August 1994 **LIDS-TH-2260**

Research Supported By:

Army Research Office
Grant ARO DAAL03-92-G-0115

Air Force Office of Scientific Research
Grant AFOSR F49620-93-1-0604

Office of Naval Research
Grant ONR N00014-91-J-1004

Air Force Office of Scientific Research
Grant AFOSR E4920-92-0002

Wavelet Transform-Based Multi-Resolution Techniques For Tomographic Reconstruction And Detection

Mickey Bhatia

August 1994

LIDS-TH-2260

Sponsor Acknowledgments

Army Research Office
Grant ARO DAAL03-92-G-0115

Air Force Office of Scientific Research
Grant AFOSR F49620-93-1-0604

Office of Naval Research
Grant ONR N00014-91-J-1004

Air Force Office of Scientific Research
Grant AFOSR E4920-92-0002

Wavelet Transform-Based Multi-Resolution Techniques For Tomographic Reconstruction And Detection

Mickey Bhatia

This report is based on the unaltered thesis of Mickey Bhatia submitted to the Harvard-M.I.T. Division of Health Sciences and Technology, and the Department of Nuclear Engineering in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Radiological Sciences at the Massachusetts Institute of Technology in September 1994.

This research was conducted at the M.I.T. Laboratory for Information and Decision Systems with research gratefully acknowledged by the above mentioned sponsor(s).

Laboratory for Information and Decision Systems
Massachusetts Institute of Technology
Cambridge, MA 02139, USA

Wavelet Transform-Based Multi-Resolution Techniques For Tomographic Reconstruction And Detection

by

Mickey Bhatia

Submitted to the Harvard-M.I.T. Division of Health Sciences and
Technology, and the Department of Nuclear Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Radiological Sciences
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
September 1994

Author ....
Harvard-M.I.T. Division of Health Sciences and Technology, and the Department of
Nuclear Engineering
August 5, 1994

Certified by ....

Alan S. Willsky

Professor, Electrical Engineering and Computer Science

Thesis Supervisor

Certified by ....

William C. Karl

Research Scientist, Laboratory for Information and Decision Systems

Thesis Supervisor

Accepted by ....
Roger G. Mark
Director, Harvard-M.I.T. Division of Health Sciences and Technology

Accepted by ....
Allan F. Henry
Chairman, NED Committee on Graduate Students

Wavelet Transform-Based Multi-Resolution Techniques For Tomographic Reconstruction And Detection

by

Mickey Bhatia

Submitted to the Harvard-M.I.T. Division of Health Sciences and Technology, and
the Department of Nuclear Engineering
on August 5, 1994, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Radiological Sciences

Abstract

In this thesis we have presented multi-resolution approaches to the problem of reconstruction and detection of objects from tomographic measurements. Such problems are commonly encountered in many applications including medical imaging, non-destructive testing and evaluation, astronomy, geophysics, and oceanography. We have focused our efforts mainly on the case where the tomographic data are sparse and noisy. In this case the Radon transform derived results based on continuous and noise-free data break down and the reconstruction and detection tasks become more challenging. We have demonstrated the effectiveness of the multi-resolution framework for regularization of such ill-posed reconstruction and detection problems, and in particular have developed highly efficient and, at the same time statistically optimal multi-resolution-based algorithms for reconstruction and detection of objects from both sparse as well as noisy data. The specific contributions of the thesis are:

- A framework for multi-resolution representation of objects *matched* to tomographic data
- A *fast*, iterative multi-resolution reconstruction method in the above framework
- Statistical prior models for the object constructed directly in the multi-resolution data-domain that lead to efficient regularized reconstructions that are *no more complex* than the corresponding unregularized reconstructions
- A method for obtaining multi-resolution reconstructions from dense data with the *same* computational complexity as the highly efficient filtered back-projection method
- A *fast* method for discrimination of fractal fields directly from sparse and noisy data
- A *fast* data-domain method for detection of anomalies superimposed on a fractal-textured background

Thesis Supervisor: Alan S. Willsky
Title: Professor, Electrical Engineering and Computer Science

Thesis Supervisor: William C. Karl
Title: Research Scientist, Laboratory for Information and Decision Systems

Acknowledgments

First of all I thank my thesis supervisors Alan Willsky and Clem Karl. Not only are these individuals one of the brightest people I have ever met, but they are also one of the coolest. This thesis would not have been possible without the guidance I received from both Alan and Clem. I also thank Alan for showing enough confidence in me by taking me on as a graduate student and for keeping me on an uninterrupted supply of money. I hope to repay him one day by buying him the season tickets for the New York Knicks, the great team that he cherishes the most. I also thank my thesis committee members Jacquelyn Yanch and Gordon Brownell for their input and support.

This thesis benefited substantially from the technical and moral support provided by my colleague and officemate Eric Miller. Eric is extremely bright and at the same time is a great person, and I am fortunate that I got a chance to interact with him. I thank Meryl for keeping Eric a bit sane, and also for providing the animal crackers that Seema ate. Despite this, Seema (with two e's) Jaggi is a wonderful person and I thank her for the support she provided me, both technical and moral. I am thankful to Paul Fieguth for his friendship, technical help and also for the stuff that he baked for me. I thank Hamid (pronounced as Haymid) Krim, Bill Irving, Peyman Milanfar, and Sekhar Ramanathan for the many technical discussions I had with them. I also thank Lori Belcastro for being a good friend, and Dharmendar Singh for all his help and support.

Finally, I thank my family back home in India, especially my mom Raj Bhatia, for having to suffer through my not being close to her for these many years. I am sure she would have suffered more if I was close to her, but brain-washed as these Indian mothers are, she would probably not buy this.

This thesis is dedicated to the memory of my father, the late Dr. R. C. Bhatia.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In this thesis we present a multiresolution approach to the problem of reconstruction and detection of objects from tomographic measurements. Such problems are commonly encountered in many applications including medical imaging, non-destructive testing and evaluation, astronomy, geophysics, and oceanography [39, 44]. The purpose of this chapter is to present an outline of the thesis, and also to highlight the main contributions made by this thesis to the field of tomography. The chapter is organized as follows. In Sections 1.1 and 1.2 we briefly state the problem of reconstruction and detection of objects from tomographic data. We mention a few typical applications where these problems arise, and describe the existing, commonly used methods to solve these problems. In Section 1.3 we discuss the factors that motivated us to develop a multiresolution framework to tackle such problems. We then highlight the main contributions of our research in Section 1.4. This is followed by an outline of the thesis in Section 1.5 where we give a brief description of the following chapters. Finally, we devote Section 1.6 to differentiate our work from other related work in multiresolution tomography.

1.1 Object reconstruction from tomographic data

The problem of reconstructing an object from tomographic measurements is also referred to as *tomography* [8,39,44,55]. Here the goal is to reconstruct the cross-section of an object given measurements that are *strip-integrals* of some property of the object. For example, in *transmission tomography* using X-rays, the measurements consist of integrals of the attenuation coefficient of the object for X-rays along strips which represent the path of the X-rays through the object. In the case of *emission tomography*, the measurements are the strip-integrals of the concentration of some radioisotope that has been introduced into the object prior to imaging. If the strips over which the tomographic data have been integrated are thin (and thus are approximated by line integrals), then the data are said to be the *Radon transform* of the object [39,44]. Further, the strip-integral data at different locations, but at the same orientation, are collectively referred to as a *projection*.

The need for tomography arises in a variety of applications. For example, in the medical field, tomography provides a non-invasive method for detecting both anatomical and functional abnormalities within the human body [44]. In an example of non-destructive testing and evaluation application, tomography is used to inspect materials for defects both during and after production [1,7,20,25,44,46,47,71,73]. In another example, the age of a tree may be non-destructively evaluated through tomography [62]. In oceanography, the temperature of the various layers of the ocean is determined by acoustic tomography [48,61]. In general, tomography can be classified into two categories depending on whether the strips over which the tomographic data have been integrated are straight or curved [44]. In *diffraction tomography*, for example using acoustic waves [23,56,60], the strips are curved and are defined by the path of the probing entity. In this thesis, however, we only concern ourselves with a second category of tomography, namely *non-diffraction tomography*, in which the strips are straight. Examples of this latter category include the commonly used imaging modalities like CT, MRI, PET, and SPECT [39,44]. Further, we limit ourselves to the reconstruction of 2-D cross-sections from 1-D projections.

Next, we turn our attention to the conventional reconstruction techniques for tomography. In general, the quality of the reconstruction in tomography is a function of the quality and the quantity of the strip-integral data [8,16,45,65,77]. If the data are noise free and dense (i.e. the strips are thin and the projections cover the object at a large number of different angles) then fast algorithms based on the exact Radon inversion formula, like the *filtered back-projection* (FBP), can be used to reconstruct the object.

The FBP [8, 39, 44, 51] is the most commonly used method for image reconstruction from high quality and dense tomographic data. This reconstruction technique is based on an important result for Radon transform, which states that the reconstruction obtained by back-projection of the projection data is a blurred version of the object, with the blurring function given by $1/|\omega|$ in the frequency domain. Thus the object can be reconstructed by applying a 2-D $|\omega|$ filter on the back-projected image. Due to another property of the Radon transform, this reconstruction is identical to the one obtained by first filtering the projection data at all angles by a 1-D $|\omega|$ filter, and subsequently back-projecting these filtered projections. This latter reconstruction technique is referred to as the FBP, and the 1-D filtering operation on the projection data is commonly known as the *ramp-filtering operation*. As mentioned previously, a major limitation of the FBP technique is that it fails to produce reasonable reconstructions if the projection data are sparse and/or noisy.

If the projection data are sparse, then one has to resort to matrix based reconstruction methods [52–54]. These techniques require more computations than the FBP since solutions of very large, generally ill-conditioned systems of equations have to be computed. A first example of such matrix based reconstruction techniques is the *algebraic reconstruction technique* (ART). In ART [37–40, 44, 75], the object is expanded in the standard rectangular pixel basis. The coefficients of expansion of the object in this basis are related to the strip-integral data through a matrix which is sometimes referred to as the *projection matrix*. The projection matrix, although very large, is *sparse* because each strip intersects only a small number of object pixels. This sparsity is exploited by a *Kaczmarz-like method* which is utilized to solve the matrix equation for the object expansion coefficients. The Kaczmarz method is an iterative method for solving a system of equations where an initial guess is cyclically projected[1]

---

[1] This should not be confused with the term *projection* that is frequently used (including in this thesis) for tomographic data.

onto the hyperplanes represented by the rows of the system matrix (the projection matrix in this case). If the system of equations is consistent, then the Kaczmarz method converges to the minimum norm solution given by the pseudo-inverse of the system of equations [75]. The computational complexity of the Kaczmarz method is proportional to the number of non-zero elements in the system matrix, and the convergence depends critically on the order in which the rows of the system matrix are accessed, and also on the choice of some relaxation parameter (to be explained later). These choices have a substantial effect on the convergence because the projection matrix is quite ill-conditioned. In contrast, if the system matrix were well-conditioned, then the ordering scheme and the relaxation parameter would have little effect on the convergence of the Kaczmarz algorithm.

A second matrix based method for sparse data tomography problem is the *natural pixel* (NP) method [9,10,24]. Here the object is expanded in a basis using the same strips along which the projection data are collected, resulting in object expansion coefficients that reside in the projection domain. This is in comparison to ART where the expansion coefficients (i.e. object pixel values) are defined in the object space. As a result, the NP representation is free of the model errors which arise in ART due to the rectangular pixel representation. This benefit comes with a cost however: whereas the projection matrix in ART is sparse, the corresponding system matrix in the NP method, relating the object expansion coefficients and the projection data, is *full*. This is because the elements of the NP system matrix are the areas of intersection of the data strips, and most of these are non-zero. Thus if the Kaczmarz method is used to solve the system of equations arising in the NP representation, it results in a computationally intensive reconstruction algorithm. Besides, as in ART, the NP matrix is also quite ill-conditioned which implies that the usual problems of choosing an appropriate ordering scheme and the relaxation parameter are encountered if the Kaczmarz method is used.

Finally, if the projection data are noisy, the standard reconstruction techniques (FBP, ART, and NP) described above yield unacceptable results. This generally reflects the fact that more degrees of freedom are being sought in the reconstruction than are really supported by the data, and hence some form of regularization is required [3, 21, 27, 28, 33, 34, 36, 39, 44, 49, 50, 55, 69]. Conventionally, this problem of reconstruction from noisy projection data is regularized by one of the following techniques. First, the FBP ramp filter may be rolled off at high frequencies thus attenuating high frequency noise at the expense of not reconstructing the fine scale features in the object. This results in a fast, though ad hoc, method for regularization if we have dense data. Similarly, in ART, a fast, ad hoc form of regularization is achieved by stopping the iterations prematurely, before an excessive fitting of the solution to the noisy data occurs. In contrast to the above two ad hoc approaches, an alternate form of regularization is to solve for a maximum aposteriori probability (MAP) estimate [76] of the object based on a 2-D (spatial) Markov random field (MRF) prior model. This latter approach results in a statistically regularized reconstruction which allows the inclusion of prior knowledge in a systematic way, but leads to optimization problems which are extremely computationally intensive.

1.2 Object detection from tomographic data

In some applications using tomographic data, the goal is not to reconstruct the entire cross-section, but is to simply detect the presence of an object. Example of such applications include the detection of cracks and other defects in materials [25, 73], and the detection of lesions in medical images [30]. The conventional methods for detecting objects from the projection data fall under two broad categories. In the first category, the entire cross-section is first reconstructed using any of the conventional techniques, and then post-processing is done on the reconstructed image to detect the object (for example as in [30]). Leaving the computational issues aside, a major disadvantage of this approach is that when the projection data are sparse and/or noisy, the reconstruction of a complete image generally suffers from severe artifacts and distortion that can easily obscure the anomaly or result in false alarms. These reconstruction artifacts can be removed by appropriate regularization at the expense of additional computational burden. This approach can, however, smooth out the anomaly and hence once again make it undetectable. In the second category of detection methods [66], these problems associated with the reconstruction of the entire field are avoided by developing detection techniques directly in the data-domain. These methods, however, assume that the objects are parameterized (for example by size, shape, location, and number) and thus are limited in scope.

We also include another application in this category of object detection, in which a fractal object is to be detected based on it's fractal dimension. This application is motivated from the medical field, where a change in fractal dimension is used to differentiate between normal and abnormal conditions in many different contexts including diagnosis of liver abnormalities [11]. Conventionally, for this application, the fractal dimension is calculated by estimating the slope of the power spectrum of the averaged projection data from all angles. Again, a major disadvantage of this approach is that it may give inaccurate results when the projection data are incomplete and/or noisy, and a more nearly optimal method may do better.

1.3 Motivation for a multiresolution framework for reconstruction and detection

We next discuss the factors that motivated us to develop a multiresolution[2] approach for solving the problems of reconstruction and detection of objects from tomographic data. In general, a multiresolution framework is desirable for the solution of these problems for a variety of reasons.

First, the multiresolution framework is a natural choice for regularizing ill-posed reconstruction problems where the projection data are incomplete and/or noisy. If these ill-posed problems are solved at the finest resolution, the resulting solution suffers from severe artifacts because of the attempt to reconstruct more degrees of freedom than are supported by the data. However, a reconstruction at a coarser scale results in an implicitly regularized solution, and is obtained at no additional

---

[2] In this thesis we use the terms *multiresolution* and *multiscale* interchangeably with the understanding that high resolution corresponds to fine scales.

cost in a multiresolution framework for image reconstruction. Further, the multiscale framework lends itself to simple statistical prior models for the object, constructed directly in the multiscale domain, that capture the intuition that for ill-posed reconstruction problems the lower resolution reconstructions are more reliable that their higher resolution counterparts. Later we will see that not only are these models conceptually simple, but they also result in fast explicit regularization techniques. In fact, the MRF prior models mentioned earlier in context of conventional regularization, attempt to impose a similar smoothness constraint in the object space. However MRF models result in reconstruction algorithms that are much more computationally complex than our regularized reconstruction methods.

A second motivating factor for using the multiscale framework is that the projection data under consideration may be naturally acquired at multiple resolutions, e.g. if data from detectors of differing resolutions are used. In fact this scenario is common in medical imaging where multi-modality data, for example from CT, MRI and nuclear medicine, are often used to obtain the best information about the object. A specific example of this is the case where high resolution information from anatomical images is used to improve the reconstruction of low resolution functional images [31]. Further, in addition to the data, the phenomenon under investigation may itself be naturally multiscale. For example, in the medical field self-similar or fractal models have been effectively used for the liver and lung [11, 12, 78].

In addition, it may be that, even if the data and phenomenon are not multiscale, our ultimate objectives are multiresolution in some way. For example, even though our data may be acquired at a fine level we may actually only care about aggregate or coarse scale quantities of the field. Such is often the case in ocean acoustic tomography or functional medical imaging. Conversely, if we are only interested in imaging high frequency details within the object (for example, boundaries) [72], then we may wish to directly obtain these features by extracting only the finer scale information in the data. Or, indeed, it may be that we want to use different resolutions in different areas – e.g. in nondestructive evaluation of aircraft we may want to look for general corrosion over an entire plane, but focus attention on certain suspect rivets to look for cracks [1]. Using conventional techniques we would first have to reconstruct the entire field and then use post-processing to extract such features.

A final compelling motivation for the use of multiresolution methods is that they lead to extremely efficient statistically-optimal algorithms for object detection directly from the projection data. As we show later, this is due to the fact that the data covariance matrices corresponding to an important class of (object) background texture models, become extremely sparse in the multiscale framework.

1.4 Main contributions of the thesis

In this section we highlight the significant contributions made by the thesis to the problem of reconstruction and detection of objects from tomographic data.

The first contribution of this thesis is the development of extremely efficient techniques for multiresolution object reconstruction from both dense and sparse tomographic data. Specifically, for dense data, the reconstructions at multiple resolutions are obtained with the same computational complexity as the highly efficient FBP; however in contrast to FBP, our algorithms also provide reconstructions at multiple resolutions and remain just as efficient when they are regularized to account for noise. In the case of sparse data, these multiresolution reconstructions are obtained much faster than the conventional sparse data reconstruction techniques, namely the ART and NP approaches.

The second contribution of this thesis is to present an iterative technique for reconstruction from sparse data, which is an alternative to the conventional ART. This technique produces reconstructions which are insensitive to the choice of the ordering scheme and the relaxation parameter. The latter are critical for convergence in ART and have to be carefully selected. Further, since the iterations in our technique are performed in a multiscale extension of the NP domain, the resulting reconstruction is free of the model errors which arise in ART due to the rectangular pixel representation. Finally, the reconstructions at multiple scales can be accessed from our iterative technique at no additional cost.

The third major contribution of this thesis is the development of a particular class of stochastic prior models for ill-posed reconstruction problems. These prior models are specified directly in the multiscale projection domain and result in extremely efficient estimation algorithms. In particular, our regularized reconstructions based on these prior models are obtained with the same computational complexity as the unregularized reconstructions. Also, for the case of dense data, the regularization using the above class of prior models is equivalent to a particular way of rolling-off of the ramp filter. This provides a statistical interpretation to the ad hoc rolling-off of the ramp filter in the FBP, and also provides a recipe for designing statistically-optimal filters to provide such roll-off.

The final major contribution of this thesis is the development of efficient, multiscale data domain techniques for object detection. We consider both of the detection problems mentioned earlier, namely the problem of detecting the presence of an object or anomaly in a textured background and also the problem of discriminating a fractal object based on it's fractal dimension. In the first case, we concern ourselves with identifying anomalies superimposed on a fractal background. We develop fast, statistically-optimal techniques for zooming-in on the anomaly. In the second case, we devise a method that is able to discriminate between two fractal fields with different fractal dimensions, from sparse data even in the presence of substantial amount of measurement noise. This discrimination is based on fast likelihood calculations for multiscale models for the hypothesized fields.

1.5 Thesis outline

The thesis is organized as follows. In Chapter 2 we give an overview of the background material that is essential for the development of our multiresolution framework for reconstruction and detection. We start by mathematically formulating the tomography problem and then present the theory underlying the conventional reconstruction techniques, namely FBP, ART, and NP. Subsequent to this, we present some results from estimation theory that we use extensively in the following chapters. After this we introduce the concept of wavelet transform-based multiresolution representations of signals and, finally, we close the chapter by describing other related work in multiresolution tomography.

In Chapter 3 we develop our multiresolution framework for tomographic reconstruction. We do this by building on the NP representation through a multiscale transformation of the NP strips. As mentioned previously, the standard NP system matrix, relating the object expansion coefficients and the projection data, is *full.* The use of wavelet bases results in a transformation matrix which is *sparse*. In addition, the coarsest scale elements of this matrix capture any ill-conditioning arising from the geometry of the imaging system. We exploit this feature to partition the multiscale system matrix by scales and obtain a reconstruction procedure that only requires inversion of a well-conditioned and sparse matrix. Moreover, the different scale components of the proposed multiscale reconstruction method induce a corresponding multiscale representation of the underlying object and, in particular, provide estimates of (and thus information about) the field or object at a variety of resolutions at *no additional cost.* This also provides a natural framework for explicitly assessing the *resolution-accuracy tradeoff* which is critical in the case of noisy data. In contrast to the standard NP method, we are able to extend our multiscale reconstruction technique to the case of noisy projections to obtain a statistically regularized, multiscale MAP object estimate. We do this by realizing that for ill-posed problems the lower resolution (i.e. the coarser scale) reconstructions are often more reliable than their higher resolution counterparts and by using prior statistical models constructed *directly in the multiscale domain* which capture such intuition. Our multiscale MAP regularized reconstructions are no more computationally intensive than our unregularized multiscale reconstructions. Finally, we exploit the sparsity of the multiscale system matrix to develop an extremely efficient method for reconstructing the fine scale features (for example, edges and boundaries) in the object.

In Chapter 4 we concern ourself with the development of an efficient reconstruction algorithm that takes advantage of the sparsity and conditioning of the multiscale system matrix from Chapter 3. Specifically, we use the Kaczmarz method to solve the sparse and well-conditioned system of equations that arises in our multiresolution NP framework. We refer to the resulting reconstruction technique as the multiscale projection-domain ART (MPART) and present it as an alternative to the conventional ART. We show that the MPART reconstruction is relatively insensitive to the ordering scheme and the relaxation parameter, which are critical for convergence in ART. We further show that, even though the sparsity of the multiscale system matrix and the projection matrix of ART are similar, the smaller magnitude elements in the former can be thresholded to zero to obtain reconstructions that are a reasonable approximation to the true reconstruction. These approximate MPART reconstructions are obtained with substantially less computations and storage than the exact reconstructions. We show that similar approximate ART reconstructions obtained by thresholding the smaller elements in the projection matrix suffer from severe artifacts. Finally, in the case of noisy projection data, we develop an extension to MPART that computes the MAP estimate from Chapter 3. This MAP estimate is obtained with essentially the same computational and storage complexity as the unregularized MPART reconstruction.

In Chapter 5 we specialize our multiscale framework from Chapter 3, which is geared towards sparse data, to yield fast reconstructions from dense data. We do this by operating on the ramp filter matrix of the FBP rather than the NP system matrix. As it turns out, the matrix representation of the resulting multiscale ramp filtering operator is approximately diagonal. This enables us to formulate an efficient multiscale tomographic reconstruction technique from complete data where the reconstructions at multiple scales are obtained with the same computational complexity as that of the FBP reconstruction method. Further, in the case of noisy but dense data, we show that the MAP estimate developed in Chapter 3 is equivalent to a rolling-off of the FBP ramp filter in a statistically-optimal manner. Again, this MAP solution is obtained with the same computational complexity as the FBP method.

In Chapter 6 we concern ourself with the likelihood discrimination of fractal-textured objects with different fractal dimensions. We show that the data covariance matrix corresponding to these objects is extremely sparse in the multiscale domain, thereby leading to fast likelihood calculations. Further, we use the recently introduced class of multiscale stochastic models defined on trees to realize accurate approximations of the data covariance matrices corresponding to fractal fields. These tree-like models have the advantage that they lead to fast likelihood calculations. We show that the resulting likelihood method is able to discriminate fractal fields from sparse data even in the presence of a substantial amount of noise.

In Chapter 7 we consider the problem of detecting anomalies that are superimposed on a fractal-textured background. Rather than solving this problem in the object space, we tackle this problem in the multiscale data domain. We first calculate the data covariance matrix corresponding to the fractal background, and then perform a chi-square test on the data to detect anomalies. We show that, since the multiscale domain results in an extremely sparse fractal-field data covariance matrix, a fast way of localizing the anomaly to a few candidate pixels is to perform a chi-square test on the data by using only the diagonal elements of this covariance matrix. Once the candidate pixels have been identified, a maximum-likelihood (ML) estimate of the field intensity is obtained at only these candidate pixel locations. This results in an extremely fast and, at the same time, a nearly statistically-optimal method for zooming-in on the anomaly.

We conclude the thesis in Chapter 8, in which we also identify a few topics for future research.

1.6 Other work related to this thesis

Wavelets have been recently applied for tomographic image reconstructions by other researchers as well. Our work contrasts with these other multiscale tomography approaches which either concentrate on the complete data tomography problem [64,67] or assume prior knowledge of the object edges to reconstruct an object from incomplete data [68]. In addition, in the approaches [64,67,68] the object is expanded in a 2-D wavelet basis for the original spatial domain and the resulting coefficients of this expansion are then calculated from the projection data. In contrast, in our multiscale approach which is fundamentally based on the incomplete data NP framework, a 2-D multiscale representation is obtained by a *1-D wavelet expansion* of the strips. This has the advantage that our multiscale basis representation of the object is closer to the measurement domain than other multiscale representation techniques. One consequence is that our algorithms for multiscale reconstruction are extremely efficient. Another consequence is that our framework also allows for the simple and efficient solution of statistically regularized problems at no additional cost when the projection data are noisy.

Besides the above mentioned work, DeStefano and Olson [22], and Berenstein and Walnut [5] have also used wavelets for tomographic reconstruction problems, in particular to localize the Radon transform in even dimensions. Through this localization the radiation exposure can be reduced when a local region of the object is to be imaged. Our multiresolution framework for tomography contrasts with the work in [22] and [5] in that the latter approaches expand the object in a 2-D wavelet basis and, further, do not provide a framework for multiscale reconstruction, which is the central theme here. In addition, our reconstruction procedure also localizes the Radon transform, though we do not stress this particular application in this thesis.

Chapter 2

Preliminaries

In this chapter we present the background material that is essential for understanding the development of our multiresolution framework for reconstruction and detection of objects from tomographic data. The chapter is organized as follows. In Section 2.1 we present a mathematical formulation of the image reconstruction problem in tomography. Following this, in Section 2.2 we describe a few conventional methods for image reconstruction, namely the filtered back-projection (FBP) reconstruction technique, the algebraic reconstruction technique (ART), and the natural pixel (NP) method. In Section 2.3 we state relevant results from detection and estimation theory. This is followed by Section 2.4 where we describe the commonly used methods for obtaining regularized reconstructions from noisy data. In Section 2.5 we present the wavelet-transform based technique for multiresolution representation of signals. Finally, in Section 2.6 we summarize other existing work in multiresolution tomography. Sections 2.A and 2.B contain certain technical details.

2.1 Tomographic measurements

In tomography the goal is to reconstruct an object or a field, f, from strip-integral data [8,39,44,55]. For simplicity, in this paper we only consider the 2-D parallel-beam imaging geometry. Here the data consists of parallel, non-overlapping strip-integrals through the object at various angles (refer to Figure 2-1). Each angular position corresponds to a specific source-detector orientation. Suppose we have N_θ positions between 0^0 and 180^0 and N_s parallel strip-integrals at each angular position. Let us label the observation corresponding to strip ℓ at angular position k by $y(m)$, where $m = (k-1)N_s + \ell$. Furthermore, let $S_m(u,v)$ be the indicator function of the strip-integral corresponding to this observation so that S_m has value one within that strip and zero otherwise. Given this notation, the measurements without noise are given by:

$$y(m) = \iint_{\Omega} f(u,v)\, S_m(u,v)\, du\, dv, \qquad m = 1, \ldots, N_\theta N_s \tag{2.1}$$

where (u,v) are the usual rectangular spatial coordinates, and the integration is carried over a region of interest Ω which represents the field-of-view of the imaging system. In tomographic literature, the term *projection* is commonly used to refer to the collection of strip-integral data at the same angular position. Thus the projection y_k at angle k is defined by the following N_s vector[1]:

$$y_k = \begin{bmatrix} y((k-1)N_s + 1) \\ y((k-1)N_s + 2) \\ \vdots \\ y((k-1)N_s + N_s) \end{bmatrix}. \tag{2.2}$$

To summarize, in tomography the image reconstruction problem amounts to solving for the object $f(u,v)$ from (2.1), given the strip-integral data $\{y(m), m = 1, \ldots, N_\theta N_s\}$, or alternatively the projection data $\{y_k, k = 1, \ldots, N_\theta\}$.

2.2 Conventional reconstruction techniques

We turn our attention next to a few commonly used techniques for image reconstruction from tomographic data. We first describe two matrix-based techniques for reconstruction from sparse data, namely the ART and the NP reconstruction method.

---

[1] In subsequent sections we will use the same notation y_k to denote the measurements in (2.2) with additive measurement noise.

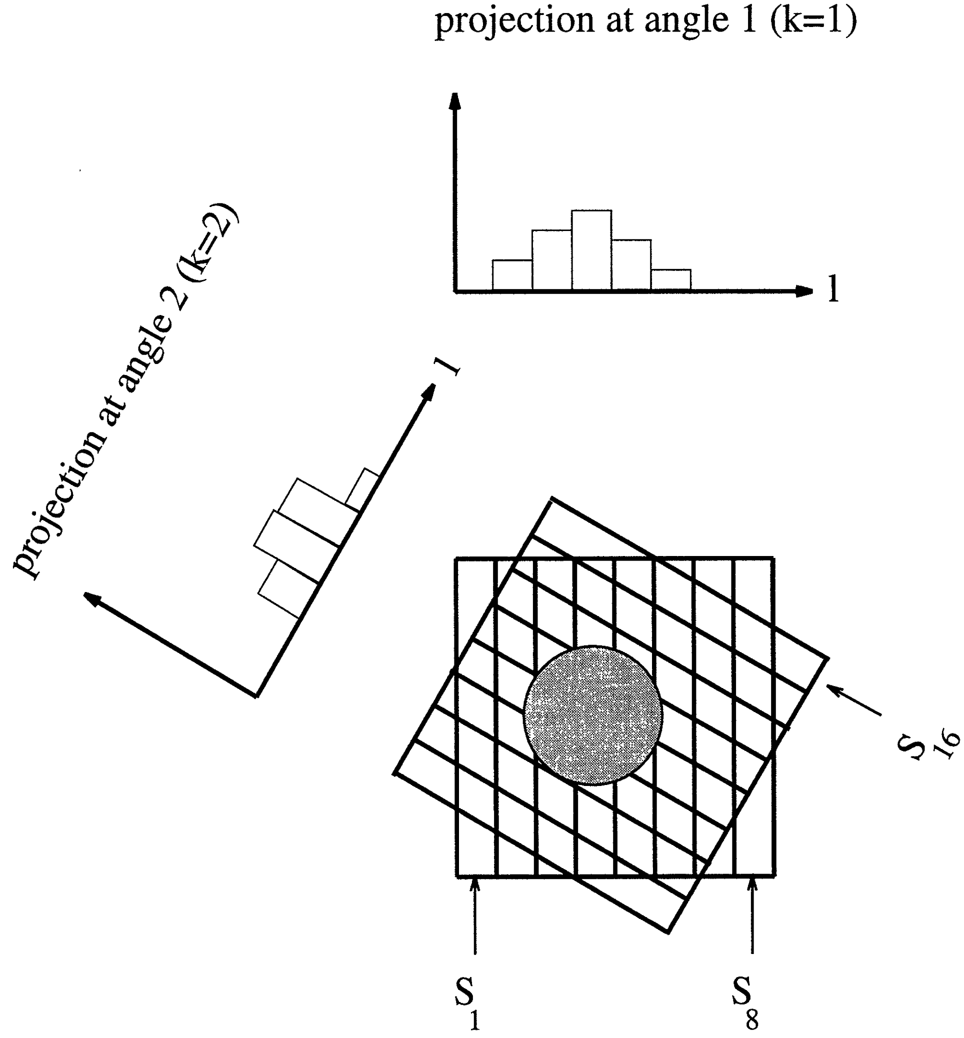


Figure 2-1: The strip-integral measurements of an object (shaded) at two different angular positions ($k = 1$ and $k = 2$). Also shown are three basis functions, S_1, S_8, and S_{16}, which are the indicator functions of the corresponding strips. Each angular projection is composed of $N_s = 8$ strips in this example.

We then move on to the FBP technique which yields accurate reconstructions only from dense data.

2.2.1 The algebraic reconstruction technique (ART)

In ART [37–40, 44], the object $f(u,v)$ is expanded in the standard rectangular pixel basis:

$$f(u,v) = \sum_{n=1}^{N_p} f(n) b_n(u,v) \tag{2.3}$$

where $b_n(u,v)$ is the indicator function of pixel n, $f(n)$ is the coefficient in this expansion of the object corresponding to basis function $b_n(u,v)$, and N_p are the total number of pixels used in the object representation. For simplicity, we assume that the object is represented on a $N_s \times N_s$ square pixel grid (where recall that N_s is the number of parallel strips in each angular projection), and hence $N_p = N_s^2$. Now by substituting the ART object representation equation (2.3) into the strip-integral data expression (2.1), we obtain:

$$y(m) = \sum_{n=1}^{N_s^2} T_{mn} f(n), \qquad m = 1, \ldots, N_\theta N_s \tag{2.4}$$

with

$$T_{mn} = \iint_{\mathbf{\Omega}} b_n(u,v)\, S_m(u,v)\, du\, dv, \qquad m = 1, \ldots, N_\theta N_s;\ n = 1, \ldots, N_s^2, \tag{2.5}$$

where $\mathbf{\Omega}$ is the field-of-view of the imaging system. Finally, by collecting together the data in (2.4), we get the following overall observation equation:

$$y = Tf, \tag{2.6}$$

where the overall data vector y has length $N_\theta N_s$, f is a N_s^2 vector containing the object expansion coefficients $\{f_n, n = 1, \ldots, N_s^2\}$, and T is a $N_\theta N_s \times N_s^2$ *projection matrix*, the (m,n)-th element of which is given by T_{mn} (see (2.5)). The tomography problem thus reduces to solving for the object coefficients f from (2.6), given the data

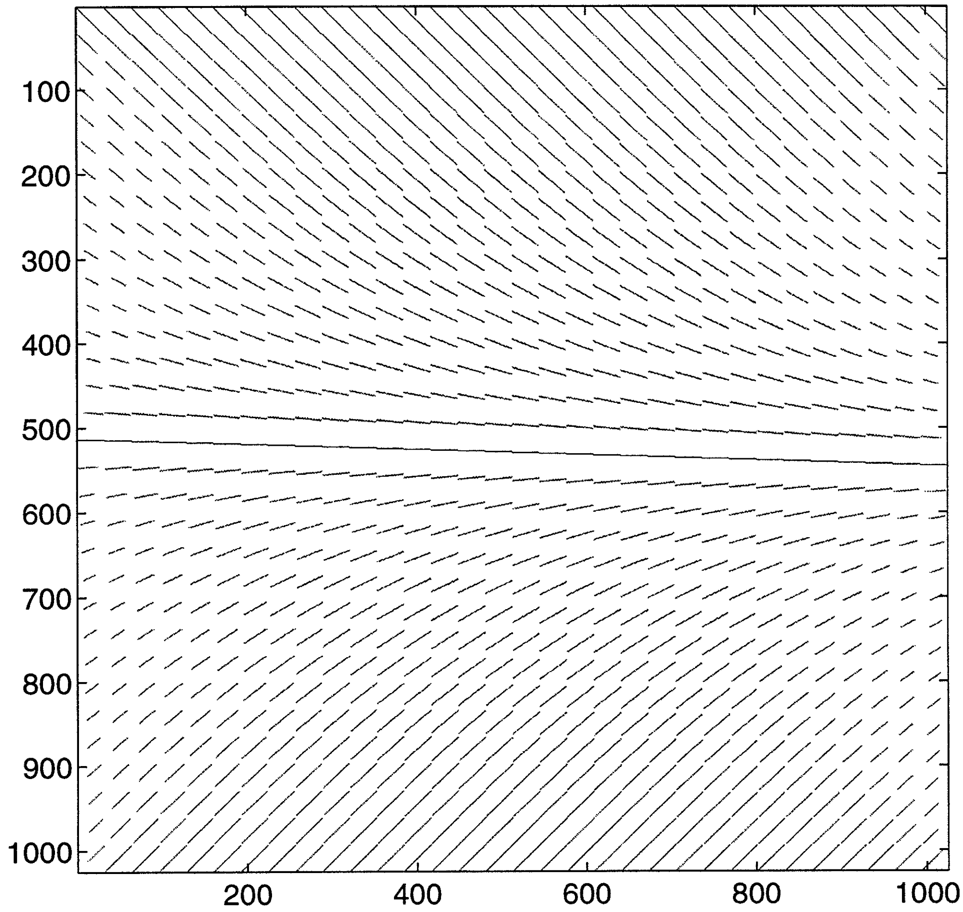


Figure 2-2: The grayscale plot (black corresponds to the maximum value and white to the minimum) of the projection matrix T for an imaging geometry defined by $N_\theta = N_s = 32$. The matrix is of size 1024 × 1024. Note that T is sparse.

y which are gathered along strips in T. We refer the reader to Figure 2-2 where we show the 1024 × 1024 projection matrix[2] T corresponding to an imaging geometry with $N_\theta = N_s = 32$. Note that the projection matrix, although very large, is *sparse* because each strip intersects only a small number of object pixels.

This sparsity is exploited by a Kaczmarz-like[3] projection[4] method that is used to solve the system of equations (2.6) for the object coefficients f. This method consists of cyclically projecting an initial guess for f, say $f^{(0)}$, onto the hyperplanes

---

[2] Refer to Section 2.A for details on the formation of the projection matrix T.

[3] The standard Kaczmarz method has no provision for relaxation. The variant used by ART, however, incorporates relaxation. For clarity, we will refer to the latter simply as the Kaczmarz method with the understanding that it includes relaxation.

[4] This is not to be confused with the term projection that is frequently used (including in this thesis) for tomographic data.

defined by the observation equations (2.4). In the first iteration all the $N_\theta N_s$ observation equations are cycled through, and the second iteration is started by projecting the result of the first iteration onto the hyperplane defined by the first observation equation $y(1) = \sum_n T_{1n} f(n)$. This process is repeated until convergence is achieved. Specifically, the following recursion is used for calculation of the object coefficients:

$$f^{(i+1)} = f^{(i)} + r^{(k)}\left(\frac{y(k) - T(k)f^{(i)}}{T(k)T(k)^T}\right)T(k)^T, \qquad k = (i)\mathrm{mod}(\mathrm{N}_\theta \mathrm{N}_\mathrm{s}) \tag{2.7}$$

where $f^{(i)}$ is the ith estimate of f, $T(k)$ refers to the k-th row of T, and $r^{(k)}$ is the relaxation parameter for the k-th iteration[5], with $0 \leq r^{(k)} \leq 2$. Once convergence is achieved for the solution of the object coefficients f, the reconstruction is obtained from (2.3). It can be easily seen from (2.7) that the computational complexity per iteration of ART is proportional to the number of non-zero elements in the matrix T. Also, since only a row of T is required for each step of the recursion (2.7), the Kaczmarz method results in a storage efficient algorithm [39].

We discuss next the parameters in the ART recursion (2.7) that have significant influence on the rate of convergence of the solution f [40]. The first choice is of the ordering scheme used for rows of T in the recursion (2.7). As (2.7) currently stands, the rows of T are accessed in a *sequential* order. However, if at each step the row is accessed that has the largest angle to the subspace spanned by all the rows of T used in the previous steps of the same iteration, the convergence can be improved. This is exactly the motivation behind a recent ordering scheme proposed by Herman and Meyer [40]. However, an improved convergence is seen even with a *random* ordering scheme [40], where the rows of T are accessed in a random fashion. Another critical choice for convergence is that of the relaxation parameter $r^{(k)}$ for iteration k. This value is determined by experimentation on test images which are similar in some sense to the image that is to be reconstructed [40]. We point out to the reader that if the system of equations (2.6) is consistent, and if the matrix T is well-conditioned, then

---

[5] In the standard Kaczmarz method no relaxation is used, i.e. the relaxation parameter $r^{(k)}$ is set to one for all iterations k. As a result (2.7) corresponds exactly to projection of $f^{(i)}$ onto the hyperplane defined by $y(k) = T(k)f$ only if $r^{(k)} = 1$.

the choice of the ordering scheme for the rows of T, and the relaxation parameter will have no effect on the convergence of the reconstruction. In practice, since T is not well-conditioned, these choices are critical for convergence. This gives us a preview of things to come. In Chapter 4 we will present an efficient method for isolating and removing the ill-conditioning in the projection matrix T. As a result, we will perform the Kaczmarz recursions of the type (2.7) on a system of equations where the system matrix is well-conditioned. This will result in our modified ART algorithm being relatively insensitive to the ordering scheme and to the choice of the relaxation parameter. This gain will *not be* at the expense of any additional computations per iteration, since our well-conditioned matrix will have a similar sparsity as that of the ART projection matrix[6] T.

Finally, there are two important sources of inconsistency in the system of equations (2.6) used by ART. The first is due to the presence of measurement noise. The second, which has been a source of some controversy in the past [37,38], is due to the fact that there is a mismatch between the rectangular pixel representation (2.3) used by ART and the strip-integral measurements of the object. In particular, reconstruction errors result in ART because the object estimate is required to have the same value within each pixel, whereas the measurements are a function of the variations of object within the pixel. The NP representation [9,10], which we discuss next, was introduced to remove this second inconsistency arising due to the standard rectangular pixel object representation in ART.

2.2.2 The natural pixel (NP) object representation

In the natural pixel object representation [9,10], the object is represented in the same basis functions (i.e. strips $\{S_m(u,v)\}$) along which the strip-integral data y are

---

[6] As an example, for an imaging geometry with $N_\theta = N_s = 32$, the projection matrix T has a condition number of 2.3×10^{16}. Our method for removing the ill-conditioning, however, results in a matrix with a condition number of 1.3×10^{3}. In addition, since the two matrices have a similar percentage of non-zero elements ($\sim 6.5\%$), the computational complexity per iteration of the Kaczmarz method is same for the two cases.

collected. Thus:

$$f(u,v) = \sum_{n=1}^{N_\theta N_s} x(n) S_n(u,v), \tag{2.8}$$

where $x(n)$ is the coefficient of expansion of the object in strip $S_n(u,v)$. Now by substituting the NP object representation equation (2.8) into the strip-integral data expression (2.1), we obtain:

$$y(m) = \sum_{n=1}^{N_\theta N_s} C_{mn} x(n), \qquad m = 1, \ldots, N_\theta N_s \tag{2.9}$$

with

$$C_{mn} = \iint_{\mathbf{\Omega}} S_n(u,v)\, S_m(u,v)\, du\, dv, \qquad m = 1, \ldots, N_\theta N_s;\ n = 1, \ldots, N_\theta N_s, \tag{2.10}$$

where $\mathbf{\Omega}$ is the field-of-view of the imaging system. Finally, by collecting together the projection data in (2.9), we get the following overall observation equation:

$$y = Cx, \tag{2.11}$$

where y is the overall data vector of length $N_\theta N_s$, x is a $N_\theta N_s$ vector containing the expansion coefficients $\{x_n, n = 1, \ldots, N_\theta N_s\}$, and C is a $N_\theta N_s \times N_\theta N_s$ matrix, the (m,n)-th element of which is given by C_{mn} (see (2.10)). Thus in the NP representation, the tomography problem reduces to solving for the object expansion coefficients x from the data y according to (2.11), and then obtaining the reconstruction $f(u,v)$ from (2.8). We point out to the reader that the NP representation is exactly matched to the data since it uses the strips, rather than square pixels, as basis functions (c.f. (2.8)). It is due to this very reason that the NP representation is free of the pixel errors that arise in ART due to the rectangular pixel object representation. Finally, even though no assumption about the discretization of the object on a rectangular pixel grid is needed to solve for x from (2.11), once the coefficients x have been calculated, a discretization of (2.8) is needed to display the reconstruction. The reader can easily verify that the discretization of (2.8) on a $N_s \times N_s$ pixel grid results in the reconstruction:

$$f = T^T x, \tag{2.12}$$

where the (m, n)-th element of T is given by (2.5), and f is the vector containing the object expansion coefficients in the rectangular pixel basis (c.f. (2.3)). Note that (2.12) can be interpreted as the back-projection operation [44] where the object coefficients x are back-projected along the basis functions T to obtain the reconstruction f.

Unfortunately, solution of the system of equations (2.11) for the coefficients x, as required by the NP method, is complicated by the fact that the NP matrix C is large and *full* (refer to Figure 2-3 for a picture[7] of C for an imaging geometry with $N_\theta = N_s = 32$). The elements of C are the areas of intersection of the strips defined by the basis functions T (see (2.10) and Figure 2-4). Most of these areas are not zero and hence the matrix C is *full*. The large size of C makes it difficult to solve for x directly from (2.11). In [9,10,24] this problem is circumvented by two different approaches. The first is to use iterative techniques and a suitable initial value to solve for x. However, since the NP matrix C is not sparse, these iterative techniques result in reconstruction algorithms that are inefficient. The second approach is to concentrate on some specific imaging geometries which result in a matrix C that can be directly inverted in a computationally efficient manner. The problem with the second approach is that these imaging geometries may not be practical. A final major difficulty in obtaining the NP reconstruction is that there is an inherent non-uniqueness in the NP object representation arising from it's tie to the data acquisition process, which results in C being rank deficient or at best being badly conditioned for most imaging geometries[8]. None of the NP related work [9,10,24] discusses this conditioning issue brought on by the non-uniqueness of the NP representation.

We discuss next the origin of the non-uniqueness in the NP representation that is responsible for the ill-conditioning of C. To simplify the discussion, we assume $C = TT^T$ which, apart from discretization effects, is always true. Further, we partition

---

[7] Refer to Section 2.A for details on the formation of the NP matrix C.

[8] For example, for an imaging geometry with $N_\theta = N_s = 32$, the 1024×1024 matrix C has a condition number of 1.4×10^5.

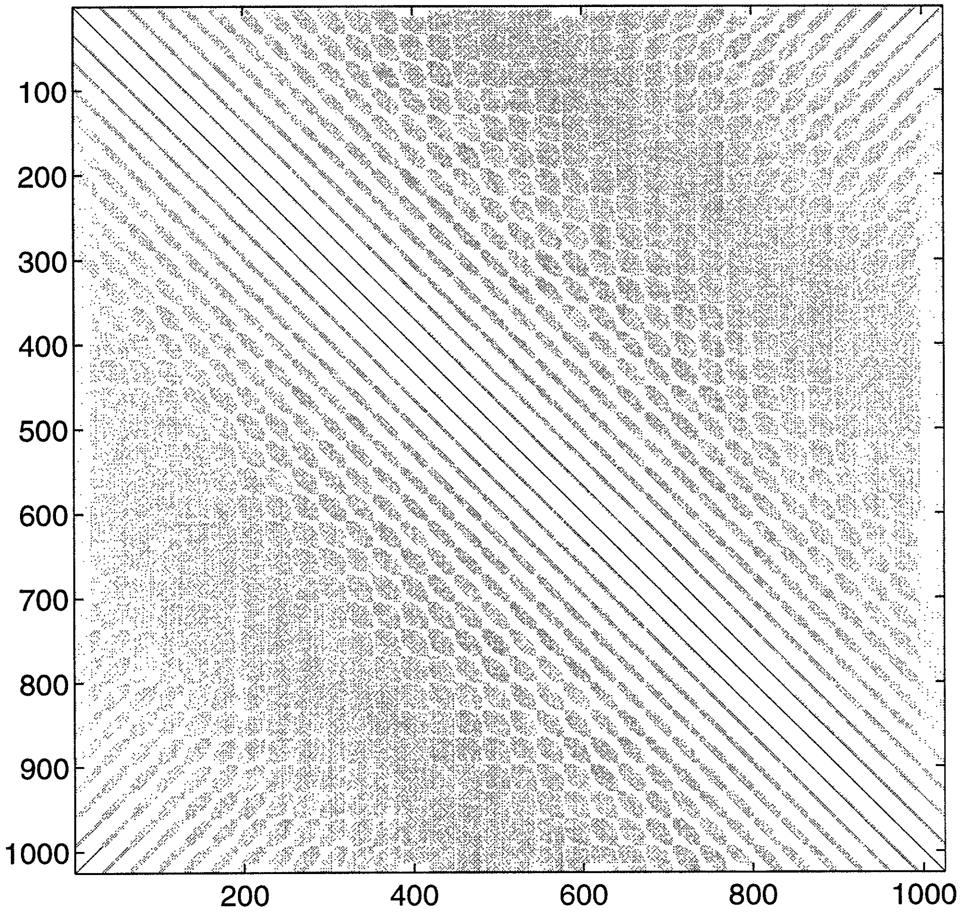


Figure 2-3: The grayscale plot (black corresponds to the maximum value and white to the minimum) of the NP matrix C for an imaging geometry defined by $N_\theta = N_s = 32$. The matrix is of size 1024 × 1024. Note that C is full.

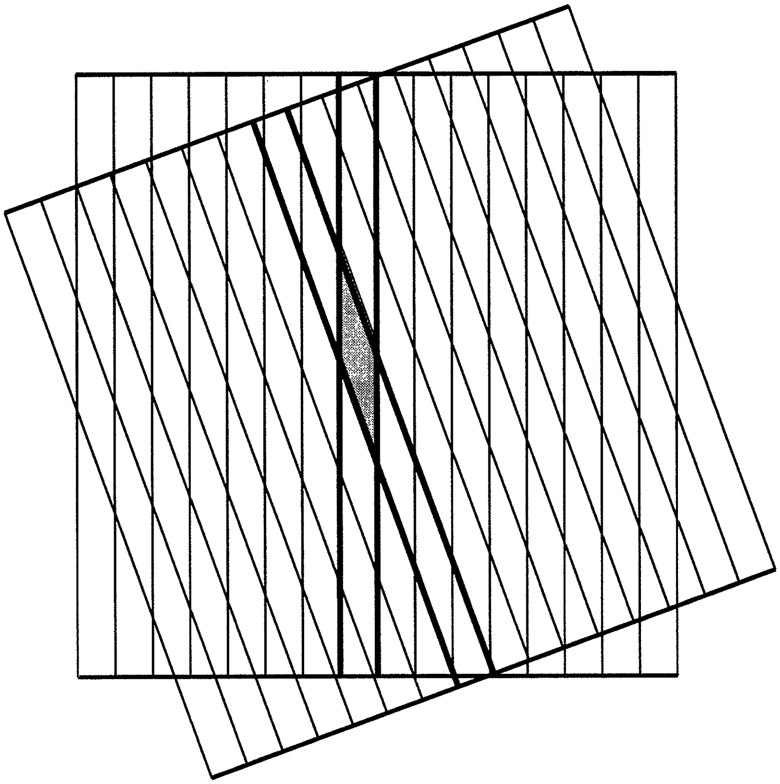


Figure 2-4: The elements of the matrix C are the areas of intersection of various strips. One such area of intersection, corresponding to two strips delineated by bold lines, is shown shaded. The matrix C is full as most of these areas are non-zero.

the projection matrix T into components at different angles, corresponding to the following partitioning of the observation equation (2.6):

$$\left[\begin{array}{c} y_1 \\ \hline y_2 \\ \hline \vdots \\ \hline y_{N_\theta} \end{array}\right] = \left[\begin{array}{c} T_1 \\ \hline T_2 \\ \hline \vdots \\ \hline T_{N_\theta} \end{array}\right] f, \tag{2.13}$$

where T_k is an $N_s \times N_s^2$ matrix, the rows of which represent the discretized strips at angle k, and recall that y_k are the projection data at angle k (c.f. (2.2)).

The ill-conditioned nature of the matrix C can now be easily understood at an intuitive level if one assumes an infinite field-of-view for the imaging geometry (rather than the finite rectangular field-of-view we show in Figure 2-1) so that edge effects are absent. Recall that the columns of T_k^T are the basis functions of the NP representation at angle k, c.f. (2.13). Now the sum of all the columns of $T_{k_1}^T$ equals the sum of all the columns of $T_{k_2}^T$ since both correspond to the same indicator function of the field-of-view. This simply reflects the physical fact that both of these sums correspond to the DC value of the object field. Thus the same underlying object can be represented in a variety of ways, corresponding to different allocations of its DC component to the different angular basis sets. Hence the representation (2.12) is non-unique, T^T does not have full column rank, and $C = TT^T$ is not invertible. For the case of our finite field-of-view the above discussion is exact only if k_1 and k_2 correspond to projections at 0^0 and 90^0 due to nonuniform edge effects. Even when the projections are not exactly at right angles however, while not dropping rank, C is quite ill-conditioned.

The above discussion provides us with a preview of things to come. In Chapter 3, we use wavelet bases to transform the NP strips S into a multiscale framework. The use of wavelet bases, in addition to providing a multiscale framework, enables us to overcome the above limitations of the NP reconstruction. In this wavelet transformed domain the multiscale system matrix corresponding to C is *sparse*. Further, the coarsest multiscale basis function at any angle k turns out to be the sum of all the columns of T_k^T, responsible for the ill-conditioning of the resulting multiscale matrix. We exploit this feature to partition the multiscale system matrix by scales to obtain a reconstruction procedure that requires inversion of only a well-conditioned and sparse matrix.

2.2.3 The filtered back-projection (FBP) technique

The filtered back-projection (FBP) reconstruction technique [8,39,44,51] is the most commonly used method for image reconstruction from dense tomographic data. It is based directly on the Radon inversion formula which is valid (i.e. yields exact reconstructions) only when a continuum of noise-free line integral projections from all angles are used [44]. In practice, as indicated in (2.1), we only have access to sampled projection data which are collected using strips of finite width. In this case, the quality of the FBP reconstruction is a function of the quality and fineness of the corresponding projection data used. We refer the reader to [8,16,45,65] for details on sampling requirements for the Radon transform.

In the FBP reconstruction, the object is expanded in a non-orthogonal basis that is closely related to the data acquisition process and the coefficients of this expansion are then found from local processing of the data in each angular projection. In particular, as in the NP method, the estimated object is represented as a linear combination of the same functions $S_m(u, v)$ along which the projection data are collected. Similar to (2.12), a discretized version of this representation may be obtained as:

$$f = T^T x = \sum_{k=1}^{N_\theta} T_k^T x_k, \tag{2.14}$$

where the N_s vector x_k contains the object coefficient set at angle k, and T_k is the block of the projection matrix associated with discretized strips at angle k (see (2.13)). Note that, as in the NP case, (2.14) can be interpreted as the back-projection operation where the object coefficients x_k are back-projected along the basis functions T_k at each angle k and then the contributions from all N_θ angles are added to get the overall reconstruction f.

To complete the reconstruction the coefficients x_k must now be determined. The standard FBP method calculates them for each angle k according to the Radon inversion formula by filtering the projection data y_k at that particular angle with a ramp filter [44]. Thus, for a fixed angle k:

$$x_k = R\, y_k, \tag{2.15}$$

where the matrix R captures this ramp-filtering operation (refer to Figure 2-5 for a picture of R for an imaging geometry with $N_s = 64$, and to Figure 2-6 which is the fast Fourier transform (FFT) of a central row of R). Thus (2.14) and (2.15) together represent the two operations used in the standard FBP reconstruction.

By collecting the object coefficient and data vectors at different angles we obtain the following overall equation which reflects the identical and independent processing from angle to angle performed on the projection data by the FBP method:

$$x = \begin{bmatrix} R & 0 & \dots & 0 \\ 0 & R & \dots & 0 \\ & & \ddots & \\ 0 & 0 & \dots & R \end{bmatrix} y. \tag{2.16}$$

An important point to note in the above equation is that the matrix R is fixed and is not a function of the imaging system (i.e. the quality and the quantity of the acquired projection data). It is this simple, regular structure in the explicit formula (2.16) that results in the computational efficiency of FBP. Note that in the NP reconstruction, since the matrix C is full, the processing of data is not independent from angle to angle, in contrast to FBP. The *advantage* of the NP reconstruction over the FBP is that since the matrix C is calculated for each specific acquisition geometry, the reconstruction can be customized for any imaging system, so that a complete set of angular projection data is neither assumed nor required for adequate reconstruction as it is in FBP. Recall that the FBP method, which is based on the Radon inversion formula, yields exact reconstructions only when a continuum of noise-free line-integral

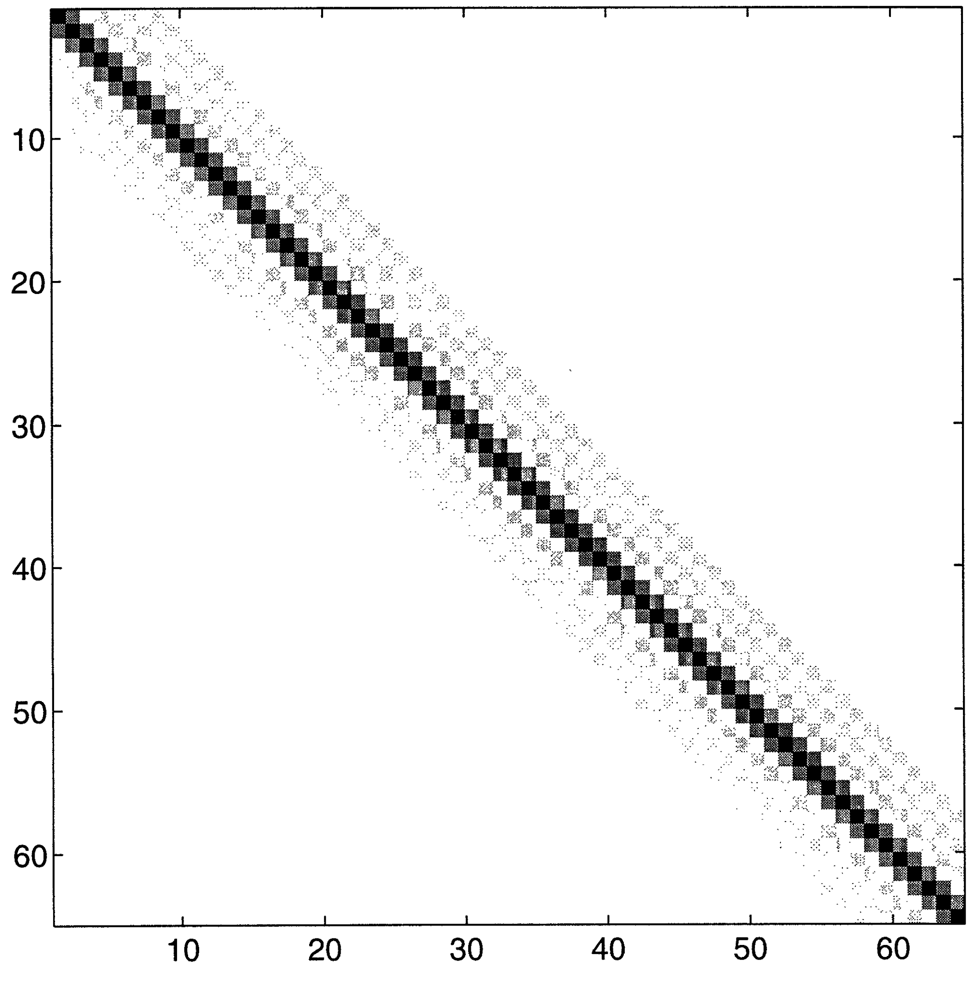


Figure 2-5: The grayscale plot (black corresponds to the maximum value and white to the minimum) of the FBP ramp-filter matrix R for an imaging geometry defined by $N_s = 64$. The matrix is of size 64 × 64.

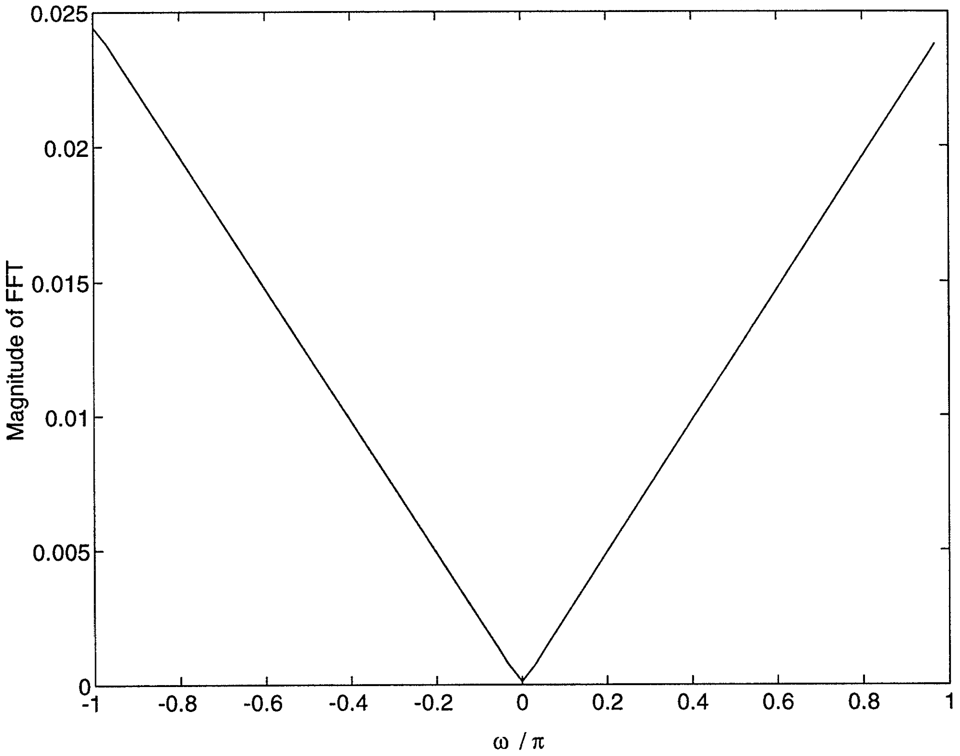


Figure 2-6: The FFT of a central row of the FBP ramp-filter matrix R for an imaging geometry with $N_s = 64$.

data are available.

2.3 Results from estimation and detection theory

In this section, we mention a few relevant results from the theory of estimation and detection of signals from noisy data. We use these results extensively, first while developing regularization techniques for reconstruction from noisy data, and subsequently while developing detection algorithms in the multiresolution framework.

2.3.1 Maximum-likelihood (ML) and maximum-aposteriori probability (MAP) estimation

Consider noisy observations g of a signal h, of the form:

$$g = Ah + n, \qquad n \sim \mathcal{N}(0, \Lambda_n) \tag{2.17}$$

where n is the additive noise vector, and the notation $z \sim \mathcal{N}(m, \Lambda)$ denotes a Gaussian distribution of mean m and covariance matrix Λ. We thus assume the noise vector n to be Gaussian, with mean 0 and covariance Λ_n.

The *maximum-likelihood (ML) estimator* of h, given noisy observations g and the distribution for noise n, is given by [76]:

$$\begin{aligned} \widehat{h}_{\mathrm{ML}} &\triangleq \arg\max_h \left[p(g|h)\right] \\ &= \arg\min_h \left[(g - Ah)^T \Lambda_n^{-1} (g - Ah)\right] \\ &= (A^T \Lambda_n^{-1} A)^{-1} A^T \Lambda_n^{-1} g, \end{aligned} \tag{2.18}$$

where $p(.)$ refers to the probability density. The estimation error in the ML case is defined as:

$$e_{\mathrm{ML}} \triangleq h - \widehat{h}_{\mathrm{ML}}, \tag{2.19}$$

and it can be shown that the ML estimator is unbiased, i.e.

$$b_{\mathrm{ML}} \triangleq E[e_{\mathrm{ML}}|h] = 0, \tag{2.20}$$

where the notation $E[z]$ refers to the expected value of z. Finally, the error covariance for the ML estimate is given by:

$$\begin{aligned} \Lambda_{\mathrm{ML}} &\triangleq E[(e - b_{\mathrm{ML}})(e - b_{\mathrm{ML}})^T|h] \\ &= (A^T \Lambda_n^{-1} A)^{-1}. \end{aligned} \tag{2.21}$$

The ML estimator assumes no statistical prior knowledge about the signal h. The *maximum-aposteriori probability (MAP) estimate*, which we discuss next, provides a method for incorporating such prior information.

Let us assume the signal h to be random, uncorrelated with noise n, and distributed according to $h \sim \mathcal{N}(0, \Lambda_h)$. Given this prior knowledge, the MAP estimate[9] of h is obtained as [76]:

$$\begin{aligned} \widehat{h}_{\mathrm{MAP}} &\triangleq \arg\max_h \,[p(h|g)] \\ &= \arg\max_h \,[p(g|h)p(h)] \\ &= \arg\min_h \,[(g - Ah)^T \Lambda_n^{-1}(g - Ah) + h^T \Lambda_h^{-1} h] \\ &= (A^T \Lambda_n^{-1} A + \Lambda_h^{-1})^{-1} A^T \Lambda_n^{-1} g. \end{aligned} \tag{2.22}$$

The estimation error in the MAP case is defined as:

$$e_{\mathrm{MAP}} \triangleq h - \widehat{h}_{\mathrm{MAP}}, \tag{2.23}$$

and, once again, it can be shown that the MAP estimator is unbiased, i.e.

$$b_{\mathrm{MAP}} \triangleq E[e_{\mathrm{MAP}}] = 0. \tag{2.24}$$

---

[9] Note that since we assume Gaussian distributions, the MAP estimate in our case is equivalent to the linear least squares estimate.

Finally, the error covariance for the MAP estimate is given by:

$$\begin{aligned}\Lambda_{\mathrm{MAP}} &\triangleq E[(e - b_{\mathrm{MAP}})(e - b_{\mathrm{MAP}})^T] \\ &= (A^T \Lambda_n^{-1} A + \Lambda_h^{-1})^{-1}. \qquad (2.25)\end{aligned}$$

2.3.2 Binary hypothesis testing

In binary hypothesis testing [76], the goal is to discriminate between two hypotheses H_0 and H_1, with prior probabilities P_0 and P_1 respectively, based on information provided by some measurement vector g. In the *Bayes risk framework*, this discrimination is achieved by minimizing a risk which penalizes the various errors of mis-classification in proportion to a cost associated with the errors. Specifically, in the Bayes framework, the binary hypothesis testing problem amounts to finding a set Z_1 such that, if we decide H_1 in case $g \in Z_1$, the risk is minimized. If c_{ij} is the cost of deciding H_i when H_j is true, then the *Bayes risk*, which is to be minimized, is written as:

$$E(c) \triangleq c_{00}P_0 + c_{01}P_1 + (c_{10} - c_{00})P_0 P_F - (c_{01} - c_{11})P_1 P_D \qquad (2.26)$$

where P_D, the conditional probability of detection, is defined as:

$$P_D \triangleq \int_{Z_1} dg\, p(g|H_1), \qquad (2.27)$$

and P_F, the conditional probability of false alarm, is defined as:

$$P_F \triangleq \int_{Z_1} dg\, p(g|H_0). \qquad (2.28)$$

The discrimination strategy that minimizes the risk (2.26) is given by the following likelihood ratio test:

$$\begin{aligned}&\text{Decide } \mathrm{H}_1 \text{ if}: \; L(g) = \frac{p(g|H_1)}{p(g|H_0)} \geq \frac{(c_{10} - c_{00})P_0}{(c_{01} - c_{11})P_1} \triangleq \eta \\ &\text{else decide } \mathrm{H}_0. \qquad (2.29)\end{aligned}$$

In case the prior probabilities P_0 and P_1 of the two hypotheses are equal, and the cost assignment is such that $c_{00} = c_{11} = 0$ and $c_{01} = c_{10}$, the likelihood ratio strategy (2.29) reduces to the following *maximum-likelihood decision rule*:

$$\begin{aligned} &\text{Decide } \mathrm{H}_1 \text{ if}: \; p(g|H_1) \geq p(g|H_0) \\ &\text{else decide } \mathrm{H}_0. \end{aligned} \tag{2.30}$$

Finally, the maximum-likelihood decision rule (2.30) is equivalent to the following *log-likelihood rule*:

$$\begin{aligned} &\text{Decide } \mathrm{H}_1 \text{ if}: \; \ln(p(g|H_1)) \geq \ln(p(g|H_0)) \\ &\text{else decide } \mathrm{H}_0. \end{aligned} \tag{2.31}$$

We consider next the form of the log-likelihood rule (2.31) for a particular case where the two hypotheses H_0 and H_1 are given by:

$$\mathrm{H}_0 \;:\; g = Ah_0 + n, \qquad n \sim \mathcal{N}(0, \Lambda_n),\; h_0 \sim \mathcal{N}(0, \Lambda_0), \tag{2.32}$$

$$\mathrm{H}_1 \;:\; g = Ah_1 + n, \qquad n \sim \mathcal{N}(0, \Lambda_n),\; h_1 \sim \mathcal{N}(0, \Lambda_1), \tag{2.33}$$

where the noise n is uncorrelated with both h_0 and h_1. In this case, the log-likelihood rule (2.31) reduces to:

$$\begin{aligned} \text{Decide } \mathrm{H}_1 \text{ if}: \quad & -\frac{1}{2}\ln|A\Lambda_1 A^T + \Lambda_n| - \frac{1}{2}g^T(A\Lambda_1 A^T + \Lambda_n)^{-1}g \\ \geq \; & -\frac{1}{2}\ln|A\Lambda_0 A^T + \Lambda_n| - \frac{1}{2}g^T(A\Lambda_0 A^T + \Lambda_n)^{-1}g \\ \text{else decide } \mathrm{H}_0, & \end{aligned} \tag{2.34}$$

where the notation $|Z|$ refers to the determinant of matrix Z.

2.4 Conventional regularization techniques for reconstruction from noisy data

In this section we describe the commonly used regularization techniques for reconstruction from noisy projection data. These techniques fall into two categories depending on whether they are ad hoc or statistically-based. As is expected, the ad hoc techniques are much faster, but do not provide a statistically-optimal framework for selecting the regularization parameters.

2.4.1 Ad hoc regularization techniques

The most familiar example of an ad hoc regularization technique is provided by the FBP where, in presence of measurement noise, the ramp filter is rolled-off at high frequencies [3, 39, 44] (see Figure 2-7). This is called *apodization* [3] and several different windows are typically used for this purpose, for example Hann, Hamming, Parzen, Butterworth etc. [50]. The assumption is that most of the object energy occurs at low frequencies while the most disturbing noise-derived artifacts occur at high frequency. The high frequency roll-off thus attenuates these components at the expense of not reconstructing the fine scale features in the object. Since the overall procedure is essentially the same as the original FBP method, the result is a fast though ad hoc, method for regularization.

Another ad hoc form of regularization is seen in ART, where the iterations are stopped prematurely before an excessive fitting of the solution to the noisy data occurs. This results in a smoothed, and thus a regularized reconstruction.

