavor have been obtained with the use of the concept of *switching components*. Formally we have

Definition 1 *Let \mathcal{P} be a compact set in the plane. Denote by $\mathcal{P}_{(a,0)}$ and $\mathcal{P}_{(0,b)}$ its horizontal and vertical translations by a and b , respectively. We say that a compact set \mathcal{O} in the plane admits a **switching component** if there exist two real numbers $a \neq 0$ and $b \neq 0$, and four sets \mathcal{P} , \mathcal{Q} , \mathcal{R} , and \mathcal{S} such that*

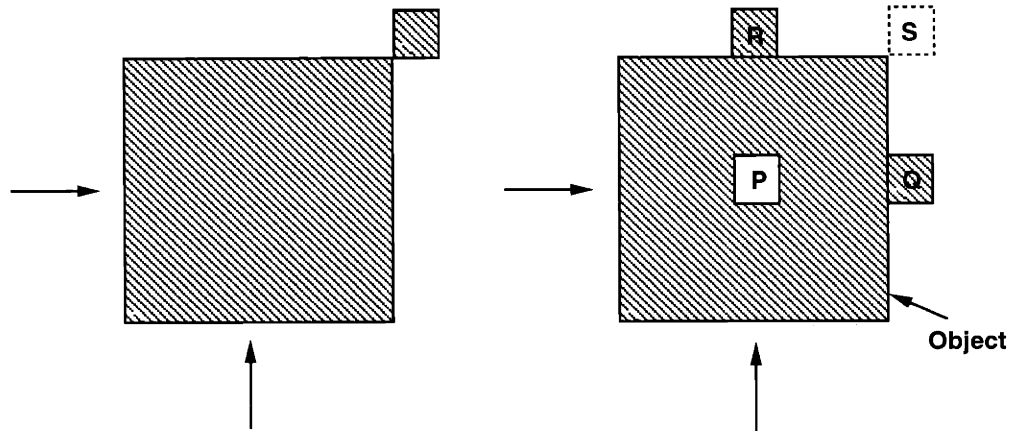


Figure 2-3: Two objects having identical vertical and horizontal projections

1.

$$Q \cup R \subseteq \mathcal{O} \quad (2.32)$$

2.

$$(\mathcal{P} \cup \mathcal{S}) \cap \mathcal{O} = \emptyset \quad (2.33)$$

3.

$$Q = \mathcal{P}_{(a,0)} \quad (2.34)$$

$$R = \mathcal{P}_{(0,b)} \quad (2.35)$$

$$S = \mathcal{P}_{(a,b)} \quad (2.36)$$

The shaded object on the right hand side of Figure 2-3 illustrates a set with a switching component and the sets defined above are also shown. The shaded object on the left hand side of Figure 2-3 shows an object which has identical projections from horizontal and vertical directions as the object on the right hand side of Figure 2-3. The corresponding picture of a switching component (i.e. a digital analog of the

set \mathcal{P}) for the digital case is either of the two patterns

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad (2.37)$$

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad (2.38)$$

Any digital object containing either of the two patterns is not uniquely determined by discrete projections in the horizontal and vertical directions [14]. Note that this is because the above two patterns have identical projections in the horizontal and vertical directions. The following result can now be formulated.

Theorem 4 ([52]) *A compact plane set of nonzero measure is uniquely determined by its two projections g_x and g_y in the orthogonal directions x and y iff it has no switching component.*

In [25], Fishburn et al. have derived necessary and sufficient algebraic conditions for uniquely determined sets in terms of *additive sets*. An additive set is defined as follows.

Definition 2 *A compact subset \mathcal{O} of the plane is said to be additive with respect to the directions x and y if there exist bounded measurable functions C_x and C_y where $C_x(x, y) = C_x(x)$ and $C_y(x, y) = C_y(y)$ depend only on the directions x and y respectively, such that*

$$(x_0, y_0) \in \mathcal{O} \Leftrightarrow C_x(x_0) + C_y(y_0) \geq 0 \quad (2.39)$$

For example, the unit disk in the plane is an additive set. Simply take $C_x(x) = \frac{1}{2} - x^2$ and $C_y(y) = \frac{1}{2} - y^2$. The following result is proved in [25] for sets in \mathcal{R}^2 .

Theorem 5 (Fishburn, et al.) *Let \mathcal{O} be a compact subset of the plane. Then the following conditions are equivalent.*

1. \mathcal{O} is uniquely determined by two projections in the directions x and y ,

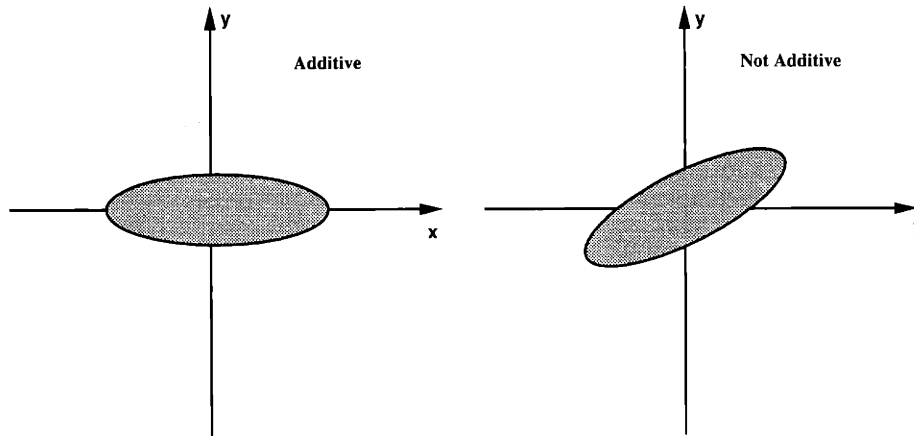


Figure 2-4: An additive vs. a non-additive ellipse with respect to the coordinate axes

2. \mathcal{O} is additive with respect to the directions x and y .

As an example of an additive set, consider the case of an ellipse centered at the origin and aligned with the orthogonal coordinate axes x and y , as illustrated in Figure 2-4. This ellipse is an additive set with respect to (x, y) . Any ellipse which is not aligned with these coordinate axes, however, is *not* an additive set with respect to these directions. One can easily verify this by writing down the equation for the tilted ellipse. It should be noted that the above result is *not* true in higher dimensions as pointed out in [48, 49].

The binary object reconstruction problem from few projections has also been studied when the directions of the two projections are arbitrary [51]. The problem has been transformed into the orthogonal case using a suitable affine transformation. The case where three projections are available is treated in some depth in [26] and [31]. An interesting consequence of a result proved by Giering in [31] is that *any* ellipse is uniquely determined with respect to *any three* projections taken from noncongruent directions. Karl [47] has also shown that an ellipsoid is uniquely determined by three *silhouette* projections from noncongruent directions. We are able to prove the same result in the framework of *moments* as discussed in Chapter 4. The following general result is due to Giering [31].

Theorem 6 (Giering) *Given a convex planar body \mathcal{O} , three directions may be cho-*

sen so that the projections of \mathcal{O} in those directions determine \mathcal{O} uniquely.

The operative word in the above result is *convex*. The assumption of convexity has lead to stronger results about objects uniquely determined from projections from few directions [28, 31]. The convexity assumption has also paved the way for results regarding reconstructability from fan-beam (FB) projections taken from a few points [52, 23]. The root of these investigations lies in the seminal work of Hammer [36]. We show in Chapter 6 that any simple binary polygonal object with N sides is uniquely determined by $2(N - 1)$ projections from distinct (but otherwise arbitrary) directions, thereby significantly strengthening existing results on the reconstruction of binary polygonal shapes from few projections.

Chapter 3

Reconstruction of Finitely Parameterized Objects

3.1 Introduction

In this chapter, we study the reconstruction of finitely parameterized objects from noisy (and possibly sparse) samples of their Radon-transform projections. The study of finitely parameterized objects is motivated by the fact that in many applications of tomographic signal processing, the final aim is often the extraction of a rather small set of geometrically-based features from a set of projection data. In such instances, a full pixel-by-pixel reconstruction of the object is a rather inefficient and non-robust approach. Classical reconstruction algorithms such as Filtered-backprojection (FBP) and Algebraic Reconstruction Techniques (ART) aim to reconstruct the intensity values of each pixel in the image with little regard to the quality and quantity of the available data [40]. Not surprisingly, they fail to produce acceptable reconstructions when the data is very noisy and/or sparse. This can be attributed to the fact that these classical algorithms in essence try to extract more information from the given data set than it actually contains. i.e. too many degrees of freedom are allowed in the reconstruction. In our effort to address this issue, we have proposed the use of simple *geometric* priors in the form of finitely parameterized objects. The assumption that the object to be reconstructed is finitely parameterized allows for the tomographic

reconstruction problem to be posed as a finite (relatively low-dimensional) parameter estimation problem. A more detailed survey of this literature was given in Chapter 1.

If we further assume, as we have done in the latter part of this chapter, that the number of such parameters is also an unknown, we can formulate the reconstruction problem as a Minimum Description Length estimation problem which provides for an automatic (data-driven) method for computing the optimal parameterized objects with the “best” number of parameters, given the data. This is, in essence, an information-theoretic criterion which gives us a direct way to estimate as many parameters as the information content of the data allows us to, and thus control the greed factor.

Our approach provides a statistically *optimal* Maximum Likelihood (ML) formulation for the direct recovery of *objects* from the projection data in the presence of *noise*. We also provide an automatic mechanism for identifying the statistically optimal number of object parameters, and thus information, in a given data set. The statistically optimal ML formulation leads to an optimization problem that is *non-linear* and has many local extrema. An appropriate initial guess is thus essential for its iterative solution. We thus provide a simple procedure to generate an appropriate initial guess, based on moment estimates of the object, formed from the original data. The direct estimation of the moments of the underlying image from measured projections is a topic which we will address in Chapter 4 in detail. As we will briefly allude to in this chapter, this moment estimation problem is a straightforward linear estimation problem. Our use of estimates of the underlying moments in the generation of an initial guess is based on a fundamental property of the Radon transform which relates the moments of a function $f(x, y)$ to the moments of its projections $g(t, \theta)$. As we will see in Chapter 4, this fundamental property will also play a key role in the development of reconstruction algorithms which we present in Chapters 5 and 6.

The organization of this Chapter is as follows. In Section 3.2 we introduce the basic definitions and assumptions and pose the general problem which we intend to

solve. We also discuss the particular statistical formulations of the reconstruction problem which we use. In Section 3.2.3 we discuss our novel technique for computing a good initial guess for the nonlinear optimization problems that result from our formulations. Section 3.3 contains basic performance results and robustness studies for various scenarios. Section 3.4 contains our conclusions.

3.2 The Reconstruction Problem

As defined in (2.1) of Chapter 2, the Radon transform of a function $f(x, y)$ defined over a compact domain of the plane \mathcal{O} is given by

$$g(t, \theta) = \iint_{\mathcal{O}} f(x, y) \delta(t - \omega \cdot [x, y]^T) dx dy. \quad (3.1)$$

Here we assume the existence of a set of parameters that uniquely specify the function f . It is the estimation of these parameters that is the concern of this chapter. Let us stack the set of parameters that uniquely define f in a vector V . In what follows, we shall then consider the function f and its set of defining parameters as interchangeable. We will assume throughout that the data available to us are discrete samples of g which are corrupted by Gaussian white noise of known intensity. In particular, our observations will be given by

$$Y_{i,j} = g(t_i, \theta_j, V^*) + w(t_i, \theta_j), \quad (3.2)$$

for $1 \leq i \leq n$, $1 \leq j \leq m$ where V^* is the true object we wish to reconstruct. The variables $w(t_i, \theta_j)$ are assumed to be independent, identically distributed (i.i.d.) Gaussian random variables with variance σ^2 . We will denote by Y the vector obtained by stacking all such observations.

In the classical framework, the parameter vector V consists of the pixel values of all pixels in the image to be reconstructed. For example, the parameter vector V here would contain over 16,000 elements for a 128×128 image. The FBP technique would estimate all of these parameters. For many applications, creation of this dense

visual field is not really necessary. For instance, in ocean acoustic tomography [65, 50], as discussed in Chapter 2, the size, location, and shape of a cold core ring are important quantities to be estimated within a gulf stream. Since a cold core ring can be considered a binary object, these parameters can be easily extracted from tomographic data directly without reconstruction of the entire pixel field. The classical approach, however, is to perform the full reconstruction of the image and then try to locate and measure objects, such as cold core rings, in the resulting image. If we only desire knowledge of the size and location of a localized occlusion, that is to say essentially three numbers, FBP is then estimating roughly 15,997 parameters too many!

Clearly, the estimation of all the pixel values is not only an inefficient use of the data, but in high noise and sparse data situations it can also produce spurious features known as “ghosts” that are bound to complicate the subsequent processing. In particular, such spurious features reflect the fact that we have *colored* the observation noise. This coloring of the noise, as reflected in reconstruction artifacts, may not be severe when we have high SNR and full-coverage data, yet can be quite limiting when these conditions are not met. In general, there may arise situations where the quality or quantity of the available data will simply not support the reliable estimation of all the parameters represented by the pixel values. In such cases, even though our interest may not be fundamentally geometric, our approach provides a rational and statistically precise way of reducing the degrees of freedom of the problem and hence of focusing the available information.

3.2.1 Maximum Likelihood Approach

In our approach, the original data in (3.2) is used to directly estimate the parameters V of a geometric parameterization in a statistically optimal way. The dimension of the parameter vector V is determined by the level of detail that one can extract from the sparse and noisy data. For clarity, we first consider the case where a fixed and known number of parameters is assumed. In this case, the Maximum Likelihood (ML) [100] estimate, \hat{V}_{ml} , of the parameter vector V is given by that value of V which

makes the observed data most likely. In particular, using the monotonicity of the logarithm:

$$\hat{V}_{ml} = \arg \max_V \log [P(Y|V)], \quad (3.3)$$

where $P(Y|V)$ denotes the conditional probability density of the observed data set Y given the parameter vector V . It is well-known that given the assumption that the data is corrupted by i.i.d. Gaussian random noise, the solution to the above ML-estimation problem is precisely equivalent to the following Nonlinear Least Squares Error (NLSE) formulation

$$\hat{V}_{ml} = \arg \min_V \sum_{i,j} \|Y_{i,j} - g(t_i, \theta_j, V)\|^2. \quad (3.4)$$

The formulation (3.4) shows that, in contrast to the linear formulation of classical reconstruction methods, the ML tomographic reconstruction approach, while yielding an optimal reconstruction framework, generally results in a highly nonlinear minimization problem. It is the nature of the dependence of g on the parameter vector V that makes the problem nonlinear. Being nonlinear, the problem is plagued by numerous extraneous local extrema, making the issue of computing a good initial guess for nonlinear optimization routines an important one. As we will discuss in Section 3.2.3, the geometric information carried in the projections can be conveniently extracted as the solution to a simple linear estimation problem in the form of moments of the underlying image. These moments can in turn be used to easily compute a good initial guess for the optimization algorithms.

Finally, note that if additional explicit geometric information is available in terms of a prior probabilistic description of the object vector V , then a Maximum-A-Posteriori estimate of V may be computed as follows:

$$\hat{V}_{map} = \arg \max_V \log [P(V|Y)] \quad (3.5)$$

In this work we concentrate on the ML problem given in (3.3) and its extensions, though application of our results to the MAP formulation is straightforward.

3.2.2 Minimum Description Length

In the previous ML discussion we assumed we had prior knowledge of the number of parameters describing the underlying object. Without this knowledge, we can consider the *Minimum Description Length* (MDL) principle [75]. In this approach, the cost function is formulated such that the global minimum of the cost corresponds to a model of least order that explains the data best. The MDL approach in essence extends the Maximum Likelihood principle by including a term in the optimization criterion that measures the model complexity. In the present context, the model complexity refers to the number of parameters used to capture the object in question. Whereas the ML approach maximizes the log likelihood function given in (3.3), the MDL criterion maximizes a modified log likelihood function, as follows:

$$\hat{V}_{MDL} = \arg \max_{V,N} \left\{ \log [P(Y|V)] - \frac{N}{2} \log(d) \right\}, \quad (3.6)$$

where $d = mn$ is the number of samples of $g(t, \theta)$ and N refers to the number of parameters defining the reconstruction. Roughly speaking, the MDL cost is proportional to the number of bits required to model the observed data set with a model of order N , hence the term *Minimum Description Length*. Under our assumed observation model (3.2), the MDL criterion (3.6) yields the following nonlinear optimization problem for the optimal parameter vector \hat{V}_{MDL} :

$$\hat{V}_{MDL} = \arg \min_N \min_V \left\{ \sigma^{-2} \sum_{i,j} \|Y_{i,j} - g(t_i, \theta_j, V)\|^2 + N \log(d) \right\}, \quad (3.7)$$

where the optimization is now performed over both V and the number of parameters N . Note that the solution of the inner minimization in (3.7) essentially requires solution of the original ML problem (3.3) or (3.4) for a sequence of values of N . Thus the optimization problem (3.7), being an extension of (3.4), is also highly nonlinear and routines to solve it will similarly require proper initializations to avoid being stuck in local minima.

In summary, our proposed approach, in its most general form, is a statistically

optimal technique for the extraction of specific geometric features or objects directly from the projection data with rational focusing of sparse and noisy information. Furthermore, via the MDL criterion it is equipped to compute the optimal number of features (parameters) that may be extracted from a given noisy and possibly sparse projection data set. While being a statistically optimal approach, the resulting optimization problems are highly nonlinear and require appropriate initialization for their solution. To this end we will present a simple robust method to generate such an initial estimate.

For the sake of focus in what follows, we will concentrate our attention on binary, polygonal objects. We define these as objects taking the value 1 inside a simple polygonal region and zero elsewhere.

Unless otherwise stated, we assume from here on that the *matrix*¹ V contains the vertices of an N -sided binary, polygonal region as follows:

$$V = \left[V_1 \mid V_2 \mid \cdots \mid V_N \right], \quad (3.8)$$

where $V_i = [x_i, y_i]^T$ denote the Cartesian coordinates of the i^{th} vertex of the polygonal region arranged in the counter-clockwise direction (See Figure 3-1). The tomographic reconstruction problem for these objects can now be stated as the problem of optimal estimation of the vertices V given noisy samples of the projections denoted by $Y_{i,j}$ as in (3.2).

3.2.3 Algorithmic Aspects – Computing A Good Initial Guess

Given the highly nonlinear nature of the dependence of the cost function in (3.4) on the parameters in V , it appears evident that given a poor initial condition, typical numerical optimization algorithms may converge to local minima of the cost function. Indeed, this issue is a major obstacle to the use of a statistically optimal though nonlinear approach such as given in (3.3) or (3.6). In this section we describe a

¹We use a matrix of parameters rather than a vector in what follows for notational ease of explanation in the algorithms to follow, though this is not essential.

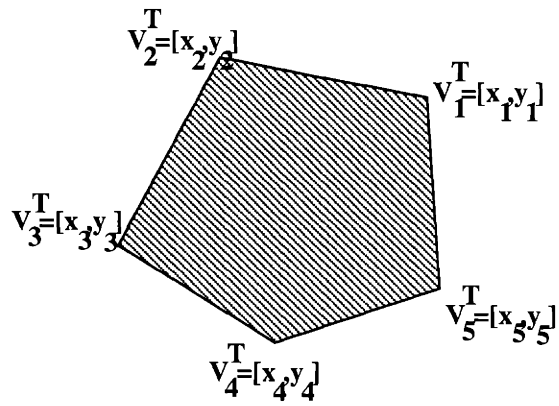


Figure 3-1: A binary, polygonal object

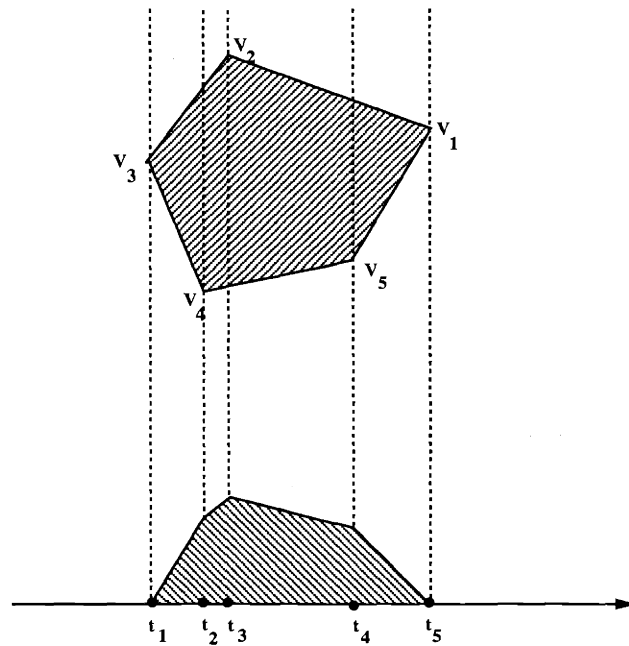


Figure 3-2: A projection of a binary, polygonal object

method for using the projection data to directly compute an initial guess that is sufficiently close to the true global minimum as to result in convergence to it in nearly all cases. We do this by estimating the moments of the object directly from the projection data and then using (some of) these moments to compute an initial guess.

In considering the use of moments as the basis for an initialization algorithm, one is faced with two important issues. The first is that, as we show in Chapter 4, estimating the moments of a function from its projections is a relatively easy task. The reconstruction of a function from a finite number of moments, however, is in general a highly ill-posed problem even when these moments are exactly known [95]. Furthermore, in our framework the moments are estimated from noisy data, and hence are themselves noisy. In fact, as higher and higher order moments are estimated, the error in the estimates of these moments becomes larger. Our approach avoids these moment related difficulties by using the moments only to guide an initial coarse estimate of the object parameters for subsequent use in solution of the nonlinear ML or MDL problems. This initial estimate, in turn, itself mitigates the difficulties associated with the nonlinearities of the optimal statistical approaches. In particular, the amount of computation involved in arriving at an initial guess using our moment-based method is far smaller than the amount of computation (number of iterations) required to converge to an answer given a poor initial guess, especially since a poor initial guess may converge to a local minimum and never reach the basin of the global minimum. Further, the parameterization of the objects serves to regularize and robustify the moment inversion process [95, 86, 70, 2].

Our method of using moments to generate an initial guess is based on the following set of observations. First, let μ_{pq} , $0 \leq p, q$ denote the moment of $f(x, y)$ of order $p + q$ as given by:

$$\mu_{pq} = \iint x^p y^q f(x, y) dx dy \quad (3.9)$$

In particular, note that the moments up to order 2 have the following physical relationships. The zeroth order moment μ_{00} is the area of the object, the first order

moments μ_{01} and μ_{10} are the coordinates of the center of mass of the object scaled by the area, and the second order moments μ_{02} , μ_{11} , μ_{20} are used to form the entries of the inertia matrix of the object. Thus these moments contain basic geometric information about object size, location, and elongation and orientation that, if available, could be used to guide our initialization of the nonlinear optimization problems (3.4) or (3.7). Our first aim then is to estimate them directly from the noisy projection data. To that end, as we showed in Chapter 2, we have

$$\int_{-\infty}^{\infty} g(t, \theta) t^k dt = \iint_{\mathcal{R}^2} f(x, y) [x \cos(\theta) + y \sin(\theta)]^k dx dy. \quad (3.10)$$

By expanding the integrand on the right hand side of (3.10), it becomes apparent that the moments of the projections are *linearly* related to the moments μ_{pq} of the object. In particular, specializing (3.10) to $k = 0, 1, 2$, and noting that $f(x, y)$ is an indicator function when the objects in question are binary, we arrive at the following relationships between μ_{pq} , $0 \leq p + q \leq 2$ and the projection $g(t, \theta)$ of $f(x, y)$ at each angle θ :

$$\mu_{00} = \int g(t, \theta) dt \equiv \mathcal{H}^{(0)}(\theta) \quad (3.11)$$

$$\begin{bmatrix} \cos(\theta) & \sin(\theta) \end{bmatrix} \begin{bmatrix} \mu_{10} \\ \mu_{01} \end{bmatrix} = \int g(t, \theta) t dt \equiv \mathcal{H}^{(1)}(\theta) \quad (3.12)$$

$$\begin{bmatrix} \cos^2(\theta) & 2 \sin(\theta) \cos(\theta) & \sin^2(\theta) \end{bmatrix} \begin{bmatrix} \mu_{20} \\ \mu_{11} \\ \mu_{02} \end{bmatrix} = \int g(t, \theta) t^2 dt \equiv \mathcal{H}^{(2)}(\theta) \quad (3.13)$$

Thus if we have projections at three or more distinct, known angles we can estimate the moments of up to order 2 of the object we wish to reconstruct. As can be seen from (3.11), (3.12), and (3.13), and as described in Chapter 4, the computation of these moments is a *linear* calculation, making their estimation from projections straightforward. Since, in general, many more than three projections are available, the estimation of these moments determining the area, center, and inertia axes of

the object is overdetermined. The result is a robustness to noise and data sparsity through a reduction in the noise variance of the estimated moments. In particular, we can stack the moments $\mathcal{H}^{(k)}(\theta_i)$ we obtain from the projections at each angle θ_i to obtain the following overall equations for the μ_{pq} :

$$\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \mu_{00} = \begin{bmatrix} \mathcal{H}^{(0)}(\theta_1) \\ \vdots \\ \mathcal{H}^{(0)}(\theta_m) \end{bmatrix} \quad (3.14)$$

$$\begin{bmatrix} \cos(\theta_1) & \sin(\theta_1) \\ \vdots & \vdots \\ \cos(\theta_m) & \sin(\theta_m) \end{bmatrix} \begin{bmatrix} \mu_{10} \\ \mu_{01} \end{bmatrix} = \begin{bmatrix} \mathcal{H}^{(1)}(\theta_1) \\ \vdots \\ \mathcal{H}^{(1)}(\theta_m) \end{bmatrix} \quad (3.15)$$

$$\begin{bmatrix} \cos^2(\theta_1) & 2 \sin(\theta_1) \cos(\theta_1) & \sin^2(\theta_1) \\ \vdots & \vdots & \vdots \\ \cos^2(\theta_m) & 2 \sin(\theta_m) \cos(\theta_m) & \sin^2(\theta_m) \end{bmatrix} \begin{bmatrix} \mu_{20} \\ \mu_{11} \\ \mu_{02} \end{bmatrix} = \begin{bmatrix} \mathcal{H}^{(2)}(\theta_1) \\ \vdots \\ \mathcal{H}^{(2)}(\theta_m) \end{bmatrix} \quad (3.16)$$

Using these equations we can easily calculate the maximum likelihood estimates of the moments of the object μ_{pq} , $0 \leq p + q \leq 2$. For a detailed discussion of the problem of optimal estimation of moments from projections, see Chapter 4. Let us denote these ML moment estimates by $\hat{\mu}_{pq}$.

Now that we have estimates of the moments of up to order 2 of the object, and thus estimates of its basic geometric structure, we need to convert this information into a suitable object for use in initializing the nonlinear optimization problem (3.4) or (3.7). The initial guess algorithm outlined next uses these ML estimates of the low-order moments $\hat{\mu}_{pq}$, obtained from the noisy projection data, to produce a polygon which has moments up to order 2 which are close to (or in some cases equal to) those which were estimated from the projection data. The resulting polygon, which will be used as our initialization, should thus have the same basic geometric structure as the underlying object.

Recall that in this process of generating an initial object from the moment data we want to avoid the difficulties usually associated with obtaining an object from a

set of its moments. For this reason, the initial polygon we will use is simply obtained as the affine transformation of a reference object $V_{\text{ref}}(N)$, which is a centered regular N -gon of unit area. For a given choice of number of sides N , the reference object we use is given by:

$$V_{\text{ref}}(N) = \frac{1}{\sqrt{\frac{N}{2} \sin(\frac{2\pi}{N})}} \left[\begin{array}{c|c|c|c} \cos(0) & \cos(\frac{2\pi}{N}) & \cdots & \cos(\frac{2\pi(N-1)}{N}) \\ \sin(0) & \sin(\frac{2\pi}{N}) & \cdots & \sin(\frac{2\pi(N-1)}{N}) \end{array} \right] \quad (3.17)$$

The affine transformation of this reference object, which will be generated from the estimated moment set, consists of a uniform scaling, a stretching along the coordinate axes, a rotation, and finally a translation, and simply corresponds to the following transformation of the underlying spatial coordinates of the reference object

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = L \begin{bmatrix} x \\ y \end{bmatrix} + C. \quad (3.18)$$

In particular, given the form of $V_{\text{ref}}(N)$ in (3.17), this yields the following equation for the family of possible initial objects V_{init} :

$$V_{\text{init}} = L V_{\text{ref}}(N) + \left[C \mid \cdots \mid C \right] \quad (3.19)$$

The set of all such affine transformations of $V_{\text{ref}}(N)$ we term the **affinely regular** N -gons. In the absence of noise, the initial guess algorithm we detail will exactly match the given estimated moments if the underlying object itself happens to be affinely regular. As shown in Appendix 3-B, if the underlying object is not affinely regular itself, the algorithm will not necessarily produce an N -gon exactly matching its moments, even in the absence of noise, though it will usually be close. Of course, in the presence of noise, the estimated moments themselves are not exact and thus, while we would hope to get close, our resulting initial N -gon will never exactly match the true moments of the underlying object anyway.

Given that we will restrict ourselves to initial objects of the form (3.19), let us consider how we might choose the parameters of the transformation L and C to match

a given estimated moment set $\hat{\mu}_{pq}$, $0 \leq p+q \leq 2$. Using (3.19) and (3.17) to calculate the moments of V_{init} up to order 2 we obtain the following relationships:

$$\mu_{00}(V_{\text{init}}) = |\det(L)| \quad (3.20)$$

$$\begin{bmatrix} \mu_{10}(V_{\text{init}}) \\ \mu_{01}(V_{\text{init}}) \end{bmatrix} = |\det(L)| C \quad (3.21)$$

$$\begin{bmatrix} \mu_{20}(V_{\text{init}}) & \mu_{11}(V_{\text{init}}) \\ \mu_{11}(V_{\text{init}}) & \mu_{02}(V_{\text{init}}) \end{bmatrix} = |\det(L)| (k_N LL^T + CC^T) \quad (3.22)$$

where $\mu_{pq}(V_{\text{init}})$ is the pq -th moment of V_{init} and $k_N = 1/(4N \tan(\pi/N))$. For explicit derivations of these relations, see Appendix 3-B. Thus to match $\mu_{00}(V_{\text{init}})$, $\mu_{10}(V_{\text{init}})$, $\mu_{01}(V_{\text{init}})$ with their estimated counterparts the first two conditions require:

$$|\det(L)| = \hat{\mu}_{00} \quad (3.23)$$

$$C = \frac{1}{\hat{\mu}_{00}} \begin{bmatrix} \hat{\mu}_{10} \\ \hat{\mu}_{01} \end{bmatrix} \quad (3.24)$$

The first condition simply corresponds to a scaling of $V_{\text{ref}}(N)$ so that its area matches the estimated one. The second condition shows that the term C in the transformation (3.18) should correspond to a translation of $V_{\text{ref}}(N)$ to the estimated center of mass of the object. These two conditions ensure that we match the estimated area and center of mass location.

Now, after some manipulation, (3.22) implies that to match the estimated second order moments $\hat{\mu}_{pq}$, $p+q=2$, we must have:

$$LL^T = \frac{1}{\hat{\mu}_{00} k_N} \hat{\mathcal{I}} \quad (3.25)$$

where $\hat{\mathcal{I}}$ is the matrix of estimated central moments defined by:

$$\hat{\mathcal{I}} = \begin{bmatrix} \hat{\mu}_{20} & \hat{\mu}_{11} \\ \hat{\mu}_{11} & \hat{\mu}_{02} \end{bmatrix} - \frac{1}{\hat{\mu}_{00}} \begin{bmatrix} \hat{\mu}_{10} \\ \hat{\mu}_{01} \end{bmatrix} \begin{bmatrix} \hat{\mu}_{10} & \hat{\mu}_{01} \end{bmatrix} \quad (3.26)$$

In particular, this condition implies another constraint on $\det(L)$ independent of (3.23), which we will not, in general, be able to satisfy. Specifically, a necessary condition for finding an L satisfying both (3.25) and (3.23) is that:

$$\det(\hat{\mathcal{I}}) = \hat{\mu}_{00}^4 k_N^2 \quad (3.27)$$

where the expression on the left is the determinant of the estimated central second moments. Actually the condition (3.27) is also sufficient, as shown in Appendix 3-B. Clearly, this condition will not, in general, be satisfied and we will be unable to exactly match the estimated second moments. In fact, the objects that do meet this constraint, and thus whose moments we can exactly match, are precisely the elements of the set of affinely regular N -gons. Geometrically, this situation reflects the limitation of our restricted object class (3.19), i.e. the set of affinely regular N -gons. Within this class, for a given object area we are constrained as to the “size” of the corresponding inertia matrix we may have, where inertia size is measured as the product of the principle central inertia moments (eigenvalues of the central inertia matrix). For example, whereas the objects of our class will always be convex polygons, for a given area we can always obtain nonconvex objects of greater inertia by “moving area outward,” as in a dumbbell.

The condition (3.25) can also be viewed as implying a different scaling on L , needed to obtain a perfect match to the inertia condition. In general, we thus have a choice of picking this scaling of L to satisfy either the inertia condition (3.25) or the area condition (3.23). Since the area condition (3.23) is both a more robustly estimated and a more fundamental geometric quantity, we choose to enforce this condition in the algorithm to follow. We then choose L so that the resulting central inertia matrix of V_{init} has the same principle axes directions and has its principle inertias in the same ratio as those estimated from the data as found in $\hat{\mathcal{I}}$. We accomplish these goals by using a square root of the matrix $\hat{\mathcal{I}}$ normalized to have unit determinant for the form of L , then scaling the result to match the determinant condition (3.23). Thus we sacrifice overall scaling of the inertia matrix in favor of matching of the

estimated area. We have proved, in Result 4 of Appendix 3-B, that the normalized error incurred in matching the inertia matrix of a general n -gon by an affinely-regular n -gon is always less than 100%.

Collecting the above steps and reasoning, the overall algorithm is given by the following:

Algorithm 1 (Initial Guess)

1. Compute the optimal (ML) estimates of the moments up to order 2 ($\hat{\mu}_{00}$, $\hat{\mu}_{10}$, $\hat{\mu}_{01}$, $\hat{\mu}_{20}$, $\hat{\mu}_{11}$, and $\hat{\mu}_{02}$) from the raw projection data using (3.14)–(3.16).
2. Construct an N -sided regular polygon centered at the origin with vertices chosen as the scaled roots of unity in counter-clockwise order so that they lie on the circle of radius $1/\sqrt{\frac{N}{2} \sin(\frac{2\pi}{N})}$. This polygon has unit area and is given by.

$$V_{\text{ref}}(N) = \frac{1}{\sqrt{\frac{N}{2} \sin(\frac{2\pi}{N})}} \begin{bmatrix} \cos(0) & \cos(\frac{2\pi}{N}) & \cdots & \cos(\frac{2\pi(N-1)}{N}) \\ \sin(0) & \sin(\frac{2\pi}{N}) & \cdots & \sin(\frac{2\pi(N-1)}{N}) \end{bmatrix} \quad (3.28)$$

3. Compute the translation C , obtained as the estimated object center of mass:

$$C = \frac{1}{\hat{\mu}_{00}} \begin{bmatrix} \hat{\mu}_{10} \\ \hat{\mu}_{01} \end{bmatrix} \quad (3.29)$$

4. Form the estimated central inertia matrix $\hat{\mathcal{I}}$ as follows:

$$\hat{\mathcal{I}} = \begin{bmatrix} \hat{\mu}_{20} & \hat{\mu}_{11} \\ \hat{\mu}_{11} & \hat{\mu}_{02} \end{bmatrix} - \frac{1}{\hat{\mu}_{00}} \begin{bmatrix} \hat{\mu}_{10} \\ \hat{\mu}_{01} \end{bmatrix} \begin{bmatrix} \hat{\mu}_{10} & \hat{\mu}_{01} \end{bmatrix} \quad (3.30)$$

The entries of this matrix are the (estimated) central second moments.

5. If $\hat{\mathcal{I}}$ is not positive definite, set $L = \sqrt{\mu_{00}}I$ to match areas, and goto step 8. Otherwise proceed to step 6.

6. Perform an eigendecomposition of the normalized matrix $\hat{\mathcal{I}}$ as follows:

$$\frac{\hat{\mathcal{I}}}{\sqrt{\det(\hat{\mathcal{I}})}} = U \begin{bmatrix} \lambda & 0 \\ 0 & 1/\lambda \end{bmatrix} U^T \quad (3.31)$$

where we have assumed that the eigenvalues are arranged in descending order and U is orthogonal so that $\det(U) = \pm 1$. Note that the eigenvalues are reciprocals of each other since we have scaled the left hand side so that its determinant is 1.

7. Form the linear transformation L as a scaled square root of $\hat{\mathcal{I}}$ as follows:

$$L = \sqrt{\hat{\mu}_{00}} U \begin{bmatrix} \sqrt{\lambda} & 0 \\ 0 & 1/\sqrt{\lambda} \end{bmatrix}. \quad (3.32)$$

Note that $\det(L) = \hat{\mu}_{00}$ as desired. Depending on whether U is a pure rotation or a rotation followed by a reflection, it will have determinant $+1$ or -1 , respectively.

8. The initial guess V_{init} is now obtained by applying the scaling, stretching, and rotation transformation L and the translation C to the reference object $V_{\text{ref}}(N)$ via the coordinate transformation $[x', y']^T = L[x, y]^T + C$. Because of the form of $V_{\text{ref}}(N)$ this operation yields:

$$V_{\text{init}} = L V_{\text{ref}}(N) + \left[C \mid \dots \mid C \right] \quad (3.33)$$

Note that the eigenvalue λ of the unit determinant matrix calculated in step 7 gives the eccentricity of the underlying object while the corresponding eigenvector give its orientation. Also note that, in the presence of noise, the estimated central inertia matrix for the object $\hat{\mathcal{I}}$ may not be strictly positive definite and hence may not correspond to the inertia matrix of any object at all. In such instances, the algorithm refrains from the use of these moments of order 2 and computes an initial guess based only on the estimated area and center of mass.

Also note that if the matrix L , computed above, is replaced by $L' = LT$ for any orthogonal T , the resulting quantity $L'L'^T$ satisfies

$$LL^T = L'L'^T. \quad (3.34)$$

Hence, although the initial guess generated by the above algorithm is unambiguous and unique in the sense that the square root of $\widehat{\mathcal{I}}$ is unique, an infinity of other initial guesses having the same moments up to order 2 may be generated by replacing L by LT and allowing T to range over the set of all 2×2 orthogonal transformations. A precise characterization of the set of all affinely regular N -gons with the same moments up to order 2 is, in fact, given in the following result which we prove in Appendix 3-B.

Result 2 *Consider the 2×2 symmetric, strictly positive definite matrix \mathcal{I} with $\det(\mathcal{I}) = k_N^2 \mu_{00}^4$. Let*

$$E_O = \frac{\alpha_N}{\mu_{00} k_N} \mathcal{I} \quad (3.35)$$

$$E_I = \frac{\beta_N}{\mu_{00} k_N} \mathcal{I} \quad (3.36)$$

where

$$\alpha_N = 1 / \left(\frac{N}{2} \sin\left(\frac{2\pi}{N}\right) \right) \quad (3.37)$$

$$\beta_N = 1 / (N \tan(\pi/N)) \quad (3.38)$$

Define the two ellipses \mathcal{E}_O and \mathcal{E}_I , respectively, as follows

$$\mathbf{z}^T E_O^{-1} \mathbf{z} = 1 \quad (3.39)$$

$$\mathbf{z}^T E_I^{-1} \mathbf{z} = 1. \quad (3.40)$$

For a fixed integer $N \geq 3$ let C_1 denote the set of all N -gons whose first three moment sets are μ_{00} , 0 , 0 , \mathcal{I} ; and let C_2 denote the set of all N -gons with vertices on \mathcal{E}_O and

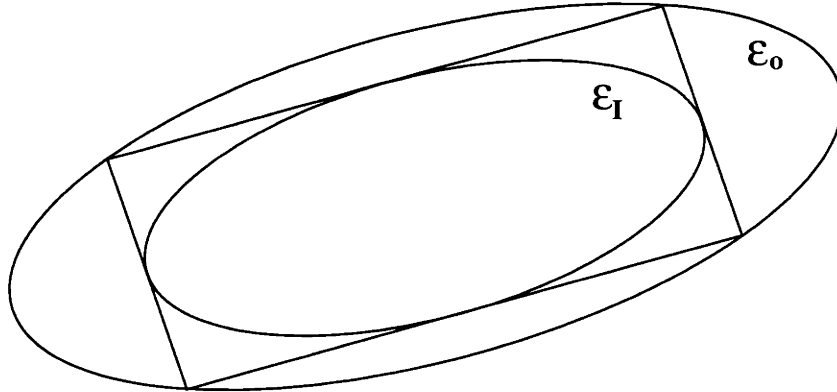


Figure 3-3: Illustration of Result 2

with sides tangent to \mathcal{E}_1 at their midpoints. Then

$$C_1 = C_2 \tag{3.41}$$

This result states that the class of all affinely regular N -gons with a given set of moments up to order 2 is given by the class of N -gons whose vertices are on a fixed ellipse and whose sides are tangent to a second ellipse which is confocal with the first. (See Figure 3-3 for an example). The ellipses are uniquely determined by the value of the given moments. In order to simplify the Initial Guess algorithm, we do not search further over this set of all possible initial guesses. We simply use the output of the Initial Guess Algorithm described above as the starting guess for our nonlinear optimization routines.

This algorithm begs the question of how one would incorporate higher order moments in this process to perhaps construct a better initial guess. In Chapter 5 we describe how to reconstruct an image from its estimated moments of higher order; and in Chapter 6, we describe how to reconstruct a polygonal region from its estimated moments of higher order using array processing techniques.

3.3 Experimental Results

In this section we present some performance studies of our proposed algorithm with simple polygonal objects as prototypical reconstructions. One may expect that our algorithms work best when the underlying object (that from which the data is generated) is itself a simple binary polygonal shape. While this is true, we will also show that our algorithms perform consistently well even when the underlying objects are complex, non-convex, and non-polygonal shapes.

First we demonstrate reconstructions based on the ML criterion. In these reconstructions we use the parameters of the true polygon as the initial guess to ensure that the solution produced by the algorithm is actually the ML estimate, i.e., at the global minimum of the ML cost function. Typical reconstructions are shown along with average performance studies for a variety of noise and data sampling scenarios. In particular, we show that the performance of our algorithms is quite robust to noise and sparsity of the data, significantly more so than classical reconstruction techniques. To demonstrate this point, reconstructions using our techniques and the classical FBP are provided.

Next we demonstrate how the MDL criterion may be used to optimally estimate the number of parameters (sides) N directly from the data. We solve these MDL problems by solving the ML problem for a sequence of values of N . To initialize each of these ML problems, a regular polygon of the desired number of sides with the same area and centroid is used. This initialization tries to ensure us that the actual ML solution corresponding to the number of sides in question is most likely being found. The robustness of the MDL approach and its ability to capture the shape information in noisy data when the underlying object is not polygonal is also shown through polygonal reconstruction of more complicated shapes.

In order to quantify some measure of performance of our proposed reconstruction algorithms, we first need to define an appropriate notion of signal-to-noise ratio (SNR). We define the **SNR per sample** as

$$\text{SNR} = 10 \log_e \frac{\sum_{i,j} g^2(t_i, \theta_j)/d}{\sigma^2}, \quad (3.42)$$

where $d = m \times n$ is the total number of observations, and σ^2 is the variance of the i.i.d. noise w in the projection observations (3.2) ².

In all our simulations the reconstruction error is measured in terms of the percent Hausdorff distance [6] between the estimate and the true polygon or shape. The Hausdorff metric is a proper notion of “distance” between two nonempty compact sets and it is defined as follows. Let $d(p_1, S)$ denote the minimum distance between the point p_1 and the compact set S :

$$d(p_1, S) = \inf\{\|p_1 - p\| \mid p \in S\}. \quad (3.43)$$

Define the ϵ -neighborhood of the set S as

$$S^{(\epsilon)} = \{p \mid d(p, S) \leq \epsilon\}. \quad (3.44)$$

Now given two non-empty compact sets, S_1 and S_2 , the Hausdorff distance between them is defined as:

$$\mathcal{H}(S_1, S_2) = \inf\{\epsilon \mid S_1 \subset S_2^{(\epsilon)} \text{ and } S_2 \subset S_1^{(\epsilon)}\} \quad (3.45)$$

In essence, the Hausdorff metric is a measure of the largest distance by which the sets S_1 and S_2 differ. The *percent* Hausdorff distance between the true object S and the reconstruction \hat{S} is now defined as

$$\text{Percent Error} = 100\% \times \frac{\mathcal{H}(\hat{S}, S)}{\mathcal{H}(\emptyset, S)} \quad (3.46)$$

where \emptyset denotes the set composed of the single point at the origin, so that if S contains

²Note that the logarithm in the definition of SNR is the natural logarithm. Although this is unconventional, we have consistently used the same definition of SNR through the thesis. The usual (dB) definition of the SNR with base 10 logarithm can be obtained from our definition by dividing by 2.3

the origin, $\mathcal{H}(\emptyset, S)$ is the maximal distance of a point in the set to the origin and thus a measure of the set's size.

3.3.1 ML Based Reconstruction

Here we present examples and performance analyses of the ML based reconstruction method (3.4). For these experiments an initial condition equal to the true object was used to ensure us of obtaining the actual ML estimate.

Sample Reconstructions

In Figures 3-4 and 3-5 are optimal reconstructions of a triangle and a hexagon, respectively, based on the ML criterion. The true polygon, in each case is depicted in solid lines while the estimate is shown in dashed lines. For both objects, 1000 noisy projection samples were collected in the form of 50 equally spaced projections in the interval $(0, \pi]$ ($m=50$), and 20 samples per projection ($n=20$). The field of view (extent of measurements in the variable t) was chosen as twice the maximum width of the true object in each case. For each of these data sets the variance of the noise in (3.2) was set so that the SNR given by (3.42) was equal to 0. The typical behavior of the optimal ML based reconstructions in the projection space can be seen in Figure 3-6, which corresponds to the hexagon of Figure 3-5. The top image of this figure shows the underlying projection function $g(t_i, \theta_j, V^*)$ of (3.2) for the hexagon, while the middle image shows the noisy observed data $Y_{i,j}$. The object is difficult to distinguish in the noise in the image. The bottom image shows the reconstructed projections corresponding to the optimal estimate $g(t_i, \theta_j, \hat{V})$, which is virtually indistinguishable from those corresponding to the true object. Figure 3-7 shows the best³ FBP reconstruction of the hexagon used in Figure 3-5 based on 4096 projection samples of the same SNR (0) (64 angles with 64 samples per angle). For comparison, the reconstruction using our algorithm is shown in Figure 3-8.

³The fact that this is the "best" FBP is somewhat subjective. We tried many different filters and visually, the best reconstruction was obtained with a Butterworth filter of order 3 with 0.15 normalized cutoff frequency.

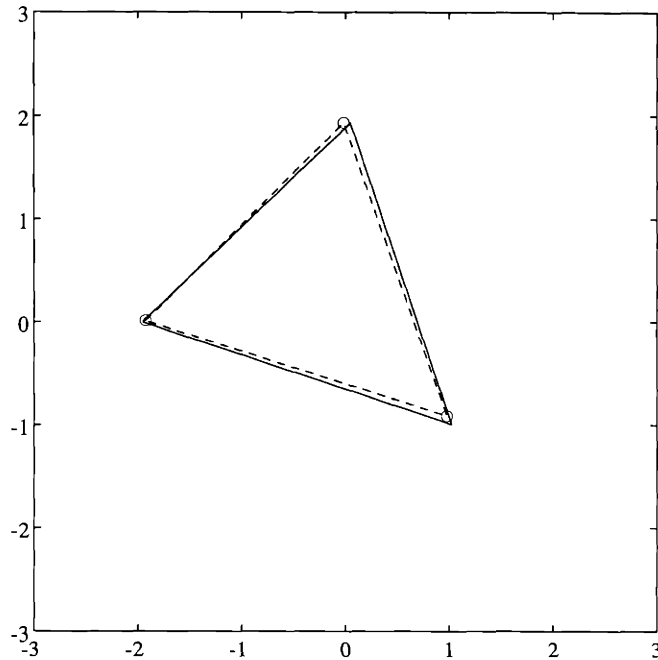


Figure 3-4: Sample reconstruction of a triangle at 0 SNR 50 views, 20 samples/view: True(-), Reconstruction(- -)

In contrast to the ML-based reconstruction, the details of the hexagon are corrupted in the FBP reconstruction. In addition, there are spurious features in the FBP reconstructions and perhaps most importantly, to extract a binary object from the FBP reconstruction, we would need to threshold the image or perform edge detection on it. Neither of these postprocessing steps are easily interpretable in an optimal estimation framework and, of course, they incur even more computational costs.

Note that the number of samples per projection used in this reconstruction is actually *more* than the number used to produce the ML-based reconstruction in Figure 3-5. The increase in sampling was necessary because CBP produces severe artifacts if the number of views exceeds the number of samples per view. This behavior is called the “incomplete data” effect [40]. The ML approach has no such difficulties, as we will see in the next section where we examine performance.

Effect of Noise on Performance

The average performance of the ML based reconstructions is presented through several Monte-Carlo studies. Again, for these experiments an initial condition equal to the

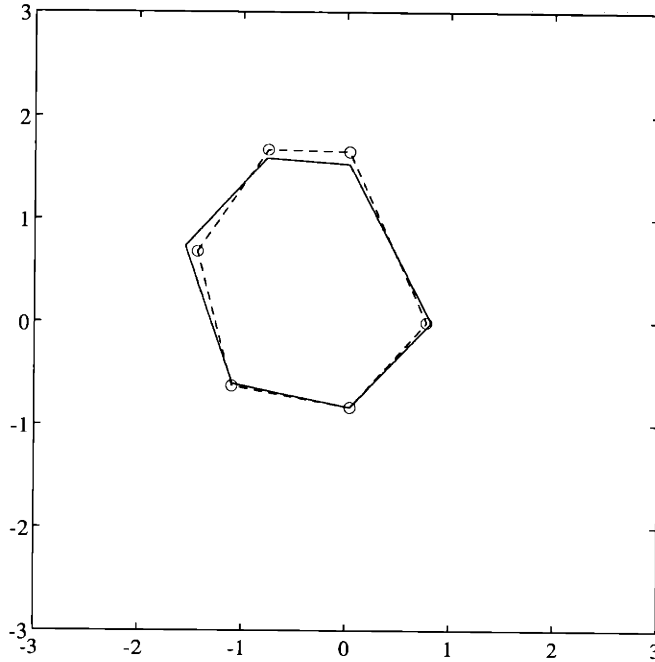


Figure 3-5: Sample reconstruction of a hexagon at 0 SNR 50 views, 20 samples/view : True(-), Reconstruction(- -)

true object was used in each case to ensure us of obtaining the actual ML estimates. The first study establishes average reconstruction errors at various SNR's for a fixed number of data samples. The purpose of these simulations is to demonstrate that the ML-based reconstructions are robust to the quality of the data used for a wide range of SNR's. The same two polygons as in Figures 3-4 and 3-5 were chosen as the underlying objects. Again, in each case, 1000 samples of the projections of these objects were collected in the form of 50 equally spaced projections in the interval $(0, \pi]$ ($m=50$), and 20 samples per projection ($n=20$) while the field of view (extent of measurements in the variable t) was chosen as twice the maximum width of the object in each case. The samples $g(t_i, \theta_j, V^*)$ were then corrupted by Gaussian white noise $w(t_i, \theta_j)$ of different intensities to yield data sets at several SNR's. At each SNR, 100 reconstructions were done using independent sample paths of the corrupting white noise. The average reconstruction error was then computed and is displayed against the SNR in Figure 3-9. The error bars denote the 95% confidence intervals for the computed mean values.

The percent error in these reconstructions increases with decreasing SNR, as one

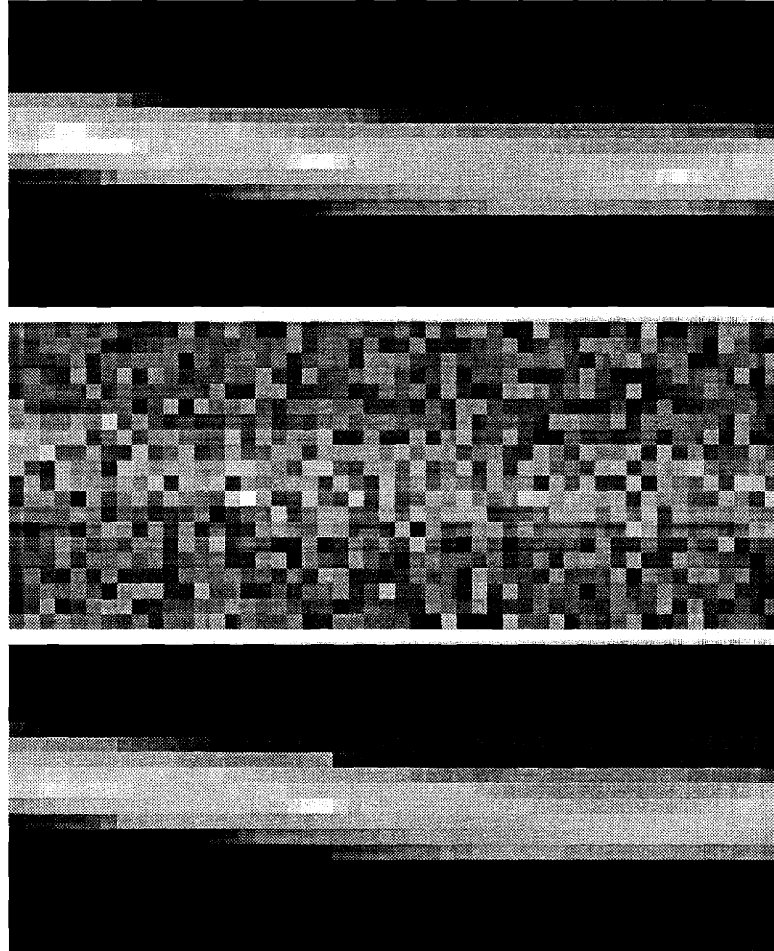


Figure 3-6: From Top to Bottom: Sinograms with 50 projections and 20 samples per projection of I) Noiseless Hexagon, II) Noisy data at SNR of 0 dB III) Reconstructed Hexagon. In each of these images, the horizontal axis is θ , the vertical axis is t , and the intensity values are the values of the corresponding projections mapped to the grayscale range of $[0, 255]$

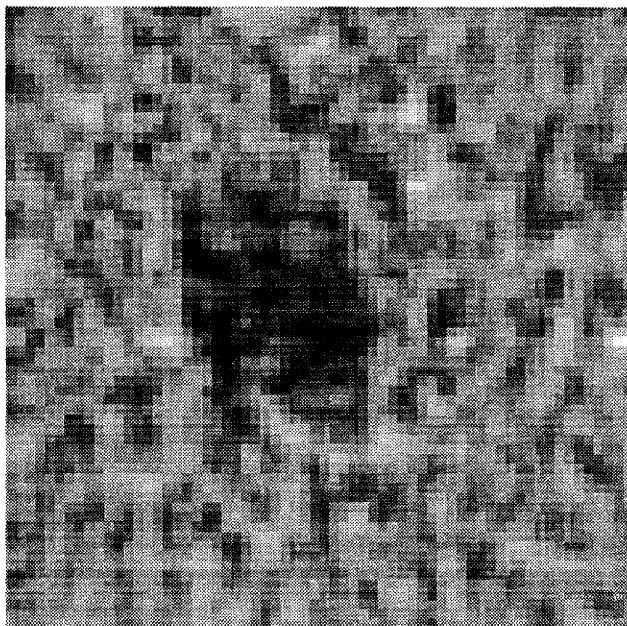


Figure 3-7: Sample reconstruction of a hexagon at 0 SNR using FBP: 64 views, 64 samples per view

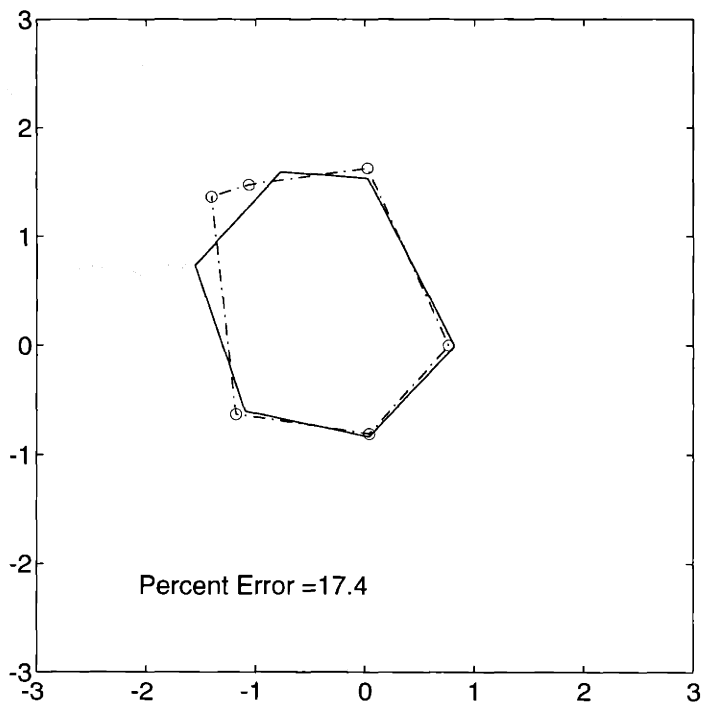


Figure 3-8: Sample reconstruction of a Hexagon at 0 SNR 64 views, 64 samples/view: True(-), Reconstruction(- -)

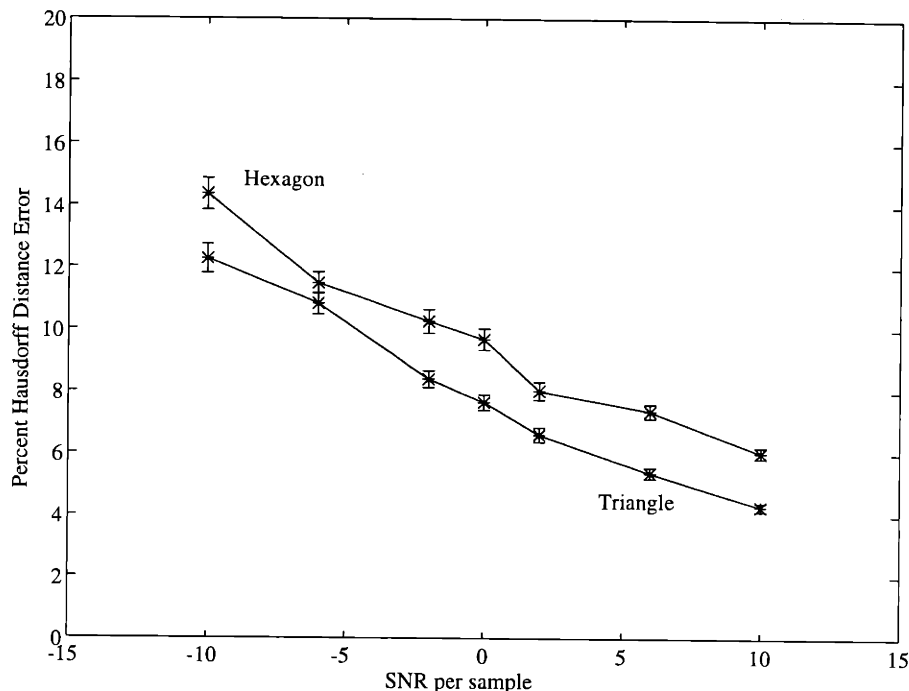


Figure 3-9: Mean performance curves for ML reconstructions of a triangle and a hexagon

would expect. In fact, the graph shows that, at least in the examined SNR range of -10 to +10, the relation between between percent error and SNR is roughly linear in the cases of the triangle and the hexagon. This suggests that the performance of our algorithm does not degrade very quickly with decaying SNR, demonstrating the robustness to noise of such object-based optimal ML estimates.