2.4.2 Statistically-based regularization techniques

A first example of the statistically-based regularization techniques is seen in ART, where a slightly modified version of the unregularized iterative algorithm is used to calculate an MAP estimate of the object pixels [39]. Specifically, the following model

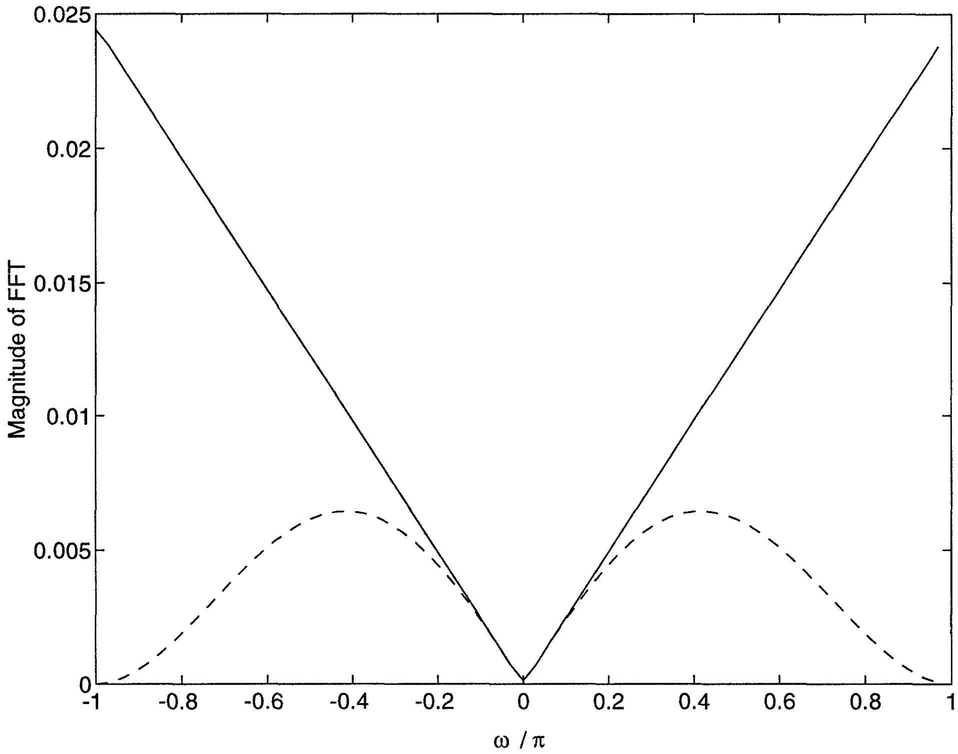


Figure 2-7: The ramp-filter (solid line) for $N_s = 64$, and the rolled-off version (broken line). The roll-off is according to the Hann window.

is considered for noisy observations (the noisy version of (2.6)):

$$y = Tf + n, \qquad\qquad n \sim \mathcal{N}(0, \lambda_n I_{N_\theta N_s}) \tag{2.35}$$

where I_z refers to a $z \times z$ identity matrix. The noise vector n is thus assumed to be zero-mean and uncorrelated. The MAP estimate of the object pixels f is then calculated from (2.35) based on a prior model $f \sim \mathcal{N}(0, \lambda_f I_{N_s^2})$ which corresponds to assuming that the pixel values of f are uncorrelated. This MAP estimate is given by (c.f. (2.22)):

$$\begin{aligned}\hat{f}_{\mathrm{MAP}} &= \arg\min_f \, [(\frac{1}{\lambda_n})(y - Tf)^T(y - Tf) + (\frac{1}{\lambda_f})(f^T f)] \quad \text{i.e.,} \\ &= [(\frac{1}{\lambda_n})T^T T + (\frac{1}{\lambda_f})I]^{-1}(\frac{1}{\lambda_n})T^T y, \end{aligned} \tag{2.36}$$

and is obtained by the modified ART algorithm with essentially the same computational and storage requirements as the original, unregularized ART reconstruction.

A second class of statistically-based regularization techniques calculates the MAP estimate of the object pixels f by using more involved, correlated, *Markov random field (MRF) prior models* for the object [27, 28, 33, 49]. These prior models impose local smoothness constraints on the object pixels by penalizing reconstructions where the reconstructed pixel values fluctuate wildly in localized regions. The MRF class of prior models is captured by the following *Gibbs probability distribution*:

$$f \sim c \exp(-\beta U(f)), \tag{2.37}$$

where c and β are constants, and $U(f)$ is an energy function that specifies the correlations between the pixel values.

As an example, for a particular nearest-neighbor MRF prior model suggested by Geman and McClure [28], the energy function has the form:

$$U(f) = \sum_{[s,t]} \Phi(f_s - f_t) + \sum_{\{s,t\}} \frac{1}{\sqrt{2}} \Phi(f_s - f_t), \tag{2.38}$$

where $[s, t]$ denotes pixels that are horizontally or vertically adjacent, $\{s, t\}$ denotes pixels that are diagonally adjacent, and

$$\Phi(z) = \frac{-1}{1 + (z/\delta)^2}, \tag{2.39}$$

where δ is a constant. Note that, since $\Phi(z)$ initially increases with z but subsequently levels-off asymptotically at large values of z, the energy function (2.38) penalizes large differences in nearest-neighbor pixel values, but still allows abrupt transitions to occur without excessive penalty.

Finally, the use of the MRF prior models (2.37) results in MAP estimates that are extremely computationally intensive. As an example, the MAP estimate corresponding to the MRF prior model (2.37) and the noise model (2.35) has the form:

$$\hat{f}_{\mathrm{MAP}} = \arg\min_{f} \left[\left(\frac{1}{\lambda_n}\right)(y - Tf)^T (y - Tf) + \beta U(f) \right] \tag{2.40}$$

and is obtained by using elaborate Monte Carlo techniques like the computationally intensive Metropolis algorithm [27].

2.5 Wavelet transform-based multiresolution representation

In this section we present a brief summary of the wavelet-based multiresolution representation of 1-D and 2-D signals. The multiresolution techniques in this thesis require only the representation of 1-D signals. However, we also discuss the form of the 2-D multiresolution representation, which we use in the next section to bring out the differences between our work in multiresolution tomography and other existing work, almost all of which uses the 2-D wavelet representation. In either case, we do not intend this as a complete treatment of the topic and intentionally suppress many details. The interested reader is referred to any of the many papers devoted to this topic, e.g. the excellent paper [59]. Finally, in this section, we use the terms *multires-*

olution and *multiscale* interchangeably with the understanding that high resolution corresponds to fine scales.

2.5.1 Representation of 1-D signals

Let $L^2(R)$ denote the vector space of measurable, square-integrable, 1-D signals $x(u)$, and let Z denote the set of integers. A multiscale approximation of $L^2(R)$ is a sequence of subspaces $\{V_j\}_{j \in Z}$ with $\ldots V_{-1} \subset V_0 \subset V_1 \ldots$ having the interpretation that the projection $A_j x(u)$ of $x(u)$ on V_j gives the approximation of $x(u)$ at scale j. The scales become finer with increasing j. Now, as shown in [59], there exists a unique function $\phi(u) \in L^2(R)$, called the *scaling function*, such that for each scale j, $\{\phi_{j,l}(u) = \sqrt{2^j}\phi(2^j u - l)\}_{l \in Z}$ is an orthonormal basis of V_j. Thus the approximation of the function at scale j can be written as:

$$A_j x(u) = \sum_l x^{(j)}(l)\phi_{j,l}(u) \tag{2.41}$$

with

$$x^{(j)}(l) = < x(u), \phi_{j,l}(u) > \tag{2.42}$$

where $< ., . >$ refers to the inner product operation. The difference in information between the approximation of the signal at successive scales j and $j+1$ is captured by the ***detail signal*** at scale j. This detail signal is obtained as the projection $D_j x(u)$ of $x(u)$ on the subspace O_j which represents the orthogonal complement of V_j in V_{j+1}. Similar to the scaling function, there exists a function $\psi(u) \in L^2(R)$ called an *orthogonal wavelet*, such that for each scale j, $\{\psi_{j,l}(u) = \sqrt{2^j}\psi(2^j u - l)\}_{l \in Z}$ is an orthonormal basis of O_j. Thus the detail signal at scale j, capturing the difference in information between $A_j x(u)$ and $A_{j+1} x(u)$, can be written as:

$$D_j x(u) = \sum_l \xi^{(j)}(l)\psi_{j,l}(u) \tag{2.43}$$

where

$$\xi^{(j)}(l) = < x(u), \psi_{j,l}(u) > . \tag{2.44}$$

As discussed in [59], there exist functions h and g satisfying:

$$\begin{aligned}\phi(u) &= \sum_l \sqrt{2}h(l)\phi(2u-l) \\ \psi(u) &= \sum_l \sqrt{2}g(l)\phi(2u-l).\end{aligned} \tag{2.45}$$

The discrete approximation and detail coefficients at scale j, $x^{(j)}$ and $\xi^{(j)}$ respectively, can then be obtained from the next finer scale discrete approximation coefficients, $x^{(j+1)}$, by convolution with h and g followed by down-sampling by a factor of 2:

$$\begin{aligned}x^{(j)}(l) &= (h * x^{(j+1)}(2l)) \\ \xi^{(j)}(l) &= (g * \xi^{(j+1)}(2l)),\end{aligned} \tag{2.46}$$

where $*$ refers to 1-D convolution. If $x^{(j)}$ is the vector containing the sequence $x^{(j)}(l)$ and $\xi^{(j)}$ is the corresponding vector of detail added in proceeding to the next finer scale, then the recursions (2.46) are equivalent to the following *wavelet analysis equations* [74]:

$$\begin{aligned}x^{(j)} &= L(j)\,x^{(j+1)}, \\ \xi^{(j)} &= H(j)\,x^{(j+1)},\end{aligned} \tag{2.47}$$

where $L(j)$ and $H(j)$ are matrices (linear transformations) which depend on the sequences h and g, and hence on the particular scaling function and wavelet that are used for multiscale representation. The operators $L(j)$ and $H(j)$ correspond roughly to low and high pass filters, respectively, followed by down-sampling by a factor of 2.

Beyond the recursive computation of the representation (2.46), the finer scale discrete approximation $x^{(j+1)}$ can be synthesized from the next coarser scale discrete approximation and detail coefficients, $x^{(j)}$ and $\xi^{(j)}$ respectively, by first up-sampling by a factor of 2 followed by convolution with h and g:

$$x^{(j+1)}(l) = \sum_k h(2k-l)x^{(j)}(k) + \sum_k g(2k-l)\xi^{(j)}(k). \tag{2.48}$$

The above recursion (2.48) is equivalent to the following *wavelet synthesis equation* [74]:

$$x^{(j+1)} = L^T(j)\,x^{(j)} + H^T(j)\,\xi^{(j)}, \tag{2.49}$$

where $L^T(j)$ and $H^T(j)$ are the transposes (i.e. adjoints) of the same operators defined in connection with (2.47). The operators $L^T(j)$ and $H^T(j)$ serve to interpolate the "low" and "high" frequency (i.e. approximation and detail) information, respectively, at one scale up to the next finer scale. The vector $\xi^{(j)}$, containing the information added in going from scale j to $j+1$, is composed of the *wavelet transform coefficients* at scale j.

Till now we have considered the representation of a continuous signal $x(u)$. We next describe how the continuous-signal results developed so far can be used to obtain the multiresolution representation of a discrete signal. To begin, let us assume that we do not have access to the continuous signal $x(u)$ but rather to a finite number N of its samples. We assume these samples to be the approximation coefficients $x^{(J)}$ of $x(u)$ at some finest resolution J. For convenience we assume N to be a power of 2 so that $J = \log_2(N)$. Starting from a "coarsest" approximation $x^{(0)}$ (usually taken to be the average value of the signal) then, it is possible to recursively and efficiently construct the different scale approximations through (2.49) by using the complete set of wavelet coefficient vectors $\{\xi^{(j)}\}$. This layered construction is shown graphically in Figure 2-8a, where our approximations are refined though the addition of finer and finer levels of detail as we go from right to left until the desired scale of approximation is achieved. In particular, the original signal x is obtained by adding all the inter-scale detail components $\xi^{(j)}$ to the initial approximation $x^{(0)}$. For a given signal x the complete set of these elements uniquely captures the signal and thus corresponds to a simple change of basis. In addition, note from Figure 2-8 that the intermediate approximation $x^{(j)}$ at scale j is generated using only the corresponding subset of the complete wavelet coefficient set (e.g. to obtain $x^{(2)}$ we use only $x^{(0)}$, $\xi^{(0)}$, and $\xi^{(1)}$). The representation of this intermediate approximation at the original finest scale can be found by repeated interpolation of the information in $x^{(j)}$ through

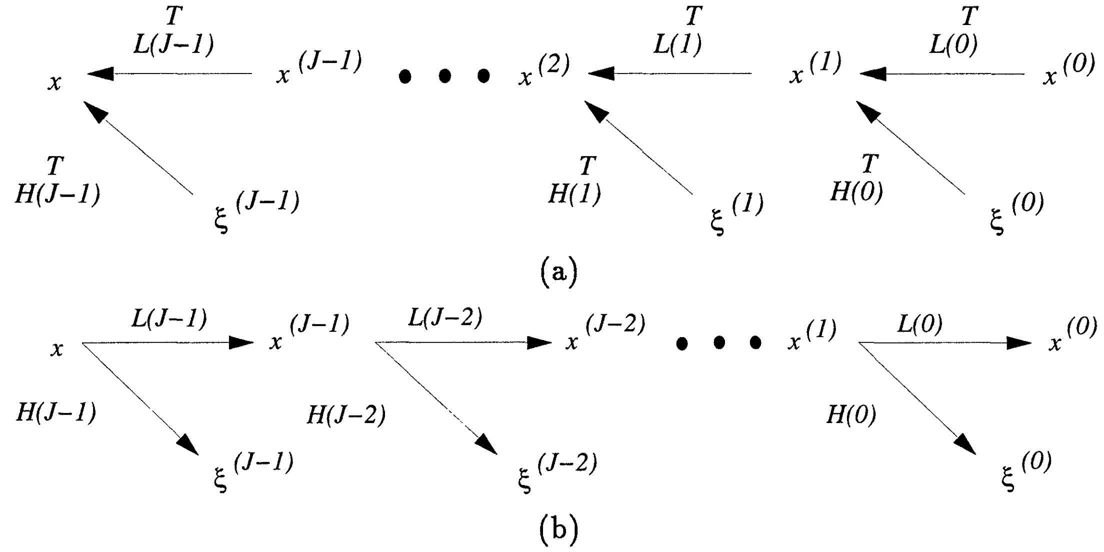


Figure 2-8: (a) Tree diagram for wavelet transform synthesis. We start from a coarsest approximation $x^{(0)}$ on the right and progressively add finer levels of detail $\xi^{(j)}$ as we proceed to the left, thus refining the original approximation to the signal. The original (finest scale) sequence is obtained as the final output on the left. (b) Tree diagram for wavelet transform analysis. Starting from a finest level signal x in the left we recursively peel off layers of detail $\xi^{(j)}$ as we proceed to the right and the next coarser scale representation $x^{(j)}$.

the application of $L^T(j')$, $j' \geq j$. This interpolation up to the finest scale corresponds to effectively assuming that additional, finer scale, detail components $\xi^{(j')}$, $j' \geq j$ are zero in our representation of the signal. It is such intermediate scale approximations and the detail necessary to go between them that give the wavelet transform its natural multiscale interpretation, and indeed we exploit such interpretations in the following chapters to obtain induced multiscale *object* representations.

In addition to the recursive computation of the approximations, it is also possible to compute the components of the *representation* itself (i.e. the wavelet coefficients) recursively through (2.47) by exploiting the same multiscale structure. Figure 2-8b shows how these wavelet coefficient vectors at each scale are obtained by "peeling off" successive layers of detail as we proceed from finer to coarser scales (left to right in the figure). This recursive structure yields algorithms for the computation of the wavelet transform coefficients that are extremely efficient. For convenience in the development to follow, we will capture the overall operation which takes a vector x containing a discrete signal to the vector ξ containing all of its corresponding wavelet transform elements $\{\xi^{(j)}\}$ and $x^{(0)}$ by the matrix operator W as follows:

$$Wx = \begin{bmatrix} \xi^{(J-1)} \\ \hline \vdots \\ \hline \xi^{(0)} \\ \hline x^{(0)} \end{bmatrix} \triangleq \xi. \tag{2.50}$$

Since the transform is invertible and the wavelet basis functions are orthonormal, it follows that W^{-1} exists and further that W is an orthogonal matrix, i.e. that $W^{-1} = W^T$. Finally, because of the highly recursive structure of the multiscale representation algorithms in the wavelet domain, the multiscale representation operation (2.50) can be performed in an extremely efficient manner ($\mathcal{O}(N)$) [6].

From the above discussion, the matrix W captures the operation of the operators $L(j)$ and $H(j)$, and thus depends on the underlying chosen wavelet. In our work in this paper, in addition to the Haar wavelet[10] we will use wavelets from an especially popular family of these functions due to Daubechies [18], the separate elements of which are denoted D_n, where n is an indication of the support size of the corresponding filters contained in $L(j)$ and $H(j)$. Finally, since our signals are of finite length, we need to deal with the edge effects which occur at the ends of the interval in the wavelet transform. While there are a variety of ways in which to do this, such as modifying the wavelet functions at the ends of the interval in order to provide an orthogonal representation over the interval [17], we have chosen here to use one of the most commonly used methods, namely that of cyclically wrapping the interval which induces a circulant structure in $L(j)$ and $H(j)$ [29, 59]. While this does introduce some edge effects, these are of negligible importance for the objectives and issues we wish to emphasize and explore and for the applications considered here. Further, the methods we describe can be readily adapted to other approaches for dealing with edge effects as in [17] and the references contained therein.

As noted above, the intermediate approximations $x^{(j)}$ and their finest scale representation may be obtained by using only part of the full wavelet coefficient set during synthesis, effectively assuming the finer scale detail components are zero. For convenience in the discussion to follow we capture this partial zeroing operation in the matrix operator $A(j)$, which nulls the upper $(J-j)$ sub-vectors of the overall wavelet vector ξ and thus retains only the information necessary to construct the *approximation* $x^{(j)}$ at scale j:

$$A(j) \triangleq \text{block diag}\left[0_{(2^J-2^j)},\ I_{(2^j)}\right], \tag{2.51}$$

where 0_p is a $p \times p$ matrix of zeros and I_q is a $q \times q$ identity matrix. Also it will prove convenient to define a similar matrix operator $D(j)$, that retains only the information in ξ pertaining to the *detail* component at scale j by zeroing all but the sub-vector

---

[10] We refer the reader to Section 2.B where show the form of W for the Haar case.

corresponding to $\xi^{(j)}$:

$$D(j) \triangleq \text{block diag}\left[0_{(2^J-2^{j+1})},\ I_{(2^j)},\ 0_{(2^j)}\right]. \tag{2.52}$$

Finally, with these definitions note that we have the following scale recursive relationship for the partially zeroed vectors, in the spirit of (2.49):

$$A^{(j+1)}\xi = A(j)\,\xi + D(j)\,\xi. \tag{2.53}$$

2.5.2 Representation of 2-D signals

We briefly describe next the multiresolution representation of 2-D signals [59]. We only state the form of the representation and refer the reader to [59] for implementational details. Further, we consider only the special class of separable multiscale representation where the 2-D approximation vector spaces are decomposed into a tensor product of two identical 1-D approximation vector subspaces [59].

Let $L^2(R^2)$ denote the vector space of measurable, square-integrable, 2-D signals $f(u,v)$, and let Z denote the set of integers. A multiscale approximation of $L^2(R^2)$ is a sequence of subspaces $\{V_j\}_{j\in Z}$ with $\ldots V_{-1} \subset V_0 \subset V_1 \ldots$ having the interpretation that the projection $A_j f(u,v)$ of $f(u,v)$ on V_j gives the approximation of $f(u,v)$ at scale j. Once again, as in the 1-D case, the scales become finer with increasing j. Now, there exists a unique function $\phi(u,v) = \phi(u)\phi(v) \in L^2(R^2)$, called the *scaling function*, such that for each scale j, $\{\phi_{j,l,m}(u,v) = 2^j\phi(2^j u - l, 2^j v - m) = \phi_{j,l}(u)\phi_{j,m}(v)\}_{(l,m)\in Z^2}$ is an orthonormal basis of V_j. Thus the approximation of the signal at scale j can be written as:

$$A_j f(u,v) = \sum_l \sum_m f^{(j)}(l,m)\phi_{j,l}(u)\phi_{j,m}(v), \tag{2.54}$$

with

$$f^{(j)}(l,m) = < f(u,v), \phi_{j,l}(u)\phi_{j,m}(v) > . \tag{2.55}$$

The difference in information between the approximation of the signal at successive scales j and $j+1$ is captured by the *detail signal* at scale j. This detail signal is obtained as the projection $D_j f(u,v)$ of $f(u,v)$ on the subspace O_j which represents the orthogonal complement of V_j in V_{j+1}. If $\psi(u) \in L^2(R)$ is the 1-D wavelet corresponding to the scaling function $\phi(u)$, then the following three wavelets can be defined:

$$\begin{aligned} \psi^1(u,v) &= \phi(u)\psi(v) \\ \psi^2(u,v) &= \psi(u)\phi(v) \\ \psi^3(u,v) &= \psi(u)\psi(v), \end{aligned} \tag{2.56}$$

such that for each scale j, $\{\psi^1_{j,l,m}(u,v) = 2^j\psi^1(2^j u - l, 2^j v - m), \psi^2_{j,l,m}(u,v) = 2^j\psi^2(2^j u - l, 2^j v - m), \psi^3_{j,l,m}(u,v) = 2^j\psi^3(2^j u - l, 2^j v - m)\}_{(l,m)\in Z^2}$ is an orthonormal basis of O_j. Thus the detail signal at scale j, capturing the difference in information between $A_j f(u,v)$ and $A_{j+1} f(u,v)$, can be written as:

$$\begin{aligned} D_j f(u,v) = & \sum_l \sum_m \rho_1^{(j)}(l,m)\psi^1_{j,l,m}(u,v) \\ & + \sum_l \sum_m \rho_2^{(j)}(l,m)\psi^2_{j,l,m}(u,v) \\ & + \sum_l \sum_m \rho_3^{(j)}(l,m)\psi^3_{j,l,m}(u,v), \end{aligned} \tag{2.57}$$

where

$$\begin{aligned} \rho_1^{(j)}(l,m) &= < f(u,v), \psi^1_{j,l,m}(u,v) > \\ \rho_2^{(j)}(l,m) &= < f(u,v), \psi^2_{j,l,m}(u,v) > \\ \rho_3^{(j)}(l,m) &= < f(u,v), \psi^3_{j,l,m}(u,v) > . \end{aligned} \tag{2.58}$$

The images $\rho_1^{(j)}$, $\rho_2^{(j)}$, and $\rho_3^{(j)}$ capture the object details in the vertical, horizontal, and diagonal directions respectively and, along with the approximation image $f^{(j)}$, define the separable 2-D multiscale representation.

2.6 Existing multiresolution tomographic frameworks

In all of the existing wavelet-based multiresolution methods for tomography [67, 68], the object is represented in a separable 2-D wavelet basis of the type described in the last section[11]. The coefficients of expansion in this representation are then determined from the projection data by assuming that the data are dense and of high quality so that the various Radon transform results hold.

Specifically, in order to obtain a 2-D multiresolution representation for the object (or the reconstruction) the approximation coefficients $f^{(j)}$ (2.55) and the three detail coefficients $\rho_1^{(j)}$, $\rho_2^{(j)}$, and $\rho_3^{(j)}$ (2.58) have to be determined from the projection data. By a slight change of variables in (2.55) and (2.58), these approximation and detail coefficients can be seen to be a result of the 2-D convolution of the object $f(u,v)$ with a corresponding scaling function or a wavelet:

$$\begin{aligned}
f^{(j)}(l,m) &= (f(u,v) ** \phi_{j,0}(-u)\phi_{j,0}(-v))(2^{-j}l, 2^{-j}m) \\
\rho_1^{(j)}(l,m) &= (f(u,v) ** \phi_{j,0}(-u)\psi_{j,0}(-v))(2^{-j}l, 2^{-j}m) \\
\rho_2^{(j)}(l,m) &= (f(u,v) ** \psi_{j,0}(-u)\phi_{j,0}(-v))(2^{-j}l, 2^{-j}m) \\
\rho_3^{(j)}(l,m) &= (f(u,v) ** \psi_{j,0}(-u)\psi_{j,0}(-v))(2^{-j}l, 2^{-j}m), \qquad (2.59)
\end{aligned}$$

where $**$ refers to the 2-D convolution. Now, it is simple to see from (2.59) as to how the approximation and detail coefficients can be calculated from the projection data. Before we explain this however, let us state an elementary result for the Radon transform of a convolution of two 2-D functions. This result says that the projection of a 2-D convolution of two functions is the same as a 1-D convolution of their individual projections. Thus, in order to calculate the approximation and detail coefficients in (2.59) which define the 2-D multiresolution representation of the reconstruction, all we

---

[11] The only exception to this is the work in [64] where a radial wavelet basis is used instead of the separable wavelet basis. In that case, the discussion in this section still holds with a minor modification, namely that the radial wavelet basis only admits a single detail coefficient rather than the three detail coefficients in case of the separable wavelet basis (refer to (2.58)).

have to do is to convolve the object projection data with projections of the separable scaling function and wavelets from (2.59), prior to ramp-filtering and back-projection. Note however that this multiresolution framework is only valid in the dense and noise less data scenario since it assumes that the Radon transform results hold. Thus the existing multiscale reconstructions all of which are based on the framework described above, break down in the case when the data are sparse and/or noisy.

In this thesis we develop an efficient reconstruction technique that provides reconstructions at multiple scales from both sparse as well as noisy data. Our multiscale reconstruction technique is based on the natural pixel (NP) object representation for tomography. Since the NP representation results in a matrix based reconstruction method which uses no restrictive assumptions on the quantity of the data, our multiscale reconstruction technique is valid for the case of sparse data. Further in the case that the projection data are dense, we are able to use the Radon transform results to specialize our already efficient NP-based multiscale reconstruction technique such that it yields reconstructions at multiscale scales with even less computations. In particular, our multiscale reconstructions from dense data require less computations than those obtained from the existing methods that we just described. Finally, we extend both our sparse as well as dense data multiscale reconstruction techniques to yield statistically optimal multiscale estimates from noisy data with no additional computations. This efficiency is due to the fact that in our multiscale representation the object coefficients "live" in the projection (i.e. the data) domain. As we will see later, an important and powerful class of statistical prior models for the object is naturally captured in this multiscale projection domain. This in turn results in simplified expressions for the estimators, which can thus be solved for in an efficient manner in our multiscale framework.

2.4 Some practical considerations

For the examples in this thesis we do not generate the projection matrix T by calculating the areas of intersection of the strips with the pixels, but for the sake of computational efficiency we use the Donner Reconstruction Library [50] to generate the projection matrix. Specifically, we obtain the columns of T from the Donner Library by reading out the projection data y corresponding to a choice of impulse functions for the pixel coefficients f. However, since this method only yields an approximate projection matrix T, the NP system matrix C generated by explicitly forming the product TT^T in this case suffers from gross inaccuracies. To avoid this problem, we form C by explicitly calculating the areas of intersection between the various NP strips. In general we observe that if the projection data are generated according to the Donner Reconstruction Library, and if the matrix C that is formed explicitly by calculating the areas of intersections is used for the NP reconstruction, a mis-alignment error of the order of one pixel generally occurs in the resulting reconstruction.

2.B Multiscale representation matrix W for the Haar case

We show below the form of the multiscale representation matrix W for the Haar wavelet by assuming the finest scale $J = 3$ (see (2.50)):

$$W = \frac{1}{\sqrt{2}} \left[\begin{array}{cccccccc}
1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\
\hline
\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\
\hline
\frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\
\hline
\frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2}
\end{array} \right] \qquad (2.60)$$

Chapter 3

Using Natural Wavelet Bases And Multiscale Stochastic Models For Tomographic Reconstruction

3.1 Introduction

In this chapter we consider the solution of ill-posed tomographic reconstruction problems where the projection data are noisy and sparse. Recall from Chapter 2 that the conventional techniques for image reconstruction from sparse data (for example, the natural pixel (NP) method) are computationally intensive. In contrast, our multiscale reconstruction technique yields fast reconstructions from sparse data and can be extended to yield statistically optimal reconstructions from noisy data with very little added computational complexity. In addition, our multiresolution framework for tomographic reconstruction is natural or desirable if the ultimate objectives are multiresolution in some way, for example if the interest is not to fully reconstruct the field but to gather information about aggregate (i.e. coarse scale) or fine scale (for example, boundaries) features of the field. Using conventional techniques we would first have to reconstruct the entire field and then use post-processing to extract such features.

To develop our multiscale reconstruction technique we start with the natural pixel (NP) object representation [9, 10] (c.f. Section 2.2.2) which was originally developed for the sparse data tomography problem. The NP representation results in a matrix based reconstruction method which has the advantage that the resulting reconstructions are devoid of many of the sparse data artifacts present in the FBP reconstruction. The disadvantage of the NP reconstruction, or matrix based reconstruction methods in general, is that solutions of very large, generally ill-conditioned systems of equations are required.

In this chapter, we build on the NP approach by using wavelet bases to transform the NP strips. The standard NP system matrix, relating the input (the object coefficients) and the output (the projection data), is *full*. The use of wavelet bases results in a transformation matrix which is *sparse*. In addition, the coarsest scale elements of this matrix capture any ill-conditioning arising from the geometry of the imaging system. We exploit this feature to partition the multiscale system matrix by scales and obtain a reconstruction procedure that only requires inversion of a well-conditioned and sparse matrix. The use of wavelet bases also enables us to formulate a multiscale tomographic reconstruction technique wherein the object is reconstructed at multiple scales or resolutions. The overall reconstruction is obtained by combining the reconstructions at different scales.

Noisy imaging problems arise in a variety of contexts (c.f. Chapter 1) and in such cases standard matrix based reconstruction methods (including NP) often yield unacceptable results. These situations generally reflect the fact that more degrees of freedom are being sought than are really supported by the data and hence some form of regularization is required. In contrast to the standard NP method, we are able to extend our multiscale reconstruction technique in the case of noisy projections to obtain a statistically regularized, multiscale maximum aposteriori probability (MAP) object estimate. We do this by realizing that for ill-posed problems the lower resolution (i.e. the coarser scale) reconstructions are often more reliable than their higher resolution counterparts and by using prior statistical models constructed *directly in the multiscale domain* which capture such intuition. Our multiscale MAP regularized reconstructions are no more computationally intensive than our unregularized multiscale reconstructions.

Finally, while the work here focuses on the case of sparse data, when complete data are available, additional efficiencies may be obtained through the use of explicit Radon transform inversion formulas, such as FBP. In such complete data cases, the multiscale methodology described herein may be applied using the FBP method as a starting point to obtain both unregularized and regularized multiscale reconstructions with the same computational complexity as the FBP reconstruction. Such application is described in Chapter 5.

The chapter is organized as follows. In Section 3.2 we review the relevant results from the NP method. This is followed by Section 3.3 where we develop the theory behind our wavelet transform-based multiscale reconstruction method starting from the NP object representation. In Section 3.4 we build on this framework to provide a method for obtaining MAP regularized reconstructions from noisy data. The conclusions are presented in Section 3.5. Appendices 3.A– 3.F contain certain technical details.

3.2 Review of the natural pixel (NP) reconstruction results

In this section we briefly review the relevant results from Chapter 2. First, the measurements in tomography are of the form:

$$y(m) = \iint\limits_{\Omega} f(u,v)\, S_m(u,v)\, du\, dv, \qquad m = 1, \ldots, N_\theta N_s, \tag{3.1}$$

and can be stacked to form an overall observation vector:

$$y = \left[\begin{array}{c} y_1 \\ \hline y_2 \\ \hline \vdots \\ \hline y_{N_\theta} \end{array} \right], \tag{3.2}$$

where y_k, the projection at angle k, is defined as:

$$y_k \triangleq \begin{bmatrix} y((k-1)N_s + 1) \\ y((k-1)N_s + 2) \\ \vdots \\ y((k-1)N_s + N_s) \end{bmatrix}. \tag{3.3}$$

In the NP representation, the object is expanded in the same strips S along which the projection data are collected:

$$f(u,v) = \sum_{n=1}^{N_\theta N_s} x(n) S_n(u,v). \tag{3.4}$$

The above representation results in the following observation equation:

$$y = Cx, \tag{3.5}$$

where x is the overall object coefficient vector containing the coefficients $\{x(m), m = 1, \ldots, N_\theta N_s\}$, and can be partitioned into different angle components as follows:

$$x = \left[\begin{array}{c} x_1 \\ \hline x_2 \\ \hline \vdots \\ \hline x_{N_\theta} \end{array} \right]. \tag{3.6}$$

The elements of the matrix C are the areas of intersection of the various strips S. In particular, the (m, n)-th element of C is given by:

$$C_{mn} = \iint\limits_{\Omega} S_n(u,v)\, S_m(u,v)\, du\, dv, \qquad m = 1, \ldots, N_\theta N_s;\ n = 1, \ldots, N_\theta N_s. \quad (3.7)$$

Further, as we had mentioned in Chapter 2, the matrix C can be written as:

$$C = TT^T, \quad (3.8)$$

where T is the projection matrix containing the discretized strips (c.f. (2.5)).

In the NP method, the reconstruction problem amounts to solving for the coefficient vector x from the observation equation (3.5), given the overall data vector y. Once x is determined, the reconstruction is obtained via the back-projection specified by the representation (3.4). Even though no assumption about discretization of the object f on the rectangular pixel grid is needed to solve for x from (3.5), a discretization of the representation (3.4) is needed to display the reconstruction once the coefficients in x have been determined. This discretization of (3.4) results in the following pixel coefficients for the object (c.f. (2.12)):[1]

$$f = T^T x = \sum_{k=1}^{N_\theta} T_k^T x_k, \quad (3.9)$$

where T_k is the block of the projection matrix associated with strips at angle k. Finally, it is not easy to solve the observation equation (3.5) for the coefficient vector x, since the NP matrix C is large ($N_\theta N_s \times N_\theta N_s$), full, and ill-conditioned.

---

[1] Note that while in (2.12) we discretize the object on a $N_s \times N_s$ pixel grid, this is not essential and indeed the discretization may be independent of the acquisition process.

3.3 A multiscale approach to natural pixel (NP) reconstruction

3.3.1 Multiscale transformation of the NP strips

The multiscale reconstruction is motivated by the following observations. Recall that the elements of the NP matrix C are the areas of intersection of the strip functions S (c.f. (3.7)). If we are able to modify these strips so that they are mostly orthogonal, then the corresponding areas of intersection will be nearly zero, resulting in a sparse matrix and a simplified solution for the underlying object coefficients x according to the observation equation (3.5). Specifically, suppose that we are able to modify the strips such that they have the form shown in Figure 3-1. Each strip is a linear combination of two NP strips, one given a positive weight and the other negative. The new matrix relating the object coefficients and the projection data, according to the above choice of strips, will have as its elements the (signed) areas of intersections of the newly defined strips. It is clear from Figure 3-1 that most of these elements will be zero due to the cancellation of the positive and the negative terms. Only those elements that correspond to strip intersections near the edge of the field-of-view will be non-zero. Thus we can expect this new matrix to be *sparse* with the degree of sparsity increasing with the size of the field-of-view (since the fraction of intersections near the edge decreases with increasing field size).

The above redefinition of the basis strips with positive and negative weights is reminiscent of the finest level of a Haar transform and in fact we can imagine repeating the process at other levels as well. However, an important point to note is that here the Haar transform is taken only in one direction, i.e. the direction perpendicular to the long axis of the strip. This is the key to our multiscale reconstruction method: we expand the basis functions $\{T_k\}$ in a 1-D wavelet basis which then induces a corresponding 2-D multiscale object representation[2]. For a projection at a fixed angle

---

[2] As explained in Section 2.6, our multiscale object representation is fundamentally different from previous multiscale-related representations for tomography (for example, [64, 67, 68]). In these approaches a direct 2-D expansion of the object (i.e. a 2-D wavelet transform) is used, the coefficients

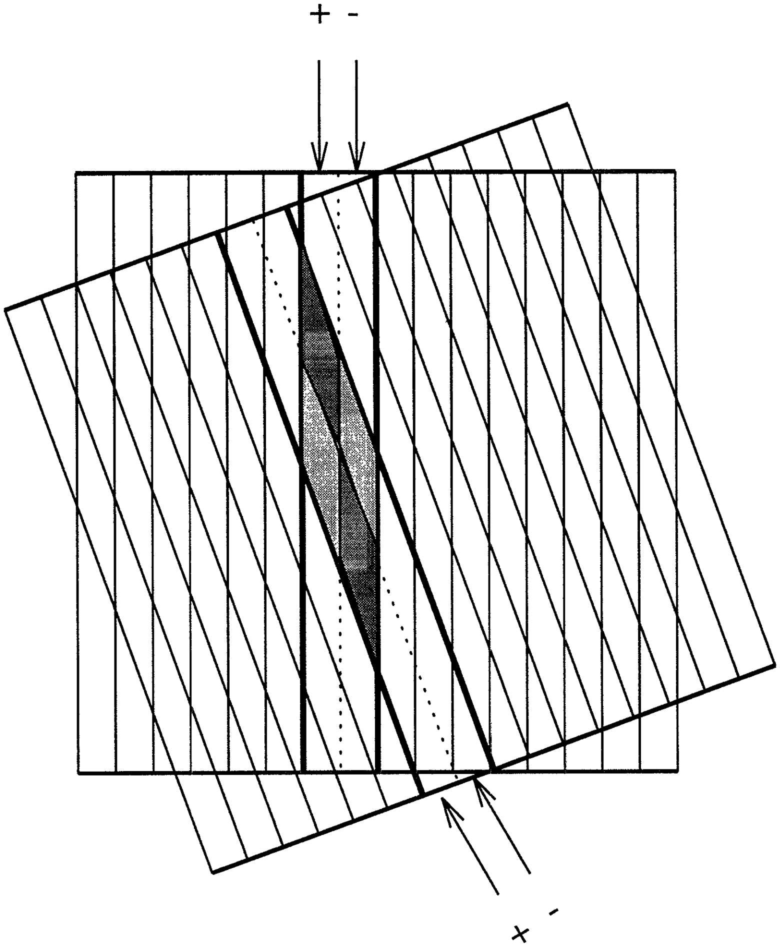


Figure 3-1: Each of the two delineated strips from Figure 2-4 are broken down into two substrips, having a positive and a negative weight respectively. The area of intersection of the two strips is zero in this case due to the cancellation of the positive (lightly shaded) and the negative (darkly shaded) terms.

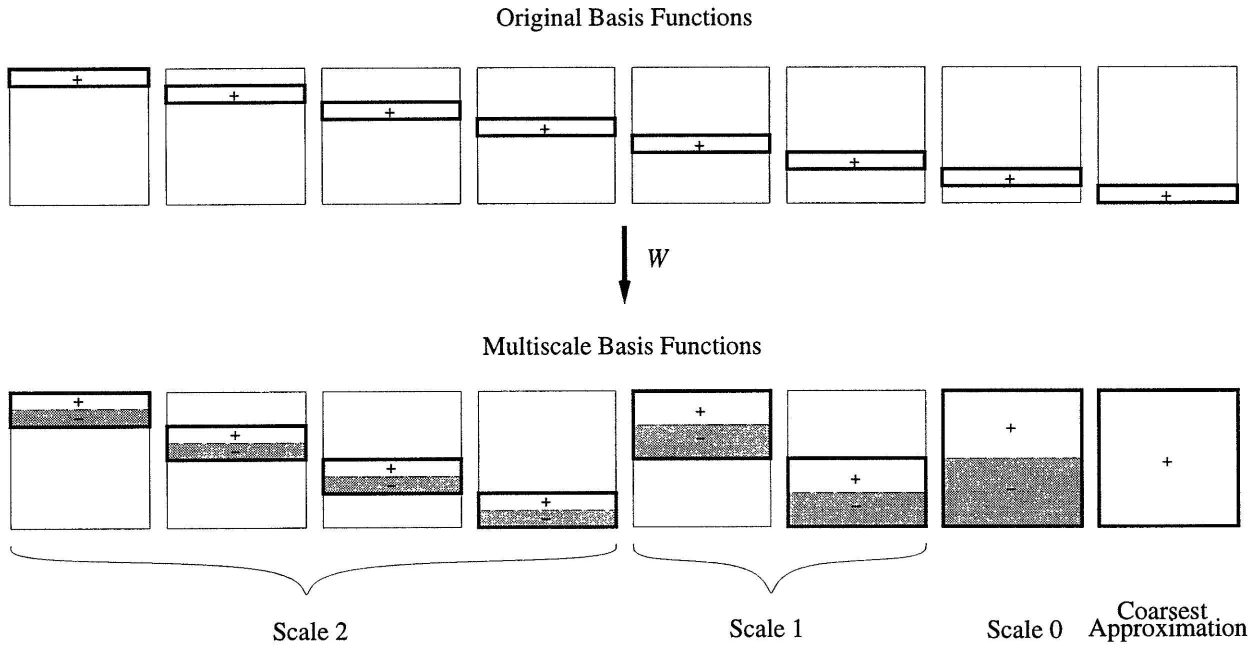


Figure 3-2: Original NP strip basis functions contained in T_k (shown in the top half of the figure) and multiscale transformed strip basis functions of $\mathcal{T}_k$ (shown in the bottom half of the figure) for a fixed angle k. The multiscale basis functions are grouped into different scale components based on their spatial extent. The Haar wavelet is used for multiscale decomposition in this example. The heavy boundaries indicate the extent of support of the corresponding basis function and the shading and $+/-$ indicate the sign of the basis function over the region.

k (and $N_s = 8$), the full Haar transform of the original strip basis functions shown in Figure 3-2 that includes contributions from all levels will look as shown in the bottom half of Figure 3-2. A notion of scale emerges from the use of the Haar transform. The original strips have been broken down into a series of strips at multiple scales having positive and negative weights. The finest scale involves strips that have twice the width of the original strips and the coarsest scale involves strips extending over the entire field-of-view. We call the above transformed strip functions the *natural wavelet basis* because of the adaptation of the natural pixel representation. While we have used the Haar wavelet in the multiscale transformation described above and shown in Figure 3-2, one can imagine using a more general wavelet for the same purpose, as discussed next.

---

of which are then calculated from the projection data. Refer to Section 2.6 for further details.

In particular, let W be a matrix representation of the linear operator that performs a 1-D orthonormal wavelet transform on a discrete sequence of finite length N_s (as described in Section 2.5.1) so that $W^{-1} = W^T$. Further, let $W_b = \text{block diag}(W)$ be a block-diagonal matrix with N_θ blocks along the diagonal, all equal to W (so that again $W_b^{-1} = W_b^T$). We define our general multiscale transformation of the strip basis functions as:

$$\mathcal{T} \triangleq W_b T, \tag{3.10}$$

where the matrix $\mathcal{T}$ contains the multiscale strip basis functions at all the different angles:

$$\mathcal{T} = \left[\begin{array}{cccccc} \mathcal{T}_1^T & \mathcal{T}_2^T & \ldots & \mathcal{T}_k^T & \ldots & \mathcal{T}_{N_\theta}^T \end{array} \right]^T, \tag{3.11}$$

with the matrix $\mathcal{T}_k$ containing the discretized multiscale strips at angle k, is given by:

$$\mathcal{T}_k = W T_k. \tag{3.12}$$

Before proceeding, let us consider the transformed bases functions contained in $\mathcal{T}_k$ in more detail. Recall from Section 2.2 that the rows of T_k are composed of the (discretized) original strip basis functions at angle k along which the data are collected. Similarly the rows of the transformed matrix $\mathcal{T}_k$ contain the corresponding (discretized) multiscale object strip basis functions at angle k. The wavelet transform operator matrix W, acting identically on each column of T_k, thus forms the new multiscale basis functions at that angle from linear combinations of the corresponding original strip functions, where these linear combinations correspond precisely to a *1-D* wavelet transform perpendicular to the projection direction. This transformation of the basis functions is shown schematically in Figure 3-2 (which corresponds to the case of the Haar wavelet). The original strip basis functions (rows of T_k) are illustrated in the top half of the figure, while the corresponding collection of multiscale basis functions (rows of $\mathcal{T}_k$) are shown in the bottom half. Note that the number of basis elements in the original (top half) and the multiscale (bottom half) framework are the same, as they must be since the multiscale framework involves an orthonormal change of basis.

Suppose now we define the vectors:

$$\eta \triangleq W_b y, \tag{3.13}$$

$$\xi \triangleq W_b x, \tag{3.14}$$

which contain the stacked set of wavelet coefficients of the projection data $\eta_k = W y_k$ and the object coefficients $\xi_k = W x_k$ at each angle k. Then by applying (3.10) to (3.5) we get the following relationship between the multiscale representation of the object coefficients, ξ, and the multiscale representation of the data, η:

$$\eta = \mathcal{C}\xi, \tag{3.15}$$

where the multiscale system matrix $\mathcal{C}$ is given by:

$$\mathcal{C} = W_b C W_b^T = W_b(TT^T)W_b^T = \mathcal{T}\mathcal{T}^T. \tag{3.16}$$

Note that (3.16) implies that the elements of the transformed matrix $\mathcal{C}$ are the (signed) areas of intersection of the various multiscale basis functions $\mathcal{T}$. From our previous discussion we expect $\mathcal{C}$ to be *sparse* if W reflects the use of the Haar wavelet for multiscale decomposition. Later we will see that the use of any compactly supported Daubechies wavelet results in approximately the same sparsity as that achieved in the Haar case.

Finally, by combining (3.14) and (3.9) we obtain the following representation of the object in the multiscale domain:

$$f = \mathcal{T}^T \xi = \sum_{k=1}^{N_\theta} \mathcal{T}_k^T \xi_k. \tag{3.17}$$

Note that (3.17) can be interpreted as the back-projection operation where the multiscale coefficients ξ_k are back-projected along the multiscale basis functions $\mathcal{T}_k$ at each angle k, and then the contributions from all N_θ angles are added to get the overall reconstruction f. Before proceeding, we note that the multiscale object representation (3.17) permits us to define object reconstructions at multiple scales, which we discuss in Section 3.3.4.

For the development to follow it will prove convenient to order the multiscale vectors η and ξ according to scales rather than projection angles, with the finest scale detail terms from all projections grouped first and the coarsest scale approximation terms grouped last. We refer the reader to Appendix 3.A for details on the ordering schemes, and simply state here that the re-arrangement of (3.15) results in the following scale ordered and partitioned observation equation:

$$\left[\begin{array}{c} \eta^d \\ \hline \eta^a \end{array}\right] = \left[\begin{array}{c|c} \mathcal{C}_{dd} & \mathcal{C}_{da}^T \\ \hline \mathcal{C}_{da} & \mathcal{C}_{aa} \end{array}\right] \left[\begin{array}{c} \xi^d \\ \hline \xi^a \end{array}\right] = \mathcal{C}_s \left[\begin{array}{c} \xi^d \\ \hline \xi^a \end{array}\right], \tag{3.18}$$

where the vectors η^d and ξ^d contain all the detail terms at various scales and angles and are of length $N_\theta(N_s - 1)$, and the vectors η^a and ξ^a contain the coarsest scale approximation, i.e. the DC, terms at all angles (one for each angle) and are of length N_θ. Thus $\mathcal{C}_s$ is simply a scale reordered version of $\mathcal{C}$. The upper left block $\mathcal{C}_{dd}$ is a $N_\theta(N_s - 1) \times N_\theta(N_s - 1)$ symmetric matrix, the elements of which are the areas of intersection of the detail basis functions (i.e. strips) at various scales and angles. The lower right block $\mathcal{C}_{aa}$ is a $N_\theta \times N_\theta$ symmetric matrix the elements of which are the areas of intersection of the coarsest scale basis function at each angle. Finally, the off-diagonal block $\mathcal{C}_{da}$ is a $N_\theta \times N_\theta(N_s - 1)$ matrix, the elements of which are the areas of intersection of the coarsest scale approximation basis functions and the detail basis functions at various scales.

3.3.2 Multiscale matrix sparsity calculations

The degree of sparsity of $\mathcal{C}$ and $\mathcal{C}_s$ is exactly the same because the elements in these matrices are the same modulo a permutation. Figure 3-3, right, shows $\mathcal{C}_s$ for an imaging geometry with $N_\theta = 32$ angular positions and $N_s = 32$ strips per each angular position. We have used the Haar wavelet for this multiscale representation.

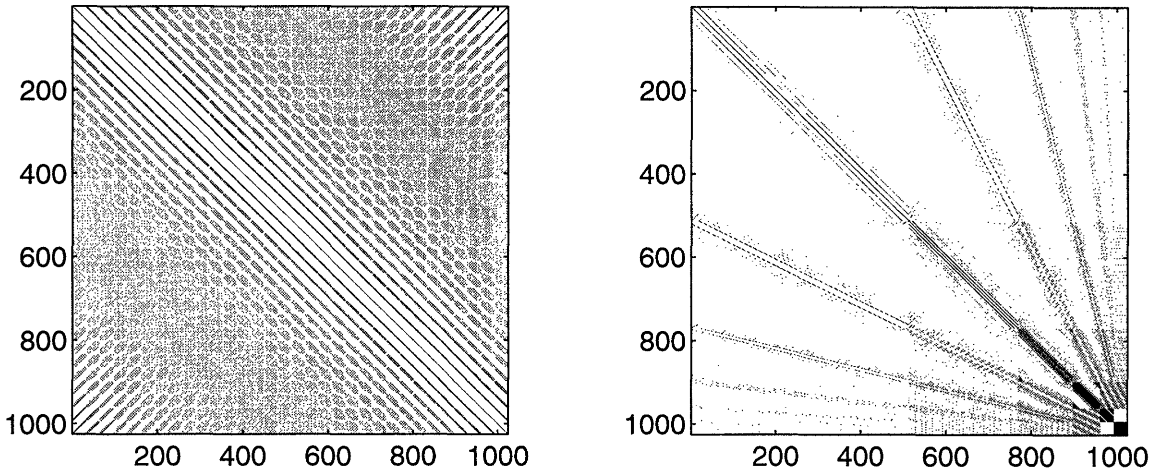


Figure 3-3: The grayscale plots (black corresponds to the maximum value and white to the minimum) of: Left: The natural pixel system matrix, C; Right: The multiscale scale ordered system matrix, C_s, for an imaging geometry with $N_\theta = N_s = 32$. The Haar wavelet is used for multiscale decomposition. The matrices are thresholded to display elements that are equal to or greater than 2% of the absolute maximum.

Comparing with Figure 3-3, left, which shows the corresponding NP system matrix, C, we see that C_s (hence C) is considerably sparser than C. From the figure, most of the non-zero elements in C_s correspond to the coarser scale terms in the lower right part of the matrix where field-of-view edge effects are more pronounced. We had claimed earlier that the degree of sparsity of C increases as the size of the field-of-view increases. This claim is validated by Figure 3-4 where we plot the degree of sparsity of C as a function of the size of the field ($N_\theta = N_s$). We measure the degree of sparsity by the percentage of elements in the matrix C (or C_s) which are equal to or below a certain percentage of the absolute maximum. Figure 3-4 reports the sparsity calculations for three different values of threshold, namely 0.0%, 0.5% and 2.0% of the absolute maximum. It is empirically observed that setting all values in C below a threshold of 2.0% to zero makes no visible difference to the reconstructions. From Figure 3-4, we see that for the case of $N_\theta = N_s = 128$ and a threshold of 2.0%, C is 98.75% sparse (or, equivalently, 1.25% full). In Figure 3-5, we show the degree of sparsity of C for $N_\theta = N_s = 32$ achieved by the Haar wavelet and the Daubechies wavelets D_3 and D_8. From the figure we see that the number of elements that are exactly zero decrease as wavelets with larger support are used. This is expected because the

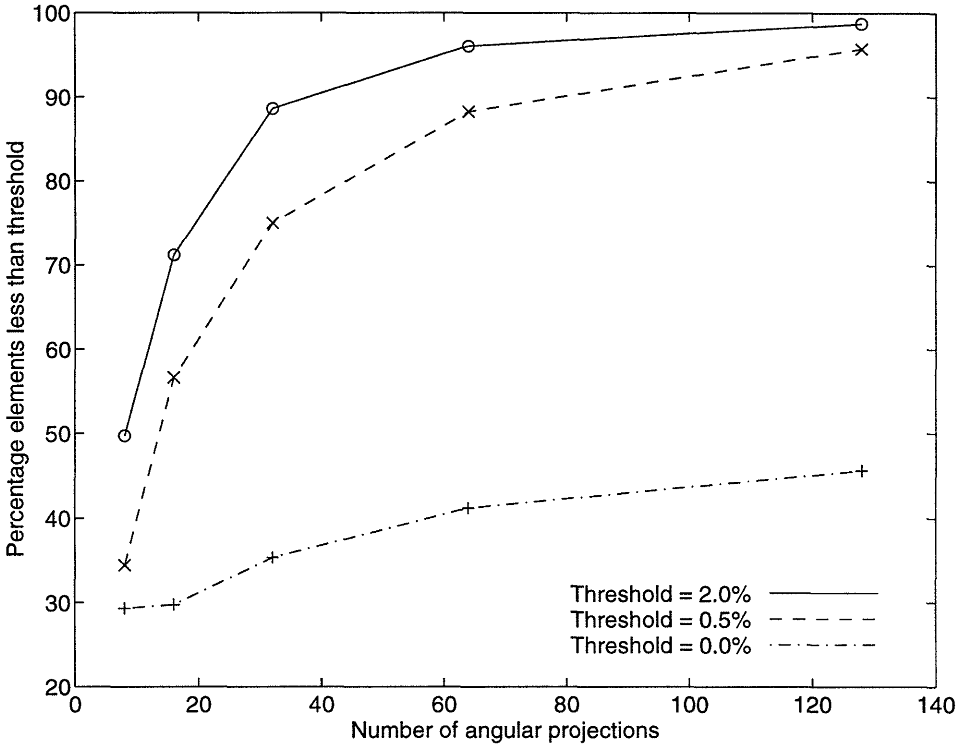


Figure 3-4: The degree of sparsity (percentage of elements below a threshold) of multiscale system matrix, $\mathcal{C}$, as a function of N_θ ($=N_s$). The Haar wavelet is used here for multiscale decomposition. Setting all elements in $\mathcal{C}$ below 2.0% of the maximum to zero makes no visible difference to the reconstructions.

number of incomplete strip intersections near the edge of the field-of-view, resulting in a non-zero value for the corresponding elements of $\mathcal{C}$, increase for wavelets with larger support. However, since a threshold of 2.0% does not affect the reconstructions, the effective sparsity achieved by the Haar, D_3 and D_8 is approximately the same.

3.3.3 Multiscale object coefficient determination

Let us turn our attention next to the calculation of the object coefficients ξ which, through the back-projection equation (3.17), specify the reconstruction f. Recall that the vector ξ consists of the object detail and approximation coefficients, ξ^d and ξ^a respectively, which are related to the projection data through (3.18). For clarity, let us first consider the ideal case of an infinite field extent where edge effects are absent. In this case, as argued in Section 2.2.2, the NP matrix C is rank deficient due to the non-uniqueness of the NP representation. Thus the multiscale matrix C_s

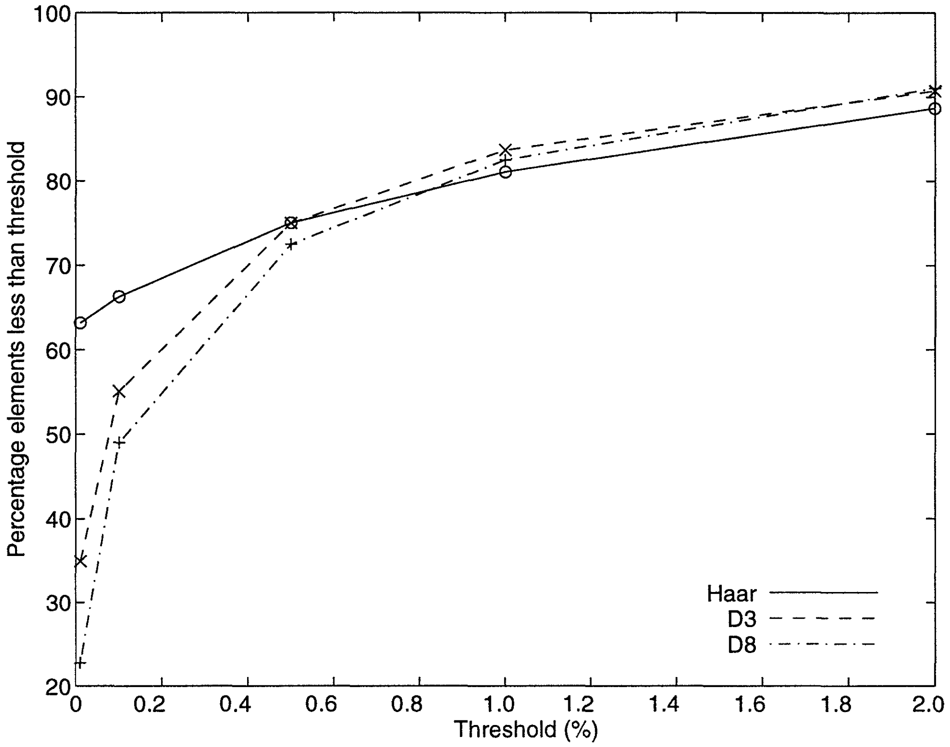


Figure 3-5: The degree of sparsity (percentage of elements below a threshold) of multiscale system matrix, C, as a function of threshold, for different wavelets. Here $N_\theta = N_s = 32$. Setting all elements in C below 2.0% of the maximum to zero makes no visible difference to the reconstructions.

(or equivalently $\mathcal{C}$) is also rank deficient since it is related to C through a change of basis. Since a unique solution does not exist in this case, a rational approach is to seek the minimum norm solution to (3.5) or, in the multiscale domain, to (3.18). The NP matrix equation (3.5) represents a large and full system of equations and so it is difficult in practice to find the minimum norm solution in this case. The multiscale relationship (3.18), however, has a structure that can be exploited to simplify the computations. In particular, first note that in the ideal case with no edge effects, the elements of $\mathcal{C}_{da}$, capturing the areas of intersection between the coarsest scale approximation and finer scale detail basis functions, are identically zero[3]. Further, one can show that the matrix $\mathcal{C}_{dd}$ has full rank, while the matrix $\mathcal{C}_{aa}$ is rank deficient[4]. This is hardly surprising in view of our earlier discussion in Section 2.2.2 since we have grouped all the terms contributing to the DC value of the object (and hence to the ill-conditioning of C or C_s) in the $\mathcal{C}_{aa}$ block. Thus our multiscale transformation has served to "compress" and isolate the non-uniqueness that is present in the NP representation. Now, with $\mathcal{C}_{da} = 0$, the minimum norm solution to (3.18) is easily found since the detail and the approximation equations decouple. In particular, the minimum norm solution is given by:

$$\xi^d = \mathcal{C}_{dd}^{-1}\eta^d, \tag{3.19}$$

$$\xi^a = \mathcal{C}_{aa}^{+}\eta^a, \tag{3.20}$$

where $\mathcal{C}_{aa}^{+}$ is the pseudo-inverse of $\mathcal{C}_{aa}$ [32]. Thus (3.19) captures the unique part of the solution and (3.20) provides a particular distribution of the DC components of the object over the coefficients in ξ^a. Before proceeding we note that since $\eta^d = \mathcal{C}_{dd}\xi^d$ represents a large but sparse system of equations we will not, in practice, find the solution to this system by explicitly calculating the inverse of $\mathcal{C}_{dd}$ (as suggested by (3.19)) but rather we will exploit it's sparse structure and use any of the methods

---

[3] This is because these elements represent the area under each wavelet (which is zero) due to the fact that the coarsest scale approximation basis functions are just indicator functions over the field-of-view.

[4] In fact for this ideal case it is easy to see that $\mathcal{C}_{aa}$ is an $N_\theta \times N_\theta$ matrix of ones scaled by a constant.

created especially to solve such sparse systems [32]. In fact, in Chapter 4 we will use the Kaczmarz method (c.f. Section 2.2.1) to solve the system of equations $\eta^d = C_{dd}\xi^d$. This method will exploit the sparsity of C_{dd} to yield an extremely efficient iterative reconstruction algorithm in the multiscale framework. Further, since C_{dd} is well-conditioned, the reconstructions obtained by the Kaczmarz method will not be affected by the choice of the ordering scheme or the relaxation parameter. Recall that this choice is critical for convergence in the algebraic reconstruction technique (ART) [40].

Finally, if we assume that the object is completely contained in the field-of-view, then we obtain the following simplification of (3.20) for ξ^a (c.f. Appendix 3.B for details):

$$\xi^a = (\frac{1}{N_\theta N_s})\eta^a = (\frac{\mu(f)}{N_\theta N_s \sqrt{N_s}})1_{N_\theta}, \tag{3.21}$$

where 1_p refers to a vector of length p with all elements equal to unity, and $\mu(f)$ is the total mass under the object and is defined as:

$$\mu(f) = \iint_{\mathbf{\Omega}'} f(u,v)\, du\, dv, \tag{3.22}$$

where $\mathbf{\Omega}'$ is the region of support of the object, which is assumed to be completely contained in the field-of-view of all N_θ projections.

The development to this point has focused on the ideal case in which field-of-view edge effects are absent. In any practical situation the field is finite and such effects arise. However, as we discuss below, they have minimal impact on the preceding development. First, due to these edge effects the off-diagonal block C_{da} in (3.18) is no longer zero. It will in general have a few non-zero elements corresponding to the areas of intersection of the basis functions near the domain boundary (see Appendix 3.C for numerical bounds on the absolute values of elements in C_{da} for the Haar case). Even though a variety of methods exist for modifying the solutions (3.19) and (3.20) to account for this neglected coupling[5], our experience is that practical reconstructions

---

[5] For example, in Appendix 3.E we use the matrix inversion lemma [41] to refine the estimate of ξ^d in (3.20) by introducing the first and second order C_{da} coupling.

based on (3.19) and (3.20) (which assume $\mathcal{C}_{da} = 0$) are visually indistinguishable from ones where a correction is made for the coupling. As a result we use (3.19) and (3.20) in our multiscale algorithm and for all the reconstructions we present in this chapter.

The other impact the inclusion of edge effects has is to change the structure of $\mathcal{C}_{aa}$. In particular $\mathcal{C}_{aa}$ is no longer truly singular (unless we take views exactly 90^0 apart) though it is nearly so. In any case, $\mathcal{C}_{aa}$ is still a circulant matrix with row sums nearly equal to the case when the edge effects are neglected (c.f. Appendix 3.D), and hence (3.21) is still valid. Finally, since the edge effects have no significant impact on the matrix $\mathcal{C}_{dd}$, the latter still has full rank and is well-conditioned.

From the preceding discussion we see that the inclusion of edge effects corresponding to a practical imaging geometry have no significant influence on the structure of the reconstruction algorithm that we had developed for the ideal geometry. We thus present next the algorithm for obtaining our NP-based multiscale reconstructions that uses the ideal approximation where the edge effects are ignored.

Algorithm 1 (Multiscale Reconstruction)

1. *For a given choice of wavelet, form the sparse multiscale matrix* $\mathcal{C} = W_b\, C\, W_b^T$ *(the multiscale counterpart of the original NP matrix* C*). Re-arrange the elements in* $\mathcal{C}$ *according to scales (c.f. Appendix 3.A) to form* $\mathcal{C}_s$*. The block* $\mathcal{C}_{dd}$ *in* $\mathcal{C}_s$ *is associated with the detail-detail coupling (c.f. (3.18)).*

2. *Find the multiscale observations* η *by taking the 1-D wavelet transform of the projection data,* $\eta = W_b y$*. Re-arrange the elements in* η *according to scales to obtain* η^d *and* η^a*, the multiscale detail and the coarsest scale approximation data sub-vectors respectively (c.f. Appendix 3.A).*

3. *Calculate the coarsest scale approximation object coefficients* $\xi^a = (1/(N_\theta N_s))\eta^a$*.*

4. *Calculate the multiscale detail object coefficient vector* ξ^d *by solving the sparse and well-conditioned system of equations* $\eta^d = \mathcal{C}_{dd}\xi^d$*.*

5. *Calculate the overall observation vector* ξ *by re-arrangement of* ξ^a *and* ξ^d *according to projection angles (c.f. Appendix 3.A). The k-th block* ξ_k *of* ξ *is now*

associated with angle k (c.f. (3.6)).

6. *For each angle k, back-project ξ_k along the corresponding multiscale basis functions $\mathcal{T}_k$, $\mathcal{T}_k^T \xi_k$.*

7. *Combine the object contributions from the individual back-projections at each angle to obtain the overall reconstruction, $f = \sum_k \mathcal{T}_k^T \xi_k$.*

3.3.4 Object reconstruction at multiple scales

So far we have simply transformed the representation of the original finest scale object f (c.f. (3.17)). But the discussion in preceding sections together with the development in Section 2.5 suggests how to use our new multiscale representation ξ_k and corresponding basis functions $\mathcal{T}_k$ to obtain a multiscale decomposition of the object estimate in the original space. Such a multiresolution decomposition can be obtained through (3.17) by using a series of approximations $x_k^{(j)}$ to x_k at successively finer scales, thereby inducing a series of corresponding approximate representations of the object. In particular, we *define* the j-th scale approximation $f^{(j)}$ to f as:

$$f^{(j)} \triangleq \sum_{k=1}^{N_\theta} \mathcal{T}_k^T \left(A(j)\, \xi_k \right), \tag{3.23}$$

where recall that the elements $(A(j)\, \xi_k)$ of the j-th scale approximation $x_k^{(j)}$ are obtained by zeroing the finer scale components in the vector of 1-D wavelet transform coefficients of ξ_k, as discussed in Section 2.5 (c.f. (2.51)). Thus the approximation $f^{(j)}$ uses only the j coarsest scale components of the full vector ξ_k. Similarly, by $\Delta f^{(j)}$ we denote the additional detail required to go from the object approximation at scale j to that at scale $(j+1)$, which is given by:

$$\Delta f^{(j)} \triangleq \sum_{k=1}^{N_\theta} \mathcal{T}_k^T \left(D(j)\, \xi_k \right), \tag{3.24}$$

where recall that the detail vector $(D(j)\, \xi_k)$ is obtained by zeroing all but the corresponding level of detail $\xi_k^{(j)}$ in ξ_k (c.f. (2.52)). Combining the object approximation and detail definitions (3.23) and (3.24) with the scale recursive relationship (2.53) we see that the *object itself* satisfies the following scale recursive relationship, whereby the object approximation at the next finer scale is obtained from the approximation at the current (coarser) scale through the addition of the incremental detail at this scale, just as for the 1-D case treated in Section 2.5:

$$f^{(j+1)} = f^{(j)} + \Delta f^{(j)}. \tag{3.25}$$

Note that our multiscale object representation given in (3.23) and corresponding scale recursive construction (3.25) is induced naturally by the structure of the individual 1-D wavelet-based multiscale decompositions at each angle k and is *not* simply a 2-D wavelet transform of the original object estimate f. In other words, we are not simply relating the coefficients of a 2-D multiscale decomposition of f based in the original object domain to those of a 1-D decomposition of the data in the projection domain, but rather we are allowing a multiscale projection domain decomposition to *induce* a corresponding, and thus naturally well matched, multiscale object representation. In particular, the j-th scale approximation of the object $f^{(j)}$ is created as a linear combination of the corresponding j coarsest multiscale basis functions (c.f. Figure 3-2) summed over all angles k (note that the coefficients finer than level j in $(A(j)\,\xi_k)$ are zero and use the object definition (3.23)). As can be seen, our resulting object representation lives close to the projection domain in which data is gathered, with advantages in efficiency as we will see.

Beyond simply finding a finest scale object estimate as described in Algorithm 1, we thus also have a method to reconstruct the underlying object at *multiple resolutions* through (3.23), (3.24) and (3.25) and thus for easily obtaining information about the object at multiple scales. In particular, if an approximation $f^{(j)}$ at scale j is desired, then in Algorithm 1 we need only replace ξ_k by $(A(j)\,\xi_k)$ in Step 6 and 7. In particular, this simply amounts to zeroing detail components in ξ_k which are finer than scale j. Further, if instead we want to reconstruct the detail $\Delta f^{(j)}$ added at a particular scale, we need only replace ξ_k by $(D(j)\,\xi_k)$ in Step 6 and 7 of Algorithm 1. Similarly, this

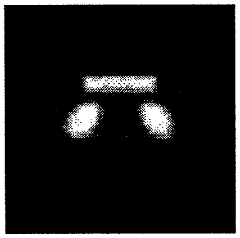


Figure 3-6: The grayscale plot (white corresponds to the maximum value and black to the minimum) of the 32 × 32 phantom used for reconstructions.

simply amounts to zeroing all but the desired scale of detail $\xi_k^{(j)}$ in ξ_k.

3.3.5 Examples

In Figure 3-8 we show reconstructions $f^{(j)}$ at various scales j, of the 32×32 phantom shown in Figures 3-6 and 3-7, from projection data collected at $N_\theta = 32$ angles with $N_s = 32$ strips per angular projection. The D_3 wavelet is used for multiscale reconstruction (i.e. in the definition of W). For this reasonably dense data case we expect the FBP and the finest scale, multiscale reconstructions to be similar since the NP and the FBP solutions converge in case of dense data. This is precisely what is seen in Figure 3-8 and confirmed in Figure 3-9, which shows a section through the reconstructions. Finally, in line with the multiscale nature of our reconstructions, notice that the information about the phantom becomes more focused as we proceed from coarse to fine scales. In Figure 3-10 we show the corresponding detail reconstructions $\Delta f^{(j)}$ at various scales. Note that the finest scale detail reconstruction $\Delta f^{(4)}$ (bottom row, middle in the figure) contains information about the edges and boundaries in the phantom.

In Figure 3-11 we show an example of a sparse data case. Here we reconstruct the phantom at different scales using $N_\theta = 5$ angular projections with $N_s = 32$ strips in each projection, and the D_3 wavelet for multiscale decomposition. We also show the corresponding FBP reconstruction for comparison. In this sparse data case our NP-based multiscale reconstructions are free of many of the finest scale artifacts which arise in the FBP reconstruction. The figure also illustrates the resolution-accuracy tradeoff that is inherent in reconstructions from sparse data, wherein the coarse scale

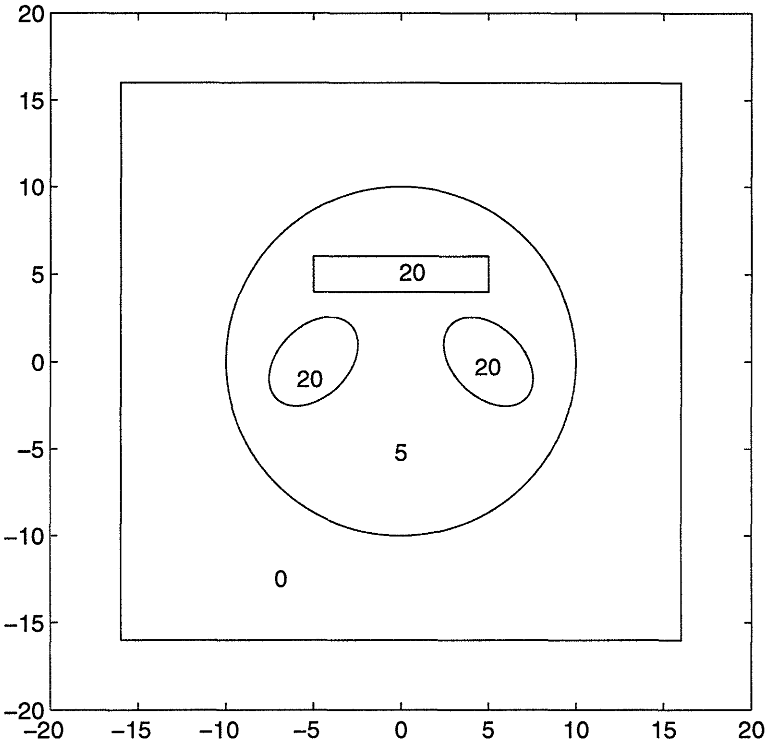


Figure 3-7: The intensity map of the 32×32 phantom shown in Figure 3-6. The numbers shown are intensities of respective regions.