Effect of Sampling on Performance

Here the performance of our ML-based estimates with respect to both the number of available data samples and their distribution is studied. One would naturally expect that as the number of available data points decreases, the reconstruction error should increase. The main aim of these simulations is to demonstrate that the ML-based reconstructions are robust to both the quantity and the distribution of data over a wide range of SNR's. In particular, reasonable estimates are produced even with a drastic reduction of data and, unlike the behavior seen in FBP reconstructions, the ML estimates display no catastrophic degradation as the samples per angular view are reduced. The *relative* sensitivity of the ML estimates to density of samples of

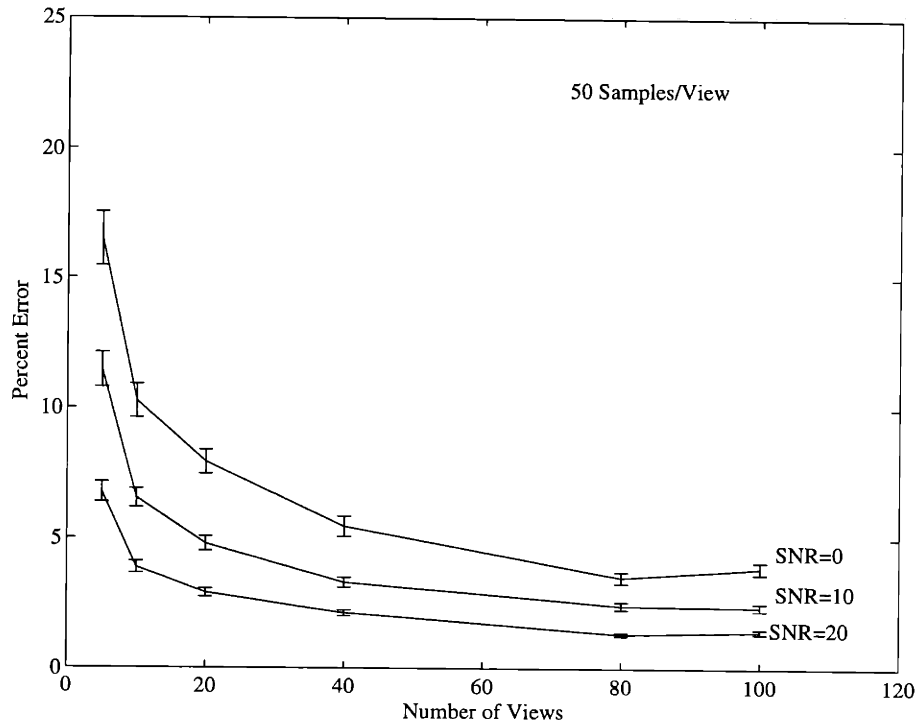


Figure 3-10: Performance as a function of Number of Views

$g(t, \theta, V^*)$ in t and θ is also discussed, providing information of use for the design of sampling strategies.

The true hexagon used in Figure 3-5 was again used as the underlying object. As before, an initial condition equal to the true object was used for each of experiments to ensure us of obtaining the actual ML estimates. A series of Monte-Carlo simulations (50 runs for each sampling configuration) were then performed at various SNR's to observe the effect of sparse projections and sparse sampling in each projection. In Figure 3-10, the percent Hausdorff reconstruction error is plotted versus the number of angular views for SNR's of 0, 10, and 20, while the number of samples per view was fixed at 50. With a modest 50 samples per view, all three curves fall below 10% reconstruction error when the number of views is greater than about 10. This is only 500 total observations, many of which do not contain the object at all (since the field of view is twice as large as the object). Furthermore, as the number of angular views is decreased from 100 to 10, only a marginal increase in the reconstruction error is observed. These observations testify to the robustness of optimal ML estimates with respect to the number of views.

In Figure 3-11, the dual case is presented. In this figure the percent Hausdorff

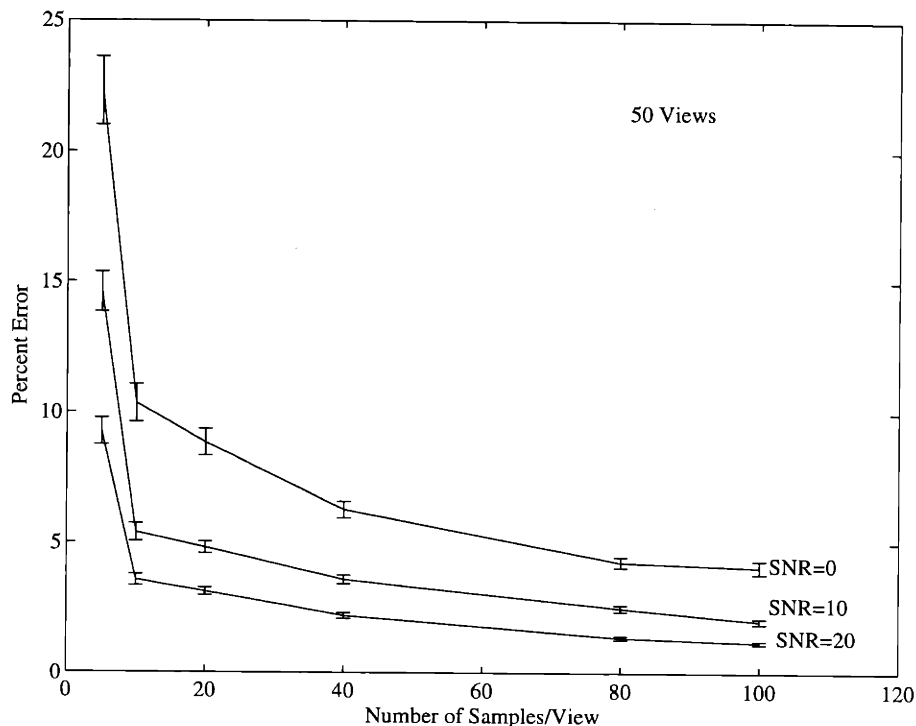


Figure 3-11: Performance as a Function of Number of Samples per View

reconstruction error is plotted versus the number of samples per view for SNR's of 0, 10, and 20, while the number of angular views was fixed at 50. With 50 angular views, all curves fall below 10% reconstruction error when the number of samples per view is greater than only 10. Also, as the number of samples per view is decreased from 100 to 10, again only a marginal increase in the reconstruction error is observed. This behavior shows that the optimal ML estimates are robust with respect to the number of samples per view. Note that for a fixed sampling strategy, the reconstruction error increases only slightly as the SNR is decreased over a wide range. For instance, in Figure 3-10, with 40 angular views and 50 samples per view, the percent error is reduced only about 5% while the SNR goes from 0 to 20.

Finally, it is noteworthy that the reconstruction error enjoys a dramatic improvement for all SNR's (0, 10, and 20) when the number of samples per view is increased from 5 to 10. This improvement is more significant than that observed in Figure 3-10 when the number of views is increased from 5 to 10. This behavior indicates that in a scenario where only a small (fixed) number of sample points can be collected, it is more beneficial to have more samples per view rather than more views.

3.3.2 MDL Reconstructions

Here we will examine reconstruction under the MDL criterion of (3.7) where we now assume that the number of sides of the reconstructed polygon is unknown. In particular, the reconstruction experiments for the hexagon in Figure 3-5 were repeated at SNR=0 assuming no knowledge of the number of sides. The MDL criterion was employed to estimate the optimal number of sides. Figure 3-12 shows a plot of the MDL cost corresponding to the expression in (3.7) vs. the number of sides for a sample reconstruction of the hexagon in Figure 3-5. It can be seen that the minimum occurs at $N = 6$, demonstrating that the optimal MDL reconstruction will consist of 6-sides. Indeed this number coincides with the true number of sides of the underlying object. The optimal MDL estimate is thus exactly the optimal ML estimate for this data set presented before. Note that as in the ML algorithm, it is important to find a good initial guess for the MDL algorithm as well. The problem is twofold. First, a reasonable guess must be made as to the appropriate *range* of the number of sides. We picked a fairly small range for the number of sides of the reconstruction; typically, 3 to 10 sides. Next, for each number of sides in this range, the Initial Guess algorithm was used to produce an initial guess for the optimization routine. The procedure for selecting the range of number of sides is ad hoc, but was shown to be reliable in the sense that for our simulations, the MDL cost never showed local or global minima for number of sides larger than 10.

3.3.3 Polygonal Reconstruction of Non-polygonal Objects

In this section we wish to study the robustness of MDL-based estimates when the underlying, true object is non-polygonal. First we examine the case of an elliptical object. We use the MDL formulation presented in the previous section and study the behavior of the optimal reconstructions at two different SNR's. To this end, let the true object (that which generated the data) be a binary ellipse whose boundary is given by:

$$\left\{ x, y \mid \left(x - \frac{1}{2}\right)^2 + \frac{\left(y + \frac{1}{2}\right)^2}{9/4} = 1 \right\}. \quad (3.47)$$

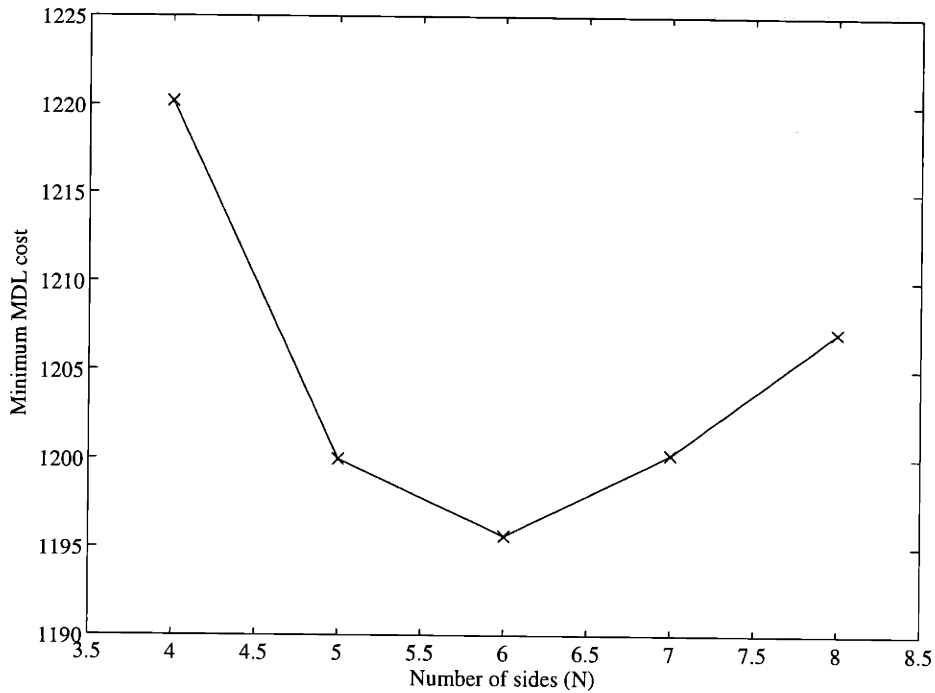


Figure 3-12: Cost vs number of sides for the hexagon in Figure 3-5

The above relation defines an ellipse centered at the point $(1/2, -1/2)$ whose major and minor axes are aligned with the coordinate axes with lengths 1 and $3/2$, respectively.

One thousand (1000) noisy samples of the Radon transform of this ellipse were generated ($m=50$ equally spaced angular views in $(0, \pi]$, and $n=20$ samples per view) at SNR's of 0 and 5 respectively for 50 different sample paths of the corrupting noise. For each set of data, reconstructions were performed using the ML algorithm with 3, 4, 5, 6, 7, and 8 sides. The MDL cost in (3.7) was then computed for each of these reconstructions. The ensemble mean of this cost over the 50 runs, for each value of N , is presented in the top part of Figure 3-13. The error bars denote the 95% confidence intervals for the computed mean values. The top left curve corresponding to the SNR= 5 case displays its minimum at $N = 5$. This behavior indicates that the average optimal MDL reconstruction uses 5 sides at this noise level. A corresponding typical such five-sided reconstruction of the ellipse is displayed on the lower left plot of Figure 3-13 together with the true ellipse. The upper right curve corresponding to the SNR= 0 case displays its minimum at $N = 6$ which indicates that the average optimal MDL reconstruction for this case uses 6 sides. The MDL cost curve for this

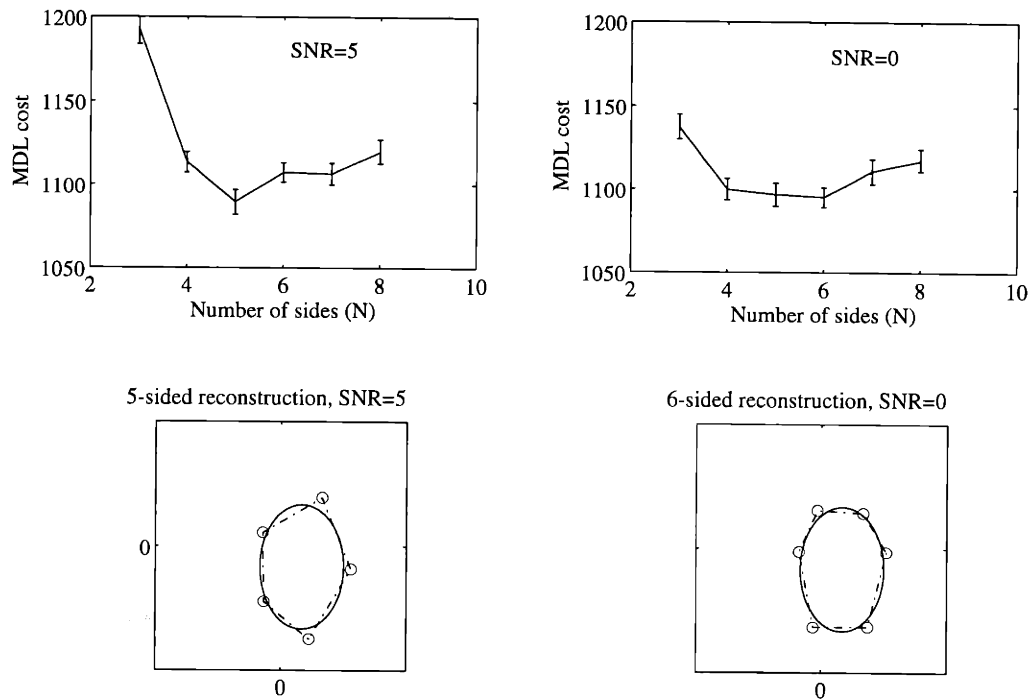


Figure 3-13: Minimum MDL costs and Sample Reconstructions for an Ellipse

lower SNR case has now become quite flat however, showing that the reconstruction with N from 4 to 6 are all about equally explanatory of the data. Although the curves for both cases demonstrate the ability of the MDL procedure to capture the shape's complexity through its choice of N , this behavior suggests that with increasing noise intensity, an MDL-based estimate becomes less sensitive to the *precise level* of complexity of the reconstruction, as we would expect. Apparently, in high noise situations the differences between these reconstructions that would be apparent in high SNR scenarios are masked. As the noise level increases, these fine distinctions are unimportant or *not supported by the data*. A typical 6-sided reconstruction is also displayed in the lower right plot of Figure 3-13 along with the true ellipse.

As another example of the robustness of the ML-based estimates when the underlying object is non-polygonal, we present a sample reconstruction of a nonpolygonal object that is also *non-convex*. In Figure 3-14 a typical reconstruction of this object is presented based on 20 equally spaced projections with 50 samples per projection, at a signal to noise ratio of 10. To ensure that the reconstruction was not a local minimum, the vertices of the initial guess were chosen by eye at points where the underlying object had curvature extrema. Furthermore, the number of sides was

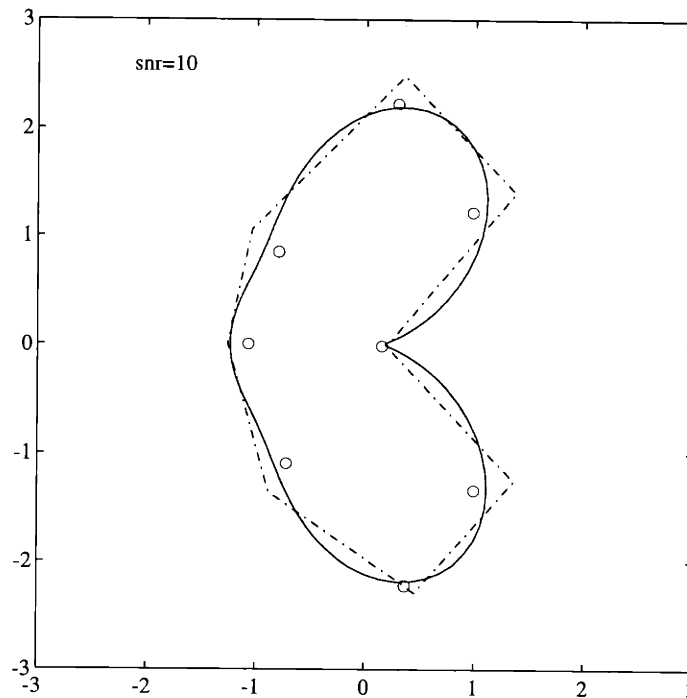


Figure 3-14: True object (-), reconstruction (-.), initial guess (o) picked by eye

picked arbitrarily according to the number of apparent curvature extrema of the true object. Figure 3-15 shows a reconstruction of the same kidney-shaped object at SNR of 10 with the initial guess chosen by the Initial Guess Algorithm. Figure 3-16 contains the reconstruction produced by FBP using the same data set. As in our other examples, the underlying object has been captured more accurately than the FBP reconstruction and without spurious features through the use of our algorithm.

3.3.4 Initial Guess Algorithm

In this section we present some sample reconstructions and performance plots for which we use the Initial Guess algorithm for generating a starting point to the non-linear optimization (3.3). To study the average performance of the ML algorithm using the output of the Initial Guess Algorithm, a Monte-Carlo simulation was done for the reconstruction of the hexagon shown in Figure 3-5. 100 reconstructions were carried out for different realizations of the noise at various SNR's, each with 1000 projection samples as before (50 projections and 20 samples per projection). For each signal to noise ratio, on average less than 5 percent of the reconstructions (i.e. 5 out of 100 sample reconstructions) had very large reconstruction errors (we call these

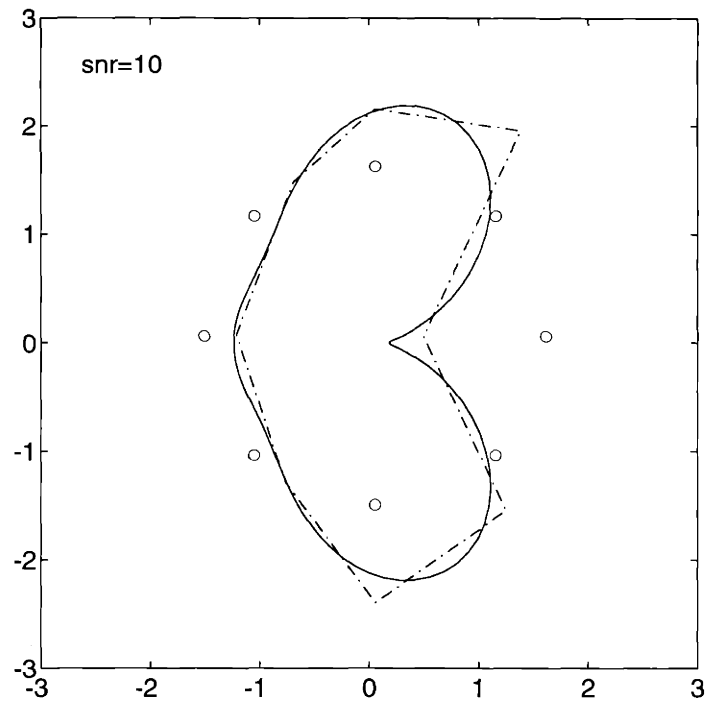


Figure 3-15: True object (-), reconstruction (-.), initial guess (o) picked using Initial Guess Algorithm

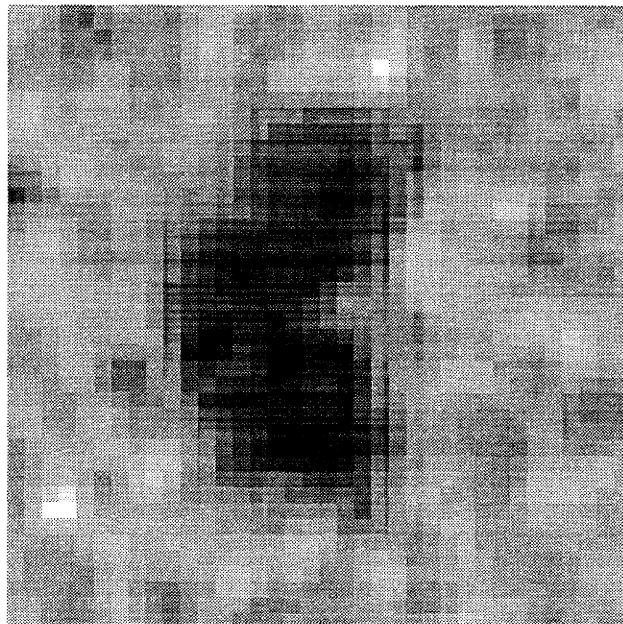


Figure 3-16: FBP Reconstruction of Non-polygonal, Non convex Object: 3rd order Butterworth filter with 0.15 normalized cutoff frequency

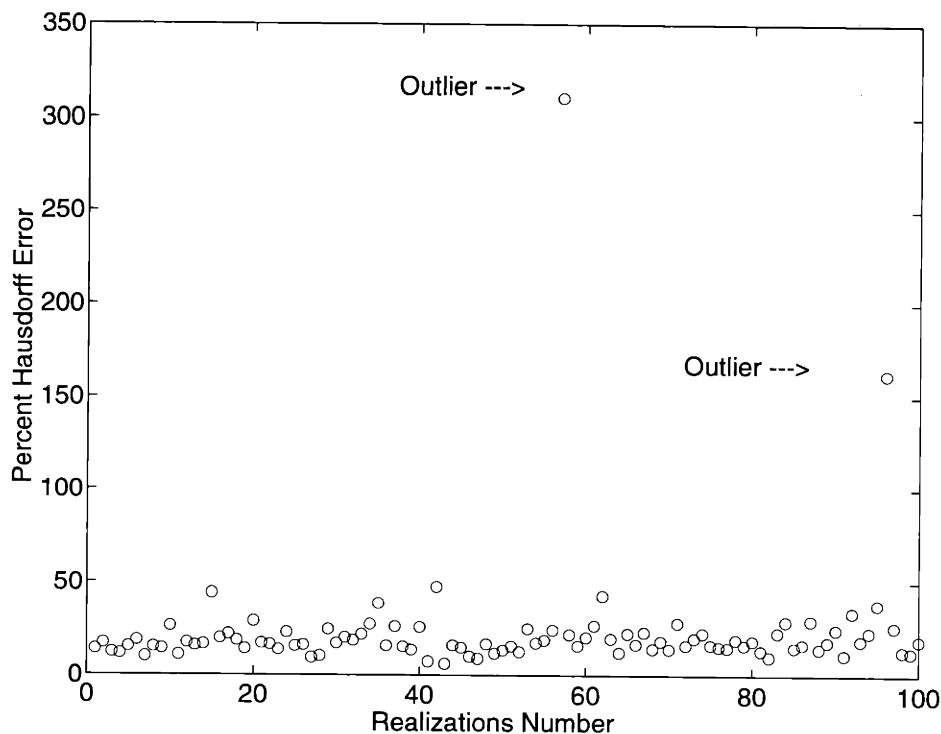


Figure 3-17: A sample path of the reconstruction error at SNR=0

instances *outliers*). Figure 3-17 shows the reconstruction errors for 100 realizations of the noise at SNR=0. The outliers are clearly visible.

Figure 3-17 indicates that in a few instances, the reconstructions were essentially at local minima very far from the global minimum of the cost. In our experience, these outliers occur most frequently when poor estimates of the moments of order 2 are obtained from the noisy data. Note that the second order moments are used in the Initial Guess Algorithm only if the corresponding inertia matrix obtained from them is strictly positive definite. Even though the elements of this estimated inertia matrix (the second order moments) are normally distributed, the eigenvalues of this matrix (which along with the eigenvectors determine inertial axes and whether the inertia matrix is positive definite) are not. Hence, it appears that the outliers shown in Figure 3-17 are a direct result of a somewhat “heavy-tailed” distribution of the eigenvalues which results when the elements of the underlying matrix are normally distributed, though this has not been analytically verified. So in summary, the outliers occur in those rare instances when the estimated inertia matrix happens to be a very poor estimate, but yet positive definite (and hence used in the Initial Guess Algorithm). This phenomenon, in turn, seems to occur when relatively few samples

per projection are available. To put it differently, the estimates of moments of order 2 are more sensitive to the number of samples per view than the number of views. Consequently, the performance of the overall ML algorithm using the Initial Guess algorithm's output as a starting point, is also more sensitive to the number of samples per view rather than the number of views. This is consistent with the conclusions of Section 3.3.1 where we discussed the effects of different sampling strategies on the performance of the ML algorithm.

Figure 3-18 shows the mean percent error in the Monte-Carlo runs after the removal of the outliers. The outliers were simply removed from the ensemble and the results of the remaining realizations were averaged to yield the values in Figure 3-18. That is to say that if 3 out of 100 realizations led to outliers, then only those 97 results which seemed "reasonable" were used in computing the ensemble average. Whether the result of a run was deemed reasonable or not was decided by comparing the resulting percent error to the ensemble median reconstruction error for all 100 runs. In particular, if the percent error for a particular run was larger than one standard deviation away from the *median*, the run was declared an outlier. In the case of Figure 3-17, using the computed median value of 17.2, and standard deviation of 33.2 a threshold level of 50.4 was chosen above which outliers were declared.

The mean percent error after this outlier removal process is plotted here in Figure 3-18 to show the average performance without the effect of the outliers. It can be seen, upon comparing Figure 3-18 with the corresponding performance curve in Figure 3-9 that the performance of the ML algorithm using the output of the Initial Guess algorithm still suffers about a 5 % penalty even after discounting the obvious outliers. This means that instances of convergence to local minima still occur, but note that the average performance after the removal of outliers is not significantly different from the average performance with the actual polygon as the initial guess. From this observation, we conclude that the initial guess algorithm does not always lead to convergence to the global maximum, but it almost always leads at least to a local minimum that is fairly close to the global minimum of the ML cost function. Typical reconstruction at local minima which are close to the global minimum of the cost are

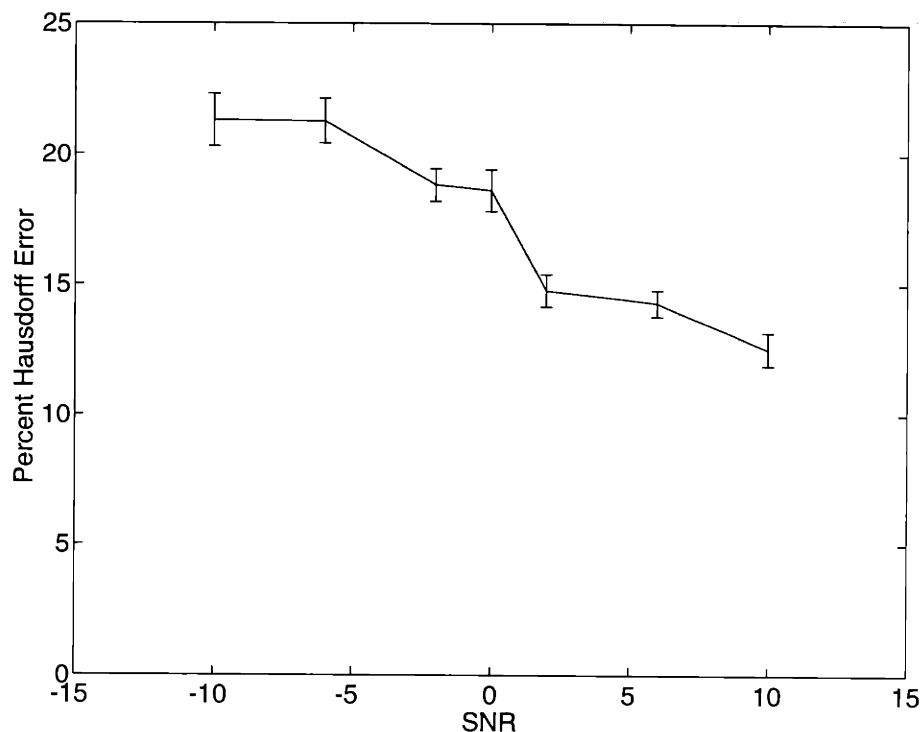


Figure 3-18: Percent error for Hexagon vs SNR after outlier removal

shown in Figure 3-19 for SNR=0 and in Figure 3-20 for SNR=10.

3.3.5 Conclusions from Experiments

Several conclusions may be drawn from the experimental results presented in this section. The optimal estimates based on the ML criterion produce reconstructions that are highly robust to noise and sparsity and distribution of data. This behavior is essentially a direct consequence of the fact that ML-based techniques focus all of the available information to the task at hand as represented in the parameters of underlying object. In contrast, classical approaches, such as the FBP algorithm must spread this information over all the pixels in an image. Extending the ML approach through the MDL principle allows for the automatic determination of the optimum number of parameters needed to describe the given data set. The experiments verify the utility of this method. The ML-based approaches are also able to produce reconstructions for a wide variety of objects. The drawback of such statistically optimal ML and MDL approaches is that, in contrast to the linear formulations of classical reconstruction algorithms, they lead to highly non-linear optimization problems. This fact makes the issue of computing a good initial guess to the nonlinear optimization routines an

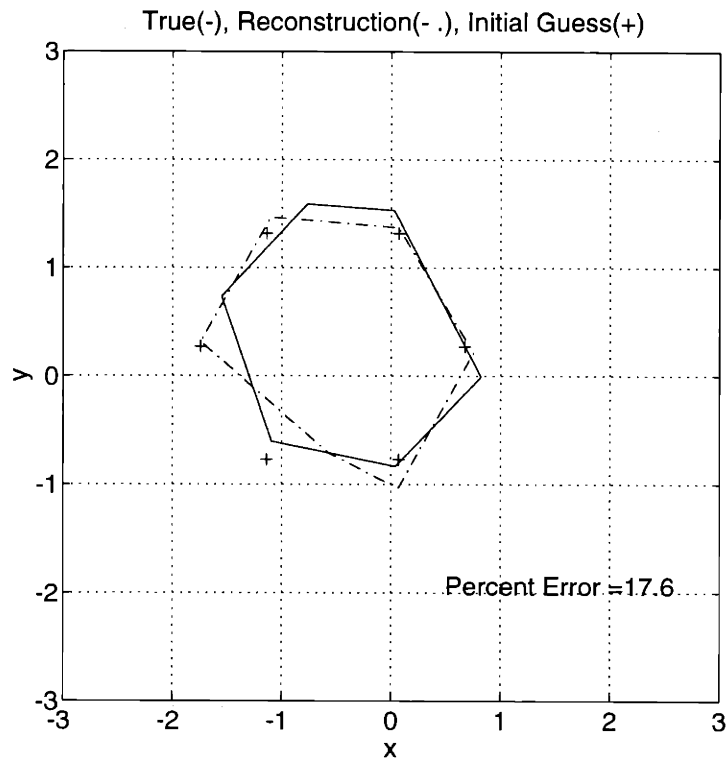


Figure 3-19: A typical reconstruction at a local minimum with SNR=0

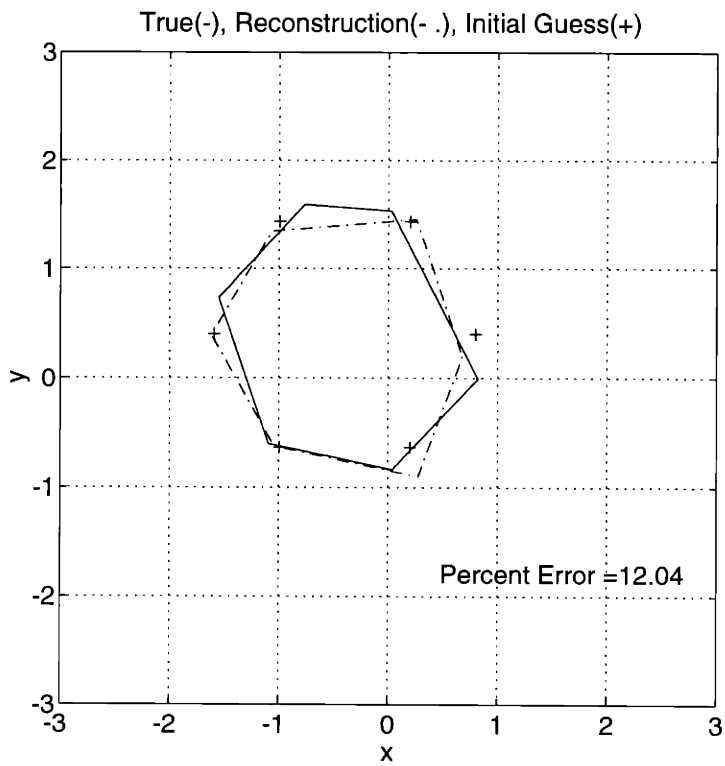


Figure 3-20: A typical reconstruction at a local minimum with SNR=10

important one. We provided a way to circumvent this difficulty through a robust and simple initial guess algorithm based on the estimated moments of the object. In particular, the coarse geometric information carried in the projections can be easily extracted in the form of moments, which are then used to generate the initial guess. The efficacy of this algorithm over a wide range of SNR's was demonstrated.

3.4 Conclusions

In this Chapter, we have studied statistical techniques for the reconstruction of finitely parameterized geometric objects. In particular, we focused on the reconstruction of binary polygonal objects. The reconstruction of such objects was posed as a parameter estimation problem for which the Maximum Likelihood technique was proposed as a solution. In contrast to the classical techniques, such as FBP, the ML based reconstructions showed great robustness to noise and data loss and distribution. The drawback of such ML-based formulations is that the resulting optimization problems are highly non-linear and thus a good initial guess is necessary to ensure convergence of optimization routines to the true ML estimate or to at least a local minimum near the true ML estimate. To this end, an algorithm was presented for computing such a reasonable initial guess using moments of the object which are estimated directly from the projection data. While estimation of a function from its moments is, in general, a difficult and ill-posed problem, we avoid these problems by using the noisy estimated moments only to guide a coarse object estimate. This estimate, in turn, mitigates the difficulties associated with the non-linearities of the optimal ML statistical approach. The efficacy of this moment based initial guess algorithm was demonstrated over a wide range of SNR's.

If the number of parameters describing the underlying object are not known, a Minimum Description Length criterion can be employed that simply generalizes the ML framework to penalize the use of an excessively large number of parameters for the reconstruction. The MDL approach was shown to work successfully in estimating the number of sides and the underlying object itself for low signal-to-noise ratio

situations and for a variety of sampling scenarios. It was further demonstrated that if the underlying object is not polygonal, but still binary, the proposed MDL algorithm is still capable of producing polygonal reconstructions which reasonably capture the object shape in the presence of high noise and sparsely sampled data.

In this work we have focused on the reconstruction of binary polygonal objects parameterized by their vertices. The ML and MDL-based techniques used here may also be applied to more general object parameterizations. In particular, while we used the (estimated) moments of the object only as the basis for generation of an initial guess, it is possible to actually parameterize the entire object through its moments, as we will show in Chapter 6. More generally, a “well-behaved” function is completely determined by the entire set of its moments. In reality we will only have access to a finite set of these moments and these numbers, being estimates, will themselves be inexact and noisy. While estimation of the moments of a function based on its projections is, conveniently, a linear problem, inversion of the resulting finite set of moments to obtain the underlying function estimate is a difficult and ill-posed problem. These observations suggest a spectrum of ways in which to use moments in our reconstruction problems. At one extreme, only a few moments are used in a sub-optimal way to generate a simple initialization for solution of a hard, non-linear estimation problem. At the other extreme, the moments are themselves used in an optimal reconstruction scheme. In Chapter 5, we also consider the use of estimated moments as constraints in variational formulations of general tomographic reconstruction problems.

3-A Alternate Parameterizations

A finitely parameterized object may be represented in many different, but equivalent ways. Some parameterizations may yield more numerically robust algorithms while others may prove to be more geometrically intuitive. A case in point is that of binary polygonal objects. The parameterization of binary polygonal objects in terms of their vertices is only one of many possible parameterizations. We have chosen this

particular parameterization because it seems natural, and in fact it is rather easy to compute the projections of a binary polygonal object in terms of its vertices. It does, however, lead to highly nonlinear and complex optimization problems as we have seen. We sought other parameterizations of binary polygonal objects such as those in terms of normals to the sides, the slopes and intercepts of the sides, and also the polar representation of the vertices; but they all essentially led to rather complicated cost functions with no apparent advantage. In some instances, however, other geometrically informative parameterizations are possible. One such parameterization is the affine representation. In particular, the affine representation of triangles is rather interesting. We briefly describe this representation in this section.

3-A.1 The Affine Representation of Triangles

We wish to consider an affine representation of all non-degenerate triangles in the plane in this section. More specifically, we can show that every non-degenerate triangle in the plane can be represented (up to a permutation of the vertices) by some affine transformation of a reference triangle $V_o(N)$ which we have picked to be an equilateral triangle centered at the origin with vertices at the third roots of unity in counter-clockwise order:

$$V_o(N) = \begin{bmatrix} 1 & -1/2 & -1/2 \\ 0 & \sqrt{3}/2 & -\sqrt{3}/2 \end{bmatrix}. \quad (3.48)$$

We have the following result.

Result 3 *For every non-degenerate triangle V in the plane, there exists a unique pair (L, C) such that*

$$V = LV_o(N) + [C \mid C \mid C] \quad (3.49)$$

where L is an invertible 2×2 matrix, and C is a 2×1 vector.

Proof: It is easy to check that the unique pair is given by

$$L = (2/3)VV_o^T(N) \quad (3.50)$$

$$C = (1/3)V_s \quad (3.51)$$

where $s = [1, 1, 1]^T$. Since V and $V_o(N)$ are both assumed to be non-degenerate triangles, L will have rank 2, and is therefore invertible. \square

Note that the expression for C is simply the arithmetic average of the coordinates of the vertices. For a binary triangle, this corresponds exactly to the center of mass of V . It is interesting to note that a representation of the form (3.49) will not yield arbitrary N -gons in the plane as affine transformations of a reference, regular N -gon.

Now consider the Singular Value Decomposition (SVD) of the matrix L . i.e.

$$L = L_\phi S_\lambda L_\gamma \quad (3.52)$$

Where L_γ , and L_ϕ are orthogonal transformations which can be interpreted as rotations followed by reflections in the plane, while S_λ is diagonal and strictly positive definite which can be interpreted as an isotropic scaling in the plane. Hence we have

$$V = (L_\phi S_\lambda L_\gamma)V_o(N) + [C \mid C \mid C]. \quad (3.53)$$

Therefore in geometric terms, any triangle in the plane (up to a permutation of vertices) can be obtained by a series of rotations, reflections, and a stretching of $V_o(N)$.

Some interesting observations can be made regarding the information one obtains from the SVD of L . The condition number of L is simply the ratio of the largest to smallest singular values of L . We term square root of this number the **eccentricity** of the triangle V . Given the above parameterization of triangles in the plane, it is natural to ask whether similar representations exist for simple polygons with N sides ($N \geq 4$). As pointed out earlier, an affine transformation of the form given for the triangle will clearly not work since there are simply not enough degrees of

freedom. An alternative is to split the simple N -gon into triangular pieces, each of which may be described by an affine transformation as in (3.49). There clearly need to be constraints among these affine transformations that guarantee the end result to be a simple polygon. In the interest of the focus of this chapter, we will not address these questions here further.

3-B Theoretical Results on the Initial Guess Algorithm

In this section we present some theoretical justification for the initial guess algorithm. To start, we state some elementary properties of unit area polygons $V_{\text{ref}}(N)$ whose vertices are the scaled N^{th} roots of unity (in counter-clockwise direction). Explicitly,

$$V_{\text{ref}}(N) = \frac{1}{\sqrt{\frac{N}{2} \sin(\frac{2\pi}{N})}} \left[\begin{array}{c|c|c|c} \cos(0) & \cos(\frac{2\pi}{N}) & \cdots & \cos(\frac{2\pi(N-1)}{N}) \\ \hline \sin(0) & \sin(\frac{2\pi}{N}) & \cdots & \sin(\frac{2\pi(N-1)}{N}) \end{array} \right] \quad (3.54)$$

Invoking Result 6 of Chapter 4, it is a matter of some algebraic manipulations to show that the regular polygon $V_{\text{ref}}(N)$ has moments of up to order 2 given by

$$\mu_{00}(V_{\text{ref}}(N)) = 1 \quad (3.55)$$

$$\mu_{10}(V_{\text{ref}}(N)) = \mu_{01}(V_{\text{ref}}(N)) = 0 \quad (3.56)$$

$$\mu_{20}(V_{\text{ref}}(N)) = \mu_{02}(V_{\text{ref}}(N)) = \frac{1}{4N \tan(\frac{\pi}{N})} = k_N \quad (3.57)$$

$$\mu_{11}(V_{\text{ref}}(N)) = 0 \quad (3.58)$$

Now let V_{init} be an affine transformation of V_{ref} as

$$V_{\text{init}} = LV_{\text{ref}}(N) + [C \mid C \mid \cdots \mid C] \quad (3.59)$$

for some linear transformation L and some 2×1 vector C . Let $\mathcal{O}(V_{\text{init}})$ denote the closed, binary polygonal region enclosed by the N -gon V_{init} . Now by considering the change of variables $\mathbf{z} = L\mathbf{u}$, and dropping the explicit dependences on N we have

$$\mu_{00}(V_{\text{init}}) = \iint_{\mathcal{O}(V_{\text{init}})} d\mathbf{z}, \quad (3.60)$$

$$= \iint_{\mathcal{O}(V_{\text{ref}})} |\det(L)| d\mathbf{u}, \quad (3.61)$$

$$= \mu_{00}(V_{\text{ref}}) |\det(L)| = |\det(L)| \quad (3.62)$$

Similarly, we get

$$[\mu_{10}(V_{\text{init}}) \ \mu_{01}(V_{\text{init}})]^T = \left(L[\mu_{10}(V_{\text{ref}}) \ \mu_{01}(V_{\text{ref}})]^T + C \right) |\det(L)| = |\det(L)| C \quad (3.63)$$

and

$$\mathcal{I}(V_{\text{init}}) = (L\mathcal{I}(V_{\text{ref}})L^T + CC^T) |\det(L)| = (k_N LL^T + CC^T) |\det(L)| \quad (3.64)$$

where for any N -gon V we write

$$\mathcal{I}(V) = \begin{bmatrix} \mu_{20}(V) & \mu_{11}(V) \\ \mu_{11}(V) & \mu_{02}(V) \end{bmatrix}. \quad (3.65)$$

This proves relations (3.20), (3.21), and (3.22). We next establish an explicit description of the set of all affinely regular N -gons with a fixed set of moments up to order 2. In order to do this, we first need to prove a lemma.

Lemma 1 *For every N -gon V with moments μ_{00} , $\mu_{10} = 0$, $\mu_{01} = 0$, μ_{20} , μ_{11} , μ_{02} , such that the inertia matrix \mathcal{I} satisfies $\det(\mathcal{I}) = k_N^2 \mu_{00}^4$, there exists a matrix L , unique up to some orthogonal transformation, such that $V = LV_{\text{ref}}$.*

Proof: The assumptions that $\mu_{10} = 0$ and $\mu_{01} = 0$ are made without loss of generality and to facilitate the presentation of the proof. Having said this, we define L as the scaled (unique) square root of \mathcal{I} as follows. First, write the following eigendecomposition

$$\frac{\mathcal{I}}{\sqrt{\det(\mathcal{I})}} = US^2U^T, \quad (3.66)$$

where U is orthogonal and S has unit determinant. Define L as

$$L = \sqrt{\mu_{00}}US. \quad (3.67)$$

The moments of $V = LV_{\text{ref}}$ are then given by

$$\mu_{00}(V) = \mu_{00} \quad (3.68)$$

$$\mu_{10}(V) = \mu_{01}(V) = 0 \quad (3.69)$$

$$\mathcal{I}(V) = k_N LL^T |\det(L)| \quad (3.70)$$

Note that

$$\det(\mathcal{I}) = k_N^2 \mu_{00}^4 \quad (3.71)$$

as required. If L is replaced by LT where T is any 2×2 orthogonal transformation, the same moments are obtained. Hence the lemma is established. \square

Given this lemma, we obtain an interesting geometric representation of all affinely regular N -gons that have a prespecified set of moments of up to order 2. This characterization is given by Result 2, on page 56, which we prove next.

Proof of Result 2: First consider an N -gon $V \in C_1$. V has moments $\mu_{00}, 0, 0, \mathcal{I}$ and therefore, by Lemma 1, there exists an L given by (3.66) and (3.67), unique up to some orthogonal matrix T_1 such that we can write

$$V = (LT_1)V_{\text{ref}}(N) \quad (3.72)$$

Let us denote the N -gons V and $V_{\text{ref}}(N)$ explicitly in terms of their columns as

$$V = [v_1 \mid v_2 \mid \cdots \mid v_N] \quad (3.73)$$

$$V_{\text{ref}}(N) = [w_1 \mid w_2 \mid \cdots \mid w_N] \quad (3.74)$$

so that

$$v_j = LT_1 w_j. \quad (3.75)$$

It is easy to show from the definition of $V_{\text{ref}}(N)$ that

$$w_j^T w_j = \alpha_N, \quad (3.76)$$

$$(w_{j+1} + w_j)^T (w_{j+1} + w_j) = 4\beta_N. \quad (3.77)$$

Now to show that $V \in C_2$, we prove that

$$v_j^T E_O^{-1} v_j = 1 \quad (3.78)$$

$$\frac{(v_{j+1} + v_j)^T}{2} E_I^{-1} \frac{(v_{j+1} + v_j)}{2} = 1 \quad (3.79)$$

for $j = 1, 2 \dots N$.⁴ Using (3.66) and (3.67), we can write

$$v_j^T E_O^{-1} v_j = \frac{\mu_{00} k_N}{\alpha_N} v_j^T \mathcal{I}^{-1} v_j. \quad (3.80)$$

$$= \frac{\mu_{00} k_N}{\alpha_N} \frac{1}{k_N \mu_{00}^2} \mu_{00} w_j^T T_1^T S U^T U S^{-2} U^T U S T_1 w_j. \quad (3.81)$$

$$= \frac{1}{\alpha_N} w_j^T w_j. \quad (3.82)$$

$$= 1. \quad (3.83)$$

Similarly,

$$\begin{aligned} \frac{(v_{j+1} + v_j)^T}{2} E_I^{-1} \frac{(v_{j+1} + v_j)}{2} &= \frac{1}{4} \frac{\mu_{00} k_N}{\beta_N} \frac{1}{k_N \mu_{00}^2} \mu_{00} (v_{j+1} + v_j)^T \mathcal{I}^{-1} (v_{j+1} + v_j) \\ &= \frac{1}{4\beta_N} (w_{j+1} + w_j)^T (w_{j+1} + w_j) \end{aligned} \quad (3.84)$$

$$= 1 \quad (3.85)$$

Hence, $V \in C_2$.

⁴Note that, by convention, $N + 1 = 1$

Now assume that $V \in C_2$. Then, by assumption,

$$v_j^T E_O^{-1} v_j = 1 \quad (3.86)$$

$$\frac{(v_{j+1} + v_j)^T}{2} E_I^{-1} \frac{(v_{j+1} + v_j)}{2} = 1 \quad (3.87)$$

Define the vertices of the related N -gon Z as

$$z_j = \frac{1}{\sqrt{\alpha_N \mu_{00}}} S^{-1} U^T v_j. \quad (3.88)$$

where S and U are given by the normalized eigendecomposition of \mathcal{I} given in (3.66).

Writing (3.86) and (3.87) in terms of z_j , after some algebraic manipulations, we get

$$z_j^T z_j = 1 \quad (3.89)$$

$$\frac{(z_{j+1} + z_j)^T}{2} \frac{(z_{j+1} + z_j)}{2} = \cos^2\left(\frac{\pi}{N}\right) \quad (3.90)$$

From these identities, again with some algebraic manipulation, it easily follows that Z is equilateral. Specifically,

$$\|z_{j+1} - z_j\|^{1/2} = 2 \sin\left(\frac{\pi}{N}\right). \quad (3.91)$$

Since the above identities show that the N -gon Z is a regular N -gon inscribed in the unit circle, then it must be related to $V_{\text{ref}}(N)$ through a scaling and some orthogonal transformation T_2 . In particular,

$$Z = \sqrt{\frac{N}{2} \sin\left(\frac{2\pi}{N}\right)} T_2 V_{\text{ref}}(N) \quad (3.92)$$

This in turn shows that

$$\frac{1}{\sqrt{\alpha_N \mu_{00}}} S^{-1} U^T V = \sqrt{\frac{N}{2} \sin\left(\frac{2\pi}{N}\right)} T_2 V_{\text{ref}}(N) \quad (3.93)$$

or after solving for V and simplifying,

$$V = \sqrt{\mu_{00}} U S T_2 V_{\text{ref}}(N). \quad (3.94)$$

Letting $L = \sqrt{\mu_{00}} U S$, we obtain

$$V = (L T_2) V_{\text{ref}}(N). \quad (3.95)$$

This last identity implies that V has moments $\mu_{00}(V) = \mu_{00}$, $\mu_{10}(V) = \mu_{01}(V) = 0$, and

$$\mathcal{I}(V) = k_N L L^T |\det(L)| \quad (3.96)$$

with $\det(\mathcal{I}(V)) = k_N^2 \mu_{00}^4$. Hence $V \in C_1$ and the result is established. \square

Note that the ratios α_N/k_N and β_N/k_N can be simplified to get

$$\frac{\alpha_N}{k_N} = \frac{4}{\cos^2(\pi/N)}, \quad (3.97)$$

$$\frac{\beta_N}{k_N} = 4. \quad (3.98)$$

If \mathcal{I} is not the inertia matrix of an affinely regular N -gon, then the L constructed in the Initial Guess Algorithm will not have have the prescribed inertia matrix. We are, however, able to bound the approximation error from above and below in the following way.

Result 4 *Suppose that the moments μ_{00} , $\mu_{10} = \mu_{01} = 0$,*

$$\mathcal{I} = \begin{bmatrix} \mu_{20} & \mu_{11} \\ \mu_{11} & \mu_{02} \end{bmatrix} \quad (3.99)$$

are given such that $\det(\mathcal{I}) = k_N^2 \mu_{00}^4 + \epsilon > 0$. Define

$$L = \sqrt{\mu_{00}} U S \quad (3.100)$$

where

$$\frac{\mathcal{I}}{\sqrt{\det(\mathcal{I})}} = US^2U^T \quad (3.101)$$

is the normalized eigendecomposition of \mathcal{I} . Then the normalized Frobenius-norm error is given by

$$\frac{\|\mu_{00}k_N LL^T - \mathcal{I}\|_F}{\|\mathcal{I}\|_F} = \left| 1 - \frac{k_N \mu_{00}^2}{\sqrt{k_N^2 \mu_{00}^4 + \epsilon}} \right| \quad (3.102)$$

Proof: Letting $A = \mu_{00}k_N LL^T$ and $B = \mathcal{I}$, we can write

$$A = aUS^2U^T, \quad (3.103)$$

$$B = bUS^2U^T, \quad (3.104)$$

where

$$S^2 = \begin{bmatrix} \lambda & 0 \\ 0 & 1/\lambda \end{bmatrix}, \quad (3.105)$$

$$a = k_N \mu_{00}^2, \quad (3.106)$$

$$b = \sqrt{k_N^2 \mu_{00}^4 + \epsilon}. \quad (3.107)$$

Hence we have

$$\|A - B\|_F = \|(a - b)S^2\|_F = |a - b| \sqrt{\lambda^2 + \frac{1}{\lambda^2}} \quad (3.108)$$

$$\|B\|_F = |b| \sqrt{\lambda^2 + \frac{1}{\lambda^2}}. \quad (3.109)$$

Hence,

$$\frac{\|A - B\|_F}{\|B\|_F} = \frac{|a - b|}{|b|} = \frac{|\sqrt{k_N^2 \mu_{00}^4 + \epsilon} - k_N \mu_{00}^2|}{\sqrt{k_N^2 \mu_{00}^4 + \epsilon}} = \left| 1 - \frac{k_N \mu_{00}^2}{\sqrt{k_N^2 \mu_{00}^4 + \epsilon}} \right| \quad (3.110)$$

which establishes the result. \square

We have plotted the expression for the relative error in Figure 3-21 for $N = 3$ and

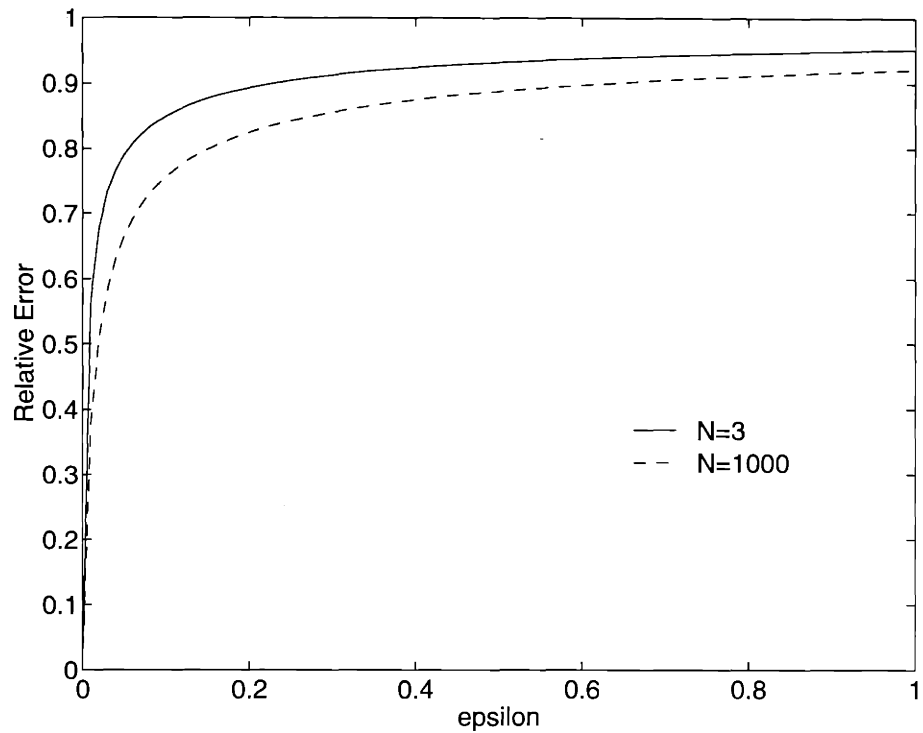


Figure 3-21: Relative error in matching second order moments using the Initial Guess Algorithm

$N = 1000$, and assuming that $\mu_{00} = 1$. This figure shows that although the relative error grows quite fast as ϵ is increased, it never exceeds the maximum of 1 (i.e. 100 percent) for a fixed μ_{00} . Also, the relative errors for different number of sides are seen to be very close.

Chapter 4

Moments and the Radon Transform

4.1 Introduction

In this chapter we discuss a fundamental property of the Radon transform which relates the moments of a function $f(x, y)$ to those of its Radon transform $g(t, \theta)$. We exploit this property to develop a Maximum Likelihood framework for the estimation of the moments of a function ¹ from noisy and possibly sparse measurements of its line-integral projections. We show that this framework provides one with a direct and statistically optimal methodology for the extraction of geometric information about an image directly from projection data. This geometric information can be used in several distinctly different ways. As we discussed in Chapter 3, these estimated moments can be used to construct good initial guesses for highly nonlinear optimization problems arising in parameterized, model-based, optimal tomographic reconstruction algorithms for binary polygonal objects. As we shall describe in Chapter 5, these estimated moments can also be used to arrive at robust and computationally efficient variational reconstruction algorithms. In addition, from the study of the relationship between the moments of a function and those of its Radon transform we gain

¹As mentioned in Chapter 1, we will use the terms “function” and “image” interchangeably.

much insight into the structure of the range and null-space of the Radon transform operator and, in fact, arrive at a decomposition of this operator which yields new interpretations for a number of classical tomographic reconstruction algorithms. From a statistical standpoint, we show that the moments of the noisy projections of an image constitute a set of sufficient statistics for the optimal estimation of the underlying image. This fact provides a stochastic justification for estimation and reconstruction based on moments. Finally, we show that the study of the explicit relationship between the moments of an image and those of its Radon transform yields insights as to what sampling (projection) strategies are generically the best in terms of yielding the best-conditioned moment estimation problem.