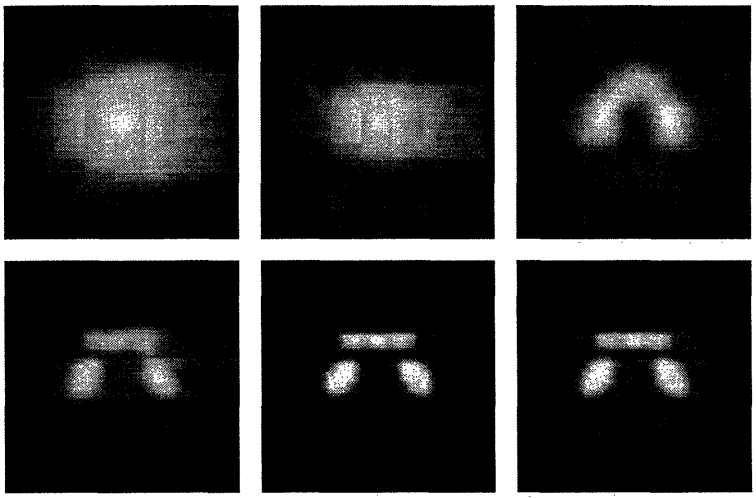


Figure 3-8: Reconstructions at various scales using the D_3 wavelet and $N_\theta = N_s = 32$. Top row, left: $f^{(1)}$. Top row, middle: $f^{(2)}$. Top row, right: $f^{(3)}$. Bottom row, left: $f^{(4)}$. Bottom row, middle: $f^{(5)}$. Bottom row, right: FBP reconstruction.

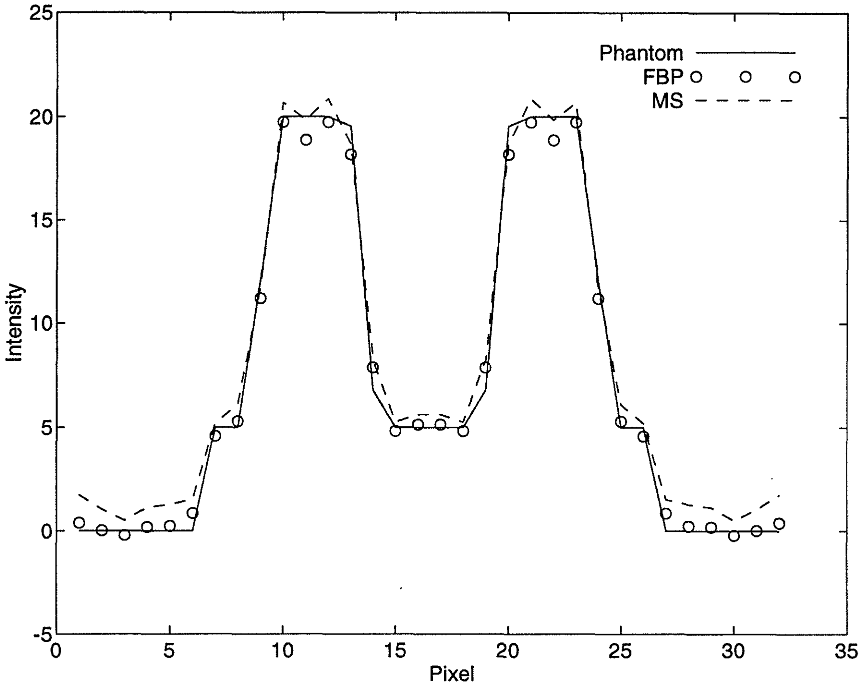


Figure 3-9: A horizontal section through the phantom (solid line), the FBP reconstruction (circles), and the multiscale reconstruction (broken line).

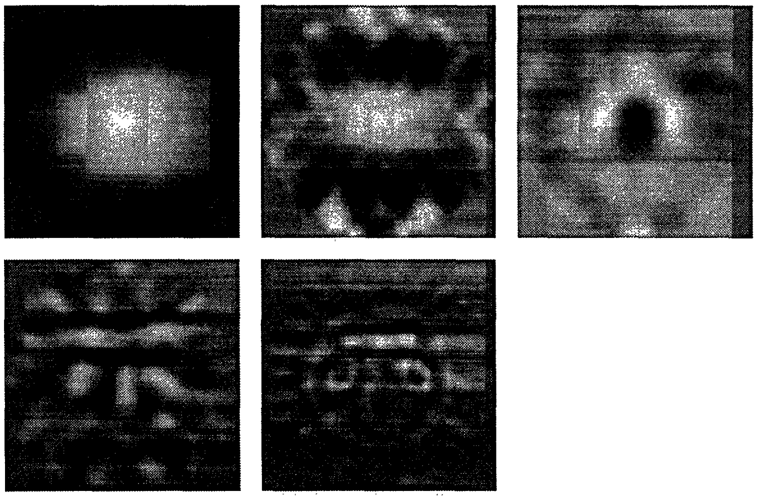


Figure 3-10: Detail reconstructions at various scales, $\Delta f^{(j)}$, using $N_\theta = N_s = 32$ and the D_3 wavelet. Top row, left: $\Delta f^{(0)}$. Top row, middle: $\Delta f^{(1)}$. Top row, right: $\Delta f^{(2)}$. Bottom row, left: $\Delta f^{(3)}$. Bottom row, middle: $\Delta f^{(4)}$.

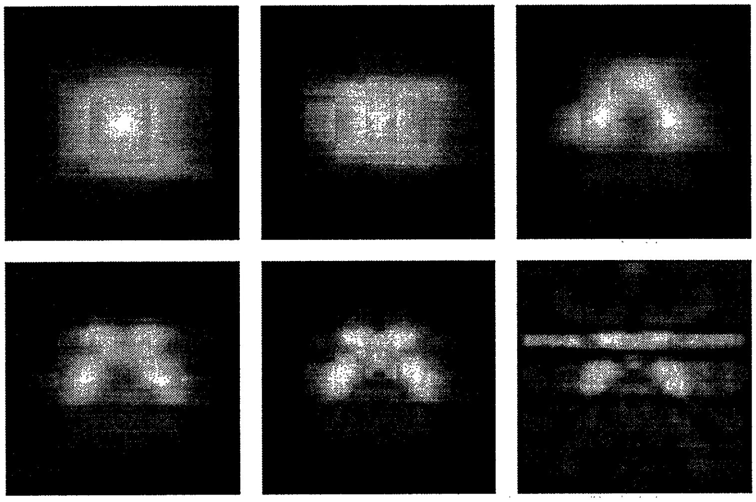


Figure 3-11: Reconstruction at various scales using the D_3 wavelet and $N_\theta = 5$ and $N_s = 32$. Top row, left: $f^{(1)}$. Top row, middle: $f^{(2)}$. Top row, right: $f^{(3)}$. Bottom row, left: $f^{(4)}$. Bottom row, middle: $f^{(5)}$. Bottom row, right: FBP reconstruction.

reconstructions have less artifacts due to sparse data effects at the expense of reduced resolution.

Finally, in Figure 3-12, we show the finest scale detail reconstruction obtained by making the assumption that $\mathcal{C}_{dd}$ is an identity matrix. Such an assumption neglects all cross-scale and cross-angle terms in $\mathcal{C}_{dd}$ resulting in an even simpler inversion procedure. In the same figure we also show for comparison the reconstruction based on finest scale detail from Figure 3-10 which uses no approximation of $\mathcal{C}_{dd}$. We can see from the figure that if the goal is edge reconstruction, it is enough to approximate $\mathcal{C}_{dd}$ by an identity matrix. This reduces the computational complexity even further. Such an edge oriented reconstruction, based on an identity assumption for $\mathcal{C}_{dd}$ in the multiscale framework, only requires 1-D wavelet transformation of the strip integral data and subsequent back-projection of just the fine scale coefficients.

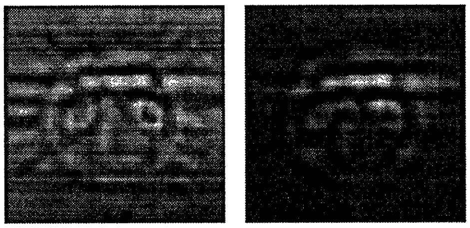


Figure 3-12: Finest scale detail reconstructions using $N_\theta = N_s = 32$, and the D_3 wavelet. Left: Using entire $\mathcal{C}_{dd}$, Right: Assuming $\mathcal{C}_{dd}$ to be the identity matrix.

3.4 Regularized multiscale natural pixel (NP) reconstructions

In this section we consider the estimation of an object f from *noisy* projection observations. We extend our multiscale reconstruction method presented in Section 3.3 to obtain statistically regularized estimates. This regularized solution is obtained by first solving for the maximum aposteriori probability (MAP) estimate [76] of the multiscale object coefficient vector, ξ, based on observations (3.15) and a certain naturally derived multiscale prior model and then back-projecting these multiscale coefficient estimates along the corresponding multiscale basis functions as before.

In the presence of noise our original observations (3.1) become:

$$y(m) = \iint_{\Omega} f(u,v)\, S_m(u,v)\, du\, dv + n(m), \qquad n(m) \sim \mathcal{N}(0,\lambda), \tag{3.26}$$

where the noise $n(m)$ is taken as additive, white Gaussian, and the notation $z \sim \mathcal{N}(m, \Lambda)$ denotes a Gaussian distribution of mean m and covariance Λ. By substituting the object representation equation (3.4) in (3.26), as was done in the noise-free case, we obtain the following equation:

$$y(m) = \sum_{i=1}^{N_\theta N_s} C_{mi} x(i) + n(m). \tag{3.27}$$

Finally, by collecting together the data in (3.27), we obtain the following overall observation equation (the noisy version of (3.5)):

$$y = Cx + n, \qquad\qquad n \sim \mathcal{N}(0, \lambda I_{N_\theta N_s}) \tag{3.28}$$

where $I_{N_\theta N_s}$ refers to a $N_\theta N_s \times N_\theta N_s$ identity matrix.

The multiscale decomposition of (3.28) using (3.14) and (3.16), followed by rearrangement in scales as described in Section 3.3.1, results in the following partitioned equation (the noisy version of (3.18)):

$$\left[\begin{array}{c} \eta^d \\ \hline \eta^a \end{array}\right] = \left[\begin{array}{c|c} C_{dd} & C_{da}^T \\ \hline C_{da} & C_{aa} \end{array}\right] \left[\begin{array}{c} \xi^d \\ \hline \xi^a \end{array}\right] + \left[\begin{array}{c} \nu^d \\ \hline \nu^a \end{array}\right] \tag{3.29}$$

where the vector $[(\nu^d)^T | (\nu^a)^T]^T \sim \mathcal{N}(0, \lambda I_{N_\theta N_s})$ contains the (similarly partitioned) elements of the multiscale noise vector $\nu = W_b n$. Note that the assumption of uncorrelated noise from angle to angle and strip to strip in the original projection domain results in uncorrelated noise from angle to angle and multiscale strip to multiscale strip in the multiscale domain, since W_b is an orthonormal transformation.

Recall from Section 3.3.3 that in the ideal case with no edge effects the off-diagonal blocks C_{da} in (3.29) are identically zero. More generally, as mentioned in Section 3.3.3, due to finite field effects these blocks are not exactly zero, however reconstructions which assume these to be zero are visually indistinguishable from those that do not. As a result, in the development to follow we neglect these field-of-view edge effects and assume $C_{da} = 0$ in our formulae.

3.4.1 Prior model for the multiscale object coefficients

In order to find the MAP estimate of the multiscale object coefficients ξ, we need a prior statistical model for these quantities. We describe next a prior model that captures significant information about the underlying object and at the same time results in estimation problems that are computationally tractable. In particular, we assume ξ^d and ξ^a to be distributed according to:

$$\left[\begin{array}{c} \xi^d \\ \hline \xi^a \end{array}\right] = \mathcal{N}\left(0, \left[\begin{array}{c|c} \Lambda_d & 0 \\ \hline 0 & \Lambda_a \end{array}\right]\right) \tag{3.30}$$

i.e. we assume ξ^d and ξ^a to be Gaussian, zero mean, independent, with variance Λ_d and Λ_a respectively. For the variance of the approximation coefficients ξ^a, capturing the prior DC behavior of the object, we choose

$$\Lambda_a = (\frac{1}{\epsilon}) I_{N_\theta} \tag{3.31}$$

with ϵ sufficiently small (i.e. $\epsilon \longrightarrow 0$) to prevent a bias in our estimate of the average (DC) behavior of the coefficients ξ, letting them be determined instead by the data.

For the variance of the detail object coefficients ξ^d we use a self-similar prior model which is obtained by choosing the elements of the detail vector $\xi_k^{(j)}$ (i.e. the wavelet coefficients) at angle k and scale j as independent, $\mathcal{N}(0, \sigma^2 2^{-\rho j})$ random variables [80]. The parameter ρ determines the nature, i.e. the texture, of the resulting self-similar process while σ^2 controls the overall magnitude. This model says that the variance of the detail added in going from the approximation at scale j to the approximation at scale $j+1$ decreases geometrically with scale. If $\rho = 0$ the resulting finest level representation (the elements of x_k) corresponds to samples of white noise (i.e. the components of x_k are completely uncorrelated), while as ρ increases the components of x_k show greater long range correlation. Such self-similar models are commonly and effectively used in many application areas such as modeling of natural terrain and other textures, biological signals, geophysical and economic time series, etc. [11, 12, 58, 78, 80]. In addition, since the *observation noise* power is uniform across scales or frequencies, the geometrically decreasing variance of the prior model implies that the projection data most strongly influences the reconstruction of coarse scale features and the prior model most strongly influences the reconstruction of fine scale features. This reflects our belief that the fine scale behavior of the object (corresponding to high frequencies) is the most likely to be corrupted by noise. The above self-similar prior model results in a diagonal covariance matrix Λ_d for the detail coefficients ξ^d, the elements of which depend on the regularization parameters ρ and σ^2, i.e. the texture and the overall magnitude, respectively. In particular, if $J = \log_2(N_s)$ is the scale at the finest level then:

$$\Lambda_d = \sigma^2 \text{block diag} \left(\, 2^{-\rho(J-1)} I_{2^{J-1}N_\theta} \;\mid\; 2^{-\rho(J-2)} I_{2^{J-2}N_\theta} \;\mid\; \ldots \;\mid\; I_{N_\theta} \right) \tag{3.32}$$

3.4.2 NP Based Multiscale MAP Estimate

The MAP estimates of ξ^d and ξ^a based on the observations (3.29), the prior model (3.30–3.32) (with $\epsilon \to 0$), and assuming $\mathcal{C}_{da} = 0$ are given by:

$$\widehat{\xi}^d = (\lambda \Lambda_d^{-1} + \mathcal{C}_{dd}^2)^{-1} \mathcal{C}_{dd} \eta^d, \tag{3.33}$$

$$\widehat{\xi}^a = \mathcal{C}_{aa}^{+} \eta^a, \tag{3.34}$$

where in (3.34) we have used the fact that as $\epsilon \to 0$, $(\lambda \epsilon I_{N_\theta} + \mathcal{C}_{aa}^2)^{-1} \mathcal{C}_{aa} \to \mathcal{C}_{aa}^{+}$ (we justify this in Appendix 3.F). Note that the estimate of the approximation coefficients ξ^a from (3.34) is the same as we had earlier in the unregularized case (3.20). Thus the only change in the reconstruction algorithm from the noise-less case is that a different expression for $\widehat{\xi}^d$ as given in (3.33) is now to be used for calculation of an estimate of the detail vector ξ^d. Further, since $(\lambda \Lambda_d^{-1} + \mathcal{C}_{dd}^2)$ is still a sparse matrix due to the compression achieved in the multiscale domain, fast and efficient recursive algorithms [63] can be used to solve for $\widehat{\xi}^d$ in (3.33). In fact, in Chapter 4 we will use the variant of ART due to Herman [39] (c.f. Section 2.4.2) that computes the MAP estimate corresponding to white observation noise and an uncorrelated prior model, to solve for $\widehat{\xi}^d$ from (3.33). This solution will be obtained with the same computational and storage requirements as the solution for the unregularized coefficients ξ^d from (3.19). Finally note that in obtaining the MAP estimates (3.33) and (3.34) we have assumed $\mathcal{C}_{da}$ to be 0. The inclusion of the effects of this neglected coupling into the MAP estimates is straightforward, as discussed in Section 3.3.3.

We present next the algorithm for obtaining the MAP regularized multiscale reconstructions.

Algorithm 2 (Regularized Multiscale Reconstruction)

1. *For a given choice of wavelet, form the sparse multiscale matrix* $\mathcal{C} = W_b\, C\, W_b^T$ *(the multiscale counterpart of the original NP matrix* C*). Re-arrange the elements in* $\mathcal{C}$ *according to scales (c.f. Appendix 3.A) to form* $\mathcal{C}_s$*. The block* $\mathcal{C}_{dd}$ *in* $\mathcal{C}_s$ *is associated with the detail-detail coupling (c.f. (3.18)).*

2. *Choose the model parameter* λ *specifying the variance of the observation noise process.*

3. *Choose the multiscale prior model parameters* σ^2 *and* ρ *specifying the magnitude and the texture of the model, respectively, and generate the prior covariance matrix* Λ_d *through (3.32).*

4. *Find the multiscale observations* η *by taking the 1-D wavelet transform of the projection data,* $\eta = W_b y$*. Re-arrange the elements in* η *according to scales to obtain the* η^d *and* η^a*, the multiscale detail and the coarsest scale approximation data sub-vectors respectively (c.f. Appendix 3.A).*

5. *Calculate the MAP estimate of the coarsest scale approximation object coefficients,* $\widehat{\xi}^a = (1/(N_\theta N_s))\eta^a$*.*

6. *Calculate the MAP estimate of the multiscale detail object coefficient vector* $\widehat{\xi}^d$ *by solving the sparse and well conditioned system of equations* $(\lambda\Lambda_d^{-1} + \mathcal{C}_{dd}^2)\eta^d = \mathcal{C}_{dd}\xi^d$*.*

7. *Produce the MAP estimate* $\widehat{\xi}$ *of the overall observation vector* ξ *by re arrangement of* $\widehat{\xi}^a$ *and* $\widehat{\xi}^d$ *according to projection angles (c.f. Appendix 3.A). The block* $\widehat{\xi}_k$ *in* $\widehat{\xi}$ *is associated with angle* k *(c.f. (3.6)).*

8. *For each angle* k*, back-project* $\widehat{\xi}_k$ *along the corresponding multiscale basis functions* $\mathcal{T}_k$*,* $\mathcal{T}_k^T \xi_k$*.*

9. *Combine the object contributions from the individual back-projections at each angle to obtain the overall MAP reconstruction,* $\widehat{f} = \sum_k \mathcal{T}_k^T \widehat{\xi}_k$.

As before, we may also easily obtain regularized reconstructions of the object at multiple resolutions by using (3.23) and (3.24) together with the MAP coefficient estimates $\widehat{\xi}_k$. In particular, to obtain the approximation $\widehat{f}^{(j)}$ at scale j then we need only replace $\widehat{\xi}_k$ by $(A(j)\,\widehat{\xi}_k)$ (corresponding to simply zeroing some of the terms in $\widehat{\xi}_k$) in Step 8 and 9. Similarly, the corresponding object detail components $\Delta\widehat{f}^{(j)}$ at scale j may be obtained by using $(D(j)\,\widehat{\xi}_k)$ in place of $\widehat{\xi}_k$ in these steps.

3.4.3 Examples

We next show the reconstructions obtained using our regularized multiscale method in the presence of noise. The noise-free projection data are generated from the phantom of Figure 3-6 and are then corrupted through the addition of independent, zero-mean Gaussian noise to yield our observations y. We use a signal-to-noise ratio (SNR) of 5 dB for our example reconstructions, defined as:

$$\text{SNR (dB)} = 10 \log \frac{\|Tf\|^2}{\lambda_n N_\theta N_s}, \tag{3.35}$$

where λ_n is the variance of the additive noise, and Tf are the noise-free projection data. Finally, in all multiscale reconstructions we show here the Daubechies D_3 wavelet is used in the definition of the multiscale decomposition matrix W.

The first example, shown in Figure 3-13, demonstrates reconstructions from noisy data using the *unregularized* multiscale approach of the previous section. The projection data are collected at $N_\theta = 32$ angles with $N_s = 32$ strips per angular projection. The variance of the additive noise used is $\lambda_n = 3.61 \times 10^3$ which results in a SNR value of 5 dB. This figure shows the various scale approximate reconstructions $f^{(j)}$ corresponding to the unregularized Algorithm 1 for the complete range of scales $j = 1, \ldots, 5$. As before, the approximations become finer from left to right and top to bottom (so that the upper left frame is $f^{(1)}$ and the bottom middle frame corresponds to $f^{(5)}$). The bottom right frame shows the standard FBP reconstruction

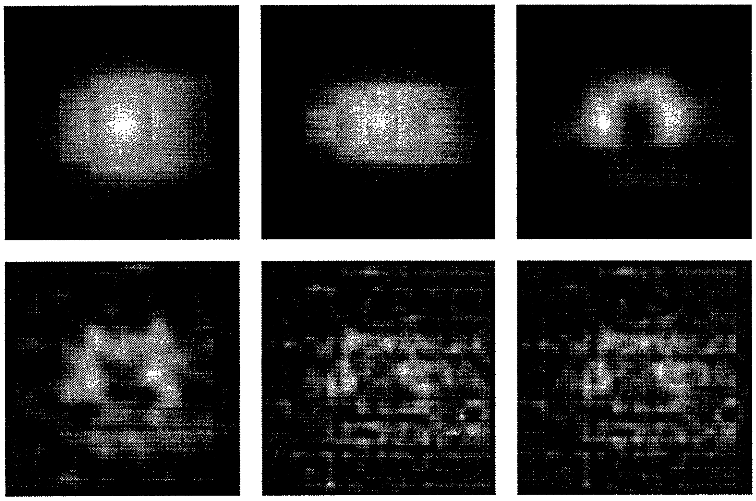


Figure 3-13: Reconstructions at various scales using the D_3 wavelet and $N_\theta = N_s = 32$ from 5 dB SNR projection data. Top row, left: $f^{(1)}$. Top row, middle: $f^{(2)}$. Top row, right: $f^{(3)}$. Bottom row, left: $f^{(4)}$. Bottom row, middle: $f^{(5)}$. Bottom row, right: FBP reconstruction.

based on the noisy data. The figure illustrates the resolution-accuracy tradeoff inherently captured in the multiscale framework and confirms the point that even in the unregularized case, information from noisy observations can be focused by stopping the reconstruction at a coarse scale, for example scale 4 (bottom row, left in the figure). The finer scale detail contribution $\Delta f^{(4)}$ is evidently mainly noise which obscure the object features.

We next show estimates generated by our multiscale MAP regularized method. Figure 3-14 shows the FBP, unregularized finest level multiscale, and MAP regularized finest level multiscale reconstructions from projection data collected at $N_\theta = 32$ angles with $N_s = 32$ strips per angular projection. The variance of the additive noise used is $\lambda_n = 3.61 \times 10^3$ which results in a SNR value of 5 dB. In both the standard FBP and the unregularized finest level multiscale reconstruction the noise almost completely obscures the object. In contrast, in the regularized solution (which is obtained with essentially the same computational complexity as the unregularized solution) the details of the object are now visible. In both of the regularized reconstructions the overall magnitude of the prior model $\sigma^2 = \lambda_n = 3.61 \times 10^3$ and the variance of the model observation noise $\lambda = \lambda_n = 3.61 \times 10^3$, however the reconstructions differ in the

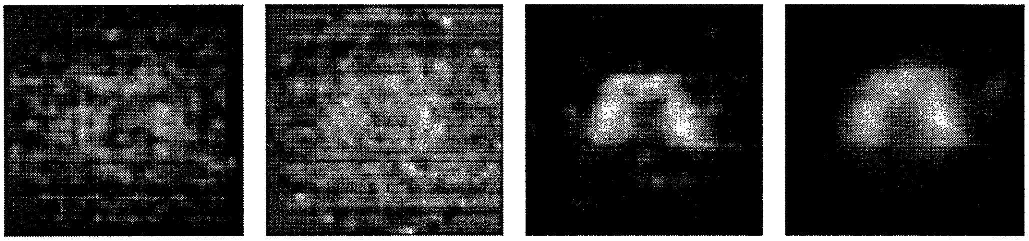


Figure 3-14: Reconstructions with $N_\theta = N_s = 32$ and 5 dB SNR projection data. From left (a) FBP reconstruction, (b) Unregularized finest level multiscale reconstruction, (c) Regularized finest level multiscale reconstruction with $\lambda = \sigma^2 = 3.61 \times 10^3$, and $\rho = 0.5$, and (d) Regularized finest level multiscale reconstruction with $\lambda = \sigma^2 = 3.61 \times 10^3$, and $\rho = 1$.

decay rate of detail variance across scales, ρ. From Figure 3-14 and Figure 3-15, which shows a section through the reconstructions, we see that as is expected, an increased regularization (i.e. smoothness) results when the value of ρ is increased from 0.5 to 1 (corresponding to a smoother prior texture) keeping other parameters fixed.

Figure 3-16 shows the FBP, and the finest scale unregularized and the finest scale MAP regularized multiscale reconstructions from sparse ($N_\theta = 5$, $N_s = 32$) and noisy (SNR 5 dB, $\lambda_n = 3.65 \times 10^3$) projection data. Again, note the ability of the multiscale regularization algorithm to pull out features of the phantom even when the data is extremely limited in both quality and quantity.

3.5 Discussion

In this chapter we have developed a multiscale tomographic reconstruction technique based on the natural pixel (NP) approach, that provides reconstructions from sparse data yet is extremely efficient from a computational standpoint. Our multiscale technique is different from other multiscale methods in the following respects. First, our 2-D multiscale object representation is naturally induced by expanding the NP basis functions (i.e. strips) in a *1-D wavelet basis.* This is in contrast to other multiscale reconstruction techniques which *begin* with a 2-D object representation obtained from a full 2-D wavelet decomposition of the object space. These techniques must subsequently relate the inherently 1-D projection data to these fundamentally 2-D object

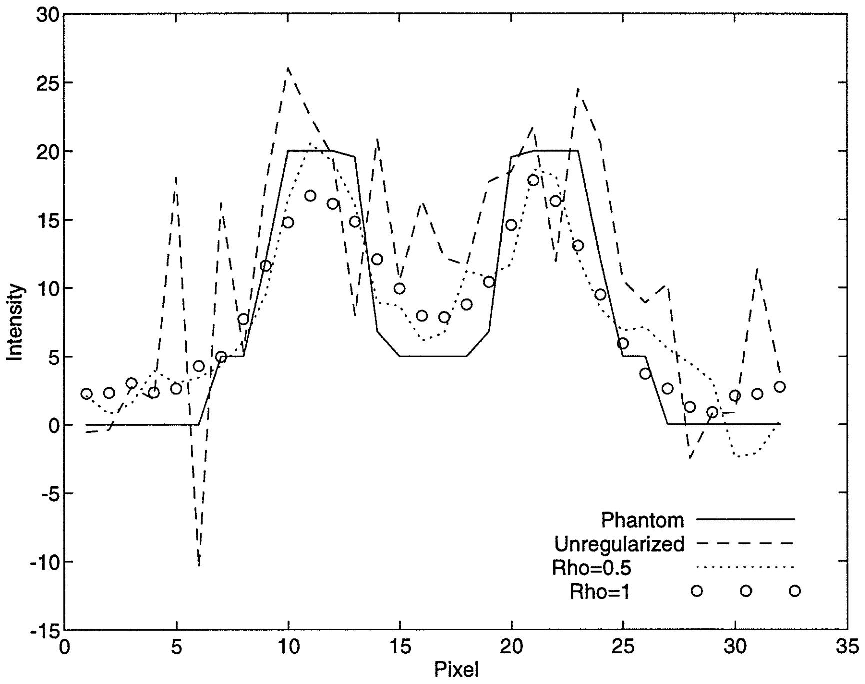


Figure 3-15: A horizontal section through the phantom (solid line) and the various reconstructions of the previous figure.

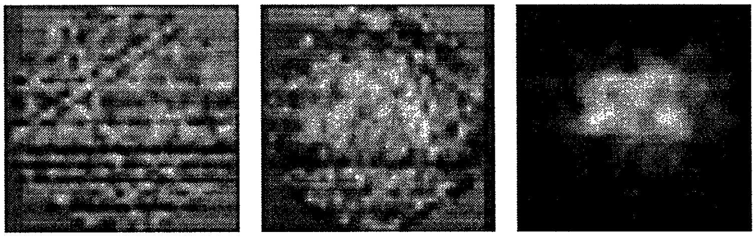


Figure 3-16: Reconstructions with $N_\theta = 5$, $N_s = 32$ and 5 dB SNR projection data. From left (a) FBP reconstruction, (b) Unregularized finest level multiscale reconstruction, and (c) Regularized finest level multiscale reconstruction with $\lambda = \lambda_n = 3.65 \times 10^3$, $\rho = 0.35$ and $\sigma^2 = 1.83 \times 10^3$.

coefficients. In contrast, the multiscale representation resulting from our approach, arising as it does from the projection strips themselves, is much closer to the measurement domain. The result is a highly efficient method to compute our multiscale object coefficients.

In addition, based on this wavelet-based multiscale framework, we have proposed a statistically-based multiresolution MAP estimation algorithm. This method provides statistically regularized reconstructions from noisy data, and does so at multiple resolutions, *at no more effort than is required for the standard unregularized method.* This approach, based on the construction of prior models directly in scale-space, allows for the inclusion of natural, self-similar prior models into the reconstruction process. These prior models not only impose a classic smoothness constraint on the resulting reconstruction, but also accurately represent several textures including natural terrain and biological structures such as liver and lung. We have showed that such self-similar prior models are captured very easily in our multiscale framework by assuming that the multiscale object coefficients are uncorrelated from angle to angle and from multiscale strip to multiscale strip.

Finally, while the prior model for the multiscale object coefficients that we use in this chapter assumes these coefficients to be independent from angle to angle, we would intuitively expect the coarse scale object coefficients at different projection angles to actually be highly correlated with each other, and further for this correlation to decrease at finer scales. Such generalizations of the self-similar prior model used in this chapter are briefly discussed in Chapter 8.

3.A The angle and scale ordering schemes

The wavelet transformed data and object coefficient vectors η and ξ, respectively, are ordered according to projection angles as shown below:

$$\eta = \left[\begin{array}{c} \eta_1 \\ \hline \eta_2 \\ \hline \vdots \\ \hline \eta_{N_\theta} \end{array} \right], \tag{3.36}$$

$$\xi = \left[\begin{array}{c} \xi_1 \\ \hline \xi_2 \\ \hline \vdots \\ \hline \xi_{N_\theta} \end{array} \right], \tag{3.37}$$

where η_k and ξ_k are the wavelet transformed data and object coefficient vectors, respectively, at angle k. These vectors are of the form (c.f. (2.50)):

$$\eta_k = \left[\begin{array}{c} \eta_k^{(J-1)} \\ \hline \vdots \\ \hline \eta_k^{(0)} \\ \hline y_k^{(0)} \end{array} \right], \tag{3.38}$$

$$\xi_k = \left[\begin{array}{c} \xi_k^{(J-1)} \\ \hline \vdots \\ \hline \xi_k^{(0)} \\ \hline x_k^{(0)} \end{array} \right], \tag{3.39}$$

where $\eta_k^{(j)}$ and $\xi_k^{(j)}$ are vectors of length 2^j and contain the scale j detail coefficients in η_k and ξ_k respectively, and $y_k^{(0)}$ and $x_k^{(0)}$ are the corresponding coarsest scale approximation coefficients.

Let P be the orthogonal permutation matrix (i.e. $P^T = P^{-1}$) which when applied to the angle-ordered vectors η and ξ, results in their scale-ordered counterparts η_s

and ξ_s, respectively:

$$\begin{aligned} \eta_s &= P\eta, && (3.40) \\ \xi_s &= P\xi. && (3.41) \end{aligned}$$

The scale-arranged multiscale data and object coefficient vectors η_s and ξ_s, respectively, are of the form:

$$\eta_s = \left[\begin{array}{c} \eta^{(J-1)} \\ \hline \vdots \\ \hline \eta^{(0)} \\ \hline \eta^{a} \end{array} \right], \tag{3.42}$$

$$\xi_s = \left[\begin{array}{c} \xi^{(J-1)} \\ \hline \vdots \\ \hline \xi^{(0)} \\ \hline \xi^{a} \end{array} \right], \tag{3.43}$$

where the vectors $\eta^{(j)}$ and $\xi^{(j)}$ contain the scale j data and object coefficients, respectively, from all N_θ angles:

$$\eta^{(j)} = \left[\begin{array}{c} \eta_1^{(j)} \\ \hline \eta_2^{(j)} \\ \hline \vdots \\ \hline \eta_{N_\theta}^{(j)} \end{array} \right], \tag{3.44}$$

$$\xi^{(j)} = \left[\begin{array}{c} \xi_1^{(j)} \\ \hline \xi_2^{(j)} \\ \hline \vdots \\ \hline \xi_{N_\theta}^{(j)} \end{array} \right], \tag{3.45}$$

and η^a and ξ^a contain the corresponding coarsest scale approximation coefficients from all angles:

$$\eta^a = \left[\begin{array}{c} y_1^{(0)} \\ \hline y_2^{(0)} \\ \hline \vdots \\ \hline y_{N_\theta}^{(0)} \end{array}\right], \tag{3.46}$$

$$\xi^a = \left[\begin{array}{c} x_1^{(0)} \\ \hline x_2^{(0)} \\ \hline \vdots \\ \hline x_{N_\theta}^{(0)} \end{array}\right]. \tag{3.47}$$

The multiscale scale-ordered vectors η_s and ξ_s can also be written in terms of detail and approximation components as follows:

$$\eta_s = \left[\begin{array}{c} \eta^d \\ \hline \eta^a \end{array}\right], \tag{3.48}$$

$$\xi_s = \left[\begin{array}{c} \xi^d \\ \hline \xi^a \end{array}\right]. \tag{3.49}$$

Finally, the scale-ordered version of the multiscale observation equation (3.15) is obtained as follows:

$$\begin{aligned} P\eta &= (P\mathcal{C}P^T)P\xi, \quad \text{i.e.} \\ \eta_s &= \mathcal{C}_s\xi_s. \end{aligned} \tag{3.50}$$

3.B Justification of (3.21)

If there are no edge effects, all elements of the $N_\theta \times N_\theta$ matrix $\mathcal{C}_{aa}$ are equal to N_s. Thus the row sum of $\mathcal{C}_{aa}$, which is same for all the rows, is given by:

$$r = N_\theta N_s. \tag{3.51}$$

If the object is completely contained in the field-of-view, then all N_θ coarsest approximation data coefficients are equal, and are given by:

$$\eta^a = (\frac{\mu(f)}{\sqrt{N_s}})1_{N_\theta}, \tag{3.52}$$

where 1_p refers to a vector of length p with all elements equal to unity, and $\mu(f)$ is the total mass under the object (c.f. (3.22)). Finally, as proved in Section 3.B.1, if $\mathcal{C}_{aa}$ is circulant, as is the case here, then:

$$\mathcal{C}_{aa}^{+}1_{N_\theta} = (\frac{1}{r})1_{N_\theta}, \tag{3.53}$$

where r is the row sum of $\mathcal{C}_{aa}$. Now (3.21) is obtained by combining (3.51)–(3.53).

3.B.1 Proof of (3.53)

Theorem 1 *If $\mathcal{C}_{aa}$ is a $N_\theta \times N_\theta$ symmetric, circulant matrix, with row sum equal to r, i.e.*

$$\mathcal{C}_{aa}1_{N_\theta} = r1_{N_\theta}, \tag{3.54}$$

where 1_p refers to a vector of length p with all elements equal to unity, then the row sums of the pseudo-inverse of $\mathcal{C}_{aa}$ are all equal to $1/r$, i.e.

$$\mathcal{C}_{aa}^{+}1_{N_\theta} = \frac{1}{r}1_{N_\theta}, \tag{3.55}$$

where $\mathcal{C}_{aa}^{+}$ refers to the pseudo-inverse of $\mathcal{C}_{aa}$.

Proof Since the matrix $\mathcal{C}_{aa}$ is circulant, it can always be *diagonalized* as follows [19]:

$$\mathcal{C}_{aa} = F^{*}DF, \tag{3.56}$$

where * denotes complex conjugation, and F is the $(N_\theta \times N_\theta)$ matrix representation of the discrete Fourier transform operation:

$$F = \frac{1}{\sqrt{N_\theta}} \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & \exp\left[-j(2\pi/N_\theta)\right] & \dots & \exp\left[-j(2\pi/N_\theta)(N_\theta - 1)\right] \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \exp\left[-j(2\pi/N_\theta)(N_\theta - 1)\right] & \dots & \exp\left[-j(2\pi/N_\theta)(N_\theta - 1)^2\right] \end{bmatrix}. \tag{3.57}$$

The matrix F has the following properties:

$$F = F^T, \tag{3.58}$$

and

$$F^* F = F F^* = I. \tag{3.59}$$

The matrix D in (3.56) is diagonal. We refer to the diagonal elements of D as $\{d_i, i = 1, \dots, N_\theta\}$, i.e.

$$D = \operatorname{diag}(d_1, d_2, \dots, d_{N_\theta}). \tag{3.60}$$

The constant row sums of C_{aa} can now be obtained as:

$$r 1_{N_\theta} = C_{aa} 1_{N_\theta} = (F^* D F) 1_{N_\theta} = F^* D \begin{bmatrix} \sqrt{N_\theta} \\ 0 \\ \vdots \\ 0 \end{bmatrix} = d_1 1_{N_\theta}, \tag{3.61}$$

implying that:

$$d_1 = r. \tag{3.62}$$

Since the elements of C_{aa} are real and positive, so is the row sum r, implying that d_1 is also *real and positive*.

The Takagi factorization of $\mathcal{C}_{aa}$ [41], which is a special form of the singular value decomposition for symmetric matrices, is given by:

$$\mathcal{C}_{aa} = UMU^T, \tag{3.63}$$

where the columns of U are the eigenvectors of $\mathcal{C}_{aa}\mathcal{C}^*_{aa}$, and M is a diagonal matrix, the diagonal elements of which are the non-negative square roots of the eigenvalues of $\mathcal{C}_{aa}\mathcal{C}^*_{aa}$. It can be shown from (3.56) that:

$$(\mathcal{C}_{aa}\mathcal{C}^*_{aa})F = F(DD^*) \tag{3.64}$$

which implies that:

$$U = F, \tag{3.65}$$

and

$$M = (DD^*)^{1/2}. \tag{3.66}$$

From (3.58), (3.63), (3.65), and (3.66), the Takagi factorization of $\mathcal{C}_{aa}$ reduces to:

$$\mathcal{C}_{aa} = F(DD^*)^{1/2}F. \tag{3.67}$$

Finally from (3.60), (3.62), (3.66), (3.67), and the definition of pseudo-inverse:

$$\mathcal{C}^+_{aa}1_{N_\theta} = (F^*M^+F^*)1_{N_\theta} = F^*M^+\begin{bmatrix}\sqrt{N_\theta}\\0\\\vdots\\0\end{bmatrix} = (\frac{\sqrt{N_\theta}}{|d_1|})F^*\begin{bmatrix}1\\0\\\vdots\\0\end{bmatrix} = (\frac{1}{r})1_{N_\theta}. \tag{3.68}$$

This completes the proof.

3.C Demonstration that the elements in $\mathcal{C}_{da}$ are mostly negligible for the Haar case

Suppose the NP strips are given by:

$$\{S_m, \quad m = 1, \ldots, N_\theta N_s\}, \tag{3.69}$$

and suppose the corresponding multiscale detail strips are given by:

$$\{\mathcal{Z}_{k,n}^{(j)}, \quad k = 1, \ldots, N_\theta; \; j = 1, \ldots, J-1; \; n = 1, \ldots, 2^j\}, \tag{3.70}$$

and the coarsest scale approximation strips are given by:

$$\{Z_k^{(0)}, \quad k = 1, \ldots, N_\theta\}, \tag{3.71}$$

where k is the angular position, j is the scale, n is the shift within the scale, and $J = \log_2(N_s)$ is the finest scale. Given the definitions above, the elements of the matrix $\mathcal{C}_{da}$ are the areas of intersection of the multiscale detail strips $\{\mathcal{Z}_{k,n}^{(j)}\}$ and the coarsest scale approximation strips $\{Z_k^{(0)}\}$. We use the notation:

$$< Z_k^{(0)}, \mathcal{Z}_{l,n}^{(j)} > \tag{3.72}$$

to refer to the area of intersection of the coarsest scale approximation strip at angle k and the n-th detail strip at scale j and angle l.

It can be shown by elementary geometry that for the Haar case the maximum absolute value in $\mathcal{C}_{da}$ corresponds to $< Z_{k_1}^{(0)}, \mathcal{Z}_{k_2,1}^{(1)} >$ or $< Z_{k_1}^{(0)}, \mathcal{Z}_{k_2,2}^{(1)} >$, where the angular positions k_1 and k_2 are separated by 45^0. This value is equal to $(1-5\sqrt{2}/8)N_s = 0.12N_s$ (recall that the maximum value in $\mathcal{C} = N_s$). The majority of the terms in $\mathcal{C}_{da}$ correspond, however, to the areas of intersection with fine scale strips, and are even much smaller. As an example, for projections k_1 and k_2, and for a scale j such that:

$$j \geq 1 - \log_2(2\beta - \alpha), \tag{3.73}$$

where

$$\begin{aligned} \alpha &= \sin\theta + \cos\theta - 1, & (3.74) \\ \beta &= \sin\theta\cos\theta, & (3.75) \end{aligned}$$

and θ is the angle between k_1 and k_2 ($0^0 < \theta < 90^0$)[6], the following bound applies:

$$| < Z_{k_1}^{(0)}, \mathcal{Z}_{k_2,*}^{(j)} > | \leq (2^{j/2})(\frac{p}{4})(\frac{\alpha}{\beta}) \left[\sqrt{1 + \frac{8p}{\alpha} + \frac{16p^2}{\alpha^2}} - 1\right] N_s, \tag{3.76}$$

where

$$p = (0.5)^{j+1}. \tag{3.77}$$

In Figure 3-17 we plot the variation of the bound (3.76) on the elements of C_{da} as a function of angle θ and scale j. We point out that the bound expression in (3.76) is only valid for a certain angular range for each scale j (c.f. (3.73)). In particular for the case $j = 1$ the bound is not applicable for any value of θ. As a consequence, we have set the bound values outside the permissible angular range to zero in the figure. One can see from the figure that the bound achieves a maximum value at the coarse scale $j = 2$, which is approximately 7% of the absolute maximum in C_{da}. For fine scales, however, the bound value is much lower, implying that most of the fine scale elements in C_{da} are negligible.

---

[6] It is enough to consider $0^0 < \theta < 90^0$, as for $90^0 < \theta < 180^0$ the same bounds apply. For $\theta = 0^0$ or 90^0, the areas of intersection $< Z_{k_1}^{(0)}, \{\mathcal{Z}_{k_2,*}^{(j)}\} >$ are identically zero.

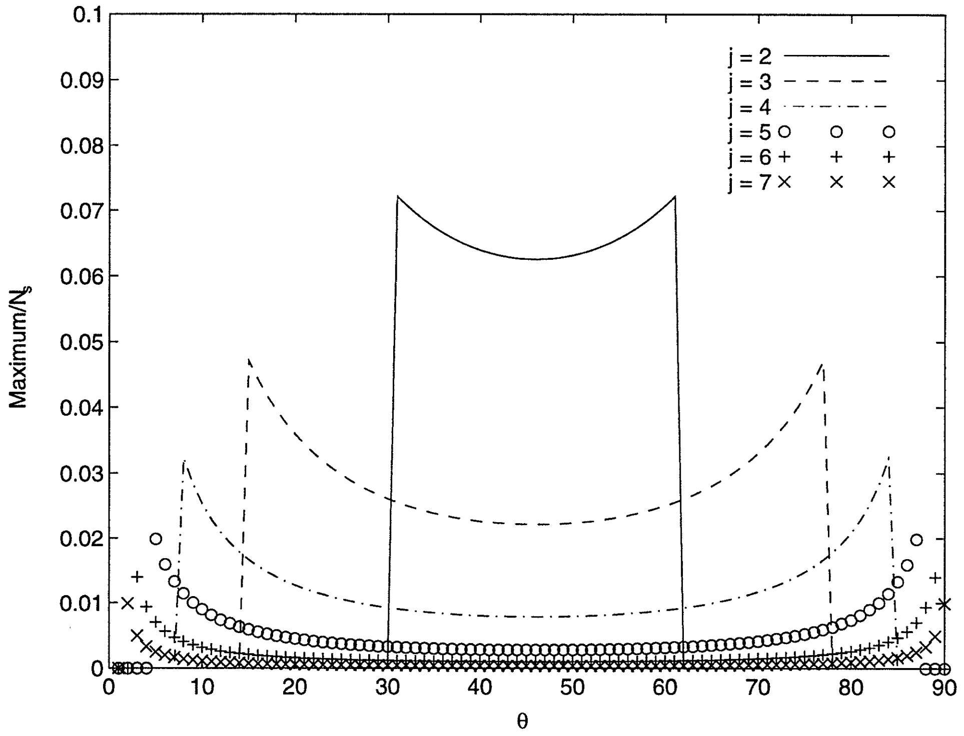


Figure 3-17: The plot of the variation of the bound (3.76) on the elements of C_{da} as a function of angle θ and scale j. The bound expression in (3.76) is only valid for a certain angular range for each scale j (c.f. (3.73)). In particular for the case $j = 1$ the bound is not applicable for any value of θ. As a consequence, we have set the bound values outside the permissible angular range to zero in the figure.

3.D The matrix $\mathcal{C}_{aa}$ for the finite-extent rectangular field-of-view

The matrix $\mathcal{C}_{aa}$ has the following form:

$$\mathcal{C}_{aa} \triangleq \begin{bmatrix} < Z_1^{(0)}, Z_1^{(0)} > & < Z_1^{(0)}, Z_2^{(0)} > & < Z_1^{(0)}, Z_3^{(0)} > & \ldots & < Z_1^{(0)}, Z_{N_\theta}^{(0)} > \\ < Z_2^{(0)}, Z_1^{(0)} > & < Z_2^{(0)}, Z_2^{(0)} > & < Z_2^{(0)}, Z_3^{(0)} > & \ldots & < Z_1^{(0)}, Z_{N_\theta}^{(0)} > \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ < Z_{N_\theta}^{(0)}, Z_1^{(0)} > & < Z_{N_\theta}^{(0)}, Z_2^{(0)} > & < Z_{N_\theta}^{(0)}, Z_3^{(0)} > & \ldots & < Z_{N_\theta}^{(0)}, Z_{N_\theta}^{(0)} > \end{bmatrix}, \tag{3.78}$$

where the notation $< a, b >$ refers to the inner product of, or equivalently, the area of intersection of a and b, and $Z_k^{(0)}$ refers to the coarsest scale approximation strip at angle k. The diagonal elements of $\mathcal{C}_{aa}$, $\{< Z_k^{(0)}, Z_k^{(0)} >, \;\; k = 1, 2, \ldots, N_\theta\}$, represent the area of the field-of-view at different angles, and are all equal. Further $\mathcal{C}_{aa}$ is symmetric (i.e. $< Z_{k_1}^{(0)}, Z_{k_2}^{(0)} > = < Z_{k_2}^{(0)}, Z_{k_1}^{(0)} >$) due to the commutivity property of the inner product operation. Also, by elementary geometry:

$$< Z_{k_1}^{(0)}, Z_{k_2}^{(0)} > = < Z_1^{(0)}, Z_{k_2-k_1+1}^{(0)} >, \tag{3.79}$$

and, if θ is the angle between projection 1 and k,

$$< Z_1^{(0)}, Z_k^{(0)} > = \begin{cases} \left(\dfrac{\sin\theta + \cos\theta - 1}{\sin\theta\cos\theta}\right) N_s & \text{if } k \text{ such that } 0^0 < \theta < 90^0, \\ N_s & \text{if k such that } \theta = 0,\ 90^0, \\ < Z_1^{(0)}, Z_{N_\theta+2-k}^{(0)} > & \text{if } k \text{ such that } 90^0 < \theta < 180^0. \end{cases} \tag{3.80}$$

Finally, it can be shown from (3.79) and (3.80) that $\mathcal{C}_{aa}$ is circulant.

From (3.80), $< Z_1^{(0)}, Z_k^{(0)} >$ achieves a minimum value equal to $2(\sqrt{2} - 1)N_s = 0.83N_s$ when k corresponds to a projection at 45^0 or 135^0. The maximum value of $< Z_1^{(0)}, Z_k^{(0)} >$ is N_s when k corresponds to projections at 0^0 or 90^0.

The constant row sum r of $\mathcal{C}_{aa}$ is given by:

$$r = \sum_{k=1}^{N_\theta} < Z_1^{(0)}, Z_k^{(0)} > . \tag{3.81}$$

From (3.81) and (3.80), the row sum r is equal to $14.2N_s$ for $N_\theta = 16$, and is equal to $28.3N_s$ for $N_\theta = 32$. For most applications, it is enough to approximate r as:

$$r \approx N_\theta N_s. \tag{3.82}$$

Note that the above row sum is the same as the row sum of $\mathcal{C}_{aa}$ for the ideal geometry where there are no edge effects.

3.E Introduction of $\mathcal{C}_{da}$ coupling in the estimate of the multiscale object coefficients

We first calculate the approximation coefficients η^a by neglecting the $\mathcal{C}_{da}$ coupling (i.e. by assuming $\mathcal{C}_{da} = 0$). Recall that for the ideal, infinite-extent field-of-view case, all elements in $\mathcal{C}_{aa}$ were equal. This circulant structure of $\mathcal{C}_{aa}$ was responsible for a simplified expression for ξ^a in the ideal case. Now, as shown in Section 3.D, the matrix $\mathcal{C}_{aa}$ remains circulant in the finite-extent rectangular field-of-view case. Hence the same techniques as in the ideal, infinite-extent case can be used to calculate ξ^a. In particular,

$$\xi^a = (\frac{1}{r})\eta^a = (\frac{\mu(f)}{r\sqrt{N_s}})1_{N_\theta} \tag{3.83}$$

where 1_p refers to a vector of length p with all elements equal to unity, r is the constant row sum of the circulant matrix $\mathcal{C}_{aa}$, and $\mu(f)$ is the total mass under the object. Now, as shown in Section 3.D, the row sum $r \approx N_\theta N_s$, which results in the following expression for ξ^a (the same as the one we had for the ideal case):

$$\xi^a = (\frac{\mu(f)}{N_\theta N_s \sqrt{N_s}})1_{N_\theta}. \tag{3.84}$$

We next calculate the detail coefficients ξ^d by using the expression (3.84) for the approximation coefficients ξ^a, and by incorporating the C_{da} coupling. To begin, once the approximation coefficients have been determined, the observation equation (3.18) has the following form:

$$\left[\begin{array}{c} \eta^d \\ \hline (\frac{\mu(f)}{\sqrt{N_s}})1_{N_\theta} \end{array}\right] = \left[\begin{array}{c|c} C_{dd} & C_{da}^T \\ \hline C_{da} & C_{aa} \end{array}\right] \left[\begin{array}{c} \xi^d \\ \hline (\frac{1}{N_\theta N_s})(\frac{\mu(f)}{\sqrt{N_s}})1_{N_\theta} \end{array}\right] \tag{3.85}$$

$$= \left[\begin{array}{c|c} C_{dd} & s \\ \hline C_{da} & r1_{N_\theta} \end{array}\right] \left[\begin{array}{c} \xi^d \\ \hline (\frac{1}{N_\theta N_s})(\frac{\mu(f)}{\sqrt{N_s}}) \end{array}\right], \tag{3.86}$$

where s a vector containing the row sums of C_{da}^T, and recall that $r \approx N_\theta N_s$ is the row sum of C_{aa}. By adding the last N_θ equations in the above matrix partitioned equation, we get:

$$\left[\begin{array}{c} \eta^d \\ \hline (\frac{\mu(f)N_\theta}{\sqrt{N_s}}) \end{array}\right] = \left[\begin{array}{c|c} C_{dd} & s \\ \hline s^T & N_\theta^2 N_s \end{array}\right] \left[\begin{array}{c} \xi^d \\ \hline (\frac{1}{N_\theta N_s})(\frac{\mu(f)}{\sqrt{N_s}}) \end{array}\right]. \tag{3.87}$$

The following expression for the detail coefficients ξ^d, which incorporates the C_{da} coupling, can now be obtained by applying the matrix inversion lemma [41] to the matrix partitioned equation (3.87):

$$\xi^d = C_{dd}^{-1}\eta^d - C_{dd}^{-1}s \left(\frac{s^T C_{dd}^{-1}\eta^d - (\frac{\mu(f)N_\theta}{\sqrt{N_s}})}{s^T C_{dd}^{-1}s - N_\theta^2 N_s} \right). \tag{3.88}$$

Note that if there is no C_{da} coupling then $s = 0$, and we obtain our ideal estimate $\xi^d = C_{dd}^{-1}\eta^d$.

We next present an example that demonstrates the fact that the accuracy of the finest scale multiscale reconstruction improves with the introduction of the C_{da} coupling. For this example we consider the reconstruction of the 32 × 32 phantom shown in Figures 3-6 and 3-7 both with and without the introduction of the C_{da} coupling, from projection data gathered at $N_\theta = N_s = 32$. In Figure 3-18 we show a

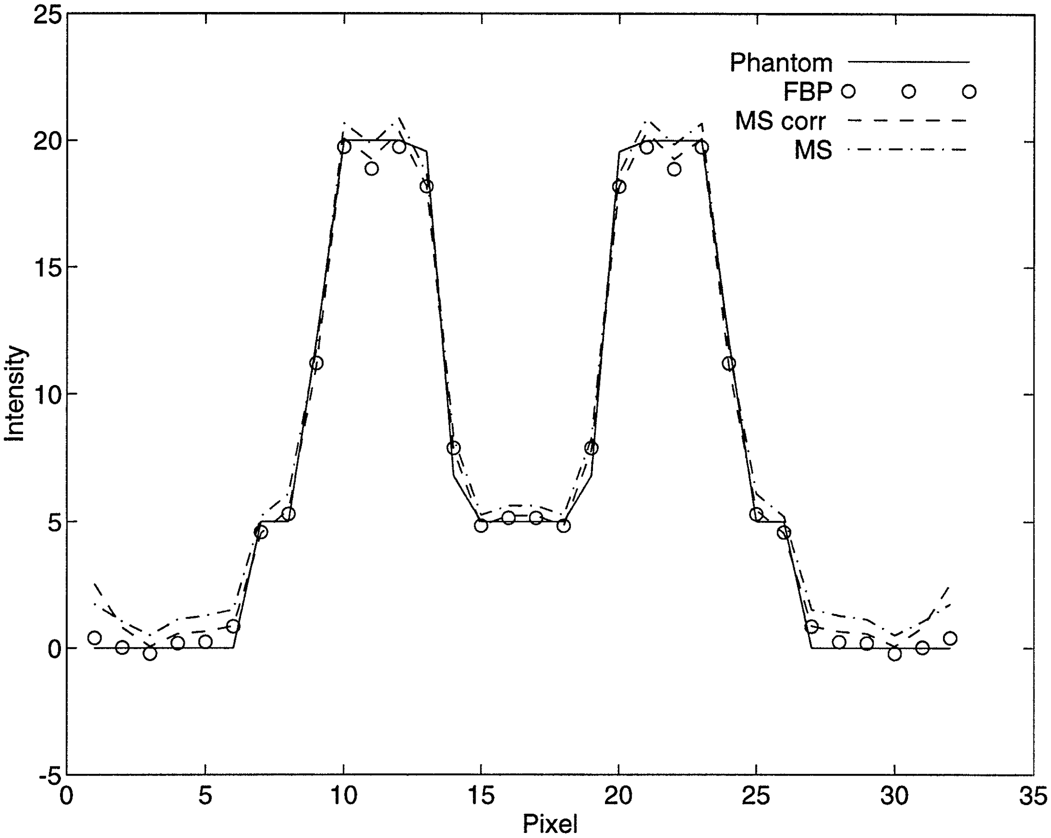


Figure 3-18: A horizontal section through the phantom (solid line), the FBP reconstruction (circles), and the multiscale reconstructions obtained with (dashed line) and without (dash-dotted line) the introduction of the $\mathcal{C}_{da}$ coupling.

central section through both the multiscale reconstructions of this phantom which are obtained by neglecting the $\mathcal{C}_{da}$ coupling (dash-dotted line) and by incorporating the $\mathcal{C}_{da}$ coupling (dashed line) respectively. Note from the figure that the reconstruction obtained by incorporating the coupling is closer to both the phantom as well as the FBP reconstruction than the one which ignores this coupling.

3.F Justification of (3.34)

The MAP estimate of ξ^a corresponding to the observation equation (3.29) (assuming $\mathcal{C}_{da} = 0$):

$$\eta^a = \mathcal{C}_{aa}\xi^a + \nu^a, \qquad \nu^a \sim \mathcal{N}(0, \lambda I_{N_\theta}), \tag{3.89}$$

and the prior model (c.f. (3.31)):

$$\xi^a \sim \mathcal{N}(0, \Lambda_a = (\frac{1}{\epsilon})I_{N_\theta}), \tag{3.90}$$

is given by:

$$\widehat{\xi}^a = (\lambda\epsilon I_{N_\theta} + \mathcal{C}_{aa}^2)^{-1}\mathcal{C}_{aa}\eta^a. \tag{3.91}$$

Since $\mathcal{C}_{aa}$ is a symmetric, circulant matrix, it can be written as (c.f. (3.67)):

$$\mathcal{C}_{aa} = FMF \tag{3.92}$$

where F is the matrix representation of the discrete Fourier transform operation, and M is a real, diagonal matrix. Since $\mathcal{C}_{aa}$ and W are real matrices, we have:

$$\mathcal{C}_{aa}^* = F^*MF^* = \mathcal{C}_{aa}, \tag{3.93}$$

and

$$\mathcal{C}_{aa}^2 = \mathcal{C}_{aa}^*\mathcal{C}_{aa} = F^*MF^*FMF. \tag{3.94}$$

By substituting (3.93) and (3.94) in the MAP estimate (3.91), and using (3.59), we get:

$$\begin{aligned}
\widehat{\xi}^a &= (\lambda\epsilon I_{N_\theta} + F^*MF^*FMF)^{-1}F^*MF^*\eta^a \\
&= (\lambda\epsilon I_{N_\theta} + F^*M^2F)^{-1}F^*MF\eta^a \\
&= (F^*(\lambda\epsilon I_{N_\theta} + M^2)F)^{-1}F^*MF^* \\
&= F^*(\lambda\epsilon I_{N_\theta} + M^2)^{-1}MF^*\eta^a \\
&= G\eta^a, \qquad (3.95)
\end{aligned}$$

where

$$\begin{aligned}
G &\triangleq F^*(\lambda\epsilon I_{N_\theta} + M^2)^{-1}MF^* \qquad (3.96) \\
&= F^*QF^*, \qquad (3.97)
\end{aligned}$$

with

$$Q \triangleq (\lambda \epsilon I_{N_\theta} + M^2)^{-1} M. \tag{3.98}$$

Assume that the diagonal matrix M has first c diagonal entries non-zero corresponding to a rank of c for $\mathcal{C}_{aa}$:

$$M = \mathrm{diag}(m_1, m_2, \ldots, m_c, 0, \ldots, 0). \tag{3.99}$$

By substituting (3.99) in (3.98) we obtain:

$$Q = \mathrm{diag}(\frac{m_1}{m_1^2 + \lambda\epsilon}, \frac{m_2}{m_2^2 + \lambda\epsilon}, \ldots, \frac{m_c}{m_c^2 + \lambda\epsilon}, 0, \ldots, 0). \tag{3.100}$$

Now if $\epsilon \to 0$, then from (3.100):

$$Q \to M^+, \tag{3.101}$$

and from (3.97):

$$G \to F^* M^+ F^*. \tag{3.102}$$

But from (3.92):

$$\mathcal{C}_{aa}^+ = F^* M^+ F^*. \tag{3.103}$$

Thus, from (3.95), (3.102) and (3.103), as $\epsilon \to 0$:

$$\widehat{\xi}^a \to \mathcal{C}^+ \eta^a. \tag{3.104}$$

Chapter 4

A Wavelet Transform-Based Multiscale Projection-Domain Algebraic Reconstruction Technique (MPART)

4.1 Introduction

Recall that in Chapter 3 we had developed a multiscale reconstruction technique based on the natural pixel (NP) representation for tomography. This multiscale technique had resulted in a sparse system of equations which, in addition, could be naturally partitioned by scales such that the reconstruction procedure required the solution of only the well-conditioned block. In this chapter we develop an efficient iterative technique to solve such a sparse and well-conditioned system of equations that arises in the multiscale framework. This iterative technique enables us to obtain reconstruction estimates at multiple scales from sparse and/or noisy data with very little storage or computational complexity.

Specifically, in this chapter we use a Kaczmarz-like method (c.f. Chapter 2) to solve the sparse and well-conditioned system of equations that arises in our NP-based multiscale framework. Recall that the computational complexity of the Kaczmarz method is equal to the number of computations required in each iteration times the number of iterations needed for convergence. Further, the computations required per iteration are proportional to the number of non-zero elements in the corresponding system matrix, and the number of iterations needed for convergence are a function of the condition number of the same matrix. Thus, since the multiscale system matrix is both sparse as well as well-conditioned, the Kaczmarz method provides an extremely efficient technique for obtaining the solution of the multiscale system of equations.

We point out that the same Kaczmarz method that we implement in this chapter arises as the basis of the algebraic reconstruction technique (ART) where it is used to solve for the object pixel coefficients from the system of equations defined by the projection matrix. As a result we consider our iterative method as an alternative to ART and refer to it as the multiscale projection-domain algebraic reconstruction technique (MPART)[1]. The MPART, however, holds several advantages over ART. The NP representation used in MPART is free of the model errors arising in the rectangular pixel object expansion used by ART. Moreover, while the projection matrix of ART is generally ill-conditioned, MPART uses a relatively well-conditioned matrix. This amounts to the MPART reconstruction being relatively insensitive to the relaxation parameter and the ordering scheme, which are critical for convergence in ART. Further, even though the ART and the MPART system matrices have approximately similar sparsity, the smaller magnitude elements in the latter can be effectively thresholded to zero without affecting the quality of the resulting reconstruction. This implies that similar quality ART and MPART reconstructions may be obtained by using substantially fewer elements in the MPART matrix than in the ART matrix. This translates into MPART requiring substantially less computations per iteration than in the case of ART. Further, MPART yields reconstruction estimates at multiple scales for free, which can be used for object feature extraction directly from the

---

[1] The term *projection-domain* is used to describe the algorithm because the NP method represents the object in the domain where the strip-integral data (commonly referred to as the projection data) resides.

strip-integral data.

Finally, Herman [39] has proposed a variant of ART that computes the maximum aposteriori probability (MAP) solution for the object coefficients from noisy measurement data, based on a prior model that assumes these coefficients to be uncorrelated. This regularized solution is obtained with essentially the same computational complexity and storage requirements as the original ART. Similarly, in the presence of noise, we are also able to extend our MPART to yield regularized reconstructions with no additional computations or storage. These regularized reconstructions are based on the same self-similar prior model for the multiscale object coefficients that we had developed in Chapter 3. Recall that such a self-similar model is captured in our framework by assuming that the multiscale object coefficients are uncorrelated, and also that the variance of the coefficients decreases geometrically from coarse to fine scales. This prior model not only imposes the classical smoothness constraint, but also results in fractal-like models for the object. Such models have been used extensively to capture terrain and biological structures such as liver and lung [11,12,78,80].

The chapter is organized as follows. In Section 4.2 we review the relevant ART and the NP results from Chapter 2. In Section 4.3 we outline the main points of our NP-based multiscale object representation from Chapter 3 and develop MPART. We then present some example reconstructions illustrating the salient features of MPART. In Section 4.4 we describe the extension of MPART that generates regularized reconstructions in the presence of measurement noise. Section 4.5 concludes this chapter.

4.2 Background

In this section we briefly review the relevant algebraic reconstruction technique (ART) results and also the results from the natural pixel (NP) reconstruction technique that we had presented in Chapter 2. These results will prove to be essential for the development and understanding of MPART.

4.2.1 Tomography setup

Recall that in tomography the observation corresponding to strip ℓ at angular position k is given by:

$$y(m) = \iint\limits_{\Omega} f(u,v)\, S_m(u,v)\, du\, dv, \qquad m = 1, \ldots, N_\theta N_s, \tag{4.1}$$

where $S_m(u,v)$, with $m = (k-1)N_s + \ell$, is the indicator function of the strip integral corresponding to the observation $y(m)$. For simplicity, as in the previous chapters, we assume a 2-D parallel-beam imaging geometry. Here the data consists of parallel, non-overlapping strip integrals through the object at various angles (c.f. Figure 2-1). Recall that such an imaging geometry is defined by N_θ angular positions and N_s parallel strip integrals at each angular position.

4.2.2 Conventional ART

In ART [39, 44], the object $f(u,v)$ is expanded in a standard, $N_s \times N_s$ rectangular pixel basis:

$$f(u,v) = \sum_{n=1}^{N_s^2} f(n) b_n(u,v), \tag{4.2}$$

where $b_n(u,v)$ is the indicator function of pixel n, and $\{f(n)\}$ are the pixel coefficients of the object. Now by substituting the ART object representation equation (4.2) into the expression for the strip-integral data (c.f. (4.1)), we obtain the following expression for the overall observation vector y which contains the set of strip integral data $\{y(m), m = 1, \ldots, N_\theta N_s\}$:

$$y = Tf, \tag{4.3}$$

where T is the $N_\theta N_s \times N_s^2$ system *projection matrix*, the (m,n)th element of which is given by:

$$T_{mn} = \iint\limits_{\Omega} b_n(u,v)\, S_m(u,v)\, du\, dv, \qquad m = 1, \ldots, N_\theta N_s;\ n = 1, \ldots, N_s^2. \tag{4.4}$$

We refer the reader to Figure 2-2 where we show a picture of the 1024×1024 projection matrix T for an imaging geometry with $N_\theta = N_s = 32$. Note that the projection matrix, although very large, is *sparse* because each strip intersects only a small number of object pixels.

The sparsity of the projection matrix T is exploited in ART by the Kaczmarz projection[2] method that is used to solve the system of equations (4.3) for the object coefficients f. This method cyclically projects an initial guess for f, say $f^{(0)}$, onto the hyperplanes defined by:

$$\{f | T(k)f = y(k)\}, \tag{4.5}$$

for $k = 1, \ldots, N_\theta N_s$, where $T(k)$ is the k-th row of T, and $y(k)$ is the k-th data sample. To begin, the initial guess $f^{(0)}$ is projected onto the hyperplane corresponding to $k = 1$ in (4.5) to obtain the solution $f^{(1)}$. This solution is in turn projected onto the hyperplane corresponding to $k = 2$ to obtain the solution $f^{(2)}$. This process is repeated until all $N_\theta N_s$ hyperplanes have been exhausted and a solution $f^{(N_\theta N_s)}$ has been obtained. This completes the first iteration in ART. The second iteration is started by projecting the result of the first iteration onto the hyperplane represented by $k = 1$ in (4.5). These iterations are continued until convergence is achieved.

Specifically, the following recursion is used in ART to obtain the $(i+1)$-th estimate of the object coefficients from the corresponding i-th estimate:

$$f^{(i+1)} = f^{(i)} + r^{(k)} \left(\frac{y(k) - T(k)f^{(i)}}{T(k)T(k)^T} \right) T(k)^T, \qquad k = (i)\mathrm{mod}(\mathrm{N_\theta N_s}), \tag{4.6}$$

where $f^{(i)}$ is the i-th estimate of f, $T(k)$ refers to the k-th row of T, and $r^{(k)}$ is the relaxation parameter for the k-th iteration[3], with $0 \leq r^{(k)} \leq 2$.

Recall from Chapter 2 that the computational complexity per iteration of ART is proportional to the number of non-zero elements in the projection matrix T. Recall

---

[2] This is not to be confused with the term projection that is frequently used (including in this thesis) for tomographic data.

[3] The recursion (4.6) corresponds exactly to a projection of $f^{(i)}$ onto the hyperplane defined by (4.5) only if $r^{(k)} = 1$. This case with $r^{(k)} = 1$ corresponds to the standard Kaczmarz method. In general other choices of $r^{(k)}$ are used to speed convergence.

further that the number of iterations required for convergence of the solution $f^{(i)}$ in ART is critically dependent on the condition number of T. The convergence is fast if the matrix T is well-conditioned. On the other hand, if T is ill-conditioned, the convergence is slow and, in addition, is critically dependent on the choice of the ordering scheme used for rows of T, and also on the relaxation parameter [40]. Note that as (4.6) currently stands, the rows of T are accessed in a *sequential* order. However, if at each step the row accessed is that which has the largest angle to the subspace spanned by all the rows of T used in the previous steps of the same iteration, the convergence can be improved. This is exactly the motivation behind a recent ordering scheme proposed by Herman and Meyer [40]. However, an improved convergence is seen even with a random ordering scheme, where the rows of T are accessed in a random fashion. To obtain a fair comparison between ART and MPART, we will present ART reconstructions using not only a sequential ordering scheme but also a random ordering scheme and the ordering scheme proposed by Herman and Meyer. The second critical parameter affecting convergence is the relaxation parameter $r^{(k)}$ for iteration k. This value is typically determined by experimentation on test images which are similar in some sense to the image that is to be reconstructed [40].

In Section 4.3 we will present an efficient method for isolating and removing the ill-conditioning in the projection matrix T. As a result, we will perform Kaczmarz recursions of the type (4.6) on a system of equations where the system matrix is well-conditioned. This will result in our modified ART algorithm being relatively insensitive to the ordering scheme and to the choice of the relaxation parameter. This gain will *not be* at the expense of any additional computations per iteration, since our well-conditioned matrix will have a similar sparsity as that of the ART projection matrix[4] T.

---

[4] As an example, for an imaging geometry with $N_\theta = N_s = 32$, the projection matrix T has a condition number of 2.3×10^{16}. Our method for removing the ill-conditioning, however, results in a matrix with a condition number of 1.3×10^{3}. In addition, since the two matrices have a similar percentage of non-zero elements ($\sim 6.5\%$), the computational complexity per iteration of the Kaczmarz method is same for the two cases.

4.2.3 The Natural Pixel Object Representation

There are two important sources of inconsistency in the system of equations (4.3) used by ART (c.f. Chapter 3). The first is due to the presence of measurement noise. The second is due to the fact that the form of the rectangular pixel representation (4.2) used by ART is inadequate for accurate representation of the measurements of a real physical object. In the NP method [9, 10] this second inconsistency arising due to the standard rectangular pixel object representation in ART, is removed by representing the object in the same functions (i.e. strips $\{S_m(u,v)\}$) along which the strip integral data y are collected. In particular:

$$f(u,v) = \sum_{n=1}^{N_\theta N_s} x_n S_n(u,v), \tag{4.7}$$

where x_n is the coefficient of expansion of the object corresponding to strip $S_n(u,v)$. Now by substituting the NP object representation equation (4.7) into the expression for the strip-integral data (c.f. (4.1)), we obtain the following expression for the overall observation vector y which contains the set of strip integral data $\{y_m, m = 1, \ldots, N_\theta N_s\}$:

$$y = Cx, \tag{4.8}$$

where C is a $N_\theta N_s \times N_\theta N_s$ matrix, the (m,n)-th element of which is given by:

$$C_{mn} = \iint_{\Omega} S_n(u,v)\, S_m(u,v)\, du\, dv, \qquad m = 1, \ldots, N_\theta N_s;\ n = 1, \ldots, N_\theta N_s. \tag{4.9}$$

Finally, even though no assumption about the discretization of the object on a rectangular pixel grid is needed to solve for x from (4.8), once the coefficients x have been calculated, a discretization of (4.7) is needed to display the reconstruction. As a result of this discretization the reconstruction is obtained as:

$$f = T^T x, \tag{4.10}$$

where T is the projection matrix containing the discretized strips (c.f. (4.4)), and f is the vector containing the object expansion coefficients in the rectangular pixel basis (c.f. (4.2)).

As mentioned earlier, the NP representation is free of the model errors which arise in the standard rectangular pixel representation of the object. However, this benefit comes with a cost – namely the NP matrix C is *full*[5]. The Kaczmarz method for solving the NP system of equations (4.8) thus results in a computationally inefficient algorithm. In the next section we transform the strip-integral data y and the object coefficients x in a wavelet-derived multiscale basis. Recall from Chapter 3 that this results in a *sparse* multiscale system matrix. In addition, any ill-conditioning in the NP matrix C is compressed into a few coarse scale elements of the multiscale matrix. We exploit this feature to partition the multiscale system matrix by scales such that the Kaczmarz method utilizes a well-conditioned matrix. This results in our reconstruction algorithm being relatively insensitive to the relaxation parameter and the ordering scheme, which are critical for convergence in ART.

4.3 Multiscale Projection Domain ART

4.3.1 Theory

We begin the development of MPART by outlining the main results from Chapter 3 regarding the multiscale transformation of the NP representation. Let W be a $N_s \times N_s$ matrix which when applied to a 1-D vector of length $N_s = 2^J$, results in a vector containing the wavelet representation of the original 1-D vector (c.f. Section 2.5)[6]. The matrix W represents the choice of a specific wavelet and scaling function that are used for multiscale representation. We assume these to be orthonormal which results in W being orthogonal, i.e. $W^{-1} = W^T$. If we define a block diagonal matrix W_b

---

[5] The elements of C are given by the areas of intersection of the strips (see (4.9)) – since most of these areas of intersection are non-zero, C is full. We refer the reader to Figure 2-3 which shows a picture of C for an imaging geometry with $N_\theta = N_s = 32$.

[6] Recall that the wavelet representation operation can be performed in an extremely efficient manner. In particular, representation of a vector of length N_s can be obtained in $O(N_s)$ multiplications.

consisting of N_θ blocks along the diagonal, all equal to W, then W_b is also orthogonal, with $W_b^{-1} = W_b^T$. This matrix W_b when applied to a 1-D vector of length $N_\theta N_s$ (for example the data vector y or the object coefficient vector x) results in a wavelet representation of all N_θ components of the vector.

Now, given W_b, the NP observation equation (4.8) can be transformed to a wavelet representation as follows:

$$\begin{aligned} W_b y &= (W_b C W_b^T)(W_b x), \text{ i.e.,} \\ \eta &= \mathcal{C}\xi, \end{aligned} \tag{4.11}$$

where $\eta = W_b y$ is the wavelet representation of the data vector, and $\mathcal{C}$ is the transformed system matrix which contains elements at multiple scales. It was shown in Chapter 3 that most of these elements are zero, resulting in a *sparse* $\mathcal{C}$. Specifically, the elements in $\mathcal{C}$ representing the fine scale couplings between η and ξ, are mostly zero. However the sub-matrices of $\mathcal{C}$ representing the coarse scale couplings have a larger fraction of non-zero elements. But since the fraction of fine scale terms in $\mathcal{C}$ is much larger then the coarse terms, the overall sparsity of $\mathcal{C}$ is governed mostly by the former. Equation (4.11) can be rearranged in scales (with the finest scale detail terms from all angles placed first and the coarsest scale approximation terms placed last) and written in a partitioned form as:

$$\left[\begin{array}{c} \eta^d \\ \hline \eta^a \end{array}\right] = \left[\begin{array}{c|c} \mathcal{C}_{dd} & \mathcal{C}_{da}^T \\ \hline \mathcal{C}_{da} & \mathcal{C}_{aa} \end{array}\right] \left[\begin{array}{c} \xi^d \\ \hline \xi^a \end{array}\right] \approx \left[\begin{array}{c|c} \mathcal{C}_{dd} & 0 \\ \hline 0 & \mathcal{C}_{aa} \end{array}\right] \left[\begin{array}{c} \xi^d \\ \hline \xi^a \end{array}\right], \tag{4.12}$$

where the vectors η^d and ξ^d contain all the detail terms at various scales and angles and are of length $N_\theta(N_s - 1)$, and the vectors η^a and ξ^a contain the coarsest scale approximation, i.e. the DC terms at all angles (one for each angle) and are of length N_θ. The upper left block $\mathcal{C}_{dd}$ is a $N_\theta(N_s - 1) \times N_\theta(N_s - 1)$ symmetric matrix, the elements of which are the areas of intersection of the detail basis functions (i.e. strips) at various scales and angles. The lower right block $\mathcal{C}_{aa}$ is a $N_\theta \times N_\theta$ symmetric matrix the elements of which are the areas of intersection of the coarsest scale basis function at each angle. Finally, the off-diagonal block $\mathcal{C}_{da}$ is a $N_\theta \times N_\theta(N_s - 1)$ matrix, the elements of which are the areas of intersection of the coarsest scale approximation basis functions and the detail basis functions at various scales, and are *negligible*.

It was shown in Chapter 3 that there exists a non-uniqueness in the NP representation, and that this non-uniqueness is fully captured in the multiscale framework by the sub-matrix $\mathcal{C}_{aa}$. Thus while the multiscale matrix $\mathcal{C}$ is ill-conditioned, the sub-matrix $\mathcal{C}_{dd}$ is relatively well-conditioned. As an example, for the parallel beam geometry shown in Figure 2-1 with $N_\theta = N_s = 32$, the condition number of sub-matrix $\mathcal{C}_{dd}$ is 100 times less[7] than that of the matrix C. Moreover, the sub-matrices in $\mathcal{C}_{dd}$ corresponding to fine scale details are much better conditioned than their coarse scale counterparts. This is shown in Figure 4-1 where the sub-matrix in $\mathcal{C}_{dd}$ corresponding to finest scale detail (i.e. scale 4) couplings between η^d and ξ^d has a condition number of 10.17, and this condition number gradually increases as the coarser scale detail couplings are included. We also refer the reader to Figure 3-3 for a picture of the matrix

$$\mathcal{C}_s \triangleq \left[\begin{array}{c|c} \mathcal{C}_{dd} & \mathcal{C}_{da}^T \\ \hline \mathcal{C}_{da} & \mathcal{C}_{aa} \end{array}\right] \tag{4.13}$$

which, as mentioned earlier, is a scale-rearranged version of the multiscale matrix $\mathcal{C}$. One can see from Figure 3-3 that the fine scale sub-matrices in $\mathcal{C}$ are extremely sparse, with the sparsity decreasing at coarser scales.

Similar to the observation equation, the NP back-projection equation (4.10) can also be transformed to a wavelet representation:

$$f = (W_b T)^T (W_b x) = \mathcal{T}^T \xi, \tag{4.14}$$

where $\xi = W_b x$ is the wavelet representation of the object coefficient vector, and $\mathcal{T} = W_b T$ are the transformed basis functions[8].

To summarize, in the wavelet-transformed NP representation the tomography

---

[7] The matrices $\mathcal{C}$ and $\mathcal{C}_{dd}$ have condition numbers of $140,000$ and $1,287.50$, respectively, in this example.

[8] Refer to Chapter 3 for an example of these transformed basis functions.

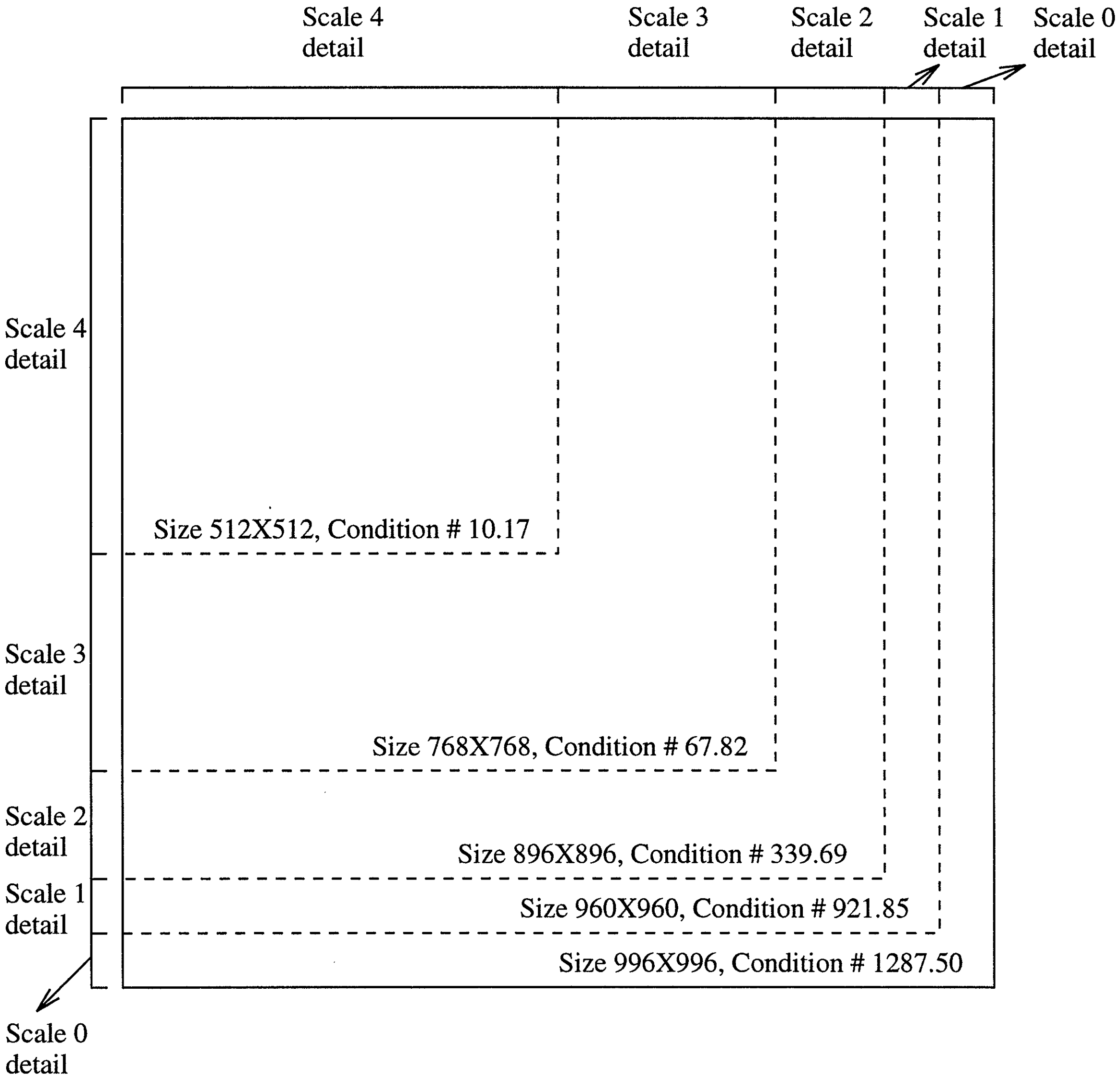


Figure 4-1: The condition number of various multiscale sub-matrices in C_{dd}.

problem reduces to solving for the detail and approximation object coefficients, ξ^d and ξ^a respectively, from (4.12), and subsequently back-projecting these along the corresponding transformed basis functions according to (4.14). Since the off-diagonal block $\mathcal{C}_{da}$ in the wavelet-transformed observation equation (4.12) is effectively zero, the detail and approximation equations decouple to yield a *small* ($N_\theta \times N_\theta$) *ill-conditioned* system of equations:

$$\eta^a = \mathcal{C}_{aa}\xi^a, \tag{4.15}$$

and a *large* ($N_\theta(N_s-1) \times N_\theta(N_s-1)$), *sparse* and *well-conditioned* system of equations:

$$\eta^d = \mathcal{C}_{dd}\xi^d. \tag{4.16}$$

Since (4.15) is ill-posed, we solve for ξ^a, the object approximation coefficients at the coarsest scale, by calculating the minimum-norm solution to (4.15). This solution is given by[9] (c.f. Chapter 3):

$$\xi^a = \mathcal{C}_{aa}^{+}\eta^a \approx (\frac{\mu(f)}{N_\theta N_s\sqrt{N_s}})1_{N_\theta}, \tag{4.17}$$

where $\mathcal{C}_{aa}^{+}$ is the pseudo-inverse of $\mathcal{C}_{aa}$, 1_{N_θ} refers to a vector of length N_θ with all elements equal to unity, and $\mu(f)$ is the total mass under the object. What remains now is to solve for the detail vector ξ^d from (4.16). Since $\mathcal{C}_{dd}$ is sparse, we use the Kaczmarz method to iteratively solve for ξ^d:

$$\xi^{(i+1)} = \xi^{(i)} + r^{(k)}(\frac{\eta(k) - \mathcal{C}(k)\xi^{(i)}}{\mathcal{C}(k)\mathcal{C}(k)^T})\mathcal{C}(k)^T, \qquad k = (i)\mathrm{mod}(\mathrm{N}_\theta(\mathrm{N}_s - 1)), \tag{4.18}$$

where $r^{(k)}$ is the relaxation parameter for the k-th iteration with $0 \leq r^{(k)} \leq 2$, and, for notational simplicity, we simply refer to the k-th row of $\mathcal{C}_{dd}$ as $\mathcal{C}(k)$, to the k-th element in η^d as $\eta(k)$, and to the ith estimate of ξ^d as $\xi^{(i)}$.

Thus to summarize, our MPART reconstruction consists of solving for the coarsest scale object coefficients from (4.17), solving for the detail object coefficients by using

---

[9] This assumes that the object is completely contained in the field-of-view.

the Kaczmarz method (4.18), and back-projecting these approximation and detail coefficients along the corresponding transformed basis functions according to (4.14). Note that in the recursion (4.18) the rows of $\mathcal{C}_{dd}$ are accessed in a sequential order. As in the case of ART, we will also present MPART reconstructions obtained by accessing the rows of $\mathcal{C}_{dd}$ in a random order. However, since $\mathcal{C}_{dd}$ is well-conditioned, the rate of convergence of the MPART reconstruction using this random ordering scheme will not be much different than the one obtained by using sequential ordering.

We point out that, as suggested by Figure 4-1, there is a tradeoff between the rate of convergence in MPART and the number of scales desired in the reconstruction. In particular, the convergence if extremely fast if only the fine scale features in the reconstruction are desired. This is because the block of the matrix $\mathcal{C}$ that represents the coupling between fine scale coefficients in the data and in the object, is extremely well-conditioned. However this condition number gradually increases as coarser couplings are included, thereby leading to slower convergence if coarser scale information in the object is also to be reconstructed. Motivated by this, we propose in Chapter 8 a modification of MPART where a faster convergence rate can possibly result through the introduction of a scale-recursive structure. In any case, even if information at all scales is desired, and even if the scale-recursive structure is not used, the convergence in MPART is still much faster than ART because the overall MPART matrix $\mathcal{C}_{dd}$, including contribution from all scales, is much better conditioned than the ART projection matrix T.

Finally, the form of the back-projection (4.14) induces a multiscale object representation. As an example, if only a coarse estimate of the reconstruction is desired then only the coarse scale object coefficients ξ have to be back-projected. Similarly, if the interest is to reconstruct only the fine scale features in the object (for example, edges and boundaries), then only the fine scale object coefficients are to be back-projected. To motivate this, we will present an example in the next section where we will reconstruct an object at multiple scales. Again, we refer the reader to Chapter 3 for details on the mathematical formulation of the multiscale representation.

4.3.2 Examples

We next present some example ART and MPART reconstructions. In all of the reconstructions we present in this paper, we start the ART and MPART iterations with a zero initial guess, and we measure the fidelity of the reconstructions by calculating the relative reconstruction error, defined as:

$$\text{Relative reconstruction error} = \frac{||\hat{f} - f||^2}{||f||^2}, \tag{4.19}$$

where f is the vector of rectangular pixel expansion coefficients of the object, and $\hat{f}$ is the vector of corresponding reconstructed coefficients. The latter are calculated from (4.3) in the case of ART and from (4.14) in the case of MPART. Finally, in MPART, we select the wavelet representation matrix W such that it reflects the use of the Daubechies D_3 wavelet [18].

In the next few examples we demonstrate that the MPART reconstructions can be obtained with substantially less computations that the corresponding ART reconstructions. To begin, recall that the computational complexity of the Kaczmarz method is proportional to the product of the number of computations required in each iteration, and the number of iterations needed for convergence. Further, the computations required per iteration are proportional to the number of non-zero elements in the corresponding system matrix. Thus MART can be made more efficient relative to ART by either accelerating the convergence of the solution or by making the system matrix more sparse. In fact, as we demonstrate next, the efficiency of MPART arises as a result of both of these factors. First we show examples which demonstrate that the MPART reconstructions converge faster than the corresponding ART reconstructions. We subsequently show that, even though the ART and the MPART system matrices have approximately similar sparsity, the smaller magnitude elements in the latter can be effectively thresholded to zero without affecting the quality of the resulting reconstruction. This implies that similar quality ART and MPART reconstructions are obtained by using substantially less elements in the MPART matrix than in the ART matrix. This translates into MPART requiring substantially fewer computations per iteration than in the case of ART.

To begin, we present an example that demonstrates that the MPART reconstructions converge faster than the ART reconstructions, and that the rate of convergence in MPART is not significantly affected by the choice of the relaxation parameter and the ordering scheme. Figures 4-2 and 4-3 show a 32×32 phantom, and Figures 4-4 and 4-5 show ART and MPART reconstruction errors for this phantom as a function of iteration number for different values of relaxation parameter. The relaxation parameter is kept constant from iteration to iteration[10]. Besides presenting the ART results for sequential ordering, Figure 4-4 also shows the ART error values for random ordering as well as the ordering scheme recently proposed by Herman and Meyer [40]. These ordering schemes have been known to speed the convergence of ART and results using these are included to obtain a fair comparison between ART and MPART. Similarly, in Figure 4-5 we show MPART error values for both sequential and random ordering schemes. From Figures 4-4 and 4-5 we note that the reconstruction error in MPART converges in a fewer number of iterations than in ART. Further the steady state relative error value ($= 0.1$) in MPART is half the corresponding value ($= 0.2$) in the case of ART. In addition, one can observe that the rate of convergence in MPART is insensitive to both the relaxation parameter as well as the ordering scheme. This is in contrast to ART where, from Figure 4-4, the reconstruction error converges at different rates depending on the choice of relaxation parameter and the ordering scheme. Finally, since the sparsity of the ART and MPART matrices T and C_{dd} are approximately the same (93.9% and 93.5% respectively), the ART and MPART reconstructions shown in this example are obtained with similar computational complexity per iteration.

The next few examples demonstrate that the smaller magnitude elements in the MPART system matrix can be thresholded to zero without affecting the quality of the

---

[10] The convergence of ART can be significantly improved if instead of using a constant relaxation parameter for different iterations, the relaxation parameter is optimized as a function of iteration number by using test images that are representative of the image that is to be reconstructed. One such optimization scheme is described in [40]. However, for the purpose of this paper we do not vary the relaxation parameter with iteration since one of the goals of this paper is to show that the convergence of MPART is relatively insensitive to the relaxation parameter.

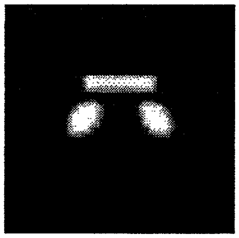


Figure 4-2: The grayscale plot (white corresponds to the maximum value and black to the minimum) of a 32×32 phantom that is used for comparing ART and MPART reconstructions. The projection data are gathered using a parallel beam geometry with $N_\theta = N_s = 32$.

corresponding reconstruction. A similar thresholding in case of the ART projection matrix results in reconstructions that suffer from severe artifacts. As mentioned earlier, this translates into MPART requiring substantially less computations per iteration than ART.

Figure 4-6 shows ART (using the ordering scheme of Herman and Meyer) and MPART reconstruction errors as a function of iteration number for a relaxation parameter value of 0.5. The error values are shown both for exact ART and MPART reconstructions (where the exact matrices T and C_{dd} are used) and for approximate reconstructions where the smaller magnitude elements in the corresponding system matrices are thresholded to zero. We threshold the two matrices such that they result in a similar sparsity (97.4% and 97.3%) and hence in similar computational complexity per iteration. We note from the figure that, as compared to ART, there is a *graceful degradation* in the MPART reconstruction when the smaller magnitude elements in the corresponding matrices are neglected. Since the computational complexity per iteration of the Kaczmarz method is proportional to the number of non-zero elements in the system matrix, this approximate reconstruction is obtained with far fewer computations than the exact reconstruction. Specifically, in this example 2.5 times less computations are required when using the approximate matrices instead of the exact counterparts.

Figures 4-7 and 4-8 show a 64×64 phantom, and Figure 4-9 shows the errors in ART and MPART reconstructions for this phantom after 6 iterations as a function of the relaxation parameter. The matrices T and C_{dd} are thresholded to a similar

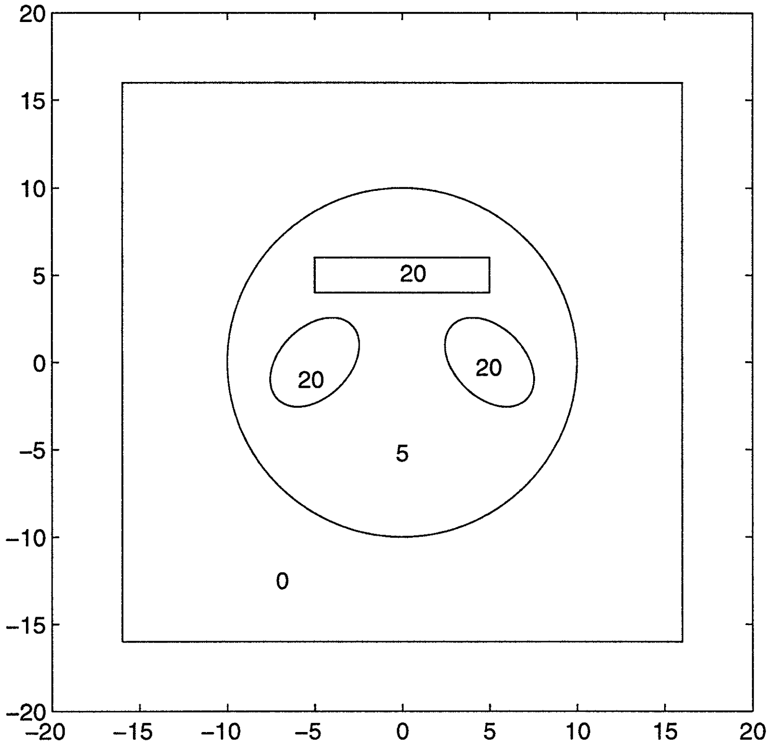


Figure 4-3: The intensity map of the 32 × 32 phantom shown in Figure 4-2. The numbers shown are intensities of respective regions. The projection data are gathered using a parallel beam geometry with $N_\theta = N_s = 32$.

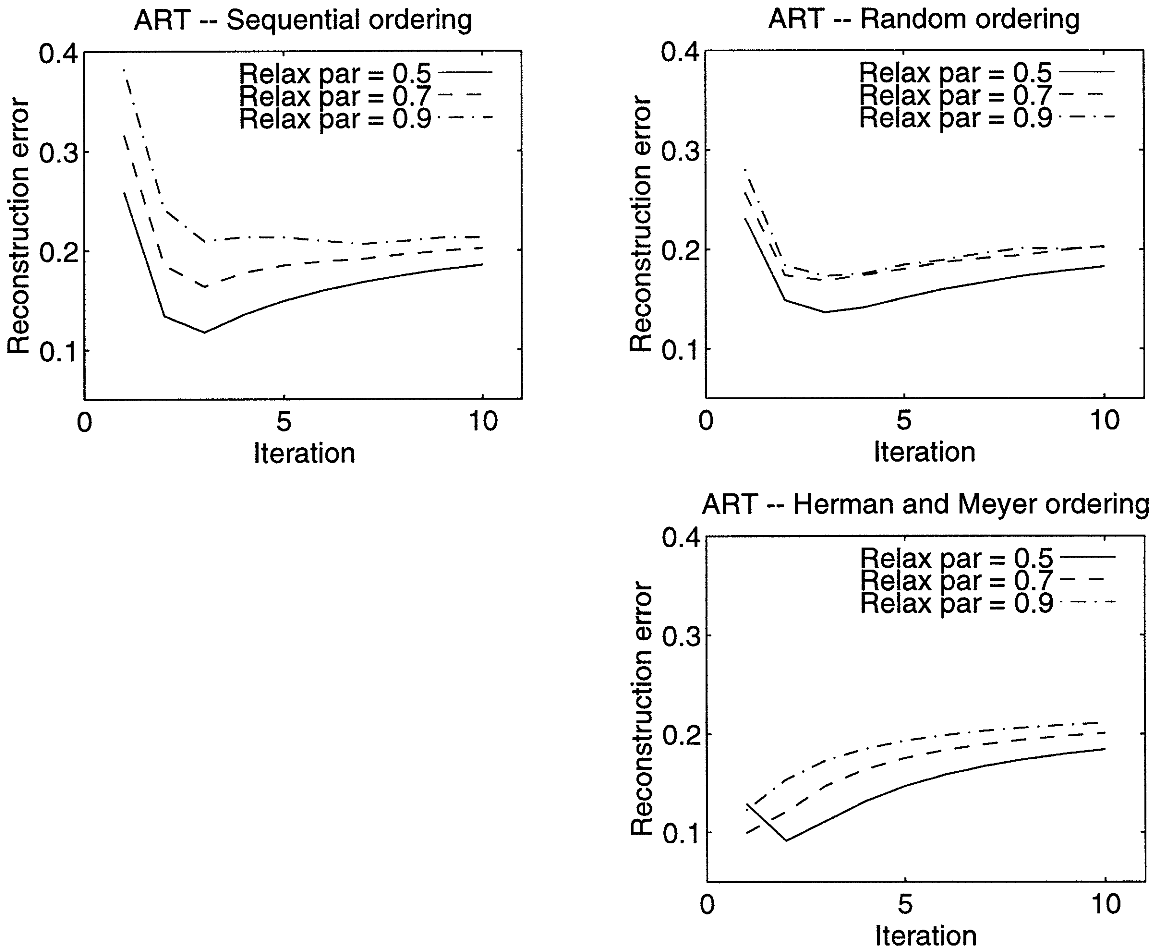


Figure 4-4: The reconstruction error values for ART as a function of iteration number for different values of the relaxation parameter. The 32×32 phantom shown in Figure 4-2 is used for reconstruction. The relaxation parameter is assumed constant in between iterations. The ART matrix T is 93.9% sparse.

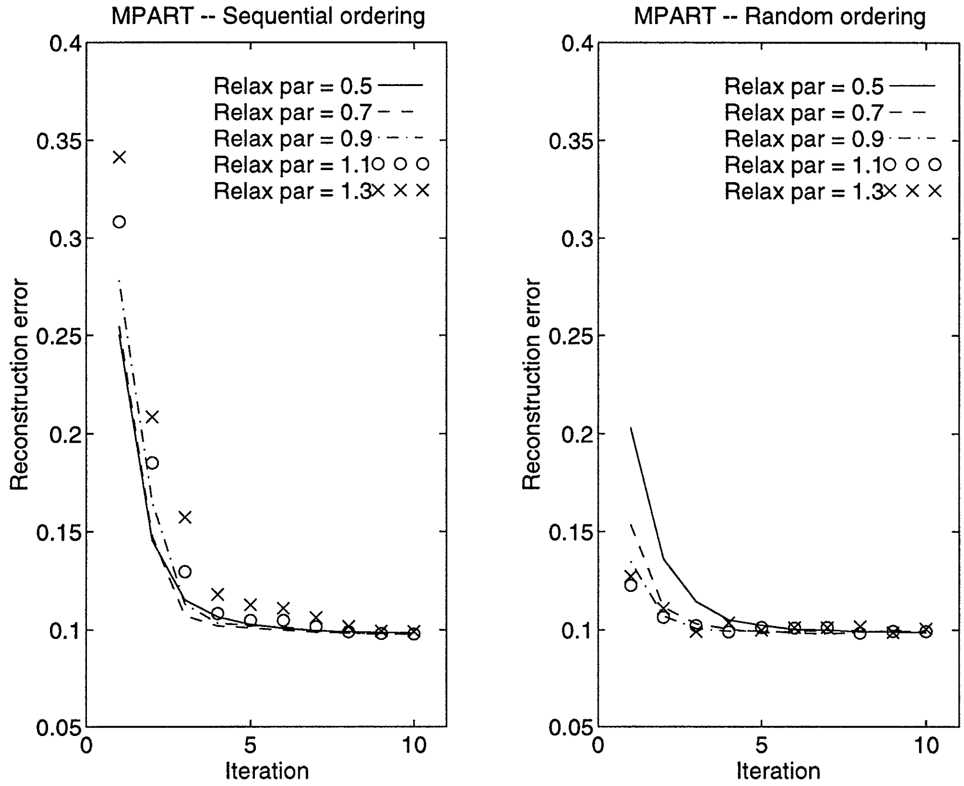


Figure 4-5: The reconstruction error values for MPART as a function of iteration number for different values of the relaxation parameter. The 32×32 phantom shown in Figure 4-2 is used for reconstruction. The relaxation parameter is assumed constant in between iterations. The MPART matrix C_{dd} is 93.5% sparse.

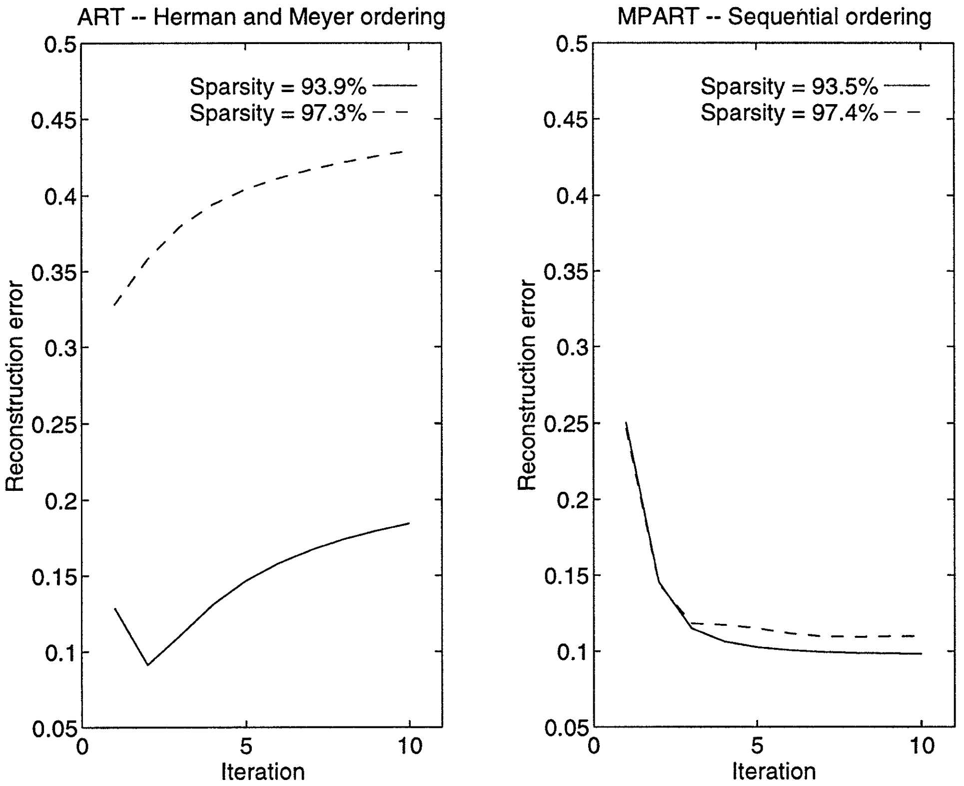


Figure 4-6: The reconstruction error values for ART and MPART as a function of iteration number for the 32×32 phantom shown in Figure 4-2. The value of relaxation parameter used is 0.5. Results are presented both for exact ART and MPART matrices (sparsity 93.9% and 93.5% respectively) and for approximate matrices (sparsity 97.3% and 97.4% respectively).

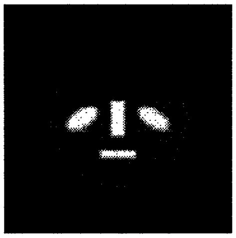


Figure 4-7: The grayscale plot (white corresponds to the maximum value and black to the minimum) of a 64×64 phantom that is used for comparing ART and MPART reconstructions. The projection data are gathered using a parallel beam geometry with $N_\theta = N_s = 64$.

sparsity value so that these methods have similar computational complexity per iteration. It can be seen from Figure 4-9 that the MPART method results in much lower reconstruction errors than ART. In addition, the reconstruction errors in MPART are once again insensitive to the relaxation parameter. In Figure 4-10 we show a section through the "best" approximate ART and MPART reconstructions (i.e. reconstructions obtained using the relaxation parameter that results in least error) after 6 iterations, corresponding to the phantom in Figure 4-7. Note that the MPART reconstruction is much closer to the phantom than the corresponding ART reconstruction. The reconstructed pixel values in ART "overshoot" the phantom and, as a consequence, the ART reconstruction in this case suffers from severe streaking artifacts. Finally, we note that in the approximate ART reconstructions (where we threshold the lower magnitude elements in the projection matrix T to zero) the random ordering scheme and the ordering scheme of Herman and Meyer result in very similar error values. An example of this is seen in Figure 4-9. Thus from now on we only use the random ordering scheme for approximate ART reconstruction results.

Figures 4-11 and 4-12 show a 128×128 phantom, and Figure 4-13 shows the errors in ART and MPART reconstructions for this phantom after 5 iterations as a function of the extent of approximation on the corresponding matrices T and C_{dd}. As before, we approximate these matrices by thresholding the smaller magnitude elements to zero and measure the extent of approximation by the percentage of non-zero elements left in the matrices after the thresholding operation. The random and sequential ordering schemes are used respectively for ART and MPART reconstructions. From

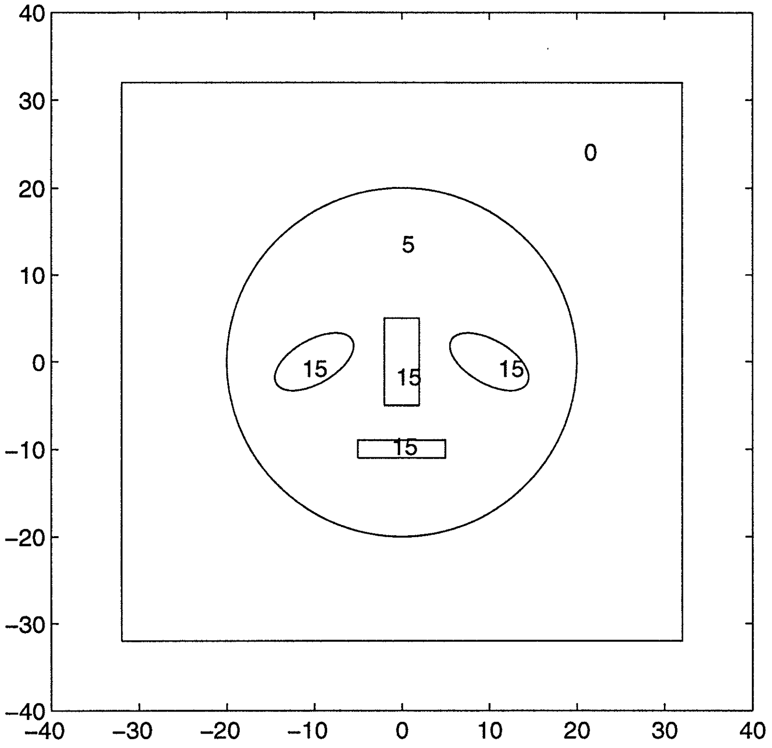


Figure 4-8: The intensity map of the 64 × 64 phantom shown in Figure 4-7. The numbers shown are intensities of respective regions. The projection data are gathered using a parallel beam geometry with $N_\theta = N_s = 64$.

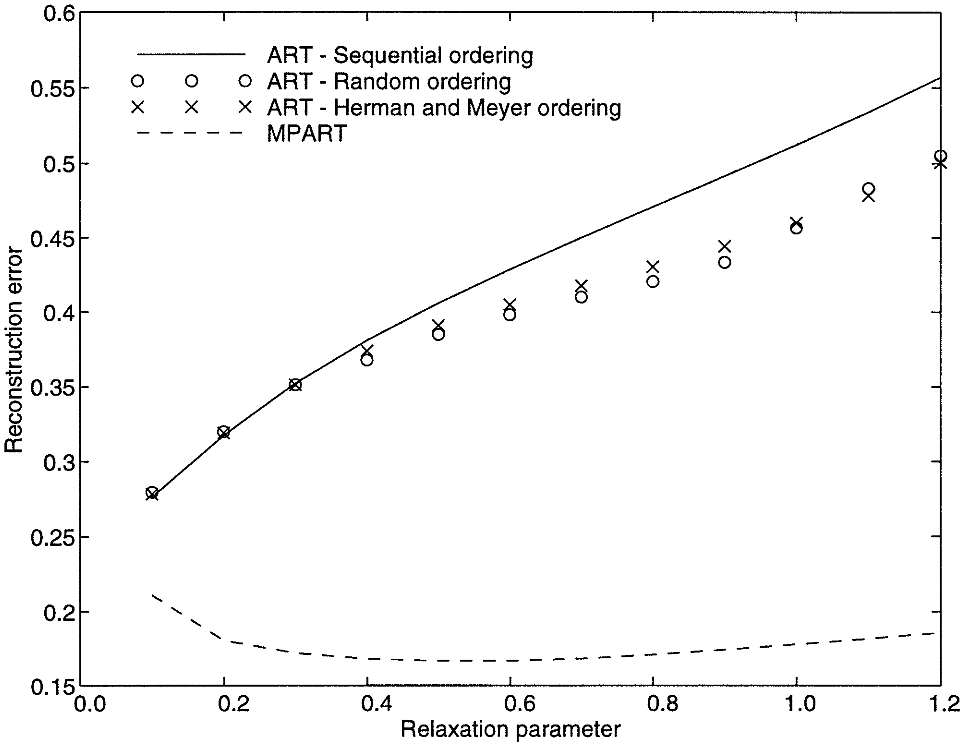


Figure 4-9: The reconstruction error values for ART and MPART for the 64 × 64 phantom (c.f. Figure 4-7) after 6 iterations as a function of relaxation parameter. Thresholds are applied to the ART matrix T and the MPART matrix $\mathcal{C}_{dd}$ such that the resulting matrices have the same sparsity (1.5% non-zero elements) and hence the two reconstructions are obtained with the same computational complexity.

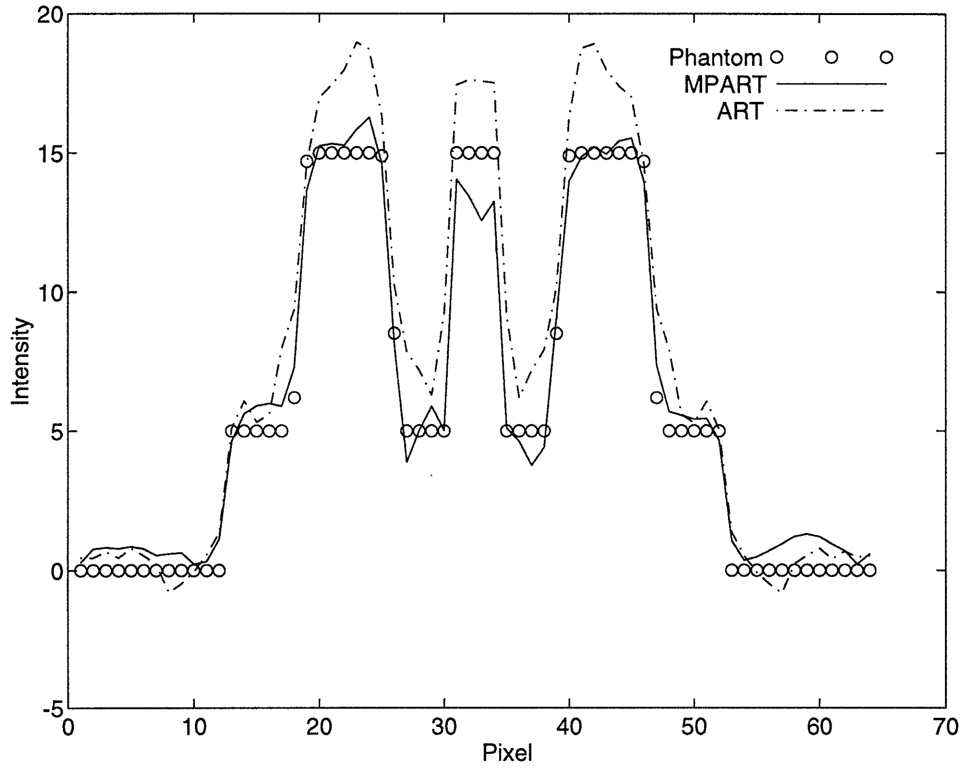


Figure 4-10: A section through the "best" approximate ART and MPART reconstructions (i.e. reconstructions obtained by using the relaxation parameter that results in least error) after 6 iterations, corresponding to the phantom in Figure 4-7.

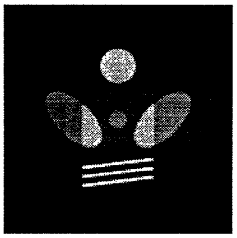


Figure 4-11: The grayscale plot (white corresponds to the maximum value and black to the minimum) of a 128×128 phantom that is used for comparing ART and MPART reconstructions. The projection data are gathered using a parallel beam geometry with $N_\theta = N_s = 128$.

Figure 4-13 we see that, as expected, the ART and MPART error values converge if the corresponding exact matrices are used for reconstruction. However, the MPART error values increase very slightly as more approximate versions of the matrix $\mathcal{C}_{dd}$ are used for reconstruction. This is in comparison to ART where the reconstruction error increases significantly with each approximation to the projection matrix T. Thus, as an example, the MPART reconstruction obtained by using 0.1% elements in $\mathcal{C}_{dd}$ has the same error as the ART reconstruction obtained by using 1% elements in T. The former, however, requires 10 times less computations and storage than the latter.

So far, for the purpose of comparison with ART, we have only presented the MPART reconstructions at the finest scale. However, as mentioned before, there exist applications where either a coarse scale reconstruction is desired, or the aim is to simply reconstruct the edges or boundaries in the object. This can be easily incorporated into the framework of MPART without any additional computations by back-projecting the object coefficients only at the desired scale(s) in (4.14). An example of this is presented in Figure 4-14 where we show the reconstruction of the 128×128 phantom of Figure 4-11 at multiple scales after 5 iterations of MPART. The value of relaxation parameter used is 0.5, and the matrix $\mathcal{C}_{dd}$ is thresholded such that it contains 0.76% non-zero elements. Note the focusing of information as the reconstructions proceed from coarse to the fine scales. In Figure 4-15 we show the finest scale detail reconstruction of the same phantom. This reconstruction represents the difference in information between the finest and next to finest scale reconstructions of Figure 4-14, and captures the edges and boundaries in the phantom.

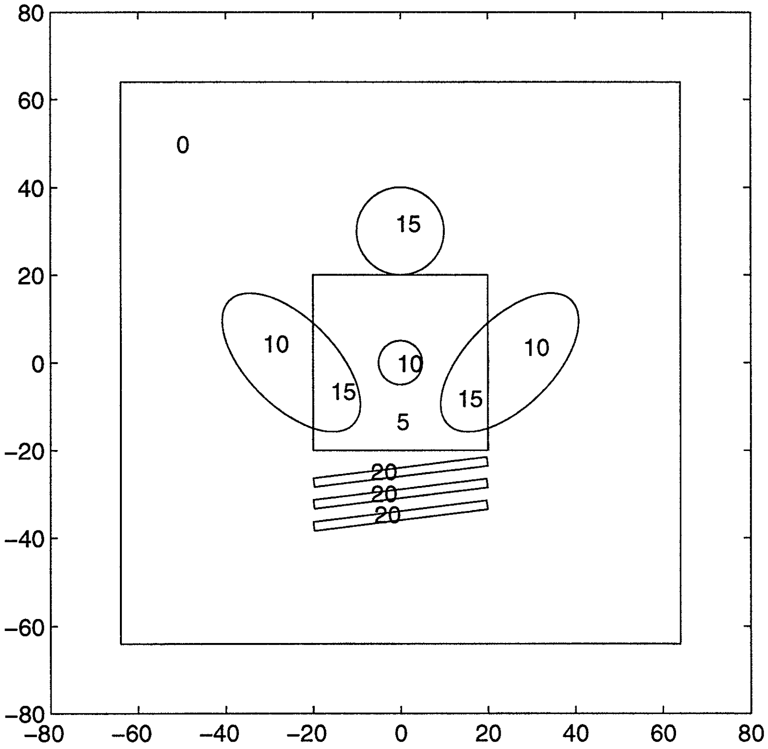


Figure 4-12: The intensity map of the 128 × 128 phantom shown in Figure 4-11. The numbers shown are intensities of respective regions. The projection data are gathered using a parallel beam geometry with $N_\theta = N_s = 128$.

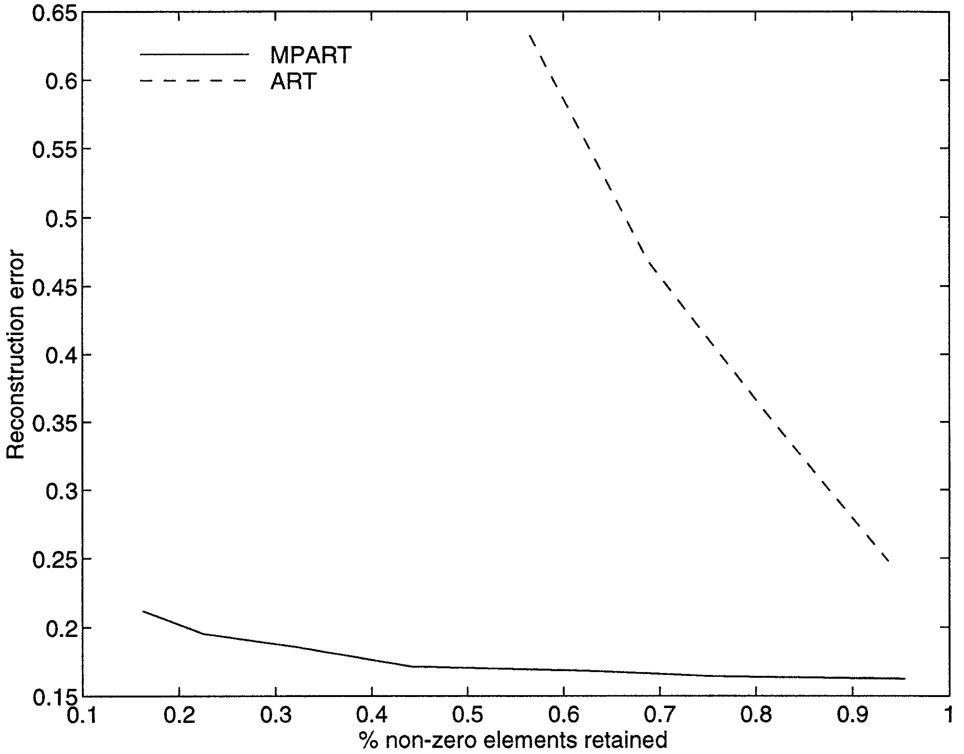


Figure 4-13: Reconstruction errors for ART (random ordering) and MPART (sequential ordering) after 5 iterations as a function of the percentage of non-zero elements in T and C_{dd}. The lower magnitude elements in these matrices are thresholded to zero. The 128 × 128 phantom shown in Figure 4-11 is used for reconstruction, and a value of 0.5 is used for the relaxation parameter.

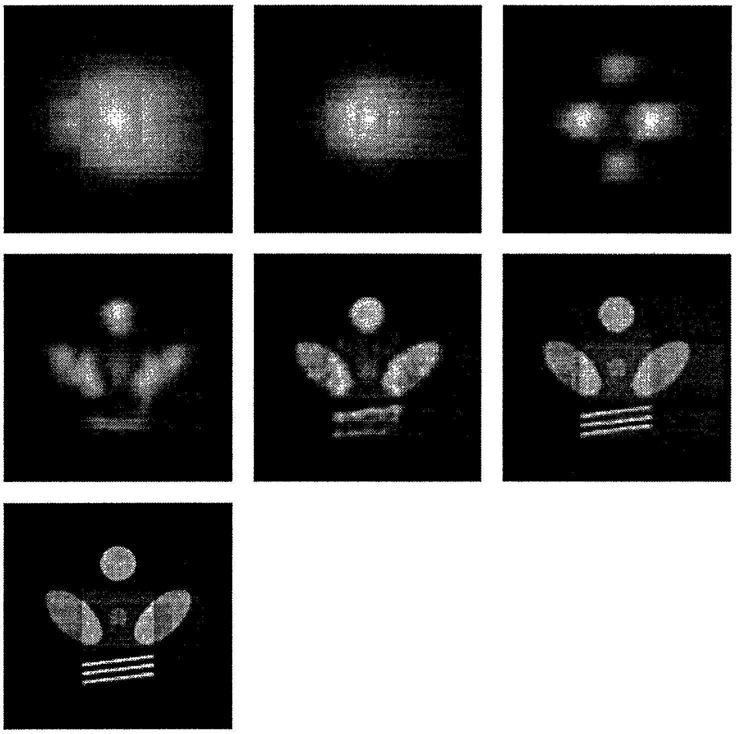


Figure 4-14: MPART reconstruction of the 128 × 128 phantom at multiple scales after 5 iterations. The value of relaxation parameter used is 0.5, and the matrix C_{dd} is thresholded to 0.76% non-zero elements. The reconstructions progress from coarse to fine scale from left to right and top to bottom. The reconstruction in the bottom row is at the finest scale.

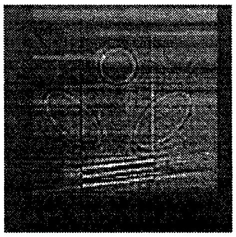


Figure 4-15: The finest scale detail reconstruction of the 128 × 128 phantom. This reconstruction represents the difference in information between the finest scale and the next to finest scale reconstructions from the previous figure.

4.4 Extension of MPART to account for noisy data

We next present an extension of MPART that computes the maximum aposteriori probability (MAP) reconstruction estimate for the transformed object coefficients ξ from noisy data, with essentially no extra computational or storage requirements. This estimate is based on a fractal-like prior model for the object which is easily specified in the wavelet transform framework. Again, we refer the reader to Chapter 3 for details about this prior model – here we deal mainly with the implementational aspects.

We model the noisy data as:

$$y(m) = \iint_{\Omega} f(u,v)\, S_m(u,v)\, du\, dv + n(m), \qquad m = 1, \ldots, N_\theta N_s, \tag{4.20}$$

where we assume the additive noise $\{n(m), m = 1, \ldots, N_\theta N_s\}$ to be Gaussian, zero-mean, uncorrelated, with variance λ. After discretization of (4.20) on a $N_s \times N_s$ pixel grid followed by wavelet transformation, and coupled with the assumption that the elements of the off-diagonal block $\mathcal{C}_{da}$ (c.f. (4.12)) are negligible, we obtain the following observation equation (the noisy version of (4.12)):

$$\left[\begin{array}{c} \eta^d \\ \hline \eta^a \end{array}\right] = \left[\begin{array}{c|c} \mathcal{C}_{dd} & 0 \\ \hline 0 & \mathcal{C}_{aa} \end{array}\right] \left[\begin{array}{c} \xi^d \\ \hline \xi^a \end{array}\right] + \left[\begin{array}{c} \nu^d \\ \hline \nu^a \end{array}\right], \tag{4.21}$$

where the vector $[(\nu^d)^T | (\nu^a)^T]^T \sim \mathcal{N}(0, \lambda I_{N_\theta N_s})$ contains the elements of the multi-scale noise vector $\nu = W_b n$. To find the MAP estimate of ξ^d and ξ^a we also need a prior statistical model for these quantities. We assume these to be distributed according to:

$$\left[\begin{array}{c} \xi^d \\ \hline \xi^a \end{array}\right] \sim \mathcal{N}\left(0, \left[\begin{array}{c|c} \Lambda_d & 0 \\ \hline 0 & \Lambda_a \end{array}\right]\right), \tag{4.22}$$

i.e. we assume ξ^d and ξ^a to be Gaussian, zero mean, independent, with variance Λ_d

and Λ_a respectively. For the variance of the approximation coefficients ξ^a, capturing the prior DC behavior of the object, we choose

$$\Lambda_a = (\frac{1}{\epsilon}) I_{N_\theta} \tag{4.23}$$

with ϵ sufficiently small (i.e. $\epsilon \to 0$) to prevent a bias in our estimate of the average (DC) behavior of the coefficients ξ, letting them be determined instead by the data. For the variance of the detail object coefficients ξ^d, as we showed in Chapter 3, a reasonable choice is to assume Λ_d to be a diagonal matrix, with the diagonal elements decreasing geometrically from coarse to fine scales. This not only imposes a classic smoothness constraint on the resulting solution $\widehat{\xi}^d$ [58], but also results in a fractal-like prior for the object f [80]. These priors have been used effectively to model natural terrain as well as biological structures such as liver and lung [11, 12, 78, 80]. Specifically, a fractal-like prior for the detail object coefficients is obtained by choosing Λ_d as (c.f. Chapter 3):

$$\Lambda_d = \sigma^2 \text{block diag} \left(\begin{array}{c|c|c|c} 2^{-\rho(J-1)} I_{2^{J-1} N_\theta} & 2^{-\rho(J-2)} I_{2^{J-2} N_\theta} & \ldots & I_{N_\theta} \end{array} \right), \tag{4.24}$$

where σ^2 is the overall magnitude, and ρ controls the texture of the resulting prior model.

The MAP estimates of ξ^d and ξ^a based on the observations (4.21), and the prior model (4.22-4.24) (with $\epsilon \to 0$), are given by (c.f. Chapter 3):

$$\widehat{\xi}^a = C_{aa}^{+} \eta^a, \tag{4.25}$$

$$\widehat{\xi}^d = \arg\min_{\xi^d} \left[\frac{1}{\lambda}(\eta^d - C_{dd}\xi^d)^T(\eta^d - C_{dd}\xi^d) + (\xi^d)^T \Lambda_d^{-1} \xi^d\right]. \tag{4.26}$$

Note that the MAP estimate for the approximation coefficients ξ^a in (4.25) results in a similar expression as we had used earlier for the case of no noise (4.17). Thus the only calculation that is now different is that of the detail coefficients, which are obtained from (4.26). We next describe an efficient method for calculation of the detail coefficients in (4.26). This implementation is based on a method due to Herman [39].

In [39] Herman describes an efficient method that calculates the MAP estimate of a vector h, given observations of the form:

$$g = Mh + q, \qquad q \sim \mathcal{N}(0, \lambda I), \tag{4.27}$$

and the following prior model:

$$h \sim \mathcal{N}(0, I). \tag{4.28}$$

Specifically, it is shown in [39] that the MAP estimate of h from (4.27) and (4.28) is the same as the minimum norm solution for h from the following equation:

$$\left[\begin{array}{c|c} I & (\sqrt{\frac{1}{\lambda}})M \end{array}\right] \left[\begin{array}{c} u \\ \hline h \end{array}\right] = (\sqrt{\frac{1}{\lambda}})g, \tag{4.29}$$

where u is a dummy variable. Thus in [39] the MAP estimate of h is obtained by using the Kaczmarz method to solve the augmented system of equations (4.29). This solution is obtained with essentially the same computational complexity and storage requirements as the original unregularized estimate. Further, the computations required per iteration of this method are again proportional to the number of non-zero elements in the matrix M, and the conditioning of M dictates whether the choice of the ordering scheme and the relaxation parameter has significant affect on convergence.

Recall that our detail observations are of the form:

$$\eta^d = \mathcal{C}_{dd}\xi^d + \nu^d, \qquad \nu^d \sim \mathcal{N}(0, \lambda I_{N_\theta(N_s-1)}), \tag{4.30}$$

and the prior model for the detail object coefficients in our case is given by:

$$\xi^d \sim \mathcal{N}(0, \Lambda_d), \tag{4.31}$$

where Λ_d is the diagonal matrix reflecting the fractal prior. Now the observation (4.30) and the prior model (4.31) in our case can be cast into the form (4.27) and

(4.28) by making the following associations:

$$\begin{aligned} M &= C_{dd}\Lambda_d^{1/2}, && (4.32)\\ h &= \Lambda_d^{-1/2}\xi^d, && (4.33)\\ g &= \eta^d. && (4.34) \end{aligned}$$

Once these associations have been made, the efficient method due to Herman can then be used to solve for the MAP estimate of h. Once this value is determined, the MAP estimate of ξ^d is calculated from (4.33). Note that since C_{dd} is sparse and Λ_d is a diagonal matrix, the pre and post-processing steps in (4.32) and (4.33) require only a very few additional computations. The benefit of our regularization approach, once again, lies in the fact that extremely sparse but accurate approximations to C_{dd} can be formed by thresholding the smaller magnitude elements to zero. The use of these approximate matrices substantially lowers the computation count per iteration. Further since C_{dd} is well-conditioned, the matrix $M = C_{dd}\Lambda_d^{1/2}$ is also well-conditioned. This translates into a fast convergence to the MAP estimate in our case, with a rate of convergence that is not significantly affected by the choice of the ordering scheme and the relaxation parameter.

Figure 4-16 shows regularized reconstructions of the 64×64 phantom in Figure 4-7 from noisy data ($N_\theta = N_s = 64$) after 5 iterations. The matrix C_{dd} is thresholded to 1.5% non-zero elements, and a value of 0.5 is used for the relaxation parameter. The noise-less strip-integral data are first generated from the phantom and are then corrupted through the addition of Gaussian, zero-mean, uncorrelated noise such that the signal-to-noise (SNR) ratio of the resulting noisy data is 5 dB. We calculate the SNR according to the formula:

$$\mathrm{SNR(dB)} = 10 \log \frac{||Tf||^2}{\lambda_n N_\theta N_s}, \qquad (4.35)$$

where Tf is the noise-free data, $N_\theta N_s$ is the length of the data vector, and λ_n is the variance of the additive noise. For the phantom used in this example, and for

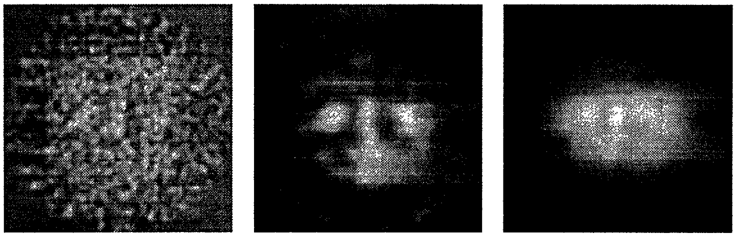


Figure 4-16: Reconstructions of the 64 × 64 phantom from 5 dB SNR data after 5 iterations. The matrix C_{dd} is thresholded to 1.5% non-zero elements, and a value of 0.5 is used for the relaxation parameter. The variance of the additive noise $\lambda_n = 8.5 \times 10^3$. Left: Unregularized MPART reconstruction. Middle: Regularized MPART reconstruction with $\lambda = 8.5 \times 10^3$, $\rho = 0.5$, and $\sigma^2 = 3.5$. Right: Regularized MPART reconstruction with $\lambda = 8.5 \times 10^3$, $\rho = 1$, and $\sigma^2 = 3.5$.

a SNR of 5 dB, $\lambda_n = 8.5 \times 10^3$. In Figure 4-16 we show the unregularized MPART reconstruction, as well as two regularized reconstructions corresponding to values of 0.5 and 1, respectively, for the texture parameter ρ in (4.24). The value of the regularization parameter σ^2, which governs the overall magnitude of the prior model (4.24), is kept constant at 3.5, and the variance of the model noise λ is chosen to be equal to $\lambda_n = 8.5 \times 10^3$ in these regularized reconstructions. We can see from the figure the ability of the regularized solution to pull out features even in the presence of substantial amount of noise. Also, as is expected, an increasing regularization can be seen to result when ρ is increased from 0.5 to 1, keeping σ^2 constant.

4.5 Discussion

In this chapter we have used the Kaczmarz method to iteratively solve the system of equations that arises in the NP-based multiscale framework. The Kaczmarz method takes advantage of both the sparsity and the well-conditioned nature of the multiscale equations and as a consequence results in an extremely efficient method for obtaining reconstructions from sparse and/or noisy data. Since the conventional algebraic reconstruction technique (ART) is also based on the Kaczmarz method, we have presented our multiscale iterative technique as an alternative to ART and we have thus referred to the latter as the multiscale projection-domain ART (MPART). In MPART, the object is expanded in a wavelet-transformed natural pixel basis and the expansion coefficients are related to the projection data through a multiscale system matrix which has approximately the same sparsity as the projection matrix used in ART. However, we are able to partition the multiscale system matrix by scales such that the Kaczmarz method operates on a relatively well-conditioned matrix. This is in contrast to ART where the Kaczmarz method operates on the system projection matrix which is quite ill-conditioned. This translates into the MPART reconstructions being relatively insensitive to the system matrix row ordering schemes and to the choice of the relaxation parameter. These factors are critical for convergence in ART and have to be carefully selected. Since the sparsity of the multiscale system matrix in MPART and the system projection matrix in ART are similar, the MPART and ART reconstructions are obtained with similar computational complexity per iteration. However, we are able to obtain approximate MPART reconstructions which come exceedingly close to the true object, by thresholding the lower magnitude elements in the multiscale system matrix to zero. These approximate MPART reconstructions are obtained with substantially less computations than the exact reconstruction. A similar approximation applied to the projection matrix results in ART reconstructions that suffer from severe artifacts. Another advantage of using MPART is that it yields reconstruction estimates at multiple scales with no added computational complexity over the full reconstruction of the object. These estimates can be used for object feature detection directly from projection data. Further, the convergence in MPART is extremely fast if only the fine scale features in the reconstruction are desired since the fine scale block of the multiscale matrix is extremely well-conditioned and at the same time is extremely sparse. Finally, we have extended our MPART to yield regularized reconstructions in the presence of noise. These regularized MPART reconstructions are obtained with essentially no additional computational or storage complexity, and are based on the fractal-like prior model for the object that we had developed in Chapter 3.

Chapter 5

An FBP-Based Multiresolution Reconstruction Technique

5.1 Introduction

In this chapter we present a multiresolution approach to the problem of reconstructing an object from *dense* tomographic projections. Recall that in Chapter 3 we had developed a NP-based multiresolution framework for reconstruction of objects from sparse data. This framework had enabled us to obtain multiresolution reconstructions from sparse data with substantially less computations than the original NP method. If the projection data are dense, however, additional efficiency may be gained by adapting our multiresolution framework from Chapter 3 to the filtered back-projection (FBP) reconstruction technique, the latter being the most commonly used method for image reconstruction from dense tomographic data (c.f. Section 2.2.3).

The FBP and the NP methods share the same object representation, i.e. the object is expanded in the same strips along which the projection data are collected. The only difference between the two methods is how the expansion coefficients are calculated from the projection data. In the NP method this calculation is complicated by the fact that the coefficients at a particular angle not only depend on the data at that angle, but also on the data at all other angles. In the FBP method, however, the dense nature of the data results in a simplified expression for the calculation of these coefficients. Specifically the coefficients at a particular angle are obtained by filtering the projections at the *same angle* with a ramp filter. Since the object representation is the same in the NP and the FBP, our NP-based multiscale object representation from Chapter 3 is still valid in the FBP case. However, in the FBP framework, we use a multiscale transformation of the ramp-filtering operation to obtain the object coefficients of the multiscale representation. As it turns out, the matrix representation of the resulting multiscale filtering operator is approximately diagonal. This enables us to formulate an efficient multiscale tomographic reconstruction technique that has the same computational complexity as that of the FBP reconstruction method. Perhaps more significantly, however, the different scale components of our proposed multiscale reconstruction method induce a corresponding multiscale representation of the underlying object and, in particular, provide estimates of (and thus information about) the field or object at a variety of resolutions at *no additional cost.* This provides a natural framework for explicitly assessing the *resolution-accuracy tradeoff* which is critical in the case of noisy data.

Noisy imaging problems arise in a variety of contexts (c.f. Chapter 1) and in such cases standard techniques such as FBP often yield unacceptable results. These situations generally reflect the fact that more degrees of freedom are being sought than are really supported by the data and hence some form of regularization is required. Conventionally, as we had mentioned in Section 2.4, the problem of reconstruction from noisy projection data is regularized by one of the following two techniques. First, the FBP ramp filter may be rolled off at high frequencies thus attenuating high frequency noise at the expense of not reconstructing the fine scale features in the object [3, 50]. This results in a fast, though ad hoc, method for regularization. The other common method for regularization is to solve for a maximum aposteriori probability (MAP) estimate of the object based on a 2-D (spatial) Markov random field (MRF) prior model [34, 69] in the original object domain. This results in a statistically regularized reconstruction which allows the inclusion of prior knowledge in a systematic way, but leads to optimization problems which are extremely computationally intensive. In contrast to these methods, we are able to extend our multiscale reconstruction technique in the case of noisy projections to obtain a multiscale MAP object estimate which, while retaining all of the advantages of statistically-based approaches, is obtained with the *same* computational complexity as the FBP reconstruction. This MAP estimate is based on the same self-similar prior model for the multiscale object coefficients as the one we had used in the NP case. Recall that such a self-similar model is easily captured in the multiscale framework by a diagonal covariance matrix for the multiscale object coefficients, and that it is this diagonal structure that is responsible for the efficiency of the MAP estimate in the NP case. Similarly, in the FBP case, the use of the above self-similar prior statistical model also leads to efficient multiscale MAP estimates.

The chapter is organized as follows. In Section 5.2 we review the relevant results from the FBP reconstruction technique. In Section 5.3 we develop the theory behind our wavelet-based multiscale reconstruction method starting from the FBP object representation. In Section 5.4 we build on this framework to provide a fast method for obtaining MAP regularized reconstructions from noisy data. The conclusions are presented in Section 5.5. Appendix 5.A contains certain technical details.

5.2 Review of the filtered back-projection (FBP) reconstruction results

In this section we briefly review the relevant FBP results from Chapter 2. First, the strip-integral data in tomography are of the form:

$$y(m) = \iint_{\Omega} f(u,v)\, S_m(u,v)\, du\, dv, \qquad m = 1, \ldots, N_\theta N_s, \tag{5.1}$$

where f is the object, $S_m(u,v)$ is the indicator function of the m-th strip, N_θ is the number of angular positions, and N_s is the number of strips in each angular position. The strip-integral data corresponding to same angular position are collectively referred to as a *projection*. The projection y_k at angle k is thus defined as:

$$y_k \triangleq \begin{bmatrix} y((k-1)N_s+1) \\ y((k-1)N_s+2) \\ \vdots \\ y((k-1)N_s+N_s) \end{bmatrix}. \tag{5.2}$$

The discretization of the observation equation (5.1) on a $N_s \times N_s$ pixel grid results in the following expression for the projection y_k (c.f. (5.2)):

$$y_k = T_k f, \tag{5.3}$$

where the rows of the matrix T_k contain discretized strips at angle k.

The FBP method uses the same object representation as the NP method. In particular, the object is expanded in the same strips S along which the projection data y are collected:

$$f(u,v) = \sum_{n=1}^{N_\theta N_s} x(n) S_n(u,v). \tag{5.4}$$

The discretization of (5.4) on a $N_s \times N_s$ pixel grid results in the following expansion coefficients (i.e. pixel values) for the object:

$$f = \sum_{k=1}^{N_\theta} T_k^T x_k, \tag{5.5}$$

where x_k are the object coefficients at angle k.

In the FBP method, which yields exact reconstructions only from noise-free dense data, the object coefficients x_k at each angle k are obtained by ramp-filtering of the projection y_k at the same angle. Thus:

$$x_k = R y_k, \tag{5.6}$$

where R is the matrix representation of the ramp-filtering operation. Once the expansion coefficients $\{x_k\}$ have been calculated, the reconstruction f is obtained through the back-projection specified in (5.5). Finally, in case of noisy projection data, regularized FBP reconstructions are obtained by rolling-off the ramp filter in R at high frequencies in an empirical manner.

5.3 The Multiscale Reconstruction Technique

In this section we derive our FBP-based multiscale reconstruction technique. As was done for the NP case in Chapter 3, we start by applying the wavelet-derived multiscale change of basis W to the discretized strips T_k in (5.5), which induces a natural multiresolution object representation. Since the FBP method shares the same object representation (5.5) as the NP method, this multiresolution representation is the same as the one we had obtained for the NP case. The difference between the two multiscale techniques, however, lies in the calculation of the coefficients of the multiresolution representation. Recall that in the NP framework this calculation required solving the large system of equations (2.11) where the coefficients at a particular angle were not only related to the projection data at that angle, but also to the data at all other angles. In the FBP framework we exploit the simplified structure of (5.6) to obtain the multiresolution coefficients at a particular angle from corresponding multiscale versions of the data at the *same* angle. Taken together these two components define a multiscale reconstruction algorithm, analogous in structure to the FBP method. In contrast to conventional multiresolution tomography approaches, which start with a decomposition of the object in a 2-D wavelet basis and attempt to then find the resulting 2-D coefficients, our method works directly in the projection domain. The multiscale nature of our object representation in the 2-D or spatial domain arises naturally from the original FBP definitions and our multiscale decomposition of the back-projection functions T_k, and thus we retain the simplicity and efficiency of the popular FBP method.