The explicit relationship between the moments of a function and those of its Radon transform has been known and used extensively in the mathematics community for some time [39, 30, 33]. The use of this relationship has basically been limited, however, to mostly theoretical considerations. Practical applications of this relationship have rarely been used outside of the mathematics community [58, 66]. A framework has, to date, not been established for exploiting this relationship for estimation and reconstruction purposes from a practical point of view. In this chapter, we intend to provide such a framework. We use this framework in Chapters 3, 5, and 6 to develop novel reconstruction algorithms.

The organization of this chapter is as follows. In Section 4.2 we present the basic definitions of various types of moments of an image (a function) and also describe some notation which will be used in what follows. Section 4.3 contains fundamental results relating the moments of a function to the moments of its projections and also results regarding the number of uniquely computable moments from a finite number of projections. In Section 4.4 we present new interpretations of classical reconstruction algorithms based on the results of Section 4.3. We discuss the optimal estimation of the moments of a function from noisy measured projections in Section 4.5. Section 4.6 describes other useful results arising from the study of the moments of the Radon transform of a function. Finally, to illustrate the main concepts of this chapter, we present some examples and experiments in Section 4.7.

4.2 Basic Definitions

The moments of a function $f(x, y)$ are the coefficients of the expansion of f in any basis. In what follows, we shall assume that f is a square-integrable function with compact support \mathcal{O} contained in the closed unit disk D . We shall denote this by $f \in L^2(D)$. We next define the geometric, orthogonal, and complex moments of f .

Definition 3 (Geometric Moments) *Let $f(x, y) \in L^2(D)$ be defined over a compact subset of the plane $\mathcal{O} \subset D$. The **geometric moments** of f over \mathcal{O} are defined as the set of real numbers μ_{pq} given by*

$$\mu_{pq} = \iint_{\mathcal{O}} f(x, y) x^p y^q dx dy \quad (4.1)$$

for all integers $p, q \geq 0$.

Note that μ_{00} is the area of \mathcal{O} while μ_{10}/μ_{00} and μ_{01}/μ_{00} give the x and y coordinates of the center of mass of \mathcal{O} , respectively. Second order moments of f give rise to other geometric information; namely, the inertial properties of f (for $p + q = 2$). Higher order moments of f yield more detailed information about f . However, these higher moments are not easily interpreted as tangible geometric quantities.

Definition 4 (Orthogonal Moments) *The **orthogonal (Legendre) moments** of f over \mathcal{O} , denoted by λ_{pq} , are defined by*

$$\lambda_{pq} = \iint_{\mathcal{O}} f(x, y) P_p(x) P_q(y) dx dy, \quad (4.2)$$

where $P_k(\cdot)$ denotes the k^{th} order normalized Legendre polynomial over the closed interval $[-1, 1]$.

Explicitly, the k^{th} -order normalized Legendre polynomial is given by

$$P_k(x) = \sum_{i=0}^k \beta_{ik} x^i = \sqrt{\frac{2k+1}{2}} \frac{1}{2^k k!} \frac{d^k}{dx^k} (x^2 - 1)^k. \quad (4.3)$$

The family of Legendre polynomials $\{P_k(x) \mid k \geq 0\}$ form a complete orthonormal basis over the interval $[-1, 1]$, i.e.

$$\int_{-1}^1 P_p(x)P_q(x)dx = \delta_{pq} \quad \forall p, q \geq 0, \quad (4.4)$$

where

$$\delta_{pq} = \begin{cases} 1 & \text{for } p = q \\ 0 & \text{for } p \neq q \end{cases} \quad (4.5)$$

Therefore, the family of functions $\{P_p(x)P_q(y) \mid p, q \geq 0\}$ forms a complete orthonormal set of functions over the square given by $[-1, 1] \times [-1, 1]$.

Since the Legendre Polynomials span the set of functions $\{x^k \mid k \geq 0\}$, it is natural to expect that the Legendre moments of $f(x, y)$ should be linearly related to the geometric moments of f . This is indeed the case and the explicit relationship is given by the following expression.

$$\lambda_{pq} = \sum_{i=0}^p \sum_{j=0}^q \beta_{ip} \beta_{jq} \mu_{ij} \quad (4.6)$$

Hence, a given set of orthogonal moments depends only on geometric moments of the same order and lower, and conversely.

Definition 5 (Complex Moments) *The complex moments of f over \mathcal{O} , denoted by c_{pq} , are defined by*

$$c_{pq} = \iint_{\mathcal{O}} f(x, y) z^p \bar{z}^q dx dy \quad (4.7)$$

where $z = x + iy$, $\bar{z} = x - iy$, and $i = \sqrt{-1}$.

The complex moments of order k are linear combinations, with complex coefficients, of the geometric moments of the same order. More precisely, we have [97]:

$$c_{pq} = \sum_{r=0}^p \sum_{s=0}^q \binom{p}{r} \binom{q}{s} i^{(p+q)-(\tau+s)} (-1)^{q-s} \mu_{r+s, (p+q)-(\tau+s)}. \quad (4.8)$$

The complex moments and their relationship with the geometric moments will be used in Chapter 6 when the reconstruction of binary polygonal objects from their moments is discussed. We hence defer further discussion of the complex moments to Chapter 6.

For the sake of notational convenience, we group the moments of a given order k and define the k^{th} order geometric and orthogonal moment vectors respectively as follows:

$$\mu^{(k)} = [\mu_{k,0}, \mu_{k-1,1}, \dots, \mu_{0,k}]^T \quad (4.9)$$

$$\lambda^{(k)} = [\lambda_{k,0}, \lambda_{k-1,1}, \dots, \lambda_{0,k}]^T \quad (4.10)$$

Note that each of these vectors contains all the moments whose indices add up to k . These moments are arranged in this particular way to ease the exposition in what follows.

Vectors of the geometric and orthogonal moments *up to order* N are defined using the moment vectors $\mu^{(k)}$ and $\lambda^{(k)}$ respectively for $k = 0, 1, \dots, N$ as follows:

$$\mathcal{M}_N = \begin{bmatrix} \mu^{(0)} \\ \mu^{(1)} \\ \vdots \\ \mu^{(N)} \end{bmatrix}, \quad \mathcal{L}_N = \begin{bmatrix} \lambda^{(0)} \\ \lambda^{(1)} \\ \vdots \\ \lambda^{(N)} \end{bmatrix} \quad (4.11)$$

By virtue of (4.6), we then have a linear relation between the vectors \mathcal{M}_N and \mathcal{L}_N which we denote as

$$\mathcal{M}_N = T_N \mathcal{L}_N \quad (4.12)$$

where T_N is a lower triangular, invertible, $\frac{(N+1)(N+2)}{2} \times \frac{(N+1)(N+2)}{2}$ matrix. For $N = 2$, for instance, T_N is given by:

$$T_2 = \begin{bmatrix} 2.0000 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1.1547 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1.1547 & 0 & 0 & 0 \\ 0.6667 & 0 & 0 & 0.5963 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.6667 & 0 \\ 0.6667 & 0 & 0 & 0 & 0 & 0.5963 \end{bmatrix} \quad (4.13)$$

For a fixed angle θ , we can define the k^{th} order geometric and orthogonal moments of the projection $g(t, \theta)$ analogously to the moments of f as follows.

$$H^{(k)}(\theta) = \int_{-1}^1 g(t, \theta) t^k dt, \quad (4.14)$$

$$G^{(k)}(\theta) = \int_{-1}^1 g(t, \theta) P_k(t) dt, \quad (4.15)$$

where from the definition of the Legendre polynomials, it follows that

$$G^{(k)}(\theta) = \sum_{i=0}^k \beta_{ik} H^{(i)}(\theta). \quad (4.16)$$

We define the vectors of moments up to order N of $g(t, \theta)$ similarly as

$$\mathcal{H}_N(\theta) = \begin{bmatrix} H^{(0)}(\theta) \\ H^{(1)}(\theta) \\ \vdots \\ H^{(N)}(\theta) \end{bmatrix}, \quad \mathcal{G}_N(\theta) = \begin{bmatrix} G^{(0)}(\theta) \\ G^{(1)}(\theta) \\ \vdots \\ G^{(N)}(\theta) \end{bmatrix}. \quad (4.17)$$

Through (4.16) we have:

$$\mathcal{G}_N(\theta) = B_N \mathcal{H}_N(\theta) \quad (4.18)$$

where

$$B_N = \begin{bmatrix} \beta_{00} & 0 & 0 & \cdots & 0 \\ \beta_{01} & \beta_{11} & 0 & \cdots & 0 \\ \beta_{02} & \beta_{12} & \beta_{22} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \beta_{0N} & \beta_{1N} & \beta_{2N} & \cdots & \beta_{NN} \end{bmatrix}, \quad (4.19)$$

where β_{ij} are the coefficients of the normalized legendre polynomial $P_j(t)$ of order j .

In the next section, we prove that the vectors $\mathcal{G}_N(\theta)$ and $\mathcal{H}_N(\theta)$ are, respectively, linearly related to the vectors \mathcal{L}_N and \mathcal{M}_N .

4.3 Basic Results

In this section we establish the basic relationship between the moments of f and the moments of its Radon transform. We show that this relationship essentially yields a decomposition of the Radon transform operator in terms of linear operators mapping a function and its projections to their respective moments. In the case where a finite number of viewing angles are given along with complete, continuous data in the variable t , we derive analytic constraints on the maximum number of moments of a function f that may be *uniquely* determined from a finite set of (noiseless) projection measurements of f .

4.3.1 Relating the Moments of f to the Moments of $g = \mathfrak{R}f$

To arrive at the explicit relationship between the moments of f and the moments of g , recall Theorem 1 of Chapter 2 on page 29. From this theorem, we have that if $F(t)$ is any square-integrable function over $[-1, 1]$, the following is true.

$$\int_{-1}^1 g(t, \theta) F(t) dt = \iint_{\mathcal{O}} f(x, y) F(x \cos(\theta) + y \sin(\theta)) dx dy \quad (4.20)$$

The expressions relating the geometric and orthogonal moments of f to the geometric and orthogonal moments of g are arrived at by letting $F(t) = t^k$ and $F(t) = P_k(t)$,

respectively. First let $F(t) = t^k$ and expand the resulting term $(\cos(\theta)x + \sin(\theta)y)^k$ on the right hand side of (4.20) according to the Binomial Theorem. This yields:

$$H^{(k)}(\theta) = \int_{-1}^1 g(t, \theta) t^k dt = \sum_{j=0}^k \binom{k}{j} \cos^{k-j}(\theta) \sin^j(\theta) \mu_{k-j,j}, \quad (4.21)$$

where $H^{(k)}(\theta)$ is simply the k^{th} order geometric moment of the *projection* function $g(t, \theta)$ for a fixed angle θ . Hence, we have arrived at an explicit expression relating the moments of f to those of g . Defining

$$D^{(k)}(\theta) = [\gamma_{k,0} \cos^k(\theta) \mid \gamma_{k,1} \cos^{k-1}(\theta) \sin(\theta) \mid \cdots \mid \gamma_{k,k} \sin^k(\theta)], \quad (4.22)$$

$$\gamma_{k,j} = \binom{k}{j}, \quad (4.23)$$

the identity (4.21) can be concisely expressed as

$$H^{(k)}(\theta) = D^{(k)}(\theta) \mu^{(k)}. \quad (4.24)$$

Note that this expression states that moments of order k of the projections depend *only* on moments of order k , $\mu^{(k)}$, of $f(x, y)$. Now invoking this last expression for $k = 0, 1, \dots, N$, we can write

$$\begin{bmatrix} H^{(0)}(\theta) \\ H^{(1)}(\theta) \\ \vdots \\ H^{(N)}(\theta) \end{bmatrix} = \begin{bmatrix} D^{(0)}(\theta) & 0 & \cdots & 0 \\ 0 & D^{(1)}(\theta) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & D^{(N)}(\theta) \end{bmatrix} \begin{bmatrix} \mu^{(0)} \\ \mu^{(1)} \\ \vdots \\ \mu^{(N)} \end{bmatrix} \quad (4.25)$$

$$\mathcal{H}_N(\theta) = \mathcal{D}_N(\theta) \mathcal{M}_N \quad (4.26)$$

where the obvious association is made in the last equation and \mathcal{M}_N is defined in (4.11). By invoking (4.12) and (4.18) we arrive at the following relationship between

the orthogonal moments of f and the orthogonal moments of g .

$$\mathcal{G}_N(\theta) = B_N \mathcal{D}_N(\theta) T_N \mathcal{L}_N \quad (4.27)$$

In the above expression, N is a finite integer. The relationship (4.27), however, can be expressed similarly when all the moments of f and g are considered. (That is to say, letting $N = \infty$.) To this end, we need to define the appropriate Hilbert spaces over which the relevant quantities exist. Referring to Figure 4-1, consider the function $f \in L^2(D)$ and its Radon transform $g(t, \theta) \in \mathcal{C} \subset L^2([-1, 1] \times [0, 2\pi])$ where \mathcal{C} denotes the set of valid Radon transforms defined by Theorem 2 of Chapter 2 on page 31. Let Ω denote the operator mapping f onto its set of orthogonal moments $\{\lambda_{pq}\}$ defined as follows.

$$\Omega : f \in L^2(D) \rightarrow \mathcal{L} \in \mathcal{Ra}(\Omega) \subset l^2, \quad (4.28)$$

where \mathcal{L} is simply the moment vector \mathcal{L}_N with $N = \infty$, and l^2 denotes the Hilbert space of square-summable infinite sequences of real numbers. Similarly, define the operator M mapping $g(t, \theta)$ onto its orthogonal moments $\{G^{(k)}(\theta)\}$ for each fixed θ as follows.

$$M : g \in \mathcal{C} \subset L^2([-1, 1] \times [0, 2\pi]) \rightarrow \mathcal{G}(\theta) \in \mathcal{Ra}(M) \subset l^2([0, 2\pi]), \quad (4.29)$$

where $\mathcal{G}(\theta)$ is simply the moment vector $\mathcal{G}_N(\theta)$ with $N = \infty$.

Now dropping the explicit dependence of the quantities in (4.27) on N we can write

$$\mathcal{G} = A(\theta) \mathcal{L} \quad (4.30)$$

$$A(\theta) = B \mathcal{D}(\theta) T, \quad (4.31)$$

where the operators B , $\mathcal{D}(\theta)$, and T are interpreted simply as (infinite-dimensional) operator counterparts of the quantities in (4.27), and the operator $A = B \mathcal{D} T$ is de-

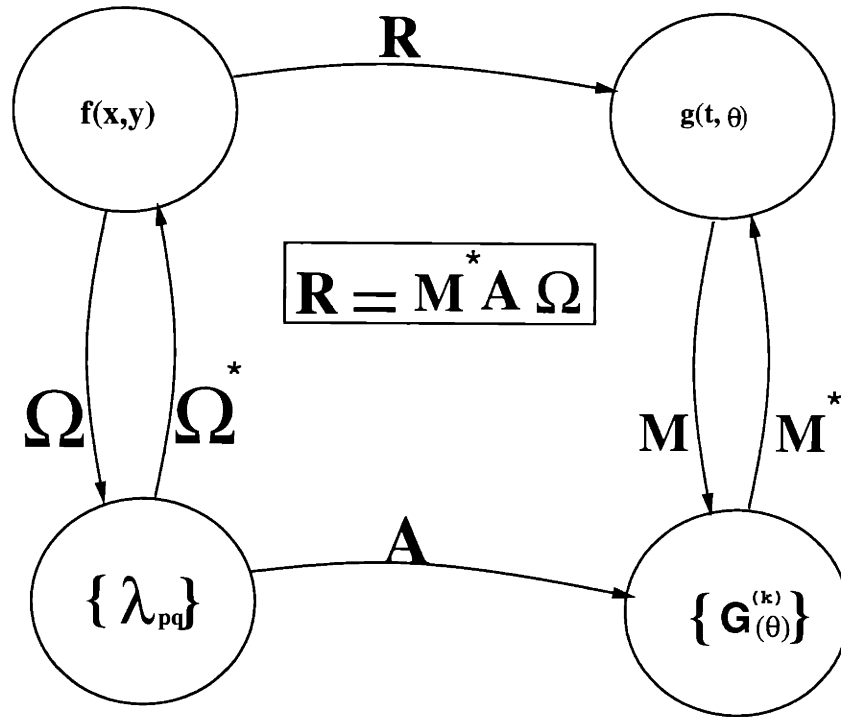


Figure 4-1: A diagram of the decomposition of the Radon transform

defined for convenience. As is now apparent from the commutator diagram shown in Figure 4-1, the *linear* interdependence between the orthogonal moments of f and the orthogonal moments of g naturally results in a decomposition of the Radon transform operator $\mathfrak{R} = M^* A \Omega$. Note that in Figure 4-1, the respective inverses of the operators M and Ω are displayed as their adjoints M^* and Ω^* . This is a consequence of the fact that both of these operators essentially amount to an orthonormal transformations applied to functions in their respective domains. In particular, both operators are isometric (from Parseval's Theorem [62]) and surjective (or onto their respective ranges $\mathcal{R}a(M)$ and $\mathcal{R}a(\Omega)$) so that both of these operators are unitary. (i.e. their inverses coincide with their adjoints.)

4.3.2 Analytic Constraints on the Number of Uniquely Computable Moments

Given the linear relationship between the moments of a function and those of its Radon transform (4.26), in this section we wish to address the following question:

“Given a finite number p of noise-free projections at distinct angles, what is the maximum N for which \mathcal{M}_N can be uniquely determined?” This is in essence a way of asking to what extent a finite number of projections uniquely determines a function $f(x, y)$. This issue is of practical importance since in many instances such as Ocean Acoustic Tomography [65, 50] where due to a variety of practical difficulties, such as size of objects being imaged, only a very limited set of projection data is available.

The question of reconstructability² from a finite number of projections has been studied rather carefully in the mathematical literature on tomography for the particular case of binary polygonal objects [101, 52, 22, 27, 26]. Results are also available for the general case [90], where it is shown that no finite number of projections will, in general, suffice to uniquely determine a function. The question we have posed in this section in essence seeks to find that “part” of a function that *is* uniquely determined by a finite number of projections from different directions.

In this section, we will prove that one may uniquely recover the first p geometric moments of a function $f(x, y) \in L^2(D)$ from a fixed number p of Radon-transform [39] projections from non-congruent directions. We further show that one **can not uniquely** recover any higher order moments of $f(x, y)$ from such limited information. Note that the operative word here is “unique”. As we shall see, the fact that only the first p moments of f are uniquely defined by p distinct projections does not mean that this given data set is void of information about higher order moments of f . The importance of this result lies in the fact that it directly shows to what extent a limited number of projections of a function determine the function uniquely. This, in essence, is a precise notion of the (unique) geometric complexity that a limited set of projections can support. We will use the results of this section directly in the

²Reconstructability refers to a situation where a function can be uniquely recovered from only a finite number of its projections.

development of optimal estimation algorithms for computing the moments of $f(x, y)$ from a finite number of noisy projections later in this chapter.

A function $f(x, y) \in L^2(D)$ is uniquely and completely determined by the complete set of its geometric or orthogonal moments $\mu^{(k)}, \lambda^{(k)}$, for all $k \geq 0$ [97, 96]. If only a finite number of moments of f are given, then an infinity of functions exist that match those moments. If the collection of all such functions is considered as a subset of $L^2(D)$, then in essence, we seek to find the one function that can be uniquely identified given a finite set of projections $g(t, \theta_i)$ for $i = 1, \dots, m$. We assume that we are given (noise-free) integral projections of $f(x, y)$ at a fixed number p of angles θ_i . Essentially we are given “cuts” of the Radon transform $g(t, \theta)$ of $f(x, y)$ at a finite number of θ . Note that each projection $g(t, \theta_i)$ itself is uniquely and completely defined by the complete set of its geometric moments $H^{(k)}(\theta_i)$ for all $k \geq 0$. In Section 4.2 we showed that $\mathcal{H}_N(\theta)$ and \mathcal{M}_N are related as follows:

$$\mathcal{H}_N(\theta) = \mathcal{D}_N(\theta)\mathcal{M}_N \tag{4.32}$$

Now consider the problem where we observe a number p of (noise free) projections $g(t, \theta_i)$, $i = 1, \dots, p$, and wish to uniquely recover as much of $f(x, y)$ as possible. We look at this problem in the following way: we observe $H^{(k)}(\theta_i) \forall k, i = 1, \dots, p$ and wish to uniquely recover as many of the $\mu^{(j)}$ as possible. Note that we treat moments of a given order, j , as a unit, e.g. even though $\mu^{(1)}$ consists of two numbers, we consider it determined only if both are uniquely determined. Our main result is the following:

Result 5 (p Moments From p Projections) *Given line integral projections of $f(x, y)$ at p different angles θ_i in $[0, \pi)$, one can uniquely determine the first p geometric moment vectors $\mu^{(j)}$, $0 \leq j < p$ of $f(x, y)$. Furthermore, this can be done using only the first p order geometric moments $H^{(k)}(\theta)$, $0 \leq k < p$ of the projections. Conversely, moments of $f(x, y)$ of higher order cannot be uniquely determined from p projections.*

Proof: Let us extract, from each projection, a finite but arbitrary number N of geometric moments of that projection. This simply amounts to computing $\mathcal{H}_N(\theta_i)$ for $i = 1, \dots, p$. Note that since each projection is completely characterized by the complete set of its moments, we recover all the information in our original projections as $N \rightarrow \infty$.

Now consider (4.26) and suppose that we stack up our moment observations at the different angles to obtain:

$$\begin{bmatrix} \mathcal{H}_N(\theta_1) \\ \mathcal{H}_N(\theta_2) \\ \vdots \\ \mathcal{H}_N(\theta_p) \end{bmatrix} = \begin{bmatrix} \mathcal{D}_N(\theta_1) \\ \mathcal{D}_N(\theta_2) \\ \vdots \\ \mathcal{D}_N(\theta_p) \end{bmatrix} \mathcal{M}_N \quad (4.33)$$

$$\mathbf{H}_N = \mathbf{D}_N \mathcal{M}_N \quad (4.34)$$

so that the matrix \mathbf{D}_N relating \mathcal{M}_N and the projections is $pN \times \frac{(N+1)(N+2)}{2}$. Clearly, if \mathbf{D}_N has full column rank, then \mathcal{M}_N is uniquely determined. Now we may rearrange the rows of (4.33), grouping together the moments of the same order from all projections, without changing the column rank of \mathbf{D}_N . This operation yields the following

equivalent equation:

$$\begin{bmatrix} H^{(0)}(\theta_1) \\ H^{(0)}(\theta_2) \\ \vdots \\ H^{(0)}(\theta_p) \\ \hline H^{(1)}(\theta_1) \\ H^{(1)}(\theta_2) \\ \vdots \\ H^{(1)}(\theta_p) \\ \hline \vdots \\ \hline H^{(N)}(\theta_1) \\ H^{(N)}(\theta_2) \\ \vdots \\ H^{(N)}(\theta_p) \end{bmatrix} = \begin{bmatrix} D^{(0)}(\theta_1) & \dots & & & \\ D^{(0)}(\theta_2) & \circ & \dots & & \circ \\ \vdots & & & & \\ D^{(0)}(\theta_p) & & & & \\ \hline & & D^{(1)}(\theta_1) & & \\ \circ & & D^{(1)}(\theta_2) & \ddots & \vdots \\ & & \vdots & \ddots & \\ & & D^{(1)}(\theta_p) & & \\ \hline \vdots & & \ddots & \ddots & \circ \\ \hline & & & & D^{(N)}(\theta_1) \\ \circ & \dots & \circ & & D^{(N)}(\theta_2) \\ & & & & \vdots \\ & & & & D^{(N)}(\theta_p) \end{bmatrix} \begin{bmatrix} \mu^{(0)} \\ \mu^{(1)} \\ \vdots \\ \mu^{(N)} \end{bmatrix} \quad (4.35)$$

$$\begin{bmatrix} \tilde{H}_p^{(0)} \\ \tilde{H}_p^{(1)} \\ \vdots \\ \tilde{H}_p^{(N)} \end{bmatrix} = \begin{bmatrix} \tilde{D}_p^{(0)} & \circ & \dots & \circ \\ \circ & \tilde{D}_p^{(1)} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \circ \\ \circ & \dots & \circ & \tilde{D}_p^{(N)} \end{bmatrix} \begin{bmatrix} \mu^{(0)} \\ \mu^{(1)} \\ \vdots \\ \mu^{(N)} \end{bmatrix} \quad (4.36)$$

where $\tilde{H}_p^{(k)}$ is thus the collection of moments of order k from all the projections and $\tilde{D}_p^{(k)}$ is a matrix relating moments of order k of the object to moments of order k in each of the p projections. Now since the overall matrix is block diagonal, the problem of determining each $\mu^{(k)}$ decouples so that $\tilde{H}_p^{(k)} = \tilde{D}_p^{(k)}\mu^{(k)}$ for any k and p . In fact, we can determine a particular $\mu^{(k)}$ uniquely if and only if the corresponding matrix block:

$$\tilde{D}_p^{(k)} = \begin{bmatrix} D^{(k)}(\theta_1) \\ D^{(k)}(\theta_2) \\ \vdots \\ D^{(k)}(\theta_p) \end{bmatrix} \quad (4.37)$$

has full column rank. Let us now examine the conditions when $\tilde{D}_p^{(k)}$ will have full column rank. Substituting for $D^{(k)}(\theta_i)$ from (4.22) we find that $\tilde{D}_p^{(k)}$ is of the form:

$$\tilde{D}_p^{(k)} = \begin{bmatrix} \gamma_{k,0} \cos^k(\theta_1) & \cdots & \gamma_{k,k-1} \cos(\theta_1) \sin^{k-1}(\theta_1) & \gamma_{k,k} \sin^k(\theta_1) \\ \gamma_{k,0} \cos^k(\theta_2) & \cdots & \gamma_{k,k-1} \cos(\theta_2) \sin^{k-1}(\theta_2) & \gamma_{k,k} \sin^k(\theta_2) \\ \vdots & \vdots & \vdots & \vdots \\ \gamma_{k,0} \cos^k(\theta_p) & \cdots & \gamma_{k,k-1} \cos(\theta_p) \sin^{k-1}(\theta_p) & \gamma_{k,k} \sin^k(\theta_p) \end{bmatrix} \quad (4.38)$$

and the matrix $\tilde{D}_p^{(k)}$ is $p \times (k+1)$. Now $\tilde{D}_p^{(k)}$ will have full column rank if and only if its columns are independent. Note that we must, at least, have $p > k$. The columns will be independent if and only if there is no set of α_i (not all zero) such that:

$$\alpha_0 \cos^k(\theta_i) + \alpha_1 \cos^{k-1}(\theta_i) \sin(\theta_i) + \cdots + \alpha_{k-1} \cos(\theta_i) \sin^{k-1}(\theta_i) + \alpha_k \sin^k(\theta_i) = 0, \quad (4.39)$$

for $1 \leq i \leq p$. In particular, this will be true (when $p > k$) if the homogeneous trigonometric polynomial of order k defined by (4.39) has at most k roots θ_i in $[0, \pi)$. This motivates the following lemma, which we prove in Appendix 4-A.

Lemma 2 (Roots of Homogeneous Trigonometric Polynomial of order k) *The k^{th} order homogeneous trigonometric polynomial equation given by:*

$$\alpha_0 \cos^k(\theta) + \alpha_1 \cos^{k-1}(\theta) \sin(\theta) + \cdots + \alpha_{k-1} \cos(\theta) \sin^{k-1}(\theta) + \alpha_k \sin^k(\theta) = 0 \quad (4.40)$$

has at most k distinct roots in $[0, \pi)$.

Given this lemma we see that $\tilde{D}_p^{(k)}$ will have full column rank if and only if $p > k$, i.e. the number of projections p is greater than the order k of the moment vector $\mu^{(k)}$ that we are interested in. In particular, we can achieve this full column rank if $p = k + 1$, so we use no more than the first p moments of the projection. Also note that this implies that if $\tilde{D}_p^{(n)}$ is of full column rank for some n then so is $\tilde{D}_p^{(k)}$ for all $k < n$.

Now the preceding arguments essentially show the first two statements of the theorem. Given p projections the matrices $\tilde{D}_p^{(k)}$ will have full column rank for $0 \leq k < p$. Thus we can uniquely find the moments $\mu^{(k)}$ from the corresponding $\tilde{H}_p^{(k)}$

and $\tilde{D}_p^{(k)}$. In particular, note that due to the block diagonal structure of (4.36), using higher order moments of the projections is of no help in determining a given moment $\mu^{(k)}$. □

Remarks:

We have shown that to uniquely specify the first p moments of a function $f(x, y)$ one needs at least p integral projections at distinct angles in the interval $[0, \pi)$, and that this number is both necessary and sufficient. It is important to note that this result does not mean that the given p projections are void of information about the higher order moments of f . To make this concrete, consider the case where only one projection $g(t, \theta_1)$ of the function f is given at some fixed angle θ_1 . Then our result states that only the integral of f (i.e. its zero-th order moment) is uniquely determined from this projection. Even though this is true, the shape of $g(t, \theta_1)$ does tell us something about what f and its region of support look like. For example if f is an indicator function over a simply-connected region of the plane and $g(t, \theta_1)$ has a very narrow region of support, we can conclude that the underlying object is eccentric in the direction of θ_1 . Conversely, if $g(t, \theta_1)$ has a wider region of support, we can assume that the underlying object must be wide in the θ_1 direction. This is a trivial instance of how a single projection uniquely determines the area of the underlying object but also gives some (non-unique) idea of its location and orientation (first and second-order moment characteristics).

4.3.3 Specific Results for Binary Polygonal Objects

Here we present some results regarding the moments of binary polygonal objects and the moments of their Radon transforms. The results in this particular section may not seem immediately applicable, but we shall invoke them in other chapters of this thesis.

If the function $f(x, y)$ is the indicator function of a binary N -sided polygonal region \mathcal{O} , then the geometric moments of f can be written as algebraic polynomials

in the coordinates of the vertices of \mathcal{O} as follows.

Result 6 (Moments of Polygons) *The geometric moments μ_{pq} of a binary, N -sided, polygonal object with vertices at (x_j, y_j) , $1 \leq j \leq N$, are given by*

$$\mu_{pq} = \frac{1}{(p+1)2^{p+q+1}} \sum_{j=1}^N (x_{j+1} + x_j)^{p+1} (y_{j+1} + y_j)^q (y_{j+1} - y_j) \quad (4.41)$$

$$= \frac{-1}{(q+1)2^{p+q+1}} \sum_{j=1}^N (x_{j+1} + x_j)^p (y_{j+1} + y_j)^{q+1} (x_{j+1} - x_j), \quad (4.42)$$

with $x_{N+1} = x_1$, and $y_{N+1} = y_1$.

Proof: See Appendix 4-B ³. As we have seen in Chapter 3, this result is useful in studying the moments of binary polygonal objects when their vertices are given explicitly. In particular, this result is invoked in the proof of Lemma 1 of Chapter 3.

On a different note, the application of Result 5 to binary polygonal objects yields some interesting results. A fascinating theorem due to Davis [18, 19] states that a triangle in the plane is uniquely determined by its moments of up to order 3. i.e. $\{\mu^{(0)}, \mu^{(1)}, \mu^{(2)}, \mu^{(3)}\}$. Furthermore, Davis has, in essence, provided an explicit algorithm for reconstructing the triangle from this set of numbers. Our result would imply that exactly 4 projections are sufficient to determine this set of moments. Hence, together with the work of Davis, our result provides a closed form solution to the problem of reconstructing a binary triangular region in the plane from only 4 tomographic projections in the absence of noise. In fact, as we will show in Chapter 6, we have improved Davis' result to show that the vertices (and in some instances the interior) of any nondegenerate simply-connected n -gon are uniquely determined by its geometric (or orthogonal) moments up to order $2n - 3$. (i.e. \mathcal{M}_{2n-3}) This generalization improves on known theoretical results on the reconstructability of binary objects from few (noiseless) projections as discussed in Section 2.2.2 of Chapter 2.

³This result was independently proved by the author and the authors of [78]

4.4 New Interpretations of Classical Reconstruction Algorithms

In this section we present some new interpretations of classical tomographic reconstruction algorithms from noisy projection data based on the results presented in Section 4.3.

Let us consider the following measurement equation for the noisy Radon transform of the function f

$$Y(t, \theta) = \mathfrak{R}f + e(t, \theta) \quad (4.43)$$

where $e(t, \theta)$ is assumed to be white Gaussian noise of intensity σ^2 . (Here we are assuming that Y is given for all $t \in [-1, 1]$ and $\theta \in [0, 2\pi]$.) Then the ML estimate of f based on this data can be written as

$$\hat{f}_{ml} = \arg \min_f \|Y - \mathfrak{R}f\|^2, \quad (4.44)$$

where the norm on the right hand side is that of $L^2([-1, 1] \times [0, 2\pi])$. Denoting the adjoint of the Radon transform by \mathfrak{R}^* , the ML estimate of f is given by

$$\hat{f}_{ml} = (\mathfrak{R}^*\mathfrak{R})^{-1}\mathfrak{R}^*Y. \quad (4.45)$$

This solution corresponds to the classical Convolution-Backprojection (CBP) algorithm [40], where \mathfrak{R}^*Y is the backprojection operation and $(\mathfrak{R}^*\mathfrak{R})^{-1}$ essentially amounts to filtering the resulting backprojection with a ramp (differentiator) filter [67].

Recall from Figure 4-1 that the Radon transform has the following decomposition

$$\mathfrak{R} = M^*A\Omega \quad (4.46)$$

Substituting this decomposition into (4.45) we obtain

$$\hat{f}_{ml} = \Omega^* \underbrace{(A^*A)^{-1}A^*MY}_{\hat{\mathcal{L}}_{ml}} \quad (4.47)$$

Now note that the $\hat{\mathcal{L}}_{ml} = (A^*A)^{-1}A^*MY$ is nothing but the ML estimate of (all) the moments of f , from the given noisy projection data, while Ω^* is the operator which reconstructs a function f from its full set of moments ⁴. Hence, the classical CBP algorithm can be interpreted as a two-step process whereby ML estimates of the moments of f are first estimated from the raw projection data and then the estimate \hat{f}_{ml} is reconstructed by inverting the operator Ω . Similar statements can be made if a prior is given for f . In particular, if a Gaussian Random field prior with zero mean and variance σ^2/γ is given, the corresponding Maximum A-Posteriori or MAP (regularized) solution is obtained by finding the solution of

$$\min_f \|\mathfrak{R}f - Y\|^2 + \gamma\|f\|^2. \quad (4.48)$$

It is easy to show that this solution is given by

$$\hat{f}_{map} = (\mathfrak{R}^*\mathfrak{R} + \gamma I)^{-1}\mathfrak{R}^*Y, \quad (4.49)$$

which corresponds to the classical *Filtered* Back-Projection algorithm [40] where $(\mathfrak{R}^*\mathfrak{R} + \gamma I)^{-1}$ corresponds to filtering the backprojected image with a low-pass filter whose order and cutoff frequency are set by picking an appropriate value of the regularization parameter γ . The decomposition of the Radon transform shown in Figure 4-1 shows that this operation is in essence equivalent to computing a regularized inverse of the operator A . i.e.

$$\hat{f}_{map} = \Omega^*(A^*A + \gamma I)^{-1}A^*MY \quad (4.50)$$

In the next section, we develop Maximum Likelihood algorithms for the estimation of the moments of a function f from noisy measurements of its projections. Later in Chapter 5 we discuss robust and numerically efficient variational algorithms for

⁴Note that here we make the assumption that the estimated moment vector $\hat{\mathcal{L}}_{ml}$ is consistent, i.e. that it is in the range of the operator Ω . Due to noise, this may not be so; and we shall deal with this case in Chapter 5.

computing the regularized inverse of the moment operator Ω from a finite number of estimated orthogonal moments.

4.5 Optimal Estimation of Moments From Noisy Projections

In this section we develop the Maximum Likelihood estimator for the moment vectors \mathcal{M}_N and \mathcal{L}_N of an image $f(x, y)$ from noisy measurements of its Radon transform $g(t, \theta)$. These estimates, along with their computed covariances will be used in Chapter 5 in regularized variational algorithms to reconstruct the underlying image. In Chapter 6, we will use the estimated geometric moments of a binary polygonal object to compute ML estimates of its complex moments. These moments will subsequently be used in novel array processing-type algorithms to reconstruct the vertices of the underlying polygon.

In the development that follows, we present the full details of the derivation of the ML estimator for the case where only a finite number of views are given. In the case where a complete, continuous set of views is available, the development is limited to the final formulae which are obvious extensions of the discrete angle cases. In either case, we assume that complete and continuous data is given in the variable t for each projection⁵. In the discussion of the covariance structure of the estimated moments, the complete-view case can lead to some nice closed-form expressions which are not available for the discrete-view case. To this end, most of the discussion on covariance structures will be centered around the continuous, complete-view scenario.

4.5.1 Measurement and Noise Model

As is assumed throughout this thesis, the projections $g(t, \theta)$ are assumed to be corrupted by zero-mean Gaussian white noise of some known intensity σ^2 , which we shall

⁵Obviously, this is an idealization of the practical scenario. We are essentially assuming that the samples in t are obtained at sufficiently many points to justify the assumption that this data is available continuously in t .

denote by $\mathcal{N}(0, \sigma^2)$. That is, the data is given by

$$Y(t, \theta) = g(t, \theta) + e(t, \theta), \quad (4.51)$$

where $e(t, \theta) \sim \mathcal{N}(0, \sigma^2)$. Define the k^{th} order orthogonal moments of the quantities appearing in (4.51) as follows.

$$Y^{(k)}(\theta) = \int_{-1}^1 Y(t, \theta) P_k(t) dt \quad (4.52)$$

$$G^{(k)}(\theta) = \int_{-1}^1 g(t, \theta) P_k(t) dt \quad (4.53)$$

$$e^{(k)}(\theta) = \int_{-1}^1 e(t, \theta) P_k(t) dt \quad (4.54)$$

where $P_k(t)$ is the k^{th} order normalized Legendre polynomial defined in (4.3). From (4.51) we then have

$$Y^{(k)}(\theta) = G^{(k)}(\theta) + e^{(k)}(\theta), \quad (4.55)$$

where due to the orthonormality of the family $\{P_k(t), k \geq 0\}$, the stochastic processes $\{e^{(k)}(\theta), k \geq 0\}$ are a family of zero-mean, uncorrelated, Gaussian white noise processes of equal intensity σ^2 . In effect, the functions $e^{(k)}(\theta)$ are the coefficients of the Karhunen-Loeve expansion [100] of the white noise process $e(t, \theta)$ in the variable t . Adopting the notation

$$Y_N(\theta) = [Y^{(0)}(\theta), Y^{(1)}(\theta), \dots, Y^{(N)}(\theta)]^T, \quad (4.56)$$

$$\mathcal{G}_N(\theta) = [G^{(0)}(\theta), G^{(1)}(\theta), \dots, G^{(N)}(\theta)]^T, \quad (4.57)$$

$$e_N(\theta) = [e^{(0)}(\theta), e^{(1)}(\theta), \dots, e^{(N)}(\theta)]^T, \quad (4.58)$$

we can write

$$Y_N(\theta) = \mathcal{G}_N(\theta) + e_N(\theta), \quad (4.59)$$

with

$$e_N(\theta) \sim \mathcal{N}(\underline{0}, \sigma^2 I_{N+1}), \quad (4.60)$$

where I_{N+1} denotes the $(N+1) \times (N+1)$ identity matrix. By appealing to (4.27), we have

$$Y_N(\theta) = B_N \mathcal{D}_N(\theta) T_N \mathcal{L}_N + e_N(\theta), \quad (4.61)$$

or simply

$$Y_N(\theta) = A_N(\theta) \mathcal{L}_N + e_N(\theta) \quad (4.62)$$

which is a linear measurement equation for the orthogonal moments of $f(x, y)$ in terms of the orthogonal moments of the noisy projections.

4.5.2 Maximum Likelihood (ML) Moment Estimation

Suppose that we are given samples of $Y_N(\theta)$ at only a finite number m of distinct angles $\{\theta_j, j = 1, \dots, m\}$ in $[0, \pi)$. Then we can stack our observations given by (4.61) as follows

$$\begin{bmatrix} Y_N(\theta_1) \\ Y_N(\theta_2) \\ \vdots \\ Y_N(\theta_m) \end{bmatrix} = (B_N \otimes I_m) \begin{bmatrix} \mathcal{D}_N(\theta_1) \\ \mathcal{D}_N(\theta_2) \\ \vdots \\ \mathcal{D}_N(\theta_m) \end{bmatrix} T_N \mathcal{L}_N + \begin{bmatrix} e_N(\theta_1) \\ e_N(\theta_2) \\ \vdots \\ e_N(\theta_m) \end{bmatrix} \quad (4.63)$$

$$\mathbf{Y}_N = \mathbf{B}_N \mathbf{D}_N T_N \mathcal{L}_N + \mathbf{e}_N \quad (4.64)$$

with

$$\mathbf{e}_N \sim \mathcal{N}(\mathbf{0}, \sigma^2 I_{m(N+1)}) \quad (4.65)$$

where the symbol \otimes denotes the Kronecker product, and $I_{m(N+1)}$ denotes the $m(N+1) \times m(N+1)$ identity matrix. Hence if $\mathbf{B}_N \mathbf{D}_N T_N$ is full column rank, the ML estimate of \mathcal{L}_N is given by

$$\hat{\mathcal{L}}_N = \left(T_N^T \mathbf{D}_N^T \mathbf{B}_N^T \mathbf{B}_N \mathbf{D}_N T_N \right)^{-1} T_N^T \mathbf{D}_N^T \mathbf{B}_N^T \mathbf{Y}_N. \quad (4.66)$$

Since T_N is invertible, this expression simplifies to

$$\widehat{\mathcal{L}}_N = T_N^{-1} \left(\mathbf{D}_N^T \mathbf{B}_N^T \mathbf{B}_N \mathbf{D}_N \right)^{-1} \mathbf{D}_N^T \mathbf{B}_N^T \mathbf{Y}_N. \quad (4.67)$$

Note that in order for this expression to make sense, the term $\left(\mathbf{D}_N^T \mathbf{B}_N^T \mathbf{B}_N \mathbf{D}_N \right)$ must be invertible. Since the matrix \mathbf{B}_N is known to be invertible, the invertibility of $\left(\mathbf{D}_N^T \mathbf{B}_N^T \mathbf{B}_N \mathbf{D}_N \right)$ will only depend on whether \mathbf{D}_N has full column rank. Appealing to Result 5 we deduce that \mathbf{D}_N has full column rank, and hence $\widehat{\mathcal{L}}_N$ is well defined⁶ if, and only if, $m > N \geq 0$. i.e. the number, m , of available projections should strictly exceed the order of the highest order estimated moments.

Using the invariance property of ML estimates [81], from (4.67), and assuming the inverses exist, the ML estimate of \mathcal{M}_N is derived as

$$\widehat{\mathcal{M}}_N = T_N \widehat{\mathcal{L}}_N = \left(\mathbf{D}_N^T \mathbf{B}_N^T \mathbf{B}_N \mathbf{D}_N \right)^{-1} \mathbf{D}_N^T \mathbf{B}_N^T \mathbf{Y}_N. \quad (4.68)$$

Digressing to the continuous angle (complete-view) case, if the (noisy) projection data $Y(t, \theta)$ is available at all angles $\theta \in [0, \pi)$, by appealing to (4.61), the expression for the ML estimate of \mathcal{L}_N takes the following form

$$\widehat{\mathcal{L}}_N^* = T_N^{-1} \left(\int_0^\pi \mathcal{D}_N(\theta)^T B_N^T B_N \mathcal{D}_N(\theta) d\theta \right)^{-1} \left(\int_0^\pi \mathcal{D}_N(\theta)^T B_N^T Y_N(\theta) d\theta \right) \quad (4.69)$$

where we use the superscript * to distinguish full-view estimates from estimates obtained when only a finite number of projections are available.

4.5.3 Statistics of the Estimated Moments

The covariance matrices of the estimated orthogonal moments $\widehat{\mathcal{L}}_N$ and geometric moments $\widehat{\mathcal{M}}_N$ are readily given (when $m > N$) by:

$$Q_N \equiv \text{Cov}(\widehat{\mathcal{L}}_N) = \sigma^2 \left(T_N^T \mathbf{D}_N^T \mathbf{B}_N^T \mathbf{B}_N \mathbf{D}_N T_N \right)^{-1} \quad (4.70)$$

⁶Well defined means it exists and is unique

$$R_N \equiv \text{Cov}(\widehat{\mathcal{M}}_N) = \sigma^2 \left(\mathbf{D}_N^T \mathbf{B}_N^T \mathbf{B}_N \mathbf{D}_N \right)^{-1}. \quad (4.71)$$

By studying the complete-view case, the corresponding covariance matrices reveal some interesting structure which we examine next. In particular, one might hope that the covariance matrix of the estimated orthogonal moments is at least block-diagonal. This, unfortunately, is in general not the case. Also, the trace of the individual covariance matrices may be thought of as some “overall” measure of how poor the corresponding estimated vector is. We will show that for the complete-view case, the trace of the covariance matrix for the estimated geometric moments may be bounded below by a rather easily computed function of the trace of the inverse of this covariance matrix. To begin, note that the expressions for the covariance matrices when a complete, continuous set of projections is available is given by

$$Q_N^* \equiv \text{Cov}(\widehat{\mathcal{L}}_N^*) = \sigma^2 \left(\int_0^\pi T_N^T \mathcal{D}_N(\theta)^T B_N^T B_N \mathcal{D}_N(\theta) T_N d\theta \right)^{-1} \quad (4.72)$$

$$R_N^* \equiv \text{Cov}(\widehat{\mathcal{M}}_N^*) = \sigma^2 \left(\int_0^\pi \mathcal{D}_N(\theta)^T B_N^T B_N \mathcal{D}_N(\theta) d\theta \right)^{-1} \quad (4.73)$$

For $N = 2$, these matrices are given by

$$Q_2^* = \frac{2\sigma^2}{\pi} \begin{bmatrix} \frac{1}{4} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{3}{4} & 0 & \frac{-1}{4} \\ 0 & 0 & 0 & 0 & \frac{2}{5} & 0 \\ 0 & 0 & 0 & \frac{-1}{4} & 0 & \frac{3}{4} \end{bmatrix} \quad (4.74)$$

$$R_2^* = \frac{2\sigma^2}{\pi} \begin{bmatrix} 1 & 0 & 0 & \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & \frac{2}{3} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{2}{3} & 0 & 0 & 0 \\ \frac{1}{3} & 0 & 0 & \frac{17}{45} & 0 & \frac{1}{45} \\ 0 & 0 & 0 & 0 & \frac{8}{45} & 0 \\ \frac{1}{3} & 0 & 0 & \frac{1}{45} & 0 & \frac{17}{45} \end{bmatrix} \quad (4.75)$$

For $N > 2$, the desirable block-diagonal structure of Q_N^* breaks down.

The inverse of the covariance matrix R_N^* has a particularly neat form. We can exploit this fact to arrive at a lower bound on the trace of R_N^* in terms of the trace of its inverse. We can write

$$R_N^{*-1} = \sigma^{-2} \int_0^\pi \mathcal{D}_N(\theta)^T B_N^T B_N \mathcal{D}_N(\theta) d\theta \quad (4.76)$$

Now by referring to (4.22), (4.25) and (4.19) respectively, where $\mathcal{D}_N(\theta)$ and B_N are defined, the trace of R_N^{*-1} can be written as

$$\text{Tr}(R_N^{*-1}) = \sigma^{-2} \int_0^\pi \sum_{k=0}^N \sum_{j=0}^k \sum_{l=0}^j \left[\beta_{jk} \binom{j}{l} \cos^{j-l}(\theta) \sin^l(\theta) \right]^2 d\theta. \quad (4.77)$$

Denoting by $\mathbf{B}(\frac{a+1}{2}, \frac{b+1}{2})$ the standard Beta function [34] given by

$$(1/2)\mathbf{B}\left(\frac{a+1}{2}, \frac{b+1}{2}\right) = \int_0^{\pi/2} \sin^a(\theta) \cos^b(\theta) d\theta \quad (4.78)$$

for nonnegative integers a and b we obtain

$$\text{Tr}(R_N^{*-1}) = \sigma^{-2} \sum_{k=0}^N \sum_{j=0}^k \sum_{l=0}^j \beta_{jk}^2 \binom{j}{l}^2 \mathbf{B}\left(j-l + \frac{1}{2}, l + \frac{1}{2}\right) \quad (4.79)$$

To obtain a lower bound on the trace of the covariance matrix R_N^* itself we use the following result, the proof of which may be found in appendix 4-C.

Result 7 *Let A be an $n \times n$ real, symmetric strictly positive definite matrix. Then the following identity holds true*

$$\frac{n^2}{\text{Tr}(A^{-1})} \leq \text{Tr}(A). \quad (4.80)$$

Furthermore, the inequality can not be improved.

Given this result, we obtain the following lower bound

$$\frac{(N+1)^2(N+2)^2}{4Tr(R_N^{*-1})} \leq Tr(R_N^*). \quad (4.81)$$

Note the fact that computing the trace of R_N^* is quite a messy affair, while computing the trace of R_N^{*-1} is, as we have shown in (4.79) is rather simple via the use of the Beta function. Hence, the lower bound (4.81) serves in essence only as a quick way of getting a rough lower bound on the quality of the estimated geometric moment vector. In at least this sense, it is reminiscent of the Cramer-Rao bound for nonlinear ML estimation [100].

4.6 Other Results on Moments and the Radon Transform

So far in this Chapter we have shown that the moments of a function $f(x, y)$ can be computed from the moments of its Radon Transform $g(t, \theta)$ in a straightforward (linear) fashion. We have developed an optimal ML algorithm for the estimation of moments of a function $f(x, y)$ from noisy measurements of its Radon transform. In this section, we wish to discuss specific advantages and insights that may be gained from the study of moments of the Radon transform of a function. Specifically, we concentrate on the following four issues.

1. Not every square-integrable function $g(t, \theta)$, over $[-1, 1] \times [0, 2\pi]$, is a Radon transform of some square-integrable $f(x, y)$ defined over the unit disk. Certain well-known necessary and sufficient conditions, widely known as the Consistency Conditions (see Theorem 2 of Chapter 2), must be satisfied by $g(t, \theta)$ to be a valid Radon transform. In particular, a given data set $Y(t, \theta)$ may not be a valid Radon transform due to the effect of noise. A valid Radon transform may be obtained from Y by projecting it onto the set of valid Radon transforms. We show that this projection can be obtained from the moments of the Radon transform by simply computing the ML estimates of the parameters (moments) that define a valid Radon transform.

2. We show that computing the orthogonal moments of a noisy data set $Y(t, \theta)$ amounts to a linear transformation of the data. We will use this observation to deduce that the computed moments of the data are sufficient statistics for the estimation of a function $f(x, y)$ from this data.
3. Given a finite number of views, the explicit relationship between the orthogonal moments of the data $Y(t, \theta)$ and the (ML) estimated orthogonal moments for the underlying function $f(x, y)$ is given by (4.67). The right hand side of (4.67) depends on the inverse of the matrix $\mathbf{D}_N^T \mathbf{B}_N^T \mathbf{B}_N \mathbf{D}_N$. The condition number of this matrix, in turn, depends on the number of available views and the particular geometry of the acquisition system (i.e. the values of the angles θ_i). Hence, by studying the condition number of $\mathbf{D}_N^T \mathbf{B}_N^T \mathbf{B}_N \mathbf{D}_N$ with respect to these parameters, we can identify sampling schemes that provide the most numerically stable algorithms for the estimation of the moments of the underlying object $f(x, y)$.

4.6.1 Consistent Estimation Using Moments

Consider the problem of finding the “closest” function \hat{g} (in a sense to be made precise later) to the given data Y that is a valid Radon Transform of some object $f(x, y)$. We will henceforth refer to this problem as the *consistent estimation problem*. We shall separate the problem into three distinct cases that we distinguish by considering whether I) data is available continuously in both variables t and θ II) data is available continuously in t but discretely in θ or III) data is given discretely in both t and θ .

The t -Continuous/ θ -Continuous (CC) case

In this section we assume that the data set Y is given for all values of $t \in [-1, 1]$ and $\theta \in [0, 2\pi]$ ⁷ The explicit problem we wish to solve here is that of finding a function \hat{g} which is closest to the data Y in the L^2 sense subject to the constraint that \hat{g} is a

⁷Note that from a practical point of view, due to the symmetry of the Radon transform, data from only the interval $[0, \pi]$ can be considered complete.

valid Radon transform of some square-integrable function f with support in the unit disk. Formally we have

Problem 1

$$\min_{g \in \mathcal{C}} \|Y - g\|^2 \tag{4.82}$$

Where \mathcal{C} denotes the set of all valid Radon transforms and the norm is taken to be the L^2 norm over the Hilbert space of square integrable functions g over $[-1, 1] \times [0, 2\pi]$ given by

$$\|g\|^2 = \int_0^{2\pi} \int_{-1}^1 g^2(t, \theta) dt d\theta \tag{4.83}$$

The set \mathcal{C} is determined by the Classical result of Helgason-Ludwig [39, 59] which we presented as Theorem 2 of Chapter 2 and which we repeat here for the sake of continuity.

Theorem 7 ([39, 59]) *Let $L^2(D)$ denote the space of all square-integrable functions $f : D \rightarrow \mathcal{R}$ with support inside the closed unit disk D . In order for $g(t, \theta)$ to be the Radon transform of some function $f \in L^2(D)$, it is necessary and sufficient that*

1. $g \in L^2([-1, 1] \times [0, 2\pi])$,
2. $g(t, \theta + \pi) = g(-t, \theta)$, and
3. the integral

$$\int_{-1}^1 g(t, \theta) t^k dt \tag{4.84}$$

be a homogeneous polynomial of degree k in $\cos(\theta)$ and $\sin(\theta)$ for every positive integer k .

In order for Problem 1 to have a solution, and in particular one that is unique, it suffices to show that the set \mathcal{C} is closed and convex. We have the following Lemma.

Lemma 3 *The set \mathcal{C} is a closed and convex subset of $L^2([-1, 1] \times [0, 2\pi])$.*

Proof: See Appendix 4-D.

Given this Lemma, the solution to Problem 1 is simply given by the projection of the function $Y(t, \theta)$ onto the set \mathcal{C} . Let the projection operator onto the consistency set \mathcal{C} be denoted by $\Pi_{\mathcal{C}}$. We now show that this projection operator can be decomposed using the operator M which maps a given function in $L^2([-1, 1] \times [0, 2\pi])$ onto its set of orthogonal moments for every fixed θ as defined in (4.29). Referring to Figure 4-2, let us denote the range of this operator as $\Lambda = \mathcal{R}a(M)$ and let the image of \mathcal{C} under the operator M be denoted by $\Lambda_{\mathcal{C}}$. Furthermore, denote the projection operator onto $\Lambda_{\mathcal{C}}$ by $\Pi_{\Lambda_{\mathcal{C}}}$. Then, as illustrated in Figure 4-2, the solution to Problem 1 can be computed by the following process:

1. Compute the orthogonal moments MY of Y ,
2. Project MY onto the set $\Lambda_{\mathcal{C}}$,
3. Apply the inverse M^* of M to the result.

This process is justified by the following Result.

Result 8 *Consider the linear operator $T = M^*\Pi_{\Lambda_{\mathcal{C}}}M$ mapping $L^2([-1, 1] \times [0, 2\pi])$ onto itself. Then T is a projection operator onto the consistency set \mathcal{C} . That is,*

$$T = M^*\Pi_{\Lambda_{\mathcal{C}}}M = \Pi_{\mathcal{C}} \quad (4.85)$$

Proof: We have

$$T^* = M^*\Pi_{\Lambda_{\mathcal{C}}}^*M = M^*\Pi_{\Lambda_{\mathcal{C}}}M = T \quad (4.86)$$

$$TT = M^*\Pi_{\Lambda_{\mathcal{C}}}MM^*\Pi_{\Lambda_{\mathcal{C}}}M = M^*\Pi_{\Lambda_{\mathcal{C}}}M = T \quad (4.87)$$

Hence T is a projection. Since T is a projection, then there exists one and only one closed linear subspace $S \subset L^2([-1, 1] \times [0, 2\pi])$ such that $T = \Pi_S$ [5]. In particular, $S = \mathcal{R}a(T)$; and since $\mathcal{R}a(T) = \mathcal{C}$, the result is established. \square

As a direct consequence of this result, the solution to Problem 1 can be written in

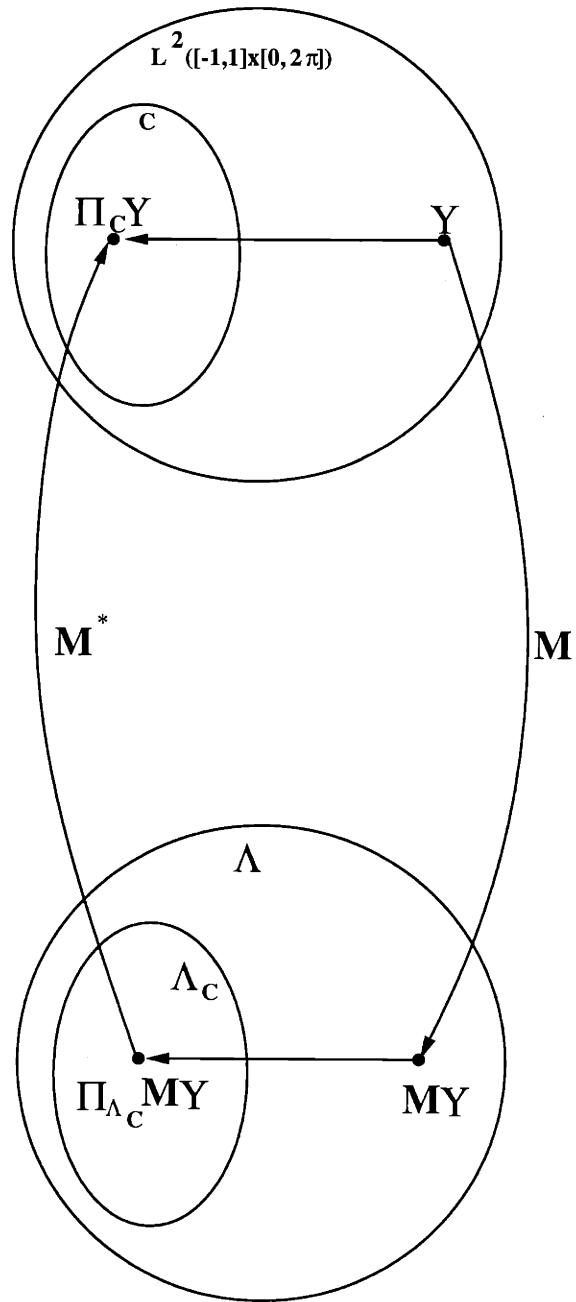


Figure 4-2: A diagram for the proof of Result 8

terms of the following expansion:

$$\hat{g}(t, \theta) = \sum_{k=0}^{\infty} \hat{G}^{(k)}(\theta) P_k(t). \quad (4.88)$$

where $\hat{G}^{(k)}(\theta)$ are simply the projections of the *orthogonal* moments of the data $Y^{(k)}(\theta)$ onto the set $\Lambda_{\mathcal{C}}$. i.e. $\hat{G}^{(k)}(\theta)$ are the image of $Y^{(k)}(\theta)$ under $\Pi_{\Lambda_{\mathcal{C}}}$. According to the consistency Theorem 7, the elements of the consistency set \mathcal{C} are those functions $g(t, \theta)$ for which $G^{(k)}(\theta)$ are sums of homogeneous polynomials of order no greater than k . The conditions of Theorem 7 show that $g \in \mathcal{C}$ if and only if the Fourier series expansion of $G^{(k)}(\theta)$ has the following form [72, 58].

$$G^{(k)}(\theta) = \sum_{\substack{n=0 \\ n+k \text{ even}}}^k a_n^{(k)} \cos(n\theta) + b_n^{(k)} \sin(n\theta) \quad (4.89)$$

$\hat{G}^{(k)}(\theta)$ is now obtained by projecting $Y^{(k)}(\theta)$ onto the set $\Lambda_{\mathcal{C}}$. That is, by solving the following optimization problem⁸:

$$\min_{G^{(k)}(\theta) \in \Lambda_{\mathcal{C}}} \|Y^{(k)}(\theta) - G^{(k)}(\theta)\|^2 \quad (4.90)$$

But $G^{(k)}(\theta) \in \Lambda_{\mathcal{C}}$ if and only if $G^{(k)}(\theta)$ has the form given in (4.89) for all $k \geq 0$. Hence, we can rewrite (4.90) as

$$\min_{a_n^{(k)}, b_n^{(k)}} \|Y^{(k)}(\theta) - \sum_{\substack{n=0 \\ n+k \text{ even}}}^k a_n^{(k)} \cos(n\theta) + b_n^{(k)} \sin(n\theta)\|^2 \quad (4.91)$$

where the minimization over $a_n^{(k)}$ and $b_n^{(k)}$ is now unrestricted.

Now recalling from (4.55) that $Y^{(k)}(\theta)$ is corrupted by white Gaussian noise, the solution of (4.90) can now be given from the Maximum Likelihood estimates of the coefficients $a_n^{(k)}$ and $b_n^{(k)}$ as follows:

⁸The norm used here is that of $L^2([0, 2\pi])$

$$\widehat{G}^{(k)}(\theta) = \sum_{\substack{n=0 \\ n+k \text{ even}}}^k \widehat{a}_n^{(k)} \cos(n\theta) + \widehat{b}_n^{(k)} \sin(n\theta) \quad (4.92)$$

where $\widehat{a}_n^{(k)}$ and $\widehat{b}_n^{(k)}$ are the respective ML estimates of the Fourier coefficients $a_n^{(k)}$ and $b_n^{(k)}$ obtained from 4.91.

The t -Continuous/ θ -Discrete (CD) case

In this section we study the analog of Problem 1 for the case where the data $Y(t, \theta)$ is given for all $t \in [-1, 1]$, but only for a finite number m of non-congruent viewing directions $\theta_j \in [0, 2\pi]$, $j = 1, \dots, m$. This analogous problem is given by ⁹

Problem 2

$$\min_{\{g(t, \theta_j), 1 \leq j \leq m\} \in \mathcal{C}_m} \sum_{j=1}^m \|Y(t, \theta_j) - g(t, \theta_j)\|^2 \quad (4.93)$$

where \mathcal{C}_m denotes the set of all collections of m projections from fixed, non-congruent, angles θ_j which are consistent.

The description of the consistency set in this case is somewhat different from the CC case. For each angle θ_j , consider the geometric moments of the projection data $Y(t, \theta_j)$ given by¹⁰

$$\Upsilon^{(k)}(\theta_j) = \int_{-1}^1 Y(t, \theta_j) t^k dt. \quad (4.94)$$

In order for these values to correspond to a set of consistent samples, $\Upsilon^{(k)}(\theta_j)$ must be the samples of a homogeneous trigonometric polynomial in $\sin(\theta)$ and $\cos(\theta)$ of order k for all $k \geq 0$. However, using Lemma 2 on page 103, we can show that the values of a homogeneous polynomial of order k in $\sin(\theta)$ and $\cos(\theta)$ can be prescribed arbitrarily at any $k+1$ points on the unit circle. That is to say, any homogeneous polynomial of order k in $\sin(\theta)$ and $\cos(\theta)$ is completely determined by its values measured at any $k+1$ (noncongruent modulo π) points on the unit circle. Hence, the condition that $\Upsilon^{(k)}(\theta_j)$ be the values of homogeneous polynomials of order k for all $k \geq 0$ are trivially

⁹The norm used here is that of $L^2([-1, 1])$

¹⁰Note that $\Upsilon^{(k)}(\theta)$ and $Y^{(k)}(\theta)$ respectively denote the geometric and orthogonal moments of $Y(t, \theta)$ and should not be confused.

satisfied for $k \geq m-1$. Therefore, the set of these conditions that are binding becomes finite, and the consistency set \mathcal{C}_m is defined by only a finite number of constraints. The question arises as to whether this finite number of conditions is also sufficient for consistency. The answer has been provided by Smith et al. [90] in the form of the following Theorem. Let $g_f(t, \theta_j)$ denote the Radon Transform projection of $f(x, y)$ at the angle θ_j .

Theorem 8 ([90]) *Let $\{Y(t, \theta_j), 1 \leq j \leq m, \theta_i \neq \theta_j \pmod{\pi} \in [0, 2\pi]\}$ be square integrable and vanish for $|t| \geq 1$, and assume that for each $k \leq m-2$, the geometric moments*

$$Y^{(k)}(\theta_j) = \int_{-1}^1 Y(t, \theta_j) t^k dt \quad (4.95)$$

of $Y(t, \theta_j)$ are the values of a homogeneous polynomial of degree k . Then for each $\epsilon > 0$ there is a square integrable function $f(x, y)$ with support in the closed unit disk of radius $1 + \epsilon$ centered at the origin such that $g_f(t, \theta_j) = Y(t, \theta_j)$ for $1 \leq j \leq m$.

This theorem provides an explicit description of the set \mathcal{C}_m in terms of constraints which involve the values of the geometric moments of the given discrete-view data set. To draw a parallel analysis with the CC case, we interpret the statement of Theorem 8 in terms of already familiar operators defined for the CC case.

Let us stack up the given projections to define the following data vector.

$$Y_m(t) = [Y(t, \theta_1) \mid Y(t, \theta_2) \mid \cdots \mid Y(t, \theta_m)]^T \quad (4.96)$$

Hence,

$$Y_m(t) : [-1, 1] \rightarrow \underbrace{L^2([-1, 1]) \times \cdots \times L^2([-1, 1])}_{m \text{ times}} \quad (4.97)$$

The operator M defined in (4.29) may be applied to each element of the vector $Y_m(t)$ to obtain the *orthogonal* moments of each projection. This can be written as $M_m Y_m(t)$ where M_m denotes the operator which maps each element of the vector $Y_m(t)$ to its

set of orthogonal moments. In particular,

$$M_m = \text{diag}(\underbrace{M, M, \dots, M}_{m \text{ times}}), \quad (4.98)$$

so that

$$M_m : \underbrace{L^2([-1, 1]) \times \dots \times L^2([-1, 1])}_{m \text{ times}} \rightarrow \underbrace{l^2 \times \dots \times l^2}_{m \text{ times}}, \quad (4.99)$$

$$M_m Y_m(t) = [MY(t, \theta_1) \mid MY(t, \theta_2) \mid \dots \mid MY(t, \theta_m)]^T, \quad (4.100)$$

where l^2 denotes the set of infinite length, square-summable, sequences of real numbers. Now denote the image of the discrete-view consistency set \mathcal{C}_m by $\Lambda_{\mathcal{C}_m}$ and denote by $\Pi_{\mathcal{C}_m}$ and $\Pi_{\Lambda_{\mathcal{C}_m}}$ the projection operators onto these respective sets. Analogously to Result 8, we have

Result 9 Consider the linear operator $T_m = M_m^* \Pi_{\Lambda_{\mathcal{C}_m}} M_m$ mapping the space $\underbrace{L^2([-1, 1]) \times \dots \times L^2([-1, 1])}_{m \text{ times}}$ onto itself. Then T_m is a projection operator onto the discrete-view consistency set \mathcal{C}_m . That is,

$$T_m = M_m^* \Pi_{\Lambda_{\mathcal{C}_m}} M_m = \Pi_{\mathcal{C}_m} \quad (4.101)$$

Proof: Note that from the definition of the operator M_m , it follows that it too is a unitary operator like M . The proof of this result then is essentially identical to the proof of Result 8. \square

Hence the solution to Problem 2 is given by

$$\hat{g}_m(t) = \Pi_{\mathcal{C}_m} Y_m(t) = M_m^* \Pi_{\Lambda_{\mathcal{C}_m}} M_m Y_m(t) \quad (4.102)$$

where

$$\hat{g}_m(t) = [\hat{g}(t, \theta_1) \mid \hat{g}(t, \theta_2) \mid \dots \mid \hat{g}(t, \theta_m)]^T \quad (4.103)$$

Due to the fact that the discrete-view consistency set \mathcal{C}_m is defined by only a finite number of constraints, the solution $\hat{g}_m(t)$ to Problem 2 will only involve the *nontrivial* estimation of a finite number of coefficients. To see this, let us decompose

the function $M_m Y_m(t)$ into two orthogonal pieces as follows

$$M_m Y_m(t) = \Pi_{\mathcal{A}} M_m Y_m(t) + (I - \Pi_{\mathcal{A}}) M_m Y_m(t) \quad (4.104)$$

where \mathcal{A} denotes the following subspace of $\underbrace{L^2([-1, 1]) \times \cdots \times L^2([-1, 1])}_{m \text{ times}}$

$$\mathcal{A} = \text{span} \left\{ \underbrace{[P_k(t) \mid \cdots \mid P_k(t)]^T}_{m \text{ times}}, 0 \leq k \leq m-2 \right\} \quad (4.105)$$

and $\Pi_{\mathcal{A}}$ denotes the projection operator onto \mathcal{A} , and $I - \Pi_{\mathcal{A}}$ denotes the projection operator onto the orthogonal complement of \mathcal{A} denoted by \mathcal{A}^\perp . The subspace \mathcal{A} contains all functions in $\underbrace{L^2([-1, 1]) \times \cdots \times L^2([-1, 1])}_{m \text{ times}}$ whose moments of order equal to, or larger than, $m-1$ are identically zero; while the subspace \mathcal{A}^\perp contains all functions in $\underbrace{L^2([-1, 1]) \times \cdots \times L^2([-1, 1])}_{m \text{ times}}$ whose orthogonal (and therefore geometric) moments of order up to, and including, $m-2$ are identically zero. Hence, all functions in \mathcal{A}^\perp trivially satisfy the constraints of Theorem 8; that is to say,

$$(I - \Pi_{\mathcal{A}}) M_m Y_m(t) \in \Lambda_{c_m}. \quad (4.106)$$

Hence, $(I - \Pi_{\mathcal{A}}) M_m Y_m(t)$ is invariant under the projection operator $\Pi_{\Lambda_{c_m}}$, i.e.

$$\Pi_{\Lambda_{c_m}} (I - \Pi_{\mathcal{A}}) M_m Y_m(t) = (I - \Pi_{\mathcal{A}}) M_m Y_m(t). \quad (4.107)$$

Now substituting the decomposition (4.104) into the right hand side of (4.102), we obtain

$$\hat{g}_m(t) = M_m^* \Pi_{\Lambda_{c_m}} \Pi_{\mathcal{A}} M_m Y_m(t) + M_m^* \Pi_{\Lambda_{c_m}} (I - \Pi_{\mathcal{A}}) M_m Y_m(t), \quad (4.108)$$

which after invoking (4.107), becomes

$$\hat{g}_m(t) = M_m^* \Pi_{\Lambda_{c_m}} \Pi_{\mathcal{A}} M_m Y_m(t) + M_m^* (I - \Pi_{\mathcal{A}}) M_m Y_m(t), \quad (4.109)$$

Note that the second term on the right hand side of (4.109) does not involve the constraint set \mathcal{C}_m at all. In fact, this term is a function which has orthogonal moments up to order $m - 2$ which are identically zero, and its remaining orthogonal moments are precisely those of the original data vector $Y_m(t)$. To explore the structure of (4.109) more carefully, we can study one component of the vector $\hat{g}_m(t)$ as follows. Equation 4.109 directly implies that for every $j = 1, \dots, m$, we must have

$$\hat{g}(t, \theta_j) = \sum_{k=0}^{m-2} \hat{G}^{(k)}(\theta_j) P_k(t) + \sum_{k=m-1}^{\infty} Y^{(k)}(\theta_j) P_k(t) \quad (4.110)$$

where

$$Y^{(k)}(\theta_j) = \int_{-1}^1 Y(t, \theta_j) P_k(t) dt \quad (4.111)$$

and $\hat{G}^{(k)}(\theta_j)$ is obtained from (4.91) and (4.92) and evaluated at $\theta = \theta_j$. Hence, similar to the CC case, the solution to Problem 2 amounts to computing Maximum Likelihood estimates of the parameters $a_n^{(k)}$ and $b_n^{(k)}$ with $n + k$ even, but with k only in the range $0 \leq k \leq m - 2$.

The discrete t , discrete θ (DD) case

In this case, we assume that the projection data is given at a finite number of angles and that, at each angle, only a finite number of samples in the variable t are given. As is pointed out in [90], as a corollary of Theorem 8, we have the following result.

Corollary 3 *A finite number of noisy projections read at a finite number of points cannot be inconsistent.*

This result states that any finite data set is consistent. It is interesting to note that ideally one would like the number of samples in t to be very large so that there is redundancy and better resolution in the data, but if we have a continuous data set in t , we have to worry about the possibility of inconsistency in the data.

4.6.2 Moments as Sufficient Statistics

In this Section we wish to highlight an important aspect of the study of moments of the Radon transform. Namely, the idea that, in general, the complete set of moments of the (noisy) Radon transform of a function f constitute a set of sufficient statistics for the estimation of the underlying function f . In fact, we show that in the case of a finite number of views with a finite number of samples per view, only a finite number of moments of the projection constitute a set of sufficient statistics. The primary reason for the study of moments as sufficient statistics is that it allows one to embed the problem of tomographic reconstruction in the context of approximation at different scales of resolution. In fact, even if the final goal is not the reconstruction of a function f , moments are useful as features for classification, diagnosis and, in general, data compression in applications such as Non-Destructive Evaluation of materials where a complete reconstruction of the underlying image may be unnecessary [40]. In such a scenario, the final goal is to answer a question such as whether a certain feature is present in the underlying object or not. In these applications, the moments of the projections can be used to estimate the moments of the underlying object and from these a statistically sound answer may be provided.

We will show that if the objective is to optimally estimate a function $f(x, y)$ from a set of noisy projection data Y , the set of orthogonal moments of Y is a sufficient statistic for f . This is to say that once we have computed the moments of Y , the original data may be discarded, for the set of moments will contain exactly the same information as the original data since this calculation of the moments is simply an orthonormal transformation of the data.

Recall the measurement equation for the noisy Radon transform

$$Y(t, \theta) = \mathfrak{R}f + e(t, \theta), \quad (4.112)$$

We again study the three cases CC, CD, and DD as in the previous section.

CC case: Let us begin by assuming that a complete, continuous set of data $Y(t, \theta)$ is given in both variables t and θ . Under the assumption that the noise process $e(t, \theta)$ is Gaussian and white with intensity σ^2 , the ML estimate of f is given by

$$\hat{f}_{ml} = \arg \max_f P(Y|f), \quad (4.113)$$

where $P(Y|f)$ denotes the conditional probability density of Y given f . A function of the data $\mathcal{S} = \mathcal{S}(Y)$ is termed a sufficient statistic for f if and only if $\mathcal{S}(Y)$ contains the same information as Y [81]. More specifically, \mathcal{S} is sufficient for f if and only if

$$P(Y|f) = c(\mathcal{S}, Y)P(\mathcal{S}|f), \quad (4.114)$$

where c is a function that is independent of f . Note that this makes intuitive sense since from (4.114), the maximization of $P(Y|f)$ with respect to f is equivalent to the maximization of $P(\mathcal{S}|f)$ because $c(\mathcal{S}, Y)$ is independent of f . The notion of sufficient statistics is much more general than our exposition. However, for the sake of focus, we will not make our definitions any more technical than they need to be. Now to show that the *infinite set* of moments of $Y(t, \theta)$, MY , are sufficient for f , we use the results of Section 4.6.1. The probability density $P(Y|f)$ can be written as

$$P(Y|f) = k_1 \exp\left(\frac{-1}{2\sigma^2} \|Y - \Re f\|^2\right), \quad (4.115)$$

where k_1 is a positive normalizing constant, and the norm used in the exponent on the right hand side is that of $L^2([-1, 1] \times [0, 2\pi])$. Now recall the definition of the operator M acting on $L^2([-1, 1] \times [0, 2\pi])$ and note that due to the fact that M is a unitary operator, (4.115) may be written as

$$P(Y|f) = k_1 \exp\left(\frac{-1}{2\sigma^2} \|MY - M\Re f\|^2\right). \quad (4.116)$$

where the norm used on the right hand side is that of $L^2([0, 2\pi])$. But the right hand side of (4.116) is simply $P(MY|f)$ and hence we have

$$P(Y|f) = P(MY|f), \quad (4.117)$$

which shows that MY is sufficient for f .

CD case: When data is given continuously in t , but only for a finite number of angles θ_j , $j = 1, \dots, m$, we write the probability density of the vector $Y_m(t)$ of all projections as

$$P(Y_m|f) = k_2 \exp\left(\frac{-1}{2\sigma^2} \sum_{j=1}^m \|Y(t, \theta_j) - g_f(t, \theta_j)\|^2\right) \quad (4.118)$$

where $g_f(t, \theta_j)$ denotes the Radon transform projection of $f(x, y)$ in the direction θ_j and the norm used in the exponent on the right hand side of (4.118) is that of $L^2([-1, 1])$. Again, similar to the CC case, due to the fact that M (and therefore M_m) is a unitary operator, we can write (4.118) as

$$P(Y_m|f) = k_2 \exp\left(\frac{-1}{2\sigma^2} \sum_{j=1}^m \|MY(t, \theta_j) - Mg_f(t, \theta_j)\|^2\right) \quad (4.119)$$

where the right hand side is now simply $P(M_m Y_m|f)$, hence showing that

$$P(Y_m|f) = P(M_m Y_m|f) \quad (4.120)$$

DD case: In the case of discrete data in both t and θ , we are given data samples $\{Y(t_i, \theta_j) : 1 \leq i \leq n, 1 \leq j \leq m\}$. The moments of the projections are now defined in terms of sums that approximate the integral expressions for these moments given in the CC and CD cases. We denote the DD moments of the data Y by $\tilde{Y}^{(k)}(\theta_j)$ and define these for $j = 1, \dots, m$ as follows

$$\tilde{Y}^{(k)}(\theta_j) = \sum_{i=1}^n P_k(t_i) Y(t_i, \theta_j) \Delta t_i \quad (4.121)$$

where $t_i \in [-1, 1]$, $\Delta t_i = t_{i+1} - t_i$, and $t_{n+1} = 1$. For each fixed j , (4.121) denotes a linear transformation of the projection samples at the angle θ_j . Written another way we have

$$\mathcal{Y}_p(\theta_j) = U_p \underline{Y}(\theta_j), \quad (4.122)$$

where

$$\mathcal{Y}_p(\theta_j) = [\tilde{Y}^{(0)}(\theta_j), \tilde{Y}^{(1)}(\theta_j), \dots, \tilde{Y}^{(p)}(\theta_j)]^T \quad (4.123)$$

$$U_p = \begin{bmatrix} P_0(t_1)\Delta t_1 & P_0(t_2)\Delta t_2 & \cdots & P_0(t_n)\Delta t_n \\ P_1(t_1)\Delta t_1 & P_1(t_2)\Delta t_2 & \cdots & P_1(t_n)\Delta t_n \\ \vdots & \vdots & & \vdots \\ P_p(t_1)\Delta t_1 & P_p(t_2)\Delta t_2 & \cdots & P_p(t_n)\Delta t_n \end{bmatrix} \quad (4.124)$$

$$\underline{Y}(\theta_j) = [Y(t_1, \theta_j), Y(t_2, \theta_j), \dots, Y(t_n, \theta_j)]^T. \quad (4.125)$$

The samples $\underline{Y}(\theta_j)$ are from an n -dimensional Gaussian density, hence we know that the minimal sufficient statistic must be a vector of length n . Hence, for each j , we must compute (at least) the numbers $\tilde{Y}^{(k)}(\theta_j)$ for $k = 0, \dots, n-1$. i.e. we must at least have $p = n-1$. To show that the set of numbers $\{\tilde{Y}^{(k)}(\theta_j), 0 \leq k \leq n-1, 1 \leq j \leq m\}$, or equivalently, the set of vectors $\{\mathcal{Y}_{n-1}(\theta_j), 1 \leq j \leq m\}$ form a set of sufficient statistics, it suffices to show that the linear transformation given by U_{n-1} is nonsingular. We first note that U_{n-1} can be decomposed as follows.

$$U_{n-1} = B_{n-1} V(t_1, t_2, \dots, t_n) \mathbf{\Delta}, \quad (4.126)$$

where

$$B_{n-1} = \begin{bmatrix} \beta_{00} & 0 & 0 & \cdots & 0 \\ \beta_{01} & \beta_{11} & 0 & \cdots & 0 \\ \beta_{02} & \beta_{12} & \beta_{22} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \beta_{0,n-1} & \beta_{1,n-1} & \beta_{2,n-1} & \cdots & \beta_{n-1,n-1} \end{bmatrix} \quad (4.127)$$

$$V(t_1, t_2, \dots, t_n) = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ t_1 & t_2 & t_3 & \dots & t_n \\ t_1^2 & t_2^2 & t_3^2 & \dots & t_n^2 \\ \vdots & \vdots & \vdots & & \vdots \\ t_1^{n-1} & t_2^{n-1} & t_3^{n-1} & \dots & t_n^{n-1} \end{bmatrix} \quad (4.128)$$

$$\mathbf{\Delta} = \text{Diag}(\Delta t_1, \Delta t_2, \Delta t_3, \dots, \Delta t_n) \quad (4.129)$$

where $\beta_{i,j}$ are the coefficients of normalized Legendre polynomials of order $i+j$ and are defined in (4.3) on page 91. The matrix B_{n-1} is non-singular since all of its diagonal elements are nonzero [94]. The matrix $\mathbf{\Delta}$ is also non-singular since by assumption, $\Delta t_i \neq 0$ for $1 \leq i \leq n$. We recognize the matrix $V(t_1, t_2, \dots, t_n)$ as a *Vandermonde* matrix whose determinant is given by

$$\det(V(t_1, t_2, \dots, t_n)) = \prod_{l>q} (t_l - t_q) \quad (4.130)$$

for $1 \leq l, q \leq n$. Hence $V(t_1, t_2, \dots, t_n)$ is also non-singular since by assumption, $t_1 < t_2 < \dots < t_n$. Therefore, U_{n-1} is nonsingular, as desired. Hence, the set of vectors $\{\mathcal{Y}_{n-1}(\theta_j), 1 \leq j \leq m\}$ is a sufficient statistic for f .

4.6.3 Better Sampling Strategies

The study of moments can be used as a means of justifying certain commonly used sampling strategies in the variables t and θ . For instance, a typical tomogram is sampled at equally (maximally) spaced angles in $[0, \pi)$, with each projection containing samples at regularly spaced intervals in t that are symmetrically placed with respect to the center of the detector plane. (This is clearly an approximation to what happens in reality with standard detector geometries that are commonly in use.) In terms of the moments we show that these strategies are, in fact, the best possible in the sense that they provide the best-conditioned linear formulations for estimating

the moments of a function (image) f from moments, in the variable t , of its noisy and sparse integral projections.

Sampling in the Variable t :

In Section 4.6.2 we showed that when a finite number of projections with a finite number of samples per projection are measured, the set of vectors $\{\mathcal{Y}_{n-1}(\theta_j), 1 \leq j \leq m\}$ defined by (4.122) is a sufficient statistic for the underlying function $f(x, y)$. The numerical stability of the computation of these sufficient statistics is directly dependent on the condition number of U_{n-1} through (4.122). Hence, by studying the condition number of U_{n-1} in terms of the partition $t_1 < t_2 < \dots < t_n$, we may gain some insight as to what the best sampling strategies in t may look like. (The more ill-conditioned U_{n-1} is, the more numerical distortion is introduced when the sufficient statistics are computed). For convenience, let us denote $V(t_1, t_2, \dots, t_n)$ simply as V_n . Note that if we assume that the samples in t are uniformly spaced, then the condition number of U_{n-1} depends only on the condition number of V_n . This is because B_{n-1} is independent of the samples in t , and under the assumption of uniformly spaced samples in t , Δ is a nonzero scalar multiple of the identity matrix and is hence perfectly conditioned. Under the assumption of uniform sampling in t , consider the L_∞ condition number of V_n which is defined as follows.

$$\kappa(V_n) = \|V_n\|_\infty \|V_n^{-1}\|_\infty \quad (4.131)$$

From [29] we quote the following result.

Theorem 9 *If the minimum of the condition number $\kappa(V_n)$ is unique as a function of the partition $t_1 < t_2 < \dots < t_n$, then this minimum is necessarily achieved when the partition is symmetric with respect to the origin. i.e.*

$$t_i + t_{n+i-1} = 0 \quad \text{for } i = 1, 2, \dots, n \quad (4.132)$$

Although it is not known whether or not the minimum of $\kappa(V_n)$ is unique, this theorem suggests that when the object is assumed to lie in the unit disk of the plane, and uniformly spaced samples are taken in the variable t , the “optimal” sampling strategy should be one where the samples are taken symmetrically about zero on the t -axis.

Sampling in the Variable θ

To study the effect of different sampling strategies in the variable θ on the numerical stability of the moment estimation problem from projections, we consider the CD case where projections from a finite number of angles are given while each projection contains a complete, continuous data set in the variable $t \in [-1, 1]$. In particular, recall the expression relating the estimates of the orthogonal moments up to order N of a function f to moments of its projections given in (4.67) on page 111 :

$$\hat{\mathcal{L}}_N = T_N^{-1} \left(\mathbf{D}_N^T \mathbf{B}_N^T \mathbf{B}_N \mathbf{D}_N \right)^{-1} \mathbf{D}_N^T \mathbf{B}_N^T \mathbf{Y}_N. \quad (4.133)$$

Note that in this expression, the condition number of $\mathbf{D}_N^T \mathbf{B}_N^T \mathbf{B}_N \mathbf{D}_N$ determines the numerical sensitivity of the estimated moment vector $\hat{\mathcal{L}}_N$ to sampling strategies in the variable θ . Figure 4-3 contains the plots of the (2-norm) condition number of $\mathbf{D}_N^T \mathbf{B}_N^T \mathbf{B}_N \mathbf{D}_N$ (in log scale) versus the number of available views. This plot shows that this condition number does not dramatically improve as the number of views are increased, while it does degrade if the number of computed moments are increased, i.e. when the size of the moment estimation problem is increased. This increase is linear, however. (This can be seen by plotting the corresponding points of each graph versus N). This fact suggests that in computing the moments of a function from those of its Radon transform, the numerical stability of the problem depends on the number of the moments being computed and is essentially insensitive to increasing the number of views. Figure 4-4 shows the condition number (in log scale) of $\mathbf{D}_N^T \mathbf{B}_N^T \mathbf{B}_N \mathbf{D}_N$ versus decreasing angular coverage for various N and with a fixed number of views of 15. In particular, the angular coverage is parameterized with an integer c , which we call

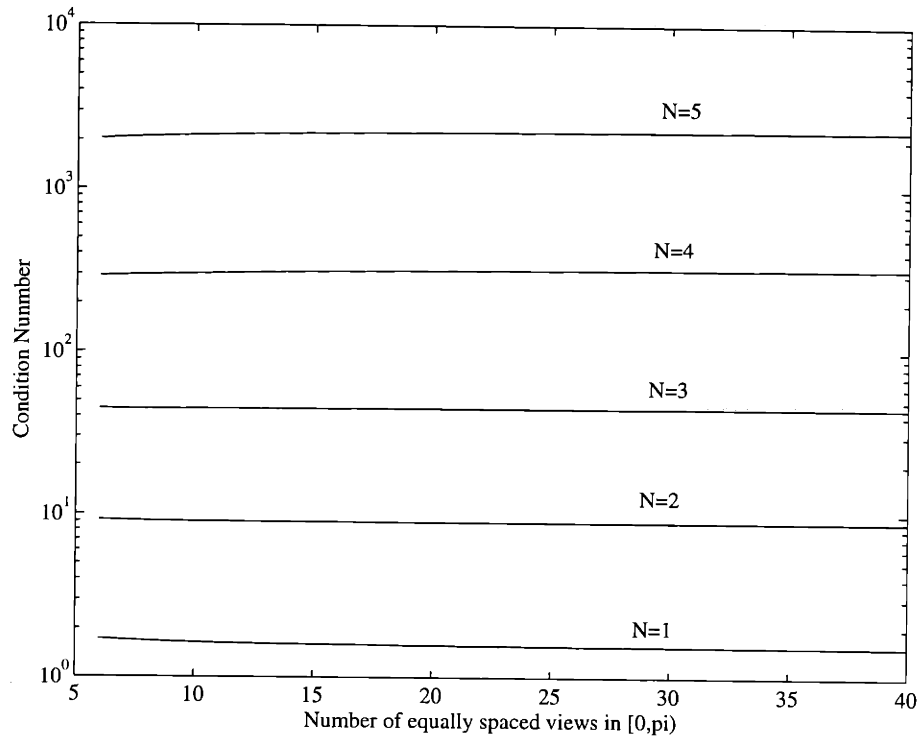


Figure 4-3: Condition Number of $\mathbf{D}_N^T \mathbf{B}_N^T \mathbf{B}_N \mathbf{D}_N$ vs the number of views for various N

the “compression parameter”, represented on the horizontal axis of this figure. The angular coverage is then taken to be from $\theta = 0$ to $\theta = \frac{\pi}{c}$ and in this interval 15 views are taken at equally spaced intervals of length $\Delta\theta = \frac{\pi}{15c}$. It can be seen that the condition number degrades rather severely as the angular coverage is reduced from $[0, \pi]$ to $[0, \frac{\pi}{10}]$. The condition number is also enlarged with increasing N , the number of moments to be computed. These observations are consistent with the practice of choosing a wide angular coverage whenever possible.

4.7 Some Examples and Experiments

In this section we present some examples and experiments to illustrate practical uses of the ideas discussed in this chapter.

4.7.1 Consistent Estimation with Constraints

In Section 4.6.1 we showed how the problem of producing a consistent Radon transform \hat{g} from a possibly inconsistent data set Y may be solved with the aid of moments.

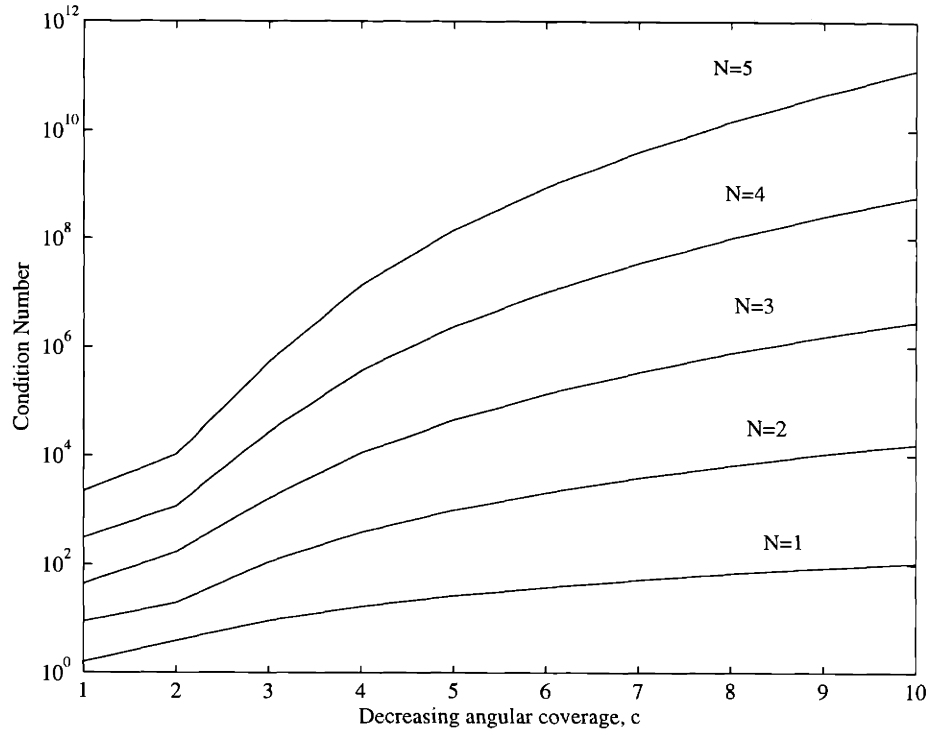


Figure 4-4: Condition Number of $\mathbf{D}_N^T \mathbf{B}_N^T \mathbf{B}_N \mathbf{D}_N$ vs decreasing angular coverage for various N

In this section, we give an illustration of this idea when additional constraints, other than the standard Radon transform constraints, are placed on the function \hat{g} . These constraints may involve explicit conditions placed on the function $\hat{f} = \mathfrak{R}\hat{g}$, such as the one we discuss next. Let us assume that we wish to solve the following specific problem.

Problem 3

$$\min_g \|Y - g\|^2 \tag{4.134}$$

subject to:

1. $g \in \mathcal{C}$

2. $f = \mathfrak{R}^{-1}g$ satisfies

$$\iint f(x, y) dx dy \geq 0 \tag{4.135}$$

Note that the second constraint can simply be written as

$$G^{(0)}(\theta) \geq 0 \tag{4.136}$$

which, in turn, defines a convex constraint set \mathcal{C}_1 over the functions g , or equivalently a convex constraint set Λ_1 in the moment space of g . Hence, the solution to Problem 3 is given by

$$\hat{g} = M^* \Pi_{\Lambda_C \cap \Lambda_1} M Y = \Pi_{\mathcal{C} \cap \mathcal{C}_1} Y \quad (4.137)$$

where $\Pi_{\mathcal{C} \cap \mathcal{C}_1}$ and $\Pi_{\Lambda_C \cap \Lambda_1}$ refer to orthogonal projections onto the (convex) sets $\mathcal{C} \cap \mathcal{C}_1$ and $\Lambda_C \cap \Lambda_1$, respectively. The projection operator $\Pi_{\Lambda_C \cap \Lambda_1}$ may be approximated by the method of successive projections [85], otherwise known as Projection onto Convex Sets (POCS). In this technique, a recursion relation is defined as follows:

$$P g_k = g_{k+1} \quad (4.138)$$

where P is an arbitrary permutation of the projection operators Π_C and $\Pi_{\mathcal{C}_1}$, such as $P = \Pi_C \Pi_{\mathcal{C}_1}$. Picking $g_0 = Y$, this iteration can be shown to converge to a consistent solution as k becomes large. This recursive algorithm can easily be recast in terms of the moments by observing that

$$\Pi_C = M^* \Pi_{\Lambda_C} M, \quad (4.139)$$

$$\Pi_{\mathcal{C}_1} = M^* \Pi_{\Lambda_1} M. \quad (4.140)$$

In general, the POCS algorithm given by (4.138) would work whenever the second constraint in the statement of Problem 3 is any convex-type constraint.

4.7.2 An Approximation of the Radon Transform

In Section 4.3.1, we showed how the Radon transform operator \mathfrak{R} may be decomposed into a product of moment operators Ω , M^* , and the operator A relating the moments of a function f to the moments of its Radon transform g as follows

$$\mathfrak{R} = M^* A \Omega. \quad (4.141)$$

This expression suggests that the Radon transform of a function f may be obtained (exactly) by first computing the (infinite set of) moments of f , then applying the operator A , and finally, applying M^* . In this example we wish to give an illustration of this idea by computing a projection of a binary object at a particular angle using only a finite number of moments. In particular, assume that we are given the function $f(x, y)$ which is the indicator function for the rectangle $\mathcal{O} = [-0.44, 0.23] \times [-0.44, 0.23]$. i.e.

$$f(x, y) = \begin{cases} 1, & (x, y) \in \mathcal{O} \\ 0, & (x, y) \notin \mathcal{O} \end{cases} \quad (4.142)$$

From (4.27) we have that for moments of a function f up to order N , for every $N \geq 0$, the moments of f and those of its projection at angle θ are related by

$$\mathcal{G}_N(\theta) = B_N \mathcal{D}_N(\theta) T_N \mathcal{L}_N. \quad (4.143)$$

where $\mathcal{G}_N(\theta)$ denotes the vector of orthogonal moments of $g(t, \theta)$ up to order N , \mathcal{L}_N denotes the vector of orthogonal moments of f up to order N , and $A_N = B_N \mathcal{D}_N(\theta) T_N$ denotes the linear operator which relates these moments through (4.27). We have numerically computed the vector \mathcal{L}_N for $N = 5, 10, 15$, and 20 . From these, we have computed $\mathcal{G}_N(\theta)$ for the same values of N and for $\theta = 0$. For each θ , the approximation of $g(t, \theta)$ over the interval $[-1, 1]$ is then given by

$$g_N(t, \theta) = \sum_{k=0}^N G^{(k)}(\theta) P_k(t), \quad (4.144)$$

where as N goes to infinity, g_N converges to g according to (4.141). To illustrate this behavior, we have plotted the function $g_N(t, \theta)$ for $\theta = 0$ and for $N = 5, 10, 15$, and 20 in Figure 4-5.

4.7.3 ML Moment Estimation Example and Performance

In this subsection we present an example of the reconstruction of the moments of a function $f(x, y)$ from noisy measured projection data. In particular, let $f(x, y)$ be

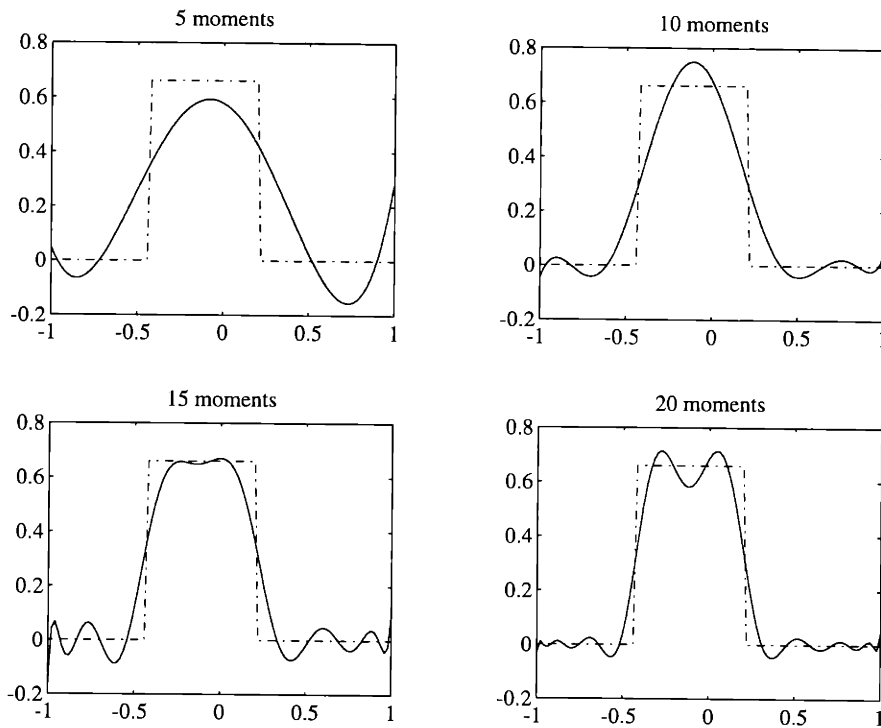


Figure 4-5: The functions $g_N(t, 0)$ for $N = 5, 10, 15,$ and 20

given by the image shown in Figure 4-6. This image is an 81×81 two-valued image with dark areas having the value of zero while the white areas have activity level of 10. So $f(x, y)$ is (10 times) the indicator function over the white areas of the image. This “MIT ellipse” is tilted at an angle of 45 degrees with the horizontal axis and has major and minor axis lengths of 0.8 and 0.24 respectively. Hence the object is contained within the unit disk centered at the origin. We consider 60 projections of this image at equally spaced angles $\theta_j \in [0, \pi)$ with 80 samples per projection at a signal-to-noise ratio of 10. The noiseless projection data along with the noisy projection data are shown in Figures 5-26 and 5-27 of Chapter 5 on pages 194 and 195, respectively.

We first compute the orthogonal moments up to order 10 of this image by directly approximating the integrals representing them over the 81×81 pixel grid on which the image is defined. Next, we use the noisy projection measurements to estimate the orthogonal moments of the image by first computing the orthogonal moments of the projections and then applying the ML estimation technique discussed in Section¹¹ 4.5.

¹¹Note that throughout Section 4.5 we assumed continuous data in the variable t . In the present example, the number of samples in t for each projection is sufficiently large so that we may assume this to be the case.

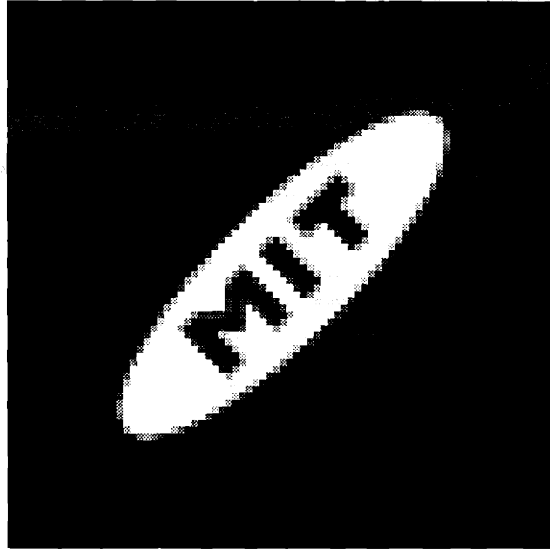


Figure 4-6: The MIT ellipse phantom

Figure 4-7 shows the elements of the vector of directly computed orthogonal moments of the image and the elements of the vector of ML estimated orthogonal moments of the image versus their index. (The elements of these orthogonal moment vectors are arranged as defined in 4.11 on page 93.) Figure 4-8 shows the difference between the respective elements of the computed and estimated values of the orthogonal moments.

To study the performance of the ML moment estimation algorithm, in Figure 4-9 we plot the trace of the covariance matrices of the estimated orthogonal moment vectors¹² $\hat{\lambda}^{(k)}$ of a given order k , defined in (4.10) on page 93, versus k and for different SNR values. As expected, the value of the trace of the covariance matrix of $\hat{\lambda}^{(k)}$ is larger for larger k . Also, as expected, the corresponding curves for worse SNR's display larger values for all moment orders.

¹²Note that for a given k , the covariance matrix of $\hat{\lambda}^{(k)}$ is simply the $(k+1)^{th}$, $(k+1) \times (k+1)$, diagonal block of the covariance matrix Q_{10} of $\hat{\mathcal{L}}_{10}$

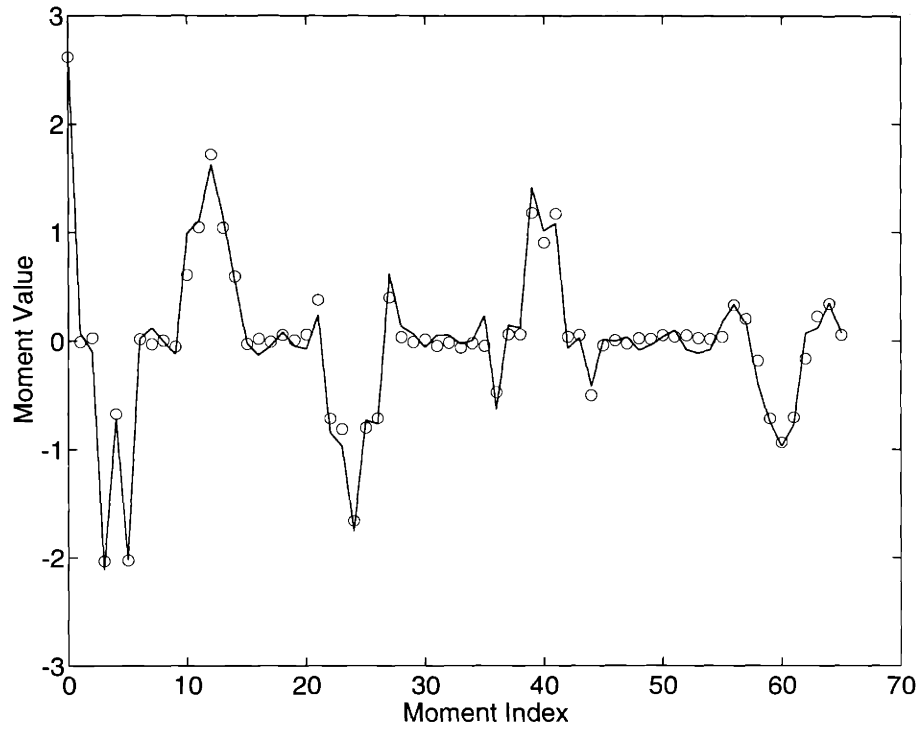


Figure 4-7: Computed (-) and ML Estimated (o) moments up to order 10 of the MIT Ellipse image versus their index

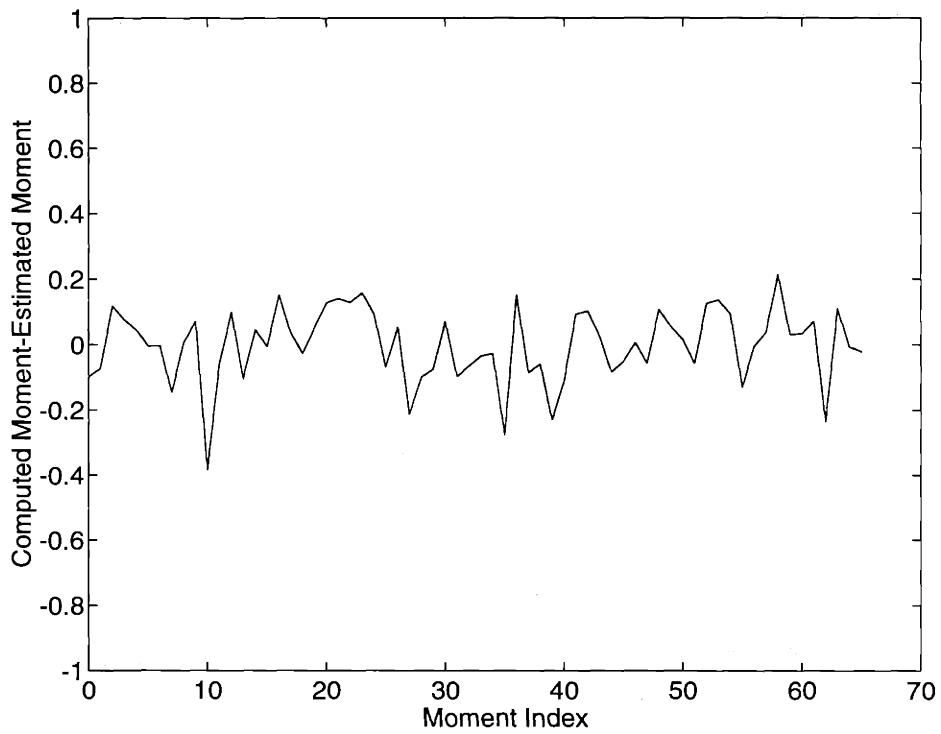


Figure 4-8: The difference values between computed and ML Estimated moments up to order 10 of the MIT Ellipse image versus their index

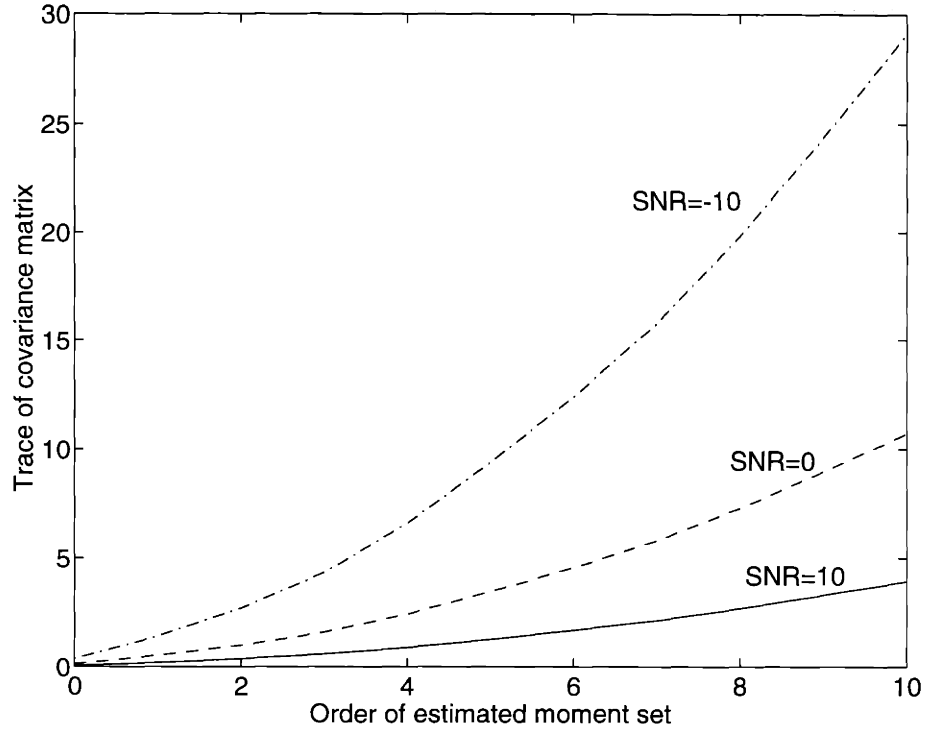


Figure 4-9: Trace of covariance matrix versus moment order up to order 10

4-A Proof of Lemma 2

Let $p(\theta)$ denote the homogeneous polynomial in question. i.e.

$$p(\theta) = \alpha_0 \cos^k(\theta) + \alpha_1 \cos^{k-1}(\theta) \sin(\theta) + \cdots + \alpha_{k-1} \cos(\theta) \sin^{k-1}(\theta) + \alpha_k \sin^k(\theta). \quad (4.145)$$

- **CASE I:** Assume that $p(\pi/2) \neq 0$. Then we can write $p(\theta)$ as

$$p(\theta) = \cos^k(\theta)q(\theta). \quad (4.146)$$

where $q(\theta)$ has no roots at $\theta = \pi/2$ and

$$q(\theta) = \alpha_0 + \alpha_1 \tan(\theta) + \cdots + \alpha_{k-1} \tan^{k-1}(\theta) + \alpha_k \tan^k(\theta). \quad (4.147)$$

Letting $u = \tan(\theta)$ we observe that the right hand side of (4.147) is simply a polynomial of order k in u . By the Fundamental Theorem of Algebra [61], this polynomial has at most k real roots. This is to say that there exist at most k values $u_i \in \mathcal{R}$ such that $q(\tan^{-1}(u_i)) = 0$. Given this, we have that the roots of

$q(\theta)$ are

$$\theta_i = \tan^{-1}(u_i). \quad (4.148)$$

We know that the function \tan^{-1} is one-to-one over the interval $[0, \pi)$. Since $q(\pi/2) \neq 0$ by assumption, it follows that there exist at most k angles $\theta_i \in [0, \pi)$ for which $q(\theta_i) = 0$.

- **CASE II:** Assume $p(\pi/2) = 0$. Let $r_0(\theta) = p(\theta)$, and define the functions $r_i(\theta)$ for $1 \leq i \leq k$ as follows.

$$r_i(\theta) = \frac{r_{i-1}(\theta)}{\cos(\theta)}, \quad \text{if } r_{i-1}(\pi/2) = 0 \quad (4.149)$$

Since $p(\pi/2) = 0$, we have

$$p(\pi/2) = \alpha_k \sin^k(\pi/2) = \alpha_k = 0. \quad (4.150)$$

Therefore we have that

$$p(\theta) = \cos(\theta) \left(\alpha_0 \cos^{k-1}(\theta) + \cdots + \alpha_{k-1} \sin^{k-1}(\theta) \right) = \cos(\theta) r_1(\theta) \quad (4.151)$$

If $r_1(\pi/2) \neq 0$, Case I shows that it has at most $k - 1$ roots in $[0, \pi)$, which together with $\cos(\theta) = 0$ give at most k roots for $p(\theta)$ in $[0, \pi)$ and we are done.

Now from the definition of $r_i(\theta)$, it is clear that if $r_{i-1}(\pi/2) = 0$, then:

$$r_i(\theta) = \alpha_0 \cos^{k-i}(\theta) + \cdots + \alpha_{k-i} \sin^{k-i}(\theta). \quad (4.152)$$

Now suppose that $r_i(\pi/2) = 0$ for $i = 0, 1, \dots, n - 1$ and $r_n(\pi/2) \neq 0$, where $1 \leq n \leq k$. Again, from the definition of $r_i(\theta)$ it follows that

$$p(\theta) = \cos^n(\theta) r_n(\theta). \quad (4.153)$$

From Case I, $r_n(\theta)$ has at most $k - n$ roots in $[0, \pi)$, which along with $\cos^n(\theta) = 0$ give at most k roots for $p(\theta)$ in $[0, \pi)$. \square

4-B Proof of Result 6

Let $P(x, y)$ and $Q(x, y)$ be two real-values functions defined over a simple region \mathcal{O} of the plane. Then by Green's Theorem [63], we have

$$\iint_{\mathcal{O}} \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dx dy = \oint_{\partial \mathcal{O}} P dx + Q dy, \quad (4.154)$$

where $\partial \mathcal{O}$ refers to the boundary of \mathcal{O} . Now define

$$Q_1(x, y) = \frac{x^{p+1}y^q}{p+1} \quad (4.155)$$

$$P_1(x, y) = \frac{x^p y^{q+1}}{q+1} \quad (4.156)$$

Letting $P = P_1$ and $Q = Q_1$ yields

$$\frac{1}{p+1} \oint_{\partial \mathcal{O}} x^{p+1} y^q dy = \frac{-1}{q+1} \oint_{\partial \mathcal{O}} x^p y^{q+1} dx \quad (4.157)$$

Now first let $P = P_1$ and $Q = 0$, and then let $P = 0$ and $Q = Q_1$, respectively in (4.154). Together, from these substitutions, we obtain

$$\mu_{pq} = \iint_{\mathcal{O}} x^p y^q dx dy = \frac{1}{p+1} \oint_{\partial \mathcal{O}} x^{p+1} y^q dy \quad (4.158)$$

$$= \frac{-1}{q+1} \oint_{\partial \mathcal{O}} x^p y^{q+1} dx \quad (4.159)$$

Since \mathcal{O} is an N -sided, simple polygonal region, we can split each of the integrals on the right hand sides of (4.158), and (4.159) into integrals over each of the sides of $\partial \mathcal{O}$. Then applying the Mean Value Theorem for integrals [62] to each of these pieces yields the desired result.

4-C Proof of Result 7

Let the eigenvalues of A be given by $\lambda_1, \lambda_2, \dots, \lambda_n$. Then we have that

$$\text{Tr}(A) = \sum_{i=1}^n \lambda_i \quad (4.160)$$

$$\text{Tr}(A^{-1}) = \sum_{i=1}^n \frac{1}{\lambda_i} \quad (4.161)$$

Let us denote by a_n and h_n the arithmetic and the harmonic means, respectively, of the eigenvalues of A . Then we have

$$a_n = \text{Tr}(A)/n, \quad (4.162)$$

$$h_n = (\text{Tr}(A^{-1})/n)^{-1}. \quad (4.163)$$

We also know that the harmonic and arithmetic means of a set of positive real numbers satisfy [38]

$$h_n \leq a_n \quad (4.164)$$

This last inequality proves the result. To see why the inequality can not be improved, let A be the identity matrix. □

4-D Proof of Lemma 3

To show convexity of the constraint set \mathcal{C} , consider two arbitrary $g_1(t, \theta) \in \mathcal{C}$ and $g_2(t, \theta) \in \mathcal{C}$, and define $g(t, \theta) = \mu g_1(t, \theta) + (1 - \mu)g_2(t, \theta)$ for $0 \leq \mu \leq 1$. Clearly, $g \in L^2([-1, 1] \times [0, 2\pi])$. We also have,

$$g(t, \theta + \pi) = \mu g_1(t, \theta + \pi) + (1 - \mu)g_2(t, \theta + \pi), \quad (4.165)$$

$$= \mu g_1(-t, \theta) + (1 - \mu)g_2(-t, \theta), \quad (4.166)$$

$$= g(-t, \theta). \quad (4.167)$$

The moments of g satisfy

$$\int g(t, \theta) t^k dt = \mu \int g_1(t, \theta) t^k dt + (1 - \mu) \int g_2(t, \theta) t^k dt, \quad (4.168)$$

The right hand side of (4.168) is a linear combination of two homogeneous polynomials of degree k in $\cos(\theta)$ and $\sin(\theta)$. Hence, the left hand side is a homogeneous polynomial of degree k . This shows convexity.