5.3.1 Multiscale Object Representation

We start by applying a multiscale change of basis, as defined by the matrix W in Section 2.5, to the original set of object coefficients x_k at each angle k to obtain an equivalent set of multiscale object coefficients as follows:

$$\xi_k = W x_k. \tag{5.7}$$

Thus, for a given choice of wavelet defining W, the vector ξ_k contains the corresponding wavelet coefficients and coarsest level approximation (i.e. the average) associated with x_k and thus forms a multiresolution representation of this signal. More importantly, by reflecting this change of basis into the original object representation (5.5), we naturally induce a corresponding *multiscale representation of the object* through the creation of a corresponding set of transformed multiscale basis functions. In particular, substituting (5.7) into (5.5) we obtain:

$$f = \sum_{k=1}^{N_\theta} (T_k^T W^T)(W x_k) \triangleq \sum_{k=1}^{N_\theta} \mathcal{T}_k^T \xi_k, \tag{5.8}$$

where $\mathcal{T}_k = W T_k$ is now a matrix representing the transformed, multiscale basis functions at angle k. Recall that (5.8) is the same multiresolution representation that we had obtained for the NP case (c.f. (3.17)). Also, recall that the rows of the transformed matrix $\mathcal{T}_k$ contain the corresponding (discretized) multiscale object basis functions at angle k. This transformation of the basis functions is shown schematically in Figure 3-2 (which corresponds to the case of the Haar wavelet). The original strip basis functions (rows of T_k) are illustrated in the top half of the figure, while the corresponding collection of multiscale basis functions (rows of $\mathcal{T}_k$) are shown in the bottom half. We may naturally group the multiscale 2-D spatial basis elements into a hierarchy of scale related components based on their support extent or spatial localization, as shown in the figure. The basis elements defining the m-th scale in such a group are obtained from the rows of $\mathcal{T}_k$ corresponding to (i.e. scaled by) the associated wavelet coefficients $\xi_k^{(m)}$ at that scale. We can see that the basis functions of these different scale components, though arising from a *1-D* multiscale decomposition, naturally represent behavior of the *2-D* object at different resolutions, directly corresponding to the different scale components contained in the transformed vector ξ_k. In particular, in defining the overall object f, the multiscale basis functions at scale m and angle k are weighted by the corresponding detail component $\xi_k^{(m)}$. The overall object is then represented by a superposition of such components at *all* angles k, as captured in the sum in (5.8).

So far we have simply transformed the representation of the original finest scale object estimate f. But, as in the NP case, a multiscale representation of the object can be obtained through the multiscale representation ξ_k and corresponding basis functions $\mathcal{T}_k$. In particular, we *define* the j-th scale approximation $f^{(j)}$ to f as:

$$f^{(j)} \triangleq \sum_{k=1}^{N_\theta} \mathcal{T}_k^T \left(A(j)\,\xi_k\right), \tag{5.9}$$

where recall that the j-th scale approximation $(A(j)\,\xi_k)$ is obtained by zeroing the finer scale components in the vector of 1-D wavelet transform coefficients of ξ_k, as discussed in Section 2.5. Similarly, by $\Delta f^{(j)}$ we denote the additional detail required to go from the object approximation at scale j to that at scale $(j+1)$, which is given by:

$$\Delta f^{(j)} \triangleq \sum_{k=1}^{N_\theta} \mathcal{T}_k^T \left(D(j)\,\xi_k\right), \tag{5.10}$$

where recall that the detail vector $(D(j)\,\xi_k)$ is obtained by zeroing all but the corresponding level of detail $\xi_k^{(j)}$ in ξ_k. Combining (5.9) and (5.10) with the scale recursive relationship (2.53) we obtain the following scale recursive relationship for the object, whereby the object approximation at the next finer scale is obtained from the approximation at the current (coarser) scale through the addition of the incremental detail at this scale:

$$f^{(j+1)} = f^{(j)} + \Delta f^{(j)}. \tag{5.11}$$

5.3.2 Multiscale Coefficient Determination

We now have a natural multiscale object representation framework through (5.9), (5.10), and (5.11) that is similar in spirit to the FBP case (5.5). To complete the process and create multiscale object estimates from data we must find the multiscale object coefficients ξ_k (which contain all the information we need). Further we desire to find these object coefficients directly from corresponding multiscale tomographic observations. Aside from simply being an evocative notion (e.g. directly relating scale-specific data features to corresponding scale-specific object characteristics), such an approach should be more efficient, in that we would expect coarse scale object characteristics to be most strongly affected by coarse or aggregate data behavior and, conversely, fine scale object characteristics to depend most strongly on fine scale data behavior. Said another way, we would expect the relationship between such multiscale data and object elements to be nearly diagonal, and this is indeed the case.

To the above ends, we perform a wavelet-based multiscale change of basis to the data sequences y_k, similar to object oriented one in (5.8), to obtain an equivalent set of multiscale observations:

$$\eta_k \triangleq W y_k, \tag{5.12}$$

where, recall, W is a matrix taking a discrete sequence to its wavelet transform. We may now easily obtain our desired direct relationship between the multiscale representation of the data at angle k in η_k and the multiscale object coefficients ξ_k at the same angle by combining the two transformations (5.7) and (5.12) together with the original FBP relation (5.6) to obtain:

$$\xi_k = \mathcal{R} \eta_k, \tag{5.13}$$

where $\mathcal{R} = WRW^T$ is the multiscale data filter, corresponding to the ramp filter R of the usual FBP case. As we show through examples later, the operator R is compressed by the wavelet operator so that $\mathcal{R}$ is nearly diagonal. Further, higher compression is achieved if Daubechies wavelets D_n with larger n are used. This observation is consistent with the observations of Beylkin et al [6], since R is a pseudo-differential operator.

5.3.3 The Overall Multiscale Algorithm

We are now in a position to present our overall multiscale reconstruction method. By comparing the FBP equations (5.5) and (5.6) to the corresponding multiscale equations (5.8) and (5.13), respectively, we see that our complete multiscale reconstruction process for estimation of f parallels that of the standard FBP reconstruction, in that identical and independent processing is performed on the multiscale data sets η_k at each angle to obtain the corresponding multiscale object coefficients ξ_k at that angle, which are then back-projected along corresponding multiscale basis functions $\mathcal{T}_k$ and combined to obtain the final object estimate. Thus our overall procedure, given next, is *no more complex than the standard FBP method.*

Algorithm 3 (Multiscale Reconstruction)

1. *For a given choice of wavelet, form the multiscale filter matrix* $\mathcal{R} = W R W^T$ *(the multiscale counterpart of the original ramp filter) to process the data at each angle.* $\mathcal{R}$ *is nearly diagonal.*
2. *For each angle* k *perform the following:*
   (a) *Find the multiscale observations* η_k *by taking the 1-D wavelet transform of the projection data at angle* k, $\eta_k = W y_k$.
   (b) *Calculate the multiscale object coefficient set* $\xi_k = \mathcal{R} \eta_k$ *by filtering the multiscale observations.*
   (c) *Back-project* ξ_k *along the corresponding multiscale basis functions* $\mathcal{T}_k$, $\mathcal{T}_k^T \xi_k$.
3. *Combine the object contributions from the individual back-projections at each angle to obtain the overall estimate,* $\sum_k \mathcal{T}_k^T \xi_k$.

Again as in the NP case, beyond simply finding a finest scale object estimate as described in Algorithm 3, however, we could also reconstruct the underlying object at

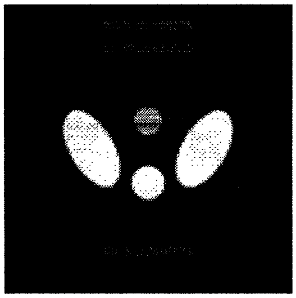


Figure 5-1: Phantom used for reconstruction experiments. The phantom is 256×256 and projections are gathered at 256 equally spaced angles ($N_\theta = 256$) with 256 strips per angle ($N_s = 256$).

multiple resolutions through (5.9), (5.10) and (5.11). In particular, if an approximation $f^{(j)}$ at scale j is desired, then in Algorithm 3 we need only replace ξ_k by $(A(j)\,\xi_k)$ in Step 2c and 3. In particular, this simply amounts to zeroing detail components in ξ_k which are finer than scale j. Further, if instead we want to reconstruct the detail $\Delta f^{(j)}$ added at a particular scale, we need only replace ξ_k by $(D(j)\,\xi_k)$ in Step 2c and 3 of Algorithm 3. Similarly, this simply amounts to zeroing all but the desired scale of detail $\xi_k^{(j)}$ in ξ_k. Note that such intermediate scale information about f can even be efficiently found by calculating *only* those elements necessary for reconstructing the scale of interest – i.e. all of ξ_k is not required. For example, if all that is required is a coarse estimate of the object and not the full reconstruction, only the coarsest elements of $\xi_k^{(j)}$ are required. Conversely if only fine scale features are to be reconstructed, then only the finest scale detail components of $\xi_k^{(j)}$ are needed.

5.3.4 Examples

We now show some examples of our multiscale reconstruction framework. Figures 5-1 and 5-2 show the 256×256 phantom used in the experiments of this section. Projection data were collected at 256 equally spaced angles ($N_\theta = 256$) with 256 strips used for each projection ($N_s = 256$).

First we show a series of approximate reconstructions using the Daubechies D_3 wavelet for the multiscale decomposition W. Figure 5-3 shows the various scale approximate object reconstructions $f^{(j)}$ for the entire range of scales $j = 1, \ldots, 8$.

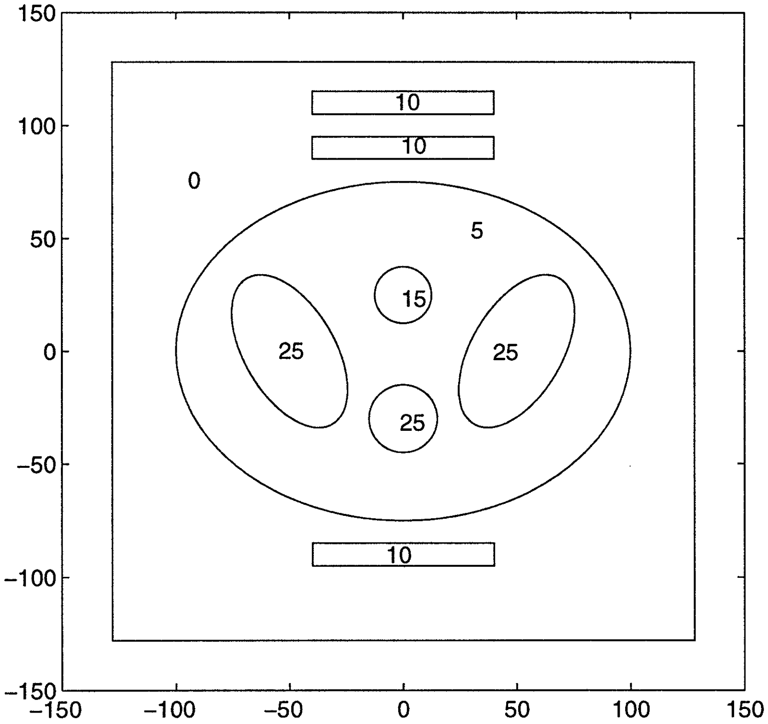


Figure 5-2: The intensity map of the 256 × 256 phantom shown in Figure 5-1. The numbers shown are intensities of respective regions. The projection data are gathered using a parallel beam geometry with $N_\theta = N_s = 256$.

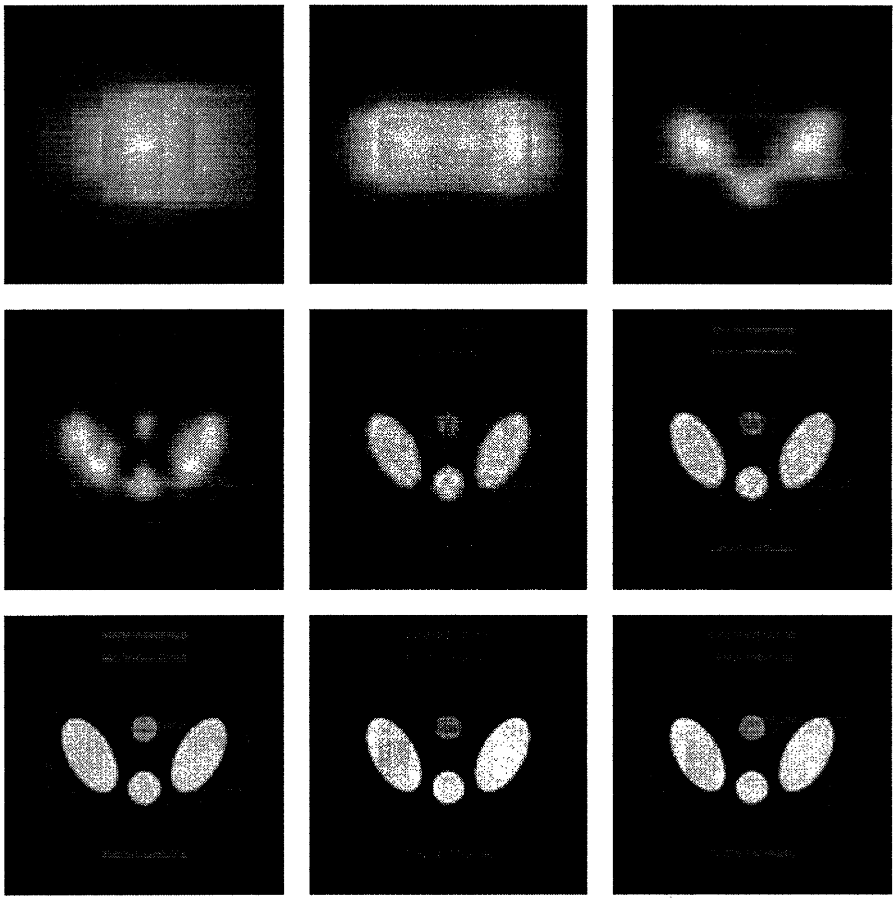


Figure 5-3: Approximation reconstructions of phantom of Figure 5-1 at various scales, using D_3 wavelet. First row, left: $f^{(1)}$. First row, middle: $f^{(2)}$. First row, right: $f^{(3)}$. Second row, left: $f^{(4)}$. Second row, middle: $f^{(5)}$. Second row, right: $f^{(6)}$. Third row, left: $f^{(7)}$. Third row, middle: $f^{(8)}$. The third row, right shows the corresponding FBP reconstruction f for comparison. The FBP reconstruction is the same as $f^{(8)}$, since this is the complete reconstruction.

The approximations get finer from left to right and top to bottom (so that the upper left frame is $f^{(1)}$ and the bottom middle frame corresponds to $f^{(8)}$). The bottom row, right shows the FBP reconstruction for comparison. Note in particular, that the finest scale approximation $f^{(8)}$ is identical to the FBP estimate f. The intermediate scale estimates demonstrate how information is gathered at different scales. For example, in the scale 3 reconstruction $f^{(3)}$ (top right in the figure) though only 8 of the full 256 coefficient elements in the vectors ξ_k are being used, we can already distinguish separate objects. By scale 4 (middle row, left) we can start to identify the separate bright regions within the central larger object, while by scale 5 this information is

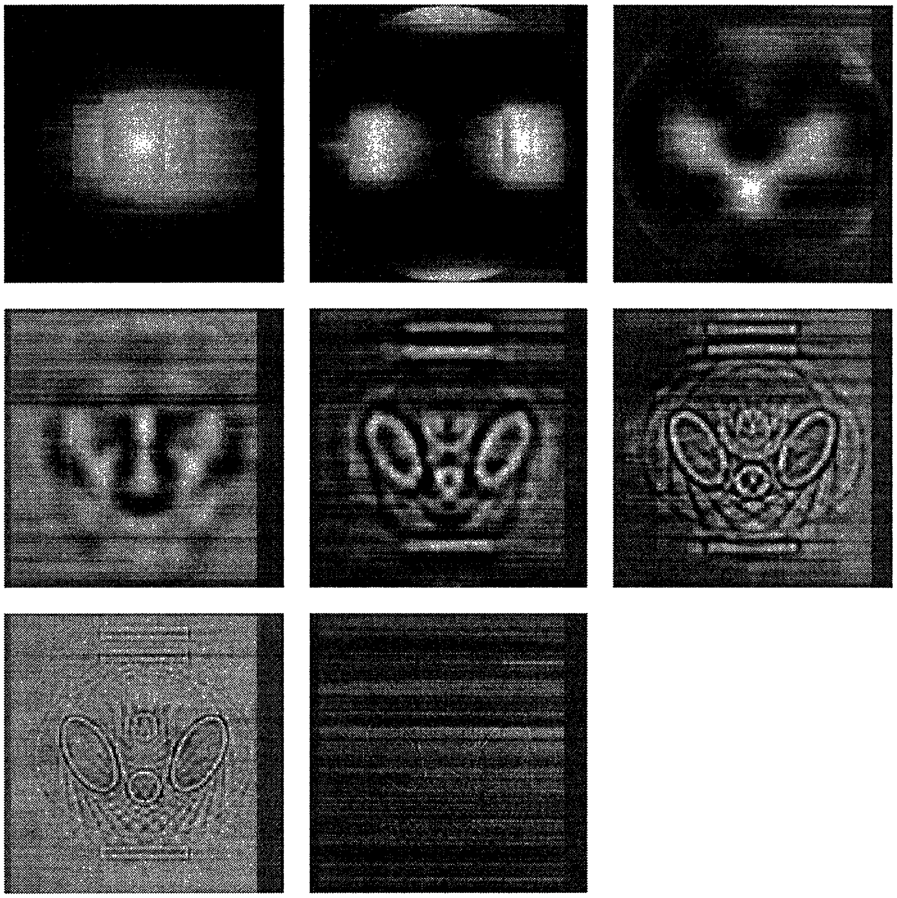


Figure 5-4: The detail added between successive scales in the reconstructions of Figure 5-3. First row, left: $\Delta f(0)$. First row, middle: $\Delta f(1)$. First row, right: $\Delta f(2)$. Second row, left: $\Delta f(3)$. Second row, middle: $\Delta f(4)$. Second row, right: $\Delta f(5)$. Third row, left: $\Delta f(6)$. Third row, middle: $\Delta f(7)$.

well localized. Even at this comparatively fine scale we are still only using about 12% of the full object coefficient set.

In Figure 5-4 we show the corresponding detail components $\Delta f^{(m)}$ for the same phantom. Again, the additive detail becomes finer going from left to right and top to bottom in the figure. Notice that the fine scale, edge based, features of the phantom are clearly visible in the $\Delta f^{(4)}$ and $\Delta f^{(5)}$ reconstructions (center row, middle and right in the figure), showing that structural information can be obtained from these detail images alone. Recall that these images provide the added information needed in going from the object approximation at one scale to that at the next finer scale (as provided in Figure 5-3).

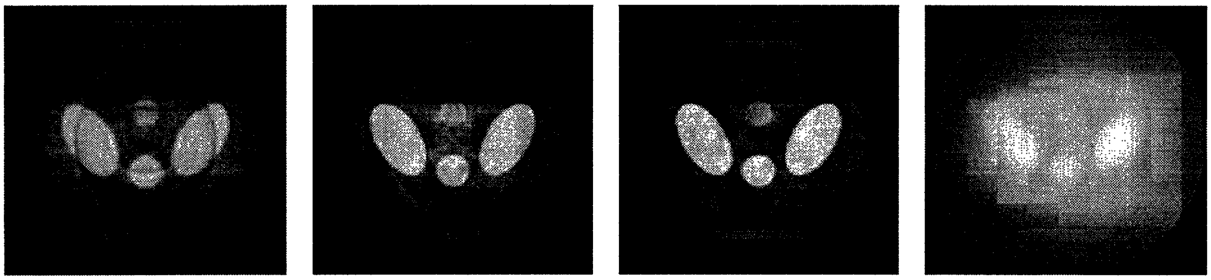


Figure 5-5: Complete finest scale multiscale reconstructions for phantom of Figure 5-1 for different approximate filtering operators. The left three frames show approximate multiscale reconstructions using only the diagonal elements of $\mathcal{R}$ corresponding to different choices of the underlying wavelet: First column = Haar. Second column = D_3. Third column = D_8. For comparison, the right-most frame shows an equivalent approximate FBP reconstruction using only the diagonal elements of R, demonstrating the superiority of the approximations based in the multiscale domain.

As we discussed earlier, the wavelet-based multiscale transformation of both the representation x_k and data y_k also serves to compress the ramp filter matrix R so that the corresponding multiscale filter matrix $\mathcal{R}$ is nearly diagonal. As we argued earlier, this reflects the fact that coarse scale object characteristics are most strongly affected by coarse or aggregate data behavior and, conversely, fine scale object characteristics tend to depend most strongly on fine scale data behavior. One consequence is that a very good approximation to the exact reconstruction procedure of Algorithm 3 can be achieved by ignoring the off-diagonal terms of $\mathcal{R}$ in (5.13). These off-diagonal terms capture both intra- and inter-scale couplings. Further, this approximation to the exact reconstruction becomes better as Daubechies wavelets D_n with larger n are used. To illustrate this point, in Figure 5-5 we show complete (finest scale) reconstructions f of the same phantom as before, based on the same projection data but using a diagonal approximation to $\mathcal{R}$ in (5.13) and Algorithm 3 for a variety of choices of the wavelet defining W. For the reconstructions we use only the diagonal elements of $\mathcal{R}$ (which account for 0.0031% of all the elements for this case) in the calculation of ξ, effectively setting all off-diagonal elements to zero. Reconstructions corresponding to Daubechies wavelets D_n with increasing n (in particular Haar or D_1, D_3, and D_8) are shown from left to right in the figure. It can be seen from the improvement in the reconstructions that the accuracy of the diagonal approximation becomes better as D_n wavelets with increasing n are used in the definition of W. In particular, the approximations can be seen to compare very favorably with the standard FBP reconstruction. For comparison we also show on the far right in Figure 5-5 a corresponding approximate FBP reconstruction obtained using a diagonal approximation to the original ramp-filter matrix R for reconstruction. It can be seen that a diagonal approximation in the multiscale domain results in far better reconstructions than a similar approximation in the original domain, indicating that the multiscale transformation of data and coefficients has served to decouple the resultant quantities.

In summary, we have formulated a 2-D multiscale object reconstruction method in terms of approximation and detail images. This method is derived from the classical FBP method and thus is well matched to reconstruction from projection data. The associated 2-D multiresolution object representation is induced by a 1-D wavelet-based change of basis to the original FBP projection space object coefficients. While the resulting representations are similar in spirit to a direct 2-D multiresolution decomposition of the original object, in that approximations are produced at a series of scales along with the detail necessary to proceed from one such approximation to the next finer one, our approach does not correspond to such a direct orthonormal decomposition. As a result it is fundamentally different from previous multiscale-related work in tomography (for example, [64]). In these approaches such a direct 2-D expansion of the object (i.e. a 2-D wavelet transform) is used to directly define the approximation and detail images, the coefficients of which are then calculated from the projection data. In contrast, all of our multiscale quantities inherently "live" in the projection domain. As a result, our representation is closer to the measurement domain than previous multiscale representations, and in particular implies that our approach is no more computationally complex than FBP. To this point we have focused on noiseless reconstructions. Next, we build on our multiscale reconstruction method to obtain a fast method for computing regularized reconstructions from *noisy* projections.

5.4 Multiscale Regularized Reconstructions

In this section we consider the estimation of an object f from complete but *noisy* projection observations. We extend our multiscale reconstruction method presented in Section 5.3 to obtain statistically regularized estimates which may be simply and efficiently computed, in particular, with no more effort than is required for the standard FBP reconstruction. This regularized solution is obtained by first solving for the maximum aposteriori probability (MAP) estimate [76] of the multiscale object coefficients, $\hat{\xi}_k$, corresponding to the same multiscale, self-similar prior model we had used in the NP case (c.f. Section 3.4), and then back-projecting these multiscale coefficient estimates along the corresponding multiscale basis functions.

The presence of noise in projection data leads to reconstructions by the FBP method that are unacceptable and thus require some form of regularization. The traditional approach has been to simply roll-off the ramp filter used in the standard FBP reconstruction at high frequencies. This is called apodization [3] and several different windows are typically used for this purpose, for example Hanning, Hamming, Parzen, Butterworth etc. [50] (c.f. Section 2.4.1). The assumption is that most the object energy occurs at low frequencies while the noise-derived artifacts occur at high frequency. The high frequency roll-off thus attenuates these components at the expense of not reconstructing the fine scale features in the object. The result is a fast, though ad hoc, method for regularization. The other traditional approach to regularizing the noisy data problem is statistically based. This method starts with a statistical model for the noisy observations based on (5.3):

$$y_k = T_k f + n_k, \tag{5.14}$$

where n_k is taken as an additive noise vector at angle k. This observation model is then coupled with a 2-D Markov random field (MRF) prior model [34,69] for f to yield a direct MAP estimate of the object $\hat{f}$ (c.f. Section 2.4.2). While statistically based, thus allowing the systematic inclusion of prior information, the 2-D spatially-local MRF prior models used for the object generally lead to optimization problems that are *extremely* computationally complex. As a result, these methods have traditionally not found favor in practical applications.

In contrast to the above two techniques, we will develop a multiscale MAP object estimate that, while retaining all of the advantages of statistically-based approaches, is obtained with the same computational complexity as the FBP reconstruction. To accomplish this, we continue to work in the projection domain, as the FBP method does, and build our statistical models there, rather than in the original object domain. As in Section 5.3, we then allow the resulting projection domain coefficients to induce a 2-D object representation through the back-projection and summation operations. To this end we start with an observation equation relating the noisy data y_k to the *FBP object coefficients* x_k, rather than the corresponding 2-D object f as is done in (5.14). Such a relationship may be found in the FBP relationship (5.6), which in the presence of noise in the data becomes:

$$y_k = R^{-1}x_k + n_k, \qquad n_k \sim \mathcal{N}(0, \Lambda_{n_k}), \tag{5.15}$$

where, recall R is the FBP ramp filter operator[1], the notation $z \sim \mathcal{N}(m, \Lambda)$ denotes a Gaussian distribution of mean m and covariance Λ. In particular, we assume $\Lambda_{n_k} = \lambda_k I_{N_s}$, i.e. that the noise is uncorrelated from strip to strip but may have different noise covariances at different angles, capturing the possibility that the data at different projections may be of differing quality (e.g. due to different sensors or imaging configurations). Further, we assume that the noise is uncorrelated from angle to angle, so that n_k is independent of n_j, $k \neq j$. This model of independent noise in the projection domain is well justified for most tomographic applications.

As in Section 5.3, for purposes of estimation we desire a relationship between multiscale representations of the data, object coefficients, and noise. Working in the multiscale transform domain will again allow us to obtain induced multiresolution

---

[1] Note that (5.15) assumes that R^{-1} exists. For the case where R represents an ideal ramp filter this will indeed *not* be the case, as this operator nulls out the DC component of a signal. For filters used in practice, however, this inverse *does* exist and the expression given in (5.15), based on such a filter is well defined. Details may be found in Appendix 5.A.

estimates of the object. Such a multiscale oriented relationship between the quantities of interest can be found by combining (5.15) with the multiresolution changes of bases (5.7) and (5.12) based on W (defined in Section 2.5) to obtain:

$$\eta_k = \mathcal{R}^{-1}\xi_k + \nu_k, \qquad \nu_k \sim \mathcal{N}(0, \Lambda_{\nu_k}), \tag{5.16}$$

where $\nu_k = Wn_k$ is the multiscale transformed noise vector at angle k with $\Lambda_{\nu_k} = W\Lambda_{n_k}W^T = \lambda_k I_{N_s}$ as its corresponding covariance. This equation relates our observed noisy multiscale data η_k to our desired multiscale object coefficients ξ_k through the multiscale filtering operator $\mathcal{R}$. Note that the assumption of uncorrelated noise from angle to angle and strip to strip in the original projection domain results in uncorrelated noise from angle to angle and multiscale strip to multiscale strip in the multiscale domain, since W is an orthonormal transformation.

5.4.1 The Multiscale Prior Model

To create a MAP estimate of the multiscale object coefficients ξ_k, we will combine the observation equation (5.16) with a prior statistical model for the desired unknown multiscale coefficient vectors ξ_k. Multiresolution *object* estimates and the detail between them can then be easily obtained by using the resulting MAP coefficient estimates $\widehat{\xi}_k$ at multiple scales in the multiscale object definitions (5.9) and (5.10), as was done previously in Section 5.3.

We use the same self-similar prior model for the object detail coefficients as the one we had used for the NP case in Section 3.4 which we review here. Recall that such a self-similar model can be obtained by assuming that the variance of the detail added in going from the approximation at scale j to the approximation at scale $j+1$ decreases geometrically with scale, i.e. by choosing $\xi_k^{(j)}$ (the wavelet coefficients at each scale) as independent, $\mathcal{N}(0, \sigma^2 2^{-\rho j})$ random variables [80]. The parameter ρ determines the nature, i.e. the texture, of the resulting self-similar process while σ^2 controls the overall magnitude. If $\rho = 0$ the resulting finest level representation (the elements of x_k) correspond to samples of white noise (i.e. are completely uncorrelated), while as ρ increases the components of x_k show greater long range correlation. Such self-similar models are commonly and effectively used in many application areas such as modeling of natural terrain and other textures, biological signals, geophysical and economic time series, etc. [11,12,58,78,80].

In addition to defining the scale varying probabilistic structure of the detail components of ξ_k, we also need a probabilistic model for the element of ξ_k corresponding to the coarsest scale approximation of x_k, i.e. $x_k^{(0)}$. This term describes the DC or average behavior of x_k, of which we expect to have little prior knowledge. As a result we choose this element as $\mathcal{N}(0, \overline{Q})$, where the (scalar) uncertainty $\overline{Q}$ is chosen sufficiently large to prevent a bias in our estimate of the average behavior of the coefficients, letting it be determined instead by the data.

In summary, we use a prior model for the components of the multiscale coefficient vectors ξ_k which is defined directly in scale-space and which corresponds to a self-similar, fractal-like prior model for the corresponding object coefficients x_k. In particular, this model is given by $\xi_k \sim \mathcal{N}(0, \Lambda_{\xi_k})$ with ξ_k independent from angle to angle and where:

$$\begin{aligned} \Lambda_{\xi_k} &= \text{block diag}\left[Q^{(J-1)}, \ldots, Q^{(1)}, Q^{(0)}, \overline{Q}\right], \\ Q^{(m)} &= \sigma^2 2^{-\rho m} I_{2^m}. \end{aligned} \tag{5.17}$$

Again, this model not only assumes that the sets of multiscale object coefficients, ξ_k, are *independent from angle to angle* but also that these coefficients are *independent from scale to scale*, that they are *independent and identically distributed within a given scale*, and finally that their *variance decreases geometrically proceeding from coarse to fine scales*. Obviously other choices may be made for the statistics for the multiscale object coefficients, and we discuss some particularly interesting possibilities in Chapter 8. The choice we have made in (5.17) while simple, is well adapted to many naturally occurring phenomenon. In addition recall that, since the *observation noise* power is uniform across scales or frequencies, the geometrically decreasing variance of this prior model implies that the projection data will most strongly influence the reconstruction of coarse scale features and the prior model will most strongly influence the reconstruction of fine scale features. This reflects our belief that the fine scale behavior of the object (corresponding to high frequencies) is the most likely to be corrupted by noise. Finally, our choice of prior model in (5.17) results in efficient processing algorithms for the solution of the corresponding MAP estimate, in particular with no more complexity than the standard FBP reconstruction.

5.4.2 The Multiscale MAP Estimate

We are now in a position to present our overall algorithm for computing a MAP [76] multiscale object estimate $\widehat{\xi}_k$. Since the data at each angle η_k and the corresponding prior model for ξ_k are independent from angle to angle, the MAP estimates of the vectors ξ_k decouple. In particular, the estimate of ξ_k at each angle, based on the observations (5.16) and the prior model (5.17) is given by:

$$\begin{aligned} \widehat{\xi}_k &= \left[\Lambda_{\xi_k}^{-1} + \mathcal{R}^{-T}\Lambda_{\nu_k}^{-1}\mathcal{R}^{-1}\right]^{-1} \mathcal{R}^{-T}\Lambda_{\nu_k}^{-1}\eta_k, \\ &\triangleq \overline{\mathcal{R}}\eta_k, \end{aligned} \tag{5.18}$$

where the regularized multiscale filter operator $\overline{\mathcal{R}}$ is defined in the obvious way. This regularized filtering matrix is exactly analogous to the unregularized filtering operator $\mathcal{R}$ of (5.13) for the noise free case. In this regularized case, however, $\overline{\mathcal{R}}$ now also depends on both the noise model Λ_{ν_k} and the prior object model Λ_{ξ_k}. If the noise variance is low relative to the uncertainty in the prior model (so $\Lambda_{\nu_k}^{-1}$ is large) then $\overline{\mathcal{R}}$ will approach $\mathcal{R}$ and the estimate will tend toward the standard unregularized one. Conversely, as the noise increases, $\overline{\mathcal{R}}$ will depend to a greater extent on the prior model term Λ_ξ and the solution will be more regularized or smoothed.

Finally, as in the noise-less case, the resulting *object estimate* $\widehat{f}$ is then obtained by back-projecting the estimated multiscale object coefficients $\widehat{\xi}_k$ along the corresponding multiscale basis functions $\mathcal{T}_k$ and combining the result. The overall structure of this regularized reconstruction parallels that of the original FBP method, and therefore is of the *same computational complexity as FBP*. In summary, our overall, efficient regularized multiscale estimation algorithm is given by the following procedure, which parallels our unregularized multiscale reconstruction algorithm:

Algorithm 4 (Regularized Multiscale Reconstruction)

1. *Find the regularized multiscale filter matrix* $\overline{\mathcal{R}}$ *(the multiscale regularized counterpart of the original ramp filter) by doing the following:*
   - (a) *For a given choice of wavelet, form the unregularized multiscale filter matrix* $\mathcal{R} = W\,R\,W^T$ *as before.*
   - (b) *Choose the model parameters* λ_k *specifying the variances of the observation noise processes and thus defining* Λ_{ν_k} *c.f. (5.16).*
   - (c) *Choose the multiscale prior model parameters* σ^2*,* ρ *and* $\overline{Q}$ *specifying the magnitude and texture of the model and the uncertainty in its average value, respectively, and generate the prior covariance matrix* Λ_{ξ_k} *through (5.17).*
   - (d) *Form* $\overline{\mathcal{R}} = \left[\Lambda_{\xi_k}^{-1} + \mathcal{R}^{-T}\Lambda_{\nu_k}^{-1}\mathcal{R}^{-1}\right]^{-1} \mathcal{R}^{-T}\Lambda_{\nu_k}^{-1}$.
2. *For each angle* k *perform the following:*
   - (a) *Find the multiscale observations* η_k *by taking the 1-D wavelet transform of the projection data at angle* k*,* $\eta_k = W\,y_k$.
   - (b) *Calculate the regularized multiscale object coefficient set* $\widehat{\xi}_k = \overline{\mathcal{R}}\,\eta_k$ *by filtering the multiscale observations.*
   - (c) *Back-project* $\widehat{\xi}_k$ *along the corresponding multiscale basis functions* $\mathcal{T}_k$*,* $\mathcal{T}_k^T\widehat{\xi}_k$.
3. *Combine the regularized object contributions from the individual back-projections at each angle to obtain the overall regularized object estimate,* $\sum_k \mathcal{T}_k^T\widehat{\xi}_k$.

As before, we may also easily obtain regularized reconstructions of the object at multiple resolutions by using (5.9) and (5.10) together with the MAP coefficient estimates $\widehat{\xi}_k$. In particular, to obtain the approximation $\widehat{f}^{(j)}$ at scale j then we need only replace $\widehat{\xi}_k$ by $(A(j)\,\widehat{\xi}_k)$ (corresponding to simply zeroing some of the terms in $\widehat{\xi}_k$) in Step 2c and 3. Similarly, the corresponding object detail components $\Delta\widehat{f}^{(j)}$ at scale j may be obtained by using $(D(j)\,\widehat{\xi}_k)$ in place of $\widehat{\xi}_k$ in these steps.

While Algorithm 4 is already extremely efficient, in that 2-D multiscale regularized object estimates are generated with no more complexity than is needed for the standard FBP method, additional gains may be obtained by exploiting the ability of the wavelet transform operator W to compress the FBP filtering operator R. Recall, in particular, that the (unregularized) multiscale filtering matrix $\mathcal{R} = WRW^T$ is nearly diagonal, with this approximation becoming better as Daubechies wavelets D_n with increasing n are used in the specification of W. Based on our assumptions, the matrices Λ_{ξ_k} and Λ_{ν_k}, specifying the prior model and observation covariances respectively, are already diagonal. If in addition $\mathcal{R}^{-1}$ were also a diagonal matrix, then from (5.18) we see that $\overline{\mathcal{R}}$ itself would be diagonal, with the result that the "filtering" in Step 2b of Algorithm 4 would simply become point by point scaling of the data. To this end we will assume that the wavelet transform W truly diagonalizes R by effectively ignoring the small, off-diagonal elements in $\mathcal{R}^{-1}$. That is, we assume[2]:

$$\mathcal{R}^{-1} \approx \operatorname{diag}(r_1, r_2, \ldots, r_{N_s}), \tag{5.19}$$

where r_i are the diagonal elements of $\mathcal{R}^{-1}$. Now let us represent the diagonal prior model covariance matrix as $\Lambda_{\xi_k} = \operatorname{diag}[p_1, p_2, \ldots, p_{N_s}]$, and recall that $\Lambda_{\nu_k} = \lambda_k I_{N_s}$. Using these quantities together with our approximation to $\mathcal{R}^{-1}$ in the specification of the estimate (5.18) yields an approximate expression for $\widehat{\xi}_k$:

$$\begin{aligned} \widehat{\xi}_k &\approx \operatorname{diag}\left(\frac{r_1}{r_1^2 + (\lambda_k/p_1)}, \frac{r_2}{r_2^2 + (\lambda_k/p_2)}, \ldots, \frac{r_{N_s}}{r_{N_s}^2 + (\lambda_k/p_{N_s})}\right) \eta_k, \\ &\triangleq \widetilde{\mathcal{R}} \eta_k, \end{aligned} \tag{5.20}$$

where the *approximate* MAP filtering matrix $\widetilde{\mathcal{R}}$ is defined in the obvious way. Our experience is that when W is defined using Daubechies wavelets of order 3 or higher (i.e. using D_3, D_4,...), the estimates obtained using $\widetilde{\mathcal{R}}$ in place of the exact regularized filter $\overline{\mathcal{R}}$ in Algorithm 4 are visually indistinguishable from the exact estimates where

---

[2] One can imagine another level of approximation where we set the off-diagonal elements of $\mathcal{R}$ itself to zero *prior* to inversion rather than those of $\mathcal{R}^{-1}$. This further approximation results in reconstructions which are visually very similar to what we obtain here.

$\mathcal{R}^{-1}$ is not assumed to be diagonal. Indeed, it is actually this approximate filtering operator $\widetilde{\mathcal{R}}$ that we use to generate the example reconstructions we show next.

Before proceeding, however, let us examine our MAP regularized filtering operator $\overline{\mathcal{R}}$ in more detail to understand how our multiscale MAP estimation procedure relates both to the standard FBP method and the empirical regularization obtained through apodization of the FBP filter. The MAP estimates $\widehat{\xi}_k$ induce corresponding estimates $\widehat{x}_k$ of the original object coefficients x_k through the change of basis (5.7) and, similarly, η_k and y_k are related through (5.12). Thus, the multiscale MAP estimation operation specified by (5.18) imposes a corresponding relationship between the original finest scale quantities $\widehat{x}_k$ and y_k, given by:

$$\widehat{x}_k = \left(W^T \overline{\mathcal{R}} W\right) y_k \triangleq R_{\text{eff}}\, y_k, \tag{5.21}$$

where the effective multiscale MAP regularized filtering matrix R_{eff} is defined in the obvious way. The effect of this MAP regularized filter can now be compared to the standard FBP or apodized ones. The behavior of the matrix operator R_{eff} can be most easily understood by examining its corresponding frequency domain characteristics. To this end, in Figure 5-6 we plot the magnitude of the Fourier transform of the central row of the effective regularized matrix R_{eff} corresponding to a variety of choices of the model or regularization parameters λ_k (the noise variance) and ρ (the decay rate across scales of the added detail variance) for fixed $\sigma^2 = 1$ (overall prior model amplitude) and $\overline{Q} = 1$ (prior model DC variance). We also plot, with heavy lines, the magnitude of the Fourier transform of the corresponding central row of the standard FBP ramp filter matrix R for comparison. From Figure 5-6, we can see that in the multiscale MAP framework regularization is basically achieved by rolling off the ramp filter at high frequencies, the same principle as used in the ad hoc, apodization regularized FBP reconstructions. We also see that decreasing the observation noise variance λ_k for a fixed prior model structure ρ, or conversely, increasing the variance of the detail added in proceeding from coarse to fine scales in the prior model (i.e. decreasing ρ) for a fixed observation noise variance λ_k, leads to decreased regularization as reflected

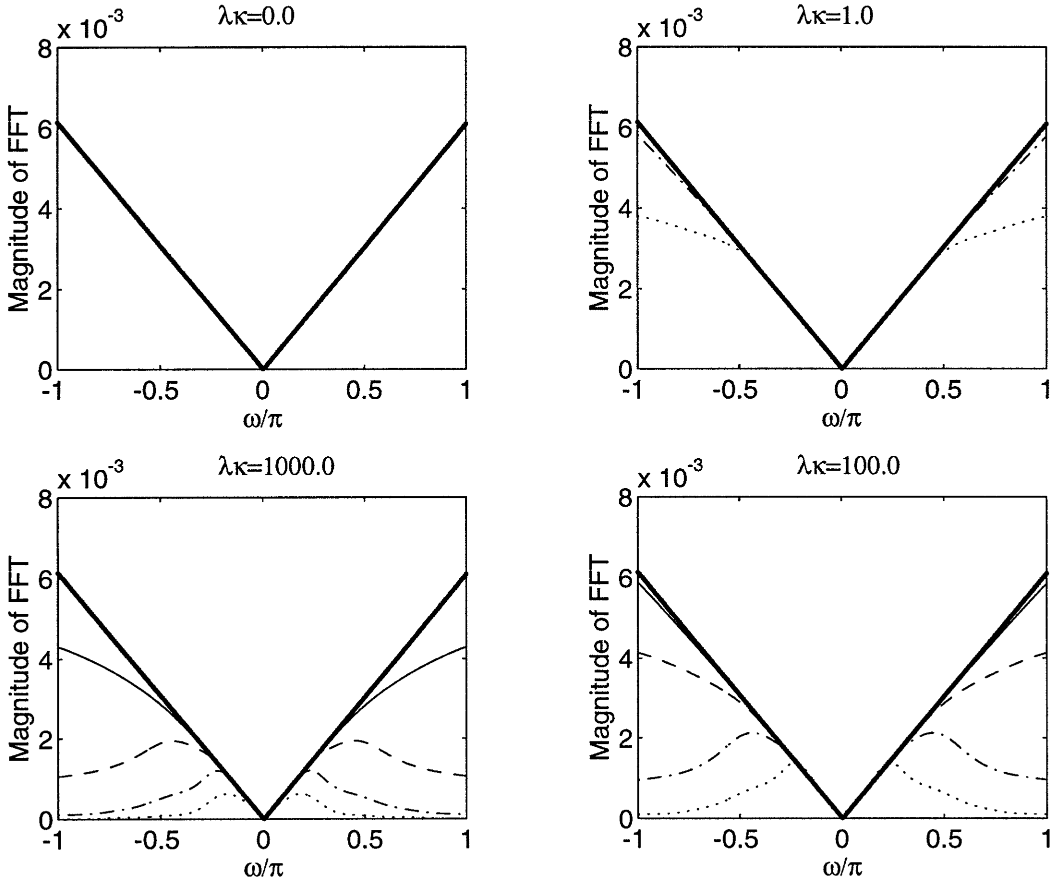


Figure 5-6: The Fourier transform of the central row of R_{eff} for different values of regularization parameters ρ and λ_k, illustrating the effect of the multiscale regularizing filter in the frequency domain. In each of the subplots, the V-shaped heavy line corresponds to the standard FBP ramp filter and the four curves from top to bottom correspond to $\rho = 0.5$ (solid line), 1.0 (dashed line), 1.5 (dashdot line) and 2.0 (dotted line) respectively (in some subplots some of the lines overlap). In all cases we fixed $\sigma^2 = 1$ (the overall prior model amplitude) and $\overline{Q} = 1$ (the prior model DC variance).

in decreased high frequency attenuation. This behavior is reasonable, in that in the first case, the data becomes less noisy while in the second the uncertainty in the prior model becomes larger. In both these cases one would want to put more reliance on the data (i.e. less regularization).

In summary then, our multiscale based regularization approach, though derived from statistical considerations and possessing all the advantages of such methods (e.g. the ability to incorporate prior knowledge in a rational way, the ability to do performance analysis and understand the relative importance of various sources of uncertainty, etc.), obtains results at no greater (and in some cases with substantially less) computational complexity than standard unregularized or ad hoc approaches. In addition, we obtain, *essentially for free*, estimates at multiple resolutions and thus the ability to extract information from data at multiple scales.

5.4.3 Examples

Next we show some examples of reconstructions using our multiscale methods in the presence of noise. The same 256 × 256 phantom shown in Figure 5-1 was used for all experiments. In each case projection data for the phantom were again generated at $N_\theta = 256$ equally spaced angles with $N_s = 256$ strips in each projection. These noise-free values were then corrupted through the addition of independent, zero-mean Gaussian noise to yield our observations. The variance λ_n of this additive noise depended on the experiment and was chosen to yield an equivalent signal-to-noise ratio (SNR) of the resulting observations, defined as:

$$\text{SNR (dB)} = 10 \log \frac{\sum_{k=1}^{N_\theta} \|T_k f\|^2}{\lambda_n N_\theta N_s} \tag{5.22}$$

where, recall, $T_k f$ is the noise-free projection data at angle k. Finally, in all multiscale reconstructions we show here the Daubechies D_3 wavelet was used in the definition of W for the reconstruction.

The first example, shown in Figure 5-7, demonstrates reconstruction from noisy data using the *unregularized* multiscale approach of Section 5.3. A value of 5.5×10^5

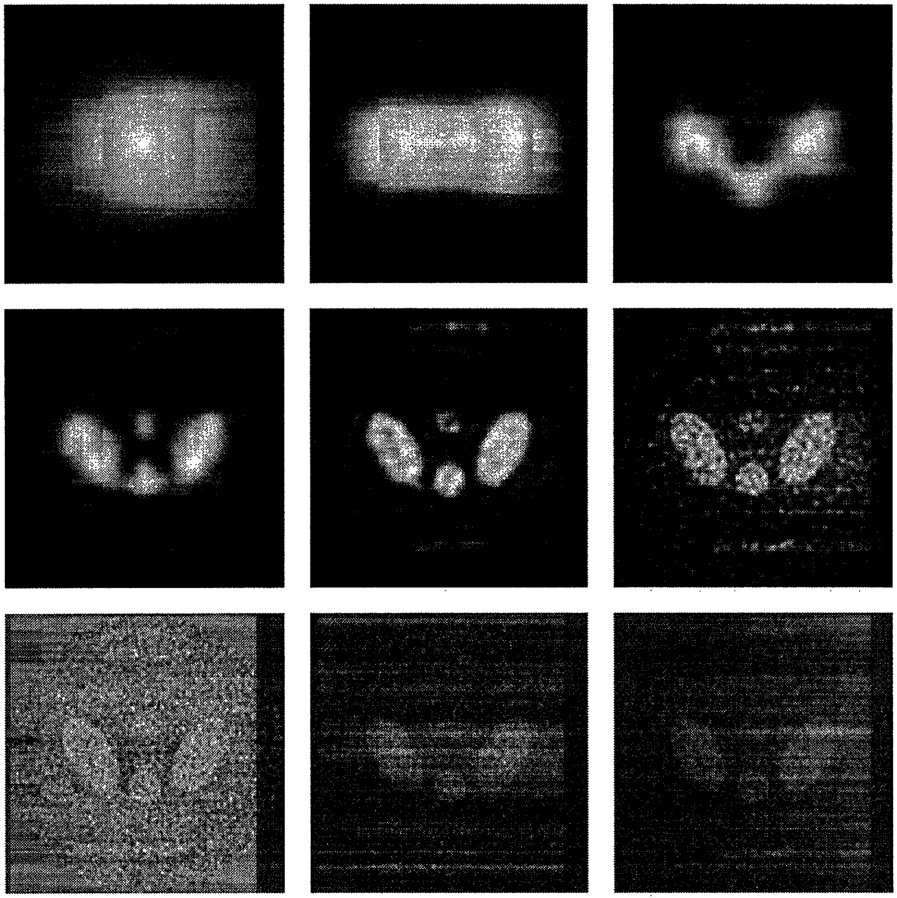


Figure 5-7: Reconstructions of phantom of Figure 5-1 from 5 dB SNR projection data based on unregularized Algorithm 3 using D_3 wavelet. Reconstructions are shown at various scales demonstrating the smoothing effect that can be achieved. First row, left: $f^{(1)}$. First row, middle: $f^{(2)}$. First row, right: $f^{(3)}$. Second row, left: $f^{(4)}$. Second row, middle: $f^{(5)}$. Second row, right: $f^{(6)}$. Third row, left: $f^{(7)}$. Third row, middle: $f^{(8)}$. The standard FBP is shown in the third row, right for comparison. The FBP reconstruction is the same as $f^{(8)}$, since this is the complete unregularized reconstruction.

was used for the variance λ_n of the added noise, which resulted in a SNR value of 5 dB. This figure shows the various scale approximate object reconstructions $f^{(j)}$ corresponding to the unregularized Algorithm 3 for the complete range of scales $m = 1, \ldots, 8$. As before, the approximations become finer from left to right and top to bottom (so that the upper left frame is $f^{(1)}$ and the bottom middle frame corresponds to $f^{(8)}$). The bottom right frame shows the standard FBP reconstruction based on the noisy data. Since $f^{(8)}$ corresponds to the unregularized complete finest scale reconstruction it is also the same as the standard FBP reconstruction based on the noisy data for this case. The figure illustrates the resolution-accuracy tradeoff inherently captured in the multiscale framework and confirms the point that even in the unregularized case, information from noisy observations can be focused by stopping the reconstruction at a coarse scale, for example scale 5 (center row, middle in the figure). The finer scale detail contributions $\Delta f^{(j)}$, $j \geq 5$ are mainly noise which obscure the object features (c.f. Figure 5-8). In particular, in the finest scale reconstruction (i.e. the standard FBP reconstruction) the object is almost completely lost in the noise.

Next we show estimates generated by our multiscale MAP regularized estimation method discussed in this section. Figure 5-9 shows the various scale approximate object reconstructions $\widehat{f}^{(j)}$ corresponding to our multiscale MAP estimate of $\widehat{\xi}_k$ using noisy data with same SNR (i.e. SNR = 5 dB) as in Figure 5-7. The MAP estimate $\widehat{\xi}_k$ was generated using the extremely efficient approximate expression (5.20), which, for the Daubechies D_3 wavelet we are using, was indistinguishable from the corresponding estimate based on the exact expression (5.18). Again the approximations become finer from left to right and top to bottom in the figure. For these reconstructions we chose the modeled observation noise variance as $\lambda_k = 5.5 \times 10^5$. For the statistical model parameters of the prior, the decay rate across scale of the added detail variance was chosen as $\rho = 1.5$, the overall magnitude of the prior was set to $\sigma^2 = 11$, and the variance of the prior model average value was $\overline{Q} = 1$. The effect of the regularization can be readily seen in its ability to suppress noise in the finest scale reconstruction. For comparison, the standard FBP reconstruction for this case is given on the bottom

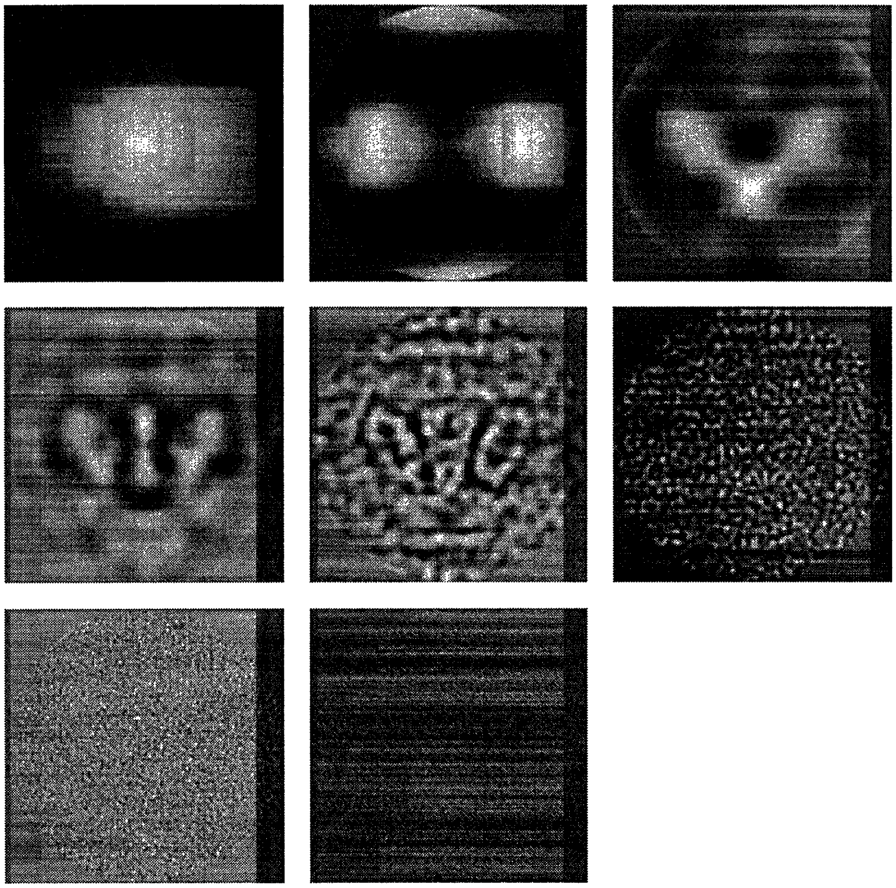


Figure 5-8: The detail added between successive scales in the reconstructions of Figure 5-7. First row, left: $\Delta f(0)$. First row, middle: $\Delta f(1)$. First row, right: $\Delta f(2)$. Second row, left: $\Delta f(3)$. Second row, middle: $\Delta f(4)$. Second row, right: $\Delta f(5)$. Third row, left: $\Delta f(6)$. Third row, middle: $\Delta f(7)$.

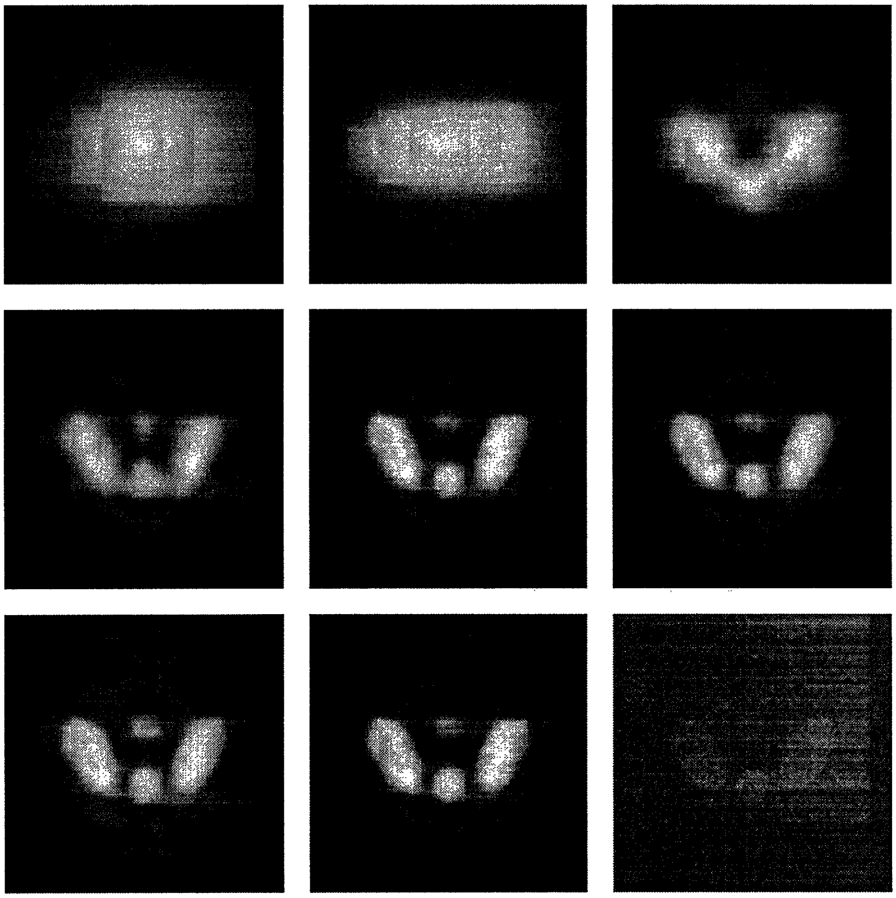


Figure 5-9: Multiscale MAP regularized reconstructions at various scales of phantom of Figure 5-1 from 5 dB SNR projection data using D_3 wavelet. The values of the statistical model parameters used are $\lambda_k = 5.5 \times 10^5$, $\rho = 1.5$, $\sigma^2 = 11$, $\overline{Q} = 1$. First row, left: $\widehat{f}^{(1)}$. First row, middle: $\widehat{f}^{(2)}$. First row, right: $\widehat{f}^{(3)}$. Second row, left: $\widehat{f}^{(4)}$. Second row, middle: $\widehat{f}^{(5)}$. Second row, right: $\widehat{f}^{(6)}$. Third row, left: $\widehat{f}^{(7)}$. Third row, middle: $\widehat{f}^{(8)}$. For comparison, the standard FBP reconstruction for this case is given in the third row, right. The improved ability of the regularized reconstructions to extract information is demonstrated.

row, right in Figure 5-9. In addition, the multiscale nature of the information focusing can be seen in the scale evolution of the estimates. In particular, there appears to be little difference between scale 5 and finer scale estimates in the figure, suggesting that little additional information is being obtained in proceeding to such finer scales, that the additional degrees of freedom being added at such finer scales are not really being supported by the data, and thus that we should stop the reconstruction at this coarser scale. This observation is confirmed in Figure 5-10 where we show the detail reconstructions corresponding to the regularized reconstructions of Figure 5-9. Further, estimates at scale 5 and coarser appear quite similar to the corresponding *unregularized* estimates in Figure 5-7, showing that these coarser scale estimates are dominated by the data and are not very dependent on the prior model at this point anyway.

Finally, in Figure 5-11, we show a series of finest scale multiscale MAP regularized reconstructions, corresponding to different choices of the prior model texture as determined by the parameter ρ. The same phantom as before is used, but we use observations with a SNR of -10 dB (much worse than used above). The standard FBP reconstruction is shown for comparison in the far right image of the figure. The object is completely lost in the FBP reconstruction at this extreme level of noise. The MAP reconstructions are shown in the first three frames of the figure, with a smoother, more correlated prior model being used as we proceed from left to right. The specific multiscale MAP model parameters were chosen as follows. The observation noise variance was chosen as $\lambda_k = 1.7 \times 10^7$. The overall prior model magnitude was set to $\sigma^2 = 17$ while the prior model DC variance was set to $\overline{Q} = 1$. The prior model texture parameter ρ took on the values $\{0.5, 1.0, 1.5\}$. The increased smoothness in the prior can be seen to be reflected in increased smoothness of the corresponding estimates. Note also the ability of the algorithm to pull out at least the global object features in the presence of this substantial amount of noise. Again, the more highly smoothed reconstructions (corresponding to higher values of ρ) appear quite similar to the coarser level, unregularized reconstructions shown previously, showing that we are really accessing the coarse level information in the data.

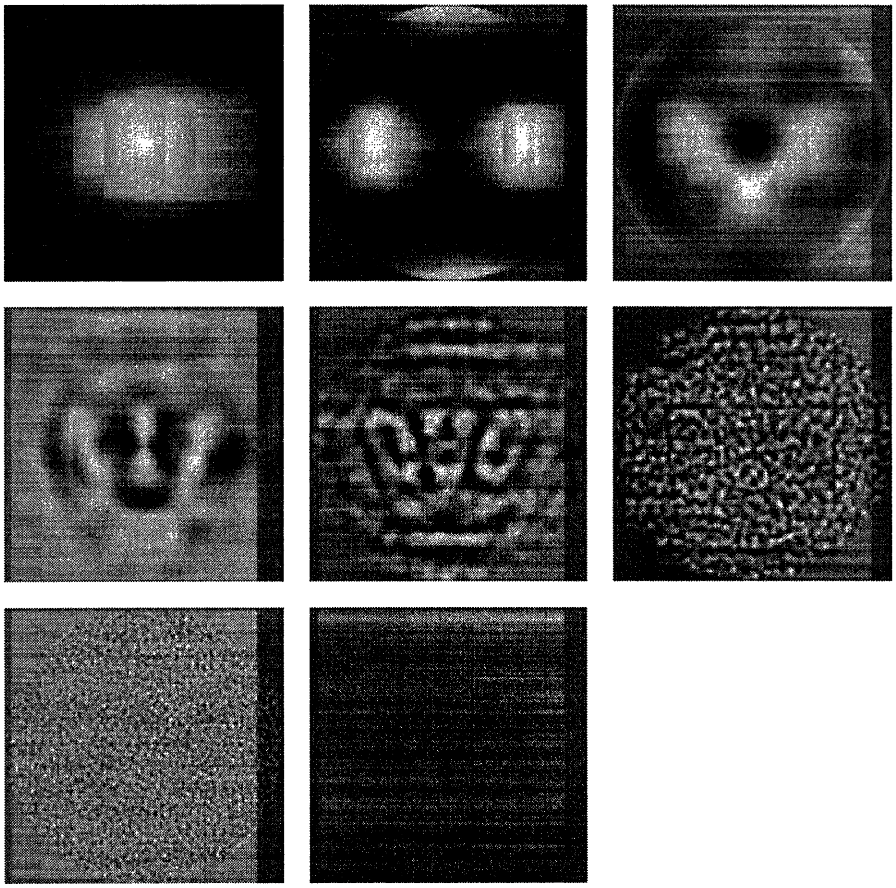


Figure 5-10: The detail added between successive scales in the reconstructions of Figure 5-9. First row, left: $\Delta\hat{f}(0)$. First row, middle: $\Delta\hat{f}(1)$. First row, right: $\Delta\hat{f}(2)$. Second row, left: $\Delta\hat{f}(3)$. Second row, middle: $\Delta\hat{f}(4)$. Second row, right: $\Delta\hat{f}(5)$. Third row, left: $\Delta\hat{f}(6)$. Third row, middle: $\Delta\hat{f}(7)$.

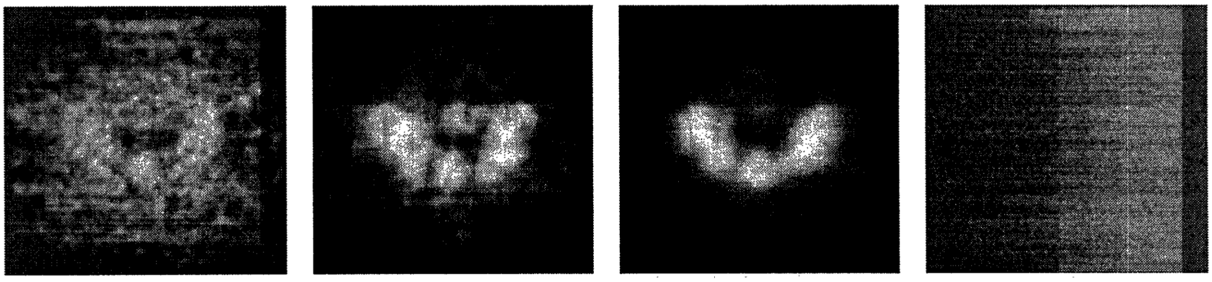


Figure 5-11: Multiscale MAP regularized reconstructions of the phantom of Figure 5-1 at the finest scale from -10 dB SNR observations for different choices of prior model texture, ρ, with $\lambda_k = 1.7 \times 10^7$, $\sigma^2 = 17$, and $\overline{Q} = 1$, are shown in the first three frames: Frame 1: $\rho = 0.5$. Frame 2: $\rho = 1.0$. Frame 3: $\rho = 1.5$. For comparison the standard FBP reconstruction is shown in the last frame on the far right.

5.5 Discussion

In this chapter we have specialized our NP-based multiscale tomographic reconstruction technique from Chapter 3 to yield fast reconstructions at multiple scales in the case when the projection data are dense. We have achieved this by adapting our multiscale framework from Chapter 3 to the efficient filtered back-projection (FBP) reconstruction technique. The result is a highly efficient method to compute our multiscale object reconstructions, in particular, no more complex than the widely used standard FBP method. Yet, unlike the FBP method, our multiscale reconstructions also provide a framework for the extraction and presentation of information at multiple resolutions from data. Further, our resulting multiscale relationships between data and object allow extremely simple *approximations* to be made to our exact relationships with *virtually no loss in resulting image quality*, thus further improving the potential efficiency of our approach. Such approximations are not possible with the standard FBP method, as they result in severe artifacts.

In addition, based on this wavelet-based multiscale framework, we have proposed a statistically-based multiresolution MAP estimation algorithm. This method provides statistically regularized reconstructions from noisy data, and does so at multiple resolutions, *at no more effort than is required for the standard FBP method.* This approach, based on the construction of prior models directly in scale-space, allows for the inclusion of natural, self-similar prior models into the reconstruction process. In contrast, conventional statistically-based regularization methods, utilizing MRF-type prior models constructed directly in (finest scale) object space, lead to extremely complex and taxing optimization problems. The result has typically been that such statistically motivated methods have been shunned in practice in favor of fast, though ad hoc, approaches. Our results provide a bridge between these two extremes. Further, in providing estimates at multiple resolutions, our results provide tools for the assessment of the resolution versus accuracy tradeoff, wherein we expect coarser scale features of data to be more accurately determined than finer scale ones. Though we did not exploit this ability in the present paper, our formulation also allowed for the possibility of combining data from projections of fundamentally different quality, through the specification of different noise variances λ_k at different angles. The resulting estimates do not correspond to a simple FBP or even rolled off FBP reconstruction, yet are easily obtain in our framework. Further, as before, our multiscale MAP approach leads to algorithms which are amenable to an additional level of approximation, with resulting improved efficiency, again at virtually no loss in corresponding reconstruction quality. Finally, the self-similar prior model that we use in this chapter assumes that the sets of multiscale object coefficients, ξ_k, are independent from angle to angle and scale to scale, and that they are independent and identically distributed within a given scale. Obviously other choices may be made for the statistics for the multiscale object coefficients, and we briefly discuss some particularly interesting possibilities in Chapter 8.

5.A Details about the formation of FBP ramp-filter matrix R

In this work we take the matrix R to represent the practical FBP filtering operator. In the ideal case, this FBP ramp filter operation is given by:

$$F_N^{-1} H_N F_N, \tag{5.23}$$

where F_N is a $N \times N$ matrix representing the 1-D Fourier transform operation on a sequence of length N, and H_N is a $N \times N$ diagonal matrix containing ideal high-pass "ramp" filter coefficients for a length N sequence. The matrix H_N has a diagonal entry of 0 since it gives zero weight to the frequency cell centered around 0. Thus the ideal ramp filter coefficient matrix H_N, and hence the matrix (5.23), is not invertible. In practice, however, to avoid dishing (i.e. interperiod interference) and DC artifacts, a filtering operator R is used which is constructed according to [44]:

$$R = M\, F_{2N}^{-1}\, H_{2N}\, F_{2N}\, M^T, \tag{5.24}$$

where F_{2N} is a $2N \times 2N$ matrix representing the 1-D Fourier transform operation on a sequence of length $2N$, H_{2N} is a $2N \times 2N$ diagonal matrix containing the corresponding ideal ramp filter coefficients, and the $N \times 2N$ zero-padding matrix M is given by

$$M = \begin{bmatrix} \mathbf{0} & I_N & \mathbf{0} \end{bmatrix}. \tag{5.25}$$

If we define H to be the equivalent $N \times N$ practical ramp filter coefficient matrix such that:

$$R = F_N^{-1}\, H\, F_N = M\, F_{2N}^{-1}\, H_{2N}\, F_{2N}\, M^T, \tag{5.26}$$

then H can be seen to be given by:

$$H = F_N\, M\, F_{2N}^{-1}\, H_{2N}\, F_{2N}\, M^T\, F_N^{-1}. \tag{5.27}$$

One can show that H is a diagonal matrix and the diagonal elements of H are almost identical to H_N except that H gives small positive weighting to the frequency cells centered around 0 (refer to [44] for a plot of the diagonal elements of H). Thus H has no 0 diagonal entry, resulting in an invertible $R = F^{-1}HF$. The only issue that remains now is of the conditioning of such a R. The above procedure for computing R results in a relatively well-conditioned matrix, with the condition number of R ranging from 24 for $N = 16$ to 389 for $N = 256$. Intermediate values of N result in condition numbers between 24 and 389. In the work of this paper we use this practical, and thus invertible, filtering operator given in (5.26) for all calculations.

Chapter 6

Discrimination Of Fractal Fields From Tomographic Data

6.1 Introduction

In this chapter we consider the problem of discrimination of fractal fields, with different fractal dimensions, directly from the noisy tomographic projection data. This application is motivated from the medical field, where a change in fractal dimension is used to differentiate between normal and abnormal conditions in many different contexts, including diagnosis of liver abnormalities [11].

The conventional method for solving this problem consists of discriminating the fractal fields based on the slopes of the averaged power spectra of their projections [11]. The rationale for this method is derived from a Radon transform result which is valid only in the case of noise-free, continuous-data. In practice, the data may be noisy and sparse and in this case the above discrimination procedure may not yield accurate results.

In order to avoid any restrictions on the quality and quantity of the projection data, we formulate our discrimination problem in a discrete hypothesis testing framework. The solution to this problem is given by the maximum-log-likelihood discrimination rule and, since we make no restrictive assumptions on the data, this solution is valid for both noisy as well as sparse projection data.

If attempted either directly in the measurement or object domain, the problem of discriminating fractal fields through likelihood calculations is, however, complicated by the fact that inverses and determinants of large, full, and generally ill-conditioned fractal-field data covariance matrices are required[1]. In principle one can imagine using an extension of results from tomographic imaging of random fields [42, 43] to find whitening transformations for the covariance matrices in order to remove these complications in the likelihood calculations. However, since the results in [42, 43] are valid only for the noise-free continuous-data case and in practice the data are sampled and may be noisy, this procedure results in an inexact whitening transformation, a fact that we demonstrate subsequently.

In this chapter we show that the above mentioned complications in the likelihood calculation, namely the necessity for calculation of determinants and inverses of large, full, and generally ill-conditioned fractal-field data covariance matrices, can be removed easily by a transformation of the data into the multiscale framework. The multiscale data covariance matrices are sparse and in addition, just as in the case of the multiscale natural pixel (NP) matrix, are naturally partitioned into ill-conditioned coarsest scale approximation blocks and relatively well-conditioned multiscale detail blocks. This natural partitioning in the multiscale framework eliminates the need for costly techniques, like the eigen-decomposition, to remove the ill-conditioning in the covariance matrices. After removing the redundancy in the coarsest scale block, the likelihood calculations in the multiscale framework require the inverse and determinant of just the relatively well-conditioned, detail blocks. Moreover we simplify our likelihood calculations even further by eliminating the need for explicit calculation of these quantities. We achieve this by using the recently introduced class of multiscale stochastic models defined on trees [4, 13–15] to realize accurate approximations of the detail block of the data covariance matrices. These tree-based models have the advantage that they lead to fast likelihood calculations [57], and thus to an efficient discrimination technique.

The chapter is organized as follows. In Section 6.2 we review a few relevant results

---

[1] This ill-conditioning is a result of the redundancy in the tomographic data representation.

that are essential for the development of our multiscale discrimination technique. In Section 6.3 we describe a particular class of statistically self-similar stochastic processes that we use to model the field. In Section 6.4 we mathematically formulate the fractal-field discrimination problem. In Section 6.5 we develop our efficient multiscale discrimination technique. We present a few examples in Section 6.6 which confirm the efficiency of our discrimination technique. Section 6.7 concludes the chapter. Appendices 6.A – 6.C contain certain technical details.