To show closure of \mathcal{C} , let $\{g_n\}$ be an arbitrary convergent sequence in \mathcal{C} . i.e. $g_n \in \mathcal{C}$ for all n . Suppose that $g_n \rightarrow g$, that is

$$\|g_n(t, \theta) - g(t, \theta)\|^2 = \iint (g_n - g)^2 dt d\theta \rightarrow 0 \text{ as } n \rightarrow \infty \quad (4.169)$$

Now consider the quantity $E \equiv \left| \iint (g_n - g) P_k(t) e^{in\theta} dt d\theta \right|$, where $P_k(t)$ denotes the k^{th} order normalized Legendre polynomial over $[-1, 1]$. The Cauchy-Schwartz inequality yields

$$E \leq \left(\iint P_k^2(t) e^{2in\theta} dt d\theta \right)^{1/2} \left(\iint (g_n - g)^2 dt d\theta \right)^{1/2}. \quad (4.170)$$

The second term on the right hand side of (4.170) vanishes as $n \rightarrow \infty$, hence $E \rightarrow 0$ as $n \rightarrow \infty$. Therefore, we have that the Fourier series expansion of the orthogonal moments $G^{(k)}(\theta)$ of g has the required form given by (4.89) in page 119. Hence, $g \in \mathcal{C}$.

□

Chapter 5

Variational Formulations for Reconstruction from Moments

5.1 Introduction

In this chapter, we discuss the reconstruction via regularization of a function $f(x, y)$ from the Maximum Likelihood estimates of its orthogonal moments. The problem of reconstructing a function from a finite number of moments is known to be highly ill-posed [95], as is the problem of inverting the Radon transform. In our framework, we have transformed the problem of inverting the Radon transform into the problem of reconstructing a function from its moments (classically known as the moment problem). The fundamental result that the algorithms in this chapter rely on is, as we showed in Chapter 4, that the ML (or MAP) estimate of an image from samples of its Radon transform corrupted by Gaussian white noise can be obtained in two distinct steps. The first step being the ML (or MAP) estimation of the moments of the underlying image from the noisy data, while the second step involves the reconstruction of the image from its estimated moments. The first step is a simple *linear* estimation problem while the second is a highly ill-posed inverse problem. In this chapter we present efficient algorithms for recovering the image from its estimated moments.

We show that there are several advantages to our two-step approach. One is

that the use of moments provides an explicit mechanism for controlling the degrees of freedom in the reconstructions. Another is the computational savings inherent to our approach. Finally, these features yield an overall efficient and versatile set of algorithms that yield reconstructions of excellent quality when compared to other available algorithms.

In our formulation of the reconstruction problem, we do not have exact knowledge of the moments due to noise; nor can we estimate all the moments of a function given finite resolution data. The algorithms we propose take these two issues into account. The reconstruction of images from their moments is not a central issue in image processing and pattern recognition theory since the use of moments there is essentially restricted to applications where extraction of distinguishing features is the end goal. As a result, little work has been done on the reconstruction of images from their moments in the image processing community [70]. Also, there has been a little work in this area within the tomography community [84]. On the other hand, the moment problem has been the subject of much work in the mathematics community for many years. For comprehensive surveys see [2, 86]. However, almost all of this work has been concentrated around questions of existence and uniqueness of solutions. As a consequence, there exists a wealth of theoretical results on the moment problem, while very little work exists on explicit, and numerically sound, algorithms for a solution. Here we propose two different techniques. The first is L^1 -divergence regularization and the second is Quadratic regularization. Variants of both of our techniques have been proposed elsewhere in the literature [95, 99, 45, 67, 72, 3] but not in the particular context that we present. We also propose novel numerical techniques for solving these variational problems and study some of their properties and extensions. We demonstrate, in this Chapter, that our proposed overall tomographic reconstruction techniques enjoy a number of advantages over classical techniques. One such advantage over the classical reconstruction techniques, such as CBP, which rely on the direct inversion of the Radon transform, is that in our framework, there is an inherent mechanism to control the degrees of freedom of the reconstructions. In specific, this is given by the number of moments used to obtain the reconstruction.

This number may, in turn, be estimated using a statistically sound technique such as the Minimum Description Length criterion. Another advantage is that geometric information can be incorporated easily and directly into our algorithms.

In this chapter we also discuss an iterative extension of these algorithms that can efficiently yield a solution to the equality constrained variants of these regularization problems. We discuss the relative merits and drawbacks of each and support these assertions with numerical examples which appear at the end of the chapter.

5.2 Regularization

In this section we present regularization techniques for the ill-posed problem of reconstructing an image from its estimated orthogonal moments. The motivation behind using the I-divergence regularization (IDR) technique is based on several ideas. The solution of the IDR problem has a particularly compact and attractive form. From this form it follows that, by picking a positive prior estimate, we are guaranteed that the solution will be a positive function. This is clearly desirable given the objective of reconstructing an image (a positive function). The particular form in which the prior enters the solution allows for the explicit incorporation of prior geometric information into the reconstruction process. Finally, by picking a uniform prior, we can arrive at a *Maximum Entropy* solution which has an interesting statistical interpretation [44].

The Quadratic Regularization (QR) technique enjoys the advantage that it yields a closed form solution which translates to obvious computational savings. It is also the preferred classical regularization technique for solving the moment problem [95]. Hence, its solution can be used as a benchmark against the performance of the IDR algorithm. The QR solution is not guaranteed to be positive so in at least this sense, it is not as well suited for image reconstruction as the IDR solution.

5.2.1 I-Divergence Regularization

The principle of Minimum I-divergence [17, 89, 91] provides a general method of inference about a function f when there exists a prior estimate of f and new information

about f in the form of moments. The principle states that, of all the functions that satisfy the moment constraints, one should pick the one \hat{f} with the least *I-divergence* $D(f, f_0)$, where this is defined as

$$D(f, f_0) = \iint \left(f(\mathbf{x}) \log\left(\frac{f(\mathbf{x})}{f_0(\mathbf{x})}\right) + f_0(\mathbf{x}) - f(\mathbf{x}) \right) d\mathbf{x} \quad (5.1)$$

where f_0 is a prior estimate of f .

The basic idea dates back to Kullback [53] who introduced *Cross Entropy* (or directed divergence) minimization for measuring the relative information content between probability density functions. I-divergence was later introduced by Csizar [17] to generalize Cross Entropy to other functions. The principle of *Maximum Entropy* [44] is equivalent to the concept of Cross Entropy in the case of uniform priors f_0 . Entropy and more recently I-divergence have a rich history of applications in pattern classification [88], spectral analysis [87], image processing [102, 35] and recently tomography [84, 74, 20, 64]. In these applications, the general problem has been posed as the following type of equality constrained optimization problem:

$$\min_f D(f, f_0) \quad \text{subject to} \quad \iint f(\mathbf{x}) \phi_{i,j}(\mathbf{x}) d\mathbf{x} = \hat{s}_{i,j} \quad (5.2)$$

In particular, in the context of tomography, the weight functions $\phi_{i,j}(\mathbf{x})$ have been chosen as appropriate delta functions so that the constraints $\hat{s}_{i,j}$ are the *noisy* measured values of the Radon transform of f [84, 74, 20, 64]. That is to say, the constraints have the form

$$\iint f(\mathbf{x}) \delta(t_i - \mathbf{x}^T \omega_j) dx dy = \hat{s}_{i,j} \quad (5.3)$$

where ω_j is the unit direction vector making an angle θ_j with the x -axis. (See the definition of the Radon transform in Chapter 2.) In fact, most of the tomography literature on the subject has been concerned with a very special case of the divergence measure $D(f, f_0)$. Namely, the case where f_0 is taken to be a uniform prior estimate of f . This corresponds to the *maximum entropy* reconstruction algorithms. Other vari-

ants of these algorithms allow for the equality constraints to be inequality constraints so that some notion of uncertainty in the measured values of $\widehat{s}_{i,j}$ can be taken into account [45]. Three important features distinguish our approach from other available algorithms mentioned above. The first is that we allow for arbitrary (but positive) prior estimates f_0 so that our approach allows for the incorporation of prior geometric information not only in terms of the moments, but also in terms of the particular prior estimate f_0 we may wish to use. The second is that we use the estimated Legendre moments instead of the actual measured values of the projections. This is to say that, in our case, the basis functions are $\phi_{i,j}(x,y) = P_i(x)P_j(y)$, where $P_i(\cdot)$ denotes the i^{th} order normalized Legendre polynomial over the interval $[-1,1]$. We do not use the estimated moments to form hard equality or inequality constraints but rather use these estimates, along with their computed covariance structure, to form a composite cost function that consists of the I-divergence term plus a quadratic form in terms of the estimated moments. Finally, and perhaps most importantly, we estimate and use the moments *and* directly incorporate their estimated covariances, thus ensuring that these data are used in a statistically optimal way. Since the covariance of the estimated moments grows without bound for higher order moments, we effectively only use the first few of the estimated moments. That is, by using moments, we are able to *focus* the information in the raw projection data, identifying a much smaller set of statistically significant quantities capturing essentially *all* information of use in reconstruction. This is in contrast to other divergence based algorithms which use measured values of the projections as in (5.3).

Each of the above three features leads to important advantages over other available techniques. By allowing general prior estimates f_0 , we are capable of reconstructing a much wider class of objects. Classical maximum entropy approaches tend to produce reconstructions that do not capture fine variations (details) particularly well. This is a direct consequence of fixing f_0 to be a uniform prior. We alleviate this problem by allowing prior estimates that reflect explicit prior knowledge about the underlying phantom such as its spatial support, locations of objects and regions of interest, and other localized effects such as edges. By using the Legendre polynomials in a com-

posite cost function as soft constraints, we have an exact statistical interpretation of the solution as the Maximum-A-Posteriori (MAP) estimate of f based on noisy measurements of its moments. We will show this shortly. Lastly, the computational savings in estimating and using a limited number of moments versus using all of the available projection measurements are significant. For instance consider a typical scenario where projections from 64 views with 64 samples per view have been collected. In this case, if all the samples are used as constraints in (5.3), this amounts to over 4000 constraints. In our framework, a typical range for the number of moments used that would yield a good reconstruction is 15 to 20. This amounts to using between $\frac{16 \times 17}{2} = 136$ and $\frac{21 \times 22}{2} = 231$ moment values in the cost function. This translates to a significantly reduced computational cost.

Formally, we define the I-divergence Regularization (IDR) cost functional as

$$J_{IDR}(f, f_0) = \gamma D(f, f_0) + \frac{1}{2}(\hat{\mathcal{L}}_N - \mathcal{L}_N(f))^T \Sigma_N (\hat{\mathcal{L}}_N - \mathcal{L}_N(f)), \quad (5.4)$$

where $\gamma \in (0, \infty)$ is the regularization parameter, and $\Sigma_N = Q_N^{-1}$ is the inverse of the error covariance matrix for the estimate $\hat{\mathcal{L}}_N$. Now consider the MAP estimate of f based on noisy measurement of its moments up to order N . Assuming that $P(f)$ is some prior probability density function on the space of functions f , the MAP cost to be minimized is given by

$$J_{map}(f) = -\log P(\hat{\mathcal{L}}_N | f) - \log P(f) \quad (5.5)$$

$$= \frac{1}{2}(\hat{\mathcal{L}}_N - \mathcal{L}_N(f))^T \Sigma_N (\hat{\mathcal{L}}_N - \mathcal{L}_N(f)) - \log cP(f) \quad (5.6)$$

where c is a normalizing constant depending only on N and Σ_N . Comparing (5.6) to $J_{IDR}(f)$, we conclude that if

$$P(f) = \frac{1}{c} \exp(-\gamma D(f, f_0)), \quad (5.7)$$

then $J_{IDR}(f, f_0) = J_{map}(f)$. The probability density function given by (5.7) is not a commonplace one. Mathematically, we can consider the function $D(f, f_0)$ as a

distance measure between the two *positive* functions f and f_0 . In fact, it is known to be the unique function that satisfies a set of axioms known as *directed orthogonality principle* [45]. These axioms are analogous to those used in defining linear projections in Hilbert spaces. The function $D(f, f_0)$ is in fact known as a *directed distance* [45]. From this point of view, the probability density function given by (5.7) is quite analogous to the standard Gaussian density, the difference being that in the Gaussian case, the exponent is basically the L^2 norm of the difference $f - f_0$.

5.2.2 Quadratic Regularization

Quadratic regularization (QR) [99, 95, 8] is based on the idea of using the standard L^2 distance as a regularization functional. This approach is a particular case of a more general class of regularization methods proposed by Tikhonov [99]. The quadratic regularization approach, in contrast to I-divergence minimization, is not guaranteed to produce positive functions as solutions, but enjoys the advantage that it is algorithmically much simpler and numerically efficient.

Prince [72] used a term penalizing non-smooth solutions to form a cost function in the projection domain. This cost function was then minimized subject to hard constraints given by the values of the first two moments of the projections. Our approach differs from that of Prince in several ways. First, the formulations we present in this chapter do not make use of a smoothness penalty term as did Prince. We will, however, allude to how such a regularization functional may be incorporated into our reconstruction algorithms in Section 7.2.

Our regularized problem is formulated in the object space, not the projection space. Perhaps more importantly, our approach does not use the moments as hard constraints hence allowing for the computed covariances of the estimated moments to directly enter the regularized cost functional. Furthermore, we typically use more than two estimated moments to compute the solution \hat{f} .

To begin let us define our QR cost function as follows.

$$J_{QR}(f, f_0) = \gamma Q(f, f_0) + (\hat{\mathcal{L}}_N - \mathcal{L}_N(f))^T \Sigma_N (\hat{\mathcal{L}}_N - \mathcal{L}_N(f)) \quad (5.8)$$

where

$$Q(f, f_0) = \int \int_{\mathcal{O}} (f(x, y) - f_0(x, y))^2 dx dy \quad (5.9)$$

This cost function has a direct statistical interpretation as the negative of the MAP log likelihood function when f is assumed to be a white Gaussian random field with a mean value of f_0 , where the relaxation parameter γ is inversely proportional to the variance of the random field. This approach has been suggested for the reconstruction of functions from a finite number of noisy moments before [95]. However, the structure of the noise in the moments has not been explicitly used in the reconstruction process. In our approach we have directly incorporated the covariance structure of the estimated moments in the second term of (5.8). We show that this formulation leads to a closed form solution which can be computed efficiently.

5.3 Solution of Regularized Problems

In this section we proceed to the explicit computation of the solutions to the proposed regularized problems. To make the presentation simpler, define the vectors $\phi_k(x, y)$ for $k = 0, 1, \dots, N$ as

$$\phi_k(x, y) = \begin{bmatrix} P_k(x)P_0(y) \\ P_{k-1}(x)P_1(y) \\ \vdots \\ P_0(x)P_k(y) \end{bmatrix}, \quad (5.10)$$

where $P_k(\cdot)$ is the k^{th} order normalized Legendre polynomial over the interval $[-1, 1]$. Also define $\Phi_N(x, y)$ as

$$\Phi_N(x, y) = \begin{bmatrix} \phi_0(x, y) \\ \phi_1(x, y) \\ \vdots \\ \phi_N(x, y) \end{bmatrix}. \quad (5.11)$$

5.3.1 I-Divergence Solution

The cost functional $J_{IDR}(f)$ can be written as

$$J_{IDR}(f, f_0) = \left(\int \int_{\mathcal{O}} f(x, y) \Phi_N(x, y) dx dy - \hat{\mathcal{L}}_N \right)^T \Sigma_N \left(\int \int_{\mathcal{O}} f(x, y) \Phi_N(x, y) dx dy - \hat{\mathcal{L}}_N \right) + \gamma D(f, f_0) \quad (5.12)$$

To find the minimum of this cost functional with respect to f we use the techniques of calculus of variations. We can find the minimum of this cost function by computing its variation $\delta J_{IDR}(f)$ with respect to f and setting this equal to zero. This variation is given by

$$\delta J_{IDR}(f, f_0) = \gamma \delta D(f, f_0) + \int \int_{\mathcal{O}} \Phi_N^T(x, y) \Sigma_N(\mathcal{L}_N(f) - \hat{\mathcal{L}}_N) \delta f dx dy = 0 \quad (5.13)$$

with

$$\delta D(f, f_0) = \int \int_{\mathcal{O}} \log\left(\frac{f}{f_0}\right) \delta f dx dy \quad (5.14)$$

In order for the variation $\delta J_{IDR}(f)$ to be equal to zero, it is necessary that the integrands in (5.13) vanish. (This is because δf is an *arbitrary* variation in f). Setting the integrands equal to zero yields

$$\gamma \log\left(\frac{f}{f_0}\right) + \Phi_N^T(x, y) \Sigma_N(\mathcal{L}_N(f) - \hat{\mathcal{L}}_N) = 0 \quad (5.15)$$

Rewriting this in terms of f we have

$$f(x, y) = f_0(x, y) \exp\left(\frac{-1}{\gamma} \Phi_N^T(x, y) \Sigma_N(\mathcal{L}_N(f) - \hat{\mathcal{L}}_N)\right) \quad (5.16)$$

The above is now an equation in f which may be solved. Note that the prior estimate f_0 enters the solution multiplicatively. This will allow us to directly introduce geometric information into the solution. We shall have more to say about this later.

The use of calculus of variations to minimize J_{IDR} and the particularly interesting form of the solution (5.16) are quite similar to those discussed in earlier works on

the subject of I-divergence minimization [45, 89, 44] with hard constraints. What is different about our approach is that the estimated moments are explicitly used to form the cost function and *not* as hard equality or inequality constraints.

Due to the form of the solution (5.16), we may convert (5.16) into a nonlinear algebraic equation in terms of the coefficient C_N defined as follows

$$C_N = \frac{-1}{\gamma} \Sigma_N(\mathcal{L}_N(f) - \hat{\mathcal{L}}_N). \quad (5.17)$$

Substituting the expression for $\mathcal{L}_N(f)$ using (5.16) we obtain an equation in terms of C_N as follows.

$$C_N = \frac{-1}{\gamma} \Sigma_N\left(\int \int_{\mathcal{O}} f_0(x, y) \exp(\Phi_N^T(x, y)C_N)\Phi_N(x, y)dx dy - \hat{\mathcal{L}}_N\right) \quad (5.18)$$

Defining

$$H(C_N) = \left(\int \int_{\mathcal{O}} f_0(x, y) \exp(\Phi_N^T(x, y)C_N)\Phi_N(x, y)dx dy - \hat{\mathcal{L}}_N\right) \quad (5.19)$$

we can write

$$C_N = \frac{-1}{\gamma} \Sigma_N H(C_N). \quad (5.20)$$

For the sake of simplicity, we propose to solve (5.20) by linear iteration [15] as follows

$$C_N(j+1) = \frac{-1}{\gamma} \Sigma_N H(C_N(j)). \quad (5.21)$$

If this iteration is stable, with a suitable initial guess, $C_N(j)$ will approach the unique solution ¹ of (5.20) as j grows large. Linear iteration (or fixed-point iteration) is one of many algorithms that may be applied to the solution of (5.20). One such algorithm is the well-known Newton's algorithm [15]. In this approach, the equation is essentially linearized about the best current guess at every step of an iteration that involves the gradient of the function $F(C_N) = C_N - H(C_N)$ as follows. We wish to solve the

¹Uniqueness follows from the fact that the minimization of the cost functional has a unique solution due to the convex nature of J_{IDR}

equation $F(C_N) = 0$. To this end, let $C_N(k)$ denote the current best estimate of the solution and write a local approximation of the function F about this value as follows

$$F(C_N) \approx F(C_N(k)) + G_F(C_N(k))(C_N - C_N(k)) = 0 \quad (5.22)$$

where

$$G_F(C_N(k)) = \left(\frac{\partial F}{\partial C_N} \right)_{C_N(k)} \quad (5.23)$$

denotes the gradient of the function F with respect to C_N , evaluated at the current best guess $C_N(k)$. Solving (5.22) for C_N , Newton's algorithm is then given by

$$C_N(k+1) = C_N(k) - G_F^{-1}(C_N(k))F(C_N(k)). \quad (5.24)$$

Stability of algorithm

To study the stability of our fixed-point iterative algorithm (5.21), we carry out a local stability analysis. To first order, we can approximate the function $H(C_N)$ by expanding it in a Taylor series around some initial value $C_N(0)$ which we assume is sufficiently close to the solution of (5.20). For simplicity, let us consider $C_N(0) = \underline{0}$ ². We then have

$$H(C_N) \approx H(\underline{0}) + \left(\frac{\partial H}{\partial C_N} \right)_{\underline{0}} C_N. \quad (5.25)$$

Note that

$$H(\underline{0}) = \mathcal{L}_N(f_0) - \hat{\mathcal{L}}_N \quad (5.26)$$

and let

$$D_0 = \left(\frac{\partial H}{\partial C_N} \right)_{\underline{0}} = \int \int_{\mathcal{O}} f_0(x, y) \Phi_N(x, y) \Phi_N^T(x, y) dx dy \quad (5.27)$$

Hence, locally, around $\underline{0}$, (5.21) can be written as

$$C_N(j+1) \approx \frac{-1}{\gamma} \Sigma_N(\mathcal{L}_N(f_0) - \hat{\mathcal{L}}_N + D_0 C_N(j)) \quad (5.28)$$

²Our numerical experiments show that $C_N(0) = \underline{0}$ is actually often sufficiently close to the solution of (5.20) to result in convergence to it

This is now a *linear* dynamic system whose stability depends on whether the matrix $(-1/\gamma)\Sigma_N D_0$ has stable eigenvalues. ³ More generally, we have the following result.

Result 10 *The iterative system (5.21) is strictly stable around the initial value $C_N(0) = \underline{0}$ if all eigenvalues of the matrix $(-1/\gamma)\Sigma_N D_0$ are contained strictly inside the unit disk; where D_0 is the Jacobian of H with respect to C_N evaluated at $C_N(0) = \underline{0}$.*

This result fixes an allowable range for the relaxation parameter γ so that the iteration (5.21) is guaranteed to be stable. Within this range, we can pick a value of γ that provides a desirable level of regularization in our solution. Note also that Result 10 is only a sufficient condition for the local stability of (5.21). The iteration may be stable even when the eigenvalues of $(-1/\gamma)\Sigma_N D_0$ are outside the unit disk. Furthermore, (5.21) is only one possible algorithm for solving (5.20). Other algorithms, such as Newton's algorithm discussed above, can be used which may have different stability properties.

5.3.2 Quadratic Solution

Again, to compute the minimum of $J_{QR}(f, f_0)$ we employ the calculus of variations. It is easy to show that the variation of $Q(f, f_0)$ is given by

$$\delta Q(f, f_0) = \int \int_{\mathcal{O}} 2(f(x, y) - f_0(x, y)) \delta f dx dy \quad (5.29)$$

which gives

$$\delta J_{QR}(f) = 2\gamma \int \int_{\mathcal{O}} (f(x, y) - f_0(x, y)) \delta f dx dy + 2 \int \int_{\mathcal{O}} \Phi_N^T(x, y) \Sigma_N (\mathcal{L}_N(f) - \hat{\mathcal{L}}_N) \delta f dx dy. \quad (5.30)$$

Setting the integrand equal to zero and solving for f yields

$$f(x, y) = f_0(x, y) - \frac{1}{\gamma} \Phi_N^T(x, y) \Sigma_N (\mathcal{L}_N(f) - \hat{\mathcal{L}}_N). \quad (5.31)$$

³Note that if $f_0(x, y)$ is picked to be a uniform prior (a constant function = c), the matrix D_0 is simply c times the identity matrix so that stability will only depend on the eigenvalues of $(-c/\gamma)\Sigma_N$

To solve this explicitly for f , we rewrite this equation in terms of C_N (which was defined in (5.17)) as follows

$$\begin{aligned} C_N &= \frac{-1}{\gamma} \Sigma_N \left(\iint_{\mathcal{O}} \Phi_N(x, y) (f_0(x, y) + \Phi_N^T(x, y) C_N) dx dy - \hat{\mathcal{L}}_N \right) \\ &= \frac{-1}{\gamma} \Sigma_N (C_N + \mathcal{L}_N(f_0) - \hat{\mathcal{L}}_N), \end{aligned} \quad (5.32)$$

Upon solving for C_N we obtain

$$\hat{C}_N = (\gamma I + \Sigma_N)^{-1} \Sigma_N (\hat{\mathcal{L}}_N - \mathcal{L}_N(f_0)), \quad (5.33)$$

which, in turn gives

$$\hat{f}_{QR}(x, y) = f_0(x, y) + \Phi_N^T(x, y) (\gamma I + \Sigma_N)^{-1} \Sigma_N (\hat{\mathcal{L}}_N - \mathcal{L}_N(f_0)). \quad (5.34)$$

Note that in contrast to the IDR algorithm, the equation (5.32) is linear in C_N and is hence solved explicitly, giving a closed-form solution for the minimizer of the cost functional $J_{QR}(f, f_0)$. This clearly translates to significant savings in terms of the computation time required to arrive at a solution when compared to the IDR algorithm.

5.3.3 Properties of the Solutions

In this section we explore some properties of the solutions obtained by using the proposed regularization schemes.

IDR Solution Properties

The I-divergence regularization problem and its solution have some interesting properties. Below we briefly mention a few of these.

- Form of the Solution and Positivity

In Section 5.3.1, we obtained the solution $\hat{f}_{IDR}(x, y)$ that minimizes the cost function $J_{IDR}(f, f_0)$. This unique solution (uniqueness due to convexity of J_{IDR})

has the general form

$$\hat{f}_{IDR}(x, y) = f_0(x, y) \exp(\Phi_N^T(x, y) \hat{C}_N) \quad (5.35)$$

where $f_0(x, y)$ is a prior estimate of f and \hat{C}_N is the solution of Equation (5.20). The compact form of this solution makes its numerical computation quite simple. In particular, as shown in Section 5.3.1, the overall solution of the variational problem of minimizing J_{IDR} amounts to solving a nonlinear set of algebraic equations in terms of the coefficient vector C_N . This is a particularly attractive feature of the IDR algorithm since it amounts to doing relatively little computation in arriving at the final solution.

Also, note that if f_0 is a positive function of x and y , then the reconstruction \hat{f}_{IDR} is necessarily a positive function as well. This is not accidental. In fact, Csizar [17] identified all distance measures that were consistent with a set of selection axioms (later interpreted as directed orthogonality in [45]) and concluded that if the functions involved are all real with both positive and negative values, then minimizing the least squares (L^2) measure is the only consistent choice; whereas, if all functions are required to be nonnegative, then the I-divergence measure is the only consistent choice. This fact is indeed a desirable property since we are reconstructing images (positive functions with greyscale intensity values between 0 and 255). The fact that we can produce positive functions as solutions, without explicitly making positivity a constraint, is one of the major advantages of using I-divergence as a regularization functional.

- Prior geometric information

The prior $f_0(x, y)$ can be used to directly incorporate geometric information into the reconstruction process. In particular, assume that after performing some geometric preprocessing on the data, such as extraction of support information [72, 54], or a preliminary parameterized reconstruction such as the polygonal reconstructions discussed in Chapter 3, an estimate is obtained of the region of the plane where the function f is nonzero (i.e. the spatial support of the

object). Then, according to this information, the prior $f_0(x, y)$ can be chosen as essentially an indicator function over this estimated region. Since the solution \hat{f}_{IDR} is composed of f_0 multiplied by another function as shown in (5.35), this in effect nulls out the part of the reconstruction that the geometric preprocessor eliminated as not being part of the spatial support of the object. This feature of the IDR algorithm is uniquely well suited to situations where it is important to concentrate the reconstruction on a particular region of interest. To be more specific, let us carry out a simple example.

Suppose that through some geometric preprocessing it is determined that the object of interest lies inside a disk of radius r centered at the origin, denoted by $D(0, r)$. We can pick $f_0(x, y)$ to reflect this information. This may be done in a variety of ways depending on the level of confidence which we place in the information provided by the geometric preprocessor. If the preprocessor tends to give very reliable information about the spatial support of the object of interest, then we may pick f_0 as follows

$$f_0(x, y) = \begin{cases} a & \text{if } (x, y) \in D(0, r) \\ \epsilon & \text{else} \end{cases} \quad (5.36)$$

where $a = \hat{\mu}_{00}/(\pi r^2)$, with $\hat{\mu}_{00}$ denoting the estimated value of $\iint_{\mathcal{O}} f(x, y) dx dy$ from the noisy projections. Here ϵ is a small but nonzero positive real number⁴.

If there is some quantifiable level of uncertainty associated with the information provided by the geometric preprocessor (say for instance uncertainty about the value of the radius of the region of interest r), this may be incorporated into the reconstruction process as a parameter that controls the transition of the values of $f_0(x, y)$ from points inside $D(0, r)$ to points outside of it. The function $f_0(x, y)$, as defined in (5.36), is discontinuous at the boundary of $D(0, r)$. The uncertainty in the predicted radius r can be translated into a smoother transition from

⁴Note that ϵ must be positive if the cost function $J_{IDR}(f, f_0)$ is to be well-defined.

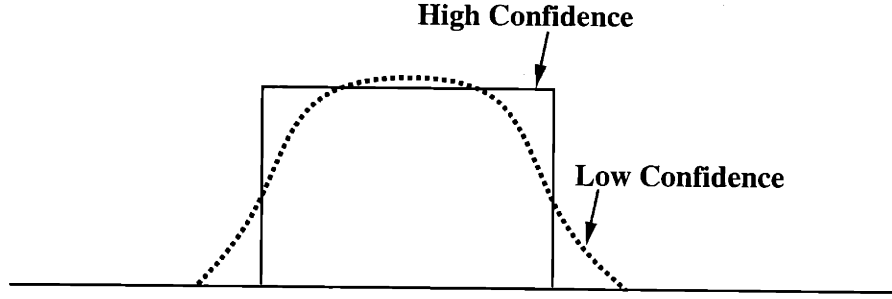


Figure 5-1: Taking uncertainty into account when picking prior

values inside the disk to those outside of it by picking $f_0(x, y)$ as illustrated in Figure 5-1. What has been suggested here is merely one way to deal with the issue of uncertainty. Many other possibilities exist. To some extent, the fact that many other approaches to incorporating prior geometric information exist in this context is an indication of the power and utility of our proposed IDR reconstruction algorithm.

- Approximate Solution

The local analysis of the equation (5.20) can be used to derive an approximate solution to this equation. This approximation is essentially the result of one iteration of Newton's algorithm discussed in Section 5.3.1.

Using the first order approximation given in (5.25), we can write

$$C_N \approx \frac{-1}{\gamma} \Sigma_N (\mathcal{L}_N(f_0) - \hat{\mathcal{L}}_N + D_0 C_N). \quad (5.37)$$

Solving this for C_N yields the approximate solution

$$\hat{C}_N \approx (\gamma I + \Sigma_N D_0)^{-1} \Sigma_N (\hat{\mathcal{L}}_N - \mathcal{L}_N(f_0)) \quad (5.38)$$

This, in turn, yields

$$\hat{f}_{IDR}(x, y) \approx f_0(x, y) \exp \left(\Phi_N^T(x, y) (\gamma I + \Sigma_N D_0)^{-1} \Sigma_N (\hat{\mathcal{L}}_N - \mathcal{L}_N(f_0)) \right). \quad (5.39)$$

5.4 Recursive Regularization

In this section we present a recursive refinement of the IDR and QR algorithms that is based on redefining the prior. In this formulation, an initial prior is chosen, and using this prior, a solution to the IDR (or QR) minimization problem is computed. This solution is then used as the prior for a new IDR cost function and the minimization is carried out again. We show that in the IDR case, the function to which the recursive regularization algorithm converges is exactly the solution of the following problem:

Problem 4

$$\min_f D(f, f_0), \quad \text{subject to } \mathcal{L}_N(f) = \widehat{\mathcal{L}}_N^{(c)} \quad (5.40)$$

where $\widehat{\mathcal{L}}_N^{(c)}$ denotes the projection, defined with respect to the inner product $\langle l_1, l_2 \rangle_{\Sigma_N} = l_1^T \Sigma_N l_2$, of $\widehat{\mathcal{L}}_N$ onto the range of the operator Ω_N . Here Ω_N denotes the operator mapping a square-integrable function $f \in L^2(D)$, with support in the unit disk, to its Legendre moments up to order N . Note that if $\widehat{\mathcal{L}}_N$ happens to be in the range $\mathcal{Ra}(\Omega_N)$ of the operator Ω_N , the constraint simply becomes $\mathcal{L}_N(f) = \widehat{\mathcal{L}}_N$.

Similarly, the recursive QR solution converges to the solution of

Problem 5

$$\min_f Q(f, f_0), \quad \text{subject to } \mathcal{L}_N(f) = \widehat{\mathcal{L}}_N^{(c)} \quad (5.41)$$

Hence, recursive regularization provides an iterative method of converting the *soft-constrained* solutions \widehat{f}_{IDR} and \widehat{f}_{QR} to *hard-constrained* solutions. If the estimated moments are not consistent, i.e. $\widehat{\mathcal{L}}_N \notin \mathcal{Ra}(\Omega_N)$, the proposed iterative algorithms implicitly compute and enforce the projection of $\widehat{\mathcal{L}}_N$ onto the set of consistent moments as hard constraints.

Two major advantages are gained from these recursive algorithms. The first is that we can improve the quality of our reconstructions by “controlling” how strictly the estimated moment information is enforced in the final solution. This “control” parameter is given by the number of iterations of the recursive algorithm that we wish to carry out. The second advantage is that the proposed algorithms *automatically* and *implicitly* compute and enforce the consistent component of the estimated moments as

hard constraints. This is particularly appealing since no explicit analytic description of the set $\mathcal{Ra}(\Omega_N)$ is known to exist [86].

Formally, let $J(f, f_0)$ denote the cost functional of either the IDR or the QR algorithms. In either case, the solution \hat{f} is obtained as

$$\hat{f} = \arg \min_f J(f, f_0) \quad (5.42)$$

Let us denote $\hat{f}_1 = \hat{f}$ and define \hat{f}_2 as

$$\hat{f}_2 = \arg \min_f J(f, \hat{f}_1). \quad (5.43)$$

In this way, we recursively have

$$\hat{f}_{k+1} = \arg \min_f J(f, \hat{f}_k). \quad (5.44)$$

This recursion forms the basis of the recursive regularization algorithms we discuss next.

5.4.1 Recursive I-divergence Regularization (RIDR)

Explicitly, the RIDR takes the form of the following recursion.

$$\hat{f}_{k+1}(x, y) = \arg \min_f \gamma_k D(f, \hat{f}_k) + (\mathcal{L}_N(f) - \hat{\mathcal{L}}_N)^T \Sigma_N (\mathcal{L}_N(f) - \hat{\mathcal{L}}_N) \quad (5.45)$$

By appealing to (5.35) the solution at each k may be written as

$$\hat{f}_{k+1}(x, y) = \hat{f}_k(x, y) \exp(\Phi_N^T(x, y) \hat{C}_N^{(k+1)}) \quad (5.46)$$

where

$$\hat{C}_N^{(k+1)} = \frac{-1}{\gamma_k} \Sigma_N (\mathcal{L}_N(\hat{f}_{k+1}) - \hat{\mathcal{L}}_N) \quad (5.47)$$

In terms of $\widehat{C}_N^{(k)}$, we may rewrite this as

$$\widehat{C}_N^{(k+1)} = \frac{-1}{\gamma_k} \Sigma_N \left(\iint_{\mathcal{O}} f_0(x, y) \Phi_N(x, y) \exp(\Phi_N^T(x, y) \sum_{j=1}^k \widehat{C}_N^{(j)}) dx dy - \widehat{\mathcal{L}}_N \right) \quad (5.48)$$

To investigate the stability of this algorithm, we make an approximation similar to (5.37). Expanding the right hand side of (5.48) in a Taylor series about $\mathbf{0}$ and solving for $\widehat{C}_N^{(k+1)}$ yields the following approximation

$$\widehat{C}_N^{(k+1)} \approx (\gamma I + \Sigma_N D_0)^{-1} \Sigma_N \left(\widehat{\mathcal{L}}_N - \mathcal{L}_N(f_0) - D_0 \sum_{j=1}^k \widehat{C}_N^{(j)} \right) \quad (5.49)$$

This approximation may be simplified significantly to yield

$$\widehat{C}_N^{(k+1)} \approx \left(I - (\gamma I + \Sigma_N D_0)^{-1} \Sigma_N D_0 \right)^k \widehat{C}_N^{(1)} \quad (5.50)$$

where an approximation to $\widehat{C}_N^{(1)}$ is given in (5.38). The linear approximation (5.50) implies a local sufficient condition for the stability of the RIDR algorithm. Formally, this condition is:

Result 11 *The recursive I-divergence regularization (RIDR) algorithm is strictly locally stable around $C_N = \mathbf{0}$ if the eigenvalues of the matrix*

$$A = I - (\gamma I + \Sigma_N D_0)^{-1} \Sigma_N D_0 \quad (5.51)$$

are strictly inside the unit disk.

Note that if the RIDR algorithm is locally stable around $\mathbf{0}$, then from (5.50) it follows that $\widehat{C}_N^{(k+1)} \rightarrow \mathbf{0}$ as $k \rightarrow \infty$. In what follows, we shall henceforth assume that the regularization factor γ is always chosen so that the RIDR algorithm is strictly locally stable around $\mathbf{0}$.

As an application of Result 11, let us assume that Σ_N is a diagonal matrix $\Sigma_N = \text{diag}[s_{ii}]$, and let us pick f_0 to be a constant prior with the value of c over the domain

\mathcal{O} . These assumptions imply that A is given by

$$A = \text{diag}[1 - (\gamma + cs_{ii})^{-1}cs_{ii}]. \quad (5.52)$$

so in order to satisfy the premise of result 11, for every $i = 1, 2, \dots, N(N+1)/2$, we must have

$$1 - (\gamma + cs_{ii})^{-1}cs_{ii} < 1. \quad (5.53)$$

Simplifying this gives the following equivalent condition.

$$\frac{\gamma}{\gamma + cs_{ii}} < 1. \quad (5.54)$$

This condition is satisfied whenever $c > 0$. This serves as an indication that the sufficient condition for the local stability of the RIDR algorithm, as given by Result 11, should not be difficult to satisfy in most cases.

5.4.2 Recursive Quadratic Regularization (RQR)

The recursion (5.44) can be written out for QR explicitly as

$$\hat{f}_{k+1} = \arg \min_f \gamma_k T(f, \hat{f}_k) + (\hat{\mathcal{L}}_N - \mathcal{L}_N(f))^T \Sigma_N (\hat{\mathcal{L}}_N - \mathcal{L}_N(f)). \quad (5.55)$$

From (5.34) we can write the solution for each k as

$$\hat{f}_{k+1}(x, y) = \hat{f}_k(x, y) + \Phi_N^T(x, y)(\gamma_k I + \Sigma_N)^{-1} \Sigma_N (\hat{\mathcal{L}}_N - \mathcal{L}_N(\hat{f}_k)). \quad (5.56)$$

Now note that the function $\Phi_N^T(x, y)\hat{\mathcal{L}}_N$ is the unique ⁵ fixed point of this iteration since it has moments exactly equal to $\hat{\mathcal{L}}_N$. Hence it follows that the solution to the recursive quadratic regularization problem is simply given by

$$\hat{f}_{RQR}(x, y) = \Phi_N^T(x, y)\hat{\mathcal{L}}_N, \quad (5.57)$$

⁵Again, uniqueness follows from convexity of the cost functional

which is, in fact, the best L^2 (i.e. least-squares error) reconstruction of f from its estimated moments.

5.4.3 Recursive Regularization and Equality-Constrained Minimization

At the outset of Section 5.4, we claimed that the RIDR and RQR algorithms solve Problems 4 and 5, respectively. In this section, we prove this assertion.

Let us consider Problem 4; the argument for Problem 5 is essentially identical. First consider the case when $\widehat{\mathcal{L}}_N \in \mathcal{Ra}(\Omega_N)$. It is well known [53, 89, 17] that a solution to Problem 4 in this case is unique and has the form⁶

$$f(x, y) = f_0(x, y) \exp(\Phi_N^T(x, y)K_N), \quad (5.58)$$

where the vector of constants K_N is chosen such that

$$\mathcal{L}_N(f) = \widehat{\mathcal{L}}_N \quad (5.59)$$

In fact, if a function of the form (5.58) exists and satisfies the constraints given by (5.59), then it is necessarily the unique solution of Problem 4. Hence, to show that \widehat{f}_{RIDR} solves Problem 4, it suffices to show that it has the form given by (5.58) and moments given by (5.59). Recall from (5.46) that at each iteration of RIDR we have

$$\widehat{f}_{k+1} = \widehat{f}_k \exp(\Phi_N^T(x, y)\widehat{C}_N^{(k+1)}). \quad (5.60)$$

This yields:

$$\widehat{f}_{k+1} = f_0 \exp(\Phi_N^T(x, y) \sum_{i=1}^{k+1} \widehat{C}_N^{(i)}), \quad (5.61)$$

Recall that we have assumed that the RIDR algorithm is stable around 0 so that $\widehat{C}_N^{(k)}$ converges to a finite limit point (namely 0). Now through 5.48, this implies that the

⁶Note that the existence of the solution is guaranteed by the assumption that $\widehat{\mathcal{L}}_N \in \mathcal{Ra}(\Omega_N)$

sum $\sum_{i=1}^{k+1} \widehat{C}_N^{(i)}$ also converges as $k \rightarrow \infty$.

Hence, as $k \rightarrow \infty$, in the limit, the RIDR solution has the same form as (5.58), with $K_N = \sum_{i=1}^{\infty} \widehat{C}_N^{(i)}$. At the fixed point of (5.60), the solution \widehat{f}_{RIDR} satisfies

$$\widehat{f}_{RIDR}(x, y) = \widehat{f}_{RIDR}(x, y) \exp(\Phi_N^T(x, y) \widehat{C}_N^{(\infty)}), \quad (5.62)$$

which, since the elements of the vector $\Phi_N(x, y)$ are linearly independent, implies that $\widehat{C}_N^{(\infty)} = 0$. This, in turn, through (5.47), implies that

$$\mathcal{L}(\widehat{f}_{RIDR}) = \widehat{\mathcal{L}}_N. \quad (5.63)$$

Therefore, $\widehat{f}_{RIDR}(x, y)$ is the unique solution of Problem 4 in the case $\widehat{\mathcal{L}}_N \in \mathcal{R}a(\Omega_N)$. Invoking the convexity of the functional $Q(f, f_0)$, a similar argument shows that \widehat{f}_{RQR} is the unique solution of Problem 5 when $\widehat{\mathcal{L}}_N \in \mathcal{R}a(\Omega_N)$.

Let us now discuss the case where $\widehat{\mathcal{L}}_N$ is inconsistent. The statement that $\widehat{\mathcal{L}}_N$ is inconsistent is equivalent to stating that it is not in the range of Ω_N . In practice, $\widehat{\mathcal{L}}_N$ may often lie outside of $\mathcal{R}a(\Omega_N)$ due to moment estimation error arising from noise in the measured projection data. We next show that the RIDR solution solves Problem 4 as follows

$$\min_f D(f, f_0), \quad \text{subject to } \mathcal{L}_N(f) = \widehat{\mathcal{L}}_N^{(c)} \quad (5.64)$$

where $\widehat{\mathcal{L}}_N^{(c)}$ is the ‘‘consistent part’’ of $\widehat{\mathcal{L}}_N$, in a sense to be made precise shortly.

Define the inner-product of two vectors l_1 and l_2 of length $\frac{(N+1)(N+2)}{2}$ (same length as $\widehat{\mathcal{L}}_N$) with respect to the matrix Σ_N as

$$\langle l_1, l_2 \rangle_{\Sigma_N} = l_1^T \Sigma_N l_2. \quad (5.65)$$

This inner product induces the norm $\|\cdot\|_{\Sigma_N}$ on the inner product space of real vectors

of length $\frac{(N+1)(N+2)}{2}$, and is defined by

$$\|l\|_{\Sigma_N} = (l^T \Sigma_N l)^{1/2}. \quad (5.66)$$

In terms of this norm, the RIDR algorithm may be described as

$$\hat{f}_{k+1} = \arg \min_f \gamma_k D(f, \hat{f}_k) + \|\hat{\mathcal{L}}_N - \mathcal{L}_N(f)\|_{\Sigma_N}^2 \quad (5.67)$$

Now define the projection operator onto $\mathcal{K} = \mathcal{Ra}(\Omega_N)$ with respect to the inner product $\langle \cdot, \cdot \rangle_{\Sigma_N}$ as

$$P_{\mathcal{K}}^{(\Sigma_N)} = \Omega_N (\Omega_N^* \Sigma_N \Omega_N)^{-1} \Omega_N^* \Sigma_N. \quad (5.68)$$

It is readily verified that this operator satisfies

$$P_{\mathcal{K}}^{(\Sigma_N)} P_{\mathcal{K}}^{(\Sigma_N)} = P_{\mathcal{K}}^{(\Sigma_N)} \quad (\text{idempotence}) \quad (5.69)$$

$$P_{\mathcal{K}}^{(\Sigma_N)T} \Sigma_N = \Sigma_N P_{\mathcal{K}}^{(\Sigma_N)} \quad (\text{self-adjoint with respect to } \langle \cdot, \cdot \rangle_{\Sigma_N}) \quad (5.70)$$

The projection operator onto the orthogonal complement of \mathcal{K} , denoted \mathcal{K}^\perp is given by

$$P_{\mathcal{K}^\perp}^{(\Sigma_N)} = I - P_{\mathcal{K}}^{(\Sigma_N)}. \quad (5.71)$$

The estimated moment vector $\hat{\mathcal{L}}_N$ can be decomposed into its consistent and inconsistent parts as follows

$$\hat{\mathcal{L}}_N = P_{\mathcal{K}}^{(\Sigma_N)} \hat{\mathcal{L}}_N + P_{\mathcal{K}^\perp}^{(\Sigma_N)} \hat{\mathcal{L}}_N = \hat{\mathcal{L}}_N^{(c)} + \hat{\mathcal{L}}_N^{(i)}, \quad (5.72)$$

where $\hat{\mathcal{L}}_N^{(c)}$ and $\hat{\mathcal{L}}_N^{(i)}$ respectively denote the consistent and the inconsistent components of $\hat{\mathcal{L}}_N$, which by definition are orthogonal with respect to the inner product $\langle \cdot, \cdot \rangle_{\Sigma_N}$. Replacing $\hat{\mathcal{L}}_N$ in (5.67) by its decomposition and invoking the orthogonality of the components we may write

$$\hat{f}_{k+1} = \arg \min_f \gamma_k D(f, \hat{f}_k) + \|\hat{\mathcal{L}}_N^{(c)} - \mathcal{L}_N(f)\|_{\Sigma_N}^2 + \|\hat{\mathcal{L}}_N^{(i)}\|_{\Sigma_N}^2 \quad (5.73)$$

Now clearly, $\mathcal{L}_N(\hat{f}_k) \in \mathcal{Ra}(\Omega_N)$ at every iteration k . Hence, the estimates \hat{f}_k do not depend on the inconsistent part of the estimated moments $\tilde{\mathcal{L}}_N^{(i)}$. Therefore, we may drop the last term on the right hand side of (5.73) without changing the solution of the optimization problem (5.73). This, in other words, implies that the RIDR algorithm converges to the solution of

$$\min_f D(f, f_0), \quad \text{subject to } \mathcal{L}_N(f) = \tilde{\mathcal{L}}_N^{(c)} \quad (5.74)$$

where

$$\tilde{\mathcal{L}}_N^{(c)} = P_K^{(\Sigma_N)} \tilde{\mathcal{L}}_N \quad (5.75)$$

is the “consistent part” of the estimated moment vector. A particularly interesting feature of this property of the RIDR algorithm is that an explicit description of the set $\mathcal{Ra}(\Omega_N)$ is never used in the algorithm. This is convenient since to our knowledge, one has not been discovered as of yet [86]!

5.5 Numerical Examples

In this section, we study the performance of our proposed algorithms using three distinct phantoms. We present experiments with various different choices of initial estimates, various number of views and at different SNR’s. To present a coherent set of experiments, we have organized this section as follows.

1. The phantom to be reconstructed in the first set of examples is shown in Figure 5-2. Here, we will show reconstructions of this phantom using the classical filtered-backprojection algorithm along with reconstructions produced by the RQR, IDR, and RIDR algorithms which use the classical FBP reconstruction as an initial estimate. Other initial estimates, namely produced by maximizing the Shannon and Burg entropy criteria, are also presented. The intent here is to show the relative performance of our proposed algorithms with respect to the choice of various different initial estimates. In addition, for the case when the FBP is used to construct an initial estimates, we also present experiments

to study the performance of our algorithms with respect to various SNR's and different number of moments used in the reconstructions.

2. The phantom to be reconstructed in the second set of examples is shown in Figure 5-17. This phantom is composed of circular regions with linearly decreasing radii and linearly increasing activity levels. The choice of this phantom was motivated by the need to study the spatial and greyscale resolution of our reconstruction algorithm. We compare the performance of the RIDR algorithm using FBP-based initial estimates to the FBP reconstructions for various cases where different number of projection views are used at a fixed SNR.
3. The phantom to be reconstructed in the last set of examples is shown in Figure 5-25. We study RIDR reconstructions of this phantom using FBP-based initial estimates at various SNR's. The choice of this phantom was motivated by two factors. First, this is the same as the canonical example used in [72], which is the direct predecessor of this thesis. The second reason for using the "MIT ellipse" image is to further study the level of geometric detail that the RIDR algorithm can extract from a given set of noisy data. We also present performance curves for mean-squared reconstruction errors versus SNR.

Throughout what follows, the SNR is defined the same way as in Chapter 3 and is given by

$$\text{SNR} = 10 \log_e \left(\frac{\sum_i \sum_j g^2(t_i, \theta_j) / (m \times n)}{\sigma^2} \right) \quad (5.76)$$

where m denotes the number of views and n denotes the number of samples per view, and σ^2 is the variance of the corrupting noise. In order to quantify the quality of the reconstructions in all examples that follow, we define the percent Mean-Squared Error (MSE) as:

$$\% \text{ MSE} = \frac{\iint (f - \hat{f})^2 dx dy}{\iint f^2 dx dy} \times 100\% \quad (5.77)$$

5.5.1 Example set I

In this section, we reconstruct the phantom shown in Figure 5-2 from estimated orthogonal moments up to order 8 via the IDR, RQR and the RIDR algorithms. We show how the choice of different initial estimates affect the reconstructions. The results of RIDR reconstructions using different initial estimates are summarized in Figure 5-16.

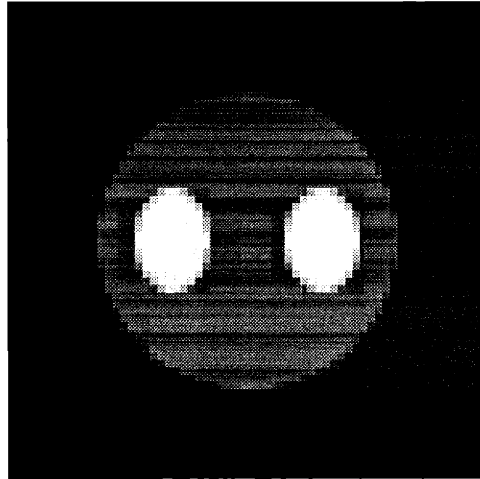


Figure 5-2: Phantom for Example set I

In this example we measured 64 projections with 64 samples per projection at various signal to noise ratios. The original noiseless and noisy sinograms are shown in Figure 5-3. To start, we present the classical FBP reconstruction produced from the data in Figure 5-4 with SNR=10.

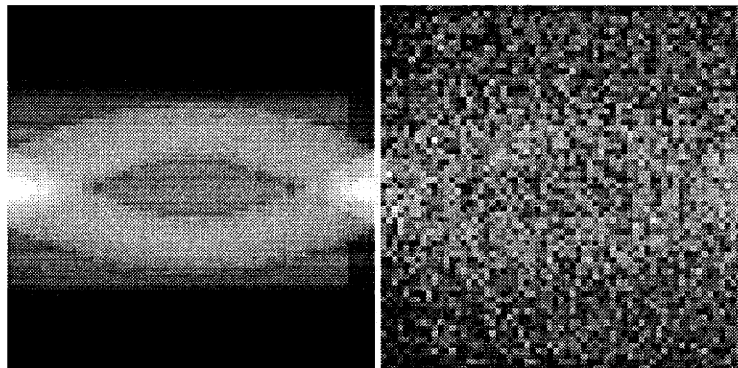


Figure 5-3: Left to Right: Sinogram of Phantom, Noisy Sinogram at SNR of 10

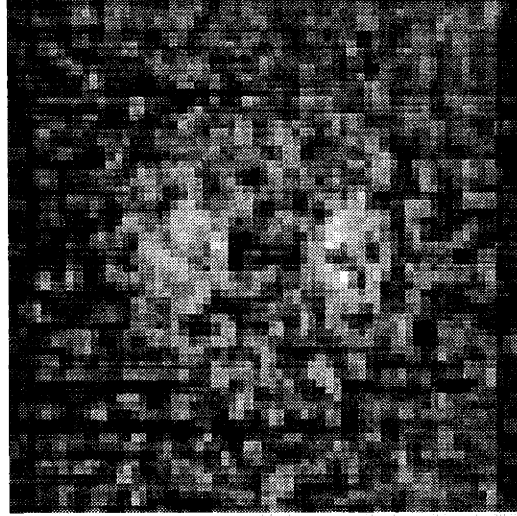


Figure 5-4: Filtered Backprojection with 3rd order Butterworth filter with 0.25 normalized cutoff frequency. SNR=10

This reconstruction is produced with a 3-rd order Butterworth filter with 0.25 normalized cutoff frequency. This reconstruction is the “best” FBP reconstruction in the sense that the filter order and cutoff frequency have been carefully chosen to produce (visually) the best reconstruction possible. In a series of experiments that follow, this FBP reconstruction is used to produce an initial estimate of the underlying phantom. The pixel values of the FBP reconstruction are, in general, not guaranteed to be positive values. Hence, in order to use the FBP reconstruction as an initial estimate, we add a number to each pixel value in the FBP image in order to maintain positivity. Furthermore, to speed up the convergence, we scaled the result so as to produce an initial estimate with integral equal to the estimated zeroth order moment. In particular, let \hat{f}_{FBP} denote the filtered-backprojection reconstruction. Then our FBP-based initial estimate f_0 is given by

$$f_0 = \alpha(\hat{f}_{FBP} + \beta) + \epsilon \quad (5.78)$$

where

$$\beta = \begin{cases} -\min_{(x,y)} \hat{f}_{FBP} & \text{if } \min_{(x,y)} \hat{f}_{FBP} < 0, \\ 0 & \text{if } \min_{(x,y)} \hat{f}_{FBP} \geq 0 \end{cases} \quad (5.79)$$

and

$$\alpha = \frac{\hat{\mu}_{00} - \epsilon A}{\iint \hat{f}_{FBP}(x, y) dx dy + \beta A}, \quad (5.80)$$

where $A = 4$ is the area of the square pixel grid $[-1, 1] \times [-1, 1]$ over which \hat{f}_{FBP} is defined, and $\epsilon \approx 10^{-16}$ is a constant value added in to maintain positivity. In our experience, the scaling α (which is entirely data-driven) produces better reconstructions with fewer iterations of the RIDR algorithm.

Quadratic Reconstruction with FBP Initial Estimate

The initial estimate f_0 for this reconstruction was chosen as the offset and scaled version of the best FBP reconstruction given by (5.78). Here we use moments up to order 8. This FBP-based initial estimate, along with the RQR reconstruction after 2 iterations and the RQR reconstruction after 8 iterations are shown in Figure 5-5. No further improvement was observed after 8 iterations. (Note that actually, the RQR solution can be computed directly in closed form without any iterations as described in Section 5.4.2) Note that while this reconstruction is an improvement over the FBP reconstruction both visually and in terms of MSE, it appears “smeared” or “diffuse” with little edge information preserved. As we shall see, the RIDR algorithm produces reconstructions which give rise to better resolution of local and global geometric information.

Uniform Prior Over the Unit Disk

The *Maximum Entropy* solution subject to the estimated moments can be arrived at by choosing a uniform prior f_0 . In this case, we pick f_0 to be a uniform image over the unit disk such that its total area is equal to the estimated zeroth order moment $\hat{\mu}_{00}$. i.e.

$$f_0(x, y) = \begin{cases} c & \text{if } (x, y) \in D, \\ 0 & \text{if } (x, y) \notin D \end{cases} \quad (5.81)$$

where

$$c = \hat{\mu}_{00}/\pi \quad (5.82)$$

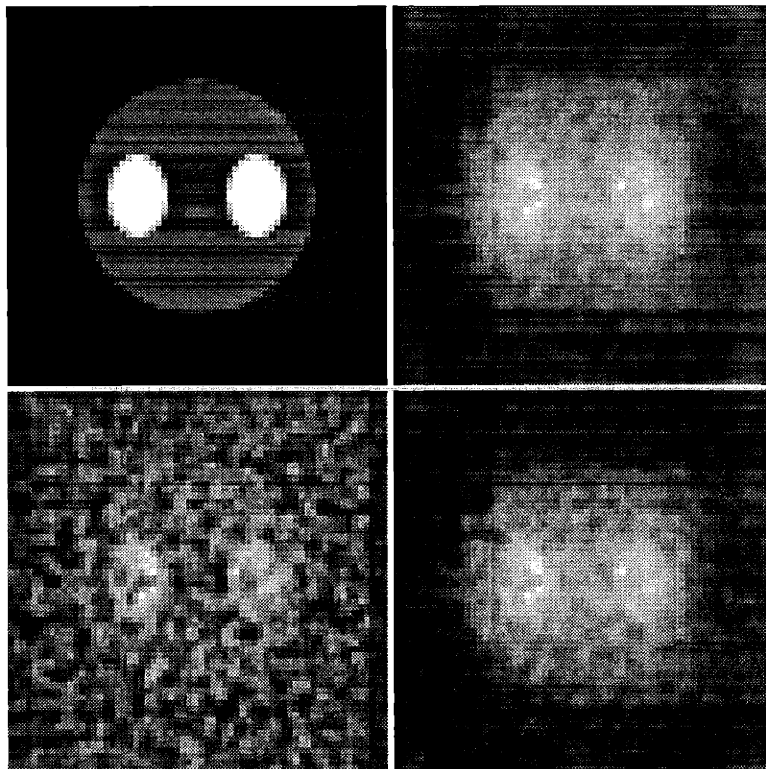


Figure 5-5: Counter-clockwise from upper left: Phantom, computed FBP initial estimate (% MSE=55.3, RQR result after 2 iterations (% MSE=17.8), RQR result after 8 iterations (% MSE=13.4). Data: 64 projections w/ 64 samples per projection at SNR= 10; moments up to order 8 were used.