6.2 Review of relevant results

In this section we briefly review the relevant results from previous chapters that are needed to develop the framework for discrimination of fractal fields from tomographic data. Recall that the strip-integral data in tomography are of the form:

$$y(m) = \iint_{\Omega} f(u,v)\, S_m(u,v)\, du\, dv, \qquad m = 1, \ldots, N_\theta N_s, \tag{6.1}$$

where f is the object, $S_m(u,v)$ is the indicator function of the m-th strip, N_θ is the number of angular positions, and N_s is the number of strips in each angular position. The discretization of the data equation (6.1) on a $N_s \times N_s$ pixel grid results in the following expression:

$$y = Tf, \tag{6.2}$$

where T is the *projection* matrix the m-th row of which is the discretized strip $S_m(u,v)$, f is the vector containing the expansion coefficients (i.e. pixels) of the object $f(u,v)$, and y is the overall vector of observations and is given by:

$$y = \begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(N_\theta N_s) \end{bmatrix}. \tag{6.3}$$

The wavelet-based multiscale representation of the observation equation (6.2) is obtained as:

$$\begin{aligned} W_b y &= W_b T f, \text{ i.e.} \\ \eta &= W_b T f, \end{aligned} \tag{6.4}$$

where, if W is the matrix that represents a N_s vector in a 1-D multiscale basis, then W_b is a block-diagonal matrix with N_θ blocks along the diagonal, all equal to W. The elements in the multiscale data vector η are arranged according to angle as follows:

$$\eta = \left[\begin{array}{c} \eta_1 \\ \hline \eta_2 \\ \hline \vdots \\ \hline \eta_{N_\theta} \end{array} \right], \tag{6.5}$$

where η_k is the multiscale data vector at angle k. This vector has the form:

$$\eta_k = \left[\begin{array}{c} \eta_k^{(J-1)} \\ \hline \vdots \\ \hline \eta_k^{(0)} \\ \hline y_k^{(0)} \end{array} \right], \tag{6.6}$$

where $\eta_k^{(j)}$ is a 2^j vector that contains the scale j detail coefficients from the projection at angle k, and $y_k^{(0)}$ is the corresponding coarsest scale approximation coefficient.

Recall further that in the natural pixel (NP) reconstruction method, the system matrix C is given by:

$$C = TT^T, \tag{6.7}$$

and is large ($N_\theta N_s \times N_\theta N_s$), full, and ill-conditioned. The transformation of C to a multiscale framework, however, results in a matrix $\mathcal{C}$:

$$\mathcal{C} = W_b C W_b^T, \tag{6.8}$$

which is sparse and, in addition, can be partitioned naturally by scales into an ill-conditioned, coarsest scale approximation block, and a well-conditioned, multiscale detail block. In particular, the blocks in C which correspond to the fine scale detail couplings between data and object coefficients, are well-conditioned. The condition number gradually increases as the coarse scale detail elements are included.

6.3 The $1/f$ fractal processes

In this chapter we use a particular class of ***statistically self-similar*** stochastic processes, namely the class of so-called $1/f$ processes[2], to model the different hypothesized textures for the field. Such self-similar models are characterized by their fractal dimension and are commonly and effectively used in many application areas such as modeling of natural terrain and other textures, biological signals, geophysical and economic time series, etc. [11,12,58,78,80]. As an example, the texture corresponding to the surface of a normal liver may be modeled as a $1/f$ process with a particular fractal dimension. This fractal dimension changes in pathological cases and thus can be used as a factor to discriminate between normal and abnormal livers [11].

The measured power spectra[3] $S_f(w_u, w_v)$ of the field f, corresponding to a $1/f$ process, obeys the following power law relationship:

$$S_f(w_u, w_v) = \frac{\sigma_f^2}{|w|^\gamma}, \tag{6.9}$$

where w is the angular frequency, σ_f^2 is a constant, and γ is the *spectral parameter*.

In this thesis we consider the simplest examples of $1/f$ processes, namely *fractional Brownian motions*. For any such process, $f(u, v)$ has Gaussian increments with mean

---

[2] The "f" in the $1/f$ processes refers to frequency, and should not be confused with the notation "f" that we use in this thesis for the object. We will attempt to avoid this confusion by using the angular frequency "w" instead of the frequency "f" wherever necessary.

[3] From here onwards in this chapter, as is conventionally done, we use the symbol "S_g" to denote the power spectral density of function g. Recall that we had used the symbol "S" for strip functions in the previous chapters. The potential for future confusion between these two definitions does not arise here, however, as we do not refer to the strip functions again in this chapter.

zero, and a variance that depends only on the distance between the increments, i.e.

$$f(u_1, v_1) - f(u_2, v_2) \sim \mathcal{N}(0, \Lambda_{\text{inc}} \propto ((u_1 - u_2)^2 + (v_1 - v_2)^2)^H), \tag{6.10}$$

where H is a parameter, with $0 < H < 1$. It can be shown that [2] the "averaged" spectral density corresponding to (6.10) satisfies (6.9), with spectral parameter[4]:

$$\gamma = 2H + 2. \tag{6.11}$$

The *fractal dimension* D which is commonly used to capture textural differences between fractal fields, is defined in terms of the parameter H as follows:

$$D \triangleq 3 - H. \tag{6.12}$$

The fractal nature in $f(u, v)$ arises from the property of statistical self-similarity. The random process $f(u, v)$ is said to be statistically self-similar if it has the same probability distribution as the scaled version $a^{-H} f(au, av)$, where a is a constant, and H is the same parameter as in (6.10). In this chapter we only consider zero-mean fractional Brownian processes, which are characterized just by their covariance matrix, or alternatively by their spectral density (6.9). In this case, $f(u, v)$ is statistically self-similar if:

$$E[f(u_1, v_1) f(u_2, v_2)] = a^{-2H} E[f(au_1, av_1) f(au_2, av_2)]. \tag{6.13}$$

It can be shown [2] that (6.13) is equivalent to the following equality that is satisfied by the power spectra of the fractional Brownian motion process (6.9):

$$S_f(w_u, w_v) = |a|^{\gamma} S_f(aw_u, aw_v), \tag{6.14}$$

implying that the fractional Brownian motion process is statistically self-similar.

---

[4] The fractional Brownian motion process is non-stationary and hence one has to resort to specialized definitions of power spectra for such non-stationary processes. The "averaged spectrum" is one such definition and we refer the reader to [2, 26] for details.

Finally, it can be shown that the increments of a fractional Brownian process exhibit a long-range correlation [79]. As a result, a persistent correlation is also seen in the process itself. It is this persistent correlation structure that results in a *full* covariance matrix for the fractional Brownian motion process.

6.4 Discrimination problem statement

In this chapter we consider the problem of discrimination of fractal fields from sparse and noisy projection data. As mentioned previously, the conventional method for solving this problem consists of discriminating the fractal fields based on the slopes of the averaged power spectra of their projections [11]. The rationale for this method is derived from a Radon transform result which states that the spectral parameter of a projection of a fractal field is equal to one plus the spectral parameter of the field itself, and that the projections at different angles are uncorrelated with each other. However this result is valid only for the noise-free, continuous-data case. In practice the data are sampled and may be noisy and in this case, as we shall see in Section 6.6, the above discrimination procedure is decidedly suboptimal. In particular, the projections at different angles are correlated due to the sampled nature of the data and one can expect that a discrimination technique that incorporates this correlation will perform better than the conventional method that assumes the projections to be uncorrelated. Further, in the presence of noise, the high frequency components of the data are mostly corrupted by noise and hence their contribution should ideally be de-weighted in the likelihood calculation. In fact, in the case of noisy data the conventional discrimination procedure is *empirically* modified in [11] by ignoring the high frequency data samples for the purpose of calculating the slope of the power spectra. The goal of this chapter is to develop a *statistically optimal* technique for discrimination of fractal fields from both sparse as well as noisy data by incorporating the correlation information between projections and also the statistical properties of the additive noise.

We next begin the development of our statistically optimal technique for discrimination of fractal fields from tomographic data. We pose the discrimination problem in a hypothesis testing framework and, for simplicity, only consider the binary case in this chapter. Specifically, let Λ_i be the covariance matrix for the pixel coefficients f of a fractional Brownian motion process $f(u,v)$ with spectral parameter γ_i (c.f. (6.9))[5]. Suppose that we have to discriminate between two fields, with spectral parameters γ_0 and γ_1 (i.e. with fractal dimensions $(4-0.5\gamma_0)$ and $(4-0.5\gamma_1)$) respectively, given noisy tomographic measurements y of the form:

$$\mathrm{H}_0 \ : \ y = z_0 + n, \tag{6.15}$$

$$\mathrm{H}_1 \ : \ y = z_1 + n, \tag{6.16}$$

where the noise free data z_i, corresponding to a field with spectral parameter γ_i, are given by:

$$z_i = Tf_i, \qquad f_i \sim \mathcal{N}(0,\Lambda_i), \qquad i = 0,1, \tag{6.17}$$

and the noise n is white, Gaussian, with $n \sim \mathcal{N}(0, \lambda I_{N_\theta N_s})$, and is uncorrelated with both f_0 and f_1. The log-likelihood test [76] for discriminating between the two fields is given by:

$$\begin{aligned} &\text{Decide } \gamma_1 \text{ if}: \ \mathcal{L}(\gamma_1) - \mathcal{L}(\gamma_0) \geq \delta \\ &\text{else decide } \gamma_0, \end{aligned} \tag{6.18}$$

where $\mathcal{L}(\gamma_i)$ is the log-likelihood function corresponding to hypothesis H_i, and δ is a threshold parameter that controls the tradeoff between the probability of detection (PD) and the probability of false alarm (PF). The log-likelihood $\mathcal{L}(\gamma_i)$ is given by:

$$\mathcal{L}(\gamma_i) = -\frac{1}{2}\ln|\Lambda_{y|H_i}| - \frac{1}{2}y^T\Lambda_{y|H_i}^{-1}y - \frac{N_\theta N_s}{2}\ln 2\pi, \qquad i = 0,1, \tag{6.19}$$

where $\Lambda_{y|H_i}$ is the data covariance matrix corresponding to the fractal field with

---

[5] We omit here the details on the generation of the fractal field $f(u,v)$ and the covariance matrix Λ_i, but refer the reader to Appendices 6.A and 6.B for more information.

spectral parameter γ_i, and is given by:

$$\Lambda_{y|H_i} = T\Lambda_i T^T + \lambda I_{N_\theta N_s}, \qquad i = 0, 1, \tag{6.20}$$

where recall that Λ_i is the covariance matrix corresponding to the fractal field with spectral parameter γ_i.

Note that the likelihood expression (6.19) requires the inverse and the determinant of the data covariance matrix $\Lambda_{y|H_i}$. These calculations are complicated by the fact that the matrix $\Lambda_{y|H_i}$ is *large* ($N_\theta N_s \times N_\theta N_s$), *full* and, in addition, is generally quite *ill-conditioned*[6]. Thus a straightforward implementation of the discrimination rule (6.18) is not possible. In the remainder of this chapter we explore transformations of the data that result in a (transformed) data covariance matrix, the inverse and determinant of which can be calculated with relatively little effort.

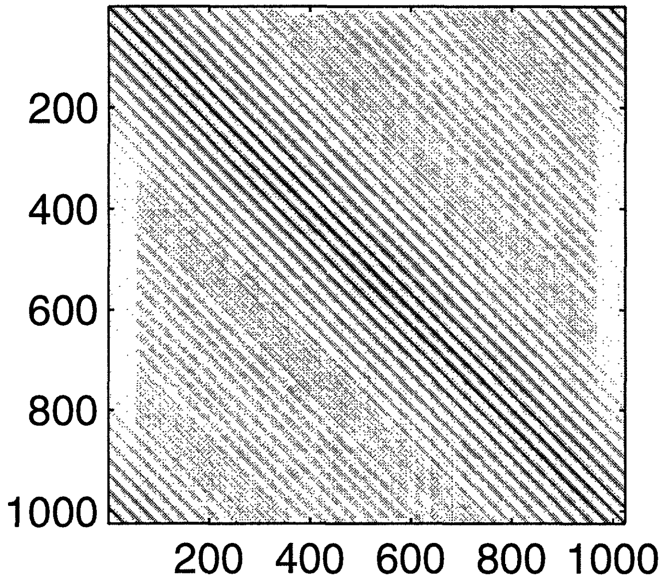
Figures 6-1 and 6-2 show the field covariance matrices Λ_i and the corresponding noise free data covariance matrices $T\Lambda_i T^T$ for spectral parameter values $\gamma_0 = 2$ and $\gamma_1 = 3$ respectively. We assume an imaging geometry with $N_\theta = N_s = 32$ so that the covariance matrices are of the size 1024 × 1024. One can easily confirm from Figure 6-2 that the noise free data covariance matrices are full. In addition, these covariance matrices do not have full rank. As mentioned previously, these features of the covariance matrices complicate the likelihood calculations.

The likelihood calculations can, however, be simplified if one can find a transformation that can be applied to the data vector in a computationally efficient manner, that has an easily computed determinant and, in addition, that results in a *diagonal* covariance matrix for the transformed data (i.e. *whitens* the data). Specifically, let M_i be a transformation such that:

$$E[(M_i y)(M_i y)^T | H_i] = M_i \Lambda_{y|H_i} M_i^T = D_i, \tag{6.21}$$

---

[6] The noise free data covariance matrix ($T\Lambda_i T^T$) is generally ill-conditioned due to the ill-conditioned nature of the projection matrix T. The covariance matrix $\Lambda_{y|H_i}$ for the noisy data (6.20)) is also generally ill-conditioned as long as the noise variance λ is not too large.



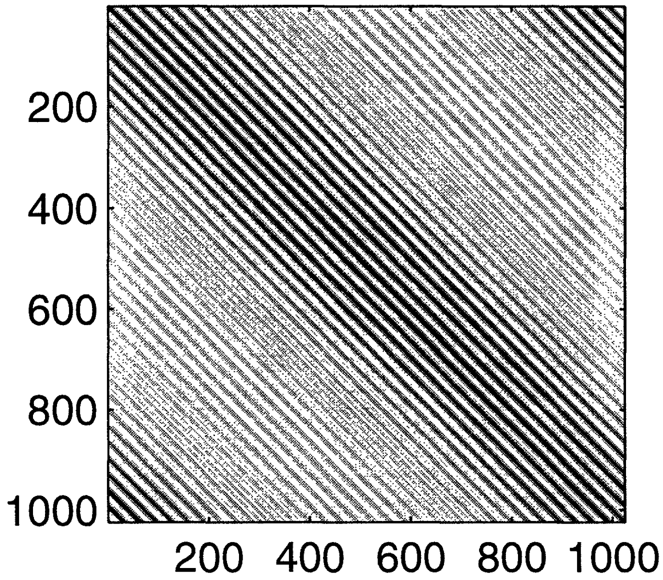
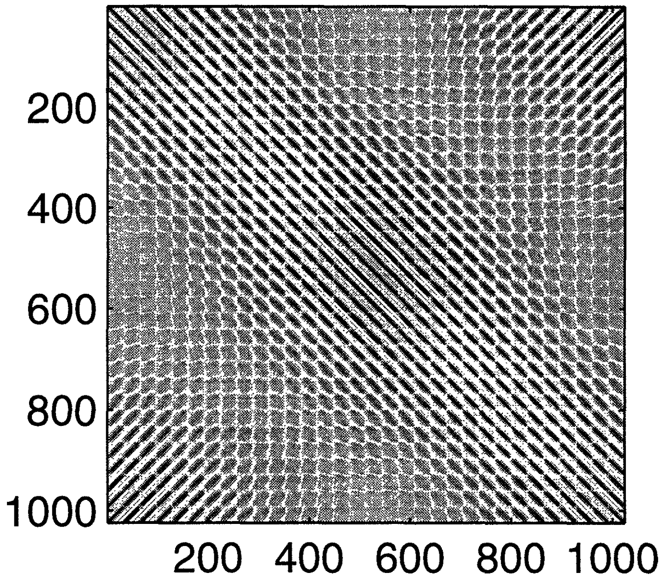


Figure 6-1: The grayscale plot (black corresponds to the maximum value and white to the minimum) of the covariance matrices for the fractal fields. Top: Λ_0 (spectral parameter $\gamma_0 = 2$). Bottom: Λ_1 (spectral parameter $\gamma_1 = 3$). The fields are 32×32, which implies that the matrices are of the size 1024×1024.



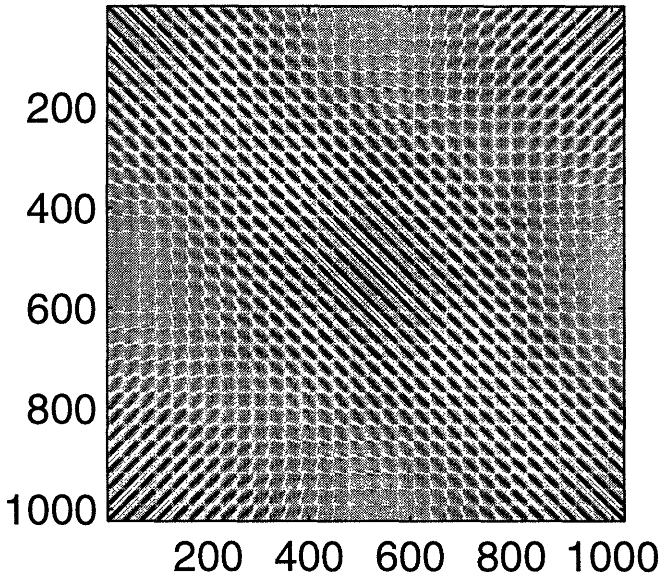


Figure 6-2: The grayscale plot (black corresponds to the maximum value and white to the minimum) of the tomographic noise free data covariance matrices corresponding to fractal fields. Top: $T\Lambda_0 T^T$ (spectral parameter $\gamma_0 = 2$). Bottom: $T\Lambda_1 T^T$ (spectral parameter $\gamma_1 = 3$). An imaging geometry with $N_\theta = N_s = 32$ is assumed, which implies that the matrices are of the size 1024×1024.

where D_i is a diagonal matrix. The likelihood $\mathcal{L}(\gamma_i)$ thus reduces to the following expression in the transform domain:

$$\mathcal{L}_t(\gamma_i) = \ln|M_i| - \frac{1}{2}\ln|D_i| - \frac{1}{2}(M_i y)^T D_i^{-1}(M_i y) - \frac{N_\theta N_s}{2}\ln 2\pi, \qquad i = 0, 1, \tag{6.22}$$

and can be computed quite easily since all that is required now is the calculation of the determinant and the inverse of a diagonal matrix. Besides, if the covariance matrix $\Lambda_{y|H_i}$ is ill-conditioned, the redundant information in the data y can be easily removed by ignoring the elements in D_i that are close to zero.

We emphasize that the challenge in the above whitening process lies in finding an appropriate whitening matrix M_i. We must be able to find this matrix in a computationally efficient manner, and also we should be able to apply it efficiently to the data vector. As an example, the eigen-decomposition of the matrix $\Lambda_{y|H_i}$ provides us with a particular whitening matrix M_i for which M_i is orthogonal so that $|M_i| = 1$. The eigen-decomposition, however, is extremely inefficient to calculate and, in general is dense so that calculating $M_i y$ is also computationally intensive. Hence the eigen-decomposition cannot be used as a practical whitening tool for large covariance matrices.

As an alternative, one can imagine using an extension of results from the existing literature on tomographic imaging of random fields [42, 43] to find the whitening matrix M_i. Specifically, it is proved in [42, 43] that the projection data corresponding to a field with spectral density 1 is whitened by a $|w|^{1/2}$ filter. This result can be generalized to fractal fields with spectral densities of the form $1/|w|^\gamma$ and, without going into details[7], we simply state here that the outcome of this generalization is that the noise free tomographic data corresponding to a fractal field with spectral parameter γ_i can be whitened by applying a $|w|^{(\gamma_i+1)/2}$ filter independently on each of the projections. This amounts to choosing M_i as a block-diagonal matrix, where all the blocks are similar and represent the discretized $|w|^{(\gamma_i+1)/2}$ filter. However, since the results in [42, 43] and in Appendix 6.C are valid only for the noise-free continuous-

---

[7] Refer to Appendix 6.C for details on this.

data case[8], this procedure results in an inexact whitening matrix when applied in practice to sampled and noisy data. That is the transformed data covariance matrix $M_i \Lambda_{y|H_i} M_i^T$ corresponding to the above choice of whitening matrix M_i is not diagonal and, in fact, deviates significantly from a diagonal matrix. In Section 6.6 we present examples that demonstrate that the discrimination technique that is based on assuming that these non-idealities are insignificant, i.e. on neglecting all off-diagonal elements of the matrix $M_i \Lambda_{y|H_i} M_i^T$ performs rather poorly, and only slightly better than a straightforward guessing approach.

Figure 6-3 shows noise free data covariance matrices for spectral parameter (i.e. γ_i) values of 2 and 3 respectively that have been transformed by using the above described procedure. An imaging geometry with $N_\theta = N_s = 32$ is assumed which implies that the covariance matrices are 1024×1024. Note from the figure that these matrices are not diagonal. In addition, these matrices do not have full rank. Now if the transformed covariance matrices were truly diagonal then, as explained previously, this singularity would not pose a major problem since all one would do in that case is ignore the elements in the diagonal matrix D_i that are equal to zero (c.f. (6.22)). However in this case, since the transformed matrices are not diagonal, the singularity is not localized and can only be removed by computing the eigen-decomposition of these matrices. This computation is costly due to the fact that the matrices are large ($N_\theta N_s \times N_\theta N_s$). The whitening procedure obtained by discretization of Radon transform thus leads to absolutely no computational gain.

Finally, note that in Figure 6-3 we have shown the performance of the conventional whitening filter for the case of *noise less* projection data. If the data are noisy, however, then the performance of such a filter is still more inaccurate. This is because any filter of the type $|w|^{(\gamma_i+1)/2}$ that is chosen to whiten the data corresponding to a $1/|w|^{\gamma_i}$ field, actually colors the white noise in the data. Specifically, in the presence of additive white noise in the data, the theoretical spectral density of the output of the whitening filter is $(|w|^{(\gamma_i+1)} + 1)$, where the term $|w|^{(\gamma_i+1)}$ is due to white noise

---

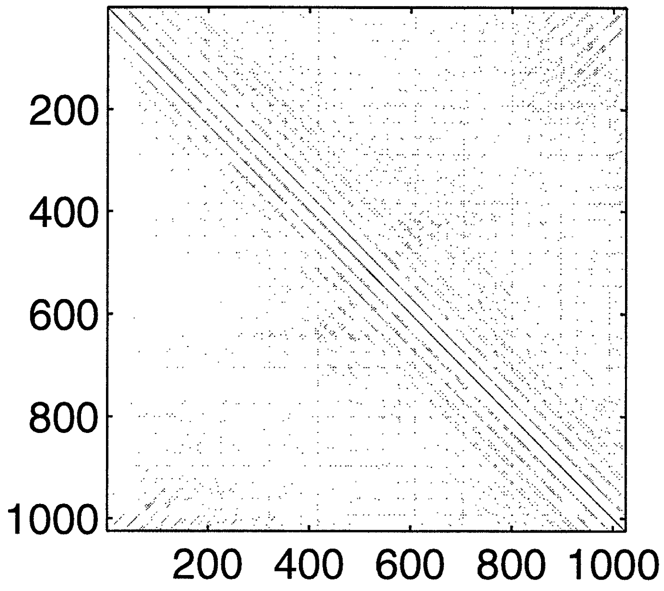
[8] This is because the results in [42, 43] assume that the various Radon transform theorems hold. These theorems, however, break down if the data are sparse and/or noisy.

which has been colored by the filter.

In the next section we will present an extremely efficient technique for whitening the projection data, using a transformation for which $|M_i| = 1$, which is valid for both sparse as well as noisy data. Specifically, we will first represent the projection data in a multiscale basis. We will then show that the form of the multiscale data covariance matrix is such that an accurate approximation of it can be realized by using a class of recently developed multiscale stochastic models on trees [4, 13–15]. These models have the advantage that they lead to a fast whitening algorithm, and hence to fast likelihood calculations [57].

6.5 A multiscale approach to discrimination of fractal fields.

In this section we develop a multiscale approach for fast discrimination of fractal fields through efficient likelihood calculations. We begin by transforming the fractal-field data into the multiscale framework. We show that this transformation results in a sparse multiscale data covariance matrix which, in addition, can be partitioned into a relatively well-conditioned multiscale detail block and an ill-conditioned coarsest scale approximation block. This natural partitioning in the multiscale framework eliminates the need for costly techniques, like the eigen-decomposition, to remove the ill-conditioning in the covariance matrix. After removing the redundancy in the coarsest scale block, the likelihood calculations in the multiscale framework require the inverse and determinant of just the relatively well-conditioned, detail block. However we simplify our likelihood calculations even further by eliminating the need for explicit calculation of these quantities. We achieve this by using the recently introduced class of multiscale stochastic models defined on trees [4, 13–15] to realize accurate approximations of the detail block of the data covariance matrices. These tree-based models have the advantage that they lead to a fast whitening algorithm [57], and thus to an efficient discrimination procedure.



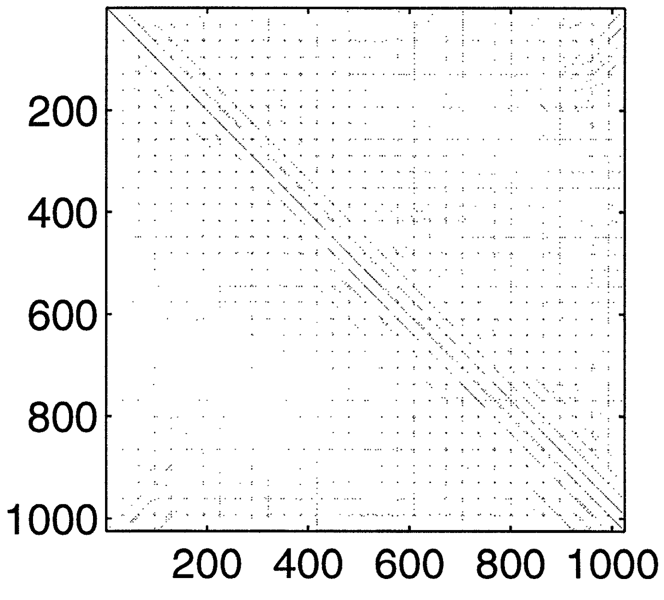


Figure 6-3: The grayscale plot (black corresponds to the maximum value and white to the minimum) of the fractal-field noise free data covariance matrices that have been whitened by using an extension of the continuous-data results in [42,43]. Top: Whitened noise free data covariance matrix for spectral parameter $\gamma_0 = 2$. Bottom: Whitened noise free data covariance matrix for spectral parameter $\gamma_1 = 3$. An imaging geometry with $N_\theta = N_s = 32$ is assumed, which implies that the matrices are of the size 1024 × 1024.

We begin the development of our discrimination technique by transforming the fractal-field tomographic data into the multiscale framework. Recall that the multiscale transformation of all N_θ angular projections contained in the data vector y, is achieved by the operation:

$$\eta = W_b y, \tag{6.23}$$

where W_b is a block-diagonal matrix with N_θ blocks along the diagonal all equal to W, and W is the multiscale representation operator for a discrete signal of length N_s. Recall further that the multiscale representation (6.23) requires very little computations ($\mathcal{O}(N_\theta N_s)$). From (6.23), the multiscale data covariance matrix corresponding to a fractal field with spectral parameter γ_i is given by:

$$\begin{aligned} \Lambda_{\eta|H_i} &= E[(W_b y)(W_b y)^T | H_i] \\ &= W_b E[z_i z_i^T] W_b^T + W_b E[n n^T] W_b^T \\ &= E[\zeta_i \zeta_i^T] + E[\nu \nu^T] \\ &= E[\zeta_i \zeta_i^T] + \lambda I_{N_\theta N_s}, \end{aligned} \tag{6.24}$$

where $\zeta_i = W_b z_i$ is the multiscale transformation of the noise free data corresponding to the field with spectral parameter γ_i, and $\nu = W_b n \sim \mathcal{N}(0, \lambda I_{N_\theta N_s})$ is the multiscale transformed noise vector. Note that the variance of the noise remains unchanged upon multiscale transformation since the matrix W_b is orthogonal. Furthermore because of orthogonality $|W_b| = 1$.

We point out to the reader that the same similarity transformation as in (6.24) when applied to the NP matrix C (c.f. (6.8)) results in a sparse multiscale matrix which can be partitioned by scales into well and ill-conditioned blocks. Now note from Figure 6-2 that the structure of the data covariance matrix $E[z_i z_i^T]$ is very similar to the NP matrix C (Figure 2-3). Thus, as in the NP case, one can expect the multiscale matrix $E[\zeta_i \zeta_i^T]$ to be sparse and, in addition, to be naturally partitioned into a relatively well-conditioned, multiscale detail block and an ill-conditioned, coarsest scale approximation block. In fact, as we show later, this is what we observe, and

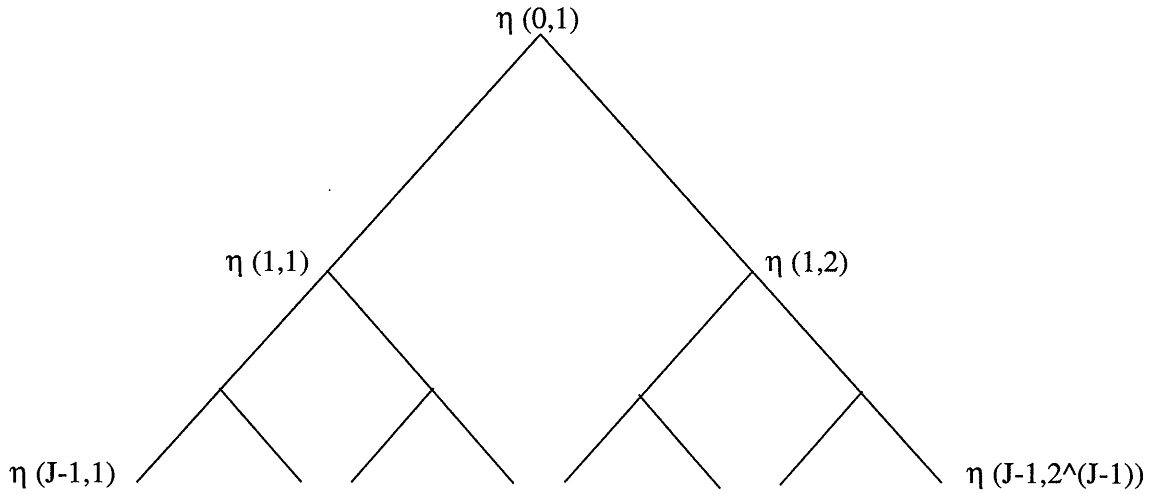


Figure 6-4: The vectors $\{\eta(j,s)\}$ can be assumed to define the nodes of a multiscale tree. These vectors are of length N_θ and contain the detail coefficients at scale j and shift s from projections at all angles.

it is this natural partitioning in the multiscale framework that helps us avoid costly eigen-decomposition computations in order to remove the ill-conditioning in the data covariance matrix.

After removing the ill-conditioning in the coarsest scale approximation block of the data covariance matrix $\Lambda_{\eta|H_i}$, the likelihood calculation in the multiscale framework requires the inverse and determinant of just the corresponding sparse and relatively well-conditioned detail block. However, we are able to develop an efficient technique for whitening the detail coefficients in the data that avoids even this determinant and inverse calculation. As a first step in the development of such a technique, we propose a particular ordering of the detail coefficients in the multiscale data vector η.

Let $\eta(j,s)$ be the N_θ vector which contains the detail coefficients at scale j and shift s from projections at all angles[9]. Now, as shown in Figure 6-4, the detail vectors:

$$\{\ \eta(j,s),\ j = 0,1,\ldots,J-1,\ s = 1,2,\ldots,2^j\ \}$$

can be assumed to define the nodes of a multiscale tree. The j-th level of the tree

---

[9] Recall that there are 2^j shifts at scale j.

consists of 2^j nodes:

$$\{\ \eta(j,s),\ s = 1, 2, \ldots, 2^j\ \}$$

which represent the 2^j shifts of the scale j detail coefficients from projections at all angles. The full detail vector η^d consists of all the vectors $\{\eta(j,s)\}$ that define the nodes of the multiscale tree in Figure 6-4:

$$\eta^d = \begin{bmatrix} \eta(0,1) \\ \eta(1,1) \\ \eta(1,2) \\ \vdots \\ \eta(j,s) \\ \vdots \\ \eta(J-1, 2^{J-1}) \end{bmatrix}. \tag{6.25}$$

Finally, to complete the multiscale representation, the coarsest scale approximation coefficients in the data have to be added to the detail vector η^d, as shown below:

$$\eta_r = P_r \eta = \left[\begin{array}{c} \eta^a \\ \hline \eta^d \end{array} \right], \tag{6.26}$$

where P_r is the (orthogonal) permutation matrix (thus with $|P_r| = 1$) that rearranges the multiscale data in η according to the tree-ordering scheme, η_r is the tree-ordered data vector, and η^a is an N_θ vector that contains the coarsest scale approximation coefficients from projection data at all angles:

$$\eta^a = \begin{bmatrix} y_1^{(0)} \\ y_2^{(0)} \\ \vdots \\ y_{N_\theta}^{(0)} \end{bmatrix}. \tag{6.27}$$

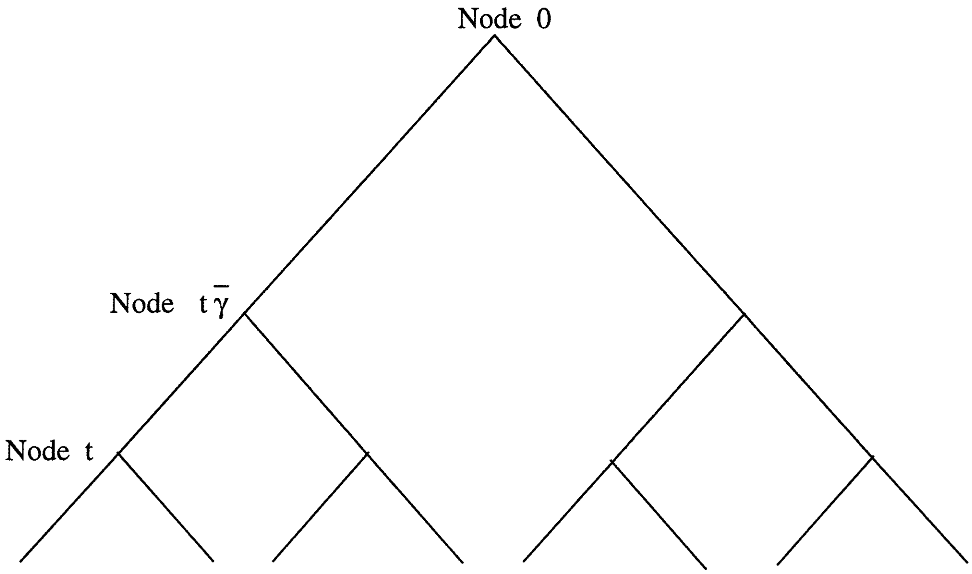


Figure 6-5: Each node of the multiscale tree is denoted by a single index $t = (j, s)$. The notation $t\bar{\gamma}$ refers to the parent node of t.

In order to simplify notation, let us denote each node of the tree by a single index:

$$t = (j, s), \tag{6.28}$$

as shown in Figure 6-5. In that case $t\bar{\gamma}$, the parent node of t on the tree, is given by:

$$t\bar{\gamma} = (j - 1, \lfloor \frac{s}{2} \rfloor), \tag{6.29}$$

where the notation $\lfloor a \rfloor$ refers to the integer part of a.

From (6.26), the covariance matrix $\Lambda_{\eta_r|H_i}$ for the tree-ordered vector η_r is given by:

$$\Lambda_{\eta_r|H_i} = P_r \Lambda_{\eta|H_i} P_r^T, \tag{6.30}$$

and can be partitioned into detail and approximation blocks as follows:

$$\begin{aligned} \Lambda_{\eta_r|H_i} &= \left[\begin{array}{c|c} E[\eta^a(\eta^a)^T|H_i] & E[\eta^a(\eta^d)^T|H_i] \\ \hline E[\eta^a(\eta^d)^T|H_i]^T & E[\eta^d(\eta^d)^T|H_i] \end{array} \right], \\ &\approx \left[\begin{array}{c|c} \Lambda_{\eta^a|H_i} & 0 \\ \hline 0 & \Lambda_{\eta^d|H_i} \end{array} \right], \end{aligned} \tag{6.31}$$

where, as in the case of the multiscale transformation of the NP matrix C, the elements of the off-diagonal blocks are negligible. Further, any ill-conditioning in $\Lambda_{\eta_r|H_i}$ is mostly concentrated in the coarsest scale approximation block $\Lambda_{\eta^a|H_i}$ and, as a result, the detail block $\Lambda_{\eta^d|H_i}$ is relatively well conditioned. As an example, in Figure 6-6 we show a histogram plot of the absolute magnitude of elements in the off-diagonal block $E[\eta^a(\eta^d)^T|H_i]$ of the covariance matrix $\Lambda_{\eta_r|H_i}$ corresponding to a fractal field with spectral parameter $\gamma_i = 2$. Note from the figure that the absolute magnitude of most of the elements in the off-diagonal block is less than two percent of the absolute maximum value in the matrix. Further for this specific example of $\gamma_i = 2$, the detail block $\Lambda_{\eta^d|H_i}$ has full rank even though the full covariance matrix $\Lambda_{\eta_r|H_i}$ is singular.

The likelihood calculation in the multiscale framework now requires the inverse of the covariance matrix $\Lambda_{\eta_r|H_i}$. This inverse can be written as:

$$\Lambda^{-1}_{\eta_r|H_i} = \left[\begin{array}{c|c} \Lambda^{+}_{\eta^a|H_i} & 0 \\ \hline 0 & \Lambda^{-1}_{\eta^d|H_i} \end{array}\right], \tag{6.32}$$

where $+$ refers to the pseudo-inverse that is required here since the approximation block $\Lambda_{\eta^a|H_i}$ is generally ill-conditioned. Thus the problem of inverting an ill-conditioned covariance matrix is easily removed in our multiscale framework and all that is required for this is the pseudo-inverse of the corresponding small $(N_\theta \times N_\theta)$ approximation block. Moreover for the examples we show in this chapter we consider the discrimination between fractal fields normalized so that they have identical expected power irrespective of the fractal dimension. This choice allows us to focus on the textural differences between the fields and not on DC power differences. This choice also implies that the coarsest scale approximation block has the same statistics under both hypotheses and thus can be ignored for the purpose of likelihood calculations without any loss in performance. We thus obtain the following expression for our multiscale log-likelihood $\mathcal{L}_m(\gamma_i)$:

$$\mathcal{L}_m(\gamma_i) = -\frac{1}{2}\ln|\Lambda_{\eta^d|H_i}| - \frac{1}{2}(\eta^d)^T\Lambda^{-1}_{\eta^d|H_i}\eta^d - \frac{N_\theta(N_s-1)}{2}\ln 2\pi, \qquad i = 0, 1, \tag{6.33}$$

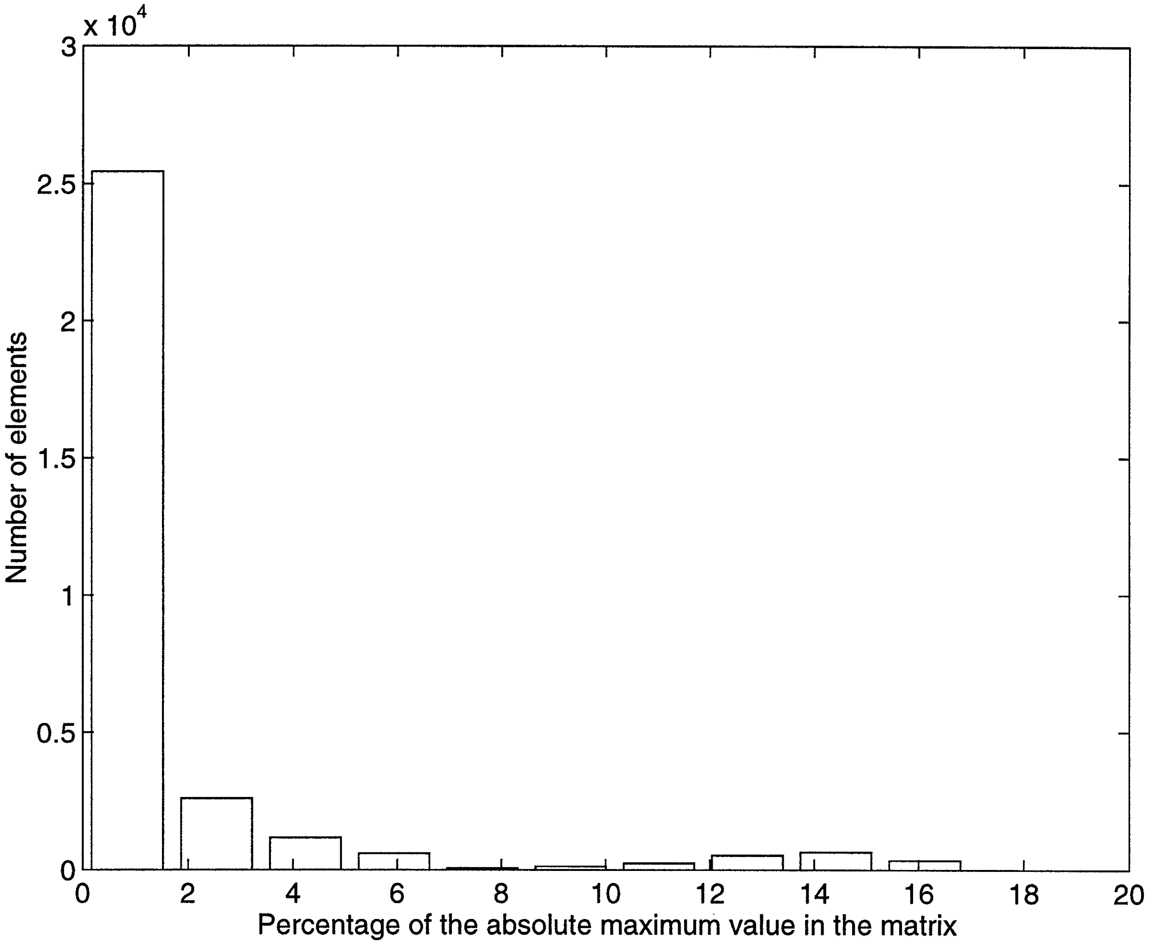


Figure 6-6: The histogram plot of the absolute magnitude of elements in the off-diagonal block $E[\eta^a(\eta^d)^T|H_i]$ of the covariance matrix $\Lambda_{\eta_r|H_i}$ corresponding to a fractal field with spectral parameter $\gamma_i = 2$. Note that the magnitude of most of the elements in the off-diagonal block is less than two percent of the maximum value in the matrix.

and the following discrimination rule:

$$\begin{aligned} &\text{Decide } \gamma_1 \text{ if: } \mathcal{L}_m(\gamma_1) - \mathcal{L}_m(\gamma_0) \geq \delta \\ &\text{else decide } \gamma_0. \end{aligned} \tag{6.34}$$

From (6.33) we see that the inverse and the determinant of the sparse and relatively well-conditioned detail matrix $\Lambda_{\eta^d|H_i}$ are now required for likelihood calculation. We describe next an efficient technique for whitening the detail coefficients η^d which eliminates the need for explicit calculation of these quantities.

To begin, note from (6.25) and (6.28) that whitening the detail vector η^d is equivalent to whitening the vectors $\eta(t)$ which are defined on the nodes of the multiscale tree in Figure 6-4. Let us assume that these vectors $\eta(t)$ correspond to observations of a dynamic state-space model in scale, as shown below:

$$\begin{aligned} \vartheta(t) &= A(t)\vartheta(t\bar{\gamma}) + B(t)\varpi(t), \\ \eta(t) &= C(t)\vartheta(t) + \nu(t), \end{aligned} \tag{6.35}$$

where $A(t)$, $B(t)$, and $C(t)$ are $N_\theta \times N_\theta$ matrices, and $\{\varpi(t)\}$ and $\{\nu(t)\}$ are independent sets of zero-mean, white, Gaussian random vectors that are independent of $\vartheta(0)$:

$$\begin{aligned} \varpi(t) &\sim \mathcal{N}(0, I_{N_\theta}), \\ \nu(t) &\sim \mathcal{N}(0, \lambda_t I_{N_\theta}). \end{aligned} \tag{6.36}$$

Now if the model (6.35) holds then, as shown in [57], there exists a fast method for whitening the vectors $\{\eta(t)\}$ with a scale recursive procedure corresponding to a whitening transformation with unity determinant. This in turn results in a fast calculation of the corresponding likelihood $\mathcal{L}_m(\gamma_i)$ (c.f. (6.33)).

The efficient, tree-based likelihood calculation procedure thus involves the following steps. First the covariance matrix $\Lambda_{\eta^d|H_i}$ is realized by choosing the appropriate model parameters $A(t)$, $B(t)$, $C(t)$, and λ_t in (6.35). Then the whitening technique developed in [57] is used to calculate the likelihood. There is a major issue, however, with this approach – the dynamics (6.35) imposes a particular structure on the covariance matrix $\Lambda_{\eta^d|H_i}$, and hence does not allow us to realize any arbitrary covariance matrix. In particular, the structure of our fractal-field data covariance matrices are such that an exact realization of these by using the model (6.35) is not possible. All hope is not lost, however, for in this case one can think of finding the model parameters such that accurate approximations $\widetilde{\Lambda}_{\eta^d|H_i}$ of the fractal-field data covariance matrices $\Lambda_{\eta^d|H_i}$ are realized.

Specifically, we find the parameters $A(t)$, $B(t)$, $C(t)$, and λ_t such that the variance of the node vectors (i.e. $E[\eta(t)\eta(t)^T]$) and their covariance with the parent (i.e. $E[\eta(t)\eta(t\bar{\gamma})^T]$) is exactly realized. In other words, the blocks in the realized covariance matrix $\widetilde{\Lambda}_{\eta^d|H_i}$ which represent $E[\eta(t)\eta(t)^T]$ and $E[\eta(t)\eta(t\bar{\gamma})^T]$, exactly match the corresponding blocks in the original covariance matrix $\Lambda_{\eta^d|H_i}$. This results in the following choice for the model parameters, and the initial condition:

$$\begin{aligned}
A(t) &= E[\zeta(t)\zeta(t\bar{\gamma})^T]E^{+}[\zeta(t\bar{\gamma})\zeta(t\bar{\gamma})^T], \\
B(t) &= \text{Real}\left\{\left(E[\zeta(t)\zeta(t)^T] - A(t)E^T[\zeta(t)\zeta(t\bar{\gamma})^T]\right)^{1/2}\right\}, \\
C(t) &= I_{N_\theta}, \\
\lambda_t &= \lambda, \\
E[\vartheta(0)\vartheta(0)^T] &= E[\zeta(0)\zeta(0)^T],
\end{aligned} \tag{6.37}$$

where $^+$ refers to the pseudo-inverse, and recall that ζ are the noise free data, and λ is the variance of the additive noise in the data.

We emphasize that the choice of model parameters according to (6.37) is not unique, but is rather a function of which blocks in the covariance matrix we wish to exactly match in the realized matrix. For example, a different expression for the model parameters is obtained if we match the covariance between the siblings rather than between the parents and their children. Given this, we claim that our choice of the model parameters according to (6.37) results in a realized covariance matrix $\widetilde{\Lambda}_{\eta^d|H_i}$

that is an accurate approximation of $\Lambda_{\eta^d|H_i}$. As an example, we show in Figure 6-7 the covariance matrix $\Lambda_{\eta^d|H_i}$ and the corresponding approximation $\widetilde{\Lambda}_{\eta^d|H_i}$ for a spectral parameter value $\gamma_i = 2$. We assume an imaging system with $N_\theta = N_s = 32$ so that the matrices are of the size $32(32-1) \times 32(32-1) = 992 \times 992$, and we assume the data to be noise free. Further, we use the Daubechies D_{10} wavelet for multiscale decomposition[10]. One can see from the figure that the covariance matrix $\Lambda_{\eta^d|H_i}$ and the tree realized version $\widetilde{\Lambda}_{\eta^d|H_i}$ are fairly similar. More important for us, however, is how well the accuracy of this approximation is reflected in the likelihood calculations. In the next section we demonstrate that the likelihoods calculated by using the tree-model model (6.35) (i.e. the realized matrix $\widetilde{\Lambda}_{\eta^d|H_i}$) are very similar to the ones calculated using the exact covariance matrix $\Lambda_{\eta^d|H_i}$. In other words, if we define $\widetilde{\mathcal{L}}_m(\gamma_i)$ to be the likelihood obtained by using the tree model, then:

$$\widetilde{\mathcal{L}}_m(\gamma_i) = -\frac{1}{2}\ln|\widetilde{\Lambda}_{\eta^d|H_i}| - \frac{1}{2}(\eta^d)^T \widetilde{\Lambda}^{-1}_{\eta^d|H_i}\eta^d - \frac{N_\theta(N_s-1)}{2}\ln 2\pi, \qquad i = 0, 1, \quad (6.38)$$

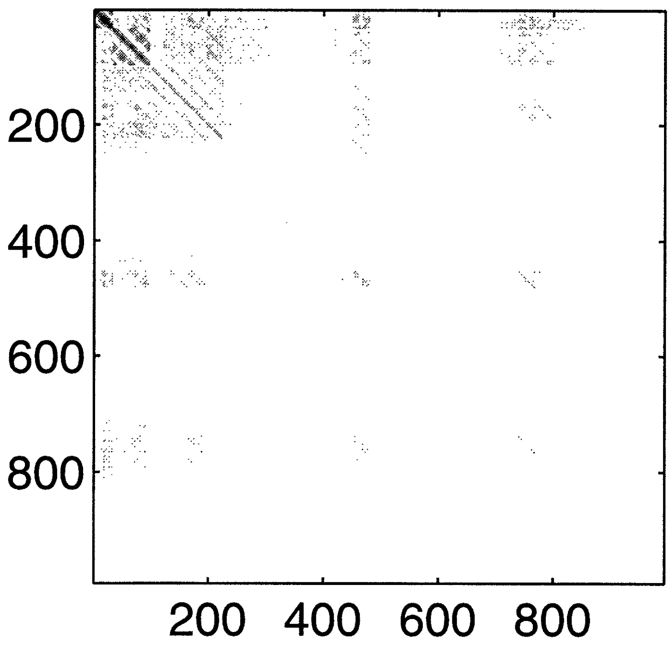
is a very good approximation to the exact likelihood $\mathcal{L}_m(\gamma_i)$ (c.f. 6.33). We point out to the reader that the expression for $\widetilde{\mathcal{L}}_m(\gamma_i)$ shown in (6.38) serves only an explanatory purpose. In reality, as mentioned previously, we use the results in [57] to efficiently whiten the vector η^d. This in turn eliminates the need for explicit calculation of the inverse and determinant of $\widetilde{\Lambda}_{\eta^d|H_i}$ as suggested by (6.38).

6.6 Examples

In this section we present examples demonstrating the effectiveness of our multiscale technique for discrimination of fractal fields. We discretize the fractal fields on a 32×32 pixel grid and assume an imaging geometry with $N_\theta = 5$ angular projections and $N_s = 32$ strips in each projection. Our choice for such a small number of angular projections (i.e. 5) is based on the following considerations. First, one of the purpose

---

[10] Our experience is that the higher order Daubechies wavelets result in a more sparse $\Lambda_{\eta^d|H_i}$, and this in turn enables a more accurate realization on the tree



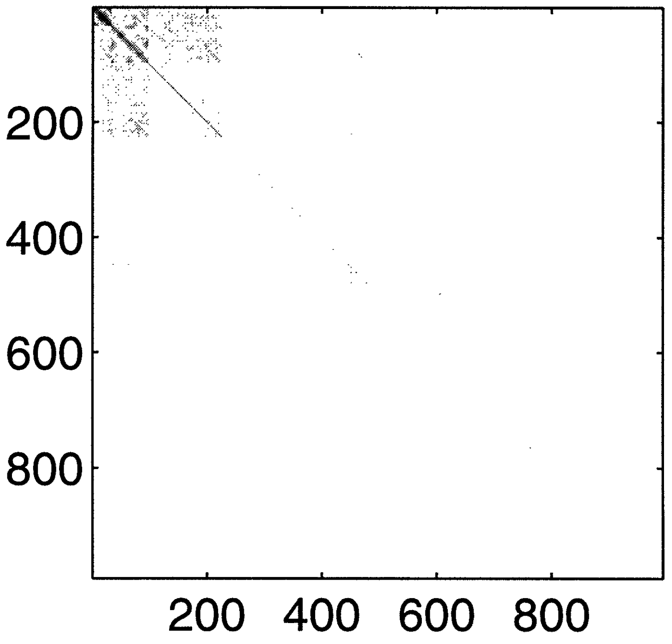


Figure 6-7: The grayscale plot (black corresponds to the maximum value and white to the minimum) of the detail covariance matrix $\Lambda_{\eta^d|H_i}$ (top) and the approximation $\widetilde{\Lambda}_{\eta^d|H_i}$ realized on the tree (bottom) for a spectral parameter value $\gamma_i = 2$. An imaging geometry with $N_\theta = N_s = 32$ is assumed, which implies that the matrices are of the size $32(32-1) \times 32(32-1) = 992 \times 992$.

of this section is to compare the performance of our tree-based approximate likelihood discriminator with the corresponding exact multiscale likelihood discriminator. Note that the exact discriminator requires the calculation of inverses and determinants of $N_\theta(N_s-1) \times N_\theta(N_s-1)$ detail covariance matrices, which becomes computationally prohibitive in the case of a large amount of data. Thus, while our approximate tree-based discriminator avoids this computational problem and thus can easily be applied to larger data sets (as in [57]), comparison to the optimal discriminator requires that we consider a problem of size that makes the exact likelihood calculation feasible. Second, the choice of a small number of angular projections enables us to test the performance of our discrimination technique by using *extremely sparse data*. Indeed one of the objectives of this chapter is to develop a discrimination technique that avoids the restriction on the quantity of the tomographic data.

In all the results we present in this section we use the Daubechies wavelet D_{10} [18] for multiscale representations. We assume the projection data to be noisy and we measure the magnitude of the additive noise by the signal-to-noise ratio, defined as:

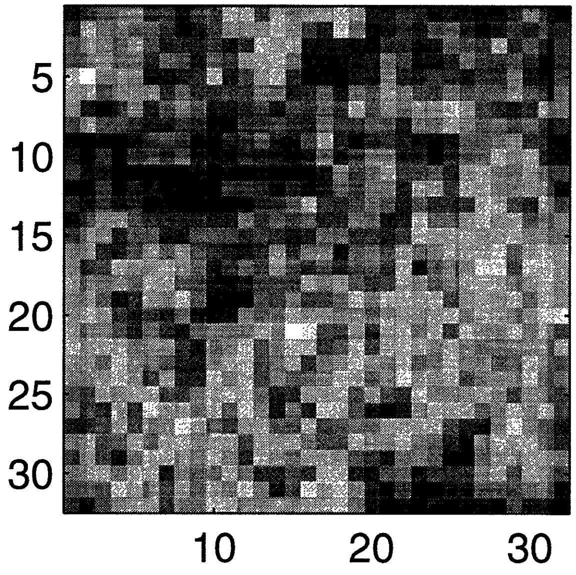
$$\mathrm{SNR(dB)} = 10\log_{10}\left[\frac{\mathrm{Trace}(T\Lambda_i T^T)}{N_\theta N_s \lambda}\right], \tag{6.39}$$

where λ is the variance of the zero-mean additive noise which we assume to be Gaussian and white, T is the projection matrix, and Λ_i is the covariance matrix for the fractal field corresponding to a spectral parameter value γ_i. Note that $T\Lambda_i T^T$ is the covariance matrix for the noise free tomographic data. We normalize the fractal-field covariance matrices Λ_i such that the noise free data covariance matrices have the same trace for all i (i.e. for all values of the spectral parameter). This implies that the expected power in the fractal-field data is same for all fractal fields irrespective of their spectral parameter (i.e. their fractal dimension).

In the first few examples we concern ourselves with discrimination of fractal fields with spectral parameter values of $\gamma_0 = 2$ and $\gamma_1 = 3$ respectively. We show one particular realization for each of these fields in Figure 6-8, and in Figure 6-9 the corresponding filtered back-projection (FBP) reconstructions from 5 dB SNR projection data is displayed. From Figure 6-9 we see that it is not possible to visually discriminate between the fractal fields from the corresponding reconstructions since the latter are mostly corrupted by noise and sparse data artifacts. However, as will be shown later, our tree-based discrimination technique will be able to distinguish between the two fractal fields from 5 dB SNR data with a probability of error that is very close to zero.

In Figure 6-10 we show likelihood results for two 32×32 fields with spectral parameters $\gamma_0 = 2$ and $\gamma_1 = 3$ respectively. We corrupt the tomographic data with additive noise such that the SNR = 5 dB. We plot the exact likelihood difference, obtained using $\Lambda_{\eta^d|H_i}$, $\mathcal{L}_m(\gamma_1) - \mathcal{L}_m(\gamma_0)$ (solid line) and the approximate difference $\tilde{\mathcal{L}}_m(\gamma_1) - \tilde{\mathcal{L}}_m(\gamma_0)$ (broken line) computed with our tree-based approximation $\tilde{\Lambda}_{\eta^d|H_i}$, for different realizations of the field and the noise. The plot in the top half of the figure shows the likelihood differences when the field is generated according to $\gamma = 2$ and the plot in the bottom half is for $\gamma = 3$. Note from the figure that the likelihood difference calculated by using the tree model tracks the corresponding exact difference very closely on an individual sample path basis and thus we expect the performance of our tree-based detection approach to be comparable to that of the optimal discriminator.

We discuss next the results of an experiment where we measure the performance of the our tree-based likelihood discriminator for different values of the threshold parameter δ that controls the tradeoff between the probability of false alarm (PF) and the probability of detection (PD). A particular measure of this performance, called the receiver operating characteristic (ROC) curve [76], is obtained by plotting PD versus PF. Besides displaying the ROC curve for our tree-based discrimination method, we also show the corresponding curve for the optimal discrimination technique employing the exact likelihood expression. Further, to compare the performance of these two techniques with the corresponding suboptimal ones, we also show the ROC curves for the conventional discrimination technique that is based on the slope of the averaged power spectra of the projections [11], and the discrimination technique obtained by using the whitening transformation (c.f. Appendix 6.C) derived from the noise-free,



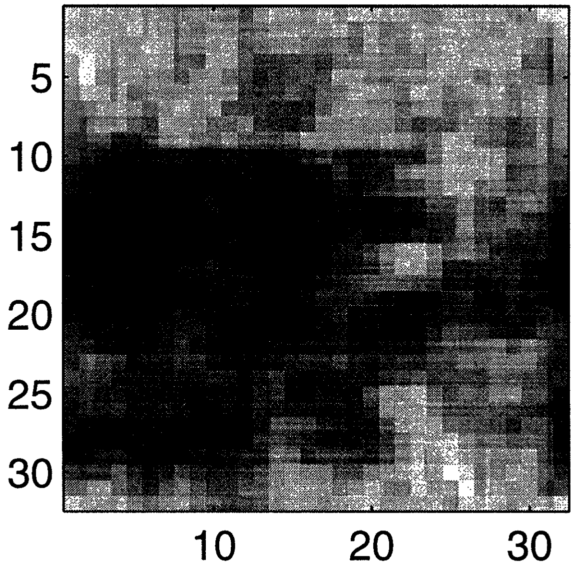


Figure 6-8: Realizations of 32 × 32 fractal fields. Top: Spectral parameter $\gamma_0 = 2$. Bottom: Spectral parameter $\gamma_1 = 3$.

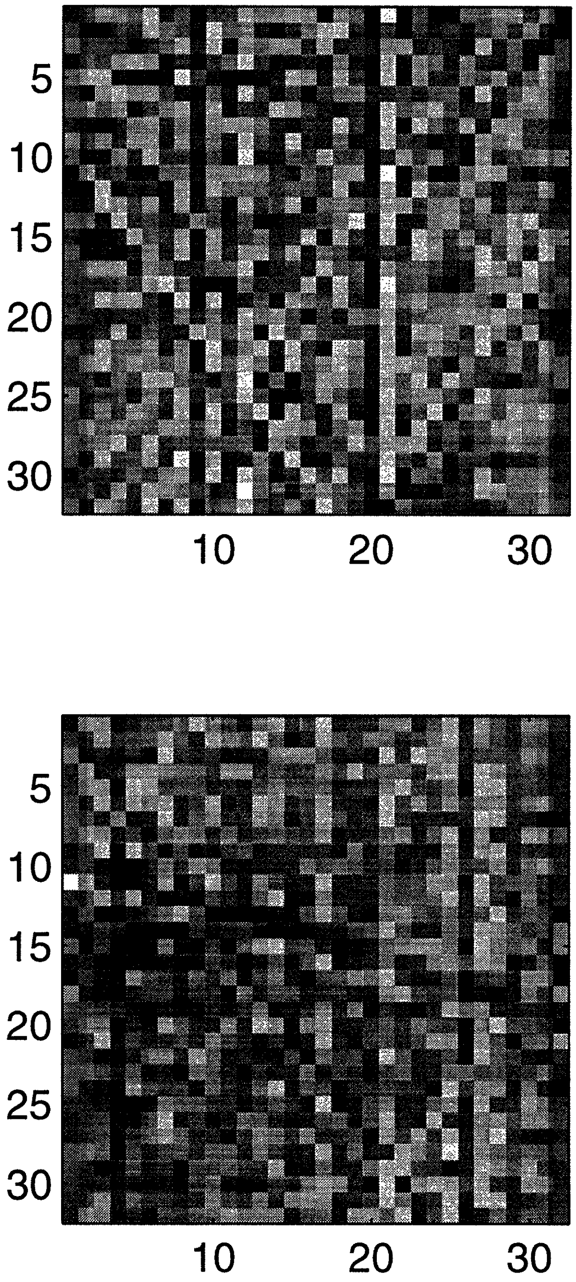


Figure 6-9: The filtered back-projection (FBP) reconstructions of the fractal fields shown in Figure 6-8 from the 5 dB SNR projection data. Top: Spectral parameter $\gamma_0 = 2$. Bottom: Spectral parameter $\gamma_1 = 3$. Note that it is not possible to discriminate between the fields from the reconstructions as the latter are mostly corrupted by sparse and noisy data artifacts.

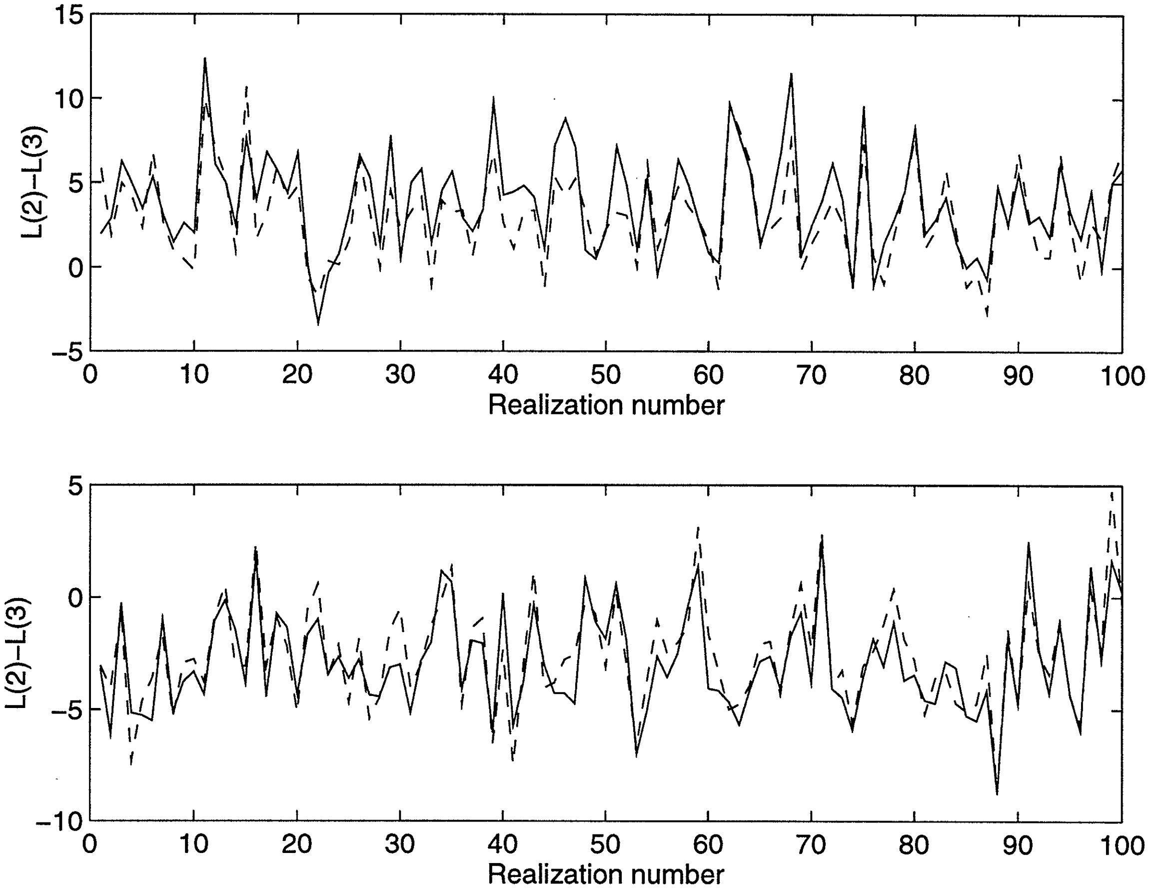


Figure 6-10: The exact likelihood difference $\mathcal{L}_m(\gamma = 2) - \mathcal{L}_m(\gamma = 3)$ (solid lines) and the approximate difference $\tilde{\mathcal{L}}_m(\gamma = 2) - \tilde{\mathcal{L}}_m(\gamma = 3)$ calculated on the tree (broken line). The fields are 32×32, the SNR = 5 dB, and an imaging geometry with $N_\theta = 5$ and $N_s = 32$ is assumed. Top: The underlying field is generated according to $\gamma = 2$. Bottom: The underlying field is generated according to $\gamma = 3$.

continuous-data Radon transform results and assuming that this transformation does indeed lead to a diagonal (i.e. whitened) covariance matrix. Before we show the ROC results, however, let us first summarize for clarity these various approaches for performing discrimination.

Recall that the exact likelihood discriminator is given by:

$$\begin{aligned} &\text{Decide } \gamma = 2 \text{ if: } \mathcal{L}_m(\gamma = 2) - \mathcal{L}_m(\gamma = 3) \geq \delta \\ &\text{else decide } \gamma = 3, \end{aligned} \tag{6.40}$$

and that the corresponding tree-based approximate discriminator is given by:

$$\begin{aligned} &\text{Decide } \gamma = 2 \text{ if: } \tilde{\mathcal{L}}_m(\gamma = 2) - \tilde{\mathcal{L}}_m(\gamma = 3) \geq \delta \\ &\text{else decide } \gamma = 3, \end{aligned} \tag{6.41}$$

where the exact likelihood $\mathcal{L}_m(\gamma_i)$ is obtained from (6.33) and the tree-based approximation $\tilde{\mathcal{L}}_m(\gamma_i)$ is obtained from (6.38).

The conventional discriminator [11] that is based on the slope of the averaged power spectra of the projections, is given by:

$$\begin{aligned} &\text{Decide } \gamma = 2 \text{ if: Weighted least squares estimate of the slope} \geq \delta \\ &\text{else decide } \gamma = 3, \end{aligned} \tag{6.42}$$

where, in this case, δ corresponds to the threshold slope parameter that is used to discriminate between the two fractal fields. Further, following the procedure outline in [11] for noisy data, we ignore a certain number of high frequency samples for calculation of the slope of the power spectra. We determine this number by trial and error and choose a value that results in the best ROC performance as we vary the threshold δ (i.e. in the maximum area under the ROC curve).

Finally, the discriminator that is based on the whitening transformation derived from the noise-free, continuous-data Radon transform results, is given by:

$$\begin{aligned} &\text{Decide } \gamma = 2 \text{ if: } \mathcal{L}_t(\gamma = 2) - \mathcal{L}_t(\gamma = 3) \geq \delta \\ &\text{else decide } \gamma = 3, \end{aligned} \tag{6.43}$$

where the likelihood $\mathcal{L}_t(\gamma_i)$ is calculated (c.f. (6.22)) in the transform domain by transforming the projection data by the matrix M_i the blocks of which represent the discretized $|w|^{(\gamma_i+1)/2}$ filter. Recall that this transformation results in an inexact whitening and hence the matrix D_i in (6.22) is not diagonal. In our examples, however, we ignore the off-diagonal elements in D_i for the calculation of the likelihood $\mathcal{L}_t(\gamma_i)$. This enables us to compare the accuracy of the whitening procedure obtained from the Radon transform results with our tree-based whitening technique.

In Figure 6-11 we show the receiver operating characteristic (ROC) curves for all four of the discrimination techniques that we just described. In these curves we have also included the so-called "guessing" line, the straight line from (PD,PF) = (0,0) to (1,1), corresponding to using only pure guessing strategies. The data for the ROC curves shown in Figure 6-11 are gathered with $N_\theta = 5$ and $N_s = 32$, and the additive noise in the data results in an SNR of 5 dB. Each point on the curve corresponds to a number of Monte Carlo simulation runs, with the number of runs chosen such that the estimates of PD and PF have relatively small standard deviation. To this end, we show in the figure the error bars that depict the 95% confidence intervals for the estimates of PD and PF. Note that these error bars are very small indicating that our PD and PF estimates are quite reliable. Finally, to enable an easy comparison of the various discrimination techniques, we plot all four ROC curves corresponding to the four different discrimination techniques on the same graph in Figure 6-12. For the sake of clarity, we do not show any error bars in this figure.

It is clear from Figures 6-11 and 6-12 that the two ROC curves corresponding to the optimal likelihood discrimination technique and our tree-based approximation are very similar. We can thus conclude that the gain in computational efficiency obtained by using the tree model is not offset by any significant loss in the ROC performance.

Further, the two figures also confirm the claims that we had made earlier namely that the discrimination techniques that are based on direct fractal dimension estimation or on whitening transformations developed for noise-free continuous-data Radon transform results break down in the case the data are sparse and noisy. In particular note that the ROC curves for both the conventional discrimination technique that is based on the slope of the power spectra, and the discrimination technique that is based on the whitening filter obtained from a discretization of Radon transform results, are only very slightly better than the guessing curve. On the other hand the likelihood based techniques perform significantly better with an area under the corresponding ROC curves that is very close to one. Further, in Table 6.1 we show the estimated spectral parameter values obtained from the conventional technique that is based on the slope of the power spectra of the projections. Note that while the actual spectral parameter values of the field are 2 and 3, the conventional technique incorrectly estimates these as 0.2515 and 0.3293 respectively. This again confirms that fact that the conventional technique fails to yield accurate results if the projection data are sparse and noisy.