Moments up to order 8 were used. The phantom, the prior estimate, the IDR and the RIDR reconstructions are shown in Figure 5-6. A value of $\gamma = 35$ was used and a total number of 10 iterations of the RIDR algorithm resulted in convergence. As can be seen, the RIDR reconstruction produces a very rough estimate of the underlying image with smooth or “flattened” edge regions. This is essentially due to the fact that the Maximum Entropy prior seeks the “flattest” reconstruction that matches the data best. As we will show next, a modestly more informative initial estimate, such as the FBP-based initial prior can yield significantly better RIDR reconstructions.

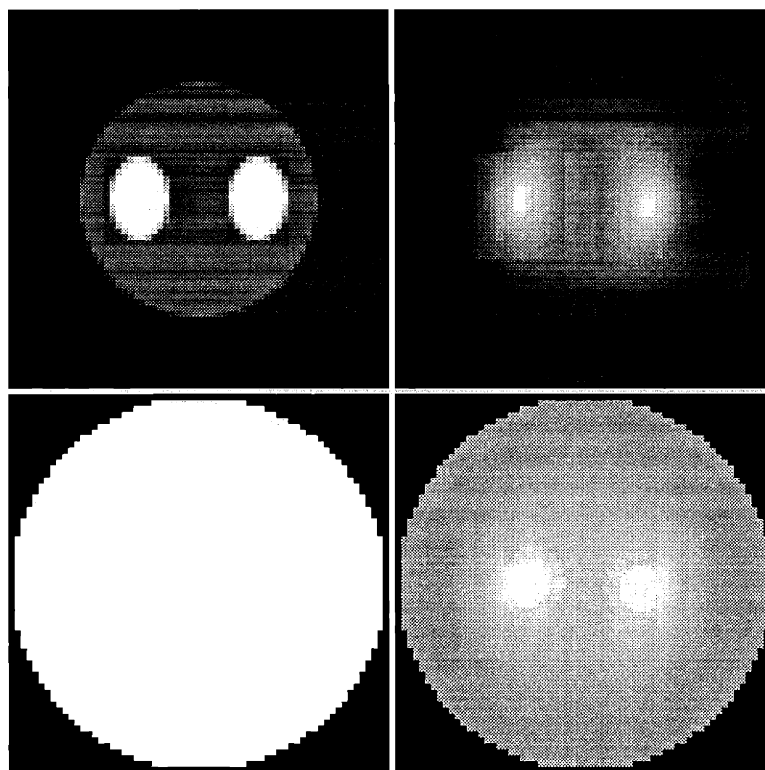


Figure 5-6: Counter-clockwise from upper left: Phantom, uniform initial estimate (% MSE=65.7), IDR solution (% MSE=55.9), RIDR solution (% MSE=15.8). Data: 64 projections w/ 64 samples per projection at SNR= 10; moments up to order 8 were used.

Burg Initial Estimate

In this case, the initial estimate f_0 was computed from the estimated moments according to the following :

$$f_0 = \arg \min_f \gamma_0 \int \int_D f - \log(f) dx dy + (\hat{\mathcal{L}}_N - \mathcal{L}_N(f))^T \Sigma_N (\hat{\mathcal{L}}_N - \mathcal{L}_N(f)), \quad (5.83)$$

with $\gamma_0 = 22$, and $N = 8$ (i.e. moments up to order 8 used). The first term in this cost function corresponds to the *Burg Entropy* of f , which is known to give “peaked” or “spikey” results [45]. This procedure was designed to avoid choosing a uniform prior as f_0 . The choice of a uniform prior produces the maximum *Shannon* entropy solution which in contrast to the Burg entropy prior, tends to suppress features and “flatten out” the reconstruction. The resulting f_0 from minimizing (5.83) has the form

$$f_0 = \frac{1}{1 + \gamma_0^{-1} \Phi_N^T(x, y) \Sigma_N (\mathcal{L}_N(f_0) - \hat{\mathcal{L}}_N)}, \quad (5.84)$$

which can be solved for f_0 according to the analogous iterative procedure to that described in Section 5.3.1. Next, the solution given by (5.84) is used directly as the initial estimate ⁷. The initial estimate, the IDR solution and the RIDR solution are shown in Figure 5-7. In this reconstruction moments up to order 8 were used with data composed of 64 projections with 64 samples per projection at SNR of 10.

FBP Initial Estimate

In this section we produce reconstructions of the phantom shown in Figure 5-2 with a prior estimate f_0 which is picked to be an offset and scaled version of the filtered-backprojection solution shown in Figure 5-4. Again moments up to order 8 were used in the reconstruction.

Due to the nonuniform nature of the FBP-based initial estimate, a rather large value of the regularization factor ($\gamma = 200$) had to be used first in order to keep the

⁷Note that this initial estimate is, in general, not guaranteed to be positive, but in this example, it happened to be positive so that no further processing was necessary to construct a “valid” initial estimate

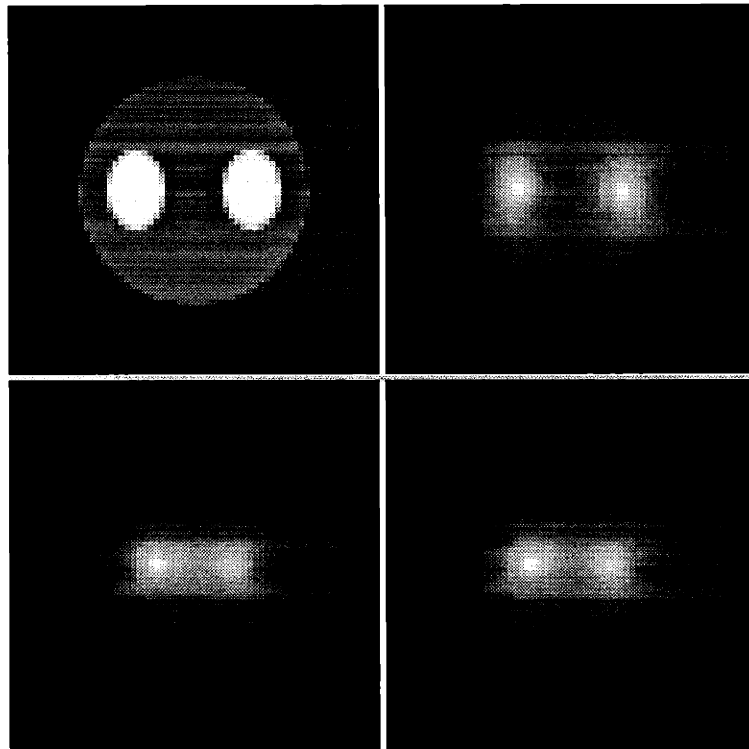


Figure 5-7: Counter-clockwise from upper left: Phantom, Initialization computed using (5.83) without support information (% MSE=39.8), IDR solution (% MSE=31), RIDR solution (% MSE=10.3). Data: 64 projections, 64 samples per projection with SNR=10; moments up to order 8 used.

IDR algorithm stable. After having computed the IDR solution with this value of γ , it was then possible to use a smaller value of $\gamma = 30$ in subsequent iterations of the RIDR algorithm. Figure 5-8 shows the initial estimate f_0 , the RIDR solution after 3 iterations, and the RIDR solution after 10 iterations. No marked improvement was observed in further iterates of the RIDR algorithm in this case. Note that the RIDR algorithm has begun to eliminate the spurious features produced by the FBP solution after only 3 iterations. After 10 iterations, it has successfully eliminated essentially all the clutter produced by FBP that was not part of the original image. It has further improved the quality of the reconstruction within the spatial support of the object. It is worth noting the fact that our reconstructions quickly null out all but a circular region tangent to the square image grid. This is a result of the assumption that all functions to be reconstructed are nonzero only inside the unit disk. Consistent with this assumption, the orthonormal Legendre polynomials $P_k(t)$ over $[-1, 1]$ were used in computing the moments of the projections from which the moments of the underlying image were estimated.

Effect of the number of moments used on performance:

To see the effect of the number of moments used in the reconstructions, we present MSE values versus number of moments used in Figure 5-9. In this graph, the MSE error of the RIDR reconstructions is shown when moments up to order 2, 5, 8, and 11 were respectively used in the reconstruction. In these reconstructions, the data was composed of 64 projections with 64 samples per projection at SNR of 10, and all used the same initial estimate which was chosen to be the FBP-based initial estimate. Figure 5-10 shows the reconstructions using 2, 5, 8, and 11 moments respectively in counter-clockwise order.

As can be seen from Figure 5-9, the use of higher order moments does improve the reconstructions. However, after some point (in this case moments up to order 8) the effect of using higher order moments is not significant, This phenomenon is essentially due to the effect of the significantly reduced weighting placed on these higher order estimated moments by the inverse of the moment estimation error covariance matrix

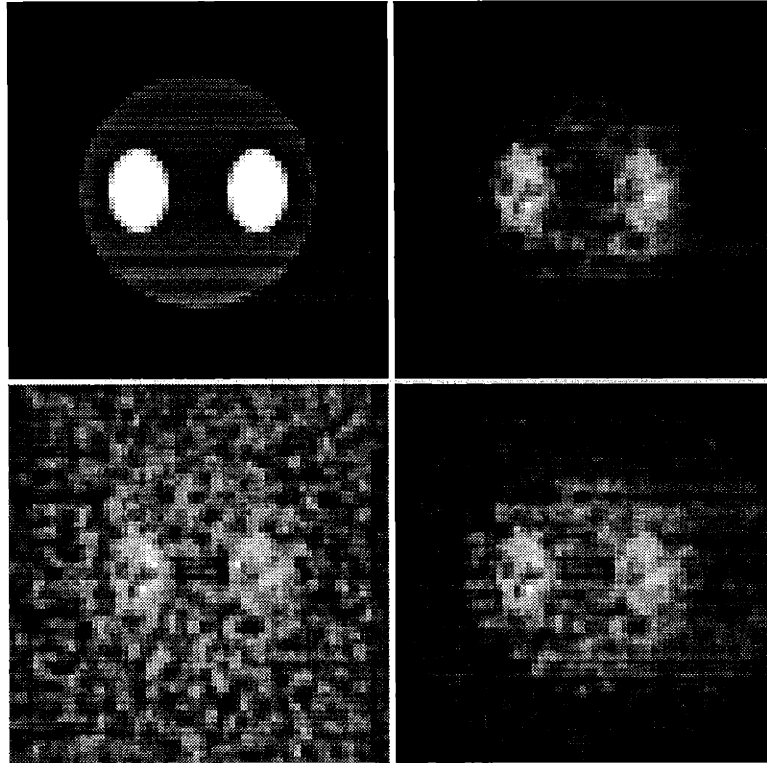


Figure 5-8: Counter-clockwise from upper left: Phantom, f_0 based on FBP (% MSE=55.3), RIDR solution after 3 iterations (% MSE=38.1), RIDR solution after 10 iterations (% MSE=11.1). Data: 64 projections w 64 samples per projection at SNR= 10; moments up to order 8 used.

which is explicitly used in the RIDR algorithm.

Effect of noise on performance:

To show the effect of noise on the quality of the reconstructions, Figure 5-11 shows the MSE values of the FBP reconstruction and the final RIDR reconstructions with FBP-based initial estimates at various SNR's. In these reconstructions, estimated moments up to order 8 were used. As can be seen, our algorithm performs significantly better than the FBP algorithm at all SNR's while at higher SNR's this better performance is more dramatic. This indicates that, in terms of MSE, the RIDR algorithm is most useful in low SNR scenarios.

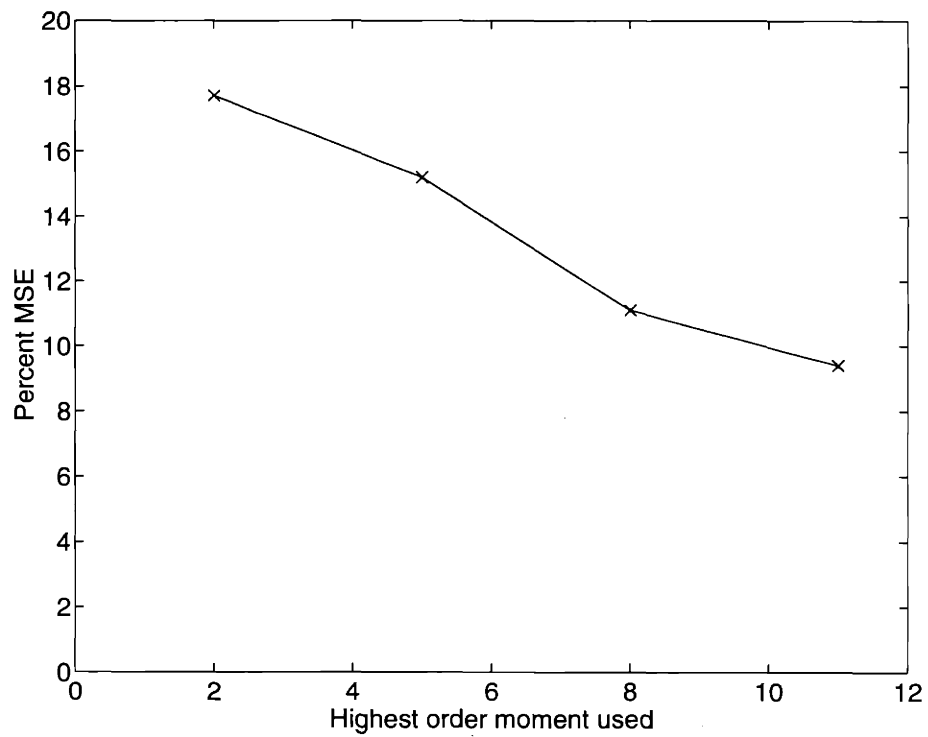


Figure 5-9: MSE versus number of moments used in reconstructing the phantom of Figure 5-2

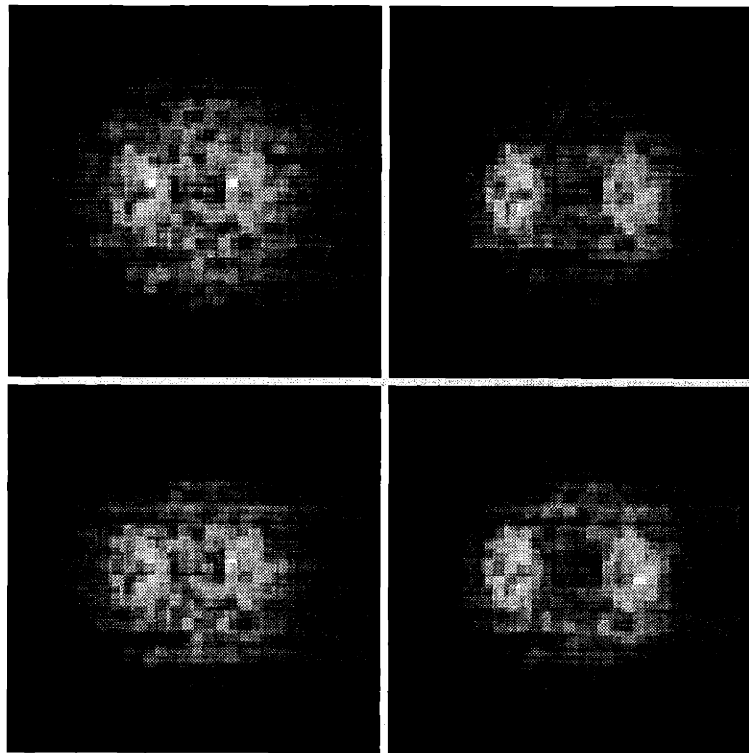


Figure 5-10: Counter-clockwise from upper left: reconstructions using moments up to order 2, 5, 8, and 11. Data: 64 projections w/64 samples per projection at SNR= 10. Initial guess was based on FBP in every case.

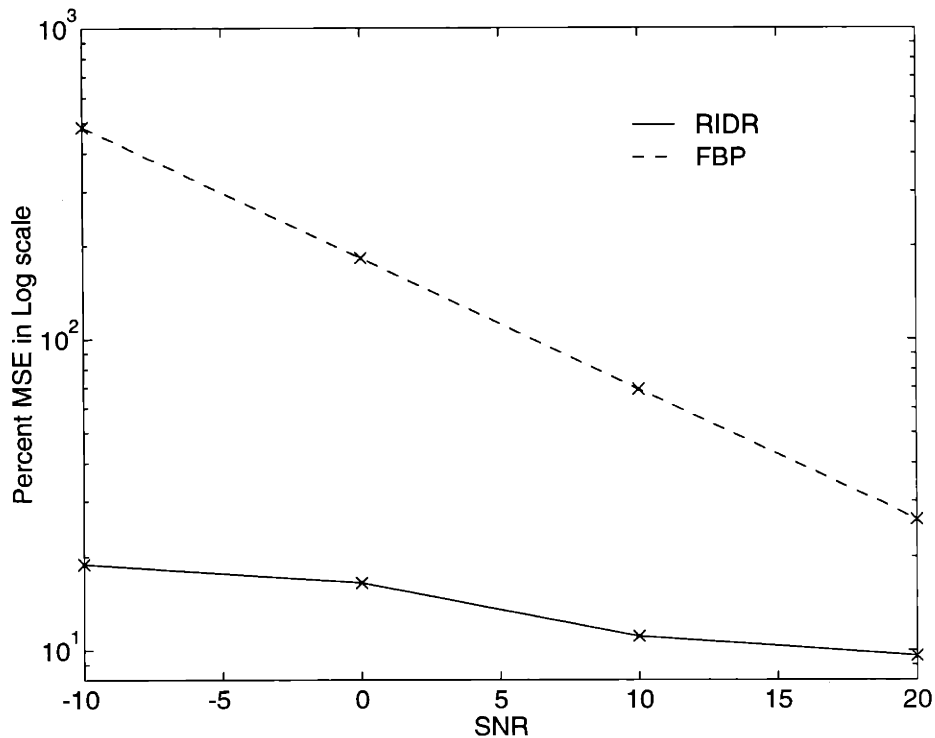


Figure 5-11: MSE versus SNR in reconstructing the phantom of Figure 5-2: Moments up to order 8 used.

Effect of Perfectly Known Support

In this subsection, we study the effect of perfectly known support on the performance of our RIDR algorithm. We combine the known support with the Burg initial estimate, and the FBP-based initial estimate. It assumed that the spatial support of the phantom is perfectly known, and this information is encoded into the initial guess by multiplying the solution of (5.84) and the FBP-based prior (5.78) by the indicator function of the supporting set. The resulting image is then normalized according to a similar procedure as in (5.78) so as to ensure having an initial estimate that is I) positive and II) has zero-th order moment equal to the estimated zero-th order moment $\hat{\mu}_{00}$. The resulting f_0 is then used to compute the IDR solution and the RIDR solution. Figure 5-12 shows the reconstructions using the Burg initial estimate with perfect support information and using moments up to order 8. The data consisted of 64 projections with 64 samples per projection at SNR of 10. It is seen that the computed f_0 gives a very coarse estimate of where the main features of the object lie. The IDR solution refines this initial estimate. The RIDR solution refines it further to an image in which the features are quite distinctly captured. The marked improvement

Image	Burg est. w/o suppt	Burg est. w/ suppt	$\hat{f}_{R IDR}$ w/o suppt	$\hat{f}_{R IDR}$ w/ suppt
% MSE	39.8	37.9	10.3	8.9
Image	FBP est. w/o suppt	FBP est. w/ suppt	$\hat{f}_{R IDR}$ w/o suppt	$\hat{f}_{R IDR}$ w/ suppt
% MSE	55.3	10.9	11.1	6.2

Table 5.1: MSE values for reconstructions using Burg prior with and without support information

over the FBP solution is quite clear.

Figure 5-13 shows the reconstructions using the FBP-based initial estimate with perfect support information and using moments up to order 8. The data again consisted of 64 projections with 64 samples per projection at SNR of 10. Again it is seen that the coarse initial estimate provided by the FBP reconstruction is improved in the final RIDR reconstruction to a point where the final reconstruction contains much of the geometric information of the underlying image. Table 5.1 shows the MSE values for the RIDR reconstructions using the Burg-based and FBP-based initial priors with and without knowledge of the support of the underlying object. It is seen that the reconstructions using the Burg estimate with and without support information differ by little in terms of the MSE values incurred. They also look rather similar in visual terms. The reconstruction using the FBP-based initial estimate with support information is somewhat better than its counterpart without support information shown in Figure 5-8 on page 177. This difference is seen both visually and in terms of the improved MSE values shown in Table 5.1.

These observations indicate that the RIDR algorithm is capable of producing better reconstructions if provided with a-priori information regarding the spatial support of the image. This information is, however, not necessary since the RIDR algorithm appears to do a fairly good job of localizing the important features of the image without a-priori support information.

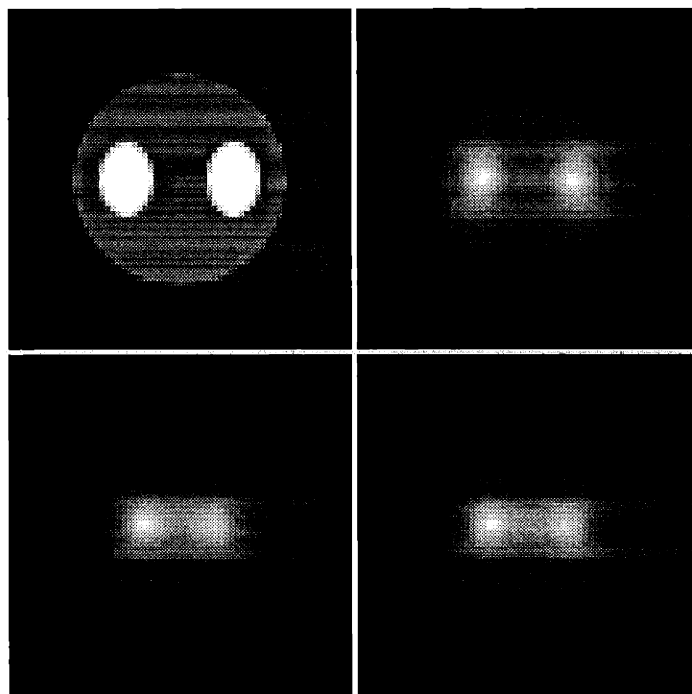


Figure 5-12: Counter-clockwise from upper left: Phantom, Initialization computed using (5.83) with support information (% MSE=37.9), IDR solution (% MSE=28), RIDR solution (% MSE=8.9). Data: 64 projections, 64 samples per projection with SNR=10; moments up to order 8 used.

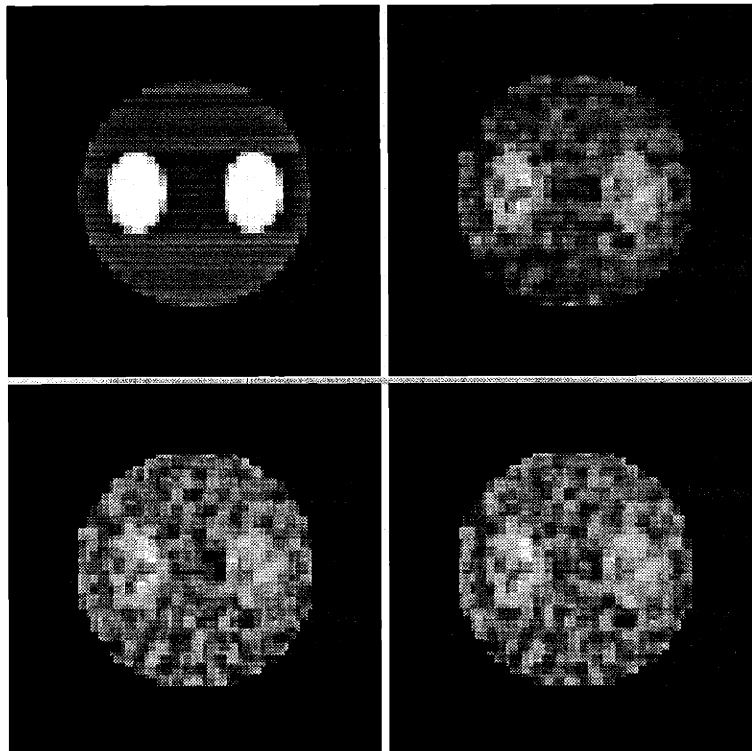


Figure 5-13: Counter-clockwise from upper left: Phantom, Initialization computed using FBP with support information (% MSE=10.95), IDR solution (% MSE=10.7), RIDR solution (% MSE=6.2). Data: 64 projections, 64 samples per projection with SNR=10; moments up to order 8 used.

Effect of Estimated Support

In this subsection we present a comprehensive example of the application of the concepts and algorithms discussed throughout this thesis. Specifically, we first use the concepts in Chapter 3 to estimate the spatial support of the underlying phantom from the noisy projection data by fitting a binary polygonal region to the data. We use the MDL criterion, in fact, to find a polygonal region of least number of sides that fits the data best. Next, we define the indicator set of this estimated region and use this to constrain the best FBP solution to the estimated spatial support. The output of this process is then used as the initial estimate, f_0 for the RIDR algorithm.

First, in order to obtain a reasonable polygonal fit to the given data, we normalize the data by its average value. i.e. we define the new data set \bar{Y} as

$$\bar{Y}(t_i, \theta_j) = \frac{\Delta t \Delta \theta}{2\pi \sum_i \sum_j Y(t_i, \theta_j)} Y(t_i, \theta_j). \quad (5.85)$$

Recall the assumption that all images in this thesis are contained in the unit disk. Hence, the above renormalization of the data is necessary to ensure that the polygonal reconstruction with unit amplitude is indeed contained inside the unit disk. Figure 5-14 shows the MDL cost values incurred versus the number of sides of the polygonal fit to the normalized data. The minimum of this graph at $N = 5$ indicates that a 5-sided polygonal fit is appropriate.

A value of $\gamma = 30$ was again applied for the regularization parameter, while moments up to order 8 were used. The data set here is still 64 views with 64 samples per view at SNR of 10. Figure 5-15 shows the phantom, the initial estimate, the solution after 3 iterations of the RIDR algorithm and the solution after 10 iterations of the RIDR algorithm. More iterations did not produce a significant difference in the reconstructed image. Moments up to order 8 were used in the reconstructions.

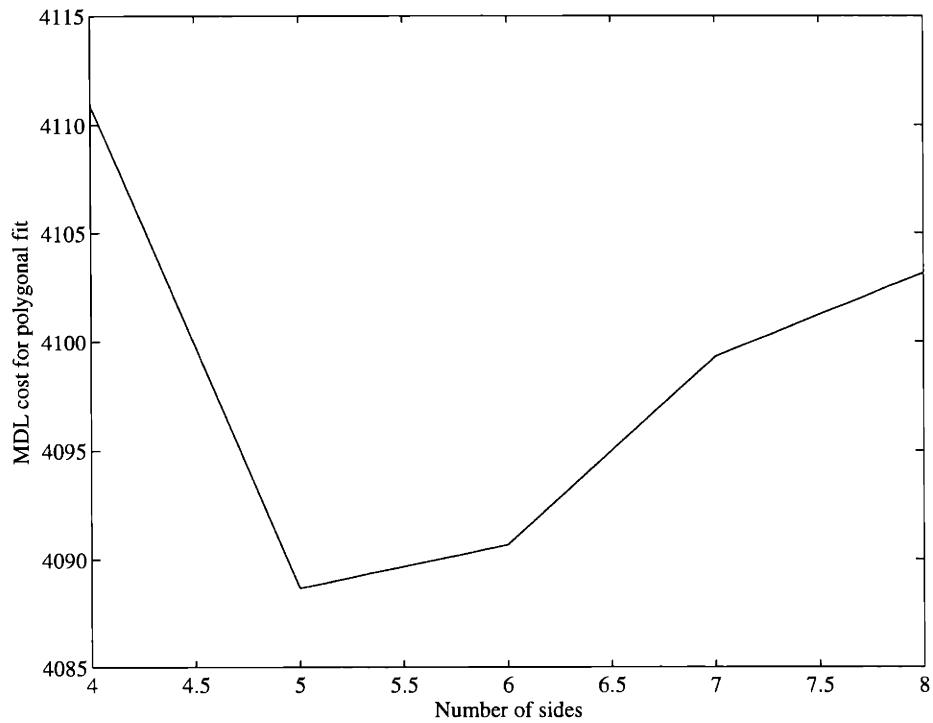


Figure 5-14: MDL cost for polygonal fit vs number of sides

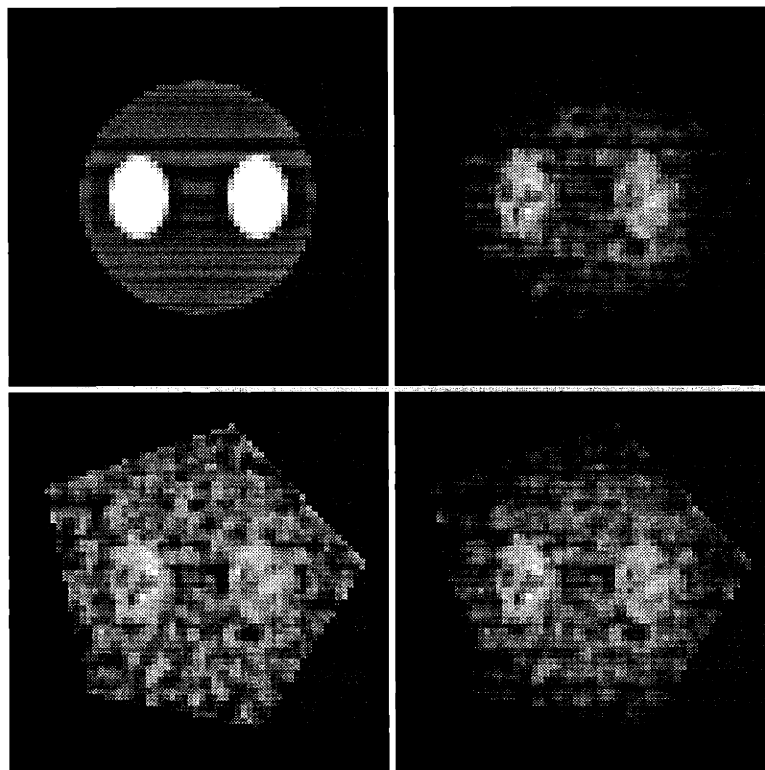


Figure 5-15: Counter-clockwise from upper left: Phantom, FBP-based initial estimate with estimated support (% MSE=24.3), RIDR result after 3 iterations (% MSE=15.3), RIDR result after 10 iterations (% MSE=9.4). Data: 64 views, 64 samples per view at SNR=10; moments up to order 8 used.

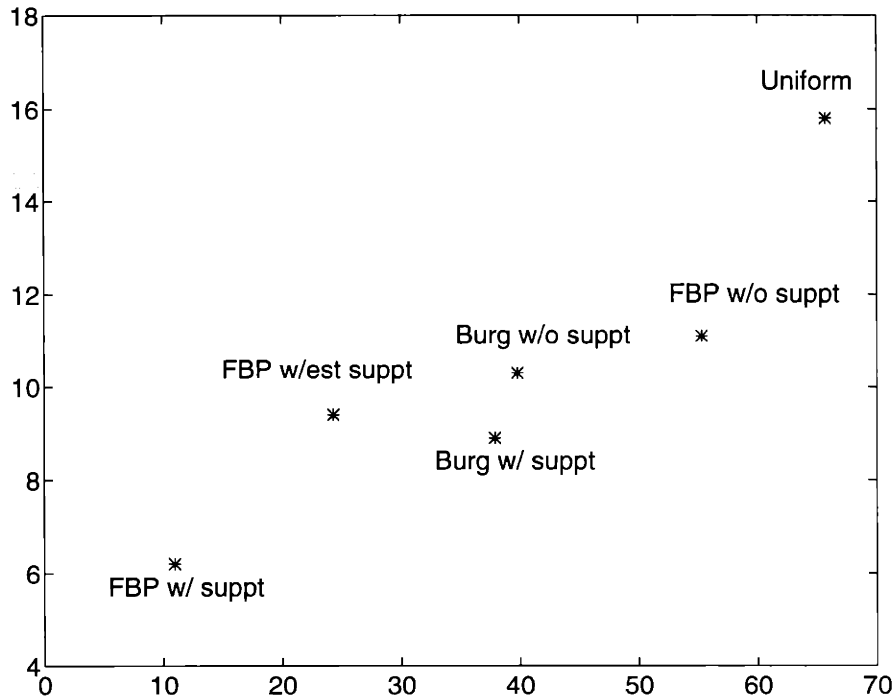


Figure 5-16: MSE of Initial Estimates vs. MSE of final RIDR reconstructions for Example set I: Moments up to order 8 used

5.5.2 Example set II

In this set of examples, we reconstruct a new phantom shown in Figure 5-17. The phantom to be reconstructed is composed of circular regions with linearly decreasing radii and linearly increasing activity levels. The choice of this phantom was motivated by the need to study the spatial and greyscale resolution of our reconstruction algorithms. We compare the performance of the RIDR algorithm to that of the filtered-backprojection algorithm for various cases where different number of projection views are available at a fixed signal to noise ratio of 10^8 . Figure 5-24 summarizes the MSE values versus the number of views for the various reconstructions we discuss next.

Due to the increased complexity of the new phantom we wish to reconstruct, moments of up to order 10 were used in the reconstructions reported in this example. In order to speed up the convergence of the RIDR algorithm and ensure stability, different value of the regularization parameter were applied at different iterations. In

⁸Note that the SNR per sample in the sinogram is kept fixed while the number of samples is increased. This means that for higher number of views, the variance of the noise corrupting the samples is proportionally increased to keep SNR per sample at a fixed value of 10

all reconstructions for this example, the (offset and scaled) best FBP reconstruction was used as the initial estimate.

32 views, SNR= 10

We first collect 32 equally-spaced views between 0 and π with 64 samples per view at SNR of 10 . The noiseless and noisy sinograms are given in Figure 5-19. Figure 5-18 shows the Phantom, the initial estimate based on the best FBP solution, the IDR and the RIDR solution after 15 iterations. The MSE values are shown in Figure 5-24. It is seen from the sample reconstruction that with 32 projections, the RIDR has been able to roughly extract the largest and the brightest of the circles. It has not, however, been able to reconstruct the circle of medium size with modest activity level which is the one on the lower right side of the phantom. This indicates that with fewer projections, the reconstructed images show less geometric and amplitude resolution, as we would expect.

64 Views, SNR = 10

Here we use 64 views with 64 samples per view at a signal to noise ratio of 10. The noiseless and noisy sinograms are displayed in Figure 5-21.

On the first pass of the RIDR algorithm (otherwise called IDR) we used $\gamma = 500$ while on the second iteration this value was reduced to $\gamma = 50$. Finally, seven more iterations were carried out with $\gamma = 15$, at which point convergence was attained. The initial FBP-based estimate, the solution after the first iteration and the final reconstruction are shown in Figure 5-20. In this case, we can see that all four circular features are accentuated in the final RIDR reconstruction, showing good geometric resolution.

128 views, SNR=10

Finally, we collect 128 equally-spaced views between 0 and π with 64 samples per view at SNR of 10. The noiseless and noisy sinograms are given in Figure 5-23. Figure 5-22 shows the Phantom, the initial estimate, the IDR and the RIDR solution

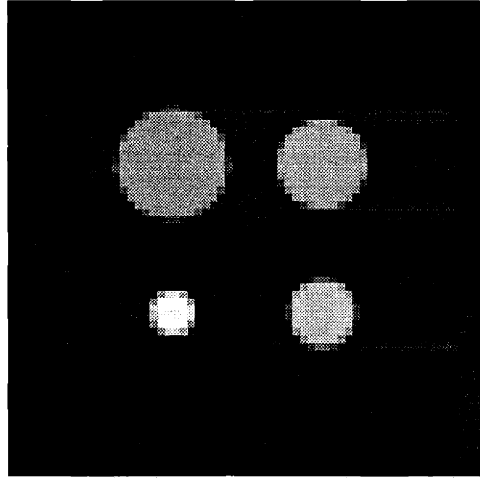


Figure 5-17: Phantom for Example set II

after 10 iterations. The MSE values are shown in Figure 5-24. It is seen that the reconstruction is improved in visual terms (except for the smallest circle), while the MSE improvement is not significant when compared to the previous case where 64 views were used.

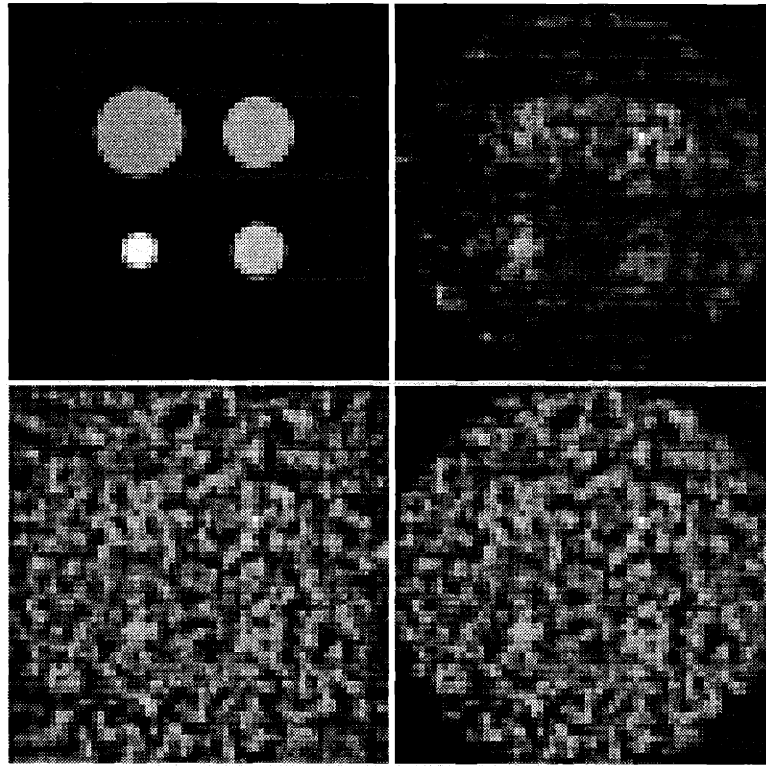


Figure 5-18: Counter-clockwise from top left: Phantom, Initial Estimate from FBP, IDR reconstruction, Final RIDR reconstruction (32 views, 64 samples per view, SNR =10, moments up to order 10 used)

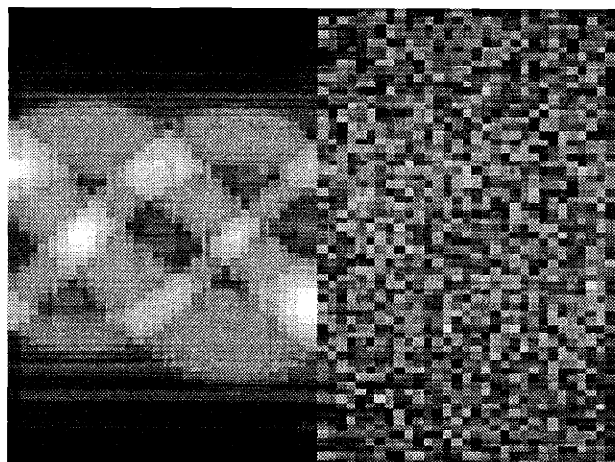


Figure 5-19: Left to right: Noiseless and noisy sinograms for phantom in Figure 5-17: 32 views, 64 samples per view, SNR =10

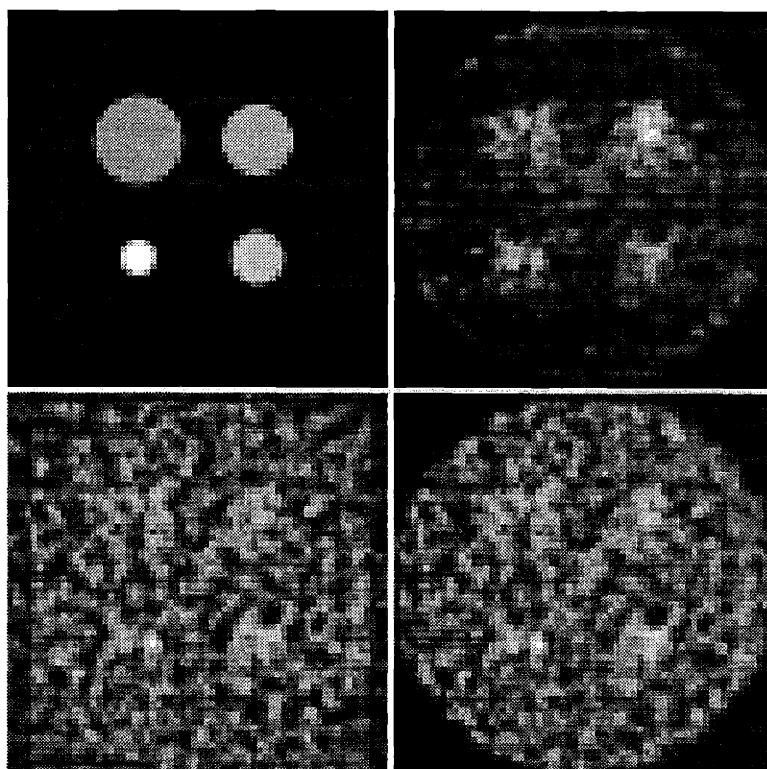


Figure 5-20: Counter-clockwise from top left: Phantom, Initial Estimate from FBP, IDR reconstruction, Final RIDR reconstruction (64 views, 64 samples per view, SNR =10, moments up to order 10 used)

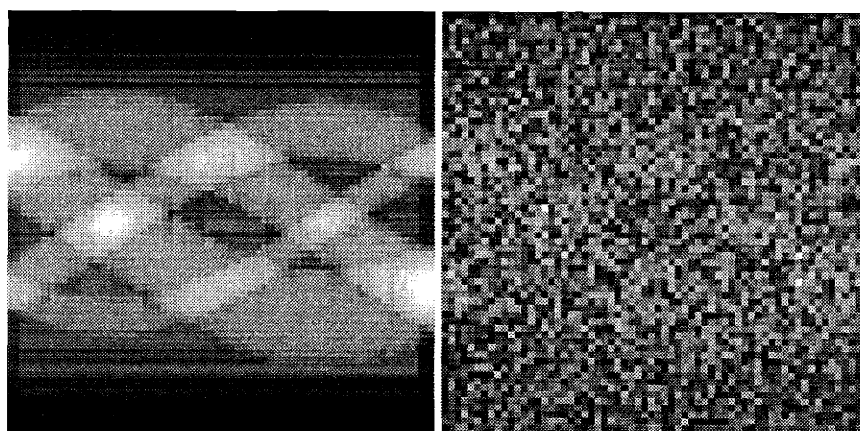


Figure 5-21: Left to right: Noiseless and noisy sinograms for phantom in Figure 5-17: 64 views, 64 samples per view SNR =10

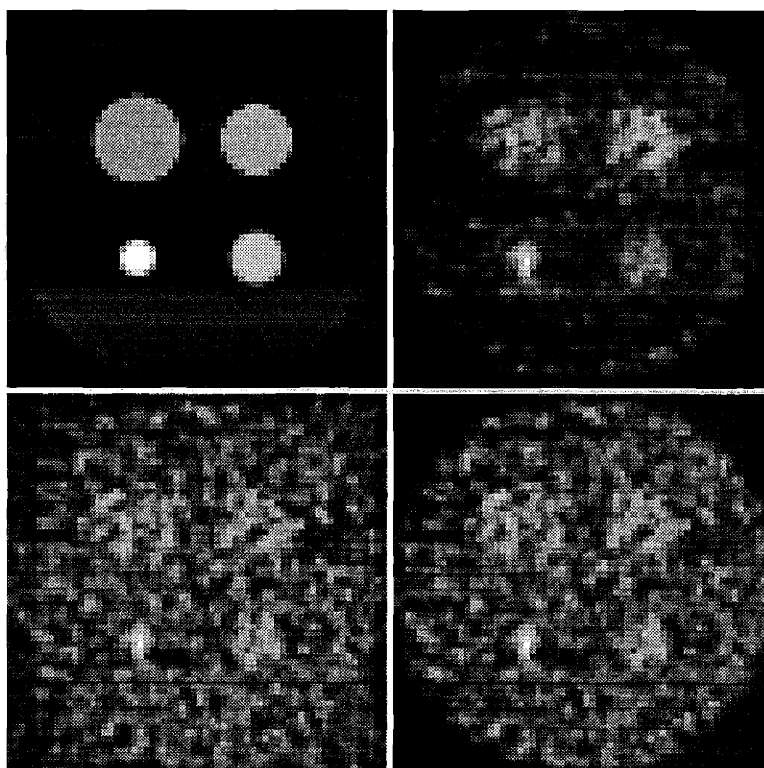


Figure 5-22: Counter-clockwise from top left: Phantom, Initial Estimate from FBP, IDR reconstruction, Final RIDR reconstruction (128 views, 64 samples per view, SNR =10, moments up to order 10 used)

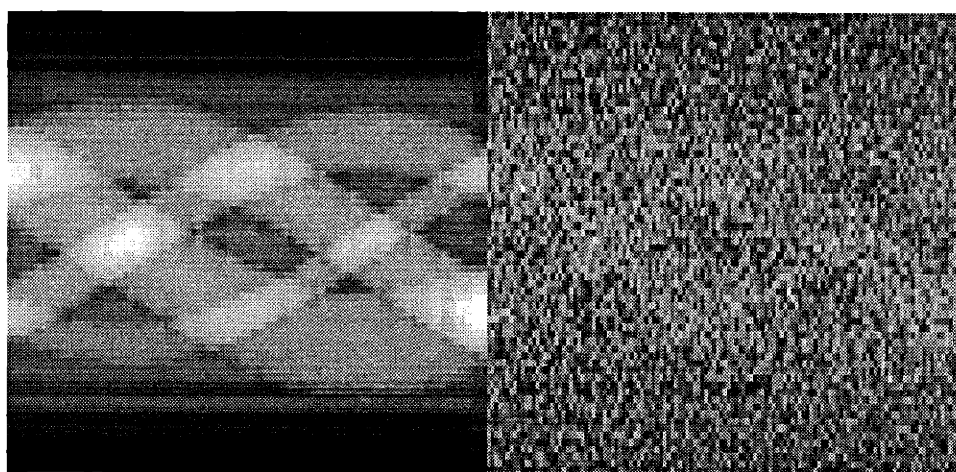


Figure 5-23: Left to right: Noiseless and noisy sinograms for phantom in Figure 5-17: 128 views, 64 samples per view, SNR =10

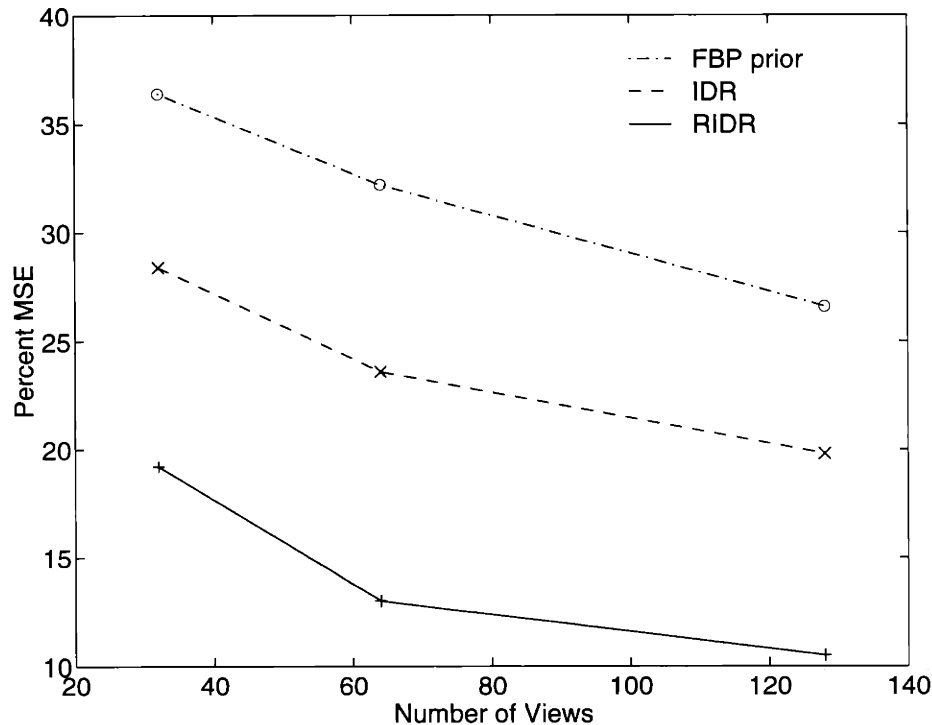


Figure 5-24: MSE versus number of views for computed FBP prior, IDR solution and RIDR solution, SNR=10, moments up to order 10 used

5.5.3 Example set III: The MIT Ellipse

In this Section, we perform reconstructions of the image of the “MIT ellipse” displayed in Figure 5-25. This image was also used by Prince in his PhD thesis [72]. This image is an 81×81 two-valued image with dark areas having the value of zero while the white areas have activity level of 10.

All reconstructions use 60 measured views at 81 samples per view. The noiseless sinogram appears in Figure 5-26. The MIT ellipse is tilted at an angle of 45 degrees with the horizontal axis and has major and minor axis lengths of 0.8 and 0.24 respectively. Hence the object is contained within the unit disk centered at the origin.

In this example we study the performance of our algorithm with respect to changes in SNR using a fixed number of moments. We used moments up to order 5 in all the reconstructions shown below and the initial estimate was based on the best FBP reconstruction in every case. Other experiments performed for the SNR values used in this subsection, and using more moments, did not show significant improvements. Hence, we show results with moments up to order 5 only.

We study 4 sets of reconstructions for which the SNR values were chosen to be

-10 , 0 , 10 , and 20 . The noisy sinograms for all cases are shown in Figure 5-27. In Figures 5-28 and 5-29 we show the corresponding FBP and IDR reconstructions, respectively. Figure 5-30 contains the final RIDR reconstructions. The MSE values of the original FBP reconstructions, along with the MSE values of the RIDR reconstructions are shown versus SNR in Figure 5-31. As this figure shows, for a fixed number of moments (up to order 5) used in the reconstruction, at low SNR's our algorithm performs significantly better than the best FBP algorithm. For high SNR cases, the improvement over the classical FBP reconstructions is, in terms of MSE, not as significant. Visually, however, our reconstructions are still superior to the FBP reconstructions. This again affirms our assertion in Example 2 that the RIDR algorithm appears to be most useful in low SNR situations while, at least visually, in high SNR scenarios it is still superior to FBP. Note that if higher order moments are used, it may be possible, depending on the geometric characteristics of the underlying image, to obtain better reconstructions over a wide range of signal to noise ratios. In the case of the MIT ellipse image, we used moments above order 5 at SNR's of -10 , 0 , 10 and 20 , and observed that the resulting reconstructions were not significantly better than those produced with moments only up to order 5. This may, in general, not be the case for a different image.

5.6 Conclusion

In this chapter we have presented variational algorithms for the reconstruction of an image from a finite number of its noisy estimated moments. In particular, we studied (I) the Divergence regularization approach (IDR), which can be interpreted as the Maximum A-Posteriori estimate of the image, and (II) the iterative Divergence regularization approach (RIDR) which enforces its solution to have moments equal to the (consistent projection of) the estimated moments. The quadratic regularization algorithm was also studied.

We showed that given a rough initial estimate of the underlying image, such as the FBP reconstruction, the IDR algorithm marginally improves the quality of the recon-

struction while RIDR produces considerably improved reconstructions at very low signal-to-noise ratios when compared to the classical filtered-backprojection reconstruction algorithm. Due to the fact that our proposed algorithms make explicit use of the covariance matrix of the estimated moments, higher order estimated moments, the estimates of which are more inaccurate, are weighed less than lower order ones. Hence our proposed algorithms essentially make use of a finite and modestly small number of moments to produce superior reconstruction. This feature, along with the overall structure of our algorithms and their observed rapid rate of convergence, demonstrate that our algorithms are also computationally efficient.

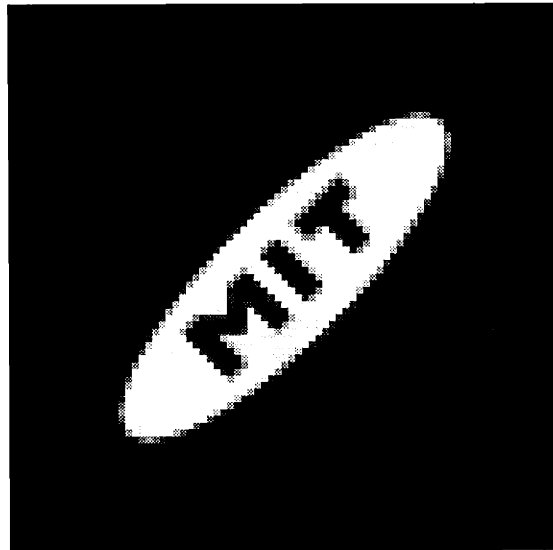


Figure 5-25: The MIT ellipse phantom

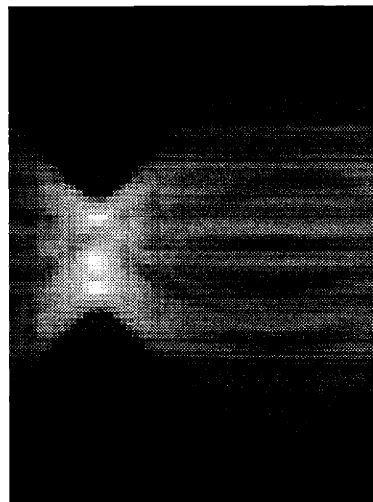


Figure 5-26: The noiseless sinogram of the MIT ellipse: 60 views, 81 samples/view

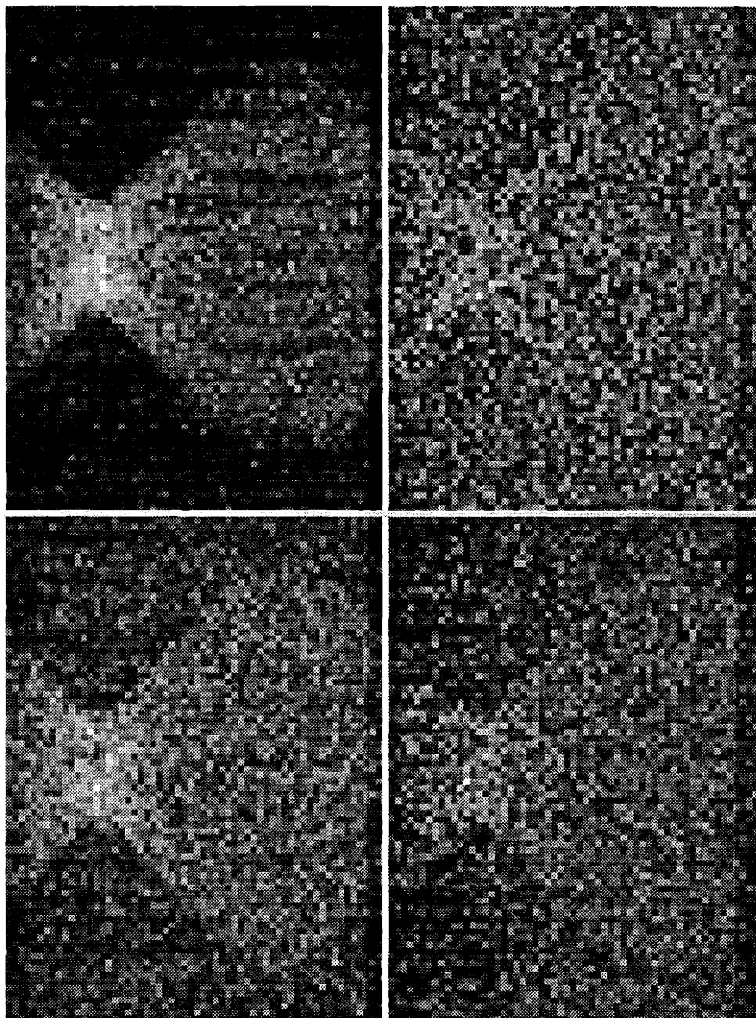


Figure 5-27: Counter-clockwise from top left:MIT ellipse sinograms at $\text{SNR} = 20, 10, 0,$ and -10 : each contains 60 views with 81 samples/view

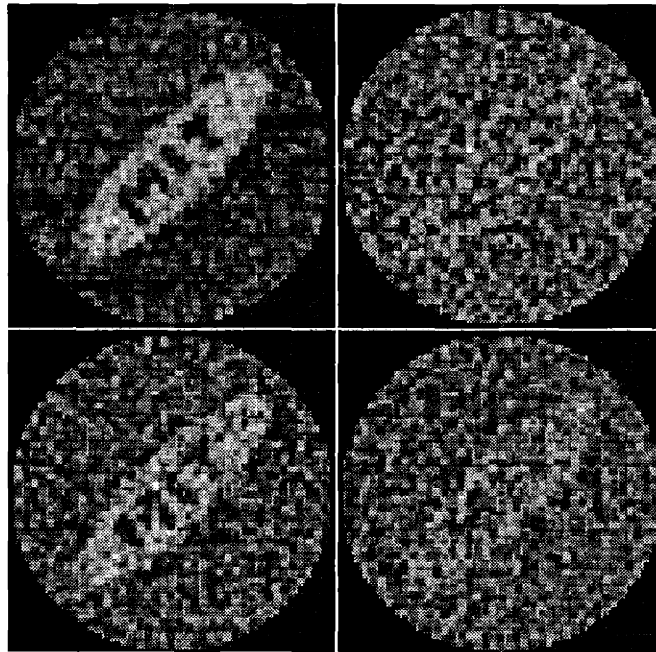


Figure 5-28: Counter-clockwise from top left:FBP reconstructions of MIT ellipse at SNR = 20, 10, 0, and -10

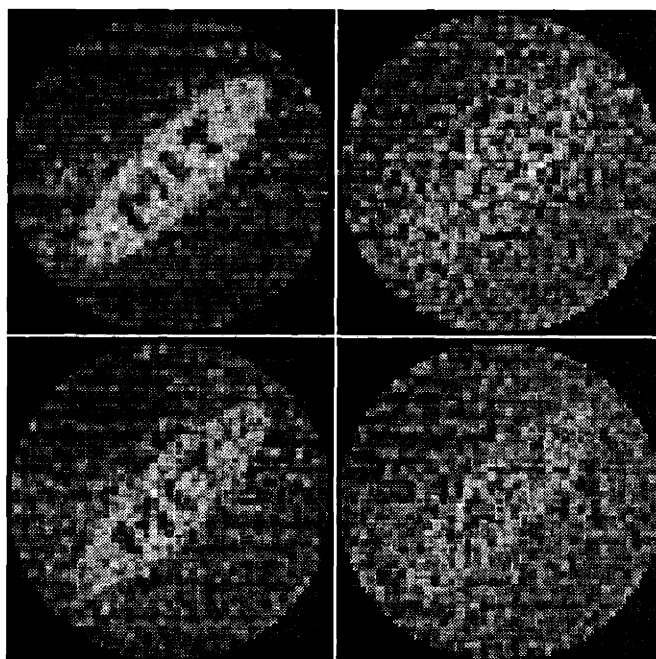


Figure 5-29: Counter-clockwise from top left:IDR reconstructions of MIT ellipse at SNR = 20, 10, 0, and -10. Moments up to order 5 used.

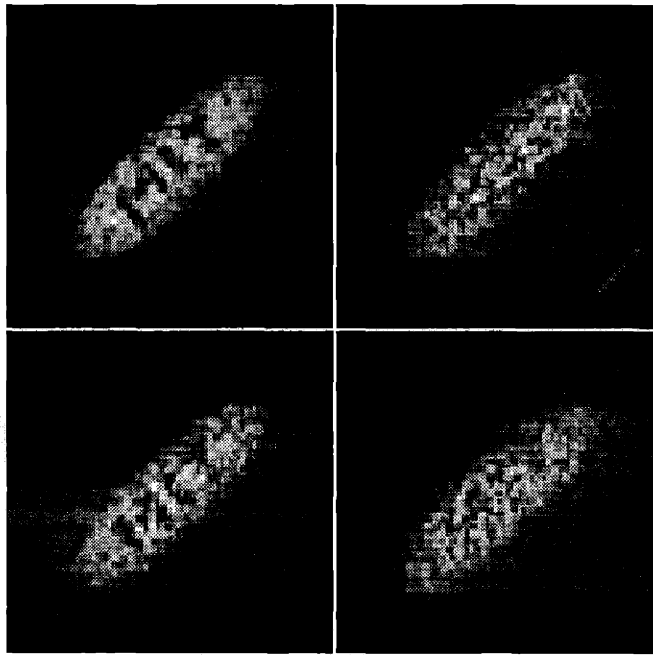


Figure 5-30: Counter-clockwise from top left: RIDR reconstructions of MIT ellipse at SNR = 20, 10, 0, and -10. Moments up to order 5 used.

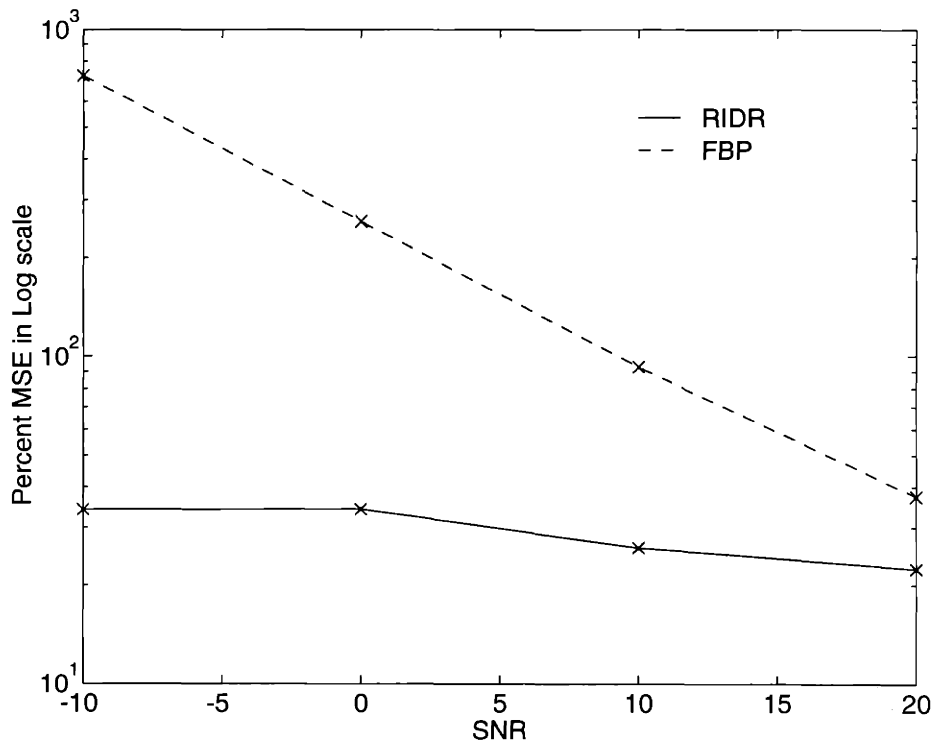


Figure 5-31: MSE versus SNR for FBP and RIDR solution with FBP-based initial estimate: Moments up to order 5 used.

Chapter 6

Reconstruction of Polygons from Moments: Connections to Array Processing

6.1 Introduction

In this chapter we present novel algorithms for the reconstruction of binary polygons from their estimated *complex* moments. We show, in fact, that this problem can be formulated as an *array processing* [81] problem, hence exposing a seemingly deep connection between the fields of tomography and array processing. This connection implies that a host of numerical algorithms such as Total Least Squares and ESPRIT [77], are now available for application to tomographic reconstruction problems. Furthermore we observe that the estimated complex moments turn out to have uncorrelated errors across moments of different order. This fact makes the statement of our reconstruction problem even more similar to the assumptions that underly a standard array processing problem.

Mathematically, our algorithms are based on the idea that the vertices of a simply-connected polygonal region in the plane are determined by a finite number of its moments. Davis [18] showed, using the Motzkin-Schoenberg formula of complex analysis [83], that a triangle in the plane is uniquely determined by its moments of up to order

3. We have generalized his result using Prony's method [41] to show that the vertices of a planar, simply-connected, n -gon are uniquely determined by its moments of up to order $2n - 3$. This implies that $2n - 2$ projections from distinct angles in $[0, \pi)$ suffice to uniquely determine the vertices of any simply-connected n -gon in the plane. This result is an improvement on theoretical results dealing with reconstructability from few projections such as those discussed in Section 2.2.2.

The organization of this chapter is as follows. In Section 6.2, we discuss the mathematical basis of reconstruction of polygonal regions from a finite number of their complex moments. In Section 6.3 we discuss the observation that the errors in estimating the complex moments of a function f from noisy projections are uncorrelated across moments of different order. In Section 6.5.1 we make direct use of Prony's method and apply ordinary and total least squares estimation algorithms to compute estimates of the vertices of polygons. In Section 6.5.2 we present a matrix pencil technique for the estimation of polygons from moments. Finally, in Section 6.6, we study the performance of the proposed algorithms and observe that the proposed algorithms do not perform well even with very little noise in the measured data. We discuss some reasons why this is so and suggest possible improvements. Due to the abundance of techniques in array processing, it is impractical, in our context, to be exhaustive in showing the applications of all the well-known algorithms of array processing to our reconstruction problem. Therefore, we have only shown some instances (namely, least squares Prony and pencil methods) of the possible applications of these algorithms. We defer further study of other algorithms and their relative performance to future research efforts.

6.2 Mathematical Background

In 1977 Davis [18] showed that any triangular region in the plane is uniquely determined by its first four complex moments. This result was derived as a corollary to a little known result in complex variable theory which he termed the *Motzkin-Schoenberg Formula*. He had worked out an alternative proof of this formula in an

earlier (1964) paper [19], where he also discussed some generalizations and applications of this result. In this section we prove a generalization of his results to show that vertices of n -sided, simply-connected polygonal regions in the plane are uniquely determined by a finite number of their complex moments.

Let T designate a triangle in the complex plane whose vertices are given by z_1 , z_2 , and z_3 . If A denotes the area of T and $h(z)$ is any analytic function in the closure of T , the **Motzkin-Schoenberg (MS) Formula** [83] states that

$$\iint_T h''(z) dx dy = 2A \det(U) / \det(V) \tag{6.1}$$

where

$$U = \begin{bmatrix} 1 & 1 & 1 \\ z_1 & z_2 & z_3 \\ h(z_1) & h(z_2) & h(z_3) \end{bmatrix} \tag{6.2}$$

$$V = \begin{bmatrix} 1 & 1 & 1 \\ z_1 & z_2 & z_3 \\ z_1^2 & z_2^2 & z_3^2 \end{bmatrix} \tag{6.3}$$

By considering triangulations of n -sided polygons, Davis [19] extended this formula to general polygons as follows.

Theorem 10 [19] *Let z_1, z_2, \dots, z_n designate the vertices of a polygon P . Then we can find constants a_1, \dots, a_n depending upon z_1, z_2, \dots, z_n , but independent of h , such that for all h analytic in the closure of P ,*

$$\iint_P h''(z) dx dy = \sum_{i=1}^n a_i h(z_i). \tag{6.4}$$

If $r \geq n$ and z_{n+1}, \dots, z_r are additional points distinct from z_1, \dots, z_n , and if there are constants b_1, \dots, b_r which depend only upon z_1, \dots, z_r such that

$$\iint_P h''(z) dx dy = \sum_{i=1}^r b_i h(z_i) \tag{6.5}$$

for all h analytic in the closure of P , then

$$b_i = a_i, \quad 1 \leq i \leq n \quad (6.6)$$

$$b_i = 0, \quad n + 1 \leq i \leq r \quad (6.7)$$

Two observations are in order about the implications of this result. First we can prove the following result for simply connected polygons ¹ using the same line of reasoning as Davis [19].

Lemma 4 *Let P be a simply-connected polygonal region. The coefficients $\{a_j\}$ in (6.4) are all nonzero if and only if P is nondegenerate.*

Proof: Using Green's theorem in the complex plane and the Cauchy-Riemann equations for analytic functions [19], the integral in (6.4) can be rewritten as

$$\iint_P h''(z) dx dy = \frac{i}{2} \int_{\partial P} h'(z) d\bar{z} \quad (6.8)$$

where $i = \sqrt{-1}$, ∂P denotes the boundary of P , and \bar{z} denotes the complex conjugate of z . The assumption that P is simply connected implies that the boundary of P consists of one piece. Hence, we can assert that this boundary is composed of n straight lines which we call s_1, s_2, \dots, s_n ; where s_j connects the vertices z_j and z_{j+1} . For convenience, we assume that the vertices z_j of P are arranged in the counter-clockwise direction in the order of increasing index. Hence, splitting the right-hand side of (6.8) into a sum of terms over the sides and using the expression for the equation of a line in the complex plane [19], we can write

$$\iint_P h''(z) dx dy = \frac{i}{2} \sum_{j=1}^n (\alpha_{j-1} - \alpha_j) h(z_j) \quad (6.9)$$

where

$$\alpha_j = \frac{\bar{z}_j - \bar{z}_{j+1}}{z_j - z_{j+1}}. \quad (6.10)$$

¹A simply connected domain is one whose boundary consists of one piece.

With some algebraic manipulation, it is not difficult to show [19] that

$$\frac{i}{2}(\alpha_{j-1} - \alpha_j) = \frac{2A_j}{(z_j - z_{j+1})(z_j - z_{j-1})} \quad (6.11)$$

where A_j is the signed, or directed area of the triangle formed by the three vertices z_{j-1} , z_j , and z_{j+1} given by

$$A_j = \frac{i}{4} \det \begin{bmatrix} z_{j-1} & \bar{z}_{j-1} & 1 \\ z_j & \bar{z}_j & 1 \\ z_{j+1} & \bar{z}_{j+1} & 1 \end{bmatrix} \quad (6.12)$$

Now comparing (6.9) to (6.4), we have that

$$a_j = \frac{i}{2} \left(\frac{\bar{z}_{j-1} - \bar{z}_j}{z_{j-1} - z_j} - \frac{\bar{z}_j - \bar{z}_{j+1}}{z_j - z_{j+1}} \right) = \frac{2A_j}{(z_j - z_{j+1})(z_j - z_{j-1})}, \quad j = 1, \dots, n \quad (6.13)$$

Hence, no a_j is zero unless the corresponding A_j is zero. This can occur if and only if P is degenerate. i.e. for some j , the triangle formed by z_{j-1} , z_j , and z_{j+1} is degenerate. \square

Note that the Equation 6.13 is an expression that depends explicitly on the vertices. This expression also implicitly depends on the way these vertices are connected by virtue of the fact that different ways of connecting them changes the relative order of the vertices that enter into the right-hand side of (6.13). For instance, the vertex z_j may be connected to z_k in one arrangement, while it may be connected to $z_{k'}$ in another, with $k \neq k'$.

A second observation is that the formula (6.4) is a *minimal* representation of the integral of h'' over P in terms of discrete values of h . To put it differently, the left-hand side of (6.4) depends *only* on the values of h at the vertices of P ; what values h takes at other points in the complex plane are completely irrelevant in this regard.

Recall, from Chapter 4, the definition of complex moments of a function f :

$$c_{kl} = \iint_{\mathcal{O}} f(x, y) z^k \bar{z}^l dx dy \quad (6.14)$$

where $z = x + iy$, $\bar{z} = x - iy$, and $i = \sqrt{-1}$. Let us define the *simple* complex moments (s-complex moments) as ²

$$c_k = c_{k,0} = \iint_{\mathcal{O}} f(x, y) z^k dx dy \quad (6.15)$$

The relationship between these moments and the geometric moments of f over \mathcal{O} is simply established as follows.

$$c_k = \iint_{\mathcal{O}} f(x, y) (x + iy)^k dx dy \quad (6.16)$$

$$= \sum_{j=0}^k \binom{k}{j} i^j \iint_{\mathcal{O}} x^{k-j} y^j f(x, y) dx dy \quad (6.17)$$

$$= \sum_{j=0}^k \binom{k}{j} i^j \mu_{k-j,j} \quad (6.18)$$

Written more simply in vector notation we have

$$c_k = u_k^T \mu^{(k)} \quad (6.19)$$

where

$$u_k = \left[\binom{k}{0} i^0, \dots, \binom{k}{k} i^k \right]^T \quad (6.20)$$

Now consider Theorem 10 and let (I) $h(z) = z^k$ and (II) $f(x, y)$ be the indicator function over a simply-connected polygonal region P of the plane. Then, this theorem states that for any nondegenerate, simply-connected n -gon, simply-connected region P in the plane, there exist nonzero complex numbers a_j such that

$$\iint_P (z^k)'' dx dy = \sum_{j=1}^n a_j z_j^k. \quad (6.21)$$

²These moments are also referred to as *harmonic* moments in the mathematics community

The left hand side of this identity can be written as

$$\iint_P (z^k)'' dx dy = k(k-1) \iint_P z^{k-2} dx dy = k(k-1)c_{k-2}. \quad (6.22)$$

Defining the numbers $\tau_k = k(k-1)c_{k-2}$, which we term *weighted complex moments* (w-complex moments), with $\tau_0 = \tau_1 = 0$, we have

$$\tau_k = \sum_{j=1}^n a_j z_j^k. \quad (6.23)$$

The w-complex moments are linear combinations of the geometric moments of the underlying polygon P which can, in turn, be directly computed from the projections of P . Hence, (6.23) is, for every k , a direct measurement equation for the w-complex moments in terms of the vertices of P . We will next show that by considering a sufficient number of such measurements for different values of the parameter k , the vertices of P may be uniquely recovered.

6.2.1 Vertices from Moments via Prony's Method

Assume that the n -gon P is simply-connected and nondegenerate, and let us consider the equation (6.23) for $k = 0, 1, \dots, 2n-1$. Written in vector form we have

$$\begin{bmatrix} \tau_0 \\ \tau_1 \\ \vdots \\ \tau_{2n-1} \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ z_1 & z_2 & \cdots & z_n \\ \vdots & \vdots & \ddots & \vdots \\ z_1^{2n-1} & z_2^{2n-1} & \cdots & z_n^{2n-1} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \quad (6.24)$$

$$\mathcal{T}_{2n} = \mathcal{V}_{2n} \mathbf{a}_n \quad (6.25)$$

where the obvious associations have been made in the last identity. We will use Prony's method [41] here to show that the vertices $\{z_i\}$ can be computed from the w-complex moment vector \mathcal{T}_{2n} given by Equation (6.25). Davis [18] showed this result for $n = 3$ (the triangular case).

Define the polynomial $P(z)$ as

$$P(z) = \prod_{i=1}^n (z - z_i) = z^n + \sum_{i=1}^n p_i z^{n-i}, \quad (6.26)$$

and consider its associated coefficient vector $p^{(n)} = [p_n, p_{n-1}, \dots, p_1]^T$. We wish to show that the coefficients of $P(z)$ can be uniquely determined from \mathcal{T}_{2n} . To this end, form the $2n \times 2n$ matrix K_{2n} from $p^{(n)}$ as follows

$$K_{2n} = \begin{bmatrix} p_n & \cdots & p_1 & 1 & & \circ \\ & & & \ddots & \ddots & \\ \circ & & p_n & \cdots & p_1 & 1 \end{bmatrix}. \quad (6.27)$$

From the definition of the polynomial $P(z)$ it follows that the matrix K_{2n} , when multiplied on the left of (6.25) annihilates it ³. That is

$$K_{2n} \mathcal{T}_{2n} = K_{2n} \mathcal{V}_{2n} \mathbf{a}_n = \underline{\mathbf{0}} \quad (6.28)$$

The identity $K_{2n} \mathcal{T}_{2n} = \underline{\mathbf{0}}$ can easily be rewritten as

$$\begin{bmatrix} \tau_0 & \tau_1 & \cdots & \tau_{n-1} \\ \tau_1 & \tau_2 & \cdots & \tau_n \\ \vdots & \vdots & \ddots & \vdots \\ \tau_{n-1} & \tau_n & \cdots & \tau_{2n-2} \end{bmatrix} p^{(n)} = - \begin{bmatrix} \tau_n \\ \tau_{n+1} \\ \vdots \\ \tau_{2n-1} \end{bmatrix} \quad (6.29)$$

$$H_n p^{(n)} = -h_n \quad (6.30)$$

To show that we can uniquely recover $p^{(n)}$ from this last identity, we must show that H_n is invertible.

Lemma 5 *The $n \times n$ matrix H_n is invertible if and only if the corresponding simply-connected polygon P is a nondegenerate n -gon.*

³This key observation is due to Prony and forms the basis of Prony's method [41]

Proof: The proof is obtained by noticing that H_n can be decomposed as

$$H_n = \mathcal{V}_n \text{diag}(\mathbf{a}_n) \mathcal{V}_n^T \quad (6.31)$$

where \mathcal{V}_n is the *Vandermonde* matrix of the vertices $\{z_i\}$ defined as follows

$$\mathcal{V}_n = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ z_1 & z_2 & \cdots & z_n \\ \vdots & \vdots & \ddots & \vdots \\ z_1^{n-1} & z_2^{n-1} & \cdots & z_n^{n-1} \end{bmatrix} \quad (6.32)$$

The matrix \mathcal{V}_n has determinant

$$\det(\mathcal{V}_n) = \prod_{i>j} (z_i - z_j) \quad (6.33)$$

which vanishes if and only if P is degenerate. Furthermore, as a consequence of Lemma 4, the elements of the vector \mathbf{a}_n are all nonzero unless P is degenerate. Hence this Lemma is established. \square

As a consequence of this lemma, the coefficients of $P(z)$ can then be uniquely determined through

$$p^{(n)} = -H_n^{-1} h_n. \quad (6.34)$$

Upon solving the polynomial equation $P(z) = 0$, the vertices of P may be recovered. In summary, we have shown the following result.

Result 12 *Let P denote a nondegenerate, simply-connected, n -sided polygonal region in the plane. The vertices of P are uniquely determined by its w -complex moments of up to order $2n - 1$.*

This result is the generalization to n -sided, simply-connected polygons of the result Davis proved in [18] where he only considered the case $n = 3$. i.e. triangles.

Several useful corollaries follow from Result 12. Recall that the w -complex mo-

ments τ_k are related to the s-complex moments c_k as

$$\tau_k = k(k-1)c_{k-2} \tag{6.35}$$

Hence, we have:

Corollary 4 *The vertices of a nondegenerate, simply-connected n -gon P in the plane are uniquely determined by its s-complex moments of up to order $2n - 3$.*

Also note that since each s-complex moment c_k depends only on the geometric moment set of the same order, $\mu^{(k)}$, it follows that

Corollary 5 *The vertices of a nondegenerate, simply-connected n -gon P in the plane are determined by its geometric moments of up to order $2n - 3$. (That is, the vector \mathcal{M}_{2n-3})*

In Chapter 4, Result 5, we showed that “Given p (line integral) projections of $f(x, y)$ at p different angles θ_i in $[0, \pi)$, one can uniquely determine the first p geometric moment vectors $\mu^{(j)}$, $0 \leq j < p$ of $f(x, y)$.” When $f(x, y)$ is taken to be the indicator function on the polygonal region P , this implies the following corollary.

Corollary 6 *Exactly $2n - 2$ projections are sufficient to uniquely determine the vertices of a plane simply-connected polygonal region P .*

6.2.2 Remarks

Result 12 and its corollaries imply that the vertices of P can be extracted from a finite number of moments. This theorem, however, does not tell us what the interior of P looks like. To put it differently, according to this theorem, from the set of moments we can decipher the locations of the vertices of P ; yet this theorem does not tell us how to “connect the dots” to form the interior of P . If the polygonal region P is assumed to be convex, this difficulty is not an issue since, in such case, only two possible ways exist of connecting the vertices. Namely, a clockwise and a counter-clockwise orientation. Either of these choices, however, yield exactly the same geometric object

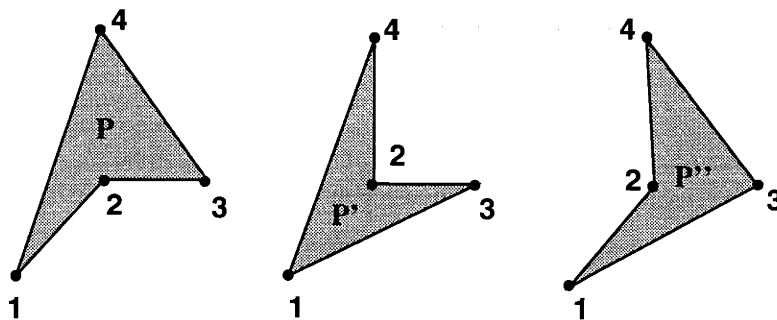


Figure 6-1: Three distinct regions corresponding to the same vertices

so that any convex n -gon is, in fact, uniquely determined by geometric moments of up to order $2n - 3$.

For nonconvex P the situation is more complicated since, in general, there is more than one way to connect the vertices, leading to multiple solutions as illustrated in Figure 6-1. However, there are only a finite number of such possibilities, with each such possibility yielding a different region. It is desirable to know whether only one of these possible regions will have the prescribed moments $\tau_0, \tau_1, \dots, \tau_{2n-1}$. The answer to this question is, in general, no. In fact, even if all the w -complex moments τ_k of a simply-connected domain are given, the interior of the domain is not necessarily uniquely specified. This question was first posed in 1975 by H. Shapiro in [4] in the following form:

“Let D_1 and D_2 be Jordan domains ⁴ such that

$$\iint_{D_1} z^k dx dy = \iint_{D_2} z^k dx dy \quad k = 0, 1, 2, \dots \quad (6.36)$$

Must we have $D_1 = D_2$?”

In 1978, a counter-example was provided by M. Sakai in [79] where he constructed simple domains bound by a finite number of piecewise circular arcs. Polygonal counter-examples were constructed by A. M. Gabrielov, V. N. Strakhov, and M. A. Brodsky and were published in the latter two authors paper [93]. These authors arrived at

⁴Jordan domains are simply-connected compact sets in the plane.

this question from considering the more general problem of uniqueness of the shape and density of plane gravitating bodies as determined from their exterior logarithmic potential. A good survey of this problem from the point of view of inverse potential theory and fluid dynamics can be found in [104] ⁵.

In certain special cases, however, the w-complex moments do uniquely specify the interior of the underlying polygon. To motivate this, note that by referring to Equation (6.13), we can see that it is reasonable to think that, in some cases, each way of connecting the vertices should result in a different set of a_j 's. This in turn, through (6.23), would imply that each way of connecting the vertices would yield a different set of w-complex moments. Another way of saying this is that, in some cases, only one of the finite number of possible n -gons formed by connecting the computed vertices will have the prescribed moments $\tau_0, \tau_1, \dots, \tau_{2n-1}$. If we can show this, then in order to reconstruct the interior of the underlying polygon, we have to search over all possibilities to find the unique one that has these prescribed moments. To make the above arguments more concrete, let us carry out a simple example.

Consider the vertices z_j for $j = 1, \dots, 4$ shown in Figure 6-1. A total of three polygons $P, P',$ and P'' can be obtained by connecting the four vertices in different ways. For convenience, let us fix a counter-clockwise orientation and let us assume that the vertices are situated in the complex plane as follows:

$$z_1 = -1 - i, \tag{6.37}$$

$$z_2 = 0, \tag{6.38}$$

$$z_3 = 1, \tag{6.39}$$

$$z_4 = i, \tag{6.40}$$

For the polygons $P, P',$ and P'' , we write:

$$\tau_k = \sum_{j=1}^4 a_j z_j^k, \tag{6.41}$$

⁵I would like to thank Chris Bishop of SUNY Stony Brook Math dept. and Prof. Pavel Etingof of Yale Math dept. for pointing out these references

$$\tau'_k = \sum_{j=1}^4 a'_j z_j^k, \quad (6.42)$$

$$\tau''_k = \sum_{j=1}^4 a''_j z_j^k, \quad (6.43)$$

and proceed to compute the coefficients a_j , a'_j , and a''_j through Equation (6.13). We obtain:

$$\begin{aligned} a_1 &= -0.1 - 0.3i, & a'_1 &= -0.6i, & a''_1 &= 0.1 - 0.3i \\ a_2 &= 0.5 - 0.5i, & a'_2 &= i, & a''_2 &= -0.5 - 0.5i \\ a_3 &= 0.5 + 0.5i, & a'_3 &= 0.4 - 0.2i, & a''_3 &= 0.9 + 0.3i \\ a_4 &= -0.9 + 0.3i, & a'_4 &= -0.4 - 0.2i, & a''_4 &= -0.5 + 0.5i \end{aligned} \quad (6.44)$$

As we can see, these coefficients are different. This shows that only one of the three possible polygons will have the prescribed set of moments τ_k for $k = 0, \dots, 7$. Note that the three possibilities in this case each share a side with the others. In general, we can show that two simply-connected n -gons P and P' , with identical vertices, which have at least one side in common must have different w-complex moments τ_k for some k . Formally, we have

Result 13 *Consider n distinct points z_1, z_2, \dots, z_n in the complex plane. Let P and P' be simply-connected, nondegenerate, n -gons generated by connecting these vertices in two distinct ways. If P and P' have at least one side in common, then for some $1 \leq j \leq n$,*

$$a_j(P) \neq a_j(P'), \quad (6.45)$$

where $a_j(P)$ and $a_j(P')$ are respectively given by

$$\iint_P h''(z) dx dy = \sum_{j=1}^n a_j(P) h(z_j) \quad (6.46)$$

$$\iint_{P'} h''(z) dx dy = \sum_{j=1}^n a_j(P') h(z_j) \quad (6.47)$$

$$(6.48)$$

Proof: We prove this result by contradiction. Assume that P and P' have at least one side in common. Without loss of generality, let us say this is the side give by

connecting the vertices z_j and z_{j+1} of P . Now if $a_j(P) = a_j(P')$, it follows from (6.13) that

$$\frac{i}{2} \left(\frac{\bar{z}_{j-1} - \bar{z}_j}{z_{j-1} - z_j} - \frac{\bar{z}_j - \bar{z}_{j+1}}{z_j - z_{j+1}} \right) = \frac{i}{2} \left(\frac{\bar{z}'_{j-1} - \bar{z}_j}{z'_{j-1} - z_j} - \frac{\bar{z}_j - \bar{z}_{j+1}}{z_j - z_{j+1}} \right) \quad (6.49)$$

where z'_{j-1} is the $j - 1^{\text{th}}$ vertex of P' . Simplifying (6.49) yields

$$\frac{\bar{z}_{j-1} - \bar{z}_j}{z_{j-1} - z_j} = \frac{\bar{z}'_{j-1} - \bar{z}_j}{z'_{j-1} - z_j} \quad (6.50)$$

It is easy to check that this last expression (6.50) implies that the vertices z_j , z_{j-1} , and z'_{j-1} must be collinear. This is a contradiction to the assumption that P and P' are nondegenerate. \square

6.3 Statistics of Estimated Complex Moments

The explicit dependence of the s-complex moments of order k on the geometric moments of the same order implies that we can compute Maximum Likelihood estimates of these moments from ML estimates of the geometric moments. Recall from (6.19) that this explicit relationship is given by

$$c_k = u_k^T \mu^{(k)} \quad (6.51)$$

where u_k is defined in (6.20). Note that computing the s-complex moments from the corresponding geometric moments involves the loss of some information. This is because each value of c_k is a complex *linear combination* of the geometric moments of the same order but the values of $\mu^{(k)}$ can not be inferred from the value of c_k . Hence, each c_k contains less information than the corresponding geometric moment vector $\mu^{(k)}$.

For $k = 0, \dots, N$ we can collectively write

$$\begin{bmatrix} c_0 \\ \vdots \\ c_N \end{bmatrix} = \begin{bmatrix} u_0^T & \cdots & \circ \\ \vdots & \ddots & \vdots \\ \circ & \cdots & u_N^T \end{bmatrix} \begin{bmatrix} \mu^{(0)} \\ \vdots \\ \mu^{(N)} \end{bmatrix} \quad (6.52)$$

$$C_N = U_N \mathcal{M}_N \quad (6.53)$$

with the obvious associations made in the last expression. Note that U_N is an $(N+1)$ by $(N+1) \times (N+2)/2$ matrix with complex elements.

Now if $\widehat{\mathcal{M}}_N$ denotes the ML estimate of the geometric moment vector \mathcal{M}_N , with covariance R_N , then the ML estimate of C_N is given by

$$\widehat{C}_N = U_N \widehat{\mathcal{M}}_N \quad (6.54)$$

The corresponding covariance matrix for \widehat{C}_N , which we name S_N , is given by

$$S_N = U_N R_N U_N^H \quad (6.55)$$

where the superscript H denotes the Hermitian transpose (or conjugate transpose) operation. To study the structure of S_N carefully, let us write U_N in terms of its real and imaginary parts as

$$U_N = U_{\Re N} + iU_{\Im N} \quad (6.56)$$

Replacing this in (6.55) yields

$$S_N = (U_{\Re N} R_N U_{\Re N}^T + U_{\Im N} R_N U_{\Im N}^T) + i(U_{\Im N} R_N U_{\Re N}^T - U_{\Re N} R_N U_{\Im N}^T) \quad (6.57)$$

This expression is valid when noisy projections from a finite number of views are given. If a continuum of views are given (in $[0, \pi)$), we write

$$S_N^* = (U_{\Re N} R_N^* U_{\Re N}^T + U_{\Im N} R_N^* U_{\Im N}^T) + i(U_{\Im N} R_N^* U_{\Re N}^T - U_{\Re N} R_N^* U_{\Im N}^T) \quad (6.58)$$

From numerical experiments, we have observed that S_N^* is real and diagonal for all N . In fact, we have observed that each term in the real part of S_N^* is diagonal and each term in the imaginary part of it is zero. We also observed that $E((\widehat{C}_N - C_N)(\widehat{C}_N - C_N)^T)$ is real and diagonal. Two important conclusions can be drawn from these facts:

Observation 1 *When a continuum of white-noise-corrupted projection data $Y(t, \theta)$ is given in both variables t and θ , the estimated s- and w-complex moments $\{\widehat{c}_k\}$, $\{\widehat{w}_k\}$ have uncorrelated estimation errors across moments of different order. Furthermore, the errors in the real and imaginary parts of these estimates are also uncorrelated.*

It is interesting that the covariance matrices for the estimated geometric and orthogonal moments failed to be diagonal (or even block diagonal) as we saw in Chapter 4; while here, the estimated s-complex moments (and consequently w-complex moments) have uncorrelated estimation errors. The implication is that the process of estimating the s-complex moments is in essence an orthogonal projection of the data $Y(t, \theta)$ onto the space of complex moments of the underlying object. This fact can be used to advantage in applications such as pattern recognition, classification, and compression of tomographic data. This implies that these moments merit attention even without regard to whether they are useful in the context of a complete reconstruction algorithm. Note that Observation 1 is not given with respect to any particular underlying object. In fact, the estimation error covariance R_N (or R_N^*) is completely independent of the underlying object. Another point worth making is that from extensive experimentation, we found that the conclusions presented in Observation 1 hold very nearly true even when only a modest number of views are given. (i.e. for S_N instead of S_N^*) For instance, given noisy data $Y(t, \theta) = g(t, \theta) + n(t, \theta)$ with $n \sim \mathcal{N}(0, \sigma^2)$,

the value of S_N^* for $N = 5$ is given by

$$S_5^* = \sigma^2 \begin{bmatrix} 0.6303 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.8405 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.8966 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.9223 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.9371 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.9466 \end{bmatrix} \quad (6.59)$$

Where the elements denoted 0 have values on the order of 10^{-16} . With only 7 equally spaced views in $(0, \pi]$, the corresponding covariance matrix is

$$S_5 = \sigma^2 \begin{bmatrix} 1.2412 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1.6812 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1.8157 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1.8864 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1.9323 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1.9655 \end{bmatrix}. \quad (6.60)$$

where the elements denoted 0 are on the order of 10^{-14} . Using a finite number of *non-equally-spaced* views tends to violate the assertions of Observation 1.

6.4 Connections to Array Processing

Array processing has been a very active field of research in the past 2 decades motivated by applications in sonar, radar, oceanography, seismology, and speech processing, to name a few. The data to be analyzed in a standard array processing application [81, 77] consists of a sum of damped complex exponentials in additive white noise. This formulation corresponds to the problem of localizing several radiating sources by observation of their signals at spatially separated sensors. More formally, the general problem is that of estimating the unknowns b_i and z_i from the

measured signals y_k given as follows

$$y_k = \sum_{j=1}^n b_j z_j^k + v_k, \quad k = 0, \dots, N-1 \quad (6.61)$$

Here, z_j denote the unknown sources, b_j denote the unknown complex amplitudes, and v_k denote (complex) white noise. In most standard array processing problems, the sources z_j are typically complex sinusoids of the form $\exp(-i\phi_j)$, but general formulations where z_j is not restricted to this form have also been studied [42, 81].

Note that by letting $\hat{\tau}_k = y_k$, and $b_j = a_j$ in (6.61), we essentially obtain a noisy version of the moment measurement equation (6.23) as follows.

$$\hat{\tau}_k = \sum_{j=1}^n a_j z_j^k + w_k. \quad (6.62)$$

The general formulation of the array processing problem is therefore the same as the formulation of the reconstruction problem of binary polygonal objects from noisy measurements of their w -complex moments. There is, however, a difference between the standard assumptions made in the array processing context regarding the noise terms v_k , and the covariance structure of the estimated w -complex moments. As we discussed in Section 6.3, when a sufficient number of views are given, the covariance matrix for the estimated s -complex moments (and hence w -complex moments) is diagonal. This is in agreement with the standard array processing assumption of white noise. This diagonal matrix however, is *not* a scalar multiple of the identity as is assumed in standard array processing problems. In fact, we know that the variance of the estimation error, w_k , grows without bound with k . Hence, the noise terms, w_k , are uncorrelated but not identically distributed.

6.5 Reconstruction Algorithms

In this section we discuss two instances of direct application of array processing algorithms to the polygon reconstruction problem from moments. An exhaustive study of all available algorithms and their relative performance is beyond the scope

of this thesis and therefore, we present only two such approaches to illustrate the main concepts. The first set of algorithms is directly based on a generalization of Prony's method. In this context, we discuss the ordinary least squares Prony (OLSP) and total least squares prony (TLSP) techniques. The second set of algorithm, known as matrix pencil techniques are variants of the well known ESPRIT (Estimation of Signal Parameters via Rotation Invariance Techniques) algorithm [77].

6.5.1 Least Squares Prony Technique

We are to estimate the parameters a_j and the vertices z_j corresponding to an n -sided polygonal region from noisy estimates of the first N w-complex moments of P (which are, in turn, estimated from noisy projections of P). For $k = 0, \dots, N - 1$, we have

$$\hat{\tau}_k = \sum_{j=1}^n a_j z_j^k + w_k \quad (6.63)$$

where, in accordance with Observation 1, we assume that w_k are (complex) Gaussian measurement errors which are uncorrelated across different k , and that the real and imaginary parts of w_k are also uncorrelated. Collecting the measurements in (6.63) into vector form we have

$$\begin{bmatrix} \tau_0 \\ \hat{\tau}_1 \\ \vdots \\ \hat{\tau}_{N-1} \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ z_1 & z_2 & \cdots & z_n \\ \vdots & \vdots & \ddots & \vdots \\ z_1^{N-1} & z_2^{N-1} & \cdots & z_n^{N-1} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} + \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_{N-1} \end{bmatrix} \quad (6.64)$$

$$\hat{\mathcal{T}}_N = \mathcal{V}_N \mathbf{a}_n + W_N \quad (6.65)$$

Applying the $N \times N$ matrix K_N to both sides of (6.65) yields

$$K_N \hat{\mathcal{T}}_N = K_N W_N. \quad (6.66)$$

which can in turn be rewritten as

$$\begin{bmatrix} \hat{\tau}_0 & \hat{\tau}_1 & \cdots & \hat{\tau}_{n-1} \\ \hat{\tau}_1 & \hat{\tau}_2 & \cdots & \hat{\tau}_n \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\tau}_{N-n-1} & \hat{\tau}_{N-n} & \cdots & \hat{\tau}_{N-2} \end{bmatrix} p^{(n)} = - \begin{bmatrix} \hat{\tau}_n \\ \hat{\tau}_{n+1} \\ \vdots \\ \hat{\tau}_{N-1} \end{bmatrix} \quad (6.67)$$

$$\widehat{H}_N p^{(n)} = -\hat{h}_N \quad (6.68)$$

The equation (6.68) forms the basis of the Least Squares Prony technique. From this equation, the parameter vector $\hat{p}^{(n)}$ is estimated and subsequently, estimates of z_i are produced by solving the polynomial equation $\hat{P}(z) = 0$, whose coefficients are the elements of $\hat{p}^{(n)}$.

The above procedure is an explicit algorithm for computing the coefficients of the polynomial $\hat{P}(z)$, the roots of which are estimates of the vertices of the underlying polygon P . This procedure works perfectly when there is no noise in the measured moments (i.e. $W_N = 0$). When noise, or measurement imprecision is introduced, however, this method performs poorly; mainly due to the sensitivity of the roots of $\hat{P}(z)$ to perturbations of its coefficients. It seems natural to use more moments so that we may achieve more stable and accurate estimates of the coefficients of $\hat{P}(z)$. In section 6.2 we showed that at least $2n - 2$ moments are necessary to uniquely recover the vertices of P (that is, $N = 2n - 1$ in (6.68)). In what follows, we shall always assume that we use $N \geq 2n - 1$. By using $N > 2n - 1$, we expect that some “averaging” effect will reduce the error in our estimates of the vertices of P .

OLSP and WLSP

The Ordinary Least Squares Prony (OLSP) method consists of computing the least squares estimate of $p(n)$ from equation (6.68) by computing the generalized inverse of the matrix \widehat{H}_N as follows.

$$\hat{p}_{ols}^{(n)} = - \left(\widehat{H}_N^H \widehat{H}_N \right)^{-1} \widehat{H}_N^H \hat{h}_N \quad (6.69)$$

With these estimated coefficients, the polynomial equation $\widehat{P}_{ols}(z) = 0$ is formed and solved to get OLSP estimates \widehat{z}_i of the vertices. Having computed these estimates, we can form the matrix $\widehat{\mathcal{V}}_N$ as defined in (6.25), and subsequently estimate the unknown vector \mathbf{a}_n as

$$\widehat{\mathbf{a}}_n = \left(\widehat{\mathcal{V}}_N^H \widehat{\mathcal{V}}_N \right)^{-1} \widehat{\mathcal{V}}_N^H \widehat{\mathcal{T}}_N \quad (6.70)$$

The estimated parameters \widehat{z}_j and \widehat{a}_j should now uniquely determine an n -sided polygonal region.

The Weighted Least-Squares Prony (WLSP) solution can also be considered. In this formulation, the inverse of the covariance matrix for \widehat{h}_N is used as a weighting factor. The resulting solution has the form

$$\widehat{p}_{wls}^{(n)} = - \left(\widehat{H}_N^H W \widehat{H}_N \right)^{-1} \widehat{H}_N^H W \widehat{h}_N \quad (6.71)$$

where W denotes the inverse of the covariance matrix for \widehat{h}_N [81].

TLSP

The equation (6.68) is an overdetermined linear parameter estimation problem of the form $Ax = b$ for the unknown $x = p^{(n)}$. The OLS procedure for estimating the desired parameters is appropriate when only the vector b is noisy. In fact, the OLS estimate coincides with the Maximum Likelihood estimate if the noise is taken to be Gaussian and white. In more general instances such as that of equation (6.68), both matrices A and b are corrupted by noise. For these cases, a more general fitting scheme called the Total Least Squares (TLS) is devised [32, 43, 81]. When the matrix A is noisy, the ordinary least squares solution $\widehat{x}_{ols} = (A^H A)^{-1} A^H b$ is no longer statistically optimal and is typically biased with increased error covariance. The TLS solution was formally introduced by Golub and Van Loan [32], although it had been used extensively before in the case of single parameter estimation. The TLS solution can, in essence, be interpreted statistically as a regularized version of the OLS solution. In particular, we can write

$$\widehat{x}_{tls} = (A^H A - \sigma_{min}^2 I)^{-1} A^H b, \quad (6.72)$$

where σ_{min} denotes the minimum singular value of the concatenated matrix $[A, b]$. An alternate way of arriving at the solution is by writing out the singular value decomposition of the matrix $[A, b]$. Let

$$[A, b] = USV^H \quad (6.73)$$

$$U = [U_s, u_{min}] \quad (6.74)$$

$$P_s = U_s U_s^H \quad (6.75)$$

where u_{min} denotes the singular vector that corresponds to the smallest singular value of $[A, b]$. The TLS solution can then be written as

$$\hat{x}_{tls} = (A^H P_s A)^{-1} A^H P_s b, \quad (6.76)$$

which is equivalent to the weighted least squares [81] solution $\hat{x}_{wls} = (A^H W A)^{-1} A^H W b$, with weight $W = P_s$. The TLS estimate of $p^{(n)}$ is then given by

$$\hat{p}_{tls}^{(n)} = -(\widehat{H}_N^H \widehat{H}_N - \sigma_{min}^2 I)^{-1} \widehat{H}_N^H \widehat{h}_N, \quad (6.77)$$

where σ_{min} is the smallest singular value of $[\widehat{H}_N, -\widehat{h}_N]$. Given this estimate, the TLS estimates for the vertices of the underlying n -gon are obtained as roots of the polynomial equation $\widehat{P}_{tls}(z) = 0$. Subsequently, TLS estimates of the vector \mathbf{a}_N can be obtained as

$$\hat{\mathbf{a}}_n = (\widehat{\mathcal{V}}_N^H \widehat{\mathcal{V}}_N - \sigma'_{min}{}^2 I)^{-1} \widehat{\mathcal{V}}_N^H \widehat{\mathcal{T}}_N, \quad (6.78)$$

where σ'_{min} is the smallest singular value of the matrix $[\widehat{\mathcal{V}}_N, \widehat{\mathcal{T}}_N]$.

6.5.2 Matrix Pencil Technique

There exist a number of matrix pencil-based algorithm in the array processing literature [42, 69, 77]. Essentially, all of these methods are based on the simple observation that there is a constant delay between any two adjacent samples in a uniformly sampled time series. This fact leads to the following idea. Let \mathbf{Y}_0 and \mathbf{Y}_1 be defined

as

$$\mathbf{Y}_0 = \begin{bmatrix} \tau_{n-1} & \tau_{n-2} & \cdots & \tau_0 \\ \tau_n & \tau_{n-1} & \cdots & \tau_1 \\ \vdots & \vdots & \ddots & \vdots \\ \tau_{2n-2} & \tau_{2n-3} & \cdots & \tau_{n-1} \end{bmatrix}, \quad (6.79)$$

$$\mathbf{Y}_1 = \begin{bmatrix} \tau_n & \tau_{n-1} & \cdots & \tau_1 \\ \tau_{n+1} & \tau_n & \cdots & \tau_2 \\ \vdots & \vdots & \ddots & \vdots \\ \tau_{2n-1} & \tau_{2n-2} & \cdots & \tau_n \end{bmatrix}, \quad (6.80)$$

These matrices can be decomposed into products of Vandermonde matrices as follows.

$$\mathbf{Y}_0 = \mathcal{V}_n \text{diag}(\mathbf{a}_n) \mathcal{W}_n^T, \quad (6.81)$$

$$\mathbf{Y}_1 = \mathcal{V}_n \text{diag}(\mathbf{a}_n) \text{diag}([z_1, \cdots, z_n]) \mathcal{W}_n^T, \quad (6.82)$$

where

$$\mathcal{W}_n = \begin{bmatrix} z_1^{n-1} & z_2^{n-1} & \cdots & z_n^{n-1} \\ z_1^{n-2} & z_2^{n-2} & \cdots & z_n^{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}. \quad (6.83)$$

The roots of the pencil of matrices $\mathbf{Y}_0 - z\mathbf{Y}_1$ are the underlying vertices z_i . These roots are the generalized eigenvalues of $\mathbf{Y}_0 - z\mathbf{Y}_1$, or simply the eigenvalues of $\mathbf{Y}_1^{-1}\mathbf{Y}_0$. If more moments than the minimum number $(2n - 1)$ are used, the matrices \mathbf{Y}_0 and \mathbf{Y}_1 are similarly formed. If a total of N such moments are used, we have

$$\mathbf{Y}_0 = \begin{bmatrix} \tau_{n-1} & \tau_{n-2} & \cdots & \tau_0 \\ \tau_n & \tau_{n-1} & \cdots & \tau_1 \\ \vdots & \vdots & \ddots & \cdots \\ \tau_{N-2} & \tau_{N-3} & \cdots & \tau_{N-n-1} \end{bmatrix}, \quad (6.84)$$

$$\mathbf{Y}_1 = \begin{bmatrix} \tau_n & \tau_{n-1} & \cdots & \tau_1 \\ \tau_{n+1} & \tau_n & \cdots & \tau_2 \\ \vdots & \vdots & \ddots & \cdots \\ \tau_{N-1} & \tau_{N-2} & \cdots & \tau_{N-n} \end{bmatrix}, \quad (6.85)$$

so that only the number of rows of these matrices is increased. In this case, the solution may be obtained by using the pseudoinverse of \mathbf{Y}_1 [81]. The vertices are then the eigenvalues of $(\mathbf{Y}_1^H \mathbf{Y}_1)^{-1} \mathbf{Y}_1^H \mathbf{Y}_0$ [42, 81].

This approach enjoys some advantages over the direct least squares techniques described in the previous section. The numerical algorithms available for generalized eigenvalue problem [32] are computationally efficient. Furthermore, this approach, unlike the OLSP or TLSP, does not require the explicit computation of the coefficients of polynomial $P(z)$ from the data, nor does it involve the computation of the roots of this, possibly high order, polynomial. This, in turn, should translate to less sensitivity with respect to noise for certain signal to noise ratios.

6.6 Numerical Examples

In this section we present some simulations to illustrate the performance of the algorithms discussed in Section 6.5. As we shall see, the proposed algorithms are quite sensitive to the number of samples and the variance of the noise in these samples. Recall that the array processing algorithms in Section 6.5 were essentially built around the assumption that the corrupting noise on the samples ($\hat{\tau}_k$) was white. In our case, this is not true. In fact, as we mentioned earlier, the variance of the samples $\hat{\tau}_k$ grows without bound as k is increased. Hence, the basic noise assumption invoked in standard array processing algorithms is violated. Hence, it is not surprising that the proposed algorithms do not perform particularly well when there is significant noise. There is no reason, however, to believe that appropriate modifications of such algorithms should not produce much improved reconstructions. We leave the investigation of such possibilities for future research.

For illustration, we apply the least squares and pencil algorithms to the reconstruction of a triangle and a quadrilateral from noisy, estimated w-complex moments. The prototypical triangle has been chosen as one with the following vertices depicted in Figures 6-4, 6-5, 6-6, and 6-7.

$$V = \begin{bmatrix} -0.4655 & 0.0082 & -0.3283 \\ 0.2201 & 0.4599 & -0.1809 \end{bmatrix} \quad (6.86)$$

The data was collected in the form of 20 projections with 500 samples per projection at a signal-to-noise ratio of 55. Note that given the high number of samples and the high SNR, we essentially have a noise-free data set. As we will see next, even with such “clean” data, typical reconstructions will have (percent Hausdorff) reconstruction errors on the order of 10 to 15 percent.

Figure 6-2 shows the performance curves for four the four algorithms: Ordinary Least Squares (OLS), Total Least Squares (TLS), Weighted Least Squares (WLS), and the Pencil Algorithm. These curves show average performance obtained by generating 100 runs of a Monte-Carlo simulation vs the number of moments used over the minimum necessary (i.e. the overfit parameter). Recall that according to Theorem 12, w-complex moments of up to order 5 are needed (at minimum) to reconstruct the triangle. Hence, an overfit parameter value of 2 corresponds to using estimated w-complex moments of up to order 7.

Figure 6-3 shows the same curves as 6-2 except overlaid to show comparisons. Note that overall, the TLS algorithm performs best, while the performance of the OLS algorithm and the Pencil algorithm exactly coincide. The WLS algorithm performs essentially the same for values of the overfit parameter larger than 2. This is due to the fact that as higher order moments are considered, these are weighted according to their inverse variance which become considerably small. The graphs show that the overfit parameter value of 3 in the TLS algorithm provides, on average, the best reconstructions. Note that somewhat surprisingly, the incorporation of even more moments does not improve the reconstruction error. This is again due to the fact that higher order moments become severely noisy and hence at some point their use

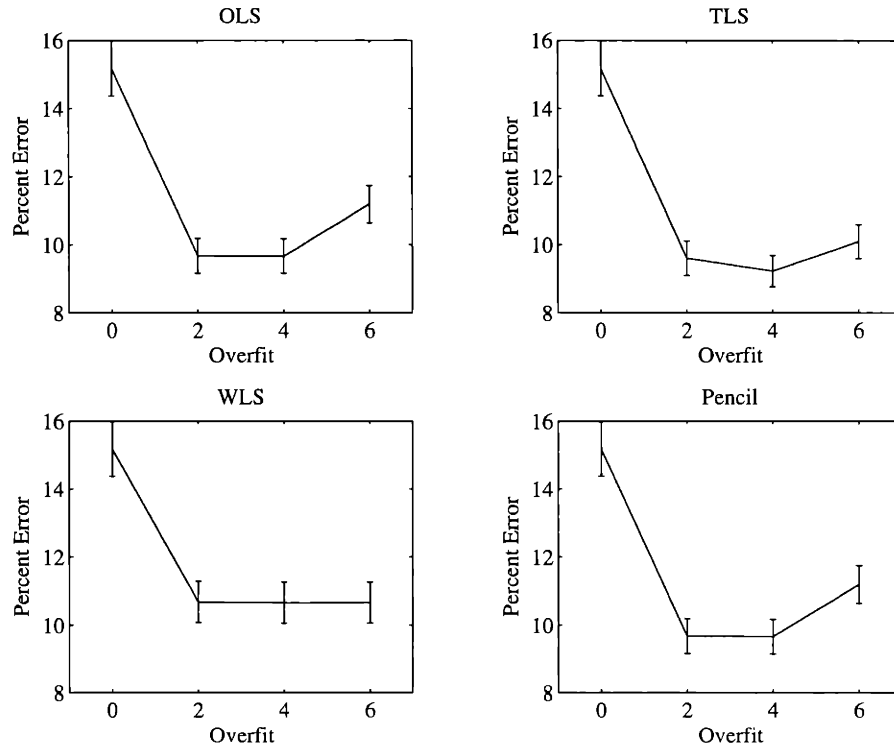


Figure 6-2: Performance curves at SNR=55

results in diminishing returns.

Figures 6-4, 6-5, 6-6, and 6-7 show sample reconstructions for all the aforementioned algorithms for values 0, 2, 4, and 6 of the overfit parameter.