Spectral parameter (Actual)	**Spectral parameter (Estimated)**
2	0.2515 ± 0.0062
3	0.3293 ± 0.0064

Table 6.1: The actual spectral parameter values for the field and the values estimated by the conventional method that is based on the slope of the averaged power spectra of the projections. The 10 highest frequency samples are ignored for the calculation of this slope. This number is determined by trial and error and is chosen to yield the maximum area under the ROC curve. The projection data are gathered according to $N_\theta = 5$ and $N_s = 32$, and the SNR is 5 dB.

We next consider the manner in which detection performance is affected by the strength of the noise corrupting the data. For these examples we again assume the projection data to be gathered according to $N_\theta = 5$ and $N_s = 32$. In Figure 6-13 we show (dashed line) a plot of the probability of discrimination error (PE) obtained by using our tree-based discrimination technique versus the signal-to-noise (SNR) ratio. We have assumed a value of $\delta = 0$ for the threshold parameter in the likelihood dis-

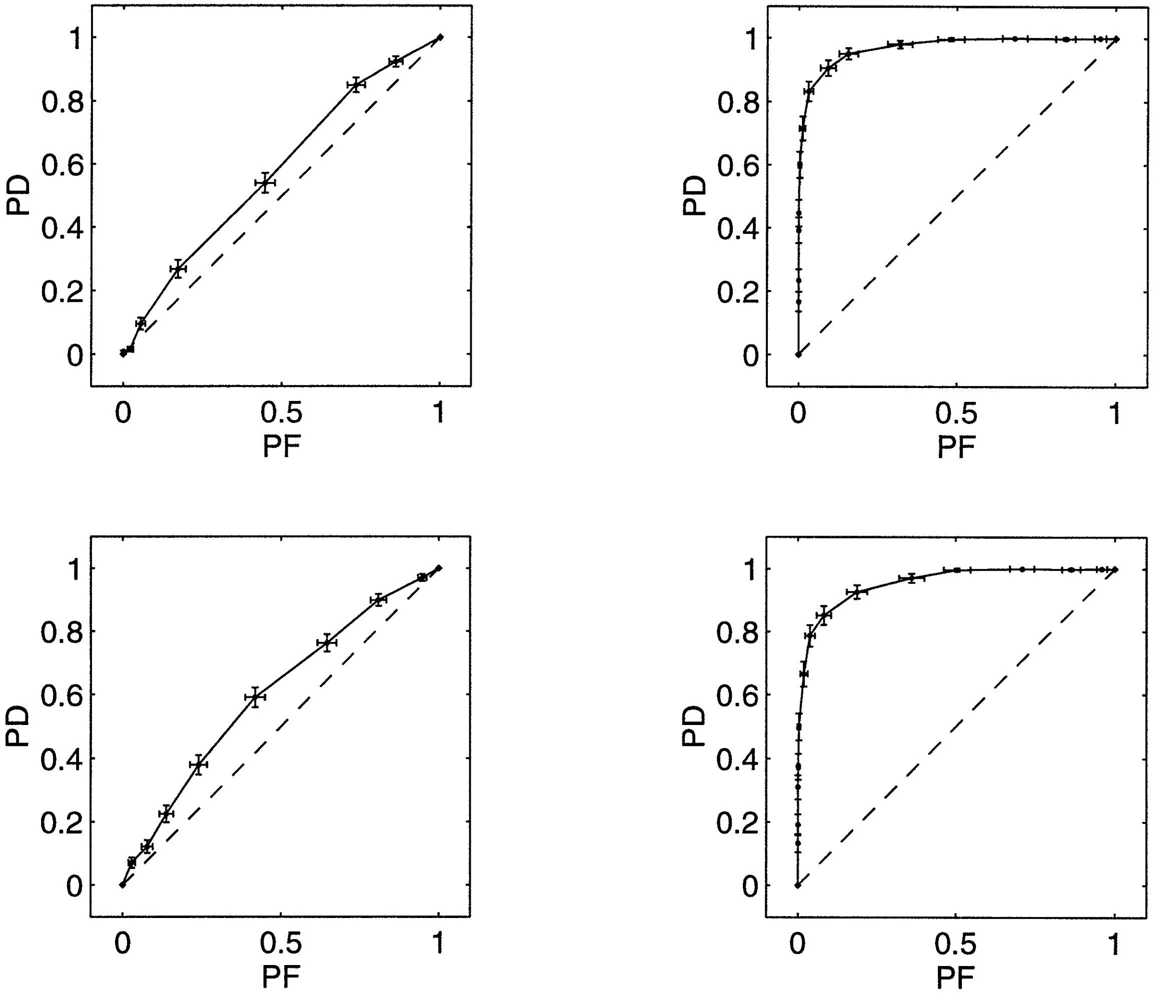


Figure 6-11: The ROC curves for the 32×32 example in Figure 6-10. The SNR = 5 dB and an imaging geometry with $N_\theta = 5$ and $N_s = 32$ is assumed. The error bars depict the 95% confidence intervals. The dashed lines in each of the subplots correspond to the guessing curve. Top right: Exact ROC curve. Bottom right: Approximate ROC curve using the tree model. Top left: The ROC curve obtained by the conventional discrimination technique that is based on the slope of the averaged power spectra of the projection data. The 10 highest frequency samples are ignored for the purpose of calculation of this slope. This number is determined by trial and error and is chosen to yield the maximum area under the ROC curve. Bottom left: The ROC curve obtained by using the whitening filter that is based on the noise-free continuous-data Radon transform results.

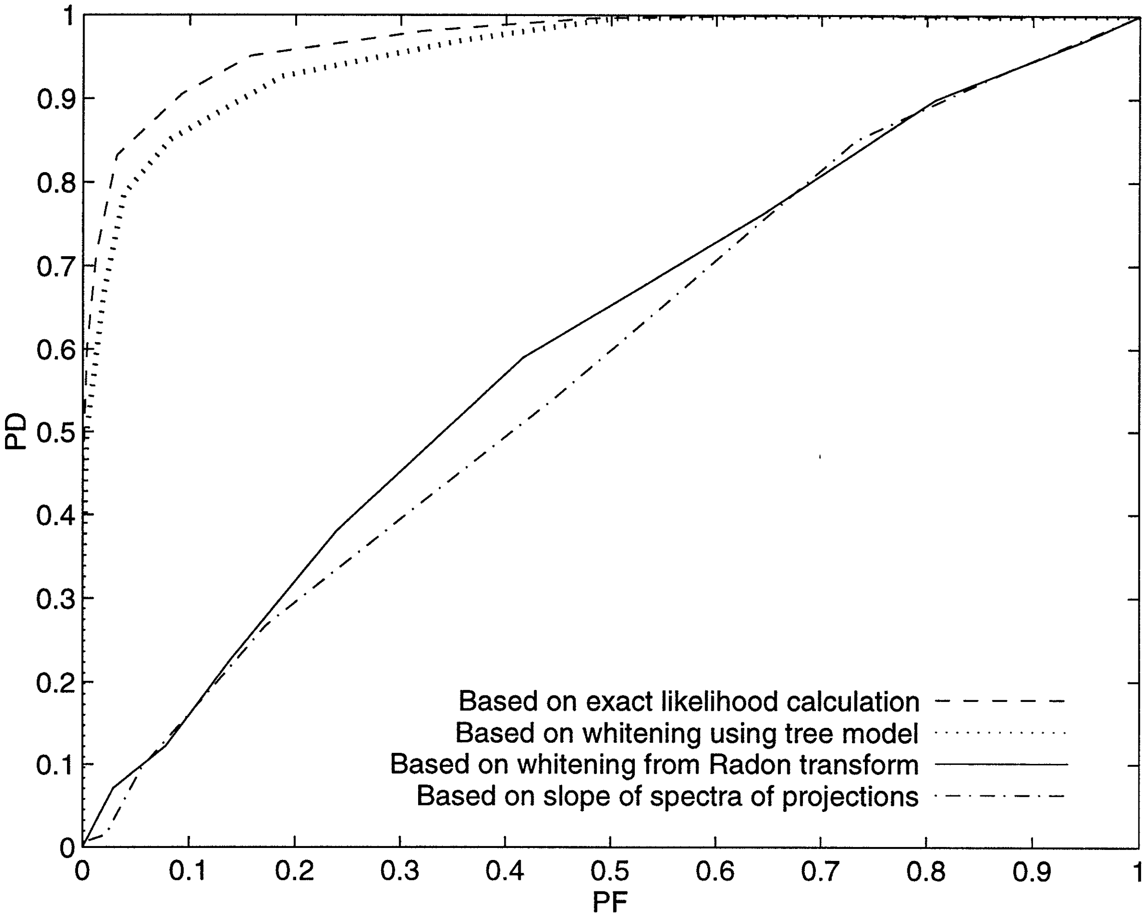


Figure 6-12: The ROC curves from Figure 6-11. The SNR = 5 dB and an imaging geometry with $N_\theta = 5$ and $N_s = 32$ is assumed. Dashed line: ROC curve for the exact likelihood test. Dotted line: Approximate ROC curve using the tree model. Dash-dotted line: The ROC curve obtained by the conventional discrimination technique that is based on the slope of the averaged power spectra of the projection data. Solid line: The ROC curve obtained by using the whitening filter that is based on the noise-free continuous-data Radon transform results.

crimination rule for these calculations. We also show in the same figure the probability of error versus SNR curve (dash-dotted line) obtained from the conventional discrimination technique that is based on the slope of the power spectra of the projections, and also the corresponding curve (solid line) obtained by using the whitening filter derived from the noise-free, continuous-data Radon transform results. It is clear from the figure that the performance of our tree-based discrimination technique even at an extremely low SNR value of -5 dB is better than that of both the other techniques at 5 dB.

So far we have only considered the discrimination of fractal fields with spectral parameter values of $(\gamma_0 = 2, \gamma_1 = 3)$. We present next in Figure 6-14 the plots of probability of error (PE) versus SNR for discrimination of fields with $(\gamma_0 = 2, \gamma_1 = 2.5)$ and also for fields with $(\gamma_0 = 2, \gamma_1 = 2.25)$. We use our tree-based likelihood discriminator with a threshold value $\delta = 0$ for these plots. In the same figure we repeat the corresponding plot from Figure 6-13 for discrimination of fields with $(\gamma_0 = 2, \gamma_1 = 3)$. Note from the figure that our discrimination technique is able to discriminate between the fields with spectral parameters $(\gamma_0 = 2, \gamma_1 = 2.5)$ from 10 dB SNR sparse $(N_\theta = 5, N_s = 32)$ data, with a probability of error that is only 0.13 (i.e. 13%). This is impressive especially since, as we show in Figure 6-15, it is difficult even to visually discriminate between these fields from the corresponding realizations. Finally, we point out that our tree-based method is able to discriminate between fields with very similar spectral parameter values $(\gamma_0 = 2, \gamma_1 = 2.25)$ with less errors than that obtained for the two suboptimal methods for spectral parameter values $(\gamma_0 = 2, \gamma_1 = 3)$ that are far apart.

6.7 Discussion

In this chapter we have considered the problem of discrimination of fractal fields from noisy and sparse projection data. The conventional techniques for solving such problems are based on Radon transform results that break down when the data are sparse and/or noisy. In order to avoid any restrictions on the quantity and the quality

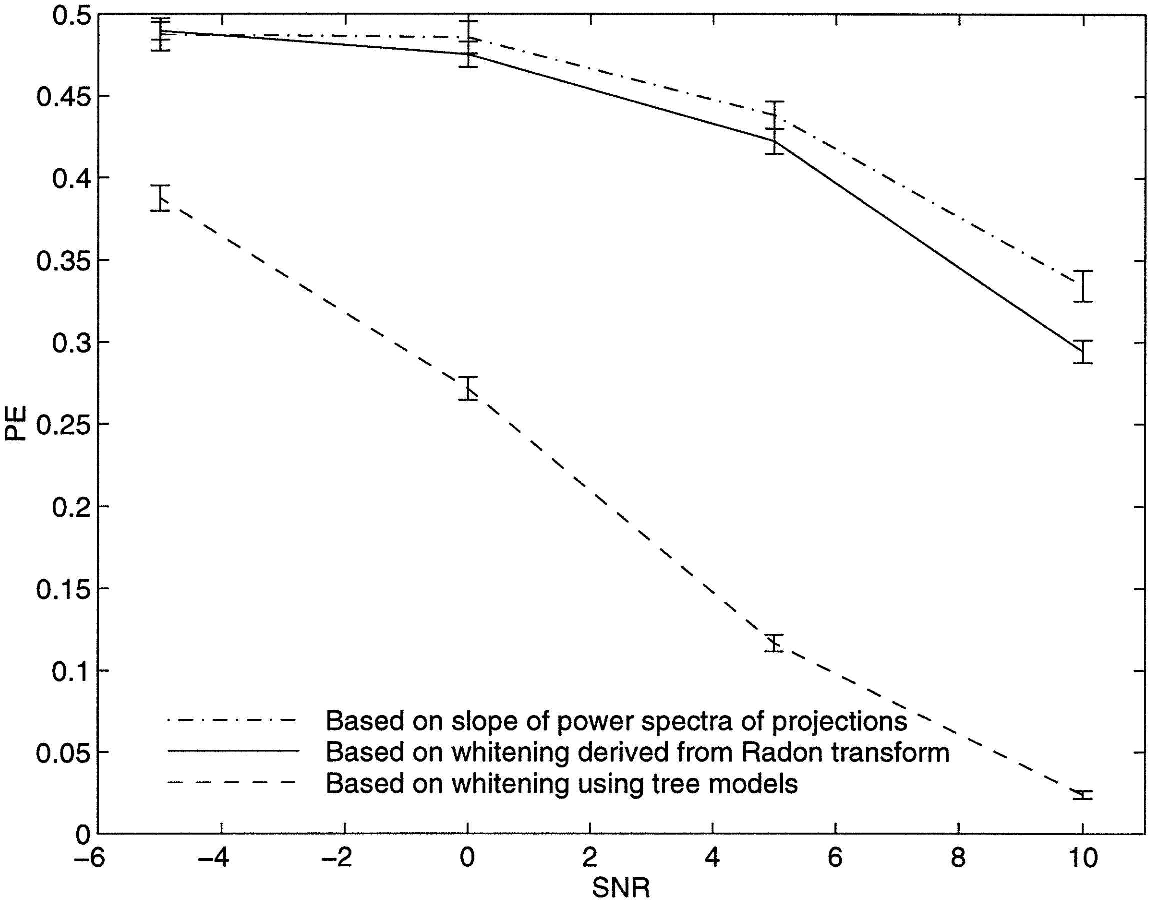


Figure 6-13: The probability of error (PE) variation with the signal-to-noise ratio (SNR) for discrimination of fields with spectral parameter values ($\gamma_0 = 2, \gamma_1 = 3$). The fields are 32×32 and an imaging geometry with $N_\theta = 5$ and $N_s = 32$ is assumed. The error bars depict the 95% confidence intervals. Dashed line: The tree-based likelihood discrimination technique. Dash-dotted line: The conventional discrimination technique that is based on the slope of the power spectra of the projections. Solid line: The discrimination technique that is based on the whitening filter derived from the noise-free, continuous-data Radon transform results.

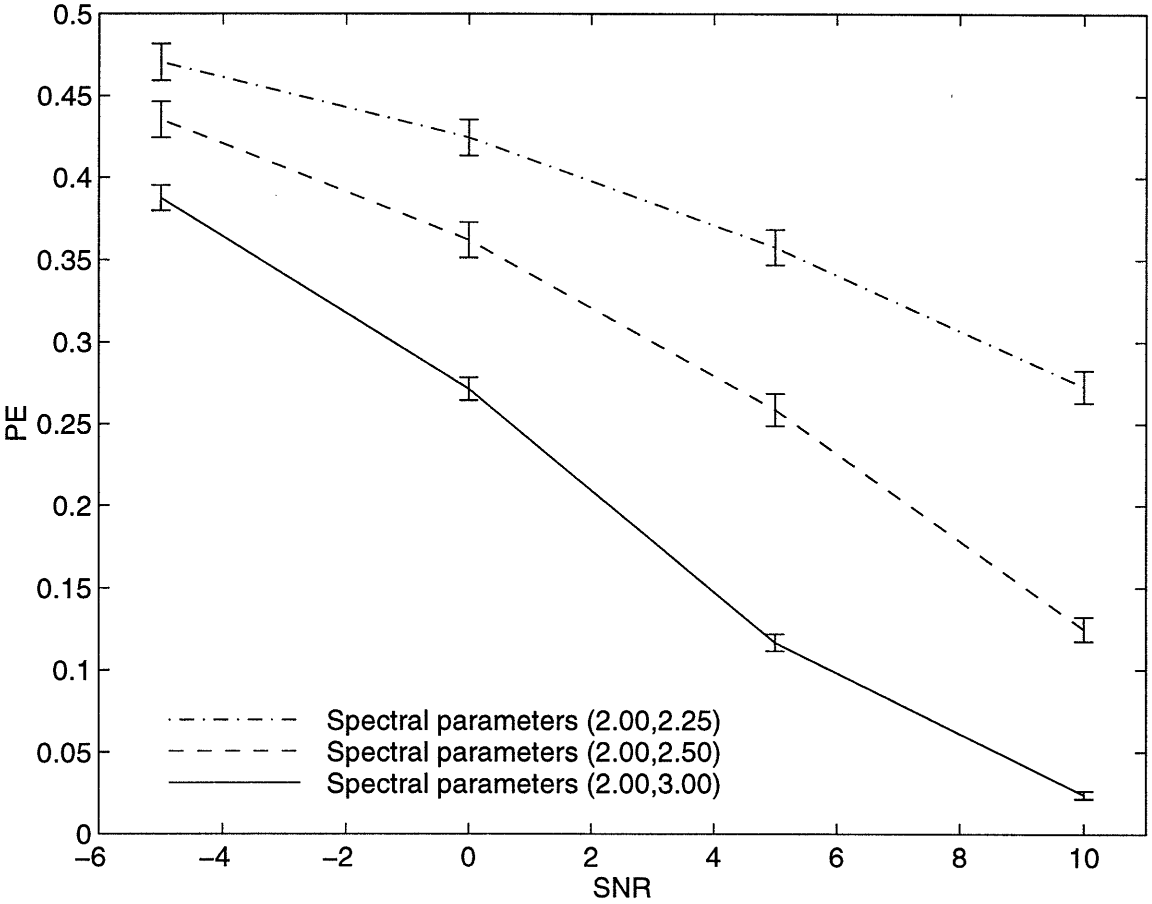


Figure 6-14: The probability of error (PE) variation with the signal-to-noise ratio (SNR) for the tree-based likelihood discrimination technique. The field is 32×32 and an imaging geometry with $N_\theta = 5$ and $N_s = 32$ is assumed. The error bars depict the 95% confidence intervals. Solid line: The two fields correspond to $(\gamma_0 = 2, \gamma_1 = 3)$. Dashed line: The two fields correspond to $(\gamma_0 = 2, \gamma_1 = 2.5)$. Dashed-dotted line: The two fields correspond to $(\gamma_0 = 2, \gamma_1 = 2.25)$.

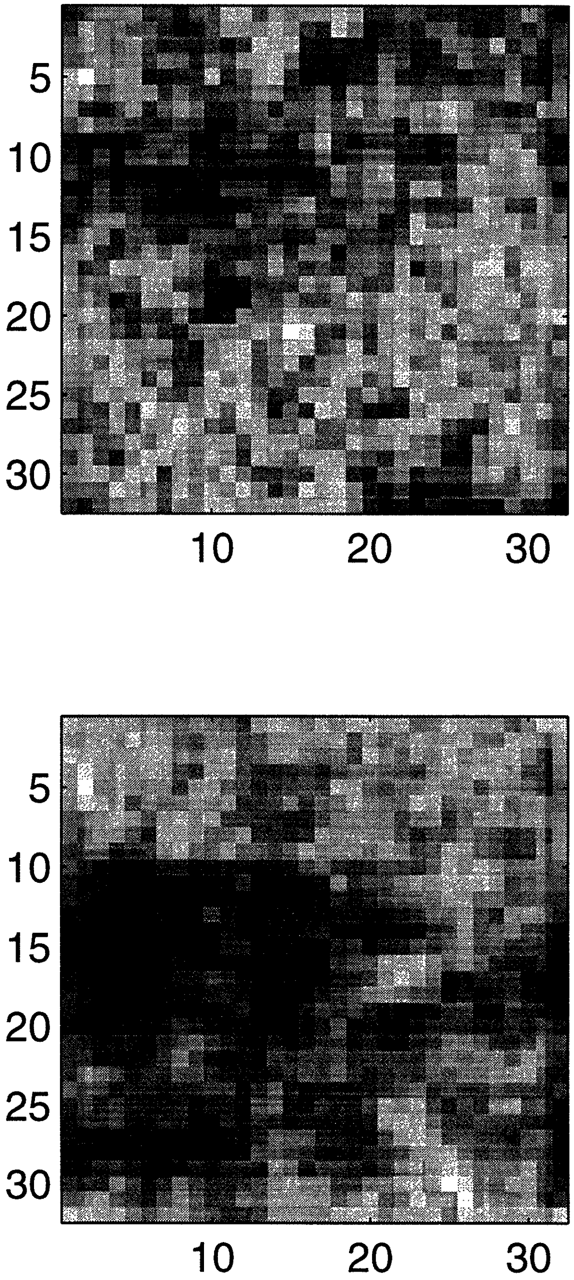


Figure 6-15: Realizations of 32 × 32 fractal fields. Top: Spectral parameter $\gamma_0 = 2$. Bottom: Spectral parameter $\gamma_1 = 2.5$.

of the projection data, we have formulated the discrimination problem in a hypothesis testing framework. The solution to this problem is then given by the statistically optimal maximum-log-likelihood discrimination rule.

The problem of discriminating fractal fields through likelihood calculations is, however, complicated by the fact that inverses and determinants of large and generally ill-conditioned fractal-field data covariance matrices are required. This ill-conditioning is a result of the redundancy in the tomographic data representation. We have shown that this redundancy in the covariance matrices is captured by the coarsest scale approximation block in the corresponding multiscale matrices, and hence can be effectively removed in the multiscale framework. Thus the likelihood calculations in the multiscale framework require just the inverses and the determinants of the well-conditioned detail blocks in the corresponding covariance matrices. However we have simplified our likelihood calculations even further by eliminating the need for explicit calculation of these quantities. We have achieved this by using the recently introduced class of multiscale stochastic models defined on trees to realize accurate approximations of the detail blocks of the covariance matrices. These tree-like models have the advantage that they lead to fast likelihood calculations, and thus to an efficient discrimination technique. Our tree-based discrimination technique is thus very close to optimal and at the same time is extremely computationally efficient.

6.A Details on the formation of the fractal-field covariance matrix Λ_i

Recall that the object $f(u,v)$ is discretized on a $N_s \times N_s$ pixel grid as follows:

$$f(u,v) = \sum_{i=1}^{N_s^2} f_i b_i(u,v), \tag{6.44}$$

where $b_i(u,v)$ is the indicator function of pixel i. We arrange the pixel coefficients f_i according to a lexicographic ordering scheme, which implies that the pixel (a_1, b_1) is represented by $i = (a_1 - 1)N_s + b_1$ in (6.44).

Let Λ denote the covariance matrix for the pixel coefficient vector f corresponding to a fractal field with spectral parameter γ. The correlation between the pixels (a_1, b_1) and (a_2, b_2) $(1 \leq (a_1, b_1, a_2, b_2) \leq N_s)$ is specified by the (m,n)-th element of Λ, where

$$\begin{aligned} m &= (a_1 - 1)N_s + b_1, & (6.45)\\ n &= (a_2 - 1)N_s + b_2. & (6.46) \end{aligned}$$

This element is obtained as:

$$\Lambda(m,n) = (\frac{1}{N_s^2}) \left[\sum_{(w_u,w_v)\in\mathcal{X}} \frac{1}{|w_u^2 + w_v^2|^{\gamma/2}} \exp\{j(w_u a + w_v b)\} \right] + (\frac{1}{N_s^2})(\frac{N_s}{2\pi})^{\gamma}, \tag{6.47}$$

where

$$\begin{aligned} a &= a_1 - a_2, & (6.48)\\ b &= b_1 - b_2, & (6.49)\\ w_u &= (\frac{2\pi}{N_s})c_u, & (6.50)\\ w_v &= (\frac{2\pi}{N_s})c_v, & (6.51) \end{aligned}$$

and

$$\mathcal{X} = \left\{(c_u, c_v) \mid c_u = -\frac{N_s}{2}+1, \ldots, \frac{N_s}{2};\ c_v = -\frac{N_s}{2}+1, \ldots, \frac{N_s}{2}\right\} \setminus \{(0,0)\}. \quad (6.52)$$

6.B Details on the generation of a fractal field

In order to generate a fractal field with spectral parameter γ, we start by calculating $\Lambda^{1/2}$, the symmetric square root of the covariance matrix Λ for the pixels corresponding to the fractal field:

$$\Lambda = \Lambda^{1/2}\Lambda^{1/2}. \quad (6.53)$$

Once $\Lambda^{1/2}$ is determined the vector f, which contains the pixel coefficients of the fractal field, is obtained as follows:

$$f = \Lambda^{1/2}\epsilon, \quad (6.54)$$

where ϵ is a vector containing the samples of a white, Gaussian distribution with zero mean and unit variance, i.e.

$$\epsilon \sim \mathcal{N}(0, I_{N_s^2}). \quad (6.55)$$

6.C Whitening transformation for continuous and noise-free fractal-field tomographic data

Here we find a whitening transformation for the projection data corresponding to a fractal field f with spectral density of the form:

$$S_f(w_u, w_v) = \frac{1}{|w|^\gamma}, \quad (6.56)$$

where γ is the spectral parameter, and we have assumed a value of 1 for the constant σ_f^2 (c.f. (6.9)). We restrict ourself here to only noise-free and continuous tomographic data. In this case there exists previous work [42,43] that finds the whitening transfor-

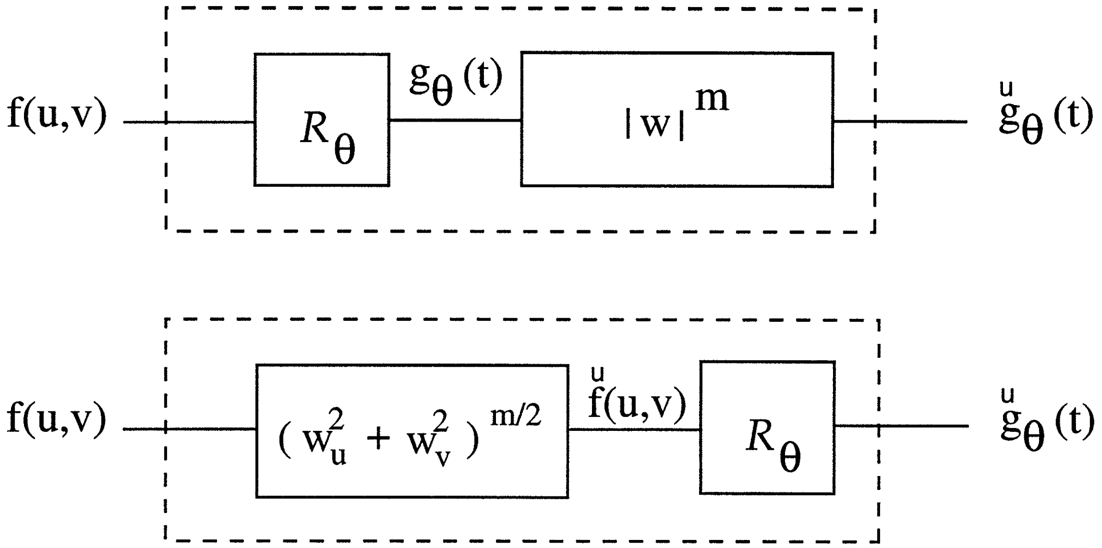


Figure 6-16: The two filtering operations described by the above two block diagrams are identical.

mation corresponding to a white noise object (i.e. an object with spectral parameter $\gamma = 0$). In the following we generalize the result in [42, 43] to arbitrary spectral parameter values.

Let us define $\mathcal{R}_\theta$ as the projection operator at angle θ such that the operation $\mathcal{R}_\theta f(u, v)$ results in $g_\theta(t)$, the projection of f at angle θ. Our goal is to find a filter with a frequency response of the type $|w|^m$ that whitens the projections $g_\theta(t)$ corresponding to a fractal field with spectral parameter γ. Obviously one can expect the filter parameter m to be related to γ and in what follows we establish a precise relationship between these two quantities. To begin, consider the two filtering operations described by the two block diagrams in Figure 6-16. These operations are identical because of a Radon transform result that states that the projection of a convolution of two 2-D functions is just the convolution of their individual projections. Now, if we denote the covariance of the fractal field f as Λ and the covariance of the filtered field $\check{f}$ as $\check{\Lambda}$, then:

$$\begin{aligned} E[\check{f}(u,v)\check{g}_\theta(t)] &= \int\!\!\int E[\check{f}(u,v)\check{f}(u',v')]\delta(u'\cos\theta + v'\sin\theta - t)\,du'\,dv' \\ &= \int\!\!\int \check{\Lambda}(u-u', v-v')\delta(u'\cos\theta + v'\sin\theta - t)\,du'\,dv' \\ &= \check{\Lambda}_r(u\cos\theta + v\sin\theta - t, \theta), \qquad (6.57) \end{aligned}$$

where the subscript "r" in $\breve{\Lambda}_r$ denotes the use of radial coordinates, and the last equality in (6.57) follows from the central slice theorem [44]. The correlation between the filtered projection data is given by:

$$\begin{aligned}
E[\breve{g}_\theta(t)\breve{g}_{\theta'}(t')] &= E[\breve{g}_{\theta'}(t')\mathcal{R}_\theta \breve{f}(u,v)] \\
&= \int\!\!\int E[\breve{g}_{\theta'}(t')\breve{f}(u,v)]\delta(u\cos\theta + v\sin\theta - t)\,du\,dv \\
&= \int\!\!\int \breve{\Lambda}_r(u\cos\theta' + v\sin\theta' - t', \theta')\delta(u\cos\theta + v\sin\theta - t)\,du\,dv,
\end{aligned} \tag{6.58}$$

where the last equality in (6.58) follows from (6.57). Finally, if S_f^r is the spectral density of the field f in radial coordinates and $S_{\breve{f}}^r$ is the same for the filtered field $\breve{f}$, then from Figure 6-16:

$$S_{\breve{f}}^r(w,\theta) = |w|^{2m} S_f^r(w,\theta), \tag{6.59}$$

which implies that:

$$\breve{\Lambda}_r(r,\theta) = \int |w|^{2m} S_f^r(w,\theta)\exp(jwr)\,dw. \tag{6.60}$$

Now by substituting (6.60) in (6.58) we obtain:

$$\begin{aligned}
&E[\breve{g}_\theta(t)\breve{g}_{\theta'}(t')] = \\
&\int\!\!\int\!\!\int du\,dv\,dw |w|^{2m} S_f^r(w,\theta)\exp(jw(u\cos\theta' + v\sin\theta' - t')) \times \\
&\delta(u\cos\theta + v\sin\theta - t) = \\
&\int |w|^{2m} S_f^r(w,\theta')\exp(-jwt')dw \times \\
&\left[\int\!\!\int \exp(jw(u\cos\theta' + v\sin\theta'))\delta(u\cos\theta + v\sin\theta - t)\,du\,dv\right].
\end{aligned} \tag{6.61}$$

The expression in square brackets in (6.61) is the Radon transform of a plane wave and can be shown to be equal to [42, 43]:

$$\exp(jwt)\frac{\delta(\theta - \theta')}{|w|}. \tag{6.62}$$

By substituting (6.62) in (6.61) we obtain:

$$E[\breve{g}_\theta(t)\breve{g}_{\theta'}(t')] = \delta(\theta - \theta') \int |w|^{2m-1} S_f^r(w, \theta) \exp(jw(t - t'))\, dw. \tag{6.63}$$

Thus the filtered projections at different angles are uncorrelated and the spectral density $S_{\breve{g}}$ of the filtered projection $\breve{g}_\theta(t)$ is given by:

$$S_{\breve{g}}(w, \theta) = |w|^{2m-1} S_f^r(w, \theta). \tag{6.64}$$

Now if the spectral density of the field satisfies a $1/f$ power law with spectral parameter γ (c.f. (6.56)) then from (6.64) the projection data may be whitened by a filter of the type $|w|^m$ where:

$$m = \frac{\gamma + 1}{2}. \tag{6.65}$$

Chapter 7

Detection Of Anomalies From Projection Data

7.1 Introduction

In this chapter we consider a commonly encountered application in tomography, in which the goal is not to reconstruct the entire field, but to simply detect the presence of an anomaly. Examples of such an application include the detection of cracks and other defects in materials [25,73], and the detection of lesions in medical images [30].

The conventional methods for detecting anomalies from projection data fall under two broad categories. In the first category, the entire field is first reconstructed using any of the conventional techniques described in Chapter 2, and then post-processing is done on the reconstructed image to detect the anomaly (for example, as in [30]). Leaving the computational issues aside, a major disadvantage of this approach is that when the projection data are sparse and/or noisy, the reconstruction of a complete image generally suffers from severe artifacts and distortion that can easily obscure the anomaly or result in false alarms. These reconstruction artifacts can be removed by appropriate regularization at the expense of additional computational burden. This approach can, however, smooth out the anomaly and hence once again make it undetectable. In the second category of detection methods [66], the problems associated with the reconstruction of the entire field are avoided by developing detection techniques directly in the data-domain. These methods, however, assume that the anomalies are parameterized (for example by size, shape, location, and number) and thus are limited in scope.

In order to avoid the above problems that are associated with detection of anomalies through the use of conventional techniques or the limitations in previous direct detection formulations, we propose here a procedure that detects anomalies directly from the projection data and, in addition, does not impose any restrictive assumptions on the anomaly. Specifically, in this chapter we develop a fast data-domain technique that detects anomalies in textured backgrounds. In particular we use $1/f$ fractal processes to model the background textures. These processes, as mentioned in the previous chapters, are accurate models for a variety of textures including terrain and several biological structures [11, 12, 58, 78, 80]. As a result, the specific detection task that we consider here is sufficiently general to encompass a wide variety of applications, from detection of lesions in medical images to detection of defects in materials.

As mentioned previously, our technique for detection of anomalies is based directly in the projection, i.e. in the data domain. As a first step of our data-domain approach we obtain the statistics for the projection data that corresponds just to the background texture. We then analyze the data to see whether it deviates significantly from these statistics. This analysis is based on a chi-square test and enables us to compute the confidence level at which an anomaly can be identified in the input field.

There are a few obstacles, however, to this data-domain detection approach that must be dealt with if we are to obtain an effective algorithm. The first issue is that without some additional structure, this technique lets us test only for the presence of an anomaly but does not enable us to localize it. The second major obstacle is that the chi-square test requires the inverse of the fractal-field data covariance matrix which, as discussed in Chapter 6 is large, full and, in general, quite ill-conditioned. Recall that we were faced with a similar difficulty in Chapter 6 where the likelihood calculations for the discrimination of fractal-fields required the inversion of the same covariance matrices. There, however, we were able to simplify things by a transformation to the multiscale framework, which not only resulted in sparse transformed covariance matrices but also that any ill-conditioning in the matrices was concentrated in the corresponding coarsest scale approximation blocks.

Along similar lines, we are able to remove the second obstacle mentioned in the preceding paragraph by transforming the data to the multiscale framework and by performing the chi-square test on the multiscale data. This transformation results in an extremely sparse multiscale covariance matrix which we are able to partition by scales in such a manner that any ill-conditioning in the matrix is concentrated in the coarsest scale approximation block. Further, by using the fact that the multiscale wavelet basis functions are localized and that the multiscale fractal-field covariance matrix is extremely sparse, and in particular is diagonally dominant, we are able to remove the first issue with the detection approach as well. In particular due to the sparse nature of the multiscale covariance matrix the various elements in the multiscale data vector are effectively uncorrelated. This, and the fact that the wavelet basis functions are localized implies that the presence of an anomaly in a particular region of the field can be established by performing a chi-square test on only those transformed data samples which correspond to the multiscale strips that intersect the test region. Further this chi-square computation is trivial since the data have been effectively uncorrelated by the multiscale transformation. Finally, after localizing the anomaly to a few candidate pixels based on the chi-square test, we go back and obtain an ML estimate of the field intensity values at only these candidate pixel locations. However, in order to obtain this ML estimate we once again take advantage of the localization property of the wavelets and consider only the data which intersect the candidate pixels. Assuming that the size of the anomaly is sufficiently small, the number of candidate pixels are a small fraction of the size of the total input field. This implies that the size of the ML estimation problem is quite small and thus the corresponding solution can be obtained in an efficient manner. The above steps together result in a highly efficient algorithm for detecting anomalies in the multiscale data-domain.

The various steps of our detection algorithm can be summarized as follows. To begin we compute the multiscale transformation of the data. We then calculate and store the contributions of each one of these data samples to the total chi-square value by assuming the multiscale data to be uncorrelated, i.e. by ignoring the off-diagonal elements in the fractal-field data covariance matrix. After this we cycle through all the pixels in the field and test to see if these are anomalous. This test is based on the approximate chi-square value that is obtained simply and efficiently by adding together the contributions from only that data which correspond to the multiscale strips that intersect the pixel. After this first pass is completed we go back and obtain an ML estimate of the intensity values of the field at only those pixel locations where, based on the results of the first pass, there is a high probability of finding an anomaly. These steps result in an extremely efficient data-domain technique for detecting anomalies in fractal textured backgrounds.

The chapter is organized as follows. In Section 7.2 we develop the theory behind our efficient multiscale discrimination technique. Then in Section 7.3 we present a few examples that demonstrate the efficacy of our detection approach. Section 7.4 concludes this chapter.

7.2 Theory

We assume that the field that is input to the imaging system consists of a fractal-textured background on which one or more anomalies may be superimposed. We characterize the anomaly by the overall region Ω_A, and for simplicity we assume an approximate knowledge of the size of this anomalous region as well as a knowledge of the minimum anomaly intensity. In practice one may have an idea of the location and the shape of the anomaly as well, and this prior information can be incorporated into the detection method to improve performance. However, for the detection method that we develop in this chapter as well as the examples that we present in the next section, we impose no restrictions on the location or the shape of the anomaly.

Specifically, let $f(u,v)$ represent a $1/f$ background fractal-field with fractal dimension D or with spectral parameter γ (c.f. Chapter 6). Further let $f_a(u,v)$ represent an anomaly such that:

$$f_a(u,v)\begin{cases} \geq E & \text{for } (u,v) \in \Omega_A \\ = 0 & \text{otherwise} \end{cases} \tag{7.1}$$

where Ω_A represents the region that contains the anomaly, and E is the minimum anomaly intensity. The field f_o that is input to the imaging system can now be written as a superposition of the background and the anomaly as follows:

$$f_o(u,v) = f(u,v) + f_a(u,v). \tag{7.2}$$

After discretization of both the background field and the anomaly on a $N_s \times N_s$ pixel grid, where recall that N_s is the number of strips in each angular projection, the discretized input field can be written in terms of the corresponding pixel coefficients as follows:

$$f_o = f + f_a. \tag{7.3}$$

The vector f that contains the background fractal-field pixels is distributed according to:

$$f \sim \mathcal{N}(0, \Lambda), \tag{7.4}$$

and the k-th element of vector f_a containing the anomalous pixels, is given by:

$$f_a(k)\begin{cases} \geq E & \text{for } k \in \Psi_A \\ = 0 & \text{otherwise} \end{cases} \tag{7.5}$$

where the set Ψ_A represents the collection of all pixels that are anomalous.

The goal of this chapter is to develop a method for detection, localization, and estimation of anomalies from projection data. To this end, the noisy projection (i.e. the strip-integral) data y corresponding to the input field f_o are given by:

$$y = Tf_o + n, \qquad\qquad n \sim \mathcal{N}(0, \lambda I_{N_\theta N_s}), \tag{7.6}$$

where T is the projection matrix the rows of which contain the discretized strips, n is the zero-mean, additive, white Gaussian noise in the data, and a parallel-beam imaging geometry is assumed with N_θ angular projections and N_s strips in each angular projection.

7.2.1 Anomaly detection

The anomaly detection problem can be formulated as follows. *Given the data y corresponding to an anomaly buried in a textured background, find the set of pixels Ψ_A that are anomalous.* However before we embark on solving this problem, we will attempt to solve a different problem instead, namely *given the data y decide if there is any anomaly buried in the input field f_o.* Once we have obtained a method for efficiently solving this second problem, we will extend it to find a solution to the set of anomalous pixels Ψ_A in the input field f_o. The second detection problem can be restated as, *given the data y find the confidence level at which the following hypothesis that there is no anomaly in the input field, can be rejected*:

$$H_0 : \qquad f_o = f. \tag{7.7}$$

We next develop an efficient method that determines the confidence level at which we can reject H_0 based on the noisy projection data y.

To begin, note that under the hypothesis H_0 (c.f. (7.7)) that there is no anomaly in the input field f_o, the noisy projection data y satisfy:

$$H_0 : \qquad y = Tf + n. \tag{7.8}$$

Thus in the case that H_0 is true, the data vector y obeys the following statistics:

$$p(y|H_0) \sim \mathcal{N}(m_{y|H_0}, \Lambda_{y|H_0}), \tag{7.9}$$

where the mean $m_{y|H_0}$ and the variance $\Lambda_{y|H_0}$ are given by:

$$\begin{aligned} m_{y|H_0} &\triangleq E[y|H_0] && (7.10) \\ &= 0, && (7.11) \\ \Lambda_{y|H_0} &\triangleq E[yy^T|H_0] && (7.12) \\ &= \left(T\Lambda T^T + \lambda I_{N_\theta N_s}\right). && (7.13) \end{aligned}$$

The presence of an anomaly can thus be established by analyzing the data y to see if there is an appreciable deviation from the statistics (7.9), (7.11) and (7.13). One particular measure of this deviation is obtained by the chi-square (χ^2) value [35]:

$$\begin{aligned} \chi^2 &\triangleq (y - m_{y|H_0})^T \Lambda_{y|H_0}^{-1} (y - m_{y|H_0}) && (7.14) \\ &= y^T \Lambda_{y|H_0}^{-1} y. && (7.15) \end{aligned}$$

The expected χ^2 value (7.15) given that the hypothesis H_0 is true (i.e. there is no anomaly in the input field f_o) is equal to:

$$\begin{aligned} E[\chi^2|H_0] &= E\left[y^T \Lambda_{y|H_0}^{-1} y\right] \\ &= \text{Trace}(I_{N_\theta N_s}) \\ &= N_\theta N_s. && (7.16) \end{aligned}$$

In addition, it can be shown [35] that under the assumption that $N_\theta N_s$ is not too small[1], the χ^2 value given that the hypothesis H_0 is true (i.e. there is no anomaly in the input field f_o) obeys the following probability law:

$$\sqrt{2\chi^2_{H_0}} - \sqrt{2N_\theta N_s} \sim \mathcal{N}(0,1). \qquad (7.17)$$

The probability law (7.17) enables us to calculate the confidence level at which H_0

---

[1] The result $\sqrt{2\chi^2} - \sqrt{2N} \sim \mathcal{N}(0,1)$ holds as long as the number N of degrees of freedom is greater than around 20 [35]. In the case that N is less than this number, the standard chi-square tables can be used instead to determine the confidence level.

can be rejected according to the following steps. First the χ^2 value is calculated from the data y by using the expression (7.15). Then the confidence level is determined from (7.17) by using the standard normal probability density tables.

The above procedure for finding the confidence level at which H_0 can be rejected is, however, complicated by the fact that the χ^2 expression (7.15) requires the inverse of the covariance matrix $\Lambda_{y|H_0}$. Recall from the discussion and examples in Chapter 6 that the fractal-field data covariance matrix $\Lambda_{y|H_0}$ is large, full, and is generally ill-conditioned due to the ill-conditioned nature of the projection matrix T. However as in Chapter 6, we can avoid the complications in the inversion of $\Lambda_{y|H_0}$ (c.f. (7.15)) by a transformation of the data to the multiscale framework. As we know from the results of Chapter 6, the multiscale transformed version of the data covariance matrix is sparse and, in addition, can be naturally partitioned by scales such that any ill-conditioning is captured by the coarsest scale approximation block. To this end, let W be an $N_s \times N_s$ matrix which when applied to a 1-D vector of length $N_s = 2^J$, results in a vector containing the wavelet representation of the original 1-D vector (c.f. Section 2.5)[2]. The matrix W represents the choice of a specific wavelet and scaling function that are used for multiscale representation. We assume these to be orthonormal which results in W being orthogonal, i.e. $W^{-1} = W^T$. If we define a block diagonal matrix W_b consisting of N_θ blocks along the diagonal, all equal to W, then W_b is also orthogonal, with $W_b^{-1} = W_b^T$. This matrix W_b when applied to the $N_\theta N_s$ data vector y results in a wavelet representation of all N_θ components of the vector as follows:

$$\eta = W_b y. \tag{7.18}$$

The multiscale vector η can be rearranged in scales and partitioned into detail and approximation blocks as follows:

$$\left[\begin{array}{c} \eta^a \\ \hline \eta^d \end{array} \right], \tag{7.19}$$

where the N_θ vector η^a contains the coarsest scale approximation terms from data

---

[2] Recall that the wavelet representation operation can be performed in an extremely efficient manner. In particular, representation of a vector of length N_s can be obtained in $O(N_s)$ multiplications.

at all angles, and the $N_\theta(N_s-1)$ vector η^d contains the corresponding multiscale detail terms. The χ^2 value in the multiscale data-domain reduces to the following expression:

$$\chi^2 = (\eta - m_{\eta|H_0})^T \Lambda_{\eta|H_0}^{-1} (\eta - m_{\eta|H_0}), \tag{7.20}$$

where $m_{\eta|H_0}$ and $\Lambda_{\eta|H_0}$ are the mean and the variance, respectively, of the multiscale data given that the hypothesis H_0 is true, i.e. that there is no anomaly present in the input field f_o. These are given by:

$$\begin{aligned} m_{\eta|H_0} &= W_b m_{y|H_0} = 0, && (7.21) \\ \Lambda_{\eta|H_0} &= W_b \Lambda_{y|H_0} W_b^T. && (7.22) \end{aligned}$$

The multiscale data covariance matrix is sparse and, in addition, can be partitioned into detail and approximation blocks as follows:

$$\begin{aligned} \Lambda_{\eta|H_0} &= \left[\begin{array}{c|c} E[\eta^a(\eta^a)^T] & E[\eta^a(\eta^d)^T] \\ \hline E[\eta^a(\eta^d)^T]^T & E[\eta^d(\eta^d)^T] \end{array}\right] && (7.23) \\ &\approx \left[\begin{array}{c|c} \Lambda_a & 0 \\ \hline 0 & \Lambda_d \end{array}\right], && (7.24) \end{aligned}$$

where any ill-conditioning in $\Lambda_{\eta|H_0}$ is concentrated in the $N_\theta \times N_\theta$ coarsest scale approximation block Λ_a and thus the $N_\theta(N_s-1) \times N_\theta(N_s-1)$ detail block Λ_d is relatively well-conditioned.

The χ^2 expression (7.20) in the multiscale domain can now be written in terms of the contribution from the detail term χ_d^2 and the approximation term χ_a^2 as follows:

$$\begin{aligned} \chi^2 &= (\eta^a)^T \Lambda_a^+ \eta^a + (\eta^d)^T \Lambda_d^{-1} \eta^d \\ &\triangleq \chi_a^2 + \chi_d^2, \end{aligned} \tag{7.25}$$

where Λ_a^+ refers to the pseudo-inverse of matrix Λ_a which is required here since the approximation block Λ_a is generally ill-conditioned[3]. Thus the transformation to

---

[3] Note that under both hypotheses, whether an anomaly is present in the input field or not, η^a

the multiscale framework lets us effectively overcome the problem of inverting an ill-conditioned matrix, and all that is required for this is the pseudo-inverse of the small $(N_\theta \times N_\theta)$ matrix Λ_a. Also note that since Λ_d is sparse and relatively well-conditioned, the detail contribution χ_d^2 to χ^2 can be calculated with very little computational complexity by first solving the sparse and well-conditioned system of equations $\Lambda_d z = \eta^d$ and subsequently by forming the product $(\eta^d)^T z$. Finally once the χ^2 value has been obtained from (7.25), the confidence level at which the hypothesis H_0 can be rejected is obtained through the use of the relationship (7.17) and the standard normal probability density tables.

To summarize, so far we have developed an efficient, multiscale-based, detection scheme that determines if the input field contains *any* anomaly. This scheme, however, does not tell us anything about the location of the anomalous pixels. In what follows next we take advantage of the sparsity of the fractal-field data covariance matrix $\Lambda_{\eta|H_0}$ in the multiscale domain as well as the fact that the wavelets are spatially localized to propose an extremely efficient technique for finding the estimate $\widehat{\Psi}_A$ of the set of anomalous pixels Ψ_A.

7.2.2 Anomaly localization and estimation

We first propose a technique that lets us decide to a certain confidence level whether there is any anomaly in a certain test region Ψ_T of the input field. To this end, note that the presence of an anomaly in the region Ψ_T will only affect that portion y_T of the data that correspond to the strips which intersect the region, and will have no effect on the rest of the data y_R. Thus we can partition the data vector y and the covariance matrix $\Lambda_{y|H_0}$ such that the χ^2 can be written in terms of contribution from y_T and y_R as follows:

$$\chi^2 \;=\; y^T \Lambda_{y|H_0}^{-1} y$$

---

does not have a component in the null space of Λ_a since the null space arises solely due to the fact that there is a non-uniqueness in the tomographic data representation in the sense that the DC component is the same in all angular projections. Thus a replacement of Λ_a^{-1} by Λ_a^{+} (c.f. (7.25)) is justified.

$$= \left[\begin{array}{c|c} y_T^T & y_R^T \end{array}\right] \left[\begin{array}{c|c} \Xi_T & \Xi_{TR} \\ \hline \Xi_{TR}^T & \Xi_R \end{array}\right]^{-1} \left[\begin{array}{c} y_T \\ \hline y_R \end{array}\right]. \tag{7.26}$$

Now if the coupling Ξ_{TR} between y_T and y_R were negligible, then the method for testing the presence of an anomaly in the region Ψ_T would simply amount to determining the confidence level from the following chi-square calculation:

$$y_T^T \Xi_T^{-1} y_T, \tag{7.27}$$

which represents the contribution from that data which corresponds to just the test region. Recall however that the covariance matrix $\Lambda_{y|H_0}$ is full, which in turn implies that the coupling Ξ_{TR} is appreciable. Due to this coupling, even though an anomaly in the region Ψ_T only affects y_T, the contribution from the rest of the data y_R can not be ignored for chi-square calculation. Thus it is not possible in the original data-domain to localize the detection of an anomaly in Ψ_T to a comparably-sized portion of the data vector in its original basis.

The preceding discussion however provides us with a clue as to how a transformation to the multiscale framework can help us with this localization. First the presence of an anomaly in the test region Ψ_T will only affect those multiscale data η_T corresponding to natural wavelet bases intersecting the test region. Since the natural wavelet bases are of finite extent, there are only a few of these bases that intersect this test region (assuming that the test region is comparatively small compared to the overall image dimensions). Further, since the multiscale covariance matrix $\Lambda_{\eta|H_0}$ is sparse, there is negligible coupling between the data which intersect the test region and the rest of the multiscale data. This implies that the presence of an anomaly in the test region Ψ_T can be established by calculating the chi-square value for only that data which intersect the test region. We next make precise the preceding intuitive ideas on localization of anomalies in the multiscale framework.

To begin, from now on, after having determined the presence or absence of an anomaly in the entire input field f_o, we ignore the coarsest scale approximation terms for the purpose of localizing the anomaly. These terms provide no help in localizing the anomaly since these correspond to the coarsest scale approximation strips that cover the entire field-of-view and hence contribute equally to all possible test regions in f_o. Further, we partition the multiscale detail data vector η^d as follows:

$$\eta^d = \left[\begin{array}{c} \eta_T \\ \hline \eta_R \end{array} \right], \tag{7.28}$$

where the vector η_T contains the multiscale data that corresponds to the test region Ψ_T, and η_R contains the rest of the elements in the detail vector η^d. The corresponding partitioning of the detail block of the multiscale covariance matrix is given by:

$$\Lambda_d = \left[\begin{array}{c|c} \Lambda_T & \Lambda_{TR} \\ \hline \Lambda_{TR}^T & \Lambda_R \end{array} \right], \tag{7.29}$$

and the detail contribution χ_d^2 to χ^2 can thus be written as follows:

$$\begin{aligned} \chi_d^2 &= (\eta^d)^T \Lambda_d^{-1} \eta^d \\ &= \left[\begin{array}{c|c} \eta_T^T & \eta_R^T \end{array} \right] \left[\begin{array}{c|c} \Lambda_T & \Lambda_{TR} \\ \hline \Lambda_{TR}^T & \Lambda_R \end{array} \right]^{-1} \left[\begin{array}{c} \eta_T \\ \hline \eta_R \end{array} \right]. \end{aligned} \tag{7.30}$$

Now the cross-covariance matrix Λ_{TR} is sparse and small in norm since the detail matrix Λ_d is diagonally dominant to begin with due to the compression achieved in the multiscale framework. We can thus expect that, to a good approximation, the presence of an anomaly in Ψ_T will have no effect on the contribution to χ_d^2 from η_R. In fact this is indeed what we observe and we present an example in Figure 7-1 to demonstrate this effect. Before we describe this example, however, let us first define through the following equation the contribution to χ_d^2 from the data η_T alone:

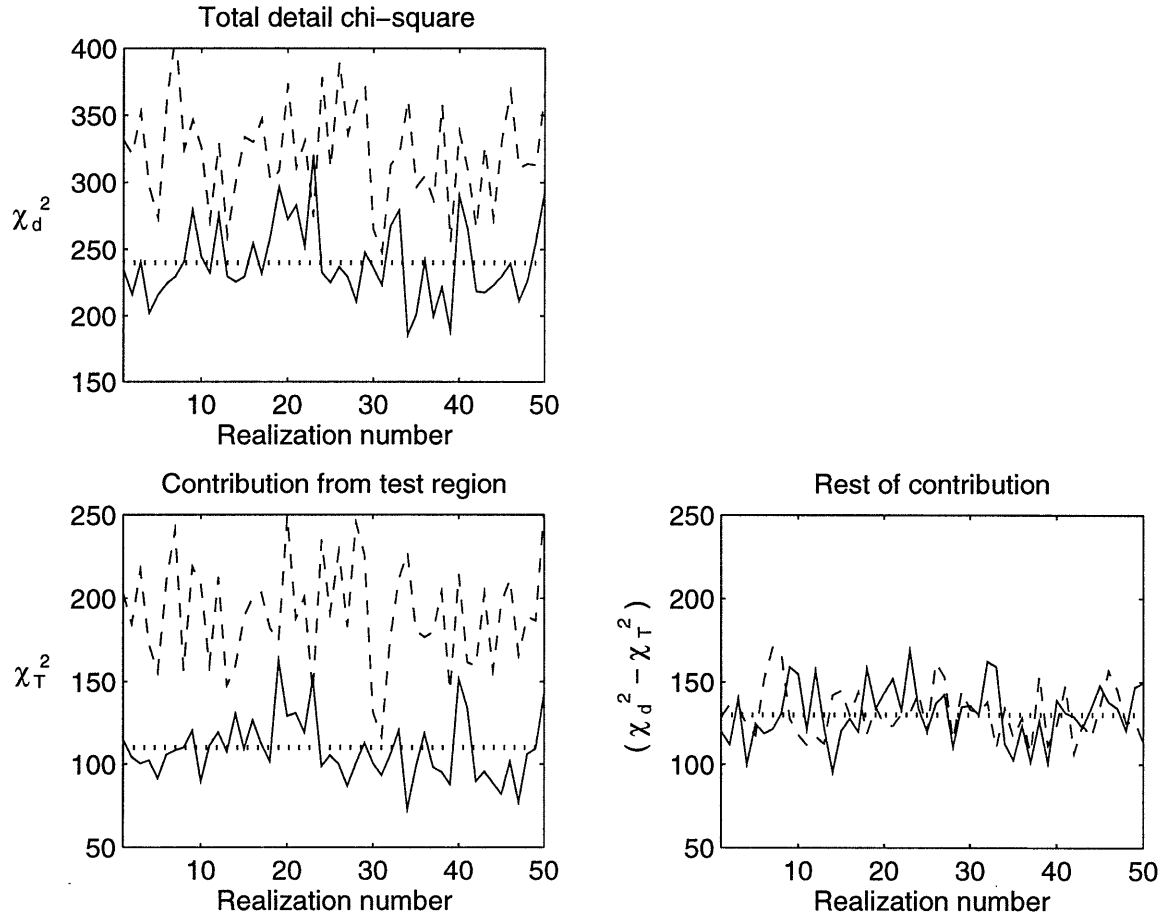
$$\chi_T^2 \triangleq \eta_T^T \Lambda_T^{-1} \eta_T. \tag{7.31}$$

Now if there is an anomaly in the test region Ψ_T then, from the above discussion, we expect the presence of the anomaly to be reflected mostly in the contribution χ_T^2 to χ_d^2. To verify this, we plot in Figure 7-1 the values χ_d^2, χ_T^2 and $(\chi_d^2 - \chi_T^2)$ corresponding to a 2×2 anomaly with a uniform intensity of 3 superimposed on a 16×16 fractal background with spectral parameter $\gamma = 2$. These values are shown for different realizations of the background texture and the additive measurement noise, which corresponds to an SNR[4] of 20 dB in this example. Further, the test region Ψ_T is chosen to coincide with the region Ψ_A that contains the anomaly. Note from the figure that the presence of an anomaly has little effect on the $(\chi_d^2 - \chi_T^2)$ value, and thus the increase in χ_d^2 due to the anomaly is entirely because of an increase in χ_T^2, the contribution from the data η_T that intersect the test region. That is $(\chi_d^2 - \chi_T^2)$ contains no useful information for determining the presence or absence of the anomaly and under either hypothesis it has essentially the same statistics, roughly corresponding to $\chi_R^2 \triangleq \eta_R^T \Lambda_R^{-1} \eta_R$.

We take advantage of this localization achieved in the multiscale framework to propose next a method that efficiently zooms in on the set of anomalous pixels Ψ_A. Specifically, in this method we use a two-step approach to anomaly localization and estimation. In the first step, we probe each individual pixel in the field for an anomaly. This probing is extremely fast and is based on an approximation $\tilde{\chi}_T^2$ to χ_T^2 where we ignore the off-diagonal elements in the diagonally dominant matrix Λ_d. The output of this first step is a set of pixels that have a high probability of being anomalous. In the second step of the detection technique, we focus our computational resources on only these candidate pixels and obtain the maximum-likelihood (ML) estimate of the intensity values at these pixel locations. However, in order to obtain this ML estimate we once again take advantage of the localization property of the wavelets and consider only those data which intersect the candidate pixels. Assuming that the size of the anomaly is sufficiently small, the number of candidate pixels is a small fraction of the size of the total input field. This implies that the size of the ML estimation problem is quite small and thus the corresponding solution can be obtained in an efficient manner. The above two steps when combined result in an efficient detection

---

[4] Refer to Chapter 6 for the precise definition of the SNR.


Total detail chi-square

Contribution from test region

Rest of contribution

Figure 7-1: The plot of the detail chi-square χ_d^2 (top left), the contribution χ_T^2 from the data that correspond to strips which intersect the test region (bottom left), and the remainder of the contribution $(\chi_d^2 - \chi_T^2)$ (bottom right), for different realizations of the background and the noise. Dashed line: Anomaly present. Solid line: No anomaly present. The dotted lines in the subplots represent the expected chi-square values given that there is no anomaly. The anomalous region Ψ_A is 2×2, and the anomaly has a uniform intensity of 3. The test region Ψ_T is chosen to coincide with Ψ_A. Note that an increase in the χ_d^2 value due to the anomaly is entirely due to an increase in χ_T^2.

procedure. We discuss next the two steps of our detection method in greater detail.

As mentioned previously, in the first step of our detection method we cycle through all the pixels in the input field and probe these individually for an anomaly. This probing is based on an approximate chi-square test where we ignore the off-diagonal elements in the diagonally dominant matrix Λ_d. This approximate chi-square value for a given test pixel p is provided by:

$$\tilde{\chi}_T^2(p) = \sum_{k \in \Psi_p} \frac{\eta(k)^2}{d(k)}, \tag{7.32}$$

where $\eta(k)$ is the k-th data sample, $d(k)$ is the k-th diagonal element of Λ_d, and Ψ_p is the set containing the indices of the multiscale detail strips which intersect pixel p. Note that since the values $(\eta(k)^2/d(k))$ can be computed and stored in memory before we begin the actual probing process, the calculation of $\tilde{\chi}_T^2$ (c.f. (7.32)) requires only the summation of the appropriate terms and is extremely fast. This first pass enables us thus to quickly cycle through all the pixels in the field and identify those pixels where there is a high probability of finding an anomaly and where the computational resources in the second step are to be focused. To this end, we store the indices of a number N_C of candidate pixels which have the N_C largest $\{\tilde{\chi}_T^2(p)\}$ values of all the pixels $\{p\}$. The choice of the number N_C is governed solely by an approximate knowledge of a prior bound on the size of the anomaly. Typically we would choose N_C conservatively so that typical anomalies of interest would consist of less than N_C pixels.

Subsequently, after the first pass is completed, in the second step of the detection algorithm we go back and calculate the maximum-likelihood (ML) estimate of the intensity values at only the N_C candidate pixel locations. However, for this ML estimation we once again use the localization property in the multiscale framework and thus only utilize those data samples arising from multiscale strips which intersect the N_C pixels. To this end, let the (N_C) vector f_C represent the intensity values at the N_C candidate pixel locations, and let the $(N_s^2 - N_C)$ vector f_R represent the intensity at the remainder of the pixels in the input field. Then f_o, the overall pixel intensity vector for the input field, can be partitioned in terms of these quantities as follows:

$$f_o = \left[\begin{array}{c} f_C \\ \hline f_R \end{array}\right]. \tag{7.33}$$

Further, if we let η_C be the multiscale data arising from the multiscale strips which intersect the N_C candidate pixels and let η_R be the rest of the multiscale data, then the following observation equation derived from (7.6) can be written in the multiscale framework:

$$\eta = \mathcal{T} f_o + \nu, \qquad \nu \sim \mathcal{N}(0, \lambda I_{N_\theta N_s}) \tag{7.34}$$

where ν is the multiscale observation noise, and the rows of the matrix $\mathcal{T}$ represent the discretized multiscale strips. The multiscale observation equation (7.34) can be written in a partitioned form as follows:

$$\left[\begin{array}{c} \eta_C \\ \hline \eta_R \end{array}\right] = \left[\begin{array}{c|c} \mathcal{T}_C & \mathcal{T}_{CR} \\ \hline 0 & \mathcal{T}_R \end{array}\right] \left[\begin{array}{c} f_C \\ \hline f_R \end{array}\right] + \left[\begin{array}{c} \nu_C \\ \hline \nu_R \end{array}\right] \qquad \begin{array}{l} \nu_C \sim \mathcal{N}(0, \lambda I_{N_{\eta_C}}) \\ \nu_R \sim \mathcal{N}(0, \lambda I_{N_{\eta_R}}), \end{array} \tag{7.35}$$

where η_C are the data which intersect the candidate pixels, η_R are the rest of the data, N_{η_C} is the size of the vector η_C, and N_{η_R} is the size of the vector η_R. As we had mentioned previously, in the second step of the detection method we take advantage of the localization achieved in the multiscale framework to obtain the ML estimate of the candidate pixel intensities f_C from only those data η_C which intersect the candidate pixels. To this end, the observation equation relating the quantities f_C and η_C can be written as follows from (7.35):

$$\eta_C = \mathcal{T}_C f_C + \mathcal{T}_{CR} f_R + \nu_C, \qquad \nu_C \sim \mathcal{N}(0, \lambda I_{N_{\eta_C}}). \tag{7.36}$$

Note from (7.36) that the ML estimation of the intensity f_C at the candidate pixel locations, from the data η_C is complicated by the fact that η_C also includes the (unknown) intensity contribution f_R from the remainder of the pixel locations in the input field. However, based on the results of the first step of our detection method, we do not expect these pixel locations to contain any anomaly, and hence the contribution from f_R is only due to the background fractal field. This fact suggests two different approaches for compensating for the intensity f_R in order to obtain the ML estimator of f_C from the data η_C. We next discuss the two ML estimators based on these two approaches.

For the first approach to obtaining an ML anomaly estimate, the observation equation (7.36) can be written as follows:

$$\eta_C = \mathcal{T}_C f_C + \varpi, \tag{7.37}$$

where ϖ is obtained by combining $\mathcal{T}_{CR} f_R$ and the noise ν_C into one single vector. Now by using the fact that the background fractal field is uncorrelated with the measurement noise, we obtain the following statistics for the vector ϖ:

$$\varpi = (\mathcal{T}_{CR} f_R + \nu_c) \sim \mathcal{N}(0, \mathcal{T}_{CR} \Lambda_{f_R} \mathcal{T}_{CR}^T + \lambda I_{N_{\eta_C}}), \tag{7.38}$$

where note that, since the pixel locations corresponding to f_R do not contain any anomaly, the matrix Λ_{f_R} is just the appropriate block of the fractal-field covariance matrix Λ (c.f. (7.4)).

Finally, the ML estimator $\widehat{f}_C$ of f_C can be calculated from the observation equation (7.37) and the noise statistics (7.38). Note, however, that the noise ϖ in (7.37) is correlated with f_C because f_C, the intensity vector at the candidate pixel locations, includes the contribution from both the anomaly as well as the fractal background, and hence is correlated with the background term $\mathcal{T}_{CR} f_R$. For our ML estimation problem, however, we neglect this correlation and simply obtain an ML estimate of f_C by assuming that ϖ and f_C are uncorrelated. This assumption results in an extremely accurate estimation of the intensity values in f_C, a fact that we demonstrate subsequently. The ML estimate of f_C from (7.37) and by assuming that ϖ and f_C

are uncorrelated is obtained by solving the following normal equation:

$$\left(\mathcal{T}_c^T \Lambda_{\varpi}^{-1} \mathcal{T}_c\right) \widehat{f}_c = \left(\mathcal{T}_c^T \Lambda_{\varpi}^{-1}\right) \eta_c. \tag{7.39}$$

Note that (7.39) represents a small ($N_c \times N_c$) system of equations and hence the corresponding ML solution $\widehat{f}_C$ is obtained with very little computation. Further, in the future we will refer to this ML estimate $\widehat{f}_C$ as the nearly-optimal ML estimate in order to distinguish it from a second sub-optimal ML estimate, which we discuss next.

In the second approach, a sub-optimal ML estimator for f_C can be obtained by ignoring the contribution from the background fractal field f_R altogether, i.e. by considering the following approximate observation equation:

$$\eta_C \approx \mathcal{T}_C f_C + \nu_C, \qquad \nu_C \sim \mathcal{N}(0, \lambda I_{N_{\eta_C}}). \tag{7.40}$$

The ML estimate $\widetilde{f}_C$ of f_C given the observation equation (7.40) is obtained by solving the following normal equation:

$$\left(\mathcal{T}_C^T \mathcal{T}_C\right) \widetilde{f}_C = \left(\mathcal{T}_C^T\right) \eta_C. \tag{7.41}$$

Note that since (7.41) represents a small ($N_C \times N_C$) system of equations, the ML solution $\widetilde{f}_C$ is obtained with very little computation. As we demonstrate in the next section, this sub-optimal ML estimator $\widetilde{f}_C$ estimates the candidate pixel intensities fairly accurately. However, as is expected, additional accuracy is obtained by using the nearly-optimal ML estimator $\widehat{f}_C$ which incorporates the statistical properties of the fractal background f_R in the ML estimation process.

Finally, once the ML estimators $\widetilde{f}_c$ or $\widehat{f}_c$ have been determined, a knowledge of E, the minimum intensity of the anomaly, can be used to threshold the pixel estimates from the corresponding estimators which are below this intensity level to zero. The pixel locations which correspond to a non-zero value then collectively constitute the solution $\widehat{\Psi}_A$ to the set of anomalous pixel locations Ψ_A. Our efficient two step approach to the detection, localization and estimation of the anomaly can now be summarized by the following algorithm:

Algorithm 5 (Anomaly Detection, Localization and Estimation)

1. *Find the multiscale noisy fractal-field data covariance matrix for the background under the null hypothesis* $\Lambda_{\eta|H_0} = W_b \Lambda_{y|H_0} W_b^T$*, where* $\Lambda_{y|H_0} = T\Lambda T^T + \lambda I_{N_\theta N_s}$ *is the noisy fractal-field data covariance matrix, and* λ *is the variance of the additive noise in the data. Note that the matrix* $\Lambda_{\eta|H_0}$ *can be calculated off-line. Define* Λ_a *and* Λ_d *to be the coarsest scale approximation and multiscale detail blocks, respectively, of* $\Lambda_{\eta|H_0}$*. Let the diagonal elements of* Λ_d *be given by* $\{d(k)\}$*.*

2. *Transform the noisy projection data into the multiscale framework* $\eta = W_b y$*. Let* η^a *and* η^d *be the coarsest scale approximation and multiscale detail blocks, respectively, of the multiscale data vector* η*.*

3. *Compute the total chi-square* $\chi^2 = (\eta^a)^T \Lambda_a^+ \eta^a + (\eta^d)^T \Lambda_d^{-1} \eta^d$ *by taking advantage of the sparsity and favorable conditioning of* Λ_d*. Find the confidence level at which an anomaly can be assigned in the input field by using the result* $\sqrt{2\,\chi^2_{H_0}} - \sqrt{2N_\theta N_s} \sim \mathcal{N}(0,1)$*. If the confidence level is greater than a certain threshold then continue, else stop and declare the no anomaly is present.*

4. *In order to determine the locations of the* N_C *candidate pixels, perform the following steps. For each pixel* p $(1 \leq p \leq N_s^2)$ *compute the approximate chi-square values as follows:*

   (a) *Find the set* Ψ_p *containing the indices of the multiscale detail strips that intersect pixel* p*.*

   (b) *Calculate the approximate chi-square value:* $\tilde{\chi}^2_T(p) = \sum_{k \in \Psi_p} (\eta(k)^2 / d(k))$*.*

5. *Find the locations of the pixels with* N_C *largest values for* $\tilde{\chi}^2_T(p)$*. These are the candidate pixels. Let the* (N_C) *vector* f_C *represent the intensity values of the field at these* N_C *pixel locations. Let the vector* f_R *represent the intensity values at the remainder of the pixels in the input field.*

6. *Find the nearly-optimal ML estimator* $\hat{f}_C$ *of* f_C*, or the sub-optimal counterpart* $\tilde{f}_C$*, according to the following steps:*

   (a) *Locate the data* η_C *that correspond to the multiscale detail strips which intersect the candidate pixels. Let* $\mathcal{T}_C$ *be that block of the multiscale projection matrix* $\mathcal{T} = WT$ *which relates* η_C *and* f_C*.*

   (b1) *If* $\hat{f}_C$ *is desired, solve the following normal equation:*
   $\left(\mathcal{T}_C^T \Lambda_w^{-1} \mathcal{T}_C\right) \hat{f}_C = \left(\mathcal{T}_C^T \Lambda_w^{-1}\right) \eta_C$.

   (b2) *If* $\tilde{f}_c$ *is desired, solve the following equation:* $\left(\mathcal{T}_C^T \mathcal{T}_C\right) \tilde{f}_C = \left(\mathcal{T}_C^T\right) \eta_C$.