The quadrilateral to be reconstructed was chosen as the polygon P' shown in Figure 6-1. Projections from 20 equally spaced angles in $[0, \pi)$ were taken with 1000 samples per view at a signal to noise ratio of 150. Reconstruction of the underlying polygon are shown with overfit parameter values of 0 in Figure 6-8. The corresponding estimated coefficients a'_i using OLS are

$$a'_1 = -0.0053 - 0.5868i,$$

$$a'_2 = -0.0308 + 1.0271i,$$

$$a'_3 = 0.4020 - 0.2422i,$$

$$a'_4 = -0.3659 - 0.1981i.$$

while the corresponding estimated coefficients using TLS are given by

$$a'_1 = -0.0053 - 0.5872i$$

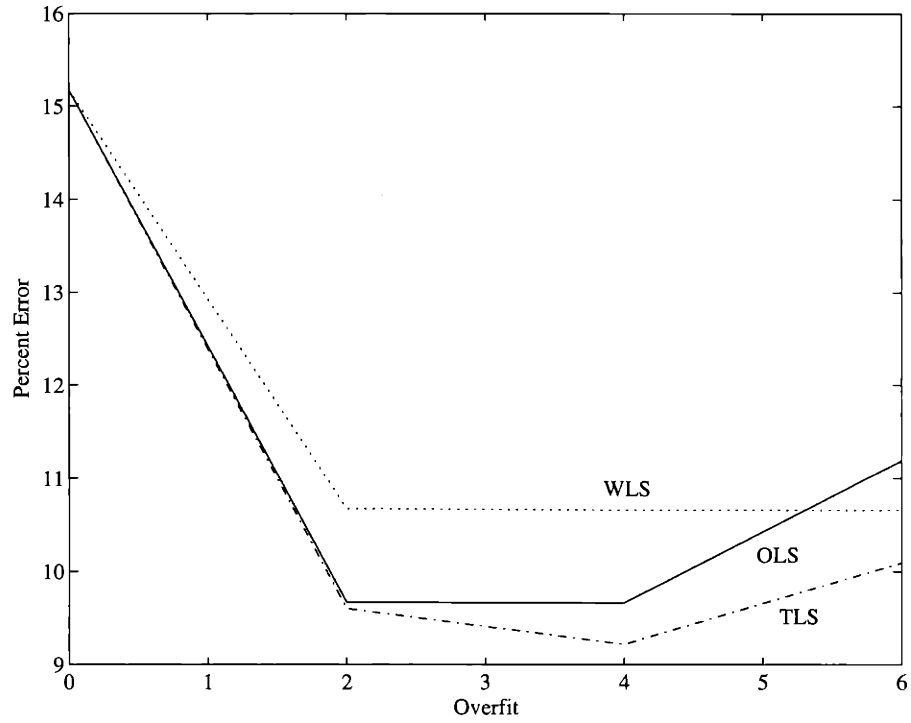


Figure 6-3: Overlaid performance curves at SNR=55

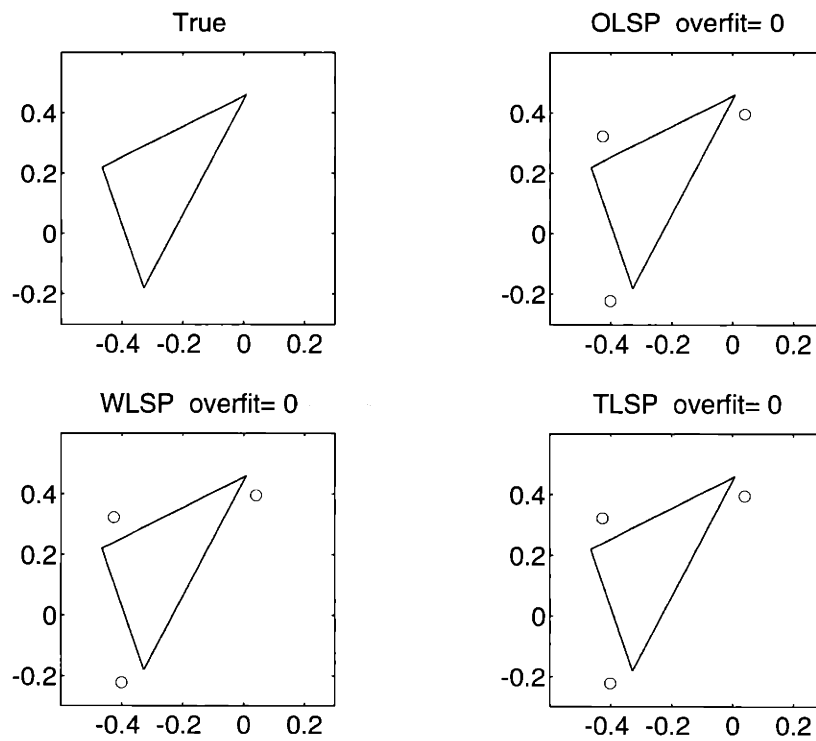


Figure 6-4: Sample reconstructions at SNR=55 solid: actual, circles: reconstructed

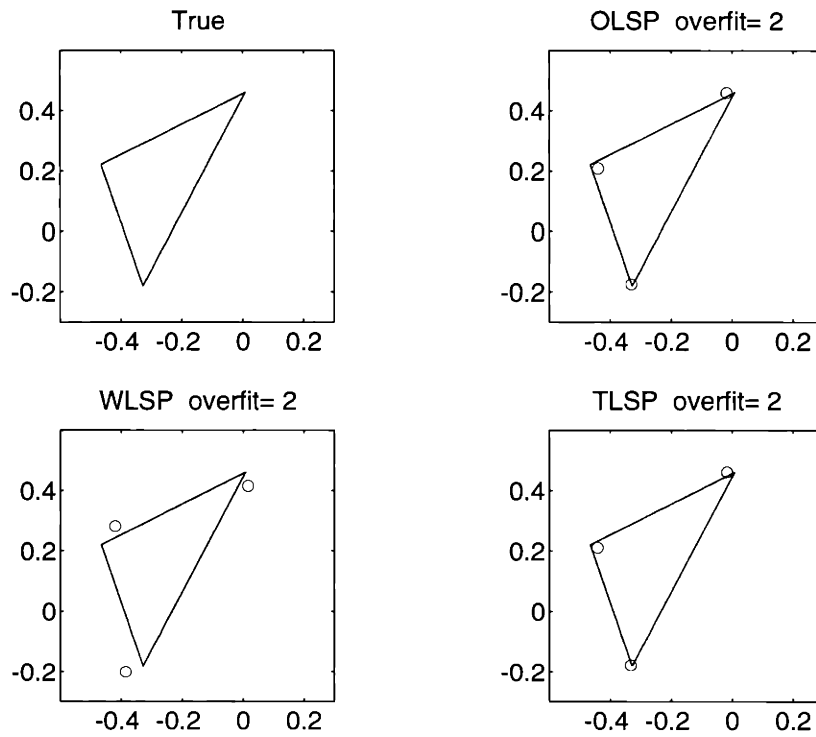


Figure 6-5: Sample reconstructions at SNR=55 solid: actual, circles: reconstructed

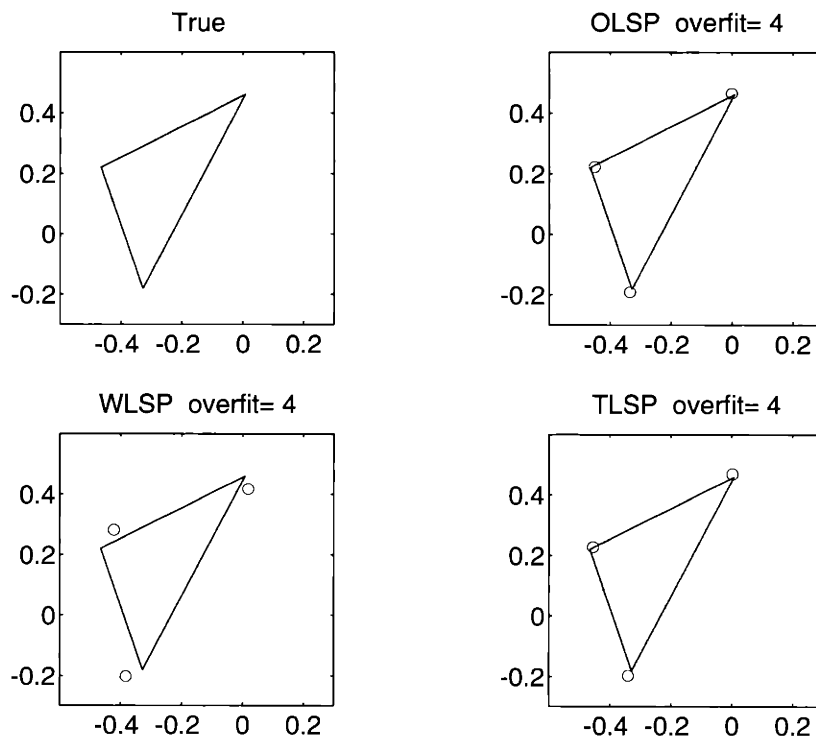


Figure 6-6: Sample reconstructions at SNR=55 solid: actual, circles: reconstructed

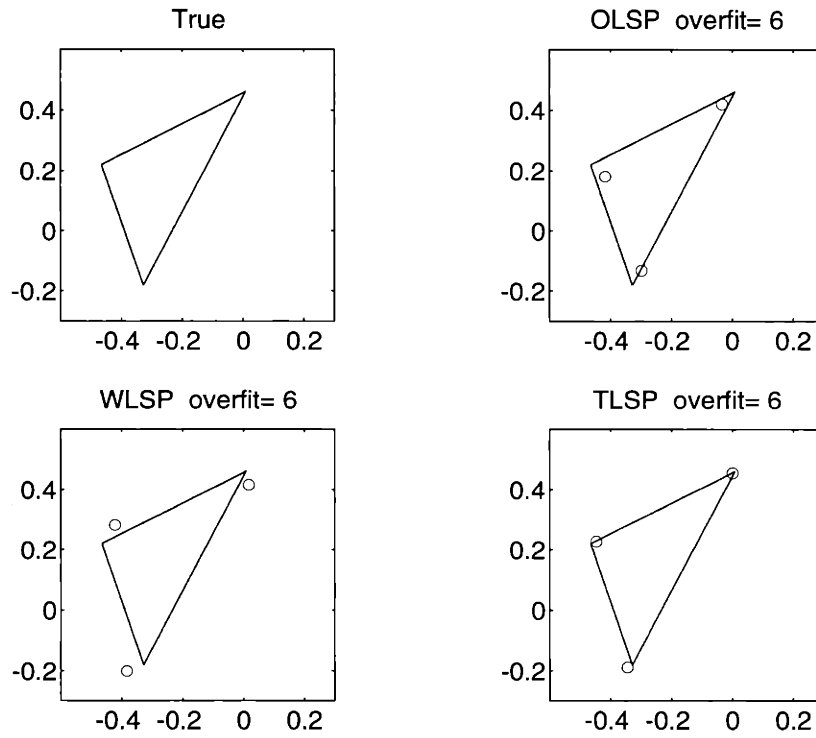


Figure 6-7: Sample reconstructions at SNR=55 solid: actual, circles: reconstructed

$$\begin{aligned}
 a'_2 &= -0.0317 + 1.0539i \\
 a'_3 &= 0.4028 - 0.2463i \\
 a'_4 &= -0.3664 - 0.2021i
 \end{aligned}$$

(6.87)

Comparing these to the values given in (6.44), we can see that they are fairly close. To decide on how to connect the given estimated vertices, we compute the coefficients a_j via formula (6.13) for each configuration, and compare these values to the estimated coefficients given above. The choice of configuration is then made according to which of the coefficient sets found using (6.13) most closely approximates the estimated coefficient set. Let us carry out this procedure for the TLS estimated vertices of the quadrilateral using overfit parameter of 0. For convenience, referring to Figure 6-1, we shall denote the configurations in which the vertices of P , P' and P'' are connected as configurations 1, 2, and 3, respectively, so that the correct configuration is number 2. The estimated coefficients using TLS, and the corresponding coefficients computed using (6.13) are shown in table 6.1 along with the l^1 norm of their difference defined

Estimated a_j	a_j config. 1	a_j config. 2	a_j config. 3
-0.0053-0.5872i	-0.1014 - 0.3998i	0.0155 - 0.6423i	0.1169 - 0.2424i
-0.0317 + 1.0539i	0.3837 - 0.4002i	0.2669 - 0.1577i	0.8735 + 0.2877i
0.4028 - 0.2463i	0.6066 + 0.4454i	0.1320 + 0.9875i	-0.4746 + 0.5421i
-0.3664-0.2021i	-0.8890 + 0.3546i	-0.4144 - 0.1875i	-0.5158 - 0.5874i
l^1 Difference	3.2075	2.2601	3.1446

Table 6.1: Estimated and Computed coefficients a_j for vertices of quadrilateral reconstructed using TLS with overfit parameter of 0

by:

$$l^1 \text{ Difference in } a_j\text{'s} = \sum_{j=1}^4 \|a_j(\text{estimated}) - a_j(\text{from (6.13)})\|. \quad (6.88)$$

As can be seen from Table 6.1, as measured by the l^1 norm, the estimated coefficients are closest to the coefficients obtained when the estimated vertices are connected according to configuration 2. Hence, our algorithm has correctly identified the underlying configuration.

With an overfit parameter of 1, the reconstructions shown in Figure 6-9 are obtained. The corresponding values of the coefficients a'_j are the same (to within 10^{-16}) as the previous case where an overfit parameter of 0 was used. The reconstructions using an overfit parameter of 2 are shown in Figure 6-10, where the estimated a'_j parameters using OLS are

$$\begin{aligned} a'_1 &= -0.0053 - 0.5868i, \\ a'_2 &= -0.0308 + 1.0271i, \\ a'_3 &= 0.4020 - 0.2422i, \\ a'_4 &= -0.3659 - 0.1981i. \end{aligned} \quad (6.89)$$

The estimates of these parameters using TLS are

$$\begin{aligned} a'_1 &= -0.0168 - 0.5774i, \\ a'_2 &= -0.0080 + 0.1819i, \end{aligned}$$

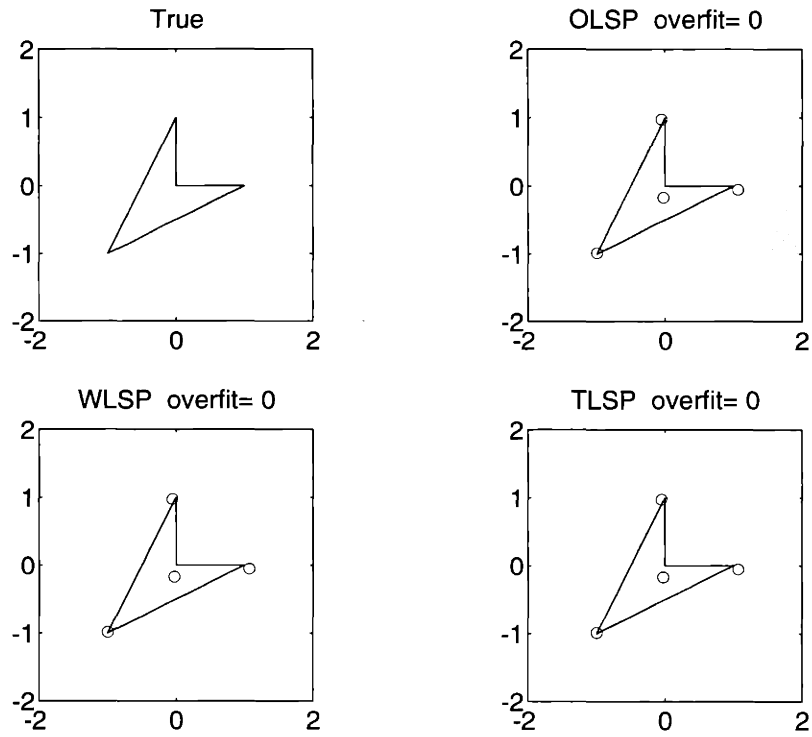


Figure 6-8: Sample reconstructions at SNR=150 solid: actual, circles: reconstructions

$$\begin{aligned}
 a'_3 &= 0.4113 - 0.1072i, \\
 a'_4 &= -0.3997 - 0.0700i.
 \end{aligned}$$

(6.90)

Let us use these last set of estimated coefficients along with the estimated vertices using TLS, with overfit parameter of 2, to decide how the estimated vertices are to be connected. We again show the values of the coefficients obtained from (6.13) and the values of the total difference of these coefficients with the estimated coefficients in Table 6.2. In this case, the algorithm has again correctly chosen configuration 2 as the solution, but note that the value of the l^1 difference of the coefficients for configurations 1 and 3 are much closer to the minimum obtained for configuration 2.

It is interesting to note that the use of higher order moments (i.e. Overfit parameter > 0) degrades the estimate of the concave vertex of the underlying object more than the others. Also, note that in this experiment, the quality of the estimates of the vertices and the parameters a'_j shows degradation as more moments (beyond the minimum necessary) are used.

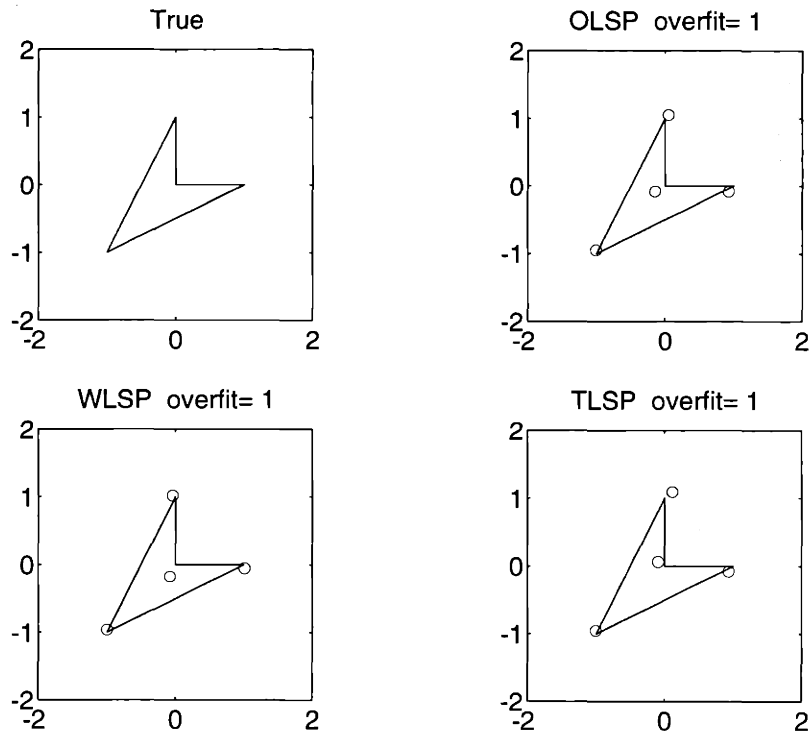


Figure 6-9: Sample reconstructions at SNR=150 solid: actual, circles: reconstructions

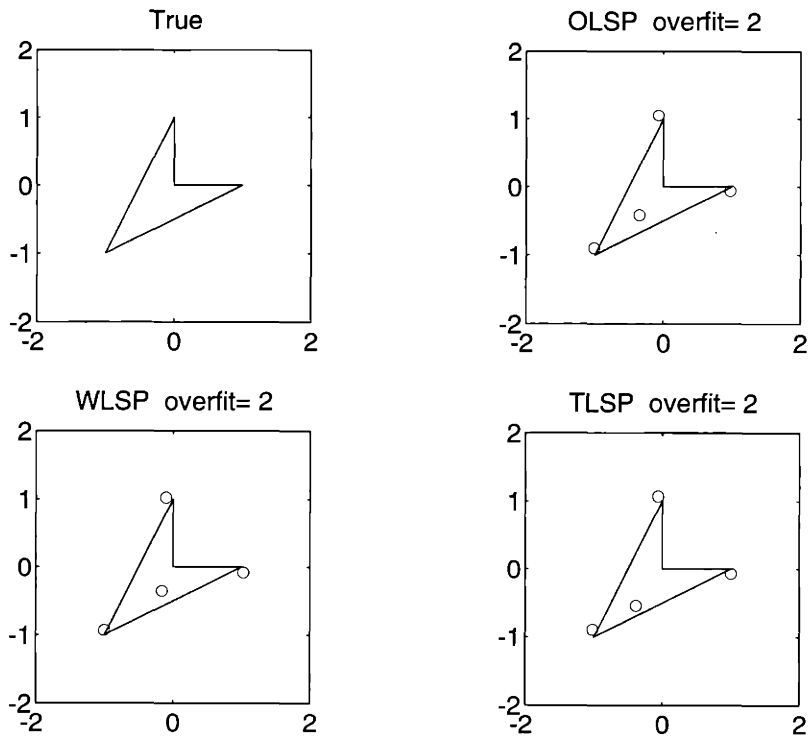


Figure 6-10: Sample reconstructions at SNR=150 solid: actual, circles: reconstructions

Estimated a_j	a_j config. 1	a_j config. 2	a_j config. 3
-0.0168-0.5774i	-0.0327 - 0.5684i	0.0452 - 0.6660i	0.0779 - 0.0976i
-0.0080 + 0.1809i	0.1250 - 0.1381i	0.0471 - 0.0405i	0.8478 + 0.3908i
0.4113- 0.1072i	0.8007 + 0.4313i	0.1092 + 0.8601i	-0.6915 + 0.4289i
-0.3997-0.0700i	-0.8930 + 0.2752i	-0.2015 - 0.1536i	-0.2342 - 0.7220i
l^1 Difference	1.6305	1.5648	3.2691

Table 6.2: Estimated and Computed coefficients a_j for vertices of quadrilateral reconstructed using TLS with overfit parameter of 2

6.7 Conclusion

In this chapter we have presented algorithms for the reconstruction of binary polygonal shapes from noisy measurements of their moments. The mathematical basis of these algorithms is the Motzkin-Schoenberg [83] formula in complex variable theory, and Prony's method. The contributions we have made in this chapter can be grouped into two categories. From a mathematical standpoint, we have improved a result due to Davis [18] which states that the vertices of a triangle are uniquely determined by its w -complex moments up to order 3. Our generalization states that the vertices of any nondegenerate, simply-connected, n -sided polygon can be determined from its w -complex moments up to order $2n - 1$. We have also shown that this number of moments is sufficient in some cases to uniquely specify the interior of the polygon.

From an estimation-theoretic viewpoint, we have established an explicit connection between the field of array signal processing and the problem of tomographic reconstruction of binary polygonal objects. We have employed array processing techniques to solve this reconstruction problem. Numerically, however, these algorithms do not seem to perform well in the presence of significant noise. Other array processing algorithms and modifications of the existing ones, we believe, may lead to significant improvements in performance. It should be kept in mind that the algorithms we have presented using a finite number of w -complex moments for the reconstruction of polygons do not make use of all the available tomographic data. In fact, the s -complex (or w -complex) moments are arrived at by forming a complex linear combination of the geometric moments of the underlying polygon. Hence, some part of the information in the estimated geometric moments is actually not being used. This may, to some

extent, also explain the performance of the proposed algorithms.

Finally, it is worth noting that aside from the array processing algorithms, the estimation problem given by (6.63) can be viewed directly as a nonlinear optimal estimation problem. We can directly write the observation equations as

$$\hat{\tau}_k = F_k(z_1, z_2, \dots, z_n) + w_k \quad (6.91)$$

where

$$F_k(z_1, z_2, \dots, z_n) = \sum_{j=1}^n a_j z_j^k \quad (6.92)$$

is a nonlinear function of the vertices *and* how they are connected. Nonlinear optimal estimation techniques can then be applied to (6.91) to estimate the underlying polygon. It would be interesting to study the performance of such algorithms. We leave this for future research efforts.

Chapter 7

Contributions, Conclusions, and Future Research

7.1 Contributions

In this section we briefly summarize the contributions made in each of the previous chapters.

7.1.1 Parametric Model-Based Reconstruction

In Chapter 3, we presented a general framework for the optimal reconstruction of finitely parameterized objects. In particular, we focused on the reconstruction of binary polygonal objects.

In this model-based framework, the tomographic reconstruction problem is viewed as a finite-dimensional parameter estimation problem. The Maximum Likelihood formalism is then used to estimate the missing parameters. In contrast to the classical techniques, such as FBP, the ML based reconstructions showed robustness to noise and data loss and distribution. The drawback of such ML-based formulations is that the resulting optimization problems are highly non-linear and thus a good initial guess is necessary to ensure convergence of optimization routines to the true ML estimate. We presented an algorithm that quickly and reliably produces a good initial guess for

the nonlinear optimization algorithm.

When the number of underlying parameters of the object was assumed to be unknown, a Minimum Description Length criterion was employed that simply generalizes the ML framework to penalize the use of an excessively large number of parameters for the reconstruction. The MDL approach was shown to work successfully in estimating the number of sides and the underlying object itself for low signal-to-noise ratio situations and for a variety of sampling scenarios.

Two major contributions of our work in the area of parametric model-based tomographic reconstruction are (I) the study of a more general class of objects than were studied before; namely binary, polygonal objects, and (II) the explicit incorporation of information-theoretic criteria for determining the number of parameters to be used in the reconstruction.

7.1.2 Direct Extraction of Geometric Information from Tomographic Data

In Chapter 4 we explicitly used the linear relationship between the moments of a function and those of its Radon transform in an estimation theoretic framework. In this framework, we showed that:

- it is possible, and quite simple, to extract geometric information about an object in the form of moments directly from (the moments of) its noisy Radon transform data and without reconstruction of the underlying image.
- the (unique) computation of the moments of a function from its Radon transform requires a minimum number of views. This, in essence provides one with a theoretical upper bound on the amount of uniquely identifiable geometric information that is contained in a limited data set.
- the linear nature of the dependence of the moments of an image on those of its Radon transform make the explicit computation of estimation error covariances possible. This allowed us to specify not only the geometric features of an image

directly from projections, but also the degree of confidence which we place on these estimates.

- the moments of the projections are sufficient statistics for the estimation of a function f from its Radon transform projections.
- the Radon transform operator \mathfrak{R} as an operator from one Hilbert space to another, can be decomposed in terms of operators which map a function and its Radon transform to their respective moments. This led to a new set of interpretations of classical reconstruction algorithms.
- the problem of estimation of moments of an image from raw projection data is numerically more stable for a particular set of sampling strategies in the projection domain. Hence, we used the analysis of the stability of this estimation problem to arrive at optimal sampling strategies.

In summary, the most significant overall contribution we have made in Chapter 4 is to highlight the importance and use of the fundamental relationship between the moments of a function and those of its Radon transform. Although this relationship is well-known in the mathematics community, it has not been used extensively in the engineering community. We hope that our work will serve to promote the careful study of the applications of this very useful property of the Radon transform.

7.1.3 Variational Formulations for Reconstruction from Moments

In Chapter 5, we presented variational algorithms for the reconstruction of a positive function (an image) from a finite set of its noisy moments. In particular, we studied Divergence-based and quadratic regularization schemes. We showed that both of these formulations have exact statistical interpretations as Maximum A-Posteriori estimates of f . Simple iterative extensions of these algorithms were shown to yield solutions to equality-constrained minimization problems.

The important contributions of Chapter 5 are

- the formulation and solution of an optimal, robust and computationally efficient algorithm for the reconstruction of a positive function from a finite number of its noisy (estimated) moments. In this formulation, we took into account the moment estimation error covariance explicitly. This represents an extension of existing algorithms for recovering a function from its moments.
- we showed that simple iterative extensions of our divergence-based regularized framework lead to solutions to highly complex nonlinear, equality-constrained optimization problems. These problems are extremely difficult to solve by other techniques since it is not possible to describe the constraint sets in question analytically. Furthermore, our algorithm is more numerically stable and analytically tractable than other existing algorithms.
- we showed that our iterative algorithms perform quite well even with very noisy and sparse data and the modest use of only a few moments. Furthermore, our framework is general enough to allow the direct incorporation of geometric prior information into the reconstruction process. As a particular instance of this, the use of the classical FBP estimate as a prior estimate was shown to yield much improved reconstructions. The classical FBP reconstructions produce very rough reconstruction of the underlying features of the image along with many anomalies that are not part of the underlying image. We have shown that using our algorithm, these rough reconstructions can be significantly enhanced and the underlying features can be extracted considerably more accurately.

Overall, the most significant contribution of this chapter is that we have presented a general framework for optimal reconstruction of images with controlled degrees of freedom that can easily incorporate prior knowledge. This framework is numerically efficient and is flexible enough to incorporate various forms of prior information. Furthermore, we have shown that, given very rough initial estimates of the underlying image, our framework is capable of producing superior reconstructions when compared to classical reconstruction algorithms.

7.1.4 Array Processing Methods for Reconstruction from Moments

In Chapter 6 we studied the reconstruction of binary polygonal objects from a finite number of their complex moments. In specific, we made the following contributions.

- we showed that the vertices of a binary polygonal region can be recovered from a finite number of its complex moments. This result extended earlier work that established the same but only for triangles.
- we showed a binary polygonal object is determined by a finite number of projections from noncongruent, but otherwise arbitrary directions. This contribution is a generalization of earlier results regarding reconstructability of binary objects from few projections.
- we presented explicit algorithms for the extraction of the vertices of a polygon from a finite number of complex moments. These algorithms are essentially borrowed from the field of antenna array processing, establishing for the first time, explicit connections between tomography and array processing.

The most important overall contribution we made in Chapter 6 is that we have discovered an interesting and potentially very useful connection between array processing and tomography.

7.2 Future Research Directions

In this section we present some ideas for further research. The first section of this chapter is concerned with a conjecture regarding explicit constraints on projection samples that may be derived from the assumption that the underlying object is binary and convex. The remaining part of this chapter is concerned with research issues that are directly related to the work performed in this thesis.

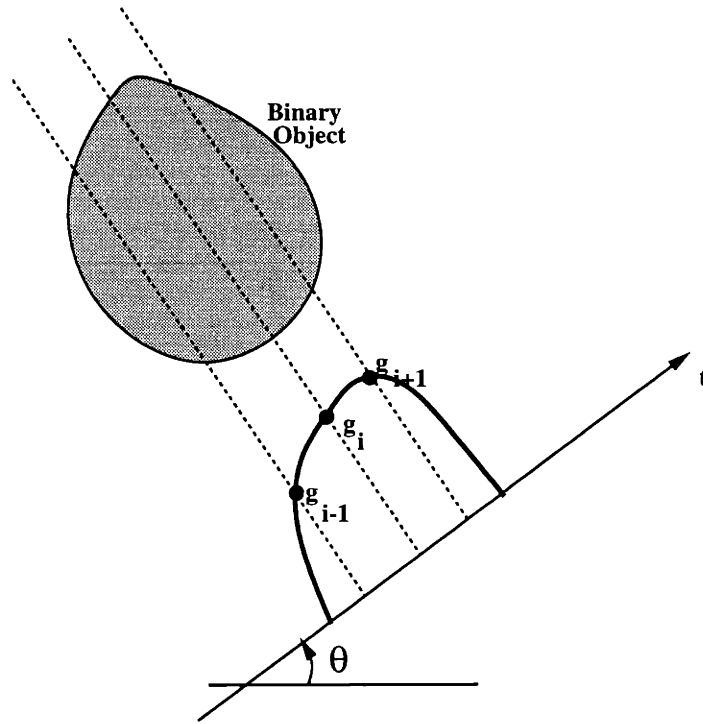


Figure 7-1: Parallel Beam Projection Data

7.2.1 Geometric Constraints and Projection Data

Projection data is typically collected according to one of two modalities. In the parallel-beam (PB) case, a fixed set of directions is given and for every direction line-integrals of the density function of the object are collected along parallel beams in that direction. See Figure 7-1. A result of Falconer [22] states that the PB projections of a compact convex set in the plane are necessarily convex function of the variable t . The converse of this result is also known to be true [21]. Hence, a binary object is convex *if and only if* its PB projections in all directions are convex functions of t . Let g_{i-1} , g_i , and g_{i+1} denote any triplet of samples of a projection, from a fixed direction, of some binary object \mathcal{O} . Then a necessary and sufficient condition for the convexity of \mathcal{O} is that for every such triplet of samples, we have

$$g_{i-1}(t_{i+1} - t_i) - g_i(t_{i+1} - t_{i-1}) + g_{i+1}(t_i - t_{i-1}) \geq 0. \quad (7.1)$$

This is simply a statement that every projection is convex and quite similar to Prince's condition for deciding whether a vector of support values is consistent or not [72]. As a corollary to (7.1), it follows that if the data samples are equally spaced along the t -axis, then (7.1) simply becomes

$$g_i \geq \frac{g_{i+1} + g_{i-1}}{2}. \quad (7.2)$$

In other words, for any triplet of equally-spaced data points, the middle one should be larger than the average of the other two. Given noisy samples $Y_j = g_j + n_j$ of the projections of a convex object, the relation (7.1) is in general not satisfied. In such instances, we can imagine constructing a constraint set defined by (7.1) and computing the (Euclidean) projection of the given data set $\{Y_j\}$ onto this set. This operation would yield a new data set $\{\hat{g}_j\}$ which would be consistent with the assumption that the underlying binary object is convex. This consistent data set may then be used in a reconstruction algorithm. Prince [72] used a similar *consistency condition* to obtain valid support vectors from noisy projection data.

The question naturally arises as to whether consistency conditions similar to (7.1) exist if the projection data is collected in Fan Beam (FB) mode. Most commercial CT scanners collect their projection data in FB format. This is to say that a number of sources emit radiation in many directions from a fixed point, as illustrated in Figure 7-2.

We define the concept of a polar projection of an object analogously to the PB projection as follows.

Definition 6 *Let the object $\mathcal{O} \subset \mathcal{R}^2$ be given, and consider the point $p \in \mathcal{R}^2$. Let the function $c(p, \phi)$ denote the total length of the (possibly disjoint) chords $\mathcal{O} \cap L(p, \phi)$, where $L(p, \phi)$ denotes a line through p making an angle of ϕ with some fixed direction. The plot of $c(p, \phi)$ in polar coordinates, with p as the pole, is termed the polar projection of \mathcal{O} with respect to p , and is denoted $\mathcal{S}(p, \phi)$.*

We propose the following conjecture.

Conjecture 1 *A compact object $\mathcal{O} \subset \mathcal{R}^2$ is convex iff its polar projections $\mathcal{S}(p, \phi)$*

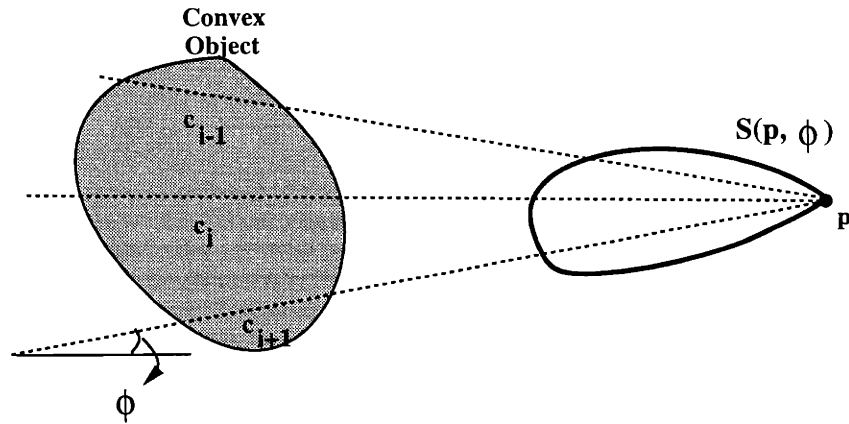


Figure 7-2: Fan-Beam projection data

are convex subsets of \mathcal{R}^2 for all $p \in \mathcal{R}^2$.

The forward implication has been implicitly proved [56] for the case where \mathcal{O} has a smooth boundary. There is very strong evidence that the reverse implication is also true. For instance, note that if the object \mathcal{O} is assumed to be star-shaped in the above conjecture, then the reverse implication is easily verified. Simply consider $\mathcal{S}(p, \phi)$ where p is the center of the object, and convexity follows immediately. This observation indicates that perhaps the hypotheses of the conjecture may even be too strong. i.e. perhaps the conjecture can be proved if $\mathcal{S}(p, \phi)$ is assumed convex for p only in some subset of \mathcal{R}^2 .

7.2.2 Alternative Variational Formulations

In Chapter 5 we discussed a number of variational formulations for the reconstructions of objects from their noisy moments. Here we present some ideas for further study in this rather rich area of research

Smoothness Constraints

A rather interesting regularization method is regularization by a smoothness constraint. In particular, the cost function we have in mind is of the form

$$J_S(f, f_0) = \gamma S(f, f_0) + \frac{1}{2}(\widehat{\mathcal{L}}_N - \mathcal{L}_N(f))^T \Sigma_N (\widehat{\mathcal{L}}_N - \mathcal{L}_N(f)) \quad (7.3)$$

where

$$S(f, f_0) = \int \int_{\mathcal{O}} \|\nabla(f - f_0)\|^2 dx dy. \quad (7.4)$$

The first term of the cost function penalizes large variations between the values of the *gradients* of f and the prior f_0 . By appealing to the calculus of variations, we can write the general form of the solution as follows.

$$\nabla^2 f = \nabla^2 f_0 + \frac{1}{\gamma} \Phi_N^T(x, y) \Sigma_N (\mathcal{L}_N(f) - \widehat{\mathcal{L}}_N) \quad (7.5)$$

To solve for f explicitly, we must first decide on boundary values for f over the spatial support of the image and then solve the elliptic partial integro-differential equation given by (7.5), subject to the chosen boundary values. Recently, a novel technique for solving such problems efficiently has been proposed in [60]. It can be shown that the smoothness term can be interpreted as a prior probabilistic model for f ; namely a Markov Random Field. In [60], the authors have used the interpretation of the smoothness constraint as a “fractal” prior to motivate regularization based on multiscale stochastic models. These models provide very efficient algorithms for solving elliptic PDE’s such as (7.5). The application of these models to the tomographic reconstruction problem, and in specific to the moment problem, appears to be quite interesting and potentially important area of research.

Mixed Regularization Terms

There is no reason to restrict the study of regularized solutions to tomographic reconstruction problems to single regularization terms as we have done in Chapter 5.

We can consider regularization functionals of the form

$$\begin{aligned}
J_M(f, f_0) &= \gamma_1 D(f, f_0) + \gamma_2 Q(f, f_0) + \gamma_3 S(f, f_0) \\
&+ (\hat{\mathcal{L}}_N - \mathcal{L}_N(f))^T \Sigma_N (\hat{\mathcal{L}}_N - \mathcal{L}_N(f))
\end{aligned} \tag{7.6}$$

where each regularization parameter γ_i is picked to enhance or suppress a particular feature of the solution or include more prior geometric information into the reconstruction algorithm. For instance, by including the term corresponding to γ_1 , as we showed, we may include information about the support of the object by letting f_0 be an indicator set over the estimated spatial support. By adding in the term corresponding to γ_3 , we enforce this condition further by making sure that the estimated image has a discontinuity at the boundary of the spatial support. Of course, the price we pay for including more regularization terms is that the resulting algorithms become more computationally intensive. To see this, note that in the general formulation (7.6), the solution \hat{f}_M is obtained as the solution of the following integro-differential equation.

$$\gamma_1 \log \left(\frac{f}{f_0} \right) + \Phi_N^T(x, y) \Sigma_N (\mathcal{L}_N(f) - \hat{\mathcal{L}}_N) + \gamma_2 \nabla^2 (f - f_0) + \gamma_3 (f - f_0) = 0 \tag{7.7}$$

7.2.3 Applications to Pattern Recognition and Data Compression

As we showed in Chapter 6, the errors in the estimated s -complex moments can (in many instances) be considered to be uncorrelated. This, in effect, implies that the computation of the s -complex moments amounts to an orthogonal projection of the noisy tomographic data. Hence, we may think of the estimated s -complex moments as estimated signatures of the underlying object which may be used to compare one set of projection data to another set for direct matching of geometric information. Statistically optimal pattern recognition algorithms may be employed here to classify, recognize, and compare geometric features, without ever performing a direct reconstruction of the underlying image. (The same may be done by using

the geometric or orthogonal estimated moments. Note that it follows from definition that the set of complex moments contains *less* information than the set of geometric or orthogonal moments.) This observation points to a connection between the two fields of tomography and pattern recognition which surely deserves attention in future research efforts in either area. The importance of this research is accentuated by the fact that many diagnostic applications of tomographic imaging rely directly on pattern analysis but in a very empirical and ad hoc way. The process can now be formalized and studied in an analytically meaningful setting.

As an illustration of the application of s-complex moments, we show that from these, algebraic quantities can be derived which are invariant to image rotation and scaling. It is easy to see from the definition of the s-complex moments that the rotation of an image only affects the phase of these moments. In particular, let c_k denote the k^{th} order s-complex moment of some image $f(x, y)$. Let c'_k denote the k^{th} order s-complex moment of a rotated version of the image $f(x, y)$. Then it is true that

$$c'_k = c_k e^{i\phi} \tag{7.8}$$

for some angle ϕ . From this it immediately follows that the magnitude of c_k is the same as the magnitude of c'_k so that the quantities $|c_k|$ (which, of course, still have uncorrelated estimation errors for different k) are invariants of the underlying image with respect to rotation. In a similar fashion, we can show that the numbers c_k/c_0 are invariant with respect to scaling of the image. Consequently, the features $|c_k|/|c_0|$ are invariant with respect to both scaling and rotation of the image.

7.2.4 Generalizations of the Motzkin-Schoenberg Formula

In [19], Davis writes:

“ In order to obtain complex formulas for integrals of the type $\int \int G(x, y) dx dy$ where G is an analytic function of two real variables x and y , we can write

$$G(x, y) = G\left(\frac{1}{2}(z + \bar{z}), \frac{1}{2i}(z - \bar{z})\right). \tag{7.9}$$

Expansion of G in a power series now suggests that one should study integrals of the form $\int \int_T \bar{z}h(z)dx dy \dots$ ”

Subsequently, he proves a generalization of the MS formula for higher derivatives of an analytic function over a polygonal region.

Theorem 11 *Given a triangle T , we can find constants a_{kj} depending on the vertices z_1, z_2 , and z_3 such that*

$$\int \int_T \bar{z}^m h^{(m+2)}(z) dx dy = \sum_{k=0}^m \sum_{j=1}^3 a_{kj} h^{(k)}(z_j) \quad (7.10)$$

for all $h(z)$ regular in the closure of T .

This study suggests that the array processing ideas may be applicable to the reconstruction of *analytic* images $f(x, y)$ defined over polygonal regions. Consider an analytic function $f(x, y)$ over a polygonal region P . Now in (7.9) write

$$G(x, y) = x^p y^q f(x, y). \quad (7.11)$$

This G is analytic and hence the comments of Davis quoted above essentially mean that the moments of an analytic image over a polygonal region yield equations of the form (7.10) that are analogous to signal measurement equations in array processing. It would be interesting to study what information can be learned about this more general class of images from measurements of their moments through array processing algorithms.

7.2.5 Other Array Processing Algorithms and Their Performance

Another important area of future work is the study of other array processing algorithms for reconstruction of polygonal objects from projections. In Chapter 6, we briefly studied the Least Squares Prony algorithm and a matrix Pencil algorithm. Other algorithms such as MUSIC [82], and Pisarenko Harmonic Decomposition [71]

need to be studied to determine if they perform better. Another approach would be to try to fit more vertices to the given moment data even when the number of underlying vertices is known. In this way we may hope to remove some of the sensitivity of the algorithms to noise. Such an algorithm would likely produce degenerate, or nearly degenerate, polygons as reconstructions from which the underlying polygon may be extracted by throwing out the “redundant” vertices (i.e. those collinear, or nearly collinear, with at least two other vertices). This kind of “overfitting” technique has been used with good results in the array processing community [81].

Variations of all the available algorithms may also be studied to take into account the fact that our array processing problem has non-identical noise distribution on the measured samples (w-complex moments). Since the variance of the estimation errors in higher order estimated moments is larger, new performance studies need to be done with this fact in mind. In particular, a variation of the Total Least Squares algorithm called constrained TLS, introduced in [1] is particularly interesting. This algorithm solves the array processing problem with assumptions that are similar in spirit to those under which the polygonal reconstruction problem is formulated. The application of the ideas in [1] to the polygonal reconstruction problem appears to be a promising line of research.

7.3 Conclusions

In this thesis, we have developed techniques for the reconstruction of images for noisy and sparse measurements of their Radon transform projections. Our techniques first estimate *geometric* features of the underlying images directly from projections and then use this information to produce estimates of the underlying image. The major areas of contribution of this thesis can be classified into four areas:

- **The optimal reconstruction of binary polygonal objects with unknown number of sides:** We showed that high quality reconstructions of these objects can be obtained from sparse and very noisy projection data. The number of underlying parameters can also be reliably estimated. We showed that for

moderate numbers of total samples, our algorithms are relatively more sensitive to the number of samples per view than number of views.

- **The direct optimal estimation of geometric information from tomographic data in the form of moments:** We developed a Maximum-Likelihood algorithm for the (linear) estimation of the moments of a function directly from projection data. We showed that the moments of the projections can be viewed as sufficient statistics for the underlying image. We also showed that the Radon transform operator can be decomposed in terms of operators that relate a function and its Radon transform to their respective moments.
- **Variational algorithms for reconstruction of images from noisy measurements of a finite number of their estimated moments:** We developed direct and iterative algorithms for the reconstruction of images from their estimated moments and showed that these algorithms are capable of producing superior reconstructions when compared to classical reconstruction algorithms. Our framework was shown to be quite flexible in that it naturally facilitates the incorporation of prior knowledge.
- **Connections to, and applications of, array processing algorithms to the reconstruction of binary polygonal objects from projections:** We showed that binary polygonal objects can be reconstructed from a finite number of their complex moments and used array processing ideas to design algorithms for doing this reconstruction. We observed that the algorithms we designed did not perform well even with very little noise in the projections. This, we believe, is due to the fact that our algorithms did not take into account the special structure of the estimation error of complex moments. We also observed that the performance of these algorithms were particularly sensitive to the number of samples per view.

We believe that the application of other specialized array processing algorithms may well lead to robustness to noise.

Overall, we hope that the contributions of this thesis to the field of tomography will serve to promote the study of a host of new image processing and reconstruction algorithms based on the extraction of geometric information.

Bibliography

- [1] T. Abatzoglou and J. M. Mendel. The constrained total least squares technique and its applications to harmonic superresolution. *IEEE Trans. Signal Processing*, 39(5):1070–1086, May 1991.
- [2] N.I. Akhiezer. *The classical moment problem and some related questions in analysis*. Hafner, New York, 1965.
- [3] U. Amato and W. Hughes. Maximum entropy regularization of Fredholm integral equations of the first kind. *Inverse Problems*, 7:793–808, 1991.
- [4] O.B. Bekker, editor. *Spaces of Analytic Functions*, volume 512 of *Lecture Notes in Mathematics*. Springer-Verlag, 1975.
- [5] S. K. Berberian. *Hilbert Space*. Chelsea Publishing Company, New York, N.Y., 1976.
- [6] Marcel Berger. *Geometry I and II*. Springer-Verlag, 1987.
- [7] M. Bergstrom, J. Litton, L. Ericksson, C. Bohm, and G. Blomqvist. Determination of object contour from projections for attenuation correction in cranial positron emission tomography. *J. Comput. Assist. Tomography*, 6:365–372, 1982.
- [8] M Bertero, C. DeMol, and E.R. Pike. Linear inverse problems with discrete data: II. stability and regularisation. *Inverse Problems*, 4:573–594, 1988.
- [9] T. Bonnesen and W. Fenchel. *Theory of Convex Bodies*. BSC Associates, Moscow, Idaho, 1987.

- [10] Y. Bresler and A. Mackovski. Estimation of 3-d shape of blood vessels from x-ray images. In *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, March 1984.
- [11] M. H. Buonocore, W.R. Brody, and A. Macovsky. A natural pixel decomposition for 2-dimensional image reconstruction. *IEEE Trans. Bio. Engr.*, 28(2):69–78, 1981.
- [12] M. H. Buonocore, W.R. Brody, A. Macovsky, and S. Wood. A polar pixel kalman filter for limited data ct image reconstruction. *Proc. SPIE*, pages 109–115, August 1979.
- [13] Y. Censor. Finite series-expansion reconstruction methods. *Proc. IEEE*, 71(3):409–419, March 1983.
- [14] S. Chang. The reconstruction of binary patterns from their projections. *Comm. of the ACM*, 14(1), 1971.
- [15] S. Conte. *Elementary Numerical Analysis*. Mc Graw-Hill, 1965.
- [16] I.J.D. Craig and J.C. Brown. *Inverse Problems in Astronomy*. Adam Higler, 1986.
- [17] I. Csiszar. I-divergence geometry of probability distributions and minimization problems. *Annals of Probability*, 3:146–158, 1975.
- [18] P. J. Davis. Plane regions determined by complex moments. *Journal of Approximation Theory*, 19:148–153, 1977.
- [19] Phillip J. Davis. Triangle formulas in the complex plane. *Mathematics of Computation*, 18:569–577, 1964.
- [20] N.J. Dusaussoy and I.E. Abdou. The extended MENT algorithm: A maximum entropy type algorithm using prior knowledge for computerized tomography. *IEEE Trans. on Signal Processing*, 39(5):1164–1180, May 1991.

- [21] Harold E. Eggleston. *Convexity*. Cambridge University Press, 1969.
- [22] K. J. Falconer. A result on the steiner symmetrization of a compact set. *Journal of the London Mathematical Society*, 2(14):385–386, 1976.
- [23] K.J. Falconer. X-ray problems for point sources. *Proc. London Math. Soc.*, 3(46):241–262, 1983.
- [24] J. A. Fessler and A. Macovski. Object-based 3-d reconstruction of arterial trees from magnetic resonance angiograms. *IEEE Trans. on Medical Imag.*, 1991.
- [25] P. C. Fishburn, J. Lagarias, J. Reeds, and L. A. Shepp. Sets uniquely determined by projections on axes i. continuous case. *SIAM J. Appl. Math.*, 50(1):288–306, 1990.
- [26] R. Gardner. Sets determined by finitely many X-rays or projections. *Pubblicazioni dell'istituto di analisi globale e applicazioni*, 50, 1990. Serie “Problemi non ben posti ed inversi”.
- [27] R. J. Gardner. Symmetrals and x-rays of planar convex bodies. *Arch. Math.*, 41:183–189, 1983.
- [28] R.J. Gardner and P. McMullen. On hammer’s X-ray problem. *Journal of London Math. Society*, 2(21):171–175, 1980.
- [29] Walter Gautschi. Optimally conditioned vandermonde matrices. *Numer. Math.*, 24:1–12, 1975.
- [30] I. M. Gelfand, M. I. Graev, and Z. Y. Vilenkin. *Generalized Functions: Volume 5, Integral Geometry and Related Problems in Representation Theory*. Academic Press, 1966. Translated by: Eugene Saletan.
- [31] O. Giering. Bestimmung von eibereichen und einkörpern durch steiner-symmetrisierungen. *Sber. Bayer. Akad. Wiss. Munchen, Math. -Nat. Kl.*, pages 225–253, 1962.

- [32] G.H. Golub and C.F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, MD, 1983.
- [33] A. B. Goncharev. Three-dimensional reconstruction of arbitrarily arranged identical particles given their projections. *Translations of Mathematical Monographs*, 81:67–95, 1990.
- [34] I.S. Gradshteyn and I.M. Ryzhik. *Table of Integrals, Series, and Products*. Academic Press, 1979.
- [35] S.F. Gull and G.J. Daniell. Image reconstruction from incomplete and noisy data. *Nature*, 272:686–690, 1978.
- [36] P.C. Hammer. Problem 2. In *Proceedings of Symposia in Pure mathematics. Volume VII:Convexity*, 1963.
- [37] K. M. Hanson and G. W. Wecksung. Bayesian approach to limited-angle reconstruction in computed tomography. *Appl. Optics*, 24:4028–4039, December 1980.
- [38] G. Hardy, J.E. Littlewood, and G. Polya. *Inequalities*. Cambridge Mathematical Library. Cambridge University Press, second edition, 1991.
- [39] Sigurdur Helgason. *Radon Transform*. Birkhauser, Boston, 1980.
- [40] G. T. Hermann. *Image Reconstruction From Projections*. Academic Press, New York, 1980.
- [41] A. Hildebrand. *Introduction to Numerical Analysis*. McGraw-Hill, New York, 1956.
- [42] Yingbo Hua and Tapan Sarkar. Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise. *IEEE Trans. on ASSP*, 38(5), May 1990.

- [43] S. Van Huffel and J. Vandewalle. The use of total least squares techniques for identification and parameter estimation. In *Proc. 7th IFAC/IFORS Symp. on Identification and System Parameter Estimation*, pages 1167–1172, York, U.K., July 1985.
- [44] Edwin T. Jaynes. On the rationale of maximum entropy methods. *Proceedings of the IEEE*, 70(9):939–952, Sept 1982.
- [45] Lee K. Jones and Charles L. Byrne. General entropy criteria for inverse problems, with applications to data compression, pattern classification, and cluster analysis. *IEEE Transaction on Information Theory*, 36(1), 1990.
- [46] A. Kak and M. Slaney. *Principles of Computerized Tomographic Imaging*. IEEE Press, 1988.
- [47] William C. Karl. *Reconstructing Objects from Projections*. PhD thesis, MIT, Dept of EECS, 1991.
- [48] H.G. Kellerer. Masstheoretische marginal probleme. *Math. Ann.*, 153:168–198, 1964.
- [49] J.H. Kemperman. Sets of uniqueness and systems of inequalities having a unique solution. Preprint.
- [50] Robert A. Knox. *Ocean Circulation Models: Combining Data and Dynamics*, volume 284 of *NATO ASI Series*, chapter Ocean Acoustic Tomography: A Primer, pages 141–188. Kluwer Academic Publishers, 1989.
- [51] A. Kuba. Reconstruction of measurable plane sets from their two projections taken in arbitrary directions. *Inverse Problems*, 7:101–107, 1991.
- [52] A. Kuba and A. Volcic. Characterization of measurable plane sets which are reconstructable from their two projections. *Inverse Problems*, 4, 1988.
- [53] S. Kullback. *Information Theory and Statistics*. Wiley, New York, 1959.

- [54] Avinash Lele. Convex set estimation from support line measurements. Master's thesis, MIT, Dept. of EECS, 1990.
- [55] Jae S. Lim. *Two-Dimensional Signal and Image Processing*. Prentice Hall, 1990.
- [56] Marco Longinetti. Some questions of stability in reconstruction of plane convex bodies from projections. *Inverse Problems*, 1:87–97, 1985.
- [57] G.G. Lorentz. A problem of plane measure. *American Journal of Mathematics*, 71:417–426, 1949.
- [58] A. K. Louis. Picture restoration from projections in restricted range. *Math. Meth. in the Appl. Sci.*, 2:209–220, 1980.
- [59] D. Ludwig. The radon transform on euclidean space. *Comm. Pure Appl. Math.*, 19:49–81, 1966.
- [60] M. Luetzgen, W. C. Karl, and A. S. Willsky. Efficient multiscale regularization with applications to the computation of optical flow. *IEEE Transactions on Image Processing*, to Appear, January 1994.
- [61] S. MacLane and G. Birkhoff. *Algebra*. Chelsea Publishers, New York, 1988.
- [62] Jerrold E. Marsden. *Elementary Classical Analysis*. Freeman and Co., 1974.
- [63] Jerrold E. Marsden. *Vector Calculus*. Freeman and Co., 1981.
- [64] Geraldo Minerbo. MENT: A maximum entropy algorithm for reconstructing a source from projection data. *Computer Graphics and Image Processing*, 10:48–68, 1979.
- [65] Walter Munk and Carl Wunsch. Ocean acoustic tomography: a scheme for larger scale monitoring. *Deep-Sea Research*, 26A:123–161, 1979.
- [66] Frank Natterer. Computerized tomography with unknown sources. *SIAM J. Applied Math.*, 43(5):1201–1212, Oct. 1983.

- [67] Frank Natterer. *Mathematics and Computer Science in Medical Imaging*, volume F39 of *NATO ASI Series*, chapter Regularization Techniques in Medical Imaging, pages 127–141. Springer Verlag, 1988.
- [68] P. Oskoui-Fard and H. Stark. Tomographic image reconstruction using the theory of convex projections. *IEEE Trans. on Med. Imaging*, 7(1), March 1988.
- [69] H. Ouibrahim. Prony, pisarenko, and the matrix pencil: A unified presentation. *IEEE Trans. ASSP*, 37(1):133–134, January 1989.
- [70] Mirosław Pawlak. On the reconstruction aspects of moments descriptors. *IEEE Trans. Info. Theory*, 38(6):1698–1708, November 1992.
- [71] V. F. Pisarenko. The retrieval of harmonics from a covariance function. *Geophysics Journal: Royal Astronomical Society*, 33, 1973.
- [72] Jerry L. Prince. *Geometric Model-based Estimation from Projections*. PhD thesis, MIT, Dept. of EECS, 1988.
- [73] J. A. Reeds and L. A. Shepp. Limited angle reconstruction in tomography via squashing. *IEEE Trans. on Medical Imaging*, 6(2):89–97, June 1987.
- [74] M.L. Reis and N.C. Roberty. Maximum entropy algorithms for image reconstruction from projections. *Inverse Problems*, 8:623–644, 1992.
- [75] Jorma Rissanen. *Stochastic Complexity in Statistical Inquiry*, volume 15 of *Series in Computer Science*. World Scientific, 1989.
- [76] David J. Rossi. *Reconstruction from projections based on detection and estimation of objects*. PhD thesis, MIT, Dept. of EECS, 1984.
- [77] R. Roy, A. Paulraj, and T. Kailath. Esprit: A subspace rotation approach to estimation of parameters of cissoids in noise. *IEEE Trans. ASSP*, ASSP-34(5):1340–1342, Oct 1986.

- [78] R. Safae-Rad, K.C. Smith, B. Benhabib, and I. Tchoukanov. Application of moment and fourier descriptors to the accurate estimation of elliptical-shape parameters. *Pattern Recognition Letters*, 13:497–508, 1992.
- [79] Makato Sakai. A moment problem on jordan domains. *Proceedings of the American Mathematical Society*, 70(1):35–38, June 1975.
- [80] J. L. C. Sanz, E.B. Hinkle, and A.K. Jain. *Radon and Projection Transform-Based Computer Vision*. Springer Verlag, 1988.
- [81] Louis L. Scharf. *Statistical Signal Processing*. Addison Wesley, 1991.
- [82] R. O. Schmidt. *A Signal Subspace Approach to Multiple emitter Location and Spectral Estimation*. PhD thesis, Stanford University, 1981.
- [83] I.J. Schoenberg. *Approximation: Theory and Practice*. 1955. Notes on a series of lectures at Stanford University.
- [84] M. Ibrahim Sezan and Henry Stark. Incorporation of a priori moment information into signal recovery and synthesis problems. *Journal of Mathematical Analysis and Applications*, 122(172-186), 1987.
- [85] M.I. Sezan and H. Stark. Tomographic image reconstruction from incomplete view data by convex projections and direct Fourier inversion. *IEEE Trans. on Medical Imaging*, MI-3:91–98, June 1984.
- [86] J.A. Shohat and J.D. Tamarkin. *The problem of moments*. American Mathematical Society, New York, 1943.
- [87] J.E. Shore. Minimum cross-entropy spectral analysis. *IEEE Trans. Acoust. Speech, Signal Processing*, ASSP-29(2):230–237, Apr. 1981.
- [88] J.E. Shore and R.M. Gray. Minimum cross-entropy pattern classification. *IEEE Trans. on Pattern Anal. and Mach. Intell.*, PAMI-4(1):11–17, Jan. 1982.
- [89] J.E. Shore and R.W. Johnson. Properties of cross-entropy minimization. *IEEE Trans. on Info. Theory*, IT-27(4), July 1981.

- [90] Kennan T. Smith, Donald C. Solomon, and Sheldon L. Wagner. Practical and mathematical aspects of the problem of reconstructing objects from radiographs. *Bulletin of the American Mathematical Society*, 83(6):1227–1271, November 1977.
- [91] D.L. Snyder, T.J. Shulz, and J.A. O’Sullivan. Deblurring subject to nonnegativity constraints. Technical report, Washington University in Saint Louis, 1991.
- [92] N. Srinivasa, K.R. Ramakrishnan, and K. Rajgopal. Detection of edges from projections. *IEEE Transaction on Medical Imaging*, 11(1):76–80, March 1992.
- [93] V.N. Strakhov and M.A. Brodsky. On the uniqueness of the inverse logarithmic potential problem. *SIAM J. Appl. Math.*, 46(2):324–344, April 1986.
- [94] G. Szego. *Orthogonal Polynomials*. American Mathematical Society, Providence, Rhode Island, 1975.
- [95] G. Talenti. Recovering a function from a finite number of moments. *Inverse Problems*, 3:501–517, 1987.
- [96] M. R. Teague. Image analysis via the general theory of moments. *J. Opt. Soc. America*, 70:920–930, 1980.
- [97] C. H. Teh and R. T. Chin. On image analysis by the method of moments. *IEEE Trans. Patt. Anal. Machine Intell.*, 10:496–513, July 1988.
- [98] Jean-Philippe Thirion. Segmentation of tomographic data without image reconstruction. *IEEE Transactions on Medical Imaging*, 11(1):102–110, March 1992.
- [99] A.N. Tikhonov and V.Y. Arsenine. *Solutions of Ill-posed Problems*. Winston/Wiley, Washington, D.C., 1977.
- [100] Harry L. Van Trees. *Detection, Estimation, and Modulation Theory: Part I*. Wiley, 1968.

- [101] A. Volcic. A three point solution to hammer's x-ray problem. *Journal of the London Mathematical Society*, 2(34):349–359, 1986.
- [102] S.J. Wernecke and Larry D"Addario. Maximum entropy image reconstruction. *IEEE Transactions on Computers*, C-26, 1977.
- [103] S. L. Wood, A. Mackovski, and M. Morf. *Computer Aided Tomography and Ultrasonics in Medicine*, chapter Reconstruction with limited data using estimation theory, pages 219–233. North-Holland Publishing Co., 1979.
- [104] L. Zalcman. *Integral Geometry*, volume 63 of *Contemporary Mathematics*, chapter Some Inverse Problems of Potential Theory, pages 337–350. American Mathematical Society, 1984.