7. *Based on an approximate knowledge of the minimum anomaly intensity* E*, threshold the pixel estimates in* $\hat{f}_c$ *(*$\tilde{f}_c$*) which are below this intensity level to zero. The pixel locations which correspond to a non-zero value in* $\hat{f}_c$ *(*$\tilde{f}_c$*) then collectively constitute the solution* $\hat{\Psi}_A$ *to the set of anomalous pixel locations* Ψ_A*.*

In the next section we present examples demonstrating the efficacy of our two-step approach to anomaly detection in the multiscale framework. In particular, we show the performance of our detection approach for both when we use the nearly-optimal ML estimator (c.f. (7.39)) as well as when we use the sub-optimal estimator (7.41) in the second step of our detection approach.

7.3 Results

In this section we present examples that demonstrate the efficacy of our two-step approach to detection. In all the examples that we present here we discretize the input field on a 32 × 32 pixel grid, and we assume the field to consist of a fractal background (spectral parameter $\gamma = 2$) with a 2 × 2 anomaly superimposed. Given that the size of the anomaly is equal to 4 pixels, we select a conservative value of $N_c = 20$ for the number of candidate pixels that are identified by the first step of our detection method. Note that the normal equations (7.39) in the final step (or it's sub-optimal counterpart (7.41)) will then be only 20 dimensional, representing a modest computational task. We further assume the projection data to be gathered according to $N_\theta = 32$ angular projections and $N_s = 32$ strips in each angular projection. Finally, we use the Haar wavelet for the multiscale decomposition of the data vector.

In the first example we demonstrate each step of our efficient two-step detection procedure. Further, we demonstrate the performance at the second step by using the nearly-optimal ML estimator (7.39) as well as the corresponding sub-optimal estimator (7.41). To this end, we show in Figure 7-2 the detection results corresponding to a particular realization of the background fractal-field. For the example shown in Figure 7-2 the 2×2 anomaly is located at coordinates $(16,16)$, $(16,17)$, $(17,16)$, and $(17,17)$ in the input field, and is assumed to have a uniform intensity of 2.5 which results in a anomaly-to-background ratio of -13 dB, defined as:

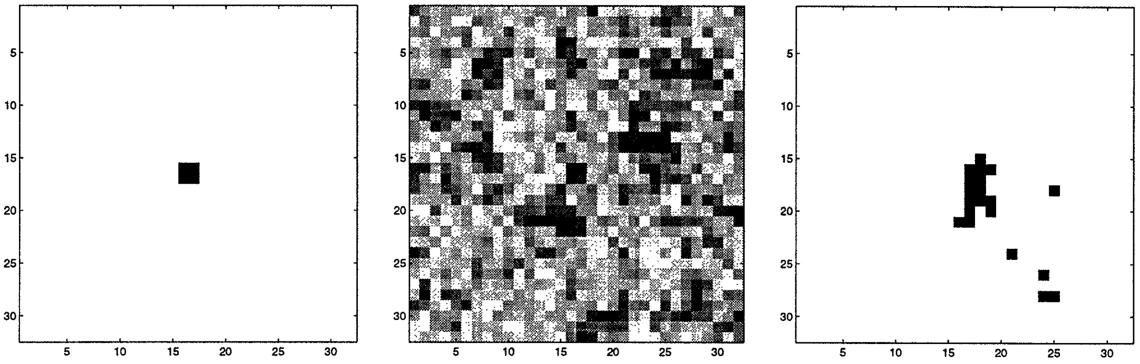
$$\text{anomaly} - \text{to} - \text{background ratio} = 10 \log \frac{f_a^T f_a}{\text{Trace}(\Lambda)}, \tag{7.42}$$

where recall that Λ is the covariance matrix for the background fractal field pixels, and f_a is the vector containing the pixel coefficients of the anomaly.

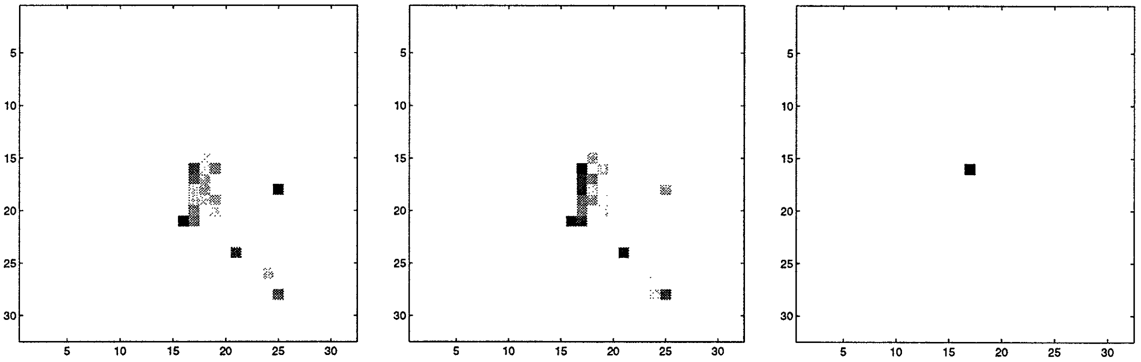
The following observations can be made from Figure 7-2. First note that the particular detection task shown in the figure is quite challenging, e.g. it is not possible for a human observer to pick out the anomaly confidently in the input field (top row, middle). Note however that, despite the difficulty of the detection task, the quick, first step (top row, right) of our two-step detection approach is able to zoom-in on the correct approximate anomaly area. This area consists of $N_c = 20$ candidate pixels which not only represent the anomaly, but also a few false alarm pixels. In the second pass, however, our detection procedure removes all of the false alarms by estimating the field intensity values at the candidate pixel locations (bottom row, middle, and bottom row, right). In particular, note that the sub-optimal ML estimator (bottom row, left) performs well in removing the false alarms. This performance is, however, improved by the nearly-optimal ML estimator (bottom row, middle) which estimates the pixel intensities to a greater accuracy than the corresponding sub-optimal estimator. This difference in performance between the sub-optimal and the nearly-optimal

ML estimators can be seen from Figure 7-3. In the figure we show the sections through the actual input field (solid lines) and the fields that are obtained by the sub-optimal (dash-dotted lines) and the nearly-optimal (dashed lines) ML estimators. Further, we show the sections through the two rows that contain the anomaly (i.e. rows 16 and 17 of the fields) and also two sample rows that correspond to just the fractal background (i.e. rows 19 and 20 of the fields). Note from Figure 7-3 that even though the pixel intensities estimated by the sub-optimal ML estimator are reasonably accurate, this accuracy is increased substantially by the use of the nearly-optimal ML estimator.

We next show the receiver operating characteristic curve (ROC) for our detection technique, corresponding to the same imaging configuration as in the examples shown in the previous figures. These curves are obtained by plotting the probability of detection (PD) versus the percentage of false alarm pixels for different values of the threshold intensity level. For the PD and false alarm calculations we embed the true 2×2 anomaly in a larger, 6×6 region (c.f. Figure 7-4). We assume a detection to correspond to the event that the ML estimate of any pixel within this 6×6 region has an intensity greater than the threshold (i.e. we declare a detection if our algorithm finds an anomalous pixel in this region). The pixels which are outside this 6×6 region and for which the ML estimate of the intensity level is greater than the threshold are considered to be false alarms. This embedding of the true anomaly in a larger region can represent a variety of situations. First, it may be that for a particular detection application a localization error of a few pixels falls within the acceptable tolerance level. Second, a larger embedding region around the anomaly may signify the fact our method is not able to detect anomalies that are smaller than a certain critical size. Finally, pixel errors in our projection-domain detection method can arise due to an imprecise knowledge of the projection operator specified by the imaging geometry. For example, recall that we had stated in Chapter 2 that the approximate projection operator that we use in this thesis for computational simplicity often results in a misalignment error of one pixel. Specifically, our choice in this chapter of an embedding region of size 6×6 for a 2×2 anomaly implies that we are willing to tolerate an anomaly localization error of 2 pixels.

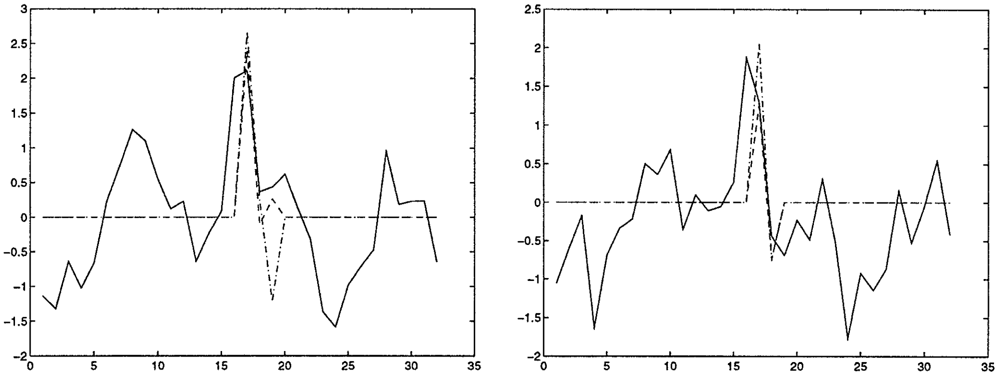


(a) Anomaly (b) Input field (c) Zoom-in step result



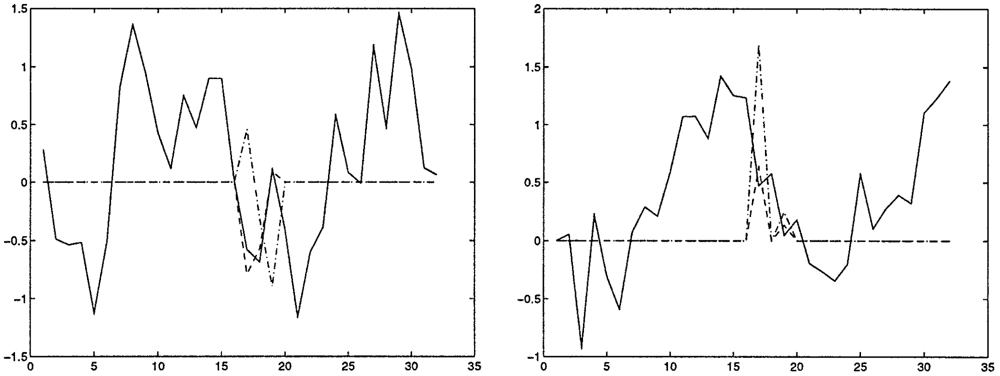
(d) Sub-optimal est. (e) Nearly-optimal est. (f) Thresholded estimate

Figure 7-2: The detection results corresponding to a particular realization of the background fractal-field (spectral parameter $\gamma = 2$). The projection data has an SNR of 20 dB and consists of $N_\theta = N_s = 32$. (a) The 2×2 anomaly with a uniform intensity of 2.5 is located at pixels $(16, 16)$, $(16, 17)$, $(17, 16)$, and $(17, 17)$. (b): The grayscale plot of the input field (black corresponds to the maximum value) consisting of the anomaly plus the fractal background. (c) The binary plot of $N_c = 20$ pixels with the largest approximate chi-square value $\tilde{\chi}^2_T$ that are identified at the end of the first step. (d) The grayscale plot of the sub-optimal ML estimate values of the N_c pixels. (e) The grayscale plot of the nearly-optimal ML estimate values of the N_c pixels. (f) The binary image displaying the result of the application of a threshold equal to 2 to the nearly-optimal ML estimate image (e).



(a) Section through row 16

(b) Section through row 17



(c) Section through row 18

(d) Section through row 19

Figure 7-3: The sections through the input field (solid line) and the fields obtained by the sub-optimal ML estimation (dash-dotted line) as well as by the nearly-optimal ML estimation (dashed line) corresponding to the detection task in Figure 7-2. The top row corresponds to sections where the anomaly is present, and the bottom row corresponds just to the background.

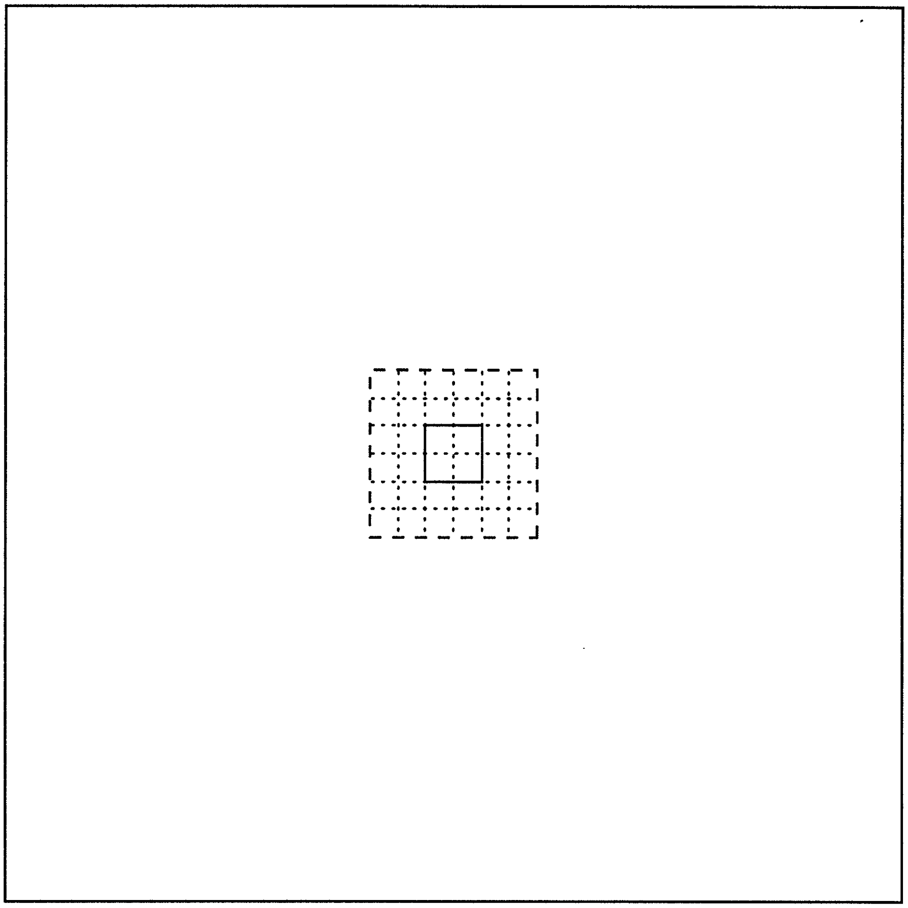


Figure 7-4: For the probability of detection and the false alarm calculations the 2×2 anomaly (center square, solid lines) is embedded in a larger, 6×6 region (dashed lines). A detection is assumed to correspond to the event that the ML estimate of any pixel within this 6×6 region has an intensity greater than the threshold. The pixels which are outside this 6×6 region and for which the ML estimate of the intensity level is greater than the threshold are considered to be false alarms.

Figure 7-5 shows the two ROC curves obtained by using the nearly-optimal and the sub-optimal ML estimators respectively. Each point on the ROC curve corresponds to a certain threshold intensity level and is obtained by averaging the number of detections and the number of false alarm pixels from a series of Monte Carlo simulation runs. Also on the ROC curves we show the error bars that depict the 95% confidence intervals. The anomaly is once again assumed to span 4 pixels for this example and is assumed to have a uniform intensity of 2.5. Note from the figure that as expected the ROC curve corresponding to the nearly-optimal ML estimator reflects a more accurate detection than that corresponding to the sub-optimal ML estimator. Note further that for the case of the nearly-optimal ML estimator our method is able to achieve a good detection rate of 87% with just 0.07% (i.e. less than one) false alarm pixel. In the case the sub-optimal ML estimator is used, an 80% detection rate corresponds to an average of $4 - 5$ false alarm pixels.

7.4 Discussion

In this chapter we have developed an efficient data-domain approach for tomographic detection of anomalies that are superimposed on textured backgrounds. The conventional methods for these problems either first reconstruct the entire field and then use post-processing to detect anomalies or assume an anomaly to be parameterized and then perform the detection procedure directly in the data-domain. The problem with the former approach is that if the data are sparse and/or noisy the reconstruction suffers from severe artifacts that makes the anomaly very difficult to detect. The latter approach, on the other hand, imposes restrictive assumptions on the anomaly and hence is limited in scope. We avoid these problems associated with the conventional detection methods in our approach by working directly in the data-domain, by imposing no restrictive assumptions on the anomaly, and finally by using the statistical properties of the data and the measurement noise in such a manner that our approach results in a nearly statistically optimal detection technique which, at the same time, is also highly efficient.

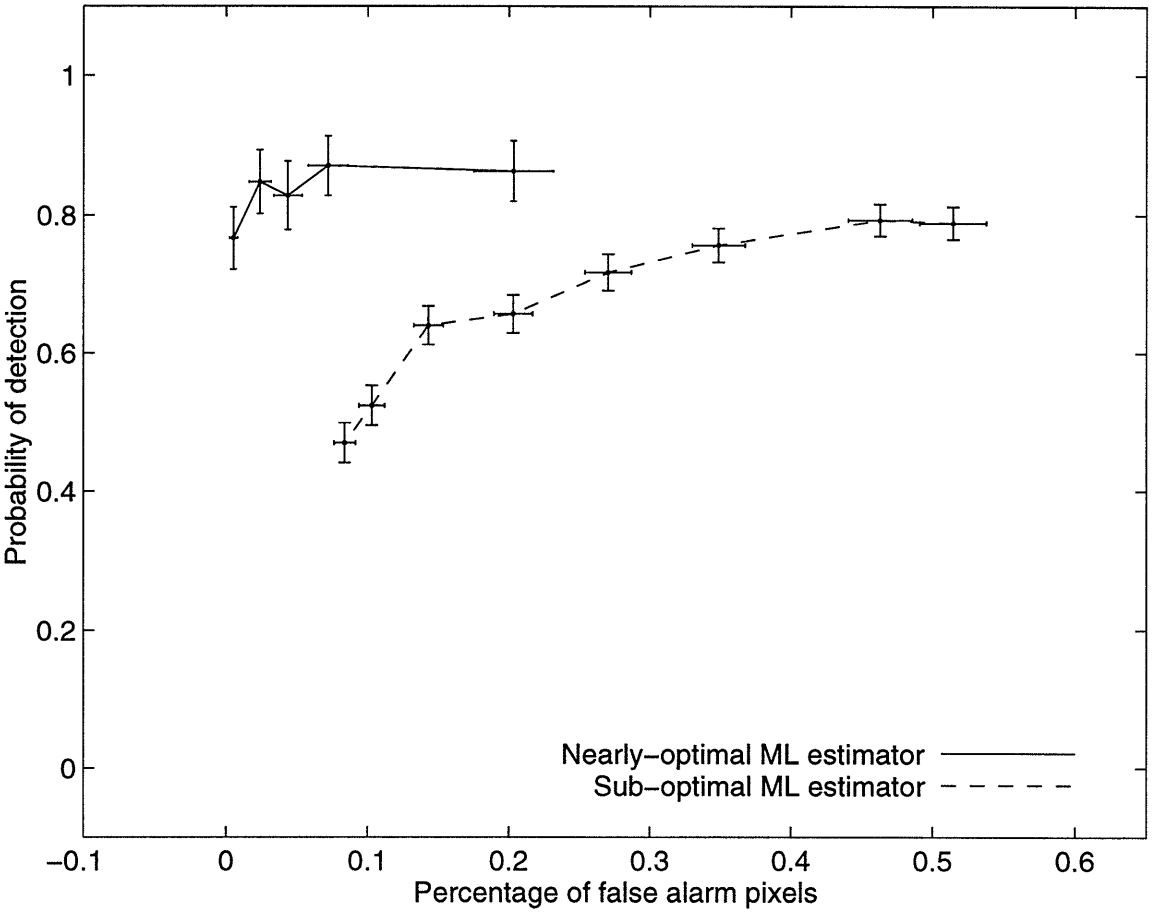


Figure 7-5: The ROC curve for the two-step detection technique, corresponding to different realizations of the background field (spectral parameter $\gamma = 2$) and the measurement noise (SNR = 20 dB). The projection data are gathered according to $N_\theta = N_s = 32$. The error bars depict the 95% confidence intervals. The anomaly is 2×2 with a uniform intensity of 2.5. Solid line: Nearly-optimal ML estimator used in the second step. Broken line: Sub-optimal ML estimator used in the second step.

Specifically, in this chapter we model the background texture as a $1/f$ fractal process and the object as a localized anomaly. We then obtain the statistics for the projection data that corresponds to just the fractal background, and perform a chi-square test on the actual data to see whether there is a significant deviation from the non-anomalous statistics. If performed in the original data-domain, however, the exact chi-square test is complicated by the fact that inversion of the large, full, and in general quite ill-conditioned fractal-field data covariance matrix is required. As in Chapter 6, we avoid this complication by transforming the data into the multiscale framework. The multiscale fractal-field data covariance matrix is sparse and, in addition, is partitioned naturally into an ill-conditioned coarser scale approximation block and a relatively well-conditioned multiscale detail block. Because of these properties of the multiscale fractal-field data covariance matrix, the chi-square calculation can thus be performed extremely efficiently in the multiscale data-domain.

We further specialize our detection technique so that not only does it detect the presence of an anomaly in the input field, but it also localizes the anomaly and estimate it's value. We do this by using the fact that the multiscale wavelet basis functions are localized and that the multiscale fractal-field data covariance matrix is extremely sparse and in particular is diagonally dominant. Specifically, in this chapter we develop an efficient two-step method for localizing the anomaly in the input field. To begin we compute the multiscale transformation of the data. Then, in the first step of our detection method, we calculate and store the contributions of each one of these data samples to the total chi-square value by assuming the multiscale data to be uncorrelated, i.e. by ignoring the off-diagonal elements in the fractal-field data covariance matrix. After this we cycle through all the pixels in the field and test to see if these are anomalous. This test is based on the approximate chi-square value that is obtained simply and efficiently by adding together the contributions from only those data which correspond to the multiscale strips that intersect the pixel. Finally, after localizing the anomaly to a few candidate pixels based on this approximate chi-square test, we go back and obtain a nearly-optimal ML estimate of the field intensity values at only these candidate pixel locations. However, in order to obtain this ML estimate we once again take advantage of the localization property of the wavelets and consider only that data corresponding to multiscale strips which intersect the candidate pixels. Assuming that the size of the anomaly is sufficiently small, the number of candidate pixels are a small fraction of the size of the total input field. This implies that the size of the ML estimation problem is quite small and thus the corresponding solution is obtained in an efficient manner. The above steps put together result in a highly efficient algorithm for detecting anomalies in the multiscale data-domain.

Chapter 8

Thesis Contributions And Topics For Future Research

8.1 Contributions

In this thesis we have explored the use of wavelet-based multiscale representation for reconstruction and detection of objects from tomographic data. We have concentrated our efforts mainly for the case when the tomographic data are sparse and/or noisy. In this case the Radon transform results based on continuous and noise-free data break down and the reconstruction and detection tasks become more challenging. We have demonstrated the effectiveness of the multiscale framework for regularizing these ill-posed reconstruction and detection problems, and in particular have developed highly efficient and, at the same time statistically optimal multiscale-based algorithms for reconstruction and detection of objects from sparse as well as noisy data. We next summarize the salient contributions made by the multiscale techniques developed in this thesis to the field of tomography.

8.1.1 Multiscale reconstruction: summary of contributions

An efficient multiscale reconstruction technique from sparse/noisy data

The first contribution of this thesis is the *development of an extremely efficient multiscale reconstruction technique from both sparse as well as noisy data.* This multiscale reconstruction technique is based on the natural pixel (NP) representation for tomography which was originally proposed for the solution of sparse data reconstruction problems. The NP representation results in a matrix-based reconstruction method where the projection data and the object coefficients are related through a system matrix that is large, full, and quite ill-conditioned. As a result the NP reconstruction technique is computationally intensive. In our multiscale method the data and the NP coefficients are transformed into the multiscale framework and are related by a multiscale system matrix that is *sparse*. In addition, we have shown that the non-uniqueness in the NP representation, which results in an ill-conditioned system matrix, is fully captured by the DC elements in the multiscale matrix. As a result we are able to partition the multiscale system matrix by scales such that the resulting reconstruction method requires the solution of just a *well-conditioned* and *sparse* system of equations, and hence is *computationally efficient*. Further, the use of a wavelet basis enables us to obtain the reconstruction estimates at multiple scales essentially for free.

An efficient iterative method for reconstruction from sparse data

The second contribution of this thesis is *an extremely efficient iterative method for reconstruction from sparse data*, which we refer to as MPART. We present this method as an alternative to another iterative method namely the algebraic reconstruction technique (ART) that is conventionally used to reconstruct objects from sparse data. In ART the object is expanded in the standard rectangular pixel basis. The object pixel coefficients are related to the data by the projection matrix, and are solved for iteratively through the use of the Kaczmarz projection method. The projection matrix in ART is, however, ill-conditioned and as a result the convergence of the solution for the pixel coefficients is slow and in addition is critically affected by the ordering scheme in which the rows of the projection matrix are accessed by the Kaczmarz method, and also on the choice of the relaxation parameter. In the iterative method that is developed in this thesis, we utilize the Kaczmarz method to solve the sparse and well-conditioned system of equations which arises in our multiscale transformed NP framework. Due to the well-conditioned nature of the multiscale system matrix, the convergence of the solution for the multiscale object coefficients in our iterative technique is *fast* and is *not critically dependent on the choice of the row ordering scheme or the relaxation parameter*. In addition, we demonstrate that our technique requires *substantially less computations per iteration* than ART. These facts when put together imply that our iterative reconstruction technique is much more efficient than ART. Further, in our iterative technique the reconstruction estimates at multiple scales are obtained essentially for free.

A class of stochastic models constructed in the multiscale data-domain

The third contribution of this thesis is the development of a particular *class of stochastic models for regularization of ill-posed reconstruction problems* where the data are noisy or sparse. These stochastic models are constructed directly in the multiscale data-domain and lead to computationally efficient estimation algorithms that are *no more complex* that the corresponding unregularized reconstruction algorithms. This is in contrast to the conventional methods for regularization in tomography which specify the prior model in the object-domain and which lead to extremely inefficient estimation algorithms. In particular the data-domain prior models that we use in this thesis not only impose a classic smoothness constraint on the object estimate, but also perfectly capture the statistics of the object in case the object is fractal-like. These fractal priors are accurate models for a variety of textures and hence are commonly encountered in tomography. The fractal prior models that we use in this thesis are captured simply and efficiently in the multiscale data-domain by a *diagonal* covariance matrix, and thus result in *efficient estimation algorithms*.

An extremely efficient multiscale reconstruction method from dense data

The final contribution of this thesis is a multiscale reconstruction method that is *specialized to yield fast reconstructions at multiple scales from dense but possibly noisy tomographic data.* We develop this method by adapting our NP-based multiscale framework to the filtered back-projection (FBP) reconstruction technique, the latter being the most commonly used method for image reconstruction from dense tomographic data. In particular, in our FBP-based technique the reconstructions at multiple scales are obtained with the *same computational complexity* as the highly efficient FBP reconstruction. We further show that the regularization obtained in this dense data case by using the class of fractal prior models constructed directly in the multiscale data-domain is equivalent to a particular way of rolling-off of the ramp filter. This provides a *statistical interpretation to the ad hoc rolling-off of the ramp filter* in the FBP, and also provides a recipe for designing *statistically optimal filters* to provide such roll-off.

8.1.2 Multiscale detection: summary of contributions

Discrimination of fractal fields directly from noisy and sparse data

The first detection task that we consider in the thesis is the *discrimination of fractal fields directly from the noisy and sparse projection data.* The conventional data-domain method for discrimination of such fields is based on the slope of the averaged power spectra of the projections and is derived from noise-free continuous-data Radon transform results. This conventional method thus breaks down in the case that the data are sparse and/or noisy. In contrast our discrimination method is *robust to both the quantity and the quality of the tomographic data.* We treat the fractal-field discrimination problem in a hypothesis testing framework. The *statistically-optimal solution* to this problem is then given in terms of the likelihoods. This likelihood calculation is, however, complicated because it requires the determinant and the inverse of the fractal-field data covariance matrix which is large, full, and in general quite ill-conditioned. Further, as we show in this thesis, any hope of whitening this matrix by using results from the existing literature is futile since these results are valid only if the data are continuous and noise-free, and hence break down in the case of sparse or noisy data. In contrast, in this thesis we have developed an efficient two-step method for whitening an accurate approximation of the fractal-field data covariance matrix. Because of this whitening we are able to efficiently compute the likelihoods and thus are able to discriminate fractal fields from noisy/sparse data with very little computational complexity. The first step of our two-step whitening approach consists of representing the projection data in the multiscale framework. We show that, just like the multiscale NP matrix, the multiscale fractal-field data covariance matrix is *sparse*, and in addition can be partitioned by scales into ill and well-conditioned submatrices. This in turn eliminates the need for computationally intensive techniques like the eigen-decomposition to remove the ill-conditioning in the covariance matrix. In the second step of our whitening approach we use the recently introduced class of multiscale stochastic models defined on trees to realize an accurate approximation of the multiscale fractal-field data covariance matrix. These tree-based models have the advantage that they lead to *fast likelihood calculations*, and thus to an *efficient method for discrimination of fractal fields*.

Detection of anomalies in textured backgrounds

The second detection problem considered in this thesis is the problem of *detecting anomalies in textured backgrounds*. The conventional methods for the solution of these problems either first reconstruct the entire field and then use post-processing to detect the anomalies or assume the anomaly to be parameterized and then perform the detection procedure directly in the data-domain. The problem with the former approach is that if the data are sparse and/or noisy the reconstruction suffers from severe artifacts that makes the anomaly very difficult to detect. The latter conventional approach, on the other hand, imposes restrictive assumptions on the anomaly and hence is limited in scope. We avoid these problems associated with the conventional detection methods in our approach by *working directly in the data-domain*, by imposing *no restrictive assumptions on the anomaly*, and finally by using the statistical properties of the data and the measurement noise in such a manner that our approach results in a nearly *statistically-optimal* detection technique which, at the same time, is also highly efficient. In particular we use fractal processes to model the background textures. These processes are accurate models for a variety of textures including terrain and several biological structures. As a result, the specific detection task that we consider in the thesis is sufficiently general to encompass a wide variety of applications, from detection of lesions in medical images to detection of defects in materials. In our data-domain detection approach we first obtain the statistics for the projection data that corresponds just to the background texture, and then analyze the actual data to see whether it deviates significantly from this statistics. This analysis is based on the chi-square test and enables us to compute the confidence level at which an anomaly can be assigned in a certain test region of the input field. We further take advantage of the sparsity of the fractal-field data covariance matrix in the multiscale framework and also the fact that the multiscale wavelet basis functions are localized to propose a highly efficient two-step approach for localizing the anomaly. In the first step we cycle through all the pixels in the input field and calculate the confidence levels at which an anomaly can be assigned in each of these pixels. This calculation is, however, based only on the diagonal elements of the fractal-field data covariance matrix and hence is extremely fast. Finally, after localizing the anomaly to a few candidate pixels based on the results of the first step of our detection method, we go back and obtain a maximum-likelihood (ML) estimate of the field intensity values at only these candidate pixel locations. However, in order to obtain this ML estimate we take advantage of the localization property of the wavelets and consider only that data which intersect the candidate pixels. Assuming that the size of the anomaly is sufficiently small, the number of candidate pixels are a small fraction of the size of the total input field. This implies that the size of the ML estimation problem is quite small and thus the corresponding solution is obtained in an efficient manner. This two-step procedure results in an extremely efficient data-domain technique for detecting anomalies in fractal textured backgrounds, and is made possible because in our wavelet-based multiscale representation the various multiscale fractal-field data are approximately decorrelated.

8.2 Topics for future research

In this section we identify and briefly discuss a few potential topics for future research that are related to this thesis.

8.2.1 Multiscale reconstruction: topics for future research

Extension to a fan-beam geometry

In this thesis we have developed a multiscale reconstruction technique that is tailored to the case of the parallel-beam imaging geometry. However an extension to the fan-beam case is fairly straightforward and is an obvious topic for future study. We next outline briefly the development of a multiscale reconstruction technique for the fan-beam geometry. To begin, consider the fan-beam geometry shown in Figure 8-1 [44]. Assume that, as in the parallel-beam case, there are N_θ angular projections and N_s fan-beam strips in each angular projection. Let C^f be the NP system matrix for the fan-beam case such that the elements of C^f are the areas of intersection of the various fan-beam strips. The multiscale reconstruction procedure for the fan-beam case can now be defined through the following equations:

$$\xi = W_b x, \tag{8.1}$$

$$\eta = W_b y, \tag{8.2}$$

$$\eta = (W_b C^f W_b^T)\xi \triangleq \mathcal{C}^f \xi, \tag{8.3}$$

where if W is the matrix which represents a N_s vector in a 1-D multiscale basis, then W_b is a block diagonal matrix with N_θ blocks along the diagonal all equal to W. Note that the only change from the parallel-beam case is the equation relating the multiscale object coefficients ξ and the multiscale data η. Recall that in the parallel-beam case η and ξ are related by the multiscale matrix $\mathcal{C} = W_b C W_b^T$ which is sparse and, in addition, can be naturally partitioned into a ill-conditioned coarsest scale approximation block and a well-conditioned multiscale detail block. Similarly, we claim that the fan-beam multiscale matrix C^f is also sparse and, in the case that the object is entirely covered by all N_θ projections, the ill-conditioning in C^f is also concentrated in the coarsest scale approximation block.

The reason why the ill-conditioning in C^f is concentrated in the coarsest scale approximation block follows from the same arguments that we had provided for the parallel-beam case. To see why C^f is sparse, consider Figure 8-2. In the figure we show two finest scale multiscale basis functions corresponding to the Haar wavelet. The finest scale elements in C^f are the areas of intersection of the finest scale basis functions of the type shown in Figure 8-2. Just as in the parallel-beam case, each such area of intersection can be broken into positive and negative contributions as shown in the figure. However, unlike the parallel-beam geometry, these positive and negative contributions do not exactly cancel each other and thus the finest scale elements in C^f are not exactly zero. Still, since in the fan-beam case the width b of the detectors is much smaller than their distance L to the source, the lines AB and DC, and AD and BC are nearly parallel. Thus the positive and negative contributions are nearly identical implying that the fine scale terms in C^f are very close to zero. Hence we expect the finest scale blocks in C and C^f to have similar sparsity. This argument however does not hold for the coarse scale terms where the basis functions are of larger extent (c.f. Figure 8-3). In this case we expect the coarse scale blocks in the parallel-beam multiscale matrix C to be more sparse than the corresponding blocks in the fan-beam matrix C^f. However, since the number of coarse scale elements is a small fraction of the total, the overall sparsity of C and C^f will not be much different.

Extension to a cone-beam geometry

A next possible topic of future research is the extension of the multiscale reconstruction technique developed in this thesis to the three-dimensional cone-beam case (c.f. Figure 8-4), a configuration which is commonly used in many practical tomog-

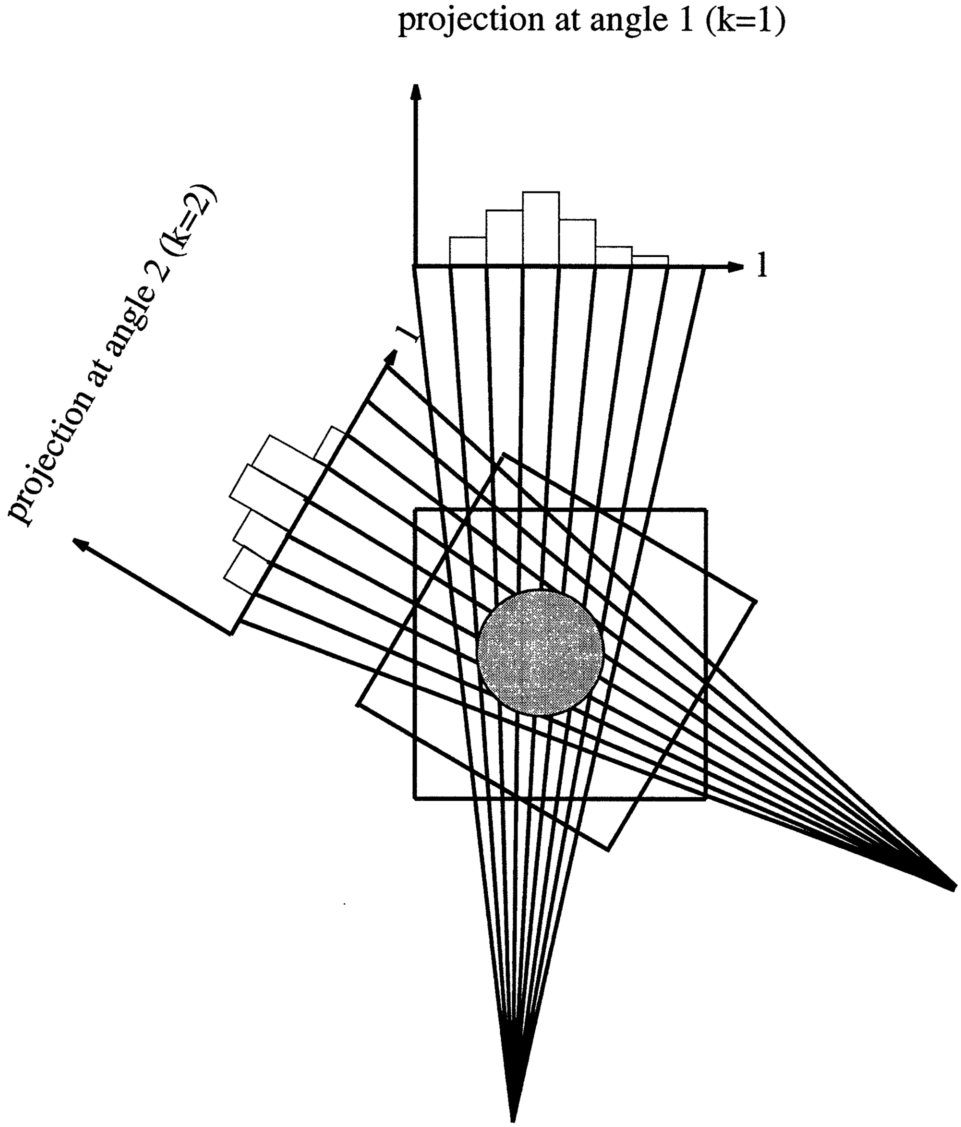


Figure 8-1: Two projections from the fan-beam imaging geometry. The circular region in the middle represents the object.

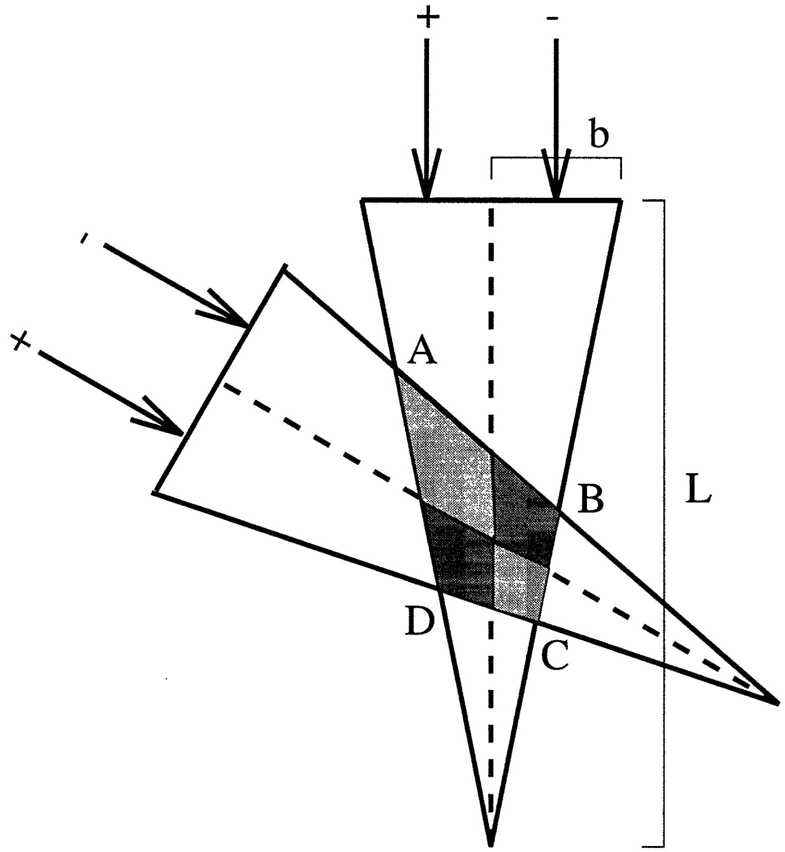


Figure 8-2: Two finest scale multiscale basis functions (corresponding to the Haar wavelet) for the fan-beam geometry. The finest scale elements in C^f are the areas of intersection of the finest scale basis functions of the type shown here. Since the width b of the detectors is much smaller than their distance L to the source, the lines AB and DC, and AD and BC are nearly parallel. This implies that the positive (darkly shaded) and negative (lightly shaded) contributions to the area of intersection are nearly identical.

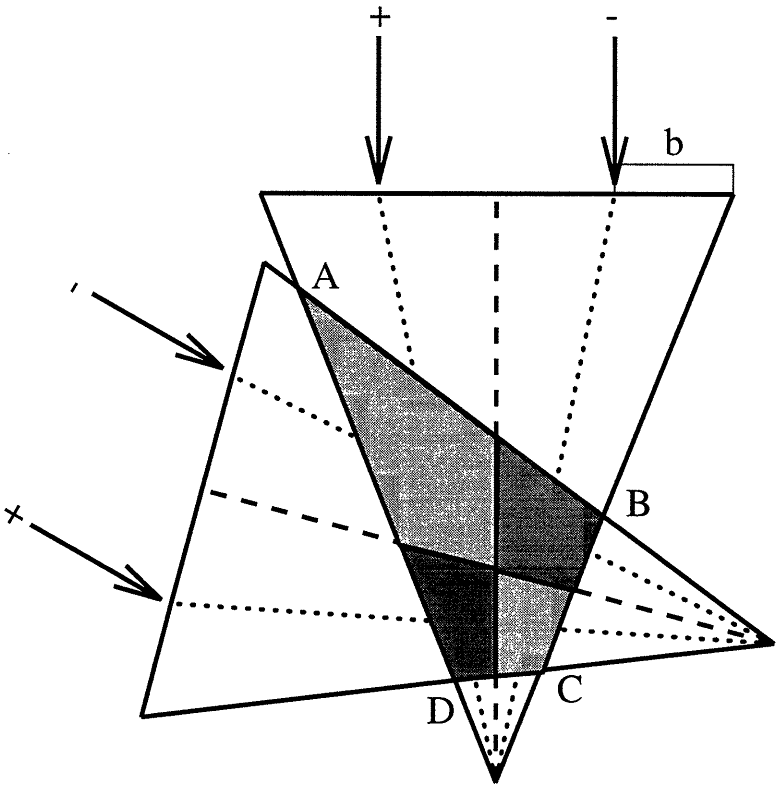


Figure 8-3: Two next to finest scale multiscale basis functions for the fan-beam geometry. The next to finest scale elements in C^f are the areas of intersection of the basis functions of the type shown here. The lines AB and DC, and AD and BC may not be nearly parallel in this case due to the large extent of the basis functions. This implies that the positive (darkly shaded) and negative (lightly shaded) contributions to the area of intersection may be different in this case.

raphy scanners. Since the projections corresponding to this configuration are now two-dimensional, a possible method for extension of our multiscale reconstruction technique to this case consists of representing each two-dimensional projection in a separable two-dimensional wavelet basis. As an example, we have shown in Figure 8-5 three fine scale basis functions corresponding to the two-dimensional separable Haar wavelet. The full two-dimensional separable Haar transform of the projection consists of inner products of the projection with basis function of the types shown in Figure 8-5 but at multiple scales and shifts. If we denote the 2-D separable wavelet transform operation by the matrix $\mathcal{W}$, then the three-dimensional multiscale reconstruction procedure is given by the following equations:

$$\begin{aligned} \xi &= \mathcal{W}_b x, & (8.4) \\ \eta &= \mathcal{W}_b y, & (8.5) \\ \eta &= (\mathcal{W}_b C \mathcal{W}_b^T)\xi, & (8.6) \end{aligned}$$

where $\mathcal{W}_b$ is a block diagonal matrix with as many diagonal blocks as there are projections, and with all diagonal blocks similar and equal to $\mathcal{W}$. Further, since the elements of the matrix C are the areas of intersection of the various cone beams, the elements in the multiscale matrix $(\mathcal{W}_b C \mathcal{W}_b^T)$ are the areas of intersection of the multiscale cone beams. By an extension of the arguments which we just presented for the fan-beam case, we can expect the finest scale elements in this matrix to be mostly zero. Thus if only the fine scales features (for example, edges and boundaries) in the object are to be reconstructed then for all practical purposes we can assume the multiscale matrix to be diagonal. This fact has a major ramification in three-dimensional angiography, which is a very important application of the cone-beam configuration. In angiography the desire is not to reconstruct the entire cross-section of the blood vessels but is just to obtain a reconstruction of their boundaries. We can expect to obtain this reconstruction in an extremely efficiently manner through the above multiscale method by assuming that the multiscale matrix is diagonal.

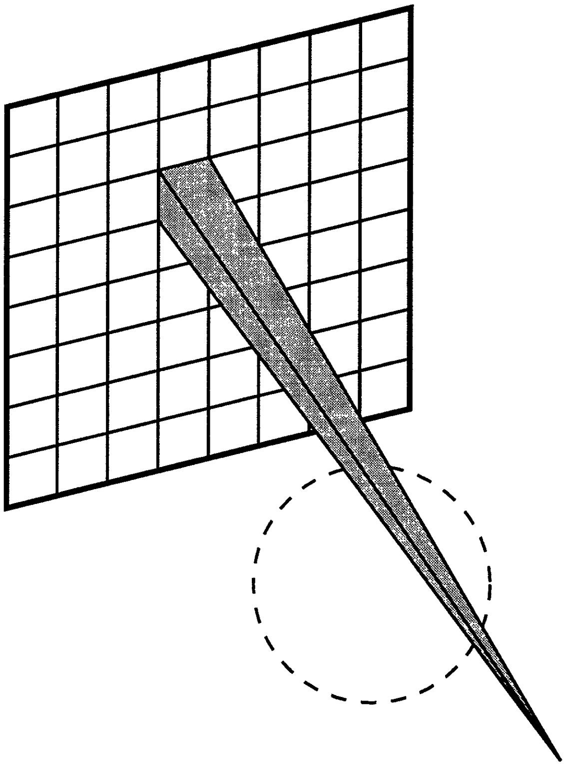


Figure 8-4: A two-dimensional projection from the three-dimensional cone-beam imaging geometry. The circular region represents the object.

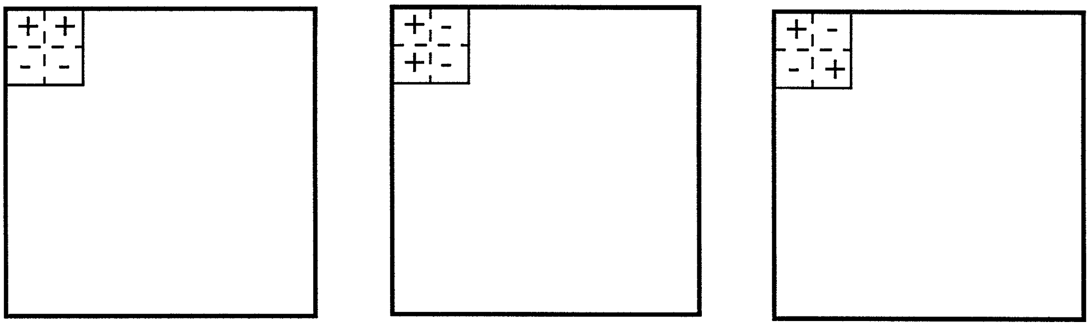


Figure 8-5: Three fine scale basis functions corresponding to the two-dimensional separable Haar wavelet. The full two-dimensional separable Haar transform of the projection consists of inner products of the projection with basis function of the types shown here but at multiple scales and shifts.

Extension to the Poisson case

We next outline the procedure for extending our multiscale reconstruction method to the Poisson case where the projection data are Poisson distributed random variables. For simplicity, we only consider here the case of two-dimensional transmission tomography[1] [70]. To begin, assume that the input photon counts are Poisson distributed with rate λ_T. The output counts recorded by the j-th detector pair are then also Poisson distributed with mean and variance given by:

$$\lambda(j) = \lambda_T \exp(-T(j)f), \tag{8.7}$$

where $T(j)$ is the j-th row of the projection matrix T, and f is the vector that contains the object pixels. It can be shown [70] that the log-likelihood expression for the Poisson case is given by:

$$\ln p(\lambda|f) = -\sum_{j=1}^{N_\theta N_s} \left[\lambda_T \exp(-T(j)f) + \lambda(j)T(j)f + \ln(\lambda(j)!)\right]. \tag{8.8}$$

Further, as shown in [70], the log-likelihood (c.f. (8.8)) can be accurately approximated as follows:

$$\ln p(y|f) = -\frac{1}{2}(y - Tf)^T D(y - Tf) + c(y), \tag{8.9}$$

where the term $c(y)$ is just a function of the data vector y the j-th element of which is given by:

$$y(j) = \ln(\frac{\lambda_T}{\lambda(j)}), \tag{8.10}$$

and D is a diagonal matrix the j-th diagonal element of which is given by:

$$d(j) = \lambda(j). \tag{8.11}$$

---

[1] The discussion presented here applies to emission tomography as well with some minor modifications.

Now by substituting the NP representation equation $f = T^T x$ in (8.9), we obtain the following expression for the log-likelihood:

$$\ln p(y|x) = -\frac{1}{2}(y - Cx)^T D(y - Cx) + c(y), \tag{8.12}$$

where the NP matrix $C = TT^T$. The MAP equation corresponding to (8.12) is given by:

$$\begin{aligned}\widehat{x} &\triangleq \arg\max_x \left[\ln p(x|y)\right] \\ &= \arg\max_x \left[\ln p(y|x) + \ln p(x)\right] \\ &= \arg\max_x \left[-\frac{1}{2}(y - Cx)^T D(y - Cx) - \frac{1}{2}x^T \Lambda_x^{-1} x\right].\end{aligned} \tag{8.13}$$

In the multiscale framework the MAP expression (8.13) transforms to:

$$\widehat{\xi} = \arg\min_\xi \left[(\eta - \mathcal{C}\xi)^T \mathcal{D}(\eta - \mathcal{C}\xi) + \xi^T \Lambda_\xi^{-1} \xi\right], \tag{8.14}$$

where the multiscale data η is given by:

$$\eta = W_b y, \tag{8.15}$$

Λ_ξ is a diagonal matrix representing a fractal prior for the object coefficients, the matrix $\mathcal{D}$ is given by:

$$\mathcal{D} = W_b D W_b^T, \tag{8.16}$$

and the multiscale matrix $\mathcal{C}$ is sparse and in addition is naturally partitioned into ill and well-conditioned blocks. The MAP expression (8.14) for the Poisson case is the same as the expression we had for the Gaussian case, except that the matrix $\mathcal{D}$ is no longer diagonal. However, since $\mathcal{D}$ is still sparse and so is $\mathcal{D}^{1/2}$, the solution $\widehat{\xi}$ to the MAP equation (8.14) can once again be obtained in an efficient manner [63]. Once this solution is obtained the reconstruction for the Poisson case is obtained as:

$$f = \mathcal{T}^T \xi = \sum_{k=1}^{N_\theta} \mathcal{T}_k \xi_k. \tag{8.17}$$

Introduction of a scale recursive structure in the multiscale iterative reconstruction algorithm

Recall that in our multiscale iterative reconstruction algorithm (MPART), we use the Kaczmarz method to solve the system of equations that arises in the multiscale transformed NP representation. This multiscale iterative method is devoid of many of the convergence problems that plague ART because, as opposed to the projection matrix used in ART, the overall multiscale system matrix is quite well-conditioned. However, in addition, the multiscale matrix has a structure that can possibly be taken advantage of to further speed the convergence in MPART. Specifically, recall that the block of the multiscale system matrix that represents the coupling between fine scale coefficients in the data and in the object, is extremely well-conditioned and sparse. However the condition number gradually increases and the sparsity decreases as coarser coupling are included. Thus one possible modification of the multiscale iterative technique can be the introduction of a scale recursive structure where the fine scale object coefficients are first solved for and then this fine scale solution is used to guide the solution of the coarse scale coefficients. Note that the fine scale coefficients can be solved for relatively quickly because the fine scale block of the multiscale matrix is extremely sparse (implying fewer computations per iteration) and well-conditioned (implying a faster convergence and hence fewer iterations to convergence)[2]. Thus one can expect that the introduction of this scale recursive structure would speed up the overall convergence of MPART.

---

[2] In particular, recall that the finest scale block is approximately diagonal.

Use of prior models that are correlated in angles

The class of fractal prior models that we consider in this thesis assume that the multiscale object coefficients are independent from angle to angle. However we would intuitively expect the coarse scale object coefficients at different projection angles to actually be highly correlated with each other, and further for this correlation to decrease at finer scales. Such a correlation structure across projection angles would help us estimate at least the coarse scale object coefficients to a good accuracy even if the projection data at certain angles are missing. The current angular independence of the multiscale object coefficients corresponds to an overall covariance structure for these variables which is block diagonal, and it is this block diagonality that is partially responsible for the extreme efficiency of our current technique. Using the proposed, more correlated prior covariance structure would correspond in this paradigm to the addition of off-diagonal terms. At first, such a proposal would seem to make things dramatically worse from a computational perspective, since we must now contend with what corresponds to the inversion of a full rather than a block-diagonal matrix. All is not lost, however, for at least two reasons, both related to the fact that we are building our prior models directly in scale space. First, the addition of only coarse scale correlations may be sufficient to regularize a given problem, with the result that only a few, low dimensional, off diagonal elements need be added to the prior model covariance (recall, at coarser scales there are far fewer model elements – e.g. at the coarsest scale there is only one per angle). These few additional coarse scale terms could then be aggregated into a slightly larger corresponding covariance block, returning us to the block diagonal case, but with one block slightly larger than the rest. More significantly perhaps, however, is that recent research has demonstrated that certain scale-based prior models (which correspond to tree structures), though corresponding to highly correlated fields, can lead to extremely efficient scale-recursive estimation algorithms [4,13–15]. The variations of the prior covariance structure and efficient solution of the resulting estimation equations constitutes a possible topic for future research.

Edge detection at multiple scales

Recall that the contributions of edges and boundaries in the field are enhanced in our multiscale detail reconstructions. One possible future research topic can be to develop post-processing techniques to extract binary edge and boundary maps from these detail reconstructions. This would provide a method for detecting edges and boundaries in the field at multiple scales directly from the projection data. Further we expect that if the detail images from our NP-based multiscale reconstruction method are utilized for this edge detection process, the resulting edge and boundary maps would have less sparse-data artifacts than the conventional projection-based edge detection technique [72] which is valid only in the case that the data are dense.

8.2.2 Multiscale detection: topics for future research

Performance study of anomaly detection method in case of sparse and noisy data

A major advantage of our anomaly detection method is that it is not necessarily limited to dense and noise less data as opposed to the conventional techniques. This is because our method incorporates the statistical properties of the data and the measurement noise in such a manner that it results in a nearly *statistically-optimal* detection algorithm. In this thesis, however, we only considered the case of anomaly detection from dense and noise less data. An obvious topic for future research is to study the performance of our detection approach for the case that the data are sparse or noisy.

Performance study of anomaly detection method in case of model mismatch

For the anomaly detection method that we developed and analyzed in this thesis, we assumed a perfect knowledge of the background texture. In particular, the background may correspond to a fractal texture but the fractal dimension may not be perfectly known. Or it may be that the background does not correspond to a fractal texture at all. In these cases of model mismatch it is not clear as to how our detection approach will perform, and consequently a thorough testing is needed.

A multiscale approach for zooming-in on the anomalous region

Recall that in the first step of our two-step approach to detection, we zoom-in on the set of candidate anomalous pixels by calculating approximate chi-square values for all the pixels by ignoring the off-diagonal elements in the fractal-field data covariance matrix. An alternate, and probably more accurate method for zooming-in on the set of anomalous pixels can be obtained as follows. We first start by selecting the test region as the entire field and perform a chi-square test on the entire multiscale data vector to determine the presence of any anomaly in the field. In case the presence of an anomaly is established to a certain confidence level, we subdivide the original test region into several smaller regions. We then perform a chi-square test in each of these smaller test regions but use only those data samples for chi-square calculation that correspond to the multiscale strips which intersect these regions. Based on the outcome of this chi-square calculation, a decision is made as to subdivide any of these smaller test regions, and this process is repeated until we reach the desired finest level.

Bibliography

[1] Alloys developed and computer-aided tomography used to study damage in aging aircraft. Materials Evaluation, February 1990.

[2] M. F. Barnsley, R. L. Devaney, B. B. Mandelbrot, H.-O. Peitgen, D. Saupe, and R. F. Voss. *The Science of Fractal Images.* Springer-Verlag, 1988.

[3] H. H. Barrett. Image reconstruction and the solution of inverse problems in medical imaging. In *Medical Images: Formation, Handling and Evaluation, Proceedings of the NATO Advanced Study Institute on the Formation, Handling and Evaluation of Medical Images*, pages 21–22. Springer Verlag, 1992.

[4] M. Basseville, A. Benveniste, K. C. Chou, S. A. Golden, R. Nikoukhah, and A. S. Willsky. Modeling and estimation of multiresolution stochastic processes. *IEEE Transactions on Information Theory*, 38(2):766–784, March 1992.

[5] C. Berenstein and D. Walnut. Local inversion of the Radon transform in even dimensions using wavelets. Technical Report CAM-21/93, Center For The Applications Of Mathematics, George Mason University, January 1993.

[6] G. Beylkin, R. Coifman, and V. Rokhlin. Fast wavelet transforms and numerical algorithms I. *Communications on Pure and Applied Mathematics*, XLIV:141–183, 1991.

[7] T. M. Breunig, S. R. Stock, A. Guvenilir, J. C. Ellliott, P. Anderson, and G. R. Davis. Damage in aligned-fibre SiC–Al quantified using a laboratory X-ray tomographic microscope. *Composites*, 24(3):209–213, 1993.

[8] R. A. Brooks and G. di Chiro. Principles of computer assisted tomography (CAT) in radiographic and radioisotopic imaging. *Phys. Med. Biol.*, 21(5):689–732, 1976.

[9] M. H. Buonocore, W. R. Brody, and A. Macovski. Fast minimum variance estimator for limited angle CT image reconstruction. *Medical Physics*, 8(5):695–702, September/October 1981.

[10] M. H. Buonocore, W. R. Brody, and A. Macovski. A natural pixel decomposition for two-dimensional image reconstruction. *IEEE Transactions on Biomedical Engineering*, BME-28(2):69–78, 1981.

[11] E. B. Cargill, H. H. Barrett, R. D. Fiete, M. Ker, D. D. Patton, and G. W. Seeley. Fractal physiology and nuclear medicine scans. In *SPIE Medical Imaging II*, volume 914, pages 355–361, 1988.

[12] C-C. Chen, J. S. Daponte, and M. D. Fox. Fractal texture analysis and classification in medical imaging. *IEEE Transactions on Medical Imaging*, 8(2):133–142, June 1989.

[13] K. C. Chou. *A Stochastic Modeling Approach to Multiscale Signal Processing.* PhD thesis, Massachusetts Institute of Technology, May 1991.

[14] K. C. Chou, A. S. Willsky, and A. Benveniste. Multiscale recursive estimation, data fusion, and regularization. *IEEE Transactions on Automatic Control*, 39(3):464, March 1994.

[15] K. C. Chou, A. S. Willsky, and R. Nikoukhah. Multiscale systems, Kalman filters, and Riccati equations. *IEEE Transactions on Automatic Control*, 39(3):479, March 1994.

[16] A. M. Cormac. Sampling the Radon transform with beams of finite width. *Phys. Med. Biol.*, 23(6):1141–1148, 1978.

[17] I. Daubechies. *Progress in Wavelet Analysis and Applications*, chapter Wavelet on the interval, pages 95–107. 1992.

[18] I. Daubechies. *Ten Lectures on Wavelets.* SIAM, 1992.

[19] P. J. Davis. *Circulant Matrices.* John Wiley & Sons, 1979.

[20] H. W. Deckman, J. H. Dunsmuir, K. L. D'Amico, S. R. Ferguson, and B. P. Flannery. Development of quantitative X-ray microtomography. In J. L. Ackerman and W. A. Ellingson, editors, *Advanced Tomographic Imaging Methods for the Analysis of Materials*, pages 97–110. Materials Research Society, November 1990.

[21] M. Defrise. Regularization techniques in medical imaging. In *Medical Images: Formation, Handling and Evaluation, Proceedings of the NATO Advanced Study Institute on the Formation, Handling and Evaluation of Medical Images*, pages 43–64. Springer Verlag, 1988.

[22] J. DeStefano and T. Olson. Wavelet localization of the Radon transform in even dimensions. In *Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis*, pages 137–140, October 1993.

[23] A. J. Devaney. Geophysical diffraction tomography. *IEEE Transactions on Geoscience and Remote Sensing*, GE-22(1):3–13, January 1984.

[24] B. Dolveck-Guilpart. Tomographic image reconstruction from a limited set of projections using a natural pixel decomposition. In P. C. Sabatier, editor, *Inverse Methods in Action, Proceedings of the Multicentennials Meeting on Inverse Problems*, pages 54–61. Springer-Verlag, November 1990.

[25] P. Engler, W. D. Friedman, M. W. Santana, and F. Chi. Process control for composites using computed tomography. In J. L. Ackerman and W. A. Ellingson, editors, *Advanced Tomographic Imaging Methods for the Analysis of Materials*, pages 123–128. Materials Research Society, November 1990.

[26] P. Flandrin. On the spectrum of fractional Brownian motions. *IEEE Transactions on Information Theory*, 35:197–199, 1989.

[27] S. Geman and D. Geman. Stochastic relaxation: Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6):721–741, November 1984.

[28] S. Geman and D. E. McClure. Bayesian image analysis: An application to single photon emission tomography. *Proc. Amer. Stat. Assoc. Stat. Comput. Section*, pages 12–18, 1987.

[29] N. H. Getz. A perfectly invertible, fast, and complete wavelet transform for finite length sequences: The discrete periodic wavelet transform. In *Proceedings of the SPIE Annual Conference on Mathematical Imaging: Wavelet Applications in Signal and Image Processing*, July 1993.

[30] M. L. Giger, K. Doi, H. MacMahon, and F-F. Yin. Image-processing techniques used in the computer-aided detection of radiographic lesions in anatomic background. In *Medical Imaging II*, volume 914, pages 635–637. SPIE, 1988.

[31] G. Gindi, M. Lee, A. Rangarajan, and I. G. Zubal. Bayesian reconstruction of functional images using registered anatomical images as priors. In A. C. F. Colchester and D. J. Hawkes, editors, *Information Processing in Medical Imaging, 12th International Conference, IPMI '91*. Springer-Verlag, July 1991.

[32] G. H. Golub and C. F. Van Loan. *Matrix Computations.* The Johns Hopkins University Press, 1991.

[33] T. A. Gooley and H. H. Barrett. Evaluation of statistical methods of image reconstruction through ROC analysis. *IEEE Transactions on Medical Imaging*, 11(2):276–283, June 1992.

[34] P. J. Green. Bayesian reconstructions from emission tomography data using a modified EM algorithm. *IEEE Transactions on Medical Imaging*, 9(1):84–93, 1990.

[35] A. Hald. *Statistical Theory, With Engineering Applications.* Wiley, 1952.

[36] K. M. Hanson. *Image Recovery Theory and Application*, chapter Bayesian and Related Methods in Image Reconstruction from Incomplete Data. Academic Press, 1987.

[37] G. T. Herman. Resolution in ART – an experimental investigation of the resolving power of an algebraic picture reconstruction technique. *J. theor. Biol.*, 33:213–223, 1971.

[38] G. T. Herman. ART: Mathematics and applications – a report on the mathematical foundations and on the applicability to real data of the algebraic reconstruction techniques. *J. theor. Biol.*, 42:1–32, 1973.

[39] G. T. Herman. *Image Reconstruction From Projections.* Academic Press, 1980.

[40] G. T. Herman and L. B. Meyer. Algebraic reconstruction techniques can be made computationally efficient. *IEEE Transactions on Medical Imaging*, 12(3):600–609, September 1993.

[41] R. A. Horn and C. R. Johnson. *Matrix Analysis.* Cambridge University Press, 1991.

[42] A. K. Jain. *Fundamentals of Digital Image Processing.* Prentice Hall, 1989.

[43] A. K. Jain and S. Ansari. Radon transform theory for random fields and optimum image reconstruction from noisy data. In *Proc. ICASSP*, pages 12A.7.1–12A.7.4, 1984.

[44] A. C. Kak and M. Slaney. *Principles of Computerized Tomographic Imaging.* IEEE Press, 1988.

[45] M. B. Katz. *Questions of Uniqueness and Resolution in Reconstruction from Projections*, volume 26 of *Lecture Notes in Biomathematics.* Springer-Verlag, 1978.

[46] A. Kettschau and J. Goebbels. Application of computer-tomography in industrial problems. In R. Guzzardi, editor, *Physics and Engineering of Medical Imaging, Proceedings of the NATO Advanced Study Institute*, number 119 in E, Applied sciences, pages 257–259. Martinus Nijhoff, 1987.

[47] J. H. Kinney, M. C. Nichols, U. Bonse, S. R. Stock, T. M. Breunig, A. Guvenilir, and R. A. Saroyan. Nondestructive imaging of materials microstructures using X-ray tomographic microscopy. In J. L. Ackerman and W. A. Ellingson, editors, *Advanced Tomographic Imaging Methods for the Analysis of Materials*, pages 81–95. Materials Research Society, November 1990.

[48] R. A. Knox. *Ocean Circulation Models: Combining Data and Dynamics*, chapter Ocean Acoustic Tomography: A Primer, pages 141–188. Number 284 in NATO ASI Series. Kluwer Academic Publishers, 1989.

[49] D. S. Lalush and B. M. W. Tsui. Simulation evaluation of Gibbs prior distributions for use in maximum a posteriori SPECT teconstructions. *IEEE Transactions on Medical Imaging*, 11(2):267–275, June 1992.

[50] Lawrence Berkeley Laboratory, University of California. *Donner Algorithms for Reconstruction Tomography, RECLBL Library Users Manual*, 1977.

[51] R. M. Lewitt. Reconstruction algorithms: Transform methods. *Proceedings of the IEEE*, 71(3):390–408, March 1983.

[52] J. Llacer. Theory of imaging with a very limited number of projections. *IEEE Transactions on Nuclear Science*, NS-26(1):596–602, February 1979.

[53] J. Llacer. Tomographic image reconstruction by eigenvector decomposition: Its limitations and areas of applicability. *IEEE Transactions on Medical Imaging*, MI-1(1):34–42, July 1982.

[54] J. Llacer and J. D. Meng. Matrix based image reconstruction methods for tomography. *IEEE Transactions on Nuclear Science*, NS-32(1):855–864, February 1985.

[55] A. K. Louis. Medical imaging: State of the art and future development. *Inverse Problems*, 8:709–738, 1992.

[56] Z. Q. Lu, M. Kaveh, and R. K. Mueller. Diffraction tomography using beam waves: Z-average reconstruction. *Ultrasonic Imaging*, 6:95–106, 1984.

[57] M. Luettgen. *Image Processing with Multiscale Stochastic Models.* PhD thesis, Massachusetts Institute of Technology, May 1993.

[58] M. Luettgen, W. C. Karl, and A. S. Willsky. Efficient multiscale regularization with applications to the computation of optical flow. *IEEE Transactions on Image Processing*, 3(1):41–64, January 1994.

[59] S. G. Mallat. A theory of multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693, 1989.

[60] R. K. Mueller, M. Kaveh, and G. Wade. Reconstructive tomography and applications to ultrasonics. *Proceedings of the IEEE*, 67(4):567–587, April 1979.

[61] W. Munk and C. Wunsch. Ocean acoustic tomography: A scheme for large scale monitoring. *Deep-Sea Research*, 26A:123–161, 1979.

[62] M. Onoe, J. W. Tsao, H. Yamada, H. Nakamura, J. Kogure, H. Kawamura, and M. Yoshimatsu. Computed tomography for measuring annual rings of a live tree. *Proceedings of the IEEE*, 71(7):907–908, July 1983.

[63] C. C. Paige and M. A. Saunders. LSQR: An algorithm for sparse linear equations and sparse least squares. *ACM Transactions on Mathematical Software*, 8(1):43–71, March 1982.

[64] F. Peyrin, M. Zaim, and R. Goutte. Multiscale reconstruction of tomographic images. In *Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis*, pages 219–222, October 1992.

[65] P. A. Rattey and A. G. Lindgren. Sampling the 2-D Radon transform. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-29(5):994–1002, 1981.

[66] D. J. Rossi. *Reconstruction From Projections Based On Detection And Estimation Of Objects.* PhD thesis, Massachusetts Institute of Technology, October 1982.

[67] B. Sahiner and A. E. Yagle. Image reconstruction from projections under wavelet constraints. *IEEE Transactions on Signal Processing*, 41:3579–3584, December 1993.

[68] B. Sahiner and A. E. Yagle. Limited angle tomography using the wavelet transform. October 1993.

[69] K. Sauer and C. Bouman. Bayesian estimation of transmission tomograms using segmentation based optimization. *IEEE Transactions on Nuclear Science*, 39(4):1144–1152, 1992.

[70] K. Sauer and C. Bouman. A local update strategy for iterative reconstruction from projections. *IEEE Transactions on Signal Processing*, 41(2):534–548, February 1993.

[71] E. Scheinman and F. L. Roder. Real-time CT of composites during destructive testing. In J. M. Carpentar, D. B. Cline, R. Lanza, and D. F. Mildner, editors, *Neutrons, X-Rays, and Gamma Rays: Imaging Detectors, Material Characterization Techniques, and Applications*, volume SPIE 1737, pages 171–183. SPIE, July 1992.

[72] N. Srinivasa, K. R. Ramakrishnan, and K. Rajgopal. Detection of edges from projections. *IEEE Transactions on Medical Imaging*, 11(1):76–80, March 1992.

[73] J. S. Steude and P. D. Tonner. Detection of free liquid in cement-solidified radioactive waste drums using computed tomography. In J. L. Ackerman and W. A.

Ellingson, editors, *Advanced Tomographic Imaging Methods for the Analysis of Materials*, pages 111–116. Materials Research Society, November 1990.

[74] G. Strang. Wavelets and dilation equations: A brief introduction. *SIAM Review*, 31(4):614–627, 1989.

[75] K. Tanabe. Projection method for solving a singular system of linear equations and its applications. *Numer. Math.*, 17:203–214, 1971.

[76] H. L. Van Trees. *Detection, Estimation, and Modulation Theory.* John Wiley & Sons, 1968.

[77] J. G. Verly and R. N. Bracewell. Blurring in tomograms made with X-ray beams of finite width. *Journal of Computer Assisted Tomography*, 3(5):662–678, October 1979.

[78] B. J. West and A. L. Goldberger. Physiology in fractal dimensions. *American Scientist*, 75:354–365, 1987.

[79] G. W. Wornell. Wavelet-based representation for the 1/f family of fractal processes. *Proceedings of the IEEE*, 81(10):1428, October 1993.

[80] G. W. Wornell and A. V. Oppenheim. Estimation of fractal signals from noisy measurements using wavelets. *IEEE Transactions on Signal Processing*, 40(3):611–623, March 1992.